THE FEASIBILITY OF MEASURING CAPACITY IN A REAL-TIME ATM SIMULATION INDEPENDENT OF SUBJECTIVE CONTROLLER WORKLOAD MEASUREMENT

Rod Gingell, EUROCONTROL, rod.gingell@eurocontrol.int, Brussels, Belgium

Kevin Corker and Erick Weber, San Jose State University, Kevin.Corker@sjsu.edu, Erick@erickweber.com, San Jose, California[1]

Abstract

We report on the results of an experiment on metric validation with a human-in-the-loop real-time ATM simulation. We then compare the results of this metric validation with a computational human performance model replication of the evaluation to assess the independence of the INTEGRA high level metrics from subjective controller workload measurements.

INTEGRA Project Goals

INTEGRA is a collaborative programme under the EUROCONTROL CARE initiative. The focus has been to harness real-time ATM simulations to obtain quantitative and objective assessments of system performance.

The primary objectives of the INTEGRA programme were:

- To quantitatively measure the benefits of the inclusion of advanced computer assistance tools in ATM systems. These benefits are measured in terms of Capacity, Safety and Efficiency with the results providing data for input to system designers and Cost Benefit Assessment (CBA) studies to support high-level management decisions.
- To provide metrics that can be deployed on simulations in EUROCONTROL and other establishments to give a common framework for the measurements allowing direct comparison of the validation results.

Real-time simulations of ATM systems involving advanced controller tools have taken place over a number of years and to a large extent the benefits expected from the tools have not been evident in the results. The rationale behind the advanced tools is that they should reduce the controller workload per aircraft thereby permitting the controller to handle more aircraft. This in turn would lead to an increase in capacity, while the monitoring and prediction capability of the tools would enable greater flight efficiency at the same time as maintaining or increasing the levels of safety. However, the controller workload was largely determined by analyzing qualitative self assessment data (questionnaires, rating scales etc.) and from the analysis of this data being performed the controllers perceived workload was not seen to be decreasing.

INTEGRA considered that real-time simulations are an essential step in the validation of new concepts and tools but recognized the limitations associated with the reliance on controller workload as a measure of system performance. These limitations arose from the fact that it was not feasible to train controllers for simulations to achieve the same level of experience and trust in the new system as they have with current systems.

INTEGRA, therefore, set out to develop measures for capacity that were quantitative and abstract in the sense that they were measured from events in the simulation and did not rely on controller self assessment. In parallel, metrics for safety and efficiency were developed as these three topics are interrelated. This paper concentrates on capacity.

Air MIDAS Project Goals

The use of the human performance modeling methodology is an effective means to study human performance in complex systems, especially in their early design phases. Human performance models characterize a human-system environment within a computational framework that predicts the performance of that system as an assessment mechanism. Many advantages exist for using human performance models in collaboration with human-in-the-loop methodologies. Specifically, the computational simulation can focus the costly

[1] San Jose State was supported by a research grant from the CARE INO programme.

human-in-the-loop simulation and can use the human-in-the-loop simulation data for model parameterization. Human performance models have been used to provide validated predictions of performance in complex operating environments ranging from highly advanced military systems [1], nuclear power plant operations [2], and advanced concepts in aviation [3]. The human performance modeling software tool, Air MIDAS, generates predictions of human operator performance in advanced alternative operational concepts. The purpose in the current study was to:

- Generate a human-system model of the advanced operational concept studied in the advanced controller tool INTEGRA study reviewed below,
- To compare the information processing and channel capacity values generated by the human performance model with similar metrics provided by INTEGRA in the assessment of human-in-the-loop studies.

INTEGRA Evaluation Trials

The INTEGRA methodology and metrics evaluation trials (the "INTEGRA trials") took place at NLR in Amsterdam between January and April 2004. The purpose of the exercise was to illustrate the application of the INTEGRA metrics and demonstrate that they reacted as expected to the different experiment configurations.

Experiment Configurations

The real-time simulation comprised two organizations, a baseline simulation without advanced assistance tools (ORG0) and a second simulation comprising a set of advanced assistance tools (ORG1).

The simulation facility was representative of an advanced, stripless ATC system. The term "stripless" refers to the fact that the system did not use conventional paper flight strips or on-screen equivalent representations. All flight plan information was accessible via track labels on the Plan View Display.

Each sector was operated by a controller team comprising a Tactical Controller (TC) and a Planning Controller (PC).

Airspace and Traffic Samples

A single measured sector was simulated. This was the Delta Upper Airspace sector (FL245-340), which is part of the Maastricht UAC. This represents an ATC sector as it is used in current day operations. The Delta sector controls en-route traffic to and from the adjacent sectors LONDON, COASTAL, MUNSTER, RUHR and MAASTRICHT.

The Traffic Samples used for the trials were created using the INTEGRA Traffic Sample Generator (TSG). The objectives of this generator are to readily develop pseudo-random but realistic traffic samples for given regions of airspace, to develop reference samples that can be used at different establishments and to enable the development of many statistically linked samples for use both in the various simulation runs and in training.

Operational Concept

The concepts and tools available for the two organizations are briefly described.

ORG0

ORG0 was designed to be representative of today's control paradigm.

The main components of the ORG0 facility are listed below:

- Plan View Display (PVD) and labels
- Sector inbound list (SIL)
- System Co-ordination (SYSCO)
- Safety Nets (SNET)
- Dynamic Flight Leg (DFL)

The roles of TC and PC were essentially the same as current control paradigms. The TC was the final Control Authority in the sector. The PC's role was to perform co-ordination with adjacent sectors and to assist the TC with the flow of traffic through the sector.

ORG1

ORG1 provided the controller with computer based tools that allow both for the detection of potential conflicts and for the editing of trajectories to support in the resolution of detected conflicts.

The ORG1 facility had all the same components as the ORG0 facility with the following additions:

- Monitoring Aids (MONA)
- Medium Term Conflict Detection (MTCD)
- Conflict Window (CW)
- Trajectory Editor

The advanced system implemented a new task division between the PC and the TC. As with current practice the TC remained the final Control Authority in the sector. The PC's role was to provide the TC with resolutions to potential conflicts within the sector. The PC made use of the Medium Term Conflict Detection and the Trajectory Editor to provide conflict resolution manoeuvres to the TC.

Subject Matter Experts (SMEs)

During the INTEGRA trials the role of the Subject Matter Experts (SMEs) was to provide both qualitative and quantitative information about the trials and the trial subject-groups to assist with subsequent analysis. This was achieved through observation and instantaneous assessment. An R/T monitor (listening mode) was provided for the SMEs so that they could develop a better understanding of the traffic situation.

During each run, the SME recorded observations against the following topics:

- Performance of scenario: Traffic Level
- Performance of subject-group: Competence (or compliance) with concept of operation; Competence with tools; Workload
- Significant incidents

Instantaneous assessment

Through use of an Instantaneous Assessment device, the SME recorded the task complexity at 2 minute intervals through the trial on a scale of 1 to 5. This score reflected the complexity of the situation over the period as judged by the SME rather than the perceived workload of the subject controllers. Each SME had a pair of the boxes, one to record the responses for the TC and one for the PC against the following scale:

1: No control activity
2: Low control activity
3: Frequent control activity
4: Constant control activity
5: Multi-tasking and prioritization of controller tasks

Capacity Metric

Overview

The INTEGRA Capacity metric operates on the following basis:

- The ATM system is described by a number of activities which are allocated to and performed by a number of actors. These actors are either human (controllers, pilots etc.) or machine processes (trajectory predictor, conflict probe etc.),
- To execute an activity an amount of processing has to be performed, this is termed Information Processing Load (IPL). The IPL for an activity can be calculated from data collected from the simulation,
- Each actor has a limit to the amount of IPL they can handle before becoming overloaded. The capacity of the system is reached when any one actor is overloaded.

To function on this basis there are some important requirements placed on the simulation namely:

- The simulation must have reached a level of maturity whereby it can be relied on to operate as designed,
- The tools must function as designed both individually and as a consistent set,
- The HMI must have been optimized for the task in terms of functionality and performance,
- The roles of the individuals and the associated operating procedures must be well defined,
- The assigned roles and procedures must be rigorously adhered to during the simulation runs.

IPL

IPL, its derivation and the determination of the maximum IPL an actor can handle is the core element of the Capacity metric. The IPL is calculated for each activity allocated to a given actor every timestep[2] throughout the simulation. The total IPL for each timestep is determined by summing the individual contributions. The value for each timestep is compared against a threshold to determine when the actor is overloaded, i.e. at capacity.

This threshold is determined through the Calibration process described below.

IPL and Workload

INTEGRA makes a clear distinction between IPL and workload when addressing human actors. This is because workload is a subjective measure and can be influenced by a number of factors[3] when dealing with new and unfamiliar systems such as those relating to advanced tools. Workload is a

[2] 2 minutes in the case of the INTEGRA trials.
[3] For example insufficient training or lack of trust in the tools.

reliable indicator for current systems as there is a wealth of historic data for linking it to the operations being performed. This is not the case, by definition, for advanced tools with aspects such as lack of trust in the tools manifesting itself by the controller checking the tools' output and hence generating higher than necessary workloads. IPL on the other hand measures only the processing load necessary to perform the specific activity and is not subjective.

Metrics Configuration

The first step in the deployment of the metric is to identify all the actors in the simulation and their respective roles. The second step is to allocate the activities to each of the roles. Table 1 illustrates this process for the INTEGRA simulation with the seven tasks identified on the left-hand side of the table and the actors for ORG0 and ORG1 identified in the column headings.

Table 1 – Allocation of activities to roles

		Actors				
		ORG0		ORG1		
		TC	PC	TC	PC	Systems
IPL Task	Flight Arrival	✓	✓	✓	✓	
	Interaction Detection	✓	✓			✓
	Resolution Planning	✓			✓	
	Resolution Implementation	✓		✓		
	Monitoring	✓		✓		✓
	Other Trajectory Changes	✓		✓		
	Co-ordination		✓		✓	

Parameters

The IPL for each of the seven identified activities is calculated from the recorded data. The algorithm allocates an amount of IPL for each activity which is dependent on the traffic situation. For example, the IPL allocated to the planning of a resolution for a potential conflict is dependent on the "difficulty" of the resolution. The "difficulty" is determined by factors such as the density of the aircraft in the area of the potential conflict and the closeness to a sector boundary. Parameters are available to weight the IPL of an activity relative to the others. The use of these parameters would be a value judgement if it was considered that some aspect of a particular simulation resulted in an increased "difficulty" of a specific activity.

Processing the Data to Determine IPL

Following the simulation runs the recorded data was processed to determine the total IPL for each timestep for each actor. To determine the potential capacity of the simulated environment it is necessary to compare the calculated IPL with the IPL limit of the respective actor.

Calibration

One of the objectives of the INTEGRA trials was to calibrate the IPL for the various actors. Calibration, in the context of INTEGRA, is the determination of the level of IPL that represents the maximum sustainable load for a particular actor. During the INTEGRA experiments Subject Matter Experts (SMEs) were used to provide data that could be used for the calibration activity.

It is necessary to perform calibration for any new human role as the threshold is likely to be different for different roles, dependent on the tasks being performed etc. However, the calibration of a role will be valid for this role irrespective of the simulation being performed. This is valid if the role is comprised of the same activities or a subset of the activities. For example the calibration performed as a result of these simulations will be valid for En-route Planning and Tactical controller roles but a re-calibration would be required for TMA controllers.

SME

The SME role has been described previously. The data collected was used to calibrate the IPLs from the Tactical and Planning Controllers in the baseline organizations (ORG0). The process applied was to map the SME score to the measured IPL using a "curve fitting" routine that is described in the next section.

SME/IPL Curve Fitting

R^2 linear regression was performed to match the SME score to the Capacity Metric IPL for both the Tactical and Planner actors. For each ORG0 simulation run, one hour's worth of data was taken by excluding the first twenty minutes and the last ten minutes. For each measured hour, there were 31 timesteps (of two minutes). For each timestep, there was one SME score value per actor and there was a Capacity Metric IPL value (sum of seven individual task values).

The curve fit was performed using all of the output for an actor and evaluating the formula ($y = m.x + c$) that matches the curve between the two types of data. The results were as follows:

Tactical: SME = 0.0275(IPL) + 2.1987
OR IPL = 36.324(SME) - 2.1987

Planner: SME = 0.0322(IPL) + 2.0787
OR IPL = 31.066(SME)- 2.0787

From these formulae, calculations of predicted IPL to match maximum SME values can be made and vice versa. For example, for the Tactical, a maximum SME Score of 5 means an IPL value of 102, whilst for the Planner it is 91 (both rounded to zero decimal places). The SME Score is limited to whole numbers from 1 to 5, so this limited scale means that the curve fit has an error of +/-0.5 on the SME Score, so an absolute maximum for IPL is calculated from using an SME Score of 5.5. Hence, for the Tactical actor this maximum IPL is 120 and for the Planner actor this maximum IPL is 106 (both rounded to zero decimal places). One further limitation to the range of the SME Score was that during the analyzed hour of a measured run, the values 1 and 2 were rarely used, hence the intercept value (c) in the formulae being a value above 2.

From the calculated SME/IPL formulae, the differences between the calculated values (i.e. predicted) and the actual values could also be analyzed (called residuals). 48% of the residual values were between (+/-0.5) and 80% between (+/-1.0).

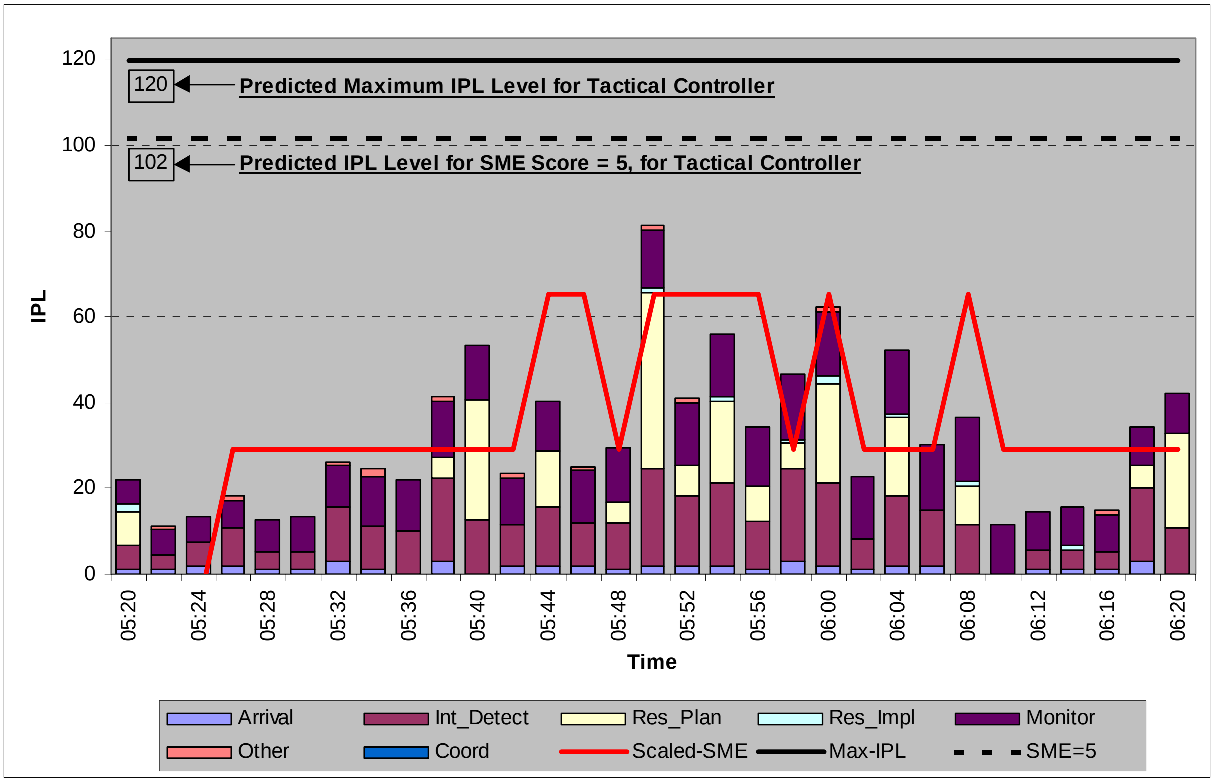

Figure 1 – Capacity Metric: Tactical – ORG0 – Run_20_A (Medium 18)

Calibration for a Non-Human Actor

Capacity of the ATM system is reached when any one of the actors (human or machine processes) reaches its processing limit. The process for determining this limit for a human actor has been described above. In the case of non-human actors the process is more straightforward. A recognized approach for specifying a loading factor has been adopted, in this case the value of 50% of processor load has been used as the desired maximum. It was seen that processor load, even for the higher traffic loads was a long way below this threshold. In fact the figures show that even if all the processes are added together the limit is not reached.

Air MIDAS Simulation Trials

We have produced a computational model of the current and advanced system and the human operators using it in order to test the procedures and information processes inherent in the concept from the perspective of human performance. The model provides a representation of the sources of variability in procedures, in the environment, and the operators to allow us to predict the response of the system, the individual, and the resource elements (aircraft etc).

The procedures modeled are the activities to be performed as a function of traffic load and assumed equipment suite. We modeled both current, ORG0, operations and the procedures assumed to be associated with the advanced assistance tools, ORG1.

Air MIDAS Data

The standard Air MIDAS [4] data sets were generated which include for the human operator the activities performed, their beginning and end time, the number interrupted and resumed, the number aborted, the decisions made, the control actions taken and the load on the operator for the visual, auditory, cognitive and motor (VACM) requirements of the tasks over the simulation runs. These loads are conceptually analogous to the Information Processing Loads (IPL) in the INTEGRA metrics and will be used for direct model-metric to human INTEGRA metric comparison.

It is worth noting however, that the Air MIDAS system uses the VACM loads as a constraint in scheduling performance. That is to say, if the loads exceed a threshold, the task associated with those loads will be deferred (on a priority-based scheduling basis) to a later performance. As will be shown, the Air MIDAS IPL loads - while following the pattern of the human INTEGRA IPL loads - are generally lower than those calculated on the human performance. This reduced overall level of Air MIDAS loads may be due to the model's fundamental scheduling process that keeps loads below an exceedence threshold.

The average VACM loads for the Air MIDAS Controllers across the conditions are provided in Table 2.

Table 2 - Average Air MIDAS processing loads for ORG0 & ORG1 runs in high traffic

VACM				
ORG0	**Visual**	**Auditory**	**Cognitive**	**Motor**
TC	4.1	1.4	6.0	1.0
PC	4.6	1.2	2.7	0.8
ORG1	**Visual**	**Auditory**	**Cognitive**	**Motor**
TC	3.6	1.7	2.7	0.9
PC	3.7	0.2	3.1	1.3

The data show the same pattern of controller workload as that illustrated in the INTEGRA Execution Phase report [5]. The workload for the tactical is reduced overall in the ORG1 condition as compared to the ORG0. The workload for the planner controller increases somewhat specifically with respect to the cognitive and motor elements of the load. Visual scanning is reduced for both controllers consistent with the reliance on the MTCD and alert function in ORG1.

The time line of loads for Air MIDAS follows the standard traffic increase and decrease pattern. The Air MIDAS loads also show a small variation (decrease) as you move from high to medium traffic.

INTEGRA Air MIDAS Metrics

The second output of the Air MIDAS system in the INTEGRA analysis is the result of using the Air MIDAS generated activity set as if it were input into the INTEGRA metrics from human operators. In these analyses, the Air MIDAS model is essentially functioning as a "fast time" simulation of actual controller behavior. These analyses will allow us to make a direct comparison of the model-generated performance compared with the human-generated performance as INTEGRA metrics.

In this report we will provide the Information Processing Load (IPL) data. In Figure 2 and Figure 3, we provide the IPL data across the simulation runs for ORG0 and ORG1 Tactical and Planning Controllers at medium traffic load.

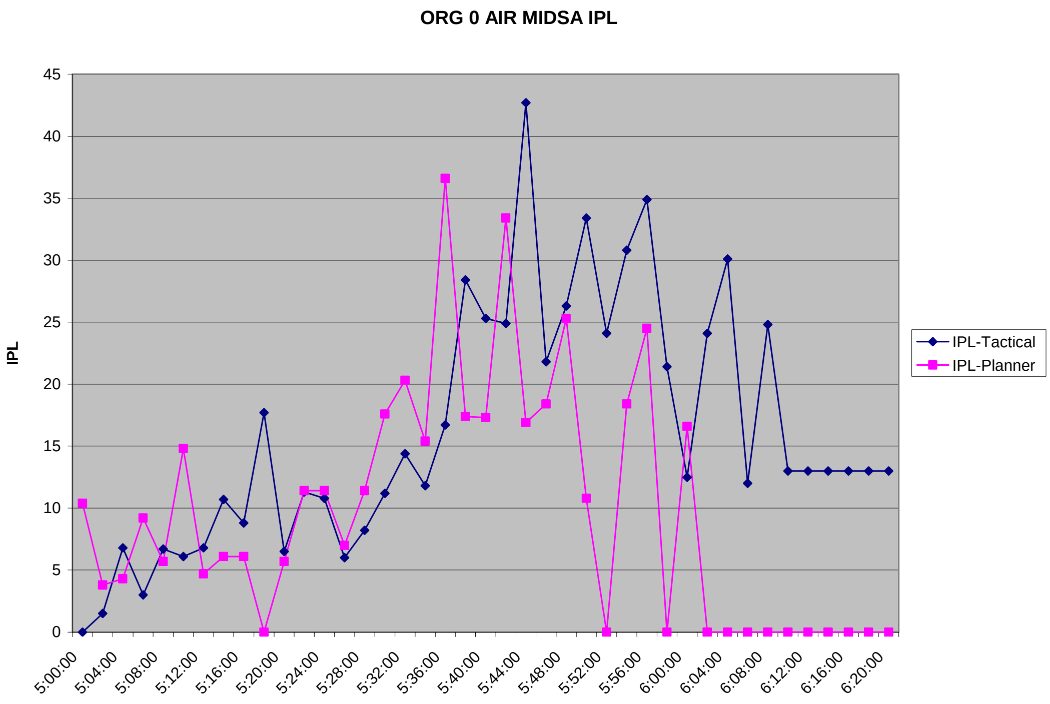

Figure 2 - Air MIDAS IPL Load in ORG0 operations

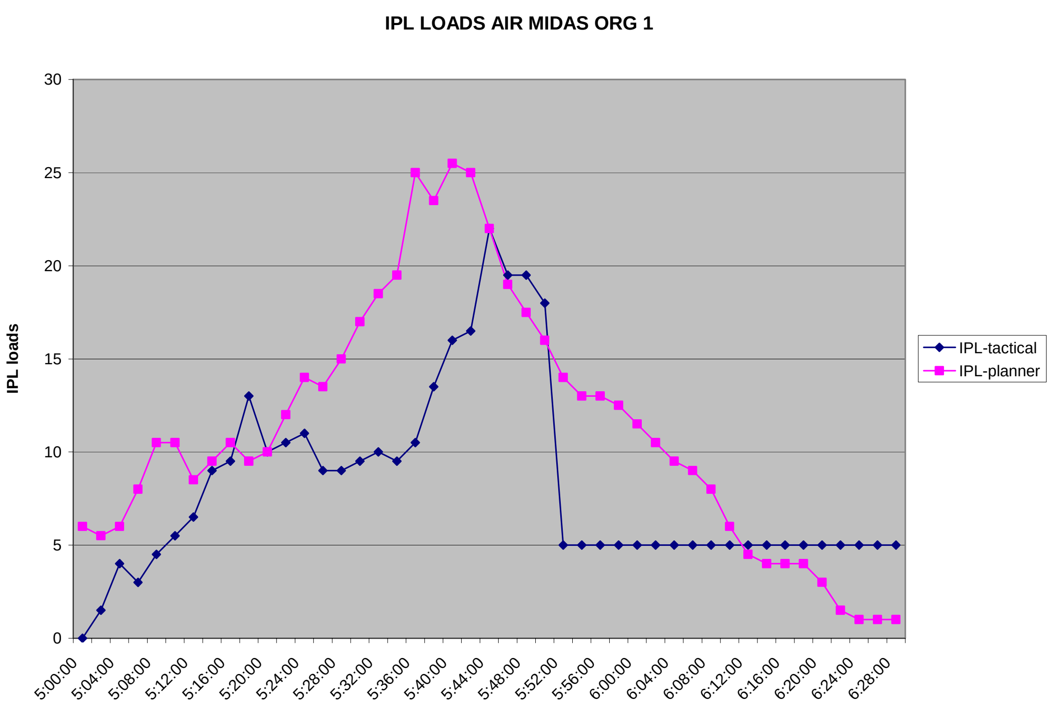

Figure 3 - Air MIDAS Operations in ORG1 operations

Air MIDAS INTEGRA IPL Data follow the same pattern in loading as the human-generated data for the same scenario. The peak level of the IPL is lower than that reported by the INTEGRA team, but the pattern of response in time and the inversion of load, with increase in Planning Controller loads under ORG1 is the same as reported by the INTEGRA team.

Summary Assessment

This analysis was run in order to see if the fast time simulation data generated by a computational model would be compatible with the results from the INTEGRA metric process. The comparability of the results supports one of the INTEGRA goals namely that the capacity metrics can derive results from a real-time simulation without recourse to assessing subjective controller workload.The Air MIDAS model assumes 'perfect' human performance i.e. operating the system in accordance with its designwhich is the basic concept underpinning the INTEGRA metric which is attempting to eliminate the influence of controller lack of trust or experience with the system.

We did not measure the non-human information processing loads for the advanced systems and so cannot comment on their values. However, since the air MIDAS system is designed to treat information processors (be they human or machine) using similar structures, there is no reason to assume that the non-human elements would not match the INTEGRA IPL values for this equipment.

One note, however, is that the calibration of information processing load with respect to controller load, both in Air MIDAS and in the INTEGRA approach is based on subject matter expertise. If a concept varies too far from that expertise, then the maximum load estimates for both systems may be suspect. Also, IPL is assessed in INTEGRA for a set of activities see Table 1. A question that arises is how much different any new system is with respect to these activities and how often the activities contributing to the IPL need to be reassessed. ORG0 and ORG1 were both En-Route environments and hence the activities were similar, albeit allocated differently between the Tactical and Planning Controller. However, if the tasks are fundamentally different, say in a TMA environment, then the activities contributing to the IPL for INTEGRA and the task structure for air MIDAS would need to be changed.

INTEGRA Conclusions

The evaluation trials have shown that the INTEGRA methodology has a number of benefits, over existing methods, when assessing simulations of advanced ATM systems. These can be summarized as follows:-

- Standardized, non-subjective, measurements that are not dependent on controller familiarity with or trust in the system,and are independent of any particular simulation platform permitting direct comparison between simulations performed at different establishments.
- Determination of quantitative data for capacity safety and efficiency rather than subjective data, resulting in data that can be utilized in the development of CBAs and for business cases to support management decisions.

References

[1] Atencio, A. , Shively, R.J., & Shankar, R. (1996).*Evaluation of Air Warrior Baselines in a Longbow Apache Helicopter Crewstation in a MIDAS Simulation*.American Helicopter Society 52nd Annual Forum, Washington D.C.

[2] Corker, K. *"Models of Human Information Requirements: When reasonable aiding systems disagree"* (invited address) Topical meeting of the American Nuclear Society. Philadelphia, Pa. June 25-29, 1995.

[3] Corker, Gore Fleming & Lane (2000). *Free Flight and the Context of Control: Experiments and Modeling to determine the impact of distributed air-ground air traffic management on safety and procedures*. 3 rd FAA EUROCONTROL International Symposium on Air Traffic Management.

[4] Corker, K. (2000). *Cognitive Models & Control: Human & System Dynamics in Advanced Airspace Operations*. (2000) in N. Sarter and R. Amalberti (Eds.) Cognitive Engineering in the Aviation Domain. Lawrence Earlbaum Associates, New Jersey.

[5] Gingell, Rod, Chris Strachan, Andy Taylor, Steve Kinnersly, Sarita Fox, 2004, *INTEGRA Metrics & Methodologies, Execution Phase – Final Report*, EUROCONTROL, www.eurocontrol.int/integra

Key Words

Metrics, Air Traffic Management Aiding Human Performance Model, Quantitative Analyses, Validation and Verification, Information Processing Loads.

Biographies

Kevin M. Corker

Kevin Corker is currently an Associate Dean of the College of Engineering and a Professor in the Industrial and Systems Engineering Department, San Jose State University. He is Director of the Human Automation Integration Laboratory (HAIL) at SJSU. He received his Ph.D. in a joint program in Cognitive Psychology and Engineering Systems from the University of California, Los Angeles. His major areas of research interest include computational cognitive modeling, human performance, human-system measurement methodologies, human-automation system integration and issues in control and display in large-scale dynamic

Rod Gingell

Rod Gingell is currently the EUROCONTROL INTEGRA Technical Manager. He is an ATM consultant with over 20 years' experience in technical consultancy and project management in the defence and Air Traffic Management domains. He has particular expertise in the area of ATM concept evaluation, real-time simulations and integration, gained through work for CAA/NATS, EUROCONTROL and the European Commission.