Model Evaluation

ACT-R, IMPRINT, and Matlab

- ACT-R, IMPRINT, and Matlab
  Three main modeling platforms with very different characteristics and capabilities

- First 'unified test problem'
  Keystroke entry task: simulation of a set of experiments by all three approaches
  Very similar results, but vastly different computational speeds

- Parameter optimization
  Real-time demonstration with Matlab's Genetic Algorithm toolbox

- Conclusions

Bengt Fornberg

University of Colorado, Boulder
Department of Applied Mathematics

Presentation at annual MURI meeting, held at Carnegie-Mellon University, Pittsburgh, September 7, 2007.

Modeling System:	Designed for:
ACT-R	cognitive modeling
IMPRINT	human and system performances in military tasks
Matlab	science and engineering applications

Very different paradigms / capabilities - however our task is to compare / evaluate in a very specialized set of applications (cognitive model development). Hence:

- Define 'unified test problems' that all three modeling systems will be applied to (using exactly the same specifications)
- Compare the accuracy of the generated simulated data vs. actual experimental data
- Compare the computational speeds in creating these simulated data sets
- Make some conclusions

We are NOT comparing the systems' overall capabilities in a broad sense (would be 'apples against oranges') - only their effectiveness on very precisely stated 'unified test problems'

- The issue not ACT-R vs. IMPRINT,
  instead is ACT-R and IMPRINT and additional possibilities (notably Matlab).

First 'unified test problem'

Keystroke data entry task: Healy, Kole, Buck-Gengler & Bourne (2004)

- Subjects type (on key pad, as quickly and accurately as possible) 4-digit numbers displayed to them; then press 'Enter' key.
- For all 32 subjects, each is shown 10 blocks of 64 numbers each.
- Different experimental versions included use of left and/or right hand, numbers repeated and non-repeated, etc.
- Timing and accuracy data recorded for each key stroke.

Model implementations / performance on this task have been reported in some detail:

ACT-R	Gonzales, Fu, Healy, Kole & Bourne (2006)
IMPRINT	Buck-Gengler, Raymond, Healy & Bourne (2007)
Matlab	Fornberg, Raymond & Best (2007)

- The ACT-R and the IMPRINT codes were developed separately;
- The Matlab code is a mathematically equivalent translation from the IMPRINT code.

ACT-R (The Adaptive Control of Thought - Rational)

- Combines strengths from artificial intelligence (AI) and neural networks (NN)
- Contains unified theory of cognition - over 30 years of cumulative improvement
- Modular architecture with interacting modules for memory, perceptual systems (vision, audition), motor systems, etc., synchronized through a central production system.
- Fatigue not naturally included in ACT-R, but has been incorporated in present model

IMPRINT 7

- Designed to simulate personnel and equipment in military tasks
- Evaluate planning efficiency, given constraints on time, accuracy, equipment functionality, human abilities, external environment.
- IMPRINT is NOT designed for cognitive modeling - however the present work shows it can be used for this purpose (although the implementation utilizes little of the software system's primary strengths).
- Present model implemented to include steps such as
  - Read and represent a number
  - Create a motor plan
  - Execute the motor plan
- Many human performance variables were assigned values from different stochastic distributions, described by parameters for which values were assigned based on general knowledge.

Matlab

- Evolved from FORTRAN in the late 1970's; now very widely used in science and engineering
- Language built around matrix syntax - runs at nearly peak hardware speed when this syntax can be fully utilized.

Accuracy comparisons

The single table below is typical of much more extensive results presented separately

(Experiment 2, Healy et al. (2004): Comparison involved 5 measures in 4 conditions across 10 blocks for a total of 200 data points)

RMSE (Root Mean Square Error) between two data sets

	Experiment	ACT-R	IMPRINT, run 2
ACT-R	0.1400	-	0.1695
IMPRINT, run 1	0.0588	0.1658	0.0455
IMPRINT, run 2	0.0628	0.1695	-
Matlab	0.0767	0.1827	0.0299

- Usually smallest variations when same model run repeatedly (or IMPRINT compared to its translation to Matlab)
- Especially in IMPRINT (and Matlab) models, possible to 'tune' parameters to fit better still with experiment

Speed comparisons (on comparable PCs)

ACT-R

Typical timing report for one simulation

```
; cpu time (non-gc) 605,045 msec (00:10:05.045) user, 392 msec system
; cpu time (gc)     251,940 msec (00:04:11.940) user,  77 msec system
; cpu time (total)  856,985 msec (00:14:16.985) user, 469 msec system
; real time  859,952 msec (00:14:19.952)
```

- **non-gc** — **Actual program execution**
- **gc (garbage collection)** — **Lisp system activity**
- **total** — **around 14 minutes**

IMPRINT

Wall-clock timing (only method available) - about **24 minutes**

Matlab

Extensive profiling options available. Sample output:

Lines where the most time was spent

Line Number	Code	Calls	Total Time	% Time	Time Plot
105	g_dist = (0.5*sum(randn(nr_c,8...	320	0.025 s	29.1%	
116	total_times = ...	320	0.017 s	20.3%	
113	CC_PH_Actual = (Counts(:,1)-Co...	320	0.010 s	11.7%	
73	Counts = cumsum([[ItemCt,Corre...	320	0.006 s	7.1%	
71	ab = 3-s1-s2;	320	0.003 s	3.5%	
All other lines			0.024 s	28.3%	
Totals			0.085 s	100%	

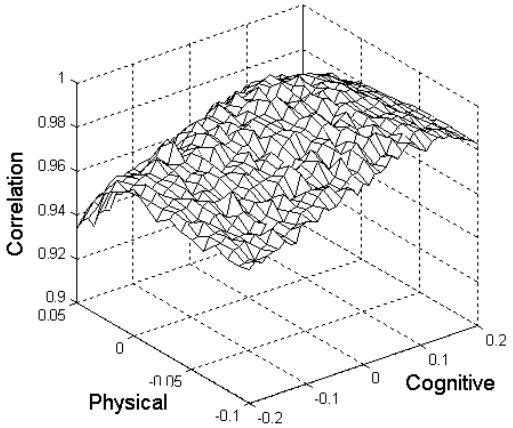
Total Matlab time: **0.085 seconds.**

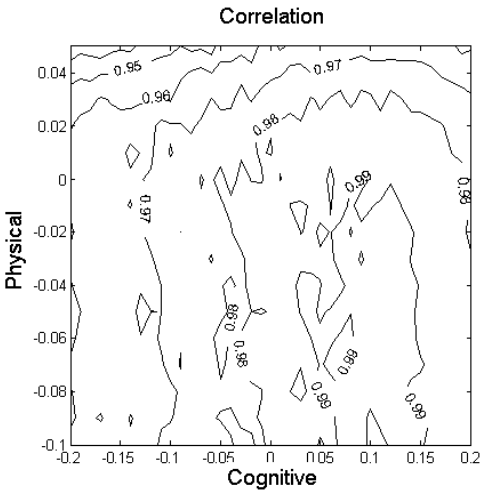
Opens major opportunities for parameter optimization.

Parameter optimizations

The Imprint / Matlab models have about a dozen parameters describing cognitive factors. We would like to vary all these in order to minimize RMSE against experiment.

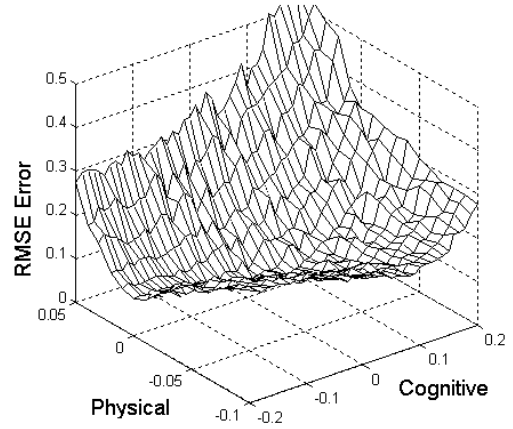
Example of 'correlation' and 'RMSE' in comparing Matlab model against experiment. ⇒

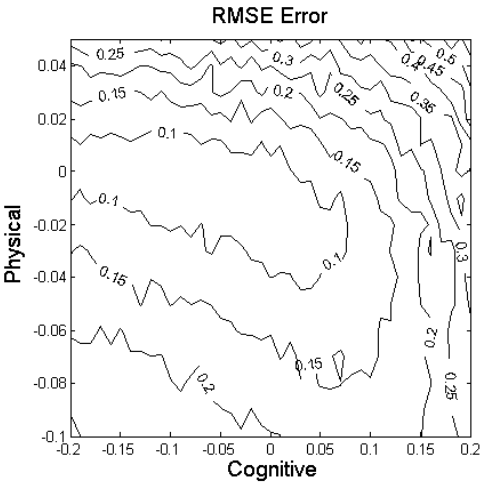




Very fast model evaluation makes possible:
Minimize RMSE by use of Genetic Algorithm

- Far more effective in many dimensions than exhaustive search
- Highly insensitive to stochastic noise





Real-time demonstration of Genetic Algorithm optimization will be shown after the conclusion slides

Conclusions

Described in this presentation:

- ACT-R and IMPRINT are very different systems, with very different built-in capabilities.
- When creating simple models in 'equation form', Matlab can offer vast speed advantages (by factors of 10,000 or more).
- This addition opens up major possibilities for global parameter optimizations (without needing to use massively parallel computer systems), likely leading to a better understanding of underlying cognitive processes.
- Additional 'unified test problems' will need to be tested before firm conclusions can be made

Plans for the future:

- Define additional 'unified test problems' of increasing complexity (next one will be the RADAR task).
- Explore further the use of genetic algorithm-based parameter optimization
- Develop more graphical tools for visualizing processes with several independent variables
- Compare models in terms of their assumptions about underlying cognitive processes
- Compare models in terms of the accuracy and range of their predictions with respect to the results of future experiments
- Compare models in terms of their ability to simulate the range of individual differences in performance