Human-Systems Integration and Air Traffic Control

Hayley J. Davison Reynolds, Kiran Lokhande, Maria Kuffner, and Sarah Yenson

Data-driven prototype evaluations and benefits analyses that iteratively feed back into the prototype design and development produce a user-friendly system that yields operational benefit. Initial requirements development and field evaluations performed at the Dallas/Fort Worth International Airport Air Traffic Control Tower helped in the design process of the air traffic control Tower Flight Data Manager prototype.

» **Development of air traffic control decision** support tools necessarily requires an iterative design process to ensure that the tool is effectively integrated into the complex and demanding operational environment. Without an iterative design process that actively incorporates user input, decision support tools fielded in the air traffic control sector could result in controller resistance to the technology and/or budget overruns caused by the need to fix problems identified by the controllers post hoc.

While the usability and acceptability of a system are necessary to achieving a successful system, by themselves, these attributes are not sufficient. A highly usable and acceptable system may be unsuccessful at achieving any operational benefit (e.g., improved efficiency, improved safety). Thus, the design process should strive to interactively achieve the operational benefits expected of the system. Lincoln Laboratory used a benefits-driven, iterative, system design process for the Tower Flight Data Manager (TFDM), a prototype of an integrated decision support system for the air traffic control tower (ATCT) environment.

Introduction

TFDM is one of the key systems in a major overhaul of the information systems across the air traffic control domain. The En Route Automation Modernization (ERAM) is the updated information system for the en route (higher altitude) ATC environment being deployed across the United States since 2009 [1]. The Terminal Automation Modernization and Replacement (TAMR), the update for the Terminal Radar Approach Control (TRACON) (arrivals and departures) ATC environment currently in development, began to be deployed in 2006. TFDM is ERAM and TAMR's counterpart in the ATCT environment. The first phase of TFDM, which will integrate multiple individual, stove-piped information sources, is expected to be field deployed in 2017. TFDM will have sustaining influence on the future of the U.S. ATC system because airports are considered the limiting factor in the current National Airspace System capacity.

A user-friendly system that yields operational benefit results from data-driven prototype evaluations and benefits analyses that iteratively feed back into the prototype design and development. In this article, initial requirements development and field evaluations at the Dallas/Fort Worth International Airport ATCT are discussed in reference to the design process of the air traffic control TFDM prototype. Nonintrusive measures used for quantitatively validating human-systems design issues included visual-gaze analysis and verbal-command sequence analysis. Behavioral validation of design issues simplifies the process to prioritize beneficial design changes. While only tested at a single ATCT, the iterative process resulted in an interim TFDM prototype that was rated by a set of 12 active air traffic controllers as both acceptable and usable in a nonoperational field demonstration in an actual ATC tower environment.

Decision Support Development Process

A benefits-driven system design process, with the human user as the key component of that system, is suggested as a means to design and evaluate a system's effect on the operational environment. As shown in Figure 1, a measurement process is first used to develop an understanding of the system's effectiveness—what aspects are successful and what are problematic? This measurement process uses both quantitative and qualitative data to define the system state. Operational data, such as system-delay statistics or environment-impact assessments, are collected, and analyses are performed with respect to hypotheses made about problems with the current operation. Concurrently, field visits allow developers to understand the operation itself, including procedures followed and system constraints. Hypotheses about the potential problems with the current system can be collected as decision support suggestions as well. These hypotheses can then be evaluated in operational data analysis.

The data-confirmed operational problems, procedures, and constraints then form the designer's "operational model" of the system. As more data are collected, from both field observations and analyses, this model is iterated upon to reflect the designer's current understanding of the system. From this operational model, a set of decision support requirements is then developed, reflecting the particular problem(s) identified in the system. A means of providing the decision support is then designed, resulting in a new human-machine interface (HMI), training regimen, and/or operational procedure. The effective implementation of the means of decision support (tool, training, procedure) should then affect the key variables of the operational system and produce a measurable operational benefit.

A critical aspect of the design is the iterative nature of this process. It is likely on the first pass of the design that the operational benefits achieved are less than satisfactory. Thus, field observations of the tool in use should illuminate possible requirements or design specification issues that can be corrected in the next iteration.

This process has been used successfully in the design of the air traffic management Route Availability Planning Tool (RAPT) that is used by ATC facilities during convective weather (thunderstorms) [2–4]. In RAPT, a simple red/yellow/green indication provides decision support to New York area air traffic managers about whether a departure route out of an airport will be impacted by weather. Despite the simple concept, years of iteration occurred to smooth out algorithmic issues, such as when and how fast colors would flicker back and forth and where the red versus yellow and yellow versus green thresholds should be drawn to yield the most effective recommendation. By collecting operational data, researchers discovered in 2008 that the main decision for the effective use of departure routes was not the determination of when to close routes, but the determination of how soon to open the departure routes after a storm had passed through. By providing a simple alteration to the RAPT HMI coupled with training, departure delay savings attributable to RAPT more than doubled in 2009.

While the RAPT prototype was deployed and iterated upon in the field (in which actual operational benefits can be assessed), other studies analyzed advances in automation through measuring operational performance metrics in a human-in-the-loop experimental environment [5–7].

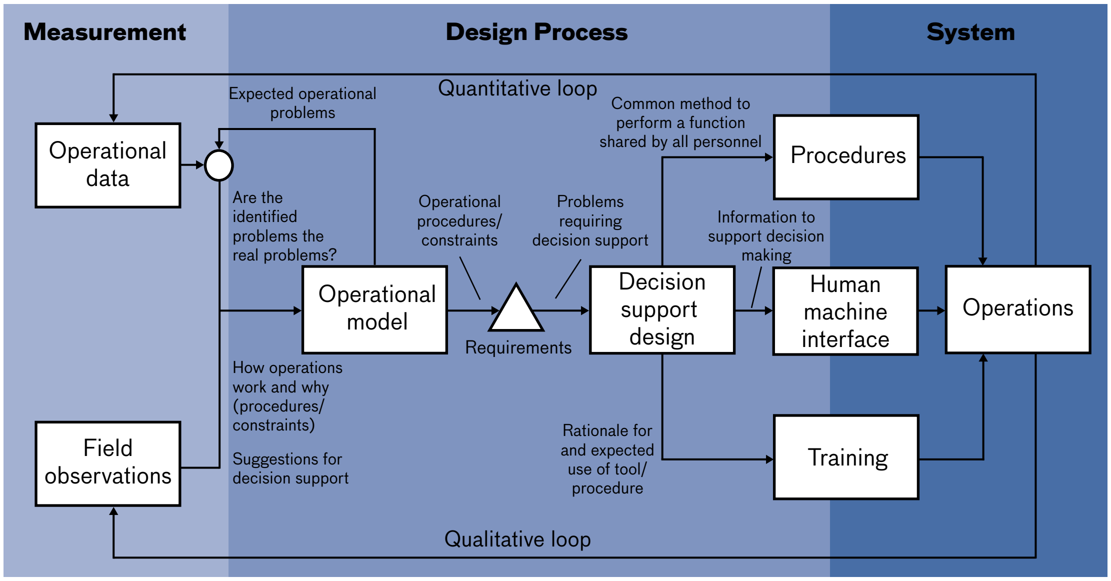


FIGURE 1. A benefits-driven, decision support design process provides a means of both qualitatively and quantitatively iterating upon a decision support system (including the human-machine interface, operational procedures, and operational training) to achieve quantifiable operational benefit.

In the RAPT program, the designers were incredibly fortunate to have had the opportunity to prototype the system in the field and to collect operational data along the way to evaluate the effectiveness of the system. A similar benefits-driven, decision support design process was followed in the TFDM prototype program. However, because of the safety-critical nature of the system, the prototype was fielded in a shadow operations environment.

Current Air Traffic Control Tower Environment

In today's ATCT environment, usually at least four operations are active in the larger tower facilities—flight data, ground control, local control, and supervisor/traffic management coordinator (TMC), hereafter called simply supervisor.

Flight-data staff ensure that each departing flight has an ATC-acceptable route. This route is shown on small paper flight strips—one for each flight—that can be written upon and resequenced by the controller. Once the flight has received a route clearance by flight data, the flight strip is then physically passed to the ground-control personnel.

Ground-control personnel issue taxi clearances to the aircraft once it pushes back from the airport gate. Ground control can view the aircraft out the ATCT window and can often see the aircraft's icon on the Airport Surface Detection Equipment (ASDE-X), a surveillance display of the airport and its surrounding airspace (approximately 5 nautical miles). The departing flights are then sequenced for departure by ground control to maximize departure efficiency while adhering to constraints (such as wake-vortex separation or traffic flow separation) affecting individual flights. Once the flights have been sequenced and have entered the departure queue, ground control passes control and the flight strip to the local-control sector.

Local control-sector then makes final adjustments to the departure sequence and clears the flights for takeoff. Local control also manages arriving flights and flights that are intending to cross active runways. The primary information source to local control is an out-the-window view of the arriving and departing flights, but the local-control staff are also supported by the information on the flight strip, the Remote Automated Radar Terminal System Control Tower Displays (RACD) (surveillance information extending approximately 20 nautical miles from the airport), and the ASDE-X display.

Overseeing the ATCT operation is the **supervisor/TMC,** who is responsible for ensuring that the airport is in the optimal configuration (direction of runway landing/departing) for the prevailing conditions (e.g., traffic flow, number of aircraft, weather conditions). The supervisor also ensures that traffic flow management restrictions are coordinated with other ATC facilities and new restrictions placed on departing aircraft are passed to ground and local control for implementation. These restrictions are

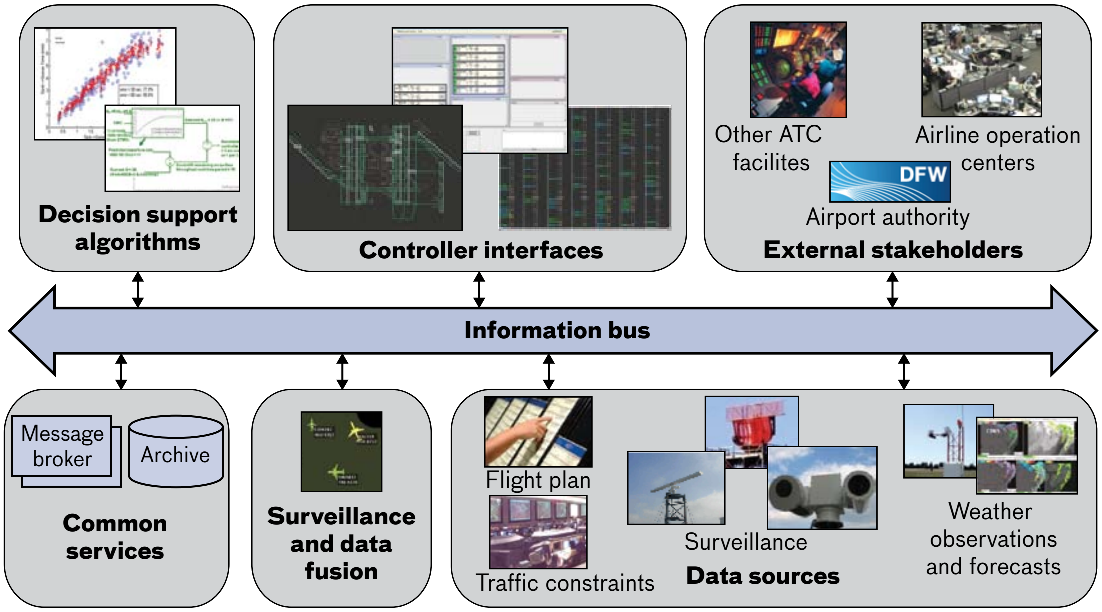


FIGURE 2. The Tower Flight Data Manager (TFDM) exhibits a net-centric architecture in which sources such as decision support algorithms publish data to an information bus, while other entities such as controller interfaces subscribe to these data to display to the users.

typed into the Information Display System (IDS) for all of the ATCT positions to see. The flight data, ground control, local control, and TMC are Certified Professional Controller (CPC) positions, while the supervisor is a front-line manager position.

Tower Flight Data Manager Prototype

The TFDM prototype is an iteration of the future ATCT surveillance, flight data, and traffic management information system. The purpose of TFDM is to integrate stove-piped information sources within the tower environment so that the integrated information is calculated to provide additional decision support capabilities for increasing departure efficiency and reducing fuel burn and emissions. Lincoln Laboratory was tasked to build a prototype of TFDM for proof-of-concept purposes and to aid in generating system design requirements for the production TFDM system. The TFDM prototype also supports the needs of the Staffed NextGen Tower program, which is an activity to explore the possibility of evolving toward a tower environment that is not reliant on out-the-window information. The TFDM prototype system, shown in Figure 2, has a net-centric information architecture, including an information bus that shares information between sensors, such as ASDE-X surveillance data, HMIs, and decision support tools.

Three human-machine interfaces were developed for the prototype: a surveillance display, a flight data manager, and a supervisor display. The Tower Information Display System (TIDS), shown in Figure 3, is an enhanced surveillance display, built upon the display of the existing tower surveillance system, ASDE-X. Controllers are able to zoom, scroll, and create individualized picture-in-picture windows on the TIDS to match their operations. The following are some specific features of the display:

- Data block colors that indicate whether a flight is airborne (cyan) or on the surface (white).
- Proportionately sized aircraft icons and aircraft icon colors to indicate the type of aircraft and whether it belongs to a small, large, or heavy weight class.
- Wake-vortex timers in orange at the end of the departure runway that indicate the time since a heavy aircraft has taken off.
- Approach bars in green at the end of the arrival runway to indicate the location of an arriving aircraft on the approach path before its icon is visible on the display.
- A video picture-in-picture window to display a camera surveillance picture of the airport surface. The video can also be set to track arriving or departing flights automatically for Staffed NextGen Tower program purposes.

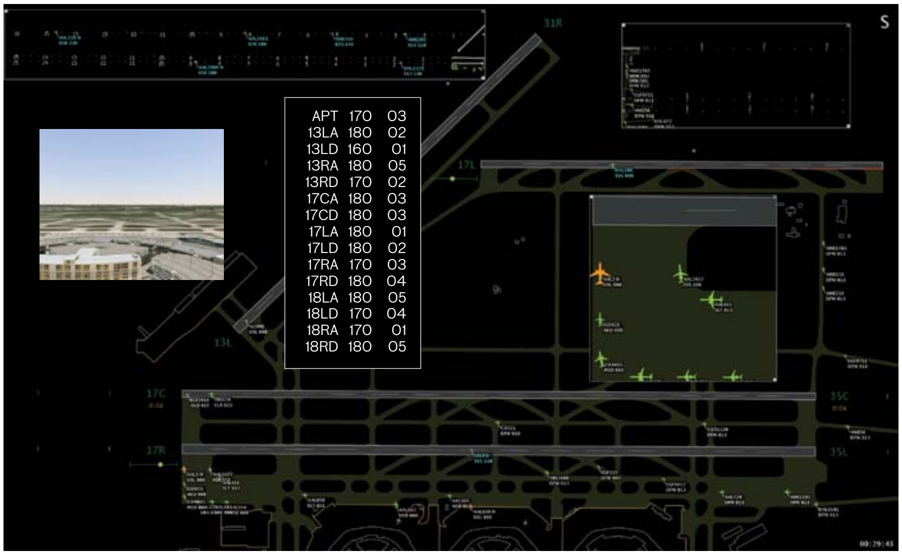


FIGURE 3. The Tower Information Display System (TIDS) is an advanced surveillance display that contains surveillance processing enhancements over the current ASDE-X surveillance display as well as interface improvements.

In addition to interface enhancements, surveillance processing improvements were made to correct surveillance position report jitter, which caused spinning stationary aircraft icons, phantom icon movement, and misleading lateral deviation of icons from the runway center line, and to stitch missing/split track reports into a cohesive, single surveillance track for an aircraft.

The Flight Data Manager (FDM), shown in Figure 4, is an electronic flight-data display that allows personnel to interact with flight-data entries via touchscreen display as well as mouse and keyboard. Surveillance and flight data are linked in TFDM, allowing a selection on the TIDS display to highlight its corresponding flight strip on the FDM and vice versa. The flight-data-entry (FDE)

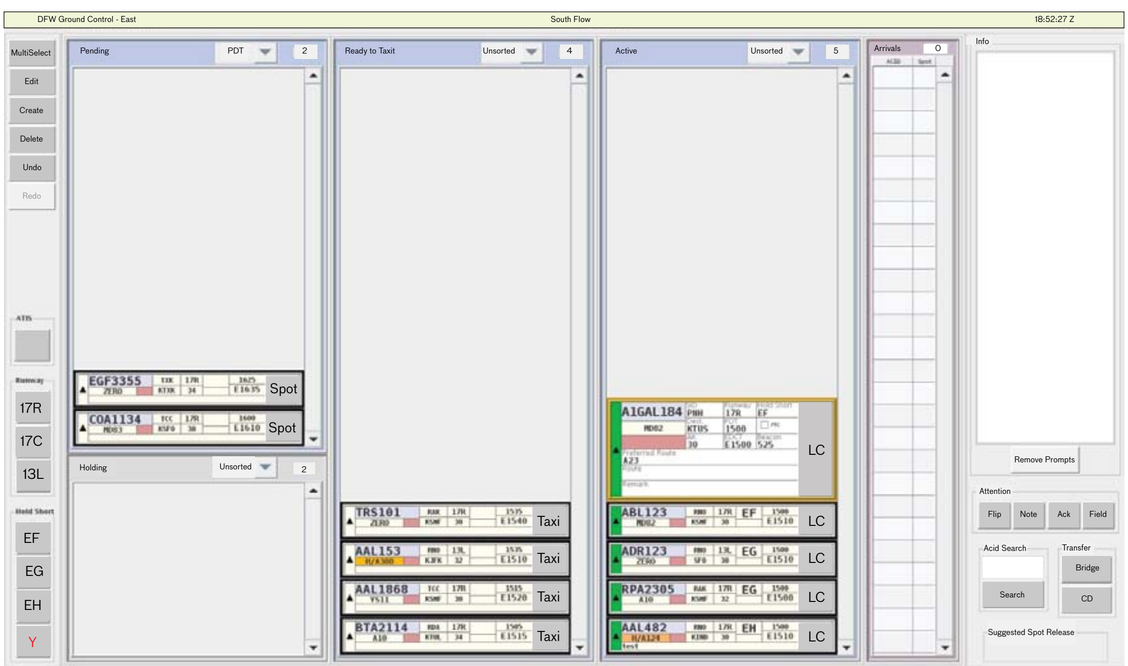


FIGURE 4. The Flight Data Manager (FDM) is a configurable, touchscreen, electronic flight-strip display that is interconnected to surveillance information to enable automatic flight-strip movement.

information layout is similar to a paper flight strip, with some minor changes. The FDE has a flight-status icon that indicates whether a flight is a departure or arrival and whether it is active (flight has pushed back from gate) or not. The FDE can be expanded to display more information about the flight, including the entirety of its filed route. Individual FDEs can be resequenced, edited, and highlighted, similar to what controllers do with their paper flight-progress strips. Hot buttons allow controllers to perform the most frequent actions on an electronic flight strip (e.g., departure runway assignment) in a simple, quick way. An information window at the right of the display provides a means to communicate alarms, alerts, and decision support prompts to the controllers. Information from surveillance allows movement of FDEs to automatically progress to queues based on aircraft position on the surface (e.g., from the leftmost "Pending" queue to the center "Ready to Taxi" queue when a flight is detected at the "Ready to Taxi" spots on the airport surface).

The supervisor display, shown in Figure 5, is a tab-based display that enables the tower supervisor to view and manage tower operations in aggregate. The supervisor is able to view expected airport demand, weather impacts on departure routes/individual departures, traffic management restrictions, and individual flight information that is also displayed on the FDMs. On the supervisor display, the supervisors/TMCs can change or schedule a change in airport configuration, add or modify traffic management restrictions, and view the effect of these changes upon the arrival/departure demand. In the screen shot in Figure 5, each timeline represents a runway at Dallas/Fort Worth International Airport (DFW), with the expected arrival times of flights in blue and expected departure times of flights in green. At the top right of the display, the current airport configuration ("Southflow") is indicated. The left panel provides the ability to change this configuration, add traffic management restrictions, and perform other actions. The tabs above the panel with the runway timelines let supervisors switch to view other information about the ATCT operation, such as the departures by assigned departure routes and any departure route's weather status.

Besides the HMIs, the TFDM prototype also includes a variety of decision support tools (DST) that establish runway assignments; departure metering, sequencing, and scheduling; airport configuration; and departure routing. The airport configuration DST provides the ability to alter or schedule a change in airport configuration. The runway assignment DST automatically assigns departures and arrivals on the basis of a basic rule set. A feature in the supervisor display allows this rule set to be altered. The departure metering DST suggests to the controller the rate at which departures should be released onto the taxiway to maintain pressure on the departure queue while minimiz-

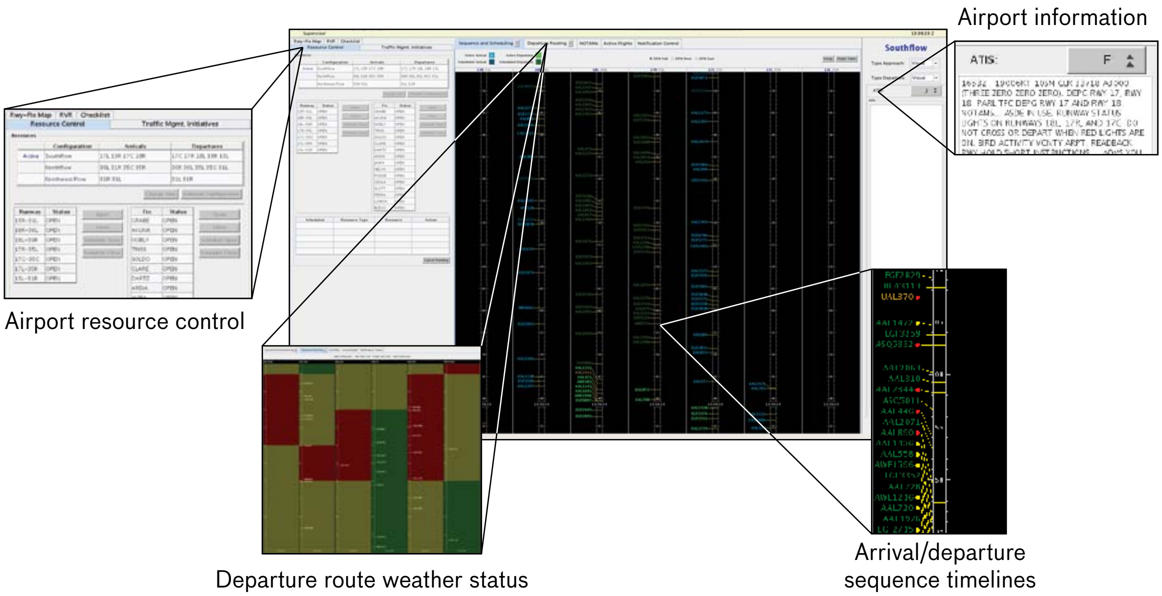


FIGURE 5. **The unique supervisor display designed for TFDM allows supervisors to monitor and adjust the tower's decision support tools to the needs of the operation.**

ing surface congestion. The sequencing and scheduling DST provides an estimated wheels-off time for departures and an estimated wheels-on time for arrivals that can be viewed on a timeline on the supervisor display. The departure routing DST estimates and displays departure-route weather blockage and overlays the estimated wheels-off times on timelines to allow the supervisor to see which departures are predicted to be affected by convective weather.

Initial TFDM Requirements Development

An initial set of system design requirements for the TFDM prototype system drew from three primary sources: TFDM operational benefits research, ATCT site visits, and input from a national TFDM user group. The TFDM benefits research revealed that 39% of surface inefficiencies were due to inefficient timing of pushback of the aircraft from the gate. Secondarily, 33% of the inefficiencies were due to inefficient sequencing of flights departing and 12% were due to inefficient airport resource planning (e.g., changing airport configuration too late to allow taxi planning) [8–9]. By using operational data (e.g., ASDE-X, FAA terminal area forecasts, and other sources), the benefits team was able to identify the key decision support requirements for which the TFDM prototype should be designed. On the basis of this research, the DSTs projected to provide the most benefit to surface inefficiency included departure metering, sequencing, and scheduling, and airport configuration capabilities.

In addition to investigating potential operational benefits, researchers conducted ATCT field site visits to better understand and model the current tower operations. Towers with distinctive operations were visited, including Boston's Logan, Chicago's O'Hare, Dallas/Fort Worth (DFW), New York's John F. Kennedy, and Newark's Liberty airports. For each of the towers visited, information was collected to form the operations model of the ATCT operation, including process models, cognitive work identification, information systems analysis, communications analysis, and procedures aggregation. By using these overlapping, but different, methods of work discovery, a reasonable operational model could be developed of the tower environment and its individual operations.

A national TFDM user group provided another source of initial requirements information for the program. Every two to three months, seven ATCT supervisors, one airline ramp supervisor, one Air Route Traffic Control Center (ARTCC) supervisor, one ATC terminal supervisor, and one National Air Traffic Controllers Association representative would provide input on the TFDM program. The main focus for the user group was a series of discussions on how potential future capabilities of TFDM could or could not be effectively implemented operationally.

From the operational model developed from the benefits analysis, field visits, and user group meetings, an initial set of TFDM functional requirements was developed. These requirements, as well as some TFDM design specifications suggested by the user group or FAA procedures, were used as a basis for the design of the initial TFDM prototype. This prototype was then evaluated and iterated in the field during two successive field demonstrations at the Dallas/Fort Worth International Airport.

Field Demonstrations and Evaluations at the Dallas/Fort Worth International Airport

Two field demonstrations at DFW were conducted to prove the concept of the TFDM prototype as well as to iterate upon the initial requirements of the TFDM system. The initial demonstration (DFW-1) in August 2010 tested the TFDM net-centric infrastructure, the TIDS, the FDM, and the runway assignment DST. In the second demonstration (DFW-2), in April/May 2011, the supervisor display and the initial iteration of the other four DSTs were introduced.

Both field demonstrations were conducted in the DFW Center Tower, an operational but not regularly used ATCT that serves as a backup for the East and West DFW Towers. In DFW-1, there were two test operations—ground control and local control. Each had a TIDS and an FDM HMI. In DFW-2, the demonstration added the evaluation of the supervisor tools, including TIDS and a supervisor display. In both demonstrations, a flight-data/clearance-delivery position was manned by the research personnel, and this position had an FDM that could modify and transfer control of electronic flight strips to ground- or local-control positions to maintain operational validity of the demonstration.

Both demonstrations were conducted as shadow operations of the DFW's East Tower operations. In a shadow operation, test protocols requested that the participants issue verbal commands and update information systems as if they were actually controlling air traffic, but the actual clearances were issued by, and separation responsibility remained in, the East Tower. Participants heard both the pilots' and East Tower controllers' communication, but neither the pilots nor East Tower controllers could hear the Center Tower participants.

Both field demonstrations were two weeks long, with three days of demonstration each week. Each day during DFW-1, a different pair of DFW CPCs switched between ground and local control throughout the day, resulting in a total of 12 controllers evaluating the system.

DFW-1 Method and Results

A significant amount of surveillance data were collected from this field demonstration. A series of flight tests with a test aircraft equipped with a differential global positioning system was conducted to assess the validity of the surveillance data presented on the TIDS display and to establish the benefits of the improved surveillance processing. By combining the airfield surveillance radar ASR-9 data with the ASDE-X data, surveillance coverage was increased to an approximately 20-nautical-mile radius of the airport. This coverage allowed flights on approach and departure to be surveilled completely without track loss on extended approaches or departures. In addition, because of the advanced surveillance processing, continuity of the flight track on the surface was maintained. Such continuity is critical to maintaining the surveillance/flight data linkage throughout the flight's time at the airport. The processing also ensured that the heading of the aircraft icon was correct on the TIDS display and remained near the centerline of the taxiways.

In DFW-1, the primary means of gathering human-systems integration data were field observations, questionnaires, and post-demonstration discussions. For each participant during the demonstration, two research personnel were assigned to answer the participant's questions during the day and take observational notes of the use of TFDM. Structured forms were used to organize and direct the observations (e.g., interface problem, performance issue, functionality liked, functionality suggestion, adaptation observed, workload, out-the-window, and benefit observed). In DFW-1, another observation form was populated to determine how the participant interacted with the system (e.g., via mouse, touchscreen, virtual keyboard, actual keyboard). These observations would aid in identifying whether designing the system for touchscreen was a fruitful endeavor. The questionnaires asked controllers to rate on a scale of 1 to 5 the accuracy, acceptability, usefulness, and usability of different aspects of the TIDS and FDM, as well as the interaction between TIDS and FDM. Open-ended questions were also included to allow controllers to document their reactions and suggestions in their own words.

A significant amount of technical data as well as human-factors data were collected from this field demonstration. In this article, only select cognitive engineering results impacting the future design are discussed.

Overall, the TFDM prototype was well-received in the DFW-1 demonstration. Over 100 individual questions were answered by the participants. Figure 6 and Figure 7 provide an overall indication of how the participants viewed the human-machine interfaces for the system. A majority of the responses were positive regarding the interfaces' role in performance of critical ATC tasks, ease of use, and benefit afforded to tower controllers.

Hundreds of observations and suggestions were documented from DFW-1. The problems experienced by at least 30% of the participants (four CPCs) centered around either TFDM's misinterpretation of current procedure (e.g., the wrong wake-vortex timer setting) or not accounting for critical key functionality of the current operation (e.g., providing the ability to "cock" or offset electronic flight strips in the strip bays). Suggestions primarily involved providing a means to invoke functionality of the systems today (e.g., "cocking," highlighting, showing beacon code) or better utilizing the electronic information functionality (e.g., suggesting hot buttons, expanding touchscreen to TIDS).

The use of the FDM touchscreen was also evaluated. Table 1 describes the results of analyzing 357 sample interactions noted for the six CPCs. Overall, the CPCs made considerable use of the touchscreen, and this observation supported the decision to continue further iterations of the TFDM prototype with a touchscreen FDM. In addition, local-control participants and then ground control used the touchscreen more frequently. This difference in usage is likely the result of the following: (1) there appeared to be less need to interact with the data on the electronic flight strips and their sequence at the local-control position, and (2) because of the speeds of imminently arriving and departing aircraft, there appeared to be more of a need at the local-control position to be "heads up" to constantly monitor the progress of these aircraft.

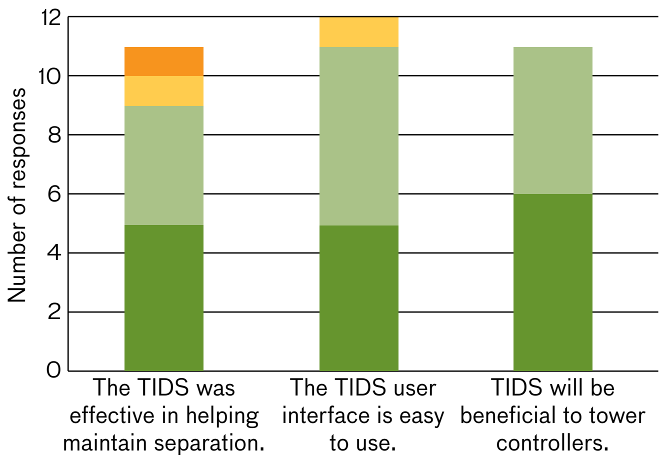


FIGURE 6. The TIDS questionnaire responses indicated a positive reception of the display to the DFW-1 controller participants.

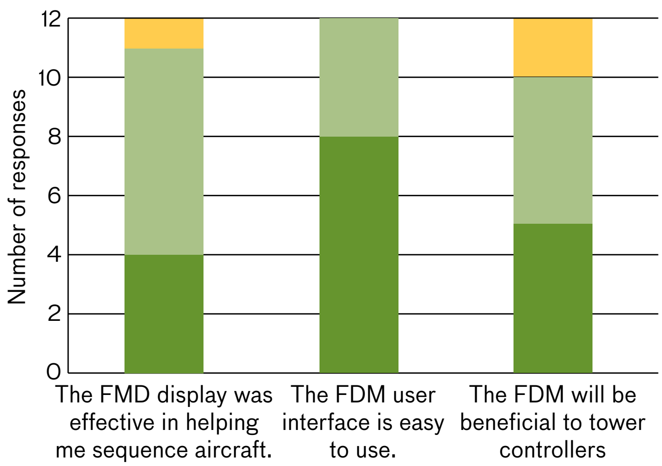


FIGURE 7. The FDM questionnaire responses indicated a positive reception of the display to the DFW-1 controller participants.

When touchscreen usage was compared with mouse usage during the DFW-1 demonstration, results indicated that touchscreen was a design specification that should remain in place.

On the basis of the DFW-1 demonstration results, a significant number of improvements were made to the TFDM prototype before the DFW-2 demonstration. All of the chief problems and suggestions were addressed. The wake-vortex timer was modified to be triggered by Boeing 757 aircraft. Means of drawing attention to both the electronic flight strip as a whole and the individual information fields were added, including the ability to "flip" an electronic flight strip (i.e., remove all information from the electronic flight strip except for its aircraft identification, which is an electronic analogy to flipping a physical flight strip over); to change the flight-status icon to an "!" as a substitute for cocking; to highlight an individual field with a yellow background; and to change text in an individual field to red. The requested hot buttons were added to the FDMs, and the beacon code was added to the expanded electronic flight strip.

Table 1: A comparison of touchscreen versus mouse use as a percentage of interactions

	Touchscreen used	Mouse used
Total	72.4%	27.6%
Total of CPCs at ground control	61.8%	38.2%
Total of CPCs at local control	86.6%	13.4%

DFW-2 Method and Results

The data-collection methods used in DFW-1 and DFW-2 varied. The methods in DFW-1 provided initial guidance on problem areas in the TFDM design that could be addressed before DFW-2. In DFW-2, steps toward understanding the TFDM system's effect on a controller's behavioral performance were taken. In addition to field observations, questionnaires, and post-demonstration discussions, a video and audio playback system was created that allowed performance analyses to occur post-demonstration. In the interest of brevity, only the questionnaire review of the new electronic flight-strip features and the video and audio playback analysis will be discussed in detail.

The means to draw attention to the electronic flight strip and its fields, which were integrated into the prototype for the DFW-2 demonstration, were received well by the participant controllers, as seen in Figure 8. Both the field-highlighting feature and the application of red text to a field in the electronic flight strip were considered useful by the participant controllers. From the participant feedback, it appears that the interface used to highlight a field and to change a field's text color requires further consideration. This feedback was also discovered in the observations that these features required too many FDM inputs to achieve the desired state. In the revisions made after the DFW-2 demonstration, the user interface that highlights fields and changes field text color was revised, and the requirement was modified to reduce the number of inputs required to apply these changes.

The post-demonstration analyses were conducted with an audio and visual playback system. Video was chosen as the most noninvasive and inconspicuous option for capturing the participants' actions during the field demonstration. Real-time screen recordings of the TIDS, FDM, supervisor display, and Cohu camera out-the-window display screens were also captured by using a tool called Epiphan. The screen shots and video captured exactly what the participants were looking at and doing on screen, and could be replayed. Audio recordings of participant controllers issuing shadow verbal commands were captured through the use of small microphones worn around their necks, and audio recordings of East Tower controller and pilot communication frequencies were provided by the DFW Tower. During playback analysis, all of these data were gathered into and synchronized together with Adobe Premiere, so that a video playback complete with participant controller, East Tower controller, and pilot audio could be heard. A clock displaying the Coordinated Universal Time of the demonstration was added to the center of the playback for timing purposes. A screenshot of this video playback system is shown in Figure 9.

Verbal-Command Analysis

During DFW-1, research personnel observed that many participant controllers issued instructions for flights before East Tower controllers did. Occasions in which participant controllers issued control commands later appeared to be caused by problems understanding the user interface or by other workload-inducing situations. It was hypothesized that the order in which participant and East Tower controllers issued verbal commands could indicate their cognitive workload. The differences between response times when similar instructions were issued by participant and East Tower controllers can aid in measuring the extent of cognitive workload because longer response times from

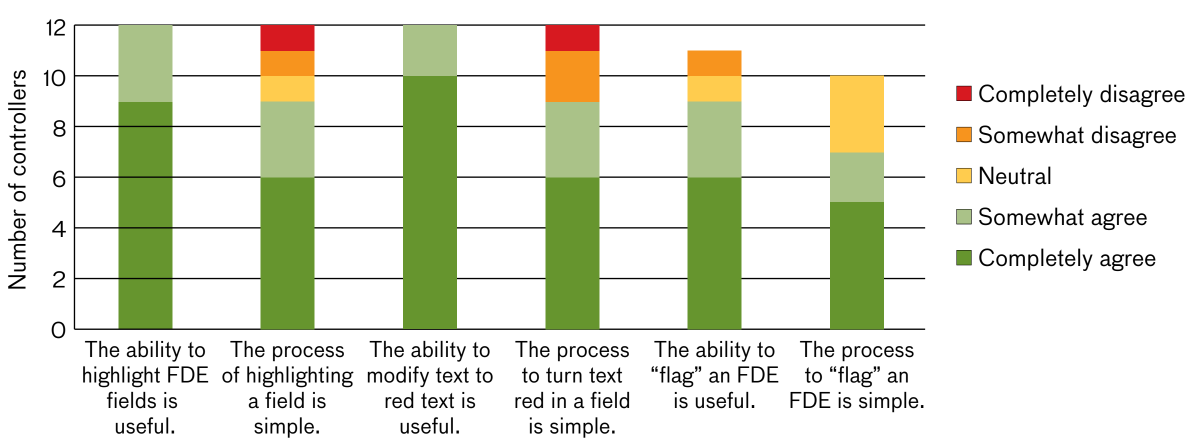


FIGURE 8. Controller questionnaires in DFW-2 indicated that the capability to highlight FDE fields and to modify text to red should remain, but that the process to do these tasks required improvement.

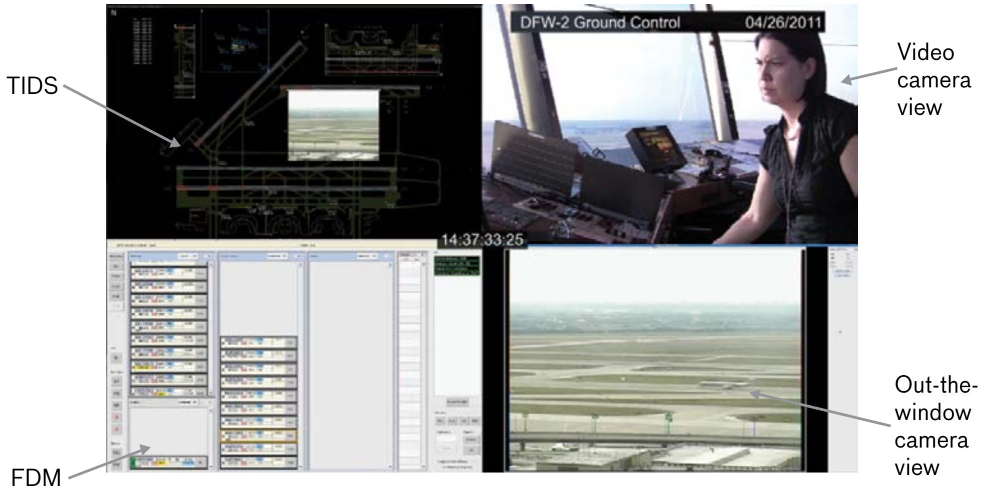


FIGURE 9. A video/audio playback system was created for post-demonstration analysis of DFW-2 data. The ability to play back the videos and screenshots simultaneously provided insight into how the controllers used the prototype and aided in identifying which features of the design required further attention.

the participant controllers indicated a larger cognitive workload for them as well.

Quantitative measures of controller responses were used to demonstrate the level of cognitive workload each participant experienced. The order in which participant and East Tower controllers issued commands for the same flights was evaluated to get the controller response rate, i.e., the percentage of instances in which participants issued a command before, after, and at the same time as (neutral to) their corresponding East Tower controller. The issuance time for each command was recorded to the nearest second with use of the onscreen clock timer. The difference in command issuance time between participant and East Tower controllers was also calculated and termed the gap time. Since shadow operations sessions were variable in length, usually ranging from 30 to 60 minutes, 5-minute data samples were selected on the basis of which had the least controller and observer interaction.

Verbal-command analysis is a novel approach to cognitive workload estimation. While our results are indicative of trends, sources of error discussed later in this article can be used to apply verbal-command analysis in a more robust fashion in future efforts to quantify ATC cognitive workload.

Verbal-command data were analyzed through two-way Welch t-tests to compare participant controllers and their East Tower counterparts. Data were also analyzed separately on individual subjects by using paired t-tests to examine differences between individual participants' respective ground and local controller roles. With these quantified results, individual participant trends were further investigated to identify specific causes of high gap times and to potentially formulate design requirements and correct specifications to address these causes.

Participants were successful at issuing the majority of their verbal commands (averaging 72% among ground and local participant controllers) before or at the same time as East Tower controllers ($t = -6.74$, d.f. = 19, $p < 0.01$). The issuance of verbal control commands is considered a secondary task indicative of cognitive workload. It appears from these results that, for a majority of the instances of control commands, participant controllers were not negatively impacted by cognitive workload issues while using the TFDM suite.

Average gap-time measurements were used to investigate instances in which verbal commands were issued second by ground and local controllers. However, no significant time differences were revealed between ground participant controllers and ground East Tower controllers nor between ground and local participant controllers. Thus, there did not appear to be any significant difference in the ability to use the TFDM suite between ground and local controller participants on the basis of verbal-command gap times.

Consolidating gap-time and response-rate data revealed individual controller trends in the form of gap-time plots. From these plots, instances in which participants had the highest gap times were correlated with actions in video recordings to reveal potential avenues of TFDM design improvement. According to the results, ground participant controllers experienced the most frequent lags in issuing instructions when interacting with electronic flight strips on the FDM, especially when searching for, moving, and editing them (23 out of 54 instances). Local participant controllers also lagged when interacting with the electronic flight strips, though to a lesser degree (8 out of 37 instances). Also, both ground and local participant controllers demonstrated instances in which instructions were issued immediately after their East Tower counterparts, prompted by hearing them, as evidenced in the video playback (19 out of 91 instances). In many of these cases, participant controllers began issuing commands as soon as they heard East Tower controllers, but before East Tower controllers finished their own commands. After examining the data, researchers determined that participant controllers clearly did not simply reiterate East Tower commands.

Visual-Gaze Analysis

Scanning data was gathered in order to quantify "heads up" and "heads down" time, and to evaluate attentional demands of the TFDM test environment. For each participant controller, five 1-minute samples of video data from each shadow operations session were analyzed to capture eye scanning behavior. The first 10 minutes of each shadow operations session were omitted from analysis in order to eliminate outlier data caused by controller unfamiliarity with the TFDM system or by controller readjustment to the system after a break. Data samples during times when participant controllers were interacting too heavily with observers were not recorded; instead, the next minute during which there was little interference was analyzed. As a result, sample minutes were not all spaced evenly apart in time.

To quantify participant controller eye movements, each potential dwell area was assigned a numerical code, called a dwell code (i.e., 1 = FDM, 2 = TIDS). For a given sampled minute, each of the controller's individual gazes was recorded as a dwell code along with its duration. An offline calibration of gaze with an ATC subject-matter expert was held to determine gaze direction. At each point at which the participant controller changed his focus, video playback was paused and the time at which the dwell began was recorded by the onscreen clock timer (centered in Figure 9). The difference between beginning and end time for these events was calculated to the nearest second and termed the dwell duration. Recording dwell codes resulted in numerical sequences that were analyzed in the form of the following variables: total dwell time per 1-minute sampling session, average single dwell length per code, and dwell frequency.

Visual-gaze data were analyzed with two-way Welch t-tests to compare behaviors of participant controllers and their East Tower counterparts. Data were also analyzed for each subject with paired t-tests to examine scanning differences between a participant controller's behavior in respective ground and local controller roles. By using quantified results, individual dwells were further investigated to identify specific causes of increased dwell duration and to formulate potential design requirements and modify specifications to address these issues.

Figure 10 indicates that participant controllers spent significantly more time heads down than heads up ($t = -17.12$, d.f. = 43, $p < 0.001$) and significantly more time was spent looking out the window ("Heads up") over "Other" dwell areas ($t = -7.87$, d.f. = 43, $p < 0.001$). Information categorized within the "Other" category comprised instances in which the participants viewed miscellaneous objects (such as their coffee) and in which they were looking at the observer. In addition to spending almost 50 seconds per sampling minute viewing heads-down information, test controllers also spent longer individual dwells viewing and gathering information from heads-down displays when compared to heads up ($t = 10.61$, d.f. = 43, $p < 0.001$). A significant amount of attention was clearly directed toward displays and information within the tower as opposed to outside of it or to miscellaneous non-informational areas. This result is natural since evaluating the TFDM display suite was the primary task requested of the participant controllers. In non-shadow operations, these measures would be particularly important to evaluate the goal of the TFDM suite to support the controllers in ensuring separation (which is currently procedurally required using out-the-window information) while minimizing heads-down time in the display suite.

Figure 10 also shows that the top three areas particularly focused on by both ground and local participant controllers were the TIDS, FDM, and heads-up areas. Total dwell-time comparisons of ground and local

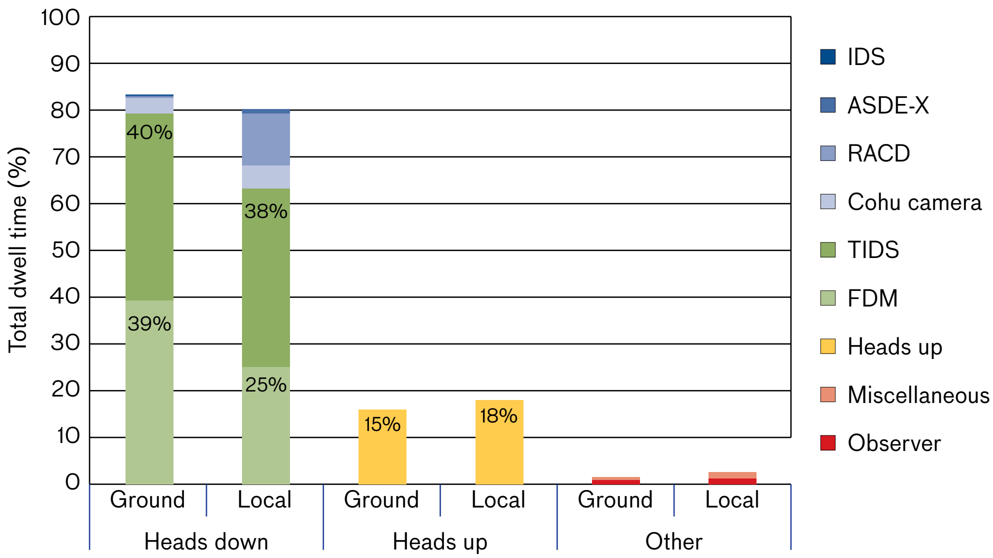


FIGURE 10. Dwell-time percentages by controller type and dwell area are shown. A majority of the visual dwell time for ground and local controller participants was spent heads-down observing the TFDM prototype and other information systems.

participant controllers revealed a statistically significant difference between ground and local participant controllers' viewing of the FDM ($t = -5.36$, d.f. = 20, $p < 0.001$) and of the RACD ($t = 13.80$, d.f. = 20, $p < 0.001$). This is an expected result because ground participant controllers spend more time editing and sequencing electronic flight strips (on the FDM) and local participant controllers spend more time monitoring arrivals and departures (on the RACD).

To elicit design recommendations from visual-gaze data, playback videos were referenced to find controller actions causing long dwell durations (defined as over 15 seconds). Ground-control participants had a higher frequency of long dwells (53 instances) compared with local-control participants (22 instances). When viewing the FDM for over 15 seconds, participant controllers were found to be spending time editing an electronic flight strip, finding one in the "Pending" bay, or using the "Search" function on the FDM. For ground participant controllers, editing electronic flight strips was the primary source of long dwells (13 out of 53). Local participant controllers, as a result of their limited role in editing flight strips in the tower environment, exhibited few long dwells when editing electronic flight strips (2 out of 22 instances). Viewing and monitoring electronic flight strips was another source of lengthy ground participant controller dwells (8 out of 53 instances); however, this step, too, had less impact on local participant controllers (3 out of 22 instances). For local participant controllers, a majority of long dwell instances were found to occur when participants were viewing/monitoring the TIDS (16 out of 22).

After the DFW-2 demonstration, a number of design issues were prioritized for post-demonstration consideration and implementation. The attention functionality required another design iteration to reduce the number of inputs to change the state of the field (e.g., highlighted, red text) in the electronic flight strip. The surveillance-based, automatic, flight-strip movement required another iteration after some flight-strip movements in DFW-2 caused the participant controllers some confusion. Further consideration of the search functionality and the editing of electronic flight strips on the FDM was also deemed useful because issues with the search functionality appeared both in the verbal-command analysis and the visual-gaze analysis.

Contributions and Next Steps

During the development process of the TFDM prototype, multiple opportunities were available to evaluate and revise its requirements and specifications. Examples of how the DFW evaluation changed the TFDM requirements and the prototype specification are shown in

Figure 11. By using field observations, benefits analyses, user groups, and field evaluations, the operational model of the tower environment and TFDM's role in it have been honed throughout the design process. After each evaluation point, the requirements for the TFDM system and the prototype specifications were modified to reflect the updated operational model.

In DFW-1, field observation data were used to revise the operational model to emphasize the importance of the ability to call attention to fields on the electronic flight strips and to classify Boeing 757s as "heavy" aircraft for wake-vortex separation requirements. The need for attention functionality was determined from the results of DFW-1, and the prototype specification for it was developed based upon the operational model of how attention functionality was used in the Dallas/Fort Worth tower (as well as other air traffic control towers visited in the field). In the DFW-2 demonstration, the participants were trained on and used the new features and improvements. Through field observations, questionnaires, and post hoc video/audio analyses, the attention functionality was evaluated. While the functionality was assessed as useful, verbal-command data analysis and visual-gaze data analysis revealed that the user interface required another iteration of design to reduce the number of inputs required to change the state of an electronic-flight-strip field. TFDM requirements were then modified to specify the number of acceptable inputs to implement attention features.

One contribution from the study was the development of nonintrusive measures of user behavior during the shadow field evaluation. Head-mounted equipment cause controllers discomfort that can result in cognitive distraction from the task at hand [10]. Using nonintrusive measures such as video recordings provides the ability to gain quantitative behavioral data without the confounding of the ecological validity of the task. One benefit to having the behavioral information is to provide a means to uniquely identify issues with the system design that were not, or could not, be identified observationally or through questionnaires. Issues with the search feature on the FDM were discovered through the playback analyses, not through recorded observations. As was discovered in this study, the information gleaned from the post hoc analyses overlapped with the information gathered through observations and questionnaires. The behavioral information also provides guidance to prioritize features included in the next development iteration.

Over 200 individual suggestions were made during the observations in the field evaluation at DFW. Without a means to effectively prioritize them, valuable development time could be spent on changes that yielded little actual benefit to the controller and his or her future work

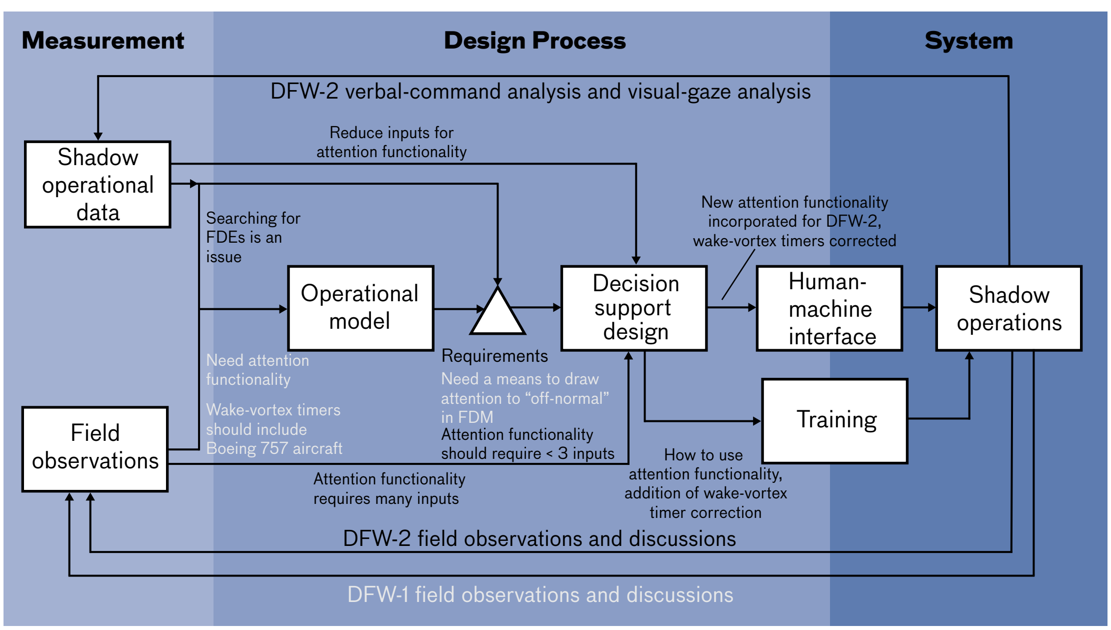


FIGURE 11. This TFDM design process model indicates selected requirements and specification modifications made as a result of the DFW-1 and DFW-2 quantitative and qualitative analyses.

process. Knowing that a design flaw was affecting the participant controller's secondary task of control command issuance or was increasing the participant controller's requirement to be heads-down to the system is a clear means of separating "required" improvements from "nice-to-haves." Through the playback analyses, cognitive engineers and systems engineers were able to focus on select issues with the FDM (searching, editing, moving electronic flight strips, surveillance-based automatic strip movement) that caused increased cognitive workload and heads-down time.

The positive feedback from the participant controllers and the TFDM user group throughout this design process, as well as the behavioral information gathered from the DFW-2 demonstration, suggests that the TFDM prototype developed thus far is acceptable and usable in the demanding environment of an ATC tower. When the participant controllers were asked whether the TIDS and FDM would be beneficial to the tower, there was strong agreement. One-hundred percent somewhat or completely agreed that TIDS would be beneficial, and 83% somewhat or completely agreed that FDM would be beneficial. TFDM appeared to exert low extra cognitive workload on the participant controllers, as proven by the fact that 73% of the control commands issued were before or at approximately the same time as their East Tower counterparts. While total heads-down time for participant controllers using TFDM appeared excessive and averaged approximately 49.1 seconds per minute (82%), other studies have reported varying measures of heads-down time, some similar to these results: 35 to 80% [11–13].

The next step is to ensure that the prototype also achieves the operational benefits (e.g., increased departure/arrival efficiency, reduced fuel burn of aircraft on the surface) for which TFDM was designed. While informative for evaluating the usability and acceptability of the prototype, the shadow operations environment was unable to support an evaluation of the TFDM prototype with respect to operational efficiency and fuel burn. Continual design to ensure operational benefit should occur through a similar iterative process using human-in-the-loop simulations followed by operational evaluations. Opportunities to evaluate potential benefits of the TFDM prototype in the future include field demonstrations planned for Washington, D.C.'s Dulles International Airport and potential TFDM benefits-related human-in-the-loop simulation studies.

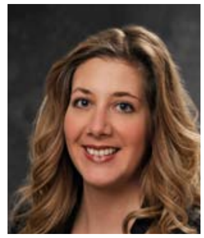
Acknowledgments

The authors wish to thank the entire TFDM and Staffed NextGen Tower teams, who were responsible for the collective design and evaluation efforts of this complex system. They would especially like to thank Lianna Hall and Dan Tennant, who were responsible for video and audio data collection and compilation described in this article. ■

References

1. J. Sheridan, "FAA Remains Quiet on ERAM Budget Overruns, Delays," AIN Online. http://www.ainonline.com/?q=aviation-news/aviation-international-news/2011-12-02/faa-remains-quiet-eram-budget-overruns-delays, 2011.
2. H.J. Davison Reynolds, R.A. DeLaura, and M. Robinson, "Field and (Data) Stream: A Method for Functional Evolution of the Air Traffic Management Route Availability Planning Tool (RAPT)," *Proceedings of the 54th Human Factors and Ergonomics Society Conference,* San Francisco, California, 2010.
3. R. DeLaura and S. Allan, "Route Selection Decision Support in Convective Weather: Case Study of the Effects of Weather and Operational Assumptions on Departure Throughput," *Proceedings of the 5th Eurocontrol/ FAA Air Traffic Management Research and Development Seminar,* Budapest, Hungary, 2003.
4. M. Robinson, N. Underhill, and R. DeLaura, "The Route Availability Planning Tool (RAPT): Evaluation of Departure Management Decision Support in New York during the 2008 Convective Weather Season," *Proceedings of the 8th USA/ Europe Air Traffic Management Research and Development Seminar,* San Francisco, California, 2009.
5. K. Swieringa, J.L. Murdoch, B. Baxley, and C. Hubbs, "Evaluation of an Airborne Spacing Concept, On-board Spacing Tool, and Pilot Interface," *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference,* Virginia Beach, Virginia, 2011.
6. T. Prevot, J.S. Mercer, L.H. Martin, J.R. Homola, C.D. Cabrall, and C.L. Brasil, "Evaluation of High Density Air Traffic Operations with Automation for Separation Assurance, Weather Avoidance, and Schedule Conformance," *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference,* Virginia Beach, Virginia, 2011.
7. G. Nagle, D. Sweet, G. Carr, V. Felipe, A. Trapani, R. Coppenbarger, and M. Hayashi, "Human-in-the-Loop Simulation of Three-Dimensional Path Arrival Management with Trajectory Error," *Proceedings of the 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference,* Virginia Beach, Virginia, 2011.
8. T.G. Reynolds, R.K. Jordan, M.A. Ishutkina, R.M. Seater, and J.K. Kuchar, "Benefits Assessment Methodology for an Air Traffic Control Tower Advanced Automation System," *Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference,* Fort Worth, Texas, 2010.

9. R.K. Jordan, M.A. Isutkina, and T.G. Reynolds, "Preliminary Benefits Pool Assessment for the Tower Flight Data Manager," MIT Lincoln Laboratory Technical Report 43PM-WX-0113, 2010.
10. R. Reisman and D. Brown,"Design of Augmented Reality Tools for Air Traffic Control Towers," *Proceedings of the 6th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference,* Washington, D.C., 2006.
11. D. Pavet, J. Garron, and I. Coullon, "Use of Paper Strips by Tower ATCO and Promises Offered by New Design Techniques on User Interface," *Proceedings of the 4th USA/Europe Air Traffic Management Research and Development Seminar,* Santa Fe, New Mexico, 2001.
12. E. Pinska, "An Investigation of the Head-up Time at Tower and Ground Control Positions," *Proceedings of the 5th Eurocontrol Innovative Research Workshop,* pp. 81-86, 2006.
13. SensoMotoric Instruments Case Study Eye Tracking: Remote Air Traffic Control. SensoMotoric Instruments GmbH, www.smivision.egts, 2011.



Hayley Davison Reynolds is a technical staff member in the Weather Sensing Group and has been conducting human-systems integration at Lincoln Laboratory since 2009. She previously worked at a human-computer interaction consultancy in Cambridge, England, performing work for Microsoft, Yahoo, and the United Kingdom's 2012 Olympic Committee. She received her doctorate from MIT in aeronautical systems and applied psychology in 2006, her master's degree from MIT in aeronautics and astronautics in 2001, and her bachelor's degree in psychology from the University of Illinois at Urbana-Champaign in 1999. Since joining the Laboratory in 2009, she has worked on multiple air traffic control decision support systems, biosurveillance systems, and a cross-Laboratory human-systems integration effort.

Kiran Lokhande is an assistant staff member in the Weather Sensing Group. She earned her bachelor's degree in engineering psychology and computer science in 2010 from Tufts University. Since joining Lincoln Laboratory, she has worked on projects addressing human-factors issues within the air traffic control domain.

Maria Picardi Kuffner is a specialist in the Surveillance Systems Group and has been conducting human factors engineering at Lincoln Laboratory since 1989, in support of both civil and military sponsors. She was awarded a master's degree in cognitive/experimental psychology from Arizona State University, sponsored by a Regents scholarship, after completing graduate studies at Harvard University and undergraduate studies at Regis College. Earlier, she was a senior engineer at Raytheon Company and a member of the technical staff at GTE Laboratories. She conducted research at the Psychophysics Laboratory of Harvard University for five years and was a programmer/analyst for the Air Force Geophysics Laboratory while an undergraduate. Kuffner is past president of the Human Factors and Ergonomics Society's New England chapter and is a private pilot.

Sarah Yenson is an associate staff member in the Surveillance Systems Group. She earned a bachelor's degree in aeronautical and astronautical engineering at MIT and a master's degree in civil engineering with a focus in transportation systems from the University of California, Berkeley. Since joining Lincoln Laboratory in 2008, she has worked on developing and testing next-generation planning, decision making, and analysis tools for the FAA and the Department of Homeland Security.