Validation of Workload Algorithms Within the Integrated Performance Modeling Environment

Joe Armstrong
CAE Professional Services
Overview

- Background
- Project Goals
- Methodology
- Project Results
- Discussion & Conclusion
Workload Measurement

- Categories of Workload Measurement
  - Workload Assessment
    - Measurement of operator workload in complex systems
  - Workload Prediction
    - Computational models
    - Performance Prediction
Algorithmic/Predictive Measures

- **Applications**
  - Diagnostic workload predictions
  - Implemented within Computational Simulations
    - Mathematical/Algorithmic
    - Task Analysis
    - Computer Simulation

- **Major Categories**
  - Discrete Event Simulations (Task Network Modelling)
  - Cognitive Architectures
Task Network Modelling
Task Network Modeling

- General Assumptions
  - Human Behaviour Modelled as Interrelated Tasks
  - Performance Values Assigned by Modeller

- Sequences managed by a discrete event simulator
  - IPME™
  - IMPRINT™
  - SAINT/MicroSAINT
Workload Algorithms
State/Resource Algorithms

- Measure level of task demands
  - Attentional Demand Analysis Algorithm (VACP)
  - Workload Index (W/Index)

- Derivations
  - VACP
    - McCracken and Aldritch (1984); Szabo & Bierbaum (1986); Bierbaum et al. (1989)
  - W/INDEX
    - Honeywell Systems and Research Center (1983)
    - North & Riley (1986)
General Model

Process Request, Develop Plan

Data Processing, Exploitation

Fusion

Dissemination

VIS – Visual
AUD – Auditory
COG – Cognitive
MOT – Psychomotor
Within
Interference
Workload
Num Tasks

Attentional Demand
W/Index

General Model

Process Request, Develop Plan

Data Processing, Exploitation

Fusion

Dissemination

VIS – Visual
AUD – Auditory
COG – Cognitive
MOT – Psychomotor
Within
Interference
Workload
Num Tasks

Attentional Demand
W/Index

General Model

Process Request, Develop Plan

Data Processing, Exploitation

Fusion

Dissemination

VIS – Visual
AUD – Auditory
COG – Cognitive
MOT – Psychomotor
Within
Interference
Workload
Num Tasks

Attentional Demand
W/Index

General Model

Process Request, Develop Plan

Data Processing, Exploitation

Fusion

Dissemination

VIS – Visual
AUD – Auditory
COG – Cognitive
MOT – Psychomotor
Within
Interference
Workload
Num Tasks

Attentional Demand
W/Index
State/Resource Algorithms

- Predictive Validity
  - Subjective Comparison
    - In-flight studies (Cain, 1997)

- Task Performance
  - Moderate correlation with Flight-Simulator data (Sarno & Wickens, 1995)
Scheduler/Performance Algorithms

- Qualitatively Different than VACP or W/INDEX
  - Measure impact of task demands and on task performance
  - Simulates scheduling of tasks based on demand

- Two Major Theories
  - IP/PCT (Hendy & Farrell, 1997)
  - POP (Farmer, 2000; Jordan & Farmer, 1995)

- New merged concept
  - POPIP (Fowles-Winkler et al., 2004)
The IP Model

- IP Model (Hendy & Farrell, 1997)
  - Impact of quantity of information and time to process
  - Limits on rate of processing
  - Concept of Time Pressure

\[
\text{Time Pressure} \propto \frac{\text{Amount of Information to be processed}}{\text{Time Available}}
\]

- Impact of Task Priority
- Short Term Memory (STM) model
The IP Model

- Validation
  - Few reported studies (Cain and Hendy, 1998)
  - Many values set at arbitrary levels
    - Some face validity
    - Lack of specific validation
The POP Model

- POP (Farmer, 2002)
  - Influence of Task Demands
    - Based on DRAWS Ratings
    - Divisible task demands
      - Input, Central, Output
  
- Task Interference
  - Structural
    - Same hand/digit
  - Overload
    - 2+ tasks demand same channel
The POP Model

- Task Deferring
  - POP determines multiple task workload within a channel

- Tasks deferred based on priority
  - All tasks can be deferred
  - All tasks are remembered, and always processed
The POPIP Model

- Hybrid POP and IP Model (Fowles-Winkler et al, 2004)
  - POP Task Interference + Deferment and Shedding of IP
    - Structural Interference (IP)
    - Non-Structural Interference (POP)
    - Shedding/Deferral (IP)
      - Applies STM model
Model Validity

- Studies
  - Validation of task ratings (Farmer et al., 1995b; Jordan et al., 1996)
  - Initial POP Validation
    - Bakan Validation (Belyavin & Farmer, 2006)
  - No Current POPIP Validation
Project Goals

- Validate Workload Algorithms
  - VACP & W/Index
  - IP/PCT
  - POP
  - POPIP
Project Goals

- Levels of Validation
  - Construct/Content Validity
    - Is the model built on appropriate principles
  - Predictive Validity
    - Do the models predict performance
Project Goals

- Task Network Validity
  - Rating Allocation
- Model Algorithms
- State Predictions
- Overall Performance
Task Environment

- Simulated Air Traffic Control
  - Aircraft Routing
  - Time Pressure
  - Collision Avoidance

- Visual Bakan Task (Belyavin & Farmer, 2006)
  - Perception and Decision Process
  - Mental Rehearsal
Methodology

- **Pilot Study**
  - Single Threaded
  - Generate Baseline Datasets
    - Bakan and ATC
  - Independent participant

- **Lab Evaluation**
  - Multi-threaded
  - Workload Variations
    - Single Threaded
    - Multi-threaded
  - Independent Participants
Task Environment

Radar Plot

ATC Schedule Window

Bakan

3
## Methodology

<table>
<thead>
<tr>
<th>Pilot Study</th>
<th>ATC Settings</th>
<th>No. of aircraft</th>
<th>Length of update interval</th>
<th>No. of airports</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td></td>
<td>5</td>
<td>9 sec.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Bakan Settings (3 Digit)</strong></td>
<td>Display Duration</td>
<td>500 ms</td>
<td></td>
<td>ISI 2000 ms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experimental Evaluation</th>
<th>ATC Settings</th>
<th>No. of aircraft</th>
<th>Length of update interval</th>
<th>No. of airports</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td>5</td>
<td>9 sec.</td>
<td>1</td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td>5</td>
<td>6 sec.</td>
<td>1</td>
</tr>
<tr>
<td><strong>Bakan Settings (3 Digit)</strong></td>
<td>Display Duration</td>
<td>500 ms</td>
<td></td>
<td>ISI 2000 ms</td>
</tr>
</tbody>
</table>
Results Overview

- Workload Data
  - Primary Validation
    - Initial Results for Mental and Physical Predictions
      - VACP, POP, POPIP
    - No IP Time Pressure

- Performance Data
  - Secondary Validation
    - ATC and Bakan Error Data
Predicted Workload

<table>
<thead>
<tr>
<th></th>
<th>Physical</th>
<th>Mental</th>
</tr>
</thead>
<tbody>
<tr>
<td>VACP</td>
<td>Underestimate</td>
<td>Single vs. Multi-Threaded</td>
</tr>
<tr>
<td>POP</td>
<td>Reasonable</td>
<td>Moderate Overestimate</td>
</tr>
<tr>
<td>POPIP</td>
<td>Reasonable</td>
<td>Moderate Overestimate</td>
</tr>
<tr>
<td>IP</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

![Graph showing predicted workload](image)

16th BRIMS Conference 26-29 March 2007
ATC Performance

MisDirection Errors

Proportion of Errors

Task

ATCLow  ATCHigh  ComboLow  ComboHigh

BRIMS Conference 26-29 March 2007
Bakan Performance

Bakan Omission Errors

Bakan Comission Errors (False Alarms)

Predicted Performance

<table>
<thead>
<tr>
<th></th>
<th>ATC Misdirections</th>
<th>Bakan Omission Errors</th>
<th>Bakan Comission Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>VACP</td>
<td>Overestimate</td>
<td>Single vs. Multi-Threaded</td>
<td>Single vs. Multi-Threaded</td>
</tr>
<tr>
<td>POP</td>
<td>Overestimate</td>
<td>Reasonable</td>
<td>Underestimate</td>
</tr>
<tr>
<td>POPIP</td>
<td>Overestimate</td>
<td>Reasonable</td>
<td>Underestimate</td>
</tr>
<tr>
<td>IP</td>
<td>Overestimate</td>
<td>Reasonable</td>
<td>Underestimate</td>
</tr>
</tbody>
</table>

16th BRIMS Conference  26-29 March 2007
Discussion

- Model Predictions
  - Dissociation between Workload and Performance
  - Demonstration of Scheduling Effects
  - Sensitivity of Models to Workload Extremes
  - Lack of Variability in Workload Data
  - Algorithm Implementation Issues
Conclusion

- Future Research
  - Continue Validation Effort
    - Iterative Process
    - Link Models to ATC Simulator
    - Expand Human Data-Capture Capability
    - Improve variability in workload predictions
  - Verify Model Implementation Issues
Thanks to the Project Team

- DRDC Toronto
  - Dr. Wenbi Wang
  - Mr. Brad Cain

- QinetiQ
  - Dr. Andrew Belyavin
  - Mr. Chris Ryder

- CAE PS
  - Mr. Joe Armstrong
  - Mr. Gerald Lai
  - Ms. Michelle Gauthier
  - Ms. Neda Faregh