Validation of an Agent Model for Human Work Pressure

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 1081a, 1081 HV Amsterdam, the Netherlands
{fboth, swjaffry, rm.van.lambalgen, sharp, treur}@few.vu.nl

[2]Force Vision Lab, Barbara Strozzilaan 362a,
1083 HN Amsterdam, The Netherlands
{rogier, michael}@forcevisionlab.nl

Abstract

Human performance can seriously degrade under demanding tasks. To improve performance, agents can reason about the current state of the human, and give the most appropriate and effective support. To enable this, the agent needs a work pressure model, which should be valid, as the agent might otherwise give inappropriate advice and even worsen performance. This paper concerns the validation of an existing work pressure model. First, human experiments have been designed and conducted, whereby measurements related to the model have been performed. Next, this data has been used to obtain appropriate parameter settings for the work pressure model, describing the specific subject. Finally, the work pressure model, with the tailored parameter settings, has been used to predict human behavior to investigate predictive capabilities of the model. The results have been analyzed using formal verification.

Introduction

In demanding working circumstances the quality of the tasks performed by a human might be severely influenced (cf. Hancock *et al.*, 1995, Hanley, 1997). Especially when tasks are performed in a critical domain, such effects are highly undesired. To improve task performance in such situations, personal assistant agents (cf. Kozieok and Maes, 1993; Mitchell *et al.*, 1994; Maheswaran *et al.,* 2003) can be used to monitor the activities of the human, and intervene in case needed. Interventions could for example take the form of assigning (part of) the tasks to other humans, or give advice regarding the performance of the task.

One crucial element in the support given by a personal assistant agent is that it should be given in appropriate circumstances: the agent should have an awareness of the state of the human. In Bosse *et al.* (2008a) a dynamical model has been presented that describes the cognitive workload experienced by humans, given knowledge of the human's characteristics in combination with the tasks that need to be performed. The model is quantitative, based upon mostly qualitative theories from Psychology, but was not validated yet using human experiments. The primary focus of this paper is to develop and implement an approach for the validation of this human work pressure model. 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 were to play a game 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 subject's 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 relate to the work pressure model have been verified against the empirical data as well.

This paper is organized as follows. First, the work pressure model is briefly explained. Thereafter, the setup of the experiment and the results of parameter estimation are shown. Next, the verification of properties against the empirical data, and finally the paper is concluded and future work is discussed.

Work pressure model

The Agent model for the Functional State (FS) of a human represents the dynamical state of a person when performing a certain task. States such as experienced pressure, motivation and exhaustion of the person are predicted, but also the performance quality and the amount of generated effort to the task..

The model is based on two different theories: 1) the cognitive energetic framework (Hockey, 1997), which states that effort regulation is based on human recourses and determines human performance in dynamic conditions; 2) The idea, that when performing sports, a person's generated power can continue on a *critical power* level without becoming more exhausted (Hill, 1993). In the FS model (cf. Figure 1) critical power is represented by the critical point: the amount of effort someone can generate without becoming more exhausted.

As input the FS model uses external factors (task demands and environment state) and personal factors (experience, cognitive abilities and personality profile), which are used to determine a person's dynamical state. In addition, it determines the relation of this state to the human's actions with respect to the task (e.g. performance quality), represented in the Task Execution State.

An example equation of the model is:

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

Here Exhaustion (*E*) builds up or reduces over time. When the generated effort (*GE*) is above the critical point (*CP*), exhaustion increases, otherwise exhaustion decreases depending on the level of recovery effort (*RE*). Parameters η and π determine the amount of increase or decrease. The function *Pos(x)* in this formula is defined as the maximum of x and 0. For more details on the model, see (Bosse *et al.*, 2008).

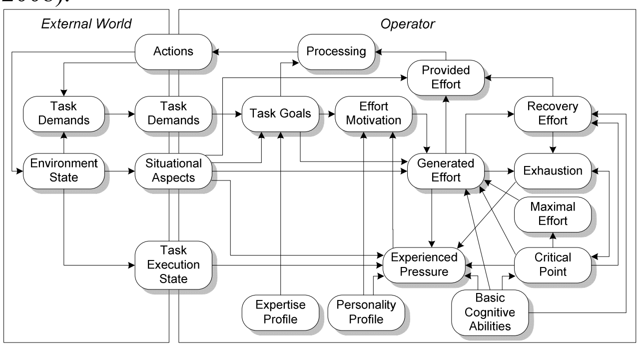

Figure 1. Agent Model for an Operator's Functional State

Experimental setup

First, an overview of the game and its participants is given. The main part of the experiment is a game which combines a shooting task and a calculation task. Thereafter, the procedure of the experiment is explained. A more detailed version can be found in Appendix A: http://www.few.vu.nl/~fboth/ICCM/appendix_A.pdf.
Finally, a description is given of how data from the experiment has been used as input for the work pressure model.

Game and Participants

In the experiment the main task is a shooting game where the goal is to get as many points as possible. Objects (friends and enemies) were falling down in different locations at different speeds. The purpose is to shoot the enemies before they hit the ground. Shooting at a missile is done by a mouse click at a specific location; the missile would then explode exactly at the location of the mouse click. The speed with which the missile reaches this location is 79.6 pixels per second. When an object is within a radius of 50 pixels of the explosion, the object is destroyed. The number of points a participant receives for hitting an enemy is proportional to the proximity of the explosion. When a participant shoots a friend or when an enemy reaches the bottom of the screen, points are lost. When a friendly object reaches the bottom of the screen points are gained. Next to each of the objects, a calculation is written on the screen. A correct calculation indicates that the object is friendly and should not be shot. An incorrect calculation indicates that the object is an enemy and should be shot before it reaches the bottom of the screen. For a demo of the shooting game, see http://www.forcevisionlab.nl/demo/missilecommand.swf.

In the study 31 persons participated (18 males, 13 females, of which 25 students). They 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 game.

Procedure

For the experiment a 2 factor within subjects design was used. Two different conditions within each participant were tested. In Bosse *et al.* (2008a), two scenarios were simulated using the model. Scenario 1 started with a low task level and continued with a high task level. Scenario 2 started with a high task level and continued with a low task level. Condition was counterbalanced over participants to correct for a possible order effect, such that participants with an odd number started with condition 2 and even numbered participants 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 (Costa and McCrae, 1992); with these questions some aspects of each participant's personality were measured, to serve as input for the personality profile of the work pressure 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 consisting of 30 trials which were equal between participants. These tests served as input for model validation (see the next subsection and Appendix A for the explanation thereof). Instructions for each test were shown on the screen. The first test was 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. Again, 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 for the experiment-game described in the previous subsection. 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.

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 work pressure model, namely personality profile, basic cognitive abilities (BCA) and expertise profile, following theories from Psychology (Matthews & Deary, 1998; Plomin &

Spinath, 2002; Rose *et al*, 2002; Salgado, 1997). Hereby, several parameters are introduced that need to be estimated by the parameter estimation approach as well. Including this, the number of parameters that should be estimated is 27. For the precise mathematical equations used, see http://www.few.vu.nl/~fboth/ICCM/appendix_D.pdf.

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 and the expertise profile together contribute to task level.

$$TaskLevel = (1.5 - Exp)\cdot SitD \qquad (1)$$

In the experiment, performance quality was measured in terms of efficiency and effectiveness. Efficiency represented the number of missiles necessary to shoot an enemy. Effectiveness was dependent on how close to the object the missile exploded (explosion fraction) and whether an enemy or friend was shot. In case of an enemy being shot:

$$Effectiveness = (1 + explosion\_fraction)/2.0 \qquad (2)$$

Effectiveness was 0 when a friend was shot or an enemy landed. When a friend landed, effectiveness was 1. Using effectiveness and efficiency, the task execution state was calculated:

$$ObjTES = (0.25\cdot efficiency + 0.75\cdot effectiveness)\cdot 2 \qquad (3)$$

Estimation of parameters

This section presents the results of parameter estimation for the work pressure model using two different methods: a gradient-based approach and an approach based on probabilistic search.

Gradient-based parameter estimation

To perform parameter estimation, a method based on the maximum likelihood principle has been applied (Sorenson, 1980). 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 $p(z|\theta)$. Furthermore, it was assumed that the measurements contained noise which is zero-mean and has a Gaussian distribution. The measurement data were represented by the random, normally distributed variable z. Such an assumption is often made for dynamic systems in many areas. The parameter vector, which makes the likelihood function most probable to obtain the measurements z ($\ldots\hat{\theta}_{ML}$ .. which maximizes the likelihood function) is called the maximum likelihood estimate; it is obtained by minimizing the error function:

$$E(\theta) = \frac{1}{2}\cdot\sum_{i=1}^{N}(z_i - y_i)^T\cdot R^{-1}\cdot(z_i - y_i) + \frac{N}{2}\cdot\ln|R| \qquad (5)$$

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

$$\hat{R} = \frac{1}{N}\cdot\sum_{i=1}^{N}(z_i - \hat{y}_i)\cdot(z_i - \hat{y}_i)^T \qquad (6)$$

The maximum likelihood estimates are consistent, asymptotically unbiased and efficient (Sorenson, 1980).

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

$$\hat{\theta}_{ML}^{\,k+1} = \hat{\theta}_{ML}^{\,k} + [\nabla_\theta^2 E(\theta)]^{-1}\cdot[\nabla_\theta E(\theta)] \qquad (7)$$

Here the first gradient is defined as

$$\nabla_\theta E(\theta) = \sum_{i=1}^{N}\left[\frac{\partial y_i}{\partial\theta}\right]^T\cdot R^{-1}\cdot(z_i - y_i) \qquad (8)$$

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 Appendix B: http://www.few.vu.nl/~fboth/ICCM/appendix_B.pdf).

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

$$\nabla_\theta^2 E(\theta) = \sum_{i=1}^{N}\left[\frac{\partial y_i}{\partial\theta}\right]^T\cdot R^{-1}\cdot\left[\frac{\partial y_i}{\partial\theta}\right] \qquad (9)$$

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:

$$err = \sqrt{\sum_{i=1}^{N}\frac{(z_i - \hat{y}_i)^2}{N}} \qquad (10)$$

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 (10).

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 (6)

6b Calculate the sensitivity coefficients $\partial y/\partial\theta$

6c Calculate the first and second gradients using the formulae (8) and (9) respectively.

6d Calculate the parameter values for the next iteration θ^{i+1} using (7)

endif

7 i = i+1

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

The algorithm was implemented in the Matlab 7 environment. The worst case complexity is estimated as O(NN·|θ|·M), where NN is the number of integration points, |θ| is the number of the estimated parameters, M is the number of outputs. The execution of an iteration took less than 2 sec on an average PC.

Simulated annealing

The Simulated Annealing method 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

$$\textit{Temperature} = \textit{computational-budget-left} \cdot \textit{error} \qquad (11)$$

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

$$\gamma=\gamma+\textit{Temperature} \cdot (1-\gamma) \cdot \textit{random\_no\_between}[0,1] \qquad (12a)$$

$$\text{or } \gamma = \gamma-\textit{Temperature} \cdot \gamma \cdot \textit{random\_no\_between}[0,1] \qquad (12b)$$

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 behaviour B;
Output: Best estimate of parameter settings θ_{BE}
1 $\theta_{BE}=\theta^1$
2 while $C \geq 0$ perform steps 3-8
3 Choose a random parameter setting θ in neighbourhood of θ_{BE} using equation (11 and 12a, 12b).
4 Calculate the output root mean square error err for θ using (10).
5 Calculate the output root mean square error err_{BE} for θ_{BE} using (10).
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}.

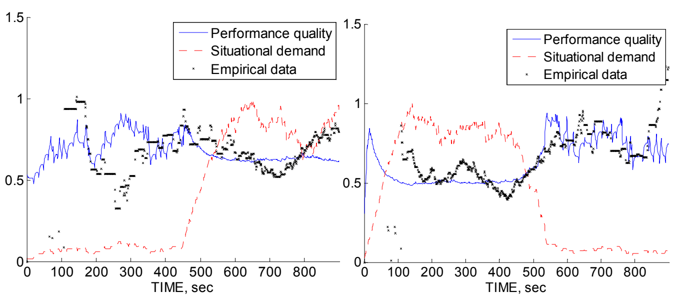

Figure 2. Empirical data and the estimated output performance quality for subject 37 for condition1 (left) and condition 2 (right)

In figure 2 performance quality for subject 37 is shown for computational budget 10000 and 900 observed human behavior. Here it should be noted that graph represents the curve generated with parameter settings producing minimum root mean square error found till the end of computational budget. The algorithm has been implemented in C++ 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(C·B), where B is the number of observed behaviors. Here it could be observed that computational complexity of this method is independent number of parameter.

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 Appendix C: http://www.few.vu.nl/~fboth/ICCM/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 Bosse *et al.* (2008a). This setting is grounded partially in the psychological literature; furthermore it ensures the desired properties of the modeled system. Figure 2 illustrates the empirical data and the estimated output performance quality for subject 37 for both conditions.

The estimation by both methods showed 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 1. 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).

Table 1. 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

This measure sets a lower bound on the standard deviation of the estimators:

$$\sigma_\theta \geq \sqrt{I^{-1}(\theta)} \qquad (13)$$

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

$$(I(\theta))_{ij} = E\left|\frac{\partial^2 \log p(z \mid \theta)}{\partial\theta_i \partial\theta_j}\right| \qquad (14)$$

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 (9); 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:

$$c_{\theta_i\theta_j} = \frac{(I(\theta)^{-1})_{ij}}{\sqrt{(I(\theta)^{-1})_{ii}\cdot(I(\theta)^{-1})_{jj}}} \tag{15}$$

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.

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 2) differ from the estimation errors (Table 1, subjects in condition 1) insignificantly (less than 10%). Furthermore, also cross-validation was performed, in which data from one of the settings were used for parameter estimation and data from the other setting were used for validation (Figure 3).

Table 2. Prediction errors of estimation by the GB and 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

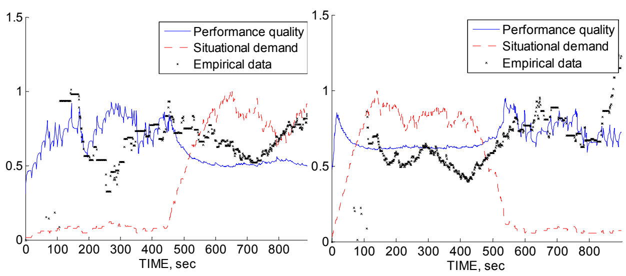

Figure 3. Predicted dynamics for subject 37 in condition 1 using the estimated parameters from condition 2 (left) and in setting 2 using the parameters from setting 1 (right)..

Verification of Properties

This section focuses on logical verification, another approach which has been used to validate the model. The idea is that properties are identified that are entailed by the work pressure model, and these properties are verified against the empirical data that has been obtained. In order to conduct such an automated verification, the properties have been specified in a language called TTL (for Temporal Trace Language, cf. Bosse *et al.*, 2008b) that features a dedicated editor and an automated checker. This predicate logical temporal 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 |=, i.e., 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 ¬, ∧, ∨, ⇒, ∀, ∃. For more details on TTL, see (Bosse et al., 2008a).

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 cross 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) ≡
∀γ:TRACE, t1:TIME, pq1:REAL
[state(γ, t1) |= has_value(performance_quality, pq1) &
∀tl:REAL, t':TIME ≥ t1 & t' ≤ t1 + d
[state(γ, t') |= has_value(task_level, tl) ⇒
[tl ≤ max_level & tl ≥ min_ level]]
⇒ ∃t2:TIME > t1, pq2:REAL
[state(γ, t2) |= has_value(performance_quality,pq2) &pq2 ≤ x * pq1]

This property has been verified using the following values: min_level is set to 20% above BCA, max_level is set to the highest task level encountered in the experiment, the duration d is set to 60 time steps (i.e. a minute real time), and x is set to 1 (i.e. performance quality should never go up, but can remain the same). These settings follow the model: in case a task level above BCA 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 (note that the formal form has been omitted for the sake of brevity):

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 cross these boundaries, then there exists a time point t2> t1 at which the performance quality is at least x * pq.*

Using the following settings: max_level at 20% below BCA, min_level is set to 0 and d and x the same as for the previous property, this property is satisfied in **45%** of the cases. 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 BCA), then the performance will get better, or at least stay the same (as there is no exhaustion).

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''.

Using a low_level of 20% below BCA, and a high_level of 20% above the cognitive abilities, this property is satisfied in **60.7%** of the cases. The property complies with the model, because a task level beyond BCA results in exhaustion leading to a worsened performance, which is not the case for a task level far below BCA. In total, **25.0%** of the cases comply with properties P1, P2, and P3.

Discussion and conclusions

To reason about the human behavior and support possibilities personal assistant agents often use (cognitive) models. To ensure that support is provided by agents in a timely and knowledgeable manner, such models should be accurate and validated. This paper contributes an approach to validate the work pressure model. In the following the performed validation steps of the approach are discussed.

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 the model validity.

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 (Yilmaz, 2006) 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 (Parkhurst et al., 2002), however often no parameter estimation is involved.

In the 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.

References

Bosse, T., Both, F., Lambalgen, R. van,, & Treur, J. (2008). An Agent Model for a Human's Functional State and Performance. In: Jain, L. *et al.* (eds.), *Proceedings of International Conference IAT'08*. (pp. 302-307). IEEE Computer Society Press.

Bosse, T., Jonker, C.M., Meij, L. van der, Sharpanskykh, A., & Treur, J. (2008). Specification and Verification of Dynamics in Agent Models. *International Journal of Cooperative Information Systems*, *18 (1)*, 167-193.

P.T. Costa Jr., & R.R. McCrae. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and the NEO Five-Factor Inventory (NEO-FFI) professional manual* (Psychological Assessment Resources). Odessa, FL.

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

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

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.

Kozierok, R., & Maes, P. (1993). A Learning Interface Agent for Scheduling Meetings. *Proceedings of the 1st International Conference on Intelligent User Interfaces* (pp. 81-88).

Maheswaran, R., Tambe, M., Varakantham, P., & Myers, K. (2003). Adjustable autonomy challenges in personal assistant agents: A position paper. *Proceedings of the AAMAS'03 Workshop on Agents and Comp. Autonomy* (pp.187–194).

Matthews, G., & Deary, I.J. (1998). *Personality traits*. Cambridge, UK: Cambridge University Press.

Mitchell, T., Caruana, R., Freitag, D., McDermott, J., & Zabowski, D. (1994). Experience with a Learning Personal Assistant. *Communication of the ACM 37(7)*, 81-91.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research 42(1)*, 107-123.

Plomin, R., & Spinath, F.M. Genetics and general cognitive ability. *Trends in Cognitive Science 6(4)*, 369-176.

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

Salgado, J.F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology 82* (1): 30–43.

Sorenson, H.W. (1980). *Parameter estimation: principles and problems*. Marcel Dekker, Inc., New York.

Yilmaz, L. (2006). Validation and verification of social processes within agent-based computational organization models. *Computational and Mathematical Organization Theory*, *12*, 283-312.