The Evolution and Implementation of Workload Algorithms in IPME

Mr. Joe Armstrong; Ms. Michelle Gauthier; Mr. Gerald Lai
CAE Professional Services
1135 Innovation Drive, Suite 200
Kanata, ON
K2K 3G7

Dr. Andy Belyavin; Mr. Chris Ryder
QinetiQ Limited
Cody Technology Park, Cody Building
Ively Road, Farnborough, Hampshire,
GU14 0LX,
United Kingdom

Mr. Brad Cain
Defence Research and Development Canada Toronto
Human System Integration Section
Human Modelling Group
1133 Sheppard Avenue West
Toronto, ON
M3M 3B9

Workload Algorithms Validation for the Integrated Performance Modeling Environment

Project Manager: Joe Armstrong [+1 613 247 0342]

Contract number: W7711-057962/001/TOR.

Scientific Authority: Wenbi Wang; DRDC Toronto, [+1 416 635 2000 ext. 3063]

This page is intentionally left blank.

Defence R & D Canada - Toronto

Contract Report CR 2006-042
DRDC Toronto
September 2008

Author

Joe Armstrong

Author

Michelle Gauthier

Author

Gerald Lai

Approved by

Wenbi Wang
Scientific Authority

Approved for release by

Kim Wulterkens
For Chair, Document Review and Library Committee

Terms of Release: The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

Abstract

This report documents a study to validate predictive workload models that are available within the Integrated Performance Modeling Environment (IPME). A literature review was conducted to assess the current state of knowledge of human workload and information processing, as well as to provide a review of the five IPME workload algorithms (Visual, Auditory, Cognitive, and Psychomotor (VACP), Workload Index (W/Index), Information Processing/Perceptual Control Theory (IP/PCT), Prediction of Operator Performance (POP), and Prediction of Operator Performance Information Processing (POPIP)). The results of the literature review indicated that, while the theories associated with human information processing are relatively mature, the predictive models of human workload integrated within IPME still require validation against human performance data. Analytical and empirical studies were then conducted within a combined Air Traffic Control (ATC) and Visual Bakan dual-task paradigm. The POP and POPIP analytical models more accurately predicted human subjective workload than did VACP and IP. The IP and POPIP analytical models predicted human performance in the Visual Bakan more accurately than did VACP and POP. All models were equally inaccurate in predicting ATC performance. Theoretical accounts of findings and practical implications for model development are discussed.

Résumé

Le présent rapport documente une étude visant à valider des modèles prédictifs de charge de travail qui sont disponibles dans un environnement intégré de modélisation des performances (IPME). Un examen de la documentation a été effectué pour permettre d'évaluer l'état actuel des connaissances de la charge de travail humaine et du traitement de l'information, ainsi que de revoir les cinq algorithmes de charge de travail IPME (VACP, W/Index, IP/PCT, POP, and POPIP). Les résultats de cet examen ont indiqué que si les théories associées au traitement de l'information humaine sont relativement à point, les modèles prédictifs de charge de travail intégrés au sein de l'IPME doivent toujours être validés par rapport au données de performance humaine. Des études analytiques et empiriques ont alors été menées en fonction du paradigme à double tâche combinée contrôle de la circulation aérienne (ATC) et tâche visuelle Bakan. Les modèles analytiques POP et POPIP ont prédit avec plus de précision la charge de travail subjective de l'être humain que ne l'ont fait les modèles VACP et IP. Les modèles analytiques IP et POPIP ont prédit la performance humaine pour une tâche visuelle Bakan plus précisément que ne l'ont fait les modèles VACP et POP. Tous les modèles ont été également imprécis dans la prédiction de la performance ATC. Les comptes rendus théoriques des conclusions et des conséquences pratiques pour l'amélioration des modèles sont présentés.

Executive Summary

This report documents the final phase of work conducted by CAE Professional Services (CAE PS) in support of PWGSC file number W7711-057962/001, titled Workload Algorithms Validation for the Integrated Performance Modelling Environment (IPME). This work was completed under contract to Defence Research and Development Canada (DRDC)-Toronto.

A literature review was conducted to assess the current state of knowledge with respect to the concepts of human workload and information processing, as well as provide a review of the five workload algorithms (VACP, W/Index, IP/PCT, POP, and POPIP) implemented within the Integrated Performance Modeling Environment (IPME). The results of the literature review indicated that, while the theories associated with human information processing are relatively mature, the predictive models of human workload integrated within IPME still require validation against human performance data. To focus the design and conduct of future validation efforts, a series of studies were conducted within a combined Air Traffic Control (ATC) and Bakan task paradigm.

In order to validate the workload algorithms within IPME, the following four questions were raised:

1. Are the five workload algorithms correctly implemented in IPME?
2. Are the correct data outputs obtained from each algorithm meeting the expectations of the theoretical constructs?
3. Is the impact of the scheduling algorithms (POP, IP/PCT and POPIP) representative of human operator performance when undertaking comparable tasks?
4. Are there differences between model performance across IPME V3 and IPME V4 for POP, POPIP, and IP/PCT?

Initial comparison of the operator workload values from the model and the observed values from a pilot study indicated that the form of the ATC model was incorrect. A suggestion for a modification to the model was submitted, approved and implemented. During the model development and testing phase, two areas were identified where the W/Index, IP/PCT and POP algorithms were not implemented correctly in IPME 3.0.25. An exploration of these issues in the newest version of IPME (4.1.3) and an update in IPME 3.0.30 demonstrated that these features had been corrected. Findings in predictive workload for the POP and POPIP models showed that they more accurately predict human subjective workload compared to VACP. Findings from the ATC and Bakan performance showed that IP and POPIP models predicted human performance in the Bakan more accurately than VACP and POP. All models were equally inaccurate in predicting ATC performance. Theoretical accounts of findings and practical implications for model development are discussed.

Sommaire

Le présent rapport documente la phase finale du travail menée par les Services professionnels de CAE (SP CAE) à l'appui du contrat de TPSGC, numéro de dossier W7711-057962/001, intitulé Validation des algorithmes de charge de travail pour un environnement intégré de modélisation des performances (IPME). Ce travail a été effectué à contrat pour Recherche et développement pour la défense Canada - Toronto (RDDC)-Toronto.

Un examen de la documentation a été effectué pour permettre d'évaluer l'état actuel des connaissances relativement aux concepts de charge de travail humaine et de traitement de l'information, et de revoir les cinq algorithmes de charge de travail (VACP, W/Index, IP/PCT, POP, and POPIP) appliqués dans un environnement intégré de modélisation des performances (IPME). Les résultats de cet examen ont indiqué que si les théories associées au traitement de l'information humaine sont relativement à point, les modèles prédictifs de charge de travail intégrés au sein de l'IPME doivent toujours être validés par rapport au données de performance humaine. Pour mettre l'accent sur la conception et la conduite des futurs efforts de validation, on a mené une série d'études en fonction du paradigme de la double tâche combinée contrôle de la circulation aérienne (ATC) et tâche visuelle Bakan.

Afin de valider les algorithmes sur la charge de travail dans le cadre d'un IPME, les quatre questions suivantes ont été posées :

5. Les cinq algorithmes de charge de travail sont-ils correctement appliqués dans le cadre d'un IPME?
6. Est-ce que les données correctes résultant de chaque algorithme répondent aux attentes des constructions théoriques?
7. L'impact des algorithmes d'horaires (POP, IP/PCT and POPIP) est-il représentatif de la performance humaine de l'opérateur lorsque des tâches comparables sont exécutées?
8. Y a-t-il des différences entre la performance des modèles POP, POPIP et IP pour l'IPME V3 et l'IPME V4?

La comparaison initiale des valeurs de charge de travail d'un opérateur à partir du modèle et des valeurs observées à partir d'une étude pilote a indiqué que la forme du modèle ATC était incorrecte. Une suggestion de modification du modèle a été présentée, approuvée et appliquée. Au cours de la phase d'amélioration et d'essai du modèle, deux secteurs ont été identifiés où les algorithmes W/Index, IP/PCT et POP n'avaient pas été appliqués correctement. Une exploration de ces questions dans la nouvelle version d'IPME (4.1.3) et une mise à jour dans l'IPME 3.0.30 ont démontré que ces caractéristiques avaient été corrigées. Les conclusions de la charge de travail prédictive pour les modèles POP et POPIP ont montré qu'ils prédisent avec plus de précision la charge de travail subjective de l'être humain comparativement au modèle VACP. Les conclusions de la performance ATC et Bakan ont montré que les modèles IP et POPIP

ont prédit la performance humaine dans la tâche Bakan avec plus de précision que les modèles VACP et POP. Tous les modèles ont été également imprécis à prédire la performance en ATC. Les comptes rendus théoriques des conclusions et des conséquences pratiques pour l'amélioration des modèles sont présentés.

Table of Contents

Table of Figures

Table of Tables

1 Introduction

The intention of this report is to document the final phase of work conducted by CAE Professional Services (CAE PS) in support of PWGSC file number W7711-057962/001, titled Workload Algorithms Validation for the Integrated Performance Modelling Environment (IPME). This work was completed under contract to Defence Research and Development Canada (DRDC)-Toronto.

1.1 Background

Of key interest to the Canadian (CA) defence and Human Factors (HF) community is the ability to develop computational models of human behaviour that operate within complex systems to compare systems performance, evaluate design alternatives for immersive and real system simulations, and predict human performance and workload prior to virtual and field-based trials of real systems (Armstrong & Lai, 2005; Armstrong & Youngson, 2004; Armstrong & Greenley, 2003; Armstrong, Brooks, & Barone, 2003). Additional research is being conducted on the efficacy of replacing human operators with human behaviour models in virtual simulations. The application of Human Behaviour Representations (HBRs) within these environments allows designers to predict system performance during development without expending the associated costs of developing complex human-in-the-loop simulations for predictive analysis.

Task network models (TNM) have been applied to the analysis of complex human-machine systems for a number of years. These models are typically used to generate estimates of task completion times, task accuracy, predictions of operator workload and task load, operator and system performance. A core assumption of the TNM paradigm is that human behaviour can be modelled as a set of interrelated tasks. The data used to drive the performance of the model (e.g., time, cognitive workload values, etc.) is assigned by a human factors analyst based on an understanding of the interaction between the operator and a specific system component and, whenever possible, empirical data.

The Integrated Performance Modelling Environment (IPME) is the most pervasive of the available TNM modelling applications currently being applied to the analysis and prediction of human behaviour within the CA defence community. IPME is a discrete-event simulation software for developing models that simulate human and system performance. It has been developed under the joint effort from Canada, the United Kingdom and the United States. IPME contains algorithms for predicting workload and the effects of performance shaping factors. Five workload models have been integrated within IPME to predict operator workload and the effects of internal and external performance shaping factors on task performance, i.e., the Visual, Auditory, Cognitive and Psychomotor (VACP), Workload Index (W/Index), Information Processing / Perceptual Control Theory (IP/PCT), Prediction of Operator Performance (POP) and Prediction of Operator Performance and Information Processing (POPIP) algorithms. These algorithms are all capable of predicting operator workload and all of them have been applied in various studies before. However, due to the different theoretical perspectives and underlying assumptions, these algorithms differ significantly in terms of the input, i.e., information needed to be fed into the algorithm, and the output i.e., workload prediction.

1.2 Workload validation

To date there has been a lack of extensive validation of the workload models within IPME from an independent source that is not associated with the development of the software. This raises concerns within the modelling community that the workload predictions from IPME may not be representative of human performance if the workload algorithms represented within IPME are not valid. The lack of objective validation becomes increasingly problematic as workload algorithms are now being combined in a single approach (i.e., in the POPIP algorithm), which may increase the difficulty in determining where the source of errors may reside. The source of the task demand data integrated within each algorithm is also a source of concern as their derivation is associated with subjective measurement techniques that also have questionable validity.

Consequently, there is a requirement to ensure that the workload algorithms within IPME are accurately modeled and are producing reliable and valid results. To focus the design and conduct of future validation efforts, a literature review was conducted to assess the current state of knowledge with respect to the concepts of human workload and information processing, as well as provide a review of the five workload algorithms implemented within IPME (see Forbes et al., 2006). The results of the literature review indicated that, while the theories associated with human information processing are relatively mature, the predictive models of human workload integrated within IPME still require validation against human performance data. Recommendations were made to establish a paradigm to facilitate the validation of the predictive workload models within IPME.

An experimental plan to validate the workload algorithms within IPME has been developed based on the results of the literature review (see Tryan et al, 2006). It is our intention in this study to systematically compare and validate the workload algorithms within IPME. Consequently, a test scenario was created and IPME models were developed using these workload algorithms. A laboratory experiment was then conducted and the results of the study were used as a benchmark for evaluating the predictions generated from IPME models. It is our goal to identify the strengths and weaknesses of each workload algorithm and generate guidelines for modeling human behaviour using IPME.

1.3 Objective

The primary objective of this report is to document the results of a study that was conducted to validate the workload algorithms within IPME.

1.4 Scope

A study was conducted to validate the workload algorithms within IPME and includes the following sections:

1. The experimental design of the validation trials that were conducted using an Air Traffic Control (ATC) task as the primary task and a Visual Bakan secondary task;
2. The methodology for the development of an IPME task network model to test the workload models;
3. The data analysis and results of the study; and

4. A discussion of the results and recommendations on how they can be used to generate guidelines for modeling human behaviour using IPME.

1.5 Relationship to other documents

The following documents are directly relevant to this report.

1.5.1 A literature review on the Evolution and Implementation of Workload Algorithms of Human Information Processing in IPME (Forbes, Darvill, Armstrong, & Banbury, 2006).

This document reports on a literature review that was conducted to assess the current state of knowledge with respect to the concepts of human workload and information processing, as well as provide a review of the five workload algorithms (VACP, W/Index, IP/PCT, POP, and POPIP) implemented within IPME. Recommendations are made to establish a paradigm that will facilitate the validation of the predictive workload models within IPME.

1.5.2 An experimental plan for the validation of the workload algorithms within IPME (Tryan, Armstrong, Ryder, & Belyavin, 2006).

This document details an experimental plan to validate the workload algorithms within IPME. This plan is based on the results of a literature review which assessed the concepts of human workload and information processing as well as the five workload algorithms implemented within IPME. The proposed experimental plan includes:

1. The experimental design of the validation trial to be conducted using an Air Traffic Control (ATC) task as the primary task and a Visual Bakan secondary task;
2. The subjective and objective performance measures to be collected and the data analysis methods to be employed; and
3. The methodology for the development of an IPME task network model to test all workload models.

2 Methodology and Procedures

2.1 Research Questions

In order to validate the workload algorithms within IPME, the following questions need to be answered:

1. Are the five workload algorithms correctly implemented in IPME?
2. Are the correct data outputs obtained from each algorithm meeting the expectations of the theoretical constructs?
3. Is the impact of the scheduling algorithms (POP, IP/PCT and POPIP) representative of human operator performance when undertaking comparable tasks?
4. Are there differences between model performance across IPME V3 and IPME V4 for POP, POPIP, and IP?

2.2 Human-in-the-Loop Experiment

2.2.1 Participants

Twelve students, men and women, from Carleton University inexperienced in air traffic control were recruited to participate in the study. Participants were screened for normal or corrected-to-normal vision. All participants were fluent in reading and writing in English and had at least 2 years of any computer experience.

2.2.2 Recruitment Process

Participants were recruited with an announcement for participation posted on the Carleton University Recruitment Board and via email. See Appendix A. Participants provided informed consent indicating that they could withdraw from the study at any time without prejudice.

2.2.3 Remuneration

Participants were selected on a voluntary basis. Each participant was compensated $125 for their participation in this study. Compensation was not dependent on the participant completing the task.

2.2.4 Material and Apparatus

The experiments took place at CAE Professional Services located at 1135 Innovation Drive, Suite 300, Kanata, ON, K2K 3G7. The participants performed the sessions within an enclosed office at CAE Professional with windows looking out into the main office area and overlooking the outdoors which will provide natural light in addition to indoor lighting. The simulator was run on a MacIntosh Power PC G3 with 512MB Ram, running OS X V10.4.4, with a 17-inch Display set to a screen resolution of 800x600 at 16-bit colour depth, and placed on a standard office desk. Participants were seated approximately 60cm from the display.

2.2.5 Experimental Tasks

Participants were required to perform two experimental tasks: a primary Air Traffic Control (ATC) task and a secondary Visual Bakan Vigilance task. In the dual task condition, the participant was asked to perform the Visual Bakan task in conjunction with the ATC task.

2.2.5.1 Air Traffic Control Task.

An Air Traffic Control (ATC) Simulation program was developed to support the validation program. The objective of the ATC task is to route each aircraft to a specified destination before the aircraft's airtime runs out. The operator must monitor the entry times and exit times presented on the program's Air Traffic Schedule and change the altitude and heading of the aircraft as required. There are two windows presented side by side, a radar window and an air traffic schedule window. See Appendix B for screen shots of the two windows. Two levels of workload (low and high) were set for the ATC task by manipulating the screen update interval (6 and 9 seconds) in the ATC simulation.

At the beginning of the experiment, aircraft were periodically added to the display, entering at the edge of the radar display randomly at one of the cardinal points, until five aircraft were present. The number of aircraft was then maintained by adding a new aircraft shortly after an aircraft left the display. For further information, see the experimental plan referenced in Section 1.5.2.

2.2.5.2 Visual Bakan Vigilance Task.

A Visual Bakan task was measured in isolation and was also presented as a secondary task during the conduct of the ATC simulation as an additional index of workload and driver of task demands. In the isolated Visual Bakan task, the participant was asked to attend to a series of random, single digits (between 0 and 9) that were continually displayed in the centre of the screen, subtending a visual angle of approximately 1.9^{o} x 2.4^{o}. Subjects were instructed to press a keyboard button (the spacebar) when an odd-even-odd sequence of numerical digits had been displayed (target stimulus). Each digit was displayed on the screen for 500 ms with an inter-stimulus interval (ISI) of 1500 ms (see Figure 1), with a 150ms overlap in responses between target strings. Target sequences appeared randomly, and with a frequency of approximately 60-70 times across a 15 minute trial.

Example sequence: 1 3 6 2 5 1 7 9 1 1 4 3 5 4 6 0 8 0
Correct response: ↑

Figure 1 Example of a stimulus sequence in the Bakan vigilance task

2.2.5.3 Dual ATC - Bakan Vigilance Task.

Finally, participants completed three 15-minute trials of the ATC task while simultaneously performing the Visual Bakan task. The dual-task approach provided a concurrent task environment whereby participants must manage the conduct of aircraft in the ATC task while responding to Visual Bakan target strings. The overall configuration of the ATC and Visual Bakan task is shown in Figure 3a in Appendix B.

2.2.6 Performance Measurements

Performance measurements were recorded and collected for all trials and conditions and were used to compare equivalent datasets generated from IPME for validation purposes. The performance measurements included the following:

2.2.6.1 Air Traffic Control Performance Measurement.

The performance measure that was collected throughout the ATC task is shown in Table 1.

Table 1 ATC task performance measures

Performance Measure	Description
Proportion of Correct Exits	Ratio of the number of correctly routed aircraft over the total number of aircraft exits.

2.2.6.2 Visual Bakan Performance Measurements.

Mean response time for successful detections of target signals, percentage of target signals detected, and percentage of false alarms (see Table 2).

Table 2 Bakan task performance measures.

Performance Measures	Explanation
Percentage of correct responses	The number of times the subject presses the spacebar when a target signal was presented over the total number of target signals presented.
Percentage of false alarms	The number of times the subject presses the spacebar when no target signal was presented over the total number of target signals presented.

2.2.7 Subjective Measurements

2.2.7.1 Workload Profile

Participants were asked during the Pilot Study to complete a paper-based Workload Profile questionnaire. The Workload Profile method for subjective workload assessment is designed with the notion that the resource dimensions put forth in Wickens (1987) multiple resource model can be used to describe the workload dimensions of a task. The workload dimensions are a representation of the task demands, which include perceptual/central processing, response selection and execution, spatial processing, verbal processing, visual processing, auditory processing, manual output, and speech output. The Workload Profile requires the participant to rate the proportion of attentional resources used for each task based on the applicable workload dimension.

2.2.7.2 NASA TLX

Participants were asked to complete an electronic version of the NASA TLX subjective workload questionnaire following training and at the completion of each experimental

condition. The NASA TLX includes a set of six sub-scales that include Mental Demand, Physical Demand, Temporal Demand, Performance, Effort and Frustration (Hart & Staveland, 1988). In Part 1 of the questionnaire, the ratings for each dimension was collected using a twenty-step bipolar scale from which a score ranging from 0 to 100 was presented (Appendix C). In Part 2 of the questionnaire, the participant was required to select from 15 possible pair-wise comparisons of the six dimensions. Participants selected the member of each pair that contributed more to the workload for that particular task. The computer tallies the number of times that each factor was selected. A uni-dimensional workload index ranging from 0 to 100 was then calculated by taking the sum of the weights of each dimension multiplied by the scale score for that dimension and dividing by 15.

2.2.8 Experimental Procedure

Participants were tested individually after being briefed according to standard Carleton's Research Ethics Board operating procedures. They were required to complete an informed consent form (see Appendix D) with a general description of the experiment and its requirements. Prior to beginning of the experiment, participants were provided with a subject information package outlining instructions for the ATC and Bakan simulations, and the NASA TLX subjective questionnaire (see Appendix D). A typical schedule for an experimental session for a subject is shown in Table 3.

Training occurred for 3 hours where each of these tasks was performed both alone and together to familiarize themselves with the controls and functions of the ATC simulator as well as gain experience responding to the Visual Bakan task. During ATC training, participants were introduced to some of the most common strategies for handling aircraft during peak loads. Some of these strategies included:

- To keep track of airplanes coming in, always keep track of the plane's call sign rather looking where it is situated in the schedule window. Do not rely on the layout of the schedule of airplanes to keep track of the aircrafts.

Subjects were told that their goal in the experiment was to route each arriving aircraft to its point of departure from the airspace in an expeditious fashion, consistent with the primary goal of aircraft safety. The avoidance of collisions with other aircraft and the ground was always the highest priority. They were told that, no matter how impossible it appeared to be to handle all the aircraft on the screen, they should try to do as much as they could.

After the training session, subjects had a break, then began the experiment. To control for task prioritization issues within the dual-task ATC-Bakan condition, the participants were told that it is critical to the conduct of the dual tasks that they approach both tasks with equal importance and were given strategies for handling both tasks simultaneously. Participants were instructed to use strategies such as:

- Constant monitoring of the Visual Bakan task – use your peripheral vision when attending to the ATC task. You can keep a running sequence of numbers in your mind by constantly repeating the presented digits.
- When there are two odd or even numbers in the Bakan, you can use this break to read some information about your aircraft
- Deal with one plane at a time and constantly check the Bakan.

Participants were required to complete five trials of the simulated ATC and Bakan tasks. Participants completed two 15 minute trials of the simulated ATC task under two workload conditions (low and high). Previous research has shown that the greatest influence on workload in the ATC simulation was the rate at which the simulation updated (Hendy, Liao and Milgram, 1997). The two levels of workload were set for the ATC task by manipulating the update interval in the ATC simulation. In the low workload condition, the update interval was set to 9 seconds while in the high workload condition the update interval was set to 6 seconds. Participants also completed one 15 minute trial of the Bakan task. Finally, participants completed two 15 minutes trials of the ATC task while simultaneously performing the Bakan task under the two workload conditions. After performing each trial, participants completed the NASA TLX questionnaires to measure their subject workload. Table 3 for the session schedule.

The participants were debriefed following the completion of all trials (See Appendix E). The entire session took approximately 5 to 6 hours to complete.

Table 3 Session Schedule

Pre-Trial Administration			
	Introduction and Informed Consent		10 min.
Training			
	Training ATC task		80 min.
	Training Bakan Task		20 min.
	Training ATC & Bakan Task		50 min.
	Questionnaire		30 min.
			190 min.
Human-In-The-Loop Experiment (trials were counter-balanced across participants)			
	ATC task – low workload	No Bakan	15 min.
	Questionnaire		5 min.
	ATC task – high workload		15 min.
	Questionnaire		5 min.
	Bakan Task		15 min.
	Questionnaire		5 min.
	ATC task – low workload	Bakan	15 min.
	Questionnaire		5 min.
	ATC task – high workload		15 min.
	Questionnaire		5 min.
			100 min
Post-Trial Administration			
	Debriefing and remuneration		10 min.
			Total Time: 5 to 6 hours

2.3 Study Design

The conduct of the ATC experiment is a (6) (Human, VACP, POP, POPIP, IP, W/Index) by (5) (Bakan, ATC low alone, ATC high alone, ATC low/Bakan, and ATC high/Bakan) factorial repeated measures design manipulating two ATC workload levels and six between group variables contrasting workload levels and task performance between the ATC and Bakan tasks and IPME models. As workload was manipulated using two

repeated measures variables and to control for possible training effects, the presentation of the workload conditions (ATC and Bakan workload conditions) were counterbalanced across participants using a partial Latin square design (Appendix I).

2.4 Methodology – IPME Models

2.4.1 IPME Workload Algorithms

Five major workload algorithms have been integrated within IPME to predict operator workload and the effects of internal and external performance shaping factors on task performance (VACP, W/INDEX, IP/PCT, POP and POPIP). For a detailed description of each workload algorithm, refer to Forbes, Darvill, Armstrong, & Banbury (2006).

Information Processing/Perceptual Control Theory (IP/PCT). IP/PCT provides a rule-based allocation of attention model and restricts multi-tasking to two concurrent tasks if these tasks draw on higher level cognitive processing. It is created based on two theoretical foundations, i.e., IP and PCT. Particularly related to workload assessment, the IP model uses the time domain for assessing the effects of task load on performance and operator workload. It introduces the concept of *time pressure* (i.e., the ratio of *time required to complete a task* to *time allowable*) as a driver of operator performance, subjective workload and errors (Hendy & Farrell, 1997). In the rest of the report, IP/PCT is also referred as IP model.

Prediction of Operator Performance (POP). The POP algorithm uses workload ratings defined for the various channels and calculates two things using the underlying Markov Process Model: Workload and Performance degradation as a time multiplier and probability of error if it is non-zero. This is applied at each instance for which the number of tasks is constant.

Prediction of Operator Performance/Information Processing (POPIP). The newly implemented POPIP merges its two predecessors, POP and IP/PCT, and uses components from both POP and IP/PCT for a combined workload algorithm that offers interference based on time pressure, and task scheduling (Fowles-Winkler, et al., 2004).

Visual, Auditory, Cognitive, and Psychomotor (VACP). This algorithm predicts operator workload using separate workload channels that include Visual, Auditory, Cognitive, and Psychomotor (manual and voice responding) channels. Workload for any given instant is predicted by summing the demands within each workload channel for all currently active tasks.

Workload Index (W/Index). The W/Index algorithm measures the resource demands imposed upon the operator within six resource channels: visual perception, auditory perception, verbal cognition, spatial cognition, manual response and speech response. Each task is decomposed within W/Index into this set of channels and weights are established representing the amount of demand required by the task for each channel. In the IPME W/Index implementation, VACP interval ratings are typically used to populate the W/Index task demands.

2.4.2 Model Development – Task Flow and Time

Computational models of the ATC and Bakan simulations were developed in IPME. A pilot test was conducted to map the task flow of human behaviour and performance in the ATC and Bakan simulation tasks. The performance data from the pilot test was used to develop the ATC model (see section 2.3.3 for more details on the model development). The Visual Bakan model was developed separated at DRDC Toronto and validated using data from other studies.

The pilot test was run using a low workload level (9 sec update) with eight subjects who did not participate in the validation experiment. The pilot tests for the ATC and Bakan tasks were each assessed in isolation, i.e. they were not combined as dual-tasks. This approach ensured that the task-data captured for the ATC and Bakan simulation tasks are representative of the baseline task performance for each domain, and are not obscured through interactions of the two tasks. Participants were monitored during the performance of each task and feedback was captured during the task through a talk-aloud procedure and following the completion of the ATC and Bakan tasks. The pilot study was also used to finalize the parameters for the two workload conditions (low, high) for the ATC simulation study.

The results of the pilot test were used to determine the task flows, time requirements to complete each task and the workload parameters for each of the five IPME workload algorithms. NASA TLX data were collected across each subject to populate the relevant workload parameters in the corresponding IPME baseline model. The performance data used to develop the ATC and Bakan IPME models are based on the *trained* performance of participants (i.e. after participants have become proficient with the ATC simulator). This ensured that training impacts were minimized in the model predictions. Upon completion of the baseline model development, the separate ATC and Bakan task networks were integrated into a single IPME task network that was representative of the dual-task environment of the ATC human-in-the-loop simulation. A scenario layer was then developed to drive the ATC and Bakan simulation

Two task flows of human behaviour and performance in the ATC and Bakan simulation tasks were derived from the pilot tests. Each task flow is described in the sections below.

2.4.2.1 Bakan Task Flow

The Bakan Task can be broken down into three main components (see Figure 2):

a. A monitoring component associated with monitoring, detecting and reading digits presented on the screen;

b. A processing stage that assessed the stimulus digit and target string

c. A recurrent component associated with the maintenance digits held in memory

d. A response component associated with the individual response.

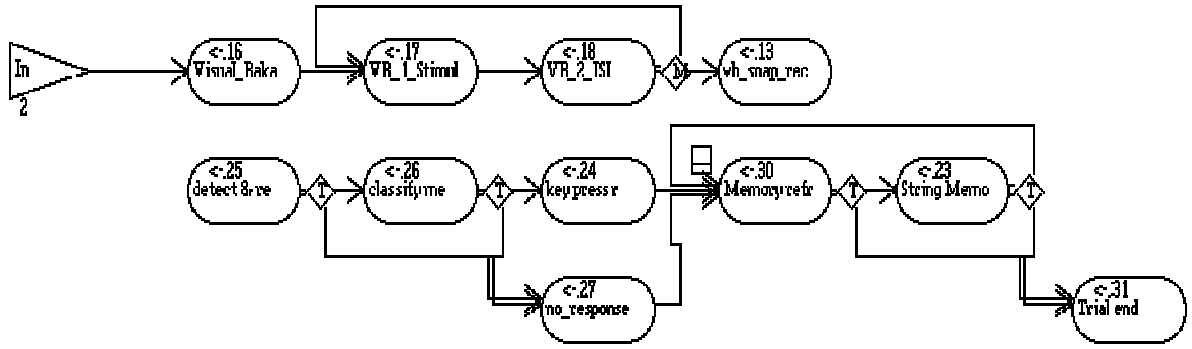


Figure 2 Human behaviour and performance in the Bakan simulation task

2.4.2.2 ATC Task Flow

The ATC task can be broken down into two main components (see Figure 3)

a. A monitoring component associated with monitoring, detecting, assessing aircrafts state; and

b. A recurrent component associated the individual response to make aircraft adjustments until desired aircraft state is achieved.

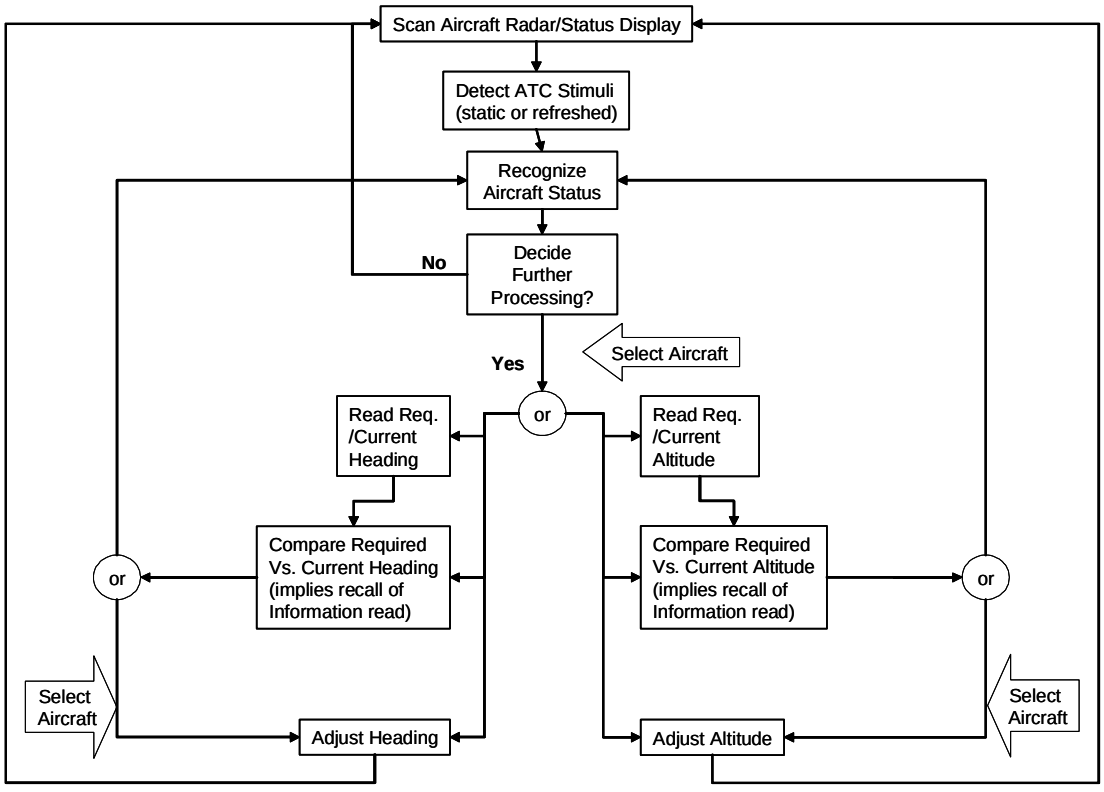


Figure 3 Human behaviour and performance in the ATC simulation task.

2.4.3 The Integrated Performance Modelling Environment

The Integrated Performance Modelling Environment is a network simulation software package for building human performance based task network models. The simulation software models operators in complex environments by assigning operators to a time

based series of discrete tasks which represent the potential interactions between the operator and the system. IPME contains a set of tools for predicting the effects of systems functions on operator workload and performance. IPME task network models are integrated as part of a complete system, in which the ATC and Bakan model was composed of:

a. **Task Network Model**: Contains the task flows, workload, and timing data which is used to run the model; also where operators are assigned to tasks within the workload model.

b. **Operator Model**: Individual models of operators that are assigned to tasks within the Task Network Model. For the purposes of the ATC and Bakan model, a single operator was used to perform each task in isolation, as well as in parallel.

c. **Environment Model**: A component model allowing for environmental factors to influence the outcome of as task network simulation. The environment model was not used in the ATC and Bakan modelling efforts.

d. **Performance Shaping Model**: A component model allowing for external factors such as physical and emotional stress to influence operator behaviour and task performance. The performance shaping model was not used in the ATC and Bakan modelling efforts.

2.4.3.1 Model Development

The network models for the ATC and Visual Bakan simulations were developed in conjunction with project team members from CAE Professional Services (Ottawa), DRDC Toronto, and QinetiQ Ltd. (London, U.K.). The majority of the ATC model was developed by CAE PS, while the Visual Bakan and the Operator model was developed at DRDC Toronto. The team members at QinetiQ provided assistance in refining these models, as well as data collection and analysis in the POP and POPIP simulations. The ATC and Visual Bakan models were developed independently and merged using IPME version 3_0_30 before executing the simulations in both isolation and parallel.

The Visual Bakan model contained two branches due to slight capability differences with the POP and IP/PCT schedulers. While attempts were made to keep the task network common for all workload models, the differences in the networks arose in an attempt to model the same logical task flow with the different approaches. The differences reflect differences in the modelling techniques with the different systems rather than an attempt to circumvent the rules entailed by each scheduler. The combined ATC and Bakan model was executed in all the experimental conditions listed in Section 2.1.

Specific model assumptions and limitations are discussed in Section 2.4.3.9.

2.4.3.2 Model Execution and Data Collection

Model execution for all ATC and Bakan conditions were run in four independent modes:

a. IPME mode with no task scheduler;

b. IP/PCT mode with IP/PCT task scheduler;

c. POP mode with POP scheduler; and

d. POPIP mode with a combined POP/IP scheduler.

The following data were collected during execution of the models:

2.4.3.3 Model Performance Measures

The following performance measures were generated in the ATC model.:

a. **Correct Exits.** Correct exits are instances when aircraft exit the radar screen along the correct track. These are expressed as a proportion from 0 to 1. The sum of the proportions for correct exits and fly out errors equals 1.

b. **Fly Out Errors.** Fly out errors are instances when aircraft exit the radar screen with incorrect track. These are expressed as a proportion from 0 to 1. The sum of the proportions for correct exits and fly out errors equals 1.

c. **Timeout Errors.** Timeout errors occur when aircraft fail to exit the radar within the allotted time. The allotted time for each plane is generated upon appearance.

d. **Collision Errors.** Collision errors occur when two or more aircraft occupy the same space.

The following performance measures were generated in the Bakan model:

a. **Hits.** Hits occur when operators respond to the presentation of a Bakan stimulus (odd, even, odd numeric sequence), and **misses represent** when operators fail to respond to the presentation of a Bakan stimulus.

b. **False Alarms.** False alarms occur when operators respond to non-Bakan stimulus (any instance when the odd, even, odd sequence is not shown), and conversely **Correct Rejections** appropriately do not respond to non-Bakan stimulus.

2.4.3.4 Model Subject Measures

The following subjective measures were collected across all ATC and Bakan conditions and analyzed depending on the simulation mode used:

a. **VACP visual demand.** Used in IPME mode analysis (no scheduler active).

b. **VACP central demand.** Used in IPME mode analysis (no scheduler active).

c. **VACP psychomotor demand.** Used in IPME mode analysis (no scheduler active).

d. **Composite W/INDEX**. Used in IPME mode analysis (no scheduler active).

e. **POP central demand.** Used in POP mode analysis (POP scheduler active).

f. **POP output demand.** Used in POP mode analysis (POP scheduler active).

g. **POPIP central demand.** Used in POPIP mode analysis (POPIP scheduler active).

h. **POPIP output demand.** Used in POPIP mode analysis (POPIP scheduler active).

2.4.3.5 Model Task Flows

The structure of the ATC and Bakan model was constructed using the network drawing tool in IPME. The ATC and Bakan models were developed independently in separate networks and then merged into one model where they could be run in isolation or in

parallel. An additional network dedicated to data capture was developed to collect operator workload and performance metrics during the simulations. Figure 4 illustrates the top level task network grouping, Figure 5 shows the task flow representation for the ATC task, and Figure 2 exhibits the Bakan network task flow.

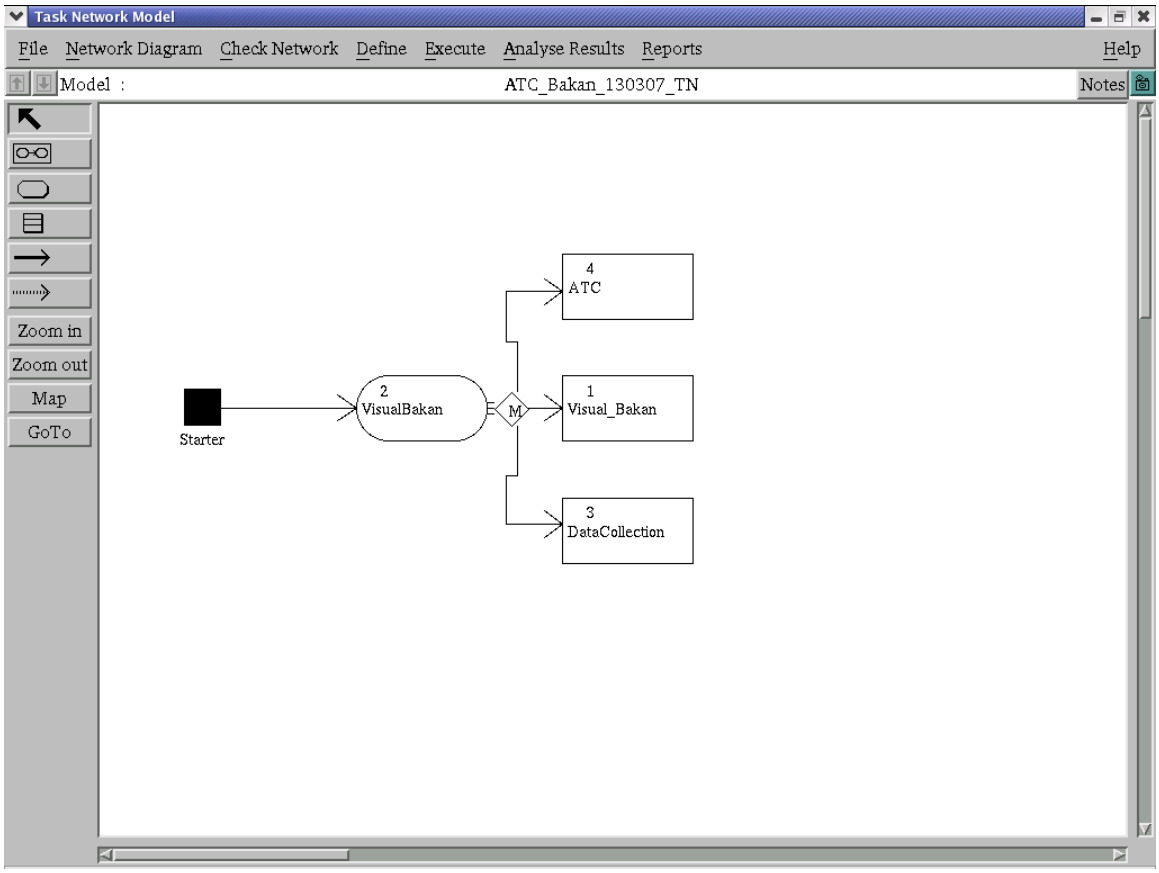


Figure 4 Top Level Task Network Grouping for ATC/Bakan Model

2.4.3.6 ATC Model Task Flow

The IPME task network was constructed similar to the operator activity analysis shown in Figure 3, although task decomposition was taken to a lower level to capture explicit psychomotor activity involved with performing the ATC task. Parameters were assigned at each task node which influences model predictions of operator performance and workload. The workload parameters were estimated based on pilot studies of human participants during baseline ATC experimentation. The task timing parameters were derived from psychophysical performance data in human factors literature (see Appendix F), and verified through comparisons with performance times drawn from the pilot studies.

The ATC model task flow can be broken down into four main components (see Figure 5). See Appendix H for a detailed description of the logical flow of one cycle in the ATC task network model.

a. A system representation component which initiates model execution, generates aircraft, and updates their attributes based on operator input;

b. A monitoring component associated with the period of time when operators have made the necessary adjustments to aircraft on the radar and are waiting for the next system refresh cycle before deciding whether additional adjustments are required;

c. A detection and decision component to locate the aircraft in question, evaluate its importance relative to other planes, and decide on the nature of adjustments required; and

d. An adjustment cycle representing selection of individual aircraft, reading the required heading and altitude, comparison of current versus required state, and selecting the appropriate control for the desired adjustment.

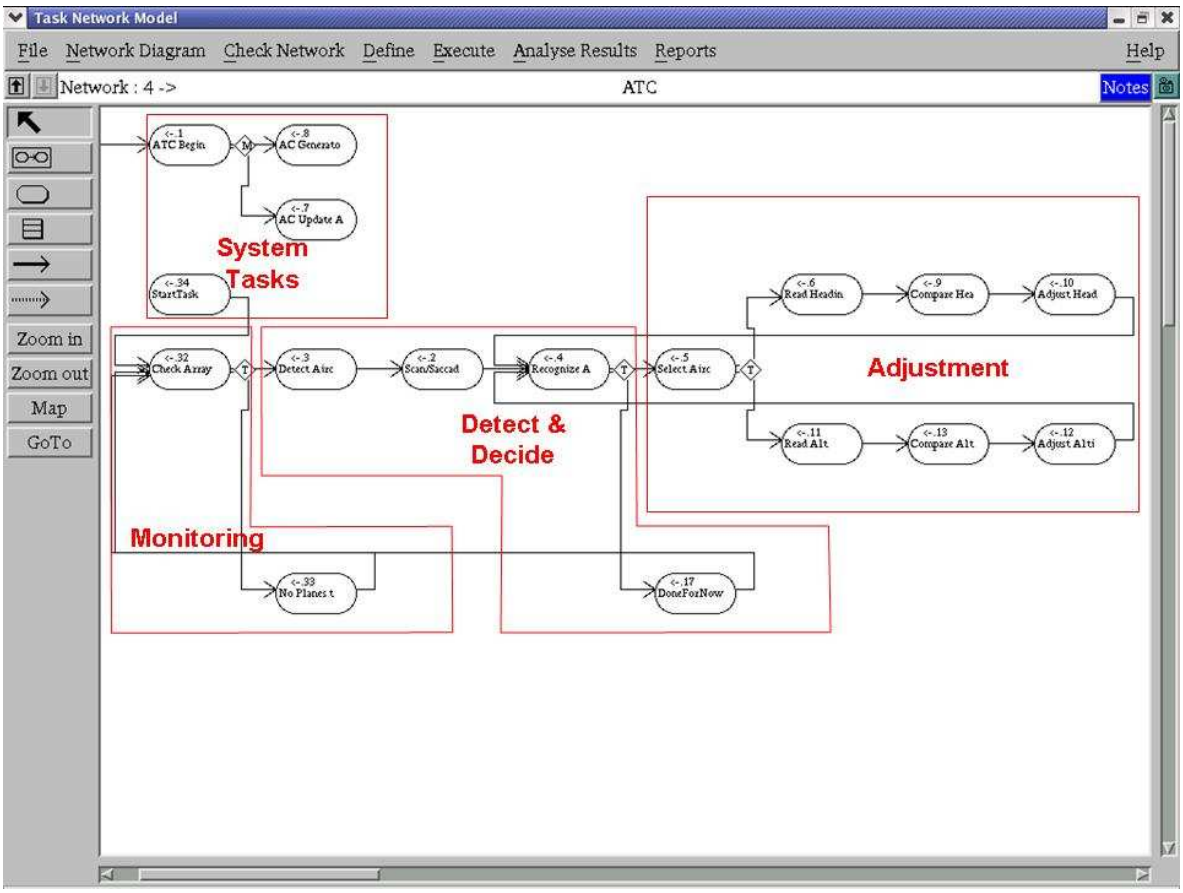


Figure 5 ATC Task flow in IPME

2.4.3.7 ATC Model Parameters

The pilot trial conducted by CAE Professional Services supported data collection efforts to obtain data on the ATC task for a single condition (constant schedule, 5 aircraft 1 airport and an update cycle of 9s) using 8 pilot-study participants. The data were then passed to QinetiQ and DRDC Toronto to assist in tuning for the baseline ATC model. The results from the pilot trial indicated that the mean participant TLX mental demand rating for the task was 0.8 and that the mean TLX physical demand rating was 0.2. Analysis from a previous study demonstrated that TLX mental demand was a good estimate of the DRAWS central rating and TLX physical demand provided a good

estimate of DRAWS output rating (Farmer et al., 1995). Model workload ratings were then assigned in concordance with the analysts best understanding of the nature of the task demands associated with the respective ATC task components as outlined in Table 5. For VACP, W/INDEX and IP/PCT ratings, workload assignments were applied to each task on the basis of the task demands in accordance with the relevant workload ratings scales.

An initial attempt at assigning the POP ratings to the task was made and the mean operator workload results corresponding to the POP workload channels derived from IPME are displayed in Table 4.

Table 4 The workload results for the initial assignment of the DRAWS ratings

Cycle time (s)	Input	Central	Output
3	0.36	0.42	0.52
6	0.28	0.32	0.40
9	0.19	0.23	0.28

The trends in the workload values appeared to be plausible but the absolute values were clearly low in comparison to the observed data from the pilot trial. Even if the maximum DRAWS ratings had been assigned to all the tasks, the average workload would not have reached the levels observed in the pilot trial. As such, a detailed analysis of the ATC model was conducted to determine whether the model structure and task flows were accurately reflecting operator task performance in the ATC tasks.

An analysis of the model revealed that the operator behaviours were only triggered once at the beginning of each update cycle. This meant that only a single plane could be updated in a given cycle. A change to the form of the model was proposed and implemented to enable corrections to be made to the heading and altitude of *multiple* planes within a given update cycle (see Figure 6). The early data seemed to suggest that the operator had very little idle time between the updates. To simulate this, a constant monitoring task was included at the end of each cycle of changes and before the next update occurred.

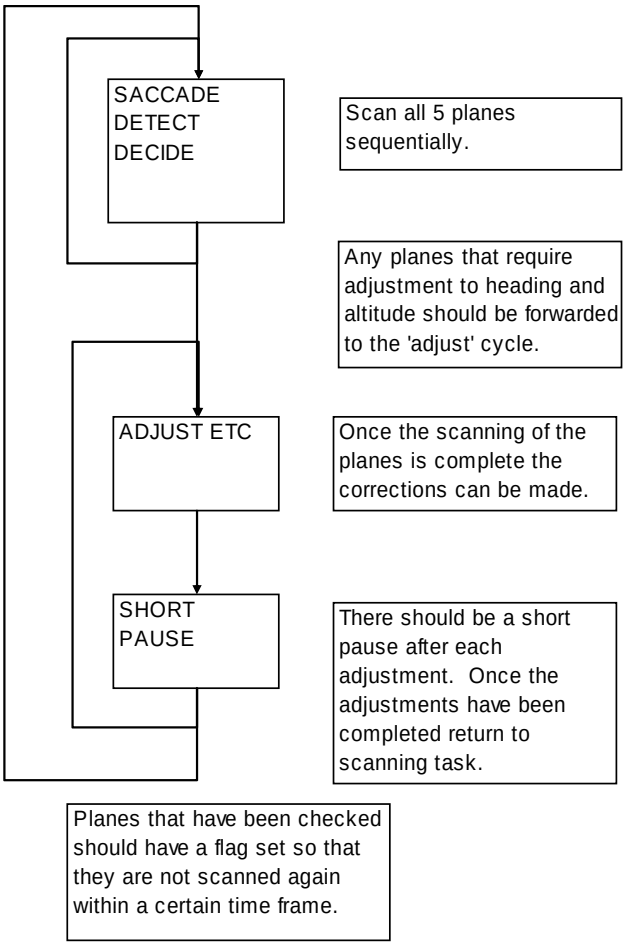


Figure 6 Proposed change to the ATC model

The final POP ratings were estimated using the following rules:

- to estimate the DRAWS ratings for each task, IPME micro-model timings for the tasks (an estimate of the active time in the task) were divided by the task timings in the model (the time available to perform the tasks), which indicated that the workload for the most of the tasks was close to 100;
- ATC tasks that have parallels with the Bakan tasks should contain similar ratings;
- similar tasks in the task flow should have similar ratings;
- tasks that occur just before a psychomotor task may contain a small amount of output demand: a pre-motor task (Belyavin & Farmer, 2006), and
- the mean workload of the ATC simulation must be similar to the mean workload observed in the pilot trial.

The final VACP, W/Index and IP ratings were derived from the associated task descriptors for the corresponding workload algorithm and are provided in Appendix G. During the course of the project, it was determined that the VACP to W/Index mappings in IPME did not correspond to a reasonable interpretation of workload. Therefore, automated VACP to W/Index mapping function in IPME was not used, and W/Index values were manually assigned to each task.

2.4.3.8 Execution Settings

There are a number of adjustable parameters in the ATC model to enable simulation of a range of experimental conditions and Table 5 provides a list of the execution settings used to simulate the experiment conducted at CAE Professional Services.

Table 5 The execution settings for the model

Parameter	Description	Values
ATC	Presence of the ATC task (0 = No, 1 = Yes)	0, 1
BAKAN	Presence of the Bakan task (0 = No, 1 = Yes)	0, 1
IPPCT	Switch for IPPCT running	0
atc_ac_max	The maximum number of planes on the radar	5
atc_cycle	The update cycle time in seconds	6, 9
no_cycles	The time in seconds before a plane can be considered for another update	15
vb_stimulus_isi	The Bakan inter stimulus interval (ISI)	1.5
vb_target_amount	The Bakan target string length (number of digits)	3
Number of runs	The number of runs	1
Number of crew samples	The number of crew samples	24
Total number of experimental conditions	The final number of experimental conditions	5 (3 ATC x 2 Bakan minus the 0,0 case)

2.4.3.9 ATC Model Implementation Limitations and Assumptions

The use of IPME to evaluate operator workload and performance has a number of limitations that are linked to the challenges associated with modelling human behaviour in the ATC task. These limitations are outlined as follows:

- The granularity of the ATC and Visual Bakan tasks were matched to ensure that the level of representation of behaviour across each task was equivalent (e.g. decision tasks, key-stroke behaviour).
- A problem was detected when running the Bakan model in IP/PCT mode: there was an index out of bounds error. QinetiQ traced this problem to the ‘Shed if late’

mechanism that appeared to generate a zero-value tag in the ending effect of one of the tasks. The solution to this problem was to use the tactical branching that was present in the POP branch of the Bakan model rather than the IP/PCT 'Shed if late' mechanism.

- An initial investigation into the model uncovered an implementation issue in the POP scheduler in IPME 3.0.25. The operator workload values were not recalculated at a task ending event. The model should recalculate the figure at the end of each task to take into account the reduction in workload due to the task finishing. The failure to recalculate the workload meant that the values would remain in the operator model beyond the end of a task until a new task started, triggering a recalculation of the values. This would lead to an overstatement of the mean workload of the operator. To prevent this, a short (70 ms), zero-workload task was inserted into the model to ensure that workload recalculation was triggered before data collection.
- DRDC Toronto determined that the W/Index implementation within IPME V3.0.25 was incorrect as it considered terms in the interaction calculations even when one of the task demands was zero, a divergence from the original model. In addition, DRDC Toronto determined that not all unique pairings of nonzero task channels were being considered in the IPME W/Index implementation. The following modifications were made to the W/Index algorithm in IPME V3.0.30 prior to the analysis:
    - Only consider terms in the interaction calculations where both task demand terms in a channel are non-zero.
    - That all unique pairings of task channels should be considered (ie. j = 1,6 NOT j = i,6 as implemented in V3.0.25)
- The decision process by which humans evaluate importance of one aircraft over another involves several layers of complexity which are subject to variations across ATC operators of different skill level and experience. These decisions can have a direct impact on predicted operator performance, and requires insight on the part of the modeller to ensure that the criteria used to drive aircraft prioritization is accurately represented. The logic currently applied in the ATC model uses a simplistic rule base which represents a trained operator, but does not account for skill increases gained from experience. As a result, there remains the potential for underestimating task performance as well as reduced performance variations between participants.
- The process by which the ATC monitoring cycle generates workload was based on a constant setting of five aircraft, and is currently limited in terms of its ability to dynamically represent task demands with varying number of aircraft.
- The monitoring loop often delays the operator task adjustment cycle due to the static nature of its duration time, which can result in poorer than expected performance.

- Cognitive processes such as memory recall, decay, confusion, and its associated effects on ATC task performance were not modelled given that these are difficult to validate without extensive effort and the identification of appropriate models from the relevant psychological literature. As such, the current implementation of the task nodes representing reading and comparison assumes that recall must be successful in order to make comparisons between the current and required goal states of aircraft.
- Airport landings were not modelled due to the substantial set of logical rules that would be required to drive and simulate such behaviour.

2.4.3.10 Bakan Model Task Flow

An IPME task network model of a Visual Bakan task was constructed based on descriptions from the literature and reports from the Defence Evaluation and Research Agency (DERA) of human studies using this task that were conducted at Cardiff University (Farmer, Jordan, Belyavin, Birch & Bunting, 1993; Farmer, Belyavin et al., 1995; Farmer, Jordan et al., 1995). The experimental setup has subjects sitting in front of a computer screen where a continual presentation of stimuli (numbers from 0 to 9) occurs in the centre of the display. Each stimulus was presented for 500ms and the inter-stimulus interval (ISI) was varied between trials to impose different levels of temporal demand. Before a trial, subjects were given a sequence of odd and even digits as a target string (for example, odd-even-odd) and the number of digits in the target string was manipulated to impose different levels of difficulty and, presumably, temporal demand. Subjects were asked to attend to the stimuli, interpreting the numbers as an even or odd digit, and when they detected the target string in the presentation, they were to respond either by tapping the table or verbally.

The Visual Bakan task was modelled by decomposing it into subtasks that we think subjects perform as shown in Figure 7. The times for each stage of the task process were derived from literature values (Card, Moran & Newell, 1983) incorporating subject variability into these times by varying operator traits. At this point, no attempt has been made to create psychological process models of each of the stages and simple engineering models were included in the subtasks to represent the processes of interpreting and responding to the stimuli. This prevented a detailed comparison of objective measures of performance, however, the objective measures were used to constrain the tuning of the Bakan Task parameters that focused on reproducing the workload ratings.

The stages of processing were broken down into a perception subtask (detect and read), a cognitive processing subtask (classify/memorize/compare) and a response subtask (key press/no response). If there was a sufficiently large ISI, a memory rehearsal task was executed to refresh the string of interpreted stimuli digits as well as the target string currently held in working memory. There was an opportunity for the memory rehearsal task to overlap with the stimulus detection task, although the rehearsal task would not start if a new stimulus was already detected. The co-occurrence of the rehearsal and detection tasks could produce some interference if the rehearsal does not finish before the next detection task occurs.

During a simulation, the operator traits were sampled to obtain characteristic times or error rates to represent between-subject variability. These traits formed the expected values for the mean times or error rates in the subtasks. The subtask time was expressed as an Ex-Gaussian distribution that was then sampled to represent within-subject variability for task completion time. A repeated measures experiment with the different Visual Bakan conditions was run using the same traits for all conditions.

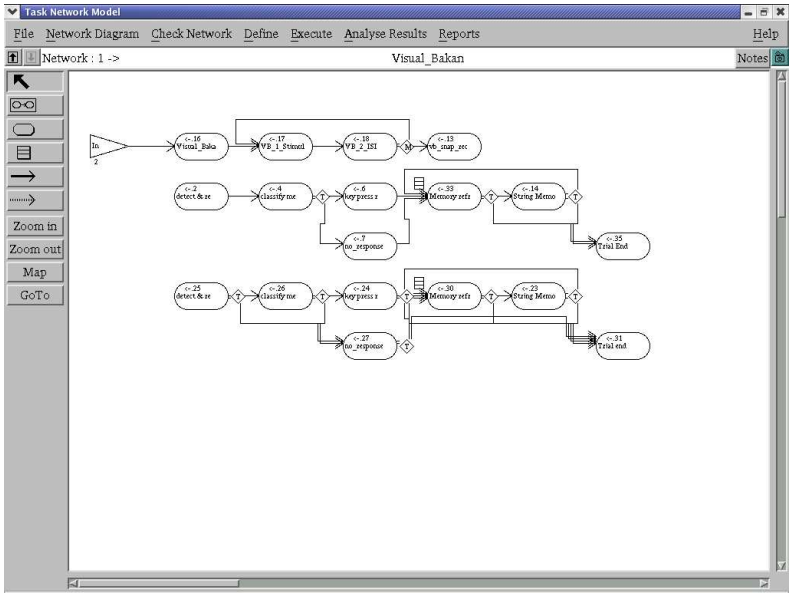


Figure 7 Bakan Task flow in IPME.

2.4.4 Translating models into IPME 4

To translate the models from IPME 3 to IPME 4, a number of minor changes were made. In IPME 4, there is a requirement for a function of type double or integer to return a value, which was never rigidly enforced in IPME 3. The difference in standards means that functions need to be explicitly stated as void in IPME 3 version of the model before translating it to IPME 4. After the translation, there were a number of minor corrections required to variable types. These translations should not have changed the behaviour of the model. To verify this, the translation was tested by statistical analysis (see Section 3.8) of IPME 3 POP results versus IPME 4 POPIP_POP results (although it was recognised that IPME 4 POPIP_POP results are only an approximation to the pure POP results).

2.4.4.1 POPIP parameters

Table 6 contains a list of the parameters available in the POPIP implementation that can be used to tune the behaviour of the scheduling algorithm. This table contains the value of these parameters for all three POPIP cases mentioned in the previous section, with reasons as to why the value was chosen.

Table 6 POPIP parameter settings

Parameter	Description	POPIP default	POPIP IP	POPIP POP
POP calculation	Enables/disables the POP task interference model	YES	NO – to remove the influence of the POP model this feature must be switched off	YES
Structural Interference	Enables/disables the IP/PCT structural interference model for visual tasks	YES	YES	NO – to remove the influence of the IP/PCT model this feature must be switched off
Critical interference	Critical value of the interference coefficient between two tasks	0.7	0.7	1.0 – set to maximum value to reduce its influence
Critical % Complete	When a task is more than X% complete the task can no longer be interrupted	70%	70%	100 – to remove the influence of the IP/PCT model this parameter must be set to its maximum value
Critical Time Pressure	When the time pressure of a task reaches this value it can no longer be interrupted	0.8	0.8	3.0 – to reduce the influence of the IP/PCT model this parameter must be set to its maximum value
Short term memory	The number of interrupted tasks that can be stored in the short term memory – extra tasks are shed	3.0	3.0	10000 – set to a high value to ensure that tasks are never shed
Task resume penalty	The additional time of restarting a task when it has been resumed	0.05	0.05	0.00 – to remove the influence of the IP/PCT model this feature must be switched off
Priority time pressure (PTP)	The task time pressure is multiplied by an importance factor to derive the task priority value.	NO	NO	YES – this feature was developed as a representation of task importance.

2.4.4.2 POPIP Implementation

The implementation issues identified in the POP and IP/PCT models were tested in IPME 4.1.3. It was found that both issues had been rectified in IPME 4.1.3 and in addition to the three POPIP modes described above, a fourth mode which used the original 'Shed if late' mechanism was executed (POP_IP default plus task shedding).

2.4.5 Comparison of POP and POPIP

A simple balances analysis of variance (ANOVA) was used to determine whether there were any significant differences between IPME 3 implementation of the POP model and IPME 4 approximation (POPIP_POP). A mixed effect linear model was implemented, containing the following main effects and interactions:

1. a fixed effect of “ipme” (version 3.0.25, version 4.1.3);
2. a fixed effect of “condition” (Bakan only, ATC low, ATC high, Bakan and ATC low, Bakan and ATC high);
3. the interaction between “ipme” and “condition”; and
4. a random effect of “crew_sample”.

The dependent variables in the model were:

1. mean POP input;
2. mean POP central;
3. mean POP output;
4. Bakan reaction time; and
5. Bakan false alarms

A lack of significant differences between the levels of the “ipme” factor would indicate that:

1. the translation had no major effects on the performance of the model; and
2. there is an adequate approximation to POP in IPME 4.

3 Results

Prior to all analyses, data was checked for outliers, missing data, and normality to verify that there were no violations in statistical assumptions**.** Missing data for two participants were imputed using the mean of two nearby points and the linear trend at that point.

3.1 Workload

IP mode predictions of Operator Mean Time pressure produced high variations across the sampled simulations and was omitted from the analysis. Given the extreme values observed, the method by which Mean Time Pressure is determined during IP mode requires further refinement.

In addition, the results of the analysis indicated a large number of low-level interactions between the modeled tasks and human performance data. Due to the scope of the current study, these interactions were not analyzed in detail. The final data associated with each condition is available for future analysis if required.

To evaluate if there is a difference between participants' subjective workload perceptions and the simulated predicted workload, the NASA/TLX overall and subscale scores and IPME model resource demand outputs were converted to a proportional scale from 0 to 1. The NASA/TLX scores were then matched to the appropriate IPME model component representing human resource demand that most represent the specific dimensions of the NASA/TLX. Table 7 shows the mapping of NASA/TLX scores to each IPME model resource demand output. The analyses were performed on the mean of subjective/predicted workload of each subject within a task condition.

Table 7 NASA/TLX scores matched to the corresponding IPME model predicted workload outputs.

NASA/TLX	VACP
Mental Demand	VACP cognitive
Physical Demand	VACP physical
NASA/TLX	POP
Mental Demand	POP central
Physical Demand	POP output
NASA/TLX	POPIP
Mental Demand	POP central
Physical Demand	POP output

3.1.1 Omission of the Workload Profile

While the Workload Profile can reveal dimensions of human workload relevant to task condition and demand, results from pilot investigations showed that it proved to be too difficult to implement and was therefore omitted from the experimental design. Anecdotal reports during the pilot study indicated that participants found it difficult to understand each dimension and how they related to task demand. Future studies should evaluate how best to implement this approach to ensure valid subjective workload scores.

3.1.2 Mental Demand

The Levene's test for equality of variances revealed that there was a violation on the assumption of homogeneity of variances on the mean mental workload scores. The Greenhouse-Geisser's epsilon adjustments was made to assume sphericity. A (5) (Task Condition) by (4) (Human, VACP, POP and POPIP) repeated measures ANOVA revealed a main effect of task condition , $F(2, 167) = 661.314$, $p<.001$, and a main effect attributable to group, $F(3, 80) = 63.411$, $p<.001$. A significant condition by group interaction effect was also found, $F(6, 167) = 50.338$, $p<.001$ (See Appendix K for SPSS output tables).

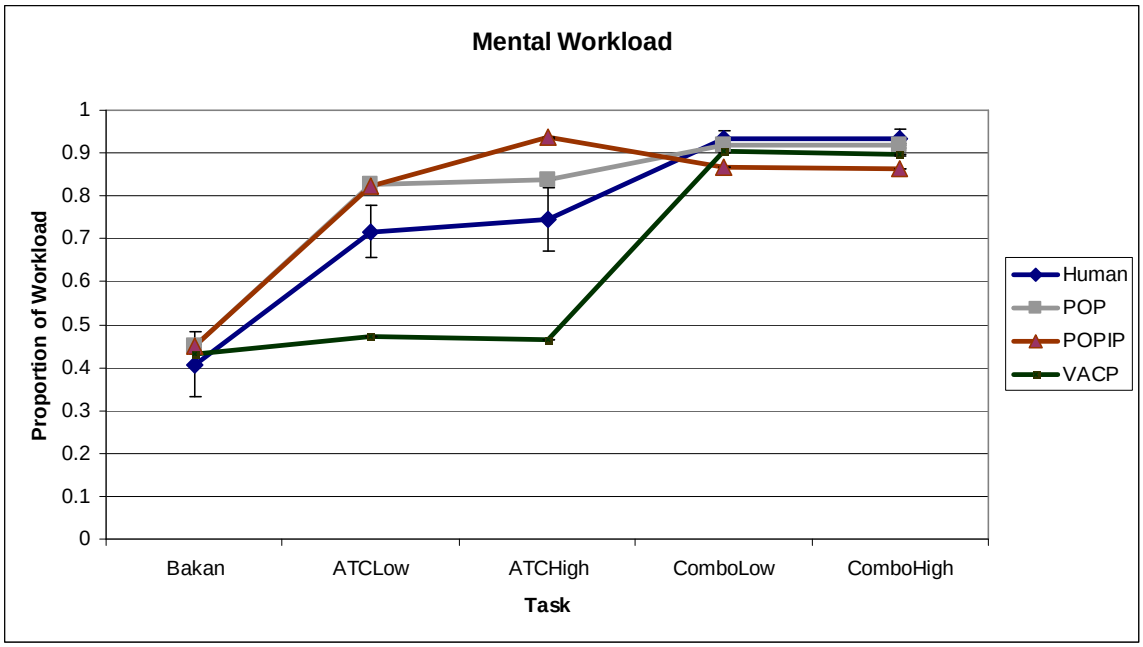


Figure 8 Mean mental workload (±*SE*) for the Human participants, POP, POPIP and VACP as a function of task condition.

Figure 8 shows that the pattern of mental workload scores varies as a function of task condition. It can also be seen that while workload generally increased over the task conditions for all groups, this trend varied as a function of group. POP and POPIP models follow a similar pattern as the human participants whereby mental workload is higher for the tasks involving the ATC simulation compared to Bakan. Note that the apparent lack of error bars for model generated workload is an artefact of the extremely small variance in workload produced by the model (see Section 4.5.2).

The trend for VACP, however, shows that mental workload in the combination task conditions is much greater compared to the single task conditions. It can also be seen that while all models are fairly accurate in predicting human mental workload in the Bakan and Combination task conditions, POP and POPIP over-predict mental workload in the ATC alone condition compared to humans, while VACP greatly under-predicts it.

In summary, these results indicate that the VACP model is not as accurate in predicting human mental subjective workload across tasks compared to the POP and POPIP models. Furthermore, POP and POPIP appear to over-predict mental workload in the ATC alone condition while VACP under-predicts compared to the human participants.

3.1.3 Physical Workload

The Levene's test for equality of variances revealed that there was a violation on the assumption of homogeneity of variances on the mean physical workload scores. The Greenhouse-Geisser's epsilon adjustments were made to assume sphericity. A (5) (Task Condition) by (4) (Human, VACP, POP and POPIP) repeated measures ANOVA revealed a main effect of task condition , $F(2,190) = 183.847$, $p<.001$, and a main effect attributable to group, $F(3,80) = 78.403$, $p<.001$. A significant condition by group interaction effect was also found, $F(7, 190) = 15.969$, $p<.001$ (See Appendix L for SPSS output tables).

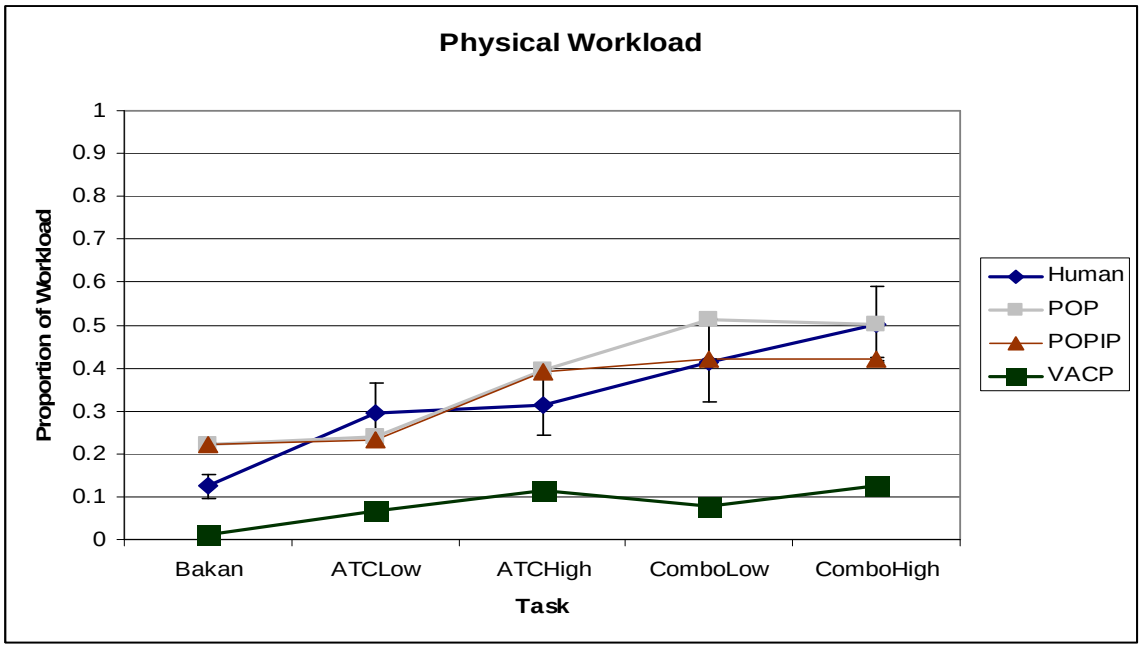


Figure 9 Mean physical workload (+*SE*) for the Human participants, POP, POPIP and VACP as a function of task condition

Figure 9 shows that the pattern of physical workload scores varies as a function of task condition. It can also be seen that workload generally increased over the task conditions for all groups. In addition, it is clear that VACP greatly under-predicts physical workload across all task conditions. It can also be seen that POP over-predicts in the Bakan, ATC High and ComboLow conditions, it under-predicts physical workload in the ATC Low condition compared to human participants. POPIP, meanwhile, over-predicts physical

workload in the Bakan and ATC High conditions and under-predicts in the ATC Low and Combo High conditions compared to humans. Interestingly, POP is very accurate in the ComboHigh condition while POPIP is very accurate in predicting physical workload in the ComboLow condition.

In summary, these results indicate that while physical workload generally increased over the task conditions for all groups, the pattern in which this trend increased varies as a function of group. It was found that VACP model is not as accurate in predicting human physical subjective workload across all tasks compared to the POP and POPIP models. Furthermore, the accuracy in POP and POPIP prediction of human physical workload is influenced by task condition.

3.1.4 Subjective Workload: W/Index, VACP and Humans

W/index values depend upon constraints of the number of concurrent tasks, which varies according to the type of task scheduling applied. Because of the difference in the schedulers, only an exploratory analysis of W/Index workload could be made. Below, in Figure 10, it can be seen that mean predictive workload for W/Index varies as a function of task condition when none of the task schedulers was invoked (IPME mode).

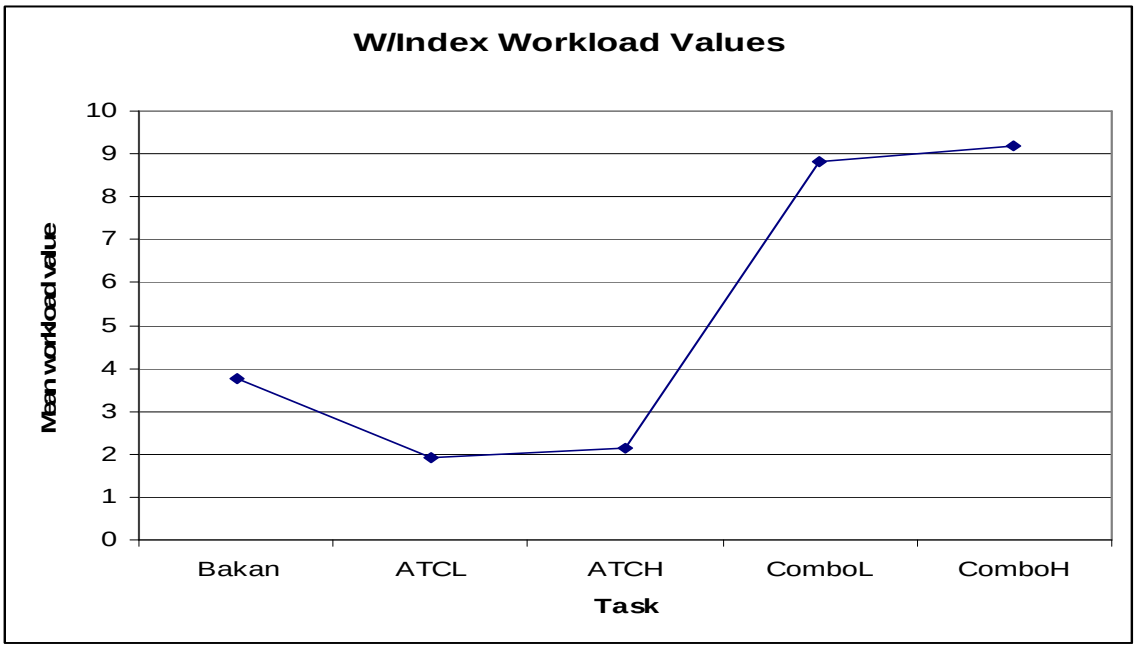


Figure 10 Mean predictive workload for W/Index as a function of task condition.

The trend in workload seen in W/index is similar to the trend seen in VACP cognitive values (see Figure 11) and Humans' composite NASA/TLX (see Figure 12) whereby the workload is significantly higher in the Combo Low and High conditions compared to the Bakan condition. W/Index differs from VACP and Humans whereby W/Index workload values decrease in the ATC low and high conditions compared to Bakan rather than increase.

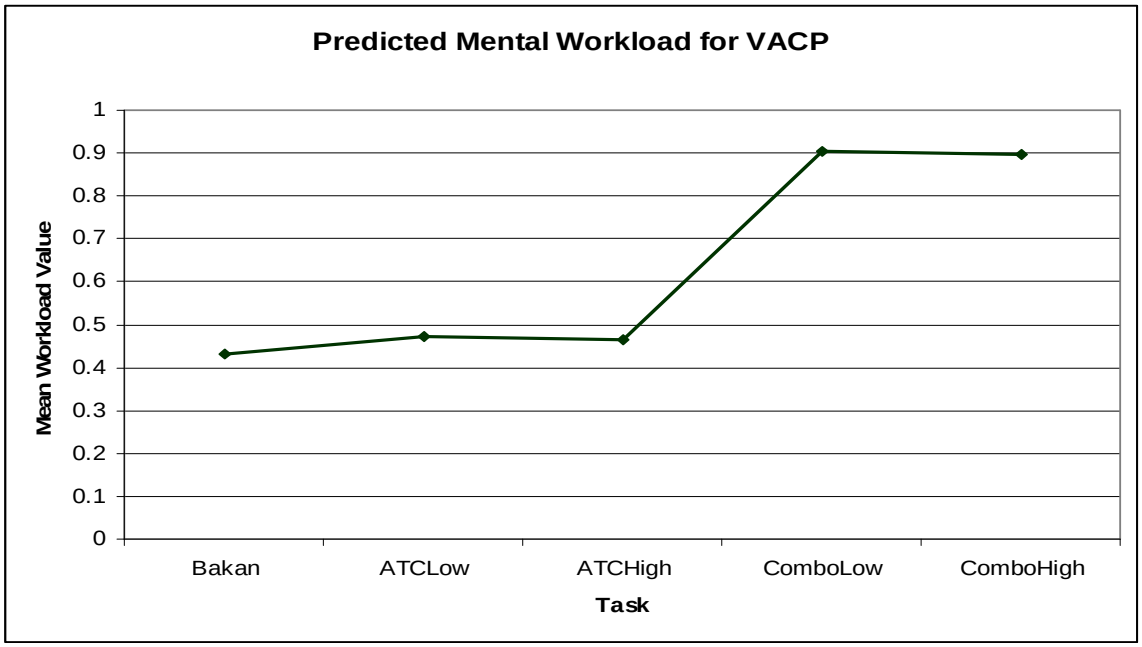


Figure 11 Mean predictive workload for VACP as a function of task condition.

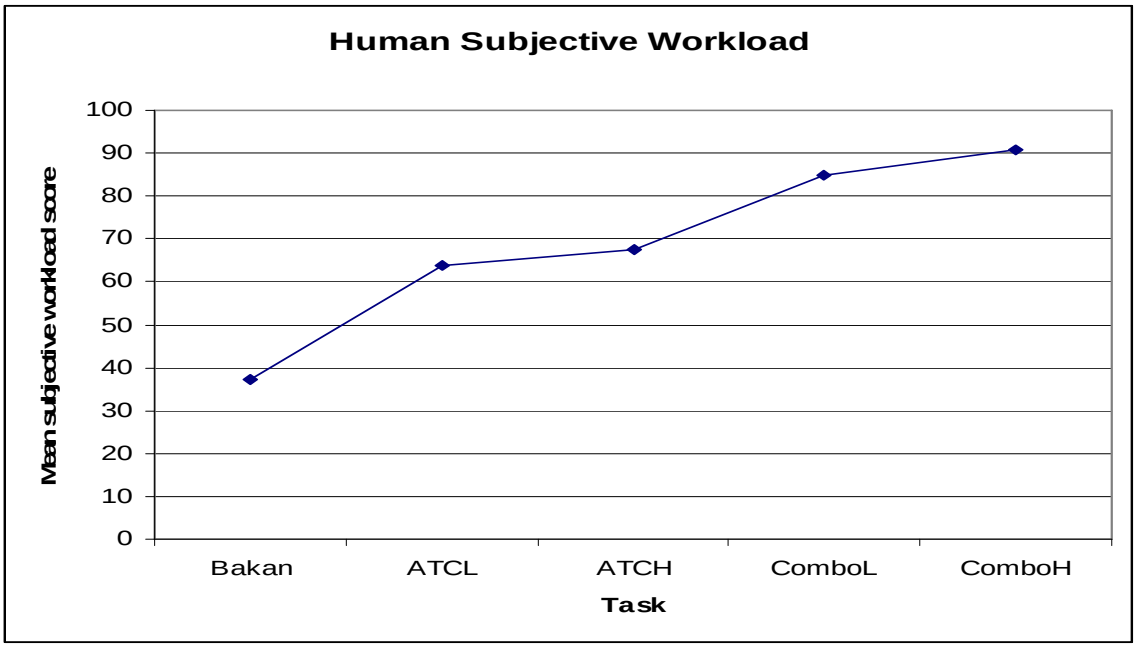


Figure 12 Mean subjective workload (composite NASA/TLX score) for Human participants as a function of task condition.

3.1.5 Summary of Workload

As expected, subjective/predictive workload increased as a function of task condition (Bakan, ATC low and ATC high alone, ATC low/Bakan and ATC high/Bakan combination). Overall, the POP, POPIP IPME models more accurately predicted human subjective workload compared to VACP. Indeed, VACP greatly under-predicted workload in both the mental and physical dimensions for the ATC task alone, however,

VACP produced reasonable estimates for the single Visual Bakan and the dual ATC-Bakan task conditions.

The pattern in which workload increased, however, was influenced by group. Indeed, it was found that POP and POPIP appear to over-predict mental workload in the ATC alone condition while VACP under-predicts compared to the human participants. Furthermore, the accuracy of POP and POPIP differ as a function of task condition. While POP is very accurate for predicting physical workload in the ComboHigh condition, POPIP is very accurate in the ComboLow condition. In addition, POP over-predicts physical workload in the Bakan, ATC High and ComboLow conditions, and under-predicts physical workload in the ATC Low condition compared to human participants. POPIP, meanwhile, over-predicts physical workload in the Bakan and ATC High conditions and under-predicts in the ATC Low and Combo High conditions compared to humans.

W/Index workload measures meanwhile follow a similar trend to VACP and Humans except that W/Index workload decreases in the ATC low and high conditions rather than increase relative to the Bakan condition.

3.2 ATC and Bakan Performance

To evaluate if there is an overall difference between participants' performance on the ATC and Bakan task and the performance produced by the models, a univariate analysis on the total performance errors were compared among groups. The total error was calculated by adding the mean errors for Bakan Misses, Bakan False Alarms, and ATC misdirection errors across all five conditions. The test of between subject variance showed a significant main effect of group ($F=23.9$, $df=4, 101$, $p < .001$) (see Figure 13). The ANOVA table can be found in Appendix M.

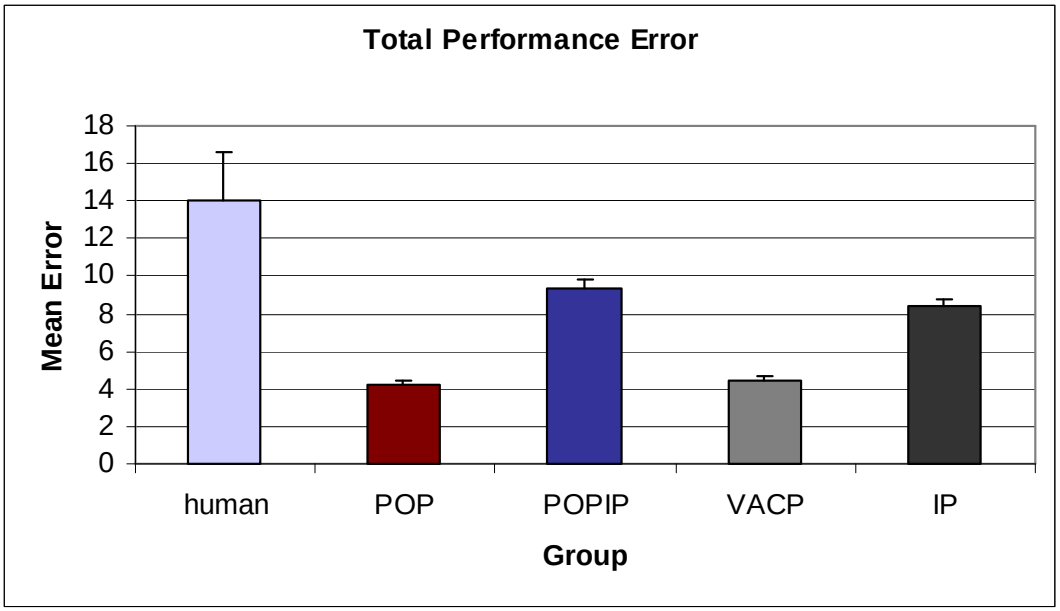


Figure 13 Estimated marginal mean for total error for the Human participants, POP, POPIP, VACP and IP.

As can be seen, all models produced significantly less errors when compared to the human participants. Among the models, the POPIP simulation produced less errors than the other groups, followed by IP, VACP and POP.

Multiple comparisons were performed to statistically test the difference between each group's performances. It was found that humans differed significantly from all simulated groups at p<.001 (lower performance or more error). POP only differed significantly from VACP at p<.008 (See Appendix L). No other significant differences were found. From the raw data we can conclude that VACP produced less errors than all other groups, followed by IP, POPIP and POP, which did not differ significantly from each other, but did differ significantly from humans by outperforming them. However, the data showed lack of sphericity and normality and equal variances could not be assumed. To account for lack of equal of variance, the Dunnett T3 posthoc test and correction were applied. See Table 8 of Multiple Comparisons comparing group performance.

Table 8 Multiple Comparisons (Dunnett T3) comparing performance of groups

	Group	Group	Mean Difference	Std Error	Sig.	95% Confidence Interval	
			Lower Bound	Upper Bound	Lower Bound	Lower Bound	Upper Bound
Dunnett T3	human	POP	9.8379*	2.56524	.023	1.1470	18.5289
		POPIP	4.6901	2.60440	.559	-4.0355	13.4157
		VACP	9.6411*	2.56434	.027	.9509	18.3313
		IP	5.6359	2.58712	.344	-3.0735	14.3453
	POP	human	-9.8379*	2.56524	.023	-18.5289	-1.1470
		POPIP	-5.1479*	.56852	.000	-6.8501	-3.4456
		VACP	-.1968	.34083	1.000	-1.1972	.8036
		IP	-4.2021*	.48323	.000	-5.6345	-2.7696
	POPIP	human	-4.6901	2.60440	.559	-13.4157	4.0355
		POP	5.1479*	.56852	.000	3.4456	6.8501
		VACP	4.9511*	.56442	.000	3.2588	6.6433
		IP	.9458	.66024	.801	-.9965	2.8881
	VACP	human	-9.6411*	2.56434	.027	-18.3313	-.9509
		POP	.1968	.34083	1.000	-.8036	1.1972
		POPIP	-4.9511*	.56442	.000	-6.6433	-3.2588
		IP	-4.0053*	.47841	.000	-5.4249	-2.5856
	IP	human	-5.6359	2.58712	.344	-14.3453	3.0735
		POP	4.2021*	.48323	.000	2.7696	5.6345
		POPIP	-.9458	.66024	.801	-2.8881	.9965
		VACP	4.0053*	.47841	.000	2.5856	5.4249

Based on observed means.

*. The mean difference is significant at the .05 level.

Interestingly, with the more stringent testing, the significance of comparisons changes. Humans are significantly different from POP and from VACP, but not from others. POP and POPIP differ significantly, POP and IP differ significantly, and VACP differs significantly from all other groups except POP.

The difference seen in the corrected and non-corrected comparisons suggests the small sample size may have contributed to the lack of equal variance among groups and that

increasing the sample size in each group might show a different picture more closely related to the comparison without post hoc corrections.

In summary, while all models greatly underpredicted error rates when compared to human performance in the ATC and Bakan tasks, only POP and VACP differed significantly from humans following conservative comparison testing. These results must be taken with caution due to the lack of equal variance among groups.

3.2.1 ATC Performance

The analyses were performed on the mean of ATC misdirection errors of each task condition. The Levene's test for equality of variances revealed that there was a violation on the assumption of homogeneity of variances on the mean ATC misdirection errors. The Greenhouse-Geisser's epsilon adjustments were made to assume sphericity. A (4) (Task Condition) by (5) (Human, VACP, POP, POPIP, and IP) repeated measures ANOVA revealed a main effect of task condition, $F(3,303) = 15.906$, $p<.001$, and a main effect attributable to group, $F(4, 101) = 44.337$, $p<.001$. A significant condition by group interaction effect was also found, $F(12,303) = 8.212$, $p<.001$ (See Appendix N for SPSS output tables).

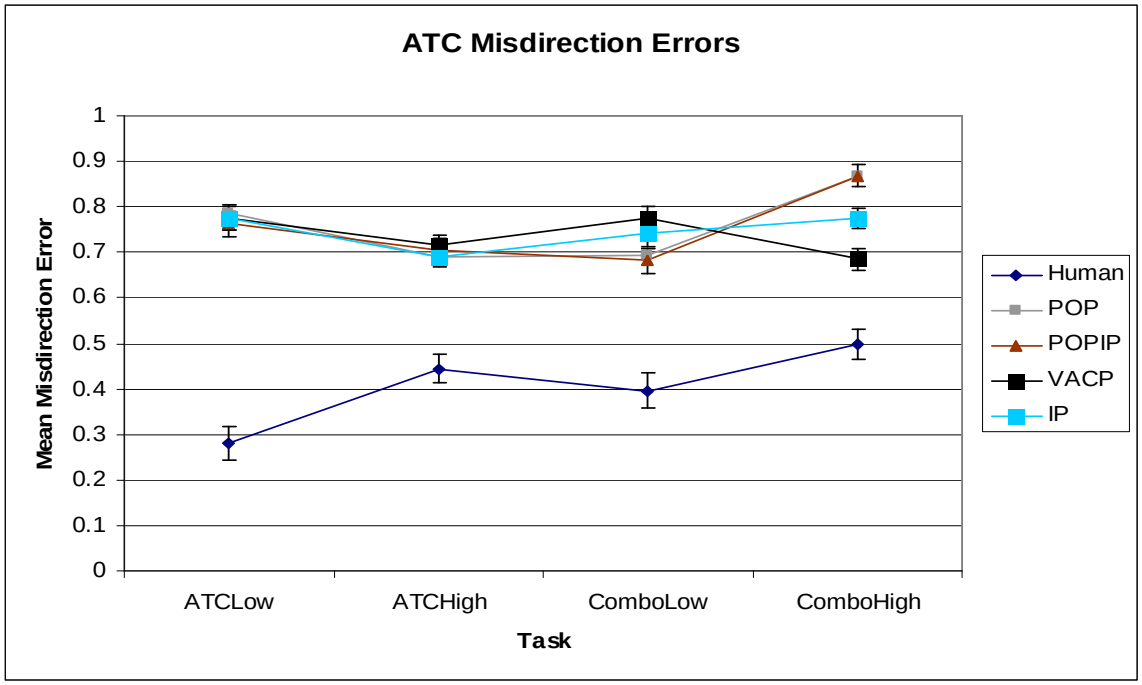


Figure 14 Mean for ATC misdirection errors for the Human participants, POP, POPIP, VACP and IP.

As can be seen from Figure 14, the pattern of ATC misdirection errors vary as a function of group and task condition. Indeed, it is clear that the models performed significantly worst on the ATC task across all conditions compared to human participants, as indicated by misdirection errors. Meanwhile, errors committed by human participants increased as a function of task. Performance on the ATC task by the simulation models however, was not consistent among models nor with human performance across task conditions. While

all models perform similarly to each other on the single ATC task, they differ in the dual-task conditions.

In summary, these results indicate that that all models greatly under-predict human performance on the ATC task. The pattern in which model performance varies across task condition meanwhile, is inconsistent with human performance.

3.2.2 Visual Bakan Performance

To evaluate if there is a difference between participants' performance on the Visual Bakan task and the simulated performance, two (3) (Task Condition) by (5) (Human, VACP, POP, POPIP, and IP) repeated measures ANOVAs were conducted on the commission and omission errors in the Bakan task.

The first analysis was performed on the Visual Bakan omission errors (Missed odd-even-odd sequences) of each task condition. The Levene's test for equality of variances revealed that there was a violation on the assumption of homogeneity of variances on the mean omission errors. The Greenhouse-Geisser's epsilon adjustments were made to assume sphericity. A main effect of task condition, $F(2,202) = 3021.777$, $p<.001$, and a main effect attributable to group, $F(4, 101) = 648.890$, $p<.001$ was observed. A significant condition by group interaction effect was also found, $F(8,202) = 596.711$, $p<.001$ (See Appendix O for SPSS output tables).

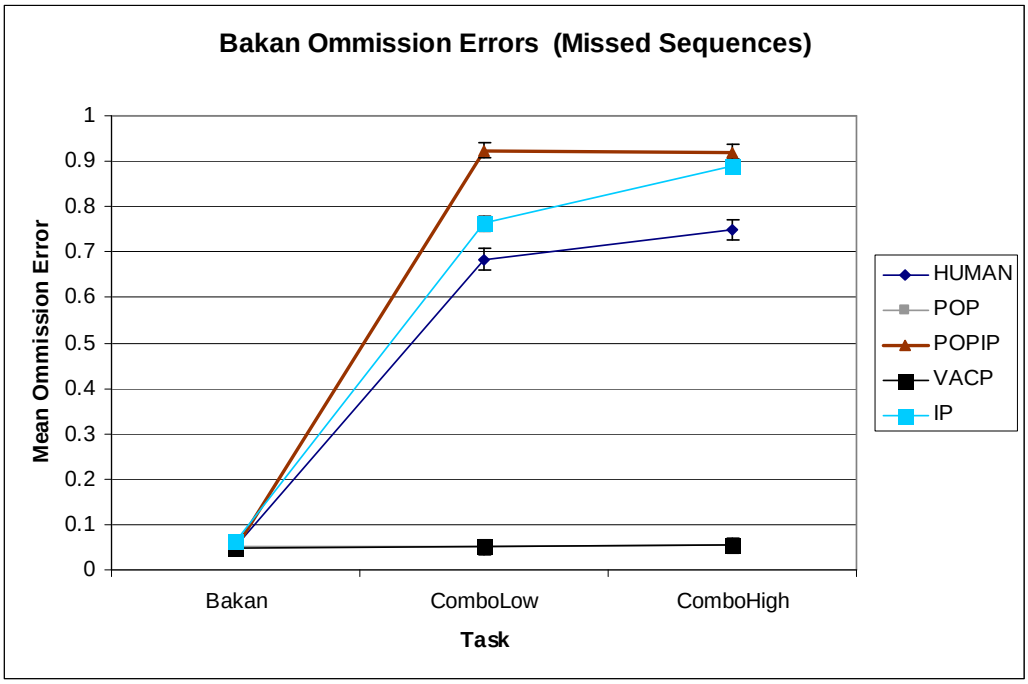


Figure 15 Mean for Visual Bakan Missed Target Sequences for the Human participants, POP, POPIP, VACP and IP.

As can be seen from Figure 15, all groups performed equally well in the Visual Bakan single task condition. Indeed, all simulation models are very accurate in predicting human performance in the single Visual Bakan condition. The pattern of omission errors vary however among the two dual-task conditions. The trend in performance among IP and POPIP models reflect a similar trend seen among human participants. Both models,

however, over-predict human errors. The performance among the POP and VACP models, meanwhile, does not vary across any condition and as a result, greatly under-predicts human errors.

The second analysis was performed on the Visual Bakan commission errors (number of false alarms) of each task condition. The Levene's test for equality of variances revealed that there was a violation on the assumption of homogeneity of variances on the mean omission errors. The Greenhouse-Geisser's epsilon adjustments were made to assume sphericity. A main effect of task condition, $F(2,202) = 23.364$, $p<.001$, and a main effect attributable to group, $F(4, 101) = 21.245$, $p<.001$ was observed. A significant condition by group interaction effect was also found, $F(8,202) = 5.573$, $p<.001$ (See Appendix P for SPSS output tables).

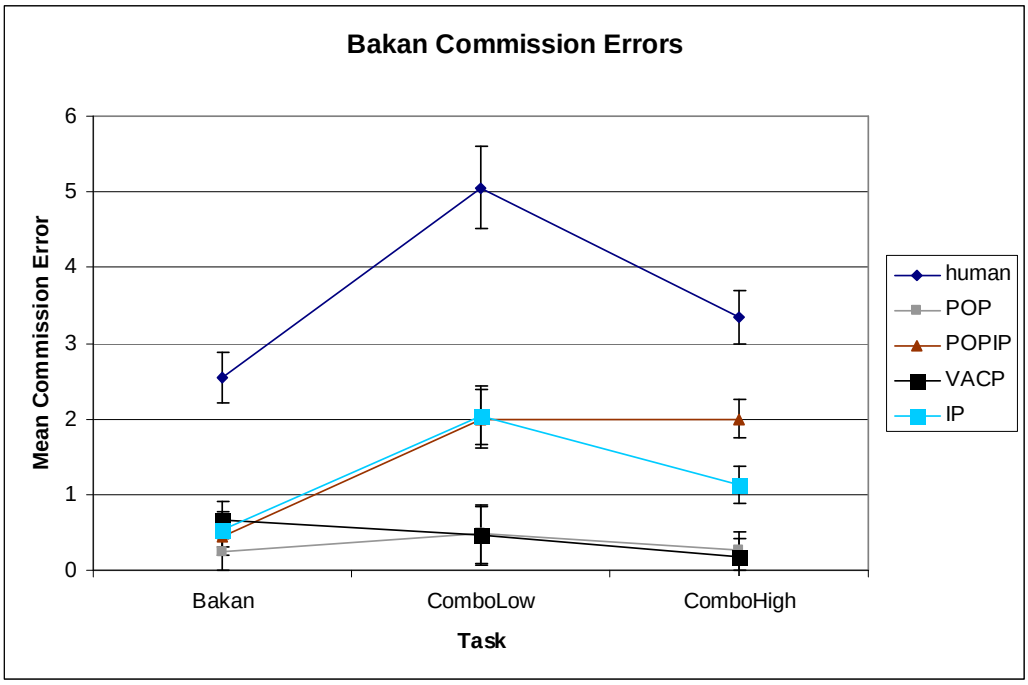


Figure 16 Mean for Bakan commission errors for the Human participants, POP, POPIP, VACP and IP.

It can be seen in Figure 16 that all models performed significantly better compared to human participants, showing fewer false alarms. It can also be seen that the pattern of errors vary as a function of task condition and group. It is clear that only IP follows a similar trend across task conditions as human participants. POPIP follows a similar trend as human for the Bakan and the low workload dual-task conditions but then deviates from humans in the high workload dual-task condition. VACP and POP models meanwhile do not deviate much across task condition.

In summary, these results indicate while IP and POPIP follow similar trends in omission (miss) and commission (false alarm) errors across task conditions as human participants, the VACP and POP models differ substantially in trend. Indeed, commission and omission errors among the VACP and POP models do not deviate across conditions.

Interestingly, while all models were very accurate in predicting human omission errors made in the Bakan single task condition, they all under-predict commission errors in the same task condition. Indeed, all models performance under-predict commission errors across all conditions. These results must be considered recalling that no underlying cognitive model of the information processing was used and that the results were derived from a stochastic engineering model of human error matching the pattern of stimuli to the Visual Bakan target.

3.2.3 Summary of Performance in the ATC and Bakan Tasks

Overall, all models underpredicted human performance in the ATC and Bakan tasks. Accuracy in performance prediction, however, varied as a function of type of error (ATC misdirection, omissions and commissions), type of model, and task condition. In the ATC task, all models greatly under-predicted human performance. The trend in performance across task condition was inconsistent with human performance for all models. In the Bakan task, all models were very accurate in predicting human omission errors made in the Bakan single task condition. In the dual-task conditions, IP and POPIP were fairly accurate in predicting omission errors but still over-predicting. Performance among VACP and POP meanwhile does not change across task conditions. In terms of commission errors, all models under-predicted human performance. Again, IP and POPIP were fairly accurate in predicting commission errors as a function of trend across tasks, but in this case, under-predicting human performance. Again, performance in commission errors among VACP and POP did not change across task conditions.

In summary, these results indicate IP and POPIP models predicted human performance in the Bakan more accurately than VACP and POP. All models were equally inaccurate in predicting ATC performance.

3.3 Comparison of IPME Version 3 vs Version 4 Models

The following section presents the results of IPME V3 and V4 comparison. A simple balances analysis of variance (ANOVA) was used to determine whether there were any significant differences between IPME 3 implementation of the POP model and IPME 4 approximation (POPIP_POP).

3.3.1 POP Input

The ANOVA table for POP input is given in Table 9. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 17 shows a difference between scheduling modes for POP demand under dual task conditions but no difference under single task conditions. In all 5 conditions, a sample of 25 subjects was used to populate the datasets.

Table 9 ANOVA table for POP Input

Factor	Degrees of freedom	F Value	P(F)
IPME version	1, 345	126.9	<0.001
Condition	4, 345	>999.9	<0.001
IPME version x Condition	4, 345	36.4	<0.001

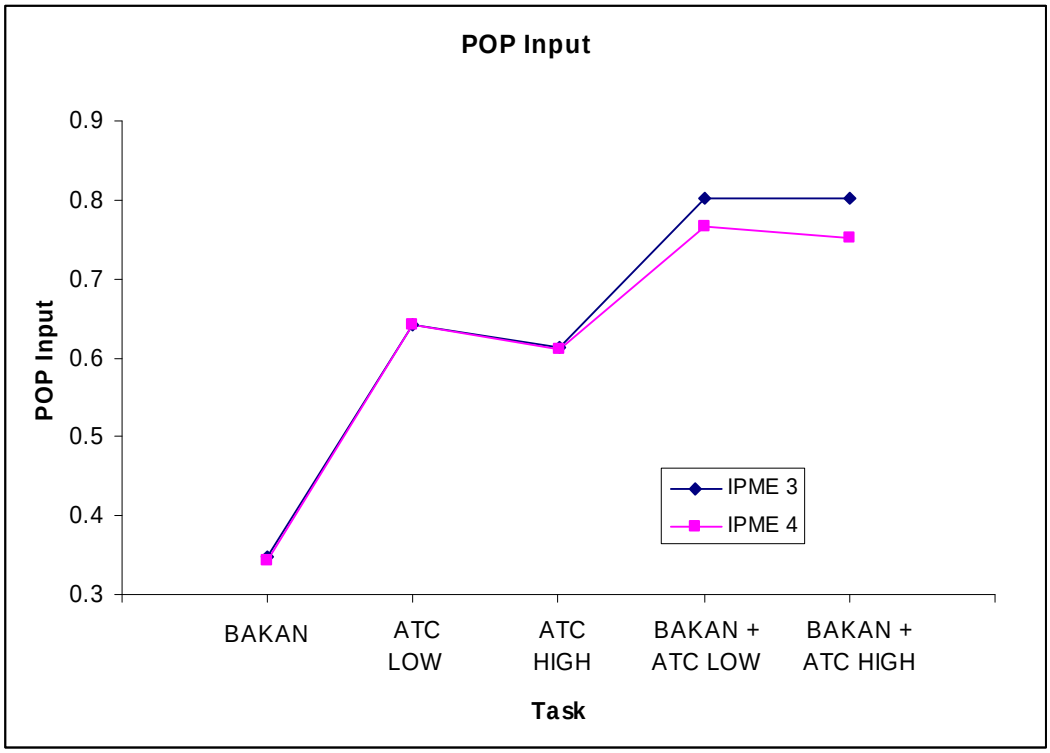


Figure 17 Mean POP Input by task and version

3.3.2 POP Central

The ANOVA table for POP central is given in Table 10. There are significant differences between the scheduling modes and the condition. Figure 18 displays little difference between IPME versions for the predicted Central demands.

Table 10 ANOVA table for POP Central

Factor	Degrees of freedom	F Value	P(F)
IPME version	1, 345	11.0	0.001
Condition	4, 345	>999.9	<0.001
IPME version x Condition	4, 345	0.6	0.647

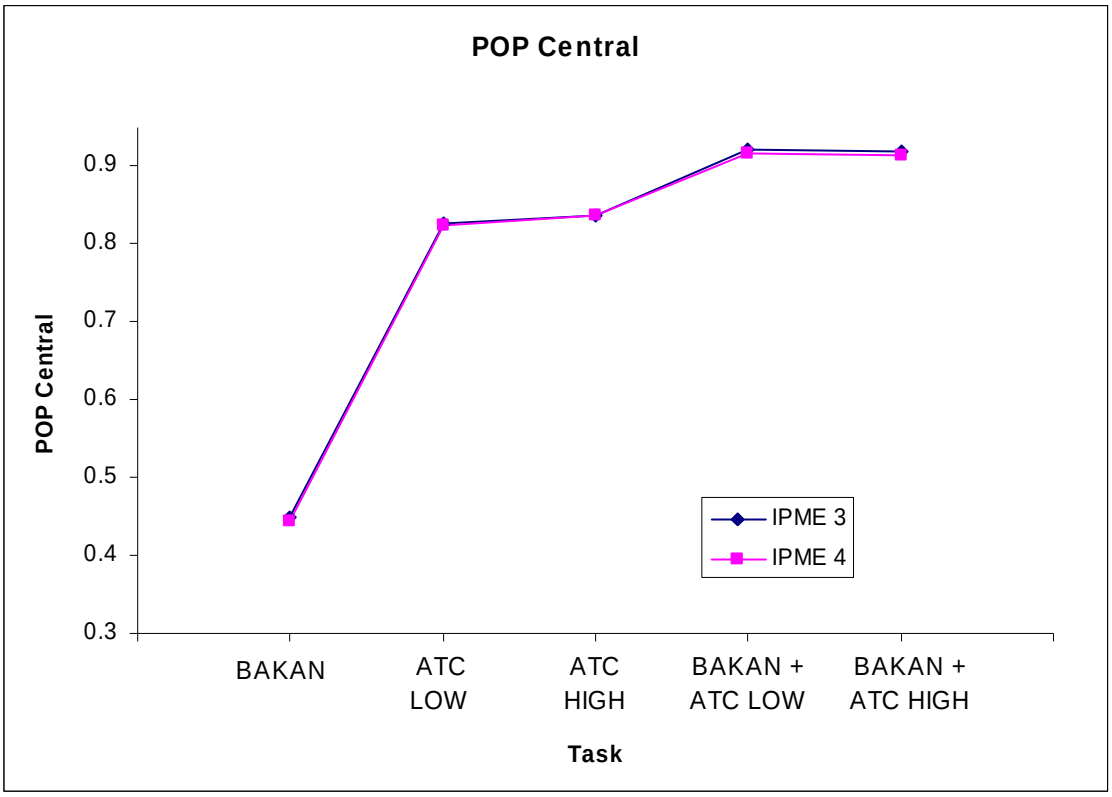


Figure 18 Mean POP Central by task and version

3.3.3 POP Output

The ANOVA table for POP input is given in Table 11. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 19 shows the difference for POP Output in the dual task conditions but little difference in the single task conditions.

Table 11 ANOVA table for POP Output

Factor	Degrees of freedom	F Value	P(F)
IPME version	1, 345	80.9	<0.001
Condition	4, 345	773.8	<0.001
IPME version x Condition	4, 345	34.8	<0.001

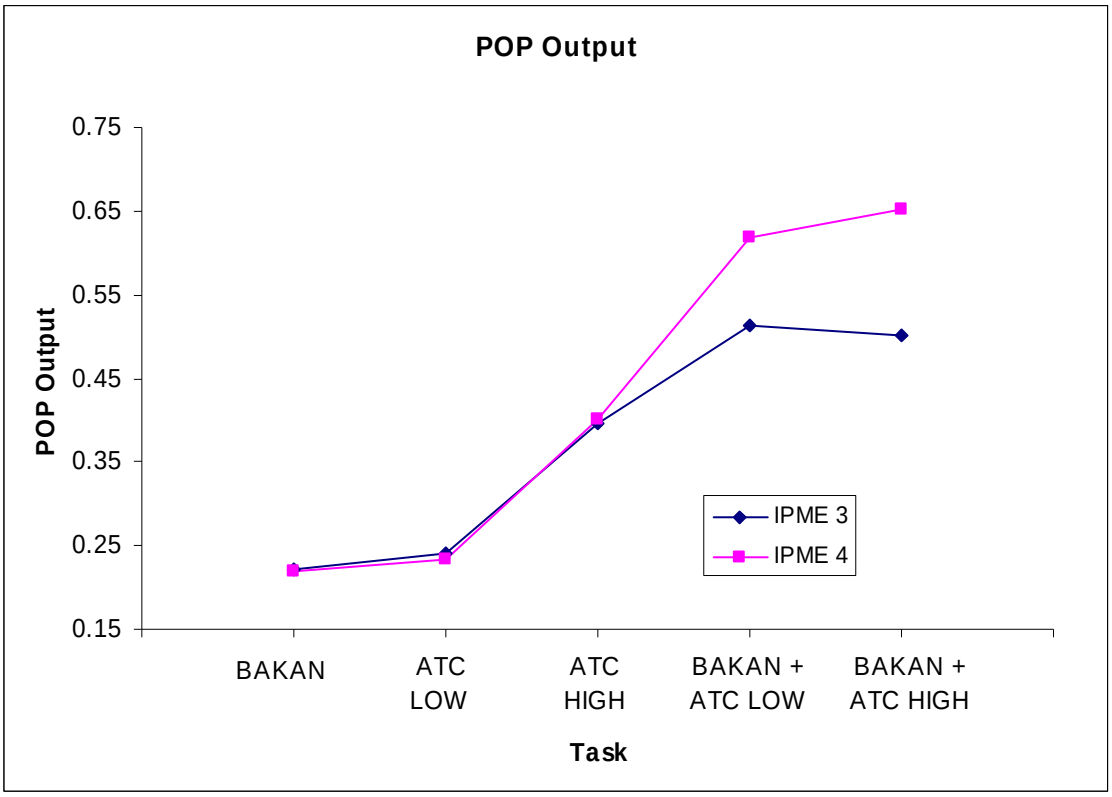


Figure 19 Mean POP Output by task and version

3.3.4 Response Time

The ANOVA table for POP input is given in Table 12. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 20 displays the mean workload graphically.

Table 12 ANOVA table for response time

Factor	Degrees of freedom	F Value	P(F)
IPME version	1, 345	762.4	<0.001
Condition	4, 345	>999.9	<0.001
IPME version x Condition	4, 345	239.6	<0.001

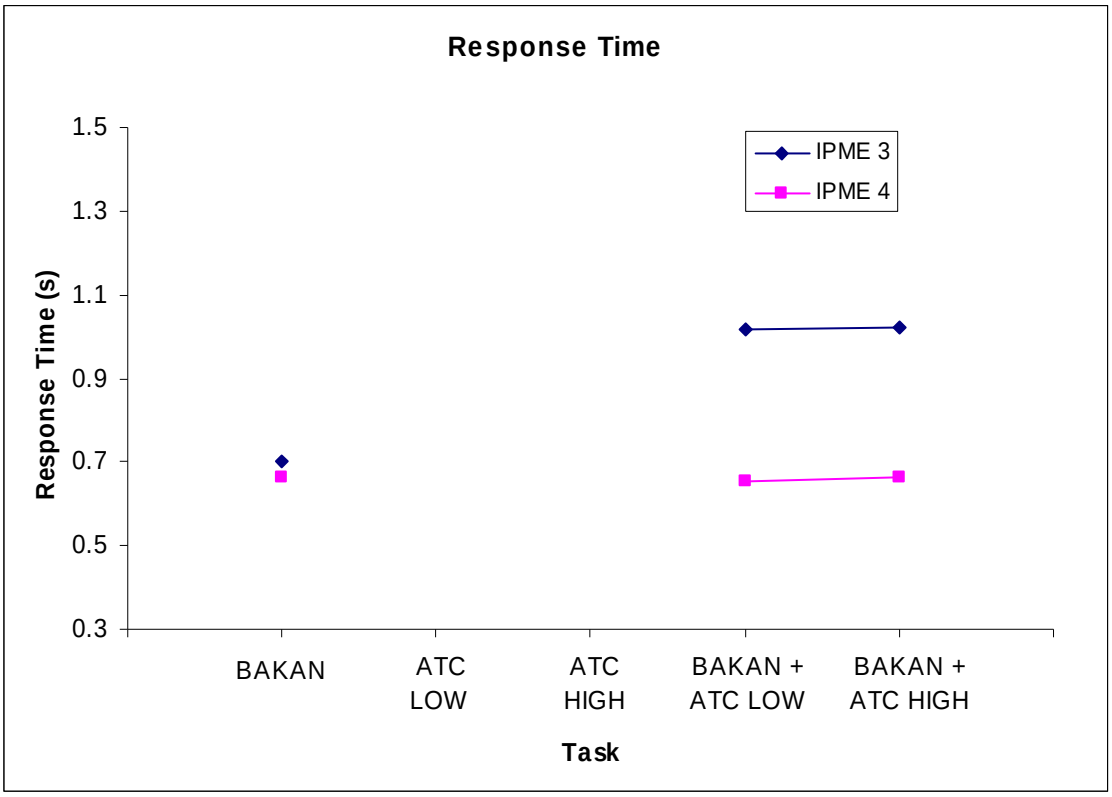


Figure 20 Mean response time by task and version

3.3.5 Miss rate

The ANOVA table for POP input is given in Table 13. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 21 displays the mean workload graphically.

Table 13 ANOVA table for miss rate

Factor	Degrees of freedom	F Value	P(F)
IPME version	2, 345	134.2	<0.001
Condition	4, 345	>999.9	<0.001
Scheduler x Condition	8, 345	81.9	<0.001

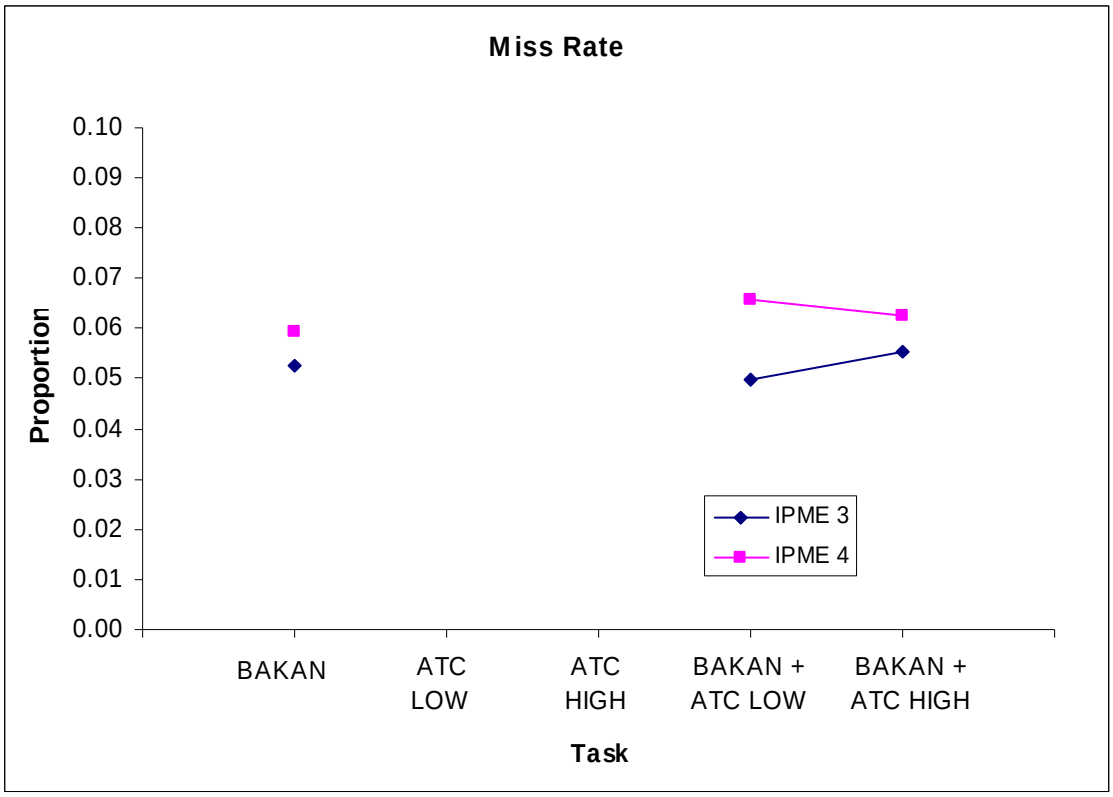


Figure 21 Miss rate by task and version

3.3.6 Conclusion

This analysis has shown that there are significant differences between the model in IPME 3.0.25 and IPME 4.1.3, and these are larger for the dual-task conditions. This suggests that the main difference between IPME 3.0.25 and IPME 4.1.3 is in the interference and scheduling rather than workload calculation. When IPME 4 was created, a significant effort was expended in ensuring that the POP calculation was consistent between the two versions and we have a significant amount of experience in translating models from IPME 3 to 4. This would indicate that the difference between the two versions is due to the implementation of scheduling in IPME 4.1.3.

3.4 Analysis of POPIP modes

Due to the evolution of IPME from Version 3 to Version 4, and the integration of the previously separate IP/PCT and POP algorithms, and assessment of the different IPME scheduler modes within the IPME V4 was desired to determine the degree to which tuning of the scheduler algorithms would produce differences in model performance. For the purposes of this report, the data presented herein should be viewed as a preliminary description of the differences in model data across scheduler modes, and should not be used to draw definite conclusions as to the manner in which the workload algorithms have been implemented.

A simple balanced ANOVA was used to explore the differences between the POPIP scheduling modes in IPME 4. A mixed effect linear model, containing the following main effects and interactions:

1. a fixed effect of “scheduler” (POPIP_default, POPIP_IP, POPIP_POP and POPIP_POP_SHED);
2. a fixed effect of “condition” (Bakan only, ATC low, ATC high, Bakan and ATC low, Bakan and ATC high);
3. the interaction between “scheduler” and “condition”, and
4. a random effect of “crew_sample”.

The dependent variables in the model were:

1. mean POP input;
2. mean POP central;
3. mean POP output;
4. mean IP time pressure;
5. Bakan reaction time; and
6. Bakan hits and misses.

3.4.1 POP Input

The ANOVA table for POP input is given in Table 14. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 22 displays the mean workload graphically.

Table 14 ANOVA table for POP Input

Factor	Degrees of freedom	F Value	P(F)
Scheduler	2, 345	134.2	<0.001
Condition	4, 345	>999.9	<0.001
Scheduler x Condition	8, 345	81.9	<0.001

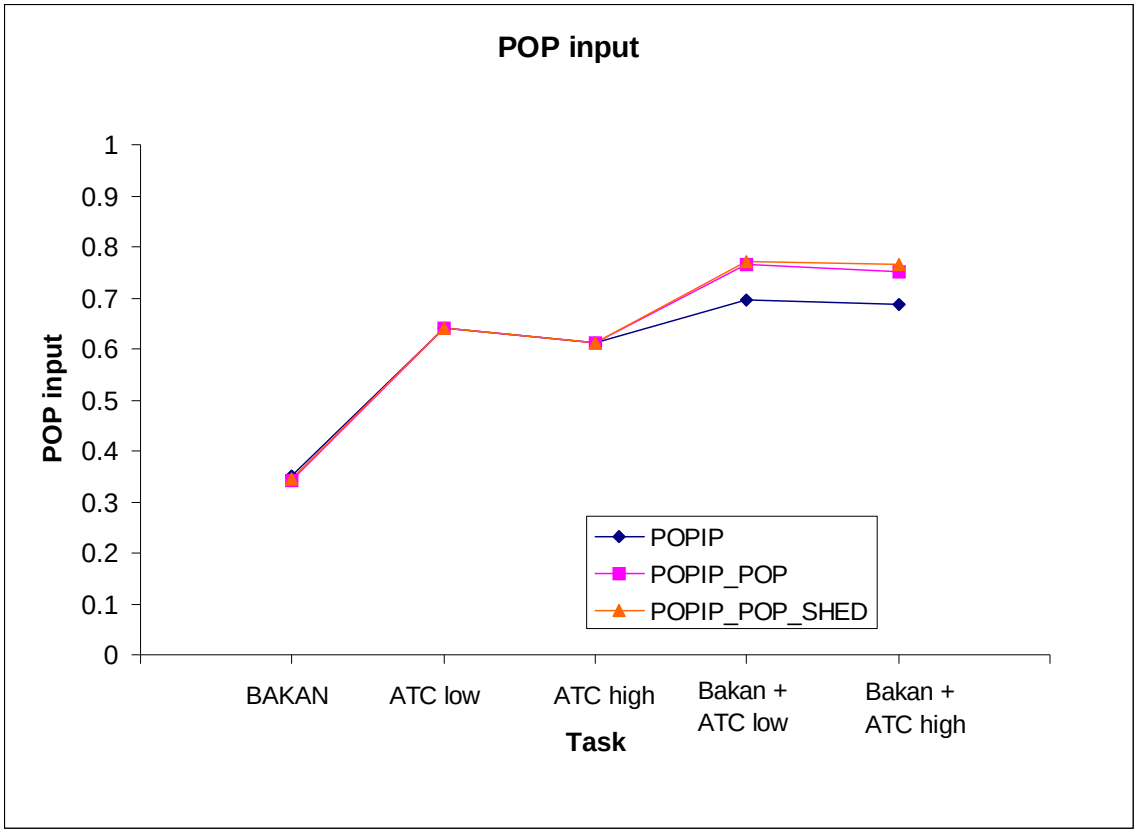


Figure 22 Mean POP input value by task and scheduling mode

The POP and POP_SHED modes are similar, but they differ from the POPIP_default mode in the dual task conditions. The most likely explanation for the difference is the inclusion of the visual interference model in the POPIP_default mode. The visual interference model represents the interference between two tasks based on the physical separation of the tasks in the operator's field of view. A large angle of separation would mean that two visual tasks could not be performed at the same time. This effect could result in a lower mean input workload in the dual task condition as visual tasks are now being processed serially rather than in parallel.

3.4.2 POP Central

The ANOVA table for POP central is given in Table 15. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 23 displays the mean workload graphically.

Table 15 ANOVA table for POP central

Factor	Degrees of freedom	F Value	P(F)
Scheduler	2, 345	179.3	<0.001
Condition	4, 345	>999.9	<0.001
Scheduler x Condition	8, 345	100.3	<0.001

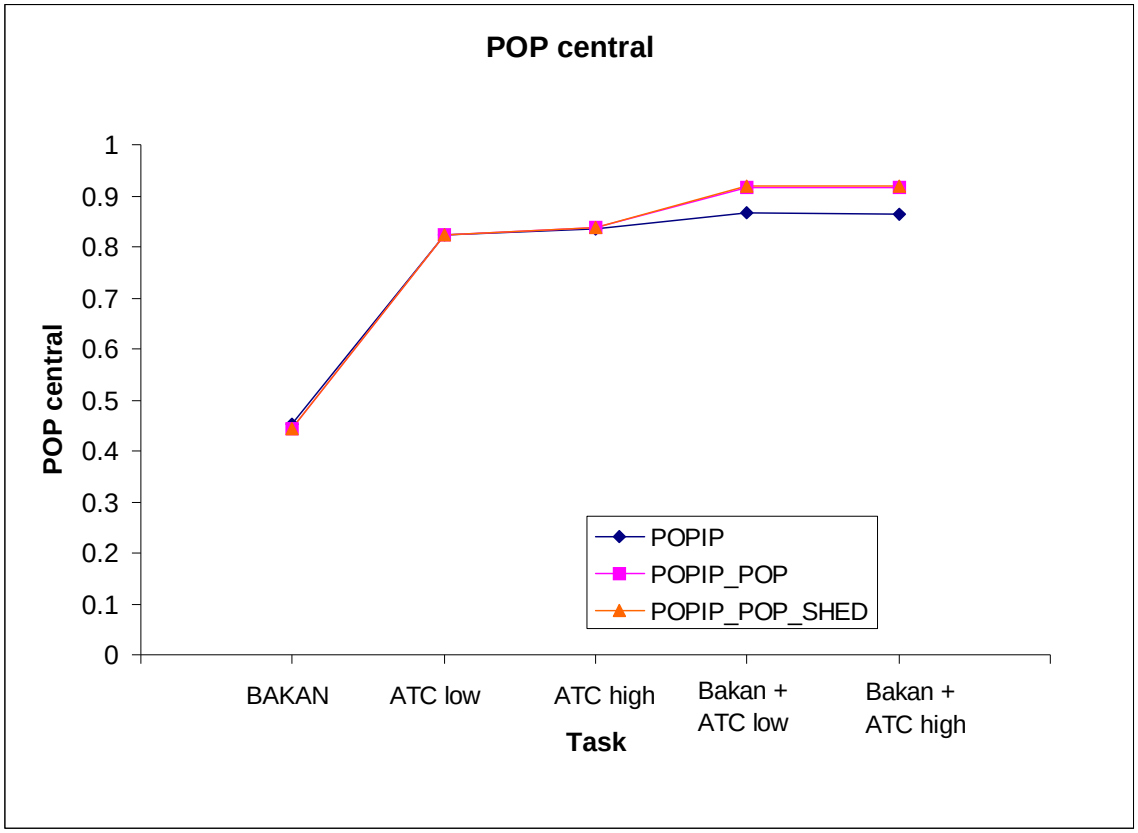


Figure 23 Mean POP central value by task and scheduling mode

3.4.3 **POP Output**

The ANOVA table for POP output is given in Table 15. There are significant differences between the scheduling modes, the condition and a significant interaction between scheduler and condition. Figure 24 displays the mean workload graphically.

Table 16 ANOVA table for POP output

Factor	Degrees of freedom	F Value	P(F)
Scheduler	2, 345	133.6	<0.001
Condition	4, 345	925.9	<0.001
Scheduler x Condition	8, 345	46.8	<0.001

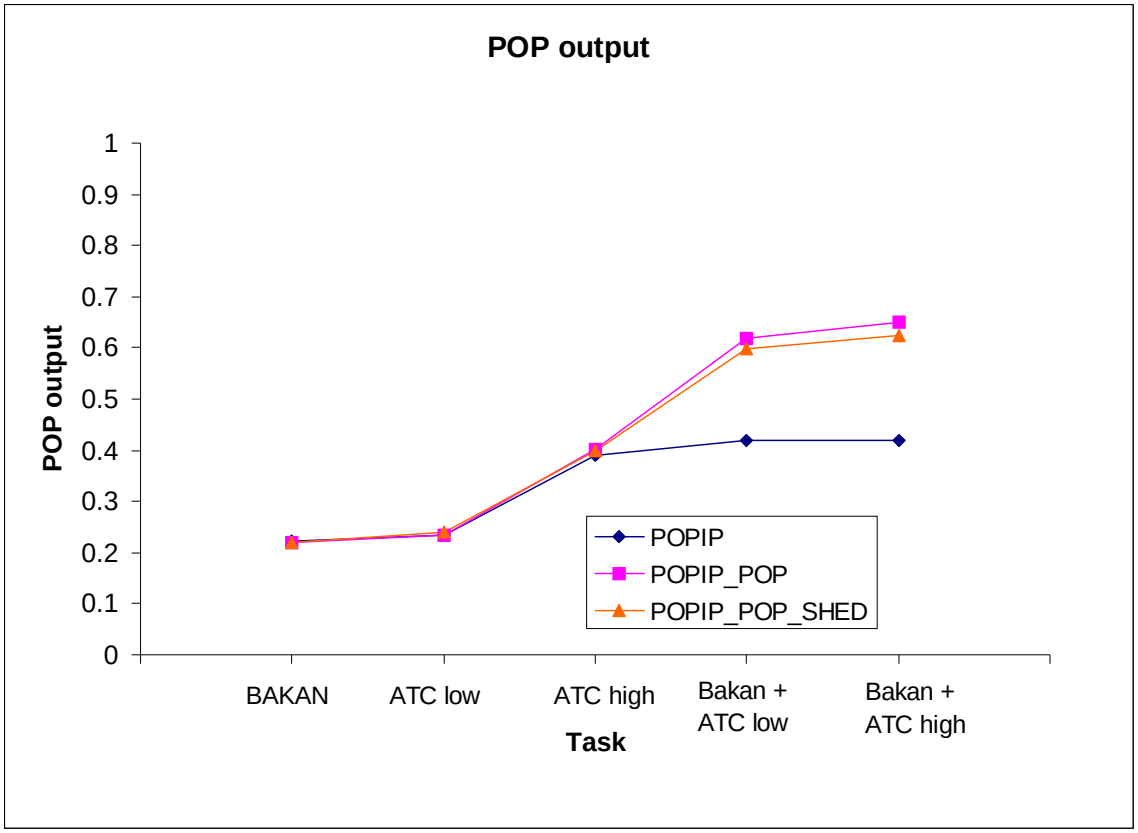


Figure 24 POP output value by task and scheduling mode

3.4.4 IP Time Pressure

The ANOVA table for IP time pressure is given in Table 17**Error! Reference source not found.**. There is a weak difference between the scheduling modes, a strong difference between the condition and a significant interaction between scheduler and condition. Figure 24 displays the mean time pressure graphically.

Table 17 ANOVA table for IP time pressure

Factor	Degrees of freedom	F Value	P(F)
Scheduler	3, 460	3.62	0.013
Condition	4, 460	88.59	<0.001
Scheduler x Condition	12, 460	2.72	0.002

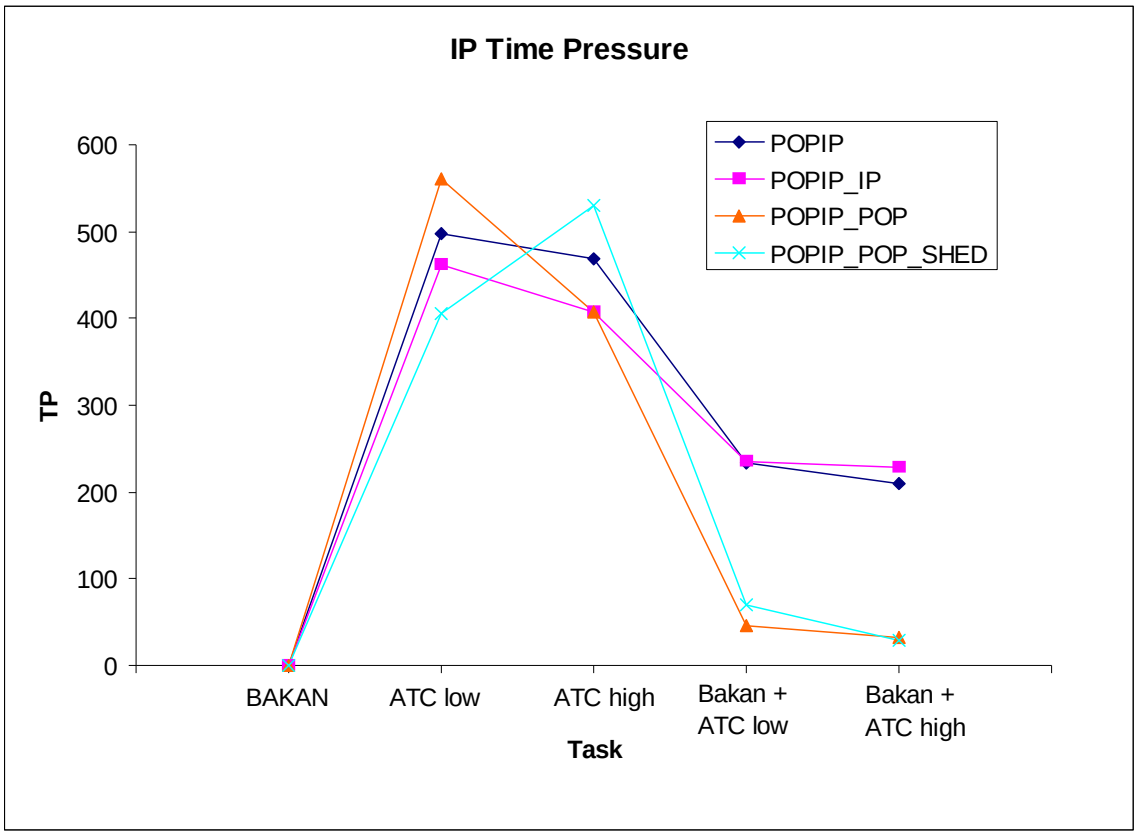


Figure 25 ANOVA table for POP output

3.4.5 Response Time

The ANOVA table for the Bakan response time is given in Table 18. The only significant effect is between conditions.

Figure 26 displays the mean response time graphically.

Table 18 ANOVA table for response time

Factor	Degrees of freedom	F Value	P(F)
Scheduler	3, 460	1.4	0.254
Condition	4, 460	>999.9	<0.001
Scheduler x Condition	12, 460	1.0	0.408

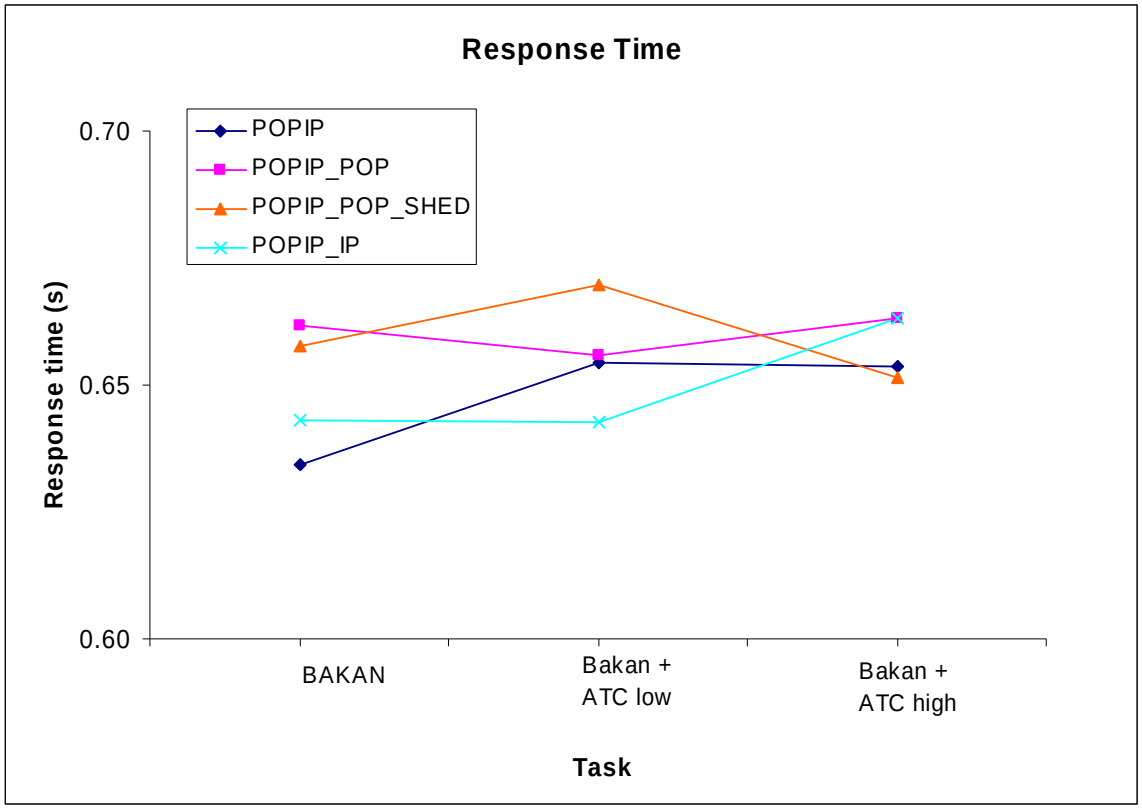


Figure 26 Response time by task and scheduling mode

3.4.6 **Bakan Probability of Miss**

The ANOVA table for the Bakan probability of missing a target string is given in Table 19. There are significant effects of scheduling mode, condition and the interaction between mode and condition. Figure 27 displays the mean probability of missing a target string graphically.

Table 19 ANOVA table for the Bakan probability of missing a target string

Factor	Degrees of freedom	F Value	P(F)
Scheduler	3, 460	>999.9	<0.001
Condition	4, 460	>999.9	<0.001
Scheduler x Condition	12, 460	>999.9	<0.001

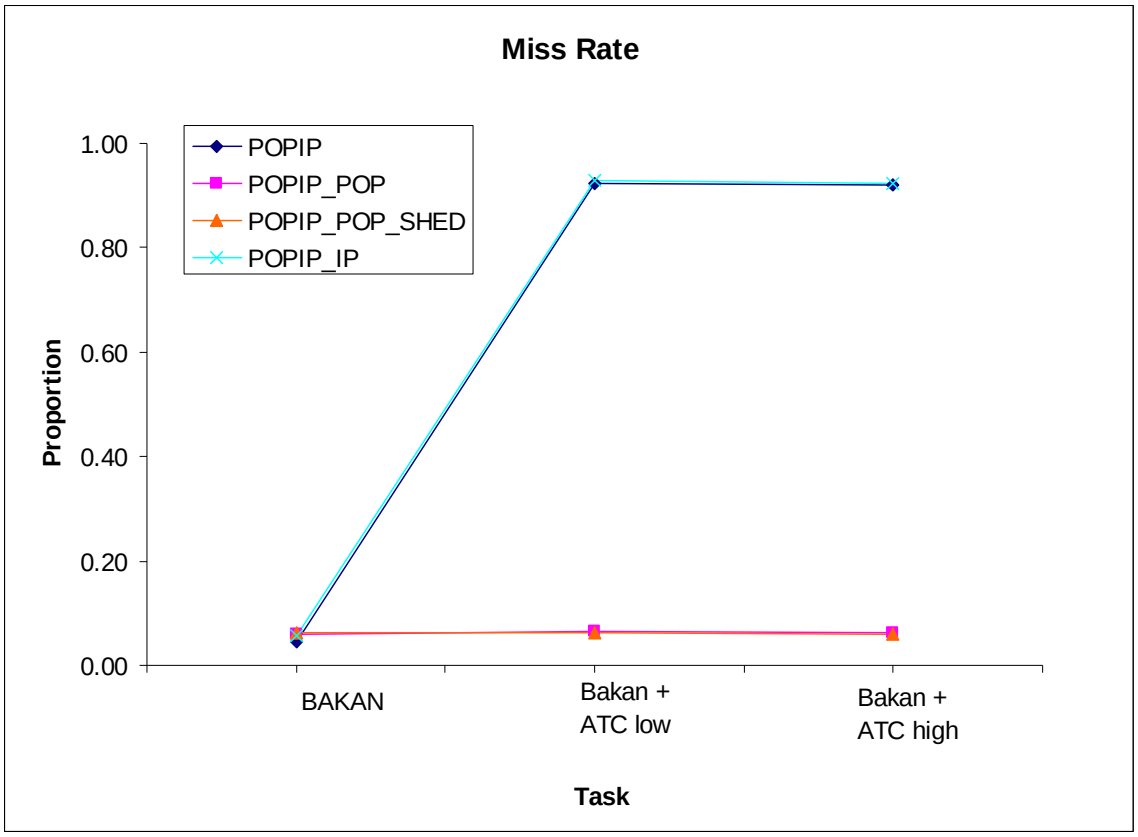


Figure 27 Probability of missing a target string by task and scheduling mode

There is a dramatic separation in the dual task condition between those scheduling modes that contain IP elements (POPIP and POPIP_IP) and those that do not (POPIP_POP and POPIP_POP_SHED).

3.4.7 Summary

This analysis has shown that there are significant differences between the POPIP scheduling modes that include the IP scheduling elements and those that do not. The most noticeable differences occur in the dual task conditions and this was most likely due to the inclusion of the visual interference model. Comparison with the observed results indicated that the POPIP mode closely reflected the magnitude and structure of the empirical data; however, to replicate the observed values a model in between the POP model and the POPIP default model would have been most suitable.

4 Discussion

This study was conducted to systematically compare the workload algorithms within IPME for evaluating the predictions generated from the models. Specifically, the research addressed the following issues: (1) Is there a difference between participants' subjective workload perceptions and the simulated workload? If so, which aspects of the human performance are most or significantly different from a comparable simulated performance? How does human perception differ from the prediction of subjective workload produced by the IPME models? (2) Is there a difference between participants' performance on the ATC and Bakan study tasks and the simulated performances? If so, which aspects of the performance are most different? And (3) Are there differences in predicted workload and performance in IPME V3.0.25 vs IPME V4.1.3.

The following section presents a discussion of the results obtained during the course of the project.

4.1 Human Participants

As expected, subjective workload increased as a function of task condition (Bakan, ATC low and ATC high alone, ATC low/Bakan and ATC high/Bakan combination). All conditions involving ATC were perceived as involving much higher workload than the Bakan task. These results suggest that the ATC task influenced subjective workload greater than the Bakan in the dual task condition.

As expected, ATC low condition was perceived as involving lower workload than the ATC high condition. Interestingly, during the conduct of the study, some participants mentioned that they preferred participating in the ATC high condition compared to the ATC low condition. They found that the interval of 9 seconds was too long a time to wait for the screen to update and they preferred the ATC high condition even though it involved higher subjective workload. This would suggest that for some participants, the long delay between update intervals in the low workload condition may have decreased vigilance to the ATC task or that additional verification or aircraft status was not being attempted.

Despite anecdotal reports, performance in the ATC task generally decreased as a function of task condition. Errors in misdirection landings of planes increased as task demand increased. Interestingly, commission errors actually decreased in the high workload dual-task condition. Participants commented that during the high workload dual-task condition, it became very difficult to simultaneously attend to the ATC and Bakan task. As such, participants may have significantly reduced the number of responses to the Bakan task when compared to the low-workload dual-task condition which accounts for the reduced commission errors in this condition.

In summary, subjective workload among human participants generally increased as a function of task condition. Performance in terms of misdirection errors in the ATC task and omission errors in the Bakan task decreased as a function of task condition. Commission errors meanwhile decreased in the high workload dual-task condition compared to the low workload dual-task condition, and did not differ significantly from the bakan condition.

4.2 IPME Models: Human Subjective Workload Prediction

Similar to human participants, predictive workload increased as a function of task condition (Bakan, ATC low and ATC high alone, ATC low/Bakan and ATC high/Bakan combination) for all IPME models. It was also found that VACP produced significantly lower predictive workload compared to POP and POPIP for both mental and physical workload. POP and POPIP meanwhile exhibited mental and physical workload outcomes similar to humans. These results suggest that POP and POPIP more accurately predict human subjective workload compared to VACP.

The pattern in which mental and physical workload increased depended on the IPME model and task condition. It was found that while all models are fairly accurate in predicting human mental workload in the Bakan and Combination task conditions, POP and POPIP over-predict mental workload in the ATC alone condition compared to humans, while VACP greatly under-predicts it. In regards of physical workload, the accuracy of POP and POPIP differ as a function of task condition. While POP is very accurate in the ComboHigh condition, POPIP is very accurate in predicting physical workload in the ComboLow condition. In addition, POP over-predicts in the Bakan, ATC High and ComboLow conditions, and under-predicts physical workload in the ATC Low condition compared to human participants. POPIP, meanwhile, over-predicts physical workload in the Bakan and ATC High conditions and under-predicts in the ATC Low and Combo High conditions compared to humans.

The trend in workload seen in W/Index is similar to the trend seen in VACP cognitive values (see Figure 11) and Humans' composite NASA/TLX (see Figure 12) whereby the workload is significantly higher in the Combo Low and High conditions compared to the Bakan condition. W/Index differs from VACP and Humans whereby W/Index workload values decrease in the ATC low and high conditions compared to Bakan rather than increase.

Of note is that the impact of the performance schedulers within IPME becomes apparent when examining the trends in W/Index and VACP data compared to the POP and IP models and related variants. As neither the W/Index nor VACP models account for the impact of time pressure, the increase in workload associated within increasing temporal demands across low and high ATC and dual-task workload conditions has no effect on W/Index or VACP predictions. It is only during the change from a single- to dual-task condition that the effect of increasing task-demands by virtue of the simultaneous processing of ATC and Bakan tasks is predicted via W/Index and VACP. These results clearly demonstrate that state-based workload algorithms will generally under-predict workload when the source of variance is temporal in nature.

In summary, while POP and POPIP more accurately predict human subjective workload compared to VACP and W/Index, all four models followed similar trends to human subjective workload as a function of task condition, particularly for the Bakan and the dual-task conditions. Interestingly, the greatest difference seen among models for predictive human subjective workload occurs in the two single-task ATC conditions. This suggests that the task-based workload values for the ATC tasks are underepresentative of human subjective workload or that the ATC TNM used was inadequate for capturing the effects experienced by the subjects under similar conditions.

4.3 IPME Version Differences

The results of IPME version analysis demonstrated that there are differences between the workload and performance outputs of IPME V3.0.25 vs. IPME V4.1.3. The majority of these differences occurred within the dual ATC and Bakan conditions. These results suggests that the main difference between IPME 3.0.25 and IPME 4.1.3 is in the interference and scheduling algorithms rather than the workload calculation. When IPME 4 was created a significant effort was expended in ensuring that the POP calculation was consistent between the two versions and the project team has a significant amount of experience in translating models from IPME 3 to 4. This would indicate that the difference between the two versions is due to the implementation of the IP scheduling component of IPME 4.1.3. Future research is required to determine if the integration of the IP scheduler serves to better predict operator performance when compared to the POP implementation alone.

4.4 Predicting Human Performance

The results presented in this report indicate IP and POPIP models predicted human performance in the Bakan task more accurately than VACP, W/Index and POP. While IP and POPIP follow similar trends in omission and commission errors across task conditions as human participants, the VACP, W/Index and POP models differ substantially in trend. Indeed, commission and omission errors among the VACP, W/Index and POP models do not deviate across conditions. In addition, while all models were very accurate in predicting human omission errors made in the Bakan single task condition, they all under-predict commission errors in the same task condition. The models also under-predict commission errors across all conditions. Accuracy in predicting ATC performance, meanwhile, was equally inaccurate for all models.

The results from this analysis provide insights into the challenges of accurately predicted performance-based measures and demonstrate the degree to which model tuning activities must be conducted to provide a stable and representative performance-based dataset. When compared to predictions of operator workload, performance-based predictions were significantly poorer. The inaccuracies in prediction can be directly attributed to the minimal performance tuning activities that occurred during the ATC model development process and the many limitations that were inherent to modelling the performance based aspects of the ATC task. When the ATC and Bakan task were then combined in the dual-task conditions, the limitations of the ATC model likely interacted with the Bakan model to further reduce both ATC and Bakan performance predictions.

Of note is that model predictions across the different algorithms for omission errors in the Bakan single-task condition were not substantially different from the human data. The accuracy with which IPME models of the Bakan task accurately predicted human performance is directly attributable to the significant tuning activities that had been applied to the development of the Bakan model by both the DRDC Toronto and QinetiQ project teams.

In the context of task-representations at a high-level of fidelity as in the case of the Bakan and ATC, tuning activities must be incorporated as a critical component of the model development activity if accurate performance-level representations are required. In contrast, the results of this study provide an initial indication that workload predictions

are perhaps less sensitive to tuning requirements if the task-sequences and task demand characteristics are relatively accurate. Further, it is important to base the model development and tuning on observations that span the range of operator behaviours so that the appropriate representation is used for predictive validation.

4.5 Theoretical Modelling Issues

4.5.1 Externally Cued Visual Detection Tasks

The original intent of the externally-cued visual detection task in accordance with the IP/PCT theory was to provide a representation of pre-attentive processes associated with the initial orientation of the sensory system to a visual signal within the environment. Upon discussions with DRDC Toronto and a review of IPME documentation, it became clear that the current representation of the externally-cued visual detection task within IPME is not immediately intuitive. If a single stimulation event occurs, an IPME Externally Cued task may be scheduled once as with any other discrete task. If, however, the stimulation persists, the Externally Cued task must be set up as a repeating task that persists until either one of two events occurs: the stimulus is detected or the stimulus ceases. The probability of detection of the stimulus is considered independent among the instances of the Externally Cued task, although a debate remains about the most appropriate means of specifying appropriate data for such tasks or whether a better representation can be implemented in the IPME discrete event framework.

4.5.2 Reduced Model Variability.

During the analysis of human and predicted workload measures it became apparent that the between-subject variability in the predicted workload measures was significantly smaller than that observed in the corresponding subjective NASA TLX measures. This is a common problem with human modelling as it is difficult to capture all factors that might affect performance variability in a model even if the essential aspects are well represented. One aspect that is not well implemented in IPME is how task demands are assigned for some of the workload models. In each case, a fixed demand rating is assigned to a given task based on the assumed nature of the cognitive or psychomotor processing associated with the specific task. As such this method of task-demands assignment provides no source of variability for between-subject differences in perceived workload. By its very nature, the corresponding NASA TLX data is assessed by human participants across experimental conditions, and incorporates between-subject variability.

To accurately predict workload, it would appear to be desirable to include some measure of the between-subject variability normally associated with human participants in the outputs of workload models. IPME could manage a task based, variable rating system through operator traits and states if sufficient data exists to quantify this source of variability. Indeed, attempts to model potential transitions of non-skilled to expert behaviour in a given task domain would likely require a methodology for representing the variations in perceived workload across novice and expert users on a task-by- task basis. The Visual Bakan developed by DRDC Toronto provides an example of how this could be done if the IPME task definition interface was modified.

In addition, the impact of extremely low variances within and across model data establishes a condition whereby ANOVA type analysis will often indicate significant differences between group means (e.g. some of the mean differences between IPME V3

and V4 models) even when mean values are very similar. Cross-model comparisons then become very difficult to interpret through the use of variance-analysis methodologies.

4.5.3 Regression Analysis and Workload vs Performance Comparisons

As previously indicated, there were many limitations inherent to the modelling activity that precluded the development of an accurate predictive performance component to the ATC models. It was therefore deemed inappropriate to conduct a regression analysis on the predicted workload and performance data. However, future efforts to better represent operator performance in the ATC task would benefit from a more thorough analysis.

4.5.4 IP/TCP Time Pressure

The Time Pressure values generated from the IP model were omitted from the analysis between human and model data. Upon an initial examination of the data output for IP, it became clear that the Instantaneous Time Pressure (ITP) values could be extremely large (several orders of magnitude) and highly variable. These large values would dominate the calculation of a mean Time Pressure, even though it was present momentarily and its contribution weighted accordingly, rendering these results un-interpretable in the context of comparisons to subjective workload ratings. ITP values were generated for IPME V3 and V4 comparisons to determine whether these inflations were observable across different implementations of the IP model and this was found to be the case.

The IP/PCT multiplier that determines the latest time for completion for a task was calculated dynamically for the Visual Bakan task and this was adopted for the ATC task as well. In instances where the time available for task completion was extremely small compared to the time required to complete the task, the instantaneous time pressure metric would produce extremely large and unrealistic values. As a result, these relatively brief moments of excessive instantaneous time pressure significantly shifted mean time pressure measures towards extreme values, thereby making interpretation of the mean time pressure metrics impossible.

Changes to both the ATC modelling approach as well as changes to the IP implementation within IPME will likely be required, and include the following recommendations:

1. The IP time pressure calculation within IPME implementation should notionally include a limit on obtained instantaneous time pressure values to reduce the likelihood that extreme values will negatively affect mean time pressure calculations. Indeed, Curry et al. (1979) assert that time pressure should never exceed 1, although this constraint is not imposed by the IP/PCT model. Such limit must be developed in conjunction with a theoretical justification for applying an absolute time pressure metric that is comparable to subjective interpretations of time pressure constraints on performance.
2. The current implementation of the IP multiplier calculation within the ATC mode is likely not a valid representation of the factors affecting human performance. In addition to defined limits on instantaneous time pressure, the variables used to determine task multipliers should be sensitive enough to reflect modifications to workload (i.e., update interval, number of planes). Given that small changes in the ratio of time available to time required can have a large impact on task deadlines,

further research must be conducted to determine appropriate variables to incorporate into the dynamic multiplier calculations within the ATC model.

3. Future models should consider the influence of overall task status and its relationship to time pressure in determining the resultant affects on behaviour. Time pressure in the current isolated ATC model has a limited impact on operator performance, as it is the set of logical rules driving aircraft selection that ultimately determines success in the ATC task. This suggests that goal states that can influence aircraft selection such as the proximity to a correct track or the desire to prevent an aircraft from leaving the radar should also have some impact on time pressure. The mechanism by which this can occur must be investigated further.

4.5.5 W/INDEX Implementation

During the model development activity, it was determined that the W/Index algorithm implementation within IPME 3.0.25 was incorrect, prompting the release of new version of IPME (V3.0.30) to address this issue. However, it was also identified that the automated VACP to W/Index mapping scheme was not founded on a theoretical basis. For example, a VACP value in the cognitive domain would either map to a W/Index Spatial or Verbal Cognition channel in a seemingly arbitrary manner. The existing IPME documentation also does not provide a description of the algorithm used to support the mapping process, and the VACP to W/Index mapping table provided in the documentation does not reflect the mappings that have been implemented. In addition, discussions with Micro Analysis & Design have not clarified this issue any further and future work must be conducted to determine appropriate justifications for the VACP to W/Index mapping algorithms. As such, the W/Index ratings used within the ATC and Bakan models were manually assigned to specific W/Index channels based on assumptions made about the relative information processing demands of a given task.

5 Conclusions

This report has documented the methodology, results and final conclusions of IPME Workload Validation Project for DRDC Toronto.

The results of this validation activity have provided a clear way-ahead for the conduct of future model development, implementation and validation activities for IPME modellers and Human Factors practitioners. These recommendations are discussed in the following sections.

5.1 Extensions to Validation Activities

The following recommendations are provided to extend and improve the current validation activities surrounding the ATC and Bakan task environments:

1. Conduct future human-in-the-loop ATC experiments across a greater range of workload conditions to provide a much greater representation of operator task behaviours and decision processes in the ATC environment.
2. Use the results from future ATC experiments to provide direct inputs into a model tuning activity for the ATC IPME model to better represent human-performance, especially in areas identified as weak in-terms of the existing representation within the current models;
3. Upon completion of the ATC tuning exercise, re-integrate the ATC and Bakan models into a dual-task condition to determine if the tuning effort has improved the models ability to predict operator performance across a range of conditions;
4. Conduct a follow-on ATC dual-task study to validate the model predictions; and
5. Integrate the ATC model within the ATC simulation environment to provide an accurate representation of the ATC environment in which the IPME ATC model operates. This integration activity would support a one-to-one match between the human and model performance data via the data-capture capabilities of the ATC simulation environment.

5.2 IPME Development

The following recommendations pertaining to the implementation of the workload algorithms within IPME are supported by the lessons learned over the course of this project:

1. Resolve the W/Index to VACP mapping issue within IPME toolset and supporting documentation. As part of this resolution, a theoretical basis for justifying the W/Index and VACP mappings must be established;
2. Clarify the use of the externally-cued visual detection tasks within the IP and POPIP modes of IPME;
3. Resolve the observed peaks in Instantaneous Time Pressure values within the IP and POPIP implementations. The course of this resolution is not clear at this time, however establishing a notional maximum ITP value may be an interim solution prior to determining the theoretical implications of large ITP values within the workload calculation; and

4. Identify and resolve the source of error within the shed-if-late behaviour.

5.3 Workload Modelling: General

The following recommendations a provided to Human Factors practitioners interested in developing workload models within IPME and are based on our collective experience in model development:

1. In the current IPME implementation of the IP/PCT theory, representation of tasks at varying levels of granularity is problematic and can lead to implausible predictions of behaviour. Integration of IP/PCT with POP into POPIP should address this issue to some degree, although further experience with POPIP is required to ensure this conclusion is valid.
2. The use of purely state-based workload predictions will likely under-predict workload effects due to temporal demands. In single-threaded environments where time-pressure may be a factor, scheduler-based algorithms may be a more appropriate alternative.

6 References

Armstrong, J., Brooks, J., and Barone, J. (2003). *Workload Modeling Support to the Maritime Helicopter Project Colume 2 (Final Report). MH Workload Modelling Results*: Report to the Department of National Defence

Armstrong, J., and Greenley, M. (2003*). Task Network Modelling Services in Support of the FAVS Project: Final Report.* Defence Research and Development Canada (DRDC).

Armstrong, J and Lai, G. (2005). *An evolutionary approach to armoured vehicle modeling*, presented at the 2005 IPME Workshop.

Armstrong, J. and Youngson, G (2004). *Results of the Build 2 Multi Mission Virtual Vehicle (MMVV) Task Network Modelling Effort*. Technical Note to the Department of National Defence.

Belyavin, A. J. and Ryder, C. (2006). *IPME workload algorithms: POP and POPIP models*. QinetiQ report, QinetiQ/06/00166

Belyavin, A. J. and Farmer, E. W., (2006) *Modelling the Workload and Performance of Psychomotor Tasks*. Proceedings of the BRIMS 2006 Conference, Baltimore, *MD, USA*

Card, S. K., Moran, T. P., and Newell, A. (1983). *The Psychology of Human-Computer Interaction*. New Jersey: Lawrence Erlbaum Associates.

Curry, R., Jex, H., Levison, W., & Stassen, H. (1979). *Final report of the control engineering group*. In N. Moray (Ed.), Mental workload: Its theory and measurement (pp. 235-253). New York: Plenum Press.

Farmer, E. W., Jordan, C. S., Belyavin, A. J., Birch, C. L., & Bunting, A. J. (1993). *Prediction of dual-task performance by mental demand ratings (No. DRA/AS/FS/CR93087/1)*. Farnborough, Hampshire: Defence Research Agency.

Farmer, E. W., Belyavin, A. J., Jordan, C. S., Bunting, A. J., Tattersall, A. J., & Jones, D. M. (1995). *Predictive workload assessment: Final report (No. DRA/AS/MMI/CR95100/1)*. Farnborough, Hampshire: Defence Research Agency.

Farmer, E. W., Jordan, C. S., Belyavin, A. J., Bunting, A. J., Tattersall, A. J., & Jones, D. M. (1995). *Dimensions of operator workload: Final report (No. DRA/AS/MMI/CR95098/1)*. Farnborough, UK: Defence Research Agency.

Forbes, K., Darvill, D., Armstrong, J., Banbury, S (2006). *The Evolution and Implementation of Workload Algorithms of Human Information Processing CR 2006-042*. Greenley & Associates, Inc.

Fowles-Winkler, A., Lorenzen, C., Belyavin, A. J., Cain, B., and Hendy, K. (2004*). A comparison of three workload methodologies: POP, IP/PCT and POPIP*. Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting, New Orleans, USA

Hart, S. G., & Staveland, L. E. (1988). *Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research*. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload***,** 139-183. Amsterdam, The Netherlands: Elsevier.

Hendy, K.C. and Farrell, P. S. E. (1997). *Implementing a Model of Human Information Processing in a Task Network Simulation Environment.* Defence and Civil Institute of Environmental Medicine, DCIEM No 97-R-71

Hendy, K.C., Liao, J. and Milgram, P. (1997), *Combining Time and Intensity Effects in Assessing Operator Information Processing Load*, Human Factors, 39(1), 30-47.

Mackenzie, S.I. (1990). *Movement Time Predictions in Human-Computer Interfaces.* Retrieved March 2007 from http://www.yorku.ca/mack/GI92.html.

Tryan, J.L., Armstrong, J., Ryder, C., and Belyavin, A. (2006). *IPME Workload Validation Experimental Plan.* Prepared for Defence Research and Development Canada (DRDC).

7 List of Abbreviations

ANOVA	Analysis of Variance
ATC	Air Traffic Control
CAE PS	CAE Professional Services
DRDC	Defence Research and Development Canada
HBM	Human Behaviour Modeling
HF	Human Factors
IPME	Integrated Performance Modeling Environment
IP/PCT	Information Processing / Perceptual Control Theory
ITP	Instantaneous Time Pressure
POP	Prediction of Operator Performance
POPIP	Prediction of Operator Performance and Information Processing
TLX	Task Load Index
TNM	Task network models
VACP	Visual, Auditory, Cognitive and Psychomotor
W/Index	Workload Index

Appendix A

Announcement for Recruitment

Investigating operator performance in a simulated multi-task environment

We are currently seeking volunteers to participate in a study to investigate operator performance in a simulated Air Traffic Control task. The results of this study shall be used to increase our understanding of human workload and information processing to better predict human workload and performance within a variety of complex systems. You will be asked to complete a variety of tasks related to Air Traffic Control and visual vigilance. Upon completing the trial, you will be paid $125.00 for your time. There are no known risks to participating in this study.

To participate, you must possess normal or corrected-to-normal vision and have no prior experience in Air Traffic Control. You must also be fluent in reading and writing in English and have at least 2 years of basic computer experience. The data collected from your participation in this research study will be maintained in the strictest confidence according to the guidelines established by Carleton University's Ethics Committee. If you are interested in participating or have further questions please contact us at info@greenley.ca.

Warmest Regards,

Research Project Coordinator
CAE Professional Services

Appendix B

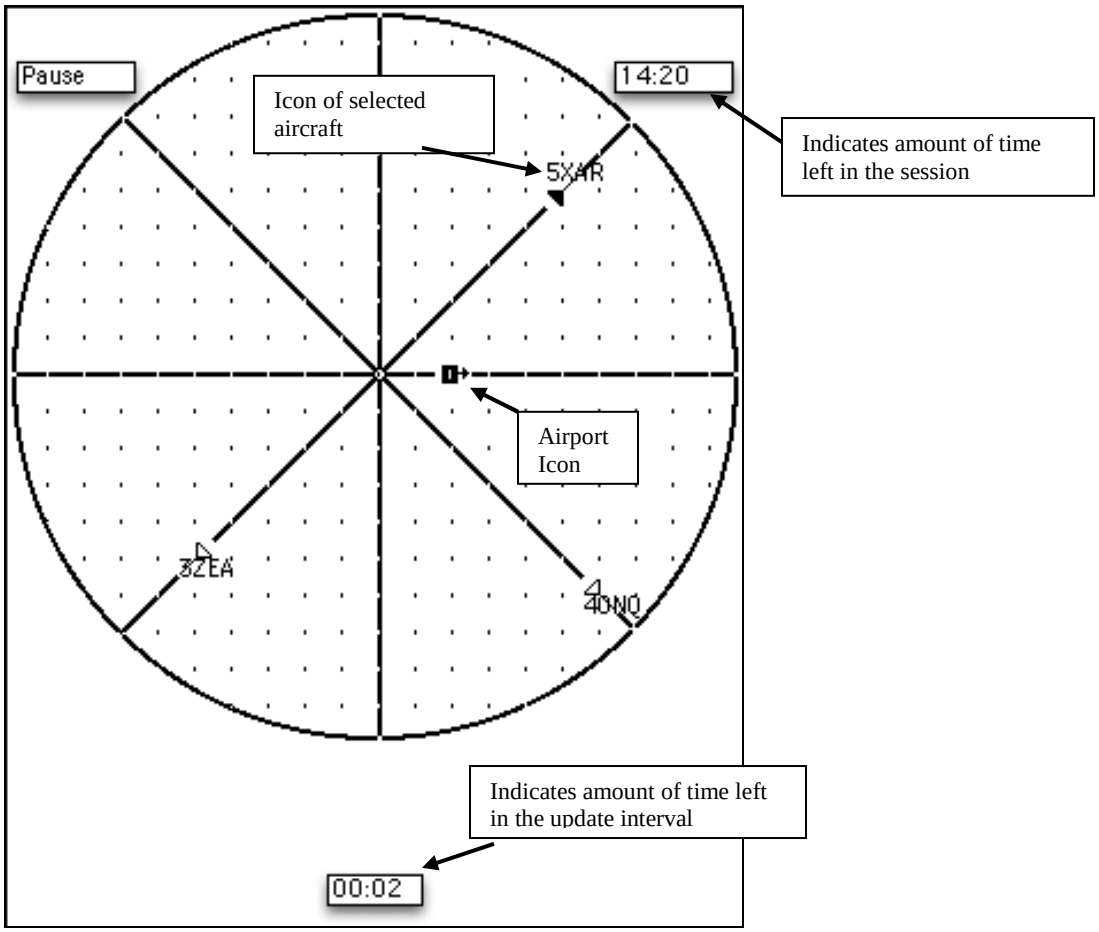


Figure 1A. The radar window. This window represents the air space in which the subject attempts to control the aircraft.

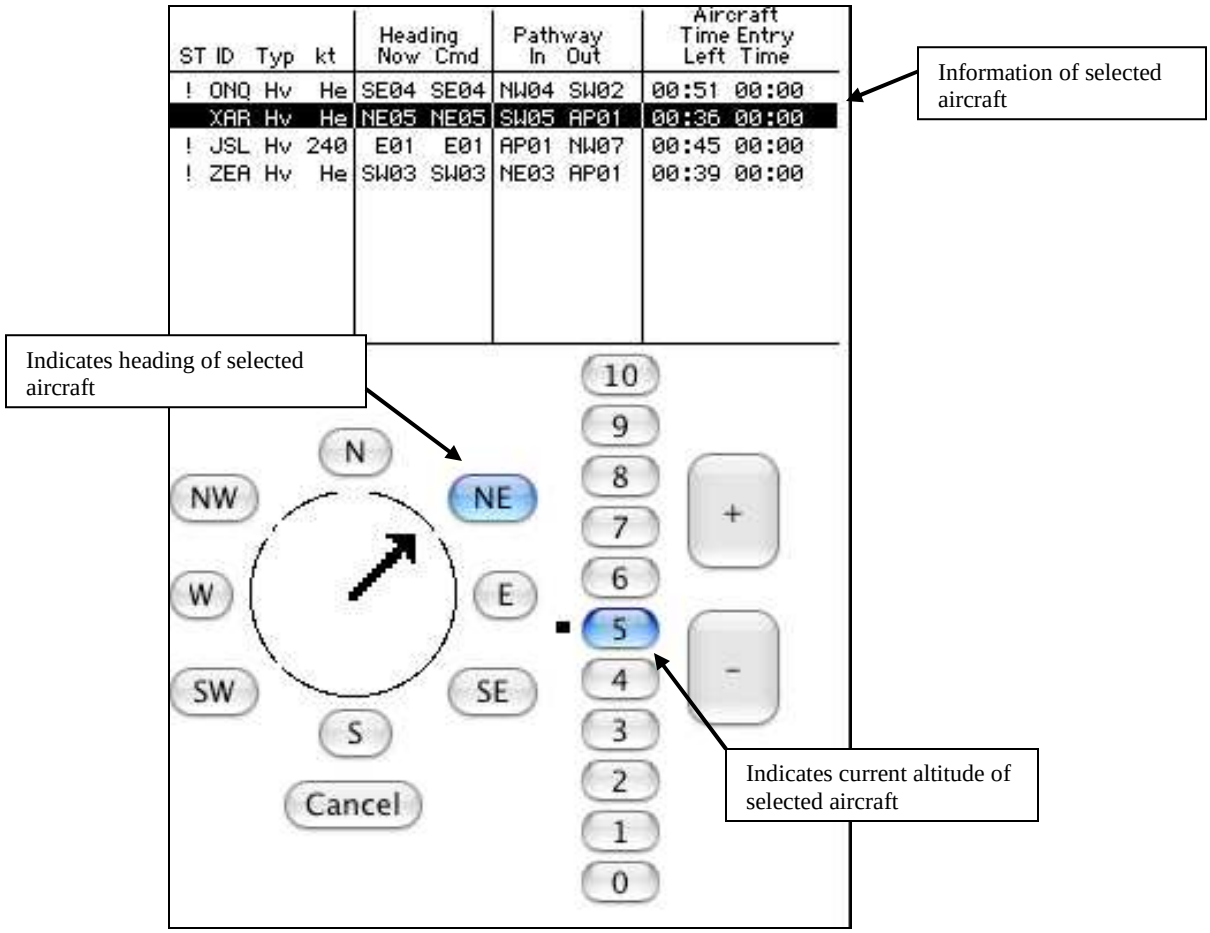


Figure 2A. The air traffic schedule window. This window provides information particular to individual aircraft. Participants also control the aircrafts within the window.

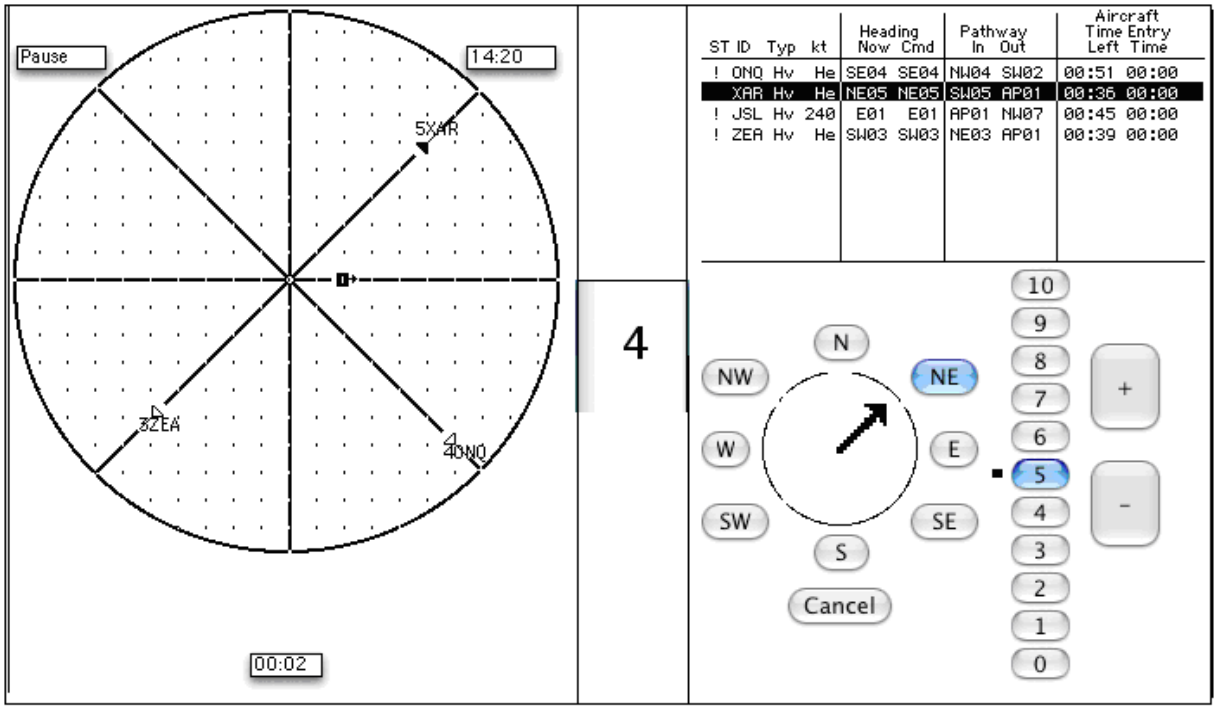


Figure 3A. Dual-task window configuration with the radar screen, Bakan digit presentation and air traffic schedule window.

Appendix C

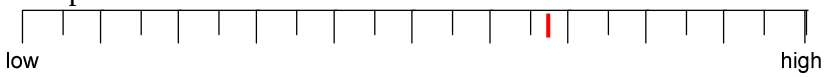
NASA TLX Questionnaire

An electronic version of the NASA TLX questionnaire was used for the present study.

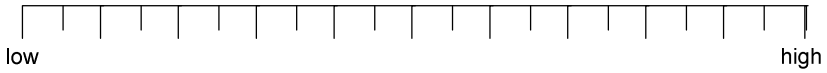
Task Questionnaire Part 1:

Place a mark on each scale at the point that best indicates your experience of the task

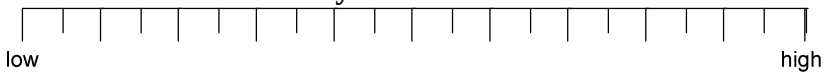
Example:



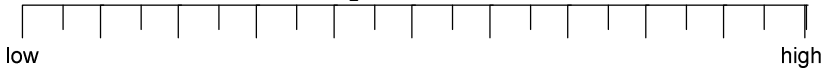
Mental Demand



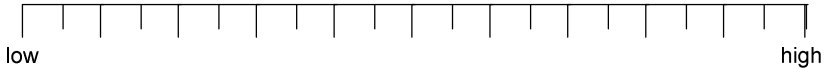
Physical Demand



Temporal Demand

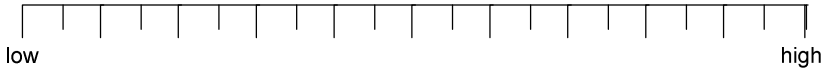


Performance



Effort

Frustration



Appendix D

Subject information package

Informed Consent Form

The information below is intended to help you understand exactly what we are asking of you. Please read this consent form carefully and ask all the questions you might have before deciding whether to participate or not in this study. Please take whatever time you need before reaching a decision. Your participation in this study is anonymous and confidential; no one will know whether you participated, nor how you performed on the tasks.

Study: A study to "investigate operator performance in a simulated multi-task environment" for Defense Research and Development Canada (DRDC) under the direction of CAE Professional Services.

Purpose: Dynamic, high risk activities in complex environments (e.g., air traffic control [ATC]) place significant cognitive demands on human operators. Due to the critical nature of these environments, the accuracy and effectiveness of operator performance can ha ve significant impact on the level of safety while performing these high risk activities. The design of complex systems used to perform these activities is critical to mediate cognitive demands and maximize operator performance and safety. The objective of this study is to investigate how people perform in a simulated multi-task environment in order to better understand the human factors involved in the design and development of effective and efficient complex systems.

Research Personnel: The following personnel are involved in this research project and may be contacted at any time:

Investigators	Role	Contact Information
Joe Armstrong, CAE Professional Services	Principal Investigator	joe.armstrong@cae.com
Michelle Gauthier, CAE Professional Services	Research Project Coordinator	michelle.gauthier@cae.com
Wenbi Wang, Defence Research and Development Canada (DRDC)	Co-Researcher	wenbi.wang@drdc-rddc.gc.ca

If any ethical concerns or complaints about this study should arise please contact Research Ethics Committee Chair (Prof. Antonio Gualtieri, 613-520-2517, e-mail ethics@carleton.ca).

Task Requirements: As a participant in this study, you will be required to complete two 15 minute trials of a simulated Air Traffic Control (ATC) task. You will then be required to complete one 15 minute trial of a visual vigilance task (Bakan). Finally, you will be required to complete two 15 minutes trials of the ATC task while simultaneously

performing the Bakan task. Before the experiment begins you will be given a 3 hour training session on the ATC simulation and Bakan tasks. Following each trial, you will have to complete a questionnaire. The entire session will take approximately 5 to 6 hours to complete.

Throughout the trials, your performance will be timed and recorded. All of the information collected during the trials and recorded through the subjective questionnaires will be kept in complete confidence. Your performance will not reflect any other higher-level cognitive ability (e.g. intelligence etc.).

Your participation in this research is completely voluntary. Please feel free to ask questions of the researcher at any point during the session. You may choose to withdraw from the study at any time. Upon completing the trial, you will be paid $125.00 for your time, regardless if you complete the session or not.

Requirements to participate in this study: In order to be able to participate in this study, you must possess normal or corrected-to-normal vision and have no prior experience in Air Traffic Control. You must also be fluent in reading and writing in English and have at least 2 years of computer experience.

Duration: The session that will last approximately 5 to 6 hours in duration.

Locale: CAE Professional Services, 1135 Innovation Dr. Suite 200**.**

Potential Risk or Discomfort: There are no known risks or harms associated with this study. Subjects may feel stressed or fatigued due to the high task demands.

Confidentiality: All the information collected in this experiment will be kept confidential and will be identified by numbered coding only. It is important to emphasize that the data collected herein **DO NOT** reflect personal skill or sensitive information of any sort. The data cannot be used to derive sensitive personal information. If the results of the study are published, your name will not be used, and information disclosing your identity will **NOT BE** released or published under any circumstance. All data will be maintained in a secure location at the offices of Greenley and Associates, Inc. Only the lead investigators as listed above shall have direct access to personal information.

Right to Withdraw: I understand that I am free to refuse to participate and may withdraw my consent at any time. Should I withdraw my consent, my participation as a subject will cease immediately. You will be paid $$ for your time regardless of session completion.

I have had the opportunity to ask questions of the investigator(s). Details of the study have been explained to me by the CAE Professional Services team, and my questions about the study have been answered to my satisfaction. I have had sufficient time to consider whether to participate in this study. I understand that my participation in this study is entirely voluntary and that I may withdraw from the

study at any time without penalty. I voluntarily consent to participate in the study "Investigating operator performance in a simulated multi-task environment".
I may obtain additional information about the project and have any additional questions answered by contacting CAE Professional Services.

_____ _____

Participant Name (Print) Participant Signature

_____

Date

To the best of my knowledge, the information in this consent form, and the information that I have provided in the response to any questions, fairly represents the project. I am committed to conducting this study in compliance with all the ethical standards that apply to projects that involve human subjects. I will ensure that the subject receives a copy of this consent form.

_____ _____

Researcher Name (Print) Researcher Signature

_____

Date

Instructions For ATC Task Simulation

The task in this experiment is an Air Traffic Control (ATC) simulation. You are an air traffic controller and you are in charge of an airspace. It is your responsibility to safely route all aircraft to their destination, either to land at the airport you are in charge of or to a pre-determined exit. That means you are in control of all aircraft arrivals, departures and over-flights within the region of a major city airport. Your job is to correctly route all aircraft to their proper exit pathway – that is, they must exist the radar screen or land at the airport in the correct heading and at the specified altitude within a certain amount of time.

How the ATC works and a Description of Screen Layout

The start of the ATC Task. Aircraft enter the screen via the screen boundary or from an airport. The aircraft will be flying at a certain heading, altitude and speed. Your job is to route the aircraft along a specific pathway so that it can exit the screen boundary or land at an airport in the correct heading and altitude.

Display windows. There are two main displays in the simulation.

Radar Screen. The first display is a simulated radar screen (*show participant Figure 1*) which shows the positions of all aircraft. Aircraft appear to "move" (that is, changes in heading and/or altitude) every time the radar screen is updated. Imagine a

radar doing a 360 degree sweeping motion of the sky (*demonstrate sweeping motion).* When the radar finishes it 360 degree rotation, the radar screen is updated. A counter at the bottom of the screen indicates how much time is left before the radar refreshes. That is, changes in the radar screen occur every time the counter at the bottom of the screen begins its new countdown interval. So, for example, the counter will countdown from :09 and reach :00, the radar will update when the counter begins the interval anew at : 09.

Aircraft Schedule Window. The second display (*show participant Figure 1*) consists of two features. The top of the window shows information about the aircraft such as the direction and altitude of the aircraft when it enters the screen, its current heading, altitude, the direction and altitude the aircraft needs to exit the radar, and the amount of time you have to do route the aircraft (*point to scheduled information in Figure 2).* As you can see, there are 8 columns in the Air Traffic Schedule window. A description of what the columns mean will be explained as we go through the procedure of the task.

The second feature of this window is the control panel to control aircraft. At the bottom of the screen, there are controls for changing the heading and altitude of the aircraft.

Guiding the aircraft

Taking control of the aircraft: An aircraft can be controlled by clicking either on the aircraft symbol on the radar screen, or on the row in the aircraft traffic schedule window which contains the information about the aircraft.

Controlling heading and altitude of aircraft: To change the aircraft's heading and altitude, you first select the aircraft then click on the appropriate heading and/or altitude keys in the control window. *Demonstrate heading and altitude and heading indicators on the schedule window.* The 5th and 6th columns are the current and commanded heading and altitude, respectively. For example, "N02" in column 5 in Figure 2 (Air Traffic Schedule) shows the highlighted aircraft is currently heading north at altitude 2 (2000 feet), while "N01" in column 6 shows the aircraft has been given a command to head north at altitude 1.

How the aircraft "moves". Aircraft can only move 1 'dot' at a time across the radar screen with every refresh rate.

Change in heading. Once you have indicated the aircraft to change its heading the aircraft will change one step (45 degrees of heading) of direction at each screen update interval. The aircraft will move 1 'dot' in the direction of its previous heading before taking up the new heading (*demonstrate using Figure 1).* For example, if an aircraft is currently heading north-east at altitude 8, and you give a command of east, it will take the aircraft 2 steps to change direction. .

Change in altitude. Once you have indicated the aircraft to change its altitude, the aircraft will change one step of altitude (1000 feet) at each screen update interval. For example, if an aircraft is currently traveling at 2000 feet and you command it to change altitude to 6000 feet it will take the aircraft 3 steps to change altitude.

Meanwhile, the direction and altitude can change simultaneously. The pilots will fly to the headings and altitudes you give them. It is your responsibility to guide them onto their paths.

Test question: So, let's take another example and you tell me how you think the aircraft will move. If the aircraft's current heading is east, at altitude 3, and you command the plane to go north at altitude 7, can you indicate where on the radar the airplane will end up and how many steps it will take for the aircraft to get to its commanded altitude? *Let participant try it themselves.*

Controlling altitude when landing at an airport: When aircraft are landing at an airport, the altitude of the aircraft must be reduced to 0 by the last update before the aircraft lands. Otherwise, the aircraft will not land and you will have to reroute the aircraft to land.

Exiting aircraft: All destinations, other than airports, are reached along one of the 8 lines that mark the cardinal compass directions of N, NE, D, SE, S, SW, W and NW. Aircraft terminating at an airport must land in the direction of the runway heading (shown by the runway symbol). So for example, if the exit pathway is N08, the aircraft must exit on the north-south cardinal line facing north, and at the appropriate altitude.

Air-time running out: If an aircraft cannot reach its destination within this time, it misses its time 'slot' for the rest of the route (necessary for coordination with other controllers or other aircraft departures). If you do not route the aircraft to its specified exit from the airspace before the amount of air-time runs out, the aircraft will proceed to blink. Blinking will beginning of the update interval immediately prior to the expiration of its' air-time.

The 9th column in Figure 2 shows the amount of time left before the aircraft must reach its destination. The last column shows the amount of time left before the aircraft comes under your control (00:00 means the aircraft has entered the radar screen or is waiting at the airport, and therefore can now be placed under your control).

The columns in the aircraft schedule window

We've already gone over the 4^{th}, 5 and 6^{th}, 9^{th} and 10^{th} columns, we will not look at what the other columns mean and how you will be using this information to do this task.

The first column contains a symbol which reflects the state of an aircraft. The following is a list of all possible states.

State	Meaning
	Aircraft under control but not yet on course (blank or space)
!	Aircraft has not yet been selected,
=	aircraft is on desired course and at the correct altitude
@	pilots had to take evasive maneuvers to avoid a collision
?	An aircraft is about to leave from an airport

The second column is the aircraft identification. The identification will always be a unique 3-letter code (e.g. EYZ).

The 3rd column is the aircraft type. Only one type of aircraft will be used for this study.

The 4th column is the current airspeed in knots. The aircraft travels at speeds of 240 and 120 knots respectively.

The 7th and 8th columns show where aircraft enter the screen and what their destination is to be, respectively. For example, "W08" in column 7 in Figure 1 (Air Traffic Schedule) shows that the highlighted aircraft entered from the west at altitude 8. "NW01" in column 8 shows that its destination is to leave the screen to the northwest at 1000ft.

A final note

The objective of the simulation is to correctly route aircraft to their *desired exit points at the required altitude and heading*, and within the allowable time interval. In achieving this goal, you should avoid various serious errors, such as near misses, collisions or running out of time to exit the aircraft. Remember, a landing aircraft should be aligned with the runway heading, otherwise a misdirection will be recorded. Those aircraft leaving the screen at a heading other than the one required will also be scored as a misdirection. In addition, aircraft that use up their allotted time while in the air will be scored as out of time. It is desirable that aircraft be handled promptly (shortest path) within the requirements for safe operation. As with all aviation operations, **safety is of prime importance (i.e. avoidance of near misses and collisions.). Remember, you are in charge of people's lives.

Across trials of the experiment, the number of aircraft under your control will vary. At times there will be very few on the radar screen, and at other times there will be a considerable number. Please do your best.

**This is a difficult task. You will be loaded to a point where perfect performance may not be possible. No matter how bad things look, do what you can. Please don't give up no matter what goes wrong!

Instructions for Bakan Visual Task

Purpose: You will be required to monitor a stream of digits in the range 0 to 9. Your job is to identify an Odd-Even-Odd sequence of numbers and provide a response to indicate that you have detected this sequence.

Method: When a set of Odd-Even-Odd digit numbers is presented in sequence, you will depress the space bar on the keyboard to indicate your response (See Figure below).

1 3 6 3 1 1 1 4 3 4 5 4
[HIT]

Instructions for NASA/TLX Questionnaire

Ratings
We are interested not only in assessing your performance but also the experiences you had during the different task conditions. Right now we are going to describe the technique that will be used to examine your experiences. In the most general sense we are examining the "workload" you experienced. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The workload contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Since workload is something is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

After performing the task, six rating scales will be displayed. You will evaluate the task by marking each scale at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "own performance" goes from "good" on the left to "bad" on the right. This order has been confusing for some people. Please consider your responses carefully in distinguishing among the task conditions. Consider each scale individually. Your ratings will play an important role in the

evaluation being conducted, thus your active participation is essential to the success of this experiment, and is greatly appreciated

Sources of Workload Evaluation

Throughout this experiment the rating scales are used to assess your experiences in the different task conditions. Scales of this sort are extremely useful, but their utility suffers from the tendency people have to interpret them in individual ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended or the performance they achieved. Others feel that if they performed well, the workload must have been low, and vice versa. Yet others feel that effort or feelings of frustration are the most important factors in workload and so on. The results of previous studies have already found every conceivable pattern of values. In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

The evaluation you are about to perform is a technique developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort vs. Mental Demands) and asked to choose which of the items was more important to your experience of workload in the task(s) that you just performed. Each pair of scale titles will appear separately on the screen. Select the Scale Title that represents the more important contributor to workload for the Specific task(s) you performed in this experiment.

Press the left button to select the top item in the pair and the right button to select the bottom item. A pointer shows which title was selected. To enter that choice press the button again and a new pair of titles will appear. If you change your mind, press the other button to cancel your first choice, and then start over.

After you have finished the entire series we will be able to use the pattern of your choices to create a weighted combination of the ratings from that task into a summary workload score. Please consider your choices carefully and make them consistent with how you used the rating scales during the particular task you were asked to evaluate. Don't think that there is any correct pattern; we are only interested in your opinions. If you have any questions, please ask them now.

Appendix E

Debriefing Form

Research of Integrated performance modeling

The ability to predict human performance in dynamic high risk activities (e.g., flying, air traffic control, driving) is extremely important considering that maintaining awareness within such complex situation can be difficult and can affect the accuracy and effectiveness of operator performance. Of key interest to the Canadian (CA) defence and Human Factors (HF) community is the ability to develop computational models of human behaviour that operate within complex systems to compare systems performance, evaluate design alternatives for immersive and real system simulations, and predict human performance and workload prior to virtual and field-based trials of real systems. Additional research is being conducted on the efficacy of replacing human operators with human behaviour models in virtual simulations. The application of Human Behaviour Representations (HBRs) within these environments allows designers to predict system performance during development without expending the associated costs of developing complex human-in-the-loop simulations for predictive analysis. It is therefore important to investigate how people perform within a complex simulated environment in order to develop effective and efficient systems.

Five algorithms based on concepts of human workload and information processing, have been developed to simulate and predict human workload and performance. Consequently, there is a requirement to ensure that the workload algorithms are accurately modeled and are producing reliable and valid datasets. The experiment was developed to investigate how well a formal model of human workload and performance in a simulated Air Traffic Control (ATC) task can predict the general performance and workload characteristics of the human operator. Human subjective and performance data of the primary and secondary tasks are then be compared with the performance of the computational model of simulated human data.

Thank you for your participation in this study. If you have any questions or comments about the study, do not hesitate to ask the facilitators before you leave, or you can contact us at the number provided below.

Phone: 1 (613) 247 – 0342

Appendix F

TIMINGS USED IN ATC TASK

Task number	Task name	Mean Time (seconds)	Std. Dev	Justification / Description
4.3	detect aircraft	vb_traits.vb_detection_time (approx. 100ms)	As per operator sampling function	Based on operator sampling in IPME. Expected value based on triangular distribution. Values drawn from Card et al. (1983). Variations based on exponential Gaussian distribution.
4.2	scan/saccade	0.1 + 0.25	(0.1 + 0.25)/6	Micro model eye movement time + eye fixation time. From IPME task manual page 9-64 (drawn from Card et al., 1983)
4.4	Gaussian aircraft status	vb_decision_time_intercept + vb_decision_time_slope*(numAC) + vb_cognitive_processor_time (approx 200 – 500ms based on # AC on radar)	As per operator sampling function	Based on operator sampling in IPME. Expected value based on triangular distribution. Values drawn from Card et al. (1983). Variations based on exponential Gaussian distribution.
4.5	select aircraft	0.230 + 0.166 x 7	(0.230 + 0.166 x 7) / 6	Mouse movement time prediction based on Fitts Law. From Mackenzie (1990), *Movement Time Prediction in Human-Computer Interfaces.* http://www.yorku.ca/mack/GI92.html . Drawn from equation (12) 'Point-Select'. Multiplier (e.g., 7) based index on difficulty generated as a function of movement distance and target width. Assumed embedded saccade times + mouse movement + click.

4.6	read heading	(0.1 + 0.25) x 4	0	Based on task flow analysis: saccade to info window from radar + discriminate correct line item + saccade to current heading column (current value) + saccade to required heading column (desired value). Each saccade step: Micro model eye movement time + eye fixation time). From IPME task manual page 9-64 (drawn from Card et al., 1983).
4.9	compare heading	0.07 + 0.35 + 0.07 + 0.07 + 0.35	0	Based on task flow analysis: cognitive process (CP) to decide to check radar after reading + saccade to radar + CP to assess required track + CP to assess current heading + CP to decide action + saccade to adjustment button. From IPME task manual page 9-54, 9-64 (drawn from Card et al., 1983). This task is up for discussion given greater complexity and dynamics (varying number of decisions and saccades) depending on current position, heading and required exit location. Further complexity in decisions would be incurred with airport landings. We would expect this task to take longer than the comparison of altitude task (4.13).
4.10	adjust heading	0.230 + 0.166 x 6	(0.230 + 0.166 x 6) /6	Same as task 4.5 ‘Select Aircraft’ with less difficulty since larger target width.
4.11	read altitude	(0.1 + 0.25) x 4	0	Same as task 4.6 ‘Read Heading’.
4.13	compare altitude	0.07 + 0.35 + 0.07 + 0.35	0	Based on task flow analysis: cognitive process (CP) to decide to check altitude indicator + saccade to altitude indicator + CP to decide action + saccade to adjustment button.
4.12	adjust altitude	0.230 + 0.166 x 6	(0.230 + 0.166 x 6) /6	Same as task 4.5 ‘Select Aircraft’ with less difficulty since larger target width.

Appendix G

VACP, W/INDEX, IP/PCT and POP Ratings for the ATC and Bakan Task

ATC TASKS VACP, W/INDEX, and IP/PCT

Task number	Task name	VACP	W/INDEX	IP/PCT
4.1	ATC Begin	No operator assigned	No operator assigned	No operator assigned
4.8	AC Generator	No operator assigned	No operator assigned	No operator assigned
4.7	AC Update Attributes	No operator assigned	No operator assigned	No operator assigned
4.34	StartTask	No operator assigned	No operator assigned	No operator assigned
4.32	Check Array	No operator assigned	No operator assigned	No operator assigned
4.3	Detect Aircraft	V= Register/Detect (1.0) C= Automatic (1.0)	Visual Perception= 1 Spatial Cognition= 1	V= Enabled; Externally Cued; Category: peripheral; Visual Area: Radar C= (1) automatized (skill-based)
4.2	Scan/Saccade	V= Locate/Align (5.0) C= Alternative Selection (1.2)	Visual Perception= 4 Spatial Cognition= 2	V= Enabled; Category: central; Visual Area: Radar C= (1) automatized (skill-based)
4.4	Recognise Aircraft Status	V= Inspect/Check (4.0) C= Evaluation/ Judgement Single (4.6)	Visual Perception= 3 Spatial Cognition= 4	V= Enabled; Category: central; Visual Area: Radar C= (5) reasoning
4.5	Select Aircraft	V= Discriminate (3.7) C= Alternative Selection (1.2) Psychomotor = 2.2	Visual Perception= 2 Spatial Cognition= 2 Manual Response= 2	V= Enabled; Category: central; Visual Area: Radar C= (1) automatized (skill-based) P= Enabled Preferred; right_hand_whole; right_hand_digit2

4.6	Read Heading	V= Read (5.9) C= Encoding/Decoding (5.3)	Visual Perception= 6 Verbal Cognition= 5	V= Enabled; Category: central; Visual Area: TextWindow C= (3) verbal (speech production)
4.9	Compare Heading	V= Locate/Align (5.0) C= Evaluation/ Judgement Several (6.8)	Visual Perception= 4 Spatial Cognition= 6	V= Enabled; Category: central; Visual Area: TextWindow C= (4) spatial (pattern recognition)
4.10	Adjust Heading	V= Locate/Align (5.0) C= Alternative Selection (1.2) P= Discrete Actuation (2.2)	Visual Perception= 4 Spatial Cognition= 2 Manual Response= 2	V= Enabled; Category: central; Visual Area: HSI C= (1) automatized (skill-based) P= Enabled Preferred; right_hand_whole; right_hand_digit2
4.11	Read Altitude	V= Read (5.9) C= Encoding/Decoding (5.3)	Visual Perception= 6 Verbal Cognition= 5	V= Enabled; Category: central; Visual Area: HSI C= (3) verbal (speech production)
4.13	Compare Altitude	V= Inspect/Check (4.0) C= Evaluation/ Judgement Single (4.6)	Visual Perception= 3 Spatial Cognition= 4	V= Enabled; Category: central; Visual Area: HSI C= (5) reasoning
4.12	Adjust Altitude	V= Locate/Align (5.0) C= Alternative Selection (1.2) P= Discrete Actuation (2.2)	Visual Perception= 4 Spatial Cognition= 2 Manual Response= 2	V= Enabled; Category: central; Visual Area: HSI C= (1) automatized (skill-based) P= Enabled Preferred; right_hand_whole; right_hand_digit2
4.33	No Planes to Adjust	V= Inspect/Check (4.0) C= Sign/Signal Recognition (3.7)	Visual Perception= 3 Spatial Cognition= 3	V= Enabled; Category: central; Visual Area: Radar C= (2) passive monitoring

4.17	DoneForNow	No operator assigned	No operator assigned	No operator assigned

ATC TASK POP RATINGS

Task ID[1]	Task name	Input	Central	Output	Other
4.3	detect aircraft	90 (visual)	90 (verbal)	0	
4.2	scan/saccade	90 (visual)	80 (verbal/spatial)	0	
4.4	recognise aircraft status	70 (visual)	90 (verbal)	0	
4.5	select aircraft	50 (visual)	85 (spatial)	90 (manual)	Interference channel – right hand whole
4.6	read heading	90 (visual)	80 (verbal)	0	
4.9	compare heading	20 (visual)	90 (verbal)	60 (manual)	
4.10	adjust heading	50 (visual)	85 (spatial)	90 (manual)	Interference channel – right hand whole
4.11	read altitude	90 (visual)	80 (verbal)	0	
4.13	compare altitude	20 (visual)	90 (verbal)	60 (manual)	
4.12	adjust altitude	50 (visual)	85 (spatial)	90 (manual)	Interference channel – right hand whole
4.33	no planes to adjust	65 (visual)	80 (verbal)	0	

[1] The task ID is the task numbers in IPME model; hence, the IDs are not in numerical order.

BAKAN TASKS IPME/IP MODE

Task number	Task name	VACP	W/INDEX	IP/PCT
1.16	Visual_Bakan _Task_ Beginning	No operator assigned	No operator assigned	No operator assigned
1.17	VB_1_Stimulus	No operator assigned	No operator assigned	No operator assigned
1.18	VB_2_ISI	No operator assigned	No operator assigned	No operator assigned
1.13	vb_snap_ record	No operator assigned	No operator assigned	No operator assigned
1.25	detect & read stimulus	V= Read (5.9) C= Sign/Signal Recognition (3.7)	Visual Perception= 6 Spatial Cognition= 3	V= Enabled; Externally Cued; Category: central; Visual Area: VisualBakanStimulus C= (2) passive monitoring
1.26	classify memorize compare	C= Evaluation/ Judgement Several (6.8)	Verbal Cognition= 6	C= (4) spatial (pattern recognition) and (5) reasoning
1.24	key press response	C= Alternative Selection (1.2) P= Discrete Actuation (2.2)	Verbal Cognition= 2 Manual Response= 2	C= (4) spatial (pattern recognition) P= Enabled Preferred; left_hand_whole; left_hand_digit1
1.27	no_response	C= Alternative Selection (1.2) P= Discrete Actuation (2.2)	Verbal Cognition= 2 Manual Response= 2	C= (4) spatial (pattern recognition)
1.30	Memory refractory period	C= Automatic (1.0)	Spatial Cognition= 1	C= (1) automatized (skill-based)
1.23	String Memory Rehearsal	C= Encoding/Decoding (5.3)	Verbal Cognition= 5	C= (3) verbal (speech production) and (4) spatial (pattern recognition)
1.31	Trial End	No operator assigned	No operator assigned	No operator assigned

1.2, 1.4, 1.6, 1.33, 1.14, 1.7, 1.35	Not used due to complications related to shed task behaviour.

BAKAN TASKS POP MODE

Task ID	Task name	Input	Central	Output	Other
1.16	Visual_Bakan_Task_ Beginning	No operator assigned	No operator assigned	No operator assigned	No operator assigned
1.17	VB_1_Stimulus	No operator assigned	No operator assigned	No operator assigned	No operator assigned
1.18	VB_2_ISI	No operator assigned	No operator assigned	No operator assigned	No operator assigned
1.13	vb_snap_ record	No operator assigned	No operator assigned	No operator assigned	No operator assigned
1.25	detect & read stimulus	70 (visual)	80 (verbal)	0	
1.26	classify memorize compare	70 (visual)	100 (verbal)	60 (vocal)	
1.24	key press response	50 (visual)	70 (spatial)	90 (manual)	
1.27	no_response	40 (visual)	70 (spatial)	90 (manual)	
1.30	Memory refractory period	0	0	0	
1.23	String Memory Rehearsal	80 (visual)	100 (spatial/verbal)	50 (manual)	
1.31	Trial End	No operator assigned	No operator assigned	No operator assigned	

Appendix H

ATC STEPTHROUGH

The following sequence can be used to represent the logical flow of one cycle in the ATC task network model:

1. Task ".1" initiates model execution.
2. Task ".8" system generates first aircraft
3. Task ".34" initiates Check Array
4. Task ".32" logical check if monitoring or review of adjustments made is required (not required in first instance)
5. Task ".3" operator detects generated aircraft
6. Task ".2" operator saccades to aircraft
7. Task ".4" operator recognizes aircraft status and decides on what adjustments to make
8. Task ".5" operator selects aircraft with mouse pointer
9. Task ".11" operator reads required altitude
10. Task ".13" operator compares current vs. required altitude
11. Task ".12" operator selects control to adjust to desired altitude with mouse pointer
12. Task ".4" operator recognizes aircraft status and decides on what adjustments to make
13. Task ".5" operator selects aircraft with mouse pointer
14. Task ".6" operator reads required heading
15. Task ".9" operator compares current vs. required heading
16. Task ".10" operator selects control to adjust to desired heading with mouse pointer
17. Task ".4" operator recognizes aircraft status and decides on what adjustments to make
18. Task ".17" logical node done for now (adjustments made to aircraft and awaiting next refresh cycle)
19. Task ".32" logical check if monitoring or review of adjustments made is required (review is required at this point)
20. Task ".3" operator detects aircraft for review
21. Task ".2" operator saccades to aircraft for review
22. Task ".4" operator recognizes aircraft status and decides on what adjustments to make

23. Task ".17" logical node done for now (adjustments appropriate, no further adjustments required)
24. Task ".32" logical check if monitoring or review of adjustments made is required (monitoring is required at this point)
25. Task ".33" monitor aircraft (repeat loop between Task .32 & .33 until a refresh cycle is detected)

Appendix I

Partial Latin Square Design

Subject	Condition Order				
	Combo ATC high	Combo ATC low	Bakan alone	ATC high alone	ATC low alone
1	5	4	1	3	2
2	1	5	2	4	3
3	2	1	3	5	4
4	3	2	4	1	5
5	4	3	5	2	1
6	5	4	1	3	2
7	1	5	2	4	3
8	2	1	3	5	4
9	3	2	4	1	5
10	4	3	5	2	1
11	5	4	1	3	2
12	1	5	2	4	3
13	2	1	3	5	4
14	3	2	4	1	5
15	4	3	5	2	1

Appendix J

Randomization of Airports

Subject	Airport Order			
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C
5	A	B	C	D
6	B	C	D	A
7	C	D	A	B
8	D	A	B	C
9	A	B	C	D
10	B	C	D	A
11	C	D	A	B
12	D	A	B	C
13	A	B	C	D
14	B	C	D	A
15	C	D	A	B

AIRPORT	ORDER
N	A
S	B
E	C
W	D

Appendix K

SPSS output tables for a (5) (Task Condition) by (4) (Human, VACP, POP and POPIP) repeated measures ANOVA on the mean of Mental Workload

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
tasks_mental	Sphericity Assumed	11.393	4	2.848	661.314	.000	.892	2645.257	1.000
	Greenhouse-Geisser	11.393	2.082	5.472	661.314	.000	.892	1376.962	1.000
	Huynh-Feldt	11.393	2.219	5.134	661.314	.000	.892	1467.471	1.000
	Lower-bound	11.393	1.000	11.393	661.314	.000	.892	661.314	1.000
tasks_mental * Group	Sphericity Assumed	2.602	12	.217	50.338	.000	.654	604.058	1.000
	Greenhouse-Geisser	2.602	6.246	.417	50.338	.000	.654	314.436	1.000
	Huynh-Feldt	2.602	6.657	.391	50.338	.000	.654	335.105	1.000
	Lower-bound	2.602	3.000	.867	50.338	.000	.654	151.015	1.000
Error(tasks_mental)	Sphericity Assumed	1.378	320	.004					
	Greenhouse-Geisser	1.378	166.573	.008					
	Huynh-Feldt	1.378	177.522	.008					
	Lower-bound	1.378	80.000	.017					

a. Computed using alpha = .05

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	207.358	1	207.358	22515.906	.000	.996	22515.906	1.000
Group	1.752	3	.584	63.411	.000	.704	190.234	1.000
Error	.737	80	.009					

a. Computed using alpha = .05

Descriptive Statistics

	Group	Mean	Std. Deviation	N
WLMBakan	human	.407419	.2668610	12
	POP	.449196	.0226045	24
	POPIP	.450744	.0181594	24
	VACP	.432958	.0116130	24
	Total	.439031	.0996651	84
WLMATCLow	human	.717358	.2070251	12
	POP	.825696	.0012364	24
	POPIP	.823207	.0011806	24
	VACP	.471438	.0016021	24
	Total	.708291	.1723930	84
WLMATCHigh	human	.744512	.2561617	12
	POP	.837933	.0012784	24
	POPIP	.835711	.0014847	24
	VACP	.463849	.0025872	24
	Total	.717071	.1887942	84
WLMComboLow	human	.933508	.0684442	12
	POP	.920554	.0042261	24
	POPIP	.866872	.0026756	24
	VACP	.903169	.0119055	24
	Total	.902100	.0355997	84
WLMComboHigh	human	.933241	.0718216	12
	POP	.919945	.0041511	24
	POPIP	.864355	.0032374	24
	VACP	.895192	.0111382	24
	Total	.898889	.0370581	84

Appendix L

SPSS output tables for a (5) (Task Condition) by (4) (Human, VACP, POP and POPIP) repeated measures ANOVA on the mean of Physical Workload

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
tasks_physical	Sphericity Assumed	3.156	4	.789	183.847	.000	.697	735.386	1.000
	Greenhouse-Geisser	3.156	2.376	1.328	183.847	.000	.697	436.885	1.000
	Huynh-Feldt	3.156	2.546	1.240	183.847	.000	.697	468.035	1.000
	Lower-bound	3.156	1.000	3.156	183.847	.000	.697	183.847	1.000
tasks_physical * Group	Sphericity Assumed	.822	12	.069	15.969	.000	.375	191.627	1.000
	Greenhouse-Geisser	.822	7.129	.115	15.969	.000	.375	113.843	1.000
	Huynh-Feldt	.822	7.637	.108	15.969	.000	.375	121.961	1.000
	Lower-bound	.822	3.000	.274	15.969	.000	.375	47.907	1.000
Error(tasks_physical)	Sphericity Assumed	1.373	320	.004					
	Greenhouse-Geisser	1.373	190.108	.007					
	Huynh-Feldt	1.373	203.664	.007					
	Lower-bound	1.373	80.000	.017					

a. Computed using alpha = .05

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Intercept	30.163	1	30.163	1101.289	.000	.932	1101.289	1.000
Group	6.442	3	2.147	78.403	.000	.746	235.210	1.000
Error	2.191	80	.027					

a. Computed using alpha = .05

Descriptive Statistics

	Group	Mean	Std. Deviation	N
WLPBakan	human	.124836	.0964862	12
	POP	.222628	.0099889	24
	POPIP	.221905	.0069352	24
	VACP	.011182	.0019339	24
	Total	.148038	.0997794	84
WLPATCLow	human	.295573	.2454862	12
	POP	.241016	.0135822	24
	POPIP	.234042	.0121426	24
	VACP	.066238	.0041235	24
	Total	.196881	.1240539	84
WLPATCHigh	human	.312843	.2413088	12
	POP	.396359	.0144186	24
	POPIP	.390406	.0179883	24
	VACP	.114009	.0045018	24
	Total	.302056	.1514613	84
WLPComboLow	human	.412329	.3127423	12
	POP	.513866	.0129574	24
	POPIP	.419861	.0128917	24
	VACP	.077928	.0050129	24
	Total	.347948	.2102367	84
WLPComboHigh	human	.501773	.3073393	12
	POP	.501678	.0113497	24
	POPIP	.419993	.0183391	24
	VACP	.126911	.0052018	24
	Total	.371277	.1949024	84

Appendix M

SPSS output tables for a univariate analysis and multiple comparison of total error by Humans, VACP, POP, POPIP and IP.

Tests of Between-Subjects Effects

Dependent Variable: TOTAL_Error

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[a]
Corrected Model	1092.017[b]	4	273.004	23.966	.000	.487	95.863	1.000
Intercept	6426.040	1	6426.040	564.113	.000	.848	564.113	1.000
Group	1092.017	4	273.004	23.966	.000	.487	95.863	1.000
Error	1150.532	101	11.391					
Total	8084.651	106						
Corrected Total	2242.548	105						

a. Computed using alpha = .05

b. R Squared = .487 (Adjusted R Squared = .467)

Descriptive Statistics

Dependent Variable: TOTAL_Error

Group	Mean	Std. Deviation	N
human	14.0351	8.84539	12
POP	4.1972	1.17864	23
POPIP	9.3451	2.45858	23
VACP	4.3940	1.15685	24
IP	8.3993	2.03830	24
Total	7.4239	4.62143	106

Pairwise Comparisons

Dependent Variable: TOTAL_Error

(I) Group	(J) Group	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
human	POP	27.882*	2.749	.000	19.947	35.817
	POPIP	25.645*	3.775	.000	14.751	36.538
	VACP	16.520*	4.042	.001	4.855	28.184
	IP	18.894*	4.715	.001	5.286	32.502
POP	human	-27.882*	2.749	.000	-35.817	-19.947
	POPIP	-2.237	3.541	1.000	-12.455	7.981
	VACP	-11.362*	3.303	.009	-20.895	-1.830
	IP	-8.988	3.823	.212	-20.022	2.046
POPIP	human	-25.645*	3.775	.000	-36.538	-14.751
	POP	2.237	3.541	1.000	-7.981	12.455
	VACP	-9.125	4.744	.579	-22.817	4.566
	IP	-6.751	4.132	1.000	-18.675	5.174
VACP	human	-16.520*	4.042	.001	-28.184	-4.855
	POP	11.362*	3.303	.009	1.830	20.895
	POPIP	9.125	4.744	.579	-4.566	22.817
	IP	2.375	1.659	1.000	-2.412	7.161
IP	human	-18.894*	4.715	.001	-32.502	-5.286
	POP	8.988	3.823	.212	-2.046	20.022
	POPIP	6.751	4.132	1.000	-5.174	18.675
	VACP	-2.375	1.659	1.000	-7.161	2.412

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Appendix N

SPSS output tables for a (4) (Task Condition) by (5) (Human, VACP, POP, POPIP and IP) repeated measures ANOVA on the mean of ATC misdirection errors.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Task	Sphericity Assumed	.486	3	.162	15.906	.000	.136
	Greenhouse-Geisser	.486	2.958	.164	15.906	.000	.136
	Huynh-Feldt	.486	3.000	.162	15.906	.000	.136
	Lower-bound	.486	1.000	.486	15.906	.000	.136
Task * Group	Sphericity Assumed	1.004	12	.084	8.212	.000	.245
	Greenhouse-Geisser	1.004	11.834	.085	8.212	.000	.245
	Huynh-Feldt	1.004	12.000	.084	8.212	.000	.245
	Lower-bound	1.004	4.000	.251	8.212	.000	.245
Error(Task)	Sphericity Assumed	3.088	303	.010			
	Greenhouse-Geisser	3.088	298.807	.010			
	Huynh-Feldt	3.088	303.000	.010			
	Lower-bound	3.088	101.000	.031			

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	182.388	1	182.388	6369.035	.000	.984
Group	5.079	4	1.270	44.337	.000	.637
Error	2.892	101	.029			

Descriptive Statistics

	Group	Mean	Std. Deviation	N
TREND(MDirATCLow)	human	.280105	.2381846	12
	POP	.785266	.0987283	23
	POPIP	.762282	.1181222	23
	VACP	.776476	.1112419	24
	IP	.774116	.1014732	24
	Total	.718576	.2017708	106
TREND(MDirATCHigh)	human	.443405	.2257529	12
	POP	.690232	.0797417	23
	POPIP	.706015	.0864245	23
	VACP	.715064	.0723752	24
	IP	.688937	.1040213	24
	Total	.671043	.1361649	106
TREND(MDirComboLow)	human	.396500	.2479225	12
	POP	.694928	.0869376	23
	POPIP	.682372	.1286005	23
	VACP	.773804	.1073939	24
	IP	.740158	.1108012	24
	Total	.686519	.1695127	106
TREND(MDirComboHigh)	human	.497619	.2687362	12
	POP	.866918	.0611870	23
	POPIP	.868840	.0608574	23
	VACP	.685216	.1028204	24
	IP	.774385	.0734779	24
	Total	.763437	.1640280	106

Appendix O

SPSS output tables for a (3) (Task Condition) by (5) (Human, VACP, POP, POPIP and IP) repeated measures ANOVA on the mean of the Bakan omission errors.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Task	Sphericity Assumed	14.083	2	7.041	3021.777	.000	.968
	Greenhouse-Geisser	14.083	1.451	9.703	3021.777	.000	.968
	Huynh-Feldt	14.083	1.525	9.232	3021.777	.000	.968
	Lower-bound	14.083	1.000	14.083	3021.777	.000	.968
Task * Group	Sphericity Assumed	11.124	8	1.390	596.711	.000	.959
	Greenhouse-Geisser	11.124	5.806	1.916	596.711	.000	.959
	Huynh-Feldt	11.124	6.102	1.823	596.711	.000	.959
	Lower-bound	11.124	4.000	2.781	596.711	.000	.959
Error(Task)	Sphericity Assumed	.471	202	.002			
	Greenhouse-Geisser	.471	146.592	.003			
	Huynh-Feldt	.471	154.063	.003			
	Lower-bound	.471	101.000	.005			

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	38.355	1	38.355	4479.733	.000	.978
Group	22.223	4	5.556	648.890	.000	.963
Error	.865	101	.009			

Descriptive Statistics

	Group	Mean	Std. Deviation	N
TREND(MISSBakan)	human	.053332	.0355051	12
	POP	.053006	.0230745	23
	POPIP	.047342	.0260901	23
	VACP	.047274	.0262778	24
	IP	.060918	.0315062	24
	Total	.052308	.0279875	106
TREND(MISSComboLow)	human	.683664	.2307792	12
	POP	.051016	.0292623	23
	POPIP	.923204	.0275699	23
	VACP	.050887	.0288471	24
	IP	.762294	.0445018	24
	Total	.472900	.3935296	106
TREND(MISSCombo High)	human	.749416	.2123965	12
	POP	.055842	.0320451	23
	POPIP	.920228	.0221690	23
	VACP	.053618	.0290574	24
	IP	.890124	.0302431	24
	Total	.510306	.4179219	106

Appendix P

SPSS output tables for a (3) (Task Condition) by (5) (Human, VACP, POP, POPIP and IP) repeated measures ANOVA on the mean of Bakan commission errors.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Task	Sphericity Assumed	61.774	2	30.887	23.364	.000	.188
	Greenhouse-Geisser	61.774	1.946	31.744	23.364	.000	.188
	Huynh-Feldt	61.774	2.000	30.887	23.364	.000	.188
	Lower-bound	61.774	1.000	61.774	23.364	.000	.188
Task * Group	Sphericity Assumed	58.937	8	7.367	5.573	.000	.181
	Greenhouse-Geisser	58.937	7.784	7.572	5.573	.000	.181
	Huynh-Feldt	58.937	8.000	7.367	5.573	.000	.181
	Lower-bound	58.937	4.000	14.734	5.573	.000	.181
Error(Task)	Sphericity Assumed	267.044	202	1.322			
	Greenhouse-Geisser	267.044	196.545	1.359			
	Huynh-Feldt	267.044	202.000	1.322			
	Lower-bound	267.044	101.000	2.644			

Tests of Between-Subjects Effects

Measure: MEASURE_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	599.973	1	599.973	161.976	.000	.616
Group	314.778	4	78.695	21.245	.000	.457
Error	374.113	101	3.704			

Descriptive Statistics

	Group	Mean	Std. Deviation	N
TREND(FalseAlarm Bakan)	human	2.540367	2.9476443	12
	POP	.260870	.4489778	23
	POPIP	.434783	.7277666	23
	VACP	.666667	.7613870	24
	IP	.541667	.6580053	24
	Total	.712117	1.3175103	106
TREND(FalseAlarm ComboLow)	human	5.049815	4.6113915	12
	POP	.478261	.6653478	23
	POPIP	2.000000	1.4142136	23
	VACP	.458333	.6580053	24
	IP	2.041667	1.6010640	24
	Total	1.675451	2.3245389	106
TREND(FalseAlarm ComboHigh)	human	3.340923	2.6498894	12
	POP	.260870	.6191924	23
	POPIP	2.000000	1.4459976	23
	VACP	.166667	.4815434	24
	IP	1.125000	.7408867	24
	Total	1.161237	1.5893826	106

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document, Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's document, or tasking agency, are entered in section 8.)

Publishing: DRDC Toronto
Performing: Contractor CAE professional services
Monitoring:
Contracting:

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parenthesis at the end of the title)

The Evolution and Implementation of Workload Algorithms in IPME (U)
(U)

4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.)

Joe Armstrong, Michelle Gauthier, Gerald Lai, Andy Belyavin, Chris Ryder, Brad Cain

5. DATE OF PUBLICATION (Month and year of publication of document.)

September 2008

6a NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.)

107

6b. NO. OF REFS (Total cited in document.)

18

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of document, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development − include address.)

Sponsoring:
Tasking:

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.)

13iq

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

W7711−057962/001/TOR

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document)

DRDC Toronto CR 2006−042

10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on the dissemination of the document, other than those imposed by security classification.)

Unlimited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11), However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited announcement

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) This report documents a study to validate predictive workload models that are available within the Integrated Performance Modeling Environment (IPME). A literature review was conducted to assess the current state of knowledge of human workload and information processing, as well as to provide a review of the five IPME workload algorithms (VACP, W/Index, IP/PCT, POP, and POPIP). The results of the literature review indicated that, while the theories associated with human information processing are relatively mature, the predictive models of human workload integrated within IPME still require validation against human performance data. Analytical and empirical studies were then conducted within a combined Air Traffic Control (ATC) and Visual Bakan dual−task paradigm. The POP and POPIP analytical models more accurately predicted human subjective workload than did VACP and IP. The IP and POPIP analytical models predicted human performance in the Visual Bakan more accurately than did VACP and POP. All models were equally inaccurate in predicting ATC performance. Theoretical accounts of findings and practical implications for model development are discussed.

(U) Le présent rapport documente une étude visant à valider des modèles prédictifs de charge de travail qui sont disponibles dans un environnement intégré de modélisation des performances (IPME). Un examen de la documentation a été effectué pour permettre d'évaluer l'état actuel des connaissances de la charge de travail humaine et du traitement de l'information, ainsi que de revoir les cinq algorithmes de charge de travail IPME (VACP, W/Index, IP/PCT, POP, and POPIP). Les résultats de cet examen ont indiqué que si les théories associées au traitement de l'information humaine sont relativement à point, les modèles prédictifs de charge de travail intégrés au sein de l'IPME doivent toujours être validés par rapport au données de performance humaine. Des études analytiques et empiriques ont alors été menées en fonction du paradigme à double tâche combinée contrôle de la circulation aérienne (ATC) et tâche visuelle Bakan. Les modèles analytiques POP et POPIP ont prédit avec plus de précision la charge de travail subjective de l'être humain que ne l'ont fait les modèles VACP et IP. Les modèles analytiques IP et POPIP ont prédit la performance humaine pour une tâche visuelle Bakan plus précisément que ne l'ont fait les modèles VACP et POP. Tous les modèles ont été également imprécis dans la prédiction de la performance ATC. Les comptes rendus théoriques des conclusions et des conséquences pratiques pour l'amélioration des modèles sont présent

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) Workload algorithm, Integrated Performance Modelling Environment