EXPLORATION OF HUMAN FACTORS ISSUES WITHIN A FUTURE SEPARATION ASSURANCE CONCEPT

Christopher Cabrall, Lynne Martin, Paul Lee and Kimberly Jobe
San Jose State University/NASA Ames Research Center, Moffett Field, CA 94035

Abstract

The human factors (HF) impact of sweeping changes in the roles of human operators, as well as the introduction of new technologies, are being studied in NASA's Next Generation Air Transportation System (NextGen). As part of a NASA funded project [1], a walkthrough technique was constructed to examine the effectiveness of using a low-cost method for looking at NextGen concepts in detail within a concrete operational context. A ground-based Separation Assurance (SA) concept was chosen as a specific example and its instantiation for a high-level en route air traffic controller position was selected as the focus.

Test run recordings from a previous study [2] provided four off-nominal events as stimuli for our walkthrough. Each event was analyzed to identify its progression due to an action by the controller or the automation. Based on reviews of four NextGen research concepts [1], [3], 18 HF themes were selected as key areas affected by introducing NextGen automation for the SA function. These encompassed cognitive and organizational topics including attention, workload and job responsibilities. A walkthrough was constructed by applying these themes as questions to relevant points in a set of events.

Six retired controllers watched each event three times. First, the event was played in real time; second, the event was stepped through and a question with a cognitive theme was asked; and third, the event was stepped through again and an organization question was asked. Participants' answers were recorded and later transcribed. Qualitative analyses selected questions that addressed the 18 themes. Results indicated valuable unique operational insights into the problems for the NextGen SA concept not previously available through human-in-the-loop simulations alone.

Given that other concept exploration methods are resource-intensive (e.g. Human in the Loop), the cognitive walkthrough was found to be a low cost and reasonably rapid method for exploring HF issues. The use of a dynamic "storyboard" to provide the stimulus for the walkthrough questions, while moving away from the original cognitive walkthrough method, was considered to be essential in the domain due to the spatial and dynamic nature of controllers' expertise.

Background

The current air transportation system is saturated, outdated, and limited in its capabilities. It relies on the use of ground-based radar, and voice radio communications on overloaded frequencies to monitor and guide individual aircraft to maintain separation and efficient traffic flow in and out of congested airports. The requirement for airplanes to fly over specific points on the ground to maintain radar and communication contact requires excessive time and limits capacity [3]. Thirteen of the 35 Operational Evolution Partnership (OEP) airports that are already at capacity (and most likely to be affected by weather) are experiencing more operations today than in 2000 and a failure to accommodate an increase in demand could have severe economic impacts [4]. Furthermore, the FAA predicts that the current day demand for air travel of around 700 million passengers will grow to more than 1 billion passengers by 2021, an approximate increase of nearly 50% [5]. The current system is unable to reduce delays at airports and will not be able to meet the anticipated demand for air travel in the future. For these reasons, the FAA has implemented an Air Traffic Control (ATC) modernizing project – the Next Generation Air Transportation System (NextGen) [3] – which will provide critically needed updates to the current system.

The NextGen satellite-based system (rather than the current ground-based system) will have tools to detect conflicts and digital communications to provide more precise, real-time flight status information to pilots and controllers. For example, the Automatic Dependent Surveillance-Broadcast (ADS-B) will enable an aircraft to constantly broadcast its position in the sky or on the runway. Area Navigation (RNAV) will enable aircraft to fly on any path (within coverage of navigation aids), allowing better access and point-to-point operations. NextGen Data Communications (Data Comm) is expected to relieve today's overloaded frequencies and increase capacity by providing another means of disseminating clearances, instructions, and advisories and handling flight crew requests and reports, allowing controllers to handle more traffic. An Advanced Airspace Concept (AAC) [6] will delegate separation assurance functions to automation systems on the ground and in the cockpit. Such an automation of separation monitoring and control is anticipated to allow airspace capacity to be significantly increased. Given the new rules and tools to be implemented for NextGen, the role of the controller is expected to change from the proactive control of today to a more reactive or supervisory role. For example, some of his/her new functions will be to decide when and what to ask the automation to do, monitor its performance on the assigned task, detect any discrepancies or failures, and to intervene, reprogram or abort the operation as necessary.

These new concepts have to be developed and tested to define tools, scope procedures, and determine operator roles. While there are laboratory studies and conceptual modeling activities to explore NextGen topics, adding to that literature of basic human factors is not where the gap exists in NextGen research. The main gap is in studying the *interaction* of the human factors issues in the context of various operational innovations of NextGen. One effective method of evaluating human factors issues for future concepts is to use Human in the Loop Simulations (HITLs).

Full-fidelity HITLs can be very costly and HITLs of NextGen concepts tend to focus on concept benefits and feasibility, rather than focus on the direct exploration of human factors issues. A low-cost method for exploring human factors issues that can be independent of HITLs, inform their design, or leverage existing results from prior HITLs could provide valuable information on the impact of human operators in NextGen concepts. The method used in this paper addresses these issues together by asking questions to gain information from a user's perspective and reframe the problems at a more operational level in an inexpensive and short time frame.

The present study was a sub-activity of a larger project that identified and prioritized human-performance issues related to NextGen operations [1]. In that project, nine human factor themes were identified and served as preliminary input to guide our current investigation of one concept in particular, SA, in a concrete operational context. A usability inspection technique – a cognitive walkthrough – which has been successfully adapted for use in advanced aircraft cockpits (see [7]), was adjusted to demonstrate the benefits of detailed HF analyses of concepts and initial tool prototypes by undertaking a focused review for an en route Air Navigation Service Provider (ANSP) as a player serving on a ground-based Separation Assurance (SA) team. The general goal of the present study was to investigate whether a cognitive walkthrough technique is an effective method for such purposes.

Purpose Statement

Our aims were two-fold;

1: to describe and test a method that could be used by concept research teams to take a "first-cut" investigation of potential HF issues,

2: to use this method to identify and explore the impact, from a user's perspective, of these HF issues prior to finalizing a tool's design.

Overview of the Cognitive Walkthrough

The cognitive walkthrough (CW) is a technique widely used in usability studies to test proposed computer interfaces (see [8] for a review as CW was being introduced). It was developed by Polson, Lewis, Rieman, and Wharton [9] based on a theory of exploratory learning which posits that an interface that is easy to learn is also easy to use, i.e., if a user who has never seen an interface can successfully navigate through and complete basic tasks, the interface is user friendly.

Briefly, the technique entails showing a potential user a mock up of a new interface and gaining their feedback through directed questioning as they indicate how they would perform various tasks that the interface is designed to facilitate. Researchers construct a storyboard that predefines the tasks along a frame of "correct actions". As the participant steps through the tasks, four lines of questioning are followed [10]:

1. Will a user try to achieve the right effect?

2. Will a user notice the correct action?

3. Will a user associate that correct action with the effect they are trying to achieve?

4. If they perform the correct action, will a user see they are progressing toward their goal?

The questions focus on participants' reasoning about the interface/ automation – what they think it is doing now, what they expect to happen when they push a button, what they think they have to do next, and so on. At each step, if the user makes an incorrect choice for their next action, this move is discussed but then the storyboard is advanced to the next step as if following a correct action, so the participant never strays far from the task-steps of interest.

To meet the conditions of the walkthrough, a researcher is required to put detailed thought into the behaviors of the prototype interface and into describing the action sequence for each task selected. Although the process of using the prototype has to be thought out in detail, the way the tool is presented does not have to be an advanced mock-up but can be a very simple presentation, as simple as a drawing of how the interface could look. Participants do not have to be users but can be design and development experts. If study participants are not potential users, time should be spent scoping user characteristics to identify what their goals and knowledge will be when using the proposed tool [10].

A limitation of the traditional cognitive walkthrough stems from its narrow focus, as ease of learning is just one attribute of usability [10]. This bias tends to push design tradeoffs in the direction of solutions that are easy to learn but may not promote other attributes like efficiency. Compared to other usability inspection methods, cognitive walkthroughs take longer to perform [8], however, it is relatively speedy compared to other concept review methods, like human-in-the-loop simulation (HITL). Jeffries, et al. [8] also found that a walkthrough identified fewer pervasive problems in their usability inspection task when compared to other methods, although it did highlight more specific and less frequently observed issues.

Advantages of the cognitive walkthrough lie in it being an exploratory inspection method [11], which allows a tool's usability to be reviewed while it is still being designed. Specifically, the degree of interface development can be minimal when a walkthrough takes place, as little as a series of hand-drawn storyboard frames that show the proposed state of the interface at each step of a task. Additionally, the technique demands few resources, so it can be performed cheaply and quickly. A researcher can draw up a storyboard and present it to a participant in an ordinary office setting without any specialized equipment. These advantages make the cognitive walkthrough a candidate technique for exploring a range of issues of concern while a tool is being designed and prototyped, and it is possible to redirect a walkthrough's focus away from "learn-ability" alone.

Capitalizing on these methodological advantages, cognitive walkthrough principles have formed the basis of many exploratory methods and have been widely adapted to suit different domains and questions. Others have followed the CW structure and varied elements to make the method better fit a task. For example, Polson and Smith [7] adapted the original method to an aviation domain and Novick and Chater [11] adapted the CW to review procedures, developing the CW-OP (Cognitive Walkthrough for Operating Procedures). The CW-OP includes five key changes to review a broader range of issues than just the usability of the candidate procedures, such as whether a procedure gives the correct direction. Furthermore, Gabrielli, Mirabella, Kimani and Catarci [12] introduced video data to give their expert participants a richer background.

Present Study CW Adaptations

Our walkthrough described below also adapts the CW method, drawing on the work of Gabrielli, et al. [12] and Novick and Chater [11] to develop an approach for a walkthrough of a NextGen (airspace) concept (Ground-based SA). Two major changes were made to the original method. Firstly, questions about human factors issues generated from a NextGen review [1] were swapped for the original four Polson et al. [9] questions to stimulate expert discussion of a wider set of issues that had been flagged as important. Like Novick and Chater [11], we were interested in human-to-human interaction as well as human-machine interaction. Topics such as training, error recovery and key issues for procedure development were also of interest. Secondly, video screen recordings of another controller's use of the prototype in an experimental setting were used in place of static storyboard events. In doing so, our participants saw a dynamic situation, and gained an impression of the time sequence and general level of activity in the ATC sector. Additionally, instead of following a line of "correct actions" per se of designed prototype use, our walkthrough participants followed the interactions and behaviors of another actual user. All of our participants still watched and commented on exactly the same events unfolding in the same way and were still asked what actions they would take themselves. However, the use of another controller's interactions created an opportunity to have our expert participants comment on another's problem solving process in addition to their own.

Simulation Automation

Concept Tools

Our ANSP participants were instructed that the environment they were evaluating was based on a NextGen concept [6] and that it had two automation tools operating for separation assurance. The first tool functioned (primarily in the background) by looking ahead in the simulation by three to 12 minutes for conflicting trajectories, identifying a conflict, and then solving it by sending a trajectory change to one of the aircraft. These avoidance maneuver solutions also included a trajectory to return the aircraft to its original route. The second tool, acting as a redundancy to the first tool, looked ahead in the simulation from zero to three minutes. It functioned in the same way as the first tool, except that it vectored aircraft away from a tactical conflict and relied on the controller to create a trajectory path to return the aircraft to its original route. In our activity we called these tools a Mid-term Separation Assurance tool (MSAT) and a Short-term Separation Assurance tool (SSAT) respectively. These tools operated by looking at a maximum of two aircraft at a time and would assign either a vertical or lateral solution to one of the aircraft via an uplink message.

Example Hypotheses

Based on our review of the general human performance issues within NextGen as well as the SA concept in particular [1], a number of issues were expected to be salient to controllers, and two are presented here as examples to substantiate our walkthrough method:

- As the ANSP is out of the loop due to the tools acting without human review, the ANSP may have trouble forming an understanding of a situation they are required to step into.

- As the automation dictates the timeframe in which the ANSP sees a problem, the ANSP may object that their decision time is constrained.

Method

Participants

Six recently retired controllers, all with more than 25 years of experience with the current air traffic control system, took part in our study as expert participants. In addition to their field experience, this group had a reasonable amount of familiarity with the initial operationalization of some of the NextGen concepts through their participation in prior HITL studies with tool prototypes.

Apparatus

Video files were originally recorded and also played back through the Camtasia Studio 6 screen recorder software. The videos were displayed to the participants on a Barco 28" LCD monitor (ISIS model) with a resolution of 2000 x 2000 pixels (see Figure 1). Additionally, pen and paper story-board frames (see Figure 3) were constructed and printed as talking point references for each of the pre-determined events. Participants' answers were recorded by an Olympus digital voice recorder (WS-311M model).

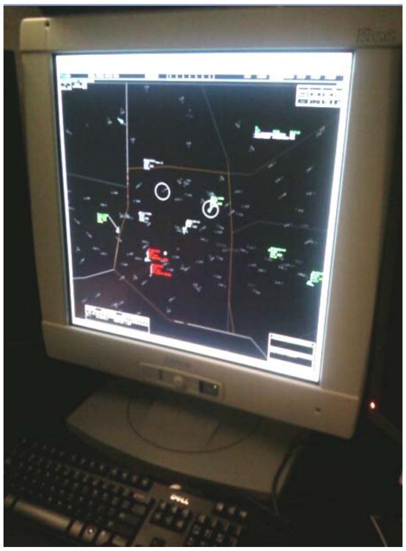

Figure 1. Full scale display to present video footage.

Walkthrough Development

Phase 1 – Walkthrough Events

In 2008, the Airspace Operations Laboratory (AOL, NASA Ames Research Center) ran a HITL simulation of nominal and off-nominal events under NextGen forecast levels of both traffic and automation [2]. These included traffic levels representative of both two times and three times the current day (which is 15-18 aircraft per sector). Throughout this simulation, recordings were taken of the controllers' scopes and were made available for use to our walkthrough activity. These silent screen recordings were reviewed by the present researchers for situations where something went awry for the controller (whether scripted or otherwise) and aircraft on the scope were detected and displayed as being in conflict for potential loss of separation minima (LoS). Such situations were judged critical for exploring a controller's point of view on a proposed future SA operational environment and ripe for human factors informed investigations/queries.

An exhaustive list of all the conflict prediction events was anticipated to be too cumbersome for fruitful discussion. Some conflicts would prove too complicated to conversationally tease apart or did not make sense for our purposes. For example, when three aircraft or more at a time were in confliction, the automation rapidly "changed its mind" in how it identified such a situation as a combination of different confliction pairs. Another common example of instances we purposefully discounted from consideration was where one of the conflicted aircraft pair was off the scope in another sector. Just over a dozen events, representing those that could reasonably drive sensible dissection from our walkthrough participants, were down-selected from the available data.

In order to be sure our walkthrough events and discussions would cover a comprehensive amount of material, a set of event parameters was constructed to characterize the range of possibilities from which we would be selecting (see Figure 2). This scenario parameter chart was developed by characterizing a conflict event and identifying the different ways that the automation and/or a controller could step in to solve the conflict. The parameters included the respective phase of flight of the two aircraft, their equipage status, the entity(ies) through which a resolution was created and lastly, the type of resolution maneuver decided upon. Working from the chart, we cross-referenced its parameters with our remaining video events and generated a preliminary selection of eight events that we felt represented a fair and interesting span of the parameters.

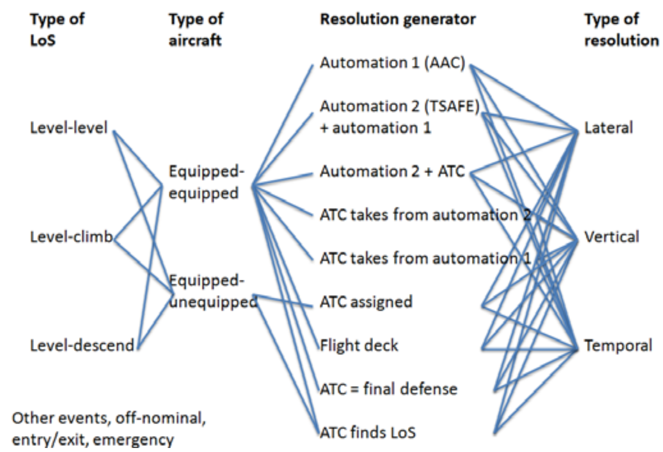

Figure 2. Scenario Parameter Chart.

Next, each of these eight events was analyzed to identify its progression and "storyboards" as collections of these steps (see Figure 3) were drawn up. Steps were defined as turning points in the events where an action or key decision was made by either the controller, the flight-deck, or the automation. For example, when an aircraft flight-deck was sent a clearance, this was counted as a step. This example is shown below in Figure 3 for Event 2, when the controller sent a message to the American Airlines (AAL) aircraft to turn left. On average, the events were broken down into twelve steps, although Event 4 had seven additional steps where a datalink malfunction was reported.

Because the audio channels from the video recordings were not available due to privacy/identification policies, breaking down each event into such detailed steps proved an invaluable process for understanding what was going on. Fortuitously, the absence of an audio track afforded a desirable level of flexibility in reconstructing what we wanted to portray to our participants, thus un-constraining ourselves from minor variances in what might have actually happened in the previous study's simulation.

These storyboards then facilitated another round of event evaluations that made it easier to see which events might not elicit enough useful or unique information (e.g., too similar in situation to another event) from those of unique interesting circumstances (e.g., a non-compliance issue, an equipment failure etc.), leaving us with a final set of four events to walk through with our participants. These events are as follows:

Event 1: Southwest Airlines (SWA) flight 339 begins a north-westerly climb that will take it through the trajectory of Atlantic Southeast Airlines (ASQ) flight 1360, which is in level flight traveling southeast. The conflict is first flagged by SSAT (the Short-term Separation Assurance Tool) at three minutes. The first resolution (for ASQ1360 to turn left) is not executed and has to be resent.

Event 2: SWA1864 enters the sector on a north-westerly climb that will take it through the trajectory of AAL140, which is in level flight traveling northeast. The conflict is first flagged by SSAT at one minute. ATC asks AAL140 to turn left and when this is possibly not enough for separation asks SWA1864 to also turn left.

Event 3: Selected from a different experimental run, this event takes place in a lower traffic level of approximately 1.5 times current day traffic. Jetlink (BTA) flight 39 and AAL711 enter the sector, in close proximity, BTA39 is flying southwest and AAL7111 is flying northwest. AAL711 is climbing through the trajectory of the BTA39. SSAT sends a resolution to BTA39 to turn left, while at the same time the controller issues a temporary altitude hold to AAL711.

Event 4: a scripted failure of the datalink capability on an aircraft, Northwest Airlink Airlines (FLG) flight 144, flying approximately west to east across the lower portion of the sector brings it into level conflict with a Northwest (NWA) 612 flying approximately south-west to north-east. An additional layer of workload is added because as the FLG144 aircraft loses its datalink, a multi-aircraft conflict is developing in the northeast of the sector. The conflict is tracked by MSAT and SSAT but is resolved by the controller because the clearance SSAT sends (repeatedly) to NWA612 goes unanswered.

Lastly, it is important to recognize that although our events came out of a prior simulation, the processes to specify which events we would select (via a scenario parameter chart) and how they would be presented (via a storyboard framework) are illustrative of how such walkthrough events might be designed from scratch and ultimately built (e.g., via animations or movie-making tools) in the absence of previously prototyped material.

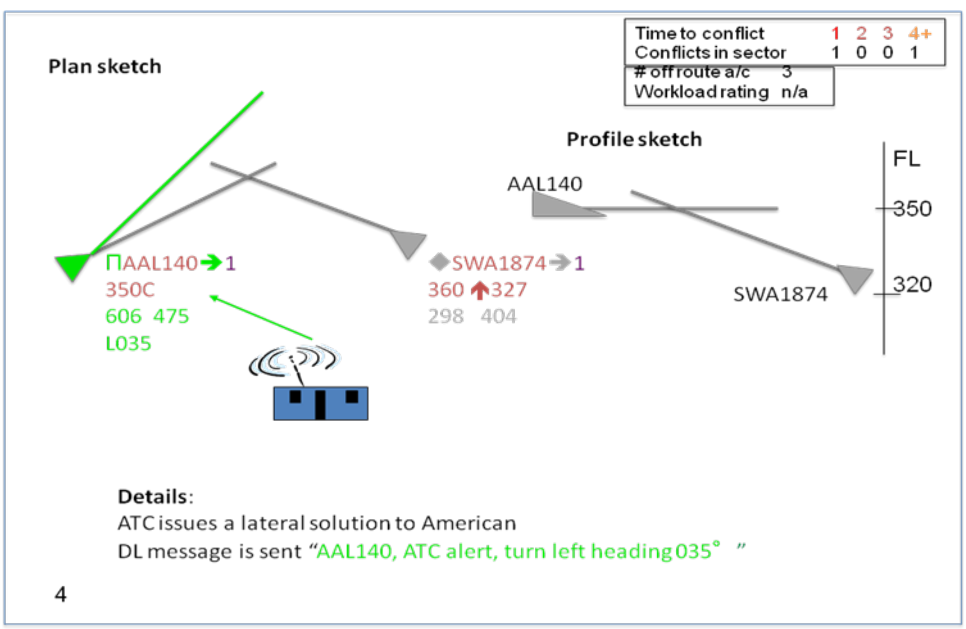

Figure 3. Storyboard of step 4 from Event 2

Phase 2 – Walkthrough Questions

A relevant question that addressed either a cognitive walkthrough or human factors theme was identified for each step in the event breakdowns. Questions were not unique—if the research group thought a question was relevant or important in more than one event or step it was asked again.

Beginning with the traditional set of cognitive walkthrough questions [10], it became apparent that it would be inappropriate to ask them at each step of our walkthrough. At our level of event analysis, too much would be repeated between some steps to warrant asking the same cognitive question again, and due to the magnitude of ground to be covered, only the most relevant cognitive walkthrough question was planned to be asked at any given step. Also, due to our participants' perspective of another person's actions, the traditional cognitive walkthrough questions were not asked verbatim but instead were rephrased or adjusted where necessary to match the walkthrough's situation.

Additionally, from the larger project [1], a set of eight categories important to an automated SA concept was used to generate questions to complement the cognitive walkthrough questions. The ninth category "recovery from error" was not included because this was the criterion on which the events had been chosen. Extending the eight categories with an additional ten subsections provided 18 points on which to probe our participants' judgment of the proposed concept (see Table 1). Initially as many relevant questions as possible were constructed for each step of each event from the human performance themes. As with the cognitive walkthrough questions, the next step included paring these down to the most relevant or fitting questions in order to keep the size of the walkthrough at a manageable granularity. Efforts were then taken to balance these questions both within and across events so that each category was equally represented. Questions from the "organizational" category, however, were asked in a separate round with fewer steps because these proved too repetitive in practice runs of the walkthrough, due to their wider concern with aspects prior to and off the scope from any given step. Overall 136 human performance themed questions were asked of each participant across our 18 categories/subtopics to complement the cognitive walkthrough questions.

Table 1. Human Performance Themes and Subtopics

NextGen Human Performance Themes	**Subtopics**
Attention	
	Monitoring
	Situation Awareness
Decision Making	
	Time Pressure
Workload	
	Task Management
Communication	
Memory	
Organizational	
	Roles
	Responsibilities
	Coordination
Interaction with Automation	
	Trust
Selection / job qualification / certification	
	Training
	Procedures

As an example, Figure 3 shows Step 4 from Event 2 where the cognitive walkthrough question was "Would you have done this?", the human factors category identified was time pressure and its question was "Is DL fast enough for this resolution?" On a subsequent round for organizational questions, the participant was asked "As you take the conflict, how should control shift from automation to human?".

Data Collection Procedure

Participants received a short orientation, which explained our purpose and introduced the sector that was the focus for the day. It was explained that all the events took place in one high altitude en-route sector (above Indiana, i.e., ZID91) in the Indianapolis (ZID) Center. This sector is a busy area, taking arrivals and departures from St Louis to the west and Louisville to the east. The automation notation on the display was briefly described but as all participants had worked with these tools in the AOL before, it was assumed that they understood the majority of icons and display elements. Beyond a list of tasks that they would be required to assume the responsibility of performing (e.g., put aircraft separated by SSAT back onto their original routes, deal with off-nominal or emergency circumstances), ANSP procedures, operating rules and jurisdiction were not formally defined for our participants because discussion of these were the aim of the explorative method.

Each of the four events was shown three times to each participant. First, the event was played in real time. Participants were free to ask questions or make comments as they watched. Second, the event was stepped through by forwarding the video to each step and then paused at a point just after the action for that step had happened. Figure 4 below shows an example of the way the problem area on the display for Event 2 looked as it was paused at the step described in Figure 3 above. One member of the team briefly described the occurrence in the step and a question with a human factors theme was asked, plus a cognitive walkthrough style question, if one had been deemed relevant for that step. In the third playback, a subset of the steps from the task breakdown was selected and an additional organization question was asked at each one of these.

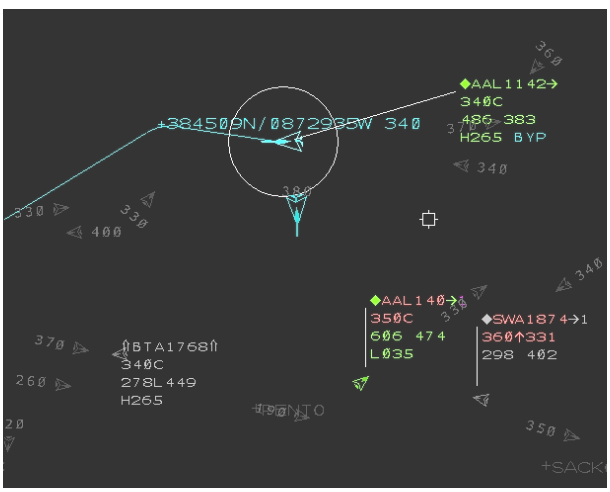

Figure 4. Snapshot from step 4 of Event 2.

Participants' answers were recorded on a small hand-held digital recorder and later transcribed. Initial analyses for this report selected 56 key questions that addressed the 18 themes. Once these questions had been transcribed and summarized, they were cross referenced with both the specific SA human factors issues and a set of general human factors issues from the NASA NextGen human performance issues report [1]. As noted before, our primary goal was not to gain answers to our questions. We wanted to find out whether and how the issues we predicted would be described by participants and to determine whether a cognitive walkthrough of events would be a useful method for collecting data about potential human factors issues of prototype systems.

Results

The exhaustive set of questions and answers exchanged during our walkthrough would be too numerous to list here in its entirety. Instead, to illustrate the method, a small sample of talking points was selected for initial analysis and reporting. Qualitative analyses that were performed involved uncovering consistency across and within our controller participant comments (i.e., where comments were made multiple times or made by multiple controllers). There were two steps to this process. First, participants' answers to questions on the same human performance topic were grouped and common themes were highlighted. Second, these themes were related to our hypotheses and an assessment was made whether they supported or opposed our initial assumptions. To illustrate this process one of our human performance topics and the relation of its themes to some of our example hypotheses is presented next.

Common Themes Within Participant Answers

Example: Attention, Situation Awareness

For participants, a handful of cues concerning aircraft speed, angle of closure and proximity, are significant and thus attention "grabbing" because they signal a rapid narrowing of solution options. From questioning our participants it became evident that controllers are pattern recognition experts and part of their skill lies in an intimate familiarity with the sectors they work. Controllers know which areas of their sectors to watch more closely, and aircraft actions that do not conform to normal patterns are very salient cues. Often sectors have "hotspots", where corridors cross or a departure flow is climbing to altitude, and participants are particularly wary of aircraft popping up here in conflict with aircraft in level flight.

Four participants highlighted the importance of localized knowledge (i.e., sector flows, pilot alertness, etc.) in shaping what they would expect of aircraft under their control. Other answers illustrated the value of retroactive (long term) memory. Because current air carrier schedules are regular and aircraft generally fly the same routes every day, controllers become familiar with the paths of the aircraft that fly through a sector during on-the-job training and by regularly working a sector. Participants said that if they were trained on a sector they would be able to remember the route to put an aircraft back onto if it had been vectored off its route for separation. These replies tied in with other questions about memory and training that highlighted the extent of current controllers' skills. Under NextGen, it is likely that controllers' required skill sets will change, but these responses underlined the value of specific situation awareness and localized knowledge.

Assessment of Common Themes Grouping

Answers to all the questions within a human performance theme were collected into paragraphs similar to these examples to give eighteen human performance theme responses. While participants' answers varied, the consistency with which they discussed some of the themes encouraged us that, in general, participants understood and interpreted our questions in the same way. One potential issue with changing the CW method to ask a wider range of questions that did not have a "correct answer" per the traditional CW framework, was that participants might interpret questions differently and answers would not be comparable or able to be summarized. That participants' answers could be grouped into common themes showed us that this was not the case, and that we had been successful in gaining information on a range of topics.

Relating themes to hypotheses

One of our two aims for this walkthrough was to take a "first-cut" investigation of potential HF issues from a user's perspective. The expert analyses described in [1] had been used prior to the walkthrough to generate possible hypotheses of human factors issues that might arise for the example ground-based SA concept and automation tools. Using the paragraphs constructed in the previous phase, each hypothesis was considered in the light of participants' responses. How the human performance responses addressed a hypothesis and whether they supported or opposed it was noted.

Example of an unsupported hypothesis:

- *As the ANSP is out of the loop, due to the tools acting without human review, the ANSP would have trouble forming an understanding of a situation they are required to step into*. From the common themes, it was concluded participants were able to follow the traffic and often predicted future conflicts before the automation flagged them, suggesting they had an awareness of the presented situations. However, our participants were controllers trained in the current day manual methods. When training is updated to encompass NextGen tools, this kind of pattern recognition may not be a skill-set that is developed by or required of incoming ANSPs. Furthermore, an operational environment that opposes or downplays the development and application of localized knowledge would be in opposition to a fundamental nature we found expressed repeatedly across our controller participants.

Example of a supported hypothesis:

- *As the automation dictates the timeframe in which the ANSP sees a problem, ANSP may object that their decision time is constrained*. This prediction was supported. If the automation flagged a conflict at or before three minutes, participants estimated that they would have enough time to solve a conflict (although three minutes was identified to be "cutting it close"). However, in three of the off-nominal events that were presented, the automation did not flag the conflict until there were two minutes or less to a loss of separation (LoS). In these cases, participants felt that they had not been given enough time to create an elegant solution to the problem, and perhaps not enough time to solve the LoS at all.

Taking the results we gained from the cognitive walkthrough as a whole provided some surprising answers to the question marks that arose from the initial general analysis of the human factors issues in SA [1]. For example, participants reported they were comfortable with the information provided and could find solutions to aircraft conflict problems given enough time. These positive results can guide the next iteration of development of SA concepts, tools, and procedures.

Some human factors areas that were suspected to cause issues *were* shown to be problematic through participants' answers. For example, there was ambiguity in the division of roles and responsibilities between human and automation in the concept. This was especially the case when participants were asked how the blame for operational errors should be apportioned. Comments underlined that clear and comprehensive procedures are needed for a SA concept – a finding that can be reported back to the concept developers to assist them in focusing their development efforts.

Assessment of Relating Themes to Hypotheses

A concern at this step of our analysis was that our specific walkthrough questions would not combine to provide answers to the broader hypotheses and concept questions that we had. Being able to relate our themes to our hypotheses to give relevant answers indicated that we had kept the essence of the issues in which we were interested in the CW questions. Addressing the study hypotheses also provides useful information for concept developers. Developers can use the themes to identify areas of their concept that require more clarification or scoping.

Discussion

In an effort to create a linkage between the human factors issues with the integrated performance of the human operators in the NextGen concepts, one is caught in a "catch 22.” It is difficult to generalize on human capabilities to perform tasks until and unless there is specification of what the tasks are. However, if the hardware/software is "cast in concrete" without some simulation to check out the human operator interactions with that hardware/software, then it is too late. There is evidence in previous major ATC automation developments of how costly that mistake can be [13].

Further exploration of these issues requires instantiated concepts, procedures, and sample traffic scenarios. HITLs provide an invaluable environment to examine these NextGen instantiations within high-fidelity situations, but unfortunately often focus on concept benefits and feasibility while also coming with great expense and investment. A smaller human factors concentrated walkthrough activity with fewer people and less simulation of the researched system can serve as an effective independent or complimentary endeavor to HITLs.

Two aspects of the images presented to our particular set of participants that were important were that the whole display was presented and that the images could be played out in real time (even though we paused the video to ask questions). The value of presenting the whole display was that participants gained an awareness of how much traffic “2x” or “3x” looked like and what kinds of events could be occurring concurrently (i.e., the size or scope of the task). The value of seeing actions in real time was that participants could estimate their workload and task / time pressures. These two factors may possibly be invaluable only in this instance because the particular ANSP position and function we looked at raised issues and concerns about these specific HF topics. But, more generally, time, workload, and task scope are certain to be HF topics of interest for most NextGen concepts.

One criticism of our particular approach to the walkthrough could be that we used previously recorded video of a working prototype as our stimulus. If such video recordings were not available and a researcher wished to use this method to inform initial prototype development, how should it be done?

Video like the excerpts played to participants could be recorded from developmental test runs prior to an actual HITL simulation. As the walkthrough process can be quickly completed it would be possible (given some preparation) to complete it during the shakedown phase of a large simulation. Alternatively, graphic images of the display, and hypothetically what procedures and problems could happen, could be built as a movie, independently of any tool, before working prototypes are available. After all, in our walkthrough, participants were not able to interact with the display at all because they were shown a recorded video, not a real-time reconstruction of the event, on a workstation.

Our walkthrough method was successful from several standpoints. First, useful and informative data was obtained and was rich enough to fulfill our aim to identify and explore the impact of HF issues in advance of a developed tool design as discussed above. Second, these data were obtained from relatively few participants. Six participants provided enough repetitions to demonstrate that answers were consistent but also offered enough variety to see that points of view did vary. Third, a small team was able to compose and conduct the research. Overall, this experience addressed our aim to describe and test a method that could be used by concept research teams to take a “first-cut” investigation of potential HF issues.

Overall, the walkthrough provided important unique insights into the human factors issues previously identified as problems for the NextGen SA concept. While some factors were found to be less problematic than expected, other suspected issues were supported with concrete examples provided from our participants. These results show the benefits of combining an environment prototyped to a purposeful level of fidelity with a cognitive walkthrough methodology as a low cost alternative or compliment to running HITLs in examining human factors issues for NextGen concepts.

References

[1] Lee, P., Sheridan, T., Poage, J., Martin, L., Cabrall, C. & Jobe, K. (2009). *Identification, characterization, and prioritization of human performance issues and research in the Next Generation Air Transportation System (NextGen)*. NASA NRA NGATS Subtopic 9 for the System Level Design Analysis and Simulation Tools element of the NASA Next Generation Air Transportation System ATM-Airspace Project. San Jose, CA: San Jose State University Foundation

[2] Prevot, T., Homola, J.R., Mercer, J.S., Mainini, M.J. & Cabrall, C.D. (2009). Initial evaluation of air/ground operations with ground-based automated separation assurance, *Proceedings of the 8th USA/Europe Air Traffic Management Research and Development Seminar*, Napa, CA.

[3] Joint Planning and Development Office (2007). *Concept of Operations for the Next Generation Air Transportation System*, Version 2.0, Washington, DC.

[4] Mohler, G. (2007). How OEP will lead FAA through the NextGen transformation. *Integrated Communications, Navigation and Surveillance Conference, 1-3 May 2007, Virginia, US*. IEEE Conference Proceeding

[5] Federal Aviation Administration (2009). *FAA Aerospace Forecasts FY 2009-2025*, Washington, DC., p.31.

[6] Erzberger, H. (2001). The automated airspace concept. *Proceedings of the fourth USA/Europe Air traffic management R&D seminar*, Santa Fe, NM, USA, December 3-7, 2001.

[7] Polson, P.G. & Smith, N. (1999). The cockpit cognitive walkthrough. In *Proceedings of the Tenth Symposium on Aviation Psychology*, Columbus, OH.

[8] Jeffries, R., Miller, J.R., Wharton, C. & Uyeda, K. (1991). User interface evaluation in the real world: a comparison of four techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*. New Orleans, Louisiana, pp: 119 – 124.

[9] Polson, P.G., Lewis, C.H., Rieman, J. & Wharton, C. (1992). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies, 36*, 741-773.

[10] Wharton, C., Rieman, J., Lewis, C.H. & Polson, P.G. (1994). The Cognitive Walkthrough Method: A practitioner's guide. In Nielsen, J. & Mack, R.L. (Eds), *Usability Inspection Methods*. New York: Wiley. Chapter 5, pp 105-140.

[11] Novick, D.G. & Chater, M. (1999). Evaluating design of human-machine cooperation: The Cognitive Walkthrough for Operating Procedures. *Proceedings of the Conference on Cognitive Science Approaches to Process Control (CSAPC 99)*, Villeneuve d'Ascq, FR, September, 1999. www.cs.utep.edu/novick/papers/cw-op.csapc99.html

[12] Gabrielli, S., Mirabella, V., Kimani, S. & Catarci, T. (2005). Supporting cognitive walkthrough with video data: A mobile learning evaluation study. ACM International Conference Proceedings Series: Vol. 11: *Proceedings of the 7th International conference on Human computer interaction with mobile devices & services*, pp 77-82.

[13] Walker, M. (2004). *Analysis for Enabling Benefits at User Request Evaluation Tool (URET) Field Sites.* MITRE Technical Report, MTR04W0000082. McLean, VA: MITRE

Acknowledgements

The authors wish to extend gratitude for the expert guidance provided by Tom Sheridan, and Jim Poage, JPL Consulting, in the analysis of human performance issues in NextGen.

This work was funded by an NRA grant through the Airspace Systems Program of the NASA NextGen Project, and facilitated by Immanuel Barshi, as its COTAR.

This work could not have taken place without the active support of the cadre of retired controllers.

Lastly, the authors are indebted to the Airspace Operations Laboratory (AOL) at NASA Ames

Research Center for the generous use of their resources and expertise.

Email Addresses

Christopher.D.Cabrall@nasa.gov

Lynne.H.Martin@nasa.gov

Paul.U.Lee@nasa.gov

Kimberly.K.Jobe@nasa.gov