Simulation of Air Traffic Controllers' Behaviour Using the Operator Choice Model

[1]Wicks, J. , [1]S. Connelly, [1]P. Lindsay, [1,2]A. Neal, [1]J. Wang and [1,2]R. Chitoni

[1]Australian Research Council Centre for Complex Systems, The University of Queensland, [2]Key Centre for Human Factors and Applied Cognitive Psychology, The University of Queensland, Email: jacki@itee.uq.edu.au

Keywords: *Human computer interaction, traffic control, human reliability assessment, reaction time, logistic modelling.*

EXTENDED ABSTRACT

The Operator Choice Model (OCM) was developed to model the behaviour of operators attending to complex tasks involving interdependent concurrent activities, such as in Air Traffic Control (ATC). The purpose of the OCM is to provide a flexible framework for modelling and simulation that can be used for quantitative analyses in human reliability assessment, comparison between human computer interaction (HCI) designs, and analysis of operator workload.

The OCM virtual operator is essentially a cycle of four processes: Scan → Classify → Decide Action → Perform Action. Once a cycle is complete, the operator will return to the Scan process. It is also possible to truncate a cycle and return to Scan after each of the processes. These processes are described using Continuous Time Probabilistic Automata (CTPA). The details of the probability and timing models are specific to the domain of application, and need to be specified using domain experts.

We are building an application of the OCM for use in ATC. In order to develop a realistic model we are calibrating the probability and timing models that comprise each process using experimental data from a series of experiments conducted with student subjects. These experiments have identified the factors that influence perception and decision making in simplified conflict detection and resolution tasks.

This paper presents an application of the OCM approach to a simple ATC conflict detection experiment. The aim is to calibrate the OCM so that its behaviour resembles that of the experimental subjects when it is challenged with the same task. Its behaviour should also interpolate when challenged with scenarios similar to those used to calibrate it. The approach illustrated here uses logistic regression to model the classifications made by the subjects. This model is fitted to the calibration data, and provides an extrapolation to classifications in scenarios outside of the calibration data. A simple strategy is used to calibrate the timing component of the model, and the results for reaction times are compared between the OCM and the student subjects. While this approach to timing does not capture the full complexity of the reaction time distribution seen in the data from the student subjects, the mean and the tail of the distributions are similar.

1. INTRODUCTION

Current models of operator behaviour are inadequate for analysing highly interleaved tasks, such as in Air Traffic Control (ATC). As Kirwan (1990) notes in his survey of Human Reliability Assessment (HRA) approaches, few practical techniques have been developed for human-error classification during risk assessment. There are even fewer practical techniques for human-error quantification. One of the best-known techniques – and in some senses most general – is the Technique for Human Error Rate Prediction (THERP) developed by Kirwan (1990). This uses event trees with recovery paths to analyse failure rates. THERP is typically used in highly proceduralized situations in which the task is broken down into a sequence of individual steps, some of which involve checks on the outcomes of previous steps. However, it is difficult to apply techniques such as THERP to tasks such as ATC in which there is no defined sequence of events.

Timeliness of actions is a critical aspect of HRA; however most HRA techniques do not adequately model it. The Human Cognitive Reliability (HCR) approach of Hannaman et al. (1985) can be used to derive a model of the probability that a task has been completed, as a function of time. HCR models assume task completion has a probability density function of a particular form - a three-parameter Weibull distribution. Although undoubtedly useful in many situations, the technique is very limited in settings, such as ATC, in which multiple interleaved processes are involved.

A number of formal and semi-formal modelling techniques have been applied specifically to ATC; however, they do not provide stochastic models. The Eurocontrol organisation has developed a very sophisticated model of the cognitive processes involved in enroute control (Kallus et al., 1999); to the best of our knowledge it has not yet been used to analyse human error in any depth. Palanque et. al. (1997) applied Petri Nets to modelling the effect on an ATC task of the introduction of a new User-Interface technology (data link). Johnson (1997) illustrates the use of formal models to support the findings of accident investigations, and illustrates the approach on an aircraft accident. None of these approaches capture the stochastic dimension of ATC.

Thus there is a need to develop a means of modelling ATC which can capture the complexity of the task and allow for realistic quantitative analyses. The Operator Choice Model provides a means of simulating the behaviour of an operator who must prioritise and address problems as they arise over time. This provides a novel approach which we are applying to the ATC domain.

2. OPERATOR CHOICE MODEL

The OCM was designed to model human behaviour in complex real-world systems. Starting from a generic model of operator behaviour, it is possible to refine the control flow to represent many individual systems. This section presents the generic OCM, and the following sections will provide an example of its application to a simplified conflict detection task in ATC.

2.1. General Form

The Operator Choice Model (OCM) represents a general framework for modelling the behaviour of an operator attending to a number of concurrent, interleaved tasks with information presented through an interface. It is based upon psychological theories of decision-making (Lindsay & Connelly, 2002). The aim of the OCM is not, however, to model human cognition; rather it is to provide a means of simulating realistic operator *behaviour* in complex tasks evolving unpredictably through time.

The basic form of the OCM is given in Figure 1 in state chart notation (Fowler et al., 1997). It separates the task of the operator into four component processes that are typically expected to occur *sequentially*:

Scan: This is the starting state where potential problems ("items") are identified. There can be a number of items to which the operator might attend; thus the Scan process requires a method for selecting one to attend to.

Classify: Once attending to an item, the operator determines whether it is a problem requiring action. If so, the operator proceeds to Decide Action; if not, the operator returns to Scan.

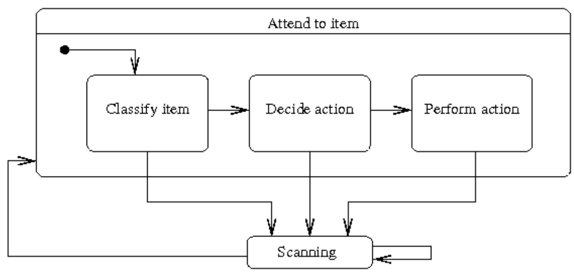

Figure 1. General form of the OCM

Decide Action: The operator develops an appropriate plan of action.

Perform Action: The operator implements the plan of action. This typically requires interactions with external devices or humans. After the actions are complete, the operator returns to Scan.

During the Decide Action and Perform Action processes, the operator might return to Scan if it is appropriate to do so.

The OCM can be applied to situations in which there are a variety of types of items to attend to. All items are incorporated into Scan, and different item types can have different Classify, Decide Action, and Perform Action definitions associated with them. When used to perform a safety assessment, the OCM provides a basis for quantitative risk analysis of complex, interleaved tasks. There is potential for operator error during each process. For example the operator might incorrectly classify an item as not requiring further action, leading to system failure. A formal analysis of erroneous operator behaviour in the OCM, applied to the ATC context, is presented in Cerone et al. (2005).

The options available to the operator within each of the four key processes are specific to the domain of application. These sub-processes can also be depicted as state diagrams. This provides a means of capturing the "flow of control" of operator activities for the particular task, and the different possible outcomes of those activities.

The different sub-states (processes and events) and transitions can be described using Hoare's CSP (Hoare, 1985). However to also capture stochastic behaviour including the time taken for each process, an extended framework is required. Continuous Time Probabilistic Automata (CTPA) provides such a framework. These describe the available states, the possible transitions between states, and specify stochastic models for the transitions and their duration (Hung and Chaochen, 1999).

3. APPLICATION TO A CONFLICT DETECTION TASK

En-route air traffic controllers regulate the speed, bearing and flight level of the aircraft within their sector, in order to maintain separation between aircraft and to manage the flow of air traffic. Key aspects of the task have been identified in a task analysis (Neal et al., 1998). One of these tasks is to ensure that aircraft maintain a minimum distance of separation at all times. Aircraft are said to be *in conflict* if their flight paths will cause them to violate separation at some future point in time

We have conducted a series of experiments in which subjects are required to monitor a sector of airspace and make judgements about whether aircraft will violate separation, and take action to prevent this occurring. The goal is to illustrate the use of the OCM in simulating human performance in a dynamic task environment. It is for this reason that we are using a highly simplified simulation. Below we describe the experimental design for the first of these experiments, and a calibration of the OCM constructed to model this task.

3.1. Conflict Detection Experiment

Twenty-seven first-year undergraduate university students, with no prior experience in ATC, were instructed to observe pairs of aircraft on a simulated ATC radar screen. Aircraft flew in straight lines at the same altitude. The subjects were asked to classify converging pairs of aircraft as conflicts or non-conflicts, according to a minimum separation requirement of 5km. Once a classification was reached the subjects clicked the mouse on a button to identify a pair as a conflict or a non-conflict.

The variables which describe each aircraft *pair*, passing through a common waypoint, are as follows:

1. Distance of Minimum Separation (DOMS). The closest distance a pair of aircraft will come to each other;
2. Time to Minimum Separation (ttms). The time until the pair reaches DOMS;
3. Angle (θ). The angle between flight paths of the two aircraft;
4. Velocity of the aircraft which passes through the waypoint first (V_1) and second (V_2)

The experiment consisted of two phases, a training phase and a test phase. The training phase was designed to expose subjects to the task and give them practice in classifying aircraft pairs. Our analyses focus on the test phase data.

A 4x3x3x3 repeated measures experimental design was used for the test phase. The levels of each experimental variable were: DOMS = 1.25km, 3.75km, 6.25km, 8.75km; ttms = 33, 66, 99 seconds; θ = 45°, 90°, 135°; (V_1, V_2) = (660km/hr, 927km/hr), (927km/hr, 660km/hr), (927km/hr, 927km/hr). One aircraft pair was presented at a time. Subjects were given 10 seconds during which they were free to respond; if they had not responded at the end of the 10 second interval, they were required to respond immediately. The simulator recorded each participant's classification response and reaction time.

4. THE OCM CALIBRATION FOR THE CONFLICT DETECTION TASK

The goal of the calibration is to achieve a form for the OCM which will simulate the distribution of behaviour of the student subjects in the classification experiment. Ideally this includes both the distribution of classifications and the distribution of reaction times.

For this task, the *Classify* process of the OCM is the most involved to calibrate. A statistical modelling method was used for this purpose and used the data from the classification test phase experiment. The *Scan*, *Decide Action*, and *Take Action* components are minimised by the design of the classification experiment. For the timing component we explored the use of a simplistic approach. It involved setting short, fixed, time intervals for each transition, and achieving a distribution of reaction times by the random number of loops that occur through the OCM, rather than by using random timing intervals themselves. This approach uses a parameter called p_{Final}, which controls, stochastically, the number of times the operator loops through the OCM before making a final decision. We show the reaction time results produced under different values for this parameter by the OCM, and compare these with the results from the student subjects.

A state chart diagram of the OCM used for the classification test phase experiment is given in Figure 2. The CTPA specification is detailed in the sub-sections below.

4.1. Scan

This process models the operator taking time to observe all possible conflicts, and either returning to Scan or selecting a pair of aircraft to attend to. The states are {scan, attend_i}, where *i* ranges over the items, in this case, aircraft *pairs*. The possible transitions are:

$scan \rightarrow scan$

$scan \rightarrow attend_i$

where *i* runs over aircraft pairs. In the test phase classification experiment, only a single pair of aircraft appear on screen at a time. Thus there is no choice between competing items and this part of the OCM is effectively circumvented. The probability models for transitions are:

$\Pr(scan \rightarrow scan) = 0$

$\Pr(scan \rightarrow attend_i) = 1$

The timing model for this is non-stochastic and is set to:

$\text{Duration}(scan \rightarrow scan) = 0\ s$

$\text{Duration}(scan \rightarrow attend_i) = 0\ s$

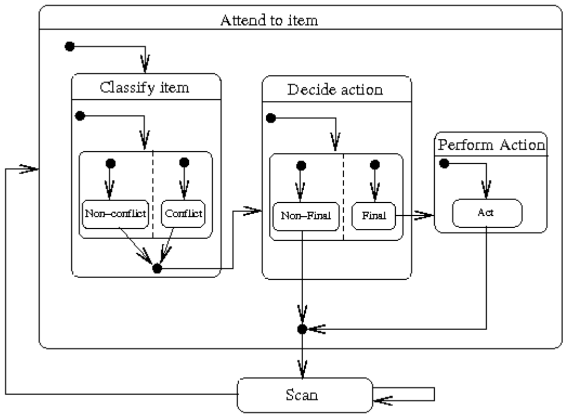

Figure 2. OCM model of ATC task

4.2. Classify

This process represents the operator attending to a pair of aircraft and making a judgement as to whether the pair is, or is not, in confict. In this version of the OCM there is no option to defer the classification and return to Scan. This is a "forced decision" model, in which the operator makes a classification in the first loop through the OCM. It is used in this situation because the operator is limited to 10 seconds in which to act. The states in the classify process are {$attend_i$, $conflict_i$, $non\text{-}conflict_i$}, and the possible transitions are:

$attend_i \rightarrow conflict_i$

$attend_i \rightarrow non\text{-}conflict_i$

In the probability model, we use *L* to denote the probability of classifying a pair as a conflict. The value of *L* is determined from experimental data, and is a function of the features of aircraft pairs (DOMS, ttms, θ, V_1, and V_2). The probability model is expressed as:

$\Pr(attend_i \rightarrow conflict_i) = L$

$\Pr(attend_i \rightarrow non\text{-}conflict_i) = 1 - L$

To obtain an appropriate function for *L*, in terms of the geometry of aircraft pairs, a logistic regression model was fitted to the experimental data. Logistic regression is used to fit models to proportions (McCullough and Nelder, 1989). The dependent variable is the probability of classifying a pair as a conflict, and the 5 independent variables are DOMS, ttms, θ, V_1, and V_2. All independent variables contributed significantly to the fit of the model in the experimental data, and there were several significant interaction terms. The model obtained was:

$L = \text{Logit}^{-1}(\beta_0 + \beta_1 \text{ DOMS} + \beta_2 \cos\theta + \beta_3 \sin\theta +$

$+ \beta_4 V_1/V_2 + \beta_5 V_2/V_1 + \beta_6\, 1/\text{ttms} + \beta_7 V_1/V_2 \sin\theta$

$+ \beta_8 V_2/V_1 \sin\theta + \beta_9 \text{ DOMS/ttms}),$ (1)

where $\text{Logit}^{-1}(x) = \exp(x)/(1+\exp(x))$, the fitted parameter values are given in Table 1.

Table 1. Parameters in logistic classification model

i	β_i	i	β_i
0	33.185	5	-18.273
1	-0.281	6	161.779
2	1.223	7	20.918
3	-51.433	8	27.584
4	-12.132	9	-29.112

The timing model sets a short, non-stochastic, duration for this process:

Duration($attend_i \rightarrow conflict_i \mid non\text{-}conflict_i$) = 0.5 s

4.3. Decide Action

The decide process for this calibration identifies if the classification of conflict or non-conflict made during the classify process is *final* or *non-final*. If it is final, the operator proceeds to Perform Action. If it is non-final, the operator returns to scanning. A classification has been made however, so that if the 10 second period runs out without a final classification, the operator is "forced" to respond with the non-final classification. The states are $\{conflict_{i,}, non\text{-}conflict_i, final_i, non\text{-}final_i, scan\}$, and the possible transitions are:

$conflict_i \rightarrow final_i$
$conflict_i \rightarrow non\text{-}final_i$
$non\text{-}conflict_i \rightarrow final_i$
$non\text{-}conflict_i \rightarrow non\text{-}final_i$
$non\text{-}final_i \rightarrow scan$

In this version of the ATC OCM we use a simple model in which the operator proceeds to a final decision stochastically with probability p_{final}. Thus, the probability of the various transitions are:

Pr($conflict_i \mid non\text{-}conflict_i \rightarrow final_i$) = p_{final}
Pr(conflict$_i$ | $non\text{-}conflict_i \rightarrow nonfinal_i$) = 1- p_{final}

The timing model sets a small amount of time for this process:
Duration($conflict_i \mid non\text{-}conflict_i \rightarrow$
$non\text{-}final_i \mid final_i$) = 0.5 *s*

If the decision is $nonfinal_i$, then the operator returns to scan with probability 1, taking 0 seconds. Thus,

Pr($nonfinal_i \rightarrow scan$) = *1*
Duration($nonfinal_i \rightarrow scan$) = *0 s*

Table 2. Comparison of Subjects and OCM in Proportion of Conflict Classifications

Condition	DOMS	Students	OCM
A	1.25	1.00	0.99
	3.75	0.92	0.86
	6.25	0.12	0.26
	8.75	0.04	0.02
B	1.25	0.85	0.91
	3.75	0.73	0.63
	6.25	0.15	0.22
	8.75	0.04	0.03
C	1.25	0.85	0.69
	3.75	0.27	0.34
	6.25	0.00	0.11
	8.25	0.00	0.03

4.4. Perform Action

The only action taken in the classification experiment is to click a button with the mouse to indicate conflict or non-conflict. The states are $\{final_i, act_i\}$, and the only possible transition in the Take Action process is:

$final_i \rightarrow act_i$

Assuming this takes one second, the probability and timing for the transition is:

Pr($final_i \rightarrow act_i$) = 1
Duration($final_i \rightarrow act_i$) = 1 s

After this the operator returns to Scan taking zero seconds:

Pr($non\text{-}final_i \rightarrow scan$) = 1
Duration($non\text{-}final_i \rightarrow scan$) = 0 s

5. COMPARISON OF HUMAN SUBJECTS AND THE OCM

In order to assess the performance of the OCM calibrated as above, it was used in simulations with the same design as the classification experiment with student subjects. We also varied the parameter p_{final} in order to compare reaction time distributions. The values chosen were: p_{final} = 0.1, 0.25, 0.5, 0.75, 0.9. Each run was repeated 1,000 times. Below we compare classification accuracy and reaction time distributions between the human subjects and the OCM.

5.1. Classification Accuracy

Results comparing the classifications of the student subjects and the OCM for a subset of the experimental design are given in Table 2. Results for the other experiments were similar. The results given in Table 2 are for the subset:

A: ttms=33, θ=45, (V_1,V_2)=(660,927);

B: ttms=66, θ=90, (V_1,V_2)=(927,660); and
C: ttms=99, θ=135, (V_1,V_2)=(927,927).

5.2. Reaction Time

Histograms of the distribution of reaction times produced by the OCM from 1,000 runs, for p_{final}=0.25 and p_{final}=0.75, are given in Figure 3. These results are for the OCM experiment with DOMS = 1.25, ttms = 66, θ = 90, and (V_1,V_2) = (927,660). The distributions for the other experiments with the same p_{final} values are similar. A key point with this calibration is that the reaction time distributions depend only on the value of p_{final}.

In order to compare reaction time distributions between the OCM and the subjects, a histogram of reaction time results for the latter is given in Figure 4. This includes the reaction times from all the experiments in the classification test phase combined.

On initial inspection the OCM reaction time distribution with p_{final} = 0.75 given in Figure 3a, bears a resemblance to the reaction time distribution for the student subjects. Indeed the mean reaction time for the students was 5.1s, and from the OCM experiment with pfinal = 0.25 was 5.1s. Further, the proportion of observations at ten seconds, which corresponds to the "tail" of the distribution, was 0.12 for the student data, and 0.10 for the OCM data. Thus, the simplistic model for timing used in the OCM produces similar results on these indicators. The way the OCM has been defined here, reaction times follow the *geometric* distribution. This distribution starts at the maximum (or "mode") and does not have the initial increase to a maximum that is seen in the reaction time distributions from the student subjects.

6. DISCUSSION

The OCM provides a means of modelling the behaviour of an operator attending to complex, interdependent tasks. It breaks up the activities of the operator into four basic processes which are performed in sequence, and enable the operator to consider each "item" one at a time.

Calibration of the OCM for application in specific domains requires human data in order to produce realistic and reliable results. Here we have illustrated calibration of the OCM based on a simplified ATC task. The approach of logistic modelling has proven useful in achieving realistic classification data. The use of a very simple approach to achieving reaction times has been explored. The distribution is achieved not by using random durations for transitions, which are part of the definition of CTPA, but rather by using random numbers of "loops" through the OCM.

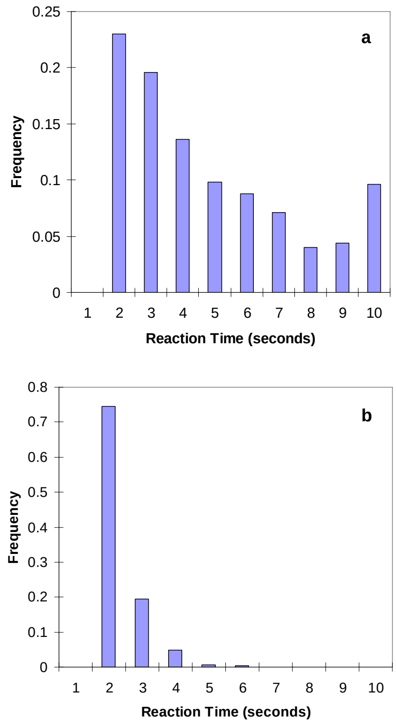

Figure 3. Reaction time for OCM experiment a: p_{final} = 0.25 b: p_{final} = 0.75

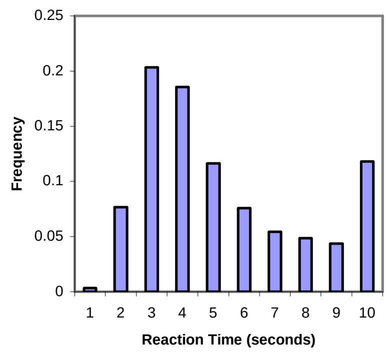

Figure 4. Histogram of reaction times from student subjects

This produces a discrete, rather than a continuous, reaction time distribution, and it fails to capture the initial increase to a maximum seen in the human experimental data. Indeed, a recent and widely used modelling approach for reaction time involves using the *ex-Gaussian* distribution

(Brown & Heathcote, 2003), which combines a Gaussian distribution and an exponential distribution. The former produces the initial peak, and the latter produces the "tail". There is no simple way to obtain a similar distribution from the OCM, using the approach of fixed time for transitions and random numbers of "loops" through the OCM.

An alternative approach is to use stochastic models for transition times directly in the CTPA. Using this approach, both the transitions followed, and the time taken for the transitions, are stochastic. Conceivably any distribution function for durations can be used, and can depend upon values of variables, in the same way as the classification function does in the calibration given here. This approach, and others, will be explored in future work. Work is also currently underway applying this modelling approach to the analysis of conflict detection performance by real air traffic controllers in the field.

Acknowledgements

We wish to gratefully acknowledge the contribution of Colin Ramsay to programming the OCM. This research was supported by a grant from the Australian Research Council and by funding from the ARC Centre for Complex Systems.

References

Brown, S. and Heathcote, A. (2003). QMLE: Fast, robust, and efficient estimation of distribution functions based on quantiles. *Behavior Research Methods, Instruments & Computers,* 35(4): 485-492.

Cerone, A., Lindsay, P.A., and Connelly, S. (2005), Formal analysis of human-computer interaction using model-checking, *3rd IEEE Conference on Software Engineering and Formal Methods*, Sept 2005.

Hannaman, G., Spurgin, A., and Lucic, Y. (1985). A model for assessing human cognitive reliability in PRA studies. *1985 IEEE Third Conference on Human Factors and Power*, pages 343–353.

Hoare, C. A. R. (1985), *Communicating sequential processes*, Prentice-Hall International.

Hung, D.V. and Chaochen, Z. (1999), Probabilistic duration calculus for continuous time, *Formal Aspect of Computing*, 11:21-44.

Johnson, C. (1997). Reasoning about human error and system failure for accident analysis. In Howard, S., Hammond, J., and Lindgaard, G., editors, *Human-Computer Interaction INTERACT '97*, pages 331–338. Chapman and Hall.

Kallus, K., van Damme, D., and Dittmann, A. (1999). Integrated task and job analysis of air traffic controllers – phase 2: task analysis of en-route controllers. Technical Report HUM.ET1.ST01.1000-REP-04, European Organisation for the Safety of Air Navigation (Eurocontrol).

Kirwan, B. (1990). Human reliability assessment. In *Evaluation of Human Work*, chapter 28. Taylor and Francis.

Lindsay, P. and Connelly, S.. Modelling erroneous operator behaviours for an air-traffic control task. In J. Grundy and P. Calder, editors, *Third Australasian User Interfaces Conference (AUIC2002)*, volume 7 of *Conferences in Research and Practice in Information Technology*, pages 43–54. Australian Computer Society, Inc, 2002.

McCullough, P. & Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edition. Chapman & Hall, London.

Neal, A., Griffin, M., Neale, M., Bamford, E., and Boag, C. (1998). *Human factors issues in the transition to a cns/atm environment: Final report.* Report to Airservices Australia, The University of Queensland.

Palanque, P., Bastide, R., and Paterno, F. (1997). Formal specification as a tool for objective assessment of safety-critical interactive systems. In Howard, S., Hammond, J., and Lindgaard, G., editors, *Human-Computer Interaction INTERACT '97*, pages 323–330. Chapman and Hall.