

N95-34761

Evaluation in Context: ATC Automation in the Field

Kelly Harwood & Beverly Sanford

Sterling Software

Introduction

Certification is defined as "attesting as certain" (Flexner & Hauck, 1983). "Certainty" however may be a rare commodity when the introduction of new technology into an existing system can "...destroy the blanket of established know-how" (Rasmussen & Goodstein, 1988; p. 179). It is impossible to foresee all emergent properties and interactions between system components and their implications. A complete set of requirements and criteria for safe and efficient system functioning is difficult, if not impossible, to define in advance of system implementation. Once the system is in an operational environment, requirements may need to be rejuvenated due to our imperfect foresight and lack of understanding. Christensen (1958) has referred to this dilemma as the "omnipresent criterion problem."

One way to tackle this dilemma is to incorporate field testing early in the system development cycle. This paper describes the field assessment process that has been applied to the development of an advanced ATC automation system, the Center/TRACON Automation System (CTAS). Field testing provides insight into the true characteristics of the system; that is, how it actually operates and any emergent properties as a function of being integrated into the operational environment. Such insight provides guidance for capturing and refining meaningful requirements for system verification and certification. By delaying field testing until late stages of development, solutions to design problems are likely to be technology driven with validation, verification, and certification relying on context-free guidelines for human-computer interaction.

Field testing conducted early during the development and demonstration phase of system development affords exploration of the user's experiences with the system in the context of their work domain. It provides the opportunity to understand the implications for system design of the interdependencies between the physical environment (lighting workplace layout), task domain (goals/functions of the domain) and work activities (social aspects of team coordination; sources of motivation and job satisfaction). The richness and complexity of these context-based factors and the relationships between them is not accessible through design guidelines or standards. Guidelines and standards cannot provide insight into effective design solutions when system performance is highly contingent on context (Meister, 1985; Gould, 1988). Early field testing promotes the development and validation of a tool as a problem-solving instrument (Woods, Roth, & Bennett, 1990), thereby increasing the likelihood of a match between the system's capabilities and its context of operation (Rasmussen & Goodstein, 1988; Bentley et al., 1992).

Human Factors Certification of Advanced Aviation Technologies
Edited by J. A. Wise, V. D. Hopkin, and D. J. Garland
Copyright © 1994 Embry-Riddle Aeronautical University Press

PRECEDING PAGE BLANK NOT FILMED

PAGE 246

The FAA TATCA Program recognizes the importance of early field testing for the development and validation of advanced ATC automation. It is presently using rapid prototyping and early field exposure as part of the development of CTAS, using on-site system evaluations with active controllers and representative traffic flows and conditions. Iterative field testing is regarded as integral to the development process, with the objective of achieving a match between the system and context for its use. This approach deviates from traditional approaches to ATC system development and will expedite a possible national deployment of CTAS. Embracing the context of the ATC domain is particularly important because of our limited knowledge of the impact of advanced information technology on controller/team job performance and the stringent requirements for maintaining ATC system safety and continuity during system transition (Harwood, 1993).

The first section of the paper provides a brief description of CTAS, followed by an overview of the field development and assessment process in the second section. In the third section, particular attention is paid to the *structured* assessments of CTAS. These assessments take a principle-driven approach, drawing on principles, perspectives, and methods from human factors engineering, cognitive engineering, and usability engineering. Activities are described that include the identification of human-centered system issues to help guide the collection and interpretation of data, method selection and tailoring, data collection, data analysis, and interpretation. Examples are provided of the types of findings that are a consequence of this development and assessment process. The fourth section discusses requirements definition and rejuvenation. This paper is not a comprehensive review of all possible methods that could be used, but rather a description of those that have been applied in tailoring a process to bring CTAS functions to a level of stability and usefulness. Emphasis is on the mechanics of executing the process, with mention made of the nuances of conducting development and assessment at an operational field site.

CTAS

CTAS is an integrated set of automation tools, designed to provide decision-making assistance to both Terminal Radar Control (TRACON) and Center controllers via planning functions and clearance advisories. CTAS consists of three sets of tools: the Traffic Management Advisor (TMA), the Descent Advisor (DA) and the Final Approach Spacing Tool (FAST). CTAS development has involved thousands of hours of laboratory simulation with controllers to refine and extend algorithms and to enhance the user interface. In order to bring system functions to a level of operational stability and to provide information to air traffic and system development organizations on a possible national deployment decision, further development and validation is being conducted at FAA ATC field sites. TMA is the first CTAS component to undergo the field development and assessment process and will be the focus of discussion for this paper. (For further information on CTAS see Erzberger & Nedell, 1989; Tobias, Volcker, & Erzberger, 1989; Davis, Erzberger, & Green, 1991; ATC Field Systems Office, 1992; Erzberger, 1993.)

TMA has been developed for use by the traffic manager at traffic management units within Air Route Traffic Control Centers and TRACON facilities. Unlike controllers, traffic managers do not control traffic directly. Instead, they monitor the demand of arrival traffic into the center and terminal areas, coordinating with TRACON, center, and tower personnel, making

decisions to balance the flow of traffic so that traffic demand does not exceed airport and airspace capacity. Traffic managers use information about the arrival flow to decide whether the traffic should be delayed or metered, to distribute the load from one area to another, and to assign departure times for aircraft departing airports within the center's airspace that will enter the arrival flow for the metered airport. Information about the traffic situation is accessed from multiple sources, such as flight strips, weather displays, operational personnel, and aircraft situation displays. Often, there is no steady state in the traffic flow; the location of a single *heavy* aircraft can disrupt the scheduled flow of traffic, as can poor weather, equipment outages, and emergencies. Given the extent of coordination required, the variety of sources of information accessed, and the dynamic and often variable state of the traffic flow, context through early field testing is crucial to ensure the robustness of TMA and its effective integration into the traffic management unit.

Representations of traffic flow are conveyed on the TMA by configurable moving timelines. Aircraft data tags move down the timelines and are color coded to portray landing schedule and sequence status information. The traffic manager can override TMA's automatically generated schedule at any time by resequencing aircraft, inserting slots for additional aircraft, or changing the airport acceptance rates. A traffic load display provides a graphical representation of various traffic load characteristics, and several configuration panels are available for modifying timeline displays and setting schedule parameters. The workstation consists of a SUN4@ SPARC™ workstation with keyboard and mouse input devices. TMA presents the traffic management coordinator with new capabilities that are a significant departure from the current traffic management system. The next section describes the process that has been applied for developing and assessing TMA at an operational field site.

Field Development and Assessment Process – Overview

Development and assessment of CTAS is currently underway at two FAA ATC field sites. This paper focuses on the development and assessment of TMA at the Center and TRACON of one of the field sites. TMA is accessible in the traffic management units at the Center and TRACON. A one-way interface with the current HOST system is available so that TMA can reflect the current traffic situation. Traffic managers reference TMA during traffic rush periods and in off times to explore its capabilities and to understand how the tool can be used to solve traffic management problems.

The field development and assessment process is geared for system refinement and enhancing our understanding of the potential impact of TMA on traffic management problem-solving and inter-facility coordination. The process also provides insight for various program objectives, such as operational procedures and requirements definition. To do this expediently, two mechanisms are in place to allow the timely transfer of information from the field site back to the primary development site at NASA-Ames. These are "unstructured" and "structured" assessments. Both are described briefly here, and key aspects of the structured assessments are elaborated further in the third section.

Unstructured assessments are performed by traffic management personnel on a daily basis, during traffic rush periods and during off-times. Here traffic managers access TMA representations, comparing data between TMA and the existing system at the Center, or with decisions made from compiling many separate sources of information together at the TRACON.

Human factors engineers and development personnel may observe TMA-use for the purpose of understanding the system, but goal-directed data collection does not occur at this time. Unstructured assessments are for traffic managers to experience the system and provide feedback without an audience or intrusion. Having the system available on a continuous basis provides exposure to a variety of traffic flow and weather situations, allowing the users to "shape" TMA-use to fit the problem-solving demands of their environment. This process is instrumental for engendering trust in the system.

Structured assessments are conducted to systematically investigate tool use and to capture the user's experience with TMA. How the traffic manager uses the tool in response to various problem-solving demands is an important gauge of the match between TMA features and functions and the context for their use. It has been argued that a major cause of system failure is a mismatch between the system capabilities and the demands and constraints of the operational environment (Bentley et al., 1992). Calibrating the match is thus a key activity during development for ensuring system success. Structured assessments are conducted by human factors engineers and development personnel and provide feedback for further development and program milestones. Methods and approaches for structured assessments are described further in the third section.

Quick transfer of information from the users to the development site and back to the users again is critical for continuity at the field site. Timely feedback to the traffic managers on their questions and suggestions during unstructured assessment is essential for maintaining their interest and involvement as well as for streamlining the development and assessment process. An electronic-mail system connecting the traffic managers at the center and TRACON to NASA-Ames and field-site development personnel has facilitated timely information transfer. Questions are addressed immediately, and design issues from structured and unstructured assessments are entered into a data-base managed at NASA-Ames. Resolution and augmentation of TMA features and functions are decided by committee, with representation from program management, developers, human factors engineers, and testing personnel. Issues are categorized and prioritized by the committee according to their pragmatic and technical implications; for example, implications for system usability and operational suitability, availability of development resources, the need for further analysis, and objectives of an upcoming structured assessment. Refined software is shipped back out to the field sites on a near-monthly basis.

Structured Assessments – Methods and Approaches

Structured assessments of TMA take a principle-driven approach, drawing on principles, perspectives, and methods from human factors engineering, cognitive engineering, and usability engineering. These fields provide a knowledge base from which methods and approaches for validating system designs may be derived. Structured assessments focus on specific aspects of the users' experience with TMA and consist of several activities:

- Issue identification
- Method selection
- Data collection
- Data interpretation
- Analysis, inferences and implications.

These activities are described next. Each activity relies heavily upon the operational context of the traffic management unit, focusing on exploration and discovery as well as assessment.

Issue Identification

Operational requirements for TMA are currently being defined and thus are not yet available for verifying the system design. This lack of guidance is compounded by the generality of ultimate criteria for ATC – namely, safe, orderly, and expeditious flow of traffic, and the general lack of knowledge regarding performance of individual and controller teams in current and future ATC environments. In the absence of requirements and criteria, there is a risk of collecting data that may be expedient but inappropriate (Parsons, 1972; Hopkin, 1980). To compensate for this knowledge gap and to systematically guide the collection and interpretation of data to support the refinement and validation of TMA, we focused on three broad categories of human-centered system issues:

- Technical usability
- Domain suitability
- User acceptance.

These three categories are also of interest to the FAA for its specification of operational requirements and for formal operational test and evaluation. Others have distinguished previously between two or three of these categories (e.g., Hopkin, 1980; Gould, 1988; FAA, 1989; Rasmussen & Goodstein, 1988). Categories and approaches for defining issues are described briefly. Further details can be found in Harwood (1993).

Technical Usability. Technical usability refers to perceptual and physical aspects of the human-computer interface such as display formatting, graphics, and human-computer dialog as well as anthropometric characteristics of the workstation. Issues in this category address the general question: Can the users extract and access the data needed to do their job? A tremendous amount of research in human factors engineering and human-computer interaction has contributed to the development of principles and guidelines for designing and evaluating human-system interfaces (See Department of Defense, 1989; Shneiderman, 1987; Smith & Mosier, 1986; Van Cott & Kincade, 1974). These principles constitute the basis for defining technical usability issues.

Domain Suitability. Simply addressing issues of interface usability does not necessarily provide insight into the suitability of the automation tool for the domain. Here it becomes necessary to address domain suitability, which refers to the content of information and representations for domain tasks as well as functions and decision-aiding algorithms. Issues in this category address the general question: Does the representation support the problem-solving requirements of the domain? In contrast to technical usability, which is driven by issues of technology utilization, domain suitability requires an understanding of the "cognitive problems to be solved and challenges to be met" (Hollnagel & Woods, 1987, p.257; Rasmussen, 1986; Rasmussen & Goodstein, 1988).

The fundamental basis for understanding the types of cognitive demands that can arise is a description of the domain in terms of the domain goals to be achieved, the relationships between these goals, and means for achieving goals (Rasmussen, 1985, 1986; Woods &

Hollnagel, 1987; Rasmussen & Goodstein, 1988). This sort of system description, in terms of a goal-means decomposition, is particularly useful for system evaluation: it guides the description of the cognitive situation that the design must support and it guards against narrowly focusing on problem-solving demands in only one aspect of the work domain.

User Acceptance. User acceptance is obviously enhanced by the ease of use and suitability of the system for supporting cognitive task requirements. Yet user acceptance also depends upon job satisfaction, professionalism, esteem and opportunities for demonstrating individual merit. Hopkin (1980, 1992) has argued that such issues are usually overlooked in the context of technology upgrades, but may have serious consequences for ultimate system safety and efficiency. Attention must thus be devoted to disclosing issues associated with the impact of new technology on ATC job satisfaction.

Context is critical for understanding the impact of new system upgrades on sources of job satisfaction and professional merit. What is satisfying and motivating about a job is as much a factor of the individual as it is the nature of the tasks and work domain. Ethnographic techniques for understanding the work environment are thus instructive for capturing valid descriptions of sources of job satisfaction. Such techniques are geared to the study of complex social settings to understand what aspects of activities are important and relevant to individuals. In general, ethnographic techniques have been recognized as essential to understanding, designing, and evaluating complex systems (Bentley, et al. 1992; Hughes, Randall, & Shapiro, 1992; Hutchins, 1992; Suchman, 1987; Whiteside, Bennett, & Holtzblatt, 1988; Suchman & Triggs, 1991).

Issue identification for TMA has been based upon hundreds of hours of observing traffic management activities, reading operational documents on traffic management, and interacting with traffic managers. Approaches for identifying issues are contextually based, that is, based upon an understanding of the physical characteristics of the environment, causal relationships between goals and functions in the task domain, and characteristics of the work activities. Focusing on only one or two of these factors risks collecting data that will not provide insight into sources of design deficiencies or provide a basis for defining meaningful human factors system requirements.

Method Selection

The operational field-site is important for gaining insight into the match between an automation tool design and its context for use. The complexity of the operational environment, with its inherent task demands and the access to operational personnel, allows discovery of unexpected feature use and assessment of the extent to which the tool will support its users. However, while testing at an ATC field site offers a unique perspective on system effectiveness, it also presents a number of constraints that preclude typical laboratory practices and techniques (Johnson & Baker, 1974).

The availability of controllers and scheduling and resource constraints can severely restrict the extent to which different conditions or system configurations can be investigated. In addition, sample sizes may be small, with the number of replications limited to a single trial. The physical environment is natural and intrusive factors are uncontrolled. Variables are driven by the system, not the experimenter, and the units for measurement are macro-units in the order of minutes. Measures are more often qualitative rather than quantitative.

Given these constraints, our expectations of field assessment must be adjusted appropriately. Field assessments provide an opportunity for capturing the users' ongoing experience with the tool, discovering how new functions will be used and where mismatches occur between the capabilities of the technology and the user's needs. Field assessment provides insight into the integration of a new automation device into an existing environment, indicating issues for transition training and operational procedures. However, field testing is only one level of system evaluation, often augmenting simulation and laboratory testing. Field testing is not a panacea, but it provides an important and necessary perspective for achieving system success.

To accommodate the constraints associated with field assessment and to maximize the opportunity of accessing the operational site, methods must be tailored accordingly. Several criteria guided the selection of methods for assessing TMA:

- Methods must capture the user's ongoing response to the system
- Methods must be sensitive to design deficiencies
- Methods must provide opportunities for discovering new strategies and system functions
- Methods must not disrupt traffic management operations.

Context-sensitive data collection techniques, that is, techniques based on observation and interpretation in the context of the user's work environment, meet these criteria (Whiteside, Bennett, & Holtzblatt, 1988). Such methods include observation and contextual interviews with active involvement of the users in the interpretation of the observations.

Field assessments of TMA, to date, have focused on capturing the traffic managers' experience with the TMA. Whiteside and his colleagues have argued for the importance of the users' experience as valuable information for engineers about users needs. The appropriateness of features and functions "...exists in the experience of the user, and experience is driven by the context in which it occurs" (Whiteside, Bennett & Holtzblatt, 1988; p. 809). Capturing the users' experience with the tool is especially important for complex, ATC automation systems, where the implications of the interactions between system components and emergent properties are largely unknown prior to implementation. When validation of system designs rests on reconciling technological possibilities with work needs, the users' experience assumes an important role.

Data Collection

Assessments are conducted in the traffic management areas at the center and TRACON. This location serves both technical and pragmatic interests. Traffic management involves extensive coordination with other traffic managers and area supervisors, communications with other facilities, and accessing and integrating information from a variety of different sources, such as weather displays, aircraft situation displays, and flight strips. Accessing TMA-use in the context of these operational activities is essential for addressing domain suitability and user acceptance. In addition, access to operational lighting conditions is desirable for validating such technical usability issues as color discrimination and readability. Lighting in the operational area is complex, with overhead lighting located in high ceilings and local lighting on work surfaces.

From a more pragmatic perspective, the location of the test area accommodates resource constraints and works well with the culture of the unit. To date, it has not been possible to

schedule participants prior to the assessments. Instead, the supervisors on duty release traffic managers when staffing and the traffic demand allows. Having the supervisors control access to the traffic managers minimizes the impact on the unit, thereby increasing acceptance of the assessment process. Supervisors release and summon traffic managers as the conditions permit. Modular organization of data-collection materials, and non-intrusive observation are also flexible to accommodate this scheduling constraint.

Several different methods are used to collect data for assessing TMA. Scenario-driven surveys using prerecorded traffic data are used to assess technical usability. Shadowing of traffic management operations is used to assess technical usability, domain suitability, and user acceptance. These methods are described next, with particular attention given to the mechanics of their execution.

Scenario-Driven Surveys. Scenarios systematically guide the traffic managers through the display and interactive features of TMA and instruct them to view or manipulate different features. Pre-recorded traffic data are used to ensure that everyone views the same traffic conditions during the exercise. Associated with each scenario are validation statements that focus on specific technical usability issues, such as color discriminability, symbol detectability, and ease of interacting with the input devices. Traffic managers indicate whether they agree or disagree with the validation statement, and space on the survey is provided for comments and suggestions. A human factors engineer sits with the traffic managers as they complete the survey, answering any questions, and observing TMA use. Scenarios generally progress from being easy and simple to more difficult and complex. This is arranged to gauge the level of understanding of basic TMA features in the implicit check of TMA proficiency. If a participant is deficient in any area, instruction is provided, and the session is treated as training instead of as TMA assessment.

Technical usability issues are assessed for all TMA modifications, new features and new functions. The initial survey of all TMA features and functions lasted 2.5-3 hours per session, and subsequent assessments have lasted 45 minutes to an hour. The modular organization of the survey allows traffic managers to resume operational traffic management duties when necessary.

Shadowing Live Operations and Contextual Interviews. Shadowing involves a traffic manager using TMA to make traffic management decisions, mirroring the operational traffic management position. The shadowing traffic manager has access to all other sources of information in the unit except for the operational traffic management system. One observer observes and queries the shadowing traffic manager, and the traffic manager's ongoing commentary is tape recorded for later analysis. Another observer watches the operational traffic manager. Here, traffic management activities and decisions are observed in a more passive mode to avoid disrupting operations. Understanding and interpreting TMA use, at both the Center and TRACON, depend upon an understanding of the operational context. The second observer is critical in this regard.

Shadow-mode operations are effective for discovering unexpected and serendipitous tool-uses and for assessing issues of technical usability, domain suitability, and user acceptance. Methods for data collection are similar at the TRACON and Center but tailored for their unique constraints. Efforts are focused on capturing the traffic managers' ongoing experience with the system using context-based interviews (cf. Whiteside, Bennett, & Holtzblatt, 1988). This technique involves observing and questioning the users about the tool as they are using it for various planning and problem-solving activities. A critical aspect of contextual interviews is

involving the users in the interpretation of their experience with the system. This aspect is discussed further in the next section on data interpretation.

An important aspect of data collection in the field is the period of acclimatization that precedes actual data collection. Prior to conducting structured assessments, we spent several weeks in the traffic management units at the Center and TRACON, simply observing operations and answering questions on the purpose of our presence and the TMA assessment process. This acclimatization period allowed the traffic managers to become comfortable with us, making our observations less intrusive. It also allowed us to work out methodology issues, (e.g., optimum observation positions, and an effective observation checklist) and allowed us to gain a deeper understanding of traffic management operations.

Subjective Ratings. Subjective ratings of a system's usefulness provide another avenue for capturing the users' experience with an automation tool. Following a traffic rush period, traffic managers rate the usefulness of various TMA features, on a scale ranging from 1 to 5, for different traffic management tasks. Ratings capture the users' cumulative experience with the tool, in contrast to the momentary experience captured by a comment made during a specific activity. As a consequence, discrepancies between ratings and comments are possible and are a cue to dig deeper: while a feature may be useful in one situation it may be perceived to be insufficient in another. Such discrepancies underscore the importance of conducting observations at different times of the day, over several days, and preferably during different seasons to capture changes in traffic flow and weather disturbances. TMA has been at the field site for over a year and assessments to date have been conducted during the summer and winter months, each lasting about a month.

Data Interpretation

Data interpretation occurs on and off the field site. Observation alone is not sufficient for exploring and assessing tool use. The observer's interpretations of the observations must be shared with the user to verify their truthfulness (Whiteside, Bennett, & Holtzblatt, 1988). Mutual understanding of the traffic managers' experience with TMA is achieved during the traffic rush and immediately following the rush. The traffic managers are questioned in a debriefing interview on feature use for various problems, their experiences with TMA and their impressions of the traffic rush. In turn, the observers' interpretations of TMA use and the traffic managers responses to questions are also verified. Specific questions and observations, during and immediately following the traffic rush, are guided by a set of general questions:

- What is/was the traffic situation?
- What decisions and planning activities are occurring/occurred?
- What information is/was accessed from TMA and non-TMA sources?
- How is/was TMA used to support various traffic management decisions?
- What information is/was lacking or hindered decisions?
- What improvements are necessary?

These questions provide a framework for systematically exploring and understanding TMA use in the context of traffic management operations. They also provide a basis for deeper probing of technical usability, domain suitability, and user acceptance issues; for example: Is the number of steps for a particular feature excessive given operational time constraints? Is

sufficient information provided for determining whether the airport acceptance rate should be changed? Is the *right* information provided to support equitable decisions? All phases of the interview are tape recorded and conducted at the TMA, in the operational area, to provide a reference for discussing and interpreting the system. The merits of video, for this purpose, have been broadly extolled. Unfortunately, we were precluded from videotaping activities in the control room.

Something that proved helpful for data interpretation was for the observers to spend time each day, off the field-site, reviewing and discussing the observations. Dovetailing the different observational perspectives was useful for identifying knowledge gaps and for recalibrating the focus to further explore unexpected discoveries of tool use and possible emerging strategies. Any outstanding questions or new interpretations were taken up with the traffic managers the next day.

Analysis, Inferences and Implications

Surveys, observations, context-based interviews, and subjective ratings provide multiple windows on the traffic managers' experience with TMA. These methods and data provide a qualitative assessment of the match between TMA features and functions and the operational context for their use. The challenge lies in elucidating a tractable set of inferences from this large amount of data. To date, the focus of the TMA development and assessment process has been on identifying design deficiencies, discovering unexpected feature uses, understanding how the tool is used for various problem-solving activities, and defining operational requirements. Analyses have been geared accordingly. Frequency counts of negative responses on surveys provide insight into deficiencies and discrepancies. Content analyses of observations and interviews, coupled with subjective ratings, also provide insight into design deficiencies and discrepancies and enhance the understanding of tool use. Analyses, inferences and implications are described next in some detail as guidance for requirements definition evolves from these insights.

Identifying Design Deficiencies

Surveys. Scenario-driven surveys directly assess technical usability issues. Analysis is straightforward, focusing on negative responses to survey validation statements, which indicate difficulties in extracting, reading, discriminating, and accessing data from the TMA. Comments made by the users during the survey suggest resolutions to these deficiencies.

Survey data can provide diagnostic insight into users' possible difficulties when using the tool to make traffic management decisions. For example, a survey finding helped account for what appeared to be less efficient decision-making during shadow-mode exercises. The survey had revealed that a particular configuration caused crowding of data. Later, during the shadow mode exercises, this finding helped pin-point why particular traffic management decisions were being altered to what appeared to be less efficient decisions: Data congestion was causing traffic managers to overlook data, leading them to alter their decisions as the traffic situation progressed. This interpretation was confirmed by the traffic managers, and the problem was remedied in re-design.

Survey data provide only a partial window on system usability. Display clutter, color coding, and data entry may be assessed differently when the users are actively engaged, using

the tool to solve traffic management problems. For example, a particular feature that required manual setting received a positive response on the survey, but negative comments during actual use. During shadow-mode operations, the traffic managers were too busy with other traffic management activities to manually re-set a feature to reflect changes in the traffic flow. Too time constrained, they had to extrapolate the actual setting, making the feature cumbersome to use. While survey data provides useful information about system usability, this example illustrates the importance of accessing the users' experience with the system from different perspectives.

Observations and Interviews. Design deficiencies are also accessible from observations and interview data collected during shadow-mode exercises. Analysis of these data is time consuming, but the richness of the findings would not be available otherwise and outweigh the cost associated with the time spent. Observations and interview data from the two observers are merged into a single chronological description of each traffic rush. Such a description is useful for capturing the context of TMA use and provides a basis for various content analyses. Content analyses are performed in order to make qualitative inferences about TMA as a potential traffic management tool.

It is important to select categories for the content analysis that reflect the objectives of the assessment. To date, traffic managers' decisions and actions have been categorized according to design deficiencies, feature use for various traffic management activities, and unexpected discoveries. (For a concise description of content analysis, see Weber, 1990.) Data interpretation with the traffic managers during the interviews greatly facilitates the categorizing of observation and interview data. Some examples of the kinds of design deficiencies that can be inferred from content analysis of interviews and observations are presented next. Feature use and unexpected discoveries are described in the following section.

Technical usability deficiencies are defined as observed or reported difficulties in accessing, interacting with, or reading data. Examples of findings include ineffective coding of information for data search, the need for labeling to reflect operations, and too many steps required to implement various functions. In some instances, usability issues revealed here support findings from the surveys; in other instances new issues are raised.

Domain suitability problems are defined as occasions where the traffic manager needs certain information that is not available, and where extracting information interferes with or hinders problem-solving or decision-making. Examples of findings include the need for organization of information on panels to reflect operational constraints, the need for display parameter settings to reflect current airport configurations, and the need for representation of specific categories of information to reflect the characteristics of the traffic flow.

User acceptance problems are not as easily accessible or apparent as usability and suitability issues because they tend to be incidental consequences of the information technology (cf Hopkin, 1980; 1992). Understanding the operational context is thus essential for identifying user acceptance problems. System upgrades can affect job satisfaction and opportunities for recognizing professional merit by either affecting what was satisfying about the job in the current system or by causing new situations to emerge that disrupt job satisfaction. Findings from the TMA assessments have provided insight into both of these possibilities. Earlier observations, prior to data collection, had revealed that an important source of job satisfaction is in making decisions and plans that strike an equitable balance of restrictions across facilities and aircraft. Findings from the assessments to date have revealed that a key source of information for ensuring equitable decisions was quite difficult to extract from TMA. This difficulty reduced the use and preference for the representation, and pointed the way for system refinement. In contrast, TMA representations have also created new situations that appear to enhance job

satisfaction. One situation is the elimination of a time-consuming counting task, and another is the pulling together of previously disparate sources of information into a single representation. Comments from the traffic managers indicate that such features provide them more time for important planning and allow them to keep up with the dynamic traffic flow situation.

In addition to helping disclose design discrepancies, content analysis of observation and interview data provide insight into feature use. This insight is important for understanding the users' needs and the extent to which they are supported by the automation's capabilities. Feature use and discovery are discussed next.

Discoveries and Description of Feature Use

The introduction of advanced technology and innovative display and interactive features, like that embodied in TMA, alters the way work is done and how problems are tackled. Understanding these changes and how the features are incorporated into the flow of work is as important for assessing the match between the user's needs and system capabilities as is the identification of design deficiencies.

Patterns of feature use elucidated from the analysis of observations and interviews led to the discovery of two different strategies for making a particular traffic management decision. One strategy involved feature use that solved the problem in a similar way to current practices; namely by accessing information that managed traffic demand at the level of individual aircraft. The other strategy solved the problem in a new and different way, by relying on representations that provided information about the aggregate traffic demand. This second strategy was an unexpected discovery. Decisions made with both strategies were equally efficient, relative to those made by the operational traffic manager using the current traffic management system. At one level, the finding of different strategies of feature use suggests that the TMA representations are flexible enough to support different traffic management styles and preferences. At another level, the finding has broader implications for operational requirements because the information needed to support different problem-solving strategies must be identified.

One of the biggest changes to traffic management as a consequence of TMA involves the level of coordination between the Center and TRACON. With TMA, both facilities now have access to the same information about the traffic demand. Observations and interviews with the traffic managers indicate that this has enabled the TRACON to coordinate proactively with the center on decisions regarding the distribution of the traffic flow. Such coordination between the two facilities is essential to avoid overloading the TRACON and to maximize airspace capacity. At a more subtle level, TMA elevates the role of the TRACON traffic managers, allowing them to be more active players in traffic management. This elevated status has obvious implications for job satisfaction.

Another change to current practices is the impact of TMA on the exchange of information between facilities. Analysis of the traffic management communications between the center and TRACON indicates that well over 50 percent of the transmissions between the facilities involves the transmission of information that is accessible from TMA. This finding suggests that certain verbal transmissions between facilities could be eliminated, augmented, or reduced by electronic sharing of information via TMA. Changes in the level of coordination and exchange of information between the Center and TRACON alter the way traffic management is performed and has implications for operational procedures and requirements.

Requirements Definition and Rejuvenation

Requirements are the services to be provided by the system and the constraints under which it must operate. Complex domains, like ATC, with their myriad interdependencies and interconnections, are difficult to understand and thus a complete set of requirements is not likely to be available prior to development. Instead, definition of requirements is likely to evolve with development, and modifications and refinements will be necessary as an understanding of the user's needs improves. Field testing conducted early in development catalyzes the requirement definition process. Identifying mismatches between the user's needs and tool features is important for refining the design and for exposing system constraints that must be captured in the requirements. This dual purpose of design deficiencies is illustrated well by the following example. A particular TMA feature had been designed to require several steps to access information. When exercised during shadow-mode operations, the feature was deemed unsuitable because immediate access to the information was needed. This not only identified a design deficiency but also exposed a system constraint that must be captured in the system requirements: immediate access to information for a particular traffic management decision.

Feature use in context is also instructive for defining and modifying system requirements. New technology can create new cognitive problems and information requirements for these problems must be defined. Similarly, capabilities may emerge when the system is used in the operational environment that were not anticipated in the conceptual design; for example, the level of coordination and information exchange between the TRACON and Center with TMA. System constraints for such capabilities must be included in the system requirements.

It has been argued that the implementation of a system and its specification should not be kept separate because, in practice, they are "inevitably intertwined." Models of system development that require their separation deviate from reality and restrict the development of effective systems (Swartout & Blazer, 1982). A similar argument can be applied to requirements definition. System implementation through field testing, early during system development, can facilitate the evolution of system requirements for complex domains like ATC. As described in this paper, field assessments help highlight system constraints, enhance our understanding of the user's needs, and provide insight into the impact of new technology on existing operational practices. While systematic analyses, feasibility studies, and system modeling are necessary precursors for requirements definition, field development and assessment in the field can help augment the process.

Summary

The process for incorporating advanced technologies into complex aviation systems is as important as the final product itself. This paper described a process that is currently being applied to the development and assessment of an advanced ATC automation system, CTAS. The key element of the process is field exposure early in the system development cycle. The process deviates from current established practices of system development – where field testing is an implementation endpoint – and has been deemed necessary by the FAA for streamlining development and bringing system functions to a level of stability and usefulness. Methods and

approaches for field assessment are borrowed from human factors engineering, cognitive engineering, and usability engineering and are tailored for the constraints of an operational ATC environment. To date, the focus has been on the qualitative assessment of the match between TMA capabilities and the context for their use. Capturing the users' experience with the automation tool and understanding tool use in the context of the operational environment is important, not only for developing a tool that is an effective problem-solving instrument but also for defining meaningful operational requirements. Such requirements form the basis for certifying the safety and efficiency of the system. CTAS is the first U.S. advanced ATC automation system of its scope and complexity to undergo this field development and assessment process. With the rapid advances in aviation technologies and our limited understanding of their impact on system performance, it is time we opened our eyes to new possibilities for developing, validating, and ultimately certifying complex aviation systems.

Acknowledgments

The authors are grateful for the support of Barry Scott, Manager FAA Engineering Field Office at NASA-Ames Research Center and David Jones, Branch Chief ATC Field Systems Office at NASA-Ames Research Center. C. Halverson, E. Murphy, S. Nowlin, and P. Stager are acknowledged for their insightful comments on earlier versions of this paper.

References

- ATC Systems Field Office. (1992). *Traffic management advisor (TMA) reference manual*. Moffett Field, CA: NASA-Ames Research Center.
- Bentley, R., Hughes, J. A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., & Sommerville, I. (1992). Ethnographically-informed systems design for air traffic control. *CSCW Proceedings*. (pp. 123-129).
- Christensen, J. M. (1958). Trends in human factors. *Human Factors*, 1 (1), 2-7.
- Davis, T. J., Erzberger, H., & Green, S. M. (1991). *Design and evaluation of air traffic control final approach spacing tool* (NASA Technical Memorandum 10287). Moffett Field, CA: NASA-Ames Research Center.
- Department of Defense. (1989). *Human engineering design criteria for military systems, equipment, and facilities* (MIL-STD-1472D). Washington, DC: Author.
- Erzberger, H. (1992). *CTAS: Computer intelligence for air traffic control in the terminal area* (NASA Technical Memorandum 103959). Moffett Field, CA: NASA Ames Research Center.
- Erzberger, H., & Nedell, W. (1989). *Design of automated systems for management of arrival traffic* (NASA Technical Memorandum 102201). Moffett Field, CA: NASA-Ames Research Center.
- Federal Aviation Administration. (1989). *FAA NAS test and evaluation program* (Order No. 1810.4A). Washington, DC: Author.

- Flexner, S. B., & Hauck, L. C. (1983). *The Random House dictionary of the English language*. Random House: New York.
- Gould, J. D. (1988). How to design usable systems. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 757-789). New York: Elsevier Science Publishers BV (North Holland).
- Harwood, K. (1993). Defining human-centered system issues for verifying and validating air traffic control systems. In J. A. Wise, V. D. Hopkin, & P. Stager (Eds.), *Verification and validation of complex systems: Human factors issues* (pp. 115-130). Berlin: Springer-Verlag.
- Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Systems*, 18, 583-600.
- Hopkin, V. D. (1980). The measurement of the air traffic controller. *Human Factors*, 22 (5), 547-560.
- Hopkin, V. D. (1992). Human factors issues in air traffic control. *Human Factors Society Bulletin*, 35 (6). Santa Monica, CA: Human Factors Society.
- Hughes, J. A., Randall, D., & Shapiro, D. (1992). Faltering from ethnography to design. *CSCW Proceedings* (pp. 115-122).
- Hutchins, E. (1991). *How a cockpit remembers its speed*. (Technical Report). San Diego: University of California, Distributed Cognition Laboratory.
- Johnson, & Baker. (1974). Field testing: The delicate compromise. *Human Factors*, 16 (3), 203-214.
- Meister, D. (1985). *Behavioral analysis and measurement methods*. New York: John Wiley & Sons
- Parsons, M. (1972). *Man machine system experiments*. Baltimore, MD: The Johns Hopkins Press.
- Rasmussen, J. (1985). The role of hierarchical knowledge representation in decision making and system management. *IEEE Transaction on Systems, Man, and Cybernetics*, 15, 234-243.
- Rasmussen, J. (1986). *Information processing and human-machine interaction: An approach to cognitive engineering*. Amsterdam: North-Holland.
- Rasmussen, J., & Goodstein, L. P. (1988). Information technology and work. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 175-201). New York: Elsevier Science Publishers BV (North Holland).
- Shneiderman, B. (1987). *Designing for the user interface*. Reading, MA: Addison-Wesley.
- Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software*. Bedford, MA: Hanscom Air Force Base, USAF Electronic Systems Division. (NTIS No. A177 198)
- Suchman, L. A. (1987). *Plans, and situated actions: The problem of human machine communication*. Cambridge, MA: Cambridge University Press.
- Suchman, L. A., & Trigg, R. H. (1991). Understanding practice: Video as a medium in reflection and design. In J. Greenbaum & M. Kyng (Eds.), *Design at Work* (pp. 65-89). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tobias, L., Volckers, U., & Erzberger, H. (1989). Controller evaluations of the descent advisor automation aid. In *Proceedings of the AIAA Guidance, Navigation, and Control Conference* (pp. 1609-1618). Washington, DC: AIAA.
- Van Cott, H. P., & Kincaid, R. G. (1972). *Human engineering guide to equipment design*. Washington, DC: U.S. Government Printing Office.

- Whiteside, J., Bennett, J., & Holtzblatt, K. (1988). Usability engineering: Our experience and evolution. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 791-817). New York: Elsevier Science Publishers BV (North Holland).
- Weber, R. P. (1990). *Basic content analysis*. Newbury Park, CA: Sage Publications.
- Woods, D. D., Roth, E. M., & Bennett, K. B. (1990). Explorations in joint human-machine cognitive systems. In S. P. Robertson, W. Zachary, & J. B. Black (Eds.), *Cognition, computing, and cooperation* (pp. 123-158). Norwood, NJ: Ablex Publishing Company.