Validation of Human Behavioural Models

Andrew Belyavin
QinetiQ Ltd
Farnborough
UK

ajbelyavin@qinetiq.com

Brad Cain
Defence Research and Development Canada
Toronto, Ontario
CANADA

brad.cain@drdc-rddc.gc.ca

ABSTRACT

Full validation of a model involves a number of steps. The first is to ensure that the model represents the required domain adequately (content validation). The second is to ensure that the principles underlying the model make reasonable use of current understanding of the problem space (construct validity). If the model meets these two criteria there is a requirement that the predictions of the model represent what happens in the "real world" to an adequate degree (predictive validity). The predictive validity of models that characterise human physiological response or low level human physical and cognitive performance can be conducted using statistical tools suitable for the analysis of interval data such as analysis of variance. When a model is developed that describes choice of course of action, an important element of human behavioural modelling, the outcomes are necessarily discrete and the volume of data available for analysis is typically smaller than desirable for validation over a broad scope. Any stream of similar decisions in a military context is likely to be aimed at maintaining the real world outcome close to a desired profile drawn up at the planning stage. In this way the process of taking decisions and monitoring their implementation is analogous to the process of tracking, embodied in such activities as driving a vehicle. The approach is applied directly to a tracking task to illustrate the interaction between a stream of decisions and outcomes and the problems of generalising the approach to more complex situations is discussed.

1.0 INTRODUCTION

Validation of human models has been the topic of a number of papers over the past decade since the team headed by Pew and Mavor (1998) published their seminal work on the state of the art of Human Behaviour Representation. Many of these papers lament the lack of validation in Human Behaviour Representation (HBR) and human performance models and while a number do directly compare predictions with observations (e.g. Foyle et al., 2005), many immediately fall back on informal, face validation: TLAR (that looks about right) or BOGSAT (bunch of guys sitting around the table: Campbell & Bolton, 2005). For many, colloquial definition of the validity of a concept or a model means accurate representation of real world events (Trochim & Donnelly, 2007). In general, absolute comparisons with the real world may not be the most appropriate starting point for addressing the validity of a model. Formal models are typically abstractions of the processes that we believe explain observed events, and therefore models often deliberately ignore aspects of the real world experience. Trying to validate a model as an accurate representation of the real world events is, in this strict sense, doomed to failure, and an alternative approach should be sought.

1.1 Problems of Validating HBR Models

There are particular challenges in the validation of HBR models. The study of HBR in constructive simulation conducted by the HFM 128 panel (Lotens et al., 2009) identified a large number of processes that have to be represented in a complete model of human behaviour, including perception, cognition, physiology and interactions between these elements. The study concluded that an important element of

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **OCT 2010**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Validation of Human Behavioural Models**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **QinetiQ Ltd Farnborough UK**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADA564696. Human Modelling for Military Application (Applications militaires de la modelisation humaine). RTO-MP-HFM-202

14. ABSTRACT
Full validation of a model involves a number of steps. The first is to ensure that the model represents the required domain adequately (content validation). The second is to ensure that the principles underlying the model make reasonable use of current understanding of the problem space (construct validity). If the model meets these two criteria there is a requirement that the predictions of the model represent what happens in the real world to an adequate degree (predictive validity). The predictive validity of models that characterise human physiological response or low level human physical and cognitive performance can be conducted using statistical tools suitable for the analysis of interval data such as analysis of variance. When a model is developed that describes choice of course of action, an important element of human behavioural modelling, the outcomes are necessarily discrete and the volume of data available for analysis is typically smaller than desirable for validation over a broad scope. Any stream of similar decisions in a military context is likely to be aimed at maintaining the real world outcome close to a desired profile drawn up at the planning stage. In this way the process of taking decisions and monitoring their implementation is analogous to the process of tracking, embodied in such activities as driving a vehicle. The approach is applied directly to a tracking task to illustrate the interaction between a stream of decisions and outcomes and the problems of generalising the approach to more complex situations is discussed.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**12**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

any model representation is the division into internal state and consequent performance. Identifying internal states has a long history in psychological theory of more than 100 years since the Yerkes Dodson law of optimal arousal was developed in 1908 (Yerkes & Dodson, 1908). A state such as arousal is fundamentally a model construct and is intrinsically unobservable. It is not possible to validate models of the evolution of such states by direct comparison with real world observations, and indirect methods have to be employed. Progress has been made with some of the state constructs by using subjective observations, such as subjective measures of alertness. It has proved possible to relate alertness to the experience of individuals in terms of sleep patterns and time of day (Belyavin and Spencer 2004) and to demonstrate that a subjective assessment made under similar conditions is reproducible – a minimum requirement for the definition of a state. Similar problems arise with the definition of elements of cognitive performance in that most of the processes cannot be observed directly. At a higher level, the same strictures apply to the elements of interactions between individuals.

1.2 Validation Criteria

To meet this challenge Cronbach & Meehl (1955) proposed a more broadly based approach to the validation of psychometric models. They proposed that validation should be conducted using three assessments of validity: Construct Validity, Content Validity and Predictive Validity. The definitions of the three validity criteria are as follows:

- Construct validity is attained if the model is built using accepted theoretical constructs about how the object in question functions or accepted abstractions of the object to be modelled are deemed suitable for the intended use.
- Content validity is attained if the range of applicability of the model, that is the range of independent variables and component models, meets the requirements criteria of its intended use and, in particular, encompasses the range of applications proposed.
- Predictive validity is attained if a model a capable of reproducing real-world observations to the required degree of fidelity for the proposed application of the model.

Construct validity is based on a Subject Matter Expert (SME) assessment of the foundations of the overall model and its components. This implies that both the individual components and their modelled interactions should be subjected to the same process. If an HBR formally models internal state and uses this to moderate some aspect of cognitive performance, the process of moderation has to be valid as well as the model of the evolution of state and the distinct model of cognitive performance.

Content validity should also be applied to each of the component models separately and to the way the components interact. The key question is whether the phenomena represented by the models span the range demanded by the requirement and whether the parameters used to define the models span a plausible space of values in that context. The majority of the judgments again have to be based on SME opinion, backed by measures where they are available.

Predictive validity is tested by comparing the output of the model with real-world observations. Ideally formal statistically methods should be employed to make the comparisons although *in extremis* SME opinion may have to be accepted. In principle, a multi-component simulation can pass the predictive validation criterion if it is able to predict the pattern of real-world data that were not used to build the model. The weakness in this logic is that any simulation involving multiple components could satisfy this criterion and yet be built with individual components that would not meet the target if considered in isolation. An HBR model is particularly vulnerable to this possibility in that there are many elements of human physiology and psychology that may be represented in a full HBR model, that are homeostatic – provide negative feedback in control systems terms – in that they tend to restore a defined state. Since the defined state will be known, it is possible to have incorrect details in these models – in terms of open-loop

properties – but the defined state is appropriately restored and in this way the overall model appears valid, although it is incorrect in detail.

In an earlier paper (Belyavin and Cain, 2009) we described the predictive validation of a whole body thermal model using experimental data that was independent of that used to construct the original model. This is a predictive model of individual thermal state and in principle the model predictions can be compared directly with observations drawn from the real world. The particular whole body thermal model subjected to validation was a rationally based model composed of sub-models of a number of distinct processes including thermal generation, thermal conduction, thermal convection through blood flow, sweating, shivering and dynamic changes in blood flow to the skin. In addition the model predictions of deep body temperature for comparison with observations are derived from models of the temperature that is measured. It was concluded that for the full range of experimental conditions assessed the model did not meet a stringent definition of predictive validity. A limited assessment of some component models was conducted and it was concluded that the thermal generation model met the criterion of predictive validity but the model of sweating did not meet the criterion fully for the set of individuals tested in the experiment.

It was possible to conduct a detailed analysis of the thermal model and its components in this earlier analysis because the validation could be based on interval measures – temperatures or sweat rates – and it is possible to employ statistical tools such as analysis of variance or multiple regression that enable the contribution of different aspects of changes in the external conditions to be assessed in detail. If the outcome measures are categorical and less frequently measured in time, it is much harder to achieve the same level of detail in the analysis and validation is more difficult. The aim of the present paper is to consider the challenge presented by validating models of those elements of human behaviour that are embodied in decision making rather than state.

1.4 Validating Decision-Making Models

Military decisions are made at a wide range of different levels and frequencies, ranging from those made by individuals involved in dismounted combat to strategic levels made at national or international level. The full range of models of HBR must include models that represent decision-making at all these levels. Validation of decisions that are intrinsically infrequent, such as strategic level decisions, is difficult because of the limited volume of data available to support predictive validation and each decision is individual in that it depends on the precise context in which is made. A frequently used approach to the modelling of human decisions that are made rapidly under time pressure is to represent these decisions by using a pattern recognition algorithm. A choice between two decisions can be described as a multivariate discrimination between the outcomes and the simplest form of such an algorithm is the application of a linear algorithm to make the choice as proposed by Fisher (1936). The criterion that determines the selection of the particular choice is expressed as a function of the perceived cost of making the wrong choice and this may depend on context and the personality of the decision-maker. The approach can be elaborated by including the quality of the perception of the variables that are the basis of the choice and non-linear choice functions can be constructed.

In whatever way decisions are modelled there is a need for replication of similar decisions if there is to be a possibility of applying statistical methods to parameterise and validate the model. It is argued in the present paper that compensatory tracking is a source of a stream of similar decisions that can be used to parameterise a simple pattern recognition decision-making model. A model of compensatory tracking behaviour is described in Belyavin and Farmer (2006) and the application of the same model to describing pilot tracking behaviour is described in Belyavin et al (2009). The tracking model is described in Section 2 and a procedure for fitting the model is described and the possible outcome measures that can be used for validation are considered. The implications of the analysis for more complex situations are discussed in Section 3.

2.0 MODELLING HUMAN TRACKING BEHAVIOUR

2.1 Background

There is a long history of development of linear control models to describe human tracking performance that represent the human processes of perception and cognition as transfer functions modified by the addition of a stochastic remnant. This formulation has been extended in the Optimal Control Model (OCM) to encompass the optimisation of the human model parameters such that an objective function comprising a weighted combination of control error and control effort can be minimized, making suitable assumptions about human performance (Baron *et al* 1970). This represents human tracking behaviour as a continuous activity described by a simple continuous control law. Direct observation of human tracking behaviour suggests that in practice the operator makes a series of discrete control decisions rather than a continuous flow of movement.

A two dimensional compensatory tracking task has been constructed in which the participant under test uses a joystick that drives X and Y velocity to cancel a velocity disturbance constructed from 6 sinusoids with wavelengths ranging from 16 seconds to 1 second. The goal of the task is to maintain a cursor within a target region at the centre of the screen. A 20 second sample of a subject's joystick control input for the X axis of the two axis compensatory tracking task is displayed in Figure 1 and it can be observed that there are short intervals for which the joystick position is constant and between these intervals there tends to be steady linear movement of the joystick.

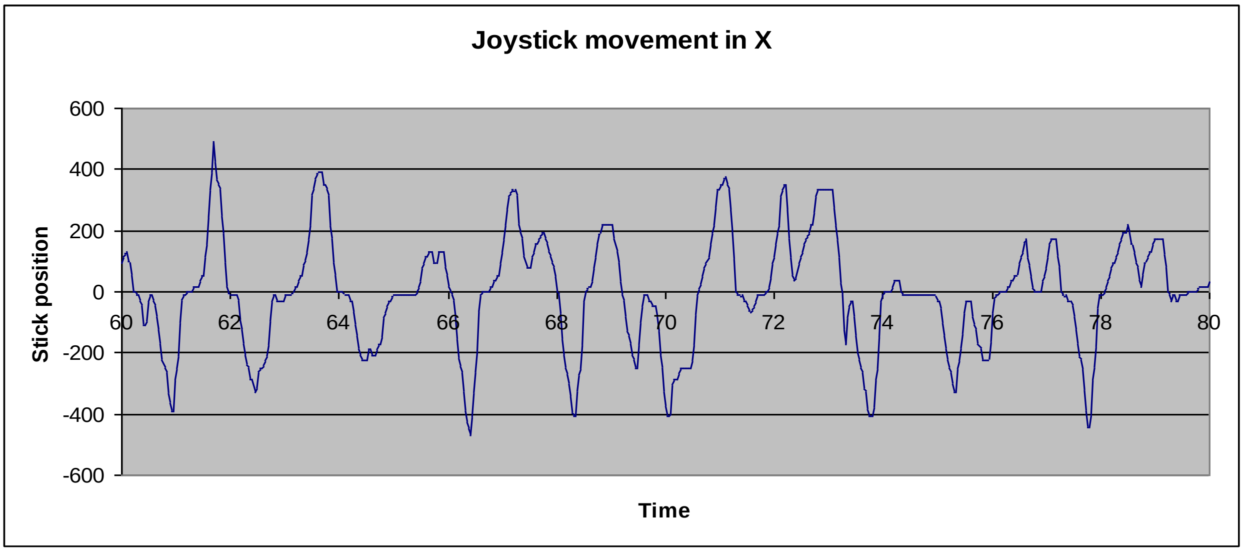

Figure 1: Joystick Position for the Control of X Position in a Compensatory Tracking Task.

2.2 Discrete Model of Human Tracking Behaviour

The model of human tracking behaviour is based on five simple assumptions that were established following analysis of observed tracking data:

1) Human control of a continuous psycho-motor task is characterised by a sequence of discrete decisions and responses to mismatches between a desired condition and the perceived current condition.

2) There is a lag between the perception of current condition and the implementation of any decision to adjust corrective action.

3) The decision to adjust corrective action is stochastic and depends on the perceived future deviation from the desired condition.

4) The corrective response is approximately linear in the perceived future deviation from the required condition and the current corrective action.

5) There is a "rest-period" between any decision or action and the subsequent assessment of the situation.

To satisfy these assumptions, the controller comprises a continuous cycle of monitoring current conditions coupled with a probabilistic decision to change the current control position. The timings of all the monitoring and control elements are based on values drawn from the human performance literature. Intra-person variability is built into the model through the representation of the variability of individual decisions in response to the external environment. Inter-person variability is represented by variation in the parameters describing the decisions made to move the controller and the amplitude of the control movements. The structure of the controller is displayed in Figure 2. The controller is constructed as a set of discrete tasks represented by the green boxes that are executed in sequence according to the logical flow. The only modification to a simple linear flow is the decision as to whether to move the control or not, represented by the green diamond. The time taken to perform each task is determined from standard human engineering data or by calibration of the model.

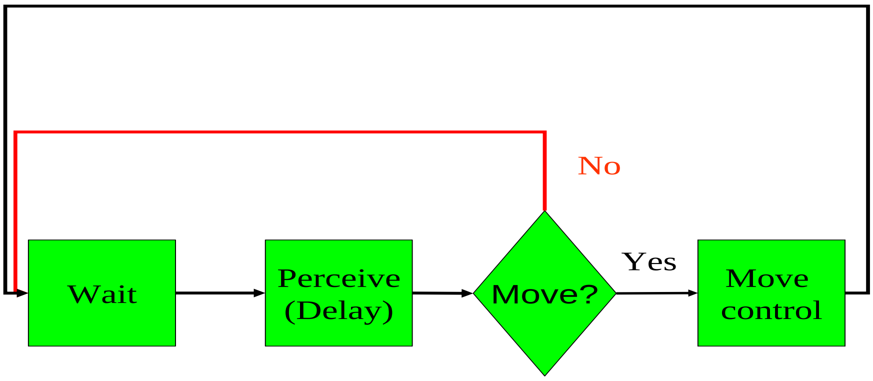

Figure 2: Task cycle for the discrete model of tracking behaviour.

The key elements of the model are the equations determining how much movement of the stick is required and whether to make the move. To test the form of the decision-making model it is assumed that the deviation of the cursor position on the screen is perceived exactly. The equation defining the perceived required control movement *dC* is defined as a modified Proportional-Differential (PD) controller in Equation (1). The additional term in current controller position was derived from preliminary analysis of tracking data as part of the initial model development (Belyavin and Farmer 2006) and the divisor of the PD term was included to improve model stability.

$$dC = \frac{\mu(Err + \eta\, dErr)}{1 + \gamma(C - C_{ref})^2} - \lambda(C - C_{ref}) \qquad (1)$$

The variable *Err* is the current deviation of the cursor from the screen centre, *dErr* is the rate of change of current deviation of the cursor from screen centre, *C* is the current joystick position, C_{ref} is the neutral position of the joystick and μ, η, λ and γ are model parameters. If γ=0 the model is exactly linear in the key decision parameters. The probability that a control movement is to be made, P, is determined by the perceived required control movement according to Equation (2), where σ and τ are model parameters. The

model form was constructed by preliminary analysis of the incidence of control movements to support the development of the Boeing 747-400 pilot model (Belyavin et al 2009).

$$P = \frac{1}{1 + \exp\left[-\sigma(| dC | - \tau) \right]} \tag{2}$$

The stochastic nature of the decision-making process ensures that the model is not exactly linear but it s close to linear in practice.

2.3 Fitting the Model Parameters and Findings

There is a large number of metrics describing performance on a tracking task that could be used for assessing whether a model matches observed performance including the Root Mean Square Error (RMSE) of cursor deviation from desired position, the spectral characteristics of joystick movement and the properties of the response in terms of the linearity of joystick response. The latter measure can be derived if the disturbance function for the tracking task comprises a combination of distinct sinusoids in that if the power spectrum of joystick movements includes power at frequencies not contained in the disturbance function the source of the power must be non-linearity in the response function. The velocity disturbance function is constructed out of a combination of 6 sinusoids and is given by the expression in Equation (3)

$$D = \sum_{i=1}^{6} r A_i \sin(r\omega_i t) \tag{3}$$

Where the values of ω_i and A_i are selected so that the peak in the power spectrum for the disturbance is at the fourth wave and the total RMSE of the integrated position disturbance is approximately independent of r. As the value of r is varied the required frequency of control movements is varied while maintaining the overall positional disturbance. The task is started from a selected large value of t so that the initial velocity disturbance is small but the waves are not in phase.

After preliminary experimentation it was concluded that the model parameters could be estimated for each participant independently by matching the linear component of the response model using the estimated gains and phases of the joystick response for the sinusoids contained in the velocity disturbance function. The tracking model is stochastic in that the “Move?” decision is determined probabilistically. It is therefore not possible to do a simple fit between deterministic model outputs and observed outputs to define model parameter values, assuming variation in the observations alone. The Nelder-Mead simplex method (Nelder and Mead 1965) was selected to perform the fit as it is well suited to the problem of fitting stochastic models in that it requires local coherence rather than precise continuity of the objective function and convergence is determined based on the variability of the objective function rather than exact reproduction of the minimum value.

The findings for an experiment involving 8 participants were summarised in Belyavin *et al* (2009). Eight participants were tested at three levels of base disturbance frequency, where the amplitude was compensated to ensure a constant root mean square error for the cursor as a result of the disturbance. The results from each participant and tracking rate were calibrated using the Nelder-Mead procedure by matching the observed and modelled gains and phases for the sinusoids with the x and y forcing functions using least squares analysis. The parameters fitted for each participant/rate combination were common values of μ, η, λ and τ for both x and y, and a value for the time taken to perform the “Wait” task displayed in Figure 1. A summary of the fitted parameter values is displayed in Table 1.

Table 1: mean values of fitted parameters for the tracking task.

Parameter	Low rate (r=0.5)	Medium rate (r=1.0)	High Rate (r=1.5)	Standard Deviation
μ	1.761	1.584	1.210	0.242
η	1.093	1.191	1.067	0.188
λ	0.414	0.545	0.656	0.104
τ	27.97	29.87	32.01	7.56
Wait time	0.130	0.085	0.102	0.029

The parameters were investigated using analysis of variance. It was concluded that Wait time and μ differed between participants ($p<0.001$) and that μ and λ differed between rates ($p<0.001$). The observed and predicted RMSE were compared for the model and observations. The findings are displayed in Figure 3 and a plot of the observed and expected RMSE for the 8 subjects for tracking at the low rate are displayed in Figure 4.

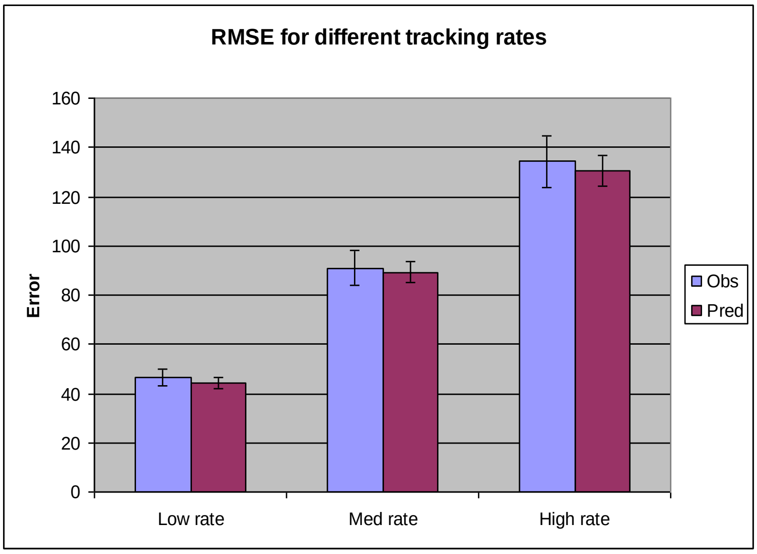

Figure 3: Comparison of observed and predicted RMSE for the three different tracking rates.

From Figure 3 it can be seen that there is generally a good match between mean observed and predicted RMSE at all three tracking rates. The spread of RMSE between individuals tends to be larger for the observed than the predicted data as shown in the standard errors displayed in Figure 3. This is confirmed from the plot of individual scores shown in Figure 4 where the observed values for the 'poor' performers tend to be larger than those predicted by the model. The model represents a systematic approach to the task and 'poor' performers may undertake the task in a different way from that proposed by the model.

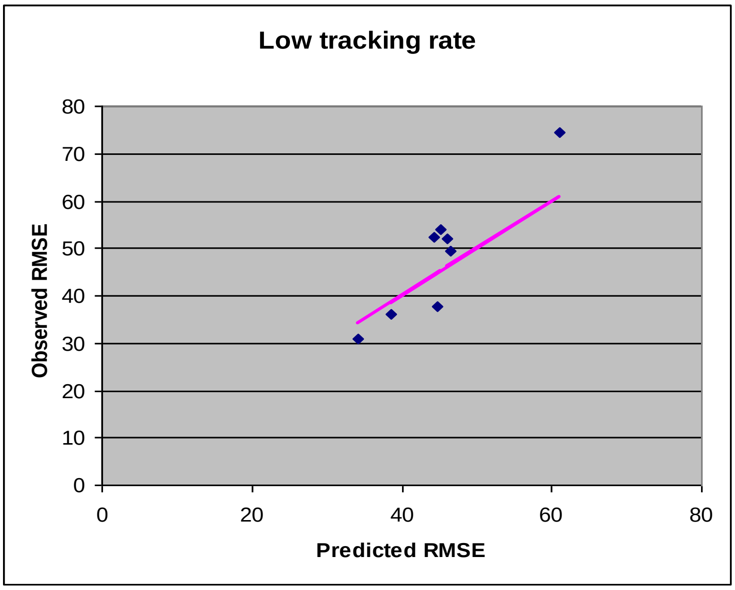

Figure 4: Observed and predicted RMSE for the 8 subjects at the low tracking rate including the Y=X line.

The general linearity of the model is broadly consistent with the observations in that the observed percentage power in the joystick response for X is 69% and that for the model is 68% and for Y the observed value is 72% and the predicted rate is 61%. A Bode plot for the observed and predicted gains for the linear component is displayed in Figure 5, plotting all three task rates on one graph.

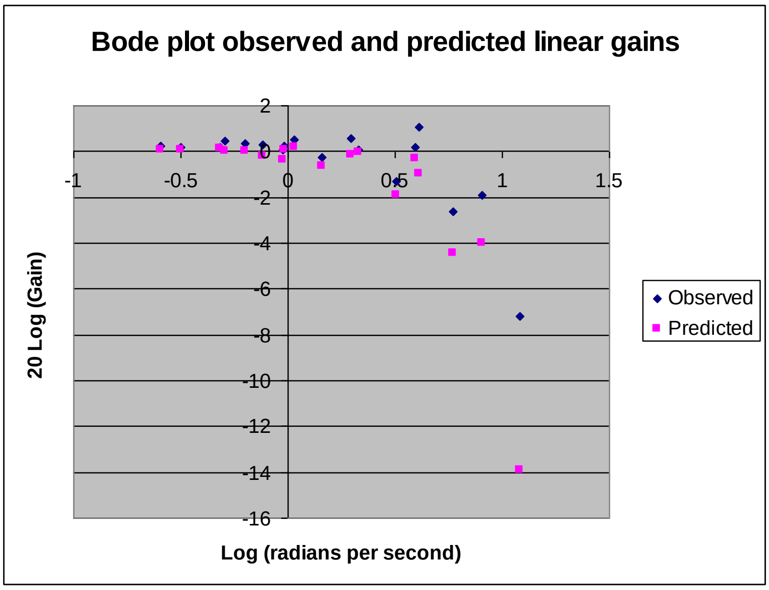

Figure 5: Observed and predicted linear gains.

For all except the highest frequency the model reproduces the observed gains for the forcing frequencies reasonably well, indicating that the structure of the model is capable of reproducing the observed pattern to a reasonable degree. The contribution of the highest frequency to the disturbance is small so that the impact of the discrepancy is low.

3.0 ISSUES IN VALIDATION

3.1 Decision-Making Model as a Description of Tracking Behaviour

From the findings described in Section 2.3 it is clearly feasible to model human performance of a tracking task as a series of discrete decisions and reproduce the general characteristics of the performance from the point of view of the widely used measure of performance RMSE. It is also clearly feasible to capture a significant fraction of the inter-individual variation using a parametric description of the decision-making procedure and the times taken to perform elements of the task. This does not constitute a demonstration of *predictive validity* since the same data set has been used to calibrate the model and test whether the model describes the observed phenomena. It supports *content validity* in that it demonstrates that the model can be parameterised to span the range of both human variation and task variation.

Neither of these findings demonstrates *construct validity* in that a discrete decision-making model is an appropriate representation of human tracking performance. It can be observed that when tracking behaviour is examined in detail it can be shown that a good description of control behaviour is that control inputs remain constant for periods which are interrupted by rapidly changing control inputs. In support of this contention, the application of the same model to the control of a Boeing 747-400 during descent to land is described in Belyavin *et al* (2009) and the model provides a reasonable reproduction of the tracking behaviour in these very different circumstances where control inputs are made less frequently than for a laboratory tracking task.

On the basis of these sets of evidence it is argued that the repeated discrete decision making model has construct and content validity as a model of human tracking performance but it has not been demonstrated that a particular model parameterisation has predictive validity. If it is accepted that the model is construct and content valid, it can be argued that a laboratory tracking task provides a continuous stream of nominally identical decisions that gives access to the investigation of models of a simple human decision making process in a systematic manner.

3.2 Nature of the Individual Decisions in the Tracking Model

For a laboratory tracking task, the individual decisions involved are likely *a priori* to be based on a simple set of observable parameters so that it is not difficult to construct a pattern that is likely to reflect that used by an experimental participant. A key element of the proposed tracking model is the way a decision is made as determined by the probability given by Equation (2). Following classical statistical decision theory, the natural way to model a pattern recognition decision is to define a criterion on the basis of costs of different types of error and to make the choice on the basis of whether the criterion is met or not. Representing the decision in a stochastic way has two effects: decisions that would not meet a strict criterion will still sometimes be made; the time at which a decision that does meet the criterion will be made is determined from a probability distribution.

The consequences of making a poor joystick move in a laboratory tracking task are relatively minor in that corrective action can always be taken later without serious compromise of overall performance. It is therefore unremarkable that inappropriate decisions can be permitted by the model without significant impact on the other measures of performance. The participant in a compensatory tracking task is acting as a negative feedback controller and so long as reasonable negative feedback is provided, overall performance is likely to be consistent with observation. It is therefore difficult to be certain as to whether such inappropriate decisions occur in practice. With a sufficiently large data set, it may be feasible to look for occurrences of supposed irrational responses, such as corrective action when none is warranted or control inputs opposite to the observed error, although the timing of perception relative to action is stochastic according to the model and this makes identification of specific events difficult. Although the

model should not be expected to reproduce these irrational responses specifically, it should reflect these behaviours in a stochastic manner if it is representing the constructs thought to give rise to the response.

From the point of view of decision timing, if the conditions that determine any criterion remain constant, the time taken to execute a decision using the probabilistic formulation will have a negative exponential distribution. This time distribution can be observed with human decisions in the laboratory and provides some indirect evidence that the model may be a plausible representation of this aspect of human decision making.

3.3 Validating Models of Decision Making in General

If the argument that a compensatory tracking task provides a stream of decisions that are handled by the human operator in the same way as any other stream of relatively low level pattern recognition decisions is accepted, some of the same models and principles should apply in other cases as well. In models of tactical conflict it is clearly feasible to define a restricted set of choices that a commander may make and then design a pattern recognition classifier to make the choice as each decision point is encountered. There are two lessons that can be drawn from the tracking model analysis. All military decision making has the objective of modifying the state of the world so that it is closer to a desired state and in that sense the management of the state of the world mirrors the activities of the negative feedback controller in the tracking task.

The analysis of the tracking model suggests that the use of a high level measure of performance such as mean RMSE alone does not reflect the variability of decision making between individuals and effort should be made to seek a range of observed streams of decisions so that any model may be tested in its ability to represent the variability. The model of the decision process in the tracking task indicates that the timing of decisions may itself be stochastic in any stream of decisions. This may be an element in any other stream of similar decisions and should be considered when validating other decision making models. While aggregate measures such as RMSE speak to the normative accuracy of a model, they obscure the plausible variability that is often desired in HBR and thus are insufficient for assessing a model’s validity. It is these unexpected excursions from normative behaviour that can result in surprise and confusion or confound systems predicated on rational, normative behaviour; incorporating such plausible variability in HBRs is expected to enrich training systems or lead to more robust systems so it is important to capture and validate these details adequately.

On this basis, assessing construct validity for models of low level decision making should include consideration of both how a choice of course of action is made and the mechanisms in the model that determine timing. Analysis of content validity should include consideration of how individual variability is represented as well as the range of external conditions. In considering predictive validity high level outcome measures can be used to provide an indication of whether a model is sound, but rigorous assessment of the timing elements of the model is likely to involve assessment of the decision pattern over time.

REFERENCES

[1] Baron, S., Kleinman, D. L., Levison, W. H. (1970) An optimal control model of human response Part II: prediction of human response in a complex task. Automatika. Vol 6. 1970. pp 371-383.

[2] Belyavin, A. & Cain, B. (2009). An example of validating models of continuous processes. Paper presented at the 18th Annual Conference on Behavior Representation In Modeling and Simulation (BRIMS), Sundance Resort, Utah. 30 Mar – 2 Apr. 8.

[3] Belyavin, A. J. & Farmer, E. W. (2006). Modeling the workload and performance of psychomotor tasks. Paper presented at the Behavioral Representation In Modeling and Simulation (BRIMS), Baltimore, MD

[4] Belyavin, A., Ryder, C. & Raabe, C. (2009). A discrete event controller as a model of pilot control behavior. Paper presented at AIAA Modeling and Simulation Technologies conference. 10 - 13 August 2009, Chicago, Illinois. AIAA 2009-5820.

[5] Belyavin, A. J. & Spencer, M. B. (2004) Modelling Performance and Alertness: The QinetiQ Approach. Aviation Space Environmental Medicine 2004; 75(3, Suppl.):A93–103.

[6] Campbell, G. E. & Bolton, A. E. (2005). HBR Validation: Integrating Lessons Learned From Multiple Academic Disciplines, Applied Communities, and the AMBR Project. In K. A. Gluck & R. W. Pew (Eds.), Modeling Human Behavior with Integrated Cognitive Architectures: Comparison, Evaluation, and Validation (pp. 365-395). Mahwah, New Jersey: Lawrence Erlbaum Associates.

[7] Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. Psychological Bulletin, 52, 281-302.

[8] Fisher, R. A. (1936). Uncertain Inference. Proceedings of the American Academy of Arts and Science, 71, 245-258.

[9] Foyle, D. C., Hooey, B. L., Byrne, M. D., Corker, K. M., Deutsch, S., Lebiere, C., *et al*. (2005). Human performance models of pilot behaviour. Paper presented at the Human Factors and Ergonomics Society 49th Annual Meeting, Santa Monica, CA. 1109-1113.

[10] Lotens, W., Allender, L., Armstrong, J., Belyavin, A., Cain, B., Castor, M., Gluck, K. A., et al. (2009). Human Behavior Representation In Constructive Simulation (No. RTO-TR-HFM128): North Atlantic Treaty Organization (NATO), Research and Technology Organization (RTO), Human Factors and Medicine (HFM).

[11] Nelder, J. A. & Mead R. (1965). A simplex method for function minimization Computer Journal, Vol. 7, pp. 308—313

[12] Pew, R. W. & Mavor, A. S. (Eds.). (1998). Modeling Human and Organizational Behavior. Applications to military simulations. Panel on Modeling Human Behavior and Command Decision Making: Representations for Military Simulations. Commission on Behavioral and Social Sciences and Education. Washington, DC.: National Research Council. National Academy Press

[13] Trochim, W. M. K. & Donnelly, J. P. (2007). The research methods knowledge base (3 ed.). Mason, OH, USA: Thomson

[14] Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology, 18, 459-482.