

Transition Behavior of Mental Load and Performance When Automation Level Changes

Z.G. Wei and P.A. Wieringa

*Man-Machine Systems Group, Laboratory for Measurement and Control
Faculty of Mechanical Engineering and Marine Technology,
Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
Email: P.A.Wieringa@wbmt.tudelft.nl*

Abstract

Human performance and mental load are investigated during a transition from a normal to an abnormal situation. The abnormal situation is simulated by a sudden loss of the automatic control of a task which is interpreted as a sudden drop in the degree of automation, DofA. The mental load perceived by the operators becomes higher due to the sudden drop of the DofA, and arrives at a peak shortly after the drop. The mental load will decrease to a low level a long period after the drop. This low level may be perceived if the operators start to operate the system at that low value of DofA. Right after the DofA drops, human performance drops considerably. As time progresses the performance recovers to a level that would have been found in a situation that the system is operated at the same low value of the DofA.

Introduction

At high degrees of automation, the role of the operator is limited to monitoring the system, rather than controlling it. However, when a failure occurs, the operator has to intervene. A failure of automated tasks can be interpreted as a reallocation of tasks. Furthermore, during this transition, the degree of automation will change from a high value directly before the failure to a low level after the failure. How will this sudden drop affect human performance? How much workload will be perceived by the operator?

To the best of the authors' knowledge, there are no reports, or experiments, that have explicitly investigated this issue. Although some researchers, such as Huey (1989), Gluckman *et al.* (1991), and Parasuraman *et al.* (1996), have investigated the effect of adaptive automation on the human operator, the transition from a high to a low DofA has not been addressed directly. Not to mention that the researches are not using a quantitative measure of the degree of automation (Wei *et al.*, 1994). It is our intention to study the behavior of human performance and mental load before and after a transition of the DofA.

According to Sheridan (1987), the human mental load may be considerably higher when the operator is required to take over manually the control of the automated subtasks due to an automation failure than that in direct manual control. During the transition, the operator may suddenly change his attention, move physically and become mentally more active to get information and to learn what is going on and what has happened. This will be a rapid transient from low to high mental load. Figure 1 displays such a hypothetical relationship between the mental load and the degree of automation. The solid curve in the upper part of the figure presents their relationship.

The mental load increases as the DofA decreases. According to the studies conducted by Wei, Macwan, and Wieringa (1997), the mental load is a linear function of the DofA for the system they studied. However, the performance may vary in the opposite direction. When the operator takes over the automated subtasks due to automation failures, the operator performance may be poorer than that if he/she controls the subtasks at the beginning. The same reasoning holds as for the mental load. After operating for a long period of time at a lower DofA, the operator gains more experience in controlling the system. The operator's mental load will decrease and the performance will recover somewhat. The solid curve in the lower part of Figure 1 presents the relationship between the performance and the DofA. We have demonstrated that the performance is a second-order polynomial of the degree of automation (Wei *et al.*, 1997).

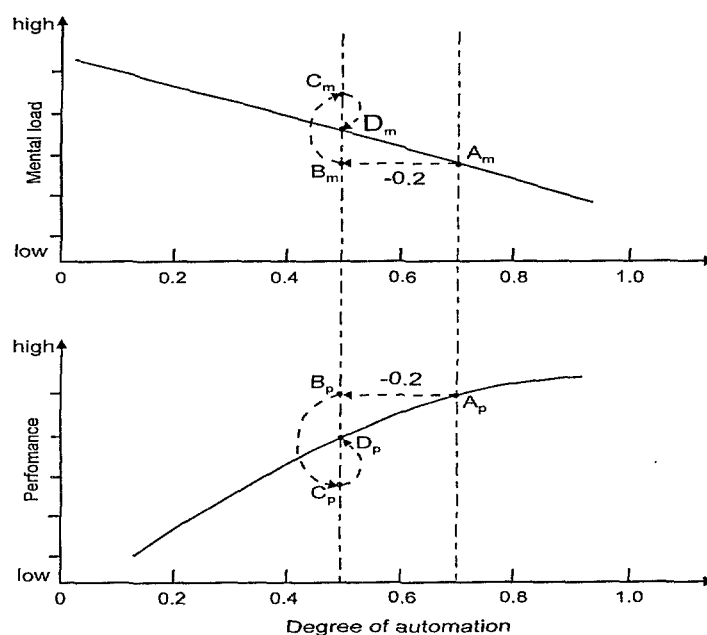


Figure 1: Hypothetical relationships between mental load, performance and degree of automation.

Based on the results on the linear system (Wei *et al.*, 1997), the experiments we conducted included a sudden drop of the DofA. The following hypotheses were tested:

1. The operator perceives a much higher mental load shortly after the degree of automation suddenly drops (from A_m to C_m in Figure 1).
2. The operator perceives a higher mental load (C_m) shortly after the DofA drops than when the system is operated for a longer time at that low DofA (D_m). Point D_m is the mental load perceived by the operator if the system is operated from the beginning at that low DofA. This means that D_m lies on the hypothetical relationship between mental load and DofA.
3. The performance reduces much when the DofA suddenly drops (from A_p to C_p).
4. The performance will recover from the low level directly after the DofA drops (C_p) to the higher level (D_p) when the system is operated for a longer time at that

low DofA. The recovered performance (D_p) can be achieved if the system is operated from the beginning at that low DofA. This means that D_p lies on the hypothetical relationship between performance and DofA.

Method

Experimental set-up

The experiment was performed using the experimental system as shown in Figure 2. The simulated system consisted of 12 first-order subsystems. All subsystems were connected in a complete forward manner with constant coupling coefficients. Each subsystem could be controlled automatically or manually as the experiment required. A proportional-integral, PI, controller was employed to execute automatic control.

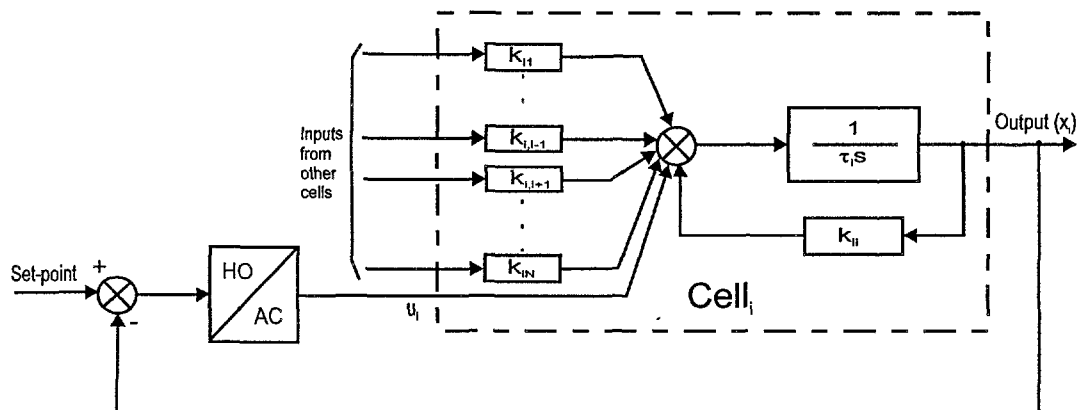


Figure 2 Cell structure in the experimental system.

The *experimental task* was to control the system to the requested Set-Point, SP. The controlled parameters were the number and the location of the automated cells, which result in different degrees of automation in the operation. The operator's control task was to generate an input for the appropriate cell(s) to bring the cell's output to the SP, and to maintain the other cells at their current set-points.

Experimental sessions

The experimental sessions were designed according to the DofAs computed in Wei *et al.* (1997). Six sessions were designed. Each session lasted 18 minutes consisting of 3 equal time intervals of 6 minutes each. Each interval had the same set of SP requests.

A session might include 2 task allocation configurations ("A" and "B" as shown in Table 1) which might have different DofAs. Within a session, the operator operated the first task allocation configuration for 9 minutes. After this period, the task allocation changed and some of the automated tasks were reallocated to the human operator. This reallocation simulated the failure of automation and induced a lower DofA.

Table 1 presents all sessions with task allocation configurations. The experiment was carried out in a sequence as presented in the table. In order to check the time effects on mental load and performance, Sessions 1 and 2 did not include a change in task

allocation. For Sessions 3 to 6, a 60% change in the DofA happened during the operation. Subjects were informed, in advance, about the dynamic task allocation, the moment when the change in task allocation happened, and the cells whose automated control will fail.

Six students (all male) from our faculty participated voluntarily as operators. They received a fixed fee for their participation plus a bonus according to their performance. Before formal sessions started, the subjects were given 1.5 hours of training.

Protocol

Figure 3 shows the time schedule and task allocation changes within a session. As can be seen, each session had 3 time intervals of 6 minutes each. The second interval was further broken into two phases with 3 minutes each. The operator operated the system with its initial task allocation configuration for 6 minutes, and then rated the Overall Mental Load, i.e. OML^I. The rating request was indicated on the screen. After the operator controlled the same configuration for another 3 minutes, the task allocation configuration changed from Configuration A to Configuration B. This reallocation was also indicated on the display. After operating Configuration B for 3 minutes, the operator was asked to rate the OML again, i.e. OML^{II}. Then, the operator controlled the system with Configuration B for another 6 minutes. At the end of the session, the operator rated the OML again, i.e. OML^{III}. The meanings of SPF₂^I, SPF₂^{II}, and oml₂^I etc. will be addressed later.

Table 1 Experimental sessions (☐: An automated cell; ☐: A manually controlled cell)

<i>Session Number</i>	<i>Task Allocation Configuration</i>	
1	A = B	☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
2	A = B	☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
3	A	☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
	B	☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
4	A	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
	B	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
5	A	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
	B	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
6	A	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐
	B	☐☐☐☐☐☐ ☐☐☐☐☐☐ ☐☐☐☐☐☐

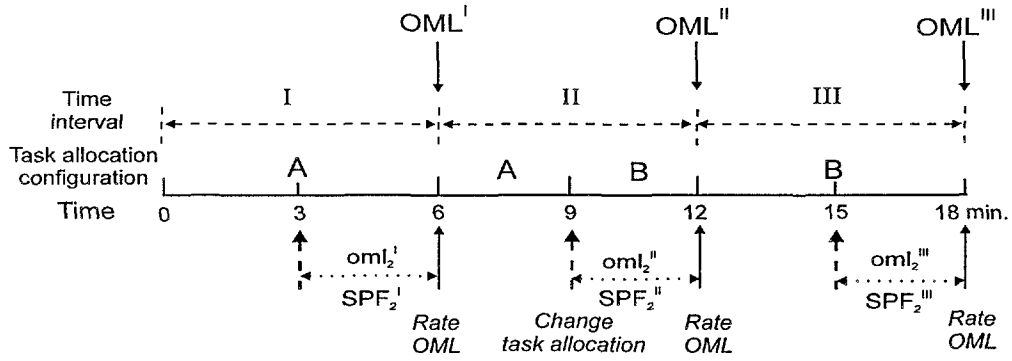


Figure 3: Time intervals and task allocation schemes for a session.

A, B: Task allocation configuration. I, II, and III: Time interval.

OML^I : OML during Interval I;

oml_2^I : OML during the last 3 minutes of Interval I.

SPF_2^I : SPF during the last 3 minutes of Interval I.

Measurements

The main variables measured during the experiment for this system are System PerFormance, SPF, and OML. These measurements and calculations were done in the same way as in Wei *et al.* (1995). SPF was calculated for three time intervals. OML for each time interval was rated based on the RSME (Zijlstra, 1993).

It is well-known that the subjective mental load instrument can not measure the mental load in real time. Thus, the mental load rated at the end of Interval II reflects the mental load level for the entire interval including the transition in DofA. It is plausible to assume that the operators are rating the whole Interval II from the last time that the mental load was assessed. It would be better to find a method to measure the mental load during the 3 minutes after the DofA changes.

In Wei *et al.* (1997), we have demonstrated that the relationship between mental load and DofA is linear. Later, we will proof that the time effect on the rated mental load can be ignored. Assuming that during the first 3 minutes of Interval II the operators operated the same system as that during Interval I, we may assume that the OML during the first 3 minutes of Interval II, noted as oml_1^{II} , should be equal to the OML perceived during Interval I, i.e. OML^I . The OML during the last 3 minutes of Interval II is noted as oml_2^{II} . Thus, the following relation exists by noting that the length of Interval II is 6 minutes:

$$\frac{3 \cdot oml_1^{II} + 3 \cdot oml_2^{II}}{6} = OML^{II}. \quad (1)$$

Since $oml_1^{II} = OML^I$, we have:

$$oml_2^{II} = 2 \cdot OML^{II} - OML^I. \quad (2)$$

Results

We performed a repeated measures analysis of variance, or ANOVA, (Norusis/SPSS Inc., 1993) for the experimental data. In the analysis, if the probability, p , is smaller than 0.05, we reject the null hypothesis.

Data for performance and mental load

Since at the beginning of the last 3 minutes of Interval II in Sessions 3 to 6 a reallocation of tasks was initiated, we focus our analysis on this period for these sessions. presents the average values of the SPF across six subjects during the last 3 minutes of three intervals, i.e. SPF_2^I , SPF_2^{II} , and SPF_2^{III} as indicated in Figure 3.

For the OML on the last 3 minutes, Table 2 lists the average values of oml_2^I , oml_2^{II} , and oml_2^{III} . As discussed before, we assume that the $oml_2^I = OML^I$ and $oml_2^{III} = OML^{III}$. The oml_2^{II} was calculated using Eq. (2). In order to present the data in a way that a larger value of the system performance indicates a better performance, the performance is presented as ten minus the system error. The ten is used to prevent the SPF from becoming a negative value. Thus, the maximum performance will not exceed 10.

Effect of time on mental load and system performance

The operator controlled the system for 18 minutes. Session 1 and Session 2 were used to check the time effect on the operator's mental load and the system performance. Table 3 shows the ANOVA results in analyzing the significance for the effect of time on SPF and OML. From Table 2, and Table 3, we observe that as the time went on, the mental load perceived by the operators for Session 1 and Session 2 increased, but not significantly, $F(2, 4) < 7.25$, $p > 0.05$. For Sessions 1 and 2 the system performance had not decreased significantly with time, $F(2, 4) < 0.9$, $p > 0.05$.

Table 2: Overall mental load and system performance during the last 3 minutes in Sessions 1 to 6

Session No	Time Interval measured	Task Allocation	DofA (TDL	oml ₂		SPF ₂	
				mean	SD	mean	SD
1	3~6	←←←← 999999	0.85	35.50	16.9	8.77	0.18
	9~12	←←←← 999999	0.85	36.17	20.4	8.52	0.66
	15~18	←←←← 999999	0.85	38.33	18.5	8.76	0.12
2	3~6	←←←← ←999←	0.17	73.00	15.8	7.60	0.55
	9~12	←←←← ←999←	0.17	86.33	28.7	7.67	0.49
	15~18	←←←← ←999←	0.17	82.50	20.1	7.51	0.46

Session No	Time Interval measured	Task Allocation	DofA (TDL	oml ₂		SPE ₂	
				mean	SD	mean	SD
3	3~6	=====	0.85	29.67	15.9	8.90	0.10
	9~12	=====	0.17	98.00	27.7	7.06	0.49
	15~18	=====	0.17	78.33	17.7	7.76	0.33
4	3~6	=====	0.67	69.50	23.8	8.85	0.13
	9~12	=====	0.16	108.1	12.2	7.29	0.37
	15~18	=====	0.16	91.33	19.2	7.39	0.44
5	3~6	=====	0.53	71.33	16.9	8.89	0.20
	9~12	=====	0.000	99.00	22.3	4.46	1.91
	15~18	=====	0.000	98.83	14.8	6.21	0.39
6	3~6	=====	0.70	48.50	16.1	8.90	0.17
	9~12	=====	0.28	103.8	37.9	7.76	0.37
	15~18	=====	0.28	83.33	15.5	7.54	0.75

Mental load and DofA

As mentioned before, the DofA dropped at least 60% halfway Interval II in Sessions 3 ~ 6. The absolute drop in DofA was larger than 0.4. When the DofA suddenly dropped from a high level to a low level the OML changed from a low level to a high level. In Sessions 3 ~ 6, subjects perceived a significantly higher mental load during Interval II than that during Interval I, $F(1, 5) > 9, p < 0.05$. The calculated oml_2^{II} together with OML^I and OML^{III} are illustrated in Figure 4. The ANOVA showed that oml_2^{II} was significantly higher than OML^I for all sessions, $F(1, 5) > 15.5, p < 0.01$. We conclude that Hypothesis 1 is confirmed by the experiment.

Table 3: ANOVA for OML and SPF in three time intervals of Sessions 1 and 2

Session	Variables	ANOVA Results	
		$F(2,4)$	p
1	Overall mental Load	1.57	0.32
	System performance	0.37	0.72
2	Overall mental Load	7.25	0.05
	System performance	0.88	0.50

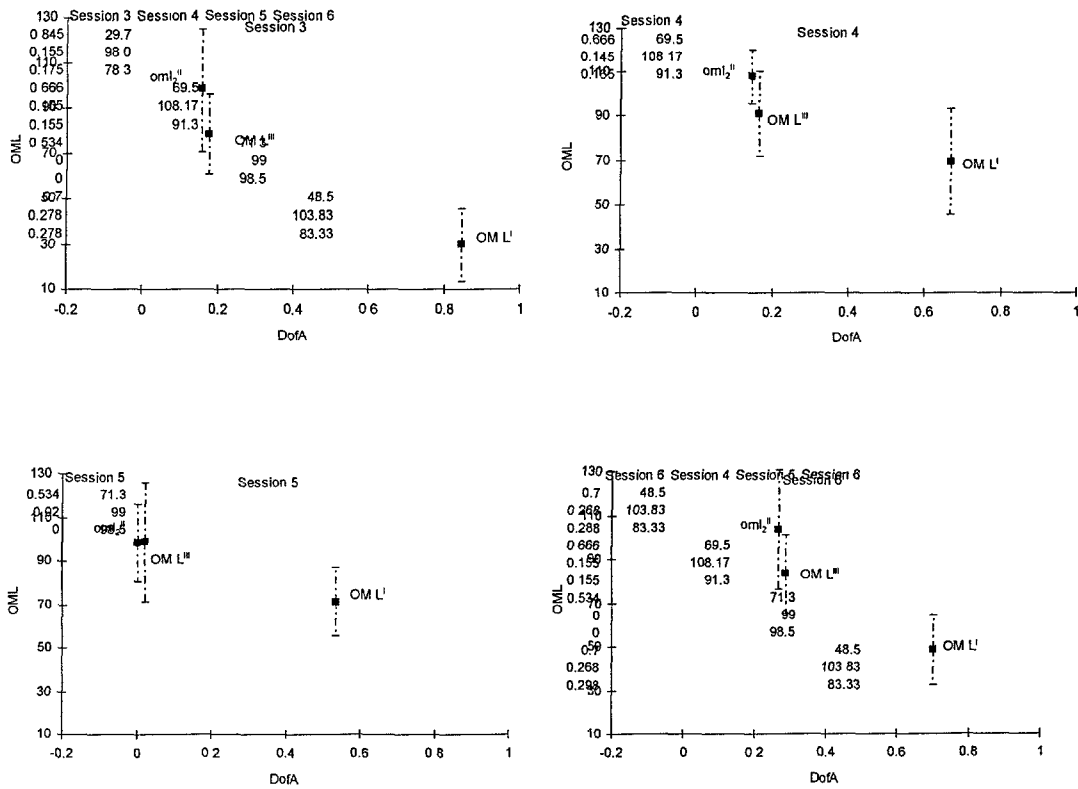


Figure 4: Variation of mental load due to a sudden drop of DofA.

Although oml_2^{II} and OML^{III} are separately plotted, they have the same DofA values.

As discussed above, when DofA dropped from a high level to a low level, the OML changed from a low level to a high level. After this change, would the OML become lower or higher? That is to ask whether the OML is higher right after the DofA drops than the OML when the operation has become stable, i.e. Hypothesis 2. As shown in Figure 4, the OML during 3 minutes after the DofA dropped, oml_2^{II} , was higher than the OML assessed 9 minutes after the drop, i.e. OML^{III} . The ANOVA test showed that oml_2^{II} was significantly higher than OML^{III} , $F(1, 5) > 5.5$, $p < 0.05$, except Session 5, $F(1, 5) = 3.38$, $p = 0.125$. Thus, a trend can be found that the OML^{III} intends to be lower than oml_2^{II} . Since the time effect can be neglected in this study, it is plausible to assume that OML^{III} is at Point D_m in Figure 1. If so, we can conclude that Hypothesis 2 is also confirmed by the experiment. This is discussed as follows.

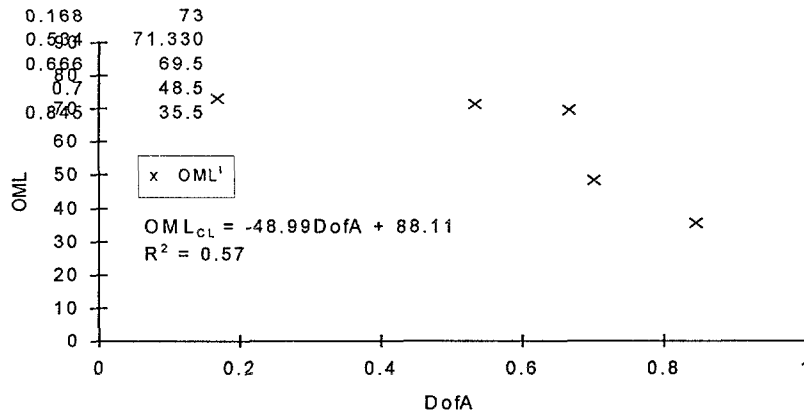


Figure 5: Relationship between overall mental load (OML^I) and DofA.

The relationship between the OML and the DofA in the case that the experimental system was operated without dynamic task allocation is plotted in Figure 5, i.e. the relationship between OML^I and DofA. According to the study in Wei *et al.* (1997), a linear polynomial is fitted to the 5 data points as presented in Figure 5.

Based on this function, the OML could be calculated for low DofA values. The low DofA values are the values during Interval III as shown in Table 2. The calculated OML, noted as OML_{CL} , and the OML^{III} for Sessions 3 to 6 are presented in Figure 6.

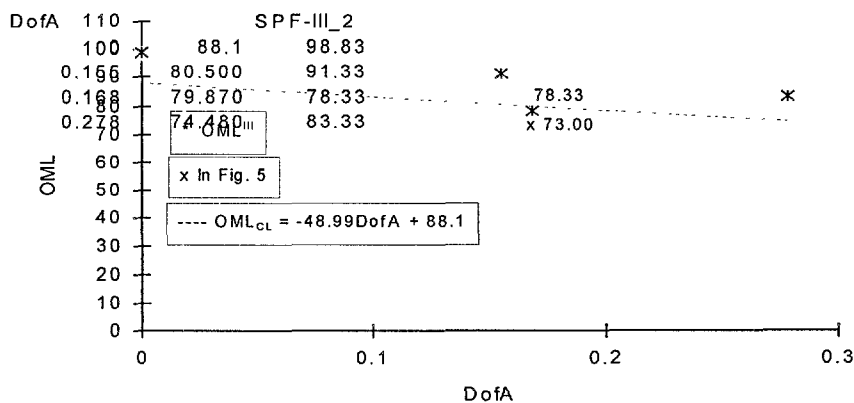


Figure 6: Comparison between the calculated OML_{CL} and the measured OML^{III} .

x: An OML^I point in Figure 5, when DofA = 0.168, $OML^I = 73.0$.

The *t-test* shows that the measured OML^{III} in Sessions 3 to 6 are not significantly different from the calculated OML at the specific values of the DofA, $p > 0.05$. We conclude that the OML, a longer period after the DofA drops to a lower value, can decrease to a level that would have been found if the operator would have been operating the system at that lower DofA from the beginning, i.e. D_m in Figure 1. So, Hypothesis 2 is proved by the experiment.

System performance and DofA

In this experiment, we investigated how the system performance changes when the DofA is suddenly changed. The hypotheses for the system performance are

Hypotheses 3 and 4. The system performance is analysed based on the SPF in the last 3 minutes of all three time intervals. These performances are compared in Figure 7.

The SPF_2^I is significantly better than SPF_2^{II} as expected ($p < 0.01$), and this confirms Hypothesis 3. SPF_2^{III} in Sessions 3, 4, and 5 is larger than SPF_2^{II} (not significantly, $F(1, 5) < 6.0, p > 0.05$). SPF_2^{II} and SPF_2^{III} in Session 6 have no significant difference, $F(1, 5) = 0.49, p = 0.52$. Based on the above analysis, we conclude that the SPF after a longer time of operation with a lower DofA intends to recover somewhat from the SPF directly after the DofA dropped.

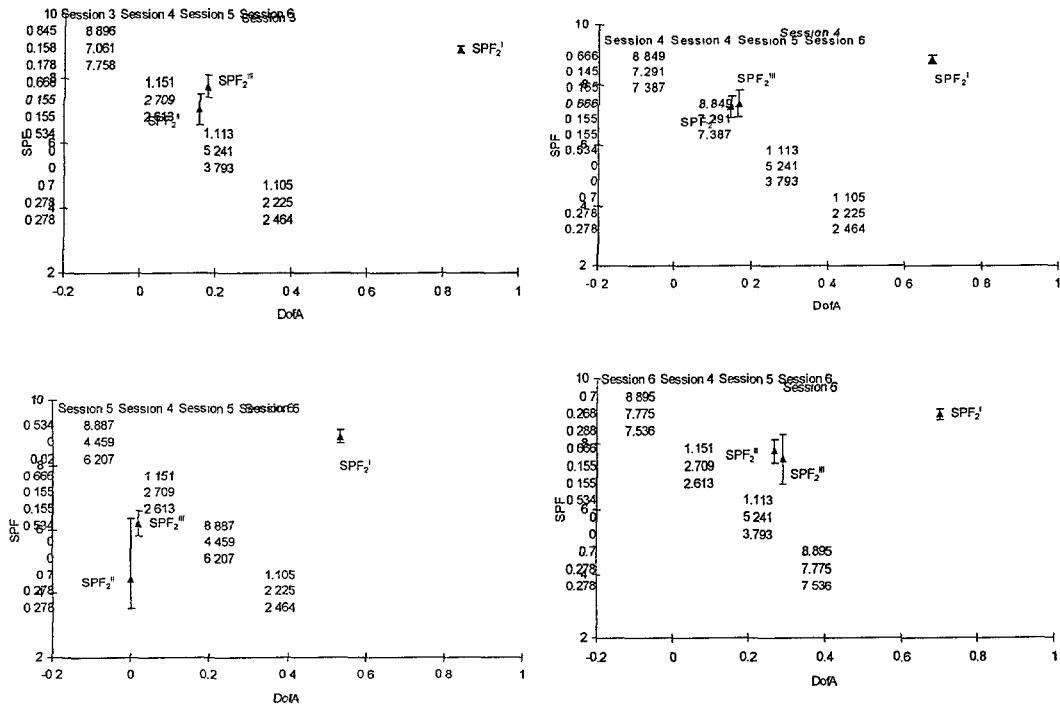


Figure 7: System performance during the last 3 minutes in each time interval. Although SPF_2^{II} and SPF_2^{III} are separately plotted, they have the same DofA values.

The relationship between SPF and DofA in the case that the experimental system was operated without dynamic task allocation is plotted in Figure 8, i.e. the relationship between SPF_2^I and DofA. A second-order polynomial, SPF_{CL} , is fitted to the 5 data points as presented in Figure 8.

Based on this function, the SPF could be calculated for low DofA values which are the values during Interval III as shown in Table 2. The calculated SPF, noted as SPF_{CL} , and the SPF_2^{III} for Sessions 3 to 6 are presented in Figure 9.

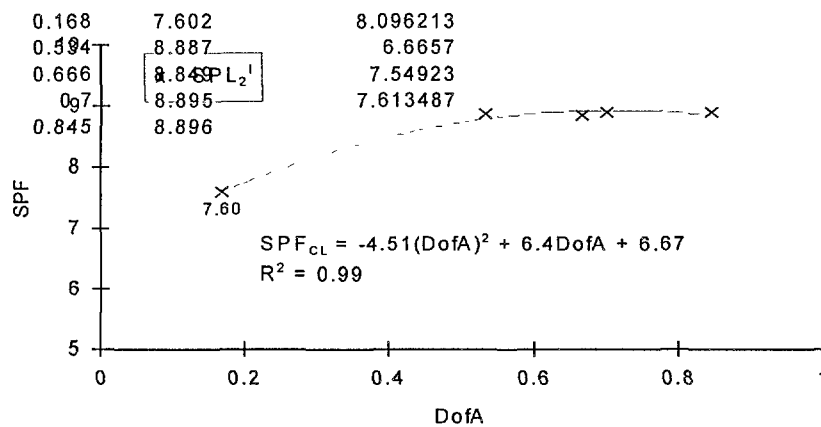


Figure 8: Relationship between performance (SPF_2^I) and DofA.

The *t*-test shows that the measured SPF_2^{III} in Sessions 3 to 6 are not significantly different from the calculated SPF at the specific values of the DofA, $p > 0.05$. We conclude that the SPF after a drop to a lower DofA can recover to a level that would have been found if the operators would have been operating the system at that lower DofA from the beginning, i.e. Point D_p in Figure 1. So, Hypothesis 4 is proved by the experiment.

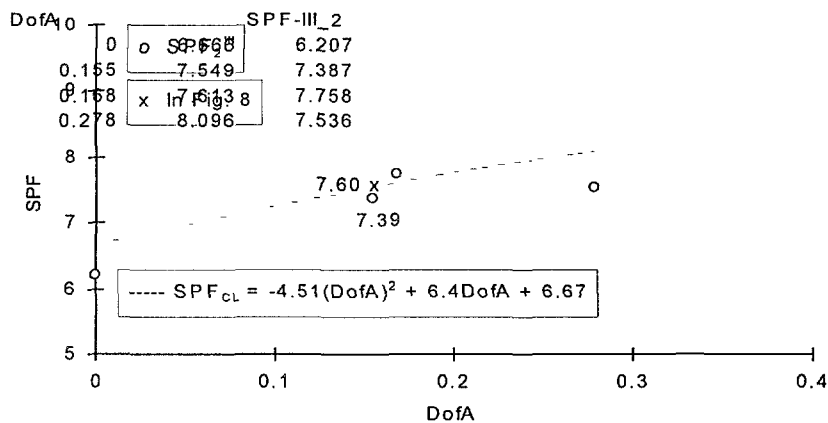


Figure 9: Comparison between the calculated SPF_{CL} and the measured SPF_2^{III} . x: An SPF_2^I point in Figure 8, when DofA = 0.17, $SPF_2^I = 7.60$.

Discussion

Many researchers have investigated the influences of automation on the human operator in situations such as: Dynamic task allocation, fault management, human interference with automation, and human use of automation (Gluckman *et al.*, 1991; Huey, 1989; Kim and Sheridan, 1995; Riley, 1994). However, the characteristics of human performance and mental load during a transition of the task allocation have not received much attention. It is hypothesised that after a transition from a DofA to a lower DofA, human operators will experience a high mental load and their performance may degrade.

The operators in our experiments did not carry out fault diagnosis and decision making tasks. The operators only needed to take over the tasks of which the

automatic control systems failed while the failure was clearly indicated. Therefore, we focus on the effect of a drop in DofA on human performance and mental load.

The experimental tasks used in this study had a low task criticality, meaning that the operator had not to worry about a large effect of the failure on system performance. System safety was not considered. Thus, the increase in mental load was affected by a change in monitoring and control of the tasks.

The performance reduced largely shortly after the DofA dropped, while after some time, the performance recovered to some degree. This is because directly after the DofA dropped, the operators faced an other system: They had to perform more tasks. After they operated this other system for a longer period of time, the operators accumulated experience so that they could improve the performance. Shortly after the DofA dropped, the operators had to change suddenly their attention and to move mentally to perform more tasks. So, the operators perceived the highest mental load right after the transition. After having gained more experience, the operators could invest less mental effort and therefore they perceived lower mental load which was still higher than that they perceived before the reallocation. The mental load increased probably because more manual tasks had to be performed.

Directly after the DofA dropped, the performance decreased to its lowest value and the mental load reached its peak. This reveals that within a short period after the DofA drops, the operators need more support from other resources, such as operator support systems, operator diagnostic tools, and from other human operators. Moreover, a dynamic task allocation, or a scheduled human intervention, is necessary for the operators to have on-line training in order to maintain their adaptability to take over the automated tasks.

The mental load immediately after the DofA dropped could not directly be measured in this study, but was calculated. To compare the mental load right after the DofA drops and when the operation becomes stable, further research is necessary using measurement instruments that can measure mental load in real time, e.g. physiological measurements.

Conclusions

This paper investigates the behavior of performance and mental load when the degree of automation suddenly drops. Based on the study, the following conclusions can be drawn:

1. When the degree of automation in supervisory control suddenly drops a much higher mental load is perceived by the operators within a short period after the drop. However, the perceived mental load will decline to a low level after the system is operated at the lower level of DofA for a longer period. The low level of mental load may be perceived if the operators start to operate the system at that low level of automation.
2. Within a short period after the degree of automation drops, the performance degrades to a low level. After the system is operated for a longer period at the low level of automation, the performance will restore to a higher level. This level can

be achieved if the operators start to operate the system at that low level of automation.

References

- Gluckman, J.P., J.G. Morrison and J. E. Deaton (1991). Complex task performance as a basis for developing cognitive engineering guidelines in adaptive automation. In: *Proc. of the Human Factors Society 35th Annual Meeting*, pp.116-120.
- Huey, B.M. (1989). *The effect of function allocation schemes on operator performance in supervisory control systems*. Unpublished Doctoral dissertation, George Mason University, USA.
- Kim, S. and T. B. Sheridan (1995). Human intervention into automatic decision-making and automatic control. In: *Proc. of the 6th IFAC/IFIP /IFORS /IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, pp.357-361, MIT, MA, USA.
- Norusis, M.J. /SPSS Inc. (1993). *SPSS for Windows- Base System User's Guide*. Release 6.0.
- Parasuraman, R., M. Mouloua, and R. Molloy (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, **38** (4), pp.665-679.
- Riley, V.A. (1994). *Human use of automation*. Unpublished Doctoral dissertation, University of Minnesota, MI, USA.
- Sheridan, T.B. (1987). Supervisory control. In: G. Salvendy (Ed.), *Handbook of Human Factors*, pp.1243-1263. Wiley, New York.
- Wei, Z.G., A.P. Macwan, J.H.M. Andriessen and P.A. Wieringa (1994). Defining degree of automation. In: P.T. Kidd and W. Karwowski (Eds.), *Advances in Agile Manufacturing*, pp. 351-354, IOS Press, Amsterdam.
- Wei, Z.G., A.P. Macwan and P.A. Wieringa (1997). A Quantitative Measure for the Degree of Automation and its Relation to System Performance and Mental Load. Accepted for publication in *Human Factors*.
- Wei, Z.G., A.P. Macwan and P.A. Wieringa (1995). Experimental investigation on mental load and task effect on system performance. In: *Proc. of the 6th IFAC/IFIP /IFORS /IEA Symposium on Analysis, Design and Evaluation of Man-Machine Systems*, pp.329-334, MIT, MA, USA.
- Zijlstra, F.R.H. (1993). *Efficiency in Work Behavior*. Ph.D. thesis, Delft University of Technology, Delft University Press, ISBN 90-6275-918-1, Delft, The Netherlands.