Acceptability and Effects of Tools to Assist with Controller Managed Spacing in the Terminal Area

Lynne Martin[1], Michael Kupfer[1], Everett Palmer[2], Joey Mercer[1], Todd Callantine[1], and Thomas Prevôt[2]

[1] San Jose State University
[2] NASA ARC: NASA Ames Research Center, Moffett Field, California, USA
{Lynne.H.Martin,Michael.Kupfer,Everett.Palmer,Joey.Mercer, Todd.Callantine,Thomas.Prevot}@nasa.gov

Abstract. In a human-in-the-loop simulation, a scheduler delivered aircraft to meter fixes in the Los Angeles terminal area with a -60 to +30 second accuracy. This study investigated whether, and how well, controllers could control aircraft to land them as close to their scheduled time of arrival (STA) as possible using speed control alone. Controllers were assigned one of three levels of tools to assist them but had to compensate for errors in the forecast winds that had not been taken into account by the scheduler. Results show that speed clearances were sufficient under all conditions to maneuver aircraft closer to their STAs. From participant reports, this form of control incurred manageable workload and two of the three levels of tools were deemed easy to use.

Keywords: decision support tools, controller managed spacing, terminal area, utility and usability.

1 Introduction

One aim of the next generation air transportation system in the USA (NextGen)[1] is to maintain a high level of (or increase) the throughput at airports and more efficiently manage the traffic in dense terminal areas. Research in both Europe and the USA has focused on developing trajectory management tools that will enable aircraft to execute efficient descents and maintain throughput [e.g., 2, 3]. This requires more precise navigation, which is accomplished through more stringent Required Navigation Performance (RNP) criteria. Although NextGen Super Density Operations (SDO) [4] are founded in scheduling tools that can organize aircraft and specify scheduled times of arrival (STA), this arrival concept still requires controllers to work the traffic as it moves through the terminal radar approach control (TRACON) area to keep each aircraft on its tightly-packed schedule.

Currently TRACON controllers manually space aircraft around merge points and for landing, absorbing the frequent delays that are incurred from this control by speed changes and tactical vectoring (horizontal path changes)[5]. While vectoring is acceptable when aircraft are flying 3D paths, when operations become trajectory-based (TBO) – where aircraft fly 4D paths – vectoring is less efficient than speed

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2011, LNAI 6781, pp. 432–441, 2011.

control and contributes significantly to trajectory prediction uncertainty, along with wind variations. Realizing the benefits of the proposed scheduling tools requires a shift in current terminal area control practices away from vectoring strategies to strategic speed and path control on lateral area navigation (RNAV) routes (i.e., time-based procedures).

For TRACON controllers, this shift to TBO will mean not only a change in the way they control aircraft but also a shift in both the salient and critical information they will need to make control decisions. While at a casual glance, changing information when the task remains fundamentally the same would seem not to pose a problem, Nunes [6] cautions that some air traffic control (ATC) decision aiding tools have had adverse effects because they removed information. That controllers are very sensitive to key pieces of data is supported by Seamster, et al. [7], who found that expert air traffic controllers were able to efficiently focus on the most critical information for control decisions. Endsley and Rodgers [8] also found that ATC experts either pass over or do not commit less important data to memory.

Given the level to which ATC skills are honed and with a change in mode of aircraft control, it is likely that TRACON controllers will need to develop new strategies for their decision-making [7] and possibly revise aspects of their mental models [6] of the way traffic moves through the airspace. To support controllers with this, tools are proposed that assist at different levels of automation [9], from displaying schedule information to suggesting solutions (advisories). By introducing tools that provide different levels of information and assistance to ATCs, a comparison of tool effectiveness was facilitated.

The objective of the current study was to determine, through a human-in-the-loop simulation, how well controllers can manage spacing of arrival aircraft on Optimized Profile Descents (OPDs) along RNAV/ RNP routes with the assistance of different advisory tools and enhanced displays designed to assist them with keeping aircraft on their 4D trajectories. The trajectory-based tools were designed to support TRACON controllers by providing information for aircraft speed control that met a time-based schedule. Rather than testing each tool individually, the study investigated whether the tools could effectively be integrated into a terminal-area controller workstation, and how well the different tools functioned in concert.

2 Methods

2.1 The Simulation: Airspace, Route Structure, Scenarios and Winds

The airspace simulated for this study was the terminal area around the Los Angeles International Airport (LAX). Aircraft in the simulation flew OPDs on merging RNAV routes to runways LAX24R and LAX25L. The RNAV routes were designed based on existing Standard Terminal Arrival Routes (STARs; e.g., RIIVR2) and approaches. Several speed and altitude restrictions were created to give a sufficiently shallow descent angle of 2.4°. This allowed speed control to be used along the OPDs. The speed restrictions supplanted tactical controller speed assignments for fly-ability and predictable flow control. Fig. 1 shows a map of the simulated airspace displaying the routes, waypoints, and sector boundaries (based on current sectors in the Southern

California (SoCal) TRACON). The simulated airspace was comprised of three feeder sectors, Zuma, Feeder and Feeder South, and two final sectors, Stadium and Downe.

Fig. 1. The LAX airspace created for the simulation

Two different scenarios were developed for the simulation. Both scenarios included 25 aircraft flying to each runway. The aircraft type mix and the traffic load distributions on the routes were selected based on an analysis of actual arrival traffic to LAX. The scenarios were built under the assumption that aircraft had been delivered to the TRACON meter fixes by en route control with no more than a 90 second nominal schedule error (up to 60s early and 30s late). However, due to the wind forecast errors, the error between the estimated time of arrival (ETA) and STA differed from that range. In addition to the standard wake spacing distances an additional buffer of 0.5 NM (Nautical Miles) was added into the scheduler to protect the wake spacing and reduce the possibility of violations.

Winds were always a headwind aligned with the landing runway from 265°. Above 20,000ft and below 1,500ft the forecast wind profile matched the actual wind profile. However between these altitudes there were two wind-forecast-error conditions where the actual wind differed from the forecast wind. This has an effect on the accuracy of the higher-level tools (see below) because their calculations take the forecast winds into account. In the “minus-bias” wind condition the forecast winds were 10kts less than the actual winds and in the “plus-bias” wind condition the forecast winds were 10kts stronger than the actual winds. A third “no-bias” condition was also used, where the actual winds were the same as the forecast.

2.2 Tools

The study was run in the Airspace Operations Laboratory (AOL) at the NASA Ames Research Center using Multi Aircraft Control System (MACS) software [10]. MACS provides an environment for rapid prototyping, human-in-the-loop air traffic simulations, and evaluation of current and future air/ground operations. Simulated aircraft were assumed to be Flight Management System and Automatic Dependent Surveillance-Broadcast-out equipped.

The controllers worked with an emulation of the Standard Terminal Area Replacement System (STARS) onto which one of three levels of decision support tools (DST) was added. The first level toolset could be implemented in the near-term NextGen; it consisted of a double-sided timeline comparing aircraft STAs with their ETAs. The timeline (TL; Fig.2.1) was referenced to an appropriate waypoint for each controller, e.g., LAX25L for the Downe position. Accompanying the timeline was an early/ late indicator displayed in the third line of an aircraft's data block (FDB) if the aircraft was five seconds or more early or late. These DSTs showed a controller how close an aircraft was to being on time and therefore the amount of time that s/he needed to create or absorb to put the aircraft on schedule.

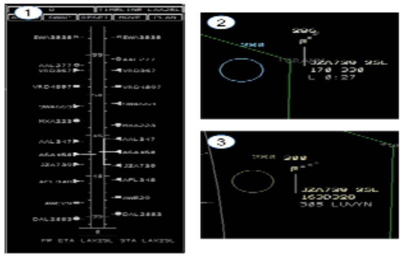

Fig. 2. In clockwise order from left: (1) timeline, (2) slot marker and early/late indicator in slot marker condition, (3) advisory and slot marker in advisory condition

The second level toolset added position prediction information in the form of a slot marker (SM) to the schedule information. The SM showed with a circle where the aircraft should be on its longitudinal arrival route in order to fly the charted speed profile in the forecast winds and arrive at the runway on its STA. This means that an aircraft in the center of its slot marker circle would be properly spaced behind its lead (providing the lead aircraft is also in the center of its slot marker). This level of support provided a target around the best schedule position for an aircraft that a controller could work towards.

The third level toolset may be envisaged in the far-term NextGen; this toolset retained the timelines and the slot marker circles but replaced the early/late indicators in the FDB with speed advisories (AD). The speed advisories suggested a speed over a distance, which, if the aircraft flew it, would put the aircraft onto its STA at the outer marker. Like the early/late indicator, the speed advisories were displayed to the controllers if an aircraft's ETA differed from its STA by five seconds or more. This DST is at a higher level of automation [9] than the other two conditions because the tool recommends a solution to the controller rather than leaving the controller to work out a solution on his or her own.

2.3 Controller Tasks, Participants, and Data Collection

The goal set for our participants was to efficiently deliver aircraft on their routes to the outer marker and runway with no wake spacing violations. The controllers were asked to manage the arrival traffic, correcting those aircraft ahead or behind schedule, and to cope with disturbances (i.e., aircraft not conforming to speed restrictions), all while dealing with the errors between forecast and actual winds. They were asked to do this using only speed control if possible. The role of the feeder controllers was to issue an approach clearance for each aircraft along an RNAV/RNP route to its assigned runway, then try to deliver the aircraft as close as possible to its STA by the sector exit point.

The final controller was tasked with further fine-tuning the traffic received from the feeder controllers to deliver the aircraft at their STAs at the outer marker.

Five air traffic controllers took part in the study. They had an average of 23.6 years of experience and had been retired for 1.9 years on average. Confederates who were also retired controllers staffed two ghost positions. The pseudo-pilots who worked the traffic were active commercial pilots or aviation students who had experience using the MACS software.

Prior to data collection, participants received three days of training in fully running practice simulations where they worked the position they would work during the data collection. Data was collected during simulation runs over five consecutive days. Each tool and wind condition combination was run under the two traffic scenarios to give a total of 18 runs. One run had to be repeated due to a procedural error. Each run lasted for an hour and immediately following it participants completed a questionnaire about that run. At the end of the data collection, participants completed another questionnaire about more general topics, and took part in a debrief discussion.

Each controller and pseudo-pilot workstation recorded a number of variables in data logs throughout every simulation run. Aircraft performance data, trajectory and flight state information as well as pilot and controller data entries were logged. As an extra task, controllers were prompted, at five-minute intervals, to give a rating of their workload for that moment between 1 ("very low") and 6 ("very high") through a workload assessment scale based on the ATWIT (Air Traffic Workload Input Technique [11]) that was embedded in the MACS software. Voice communications between controllers and pilots were through an emulation of the FAA's Voice Switching and Communication System and these communications were recorded.

3 Results

3.1 Task and Goal Achievement

A number of metrics were calculated to assess whether the controller team achieved their goals. Route conformance is one: in order to receive the benefits of OPDs aircraft are required to conform to their route with high precision. This was achieved; in all conditions route conformance within 1 NM was approximately 99.5%. Moreover, the controllers did not use vectoring techniques to manage the arrival stream, completing their tasks using speed control alone. Vertically there were very few level-offs, only 19.2% of aircraft leveled off for an average of 2.48NM. Most of these were due to deceleration for waypoint restrictions, which again means that controllers adhered to the OPDs as much as they could.

The second goal set for the controllers was to deliver aircraft with minimum spacing between a lead aircraft and its follower when crossing the runway threshold. Overall, the average inter-arrival spacing of aircraft pairs did not vary much when compared across tool conditions; they varied around 0.53 NM ($\sigma = 0.27$ NM), which reflects the spacing buffer. Throughout all of the data collection runs involving 900 aircraft and 864 aircraft pairs, there was only one wake spacing violation that was due to controller error. It took place during a timeline condition and the follower was

0.12NM closer to its leader than it should have been, suggesting controllers generally maintained a high level of traffic awareness.

A third goal was for controllers to control the aircraft in order to arrive at the runway on schedule, known as “schedule conformance”. This metric was defined as an aircraft’s ETA minus its STA (negative values indicate the aircraft was ahead of schedule). Schedule conformance indicates a combined effort from the controllers, because to achieve this goal the controller team must reduce the ETA-STA differences over the entire length of a flight through the TRACON. Compared to the initial –60 to +30 s distribution of traffic the schedule errors measured at the runway are greatly reduced. The distribution peaks around -5 s with a mean of μ=-1.21 s and standard deviation of σ = 5.21 s (Fig. 3). The curve is sharper on the left, indicating controller effort not to exceed the schedule buffer, and is wider to the right – excess spacing is somewhat inefficient but keeps aircraft separated.

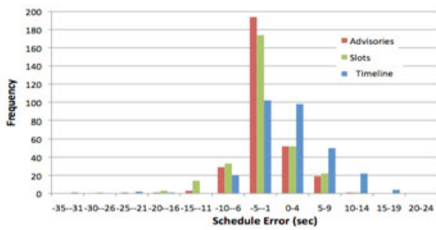

Fig. 3. ETA-STA error measured at the runway threshold broken down by tool condition

Breaking down the results by tools condition shows that the advisories and the slot markers have very similar schedule conformance, while the timeline condition has a less tight conformance to the schedule. A one-way ANOVA indicated that these conformance patterns are significantly different ($F(2,838)= 48.32$, $p=.000$). A post-hoc Tukey’s HSD test confirms that all the tool conditions are significantly different from each other at the $p<.05$ level. From this it seems the pattern of schedule conformance is most different in the timeline condition while the slot marker and advisory conditions’ patterns are more similar.

3.2 Task Load and Its Acceptability

In addition to the metrics of task achievement, another set of measures was taken to assess whether the load on the controllers was reasonable while they were completing the study tasks. Controller workload was measured in real-time using an ATWIT-based procedure [11]. All controllers perceived their workload on average as “low” to “very low” (μ = 1.85, σ = 0.53). Although the response scale offered six choices, controllers effectively used only a 3-point scale – from the raw scores, no controller ever gave a rating of 5 or 6 and rarely gave a rating of 4. Some of these workload ratings, those reported from the feeder controllers, stem from low sector traffic complexity because only arrival operations were simulated.

When participants’ workload ratings are distributed by the study’s conditions, there are only small differences between participants’ estimations of their workload (Fig. 4).

A comparison between the mean workload ratings of the three toolset conditions showed the mean rating for the timeline condition was higher than that for the speed advisory condition but only by .08 of a scale point (TL μ = 1.68, AD μ =1.60). Unsurprisingly, differences between participants' mean ATWIT ratings per run are not statistically significant when compared by tool condition.

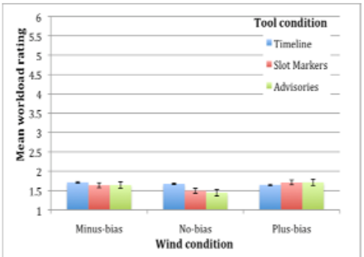

Fig. 4. Mean of ongoing workload ratings by study condition, with standard error bars

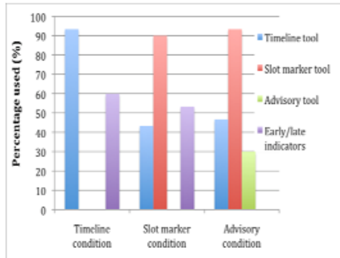

Fig. 5. Amount participants "used" the tools in each toolset condition

As a complement to the real-time workload ratings, workload data was also collected in post-run questionnaires using the NASA-TLX [12]. Controllers completed six scales that comprise this rating scheme after each run, using a ranking that ran from very low workload (1) to very high workload (7). Overall participants' average workload was between "low" and "somewhat low" (μ=2.57, σ=1.36). When the TLX scores were organized by tool condition the differences between the means were small, similar to the differences between the ATWIT means. Once again participants rated the timeline condition as having the most combined workload (μ = 2.64) and the slot marker and advisory conditions as having almost identical workload (SM μ=2.51; AD μ=2.56). Again, there were no significant differences between participants' ratings of their TLX workload. Considering that participants were only undertaking a portion of the tasks working these study sectors that they would normally have when working these sectors at SoCAL, workload ratings of "low' to "very low" are reasonable.

Controller load was assessed in a third way through the amount of communication that was required. All clearances were issued by voice, thus creating the physical load for participants and potentially a time constraint as voice clearances have to be issued serially. Across all runs aircraft received a mean of 2.5 clearances per controller. The controller of Zuma, exclusively handling aircraft flying to LAX24R, issued the most clearances on average (μ = 3.1, σ = 1.41) while, as expected, the Feeder South controller (LAX25L traffic only) issued on average the fewest clearances (μ = 1.98, σ = 1.24). At this broad level, the number of clearances issued support participants' physical load and time pressure reports; with the Zuma and Stadium controllers issuing the most clearances on average and reporting the highest mean physical load and time pressure, and the Downe and Feeder South controllers issuing the fewest clearances on average and also reporting the lowest mean workload. Comparing across the different tool conditions, the aircraft under the timeline condition received fewer clearances on average compared to the other two tool conditions (TL: $\mu = 2.18$, $\sigma = 1.4$; SM: $\mu = 2.70$, $\sigma = 1.42$; AD: $\mu = 2.64$, $\sigma = 1.37$). Comparisons of the mean clearances per aircraft do not have a statistical difference.

3.3 Tool Usage and Acceptability

Controllers were asked a number of questions about how the tools impacted the way they completed their tasks after each run. For example controllers were asked how much they used each of the four key tools and how useful they were. Note that the slot markers and the advisories were not available in every tool condition but the timelines were. Participants reported they used the timeline significantly more often ($\chi^2(2) = 8.897$, $p = 0.012$) in the timeline condition (93% of the time) when compared with the other two conditions where they reported using the timeline much less (46.6% and 43.3% of the time) (Fig. 5).

This would suggest that participants could use the timeline but it was not their first choice of tool. Although there was a statistical difference between the amount controllers said they used the slot markers in the slot marker condition versus the advisory condition, it was not a meaningful difference because controllers said they used the slot markers 93% of the time in the advisory condition and 90% of the time in the slot marker condition. However, this *is* meaningful in terms of the tools. Controllers reported they used the slot marker (as noted) 93% of the time in the advisory condition but they only used the advisories 30% of the time – which indicates they chose not to use the most advanced tool. Controllers' comments support that they preferred the slot markers over the advisories and the timelines as their "tool of choice". Controllers used the early/late indicators about the same amount in the timeline and slot marker conditions but more than they reported using the advisories that replaced them in the advisory condition.

In a post-study questionnaire, controllers were asked about how useful each tool was and how useable it was. Responses to these two questions are highly correlated ($\tau = 0.66$, $p < 0.01$) and show that, in general, if participants thought a tool was useful they also thought it had a high level of usability. The slot markers were rated as "very useful" ($\mu = 4.6$) with "high usability" ($\mu = 5$), and participants commented: "Using them helped [me] to make adjustments on aircraft" indicating that they were using the markers in the way the study intended them to be used. The same was true for the early/late indicators; participants used them as planned ("Used the early/late advisories until they would disappear then I would use the timeline for the final adjustment") and they were rated favorably as "useful" ($\mu = 4.2$) with "high usability" ($\mu = 5$). The timeline was also rated positively as both "useful" ($\mu = 4$) and "useable" ($\mu = 4.2$) and gets a third positive rating because it was the only one of the three main tools that no-one said they would have liked to have been able to turn off. However, participants did say that they found the timeline hard to use because it "took my attentions away from the radar screen." The speed advisory was rated lower and more variably than the other tools. Overall, participants said the advisories were "somewhat useful" ($\mu = 2.8$) and "somewhat usable" ($\mu = 2.75$) but, also, three participants would have liked to be able to turn them off or use them for information only.

Debrief discussions probed a little further into why the participants did not view the advisory tool as favorably as the other tools. Participants explained that the advisory tool often issued an advisory to a waypoint that was downstream of their sector, which meant (if the controller issued the advisory) that the aircraft would not be in conformance by the time it was due to be handed off to the next controller. This is at odds with current-day controller work techniques to complete all tasks related to a given aircraft before handing it off. Thus, controllers balked at issuing the speed advisory and instead tried to create a solution that would be completely implemented by the point at which they wanted to hand-off the aircraft.

4 Discussion and Conclusions

Participants completed their tasks and met the goals of the study relatively easily. The feeder controllers delivered a well-conditioned flow to the final sectors, and the final controllers merged aircraft and avoided excess spacing and wake spacing violations at the runways. They were assisted in their tasks by the support tools but the relatively low traffic load and lack of scenario complexity made the tasks easier to complete than anticipated. The lack of complexity was reflected directly in participants' low workload ratings and neither workload scale was sensitive enough to detect finer variations in controllers' workload reports that may have thrown light on whether some traffic was more difficult to manage than others and why.

Although the toolsets showed little relationship to controller performance (performance being high in all conditions), participants expressed a clear preference for the slot marker condition. They controlled the traffic by using the slot marker circles as a spatial control target with the early/late indicator, and then used the timeline for further fine-tuning. A comment controllers made more than once was that they would have liked the early/late indicator, which was displayed down to an ETA-STA error-precision of five seconds, to continue to be displayed until there was only one second of ETA-STA mismatch, i.e., until the aircraft was in its target position.

Despite being the most advanced tool offered in the study, the speed advisory was not perceived as more useful but actually as less so, and participants reported using it proportionately less than the other tools. This was because the advisories did not conform to controller norms and so controllers interpreted, rather than followed, them. Of course, interpreting advisories requires mental manipulation, which accounts for the similar levels of mental workload to other toolsets reported in this condition when it was expected to be lower. Because of this mismatch between controller work techniques, a concern that the more automated speed advisory tool may reduce controller understanding was not borne out. However, this is still potentially a hazard, as when the tool is fine-tuned to match controller strategies the potential for controllers to use the advisory without thinking may recur.

Lessons learned from this study that will carry forward into future controller managed spacing work are threefold. First, there is a need to refine the decision support toolsets, the most major being to consider allocating portions of the speed advisory on a sector-by-sector basis. These amendments have been developed and are presently being tested [13]. Second, consideration must be given to the need for controllers to maintain a deep level of understanding of the traffic that they are managing. Third, with respect to testing these more refined tools, it is possible that participants would lean on proffered tools more heavily if the scenarios were more complex or if larger disruptions to the schedule occurred.

In sum this simulation of merging terminal area arrival traffic showed that controllers, assisted by simple-to-use and informative decision support tools, were able to correct for initial schedule and wind forecasting errors and deliver aircraft on-schedule using just speed clearances. There were only small performance and workload differences when comparing between the DSTs. The preferred toolset was the slot marker toolset including timelines, slot marker circles and early/late indicators, which highlighted key information for the controllers but did not constrain their choice of strategies.

Acknowledgments. Thank you to the AOL development team and the controller subject matter experts without whom this study could not have been completed.

References

1. Joint Planning and Development Office: Concept of Operations for the Next Generation of Air Transportation System, Version 3.0. Washington, DC, October 1 (2009)
2. Coppenbarger, R., Dyer, G., Hayashi, M., Lanier, R., Stell, L., Sweet, D.: Design and test of automation for efficient arrivals in constrained airspace. In: Coppenbarger, R., Dyer, G., Hayashi, M., Lanier, R., Stell, L., Sweet, D. (eds.) International Congress of the Aeronautical Sciences, AIAA2006-774, Nice, France (September 2010)
3. Boursier, L., Favennec, B., Hoffman, E., Trzmiel, A., Vergne, F., Zeghal, K.: Integrating aircraft flows in the terminal area with no radar vectoring. In: Boursier, L., Favennec, B., Hoffman, E., Trzmiel, A., Vergne, F., Zeghal, K. (eds.) 6th AIAA Aviation Technology, Integration and Operations Conference (ATIO), Wichita, KS, September 25-27 (2006)
4. Isaacson, D.: Airspace Super Density Operations (ASDO) Concept of Operations. NASA Ames Research Center, CA (2007)
5. Shepley, J.: Near-term terminal area automation for arrival coordination. In: Eighth USA/Europe Air Traffic Management Research & Development Seminar, Napa, CA (2009)
6. Nunes, A.: The impact of automation use of the mental model: Findings from the air traffic control domain. In: Proceedings of the 47th Annual Meeting of the Human Factors and Ergonomics Society, Santa Monica, CA (2003)
7. Seamster, T., Redding, R., Cannon, J., Rider, J., Purcell, J.: Cognitive task analysis of expertise in air traffic control. International Journal of Aviation Psychology 3(4), 257–283 (1993)
8. Endsley, M., Rodgers, M.: Situation Awareness information requirements for en route air traffic control, DOT/FAA/AM-94/27. US DoT, FAA, Washington, DC (December 1994)
9. Sheridan, T.: Supervisory Control. In: Salvendy, G. (ed.) Handbook of Human Factors, pp. 1244–1268. John Wiley & Sons, New York (1987)
10. Prevôt, T.: Exploring the many perspectives of distributed air traffic management: The Multi-Aircraft Control System MACS. In: Chatty, S., Hansman, J., Boy, G. (eds.) HCI-Aero 2002, vol. 2002, pp. 149–154. AIAA Press, Menlo Park (2002)
11. Stein, E.S.: Air trafic controller workload: An examination of workload probe, DOT/FAA/CT-TN84124. Atlantic City International Aiport, NJ, FAA (1985)
12. Hart, S.G., Staveland, L.E.: Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In: Hancock, P., Meshkati, N. (eds.) Human mental workload, pp. 139–183. Elsevier, Amsterdam (1988)
13. Swenson, H., Thipphavong, J., Sadovsky, A., Chen, L., Sullivan, C., Martin, L.: Design and evaluation of the Terminal Area Precision Scheduling and Spacing System. In: Ninth USA/Europe Air Traffic Management Research & Development Seminar, Berlin, Germany (submitted)