Design and Validation of a Model for a Human's Functional State and Performance*

Tibor Bosse[1], Fiemke Both[1], Mark Hoogendoorn[1], S. Waqar Jaffry[1], Rianne van Lambalgen[1], Rogier Oorburg[2], Alexei Sharpanskykh[1], Jan Treur[1], and Michael de Vos[2]
[1] *Vrije Universiteit Amsterdam, Department of Artificial Intelligence*
De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands
{tbosse, fboth, swjaffry, mhoogen, rm.van.lambalgen, sharp, treur}@few.vu.nl
[2] *Force Vision Lab, Barbara Strozzilaan 362a*
1083 HN Amsterdam, The Netherlands
{r.a.oorburg, m.de.vos}@forcevision.nl

Abstract. This paper presents a computational model of the dynamics of a human's functional state in relation to task performance and environment. It can be used in intelligent systems that support humans in demanding circumstances. The model takes task demand and situational aspects as input and determines internal factors such as the experienced pressure, exhaustion and motivation, and how they affect performance. Simulation experiments under different parameter settings pointed out that the model is able to produce realistic behavior of different types of personalities. Moreover, by a mathematical analysis the equilibria of the model have been determined, and by automated checking a number of expected properties of the model have been confirmed. In addition to the internal validation of the model, an experiment has been designed for the purpose of external validation addressing the estimation of the for the application relevant aspects of the human process. Output from the experiment like personality characteristics and performance quality has been used to perform estimation of the parameters of the model. By the parameter estimation a set of parameter values has been identified by which an adequate representation of a person's functional state when performing a task is achieved.

Keywords: functional state model; task performance; simulation; human experimentation; parameter estimation.

1. Introduction

For a human functioning in demanding circumstances, the quality of performance may be a critical factor. Examples of persons that have to work in such situations are air traffic controllers, incident and disaster managers, or persons that monitor surveillance videos. In cases like these any performed action which is badly chosen, or simply suboptimal can lead to dramatic consequences. Therefore it often is of utmost importance to maintain a high task performance quality, for example, in the sense that it is avoided that errors and biases occur in the decision making processes.

However, as it is known from literature such as ([9], [12]) working under high pressure in demanding circumstances often entails negative effects on the human's functional state, which in turn easily may affect performance quality. To be able to reduce such negative effects, this paper presents a first step in the development of an intelligent system to provide adequate support to a human that is functioning in demanding circumstances. In order for such a system to be effective, having an estimation of the human's functional state (and its implications, such as degraded performance quality) at any point in time is a crucial requirement. This requirement is the main focus of this paper. To fulfil the requirement, a computational model is proposed that can be used to estimate a human's functional state over time. The model, which was designed in dynamical system style, takes task demand and situational aspects such as noise levels as input and determines internal factors such as the experienced

* This paper is based on work presented at the Eighth IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2008) and the 12th International Conference on Principles of Practice in Multi-Agent Systems (PRIMA 2009).

pressure, exhaustion and motivation, and how they (may) affect task performance. The idea is that such a model can be used in a software environment supporting the human. For example, when it is estimated that the human's functional state may negatively affect task performance, measures can be taken such as alerting the person, involving a colleague, or involving an automated task support system. Moreover, the model can be used to regulate the task load and/or noise level to keep experienced pressure and exhaustion between certain limits thus avoiding negative effects on performance.

In this paper the role of the proposed model in the application addressed has also been validated against empirical data. The validation has been performed by taking a number of steps. First of all, an experiment with 31 human subjects has been conducted. Hereby, the subjects had to perform a task whereby they experience different amounts of workload. Each subject was given two conditions. Using the empirical data obtained from this experiment, parameter estimation techniques have been deployed to find appropriate parameter settings for the model to accurately describe the empirical behavior in one of the conditions. Thereafter, these settings have been used to predict the behavior of the subject in the other condition. Finally, properties that follow from the proposed work pressure model have been verified against the empirical data as well.

The experimental validation proposed in this paper and the use of parameter estimation techniques allows us to investigate whether the human data can be successfully predicted by the model, similar to [17]. In this case, the main purpose of tuning the model is its use within an ambient device that aims at providing human support. If the model (equipped with the estimated parameter values) can predict the human state at a given point in time, the ambient system can give adequate support by using that prediction. Note that it does not make an exclusive claim on whether the proposed model is the one and only model satisfying this application-directed validity criterion.

This paper is organized as follows. First, the work pressure model is explained in Section 2 and it's formulas in section 3. Thereafter, Section 4 gives a description of simulations performed with the model. The model has been verified by formal analysis and automated verification (Section 5). In addition, the Experimental Setup of the task that is used for validation is described in Section 6. The explanation of parameter techniques is given in Section 7 and validation results (from parameter tuning and logical verification) are shown in Section 8. Finally, Section 9 concludes the paper and discusses future work.

2. A Computational Model for Functional State

Cognitive workload is a common term in literature on humans working in demanding circumstances and it is known to be influenced by the cognitive demands of a task ([10]) and personality factors; e.g., ([7]) states that people with low cognitive abilities suffer more from high task load than people with high abilities. Cognitive workload is seen as one of many stressors which influence the state of a human (often referred to as stress, (e.g. [12]).

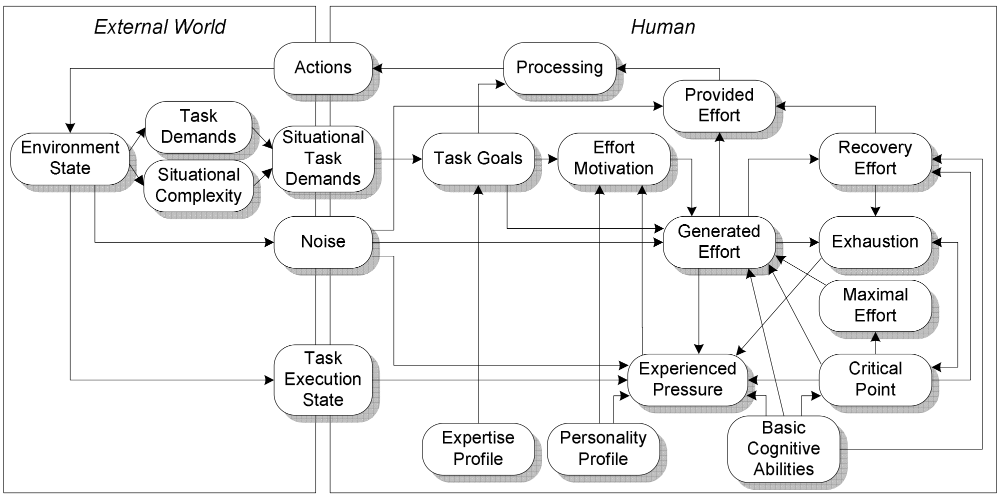


Figure 1. Model for a Human's Functional State.

Other stressors are time pressure, noise and heat, but also sleep deprivation and fatigue ([15]). These stressors are found to have a negative impact on human performance, like decision making ([15]), attention ([22]) and working memory ([9]). In the computational model for the Functional State (FS) of a human presented in this paper, cognitive workload is incorporated as a combination between the task and personal abilities. The FS model is also based on the cognitive energetic framework ([12]) that states that effort regulation depends on human recourses and that it determines human performance in dynamic conditions. Furthermore, the model is based on literature concerning exercise and sports ([11]). The idea of such literature is that a person's generated power can continue on a *critical power* level without becoming more exhausted. For the FS of a human, the exhaustion represents the mental exhaustion that occurs when too much effort contribution results in a depletion of resources.

The FS of a human (e.g., an operator in a control room) represents the dynamical state in which the person is situated. In the model presented in Figure 1, this FS is defined by a combination of exhaustion, motivation and experienced pressure, but also the amount of generated and provided effort. The FS is determined by factors from the external world (task demands and environment state) and by personal factors (experience, cognitive abilities and personality profile). In addition, the FS model shows the relation of this state to the determination of actions. However, this relation falls outside the scope of this paper and we will mainly focus on the human's functional state.

In this model, the concepts *generated effort* (the amount of effort the human generates to perform a task) and *critical point* (the amount of effort someone can generate without becoming exhausted) are adapted from Hill's concepts generated and critical power. The critical point is dynamic in time and is reduced by the amount of *exhaustion*. When the human's exhaustion is 0, the critical point is equal to the *basic cognitive ability* (which is personal to the individual).

Generated effort is influenced by the amount of effort the human wants to contribute (*effort motivation*), the amount of effort the human has to contribute (*task level*) and the amount of effort the human is able to contribute (critical point and *maximal effort*). When generated effort is below critical point, the human is able to do some recovery (*recovery effort*). This recovery effort decreases exhaustion. On the other hand, when generated effort is above critical point, the human will become more exhausted. Recovery effort and the effort the human has to contribute to the noise in the environment are extracted from the generated effort and eventually determine the effort that can effectively be contributed to the task (*provided effort*).

The motivation of a person is proportional to the task level, but depends negatively on the *experienced pressure*. Underload is represented by a level of experienced pressure below *optimal experienced pressure* (embedded in *personality profile*) and overload is represented by a level of experienced pressure above optimal experienced pressure. The impact that underload and overload have on motivation are dependent on low and high pressure sensitivities (in personality profile). Experienced pressure itself is increased by high effort (generated effort above critical point), noise, exhaustion and a performance quality below a personal optimal performance. The factors that have a negative influence on experienced pressure are low effort (generated effort below critical point) and a performance quality above optimal performance.

3. The Detailed Model

This section explains the details of the model. The implemented relations between different concepts are based on earlier findings in literature on stress and operator functional state.

There are three ***temporal relations***.

$$E(t+\Delta t) = E(t) + Pos(\eta\cdot(GE(t)-CP(t))\cdot\Delta t) - \pi\cdot RE(t)\cdot\Delta t \qquad [1]$$

In theories of cognitive energetics, the effort that is generated to perform a task can deplete the available resources, which causes an increase in mental exhaustion ([12]). Here, a depletion in resources is represented by the contribution of effort (GE) above the basic critical point (CP), similar to the Critical Power theory in physical exercise studies ([11]). In this model, Exhaustion (E) builds up or reduces over time; when the generated effort (GE) is above the critical point (CP), exhaustion increases. Exhaustion decreases depending on the amount of Recovery Effort, which adds to the amount of available resources. Parameters η and π determine the exact influence of Generated Effort and the amount of recovery on exhaustion..The function *Pos(x)* in this and other formulas is defined as the maximum of x and 0.

$$GE(t+\Delta t) = GE(t) + \beta \cdot (CCE(t)-GE(t)) \cdot \Delta t \qquad [2]$$

Formula [2] represents calculation of the amount of Generated Effort (GE). One portion of the GE at the next point in time is determined by the current contribution (*CCE*; see formulas [4] and [5] below). The other portion represents a contribution of the previous GE to allow for a smooth transition from the GE at one time point to the next. Here β is a flexibility parameter; it determines how much of the new generated effort is affected by the current contribution.

$$EP(t+\Delta t) = EP(t) + [\mu 1 \cdot Pos(EPC(t) \cdot (1-EP(t)) - \mu 2 \cdot Neg(EPC(t) \cdot EP(t))] \cdot \Delta t \qquad [3]$$

Finally, experienced pressure (EP) is determined by a temporal relation. This temporal nature of EP (similar to a person's stress level) explains why the EP does not immediately change with a shift in the current stress contribution (in this case the Experienced Pressure Change, EPC; see [6] below), but fades away or builds up.

In addition to the temporal relations, the model includes a number of ***instantaneous relations***.

$$CCE(t) = NE(t) + \varepsilon \cdot \frac{MaximalEffort(t)}{TopMaximalEffort} \cdot EM(t) \cdot (w1 \cdot (CP(t) - NE(t)) + w2 \cdot TL(t) + w3 \cdot ME(t)) \qquad [4]$$

$$TME = LCP + \zeta \cdot (BCA-LCP) \qquad [5]$$

The current contribution to the generated effort of a person to a specific task is influenced by what needs to be done (Task Level, TL), what can be done (Critical Point, CP; Maximal Effort, ME) and what a person wants to do (Effort Motivation, EM). An additional Effort (NE) is needed to deal with noise (i.e. too much sound or too much light). *TME* represents top maximal effort and is calculated by the *LCP* (lowest critical point) and *BCA* (basic cognitive abilities).

$$EPC(t) = ES \cdot E(t) - PS \cdot (PQ(t)-PN) + HES \cdot Pos((GE(t) - CP(t)) / (BCA-LCA))$$
$$- LES \cdot Pos((CP(t) - GE(t))/BCA-LCA)) + NS \ (NE(t) / MaxNE(t)) \qquad [6]$$

As stated before, the Experienced Pressure Change (EPC) can be seen as the stress contribution at a specific point in time and is used in [3] to calculate the Experienced Pressure. From the literature it is apparent that important influences on stress are fatigue, workload and noise factors (such as extreme heat, annoying sounds etc) ([8]). In addition, research has pointed out a relationship between a (perceived) suboptimal performance and stress ([5]).

In this formula, the influences mentioned above are implemented and mediated by personal sensitivity levels (S), taking the personal differences in stress into account. The first term is the influence of exhaustion; the second is the influence of the difference between current performance quality (*PQ*) and a personal performance norm (*PN*). The third and fourth terms represent the influence of generating effort above and below the critical point (the pressure of over- and underload in terms of effort). Finally, the last term represents the influence of Noise. The last three terms are first normalized by dividing them by their maximal value.

$$EM(t) = EPI(t) \cdot \left(\frac{1+\frac{1}{\gamma}}{1+\gamma \cdot e^{-\phi \cdot TL(t)}} - \frac{1}{\gamma}\right) \cdot \frac{1}{Exp(t)} \qquad [7]$$

Effort motivation is dependent on the level of Experienced Pressure a person experiences and how well a person can handle that experienced pressure. It is calculated using the current task level and the influence of experienced pressure (*EPI*), which is calculated according to [8]. Parameters φ and γ determine the shape of the (sigmoid) function.

$$EPI(t) = 1 - (HPS \cdot Pos(EP(t) - OEP) + LPS \cdot Pos(OEP - EP(t))) \quad [8]$$

The effect of experienced pressure (*EPI*) on the effort motivation is determined by the distance between the current experienced pressure and the optimal experienced pressure (*OEP*) multiplied by a personal sensitivity for high (*HPS*) and low pressure (LPS); over- and underload in terms of experienced pressure.

The formulas [9], [10], [11] and [12] consider a human's cognitive effort regulation. Since the cognitive effort regulation as outlined in the cognitive energetic framework (hockey, 1997) is similar to that of physical effort regulation, the formulas are adapted from sports and exercise literature (Hill, 1993).

$$PE(t) = GE(t) - RE(t) - NE(t) \quad [9]$$

Only part of Genarated Effort is effectively provided to the task. Generated effort is the sum of provided effort (*PE*; the net effort for task execution), recovery effort (needed for decreasing exhaustion) and noise effort (effort required to deal with noise). Therefore, to calculate the provided effort, recovery and noise effort are subtracted from generated effort.

$$RE(t) = Pos(\alpha \cdot (CP(t) - GE(t))) \cdot GE(t) \cdot (BCA - CP(t)) / BCA \quad [10]$$

When generated effort is below critical point, effort can be used to lower exhaustion. The amount of effort increases when generated effort is further from critical point (first line of the formula), and increases when the critical point is further from basic cognitive abilities (which is when the human is somewhat exhausted, second line of the formula). Parameter α represents the efficiency of recovery effort: a small percentage is lost due to inefficiency of the process.

$$CP(t) = LCP + (1 - E(t)) \cdot (BCA - LCP) \quad [11]$$

Critical point is equal to basic cognitive abilities, unless there is some level of exhaustion: then critical point decreases proportionally.

$$ME(t) = LCP + \zeta \cdot (CP(t) - LCP) \quad [12]$$

Maximal effort is ζ times critical point or equal to critical point when critical point is equal to the lowest critical point. This entails that once critical point has dropped to a low value, it is not possible anymore to generate effort that causes more exhaustion.

$$PQ(t) = PE(t) / TL(t) \quad [13]$$

Eventually, the quality of performance depends on the relation between provided effort and task level. If provided effort is lower as compared to the task level, performance quality is below 1, otherwise above 1. For example, if the task level requires an effort of 200 and a provided effort of 100 is given, the performance quality would be 0.5

4. Simulation Results

Using the formulas to determine the FS, some interesting patterns on human performance have been explored. As stated in Section 2, multiple personality factors are involved in determining the FS. Some typical patterns can be found with variation of these specific personality factors. Due to the excessive number of possible combinations, this paper shows two of the many possible persons with a different personality profile. The Matlab code of the model can be found at

http://www.few.vu.nl/~fboth/OFS/OFS.m.

The duration of the scenario for the two personalities (displayed in Figures 2 and 3) is 200 time points. The task level was 290 for the first half and 500 for the second half. Table 1 outlines the values of the different personality factors used for person 1 (perfectionist, likes high pressure) and person 2 (non-perfectionist, likes low pressure).

The results of the simulations are shown in Figures 2 and 3. Figure 2 displays the variables with respect to the task: maximal effort (ME), critical point (CP), generated effort (GE) and task level (TL). Figure 3 displays more qualitative variables: performance quality (PQ), effort motivation (EM), experienced pressure (EP) and exhaustion (EX).

Table 1. Personality profiles.

Person 1 (high pq)	*Person 2 (low pq)*
Experienced pressure change parameters	Experienced pressure change parameters
PN=1.1;	PN=0.3;
PS=0.5;	PS=0.3;
ES=0.2;	ES=0.7;
HES=0.2;	HES=0.7;
LES=0.7;	LES=0.3;
NS=0.2;	NS=0.6;
Experienced pressure influence parameters	Experienced pressure influence parameters
LPS=1.2;	LPS=0.8;
HPS=0.8;	HPS=1.2;
OEP=0.8;	OEP=0.3;

A comparison of person 1 and 2 for the task variables shows that when task level is 290, both persons have a more or less equal amount of generated effort. When task level increases to 500, person 2 generates more effort than person 1, but this effort reduces fast, which allows the critical point to recover. In contrast, the generated effort of person 1 decreases more slowly, which induces a reduction of the critical point. The critical point (and the generated effort) of person 1 therefore end at a higher level than those of person 2.

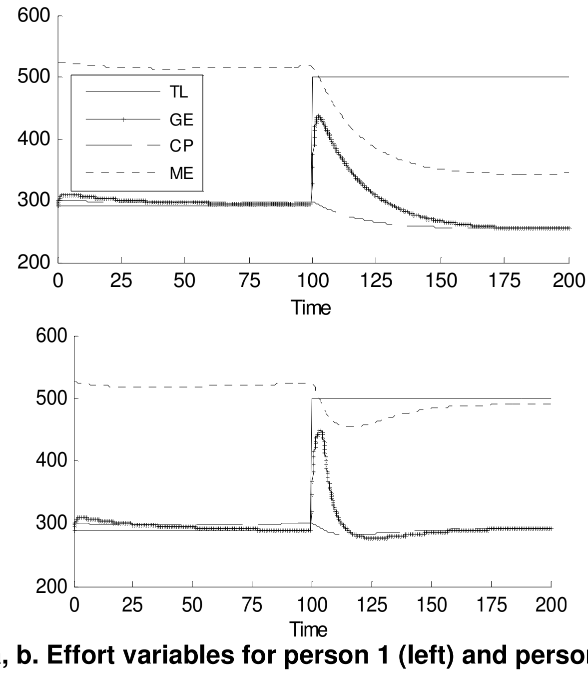


Figure 2a, b. Effort variables for person 1 (left) and person 2 (right).

Figure 3 shows the mood variables; the main difference at the start of the simulation is experienced pressure, as the optimal experienced pressure of person 1 is much higher (0.8) than the optimal experienced pressure of person 2 (0.3). Exhaustion levels of both persons remain low at the beginning. However, when the task level is increased to 500, person 1's exhaustion increases, due to the large amount of time at which person 1's generated effort is above the critical point (see Figure 2a). The motivation of person 2 decreases when task level is increased to 500, which is caused by the influence of experienced pressure being much higher than optimal experienced pressure. Person 1 also has a high experienced pressure; however, this value is close to the optimal experienced pressure so the negative effect on motivation is less.

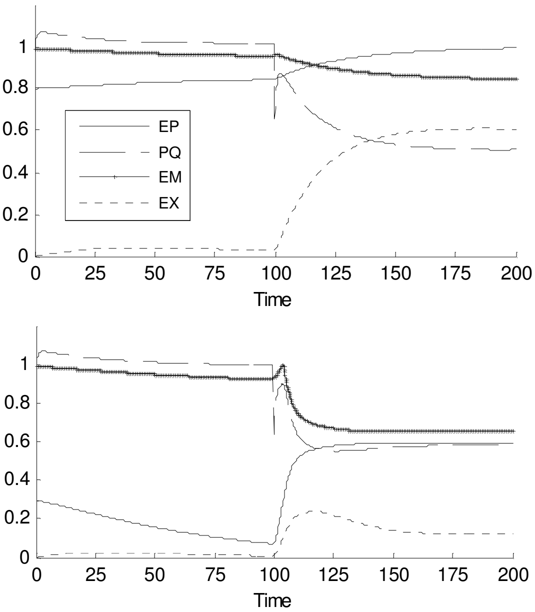


Figure 3a, b. Variables for person 1 (left) and person 2 (right).

5. Formal Analysis

In this section, the behavior of the presented model is formally analyzed, using two different approaches. In Section 5.1, the stable (equilibrium) states of the model are determined via a mathematical analysis. Next, in Section 5.2, additional dynamic properties of the model are checked via logical verification.

5.1 Mathematical Analysis

By a mathematical formal analysis the equilibria can be determined, i.e., the values for the variables for which no change occurs. Note that to this end the exogenous variables *TaskLevel* and *NoiseEffort* are assumed to have a constant value, denoted by *TL*, resp. *NE*. To obtain possible equilibrium values for the other variables, first the model is described in a differential equation form:

$$dExhaustion(t)/dt = Pos(\eta(GeneratedEffort(t) - CriticalPoint(t))) - \pi RecoveryEffort(t)$$
$$dGeneratedEffort(t)/dt = \beta(CurrentContributionEffort(t) - GeneratedEffort(t))$$

$$dExperiencedPressure(t)/dt =$$
$$\mu 1 Pos(ExperiencedPressureChange(t)).(1 - ExperiencedPressure(t)) -$$
$$\mu 2 Neg(ExperiencedPressureChange(t)).ExperiencedPressure(t)$$

Next, the equations are identified describing

$dExhaustion(t)/dt = 0$
$dGeneratedEffort(t)/dt = 0$
$dExperiencedPressure(t)/dt = 0$

Thus the following equations are found:

$GE = CP$ and $RE = 0$
$CCE = GE$
[$EP > 0$ and $EP < 1$ and $EPC = 0$] or [$EP = 0$ and $EPC \leq 0$] or [$EP=1$ and $EPC \geq 0$]

Here *GE* is the (equilibrium) value for *GeneratedEffort*, *CP* for *CriticalPoint*, *RE* for *RecoveryEffort*, *CCE* for *CurrentContributionEffort*, *EP* for *ExperiencedPressure* *EPC* for *ExperiencedPressureChange*, and similar notations are used for the other equilibrium values. Here a first set of conclusions is that an equilibrium can only occur when the generated effort equals the critical point, no recovery takes place, and the current contribution effort is equal to generated effort. Elaborating the equations further can be done by distinguishing cases (for $EPC=0$, $EP=0$ and $EP=1$) according to the last formula shown above. The first case considered is $EPC = 0$. For this case it can be derived that the values for the equilibria can be calculated by the following formulae:

$$E = [\, PeS\cdot(\, BCA - NE - PN\cdot TL)/TL - NoS\cdot NE/MNE \,] / [\, ExS + PeS\cdot (BCA\text{-}LCP) \,]$$

$$GE = CP = LCP + (1\text{-}E)\cdot(BCA - LCP)$$

$$PE = GE - NE$$

$$EM = (CP - NE)\cdot (LCP + \zeta(BCA - LCP)) \;/\; \varepsilon\cdot(LCP + \zeta(CP - LCP))\cdot(w1\ (CP - NE) + w2\ TL + w3\ (LCP + \zeta(CP - LCP)))$$

$$EP = OEP + (1 - EM \cdot Expect / logist(TL))/ HpS$$

or

$$EP = OEP - (1 - EM \cdot Expect / logist(TL))/ LpS$$

Similarly, formulae can be derived for the other two cases in the formula shown above, namely $EP = 0$ or $EP = 1$.

5.2 Logical Verification

This section addresses analysis of the model by verification of dynamic properties. Following ([2]), the dynamics of a simulation model can be studied by specifying certain dynamic statements (temporal logical expressions), that are (or are not) expected to hold and automatically verifying these statements against simulation traces. The purpose of this type of verification is to check whether the simulation model behaves as it should. A typical example of a property that may be checked is whether no unexpected situations occur, such as a variable running out of its bounds (e.g., Exhaustion < 0, or GeneratedEffort > MaximalEffort). By running a large number of simulations and verifying such properties against the resulting simulation traces, the modeler can easily locate sources of errors.

For the model of the human's functional state, a number of such dynamic properties have been formalized in the language TTL ([2]). This predicate logical language supports formal specification and analysis of dynamic properties, covering both qualitative and quantitative aspects. TTL is built on atoms referring to *states* of the world, *time points* and *traces*, i.e. trajectories of states over time. In addition, *dynamic properties* are temporal statements that can be formulated with respect to traces based on the state ontology Ont in the following manner. Given a trace γ over state ontology Ont, the state in γ at time point t is denoted by state(γ, t). These states can be related to state properties via the formally defined satisfaction relation denoted by the infix predicate |=, comparable to the Holds-predicate in the Situation Calculus: state(γ, t) |= p denotes that state property p holds in trace γ at time t. Based on these statements, dynamic properties can be formulated in a formal manner in a sorted first-order predicate logic, using quantifiers over time and traces and the usual first-order logical connectives such as ¬, ∧, ∨, ⇒, ∀, ∃. A special software environment has been developed for TTL, featuring both a Property Editor for building and editing TTL properties and a Checking Tool that enables formal verification of such properties against a set of (simulated or empirical) traces.

Various dynamic properties for the model have been formalized in TTL. Below, a number of them are introduced, both in semi-formal and in informal notation (note that they are all defined for a particular trace and time interval between tb and te):

P1a - Stability of Variable v
For all time points t1 and t2 between tb and te in trace γ
if at t1 the value of v is x1
then at t2 the value of v is between x - α and x + α
(where α is a constant).

P1a(γ:TRACE, tb, te:TIME, v:VAR) ≡
∀t1,t2:TIME x1,x2:REAL
state(γ, t1) |= has_value(v, x1) &
state(γ, t2) |= has_value(v, x2) &
tb ≤ t1 ≤ te & tb ≤ t2 ≤ te ⇒ x1-α ≤ x2 ≤ x1+α

This property can be used to verify in which situations a certain variable does not fluctuate much. It has been found, for example, that for the two traces shown in Section 4 and for α=1.0, the Critical Point remains stable between time point 450 and 500. In other words, checking P1a(trace, 450, 500, CP) was successful, where trace is one of the traces of Section 4, i.e. either trace1 (corresponding to person 1, i.e., the perfectionist) or trace2 (corresponding to person 2, the non-perfectionist).

P1b - Stability of Variable v around Value x
For all time points t between tb and te in trace γ
the value of v is between x - α and x + α (where α is a constant).

P1b(γ:TRACE, tb, te:TIME, v:VAR, x:REAL) ≡
∀t:TIME y:REAL
state(γ, t) |= has_value(v, y) &
tb ≤ t ≤ te
⇒ x-α ≤ y ≤ x+α

As a variant of P1a, property P1b can be used to check whether a variable stays around a certain value (which has to be given as an extra argument). For example, property P1b(trace, 1, 250, TL, 290) was true for the traces of Section 4 (which was expected, since the Task Level was set to 290 by hand in that period).

P2 - Monotonic Decrease of Variable v
For all time points t1 and t2 between tb and te in trace γ
if at t1 the value of v is x1 and at t2 the value of v is x2
and t1 < t2, then x1 ≥ x2

P2(γ:TRACE, tb, te:TIME, v:VAR) ≡
∀t1,t2:TIME x1,x2:REAL
state(γ, t1) |= has_value(v, x1) &
state(γ, t2) |= has_value(v, x2) &
tb ≤ t1 ≤ te & tb ≤ t2 ≤ te & t1 < t2 ⇒ x1 ≥ x2[†]

Property P2 can be used to check whether a variable decreases monotonically over a certain interval. For example, the Experienced Pressure turned out to decrease over the first half of the trace of person 2 (i.e., property P2(trace2, 1, 250, EP) succeeded).

[†] Note that a stronger variant of this (and similar) properties can be created by replacing ≥ by >.

P3 - Variable v between Boundaries
For all time points t between tb and te in trace γ
if at t the value of v is x, then min < x < max.

P3(γ:TRACE, tb, te:TIME, v:VAR, max, min:REAL) ≡
$\forall$t:TIME x:REAL
state(γ, t) |= has_value(v, x) & tb ≤ t ≤ te ⇒ min ≤ x ≤ max

This property can be used to check whether a variable stays between certain boundaries. For example, the Exhaustion should never become lower than 0 or higher than 1, which indeed turned out to be the case for the generated traces (i.e., property P3(trace, 1, 500, EX) succeeded).

P4 - Variable v1 above v2
For all time points t between tb and te in trace γ
if at t the value of v1 is x1 and the value of v2 is x2
then x1 ≥ x2.

P4(γ:TRACE, tb, te:TIME, v1, v2:VAR) ≡
$\forall$t:TIME x1,x2:REAL
state(γ, t) |= has_value(v1, x1) &
state(γ, t) |= has_value(v2, x2) &
tb ≤ t ≤ te ⇒ x1 ≥ x2

Property P4 can be used to check whether a variable value stays above (or below) another variable value during a specified interval. For example, if the model has been set up correctly, the Generated Effort should never exceed the Maximal Effort. Also this check turned out to succeed for the generated traces (i.e., property P4(trace, 1, 500, ME, GE) succeeded).

P5a - Variable v Approaches Value x
For all time points t1 and t2 between tb and te in trace γ
if at t1 the value of v is x1and at t2 the value of v is x2 and t1 < t2 then |x-x1| ≥ |x-x2|.

P5a(γ:TRACE, tb, te:TIME, v:VAR, x:REAL) ≡
$\forall$t1,t2:TIME x1,x2:REAL
state(γ, t1) |= has_value(v, x1) &
state(γ, t2) |= has_value(v, x2) &
tb ≤ t1 ≤ te & tb ≤ t2 ≤ te & t1 < t2 ⇒ |x-x1| ≥ |x-x2|

By checking property P5a, one can check whether a variable eventually approaches a certain (given) value. This can be used, among others, to check in which situations the system ends up in one of the equilibria calculated in Section 5.1. For example, for the generated traces, the case where the Experienced Pressure Change = 0 (as described in Section 5.1) can be confirmed for the end of the simulation: property P5a(trace, 475, 500, EPC, 0.0) succeeded for those traces.

P5b - Variable v Approaches Value x with Speed s
For all time points t1 and t2 between tb and te in trace γ
if at t1 the value of v is x1
and at t2 the value of v is x2
and t2 = t1+1
then s * |x-x1| ≥ |x-x2| (where s is a constant).

P5b(γ:TRACE, tb, te:TIME, v:VAR, x:REAL) ≡
$\forall$t1,t2:TIME x1,x2:REAL
state(γ, t1) |= has_value(v, x1) &
state(γ, t2) |= has_value(v, x2) &
tb ≤ t1 ≤ te & tb ≤ t2 ≤ te & t2 = t1+1⇒ |x-x1| * s ≥ |x-x2|

Property P5b is a refinement of P5a. It can be used to check not only whether a variable approaches some value, but also to determine the speed s with which this happens (where $0 < s < 1$, and a high s denotes a slow speed). For the case of the EPC described above, this s turned out to be 0.96904 for trace1 and 0.99997 for trace2[‡].

P6 - Higher variable v1 (in m1) leads to lower variable v2
If in trace γ1 at tb the value of v1 is x1
and in trace γ2 at tb the value of v1 is x2
and in trace γ1 at te the value of v2 is y1
and in trace γ2 at te the value of v2 is y2
then if x1 ≥ x2, y1 ≤ y2 and if x1 ≤ x2, y1 ≥ y2

P6(γ1, γ2:TRACE, tb, te:TIME, v1, v2:VAR) ≡
∀x1,x2,y1,y2:REAL
state(γ1, tb) |= has_value(v1, x1) &
state(γ2, tb) |= has_value(v1, x2) &
state(γ1, te) |= has_value(v2, y1) &
state(γ2, te) |= has_value(v2, y2) &
⇒ [x1≥x2 ⇒ y1≤y2] & [x1≤x2 ⇒ y1≥y2]

This property can be used to determine how variables influence each other. A specific variant of the property is P6(trace1, trace2, 1, 500, OEP, CP), which checks whether persons with a high Optimal Experienced Pressure (i.e., perfectionists, those that try to work much under stress) eventually end up with a lower Critical Point than those with a lower OEP, given the same circumstances. This property was satisfied for all of the generated traces (i.e., for the traces shown in Section 4, but also for a large number of other traces under various parameter settings).

6. Experimental Setup

In this section the experiment is described that is used for validation of the functional state model. Using the participant data from the experiment, two parameter tuning techniques are applied to find the appropriate parameters for the model. In Section 6.1 a brief overview of the participants is given. The main part of the experiment is a game that combines a motor task and a cognitive task. Section 6.2 gives a description of these tasks. In Section 6.3, the procedure of the experiment is explained. Finally, Section 6.4 describes how data from the experiment can be used as input for the work pressure model portrayed in Section 2.

6.1 Participants

In this study 31 people participated (18 males, 13 females), of which 25 students at the VU University Amsterdam. Participants ranged in age from 17 to 57 years with a mean age of 26 years. The experiment took approximately 1 hour for which participants received a voucher of 10 euro. In addition, there was a voucher of 100 euro for the one with the best score on the Experimental Task.

6.2 Experimental Task

In the experiment the main task was a simple shooting game where the goal was to get as many points as possible. A screenshot of this game is displayed in Figure 4. The object at the bottom of the screen represented the participant's weapon. Other objects (friends and enemies in the shape of a small circle) were falling down in different locations at different speeds. The purpose of the game was to shoot the enemies before they hit the ground. Participants could shoot a missile by a mouse click at a specific location; the missile would then move from the weapon to that location and explode exactly at the location of the mouse click. The speed with which the missile reached this location was 79.6 pixels per second. When an object was within a radius of 50 pixels of the explosion, the object was destroyed.

[‡] Note that lower numbers can be obtained by using t+10 in property P5b instead of t+1.

The number of points the participants received for hitting an enemy was proportional to the proximity of the explosion. When participants shot a friend and when an enemy reached the bottom of the screen, they lost -10000 points. When a friendly object reached the bottom of the screen participants received 1000 points.

In addition, next to the objects, a calculation was written on the screen. A correct calculation indicated that the object was friendly and should not be shot. An incorrect calculation indicated that the object was an enemy and should be shot before it reached the bottom of the screen.

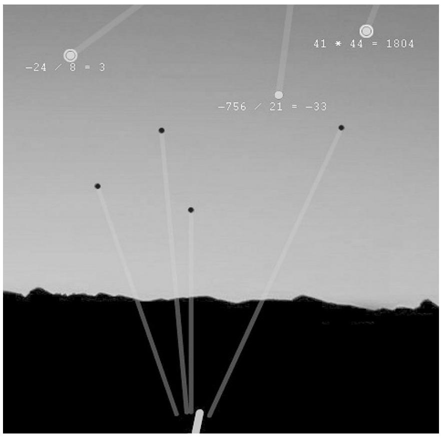


Figure 4. Screenshot of the Experimental Task.

6.3 Procedure

For the experiment a 2-factor within subjects design was used. Two different conditions within each participant were tested. The first condition (Condition 1) was similar to the scenario described in the Simulation section of this paper (Section 4). The scenario started with a low task level and continued with a high task level. This order was reversed for the second condition (Condition 2), which started with a high task level and continued with a low task level. The first condition (low-high) started with 1 object present per 10.00 seconds. After 7.5 minutes the number of objects which were presented per second (1 object per 2.25 seconds) was increased. Condition 2 (high-low) started with a high task level and after 7.5 minutes the number of objects which were presented per second was decreased. Both conditions took 15 minutes in total. Condition was counterbalanced over participants to correct for a possible order effect, such that participants with an odd number started with condition 2 and participants with an even number started with condition 1.

Participants started the experiment with filling out a personality questionnaire with questions from the NEO-PI-R and the NEO-FFI ([4]); with these questions some aspects of each participant's personality were measured, to serve as input for the personality profile of the functional state model. Neuroticism and extraversion were measured with the NEO-FFI. With the NEO-PI-R vulnerability (part of neuroticism) and ambition (part of conscientiousness) were measured.

After the questionnaire, participants performed three small tests; each consisted of 30 trials which were equal between participants. These tests served as input for model validation. The instructions for each test were shown on the screen. Participants started with a simple choice Reaction Time test (choice-RT), where a square was presented either left or right from a fixation cross at the centre of the screen. Participants had to react with either the left arrow (when the square was presented left) or the right arrow (when the square was presented right). The second test was a task where calculations were presented similar to the calculations in the calculation task of the experiment. Like in the experiment, participants had to choose whether the calculation was correct (left arrow) or incorrect (right arrow). The third small test (mouse-RT) was another Reaction Time task; here a circular target was presented somewhere on the screen. Participants had to react quickly and precisely by clicking with the mouse as close as possible to the centre.

After the three small tasks, participants practiced during 3 minutes with the experiment-game. The goal of the practice task was familiarize with the shooting and calculation tasks in the game. After practice the participants started the experiment-game with either condition 1 or condition 2, which both took 15 minutes.

6.4 From Experiment Data to Work Pressure Model

In order to validate the model, data from the experiment was used to calculate the values of several concepts of the functional state model, namely personality profile, basic cognitive abilities and expertise profile.

The *personality profile* was calculated using the personality characteristics obtained via the personality questionnaire. Ambition contributes to performance norm, as people who are more ambitious often aim to have high performance ([24]).

$$PerformanceNorm = B \cdot Ambition + (1 - 5 \cdot B) \quad [14]$$

Neuroticism contributes to performance sensitivity, as people who are more neurotic are more likely to feel pressure when their performance is not as high as they aim for (Matthews & Dairy, 1998; Rose et al., 2002)

$$PerformanceSensitivity = C \cdot Neuroticism + (0.5 - 5 \cdot C) \quad [15]$$

From (Rose et al., 2002) it was found that people who are more extravert are more easily distracted and perform less well in circumstances of underload than people who are less extravert. On the other hand, when people score high on vulnerability they are likely to perform bad in situations of overload. These two personality characteristics contribute to optimal experienced pressure.

$$OptimalExperiencedPressure = 0.1 \cdot Extraversion \cdot D - 0.1 \cdot Vulnerability \cdot (1-D) + (1-D) \quad [16]$$

$$LowPressureSensitivity = 0 \text{ when } OEP <= 0.33,\ LPS = 1 \text{ when } OEP > 0.33 \quad [17]$$

$$HighPressureSensitivity = 1 \text{ when } OEP < 0.67,\ HPS = 0 \text{ when } OEP > 0.67 \quad [18]$$

In the formulas above, B, C and D are parameters which need to be estimated.

The remaining four concepts in the personality profile have the same value for all participants: the sensitivity for noise (*noise sensitivity*; 0), the sensitivity for generated effort below critical point (*low effort sensitivity*, 0.3), the sensitivity for a generated effort above critical point (*high effort sensitivity*, 0.1) and the sensitivity for a high exhaustion (*exhaustion sensitivity*, 0.3). Note that all personal sensitivities affect the experienced pressure, except for the Optimal Experienced Pressure, which determines the influence of experienced pressure on motivation.

Basic cognitive abilities are measured using RT and accuracy data from the choice-RT task, as well as the calculation task. Cognitive ability (often referred to as general cognitive ability 'g') strongly depends on processing speed, but also psychometric tests like calculations are often used to determine 'g' ([21]). From the calculations task, the accuracy and the RT is taken:

$$Calc = \%\ correct \cdot minCalcRT / CalcRT \quad [19]$$

From the choice RT the reaction time is taken without the error trials:

$$Choice = minChoiceRT / ChoiceRT \quad [20]$$

Basic cognitive abilities follow from these two formulas:

$$BCA = (W3 \cdot Calc + W4 \cdot Choice) \cdot Z \quad [21]$$

W3, W4 and Z are parameters which need to be estimated, with the restriction that W3 and W4 add up to 1. Z is a parameter necessary to get BCA in the correct range (relative to task level and generated effort since the basic cognitive abilities is used to define the critical point in the model).

The *expertise profile* is derived from how much experience someone has with the task itself. Since the task was new to all participants, the expertise profile was calculated by using the RT and accuracy of the calculation task (since in the experiment calculations had to be solved) and those of the mouse-RT (since in the experiment, missiles had to be fired with the mouse as accurately as possible). In the formulas, W1 and W2 need to be estimated with the constraint that they add up to 1.

Calc = %correct· minCalcRT/CalcRT [22]

Mouse-RT = %dist_to_centre· minMouse-RT/Mouse-RT [23]

Exp = W1·Calc + W2·Mouse-RT [24]

Furthermore, from the experiment data the situational demands can be calculated. Although the scenarios were the same for all participants, the calculated task level could differ due to the performance quality. Therefore, Situational Demands were calculated per time step per participant. According to the model, situational demands (from the world) and the expertise profile together contribute to task level.

TaskLevel = ((1 + 0.5) – Exp)·SitD [25]

Finally, a good measure of functional state is the performance quality. In the experiment, performance quality was measured in terms of efficiency and effectivity. Efficiency represented the number of missiles necessary to shoot an enemy. Effectivity was dependent on how close to the object the missile exploded (explosion fraction) and whether an enemy or friend was shot.

When an enemy was hit, the effectivity was calculated from the explosion fraction.

Effectivity = (1+ explosion_fraction)/2.0 [26]

Effectivity was 0 when a friend was shot or an enemy landed. When a friend landed, effectivity was 1. Finally, the task execution state (Objective Task Execution State) was calculated using both effectivity and efficiency:

ObjTES = (0.25·efficiency + 0.75·effectiveness)·2 [27]

7 Estimation of Parameters

The process of determining unknown parameters of a mathematical model from (noisy) input-output data is often called *parameter estimation*. This section presents the results of parameter estimation for the work pressure model using two essentially different methods: Section 7.1 considers a gradient-based approach, whereas an approach based on probabilistic search is considered in Section 7.2. Section 7.3 presents the results of the estimation.

7.1 Gradient-based Parameter Estimation

One of the methods used commonly for parameter estimation is the least squares method. In this method an estimator of θ is chosen that minimizes the sum of the squares of the error:

$$E(\theta) = (z - y(\theta))^T \cdot (z - y(\theta)) \qquad (1)$$

here z is a vector of the measurements, y is a nonlinear vector valued function, θ is a vector of the unknown parameters. The minimization of E w.r.t. θ gives:

$$\left[\frac{\partial y(\hat{\theta}_{LS})}{\partial \theta}\right]^T \cdot (z - y(\hat{\theta}_{LS})) = 0 \qquad (2)$$

Here $\hat{\theta}_{LS}$ is the least-square estimator of θ.

Then, an iterative procedure based on the expansion (linearization) of $y(\theta)$ by Taylor's series is used to solve (24) approximately.

It is known that the least squares method lacks stability and precision, when measurement data contain noise (Sorenson, 1980). This was also observed when the method was applied to the work pressure model. The convergence of the parameter estimation was highly dependent on the choice of the initial parameter values and often was not achieved in the experiments performed.

To handle the problem of stability and the measurement noise, a method based on the maximum likelihood principle (of which the least squares method is a special case) has been used. In line with this principle a likelihood function of the measurement data and the unknown parameters is defined. This function is essentially the probability density function of the measurement data given the parameter values $p(z|\theta)$. In this case it is assumed that the measurements contain noise which is zero-mean and has a Gaussian distribution. The measurement data are represented by the random, normally distributed variable z. Such an assumption is often made for dynamic systems in many areas (Sorenson, 1980). The parameter vector $\hat{\theta}_{ML}$, which makes the likelihood function most probable to obtain the measurements z (i.e., which maximizes the likelihood function) is called the maximum likelihood estimate; it is obtained by minimizing the error function ([25]):

$$E(\theta) = \frac{1}{2} \cdot \sum_{i=1}^{N} (z_i - y_i)^T \cdot R^{-1} \cdot (z_i - y_i) + \frac{N}{2} \cdot \ln |R| \qquad (3)$$

Here the measurements obtained are discrete time, N is the number of measurements, R is the measurement noise covariance matrix. The estimate of R is obtained as:

$$\hat{R} = \frac{1}{N} \cdot \sum_{i=1}^{N} (z_i - \hat{y}_i) \cdot (z_i - \hat{y}_i)^T \qquad (4)$$

The maximum likelihood estimates are consistent, asymptotically unbiased and efficient ([25]).

The calculation of the maximum likelihood estimate is performed iteratively. The estimate value at the $(k+1)$ iteration is determined as:

$$\hat{\theta}_{ML}^{\;k+1} = \hat{\theta}_{ML}^{\;k} + [\nabla_\theta^{\,2} E(\theta)]^{-1} \cdot [\nabla_\theta E(\theta)] \qquad (5)$$

Here the first gradient is defined as:

$$\nabla_\theta E(\theta) = \sum_{i=1}^{N} \left[\frac{\partial y_i}{\partial \theta}\right]^T \cdot R^{-1} \cdot (z_i - y_i) \qquad (6)$$

For the work pressure model the expressions for the partial derivatives w.r.t. the parameters (i.e., sensitivity coefficients) have been obtained analytically (see

http://www.few.vu.nl/~fboth/OFS/appendix_B.pdf).

The analytical determination of the second gradient is more involved, therefore a Gauss-Newton numerical approximation has been used for it:

$$\nabla_\theta^{\,2} E(\theta) = \sum_{i=1}^{N} \left[\frac{\partial y_i}{\partial \theta}\right]^T \cdot R^{-1} \cdot \left[\frac{\partial y_i}{\partial \theta}\right] \qquad (7)$$

Such an approximation does not cause a significant error in the parameter estimate. Furthermore, the use of the second gradient speeds up the convergence of the estimation process significantly.

The state values of the system were calculated by numerical integration of the model equations using the 4th order Runge-Kutta method, which has proven to be both accurate and stable. The estimation error is calculated in each iteration as root mean square error:

$$err = \sqrt{\sum_{i=1}^{N} \frac{(z_i - \hat{y}_i)^2}{N}} \qquad (8)$$

The parameter estimation procedure based on the maximum likelihood principle has been implemented using the following algorithm:

Algorithm: ML-PARAMETER-ESTIMATION

Input: Initial values of the parameters θ^1, maximal number of iterations itmax; satisfactory error value err_sat; matrix of the input values U; matrix of the output values Z

Output: Maximum likelihood estimate θ_{ML}

1 i=1

2 Until i ≤ itmax perform steps 3-7

3 Calculate the current state of the system using the model equations (…)

4 Calculate the output root mean square error err^i using (8).

5 if err ≤ err_sat, then $\theta_{ML} = \theta^i$**; exit** endif.

6 if i < itmax, then

6a Calculate the noise covariance matrix *R* using (4)

6b Calculate the sensitivity coefficients $\partial y / \partial \theta$ using the formulae in the appendix

6c Calculate the first and second gradients using the formulae (6) and (7) respectively.

6d Calculate the parameter values for the next iteration θ^{i+1} using (5)

Endif

7 i = i+1

8 Find the minimum error err^m in $\{err^i \mid i=1..itmax\}$; then $\theta_{ML} = \theta^m$; **exit**.

The algorithm has been implemented in the Matlab 7 environment and applied to the work pressure model. The worst case complexity is estimated as $O(NN \cdot |\theta| \cdot M)$, where NN is the number of integration points, $|\theta|$ is the number of the estimated parameters, M is the number of outputs. The execution of each iteration took less than 2 seconds on an average PC.

7.2 Simulated Annealing

The Simulated Annealing approach ([13]) uses a probabilistic technique to find a parameter setting. In this method a random parameter setting is chosen as the best available parameter setting at the start. Then a displacement is introduced into these settings to generate a neighbor of the current parameter settings in the search space. If this neighbor is found more appropriate representation of the observed human behavior then it is marked as the best known parameter setting otherwise a new neighbor is selected to evaluate its appropriateness. The displacement in the parameter settings depends on the temperature, in case the temperature is higher, the steps will become larger. The temperature at a certain time point for the parameter settings is defined as follows

Temperature = computational-budget-left * error (9)

Here the computational budget is the number of neighbors to be tested for better approximation. The displacement in the parameter for example γ could be derived from following equations selecting any one at random.

$\gamma=\gamma+\text{Temperature} *(1-\gamma)* \text{random\_no\_between}[0,1]$ (10a)

or

$\gamma = \gamma-\text{Temperature} * \gamma * \text{random\_no\_between}[0,1]$ (10b)

The method is described as follows:

Algorithm: SA-PARAMETER-ESTIMATION

Input: Initial randomly selected values of the parameters θ^1, computational budget C; observed human behavior B;

Output: Best estimate of parameter settings θ_{BE}

1 $\theta_{BE}=\theta^1$

2 while $C \geq 0$ perform steps 3-8

3 Chose a random parameter setting θ in the neighbourhood of θ_{BE} using equation (9 and 10).

4 Calculate the output root mean square error err for θ using (8).

5 Calculate the output root mean square error err_{BE} for θ_{BE} using (8).

6 if $err \leq err_{BE}$, then $\theta_{BE} = \theta$; $err_{BE} = err$; endif.

7 Decrease C;

8 $Temperature = C * err_{BE}$;

9 **output** θ_{BE}.

The algorithm has been implemented in the C++ language and applied to the work pressure model. If C is computational budget, then the worst case complexity of the method can be expressed as $O(CB)$, where B is the number of observed behaviors. Here it could be observed that computational complexity of this method is independent number of parameter.

8 Validation Results

In this section, the results of the validation are shown. The results of both Parameter Tuning techniques are given and a formal verification (as in Section 5.2) is performed on traces of data from the Experiment.

8.1 Results of the Estimation

The gradient-based and simulated annealing methods have been applied for the estimation of 30 parameters of the work pressure model (see http://www.few.vu.nl/~fboth/OFS/appendix_C.pdf). The estimation has been performed for 31 subjects, for both experimental conditions. The initial setting of the parameters has been taken from the setting used for simulation purposes in Section 4, which is grounded partially in the psychological literature; furthermore it ensures the desired properties of the modeled system. Figure 5 illustrates the empirical data and the estimated output performance quality for subject 37 for both conditions. The estimation by both methods shows similar behavioral patterns in the output of the model. However, the gradient-based method has a better precision in comparison to the simulated annealing. The root mean square errors calculated in both parameter estimation methods are given in Table 2.

Table 2. Root Mean Square Errors of Estimation by the Gradient-Based (GB) and Simulated Annealing (SA) Methods for all Subjects in both Experimental Conditions

Error range		< 0.1	[0.1, 0.25)	[0.25, 0.4)	> 0.4
Subjects in condition 1	GB	21	11-20, 22, 24-41	-	-
	SA		40	11, 12, 22, 24-26, 30, 32-39, 41	13-18, 20, 21, 28, 29, 31
Subjects in condition 2	GB	12, 15, 18, 20, 21, 23, 27, 30	11, 13, 14, 16, 17, 19, 22, 24-26, 28, 32-41	29, 31	-
	SA	32	17, 26, 30, 31, 34. 35, 37, 40	12, 27, 38, 41	11, 13-16, 18-23, 25, 28, 29, 33, 36, 39

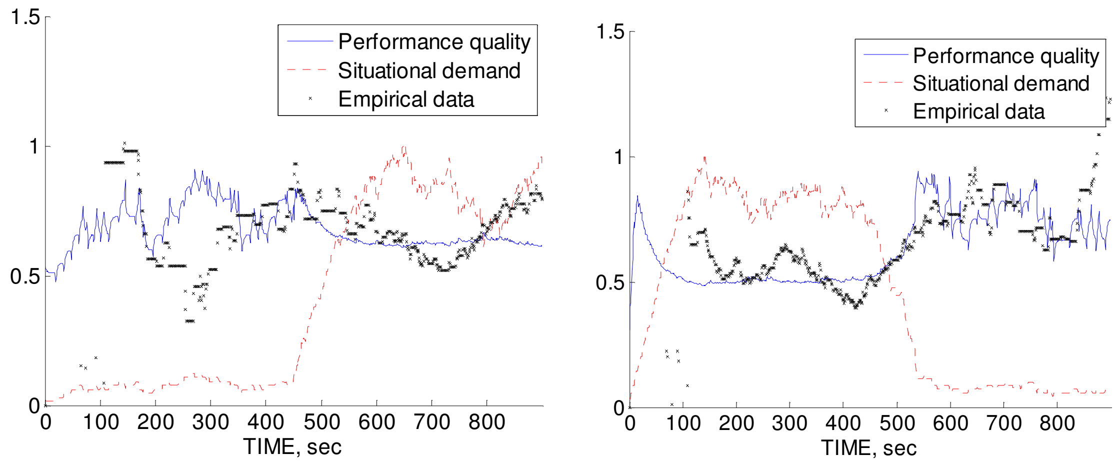


Figure 5a, b. Empirical Data and Estimated Output Performance Quality for Subject 37, for condition 1 (left) and condition 2 (right).

To evaluate the quality of estimation also other measures have been used. In particular, the Cramer-Rao bounds provide a useful measure of relative accuracy of the estimated parameters (Sorenson, 1980). This measure sets a lower bound on the standard deviation of the estimators:

$$\sigma_{\theta} \geq \sqrt{I^{-1}(\theta)} \tag{11}$$

Here $I(\theta)$ is the information matrix:

$$(I(\theta))_{ij} = E\left[\frac{\partial^2 \log\ p(z \mid \theta)}{\partial \theta_i \partial \theta_j}\right] \tag{12}$$

For efficient estimation the equality holds. Furthermore, for the maximum likelihood method, $I(\theta) = \nabla^2_{\theta} E(\theta)$, which also needs to be calculated for (7); thus no additional computation effort for the evaluation of this measure is required.

Using this measure at least 57% (70% in the best case) of the estimated parameters have been identified as accurate for all subjects in both conditions (relative standard deviation (rsd) $\leq$ 5%). Other parameters, although less accurate (5% < rsd < 40%) still have a degree of confidence.

Another useful criterion for judging the quality of the estimates is the correlation coefficients among the estimates calculated as:

$$c_{\theta_i \theta_j} = \frac{(I(\theta)^{-1})_{ij}}{\sqrt{(I(\theta)^{-1})_{ii} \cdot (I(\theta)^{-1})_{jj}}} \tag{13}$$

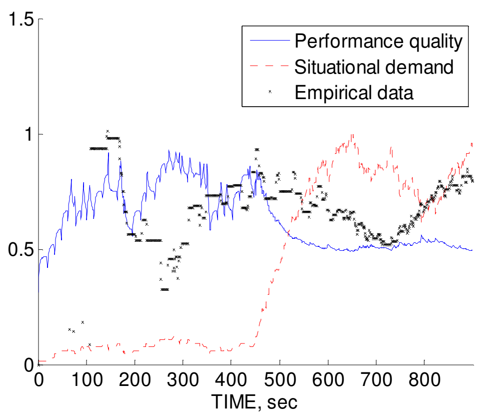
Only one significant correlation between the parameters A and ϕ has been identified.

The precision of the parameter estimation is essential for prediction of the system dynamics using the model. To examine predictive capabilities of a model, cross-validation is often used. In the cross-validation of the work pressure model the empirical data of the condition 2 have been used for the parameter estimation, whereas the data of the condition 1 were used for validation of the model with the parameter estimates obtained from the condition 1.

Table 3. Prediction Errors of Estimation by the Gradient-Based (GB) and Simulated Annealing (SA) Methods for all Subjects in Condition 1 using the Estimated Parameters from Condition 2

Error range	< 0.1	[0.1, 0.25)	[0.25, 0.4)	> 0.4
GB	21	12-20, 22, 24-30, 34-40	11, 31, 32, 41	33
SA	-	17, 26, 31, 32, 37, 40	12, 13, 22, 25, 28, 30, 34, 35, 38, 41	11, 14-16, 18-21, 29, 33, 39

The prediction quality was determined by comparing the root mean square errors for both conditions. For most of the subjects (84%) in the GB estimation prediction errors (Table 3) differ from the estimation errors (Table 1, subjects in condition 1) insignificantly (less than 10%). Notably, for a number of subjects in the SA estimation, the prediction errors are lower than the estimated errors. This indicates that the estimation was not optimal in these cases. Furthermore, also cross-validation has been performed, in which the data of setting 1 were used for the parameter estimation and the data of setting 2 were used for validation (Figure 6).



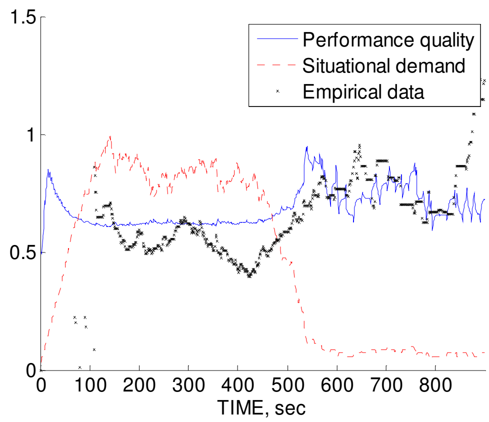


Figure 6a, b. Predicted Dynamics of the Model in setting 1 (left) and setting 2 (right).

8.2 Validation using Logical Analysis

Besides showing that parameter settings can be found so that the model accurately describes and predicts the behavior of humans as observed during the experiments, this section focuses on another approach, namely logical verification. The idea is that properties are identified that follow from the work pressure model which is being validated, and these properties are verified against the empirical data that has been obtained. This is done, using the same approach and language as explained in Section 5.2.

For validation, three main properties have been identified that follow from the work pressure model. The first property specifies that performance quality decreases in case a task level in a certain range is experienced:

P1(min_level, max_level, d, x)
*If at time point t1 the task level is tl and the performance quality pq, and tl is in the range [min_level max_level], and until t1+d the task level does not go outside of these boundaries, then there exists a time point t2> t1 at which the performance quality is at most x * pq.*

P1(min_level, max_level, d, x) ≡
$\forall\gamma$:TRACE, t1:TIME, pq1:REAL
[state(γ, t1) |= has_value(performance_quality, pq1) &
$\forall$tl:REAL, t':TIME $\geq$ t1 & t' $\leq$ t1 + d
[state(γ, t') |= has_value(task_level, tl) $\Rightarrow$[tl $\leq$ max_level & tl $\geq$ min_ level]]
$\Rightarrow$ $\exists$t2:TIME > t1, pq2:REAL [state(γ, t2) |= has_value(performance_quality, pq2) & pq2 $\leq$ x * pq1]

This property has been verified using the following values: the min_level has been set to 20% above the basic cognitive abilities, whereas the max_level is set to the maximum task level encountered in the experiment. Furthermore, the duration d is set to 60 time steps (i.e. a minute real time), and x is set to 1 (i.e. the performance quality should never go up, but can remain the same). These setting follow the model: in case a task level above the basic cognitive abilities is experienced, the human becomes exhausted, and the quality can no longer go up. Results show that this property is satisfied in **60%** of the empirical traces.

The second property concerns the opposite: in cases where there is a task level between certain boundaries, the performance quality should be at least as high as before the period:

P2(min_level, max_level, d, x)
*If at time point t1 the task level is tl and the performance quality pq, and tl is in the range [min_level, max_level], and until t+d the task level does not go outside of these boundaries, then there exists a time point t2> t1 at which the performance quality is at least x * pq.*

P2(min_level, max_level, d, x) ≡
$\forall\gamma$:TRACE, t1:TIME, pq1:REAL
[state(γ, t1) |= has_value(performance_quality, pq1) &
$\forall$tl:REAL, t':TIME $\geq$ t1 & t' $\leq$ t1 + d
[state(γ, t') |= has_value(task_level, tl) $\Rightarrow$ [tl $\leq$ max_level & tl $\geq$ min_level]]
$\Rightarrow$ $\exists$t2:TIME > t1, pq2:REAL [state(γ, t2) |= has_value(performance_quality, pq2) & pq2 $\geq$ x * pq1]

In this case, the following settings have been used: The max_level has been set to 20% below the basic cognitive abilities, whereas the min_level is set to 0. The parameters d and x has been set the same as for the previous property. This property complies to the work pressure model. In case a task level is experienced which is somewhat below the highest task level that can be handled without exhaustion building up (i.e. the basic cognitive abilities), then the performance will get better, or at least stay the same (as there is no exhaustion). It has been shown that the property is satisfied for **45%** of the empirical cases.

The final property which has been verified concerns performance quality being higher for cases whereby there is a lower task level:

P3(low_level, high_level)
In case the task level at a time point t1 is tl1, and at a time point t2 the task level is tl2, and tl1 > high_level and tl2 < low_level, then there exists a time point t' > t1 and there exists a time point t'' > t2 such that the performance quality at time point t' is lower than the performance quality at time point t''.

P3(low_level, high_level) ≡
$\forall\gamma$:TRACE, t1, t2:TIME, tl1, tl2:REAL
[state(γ, t1) |= has_value(task_level, tl1) &
state(γ, t2) |= has_value(task_level, tl2) & tl1 $\geq$ high_level & tl2 $\leq$ low_level]
$\Rightarrow$ $\exists$t', t'':TIME, pq1, pq2:REAL
[t' > t1 & t'' > t2 & state(γ, t') |= has_value(performance_quality, pq1) &
state(γ, t'') |= has_value(performance_quality, pq2) & pq1 < pq2]

Using a setting of low_level of 20% below the basic cognitive abilities, and a high_level of 20% above the cognitive abilities, this property is satisfied in **60.7%** of the cases. The property complies to the model as a task level beyond the basic cognitive abilities results in exhaustion, resulting in a worsened performance, which is not the case for a task level far below the basic cognitive abilities.

In total, **25.0%** of the cases comply to property P1, P2, and P3.

9. Discussion

In order to develop an intelligent system that supports humans in demanding circumstances, a computational model is required that describes the dynamics of a human's functional state in relation to task performance and the environment. To this end, this paper presented such a model, which was developed in dynamical system style. The model takes task demand and situational aspects as input and determines internal factors such as the experienced pressure, exhaustion and motivation, and how they (may) affect task performance. Using Matlab, a large number of simulation experiments under different parameter settings have been performed. These experiments pointed out that the model is able to produce realistic behavior of different types of personalities. Moreover, by a mathematical analysis the equilibria of the model have been determined, and by automated checking a number of expected properties of the model have been verified. For example, these checks pointed out that (at least in all generated traces), all variables stayed within their boundaries, and the calculated equilibria are confirmed. In addition, the specific hypothesis that persons with a high Optimal Experienced Pressure eventually end up with a lower Critical Point was confirmed.

Although a number of other approaches in the literature address various aspects of stress, exhaustion, or situation awareness separately ([6], [14], [19]), we are not aware of any attempts to combine all these aspects together in as much detail as the current approach.

The mathematical and automated analyses described above have been successfully performed to guarantee *internal* validity, but this does not guarantee that the model is directly applicable to real humans, and in particular which personality parameter values fit to which person. Therefore, as a next step, validation of the model (with respect to its purpose) in laboratory experiments was performed. The approach that was chosen was to offer a human certain demanding tasks, measure its performance and several physiological data, feed these data into the model, and compare the output of the model with self-reports of the participants (similar to [16]). This provides validation of the model, by providing realistic parameter settings for different types of individuals. Note that the paper does not make an exclusive claim on whether the proposed model is the one and only model satisfying this application-directed validity criterion.

The experience with the experiment was that the participants were very motivated to perform well on the main task. This was not only due to the reward; they were also enthusiastic about the game itself. In order to keep the learning effect to a minimum and to maintain the participants' concentration, every participant performed only two sessions of the 15 minute game. However, precision of parameter estimation will increase when measurements of more within-subject conditions are taken.

The results obtained for the parameter estimation are satisfactory. However, a number of parameters (35% in average) were evaluated as less accurate, and, therefore, less reliable. Partially this can be explained by a large overall number of parameters being estimated. Most of the less precise parameters have a weak relation to the measured output (e.g., noise sensitivity) Furthermore, since the empirical data were collected based on irregular events (i.e., actions of humans), some intervals contained the amount of information insufficient for estimation. Despite this, as shown in the paper, the models with estimated parameters demonstrated good predictive capabilities in the cross-validation, which is a strong indicator of validity of the model for the purpose for which it was developed, namely personalised support.

The trends as predicted by the model have also been verified against the empirical material. The results show that a reasonable percentage of the traces satisfy each of these individual properties. The combination of all three properties is however only satisfied in 25% of the cases, which can mainly be attributed to the aforementioned collection based on irregular events, making the data obtained more prone to sudden changes.

The topic of model validation received much attention in the areas of Psychology and Social Science. In particular, a validation approach from ([26]) distinguishes the validation phases similar to the ones considered in the paper (e.g., conceptual and operational validation); however, the precise elaboration of the phases is focused largely on social processes, which are not relevant for our work. Furthermore, examples of model validation are found in psychology, e.g. on the subject of visual attention ([20]), however often no parameter estimation is involved.

In future research, the considered parameter estimation methods will be extended for the case of real-time estimation, which accounts for human learning. Furthermore, a personal assistant agent will be implemented that is able to monitor and balance work pressure of the human in a timely and knowledgeable manner. By incorporating such methods, which open possibilities to further tailor the presented model to individual cases, the model can become more personalized and flexible. The authors believe that this is crucial to provide effective support to humans that operate in demanding circumstances.

10. References

[1] Bosse, T., Both, F., Lambalgen, R. van, and Treur, J. (2008). An Agent Model for a Human's Functional State and Performance. In: Jain, L., Gini, M., Faltings, B.B., Terano, T., Zhang, C., Cercone, N., and Cao, L. (eds.), *Proceedings of the Eighth IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT'08*. IEEE Computer Society Press, 2008, pp. 302-307.

[2] Bosse, T., Jonker, C.M., Meij, L. van der, Sharpanskykh, A., and Treur, J. (2009). Specification and Verification of Dynamics in Agent Models. *International Journal of Cooperative Information Systems*, vol. 18, 2009, pp. 167-193.

[3] Both, F., Hoogendoorn, M., Jaffry. S.W., Lambalgen, R. van, Oorburg, R., Sharpanskykh, A., Treur, J., and Vos, M. de (2009). Adaptation and Validation of an Agent Model of Functional State and Performance for Individuals. In: Yang, J.-J.; Yokoo, M.; Ito, T.; Jin, Z.; Scerri, P. (eds.), *Proceedings of the 12th International Conference on Principles of Practice in Multi-Agent Systems, PRIMA'09*. Lecture Notes in Artificial Intelligence, vol. 5925. Springer Verlag, 2009, pp. 595–607.

[4] Costa and McCrae, 1992. The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6.* 343-359.

[5] Davidson, R., & Henderson, R. (2000). Electronic Performance Monitoring: a laboratory investigation of the influence of monitoring and difficulty on task performance, mood state, and self-reported stress levels. Journal of Applied Social Psychology 30 (5), pp 906-920.

[6] Endsley, M. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, vol. 37, issue 1, pp. 3264.

[7] Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors,* 47, 92-101.

[8] Hancock, P.A., & Desmond, P.A. Stress, Workload and Fatigue. New Jersey: Lawrence Erlbaum Associates, 2001.

[9] Hancock, P.A., Williams, G., Manning, C.P., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The Intern. J. of Aviation Psychology, 5(1)*, 63-86.

[10] Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31(5), 519-537.

[11] Hill, D.W. (1993). The critical power concept. *Sports Medicine*, vol.16, pp. 237-254.

[12] Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biological Psychology 45*, 73-93.

[13] Ingber, L. (1993). Simulated Annealing: Practice Versus Theory. Math. Comput. Modelling 18, pp. 29-57.

[14] Kaber, D.B., & Endsley, M.R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, vol. 5, issue 2, pp. 113-153.

[15] Klein, G. (1996). The effect of acute stressors on decision making. In J. Driskell & E. Salas (Eds.) *Stress and human performance* (pp 48-88), Mahwah, NJ: Erlbaum.

[16] Koning, L. de, Maanen, P.-P. van, Dongen, K. van (2008). Effects of Task Performance and Task Complexity on the Validity of Computational Models of Attention, In: *Proceedings of the Human Factors and Ergonomics Society's 52nd Annual Meeting*, New York, USA, 2008.

[17] Lambalgen, R. van, and Maanen, P.-P. van (2010). Personalisation of Computational Models of Attention by Simulated Annealing Parameter Tuning, *Proceedings of the Fourth International Workshop on Human Aspects in Ambient Intelligence*, held in collaboration with the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE Computer Society Press, 2010.

[18] Matthews, G. and Deary, I.J. (1998). *Personality traits*. Cambridge, UK: Cambridge University Press.

[19] Neerincx, M.A., Grootjen, M. & Veltman, J.A. (2004). How to Manage Cognitive Task Load During Supervision and Damage Control in an All-Electric Ship. *IASME Transactions*, 2(1), 253-258.

[20] Parkshurst, D., Law, K., Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. Vision Research 42, 107-123.

[21] Plomin, R. & Spinath, F.M. (2002). Genetics and general cognitive ability (g). *Trends in Cognitive Sciences 6*. 169-176.

[22] Recarte, M.A., & Nunes, L.M. (2003). Mental workload while driving: effects on visual search, discrimination and decision making. *J. of Experimental Psychology: Applied, 9(2)*, 119-137.

[23] Rose, C.L., Murphy, L.B., Byard, L., & Nikzad, K. The role of the Big Five personality factors in vigilance performance and workload. European Journal of Personality 16: 185-200 (2002).

[24] Salgado, J.F. (1997). The five factor model

[25] Sorenson, H.W. (1980). *Parameter Estimation: Principles and Problems*. Marcel Dekker, Inc., New York.

[26] Yilmaz, L. Validation and verification of social processes within agent-based computational organization models. *Comput. Math. Org. Theory*, 12:283-312, 2006.