CONTROLLER-IN-THE-LOOP EVALUATION OF TRAFFIC MANAGEMENT ADVISOR (TMA) METERING DATA FORMAT AND LOCATION

D. Michael McAnulty, Ph.D., Carolina Zingale*, Ph.D., Bernardus Willems,
FAA ATO-P, *Titan Corporation
Atlantic City International Airport, New Jersey

The Traffic Management Advisor (TMA) enables en route air traffic controllers to use time-based metering to enhance airport acceptance rates. Currently, TMA data are presented on the radar display in a list that includes aircraft scheduled time of arrival and the time needed to be gained or lost to meet that time. Most facilities display this time in minutes; one in minutes and tens of seconds. We evaluated four implementations of TMA: two locations (List Only; List + in the Range Data Block [RDB]) and two time formats (Minutes Only; Minutes + Tens of Seconds). We also compared these to the miles-in-trail (MIT) method of managing traffic to determine the effects on controller scanning behavior, performance, workload, and situational awareness. Eight controllers participated in five, 40-minute scenarios, one for each condition. The participants spent less time viewing areas of the radar not containing the list in the TMA conditions compared to MIT and spent more time viewing the TMA list in the List Only conditions. They performed the metering task more effectively in the List + RDB conditions and rated that format more favorably than the List Only format. Fewer effects were found between the different time formats.

INTRODUCTION

The Federal Aviation Administration (FAA) is continually working to modernize the National Airspace System (NAS) as the demands on the system increase and as technology advances. The Traffic Management Advisor (TMA), now being deployed to Air Route Traffic Control Center (ARTCC) facilities, is designed to improve the capacity and efficiency of the NAS by enabling Air Traffic Control Specialists (ATCSs) to use time-based metering into an airport to accommodate and enhance its acceptance rate. Time-based metering has been shown to be more efficient than the current miles-in-trail (MIT) approach.

Currently, ARTCC controllers use a list on their radar display to interface with TMA. The TMA list displays the aircraft call sign, scheduled time of arrival (STA) over the metering fix, and difference in time needed to be gained or absorbed to meet the TMA-generated STA. Most ARTCCs display the time difference in minutes only, but one facility currently displays it in minutes and tens of seconds. The tens of seconds are shown as a single digit (for example, one minutes and 20 seconds is shown as “1+2”). A minus sign in front of the delay time indicates that an aircraft is behind schedule.

PURPOSE

A recent controller-in-the-loop simulation (Sollenberger, Willems, Della Rocco, Koros, & Truitt, 2004) measured controller visual scan patterns with and without the TMA list. The results indicated that controller scanning time on the radar was reduced due to the time spent scanning the list. The TMA Program Office sponsored the present study to evaluate four alternative implementations of TMA (two locations and two time delay formats) against each other and against the current MIT method of managing traffic to determine the effects on controller visual scanning, performance, workload, and situational awareness (SA).

METHOD

Participants

Eight ATCSs from five facilities currently using TMA participated in this study. All participants were nonsupervisory certified professional controllers (CPCs) who were qualified at their facilities, held current medical certificates, and were familiar with metering. The participants ranged in age from 32 to 49, with a median age of 40, and reported 6 to 26 years experience as an ATCS (median = 15). All of the participants had actively controlled traffic during the 12 months prior to the simulation, and had 11 months to 8 years of experience with TMA (median = 1 yr.). Six participants were familiar with TMA in the “minutes only” configuration. The other two participants were familiar with the “minutes and tens of seconds” configuration.

The participants indicated a high level of motivation to participate in the simulation, with a median score of 9 on a 10-point scale (1 = lowest rating; 10 = highest rating). The participants also rated their skill level, familiarity with TMA, and complexity of the typical sectors they worked as high (medians = 8). Their ratings for TMA ease of use and its effect on traffic management were more moderate (medians = 6).

Equipment

The simulation was conducted in the FAA William J. Hughes Technical Center (WJHTC) Research, Development, and Human Factors Laboratory (RDHFL). The simulation configuration consisted of the Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) and the Target Generator Facility (TGF). DESIREE emulates controller workstation (Display System Replacement [DSR]) functions and receives input from the TGF to display radar targets. The TGF also maneuvers aircraft based upon simulation pilot commands and scripted flight plan data. DESIREE incorporated the TMA prototype on the display.

The participants wore an oculometer (Applied Science Laboratories, 1991) consisting of an eye and head tracking system that allowed the researchers to measure controller eye movements under the different experimental conditions. RDHFL software can correlate what was on the dynamic display with the point of gaze information to determine the most likely fixation targets.

Airspace and Traffic Scenarios

We used a previously developed, generic, en route airspace sector that was designed to be realistic yet easy to learn (Guttman & Stein, 1997). We selected traffic scenarios from previous studies conducted at the RDHFL and modified them to meet our objectives. The scenarios mimicked the air traffic properties of the five busiest sectors in the United States. Approximately 60 aircraft were presented in each scenario, about 40 of which were arrivals and the remainder departures and overflights. The first 20 to 25 minutes of the scenarios contained a moderate level of traffic but increased after this point to reflect a push.

We prepared two basic 40-minute scenarios, one for training and the other for testing. We made four copies of each scenario and made them different only with respect to the aircraft call signs. This ensured that the scenarios were comparable in difficulty but would be less recognizable as identical.

Experimental Design

We used a repeated measures design with metering implementation as the independent variable. The five levels included (a) a baseline MIT condition without TMA, (b) TMA information presented in a list only with a time format in minutes only (LM), (c) TMA information presented in a list only with a time format in minutes and tens of seconds (LS), (d) TMA information presented in a list and in the RDB with a time format in minutes only (RM), and (e) TMA information presented in a list and in the RDB with a time format in minutes and tens of seconds (RS). We used a constrained random counterbalancing procedure in presenting the scenarios to minimize order effects.

The primary objective dependent measures were the number and duration of visual fixations on the TMA list and other areas of the radar, and metering accuracy rates. The subjective measures assessed ATCS workload and situation awareness (SA), system effectiveness and efficiency, and the acceptability of and controller preferences for the different TMA implementations. The ATCSs used two different workload rating techniques. The first was the Air Traffic Workload Input Technique (ATWIT), a real-time, unidimensional workload rating method (Stein, 1985). The participants indicated their workload level by pressing one of 10 keypad buttons labeled from 1 (very low) to 10 (very high) at 5-minute intervals throughout each scenario when prompted (i.e., tone, buttons illuminated). The second workload measurement was the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1987). The ATCSs completed this scale at the conclusion of each scenario to rate their mental and physical workload, temporal demand, frustration, effort, and performance. During each scenario, a subject matter expert (SME) unobtrusively observed the ATCS and recorded performance ratings.

Training and Scheduling

The researchers and an en route supervisory controller on detail at the RDHFL gave an inbriefing to each participant to discuss the generic airspace, any simulation-specific procedures, the different metering implementations, and how to complete all data collection procedures, including the use of the oculometer.

One participant was present during each two-day session. On the first day, we conducted the inbriefing, reviewed the participant’s rights and responsibilities, and asked the participant to sign an informed consent form. The participant then completed a short background questionnaire and began the first of five practice scenarios, one for each of the five metering implementations. On the last practice scenario, the participant wore the oculometer to become accustomed to it and to the calibration procedure. On the second day of each session, we ran the five test scenarios, with the participant wearing the oculometer in each one.

RESULTS

Visual Scanning

We evaluated the number and duration of visual fixations that the participants made on the radar scope and other areas of the workstation (e.g., desktop/keyboard). On average, the participants spent 96% of their time viewing the radar display and only 4% of their time viewing other areas (see Figure 1). A one-way repeated measures analysis of variance (ANOVA) indicated that the total amount of time spent viewing the radar display (including the TMA list) did not differ across test conditions, nor did the number of fixations.

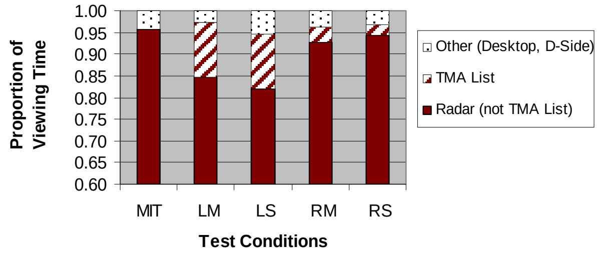

Figure 1. Proportion of time viewing the radar, TMA list, and other areas of the controller workstation.

We also evaluated the time spent and the number of fixations made on the radar after subtracting out the time spent and number of fixations made on the TMA list. The proportion of time spent viewing the remaining areas of the radar differed significantly across conditions [F (4, 28) = 94.42, p < .001], with less time spent viewing these areas in the TMA conditions compared to MIT. The amount of time spent viewing these non-TMA list areas of the radar was lowest for the LM and LS conditions. The proportion of fixations also differed significantly across test conditions [F (4, 28) = 25.09, p < .005], with the lowest proportions observed in the LM and LS conditions.

We ran a 2 X 2 repeated measures ANOVA to evaluate viewing time and number of fixations on the TMA list in the TMA conditions as a function of data location (List Only vs. List + RDB) and time format (Minutes Only vs. Minutes + Tens of Seconds). The effect of location on viewing time was significant [F (1, 7) = 116.22, p < .001]. The participants spent more time viewing the list when the data were presented in the List Only format (see Figure 2).

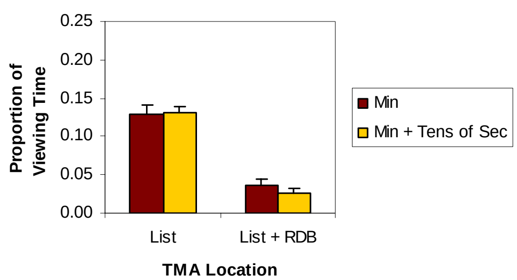

Figure 2. Proportion of time viewing the TMA list.

Likewise, the number of fixations differed significantly as a function of data location [F (1, 7) = 100.84, p < .001]. The effect of time format and the interaction of time x location were not significant for either viewing time or number of fixations.

We also evaluated the time spent viewing the arrival aircraft (see Figure 3). Similar trends were observed for the number of fixations.

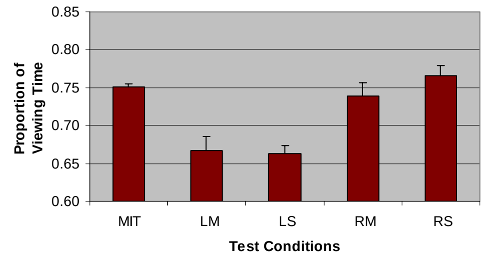

Figure 3. Proportion of time viewing arrival aircraft.

A one-way ANOVA indicated that less time was spent viewing [F (4, 28) = 21.509, p < .001] and fewer fixations were made on [F (4, 28) = 13.127, p < .001] the arrivals in the LM and LS conditions compared to MIT.

A 2 X 2 repeated measures ANOVA revealed that viewing time on arrivals was significantly lower in the List Only conditions compared to the List + RDB conditions [F (1, 7) = 47.76, p < .001], as was the proportion of fixations [F (1, 7) = 22.44, p < .005]. There was a significant interaction of time format and data location on the amount of time spent viewing arrival aircraft [F (1, 7) = 6.36, p < .05], with the highest viewing time found for the RS condition (see Figure 4). There were significantly more fixations [F (1, 7) = 5.75, p = .05] made on arrivals in the Minutes + Tens of Seconds conditions than the Minutes Only conditions (see Figure 5).

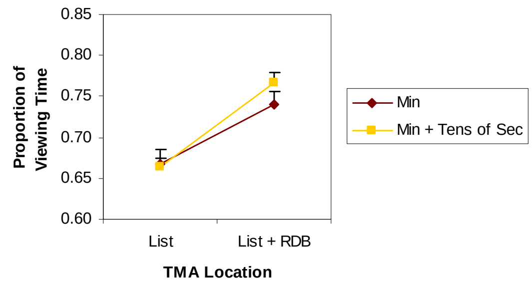

Figure 4. Proportion of time viewing arrivals.

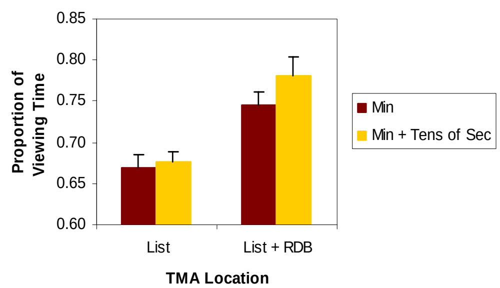

Figure 5. Proportion of fixations on arrivals.

Traffic Metering

We analyzed how effectively the participants were able to meter traffic in the four TMA conditions by calculating the proportion of aircraft that were one minute or less from their STA when exiting the sector (see Figure 6). Aircraft with delays in this range are considered to be "on time." The proportion of aircraft exiting the sector on time differed significantly as a function of data location [F (1, 7) = 7.1, $p < .05$] and time format [F (1, 7) = 4.73, $p < .1$[1]]. A greater proportion of aircraft left the sector on time when TMA data were presented in the List + RDB format and in the Minutes Only format.

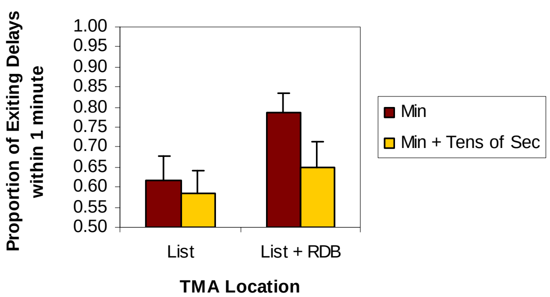

Figure 6. Proportion of aircraft delays <= 1 minute.

The trends in metering accuracy were affected by traffic density. The proportion of aircraft exiting the sector within one minute appeared higher when traffic levels were moderate and lower when traffic levels increased during the push (compare Figures 7 and 8). During moderate traffic levels (see Figure 7), metering accuracy was higher in the List + RDB conditions [F (1, 7) = 5.81, $p < .05$]. During higher traffic levels (see Figure 8), metering accuracy was higher when the TMA data were presented in Minutes Only [F (1, 7) = 5.45, $p < .1$].

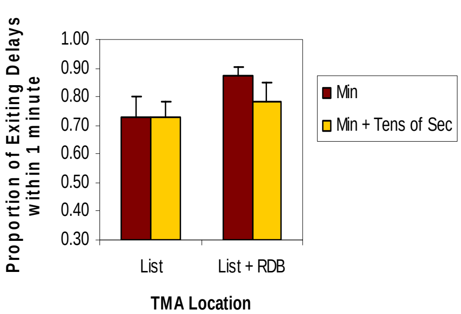

Figure 7. Metering accuracy in moderate traffic.

[1] Because of the small sample size, the power of the statistical tests to detect actual differences was limited. Therefore, we used a significance criterion of $p < .1$, rather than the more typical criterion of $p < .05$.

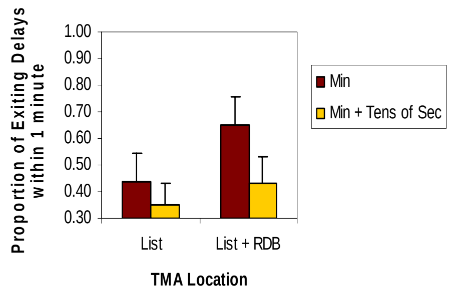

Figure 8. Metering accuracy in high traffic.

Subjective Ratings of TMA Location and Format

A 2 X 2 ANOVA revealed that participants found the TMA data easier to use when presented in the List + RDB conditions than in the List Only conditions [F (1,7) = 9.63, $p < .05$]. Average ratings on 10-point scales (1= extremely difficult; 10 = extremely easy) were 7.7 and 4.9 for the List + RDB and List Only conditions, respectively. The effect of time format was not significant. The participants also rated the TMA data easier to access when presented in the List + RDB condition [F (1, 7) = 8.31, $p < .05$]. The interaction of location x time was also significant [F (1, 7) = 9.43, $p < .05$], indicating that the perceived benefit of presenting the data in the List + RDB was more pronounced in the Minutes Only condition

The participants reported that traffic management was easier when data were presented in the List + RDB conditions than in the List Only conditions [F (1, 7) = 4.34, $p < .1$]. The interaction of location x time was also significant [F (1, 7) = 5.59, $p < .1$], indicating that the perceived benefit of presenting the data in the List + RDB was more pronounced when data were also presented in Minutes + Tens of Seconds.

The participants compared the two TMA data locations and time formats for eight ATC functions (e.g., providing timely control instructions) using a 5-point rating scale. These ratings reflected the extent to which participants found that one location or time format was "much better," "somewhat better," or "no different" than the other in supporting their ability to perform a specific task. For the location ratings, a response of "1"indicated that participants found the List Only condition "much better" than the List + RDB, "2" that the List Only was "somewhat better," "3" that there was no difference between the two locations, "4" that the List + RDB was "somewhat better" than the List Only, and "5" that the List + RDB was "much better." The median response for each of the task ratings was 4.5 - 5, indicating that participants found having the information provided in the List + RDB highly beneficial.

The evaluation of time format used a similar 5-point scale in which "1" indicated that the Minutes Only configuration was "much better" than the Minutes + Tens of Seconds format, and "5" that the Minutes + Tens of

Seconds format was "much better" than Minutes Only. The participants' reactions were variable and largely reflected a preference for the time format with which they were familiar, with median responses between 2 and 3.

Workload, Situation Awareness, and Performance

Average ATWIT workload ratings were moderate (mean responses were between 4.2 and 4.7 on the 10-point scale) and did not differ significantly across test conditions. NASA TLX ratings differed significantly in three of the six categories. Ratings of temporal demand were higher for the conditions in which data were presented in the List Only than for the baseline MIT condition [F (4, 28) = 2.33, $p < .1$]. Ratings of effort were higher in the List Only conditions than in the List + RDB conditions [F (1, 7) = 3.66, $p < .1$]. Levels of frustration were also rated higher for the List Only conditions than for MIT [F (4, 28) = 2.74, $p < .1$] and for the List + RDB conditions [F (1, 7) = 3.85, $p < .1$].

SA ratings were higher for the List + RDB conditions than the List Only conditions [F (1, 7) = 4.71, $p < .1$]. The scenarios were also rated easier to use in the List + RDB conditions than the List Only conditions [F (1, 7) = 7.92, $p < .05$], despite the fact that the scenarios were identical. ATCS performance ratings from the SMEs were high and did not differ across test conditions.

DISCUSSION

Our results indicate that displaying TMA delay time in the RDB is beneficial. The participants spent less time viewing the list and more time viewing other areas of the radar, including the arrival traffic, in the List + RDB conditions. Metering accuracy was also better when the delay time was presented in the List + RDB, and ratings of effort and frustration were lower. The participants rated the List + RDB display as "somewhat" to "much better" than in the List Only for all controller tasks. There were fewer differences found with respect to time format. When differences were found, they generally favored the Minutes Only condition, but most participants in the simulation were already familiar with this time format.

REFERENCES

Applied Science Laboratories, Inc. (1991). 4100H Eye Tracking System with Head Mounted Optics Specifications [Computer hardware]. Bedford, MA: Applied Science Laboratories, Inc.

Guttman, J. A., & Stein, E. S. (1997). *En route generic airspace evaluation* (DOT/FAA/CT-TN97/7). Atlantic City International Airport, NJ: DOT/FAA, William J. Hughes Technical Center.

Hart, S. G., & Staveland, L. E. (1987). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload.* Amsterdam: North-Holland.

Sollenberger, R. L., Willems, B., Della Rocco, P. S., Koros, A., & Truitt, T. (2004). *Human-in-the-loop simulation evaluating the collocation of the user request evaluation tool, traffic management advisor, and controller-pilot data link communications: Experiment I – tool combinations.* (DOT/FAA/CT-TN04/28). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Stein, E. S. (1985). *Air traffic controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: DOT/FAA Technical Center.