Examining the Moderating Effect of Workload on Controller Task Distribution

Paul U. Lee, Joey Mercer, and Todd Callantine

San Jose State University Foundation / NASA Ames Research Center, Mail Stop 262-4,
Moffett Field, CA 94035
{plee,jmercer,tcallantine}@mail.arc.nasa.gov

Abstract. Efforts to characterize controller workload – a key factor in limiting en route capacity – have produced mixed results. Subjective workload ratings reveal significant variations in minimum/maximum workload across individuals and show a categorical jump in perceived workload with a linear increase in aircraft count, making it difficult to predict workload limits with increased traffic. In addition, workload seems to be actively moderated by the controller to reduce monitoring tasks during high traffic/workload situations. In this paper, we examine this strategy shift by associating bookkeeping tasks and route/altitude clearances with online workload ratings. Overall, the data suggest that the controllers shed peripheral tasks related to monitoring and bookkeeping as the traffic ramps up and their perceived workload transitions from low to high. Whenever workload reached a maximum, some bookkeeping tasks were delayed and performed in "groups" after the peak traffic subsided.

Keywords: workload, task load, air traffic control, non-linear, controller strategy, situation awareness.

1 Introduction

In developing future Next Generation Air Transportation Systems (NextGen), controller workload has been identified as a key limiting factor for a significant increase in capacity and therefore has been an active field of research (e.g. [1],[2],[3]). Because workload ratings are subjective and highly prone to individual differences, some researchers have tried to replace workload with more objective metrics, such as aircraft count, number of altitude changes, number of coordination events, traffic geometry, total time in sector, etc. In related research the collective effect of all factors that contribute to air traffic control complexity was examined and termed "dynamic density" [4]. One of the key motivations for dynamic density research is to find a set of metrics that can replace current day Monitor Alert Parameters (MAPs) to predict traffic complexity and associated controller workload.

A significant caveat in substituting these objective task load metrics for workload ratings is that their relationships are often non-linear. For example, we conducted a "traffic load test" in which this non-linear relationship between workload and aircraft count was examined by asking controllers to provide feedback on the maximum

traffic they can handle and then adding a few more aircraft to see their reactions [5]. We asked controller participants to manage the challenging traffic scenarios and then increased or decreased the peak aircraft count depending on their workload feedback. The consistent results from the three sectors across participants were that adding only a few aircraft above the peak caused perceived workload to escalate from moderate to high and then to unmanageable. The left graph in Fig. 1 illustrates this point for one of the test sectors. The ten-minute peak aircraft count was 17.2, 19.9, and 22.7 aircraft for moderate, maximum, and unmanageable workload, respectively, suggesting that minor changes in peak aircraft count had significant variations in perceived workload.
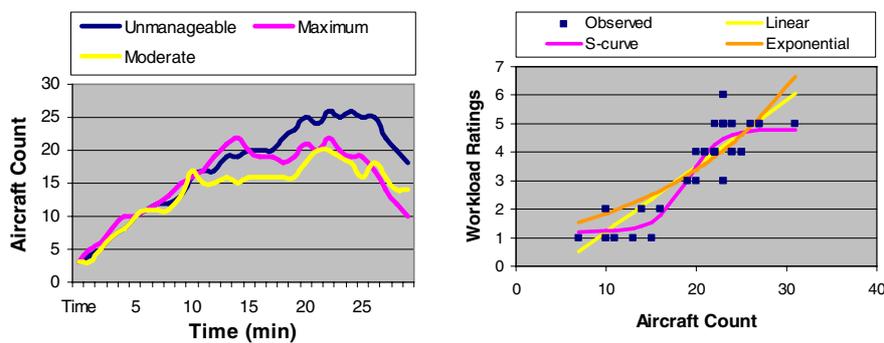

Fig. 1. Aircraft count in different traffic scenarios with moderate, maximum, and unmanageable workload during the traffic load test *(left)*; Scatter plots of workload vs. aircraft count with linear, exponential, and S-curve regression fits *(right)*

Since the peak aircraft count has emerged as one of the best predictors of workload [6], we examined the relationship between workload and aircraft count in more detail using another set of data. During this study controller participants reported their workload every five minutes during the simulation runs on a scale of 1 to 7 using a Workload Assessment Keyboard (WAK) [3]. Workload ratings were then correlated with peak aircraft count at each five minute interval. As shown in the right graph in Fig. 1, reported workload was low for aircraft count up to 16 and then quickly ramped up to high workload from 16 to 22 aircraft. Linear, exponential, and S-curve (estimating a categorical jump) functions were modeled and tested to see which function provided the best fit to the observed data. The results showed that the S-curve function captured more of the variance more consistently across multiple sectors and participants, suggesting that controllers perceive workload increases categorically with respect to the traffic count [7]. This categorical perception of workload matches well with a general phenomenon that controllers often report a low to moderate level of workload for seemingly busy traffic but report much higher workload with a few added tasks and/or minor off-nominal events once a certain traffic level is reached. In general, the ratings of different controllers across different test scenarios resulted in significant variations in minimum/maximum workload ratings as well as the shape of the curve. Given such a large variability, it would not be prudent to consider workload ratings as raw indicators of the maximum traffic that an "average" controller can handle since that number would vary significantly per

controller. However, a key advantage of the S-curve functions was that they modeled the data more consistently across different individuals [7], suggesting that S-curve functions best capture the underlying cognitive perception of perceived workload.

2   Effect of Workload on the Controller Tasks

Another challenge in correlating workload directly with task load metrics is that workload may be a causal factor that drives the task load distribution. In other words, workload is often an input factor that controllers use to actively moderate their task load distribution in order to cap one's own workload at an acceptable level. For example, controllers may try to provide efficient routings during low traffic scenarios, even if it requires more workload but abandon such practices during high traffic/workload situations in favor of actions that minimize workload.

To examine the moderating effect of workload on the types and the number of controller tasks, we re-analyzed the data from the two studies mentioned above. First, the task load – i.e. frequencies and types of controller tasks – was examined and compared with the aircraft count from the traffic load test, then the task load was compared with the associated workload ratings from the second study in a time-series plot. The results from these analyses are described below.

2.1   Task Distribution in the Traffic Load Test

Task load metrics were divided into three main categories: handoffs, clearances, and monitoring tasks. Task load was analyzed for the three test sectors, but results here will focus on a particular sector due to space limitations. Overall, the pattern of results was similar across the three sectors.

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. The handoff-related tasks therefore mirrored the aircraft count. 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 show some interesting patterns. The left graph in Fig. 2 shows the average number of times per aircraft that the controller participants toggled or adjusted the datablocks, displayed FMS routes, and displayed J-rings around the targets. Datablock 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. Toggles were used to minimize a datablock once aircraft were handed off to the next sector. Displaying FMS routes allowed the controllers to verify where the planes were going, which was important because the airspace and the traffic scenarios were unfamiliar to them. J-rings were often used to visually emphasize the 5 nm separation boundaries for aircraft that had potential conflicts with other nearby aircraft.

An interesting pattern in this graph is that as the traffic increased, these types of monitoring tasks decreased, suggesting that as controllers became busier with traffic management duties, the number of lower priority monitoring and/or bookkeeping tasks decreased.
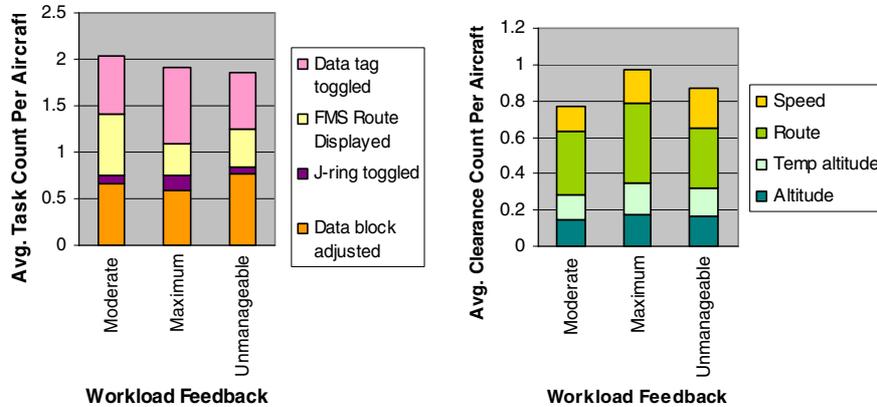
Fig. 2. Number of monitoring-related tasks (*left*) and number of clearances (*right*) per aircraft

In contrast, the number of clearances demonstrated a different pattern. The right graph in Fig. 2 shows the number of speed and route clearances controllers issued using data link, as well as the number of altitude clearances issued by voice. These results show that the number of clearances increased as the traffic increased from "moderate" to "maximum", suggesting that a slight increase in traffic count in this case created a significant increase in either the actual or perceived complexity of the traffic, leading the controllers to issue more clearances per aircraft. Interestingly enough, in the "unmanageable" traffic level, the overall number of clearances per aircraft was reduced from the "maximum" traffic level. General observations and controller feedback from the simulation runs suggest that the reduction is perhaps due to controllers abandoning clearances related to efficient traffic management to focus solely on separation related clearances.

3   Time Series Analyses of Workload vs. Task Load

Different patterns emerge with increasing traffic for the three types of task load identified in the study. Handoff-related tasks and the associated workload generally increase with aircraft count. In contrast, as the workload reaches its maximum, controllers seem to shed other bookkeeping and monitoring tasks while still trying to maintain efficient traffic flow management. However, when the traffic becomes unmanageable, they abandon efficient flow management as well and simply try to "survive" by maintaining adequate separations. By effectively shedding lower priority tasks as the traffic increases, the controllers seem to effectively keep their workload below the maximum threshold. To further examine this hypothesis, we examined the other data set to see if there is a shift in the task distribution as the subjective workload increases.

3.1   Method

For these analyses, we used high traffic simulation data that were collected during an evaluation of the En Route Free Maneuvering concept element in NASA's Distributed

Air-Ground Traffic Management (DAG-TM) project [8]. DAG-TM studies were conducted with an assumption of far-term equipage levels, including fully integrated advanced air and ground decision support tools (DSTs) with data link. The aircraft in this environment were assumed to always fly with 4-D trajectory intent which was probed for conflicts along the 4-D paths. In addition, the ground-side DSTs had 4-D route and altitude trial planning capability, which allowed the controllers to graphically construct a new conflict-free path that could be sent to the flight deck via data link. Transfer of communication was also automated and integrated with data link, sending the frequency change uplink message to the flight deck with the handoff acceptance of the next sector.

There were four experimental conditions in this study, two of which will be re-analyzed and discussed in this paper. In the first "high traffic" condition, the aircraft count reached a peak that was higher than current day MAP values, resulting in a peak workload level that was beyond the allowable level in current day operations. In a second "moderate traffic" condition, the peak aircraft count was reduced by approximately five aircraft, resulting in a comfortable moderate workload for the controller participants. Four data collection runs were analyzed per condition.

Subjective workload assessments were collected from controllers with the WAK using the Air Traffic Workload Input Technique (ATWIT) [3]. Controllers were required to rate their workload on a scale of 1 to 7 at 5-minute intervals throughout each simulation run. In the same five minute intervals, various task load metrics, such as clearances, and handoffs, as well as the average aircraft count during the interval were tabulated. The results are described in the following section.

3.2   Results

Workload vs. Aircraft Count. Fig. 3 illustrates the traffic pattern for an en route sector during the high and moderate traffic conditions, which shows that the peak aircraft count (averaged across 5-minute time span) reached 23 and 17 aircraft respectively.
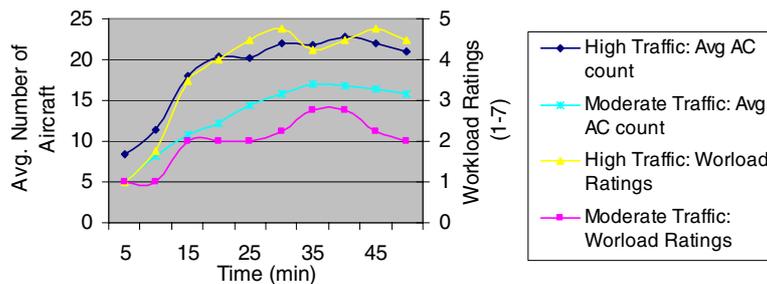

Fig. 3. Number of aircraft count vs. workload ratings for high and moderate traffic scenarios

Similar to the traffic load test results, a difference of five aircraft between the two traffic scenarios resulted in large differences in workload ratings. High traffic scenarios resulted in average workload ratings between 4 and 5, which corresponds to high workload for most controllers, as they generally reserve 6 and 7 ratings to report

situations with catastrophic failures or major re-planning such as in heavy thunderstorm or airport closures. In contrast, moderate traffic scenarios resulted in workload ratings between 2 and 3, which correspond to low to moderate workload.

Workload vs. Monitoring/Bookkeeping Tasks. Similar to task load analyses for the traffic load test, the tasks were divided into handoff-related, clearances, and monitoring/bookkeeping tasks. In this section, we focus on the monitoring/ bookkeeping tasks because our main hypothesis was that the controllers would shed less essential tasks as their workload increased.
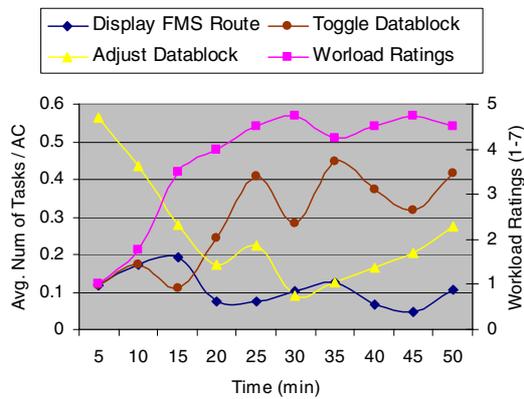

Fig. 4a. Number of "bookkeeping" tasks vs. workload ratings for high traffic scenarios
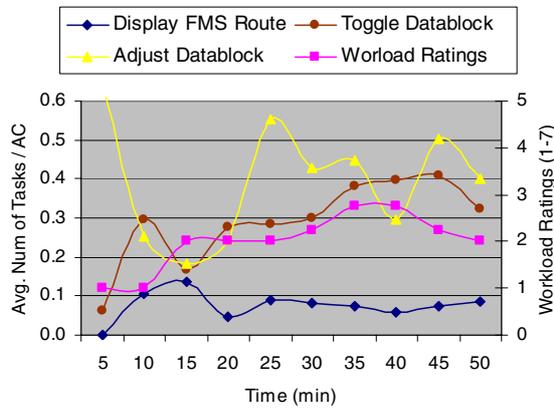

Fig. 4b. Number of "bookkeeping" tasks vs. workload ratings for moderate traffic scenarios

Only the following monitoring/bookkeeping actions that occurred with enough frequency to be analyzed: datablock adjustment, datablock toggle, and FMS route display. Each of these tasks correlates strongly with one or more bookkeeping tasks that controllers engage in to maintain situation awareness. For example, FMS routes

were displayed when controllers wanted to see their routes to resolve potential conflicts or to provide service to the aircraft. Figs. 4a and 4b suggest controllers display routes during the earlier phases of the scenarios when they have more time to plan their actions strategically. In Fig. 4a, the data in the high traffic scenarios suggest that controllers display the FMS routes for a peak of 20% of the aircraft 15 minutes into the scenario and less so thereafter. Fig. 4b shows a similar result for the moderate traffic scenarios, in which controllers display FMS routes at a peak (approx. 15% of the aircraft) at 15 minutes into the scenario. Both data show slight increase in the task frequency during dips in the workload ratings, suggesting that this task is done when workload permits.

One of the interesting results is the frequency of datablock adjustments in the high traffic scenarios. Two main reasons for datablock adjustments were 1) to organize the datablocks so that a plane entering a sector, actively being managed, or a plane ready to be shipped to the next sector each has a particular datablock orientation to remind the controllers of its current status, and 2) to keep them from overlapping on the screen. For these reasons, one would predict that datablock adjustment per aircraft would either stay relatively constant or increase slightly with increased levels of traffic. However in the *high traffic scenarios*, the data suggest that the frequency of this task sharply decreases (from about 60% of the aircraft to 10%) as the workload increases to its peak (around the 30 minute mark) and increases again as the scenario continues (see Fig. 4a). The results support the hypothesis that controllers shed this task when the workload is high, presumably because it is a lower priority task during peak workload. What is interesting, however, is that this pattern of results is not duplicated in the *moderate traffic scenarios*. Fig. 4b shows that although the frequency of datablock adjustment per aircraft dropped to approximately 20% at 15 minutes, the rate climbed back up between 30% and 50% for the rest of the scenario. Combined results suggest that high (but not moderate) workload situations reduce the rate of datablock adjustments, likely because controllers manage their workload by minimizing the frequency of peripheral tasks.

Another interesting finding is from datablock toggle task. In this simulation environment, toggling a datablock almost exclusively had one function – to minimize the datablock after the plane has been handed off to the downstream sector. This was an important task to minimize display clutter but the timing of the event was less critical. Given this understanding, it was interesting to see an "oscillation" pattern in this task that was out of phase with workload data in the *high traffic scenarios*. The result suggests that whenever the workload is at its peak, controllers delayed minimizing the datablock until the workload dipped slightly from its peak (see Fig 4a). In contrast, the oscillating pattern is not present in the *moderate traffic* condition, suggesting that this task was not delayed when the workload was not at its maximum.

Workload vs. Route/Altitude Clearances. The data used for this analysis were collected for a future operational environment that allowed route and altitude trial planning capabilities that could construct conflict-free 4-D trajectories graphically using the trackball. Controllers could then data link the conflict-free paths to the flight deck, bypassing voice clearances altogether. The only situations that required voice commands were when the clearances needed to be delivered right away or if the pilot had not responded to data link clearances. In such cases, controllers verbally assigned
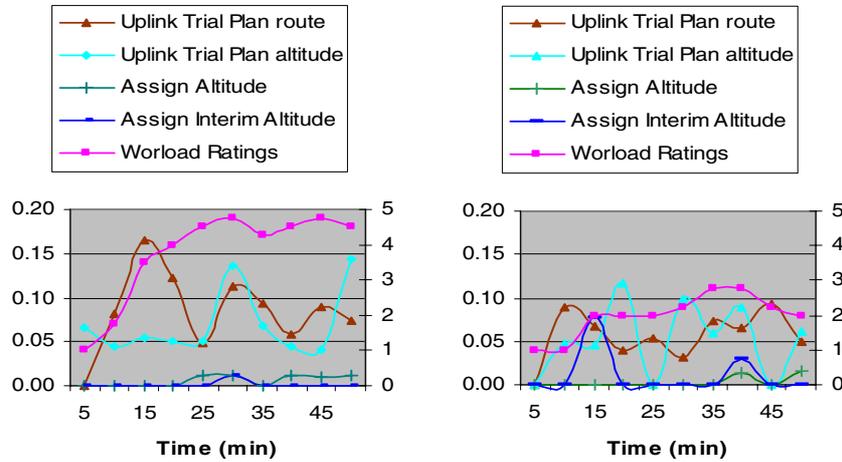
Fig. 5. Number of route and altitude clearances vs. workload ratings for high (*left*) and moderate (*right*) traffic scenarios – y-axes represent the average number of tasks per aircraft on the left and the workload ratings (1-7) on the right side of the graphs

regular and interim/temporary altitudes (shown below) as well as heading and speed changes (not shown). Fig. 5 shows the graphs for the four clearance types described above.

If the controllers minimize their workload in high traffic/workload situations by shedding lower priority and higher workload tasks, one would expect that, given a choice between a lateral and an altitude clearance, they would use the altitude clearance during high workload situation because it requires less work than a lateral route clearance due to an easier interface and less traffic complexity in the vertical dimension. On the other hand, the controllers have commented that they generally try a lateral solution first since they preferred to leave the planes at their preferred altitudes and also want to have the altitude solution available as an "out" maneuver if any last-minute maneuvers are needed. Given these two constraints, one would expect that there would be a greater number of lateral route maneuvers during the low to moderate workload situations and a shift to a greater number of altitude maneuvers during high traffic situations.

Fig. 5 suggests data trends that generally support the above hypothesis but the details of the data are difficult to interpret. As shown in the left graphs in Fig. 5, later route maneuvers in the high traffic scenarios are issued for about 15% of the aircraft at the beginning of the scenario and taper off to about 5% near the end while the altitude maneuvers are flat at about 5% of the aircraft until mid-scenario where they are issued to about 14% of the aircraft, suggesting that the route clearances are used more often during low to moderate workload situations and the altitude clearances are used most often during peak workload situations. In the moderate traffic scenarios, the lateral route clearances have higher frequencies during lower workload situations than during higher workload situations, suggesting again that the route clearances are used more often during these periods. However, altitude clearances seem to be issued periodically throughout the scenario, suggesting that they are not used specifically to

minimize workload in the moderate traffic situations. There also seem to be periodic oscillation patterns present in these data, but more analyses are needed to understand the exact nature of the oscillations.

Workload vs. Handoff-related Tasks. Similar to the traffic load test data, the number of handoffs that a controller accepts from an upstream sector and initiates to a downstream sector is directly related to the number of aircraft in their sector. The average frequency of handoff initiation and acceptance per aircraft therefore shows considerable similarity between the high and moderate traffic/workload situations (see Fig. 6). The similarity between the two graphs suggest that these tasks are important and could not be shed or delayed significantly during maximum workload situations, unlike the bookkeeping tasks and the clearances described in the previous sections.
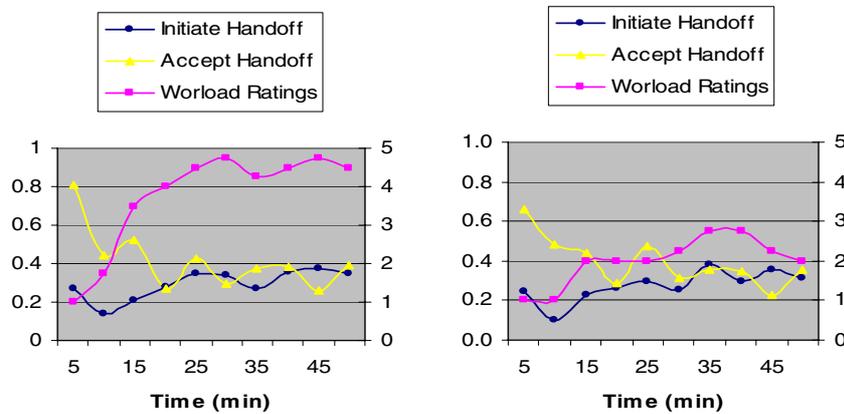

Fig. 6. Number of handoff-related events vs. workload ratings for high *(left)* and moderate *(right)* traffic scenarios – y-axes represent average number of tasks per aircraft on the left and the workload ratings (1-7) on the right side of the graph

4   Conclusion

Based on the examination of how workload affects task distribution, the results suggest a qualitative shift in the types of tasks that controllers perform in low, moderate, and high workload states. During low traffic/workload states, controllers engaged in a significantly higher percentage of "bookkeeping" activities, such as datablock adjustments, than in the high workload states. If the workload reaches the maximum such that controllers need to manage their workload by selectively shedding or delaying tasks, it seems that they shed datablock adjustments and delay toggling/minimizing datablock until they have enough time to attend to this task. None of these patterns emerge in the moderate traffic scenarios, presumably because controllers have enough mental resources to perform all of the tasks.

Examining route vs. altitude clearances, we expected controllers to issue more altitude instead of lateral route clearances during peak workload situations because

altitude clearances generally take less workload. Although the data seem to support this hypothesis, they were not conclusive.

Understanding how workload moderates task load distribution has significant potential for predicting true workload limits. If the pattern of delayed and dropped tasks show better consistency across controllers than the subjective workload ratings themselves, one can look for these patterns to indicate when the controllers are reaching their mental resource limits, which in turn could provide inputs to safety implications and capacity limits.

References

1. Athenes, S., Averty, P., Puechmorel, S., Delahaye, D., Collet, C.: ATC Complexity and Controller Workload: Trying to Bridge the Gap, In: HCI-Aero: International Conference on Human-Computer Interaction in Aeronautics, Cambridge, MA (2002)
2. Hilburn, B.G., Bakker, M.W.P., Pekela, W.D, Parasuraman, R.: The effect of free flight on air traffic controller mental workload, monitoring and system performance. In: Proceedings of the 10th International CEAS Conference on Free Flight pp. 14–1/14–12 (1997)
3. Stein, E.: Air Traffic Controller Workload: An Examination of Workload Probe, DOT/FAA/CT-TN84/24. DOT/FAA Technical Center, Atlantic City, NJ (1985)
4. Kopardekar, P., Magyarits, S.: Measurement and prediction of dynamic density. In: 5th USA/Europe Air Traffic Management R&D Seminar, Budapest, Hungary (2002)
5. Lee, P.U., Prevot, T., Mercer, J., Smith, N., Palmer, E.: Ground-side Perspective on Mixed Operations with Self-separating and Controller-managed Aircraft, In: Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD (2005)
6. Manning, C., Mills, S., Fox, C., Pfleiderer, E., Mogilka, H.: The Relationship Between Air Traffic Control Communication Events and Measures of Controller Taskload and Workload. In: The Fourth International ATM R&D Seminar ATM-2001, Santa Fe, NM (2001)
7. Lee, P.U.: A Non-Linear Relationship between Controller Workload and Traffic Count. In: 49th Annual Meeting of Human Factors and Ergonomics Society, Orlando, FL (2005)
8. Lee, P.U., Prevot, T., Mercer, J., Smith, N., Palmer, E.: Ground-side Perspective on Mixed Operations with Self-separating and Controller-managed Aircraft, In: Proceedings of the 6th USA/Europe Air Traffic Management Research and Development Seminar, Baltimore, MD (2005)