A NON-LINEAR RELATIONSHIP BETWEEN CONTROLLER WORKLOAD, TASK LOAD, AND TRAFFIC DENSITY: THE STRAW THAT BROKE THE CAMEL'S BACK

Paul U. Lee, Joey Mercer
SJSU / NASA Ames Research Center
Moffett Field, CA

Nancy Smith, Everett Palmer
NASA Ames Research Center
Moffett Field, CA

Controller workload, recognized as a significant bottleneck to capacity increase in the future National Airspace System, has been researched extensively in air traffic management. Unfortunately, subjective workload has been an unreliable predictor of a controller's ability to safely manage the traffic, leading to attempts at replacing workload with more objective metrics, such as task load (e.g. number of clearances) and traffic density (e.g. aircraft count). A significant caveat in substituting these metrics for workload ratings, however, is that their relationships are non-linear. More specifically, as traffic increases linearly, controller's perceived workload remains relatively constant until the traffic and associated task load reach a critical threshold. From this point, the workload increases at a much faster rate with each added aircraft. In an informal "traffic load test", researchers at NASA Ames Research Center manipulated the aircraft count in real-time human-in-the-loop simulations to determine the maximum traffic level at which the controllers reported the traffic to be no longer manageable. As hypothesized, traffic scenarios that were rated as moderate workload quickly became unmanageable with few additional aircraft. Feedback from the controllers further supported the non-linear nature of subjective workload. Task load data partially supported the above findings but the results were inconclusive due to differences in findings between various task load metrics. The non-linear relationship between subjective workload and aircraft count has been further examined using data from Free Maneuvering concept feasibility study in June 2004. The results showed a step-function relationship between workload and aircraft count, suggesting that controllers perceive workload as categorical. The combined results suggest that any estimation on workload should not be extrapolated linearly from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload.

Introduction

Controller workload has been a focal topic in air traffic management research (e.g. Stein 1985, Athenes, Averty, Puechmorel, Delahaye, and Collet, 2002). It is considered to be a key limiting factor to capacity increase in future air traffic operations. However, subjective workload has many undesirable characteristics. First, workload ratings have shown to have significant individual differences, making them difficult to be used as a reliable metric that can be generalized to different sectors and controllers. Furthermore, while objective metrics can be derived from traffic and sector characteristics, workload ratings are derived only after controllers work the traffic, making them difficult to be used as a predictive metric that can prevent future traffic overload.

One potential solution to this problem is to replace subjective workload with correlated objective metrics, such as peak aircraft count, traffic geometry, total time in sector, number of clearances, etc. A general approach to solving this problem is to first identify factors that are likely to correlate with workload. Then multivariate linear regression models are fitted to the observed data, followed by an elimination of factors that contribute little to the overall workload prediction. From these types of analyses, peak aircraft count has generally emerged as one of the best predictors of workload (e.g. Manning, Mills, Fox, Pfleiderer, Mogilka, 2001).

Most of these analyses assume linear correlation between workload ratings and objective metrics. This assumption seems to run counter to the subjective experience of workload. Controllers often report a low to moderate level of workload for a seemingly busy traffic but report much higher workload with few added tasks and/or minor off-nominal events once a certain traffic level is reached. In general, there seems to be a non-linear relationship between workload and objective metrics. A controller may *perceive* the workload to be low until the traffic and associated task load reach a critical point, after which s/he perceives the workload to be high.

We examined the non-linearity of workload using data that was collected during an informal "traffic load test" which established the maximum traffic that a controller can handle with advanced decision support tools. Despite the informal nature of the

study, the data provide some evidence and insight into the relationship between workload, aircraft count, and other task load metrics.

Method

Participants
Two certified professional air traffic controllers and two retired controllers/ supervisors participated in the study.

Tool Capabilities
Advanced air and ground-side decision support tools (DSTs) were integrated with Controller Pilot Data Link Communication (CPDLC) and the Flight Management System (FMS). This integration allows the controllers and the pilots to exchange 4-D trajectory information quickly and with low workload. The controller DSTs have been integrated into a high fidelity emulation of the Display System Replacement (DSR) controller workstation. In this study, all aircraft were equipped with CPDLC, FMS, and automatic dependent surveillance-broadcast (ADS-B).

Airspace
The simulation airspace included portions of Albuquerque Center (ZAB), Kansas City Center (ZKC), Fort Worth Center (ZFW), and Dallas-Fort Worth TRACON (Figure 1). Arrivals transitioned Amarillo high and Wichita Falls high from the northwest and Ardmore high from the north. The two main streams of arrivals merged at the BAMBE meter fix in the Bowie low sector. The traffic mix in Amarillo consisted of arrivals and overflights in level flight. A significant portion of Wichita Falls traffic was arrivals while Ardmore had arrivals, departures, as well as a significant number of overflights.
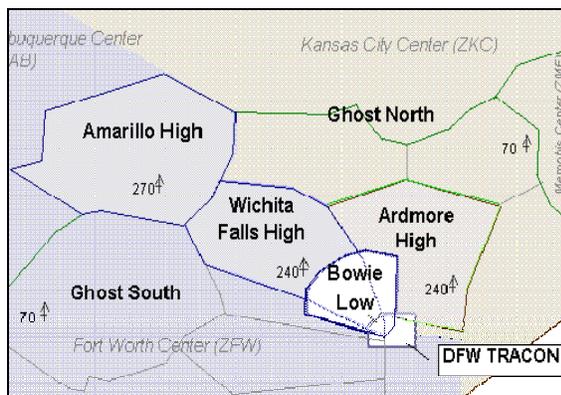

Figure 1. Simulated airspace

Procedure
The "traffic load test" was conducted to determine the maximum traffic levels that a controller can handle in each of the high altitude sectors. Each simulation run consisted of 30 – 40 minute traffic scenario, in which the traffic gradually increased during the first fifteen minutes and then sustained for the rest of the run. Ten versions of each traffic scenario were generated per sector. The aircraft count was varied at the peak traffic by an increment of two.

Each sector – Amarillo, Ardmore, and Wichita Falls – was tested one at a time. Each controller participant was paired up with a supervisor who doubled as a support controller who handled the surrounding traffic that entered or exited the test sectors. The controller participants simultaneously worked the same sector in separate parallel simulated airspaces. The controllers were given a briefing about the purpose of the study and were given training to familiarize themselves with the tools, traffic scenarios, and the overall procedures. After two days of training, the participants and the researchers discussed the definition of "unmanageable" traffic to arrive at a consensus on a common definition prior to starting the data collection runs.

For the data collection runs, a traffic level was picked based on the amount of traffic that was effectively handled during the training sessions. After working the traffic at the initial traffic level, the controller participants and the supervisors discussed and came to a group consensus on the traffic level with respect to their ability to effectively control the traffic. If they thought that the traffic was below the maximum traffic level, they worked another traffic scenario that increased the aircraft count by four, and then evaluated the new traffic scenario. If they thought the new traffic level was unmanageable, the aircraft count was decreased by two. The decision process repeated until the maximum traffic level was established for that sector. If the traffic was impossible to work, they had the option to stop the simulation run at any time. This procedure was modeled after the staircase method of establishing thresholds in psychophysical measurements (Cornsweet, 1962).

Results & Discussion

Definition of "Unmanageable" Traffic
After the training and prior to data collection, participants were asked what they would consider as "unmanageable" traffic. Surprisingly, there was a remarkable agreement among the participants in their assessments. They generally agreed that traffic

becomes unmanageable once they lose their situational awareness of the traffic situation. They also described this as losing the "flick". They described having the "flick" as having the "picture" or a plan. When they have the "flick", traffic is managed proactively to provide service rather than reactively to avoid conflicts. They felt that once they lost the "flick", safety was already compromised even if it did not result in any operational errors.

Some of the potential indicators that a controller is near the maximum traffic level are:
- handoffs are late
- can't find check-in flights easily
- reactive instead of proactive traffic control
- don't know where the planes are
- situation startles you
- service goes out the window

One controller remarked that when the traffic reached unmanageable levels during the training, he was startled to "see" an aircraft for the first time, heading for another plane in the middle of his sector. Luckily, the planes were separated by altitude but it would have resulted in a separation loss otherwise.

They also commented that near the maximum traffic level, a controller might feel that s/he is fine but one more problem – even something as simple as an altitude request – may put him/her "down the tubes". Supervisors commented that part of their job is to recognize when a controller might have reached his/her workload threshold so that they can provide relief or help before the person goes "down the tubes." They utilize the controller's body language, speech, etc., as cues for help.

Aircraft Count
The controllers worked various traffic levels during the training, which allowed them to quickly converge on the maximum traffic levels during data collection. As hypothesized, a small change in the aircraft count had a significant impact on the controller workload when the traffic was near the maximum.

For Ardmore and Wichita Falls sectors, three levels of workload – moderate, maximum, and unmanageable – were reported during data collection. As shown in Figures 2 and 3, the number of controlled aircraft was very similar between the scenarios reported as moderate and maximum levels of workload. The peak aircraft count was slightly higher in scenarios that the participants reported as unmanageable workload.
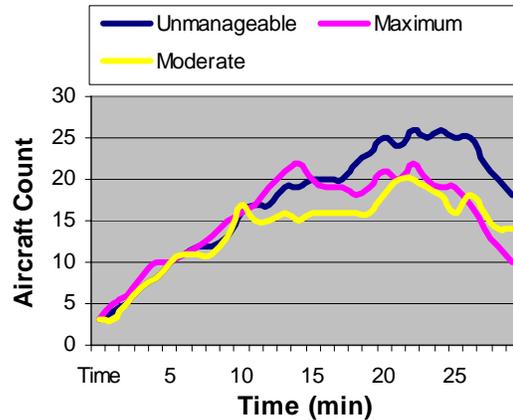

Figure 2. Controller-owned aircraft in Ardmore

The difference in aircraft count from moderate to unmanageable workload was relatively low – i.e. between 4 to 5 aircraft – suggesting that workload measurements were sensitive to minor changes in aircraft count. For the Ardmore sector, the average aircraft count during the ten minute peak was 17.2, 19.9, and 22.7 aircraft for moderate, maximum, and unmanageable workload, respectively. For the Wichita Falls sector, the average was 15, 14.7, and 18.7 for moderate, maximum, and unmanageable workload, respectively.
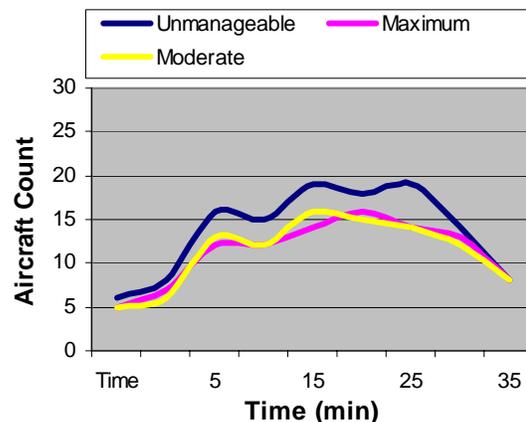

Figure 3. Controller-owned aircraft in Wichita Falls[1]

It is unclear why the moderate and maximum traffic levels had similar aircraft count in Wichita Falls sector. The task load data showed that controllers accepted more handoffs (four) and issued more clearances (3 – 11) in the maximum traffic scenario, suggesting that there were some measurable differences between the two scenarios. Further analysis is needed to understand the discrepancies

---

[1] Due to data logging problems, aircraft count was logged at every five minutes for this sector.

between task load and aircraft count in this sector.

Figure 4 shows the number of aircraft controlled in unmanageable and maximum traffic scenarios in Amarillo sector. The maximum and unmanageable traffic had 21 and 23 aircraft during the peak ten-minute duration, respectively.
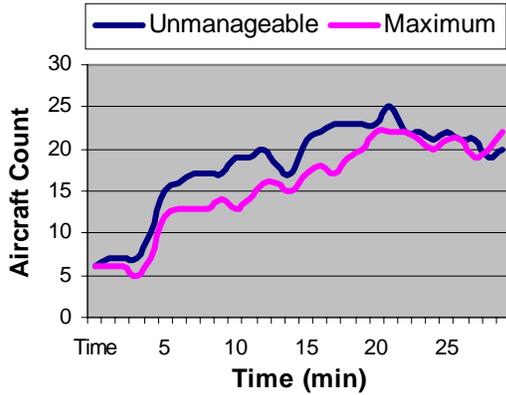

Figure 4. Controller-owned aircraft in Amarillo

Unfortunately, moderate traffic scenarios were run during training and not during data collection in this sector. Similar to Ardmore and Wichita Falls sectors, controllers reported a relatively moderate workload during training for traffic scenarios slightly below the threshold traffic, suggesting that workload increased from moderate to unmanageable with few additional aircraft.

Although the data suggest that a large change in perceived workload resulted from a small change in aircraft count, they do not directly demonstrate non-linearity in workload. However, a subsequent DAG-TM study demonstrated the non-linearity more directly. Figure 5 shows a non-linear relationship between workload and aircraft count in Amarillo high sector. During the DAG-TM study, controller participants reported their workload every five minutes during the simulation runs using a Workload Assessment Keyboard (WAK) on a scale of 1 to 7 (Stein 1985). In four simulation runs that contained maximum traffic levels, workload ratings were correlated with peak aircraft count at each five minute interval. As shown in Figure 5, reported workload was low for aircraft count up to 16 and then quickly ramped up to high workload from 16 to 22 aircraft. An S-curve, estimating a step function in workload from low to high, provided a best fit to the observed data when compared to a linear or an exponential regression line/curve (a complete analysis is in Lee, submitted).
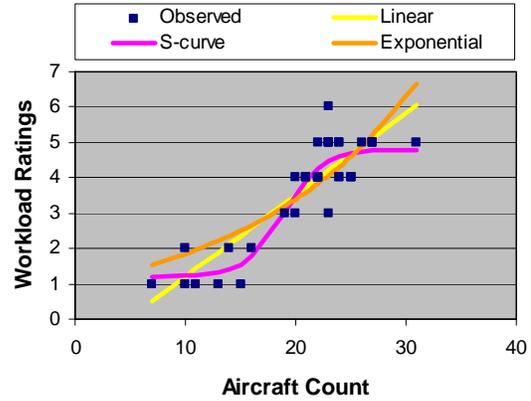

Figure 5. Workload vs. aircraft count: observed and regression fits for Amarillo High

Task Load
Controller workload was also compared to various task load metrics. A non-linear relationship between workload and task load would imply that small changes in task load would result in large changes in workload. While some of the data supported this hypothesis, others were inconclusive.

Task load metrics were divided into three main categories: handoffs, clearances, and monitoring tasks. The results of the two controller participants' performance were kept separate due to some interesting individual differences. Although task load analyses were done for all three sectors, we will focus mainly on Ardmore results in this paper due to space limitations and selectively bring in results from the other two sectors as needed. Overall, the pattern of results was similar for Ardmore, Amarillo, and Wichita Falls.

The number of handoffs that a controller accepts from an upstream sector and initiates to a downstream sector is directly related to number of aircraft in their sector. Figure 6 shows that for Ardmore sector, both controllers handled a nearly identical number of aircraft, and the number of handoffs initialized/accepted was, on average, 58, 72, and 80 for moderate, maximum, and unmanageable workload, respectively. For Wichita Falls, they were 61, 73, and 77 and for Amarillo, they were 69 and 73 for maximum and unmanageable workload. In all three sectors, the increase in the number of accepted handoffs between each traffic level were quite small (2 – 5), confirming that number of aircraft that the controllers worked were quite similar between moderate, maximum, and unmanageable traffic scenarios.
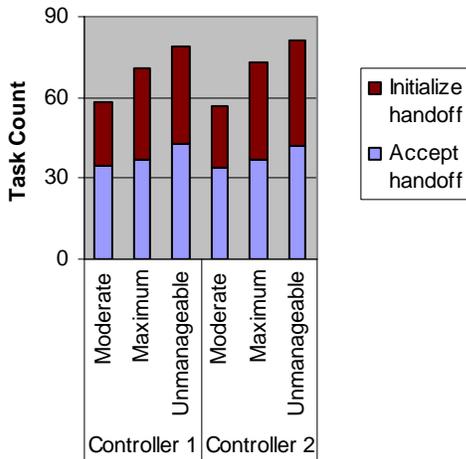
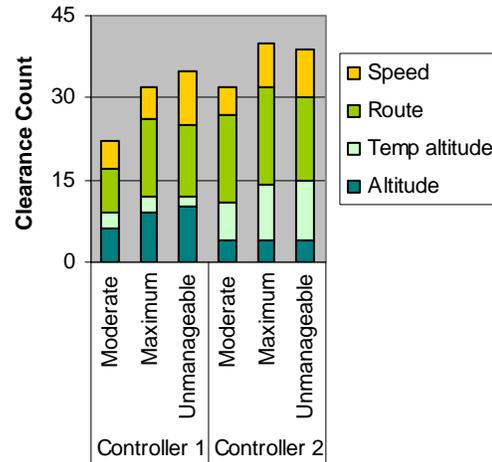Figure 6. Number of handoffs initiated and accepted for Ardmore sector

Figure 7. Number of speed, altitude, and route clearances for Ardmore sector

The number of clearances that a controller issues may be a better indicator of controller workload since it addresses not only the traffic volume but also the traffic complexity. If an aircraft flying through a sector does not increase the sector traffic complexity, controller may not need to issue any clearances to the aircraft. Figure 7 shows a number of speed and route clearances that were data linked to the flight deck, as well as the number of altitude clearances issued by voice. There were additional speed and vector clearances by voice that were not analyzed and therefore excluded in this analysis. However, over-the-shoulder observation confirmed that there were very few voice-issued vectors or speed clearances due to easy uplink of speed and 4-D route clearances via data link using advanced DSTs.

Although aircraft count data indicated a similar number of controller-owned aircraft in moderate and maximum traffic scenarios (see Figure 2), the number of clearances were greater in maximum (32 for controller 1; 40 for controller 2) than in moderate traffic (22 for controller 1; 32 for controller 2). Therefore a large increase in controller workload between moderate and maximum scenarios may be better explained by the number of clearances than by the aircraft count. However, a lack of distinct difference between the number of clearances in the maximum and unmanageable traffic scenarios limits its ability to fully explain its relationship to workload. In addition, the clearance data from Wichita Falls and Amarillo sectors did not duplicate the above results, showing only a modest increase in the number of clearances (1 – 5) between different traffic levels in all but one instance.

Controllers also engaged in various monitoring tasks. Most of the monitoring tasks were not recorded by the data collection system, but the ones that were logged show an interesting individual difference between the two controllers. Figure 8 shows the number of times the controller participants toggled or adjusted the data tags, displayed FMS routes, and displayed J-ring around the targets.
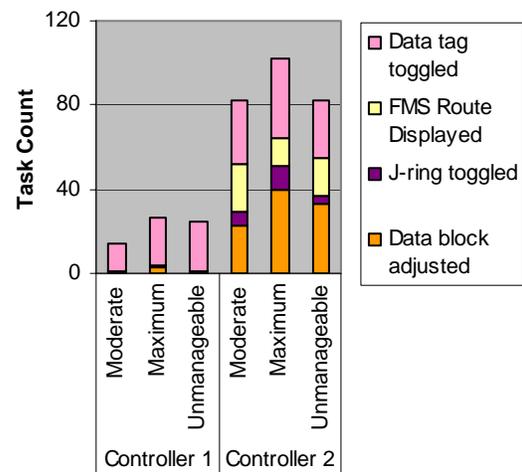

Figure 8. Number of tasks associated with monitoring for Ardmore sector

Data tag toggles and adjustments were often used as memory aids to let the controllers visually discriminate between aircraft that have been handed off, need to be attended to, etc. Display of FMS routes allowed them to verify where the planes were going, especially since the airspace and the traffic scenarios were unfamiliar to them. J-rings were often

used as additional memory aids, as well as to visually emphasize the 5 nm separation boundaries for aircraft that had potential conflicts with other nearby aircraft.

As shown in Figure 8, there was a large difference in these types of activities between the two controllers in Ardmore sector. Similarly in Amarillo and Wichita Falls, controller 2 consistently engaged in more monitoring activities than controller 1. Controller 2 also engaged in less monitoring activities in unmanageable than in maximum traffic scenarios across all three sectors, perhaps because monitoring activities were lower priority tasks that were dropped when the controller became too busy. Overall, it is interesting that these types of activities did not seem to affect their overall workload assessment since the two controller participants generally agreed on their workload in each traffic scenario despite having a large difference in these monitoring activities.

Finally, one interesting finding unique to Amarillo sector was an individual difference in the types of clearance issued by the two controller participants. As shown in Figure 9, controller 1 issued mostly lateral route amendments while controller 2 issued more altitude clearances.
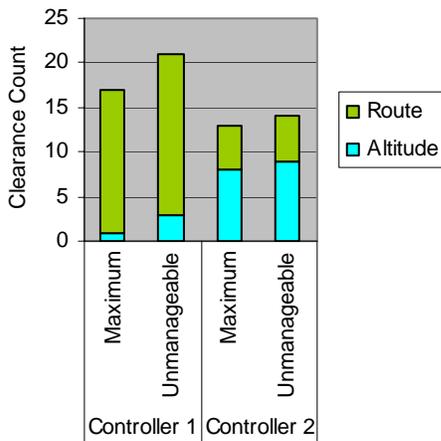

Figure 9. Number of altitude and route clearances for Amarillo sector

Controllers have commented that they try to resolve the conflicts using lateral maneuvers because 1) aircraft may be flying its preferred altitude and 2) an altitude maneuver is reserved as an "out" maneuver in case lateral maneuvers do not resolve the conflict. The data suggest that different controllers use different amount of lateral vs. vertical maneuvers in similar traffic situations.

Conclusion

There are interesting implications to the non-linear relationship between subjective workload and traffic count. First, any estimation on workload should not be extrapolated linearly from a set of workload measures taken from an experiment since the extrapolated workload is likely to significantly underestimate workload. The potential for underestimation of workload is greatest when evaluating future air traffic concepts that rely on automation to reduce task load and increase capacity. Secondly, metrics such as traffic count or task load should not be used interchangeably with subjective workload unless a better characterization of their relationship is established. Finally, non-linearity of workload implies the importance of determining the critical traffic levels that shift perceived workload from low to high. This will be a significant challenge due to individual differences in controllers' abilities and off-nominal events that can critically affect the workload. Further research is needed to understand how to accurately account for these factors.

References

Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., and Collet, C. (2002). ATC Complexity and Controller Workload: Trying to Bridge the Gap, *HCI-Aero 2002: International Conference on Human-Computer Interaction in Aeronautics,* Cambridge, MA.

Cornsweet, T. N. (1962). The staircase method in psychophysics. *American Journal of Psychology, 75,* 485-491.

Lee, P. U. (submitted). A non-linear relationship between controller workload and traffic count, *49th Annual Meeting of the Human Factors and Ergonomics Society,* Orlando, FL.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., and Mogilka, H. (2001). The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. *The Fourth International ATM R&D Seminar ATM-2001,* Santa Fe, NM.

Stein, E. (1985). Air Traffic Controller Workload: An Examination of Workload Probe, *DOT/FAA/CT-TN84/24.* DOT/FAA Technical Center. Atlantic City, NJ.