When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

van Lambalgen, R.M.

2012

Chapter 4

Relating Personality and Physiological Measurements to Task Performance Quality

This chapter appeared as:

Both, F., Hoogendoorn, M., Lambalgen, R. van, Oorburg, R., and Vos, M. de, Relating Personality and Physiological Measurements to Task Performance Quality. In: N.A. Taatgen and H. van Rijn (eds.), Proc. of the 31th Annual Conference of the Cognitive Science Society, CogSci'09. Cognitive Science Society, Austin, TX, 2009, pp. 2819-2825.

Relating Personality and Physiological Measurements to Task Performance Quality

Fiemke Both[1], Mark Hoogendoorn[1], Rianne van Lambalgen[1], Rogier Oorburg[2], and Michael de Vos[2]

[1]Vrije Universiteit Amsterdam, Department of Artificial Intelligence
De Boelelaan 1081a, 1081 HV Amsterdam, The Netherelands
Email: {fboth, mhoogen, rm.van.lambalgen}@few.vu.nl
[2] Force Vision Lab, Barbara Strozzilaan 362a, 1083 HN Amsterdam, The Netherlands
Email: {rogier, michael}@forcevisionlab.nl

Abstract. Human performance can degrade as a result of an increasing cognitive workload. Especially in domains in which good task performance is crucial these effects are unwanted, and hence, need to be avoided. In Psychological literature studies have been presented which relate a variety of measures to the workload experienced by humans. However, these experiments have not been conducted with more complex tasks. In this paper, the aim is to perform a variety of measures to see whether combinations thereof can be used to predict performance of humans under tasks with varying complexity. A dedicated experiment has been conducted with 31 subjects, of which the results have been analyzed from both a statistical as well as a temporal perspective.

Keywords: human workload, human performance, personality, physiological measures.

1. Introduction

Working under high pressure can result in negative effects on the functional state of a human (see e.g. Hancock 1995, Hockey 1997). As a result of this negative effect, the performance of the task at hand can be negatively influenced. In critical domains such degradation of performance can lead to severe consequences that might be highly undesired (Wickens, 2002) Therefore it is very useful to be able to obtain more precise knowledge about measurements that can be used to make more accurate predictions concerning the performance upon tasks (and the potential degradation thereof).

In the literature, various measurements have been proposed to measure the cognitive workload experienced by humans. Roscoe (1992) for example, shows that the heart rate of a human is a good indicator of cognitive workload (if no physical effort is performed). Also subjective measurements have been put forward, such as the NASA-TLX questionnaire (Hart and Staveland, 1988). In (Rose et al., 2002) the relation between the so-called Big Five personality factors upon vigilance performance and workload has been studied. However, the tasks studied do not concern complex tasks.

In this paper, the aim is to perform a set of measurements to see whether the performance of a task can be predicted for complex tasks as well using (a combination of) these measurements. Hereby, an experiment has been conducted whereby participants had to perform a cognitive task with varied complexity. The measurements performed consisted of: (1) personality characteristics (using the NEO-PI-R and NEO-FFI test, cf. Costa and McCrae, 1992); (2) cognitive abilities (using several tests); (3) performance quality using an objective measure; (4) the complexity of the task; (5) the NASA-TLX as a measure of perceived workload, and (6) the heart rate during the experiment.

For the data obtained, the prediction is that performance will depend on personality characteristics, as well as on heart rate and the complexity of the task. Furthermore, the subjective performance as measured with the NASA-TLX is predicted to be dependent on objective performance and personality.

For the data analysis, not only a statistical analysis is performed, but also more complex temporal properties have been analyzed. For example trends over time that influence the performance quality and perceived workload will be analyzed. In order to enable such a temporal analysis, tools from the domain of Artificial Intelligence are used (cf. Bosse *et al.*, 2008).

Using these analysis methods, the relationship between personal characteristics (like personality, cognitive abilities and expertise) and performance is investigated for complex tasks, as well as between task characteristics (e.g. the complexity of a task) and performance. These kinds of relations can be used to give adequate support, for example if a correlation would be found between heart rate and performance, a high heart rate would indicate that another human should be signaled to support the performance of the task.

This paper is organized as follows. In Section 2 the experimental setup is described. Section 3 describes the statistical non-temporal analysis whereas Section 4 descries the temporal analysis. Finally, Section 5 is a discussion.

2. Experimental Setup

2.1 Participants

In this study 31 people participated (18 males, 13 females), of which 25 students from the Vrije Universiteit. 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 participant with the best score on the experiment-task.

2.2 The Task

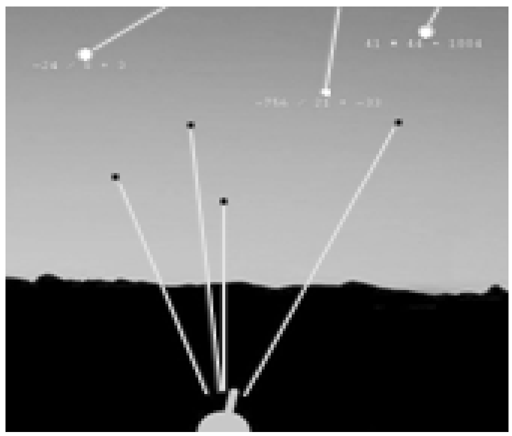

Figure 1. A screenshot of the experiment-task.

In the experiment the main task was a shooting game where the goal was to get as many points as possible. The task was performed on a 24” monitor displaying a full-screen application at 1900x1200 pixels. During the execution of the task the participant was wearing headphones to allow auditory feedback and to minimize distraction. A partial screenshot of this task is displayed in Figure 1. The object at the bottom of the screen represents the participant’s (stationary) weapon. Contacts (allies and enemies in the shape of a purple dot with a radius of 5 pixels, each accompanied by a simple mathematical equation) appear at a random location on the top of the screen and fall down to a random location at the bottom of the screen. The speed at which the contacts fall varies between 50 and 100 pixels per second. The rate at which the contacts appear varies between 10 and 20 seconds in less demanding situations and between 2.25 and 4.5 seconds in demanding situations.

The identification of a contact is performed by checking the correctness of its equation, incorrect equations correspond to enemies and correct ones to allies. Points are gained by shooting down the enemies and by allowing the allies to land. The participant can shoot a missile by executing a mouse click at a specific location; at that moment, the missile will move from the weapon and explode exactly at the location of the mouse click. The speed with which the missile reaches this location is 80 pixels per second. When a contact is within a radius of 50 pixels of the exploding missile, it is destroyed. The number of points a participant receives for destroying an enemy is proportional to the proximity of the explosion and to the distance travelled by the contact, with a maximum of 10000 points. When a participant shoots an ally or when an enemy reaches the bottom of the screen 10000 points are lost. When an allied contact reaches the bottom of the screen the participant receives 1000 points.

2.3 Procedure

For the experiment a 2 factor within subjects design was used. Two different conditions within each participant were tested. Condition 1 (easy-difficult) started with 1 object present per 10 to 20 seconds (easy). After 7.5 minutes the number of objects which were presented per second (1 object per 2.25 to 4.5 seconds) was increased and more difficult equations were given (difficult). Condition 2 (difficult-easy) started with the difficult section (as the difficult section in condition 1), after 7.5 minutes the number of objects which were presented per second was decreased (as in the easy section in condition 1). Both conditions took 15 minutes in total. The 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.

First, participants had to fill 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 in order to see if performance and stress are dependent on personality. Neuroticism and extraversion were measured with the NEO-FFI. With the NEO-PI-R vulnerability (aspect of neuroticism) and ambition (aspect of conscientiousness) were measured. When participants finished the questionnaire, electrodes were attached to the participants' body such that an ECG could be measured, and the experiment was started.

Before performing the experiment-task, participants had to perform three small tests. 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). Simple RT is often used to measure general cognitive ability (Plomin and Spinath, 2002). The second test was a task where equations were presented similar to the equations in the identification task of the experiment. As in the experiment-task, participants had to choose whether the equation 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-task. The goal of the practice task was familiarize with the shooting and identification tasks in the main experiment. After practice the participants started the experiment-task with either condition 1 or condition 2. When they finished one condition, the participants filled out the NASA-TLX (Hart and Staveland, 1988) to measure their perceived workload and performance quality. After five minutes rest, the second condition was started.

2.4 Data Analysis

All participants were analyzed using a repeated measured Analysis of Variance (ANOVA), using the Huyn-Feldt correction for violations of the sphericity assumption of the variance-covariance matrix. Personality was divided in three groups, low, middle and high, based on a significantly lower score than average (1-3), an average score (4-6) and a significantly higher than average score (7-9) and was used as a between subjects factor. In all tests, the significance level was .05.

3. Statistical Results

Table 1. Interaction of heart rate and performance with personality characteristics

	Heart_Rate				Performance			
Factor	df	F	p	partial η^2	df	F	p	partial η^2
Demands	1, 28	12.8	0.00	0.31	1, 29	19.75	0	0.41
Demands * Amb	2, 26	0.79	0.47	0.06	2, 27	4.17	0.03	0.24
Demands * Neur	2, 26	4.54	0.02	0.26	2, 27	0.67	0.52	0.047
Demands * Extr	2, 26	0.75	0.48	0.05	2, 27	1.25	0.3	0.085
Demands * Vuln	2, 26	0.42	0.66	0.03	2, 27	0.78	0.47	0.054

The performance was calculated similar to the score in the task. Performance increased when participants shot an enemy (depending on the proximity of the explosion and on the travelled distance of the enemy and on how many shots they used) and when participants landed an ally. When participants shot an ally or landed an enemy, performance decreased.

Table 1 shows the results of the ANOVA, on the dependent variable performance and heart rate and their interactions with personality characteristics. When looking at performance, a difference was found between low and high task demands. When task demands were low, performance was lower (M=0.91, SD=0.13) than when task demands were high (M=0.83, SD=0.14). The interaction between performance and ambition shows that this difference disappears for participants who scored high on ambition

(Figure 2). This difference in performance is especially apparent in the condition with low task demands, where performance of participants with a high level of ambition is lower (*M*=0.81, *SD*=0.16) than that of participants with a medium (*M*=0.92, *SD*=0.12) or a low (*M*=0.98, *SD*=0.08) level of ambition.

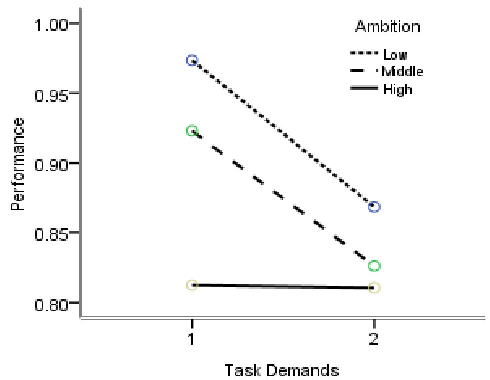

Figure 2. Mean performance as a function of task demands and level of ambition.

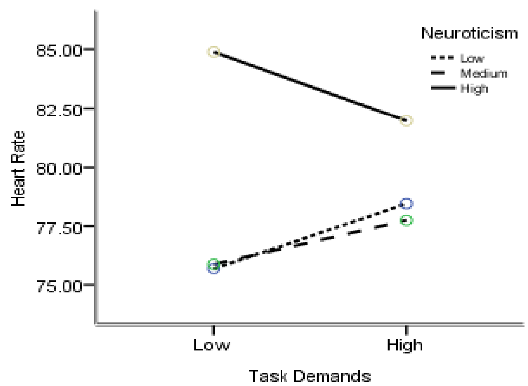

Figure 3. Mean heart rate as a function of Task Demands and Neuroticism.

The difference in heart rate between low and high task demands was significant, indicating that when task demands were low, heart rate was low (*M*=76.46, *SD*=10.47) compared to when task demands were high (*M*=78.23, *SD*=10.22). Furthermore, the significant interaction between heart rate and neuroticism points out that heart rate of participants who scored high on neuroticism was higher than heart rate of participants who scored low or medium on neuroticism. In addition, no difference in heart rate between low and high task demands was found in participants who scored high on neuroticism (Figure 3).

To further investigate the relationship between heart rate/performance and personality characteristics, a regression analysis was conducted. The interaction of neuroticism and performance was revealed in the regression analysis in the sense that neuroticism predicted performance when task demands were low (R^2=0.19, r=-0.34, $F(1, 28) = 6.61$, p<0.02). The negative correlation indicates that an increase in ambition predicted a decrease in performance quality. No significant results were found on other personality characteristics.

Table 2. Significant results for the score on the NASA-TLX on the difference between low and high task demands.

	df	F	p	partial η²
NASA-TLX Subscale				
Mental	1, 30	85.93	0.00	0.74
Physical	1, 30	45.99	0.00	0.61
Time Pressure	1, 30	156.98	0.00	0.84
Performance	1, 30	91.23	0.00	0.75
Effort	1, 30	44.75	0.00	0.6
Effort * Neuroticism	2, 28	3.022	0.07	0.18
Frustration	1, 30	54.26	0.00	0.64

In addition, it has been tested whether the equation test and mouse-RT test are a good representation of the expertise level of the participant. For this, the dependence of performance quality on these two tests has been analyzed. Also, the dependence of heart rate on expertise profile has been tested to see whether more experienced participants show a lower heart rate. No significant dependence was found between performance quality and both tests. In addition, no significant dependence was found of heart rate on the equation and mouse-RT scores.

The dependence of performance quality on heart rate has been measured to see if a higher heart rate results in lower performance. No significant relation has been found for all four parts of the two conditions or for the average performance on both low and both high parts. For the NASA-TLX, first regression analysis was conducted to see whether participant's score on the subscale "Performance' was dependent on their actual performance. The analysis revealed no significant difference, indicating that actual performance could not explain participant's score on the performance subscale of the NASA-TLX.

Furthermore, a repeated measures ANOVA was conducted to investigate the difference between low task demands and high task demands regarding the score on the NASA-TLX. In Table 2, results of the ANOVA are displayed, non-significant results excluded. The analysis revealed significant differences on all subscales between low and high task demands. On all subscales, mean score was higher when task demands were high than when task demands were low. Mean scores are represented in Table 3.

In addition, a trend in interaction was found between participant's score on Effort and their score on the personality characteristic neuroticism. This interaction shows that when participant's had a high score on neuroticism, their score on the effort scale in the condition with high task demands was high (M=18.27, SD=0.88) as compared to participants who scored low (M=13.61, SD=4.08) or medium (M=12.61, SD=2.73) on

identifier	a1:attribute	s1: {<,>}	threshold: real	x:duration	A2:attribute	s2: {<,>}	% C1	% C2	% All

neuroticism.

Table 3. Mean score on NASA-TLX as a result of task demands.

	Low Task Demands		High Task Demands	
Subscale	M	SD	M	SD
Mental Demand	5.07	3.78	12.87	3.15
Physical Demand	3.27	3.00	8.01	4.81
Temporal Demand	3.98	3.72	14.62	2.83
Performance	12	3.89	5.07	2.73
Effort	7.13	4.32	13.29	3.38
Frustration	11.87	8.75	20.81	8.62

4. Temporal Results

Besides the statistical toolkit which has been used as described in the previous section, another technique has been used to look at the temporal aspects in the measurements. Hereby, logical verification techniques have been used to perform such an automated analysis. More specific, the properties have been specified in a language called TTL (for Temporal Trace Language), (cf. (Bosse *et al.*, 2008)) 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 $\models$, comparable to the Holds-predicate in the Situation Calculus: state(γ, t) $\models$ 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 $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\forall$, $\exists$. For more details on TTL, see (Bosse *et al.*, 2008).

Table 4. Verification results

P1.1	task_level	>	300	180 sec.	heart_rate	>	0.0%	0.0%	0.0%
P1.2	task_level	>	300	180 sec.	performance_ quality	<	0.0%	0.0%	0.0%
P1.3	heart_rate	>	10% above lowest	60 sec.	performance_ quality	<	100.0%	33.%	66.7%
P1.4	heart_rate	>	20% above lowest	60 sec.	performance_ quality	<	100.0%	100.0%	100.0%
P2.1	task_level	>	-	180 sec.	heart_rate	>	33.3%	100.0%	66.7%
P2.2	task_level	<	-	180 sec.	heart_rate	<	66.7%	0.0%	33.3%
P2.3	task_level	>	-	180 sec.	performance_ quality	<	33.3%	33.3%	33.3%
P2.4	task_level	<	-	180 sec.	performance_ quality	>	100.0%	0.0%	50.0%
P2.5	heart_rate	>	-	180 sec.	performance_ quality	<	100.0%	66.7%	83.3%
P2.6	heart_rate	<	-	180 sec.	performance_ quality	>	100.0%	66.7%	83.3%
identifier	a1:attribute	s1: {<,>}	threshold: real	x:duration	A2:attribute	s2: {<,>}	% C1	% C2	% All
P1.1	task_level	>	300	180 sec.	heart_rate	>	0.0%	0.0%	0.0%
P1.2	task_level	>	300	180 sec.	performance_ quality	<	0.0%	0.0%	0.0%
P1.3	heart_rate	>	10% above lowest	60 sec.	performance_ quality	<	100.0%	33.%	66.7%
P1.4	heart_rate	>	20% above lowest	60 sec.	performance_ quality	<	100.0%	100.0%	100.0%
P2.1	task_level	>	-	180 sec.	heart_rate	>	33.3%	100.0%	66.7%
P2.2	task_level	<	-	180 sec.	heart_rate	<	66.7%	0.0%	33.3%
P2.3	task_level	>	-	180 sec.	performance_ quality	<	33.3%	33.3%	33.3%
P2.4	task_level	<	-	180 sec.	performance_ quality	>	100.0%	0.0%	50.0%
P2.5	heart_rate	>	-	180 sec.	performance_ quality	<	100.0%	66.7%	83.3%
P2.6	heart_rate	<	-	180 sec.	performance_ quality	>	100.0%	66.7%	83.3%
identifier	a1:attribute	s1: {<,>}	threshold: real	x:duration	A2:attribute	s2: {<,>}	% C1	% C2	% All
P1.1	task_level	>	300	180 sec.	heart_rate	>	0.0%	0.0%	0.0%
P1.2	task_level	>	300	180 sec.	performance_ quality	<	0.0%	0.0%	0.0%
P1.3	heart_rate	>	10% above lowest	60 sec.	performance_ quality	<	100.0%	33.%	66.7%
P1.4	heart_rate	>	20% above lowest	60 sec.	performance_ quality	<	100.0%	100.0%	100.0%

Below, the various properties that have been verified are listed, and the results thereof are shown. Hereby, formal traces have been created based upon the data obtained during the experiments. In order to reduce the computational complexity, the traces represent 60 seconds as one time point, whereby the averages of the observed values during this period have been used. First, a number of generic properties are expressed, after which they have been checked with specific values filled in. The first property specifies that if a certain attribute has a value above or below a certain threshold for a particular duration, then the value of another attribute will be positively or negatively influenced. This is expressed in property P1.

P1(a1:attribute, s1:{<, >}, threshold:real, x:duration a2:attribute, s2:{<,>})
For all time points t, if between time point t and time point t + x the value of attribute a1 is {less than, greater than} a specified threshold, and the value of a2 at time point t is v1, and the value at time t + x is v2, then v2 is {less than, greater than} v1.

Formal:
∀γ:TRACE, t:TIME, v1, v2:real[[∀t':TIME,v3:REAL[[state(γ, t') |= has_value(a1, v3)
& t' ≥ t & t' ≤ t + x] ⇒ v3 *s1* treshold] & state(γ, t) |= has_value(a2, v1)
& state(γ, t + x) |= has_value(a2, v2)]⇒ v2 *s2* v1]

The property P1 seems quite abstract, but when filling in the variables becomes quite intuitive. For example, a property stating that a task level above a certain threshold (e.g. 400) for a certain duration (e.g. 180 seconds) results in a decreasing performance quality is represented as follows: *P1(task_level, >, 400, 180, performance_quality, <).* In Table 4 an overview is given of the variants of the property above that have been verified against the empirical traces obtained from the experiment (identified with P1.x). Hereby, the first property P1.1 expresses that a high task level (for 3 consecutive minutes above 300) results in an increasing heart rate. This is seen in none of the traces. The reason for this is that as the task level is high an initial increase is seen in the heart rate, but this does not continue for the whole period. Eventually the heart rate stabilizes at a particular level. The same can be said for the second property (P1.2) which specifies that the performance quality goes down during a period of high task level. This property is also never satisfied for the same reason: the performance quality eventually stabilizes as well. Property P1.3 represents the fact that a heart rate above a certain level (in this case 10% above the lowest value measured) entails a decreasing performance quality. This property is satisfied in all traces for condition 1, whereas it is satisfied in only 33.3% of the traces belonging to condition 2. When the heart rate threshold is increased to 20% above the lowest value (P1.4) the property is satisfied for all traces.

Besides a property expressing a certain value being above a particular threshold, a property has also been expressed which checks the influence of temporal trends in certain attributes upon other attribute values. Below, the specification of this property P2 is given:

P2(a1:attribute, s1:{<, >}, x:duration, a2:attribute, s2:{<,>})
For all time points t, if between time point t and time point t + x the value of attribute a1 is {less than, greater than} the value of the attribute at the previous time point, and the value of a2 at time point t is v1, and the value at time t + x is v2, then v2 is {less than, greater than} v1.

∀γ:TRACE, t:TIME, v1, v2:real [[∀t':TIME, v3, v4:REAL[[state(γ, t')|=has_value(a1, v3)
& state(γ, t'-1) |= has_value(a1, v4)
& t' > t & t' ≤ t + x] ⇒ v3 *s1* v4]]
& state(γ, t) |= has_value(a2, v1)

& state(γ, t + x) |= has_value(a2, v2) $\Rightarrow$ v2 *s2* v1]

For example, a relation such as the fact that an increasing task level for 180 seconds results in an increasing heart rate can be expressed as follows: *P2(task_level, >, 180, heart_rate, >)*. Table 4 also shows the result of verification of various variants of the properties that have been checked (identified with P2.x). The relationship between the task level increasing (for 3 minutes) and the heart rate going up (P2.1) holds in 33.3% of the traces for condition 1, and 100% of the traces of condition 2. The latter has to do with the fact that the task level does not increase for such a long period in condition 2 (as the condition starts with high task level immediately). The relation between the task level going down, and the heart rate going down simultaneously (P2.2) is satisfied in 66.7% of the traces for condition 1, and in none of the traces of condition 2. The fact that it is satisfied in condition 1 for quite some traces can be explained by the fact that the task level in principle does not decrease for longer periods as the task level goes from low to high. Property P2.3 expresses that an increasing task level results in a decreasing performance quality. This property only holds in 33.3% of the trace, equally divided among traces for condition 1 and traces for condition 2. The opposite (P2.4: task level going down resulting in performance quality going up) holds in 0% of the traces of condition 2. In condition 1 the property is satisfied all the time since the task level does not decrease in the condition. The relation between the heart rate and the performance quality is promising: if the heart rate goes up the performance quality indeed goes down (P2.5, 83.3% of the traces), and if the heart rate goes down, the performance quality goes up (P2.6, 83.3% of the traces). Hereby, condition 1 gives the best results with an accuracy of 100% versus 66.7% for the traces concerning condition 2.

5. Discussion

The aim of this paper is to see whether a set of measurements can predict the performance accurately during a complex task. For this goal, an experiment has been conducted in which participants performed a shooting task. During the experiment, different measurements have been performed. Performance was expected to be dependent on personal characteristics (e.g. personality, expertise), physiological measurements (heart rate) and task complexity.

Results confirm that indeed task demands can predict performance. Both statistical analysis and temporal analysis (property 2.3 and 2.4) pointed out that performance decreased when task demands were high. This effect was also found in the subjective performance and effort measured with the NASA-TLX, consistent with earlier findings (e.g. Rubio et al., 2004). An unexpected result was the fact that ambitious people perform worse with low task demands than people who are less ambitious. This effect may be due to the more boring character of the low task demand parts: ambitious people could be bored easily and are therefore not motivated to perform optimally. This is contrary to the findings of Rose et al., 2002 who found no relation of ambition with performance in vigilance tasks. In future work this effect should be studied more closely.

In addition, statistical analysis shows that the average heart rate is a good predictor for the average task demands for both low and high task complexity, as heart rate goes up when task complexity increases. These results confirm that heart rate increases when more work pressure is experienced (confirming the observations found in Roscoe 1992). People who score high on neuroticism (the tendency to experience more negative emotions), do not show this effect as heart rate is high in both low and high task complexity. However, as the NASA-TLX points out, people who are more neurotic do contribute more effort to the task than people who are less neurotic when complexity is high.

In this paper, the strength of temporal analysis is shown. Not only does the temporal analysis confirm the effect of task complexity on both performance and heart rate (properties 2.1 to 2.4), it also reveals a more temporal relation between heart rate and performance. This indicates that heart rate goes up when performance goes down (properties 2.5 and 2.6), although statistical analysis fails to show such a relation. Results show the temporal character of work pressure, also emphasized by Robert & Hockey (1997) and Wilson & Russell (2003). As the temporal aspect has good predictive possibilities, it can be used as a means to provide optimal support in demanding circumstances.

The measurements that were taken to represent expertise profile, namely the equation- and mouse-RT tests are not a good representation of expertise level, since results of these tests do not predict performance quality. In addition, no relation was found between expertise profile and heart rate, which says that more experienced people do not show a lower heart rate in any part of both conditions than less experienced participants. For future work, the expertise profile needs to be redefined by possibly changing the tests so they are even more similar to the actual task.

This paper shows that it is possible to predict performance and mental/physical health from task complexity and heart rate. And although effects of personality have to be examined more thoroughly, this paper also shows that it is important to take personality into account when providing support during demanding tasks.

References

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

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

Hancock, P.A., Williams, G., Manning, C.P., and Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *Int. J. of Aviation Psy., 5(1)*, 63-86.

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

Hart, S.G., Staveland, L.E., Development of NASA-TLX: Results of empirical and theoretical research. *Human Mental Workload*, 1988.

Plomin, R., Spinath, F.M. Genetics and general cognitive ability. *Trends in Cognitive Sc. 6*(4), 369-176 (2002).

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

Roscoe, A. H. (1992). Assessing pilot workload: Why measure heart rate, HRV, and respiration? *Biological Psychology, 34,* 259–288.

Rose, C.L., Murphy, L.B., Byard, L., and Nikzad, K., The role of the Big Five personality factors in vigilance performance and workload. *European Journal of Personality* 16(3), 2002, pp. 185-200.

Rubio, S., Diaz, E., Martin, J., & Puente, J.M. (2004). Evaluation of subjective mental workload: a comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: an Int. Review* 53(1), 61-86.

Wickens, C.D. (2002). Situation awareness and workload in aviation. American Psychological Society.

Wilson, G.F., & Russell, C.A. (2003), Real-Time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors 45*, 635-644.