DECISION-MAKING IN FLIGHT WITH DIFFERENT CONVECTIVE WEATHER INFORMATION SOURCES: PRELIMINARY RESULTS.

Kara A. Latorella & James P. Chamberlain
k.a.latorella@larc.nasa.gov & j.p.chamberlain@larc.nasa.gov
Crew/Vehicle Integration Branch & Crew Systems & Operations Branch
NASA Langley Research Center, Hampton, VA 23681-0001

ABSTRACT

This paper reports preliminary and partial results of a flight experiment to address how General Aviation (GA) pilots use weather cues to make flight decisions. This research presents pilots with weather cue conditions typically available to GA pilots in visual meteorological conditions (VMC) and instrument meteorological conditions (IMC) today, as well as in IMC with a Graphical Weather Information System (GWIS). These preliminary data indicate that both VMC and GWIS-augmented IMC conditions result in better confidence, information sufficiency and perceived performance than the current IMC condition. For all these measures, the VMC and GWIS-augmented conditions seemed to provide similar pilot support. These preliminary results are interpreted for their implications on GWIS display design, training, and operational use guidelines. Final experimental results will compare these subjective data with objective data of situation awareness and decision quality.

INTRODUCTION

Eighty-five percent of the aviation accidents that occurred from 1990-1996, and nearly eighty-five percent of the accident fatalities, involved small general aviation (GA) airplanes. One major contributor to aviation accidents is weather. Desktop simulation and other laboratory experiments have demonstrated pilot error tendencies that corroborate the implication of this accident statistic (Driskill et al. 1997; O'Hare 1990; Layton et al. 1989; Potter et al. 1989; Beck 1987; Giffin & Rockwell 1982). Stimulated by the 1997 Gore commission on Aviation Safety, the NASA Aviation Safety program (AvSP) was initiated to reduce the aircraft accident rate by a factor of 5 within 10 years and by a factor of 10 within 25 years. Within AvSP, the Aviation Weather Information (AWIN) program element aims to accomplish this goal by improving weather information available to aviation users.

General aviation is particularly affected by convective weather. A survey of GA accidents from 1982 to 1993 (AOPA 1999) revealed that while only 3.5% of these accidents are directly attributed to thunderstorms, a large percentage of these accidents, 66%, resulted in fatalities. Convective weather is challenging because it can be characterized by reduced ceiling and visibility defined by instrument meteorological conditions (IMC) as well as including severe to extreme turbulence, gusts, hail, icing, lightning, and possibly severe downdrafts and microbursts. Such concomitant weather phenomena were analyzed separately in this accident analysis. Therefore the incidence of GA accidents attributed to convective activity, and the fatalities resulting from such weather systems is likely under-represented by the percentages cited for thunderstorm effects.

Currently, GA pilots rely principally on aural sources and external, "out-the-window," weather cues for weather information. Aural sources can include direct queries to Flight Service Station (FSS), Enroute Flight Advisory Service (EFAS) or "Flightwatch" (FW), and Air Traffic Control (ATC) personnel, as well as monitoring frequencies to hear other pilots' comments, queries, and the information supplied to them. Pilots can also tune automated weather information services such as AWOS/ASOS, HIWAS and ATIS to obtain a broadcast of conditions at the reporting station. Unfortunately, the information available from these sources is limited and, when weather becomes a problem, the frequencies used to obtain this information become saturated, making this information inaccessible at exactly the time it is most needed.

In addition to aural and external visual weather information, a relatively small percentage of GA pilots augment their avionics with sensor-based onboard weather hazard detection systems (e.g., Stormscope, Strikefinder) or onboard weather radar. These systems can provide improved weather awareness for severe weather hazards, however they are limited in range and accuracy (Bussolari 1994). Onboard weather radar systems are workload-intensive to use accurately (Kelly 2000), are subject to attenuation, have a limited range, and provide information that is primarily forward of the aircraft and at the aircraft's altitude (Bernays et al. 1993). While these systems show severe local weather to avoid, they do not provide the more comprehensive weather picture required to fully support strategic planning or avoidance maneuvers.

More accessible, complete, and usable weather information would benefit pilots' situation awareness, decision-making, and safety. Graphical weather is a more appropriate representation (Wickens 1984), can more effectively be integrated with other such...
information (e.g., terrain) and can be extended using symbology. Accordingly, pilots using prototype graphical weather information systems (GWIS) in static and dynamic desktop simulation experiments were more likely to acquire trend data, have a more comprehensive awareness (Potter et al. 1989), make better go/no-go decisions, rate hazard levels higher, have more confidence in weather-related decisions, make fewer calls to ground aviation weather personnel (Lind et al., 1997, 1994), and make more correct decisions with graphical, than with verbal or text alerts (Wanke et al. 1990; Wanke & Hansman 1992).

One early implementation of an uplinked radar mosaic GWIS, developed at MIT Lincoln Labs with funding from the FAA Datalink Operational Requirements Team (DLORT), had a 15 minute update rate, 6km-square resolution and employed a "lossy" algorithm (resulting in less well-defined precipitation areas) to compensate for lower available bandwidth (250bps) (Chandra et al. 1995a, 1995b). In desktop usability assessments, all subjects found the high level of lossy compression unacceptable, and some found that the medium level lacked the functional equivalence of the uncompressed image (Lind et al. 1997, 1995). When used in a GA flight test, accompanied by terminal forecasts and surface observations, and integrated with a traffic information service, subjects commented enthusiastically on the utility of this GWIS (Chandra 1995). More than 82% of subjects had positive responses to the utility of precipitation maps, surface observations, and terminal forecasts individually (Tallotta et al. 1997). All subjects had a positive overall impression of the system; 88% indicating that it would be important to make available to GA operations (Tallotta et al. 1997).

The FAA Flight Information Services Data Link (FISDL) program will soon make VHF data-linked weather information systems widely available to GA pilots. The FISDL system will provide, for no service charge, uplink of textual aviation weather products. These products include weather observations (METARS & SPECIs) and forecasts (TAFS) of terminal environments, as well as reports of severe weather conditions (SIGMETS, Convective SIGMETS, AIRMETS, and severe weather forecast alerts) and pilot reports (PIREPS). GA pilots may augment this basic information by purchasing services that will uplink graphical weather information. The graphical product of most interest is a national radar mosaic.

NavRadio (now Honeywell), in a cooperative agreement with NASA AWIN, developed such a GWIS which was subsequently selected for the FISDL program. FAA FISDL and NASA AWIN jointly funded a simulation experiment at Research Triangle Institute (RTI) to evaluate pilot weather flying with and without this GWIS. In this study, subjects were in IMC, had access to an autopilot, and were not given a present position symbol on the GWIS display. Results indicated that while this GWIS system increased awareness of the general location of convective weather, it did not improve pilot diversion decision-making (subjects did not understand the location of weather with respect to their position), increased workload for at least half the subjects, and reduced reliance on ground-based weather professionals (Yucknovicz et al. 2000). This simulation study suggests several features for GWIS's (e.g., aircraft present position symbol) and concludes that further experimentation is required to develop industry standards to appropriately design GWIS interfaces and procedures for using them.

The AWIN Convective Weather Sources (CoWS) experiment investigates how GA pilots use weather information aural, "out-the-window" visual, and GWIS-displayed cues to support in-flight decisions related to convective weather systems. This research project also addresses the effect of pilot differences in risk tolerance, personality type, and weather knowledge on these decisions. Finally, this research provides a platform for a usability assessment of one implementation of a GWIS. This paper focuses on preliminary results associated with the major objective – how GA pilots' decisions are affected by the availability of aural, out-window visual, and graphically represented weather cues.

METHODS

Apparatus. Apparatus for the CoWS experiment includes the test aircraft, tethered GWIS, and supporting ground infrastructure. NASA Langley's Raytheon B-200 Super King Air, a nine-passerenger, pressurized twin-turboprop airplane, was operated at speeds and altitudes consistent with those of the smaller, piston-engine GA aircraft used by the subject population. The GWIS includes a VHF data link (VDL) receiver, Global Positioning System (GPS) receiver, and two laptop PC's with tether cables to two small handheld display units. The display unit's screen was approximately 4 inches tall by 5 inches wide. Five bezel buttons on the right side of the unit actuate soft menu fields, and a rate-controlled joystick controls pan, zoom, and crosshairs for symbol selection (Figure 1). The unit presents lossless, nationwide radar mosaic imagery at 4-square-km resolution with a 6 minute nominal update rate assuming adequate broadcast reception, and surface weather observations (METAR) in text and symbolic form for reporting stations in the mid-Atlantic region. The display also presents contextual features (rivers, interstates, and state boundaries), airport identifiers, present position and track symbol, creation time stamp for the radar product, a scale legend, and indicates missing data. The features
and usability issues of this display will be more fully addressed in a separate report. Ground infrastructure includes prototype AWIN/Honeywell broadcast VDL transmitters located at four sites in Virginia.

Subjects. Subjects were recruited from local regional airports and through advertisement. Applicants reported their flying experience and weather exposure on a Background Questionnaire. Subject selection criteria included: an instrument rating, 10-50 flight hours within the last 90 days, and 50-1000 cross-country or 100-2000 total flight hours. In addition, participants were not selected who had worked for a scheduled air-carrier in the prior year or who had participated in the aforementioned RTI/AWIN experiment. This experiment requires 12 subjects. This paper reports preliminary results from 6 subjects.

Weather experience has been found to significantly affect weather-related decision-making and information acquisition (Wiggins & O'Hare 1995) so candidate subjects were clustered into 3 groups of “exposure experience” using cross-country hours. The midpoints of each cluster are 135 (low), 379 (medium), and 738 (high) cross-country hours respectively ($p < .0001$). Subjects were selected to form four three-member teams, each team composed of one subject from each of the clusters to balance exposure experience across flight scenarios.

Scenarios. The ideal scenario operated under Instrument Flight Rules (IFR) but in Visual Meteorological Conditions (VMC). The scenario included a departure from NASA Langley/ Langley Air Force field (LFI) on a flight path that would obliquely intercept a frontal convective system of at least moderate intensity at approximately 120nm from top-of-climb, and at an altitude above the haze layer (13000-15000 feet). Scenario definition was constrained by the location of the GWIS’s ground-based infrastructure. To accommodate this constraint and minimize training and materials, four potential IFR flight plans were developed from LFI to Hickory, NC [HKY]; Charleston, WV [CRW]; Abingdon, VA [VJI]; or Clarksburg, WV [CKB]. Subjects did not perform flying duties during these flights; a NASA test pilot served as pilot in command (PIC). To compensate for the loss of positional awareness and workload induced by not piloting, subjects were required to copy scheduled PIC reports (containing: airspeed, altitude, heading, position, next waypoint, and current time), plot position on a low altitude IFR chart, and calculate elapsed time and ground speed, and were required to note any ATC transmission affecting the flight. All subjects received scheduled aural weather information at 3 intervals, 16minute (40nm) apart. The first cue was obtained from the local HIWAS, the second from querying FW, and the third from querying ATC.

Weather Cue Conditions. Three sets of weather cue conditions were provided to subjects. The first condition represents the weather information typically available to a GA pilot in IMC. This “Traditional IMC” condition (aural) consists only of aural information. The “Traditional VMC” condition (window+aural) augments the aural cues with the visual cues provided by an “out-the-window” view. Finally the “GWIS-IMC” condition (display+aural) augments the aural cues with access to the GWIS formerly described. Opaque window coverings restricted the views of subjects in aural and display+aural conditions.

Experimental Design. The full CoWS experiment will use twelve subjects, these preliminary and partial result are based on the first six. The experimental design for the confidence ratings was: Subjects (6) x Cues (3) X Proximity (6). Since information sufficiency and workload ratings were only collected once per run, the experimental design for these analyses is simplified to: Subjects (6) x Cues (3). Cue levels are within flights to minimize the effects of weather variability over flight days. Cross-country experience level ranged within teams and was counter-balanced over teams to mitigate against concerns about generalization to the subject population and experience x flight interactions. Cue assignment to subject experience levels were counterbalanced to mitigate against cue x experience level interactions.

Protocol. A team arrived at NASA Langley in the morning, and each subject was provided with an introductory briefing, consent form, schedule, and Preliminary Questionnaire. These subjects then received a mission motivation and briefing; a local terrain, navaid and airport identifier review; a route briefing for the flight to be taken; and practice on forms and procedures to be used during the in-flight phase. Following a short break, subjects had 10...
minutes to review a textual DUATS preflight (standard weather) briefing and associated weather graphics; 10 minutes to listen twice to an audiotaped recording of a FSS preflight briefing; and an additional 10 minutes to review this preflight material. Subjects then completed the Preflight Weather Situation Awareness questionnaire. After lunch, the subject chosen to receive the GWIS received a standardized training presentation, test, and compensatory instruction on this system. This subject was also allowed to practice with the GWIS display upon reaching the aircraft. In parallel to the formal display instruction, another subject received a weather knowledge survey, and the remaining team subject received a risk tolerance test and a general personality inventory test.

The in-flight portion of the experiment started when the aircraft was approximately 120nm from the first convective weather area of moderate or greater intensity, and ended when approximately 20nm from this area. Throughout the in-flight portion, subjects were responsible for retaining positional awareness, copying relevant ATC calls, and remaining aware of weather affecting the flight. During the outbound leg of the flight, as the aircraft approached the targeted convective area, subjects were provided with a Weather Situation Awareness questionnaire six times, at roughly 8 minute (20nm) intervals. One section of this questionnaire included an interval scale to assess subjects' confidence in their weather picture. At the conclusion of the outbound leg, subjects were asked to plot their final position on the map, draw weather within 50nm of the flightpath and complete the Post-run Questionnaire. This instrument contained NASA-TLX (Hart & Staveland 1988) scales for workload assessment, asked subjects to indicate other weather sources that would have been helpful, and about their flight decisions. After completing the Post-run Questionnaire, the subject using the display was asked to complete a Usability Questionnaire and provide any additional comments. Following the flight, subjects were provided with a short debriefing questionnaire for that flight. At the conclusion of the third flight for a team, when all subjects had been exposed to the display, subjects and experimenters more fully discussed issues of experimental validity and display usability.

RESULTS

This paper summarizes data on subjects' ratings of confidence in their weather picture, information sufficiency, and workload. All ANOVAs used SPSS GLM routines and type III sums-of-squares models, with $\alpha$ set at .10.

Confidence Ratings. Ratings for “confidence in picture” were analyzed using ANOVA for effects of cue set and observation number, as well as the interaction of these factors, using subjects and flights as replications. The cues available to subjects in this experiment significantly affected their confidence in their weather picture ($p < .001$) (Figure 2). LSD post hoc comparisons demonstrated an observed difference between aural cues and window+aural cues ($p < .001$); and between aural cues and display+aural cues ($p < .001$); and that there is no statistical difference between window+aural cues and display+aural cues ($p = .582$).

Information Sufficiency. The Post-run Questionnaire has subjects indicate which of a set of possible external and onboard weather information sources they would have liked to have had access to on the just-completed flight. The sources listed included: direct questioning and monitoring FW, FSS, ATC center, departure/approach control, and tower controllers; as well as automated weather reports on ATIS and additional HIWAS and AWOS/ASOS frequencies. Onboard sources listed included: aircraft handling (kinesthetics) performance, and instrumentation readings; visual access out the front window, out the side window, a view of the wings, and over the wings; as well as current onboard weather information systems such as onboard weather radar, Stormscope, and Strikefinder. Subjects were also encouraged to note any other desired sources.

Cue set is a significant component of the variance in the number of additional sources requested ($p = .061$). Post hoc LSD analyses indicate that the number of requested sources by subjects with the aural condition significantly differed from that of the display+aural cue set ($p = .009$), and that the number of sources requested by subjects with the window+aural cue set significantly differed ($p = .049$) from that of the display+aural condition (Figure 3). Aural and window+aural cue set effects do not significantly differ ($p = .242$).
Workload Ratings. Subjects used NASA-TLX scales in the Post-run Questionnaire to rate the workload inherent in the experimental phase of the flight. Perceived performance was marginally significantly affected by cue set differences (p = .091). While LSD post hoc tests did not find pair-wise significant differences among cue set conditions, means suggest that subjects considered performance to be most hampered when only aural cues were available, and little difference between the display+aural and window+aural conditions (Figure 4). Ratings of perceived performance (p = .030) and physical workload (p = .020) were significantly affected by differences among subjects.

DISCUSSION

In another AWIN study, pilots reported that they trust weather information that they can directly observe over that which is sensor-based and displayed (Latorella et al. in press). Based on this report, and the assumption that pilots would appreciate that GWIS information could be too old to be useful, we hypothesized that confidence ratings for different cue sets would change as subjects approached the convective area. In particular, we hypothesized that subjects would be more confident with external, direct observations than with the remotely sensed displayed data when near convective weather, and least confident at all times with only aural cues. There was no such appreciable difference in confidence ratings as subjects neared the convective weather. If any trend bears watching, it is a counter-intuitive increase in confidence ratings as one nears a convective area when using only aural cues. In fact, it appears that confidence ratings for subjects using the display+aural cues, which simulates operating in IMC with a GWIS display, is indistinguishable from subjects using the window+aural, simulating “Traditional VMC.” Confidence significantly improved over the condition simulating today’s IMC (aural).

If subjects did trust their own perceptions beyond that which was displayed by the GWIS, it was not evidenced by the Information Sufficiency data either. Subjects reported the need for significant additional weather information sources when only supported by the aural cues available to subjects today. One criticism of this experiment might be that subjects were not able to acquire weather information from these aural sources on their own initiative as they would in a real flight. We scripted aural weather information acquisition for purposes of experimental control. However we offer these other facts to support the premise that the scripted scenario was not unrealistic. First, subjects typically reported in preliminary questionnaires that they would check HIWAS, AWOS/ASOS, and occasionally ATIS stations along the route, and might check in with Flightwatch halfway through the flight. Second, subjects were provided with a form of aural weather information at 16 minute intervals, which is what one might be able to optimistically expect as these frequencies might become congested in difficult weather. Finally, subjects were asked, in the debriefing questionnaire, to rate and comment on experimental validity. As a whole, subjects’ ratings over the three aural cues averaged 93%, where 100% would indicate “very representative of actual operations.” No subject mentioned that the access to aural weather information was unrealistic.

NASA-TLX results demonstrate that subjects believed their performance is significantly varied according to the cues available, and trends suggest that both external visual cues and those provided by the GWIS improved their perceived performance. After
collecting the full data set, subjects’ weather situation awareness and decision-making will also be assessed to determine if the trends suggested by these self reports are consistent with objective data. The relative insensitivity of NASA-TLX scale data to cue set effects may be due to limitations in data or artifacts of this experiment. Subjects generally rated experimental validity, including workload representativeness, highly. However it is important to recall that this study did not require subjects to actually fly the aircraft while acquiring weather information. Workload inherent in piloting and more realistic task-switching will be addressed in AWIN’s Workload and Relative Position experiment (Jones et al. 2001).

CONCLUSIONS
This research compares pilot weather awareness and decision-making when pilots have the weather cues available in a conventional IMC, a conventional VMC, and an IMC situation with the use of a GWIS. Preliminary results based on confidence ratings, information sufficiency scores, and ratings of perceived performance suggest that IMC with a GWIS provides pilots with a similar levels of support as do the out-the-window visual cues in VMC, both of which are significant improvements over that available in conventional IMC.

In addition to completing these data sets, the complete CoWS experiment will further assess the influence of cue set on weather situation awareness and decision making with more objective situation awareness and decision quality data, address preflight weather planning, the role of individual differences (risk acceptance, weather knowledge, and general personality typing) in weather flying decision-making, and usability issues associated with this GWIS’ interface and data reliability characteristics. These results will provide an empirical basis for GWIS design and operational usage guidelines.

REFERENCES

ACKNOWLEDGEMENTS
We gratefully acknowledge the dedication of the hangar staff and pilot’s office at NASA Langley, Mrs. Regina Johns of Lockheed-Martin for recruiting pilot subjects, Ms. Tracy Hunter’s administrative assistance, and the participation of experimental subjects. We also appreciate assistance with equipment from Honeywell staff in Minneapolis, MN and Olathe, KS.