1. Even highly automated systems are based on incomplete models of the world. So, human supervision will always be necessary.

_Zelfs hoog geautomatiseerde systemen zijn gebaseerd op incomplete modellen van de wereld. De menselijke supervisie zal dus altijd nodig zijn._

2. A well-performing operator concentrates not only on the non-automated tasks, but also on the automated tasks _This thesis_.

_Een operator die goed presteert concentreert zich niet alleen op de niet-geautomatiseerde taken, maar ook op de geautomatiseerde taken_ (Dit proefschrift).

3. Knowing your enemy and knowing yourself, you can always win a battle (A translation of the Chinese proverb: 知彼知己，百战百殆). This can also be applied to the human operators in human-machine systems.

_Als je je vijand en jezelf begrijpt kan je altijd een strijd winnen_ (Vertaling van het Chinese gezegde: 知彼知己，百战不殆). Dit kan ook worden toegepast op operators in mens-machine systemen.

4. If there is a perceived problem, there is probably a real problem.

_Als er een probleem wordt ervaren dan is er waarschijnlijk een echt probleem._

5. The discussion whether human beings can always command the computer or someday, the computer will command human beings is not ridicule.

_De discussie of de mens altijd de computer kan besturen of dat er een dag komt dat de computer de mens bestuurt, is niet belachelijk._

6. Employment and products are the two most important outputs an organization can contribute to the society.

_Werkgelegenheid en produkten zijn de twee belangrijkste bijdragen van een organisatie aan de maatschappij._
Stellingen (Propositions)

behorende bij het proefschrift

Mental Load and Performance at Different Automation Levels

(Mentale Belasting en Operator Prestaties bij Verschillende Graden van Automatisering)

Zhi-Gang WEI

Delft, The Netherlands, December 1997
7. Love can only sprout up from heart, but cannot be made.

*Liefde kan alleen uit het hart ontstaan, maar kan niet worden gemaakt.*

8. Having many football stars, the Netherlands cannot guarantee its team to win the World Cup championship.

*Het bestaan van vele voetbalsterren garandeert niet dat het Nederlandse team de World Cup wint.*

9. WWW really means Wait, Watch, and Worry.

*WWW betekent echt Wait, Watch, and Worry.*

10. If you don’t take time to relax, you will enjoy sickleave.

*Als je geen tijd neemt om te ontspannen, zal je tijd krijgen om ziekteverlof te genieten.*
Mental Load and Performance at Different Automation Levels
Mental Load and Performance at Different Automation Levels

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr.ir. J. Blaauwendraad,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College voor Promoties aangewezen,
op dinsdag 27 januari 1998 te 10:30 uur

door

Zhi-Gang WEI

Bachelor of Engineering
Huazhong University of Science and Technology, Wuhan, China

geboren te Songzi, Hubei, China
Dit proefschrift is goedgekeurd door de promotor:
Prof. dr.ir. H.G. Stassen

Toegevoegd promotor:
Dr. ir. P.A. Wieringa

Samenstelling promotiecommissie:

Rector Magnificus
Prof. dr.ir. H.G. Stassen
Dr. ir. P.A. Wieringa
Prof. dr. N. Moray
Prof. dr. J.H.T.H. Andriessen
Prof. ir. P.J. de Jong
Prof. dr. G.J. Olsder
Dr. F.R.H. Zijlstra
Prof. dr. ir. D. van Willigen

Technische Universiteit Delft, voorzitter
Technische Universiteit Delft, promotor
Technische Universiteit Delft, toegevoegd promotor
University of Surrey, UK
Technische Universiteit Delft
Technische Universiteit Delft
Technische Universiteit Delft
Katholieke Universiteit: Brabant
Technische Universiteit Delft (reserved member)

Mental Load and Performance at Different Automation Levels/
Zhi-Gang WEI
Thesis Delft University of Technology - with summary in Dutch
ISBN 90-370-0164-5
Subject headings: Automation, mental load, human-machine systems, supervisory control, task allocation

Cover design by Paul Breedveld.

All rights reserved. No part of this book may be reproduced by any means, or transmitted without the written permission of the author. Any use or application of data, methods and/or results etc., occurring in this report will be at the user's own risk.
To my wife Ming,
my parents, and
my son Wu
Acknowledgment

The study presented in this thesis was conducted at the Man-Machine Systems and Control Group, Laboratory for Measurement and Control, Faculty of Mechanical Engineering and Marine Technology, Delft University of Technology, the Netherlands.

I started this study as a Ph.D. student in 1993. During this study, I got great supports from my colleagues in Man-Machine Systems and Control Group. I would like to thank all the people who made the realization of this thesis possible. It is impossible to mention all the names. Nevertheless, I would like to express my sincere gratitude to the following contributors:

First of all, my “Promotor”, Prof. dr. ir. Henk G. Stassen who supervised my study strategically and supported me actively in revising, advising, and criticizing the draft versions of this thesis with great endurance to my writing style.

My “Toegevoegd promotor”, Dr. ir. Peter A. Wieringa who supervised this study continuously and in depth. His supervision, support, and stimulation made this study progress. This thesis would not have been completed without his patient guidance, fruitful advice and critical criticism.

Dr. Anil P. Macwan who played an important role during the first half period of my study. His encouragement, constructive suggestion, and careful comments on my various manuscripts made important contributions to this study.

I am grateful to Prof. N. Moray, and Dr. F.R.H Zijlstra for their comments on my study. I am also grateful to Dr. B.M. Huey (NCF, USA) for her sending me the code of the pasteurization plant.

I am indebted to ir. A.R. Paauw, MSc. student, for his interest in this study and his contributions in transplanting the pasteurization plant and doing experiments with the plant. Chapter 6 was written on the base of his master degree thesis. I am also indebted to ir. J.H.M. Andriessen for his advice in developing linear system.

My gratitude is extended to: Ir. L. Brinkman for his managerial work regarding the experiments; ir. B. Wenneker and ing. P. Valk for their support in computer hardware and software; mr. P. Ruinard for helping me to collect tremendous number of articles; drs. J. van Dielen for his support in my e-mail communications; our laboratory staff: ms. E. Arkesteijn, ms. M. Macherhi, mr. A. Gutteling, and mr. R. Overbeek.
I would never have been able to complete my thesis without support from my friends. I am pleased to give my thanks to my country fellows who studied, or are studying, at Delft, and made me happy in my leisure time.

I certainly owe my parents in China a great debt as I have left them many years. Last, but not least, I am greatly indebted to my wife, Ming, for her love, endurance, encouragement, and understanding, as well as her sacrificing her own career to support me during my study. I am also indebted to my son, Wu, for his tolerance to my frequently ignoring of his requirements when I devoted to my study.

Delft, The Netherlands  Zhi-Gang Wei
December 1997
Contents

Acknowledgment vii

Chapter 1 Introduction 1
  1.1 General 1
  1.2 Effects of an increase in the level of automation 3
    1.2.1 Positive effects of automation 3
    1.2.2 Negative influences of an increasing level of automation 3
  1.3 Definitions of basic concepts 6
  1.4 Purpose of the thesis and problem definition 7
  1.5 Outline of the thesis 9

Chapter 2 Task Allocation as a Tool in Human-Machine Systems Design 11
  2.1 Introduction 11
  2.2 Task allocation between humans and machines 13
    2.2.1 Significance of task allocation in human-machine systems design 13
    2.2.2 Task allocation principles: Their benefits and weaknesses 15
  2.3 Review of studies on task allocation methodologies 19
  2.4 Conclusions 22

Chapter 3 Evaluation Measures for Task Allocation 23
  3.1 Introduction 23
  3.2 Evaluation of task allocation 24
    3.2.1 Criteria for the evaluation of task allocation 25
    3.2.2 Evaluation approaches 26
    3.2.3 Remarks on the evaluation of task allocation 27
  3.3 Measurement of mental load 28
    3.3.1 Measurement of task mental load 30
    3.3.2 Subjective mental load measurements 32
    3.3.3 Prediction of task demand load 35
    3.3.4 Remarks on mental load assessment 37
  3.4 Summary 37

Chapter 4 A Quantitative Measure for the Degree of Automation 39
  4.1 Introduction 39
  4.2 Qualitative scales for the level of automation 40
    4.2.1 Sheridan’s scale for the level of automation 40
    4.2.2 Billings’ continuum for system control and management automation 41
4.2.3 Correspondent relation between Sheridan’s scale & Billings’ continuum 41
4.2.4 Implementation of scales of level of automation 43
4.3 Level of automation of supervisory control in industries 45
4.3.1 Level of plant automation in nuclear power plant control 45
4.3.2 Level of aircraft control automation 47
4.4 Quantitative definition of degree of automation 49
4.4.1 Motivation 49
4.4.2 Definitions 49
4.5 Approaches to estimate task weights 53
4.5.1 Task weight based on task effect on system performance 53
4.5.2 Task weight based on task demand load 54
4.5.3 Task weight based on mental load 55
4.6 Conclusions 55

Chapter 5 Determination of Degree of Automation of Linear Systems 57
5.1 Introduction 57
5.2 Method 58
5.2.1 Experimental system 58
5.2.2 Operator’s task 62
5.2.3 Experimental sessions 62
5.2.4 Subjects 62
5.2.5 System performance measurement 63
5.2.6 Estimation of weights 64
5.3 Results and discussion 67
5.3.1 Task weights based on task effect on system performance 67
5.3.2 Task weights based on task demand load 68
5.3.3 Task mental load and task weights 70
5.3.4 Computation of DofA 72
5.3.5 Relation between system performance and DofA 73
5.3.6 Relation between overall mental load and DofA 74
5.3.7 Relation between three measures of DofA 74
5.3.8 Prediction of mental load 75
5.3.9 Task mental load, system performance, and human perception 77
5.3.10 Relation between OML and system performance 78
5.3.11 RSME and NASA-TLX 80
5.3.12 Further research 80
5.5 Concluding summary 81

Chapter 6 Determination of Degree of Automation of a Pasteurisation Plant 83
6.1 Introduction 83
6.2 Method 84
6.2.1 Experimental set-up 84
6.2.2 Experimental task 88
## Contents

6.2.3 Operator's task
6.2.4 Subjects
6.2.5 Experimental sessions
6.2.6 Performance measurement
6.2.7 Estimation of task weights
6.2.8 Subjective evaluation of task allocation based on Analytic Hierarchy Process

6.3 Results and discussion
6.3.1 Task weights based on task effect on system performance
6.3.2 Task weights based on task demand load
6.3.3 Mental load and task weights
6.3.4 Computation of DofA
6.3.5 Relation between performance and DofA
6.3.6 Relation between overall mental load and DofA
6.3.7 Prediction of mental load
6.3.8 Task demand load, performance, and mental load
6.3.9 AHP subjective evaluation of task allocation decisions

6.4 Summary and concluding remarks

### Chapter 7 Effects of Sudden Loss of Automated Tasks on Human Operators
7.1 Introduction
7.2 Experiment one
7.2.1 Method
7.2.2 Results and discussion
7.3 Experiment two
7.3.1 Method
7.3.2 Results and discussion
7.4 General discussion
7.5 Conclusions

### Chapter 8 Human Monitoring Behavior of Automated Systems
8.1 Introduction
8.2 Method
8.2.1 Experimental system
8.2.2 Operator's task
8.2.3 Experimental sessions
8.2.4 Participants
8.2.5 Measurements
8.3 Results and discussion
8.3.1 Results
8.3.2 Operator's performance and interaction with the plant
8.3.3 Operator's monitoring of automation
8.3.4 Mental load from the monitoring of automation
8.4 Summary
Chapter 9  Concluding Summary  
  9.1 Conclusions from the present study  
  9.2 Implications of this study  
  9.3 Future research  

References  

Appendix 1  The original Fitts list  
Appendix 2  RSME rating form used in Chapter 5  
Appendix 3  RSME rating form used in Chapters 6 and 7  
Appendix 4  Analytic Hierarchy Process and evaluation results  
Appendix 5  A sample of pair-wise comparison tables in Chapter 6  
Appendix 6  Experimental results (Chapter 7)  
Appendix 7  Experimental results (Chapter 8)  

List of abbreviations and symbols  

Summary  
Samenvatting (Summary in Dutch)  

Curriculum Vitae
Chapter 1

Introduction

1.1 GENERAL

The level of automation in civil aviation, process industry, and manufacturing systems is continuously increasing. Higher levels of automation show an integration of supervisory control, process and system automation, production scheduling, logistics control, maintenance and system management. Automation has been employed to achieve consistent product quality and to improve production efficiency and benefits. It is also used to minimize human operator's workload, to reduce human error, and to enhance operational safety. The development of highly automated systems has been made possible as a result of the developments in computer and information technology. Although modern control systems are capable of operating a nuclear power plant or controlling an aircraft's flight, human operator activities will always be needed, ranging from the complex aircraft flight to simpler process control loops. At least in the foreseeable future, automation will not completely replace the human operator. The fact that human operators still will play an important role in the operation of an industrial plant has become a common sense among system designers, control engineers, human factors engineers, and system managers. However, the role of human operators and the nature of the operator tasks will keep changing. The role is gradually shifting from manual control to supervisory control. The operator tasks have changed from performing manual tasks, so-called doing tasks, such as activating manual switches and following an operation procedure, to performing intellectual tasks, such as diagnosis, planning and problem solving, which ask higher cognitive demands (Hollnagel, 1994). Thus, supervisory tasks increase, whereas human direct interaction with the production equipment (not including computers) decreases.

Because the human operator still will play a role in system operation, we must bear in mind that the operator may make errors during the operation. More efforts must be put into the reduction of human errors and the increase in human reliability. A higher level of automation, on the one end, may achieve this objective and obtain other benefits. On the other end, it may also have adverse effects on which Bainbridge (1983) has a comprehensive discussion. In the context of human-machine systems, automation is applied only to tasks that are fully understandable and
completely describable to system designers. The tendency has been to automate what is easiest and as much as possible and leave the rest to the operator. According to Sheridan (1992), this dignifies the human contribution and it may lead to a hodgepodge of partial automation, making the remaining human tasks less coherent, less meaningful and more complex than they need to be. Moreover, the humans may also have difficulty to perform those tasks that are not completely understood by the system designers (Rouse et al., 1992). This will result in an overall degradation of system performance. The cost of increased automation, therefore, does not necessarily result in higher benefits. Additionally, too much automation may result in poor operator performance due to too less a workload, operator’s loss of skills, absence of motivation (Bainbridge, 1983), and inadequacy of awareness of the system status (Endsley, 1995a, b).

Thus, a series of questions arise. How much automation should be used so that the human operator still feels that he is in control of the system? How can the level of automation be measured? How can automation and operator actions be balanced? How does an increase in the level of automation affect system performance, operational safety and human operator’s motivation in the control of the system? More generally speaking, today, the question is no longer whether one or another task can be automated, but, rather, whether it should be automated (Wiener and Curry, 1980).

Little is known about the effects of automation on human performance (Scherbo, 1996), although recent efforts have examined how aspects of automation affect performance by themselves and in association with other activities. It is necessary to put more efforts into investigating how the human performance is affected by automation.

In designing control centers where human operators will supervise or operate complex systems, such as aircraft cockpit, the control room of a nuclear power plant or a process plant, the allocation of tasks between human operators and machines is one of the critical steps (IEC Standard 964, 1989). These tasks include supervisory tasks as well as management tasks. The decisions on task allocation will determine the level of automation, and will influence the automation design, including the human-machine interface, organization structure, and the human jobs. An optimal allocation of tasks between human operators and machines (or computers) may balance automation and human actions in the supervision of the complex systems.

The goal of this study is to investigate the effects of a shift of the level of automation on the human operator behavior. The study focuses on the effects on human performance and human mental load. The level of automation is changed by varying task allocation between human and machine. So, the study investigates, at the same time, the effects of different task allocation decisions on the human operators.
1.2 EFFECTS OF AN INCREASE IN THE LEVEL OF AUTOMATION

In literature, the influence of an increasing level of automation on human reliability, working environment, human job design, operational safety, economic benefits, and social aspects in different complex systems has been discussed (e.g. Bainbridge, 1983; Billings, 1991; Cooley, 1984; Sheridan, 1992; Smith and Carayon, 1995; Wiener and Curry, 1980). This section presents a reduced review on the positive effects and the negative influences of the increasing level of automation.

1.2.1 Positive Effects of Automation

Automation of system functions must always be seen in relation to the functions carried out by the humans. Automation can logically serve three purposes (Hollnagel, 1994): (1) Enable the humans to do things that they could not do before, (2) substitute, by taking over system functions from the humans, and (3) amplify, allowing the human to perform more efficiently. Therefore, automation may provide, although no guarantee can be given, the following economic benefits (Smith and Carayon, 1995):

1. Lowered production costs through the use of more efficient machines,
2. reduced workforce and human labor,
3. improved product or service quality and conformity,
4. increased up-time or productive time,
5. enhanced flexibility of the production system to meet customer needs, and
6. lowered insurance costs by reducing worker risks.

Besides these economic benefits, other positive effects of automation are (Sheridan, 1992):

1. The reduced operator workload,
2. an improved task performance and reliability,
3. an increased operational and human safety, and
4. participative technology advancement.

1.2.2 Negative Influences of an Increasing Level of Automation

However, there is an emerging evidence that the increasing use of automation or inappropriate application of automation can be detrimental either to the production process or to the employees in industry. Some of the disadvantages of an increased level of automation, related to human operators and to the production process, are identified as follows:

(a) Decreasing operator's awareness of system status: The investigations conducted by Endsley and Kiris (1995) have revealed that a high level of automation in the decision making
reduces the operator’s situation awareness so that it degrades the human performance.

(b) **Unsuitable mental load and work dissatisfaction:** When a system is operated at a high level of automation, the workload imposed on the human operators is low. In this situation, the operators have too little to do and therefore they experience a very low level of workload. However, when automation fails, the operators have to take over the control of tasks that are automated before and the workload imposed on the operators may become very high. In this situation, it is easy for the operators to be overloaded. Moreover, while the human operators are happy that automation has taken away the dull monotony of repetitive manual tasks, they may not get much satisfaction from their new supervisory monitoring tasks (Sheridan, 1992).

(c) **Decreasing operator’s skills:** As a system is automated, the human operators do not have opportunities to practice their control skills. In case of automation failure, so that human intervention is required, the operators may not have enough proficiency to be able to operate the system manually.

(d) **Losing operator’s motivation in the control of the process:** If the process is run by computers, the human operators may not take their responsibilities seriously. They may feel that they are no longer responsible for what happens in the process. The computers are thought to be responsible, a rather dangerous situation! The operators will eventually lose their motivation to interact with the process. Furthermore, they may come to feel that with a high level of automation they are no longer significant contributors to the production. They develop a sense of not contributing (Sheridan, 1992).

(e) **Requiring operators to have other knowledge to operate the highly automated systems:** As the level of automation increases, the nature of tasks to be performed by the humans has also been changed, and the system may become more complex. In order to operate such a system, the human operators need more knowledge. Thus, the operators need more training. Moreover, they have to adjust to frequent technological changes, and to constant retraining.

(f) **Creating alarm inflation:** For an automated process, both the process and automatic control systems should be monitored. The alarms are induced not only by monitoring the process, but also by monitoring the automatic control systems. For a highly automated process, alarm inflation is easy to take place. The false alarm is another problem which often occurs with a high level of automation.

(g) **Increasing the system complexity:** Advanced automation technologies are becoming more and more complex. The implementation of these technologies in the process may increase the system complexity. In turn, this requires that the operators have a higher level of knowledge and capabilities to supervise the process.

(h) **Creating too much monitoring workload:** According to Sheridan (1992), automation has been motivated in some measure by desire to reduce workload. However, the requirement for human monitoring of automation itself creates workload. In some cases, people are oblivious of this.

(i) **Increasing operator’s fear of unknown and uncertainty:** As the system is highly automated, the human operators have less awareness about the working status of the system and they do
not understand the basis of what they are asked to do. The operators can become confused and distrustful if they do not know and understand the process and the allocation of function between human and machine. The introduction of new technology brings with it many fears of the unknown such as job loss, changes in the way that work will be accomplished, being able to learn how to use the new technology, the time line for being able to do the new job, and pay reductions.

(j) Overtrusting in automation: Although human trust in automation is a function of different attributes (Lee and Moray, 1992; Muir, 1994; Sheridan, 1992), the operator’s overtrusting in automation has been a problem due to the development of technology, especially computer technology. When the automation is of good quality, and when the operator is working at a high workload, operators may trust the automation too much. So, human trust in automation has become an interesting topic to be studied.

Whether human operators use automation or not, greatly depends on their degree of trust in the automation. This trust is taken as an important factor in determining operator’s intervention behavior in automated systems. If the operator’s trust in automation becomes low, the operators will override automation, and if they have too much trust, they may rely on automation and fail to override it even when it fails.

(k) Threatened unemployment: This is a factor frequently considered by the human operators. A major fear of the operators is that the automation will not only take over the work process but also the human jobs. To accomplish the same task, more supervisory control means more automation, more efficiency, less manual control, and fewer jobs for humans.

(l) Silent failures: Automation has its detecting and protecting functions against its own failures. When failures occur in automation, they tend to be “silent” without telling the human operator who is monitoring the automation, and the failures may go undetected by the operators for a long period. The human operators may be blamed after all.

(m) Increasing human operator’s responsibility: Because of the supervisory role of the human operators, a small error from the operators may make the automatic system work in a wrong way, which may result in an accident, or even worse, a hazard. The human operators have nowadays more responsibilities for the system safety than before.

It is true that there are promises in quality, benefit, and safety which motivate the new technology changes and more automation. Nevertheless, we cannot neglect the potential negative impacts which need to be reduced to achieve a satisfactory adoption of automation by the society. A detailed analysis of the negative implications of automation for the society has been addressed by Sheridan (1992).

In transportation and industry, there are many accidents that are related to a high level of automation. These accidents were associated with the decisions to rely on or not to rely on automation. It is likely to play an increasingly prominent role in future incidents as automation becomes more sophisticated and it is given more authority in the operation of complex systems (Riley, 1994). In civil aviation, many accidents result from errors in human-automation
interaction and the errors caused by human overtrust in automation. For a comprehensive review about these accidents the readers are referred to Billings (1991). The accident in 1986 at the Chernobyl nuclear power plant in the Soviet Union is another typical example in the power industry due to disabling the automated safety systems incorrectly (Riley, 1994).

1.3 DEFINITIONS OF BASIC CONCEPTS

The following basic concepts are necessary to define. Some of them are specified only for the current study.

- **Automation**: What do we mean by automation? According to Webster’s Third New International Dictionary, *automation* is the automatically controlled operation of an apparatus, a process, or a system by mechanical or electronic devices that take the place of human organs of observation, decision, and effort. In this thesis, we define:

  *Automation is the result of the allocation of a human task, such as a control task, an information processing task, or a management task, to an automatic controller, or a computer, rather than to a human operator.*

- **Level of automation**: The level of automation is used to indicate how much automation has been used, and it may be measured qualitatively and quantitatively. In different areas, the level of automation has different meanings. For example, in process industry, the level of automation indicates how much advanced control techniques have been used so that control tasks for the human operators become simple and the operators may be able to operate more control loops. In manual control, the level of automation may suggest how many manual tasks have been automated. In supervisory control, the level of automation can show how far we have gone with automation in decision making.

- **Tasks**: Tasks are actions or a set of actions performed by either human or machine for the accomplishment of a goal in the system operation. A task may be decomposed into sub-tasks.

- **Task performance**: Task performance is defined as the result of tasks carried out to satisfy an objective according to some standard.

- **Continuous task**: A continuous task is a task whose degree of accomplishment determines the magnitude of the system state, and thus the resulting system response (e.g. following set-point).

- **Discrete task**: A discrete task is a task whose degree of accomplishment does not
determine the magnitude of the resulting system change (e.g. flip switch). Since the discrete task occurs frequently, it is a main task type in the control room of the nuclear power plant, and it is becoming a major type of task in process industry due to the increase in level of automation.

- **Task analysis**: Task analysis is a detailed description of tasks, in terms of their components, to specify the detailed human activities involved, and their functional and temporal relationships. The purpose of task analysis is to identify tasks, and to provide a basis for the allocation of tasks to the human or the machine, for task execution, for workload assessment, and for defining the essential information to support the human-machine interface design (Kirwan and Ainsworth, 1992).

- **Task allocation**: Task allocation is the process of assigning tasks between humans and machines (computers) on the basis of allocation criteria, such as performance requirement, human workload, human errors, complexity, criticality, technical feasibility, and cost (Papantonopoulos, 1990). The objectives of this process include achieving performance requirements, safety requirements, and production efficiency, and balancing human workload.

- **Task demand load**: Task demand load is the workload imposed by a task on the human operator. The purpose of task demand load measurements is to compare tasks which are expected to be performed equally well from the point of view of their demands on their operators.

- **Mental load**: Mental load is used to indicate the mental load experienced, or perceived, by the human operator when he/she performs a task or a set of tasks. Mental load measurements are introduced in order to be able to choose between operators, or to provide the operator with better human support systems or better training, or to predict the overload which may cause human errors. In this thesis, task mental load is used to indicate the mental load perceived by the human in performing a task.

### 1.4 PURPOSE OF THE THESIS AND PROBLEM DEFINITION

The purpose of this thesis is to study the effect of a change in level of automation on the human operators. More precisely, the thesis focuses on the relationship between the human performance and the level of automation, and the relationship between the human mental load and the level of automation. The hypothesized relationships are presented in Figure 1.1.
Figure 1.1 Hypothetical relationships for performance vs. level of automation and for mental load vs. level of automation.

In order to carry out the study, the thesis first develops a measure to quantify the level of automation in human supervisory control. The measure is noted as Degree of Automation, DofA. Then, the study is carried out on the basis of two experimental systems. Whereas many researches studying the relation between automation and human performance, are mainly focused on the aviation industry, this thesis makes efforts to study the effects of automation in process industry.

Based on the experiments, the thesis analyzes the relationship between performance and DofA, and the relationship between mental load and DofA. At last, the thesis investigates the transition behaviors of the mental load and the performance associated with a sudden loss in DofA, and the human monitoring behavior on the automatic systems.

The thesis studies the following problems:

1. **Level of automation related to task allocation**: The measure of level of automation in supervisory control is defined on the basis of the task allocation between human and machine (how many and which tasks). If task allocation is changed during the operation for any reasons (automation failures, dynamic task allocation), the level of automation will, consequently, be changed. The level of automation in supervisory control is qualitatively defined, or measured, by many researchers (Billings, 1991; Endsley and Kiris, 1995; Sheridan, 1992).

2. **Quantitative measure of degree of automation**: This measure is still based on the decisions on task allocation between human and machine. Moreover, it takes into account both the nature of all tasks and the number of tasks which are allocated to machines. A task weight is used to measure the task nature on either system or human operator.

3. **The effect of a change of degree of automation on system performance**: The system performance is measured for systems that have different values for the DofA.
4. **The effect of a change of degree of automation on mental load:** When the DofA has different values, the amount of mental load perceived by the human operators while controlling the system is collected.

5. **Subjective evaluation of task allocation:** The human operator is asked to evaluate different task allocation decisions based on the given criteria.

6. **Transition behaviors of performance and mental load:** The changes in human performance and mental load before, shortly after, and longer time after sudden DofA drops are studied.

7. **Human monitoring on the automated subsystems:** When the human operator is operating a system by manually performing tasks, the human monitoring behavior on the automated subsystems is investigated.

### 1.5 OUTLINE OF THE THESIS

Since task allocation between human and machine (or computer) is considered as a crucial step in the design of complex human-machine systems, Chapter 2 gives an overview of task allocation based on information available in literature. The overview includes the significance and the goals of task allocation, the problems of current task allocation approaches, and studies on task allocation methodologies.

Chapter 3 addresses the evaluation issues of the decisions on task allocation. These issues deal with approaches and criteria to carry out the evaluation, and the methodologies to assess human mental load. Because the measurement of mental load is an important issue in evaluation as well as in this study, this chapter gives a broad overview of mental load definitions and measurements.

After reviewing the qualitative scales of the level of automation in supervisory control, and the level of automation in industrial systems, Chapter 4 proposes a quantitative measure for the degree of automation in supervisory control, i.e. the DofA measure, and its computation approaches. The DofA is defined as a function of the number and the nature of tasks to be performed by the operator and automation.

Chapter 5 presents the implementation of the DofA in an artificial, linear experimental system. Task effect on system performance is analyzed. Task demand load is predicted on the basis of a task complexity analysis. Subjective assessment of mental load is carried out with human subjects performing control tasks. Based on the experimental data, the effect of the change of the degree of automation on system performance and mental load is discussed, and the prediction of mental load for operating a partly-automated system is investigated based on the mental load measures associated with the individual tasks.
In Chapter 6 the DofA is applied to a simulated pasteurization plan, which is a non-linear system and is close to a real process plant. In this plant, different types of tasks (continuous vs. discrete) are defined. Similar issues as described in Chapter 5 are investigated and the results are compared with those obtained from the study on a linear system. A mathematical tool, Analytical Hierarchy Process (Saaty, 1980), is employed to carry out subjective evaluation of task allocation decisions based on a set of criteria, such as operational difficulty, situation awareness, operational confidence, too much automation, etc..

Based on the experimental systems used in Chapter 5 and Chapter 6 and the DofA measure, Chapter 7 investigates the dynamic transition characteristics of human performance and human mental load when an abrupt drop in the degree of automation occurs.

Chapter 8 uses eye-tracking measurements and the simulated pasteurization plant to study human monitoring behavior, human interaction with the controlled plant and mental load associated with monitoring automated subsystems.

The last chapter, Chapter 9, contains general conclusions drawn from this research, practical implications of the study, and recommendations for future research.
Chapter 2

Task Allocation as a Tool in Human-Machine Systems Design

The decisions on allocating tasks between human and machine determine the level of automation in the control of a complex system, and thus, task allocation has been considered as a crucial step in the design of automation. This chapter addresses the significance and the goals of task allocation in human-machine system design, presents the concepts of static and dynamic task allocation, analyzes the problems of the existing task allocation approaches, and reviews the studies on task allocation methodologies that are intended to be implemented in the design of different systems.

2.1 INTRODUCTION

Before the introduction of technological complex systems, processes were manually controlled, or in manual control. Figure 2.1 shows the concept of manual control tasks which are doing tasks. Originally, manual control did not pose too much workload on the operators because the tasks were not of a very complex nature. As technology advanced, tasks to be performed by the operator became more complex. On the one hand, there came a need for means to assist or to unburden human operators so that they could continue to adequately perform their jobs. On the other hand, human manual control has become a less important issue, and human supervisory control has become the main concept for human-machine interactions (Johannsen et al., 1994). The supervisory control tasks are predominantly intelligent and thinking tasks (Figure 2.1).

The term supervisory control is derived from the analogy between the characteristics of the supervisor’s interaction with the subordinate human staff members and the operator’s interaction with automated subsystems (Sheridan, 1987). The supervisor gives human subordinates general instructions which they in turn translate into action and report the process results to the supervisor. The supervisor of an automated subsystem may do the same. Our discussion on supervisory control will focus on the human supervisory control of the automated systems, rather than a supervisory control system constituted only by computer hardware and software (e.g. the so-called SCADA system - Supervisory Control And Data Acquisition system).
Figure 2.1 Manual control, supervisory control and automation (after IAEA, 1992).

According to Sheridan (1992), supervisory control means that human operators are intermittently programming and continually receiving information from a computer. Manual control tasks and supervisory control tasks differ substantially in terms of system complexity and task definition (Stassen et al., 1990). In supervisory control, the overall task consists of a number of subtasks, whereas in manual control direct closed loop control actions are generally involved. The tasks for the human supervisor are cognitive and include five types of subtasks as shown in Table 2.1 (Sheridan, 1992; Stassen et al., 1990). These subtasks may be categorized based on the three-level concept of human cognitive behavior, i.e. a target-oriented skill-based behavior, SBB, a procedure oriented rule-based behavior, RBB, and a goal-controlled knowledge-based behavior, KBB (Rasmussen, 1983). Stassen et al. (1990) use this qualitative model to classify human operator tasks (Table 2.1). Because more tasks are now performed by automatic controllers rather than by humans, the task on which humans focus is supervisory control via automation.

According to the definition in Ch. 1, automation is the allocation of a control task to a machine rather than to a human operator. It is considered as a means to reduce human involvement in industrial process control so as to improve the production efficiency, to reduce the production cost, and to increase the operational safety.
Table 2.1 Relationship between Human Behavior and Human Operator Tasks (after Stassen et al., 1990): The number of + indicates the significance of the relation.

<table>
<thead>
<tr>
<th>HUMAN BEHAVIOR</th>
<th>Manual Control</th>
<th>Supervisory Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervening</td>
<td>Monitoring</td>
</tr>
<tr>
<td>SBB</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>RBB</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>KBB</td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

However, on the one hand, even in highly automated systems, human intervention may be required in case unexpected situations occur. On the other hand, if automation is not applied properly, the possible result is a clumsy automation (Wiener and Curry, 1980) which usually has a negative effect. It has been proved that a pure technology-centered automation system might cause new types of human errors and new types of system breakdowns (IAEA, 1991). Moreover, there may be stages where human operators become uncomfortable with letting the computer have more and more responsibility (Sheridan, 1988). Nevertheless, it has been argued that an appropriate allocation of tasks to human and machine may avoid the occurrences of these phenomena since the results of allocation of task will definitely affect the human performance, and the design and use of automation.

2.2 TASK ALLOCATION BETWEEN HUMANS AND MACHINES

2.2.1 Significance of Task Allocation in Human-Machine Systems Design

Task allocation is considered as one of the first and probably most important issues in any human-machine system design, and the decisions about the allocation of tasks influence all the later design considerations about the system (Clegg et al. 1989; Grote et al., 1995; Johannsen, 1994). Brennan (1984) states that the allocation phase concerns the determination of the tasks to be performed by the different components and that it has been recognized for a long time as a crucial determinant of system success/failure. The importance of task allocation can be illustrated by the following human-machine system design approaches.

In a theoretical approach, Rasmussen (1986) proposes a design procedure which includes 6 phases for a supervisory control system design (Figure 2.2). Task analysis is performed in Phase 2 and Phase 3. Phase 4 demonstrates the crucial role of the task allocation between human and computer in the design process.
Figure 2.2 Phases in the design of supervisory control system (after Rasmussen, 1986).

Figure 2.3 shows the design process for the control room design in nuclear power plants (IEC Standard 964, 1989). From this figure, again one can see that allocation of tasks to man or to machine is one of the key steps in the control room design.
These examples illustrate the significance of the allocation of tasks in the development of different systems. The effectiveness with which such allocation decisions are made is a function of both the methodology employed and the philosophy underlying that methodology. It is important that allocation is carried out in a sufficiently systematic way. An appropriate allocation of tasks to human and machine should achieve the following goals (Wei and Wieringa, 1994):

- **Making humans and machines work best in a complementary manner**: An optimal task allocation between human and machine implies a proper use of the strengths of both and to achieve a result for which each one is separately incapable.
- **Optimizing the level of human workload**: In abnormal situations, the workload should not be too high. In normal operation, the workload should not be too low.
- **Deciding an appropriate level of automation**: The level of automation influences not only the level of human workload, the operational safety, the operational cost, but also humans themselves in psychological, social and cultural aspects.
- **Ensuring economy, safety, reliability of operation**: These are the ultimate goals in the system design.

### 2.2.2 Task Allocation Principles: Their Benefits and Weaknesses

Task allocation methods can be classified into two categories: **Static task allocation** and **dynamic task allocation**. The former deals control systems with all tasks, while the latter deals with only those tasks which can be performed by either humans or automatic control systems.

#### 2.2.2.1 Static task allocation

Static task allocation refers to the situation where tasks are allocated to human or to machine without changing over time and during the operation. The systematic, theoretical approaches to allocation use different criteria to allocate tasks. Static task allocation can be carried out on the basis of five main principles: (1) **Comparison allocation**: compares human capabilities with those of machines; (2) **complementary allocation**: emphasizes that tasks are allocated to make human and machine work in a complementary way; (3) **leftover allocation**: assigns as many tasks as possible to machines, tasks leftover to humans; (4) **economic allocation**: emphasizes the costs associated with developing and operating the system; and (5) **algorithmic allocation**: treats allocation as an optimization problem based on allocation criteria. The methodologies for static task allocation are developed based on the above principles. Each of them has its own weaknesses which will be discussed below. A recent trend is to combine these principles to overcome their individual disadvantages in developing a task allocation methodology (Wei and Wieringa, 1994).
Comparison allocation. This is one of the earliest attempts to develop tools for use in the allocation of tasks and it was suggested by Fitts (1951). Based on the underlying belief that human and machine abilities are directly comparable, this approach attempts to compare human capabilities and limitations with those of machines in the performance of various tasks. The relative abilities or the merit of performance and weaknesses of humans and machines are presented in the form of a list (Appendix I), the so-called Fitts list or MABA-MABA list (i.e. Men Are Better At ... Machines Are Better At). The MABA-MABA attempts to characterize those activities at which humans appear to surpass present-day machines and those activities at which present-day machines appear to surpass humans (Fitts, 1951). Based on this list, allocations can be made accordingly. This approach can be helpful and has proved the most widely reported of the available methods (Clegg et al., 1989). Many elaboration on the original Fitts List have been made. However, it is not adopted by system designers (Huey, 1989). Its generalizations regarding human and machine capabilities have led to criticisms on several grounds. First, Jordan (1963) has questioned the whole notion of comparability of human and machine abilities. He argues that humans are generally flexible but fairly inconsistent, whereas the machines are just the opposite. Hence, human performance and machine performance are not directly comparable. The Fitts approach does not allow the possibility of humans and machines having overlapping abilities and skills. Jordan proposes that humans and machines may have complementary abilities in carrying out a specific task, and this complementarity should be used to attain maximum effectiveness.

Second, Price (1985) criticizes that the assumption implicated in the MABA-MABA list is a zero-sum design for the choice between human and machine, where, if one cannot execute a task well, then the other can. This assumption does not take into account the likelihood that there are tasks that both human and machine can perform equally well, or that both cannot perform well in which case the task and possibly the overall system performance requirements should be redefined.

Third, the MABA-MABA list is too abstract to be applied to particular designs. It fails to include all human and machine variables affecting human and system performance. The Fitts approach results in a rather simplistic view of a potentially very rich problem (Rouse et al., 1992). Its exclusive consideration of relative merit of performance fails to take into account important psychological, economic, or other contextual factors. A method for task allocation should address the operational requirements of the system arising through the interaction of the system with its operational environment (Papantonopoulos, 1990). Furthermore, the Fitts list requires continuous update to keep pace with the rapid development of technology. Price (1985) also raises doubts as to validity of human performance data collected in laboratory settings for application in workplace.

Fourth, Clegg et al. (1989) point out that one of the major problems of this approach for designers is that it does not relate allocation decisions to the whole design process. Furthermore, it does not take into account the iterative nature of the system design process. The absence of usable heuristics has been the major source of complaints by designers (Kantowitz and Sorkin, 1987).
Nevertheless, in many studies and practices of task allocation, the comparison approach plays an important role (as will be shown later in Table 2.2), but based on different comparison criteria. Since Fitts list appeared, many modified lists have been suggested for the design of different systems (e.g. Clegg et al., 1989). These lists are implemented by adding more factors which influence the allocation of tasks, such as human factors, cultural and educational background of the human operators, social factors and regulating policy etc. (e.g. Bastl et al., 1990).

Meister (1985) and Papantonopoulos (1990) have proposed two quantitative comparison task allocation approaches which employ pair-wise comparison. In these approaches, weights are assigned to allocation criteria by comparing the criteria between each other. The weights are also assigned to human and machine by comparing both task performers. Arndt (1992) proposes another comparison approach where the Structure of Intellect Capability-Requirement (SOI C-R, Table 2.2) is used to quantitatively estimate the capabilities of task performers and the performance requirements. All of these quantitative approaches are useful as convenient alternatives to the qualitative methods of task allocation. At the very least these approaches force the designers to lay out the options systematically.

Complementary allocation. As mentioned earlier, Jordan suggests that humans and machines may have complementary abilities in carrying out a specific task. Thus, the complementarity of humans and automation has been taken as a principle for task allocation in automation design (Hollnagel, 1995), and it is used to compensate for the drawback of comparison allocation. Based on this principle, tasks are allocated to let automation support human control over the process, and to retain human skills. Allocation by complementarity allows for a design of partial automation as a design option (Papantonopoulos, 1990). This principle has been emphasized in the design of all industrial systems, such as aviation, manufacturing, process industry, and nuclear power plants (Bastl et al., 1990; Billings, 1990; Clegg et al., 1989; Lavis et al., 1994; Price, 1985). However, the realization of the complementation between humans and automation requires human-centered design approaches to let both humans and automation be informed about each other’s intent.

Leftover allocation. This approach is to automate as many tasks as possible and to allocate the leftover tasks to humans. This maximizes the use of technology. The erroneous assumption is that humans can do and are willing to do any type of tasks allocated to them (Baily, 1989). It does not take into account the social factors for human operators. It is easy to make a design that follows a technology-driven approach instead of following a human-centered approach. This is because the principal, practical constraints on the extent of automation are determined by the technical feasibility and cost of automating tasks. A maximum (for automation) - minimum (for the humans) principle suggested by Su (1985) has demonstrated a better result in the sense of optimization of task allocation and the improvement of system operations. However, the leftover tasks do not necessarily constitute a coherent and meaningful set of tasks for humans (Rouse et al., 1992).
Economic allocation. It emphasizes the costs associated with developing and operating the system. For each task, the problem is to determine whether it is more economical to select, train, and pay a person to perform the task, or to design, develop, and maintain automation to do the same task. This approach is still used in many systems (Bailey, 1989). The designers have to consider the overall cost of a system when making allocation decisions. It includes the cost of design and operation. The cost saved by shortening the design process may be lost many times over the life time of a system due to expenses caused by inappropriate allocation of tasks.

Algorithmic allocation. This approach treats task allocation as an optimization problem and uses different tools. An optimum value function can be used in a model to optimize the allocation (e.g. Rencken and Durrant-Whyte, 1993; Gramopadhye, et al., 1992; Moray et al., 1982). Simulation tools, such as Human Operator Simulation Model, HOS, (Streib and Wherry, 1979) and Systems Analysis of Integrated Networks of Tasks, SAINT, (Wortman et al., 1975), are used to evaluate task allocation. Mathematical tools, such as linear programming (e.g. Karmi and Herer, 1995) and Petri Net (e.g. Levis et al., 1994), are employed to carry out task allocation. These approaches enable a comprehensive and integrated view of the problem. They may use quantitative representation for tasks, automation, and human operators. Nevertheless, they fall short because they delay all design until after allocation (Rouse et al., 1992). It is not an easy job to define an optimum value function mathematically to carry out the optimization. It is difficult for this approach to take into account factors such as job satisfaction, human motivation, and social aspects of automation in the procedure of optimization, compared to the comparison allocation and the complementary allocation.

2.2.2.2 Dynamic task allocation

Dynamic allocation focuses on tasks which both human operators and machines are capable of performing. Dynamic task allocation is suggested to compensate for the drawbacks of the static allocation and to improve the overall system performance. Task allocation between humans and machines cannot be decided once and forever in a static, deterministic manner. The cognitive capabilities of humans as well as the information processing possibilities of computers suggest dynamic task allocations which are dependent on task situations as well as on the state of the human and the computerized system (Johannsen, 1994). Dynamic task allocation is a method of adaptive aiding (Chu and Rouse, 1979; Rouse, 1977). Hence, it is also called adaptive automation (Gluckman et al., 1991). Adaptive automation involves automating some of the tasks only at times when human performance needs support from automation to meet operational requirements. The goal of adaptive automation is for the operator to remain in control by providing him with aids. These aids will result in the optimal utilization of human and computer resources and will enhance the overall system performance. Instead of a fixed allocation for each task, the designer defines alternative allocations and allocation criteria. The advantages of dynamic task allocation may be summarized as follows (Chu and Rouse, 1979; Huey, 1989; Rouse, 1977; Rieger and Greenstein, 1982): (1) Dynamic task allocation lets human
retain knowledge of the system operation that would be much less likely with the static allocation of some tasks and that may be used to reallocate tasks; (2) the system can be more fault tolerant in automation failure which could not be treated if certain tasks were allocated exclusively to either human or machine; (3) human workload levels may not be as high or as low as would be encountered in automated systems with static allocation; (4) it may more effectively utilize system resources; (5) it may provide the operator with an integrated set of tasks; and (6) finally, system performance can be greatly improved. One possible drawback of dynamic task allocation is that confusion may arise over who, human or computer, should decide reallocation of a task and who should perform a given task (Huey, 1989). Dynamic task allocation has much to contribute in the area of supervisory control systems. When workload becomes too high, the automatic system should automatically pick up more workload to relieve the human operator. The best dynamic allocation will require no conscious effort from the user, i.e. the human operator (Kantowitz and Sorkin, 1987). This may become true when automation is designed to adapt to the humans instead of letting humans adapt to automation.

Because of the development of information processing, dynamic task allocation is now available. For example, artificial intelligence may become an important instrument to introduce dynamic task allocation. In recent years many efforts focus on theoretical researches of dynamic task allocation, or adaptive automation, to investigate the effects of dynamic task allocation on the human operator and to search appropriate models to carry out dynamic task allocation (Andes and Rouse, 1991; Ballas et al., 1991; Cretivits et al., 1994; Gluckman et al., 1991; Greenstein et al., 1986; Greenstein and Lam, 1985; Morrison et al., 1991). Dynamic task allocation has been found to improve overall system performance in studies of simulated aerial search (Morris et al. 1988), flight management and cockpit automation (Emerson and Reising, 1991; Hilburn et al., 1995; Parasuraman et al., 1996), air traffic control (Debbache and Abida, 1994; Debernard et al., 1992), quality inspection (Gramopadhye, et al., 1992), and process system operation (Huey, 1989).

2.3 REVIEW OF STUDIES ON TASK ALLOCATION METHODOLOGIES

Task allocation is a method to balance the effect of several factors on system performance rather than the application of fixed design rules. Especially, when the information about the system is incomplete or uncertain, the criteria are conditionally applied and there is interaction among the individual allocation decisions. Essential elements for task allocation are (Bastl et al., 1990):

1. Detailed knowledge of individual operations for safety and effective running of the system,
2. Knowledge of capabilities and limitations of humans and machines, and
3. Criteria to determine the task allocation process between humans and machines.

Besides these basic elements, Price (1985) points out that task allocation should be carried out in a systematic way to make it more reliable. Task allocation is also an iteration process.
Task allocation decisions must occur continuously throughout the design cycle. Tasks could not be allocated to humans appropriately without considering the tasks that human would be doing if they were assigned tasks in question (Rouse et al., 1992). Other factors that influence the allocation process are existing practices, feedback from experiences, regulatory factors, feasibility, cost and benefit, technical environment, policy matters, cultural and social aspects (Bastl et al., 1990). Some of these factors may lead to mandatory allocation and others to economic allocation.

Table 2.2 summarizes some of the studies on task allocation methodologies. In practice, allocation principles mentioned above are implemented in a combined manner (Clegg et al., 1989; Price, 1985; Rouse and Cody, 1986). Some task allocation methodologies may be used to perform general task allocation and others are suggested to design some specific systems.

The fact that all of the reviewed methodologies have made use of the comparison allocation principle shows that the theory of a task allocation is based on four fundamental propositions (Hollnagel, 1995; Papantopouloso, 1990). First, a process or task can be decomposed into constituent parts or functions and allocated to different task performers (i.e. human or machine). This proposition rests on an assumption that the constituents of a task are additive in a way that optimal performance of the parts results in overall optimal performance. This is called the additivity assumption (Papantopouloso, 1990). Thus, one may assume that complex tasks can be decomposed, simplified and studied by various disciplines at lower levels. However, this assumption may not always be true because a local optimal performance cannot always result in an optimal overall system performance.

A second proposition states that the responsibilities of human and machine in the operation of a human-machine system can be clearly defined. Hence, a task can be allocated to humans or to machines by selecting a suitable task performer. This proposition rests on the assumption that the conditions for manual/automatic operation can be unambiguously specified. It implies that the human can perform certain tasks better than the automation, and vice versa so that the human and the automation can be seen as complementing each other towards better overall performance. Because the transition from automatic operation to manual operation is not so simple, the responsibilities of humans and machines cannot be completely defined (Hollnagel, 1995).

The third proposition states that the capabilities of people and machines can be unambiguously described by using a small set of common characteristics. The suitable performer for a task can be selected by a comparison of relative ability according to a set of performance criteria. This makes an optimal allocation of tasks possible. This proposition is based on the assumption that human-machine interaction can be described by a limited set of criteria, and the capabilities and limitations of human and machine are essentially independent of the context. However, according to Hollnagel (1995), the capabilities of humans and machines cannot be described independent of the situations or the working conditions. Task allocation must be related to the dynamics of tasks.
<table>
<thead>
<tr>
<th>Study</th>
<th>Allocation Procedures</th>
<th>Allocation Principles</th>
<th>Intended Area of Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitts</td>
<td>MABA-MABA list</td>
<td>Comparison</td>
<td>General</td>
</tr>
<tr>
<td>Chapinis</td>
<td>1965</td>
<td>• system specification</td>
<td>Leftover</td>
</tr>
<tr>
<td></td>
<td>• task analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• tentative allocation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• evaluation of the sum of tasks to humans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rieger</td>
<td>MABA-MABA + dynamic allocation</td>
<td>Comparison Dynamic</td>
<td>General</td>
</tr>
<tr>
<td>Greinstein</td>
<td>1982</td>
<td>• task implementation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• alternative documentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• criteria and their weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• weight for alternative design</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• quantitative combination of two weights</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• allocation to the highest score alternative</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meister</td>
<td>1985</td>
<td>• Prepare for design</td>
<td>Pair-wise comparison + Quantitative</td>
</tr>
<tr>
<td></td>
<td>• Identify tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Hypothesize allocation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Test and evaluate allocation hypotheses</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Iterate the design cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>1985</td>
<td>Iteration of allocation</td>
<td>Comparison Complement Algorithm Dynamic</td>
</tr>
<tr>
<td></td>
<td>design evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rouse</td>
<td>1986</td>
<td>• Specify goals of the system</td>
<td></td>
</tr>
<tr>
<td>Cody</td>
<td>Iteration of allocation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>design evaluation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clegg</td>
<td>1989</td>
<td>• Develop requirements</td>
<td>Comparison Leftover Economic</td>
</tr>
<tr>
<td>Ravden</td>
<td>• Identify functions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corbett</td>
<td>• Make mandatory allocations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Johnson</td>
<td>• Determine provisional allocations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Check individual &amp; cumulative allocations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Allocate tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bastl</td>
<td>1990</td>
<td>• Identify global objectives and required performance</td>
<td>Comparison Complement Economic + Many other factors</td>
</tr>
<tr>
<td>Jenkinson</td>
<td>• Identify tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kossilov</td>
<td>• Classify tasks: Must automated, better automated, and should be shared</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ornshead</td>
<td>• Hypothesize allocation to man &amp; machine</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oudiz</td>
<td>• Evaluate allocations, final audit, and repeat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Papantonopoulos</td>
<td>1990</td>
<td>• Task decomposition</td>
<td>Comparison Complement Algorithmic + Quantitative</td>
</tr>
<tr>
<td></td>
<td>• Demand/Resource matching:</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 1: Identify tasks</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 2: Define performance requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 3: Define performance criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 4: March &amp; allocate to performance resources</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Task integration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arndt</td>
<td>1992</td>
<td>Intellectual Capability-Requirement comparison:</td>
<td>Comparison + Algorithmic + Quantitative</td>
</tr>
<tr>
<td></td>
<td>• Assess C-R. based on SOI C-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• compare C-R</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• analyze criticality of exceed requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• summarize comparison and analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• allocate tasks</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A fourth proposition is that operator tasks can be described as a step-by-step procedure. This proposition assumes that complete task descriptions can be produced during system design, and machines will clearly perform as specified, and so will humans. In practice, this may not be true. The human tasks cannot be described as a step-by-step procedure.

Each of the propositions discussed above has its assumption. When these assumptions cannot be satisfied during a human-machine system design, problems will appear. These call for the appearance of new allocation methodologies.

2.4 CONCLUSIONS

Task allocation between human and machine as a tool is important for human-machine systems design as well as for human operator job design, particularly in highly automated and computerized systems where humans are in supervisory control. An optimal task allocation can improve human performance and the overall system performance. A balance of human actions and automation can increase human reliability and reduce the opportunities that human operators may make errors.

From this chapter, we should learn that task allocation as well as human-machine systems development stages have to be performed in an iterative manner, and that it should be a systematic process. It should begin early in design; if not made correctly, they can be very expensive to correct (Price, 1985).

Dynamic task allocation, or adaptive automation, can be used to ensure human operators retain system knowledge and operational skills. However, it is still in its conceptual stages. Although prototypes do exist and many laboratory researches have shown the perspectives of dynamic task allocation, it will take a long time for the technology to mature. This gives us a good opportunity to begin studying the many issues about the technology before its implementation.

There are different methodologies and principles for allocation of tasks. The implementations of these methodologies are dependent on situations and systems. However, it is important for tasks to be allocated based on a complementary principle. This will make a proper use of the strengths of both humans and machines to achieve a result for which each one is separately incapable. The principles for task allocation should be combined in the design. A task orientation with task analysis and subsequent human-centered design approaches is necessary in order to achieve well balanced human-machine systems by taking the most advantage out of both, the human and the automation components (Johannsen, 1994). Further studies are needed to develop new task allocation methodologies for human-centered automation system design.
Chapter 3

Evaluation Measures for Task Allocation

First, this chapter gives an overview on the issue of evaluation of decisions on task allocation between human and machine. This issue includes approaches and criteria in order to carry out an evaluation. Then, the chapter reviews the approaches to predict the demanded workload that is imposed by tasks on the human and the methodologies to assess mental load that is perceived by the human. Both the demanded workload and the mental load are related to the decisions on task allocation.

3.1 INTRODUCTION

An appropriate allocation of tasks to human and machine in supervisory control can optimize the level of human workload, can obtain an appropriate automation level, and can balance automation and human actions. The automation level and the workload level will directly affect system performance and operational safety. The evaluation of the decisions on task allocation is one of the important steps in the task allocation process. Since task allocation is an iterative process, the evaluation may also have to be performed several times during the design process. Moreover, the evaluation process is dependent on the system to be evaluated. For different systems, one can use different evaluation strategies and different evaluation criteria. To evaluate different issues in system design, one may employ different approaches.

Because the operator task becomes more cognitive, the operator perceives mainly mental load from tasks to be performed instead of physical load. In the task allocation process for supervisory control, the mental load level becomes one of the important criteria to be considered. As Sheridan (1987) has pointed out: Although the mental load can be at issue in any human operation, it is aggravated by supervisory control. When a supervisory control system is working well in the automatic mode, the operator may have little to concern about and will experience a low mental load (Point A in Figure 3.1). When there is a failure in automation and a sudden intervention requires a human takeover of an automated task, the mental load may be considerably higher (Point C) than that in direct manual control with a low degree of automation (Point D).
Figure 3.1 Hypothetical relationship between mental load and degree of automation (derived from Sheridan, 1987; Stassen, 1981).

For the latter case it is assumed that the human operator is actively involved and is participating in the system control loop for some time. For the former case, the operator may have to undergo a sudden change from initial inattention, moving physically and mentally to get information and learn what is going on and what has happened, then making a decision on how to cope with the change (Sheridan, 1987; Stassen, 1981). This will be a quick transient from low to high mental load. This transition is illustrated in Figure 3.1.

The overload will result in the operator being unable to do his/her tasks, or at least he/she will not be able to do the tasks satisfactorily. It is also possible that the operator may be able to do his/her task acceptably, but, due to the overload, will not be able to sustain the performance over a long period of time. The level of human mental load is related to the decisions on task allocation. Hence, the human operator should have an appropriate level of mental load by allocating some tasks to him/her. This is why the question is no longer whether one or another function can be automated, but, rather, whether it should be (Wiener and Curry, 1980).

This chapter will present an overview on the evaluation of task allocation decisions. Because the mental load is one of important parameters in the evaluation of task allocation decisions, this chapter also reviews the approaches for the measurement and prediction of human mental load.

3.2 EVALUATION OF TASK ALLOCATION

After tasks have been hypothetically allocated, the important step to follow is the evaluation of the allocation decisions. "If an inadequate allocation solution is chosen, it may reveal itself during testing or operation of the system through reduced output, increased downtime, increased human error or possibly a higher than desired risk to safety" (Bastl et al., 1990). If
the results of the evaluation are unsatisfactory, the allocation must be adjusted, and the allocation process may be repeated. In order to satisfy the evaluation criteria, it is possible to redefine system performance requirements, redefine tasks, and reallocate tasks. This section identifies main aspects in the evaluation of the decisions on task allocation and presents frequently used approaches.

3.2.1 Criteria for the Evaluation of Task Allocation

According to Grote et al. (1995), the evaluation of task allocation decisions can be carried out on the basis of work-oriented criteria for task allocation, resulting jobs, and socio-technical design. When the overall system is integrated, some of the criteria have to be compromised and to be considered from the view of the overall system. However, for an integrated system, new problems may arise and the hypothetical allocation of some tasks needs to be reallocated.

For the overall system design, the ultimate requirements should be evaluated. These requirements are:

- **System performance requirements**: All system goals should be readily accomplished and the system performance requirements have to be satisfied.
- **Operational safety requirements**: The designers must, above all, ensure that safety objectives are met at all times and under all operation situations.
- **Cost effectiveness**: Both the cost of human labor and the cost of automation and maintenance should be evaluated.

As the system performance requirements are met, the safety and the cost efficiency may be reasons, in the real world, to allocate much more tasks to machines.

To meet the human factors objectives (Clegg et al., 1989), the following items are necessary to be evaluated:

- **Level of mental load**: The allocation should aim at enhancing the performance and at properly reducing the human mental load. The mental load should be at the willing-to-spend level of the operator's mental resources (Kalsbeek and Sykes, 1967).
- **Human awareness of system status and human involvement**: Human operators should be kept in the control loop so that they will not lose the awareness of system status, and operational skills. If automation fails, the operator should have adequate knowledge of the system to cope with the failures. This may be related to the level of automation, the design of the human-machine interface, and the design of the information processing system.
- **Human performance, human error and reliability**: Tasks performed by the operator should not exceed the human capacities. The decisions on task allocation between human and machine should aim at enhancing the human performance, and at reducing the opportunities for humans to make errors in order to increase human reliability. Sufficient training, adequate information presentation, and reliable and adequate human-machine interface design are necessary.
• **Human acceptance of assigned tasks, and job satisfactory:** Human operators must have a coherent set of tasks to perform and must understand tasks assigned to them. "Tasks could not be allocated appropriately without considering what human operators would be doing if they were allocated the tasks in question" (Rouse et al., 1992).

• **Human and automation working in a complementary manner:** The human operator plays a role, on the one hand, as a system supervisor with his/her creativity, intelligence, and flexibility. On the other hand, the operator acts as a back-up in case automation fails. Humans have to be kept in the control loop in order to be able to fulfill their final responsibility for production and safety although they may make errors. Both human and automation should enhance other's capabilities, and overcome other's limitations. Use of a well-balanced allocation process will optimize the contribution of both the human and the automation to overall system performance (Bastl et al., 1990). Grote et al. (1995) has presented a set of criteria for the complementary task allocation, which can be partly used to evaluate the complementarity of the decisions on task allocation.

In summary, the allocation of tasks between human and machine aims at satisfying both the human and the technical requirements. It is a human-centered process as well as a technology-driven process. As the development of the concept of human-centered automation design, new criteria may be introduced to evaluate the decisions on task allocation.

For the technology, including hardware and software, the following two aspects are necessary to be evaluated:

* **Cost and benefits:** A cost-benefit principle must exist to justify most proposals to automation.

* **Technical feasibility and climate:** Enhancing capabilities of technology may facilitate automation whereas non-availability of technology may limit what can be automated. Sometimes, automation is impossible for practical reasons or it may be forbidden to use by the regulatory organization.

Besides, other issues to be considered are human operator's cultural and educational background, and social aspects.

### 3.2.2 Evaluation Approaches

Among the approaches to evaluate the decisions on task allocation, many of them may be the so-called desk top analysis, such as the prediction of demanded workload to the human operator on the basis of a task analysis. Another approach, the simulation technique, is frequently used to let the operator perform a task or a series of tasks so that human performance and mental load can be assessed. To evaluate those aspects mentioned in Section 3.2.1, the following approaches may be considered:
• *Estimation and measurement of mental load:* The approaches to estimate the demanded load on the human operator are used to predict the possible level of human mental load. These approaches will be addressed in detail in Section 3.3.

• *Part-task and full scope simulations:* From part-task or full-task simulation by employing the human operators, human mental load, human performance as well as system performance can be investigated. Experience indicates that part-task manned simulations are of the benefit in the evaluation of early allocation decisions (Bastl et al., 1990; Rouse and Cody, 1986). Empirical data collection efforts using part-task simulations are necessary to evaluate newly designed displays, controls and procedures. Maximum benefit is obtained from a full scope, replica simulator which makes a significant contribution to human-machine interface evaluations (Bastl et al., 1990). Moreover, human acceptance and job satisfaction, human involvement in the control, and the complementarity between human and automation can be investigated.

• *Human performance model and data:* This approach aims at predicting human performance and at answering questions in the design of new displays, controls and procedures. Human performance can be measured by engineering parameters, such as executing speed, executing accuracy, processing time, production rate, error rate and training time.

• *Existing designs and operational experience:* Experience in practice is important for evaluating the use of automation. The plant operators and training staff who possess valuable knowledge, experience, and expertise, can contribute to the evaluation. If a task has been performed earlier by an automatic controller in other systems, without trouble, this task can also be performed by an automatic controller in a new design. The level of automation depends also on operational practice and the level of technological and experiential supports, such as available maintenance.

### 3.2.3 Remarks on the Evaluation of Task Allocation

It is necessary to point out that an evaluation should be performed as early as possible. If an evaluation takes place too late and a task allocation decision does not satisfy the evaluation criteria, it may be expensive to correct the decision, especially after elements of the system have already been built as hardware.

In evaluating a task allocation, three factors must be considered carefully. The first one is the evaluation team. The evaluation team should include the designers, human factors engineers, training staff, safety staff, human operators with experience of similar systems or other versions, and operators of the future systems (Bastl et al., 1990; Price, 1985). Different personnel will take care of different aspects of the evolution. For example, when human performance is evaluated, the human operators who will operate the designed system should be taken as subjects (IAEA, 1990).

The second factor is the criterion. As discussed before, there will be multiple criteria to be used to evaluate the task allocation decisions. These criteria should be used in a compromise
way. The overall goal of the task allocation process is to obtain a satisfactory balance between automation and human actions, instead of a perfect one (Bastl et al., 1990).

The third factor is the choice of the evaluation procedures and methods. This deals with how the evaluation is designed. For example, whether part-task simulations, or full scope simulations will be used.

The behavior or the performance of the operators on the basis of task allocation decisions should be evaluated not only during normal situations, but also during abnormal and emergency situations. In abnormal situations, due to a possible increase in human mental load, the operator performance may degrade, and the chance that the operator may make errors will become larger. Although Bastl et al. (1990) have argued that it is not suitable to reallocate automated tasks to the operator in the operation of the plant, the human intervention of automation when automation fails must be considered and must be evaluated (Hollnagel, 1995; Kim and Sheridan, 1995). Thus, it is possible that as the level of automation increases, more issues for the decisions on task allocation must be evaluated, and the evaluation should be done not only in a traditional manner.

3.3 MEASUREMENT OF MENTAL LOAD

While allocating tasks to humans, the designers do not only want to know whether the humans can perform these tasks acceptably. They also want to know whether the humans can sustain performance over a duty cycle, a pay period, or a career (Rouse, 1991). In many situations, the performance is taken to be acceptable even if the costs of performing are very high. Mental load is a concept for capturing the costs that humans pay (or are willing to pay) when they perform tasks. It has long been recognized as an important factor in human performance in complex systems (Moray, 1979; O’Donnell and Eggemeier, 1986; Stassen et al. 1990). In general, although not always, if the mental load is too high, performance will eventually decrease (Figure 3.2), either because humans become fatigued or because they are simply unwilling to tolerate the high mental load (Bi and Salvendy, 1994; Rouse, 1991; Stassen et al., 1990). In contrast, if mental load is too low, the operators are easily distracted and become bored, and may lose the awareness of system status. It is tempting to search for some optimal level of human mental load.

What is mental load? Many definitions have been proposed (Sheridan and Stassen, 1979). It is not the interest of this research to propose or discuss a universally acceptable definition of mental load. There is an agreement that mental load is incompletely defined, multi-faceted, and that it has a direct bearing upon an operator’s ability to maintain or reach a desired performance level (Linton et al., 1989). The four basic tenants of mental load presented by Linton et al. (1989) may help us to understand the concept of mental load:
Figure 3.2 Hypothesis for relationship between performance and mental load (after Bi and Salvendy, 1994).

(1) Mental load reflects relative, rather than absolute individual states. It depends on both the external demands, the internal capabilities, and the motivation of the individual. This relativity exists qualitatively as well as quantitatively and in the dimension of time.

(2) Mental load is not the same as the individual's performance in the face of work or tasks, nor is it synonymous with the way of measuring performance.

(3) Mental load involves the depletion of internal resources to accomplish the work. High mental load depletes these resources faster than low mental load.

(4) Individuals differ qualitatively and quantitatively in their response to mental load. There are several different kinds of task demands and corresponding internal capabilities and capacities to handle these demands. People differ in the amount of these capabilities which they possess, and their strategies for employing them.

However, according to Stassen et al. (1990), two concepts about mental load (Figure 3.3) must be distinguished. The first one is Task Demand Load, TDL, which is defined as the demand, or the anticipated mental load imposed upon humans by the task or the system to be supervised. The usefulness of this definition depends on the extent to which humans attempt to satisfy all demands (Rouse et al., 1993). Another one is Task Mental Load, TML, which is defined as the mental load experienced or perceived by humans. According to the definition, the TDL is independent of the human operator and may be called objective workload. The TML, i.e. the perceived load, is dependent on the human operator and is therefore a subjective load.
Both the TDL and the TML must be assessed in order to determine the actual importance of mental load (Rouse, 1991). The TML is assessed using human operators (Moray, 1979; O'Donnell and Eggemeier, 1986; Wierwille and Eggemeier, 1993; Zijlstra, 1993). The methods to assess TML, the so-called empirical techniques (Hill et al., 1987; Linton et al., 1989), will be reviewed in Section 3.3.1. The TDL can be assessed off-line and without the operator-in-the-loop by analytic methods based on system engineering parameters, mathematical models, and task analysis methodologies (Bi and Salvendy, 1994; Hill et al., 1987; Linton et al., 1989). The approaches to estimate TDL, the so-called analytic techniques, are presented in Section 3.3.2.

3.3.1 Measurement of Task Mental Load

As illustrated in Figure 3.3, when the human operator performs a task, he/she has to take effort to fulfill the task. This effort has been considered as the workload that the operator perceives (Zijlstra, 1993). It is also found that human’s perceptions of the state of their current situation are strongly related to their subjective assessments of their workload, which may be expressed in terms of effort required (Rouse, 1991). The methodology to assess the TML is categorized into three types (Huey, 1989; Meshkati et al., 1990; Moray, 1979; O'Donnell and Eggemeier, 1986; Rouse, 1991; Sheridan and Stassen, 1979; Stassen et al., 1990; Wierwille and Eggemeier, 1993; Zijlstra, 1993) as follows:

1. Performance measures on “primary task” and “secondary task”

   Secondary task performance measure is called reserve capacity technique. In this technique, human subjects are required to perform two (dual) tasks concurrently, one of which is given priority — the primary task. The performance is measured on the task of lower priority — the secondary task. Mental load of the operator on the primary task must be high when performance on the secondary task degrades. The drawbacks of the secondary task technique
Evaluation Measures for Task Allocation

are (Huey, 1989; Wierwille and Eggemeier, 1993): (1) It needs substantial equipment; (2) the secondary task is intrusive; (3) it is impossible to obtain a universal secondary task to be suitable to all environments; and (4) subjects may confuse the primary task with the secondary task during the experiment.

Primary task performance measure assumes that mental load must be high when primary task performance cannot be sustained. This method is usually a poor indicator of mental load since it does not reflect the variation in human resource investment due to variations in task difficulty and complexity. Moreover, despite numerous investigations on the relationship between performance and mental load, there is insufficient evidence to reject the null hypothesis that mental workload and performance are unrelated (Rouse et al., 1993; Yeh and Wickens, 1988).

2. Physiological measures

The changes of physiological variables associated with information processing and physiological state changing, such as eye activity, heart beat variations, brain activity, muscle tension, pulse wave velocity, body temperature, and so on, are said to reflect changes of mental load if two tasks with different levels of mental load yield different values of indirect measures. These measures are fairly in-intrusive, and can be used as a supplement to subjective measures. They are promising for event-related or instantaneous mental load measurements and are useful in identifying peak mental load or temporary operator overload (Wierwille and Eggemeier, 1993). They may be employed as on-line indicators of mental workload to carry out dynamic task allocation. However, these measures are usually embedded in noise. It is not the measure per se, but rather the context within which it is recorded, that determines the value of the measure (Huey, 1989).

3. Subjective measures

The human subject rated experience of high mental load is considered to be a subjective measure of mental load. The subjective measures are based on operator judgments of the mental load associated with the performance of a task or a system function. The human subjects are asked, according to some rating scales, for a direct assessment of the mental load they experienced. The assumptions of this technique are: Subjects are aware of the effort needed to cope with task demands; and subjects can perceive their own limitations and make a truthful report (Huey, 1989). The advantages of subjective measures lie on (1) easy to obtain data and simple to use, (2) least intrusive to the task performance, (3) simple equipment, (4) high face validity, (5) no specific knowledge requirement, and (6) a high level of consistency or reliability (Huey, 1989; Moray, 1982; Wierwille and Eggemeier, 1993; Zijlstra, 1993).

Sheridan and Stassen (1979) illustrate the relationships among the three measures for the mental load assessment in a control paradigm (Figure 3.4).
Figure 3.4 Alternative definitions of workload and measures (Sheridan and Stassen, 1979).

For a detailed review on performance measures and physiological measures for mental load, and guidelines for use of these approaches, the reader is referred to O'Donnell and Eggemeier (1986), Meshkati et al. (1990), and Wierwille and Eggemeier (1993). Stassen et al. (1990) present the priorities of these mental load measurement techniques. Specialists from different disciplines, i.e. psychologists, man-machine system designers, field users, and physiologists ranked these techniques. They also evaluated them against the transducer requirements, such as sensitivity, dynamics, and consistency among subjects. Moreover, a comprehensive discussion on mental load theory and its measurement can be found in Moray (1979).

As the role of the human operator becomes less to control and more to monitor complex systems, subjective measures are of increasing importance in human-machine systems and they are the dominant method for measuring mental load (Moray, 1982). Since the subjective measures are among the most sensitive, most transferable, and least intrusive techniques for mental load estimation (Wierwille and Eggemeier, 1993), they will be used in this research. In the next section, details of the chosen subjective rating technique will be discussed.

3.3.2 Subjective Mental Load Measurements

The factors known to produce a feeling of high mental load include the following: (1) Time constraints, (2) the requirement to generate a response when performing tasks with high order
transfer functions, (3) instability of the system to be supervised; (4) composite dynamics in the tasks, (5) magnitude in tracking tasks, (6) noisy signals and precision requirement in either perception or response, (7) boredom, (8) surprise to an unexpected event, (9) execution difficulty, (10) psychological inertia, and (11) hard physical work (Huay, 1989; Rouse, 1991). Among these factors, the first to sixth are covered by the definition of task complexity which will be addressed later.

The subjective measure is carried out on the basis of a rating instrument, by which the subjects are asked to rate the complexity of the task or the amount of effort that needs to be exerted while they execute a task. Several of the most frequently used techniques that were designed for implementation in a variety of systems include the well-known Cooper-Harper Scale, CH, (O'Donnell and Eggemeier, 1986), the Modified Cooper-Harper Scale, MCH, (Wierwille and Casali, 1983), the Subjective Workload Assessment Technique, SWAT, (Reid and Nygren, 1988), and the NASA Task Load Index, NASA-TLX, (Hart and Staveland, 1988).

The CH, or the MCH, has been mainly applied to assess mental load in flight experiments that incorporated several types of demand manipulations (Wierwille and Eggemeier, 1993). These scales assume one underlying dimension which ranges from “easy to fly” to “very difficult to fly”. A frequent criticism about these scales is the one-dimensional assumption (Zijlstra, 1993). Theoretically, it is hard to imagine that mental load could be expressed in one dimension. Thus, multi-dimensional rating scales, such as the SWAT and the NASA-TLX, appeared.

The SWAT is a multi-dimensional rating scale and includes three subscales: (1) Time pressure load, (2) mental effort load, and (3) psychological stress load. It has been extensively employed to assess pilot mental load and has demonstrated its sensitivity to different demand manipulations in flight environment and other systems.

The NASA-TLX is a more extensive multi-dimensional rating instrument. It has six subscales: 1) Mental task demand, 2) physical task demand, 3) time pressure, 4) successful performance, 5) effort, and 6) frustration level. According to the task performed, subjects are asked to evaluate, by pair-wise comparison, the contribution of each subscale. In this way, a weight to each subscale is assigned. Subjects have to rate each task based on these six subscales. The ratings are combined according to the weight for each subscale. The NASA-TLX has been used in many different systems and has demonstrated its sensitivity (Wierwille and Eggemeier, 1993).

Although the multi-dimensional rating scales, like the SWAT and the NASA-TLX, can provide some diagnostic information on the sources of mental load represented by the subscales, it is not easy to use in an experiment. Moreover, some experiments have demonstrated that in certain circumstances, the one-dimensional and the multi-dimensional rating scales can achieve similar results (Veltman and Gaillard, 1993; Vidulich and Tsang, 1987, and Ch. 5 in this thesis). Furthermore, the one-dimensional rating scales are easier to use in the experiment.
Figure 3.5 Rating Scale Mental Effort, RSME (after Zijlstra, 1993). The statements in RSME are presented in Dutch because the RSME was used by Dutch subjects.

A Dutch version of a one dimensional subjective rating instrument, the Rating Scale Mental Effort, RSME (Figure 3.5), has been developed by Zijlstra (1993). The RSME asks for the amount of effort that has been invested during task performance. Subjects have to respond by putting a marker on a vertical scale. The right side of the scale has statements related to mental effort such as “not effortful”, “a little effortful” and “very effortful” etc., whereas the left hand side includes numerical values from 0 to 150.

The RSME has been implemented in the assessments of mental load for humans working in different systems and by people with different backgrounds. It has been used by bus drivers and nautical officers (Zijlstra, 1993), flight jet pilots (Veltman and Gaillard, 1993), psychology students while using a railway traffic control aiding system (Neerincx, 1995), and mechanical engineering students while controlling a complex system (Wei et al., 1995a, b) or while evaluating human-machine interfaces for the control of a space manipulator (Breedveld, 1996). It has been found that when one needs a Dutch mental load scale, the RSME is a rather good one (Veltman and Gaillard, 1993). In this study, the RSME is used as a main rating scale to estimate human mental load.

Subjective measures have many advantages compared to other measures. However, there are also associated problems (Huey, 1989; Yeh and Wickens, 1988): (1) The lack of a theoretical basis to explain the results, (2) psychometric difficulties in generating interval, rather than ordinal, data scales, (3) the difficulty to compare rating results, (4) ratings yield relative, instead
of absolute, results, (5) current techniques are retrospective, and thus, some of the information may not be available or is distorted, (6) the measures depend on the content of consciousness, so subjects can only report that they are aware of their interactions with the task, (7) it is difficult to know the degree to which subjective measures can be used to account for variance in performance, and (8) it cannot be used to measure mental load in real-time or on-line, so that it is difficult to implement this technique in dynamic task allocation as an indicator of mental load.

3.3.3 Prediction of Task Demand Load

Task demand load can be obtained without an operator-in-the-loop, or real time operation. It can be assessed by analytic techniques off-line — workload predictive techniques which are applied earlier in system design. Thus, the predictive techniques are useful for concept development, preliminary system design, and evaluation of the decisions on task allocation without using task simulations associated with the human operators. The analytic techniques are described by five categories: (1) Comparability analysis, (2) mathematical models, (3) expert opinion, (4) task analysis methods, and (5) simulation models (Hill et al., 1987; Linton et al. 1989). Among these categories, mathematical models, task analysis methods, and simulation models have been systematically formalized and fully validated, and have been used in different systems and contexts. Further discussion in this review will focus on these three methods.

1) Mathematical models

The mathematical models describe relevant variables, and combine them so that workload associated effects on performance can be estimated. The basic steps for the mathematical models to model human performance are (Linton et al. 1989):

- Identify variables that influence mental load either directly or indirectly;
- determine the causal relationships of these variables;
- establish how the resulting workload predictions drive predictions of performance.

The frequently used techniques in building mathematical models are: (1) Manual control models for continuous control tasks, where the classical control theory and optimal control theory are used, (2) information theory models, and (3) queuing theory models operating within a discrete task environment. It should be stressed that the mathematical models based on the above mathematical theories are developed to describe the performance. Nevertheless, Linton et al. (1989) suggest that they are useful to estimate workload. This is because each of these techniques has parameters or components which are sensitive to some external aspects of workload or to the operator reactions to such external manipulations (Hill et al., 1987).

A number of mathematical models has been developed to model and ultimately to estimate task demand load. These models work well within the contexts in which they were developed.
According to Hill et al. (1987), the major difficulties with mathematical models are the lack of their generality and their real world applicability, unease of implementation and interpretation. A major problem with these models is the absence of the explicitly defined workload parameters. While model outputs may identify and quantify very busy periods during a given time, or particularly high periods of information transfer, it is not clear how, or whether, these phenomena relate to high workload. Another problem is that detailed system parameters must be provided to exercise these models. However, these parameters are often not available during early concept development. Linton et al. (1989) give a review on the advantages and weaknesses of mathematical models and Moray (1979) discusses the relation between mathematical models and the calculation and prediction of workload.

2) Task analysis methods

Task analysis techniques are among the most familiar and commonly used analytic methods for assessing workload in the preliminary system design process. The reasons for frequent use of these techniques are: (1) Easy to understand and to undertake; (2) task analysis is one of important steps for a new system design, e.g. for task allocation; and (3) task analyses are useful for analyst (Linton et al., 1989). There are many analytic approaches to estimate workload. The first type of them seek to uncover any situation in which the time required to complete activities exceeds the time available. This type is called time-based definition of workload. Its representative approach is the so-called Timeline Task Analysis. This approach defines the overall workload as time stress. This is expressed as a ratio of Time required, $T_r$, to perform a task over the Time available, $T_a$, yielding $T_r/T_a$. When the time to complete a task comes close to the available time, that is as $T_r/T_a$ approaches 1.0, it can be assumed that the operator will be overloaded and will have difficulty in performing the required task. This approach has been applied in many fields for the overall workload assessment, such as in the control room design of nuclear power plants (IAEA, 1992), in the overall aircraft operation, in dynamic task allocation (Rencken and Durrant-Whyte, 1993), and recently, for cognitive task load analysis (Neerinckx, 1995). Another typical approach is called Workload Index, W/INDEX (e.g. North and Riley, 1989). This approach combines mission, task, and timeline analysis with theories of attention and human performance to predict attention demands.

The second type of task analysis approaches identify individual task components and examine the complexity and conflict between competing task resources, such as the McCracken-Aldrich approach (e.g. Aldrich, et al., 1989). This approach does not only rely on the time-based definition. It breaks workload into five components: Visual, auditory, kinesthetic, cognitive, and psychomotor. The five components are assigned a scale ratings from 1 to 7 which signifies the relative degree of workload.

Salvendy and Bi (1995) have proposed a concept model of human workload prediction. This model is suggested to predict task demand load and mental load based on system engineering parameters: Task arrival rate, task complexity, task uncertainty, and task performance requirements. Bi and Salvendy (1994) report that an analytic mathematical model
has been developed based on the parameters mentioned above. This analytic model is validated on a real time scheduling simulator of an advanced manufacturing system.

3) Simulation models

The application of simulations to assess human load is a conceptual extension of traditional operator-in-the-loop simulation. The difference is that the simulation for workload estimation is expanded to use a simulated operator (a mathematical operator model), rather than a human operator. There are numerous references to simulation models that can be used for workload estimation. Among the most widely used simulation models are: Siegal-Wolf network models (e.g. Meister, 1985), SAINT/Micro SAINT—System Analysis of Integrated Networks of Task (Laughery, 1989; Laughery et al., 1986), HOS—Human Operator Simulator, and so on. Meister (1985) provides a broad overview of simulation models, most of which are computerized, and their applications.

3.3.4 Remarks on Mental Load Assessment

From the techniques for mental load estimation discussed above, we learnt that the purposes of mental load assessment are threefold. The first is to predict the task demand load of a particular system configuration before it is actually realized in production or simulation, e.g. evaluation of the decisions on task allocation. This allows the designers to discriminate among alternative task allocation decisions. The second purpose is to estimate task demand load of already existing systems. The third purpose is to measure human mental load on-line. This measurement can be used in a dynamic task allocation system to reduce excessive human mental load. A great effort has been made to develop methodologies to predict task demand load and to assess task mental load as reviewed above. However, few of the available workload measurement techniques may be considered to capture the full complexity of the mental load issue. People from different disciplines judge these techniques very differently (Stassen et al., 1990). There is no “ideal” mental load assessment methodology. Thus, to capture the mental load issue completely, for each situation, both analytical and empirical techniques are necessary, and for different situations, a combination of different overall workload assessment techniques is needed.

3.4 SUMMARY

The evaluation of the allocation decisions is one of crucial steps for task allocation in human-machine system design. There are many different approaches to carry out an evaluation. For a system design, a combination of different evaluation methodologies should be employed on the basis of the evaluation criteria and system design objectives.
Human mental load during the operation of complex human-machine systems is one of the important issues to be evaluated. When all of the decision-making, system monitoring, planning, situation assessment, and control responsibilities have been allocated between humans and machines, the question of whether the operator's mental load capacity has been exceeded is likely to determine the feasibility of the design. If the mental load level is high, a single automation failure may make the human operator's job unmanageable. It is especially important to take into account factors that will make the job unnecessarily difficult by contributing to the operator's mental load. Similar to the evaluation methodology, there are many approaches to measure or assess human mental load, but there is not a universally acceptable approach. Based on the application area and human tasks, a combination of measurement approaches is also necessary.
Chapter 4

A Quantitative Measure for the Degree of Automation

Qualitative scales of level of automation in supervisory control and their usage in studies on automation and human performance are reviewed. A quantitative measure for the level of automation, called "Degree of Automation", is proposed. Degree of automation is defined as a function of the number and the nature of tasks to be performed by the operator and the automatic control system. The computation of the degree of automation is done by employing a weighting scheme for all tasks. The weighting factor for a task is defined on the basis of how much a task affects the system performance, how much workload a task demands from and imposes on the human operator, and how much mental load that the human operator perceives during his performance of the task. The approaches for analyses of task effect on system performance, prediction of task demand workload and subjective assessment of mental load are proposed.

4.1 INTRODUCTION

Automation has been employed in industrial processes to achieve consistent product quality, to minimize operator workload, to reduce human error during execution of repetitive tasks, to improve equipment availability and production efficiency, and to enhance operational safety. Highly automated systems have become available as a result of the developments in computer and information technology. As a result, modern control systems are capable of operating a nuclear power plant or controlling an aircraft's flight from take-off to landing. Due to the demands for higher production efficiency and higher operational safety, the level of automation in system operation has considerably increased. The role of human operators has shifted from manual control to supervisory control. The nature of the operator tasks has changed from performing action control tasks to performing the intellectual tasks or cognitive tasks.

The benefits of automation are obvious. However, a high level of automation may have adverse effects. Bainbridge (1983) has discussed possible consequences of increasing the level of
automation, and Ch. 1 has reviewed the problems caused by an increasing level of automation. In the context of human-machine systems, automation is applied only to control tasks that are fully understood and that are completely described. The tendency has been to automate what is the easiest and to leave the rest to the operator. However, the human operator may also have difficulty to perform those tasks during situations that are not completely understood or foreseen by system designers (Rouse et al., 1992). This can result in an overall degradation of system performance. Furthermore, higher levels of automation imply higher system complexity which can affect both operator performance and system performance (Stassen et al., 1993). Too much automation may result in poor operator performance due to too little operator workload, operator loss of skill and awareness of the system status. A series of questions can be asked. How much automation should be used? How can the level of automation be measured? How does an increase in the level of automation affect system performance, operational safety and human mental load?

In this chapter, first, the qualitative measures for the level of automation in supervisory control of a complex system and the implementation of these measures in scientific research are briefly summarized. Then, levels of automation in the control room of nuclear power plants and in the cockpits of aircraft are analyzed. Finally, a quantitative measure for the level of automation, called “degree of automation”, is defined and the methods to obtain the parameters in the quantitative measure are proposed.

4.2 QUALITATIVE SCALES FOR THE LEVEL OF AUTOMATION

The level of automation, as a qualitative concept, has been used to indicate how far the system design has proceeded with automation (Billings, 1991; Endsley and Kiris, 1995; Kantowitz and Sorkin, 1987; Sheridan, 1992). This section reviews and compares two qualitative scales for the level of automation in supervisory control.

4.2.1 Sheridan's Scale for the Level of Automation

A qualitative scale of level of automation suggested by Sheridan (1992) is shown in Table 4.1. In his scale, Sheridan classifies the level of automation by focusing on the distribution of responsibility, or task allocation, between human and computer (Johannsen et al., 1994; Levis et al., 1994). He determines the level of automation only by the type of tasks that a human supervisor would do, such as: Decision making, execution of tasks and monitoring, compared to the tasks that are allocated to machines (computers). However, how many of these tasks will be performed by a human supervisor or by automation is not taken into account.
Table 4.1 Scale of level of automation (after Sheridan, 1992)

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1. The computer offers no assistance, human must do it all.</td>
</tr>
<tr>
<td></td>
<td>2. The computer offers a complete set of action alternatives, and</td>
</tr>
<tr>
<td></td>
<td>3. narrows the selection down to a few, or</td>
</tr>
<tr>
<td></td>
<td>4. suggests one, or</td>
</tr>
<tr>
<td></td>
<td>5. executes that suggestion if the human approves, or</td>
</tr>
<tr>
<td></td>
<td>6. allows the human a restricted time to veto before automatic execution, or</td>
</tr>
<tr>
<td></td>
<td>7. executes automatically, then necessarily informs the human, or</td>
</tr>
<tr>
<td></td>
<td>8. informs him after execution only if he asks, or</td>
</tr>
<tr>
<td></td>
<td>9. informs him after execution if it, the computer, decides to.</td>
</tr>
<tr>
<td>High</td>
<td>10. The computer decides everything and acts autonomously, ignoring the human.</td>
</tr>
</tbody>
</table>

4.2.2 Billings’ Continuum for System Control and Management Automation

A scale for the level of system control and management automation that is categorized by the degree of direct or immediate involvement of the human operator is discussed by Billings (1991) and by Endestad (1991). The level of automation is classified into seven levels as shown in Table 4.2.

Billings (1991) proposes the continuum to discuss and to define human-centered aircraft automation. Endestad (1991) uses the continuum to analyze the development and the safety of nuclear power plants with relation to the degree of automation in control, information and management automation.

4.2.3 Correspondent Relation between Sheridan’s Scale and Billings’ Continuum

The relationship between Sheridan’s scale and Billings’ continuum is displayed in Table 4.3. The following observations can be made from Table 4.3. With a high level of automation, the operator is not informed about the system’s intent and the work of automation and the system may or may not be disabled by the operator. The operator has a minor role in operation; the monitoring is limited to fault detection; goals are self defined; and the operator usually has no reason to intervene. With a low level of automation normal warnings and alerts are provided,
and routine automatic systems work. The operator has direct authority over all systems, and there is manual control and unaided decision making.

Table 4.2 Continuum for control and management automation (derived from Billings, 1991)

<table>
<thead>
<tr>
<th>Level of automation</th>
<th>Mode</th>
<th>Automation Functions</th>
<th>Human Functions</th>
</tr>
</thead>
</table>
| Very low            | DIRECT MANUAL CONTROL | • Normal warnings and alerts provided  
                     • Routine hydraulics/electric control systems | • Direct authority over all systems  
                  • Manual control and unaided decision making |
|                     | ASSISTED MANUAL CONTROL | • Automatic systems available for routine use and for housekeeping functions  
                     • Monitoring of system status and control | • Direct operator authority over all systems  
                  • Trend information available on request  
                  • Specific systems are typically automatic with operator override |
|                     | SHARED CONTROL | • Aided or enhanced control and predictor display  
                     • Advisory system available | • Operator in control through protection systems  
                  • May use advisory systems  
                  • System management manual and procedures |
|                     | OPERATIONS | • Automatic control of some system operations | • Operator commands system operation and may have limited manual functions |
|                     | MANUFACTURING | • Full automatic control of system and its operations  
                     • Intent, diagnostics and prompting functions are provided to operator | • Operator must consent to state changes, checklist execution, anomaly resolution  
                  • Manual execution of critical actions |
|                     | MAINTENANCE | • Essentially autonomous operation  
                     • Automatic reconfiguration  
                     • System informs operator and monitors responses | • Operator informed of system intent  
                  • Must consent to critical decisions  
                  • May intervene by reverting to lower level of control |
| Very high           | AUTONOMOUS OPERATION | • Fully autonomous operation  
                     • Operator not informed  
                     • System may or may not be capable of being disabled | • Human has minor role in operation  
                  • Monitoring is limited to fault detection  
                  • Goals are self-defined; human usually has no reason to intervene |
Table 4.3 Relation between Sheridan’s and Billings’ scales

<table>
<thead>
<tr>
<th>Level of Automation</th>
<th>Continuum of Control and Management Automation</th>
<th>Sheridan’s Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very low</td>
<td>Direct manual control</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Assisted manual control</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Shared control</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Management by delegation</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Management by consent</td>
<td>5, 6</td>
</tr>
<tr>
<td></td>
<td>Management by exception</td>
<td>7, 8</td>
</tr>
<tr>
<td>Very high</td>
<td>Autonomous operation</td>
<td>9, 10</td>
</tr>
</tbody>
</table>

4.2.4 Implementation of Scales of Level of Automation

Many studies where a scale for the level of automation is employed have been conducted on the effect of varying levels of automation on system performance, human performance, situation awareness, human mental load and system safety. A few examples of implementation of the scale are summarized below.

4.2.4.1 Situation-adaptive task allocation

Inagaki and his colleagues (Inagaki, 1995; Inagaki, 1993; Inagaki et al., 1994) directly use the scale of level of automation suggested by Sheridan to investigate task, or responsibility, allocation between human and automation in supervisory control of large-complex systems or semi-autonomous processes related to process safety. They have proved that the level of automation should not be fixed but must be dynamically and flexibly changeable. They argue that the levels of automation in Sheridan’s scale higher than 6 are useful for maintaining process safety, even though these levels do not satisfy the principle, “a human locus of control is required”. It is suggested that the situation-adaptive level of automation is indispensable for realizing human-centered automation.

4.2.4.2 Situation awareness

Endsley and Kiris (1995) use a scale with five levels of automation which is similar to the scale suggested by Sheridan to investigate the out-of-the-loop performance problem which leaves
human operators of automated systems handicapped in their ability to take over manual operations in the event of automation failure. The out-of-the-loop performance problem is attributed to a possible loss of skills and of situation awareness arising from vigilance and complacency problems, a shift from active to passive information processing, and change in feedback provided to the operator. The five levels they used are: A task is accomplished (a) manually without assistance, (b) by the operator with recommendations from automation in decision making, (c) by automation, but with the consent of the operator, (d) by automation, to be automatically executed unless vetoed by the operator, or (e) fully automatically, with no operator interaction. They find that the out-of-the-loop problem is significantly greater under full automation than under intermediate levels of automation because under full automation operator's situation awareness decreases greatly.

4.2.4.3 Evaluation of the effects of automation on the human operator

Kleiner et al. (1989) build a scheme that classifies systems by focusing on the contributions of human and automation to the total system, and they develop several different system configurations which result in various levels of automation. Using the scheme, they study the effects of automation on the human operator, i.e. human mental load and system performance. The scheme classifies automation into five levels: (1) Human supplies power and control, (2) automation supports for power or control, (3) automation supplies power and information, human controls, (4) automation supplies power, information and control, human monitors and/or supplies information, (5) automation supplies power, information and controls, no human monitoring required. They suggest that human-machine systems can be conceptually optimized only by developing a clear understanding of the human’s role with the system, and that human capabilities can be assessed against system demands for various levels of automation.

4.2.4.4 Dynamic task allocation or adaptive automation

In the studies on the effect of dynamic task allocation (Huey, 1989; Parasuraman et al., 1996) or adaptive automation (Gluckman et al., 1991) on human performance, different levels of automation have to be employed, because as long as the task allocation scheme is changed the level of automation will be affected. However, in these studies, the level of automation is not clearly defined. The qualitative scale for the level of automation is not suitable to measure the change of the level of automation. This is because the qualitative scale only accounts for what tasks the operator performs as can be seen in Table 4.1 and Table 4.2, and it does not take into account how many tasks the operator performs. This indicates that a quantitative measure for the level of automation is needed. Consequently, the question arises whether the level of automation could be quantitatively measured. Therefore, Wei et al. (1994) proposed a quantitative measure for the level of automation. In Section 4.4, a modified definition will be presented.
4.3 LEVEL OF AUTOMATION OF SUPERVISORY CONTROL IN INDUSTRIES

The development of technology has enhanced the implementation of automation in industries, especially in the control of large-complex systems such as manufacturing, nuclear power plants (NPP), civil aviation, and chemical processes. However, to what level the automation should reach, is always a problem for the system designers and regulatory agencies. In this section, levels of automation in the plant automation of NPPs and in flight control automation of an aircraft are evaluated along the scales described in the previous section.

4.3.1 Level of Plant Automation in Nuclear Power Plant Control

The automation in NPPs can be divided into three categories (IAEA, 1991):
1. Information integration automation: Improvement in monitoring, human-machine interface and information reliability.
2. Plant automation: Improvement in plant operation, control, reliability and maintenance.
3. Operator support and guide automation: Operating support in normal situation, abnormal situations, and maintenance.

These categories correspond to the three types of automation for human-centered automation suggested by Billings (1991), i.e. information automation, control automation, and management automation. To increase operational benefit and safety, more and more computers have been used in the control of NPPs. In several countries, such as Canada, France, and Japan, a great effort has been made to computerize NPP control rooms to a great extent (IAEA, 1990). In the computerized control room, it is easier for the operator to obtain the operational information because of the development of information integration automation. Since more plant control functions, such as plant start-up and plant shutdown, have been automated, the operator has less necessity to perform so-called hard manual tasks that ask the operator to follow a set of procedures. Moreover, a tremendous effort concerning the development of computerized operator support systems has been made. These support systems include operator decision support systems, artificial intelligence and expert systems, and smart procedures for emergency and safety operation. The operator can obtain aids from the support systems and may reduce the level of working stress (IAEA, 1990).

In the stage of development of the 1980s, NPPs were not completely automated. In most cases, the functions of monitoring, diagnosis, and selection or execution of actions were often entrusted to human operators. It is assumed that the human operator in NPPs is to decide what control actions to take and to take such actions on the basis of the information display to him/her (Sheridan, 1992). This is at Level 2 of Sheridan's scale for those control loops. As computers for decision aid and automatic control as well as for display functions are used, the level of automation in supervisory control of NPPs has been increased. The use of expert systems that
are capable of generating conclusions and offering advice, goes a step further and comes to Level 3, or 4 depending on whether the expert systems give a set of action alternatives, or whether they can also select a few alternatives, or whether they also suggest one. Having the computer decide what to do, i.e. only one alternative, goes to Level 5 and Level 6. Further, having the computer decide what to do and do it without human intervention goes to Level 7 or higher. How far to go is a question for the NPP designers as well as for the regulatory authorities (Sheridan, 1992). The development of the computerized control rooms has greatly increased the level of automation in information integration, plant control and operator support systems. Based on Sheridan's scale, Table 4.4 summarizes the levels of plant automation for main control functions that may have been achieved in some countries (Endestad, 1991; IAEA, 1990, 1991, 1992).

Table 4.4 describes the situations of NPP control rooms that are new or under construction at the end of the 1980s. Thus, the level of automation is the most recent development for the plant automation of the NPPs. In Table 4.4, one can see that the NPP plant automation in different countries has different levels. A trend is that the level of automation in the control of many functions in NPPs has been increasing. Since NPPs are considered as high risk systems and according to the main tasks performed by the operators (as addressed below), the level of automation should not exceed Level 8 in Sheridan's scale. The human operators should always be informed about the situations of the plant. The fact that the levels of automation for most of the functions are rated at a level of 7 indicates the change of the human role. Now, the operators perform mainly the monitoring tasks. The tasks that they perform become more cognitive or intelligent. Their behaviours are mainly focused on the knowledge-based-behaviour (Rasmussen, 1983; Stassen et al., 1990). The problems caused by higher level of automation as identified in the previous chapters have become more obvious. This makes the concept of human-centered automation more important in the design and the operation of the NPPs.

The level of plant automation has been evaluated. However, it is necessary to point out what the role that the human operator plays in NPP control is at today's stage of development. The main tasks to be performed by the operator in NPP operation are twofold (IAEA, 1990):
1) Satisfy safety requirements: Keeping the plant in a safe condition at all times, with no effect on the public safety, the environment, and the operators themselves.
2) Produce power at the lowest cost in accordance with production requirements.

In one word, the human operator plays a supervisory role. During the operation, the operators are retained primarily for supervisory tasks and diagnostic capability. Thus, the manual functions are to optimize the plant performance within control limits according to available knowledge, and to handle accidents using emergency procedures or expert systems (IAEA, 1991).
Table 4.4 Level of automation in NPP operation (derived from IAEA, 1990, 1991, and 1992)

<table>
<thead>
<tr>
<th>Operational Functions</th>
<th>Level of Automation based on Sheridan’s Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Canada</td>
</tr>
<tr>
<td>Start-up</td>
<td>7</td>
</tr>
<tr>
<td>Shutdown</td>
<td>7</td>
</tr>
<tr>
<td>Control rod</td>
<td>Not known</td>
</tr>
<tr>
<td>Power regulator</td>
<td>7</td>
</tr>
<tr>
<td>Pressure</td>
<td>7</td>
</tr>
<tr>
<td>Temperature control</td>
<td>7</td>
</tr>
<tr>
<td>Turbine runup</td>
<td>7</td>
</tr>
<tr>
<td>Fuelling machine control</td>
<td>7</td>
</tr>
<tr>
<td>Primary heat transportation</td>
<td>7</td>
</tr>
<tr>
<td>Boiler level control</td>
<td>7</td>
</tr>
<tr>
<td>Safety system trip</td>
<td>7</td>
</tr>
<tr>
<td>Safety system test</td>
<td>7</td>
</tr>
</tbody>
</table>

4.3.2 Level of Aircraft Control Automation

In civil aviation, the driving forces for automation are safety requirements, air pollution, economic benefits, and technological development (Sheridan, 1992). Aircraft can now automatically adjust their throttle, pitch, yaw damping characteristics. They can take off and climb to altitude autonomously, and can maintain altitude and direction in spite of wind disturbances. They can approach and land automatically in zero-visibility conditions. To perform these tasks, aircraft make use of computers programmed in supervisory fashion by pilots and ground controllers (Sheridan, 1992). The computer can assume control of almost every phase of flight in the newest generation of jet aircraft (Stix, 1991). However, a high level of automation in aircraft control has caused problems and accidents as described in Ch. 1.

Billings (1991) has pointed out that aircraft automation taxonomy includes three categories:
1. Control automation: It assists the pilot in guiding the airplane through maneuvers necessary for their safe performance (e.g. autopilot, yaw damper etc.). It includes devices devoted to the operation of aircraft subsystems. These subsystems control fuel, hydraulics, electrical power, pressurization, landing gear and brakes, etc..
2. Information automation: It includes all of displays and avionics devoted to navigation,
digital communication with air traffic control and airline operations, sensor-based environmental surveillance, and large scale on board data-based systems (e.g. electronic library systems).

3. Management automation: It refers to those functions which permit the pilot to exercise strategic, rather than simple tactical, control over the performance of a flight. Management automation directs control automation in the performance of the required functions and directs information automation to keep the pilot informed of the state of the aircraft and of the progress towards satisfying the specified goals (e.g. Flight Management System).

By focusing on control automation, levels of automation for some main functions according to Sheridan's scale and Billings' continuum for the newly introduced aircraft, e.g. Airbus A320 and Douglas MD-11, are summarized in Table 4.5 (Billings, 1991; Stix, 1991). In Billings' continuum, the level of automation indicates the highest level that automation has reached for an aircraft control function.

In many cases, the pilot can over-ride the automation and may take back to perform manual control, such as flight path control. However, the level of automation in aircraft control automation has reached a level as indicated in Table 4.5. The level of automation in flight path control, power control, and aircraft subsystems is higher than 7 showing that the level of automation in the aircraft flight and control is, generally speaking, high. As reviewed in Ch. 1, this high level of automation has caused many human errors and even worse, fatal accidents. The human errors caused alone by the pilot's over-trusting, or over-relaying, on flight path control automation have made several accidents.

While talking about the role of humans in highly automated aircraft system, it should be noted that the aviation system is not an optimal, fully controlled system. Many variables with the system are highly dynamic and not fully predictable (e.g. the severity and movement of weather system). Considering the capabilities that humans have (Appendix I), the role of humans is still very important.

<table>
<thead>
<tr>
<th>Control Functions</th>
<th>Flight path control</th>
<th>Power control</th>
<th>Landing gear</th>
<th>Aircraft subsystems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Billings' Continuum</td>
<td>Autonomous operation</td>
<td>Management by exception</td>
<td>Shared control</td>
<td>Management by exception</td>
</tr>
<tr>
<td>Sheridan's Scale</td>
<td>7 or higher</td>
<td>7 or higher</td>
<td>4 or lower</td>
<td>7 or higher</td>
</tr>
</tbody>
</table>

Table 4.5 Level of civil aircraft control automation in Airbus A320 and MD-11 (derived from Billings, 1991; Stix, 1991)
According to Billings (1991), humans must be retained in a central position in aircraft and in the aviation system even though current aircraft automation is able to perform nearly all of the continuous control tasks and most of the discrete tasks required to accomplish a flight mission and even though humans are far from perfect sensors, decision-makers and controllers.

4.4 QUANTITATIVE DEFINITION OF DEGREE OF AUTOMATION

4.4.1 Motivation

The qualitative scale of the level of automation has been discussed and used broadly as mentioned before. To supervise a system, the operator or automation has to perform a number of subtasks associated with the main tasks, such as process tuning or set-point control, intervening, monitoring, planning, diagnosis, and decision making. However, the qualitative scales do not take into account the number of tasks performed by human and by automation. Furthermore, a numerical value for the level of automation has not been provided. In this section, a measure to quantitatively take into account the level of automation is proposed. This quantitative measure will be called “Degree of Automation”, DofA, and is used to measure how much automation is used in a human-machine system. A quantitative measure for the degree of automation should take into account not only the different levels of automation that exist but also the different natures and number of tasks. It is tempting to derive at such a measure and then, as a next step, compare systems with equivalent degree of automation, with respect to their system performance, operator performance, task demand load, and robustness etc. (Stassen et al., 1993).

We believe that a quantitative index of the level of automation may be used in studies devoted to adaptive automation, or dynamic task allocation, to systematically analyze and evaluate system design and task allocation schemes between humans and machines. The DofA may be employed to measure how much automation has been used because some tasks are reallocated to either the human operator or to the machine. Such studies are aimed at keeping automation at a proper level to avoid the negative effect of automation. The DofA gives a quantitative reference that may be used to assess the effect of a change of the task allocation on human performance, system performance, and human use of automation. It may also be useful for the development of a generic model on the relationship between system performance and task complexity.

4.4.2 Definitions

4.4.2.1 General definition

Sheridan’s scale (1992), for the level of automation, considers the tasks that are allocated to human and to machine. During the allocation stage, tasks can broadly be categorised into three
groups (Price, 1985):

I. Tasks which must be allocated to the human operator;
II. tasks which can be allocated to the operator or to the machine;
III. tasks which must be automated.

Consequently, task allocation methodologies determine the minimum and maximum level of automation. The discussion of the definition of DofA will be focused on the first two categories because they affect the operator's mental load mostly. New human tasks created by automation, such as monitoring automation, are included in Category I.

A quantitative measure of the level of automation, DofA, must take into account not only the effect of how many tasks are automated, but also which types of tasks are automated. Therefore, DofA must be a function of the number of automated tasks. Not all tasks are similar with respect to task complexity, task difficulty and task criticality. So, the definition of DofA should take into account the specific tasks that are automated, i.e. the qualitative characteristics of tasks.

A simple weighting scheme is proposed to take into account the nature of tasks. The DofA will be defined as the ratio between the sum of the task weights of all tasks that are automated and the sum of the weights of all tasks. The DofA can be expressed mathematically as:

\[
\text{Do}fA = \frac{\sum_{i=1}^{N} t_i w_i}{\sum_{i=1}^{N} w_i},
\]  

(4.1)

where \(t_i\) is a task allocation indicator which is 1 if task \(i\) is automated and 0 if the task is performed manually; \(N\) is the total number of tasks; \(w_i\) is the normalized weight for task \(i\). It should be noted that \(t_i\) is subject to change if automation changes due to a redesign of the system or the effect of dynamic task allocation. A task in Eq. (4.1) can only be automated or manually controlled. If an intermediate level for a task between automatic control and manual control exists (i.e. the task can be performed partly by automatic system and partly by the human), such a task should be further broken down into subtasks which can only be automated or manually controlled.

In this model, both the control tasks and the monitoring tasks are included. However, when a task is dynamically allocated to a machine, new tasks may be created, such as the monitoring of the automated task, or the control of automatic control systems. These new tasks can be easily included in the DofA model. Taking the monitoring task as an example, we assume that the monitoring task weight may be different in the case that task \(i\) is automated or manual. Thus, we suppose:
$w_i^c$: the control task weight for a manually controlled task;

$w_i^{mm}$: the monitoring task weight in the case that the control task is manual;

$w_i^{ma}$: the monitoring task weight in the case that the control task is automated.

We also assume that the monitoring requires more effort when a task is manual than when a task is automated, i.e. $w_i^{mm} > w_i^{ma}$. When a task is automated, the difference between $w_i^{mm}$ and $w_i^{ma}$ reflects the effect of automation.

Hence, Eq. (4.1) can be rewritten as follows:

$$D of A = \frac{\sum_{i=1}^{N} \{t_i w_i^c + t_i (w_i^{mm} - w_i^{ma})\}}{\sum_{i=1}^{N} (w_i^c + w_i^{mm})} = \frac{\sum_{i=1}^{N} \{t_i (w_i^c + w_i^{mm}) - t_i w_i^{ma}\}}{\sum_{i=1}^{N} (w_i^c + w_i^{mm})},$$

(4.2)

where $w_i^{mm}$ is used in the denominator by considering $w_i^{mm} > w_i^{ma}$. It should be stressed that for an automated task, $w_i^{mm} = 0$, and for a manual task, $w_i^{ma} = 0$.

In Eq. (4.2), the parameter $w_i$ is introduced which equals:

$$w_i = w_i^c + w_i^{mm},$$

then it follows:

$$D of A = \frac{\sum_{i=1}^{N} t_i (w_i - w_i^{ma})}{\sum_{i=1}^{N} w_i}.$$

(4.3)

So, in this way we can incorporate the monitoring weight for an automated task.

A measure for a task weight can be obtained in two ways. One way is by considering how a task affects system performance. This is the system designer’s intended effect of a task on the system performance. In this way, we define the first measure of the task weight. Another way of obtaining a task weight is by considering how the task affects operator performance and mental load. The effect on the operator can be discussed from two aspects. One aspect takes into account how much workload a task will impose on the operator, i.e. the demanded workload by a task. We use the demanded workload to define the second measure of task weight. The other aspect takes into account how much workload is experienced by the operator. This is the mental load perceived by the operator. The mental load will be used to define the third measure of the task weight.
4.4.2.2 DofA based on task effect on system performance

Task Effect on System Performance, TES, is used to indicate task criticality, i.e. how much a task affects system performance. TES is defined as the system performance error caused by ignoring the execution of a task. The weight based on the TES, $w^{\text{TES}}$, is obtained by considering the value of the system performance error. When the $w^{\text{TES}}$ is used in Eq. (4.1), the DofA is an indication of the fraction of system performance controlled by automation, and is noted as DofA$^{\text{TES}}$.

4.4.2.3 DofA based on task demand load

A task, when executed by the human, imposes workload on the operator. The imposed workload is defined as Task Demand Load, TDL, (Stassen et al., 1990). The TDL is a function of the nature of the task, such as task complexity including time constraints, task uncertainty, task frequency, and task performance requirement (Salvendy and Bi, 1995; Sheridan, 1992). The TDL is inherent to a task and independent from the human. When the TDL is taken as a measure for the weight of a task, $w^{\text{TDL}}$, the DofA becomes an indication of the fraction of the maximum task demand load that is taken care of by automation, and is noted as DofA$^{\text{TDL}}$.

4.4.2.4 DofA based on mental load

When the operator performs the task, he/she perceives a workload. The perceived workload for performing a task is dependent on the human and is called Task Mental Load, TML, (Stassen et al., 1990). Thus, a measure for the TML may be used to quantify how much a task has affected the human operator. This measure will be used to derive a task weight based on TML, $w^{\text{TML}}$.

The TML is a function of a number of factors such as task complexity, task difficulty, task uncertainty, and task arrival rate (Bi and Salvendy, 1994), and also of human physical and psychological status (Zijlstra, 1993). Thus, allocation of some tasks to machines will influence the operator’s mental load. By using a TML measure, the DofA indicates the mental load that is experienced by the operator in a partly automated system relative to the mental load that would have been experienced in case the system was not automated; it is noted as DofA$^{\text{TML}}$.

The three measures of the DofA are depicted in Figure 4.1. The implementation of the DofA model using the above measures for the task weights and the relations between these DofA measures and system performance for the experimental systems will be presented in the following chapters.
Figure 4.1 Schematic presenting of three measures of DofA (HUMIF = HUman Machine InterFace).

4.5 APPROACHES TO ESTIMATE TASK WEIGHTS

The DofA has been defined by a weighting scheme in three measures. To compute the DofA, task weights, $w^{TES}$, $w^{TDL}$, and $w^{TML}$, have to be estimated or computed. In this section, the approaches to define the weighing factors for TES, TDL and TML are proposed.

4.5.1 Task Weight Based on Task Effect on System Performance

By its very definition, $w^{TES}$ is independent of control actions of the human operator for that task. To estimate $w^{TES}$, the TES must firstly be computed. Two approaches are proposed to obtain the TES for a control task.

1) Simulation technique. A system dynamic model is needed to simulate the system. In this approach, the system performance for a reference case, where all tasks are executed correctly and promptly, is compared to a case where one specific control task is ignored. For the reference case, a task is executed through a control algorithm, giving the control inputs according to the set-point requests. The measured system performance in this case is defined as reference performance. Subsequently, each task is not executed (no system input is given for the request value). Then, the system performance (called TES performance) is measured and is compared to the reference performance. In each case, the difference in system performance with respect to the reference case is used to estimate the TES associated with the task.

2) Consequence analysis. For a practical system or a simulated system, a target detection function might be defined. Thus, if a task is not performed, or not properly performed, the percentage of the targets, e.g. 30%, that will be lost, is used to assign a number for the TES, e.g. 0.3 (IAEA, 1991; LaSala, 1993). For the TES of monitoring tasks, a measure, such as
Carbonell's measure of attention cost (e.g. Moray, 1986), can be employed to calculate the TES when a monitoring task is not performed. The Carbonell's measure addresses the relation between cost and probability of ignoring the execution of a monitoring task.

After the TES is obtained for each task, the task weight can directly be defined as the value of the TES; or a normalized task weight is defined by the TES value associated with the task divided by the sum of the TES for all tasks. The implementation is described in the following chapters.

4.5.2 Task Weight Based on Task Demand Load

Task weight, $w^{\text{TDL}}$, has to be computed from the task demand load. The TDL is not dependent on the human operator. Thus, the TDL may be analyzed by analytical methods (Hill et al., 1987) and may be predicted by some system engineering parameters, such as task arrival rate, task complexity, and task uncertainty (Salvendy and Bi, 1995). The methods to estimate task workload without the-operator-in-the-loop have been reviewed in Ch. 3. Here only a few are mentioned by focusing on the prediction of task demand load.

1) Mathematical models
   
   This approach requires to identify variables that influence workload directly or indirectly, determine the lawful relationships by which these variables interact, and establish how the resultant workload predictions drive predictions of the performance.

2) Task analysis methods

   *Hierarchical Task Analysis*, HTA, may be used to identify tasks. From hierarchical structures of tasks, a number that represents the relative amount of the task workload according to the task complexity and other parameters can be assigned to a task. This number may be used as a measure of the TDL.

   *Timeline Task Analysis* is a broadly used method to estimate task demand load by a ratio of Time required, $T_r$, to perform a task over the Time available, $T_a$, yielding $T_r / T_a$, which is a measure of the TDL.

3) Expert opinion

   The TDL for a task may be rated by experts. In this way, the relative value of the TDL is obtained. However, a rating scale may need to be defined.

When the TDL of a task is obtained, the task weight can be directly defined as the value of the TDL, or a normalized task weight is defined by the TDL value associated with the task divided by the sum of the TDL for all tasks. The implementation to compute the TDL as well as $w^{\text{TDL}}$
will be presented in Chs 5 and 6 on the basis of experimental systems.

4.5.3 Task Weight Based on Mental Load

The task weight, $w^{\text{TML}}$, is human operator dependent. To compute $w^{\text{TML}}$, the mental load experienced by the operator while he/she executes a task has to be assessed. The TML may be assessed by empirical techniques with the operator-in-the-loop (Hill, et al., 1987; Wierwille and Eggemeier, 1993). A number of approaches exist to assess mental load as reviewed in Ch. 3.

When the TML for a task is assessed, the task weight can be directly defined as the value of the TML, or a normalized task weight is defined by the TML value associated with the task divided by the sum of the TML for all tasks.

After the task weights are determined based on one of the three measures, using Eq. (4.1), the DofA can be computed.

4.6 CONCLUSIONS

In this chapter, the qualitative measures for the degree of automation in supervisory control, their usage in scientific studies, and the levels of plant automation in nuclear power plants and in civil aircraft are reviewed. The automation in NPPs and in aircraft has reached a level that the human operator only plays a supervisory role and deals with optimization of the system operation and the beyond designed situations.

Motivated by the qualitative measures and the needs for studies on the effect of automation and system design, a quantitative measure for the degree of automation is proposed. The quantitative measure, DofA, is a function of the number and the nature of the tasks to be automated. This means that a quantitative definition for the degree of automation should include not only the number, but also the characteristics of the tasks to be performed.

The DofA measure is initially developed to achieve a measurement approach for the degree of automation in supervisory control (using DofA$^{\text{TDL}}$) or during system evaluation (using DofA$^{\text{TML}}$ and DofA$^{\text{TES}}$). It presents a possible direction toward the development of a practical measure for use in system design. We believe that such a quantitative index of the level of automation may be used in studies devoted to adaptive automation, or dynamic task allocation to systematically analyze and evaluate system design and task allocation schemes between human and machine.
The use of this measure will greatly depend on the definitions of tasks, the measurement of TML, the estimation of TDL, and the calculation of TES. For example, TES may not be applicable to highly critical tasks (continuous and discrete). This requires different approaches to obtain the TES value.
Chapter 5

Determination of Degree of Automation of Linear Systems

The quantitative measure of the level of automation, Degree of Automation or DofA, in supervisory control is implemented in an artificial, experimental system. Based on human controlled experiments, analyses of task effect on system performance, prediction of task demand load and assessment as well as prediction of mental load are carried out. Each experiment has a different DofA. The effect of a change of the DofA on system performance and mental load is investigated. It is found that system performance may level off when the DofA is above a certain level. The experimental data show that when the operator controls a partly automated system, his perceived mental load may be predicted from the mental load for each task component obtained from the situation where all tasks are manually controlled.

5.1 INTRODUCTION

A quantitative definition to measure the level of automation in supervisory control has been introduced in Ch. 4. Degree of Automation, DofA, is defined as a numerical parameter based on either the Task Effect on System performance, TES, the workload of a task imposed on the human operator, Task Demand Load or TDL, and the workload the human operator perceived during performance of the tasks, Task Mental Load or TML. The DofA can be used to express the level of automation as a function of the variation of task allocation between human and machine, and to quantitatively study the effect of this variation on system performance and human performance.

As defined in Ch. 4, three measures of task weights have to be derived from the system that is supervised by the operator in order to calculate the DofA. Two of the parameters, $w^{\text{TES}}$ and $w^{\text{TDL}}$, are independent of the presence of the human operator. They can be obtained off-line without the operator. Only $w^{\text{TML}}$ is obtained with the operator in the loop.
A series of experiments involving human operator control on an artificial system were conducted in the laboratory. Based on the system and the experimental results, the parameters \( w^\text{TES}, w^\text{TDL}, \) and \( w^\text{TML} \) could be assessed using the approaches proposed in previous chapters.

This chapter will demonstrate how to determine the DoFA from the experimental data. Consequently, the effects of a change in the degree of automation on system performance and on mental load will be investigated. Because the TML reflects how much workload the operator perceives from the TDL, the relationship between TDL and TML is analyzed. When TML changes, system performance may change accordingly. Thus, the relationship between system performance and TML will also be discussed.

5.2 METHOD

5.2.1 Experimental System

The experimental configuration is depicted in Figure 5.1. The experimental task is to control the system to the requested Set-Points, SP. The experimental task consists of task components or subtasks, that is, the experimental task can be further divided into a number of monitoring and control subtasks which can be allocated to the operator or to the machine. The monitoring tasks will be considered as a part of the control task since it is impossible to identify separately these task components in our experimental set-up. The system is adapted from a system developed by Stassen et al. (1993). It is implemented on a Pentium PC using a graphic language, LabVIEW (National Instruments, 1993).

![Figure 5.1 Experimental environment. Solid lines represent the signal flow within the system.](image-url)
The experimental system consists of 12 first-order systems which constitute 12 cells. Each cell has a control input \( u_i \), possible inputs from other cells, and an output (Figure 5.2). The number of cells has been based on the following considerations: (1) If the number is smaller than 8, the system complexity is too low to detect differences in operator performance (Stassen et al., 1993); (2) it has been shown that a system with 16 cells is too difficult to operate and gets too crowded on the display (Stassen et al., 1993); and (3) mimics on the display gets too small on the screen to control. Each cell has a separate cell interface as depicted in Figure 5.3. The cell interface consists of an alarm sign, an SP request call indicator, a control slide, digital and graphical displays of the SP value, actual output and predicted output. When the difference between the SP value and the actual output of a cell exceeds an allowed value (0.4), the alarm sign becomes red. When the SP value changes, the SP request call indicator becomes blue. The predicted output for a cell is calculated taking into account the dynamics of the first-order function. This output predicts the trend of the actual output.

The visualization of the 12 cell interfaces fills up a 17 inch VDT (With 16 cells, as used by Stassen et al., the size of each cell interface would become too small, making it too difficult to operate the system. Stassen et al. used a slightly different cell interface). The 12 cell interfaces are arranged on the screen in 3 rows of 4 cells. They are sequentially numbered starting at the left of the top row as shown in Figure 5.4.

The connection between the cells’ inputs and outputs can be represented by a connection matrix \( K \) defined below:

\[
K = \begin{bmatrix}
  k_{11} & k_{12} & \cdots & k_{1N} \\
  k_{21} & k_{22} & \cdots & k_{2N} \\
  \vdots & \vdots & \ddots & \vdots \\
  k_{N1} & k_{N2} & \cdots & k_{NN}
\end{bmatrix},
\]  

(5.1)
where \( k_{ij} (i, j = 1, 2, \ldots, N, i \neq j) \) is the coupling strength between cell \( i \) and cell \( j \). \( k_i \) is the gain in the feedback loop of Figure 5.2. \( N (= 12 \text{ in the experiment}) \) is the number of cells.

![Diagram](image)

Figure 5.3 Cell interface. Set-point value and predicted output are also displayed in curves. Warning, alarming, SP call indicator and output values are presented in different colors. The control input can be adjusted by either mouse and keyboard.

![Diagram](image)

Figure 5.4 Human-Machine interface. Indications for warning (WARN) and alarm (ALARM), and the cell index (e.g., C 1) and the SP call (CALL) are overlapped on the display.
Figure 5.5 Diagram of the connections among the cells used in the experiments.

In this study, we shall only consider $k_{ij} = k$ for $i > j$ and $k_{ij} = 0$ for $i < j$, indicating that two cells are connected in a forward way only (Figure 5.5). The connection matrix, therefore, becomes a lower-triangular matrix. The coupling coefficient, $k_{ij}$ has a value of 0.5 (in accordance with the conclusion by Stassen et al.; a system with such a coupling strength will be moderately difficult to control). Each cell has a time constant of 10 seconds and a steady state gain of 1 ($k_{ii}$ in Figure 5.2 = -1.0).

According to Figure 5.2 and Figure 5.5, the state-space model of the experimental system can be expressed as:

$$\dot{x} = Ax + Bu,$$  \hspace{1cm} (5.2)

where in our case $x$ is an $N \times 1$ state vector, and $u$ is an $N \times 1$ control input vector. $A$ is an $N \times N$ system matrix, and $B$ is an $N \times N$ input matrix. According to the cell structure as shown in Figure 5.2, and with $k_{ii} = 0$ for $i < j$, for cell $i$, the following equation exists:

$$\dot{x}_i = \frac{1}{\tau_i} (k_{ii}x_i + k_{i1}x_{i-1} + k_{i2}x_{i-2} + \cdots + k_{i1}x_1 + u_i).$$

Consider this equation and note that each cell has the same time constant, $\tau_1 = \tau_2 = \cdots = \tau_{12} = \tau$, the relation between the system matrix $A$ and the connection matrix $K$ can be expressed as:

$$A = \frac{1}{\tau} K,$$  \hspace{1cm} (5.3)

$$B = diag[\frac{1}{\tau}, \cdots, \frac{1}{\tau}].$$

Note that all states can be observed and are controllable. Applying Eq. (5.2) for supervisory control situations, one can define and compute system complexity (Stassen et al., 1993), and one can also compute the task effect on system performance.
The experimental set-up is such that each cell can be controlled automatically or manually depending on the protocol. To realize the automatic control function, a PI controller, consisting of the following control law, is used:

$$u_i[k] = u_i[k-1] + K_p \{e_i[k] + e_i[k-1]\} + T_i K_i e_i[k],$$  \hspace{1cm} (5.4)

where $u_i[k]$ is the control output of the PI controller; $e_i[k]$ is the error between the SP value and the cell output; $k$ stands for the $k^{th}$ sampling period, and $i$ for the $i^{th}$ cell. We have tuned the system such that the proportional coefficient $K_p = 0.25$, the integral coefficient $K_i = 0.18s^{-1}$, whereas the sampling time $T_s = 4s$.

5.2.2 Operator’s Task

As mentioned before, the experimental task was to supervise the system by tuning the SP requested. The SP requested were randomly generated. Thus, requests for an SP change appeared randomly on a particular cell and randomly in time. The operator’s monitoring task was to monitor the SP requests and values, cell outputs and alarm signs for all the manually controlled and automated cells. The operator’s control task was to generate an input for the appropriate cell(s), by means of a mouse or keyboard, and to adjust the cell’s output to the new SP, while maintaining the other cells at their current set-points. Note that all cells with a higher cell index number than the cell that receives a new SP are affected by the output of that cell.

5.2.3 Experimental Sessions

For the experiment, 28 sessions were designed, including the fully manually controlled session; each session lasted 15 minutes. Table 5.1 presents the system configurations for 19 typical sessions. In other 9 sessions, only one or two cells are manually controlled. The parameters that were changed were the number of manually controlled cells and the locations of these cells within the system configuration. The SP request for each cell, the average time interval between two subsequent SP requests (40s) and the coupling coefficients were not changed.

5.2.4 Subjects

Six students (all male) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology, with an average age of 22.8 years, participated voluntarily as operators. They received a fixed fee for their participation. Before formal sessions started, the subjects were given a one hour training session to learn to supervise the system and to get experiences in filling out the mental load assessment forms (Appendix 2). Each subject performed 28 sessions.
Table 5.1 Task allocation matrix of 19 experimental sessions. #MCC means the number of manually controlled cells. ■ represents that a cell is controlled by the operator; and □ represents that a cell is controlled by the PI controller.

<table>
<thead>
<tr>
<th>#MCC</th>
<th>Cell 1</th>
<th>Cell 2</th>
<th>Cell 3</th>
<th>Cell 4</th>
<th>Cell 5</th>
<th>Cell 6</th>
<th>Cell 7</th>
<th>Cell 8</th>
<th>Cell 9</th>
<th>Cell 10</th>
<th>Cell 11</th>
<th>Cell 12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.2.5 System Performance Measurement

For cell \( i \), the mean error, \( ME_i \), of the absolute difference between the actual output and the desired value, SP, was taken as the performance of a cell (van der Veldt, 1984), and was computed as follows:

\[
ME_i = \frac{\sum_{k=0}^{M} |x_i[k] - SP_i[k]|}{M + 1},
\]  

where \( x_i[k] \) is the output of cell \( i \) for the computing period \( k \), and \( x_i[0] = 0 \); \( SP_i[k] \) is the set-point request for cell \( i \) in the computing period \( k \), and \( SP_i[0] = 0 \). \( M \) is the total number of sampling periods, and \( i \) indicates the \( i^{th} \) cell.

Since each cell had the same nature, the mean value of all cell errors was taken as a measure for the performance of the entire system, i.e. the System Error, \( SE \). The \( SE \) was calculated by:
\[ SE = \frac{\sum_{i=1}^{N} ME_i}{N}, \]  

(5.6)

where \( N \) is the total number of cells.

In order to present the data in a way that a larger value indicates a better performance, the System Performance, \( SPF \), is presented as one minus the system error, i.e.

\[ SPF = 1 - SE. \]  

(5.7)

This implies that a mean error score of zero gives a system performance of one. In this way, error scores larger than one would cause negative performance. This did not occur. Furthermore, for each system configuration the system performance is averaged over all subjects.

5.2.6 Estimation of Weights

5.2.6.1 Task effect on system performance

In this experiment, the simulation technique as suggested in Ch. 4 was used to estimate the task effect on system performance. The system performance for a reference case (reference performance), where all control subtasks were executed correctly and promptly, was compared to the system performance for a specific case, TES performance, where one specific control subtask was not executed. The reference case was realized by automating all cells. In this case, each cell received randomly only one SP request with a magnitude of 1. Since \( ME_i \), value can be affected by the experimental length (Eq. 5.5), the experimental time for a session was 280 seconds. To let each cell have an SP request, the SP request was given to each cell separately with an average time interval of 20 seconds between two consecutive SP requests, rather than 40 seconds. Under the same conditions, 12 simulation runs were done in which for each simulation one control subtask was not executed (the associated cell was not controlled). Each simulation run started at a steady state and the TES performance was computed. The difference, \( d_i \), between the reference performance and the TES performance for each cell was used as a measure for the task effect on system performance and provided the basis for the task weights, \( w_{i}^{TES} \). The weights are computed as follows:

\[ w_{i}^{TES} = \frac{d_i}{\sum_{k=1}^{N} d_k}. \]  

(5.8)
5.2.6.2 Prediction of task demand load

As mentioned in the Chs 3 and 4, the TDL can be estimated by task analytic methods, simulation models (Hill et al., 1987) and expert opinions (Hamilton and Bierbaum, 1990; Laughery, 1989). Salvendy and Bi (1995) have proposed that the TDL can be predicted by using analytical techniques based on some system engineering parameters: Task arrival rate, task complexity, and task uncertainty. According to the experiment conducted by Bi and Salvendy (1994), task complexity plays an important role in the TDL when task uncertainty and time stress are constant. In the present study, these conditions were met and thus, the TDL was computed on the basis of task complexity only. According to Sheridan (1992), task complexity is a property of the task independent of the operator and is determined by, for instance, the number of required actions, the number or variety of stimuli, the time constraints, and the degree of overlap of multi-task demands.

According to the metric used by Hess and McNally (1986) and proposed by Weirwill and Connor (1982), the TDL can be measured by the number of control inputs used by an operator for each task. The relation between the TDL and task complexity can be determined as follow. Consider cell $i$ which has $i$-$l$ inputs from other cells (1, 2, ..., $i$-$1$) and one control input (Figure 5.5). An operator controlling the cell has to respond to the SP request, and has to compensate for the inputs from the other $i$-$l$ cells caused by control actions due to SP requests given to these cells earlier. The TDL for a response to an SP will be different from the TDL for compensatory actions. Thus, it is plausible to assume that the TDL of cell $i$ ($TDL_i$) is a function of the SP inputs and the TDL associated with the number of compensatory inputs (Figure 5.2). For simplification, we propose the TDL definition for our experimental system as follow:

$$TDL_i = TDL_{i}^{SP} + \sum_{j=1}^{i-1} TDL_{j}^{COMP}, \text{ for } i > 1 \text{ and } TDL_1 = TDL_1^{SP}. \quad (5.9)$$

where $TDL_{i}^{SP}$ is the TDL induced by an SP input and $TDL_{j}^{COMP}$ is the TDL caused by a compensatory input from the coupling disturbance induced by the output of cell $j$. It should be stressed that in this experiment $j < i$. In Eq. (5.9), we take into account only the direct coupling disturbance from other cells, i.e. the disturbance when the output of cell $j$ is directly coupled to the input of cell $i$. We do not take into account the indirect influence of cell $j$ to cell $i$. For example, cell $j$-$l$ may affect cell $i$ via cell $j$.

The normalized weight based on TDL is defined as:

$$w_{i}^{TDL} = \frac{TDL_{i}}{\sum_{k=1}^{N} TDL_{k}}, \quad (5.10)$$

where $N$ is the total number of subtasks.
The parameter $TDL_i$ can be obtained as follows. The coupling strength ($k_i$) between the output of cell $j$ and cell $i$ ($j < i$) is for all cells 0.5. The gain between an SP input and any cell is 1.0. We assume that $TDL_{SP}^i$ is the same for all cells, $TDL_{SP}$; and $TDL_{COMP}^i$ is the same for all compensatory actions, $TDL_{COMP}$. The ratio between the $TDL_{COMP}$ and the $TDL_{SP}$ is equal to the ratio between these gains, i.e. $TDL_{COMP} = 0.5 TDL_{SP}$. According to Eq. (5.9), this means that, for example, the TDL of cell $i$ that has one SP input and $i-1$ compensatory inputs is:

$$TDL_i = 1 \cdot TDL_{SP} + (i - 1) \cdot TDL_{COMP} = TDL_{SP} + (i - 1) \cdot 0.5 \cdot TDL_{SP} = 0.5(1 + i) \cdot TDL_{SP}. \quad (5.11)$$

Substituting Eq. (5.11) into Eq. (5.10), the $w_{TDL}$ for a task is:

$$w_{TDL} = \frac{0.5(1 + i) \cdot TDL_{SP}}{\sum_{k=1}^{N} \frac{0.5(1 + k) \cdot TDL_{SP}}{N(1 + k)}} = \frac{1 + i}{2(i + 1)}. \quad (5.12)$$

### 5.2.6.3 Overall task demand load

The Overall Task Demand Load, or OTDL, expressed the total TDL of all cells that are manually controlled for a given system configuration. Thus, the OTDL is defined as the sum of all TDLs of those cells that are manually controlled. If we use $t_i$ as a cell allocation indicator, i.e. $t_i = 0$ for a manually controlled cell and $t_i = 1$ for an automatically controlled cell, the OTDL is:

$$OTDL = \sum_{i=1}^{N} (1 - t_i) \cdot TDL_i. \quad (5.13)$$

### 5.2.6.4 Task weight based on task mental load

In the present study, the RSME (Zijlstra, 1993) was used to assess mental load. The RSME was presented to the subjects on paper (Appendix 2). Immediately after each session, subjects were asked to give numbers according to the rating scale that corresponded best to the experienced mental load for controlling the cells and the overall system. So, in the that case all cells were manually controlled, a subject gave 13 ratings - one for the overall system control and 12 for the TML of each cell.

The $w_{TML}$ is calculated by the TML which was experienced by the operator while controlling a cell (or a subtask). We use $TML_{i}^{FM}$ ($i = 1, 2, \ldots, N$) to indicate the TML of cell $i$ measured after the subject controlled the system fully manually. Thus, $TML_{i}^{FM}$ is used to calculate the $w_{TML}$ for a subtask. For subtask $i$, its weight is:
\[
\omega^TML_i = \frac{TML_{FM}^M}{\sum_{k=1}^{N} TML_{FM}^k}
\]  
(5.14)

5.2.6.5 Overall mental load

The Overall Mental Load, OML, is introduced to indicate the mental load experienced by the operator in operating the entire system (i.e. performing the experimental task). The OML was assessed by the operator after each experimental session. The OML experienced when the operator controls a system fully manually (FM) is indicated by the parameter \( OML_{FM} \).

5.3 RESULTS AND DISCUSSION

5.3.1 Task Weights based on Task Effect on System Performance

Figure 5.6 shows the effect of a single task on system performance for the reference system with \( N = 12 \) and \( k = 0.5 \). The reference performance (all cells automatically controlled) for this case was 0.812, indicated by a straight line in Figure 5.6. The data points on the curve (TES performance) were obtained from simulations where only one cell was not controlled, while all other cells were automatically controlled. The occurrence of the maximum for the TES performance at index number 3 in Figure 5.6 will be discussed later.

![Figure 5.6 Task effect on system performance. A larger score indicates a better performance.](image-url)
Table 5.2 Task weights from TES, $w^{\text{TES}}$, (Single task effect on system performance)

<table>
<thead>
<tr>
<th>Not Controlled Cell</th>
<th>Difference between Reference and TES Performance</th>
<th>Task Weight $w^{\text{TES}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.029</td>
<td>0.023</td>
</tr>
<tr>
<td>2</td>
<td>0.013</td>
<td>0.010</td>
</tr>
<tr>
<td>3</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>4</td>
<td>0.016</td>
<td>0.013</td>
</tr>
<tr>
<td>5</td>
<td>0.052</td>
<td>0.041</td>
</tr>
<tr>
<td>6</td>
<td>0.085</td>
<td>0.067</td>
</tr>
<tr>
<td>7</td>
<td>0.116</td>
<td>0.091</td>
</tr>
<tr>
<td>8</td>
<td>0.145</td>
<td>0.114</td>
</tr>
<tr>
<td>9</td>
<td>0.170</td>
<td>0.134</td>
</tr>
<tr>
<td>10</td>
<td>0.193</td>
<td>0.152</td>
</tr>
<tr>
<td>11</td>
<td>0.213</td>
<td>0.168</td>
</tr>
<tr>
<td>12</td>
<td>0.231</td>
<td>0.182</td>
</tr>
</tbody>
</table>

Using Eq. (5.8), a task weight based on TES, $w^{\text{TES}}$, can be calculated. Table 5.2 shows the task weights based on the Eq. (5.8). The occurrence of the minimum for the $w^{\text{TES}}$ at index number 3 in Table 5.2 will be discussed later.

5.3.2 Task Weights based on Task Demand Load

Table 5.3 presents TDL values based on Eq. (5.11) of all subtasks (cells) and the associated weights based on Eq. (5.12).

Figure 5.7 displays the relation between the OTDL and the number of manually controlled cells. From the figure, one can conclude that even for the same number of manually controlled cells, the OTDL may be different. This is because the manually controlled cells are at different locations within the system configuration.

Figure 5.8 shows the relation between the OTDL and the performance of the system controlled by the operators. The system performance decreased with an increase in OTDL.
Table 5.3 Task Demand Load and associated weights, $w^{TDL}$ (Eq. 5.12), Task Mental Load for a fully manually controlled system and associated weights, $w^{TML}$ (Eq. 5.14)

<table>
<thead>
<tr>
<th>Cell</th>
<th>TDL</th>
<th>$w^{TDL}$</th>
<th>$TML^{PM}$</th>
<th>$w^{TML}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>1.0</td>
<td>0.022</td>
<td>8.50</td>
<td>3.34</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>0.033</td>
<td>11.83</td>
<td>7.20</td>
</tr>
<tr>
<td>3</td>
<td>2.0</td>
<td>0.044</td>
<td>17.67</td>
<td>6.50</td>
</tr>
<tr>
<td>4</td>
<td>2.5</td>
<td>0.056</td>
<td>22.00</td>
<td>8.06</td>
</tr>
<tr>
<td>5</td>
<td>3.0</td>
<td>0.067</td>
<td>29.50</td>
<td>5.74</td>
</tr>
<tr>
<td>6</td>
<td>3.5</td>
<td>0.078</td>
<td>35.33</td>
<td>11.79</td>
</tr>
<tr>
<td>7</td>
<td>4.0</td>
<td>0.089</td>
<td>40.33</td>
<td>12.34</td>
</tr>
<tr>
<td>8</td>
<td>4.5</td>
<td>0.100</td>
<td>45.17</td>
<td>16.59</td>
</tr>
<tr>
<td>9</td>
<td>5.0</td>
<td>0.111</td>
<td>54.67</td>
<td>19.53</td>
</tr>
<tr>
<td>10</td>
<td>5.5</td>
<td>0.122</td>
<td>60.50</td>
<td>22.49</td>
</tr>
<tr>
<td>11</td>
<td>6.0</td>
<td>0.133</td>
<td>61.50</td>
<td>23.70</td>
</tr>
<tr>
<td>12</td>
<td>6.5</td>
<td>0.144</td>
<td>65.67</td>
<td>28.12</td>
</tr>
</tbody>
</table>

Figure 5.7 Overall task demand load vs. number of manually controlled tasks. For a fixed number of manually controlled elements, different OTDLs were found. The digits beside a data point indicate the index range of the manually controlled cells (e.g. 1-4: Cells 1, 2, 3, 4 are manually controlled).
5.3.3 Task Mental Load and Task Weights

The average OML for all subjects against the number of manually controlled cells is plotted in Figure 5.9; the standard deviations of the average OML are presented in Table 5.4. The different OML values at the same number of manually controlled cells were the result of the different locations of the manually controlled cells within the system configuration for each experimental session.

Figure 5.9 Overall mental load vs. number of manual tasks. For a fixed number of manually controlled cells, different OML were found. The digits beside a data point indicate the index range of the manually controlled cells (e.g., 1-4: Cells 1, 2, 3, 4 are manually controlled).
Table 5.4 DoA, system performance and overall mental load. (□ an automated cell; ■ a manually controlled cell.)

<table>
<thead>
<tr>
<th>Manually Controlled Cell</th>
<th>Degree of Automation, DoA</th>
<th>System Performance</th>
<th>OML (RSME)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Location</td>
<td>TES based</td>
<td>TDL based</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.977</td>
<td>0.978</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.990</td>
<td>0.967</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.994</td>
<td>0.956</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.987</td>
<td>0.944</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>0.818</td>
<td>0.856</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.967</td>
<td>0.945</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.971</td>
<td>0.934</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.964</td>
<td>0.922</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.981</td>
<td>0.900</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>0.650</td>
<td>0.723</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.948</td>
<td>0.845</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.687</td>
<td>0.666</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>0.364</td>
<td>0.490</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.840</td>
<td>0.700</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.537</td>
<td>0.534</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.540</td>
<td>0.499</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.462</td>
<td>0.467</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>0.159</td>
<td>0.301</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.635</td>
<td>0.511</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.313</td>
<td>0.335</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>0.052</td>
<td>0.155</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.350</td>
<td>0.278</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.205</td>
<td>0.166</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>0.033</td>
<td>0.055</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
If all cells were manually controlled, each subject had to give a RSME rating for each cell, and for the entire system (i.e. OML\textsuperscript{FM}). The average TML over all subjects and its standard deviation are listed in Table 5.3. The \(w^\text{TML}\) computed by Eq. (5.14) is shown in Table 5.3 as well.

### 5.3.4 Computation of DofA

The DofA was computed for different system configurations using the definition in Eq. (4.1), and for different task weight measures presented in the previous sections. Table 5.4 presents the DofAs of most experimental sessions.

It should be noted that DofA\textsuperscript{TES} in Table 5.4 has a maximum of 0.994 for cell index number 3, as has been indicated before in Figure 5.6 for TES performance, and in the Table 5.2 for \(w_3^\text{TES}\). This can be understood by considering the effect of the outputs of Cell 1 and Cell 2 on the input of Cell 3, even if this cell is not controlled. In general, when the TES performance is analyzed, an uncontrolled cell \(i\) receives \(i-1\) input signals with magnitude \(k\) from \(i-1\) automated cells. Thus the output of cell \(i\) becomes eventually \((i-1)k\). The steady state error of the uncontrolled cell \(i\) is the absolute value of the difference between the set-point request 1, which is not answered, and the inputs from the \(i-1\) cells which have the value \(k = 0.5\), i.e.:

\[
err_i = \left| 1 - (i - 1)k \right|. \tag{5.15}
\]

Hence the minimum of this function occurs at

\[
i = 1 + \frac{1}{k}, \tag{5.16}
\]

which is for our case with \(k = 0.5\): \(i = 3\). This formula predicts that for a lower value of \(k\), meaning less strong interconnection, a higher cell index number will be found for the minimum. This shows how TES depends on the system configurations.

From Table 5.4, one can find that the numerical value for the DofA obtained on the basis of the measure \(w^\text{TDL} (DofA^\text{TDL})\) and the value for the DofA obtained on the basis of the measure \(w^\text{TML} (DofA^\text{TML})\) are very close. This suggests that mental load may be predicted by task demand load which is obtained from engineering parameters related to task complexity.

The DofA\textsuperscript{TES} is less related to the other two. This difference can be explained as follows. Both \(DofA^\text{TDL}\) and \(DofA^\text{TML}\) take into account the dependency among tasks. For instance the definition of the TDL takes into account the coupling between cells. The compensatory action is also comprised in the definition of the TML in a fully manually controlled system. However, the DofA\textsuperscript{TES} does not consider the dependency among tasks because the task weights were derived from effects of single subtasks on the system performance.
5.3.5 Relation between System Performance and DofA

Relationships between the system performance and the three measures for DofA are calculated and shown in Table 5.5. A second order polynomial was used to fit the data ($R^2 > 0.92$). For the sake of the argument, Figure 5.10 shows the relationship between system performance and DofA$^{TDL}$. Similar trend lines could be found for the DofA$^{TES}$ and the DofA$^{TML}$. One may conclude from these results that when the DofA increases above a certain level, for example, DofA ≥ 0.7 (Figure 5.10), the improvement in performance becomes very small. For such high DofA, it may not be cost effective to improve the system performance furthermore. This implies that an operator can achieve system performance as good as what the completely automated system would achieve. The advantages of keeping the operator in control are clear: Monitor automated system, and stay aware of the system's main variables (Billings, 1991). These findings are in agreement with the research conducted by Endsley and Kiris (1995).

Table 5.5 Relation between system performance, overall mental load and DofAs. Note: SPF = $a \cdot (\text{DofA})^2 + b \cdot \text{DofA} + c$: Proposed equation for system performance, and OML = $b \cdot \text{DofA} + c$ for overall mental load. The regression coefficient is $R^2$.

<table>
<thead>
<tr>
<th>Degree of Automation</th>
<th>System Performance</th>
<th>Overall Mental Load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>TES Based</td>
<td>-0.58</td>
<td>1.00</td>
</tr>
<tr>
<td>TDL Based</td>
<td>-0.62</td>
<td>1.09</td>
</tr>
<tr>
<td>TML Based</td>
<td>-0.62</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Figure 5.10 Relationships between system performance, overall mental load and DofA. The digit beside a data point indicates the number of manually controlled cells.
5.3.6 Relation between Overall Mental Load and DofA

The average OML of a session (Table 5.4) is also related to the three measures for the DofA. Table 5.5 presents the relationship between the OML and the DofA using a first-order polynomial ($R^2 > 0.94$). When a second order polynomial is used to fit these trends, the $R^2$ values increase only 0.1%. This demonstrates that the relationship between the OML and the DofA can be adequately described by a straight line. The relationship between the OML and the DofA$^{TDL}$ is plotted in Figure 5.10 as well. Similar trends have been found for the other two measures of task weights.

As the DofA increases, the perceived overall mental load decreases. It should be pointed out here that an increase in the DofA does not necessarily imply an increase in the number of automated tasks. This is because each task has a different weight. Automation of a large number of tasks with small task weights may result in a smaller DofA than when a small number of tasks with large task weights are automated.

5.3.7 Relation between Three Measures of DofA

Figure 5.11 shows task weights based on TES, TDL, and TML. As can be seen, $w^{TML}$ and $w^{TDL}$ are very close. Furthermore, $w^{TES}$ and $w^{TML}$ curl around $w^{TDL}$ and intersect at Cell 7. For Cell 1 through 6, $w^{TDL} \geq w^{TML} > w^{TES}$, but for Cell 8 through 12, the order is reversed, that is, $w^{TDL} \leq w^{TML} < w^{TES}$. Consequently, the same conclusion for DofA$^{TES}$, DofA$^{TDL}$ and DofA$^{TML}$ can be drawn. This phenomenon is displayed in Figure 5.12 where DofAs for the three task weight measures are plotted for the two groups of experimental configurations. Automating control tasks for Cell 7 to 12 leads to a larger increase in DofA than automating Cell 1 to 6. As can be seen in Table 5.4, automating Cell 7 to 12 leads to a larger improvement in the system performance than automating Cell 1 to 6. This is also shown in Figure 5.10. However, when the DofA exceeds a certain level, the improvement will become smaller.

![Figure 5.11 Relationship between different task weights.](image)
5.3.8 Prediction of Mental Load

As shown in Figure 5.9 and Table 5.4, the OML perceived by the operators during control of the entire system is related to the manually controlled tasks and their number. According to Liu and Wickens (1987), the workload of a complex task may be predicted from the workload of task components, assuming that the workload should be linearly related to the workload of components and that task interference between each other is accounted for. For the simple linear system in this study, this makes us to investigate whether the OML for our system is related to its components, namely, $TML_i$, for $i = 1, 2, ..., 12$. In other words, the OML for controlling the entire system could be predictable from $TML_i$ for controlling each individual cell.

The operators were asked to assess $TML_i$ after each experimental session. The $TML_i$ in the case that the system was fully manually controlled, $TML_i^{FM}$, has been listed in Table 5.3. The question is whether the OML for controlling a partly-automated (PA) system, $OML^{PA}$, can be predicted by the $TML_i^{FM}$.

We proposed the following formula to predict the $OML^{PA}$:

$$OML^{PA}_k = \frac{OML^{FM} \sum_{i=1}^{N} \{1 - t_i(k) \cdot TML_i^{FM}\}}{\sum_{i=1}^{N} TML_i^{FM}}$$  \hspace{1cm} (5.17)

where $k = 1, 2, ..., $ for the session number;

$N$ : Total number of cells;

$OML^{PA}$ : OML for controlling a partly-automated system;
$OML^{FM}$: OML for a session in which all cells were manually controlled;

$TML_{i}^{FM}$: Task mental load in case all cells were manually controlled;

$t_{i}(k)$: Task allocation indicator, $t_{i}(k) = 1$ when cell $i$ was automated, and $t_{i}(k) = 0$ if the cell was manually controlled.

The prediction was based on the $TML_{i}^{FM}$ ($i = 1, 2, \ldots, 12$) and the corresponding OML, $OML^{FM}$. Eq. (5.17) shows a linear combination of $TML_{i}^{FM}$. $OML^{PA}$ equals 0 in the case that all cells are automatically controlled and $OML^{PA}$ equals $OML^{FM}$ in the case that all cells are manually controlled.

Table 5.6 shows the predicted $OML^{PA}$ and the $OML^{PA}$ assessed by the operators. The correlation coefficient between two is 0.977. Zijlstra (1995) has discussed that when the OML is low, the RSME becomes less sensitive. In our experiment, when the OML is below 30, the prediction is less accurate which may be due to the low sensitivity of the RSME for low mental loads. Considering the sessions with an OML higher than 30, we observe that the predicted $OML^{PA}$ is not significantly different from the assessed $OML^{PA}$ ($t$-test: $p < 0.05$). Thus, we conclude that when the predicted OML is above 30, Eq. (5.17) may be used.

**Table 5.6** Comparison between assessed and predicted overall mental load

<table>
<thead>
<tr>
<th>Number of Manually Controlled Cells</th>
<th>Assessed $OML^{PA}$</th>
<th>Predicted $OML^{PA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.83</td>
<td>1.22</td>
</tr>
<tr>
<td>1</td>
<td>21.17</td>
<td>11.01</td>
</tr>
<tr>
<td>2</td>
<td>6.00</td>
<td>5.23</td>
</tr>
<tr>
<td>2</td>
<td>28.83</td>
<td>18.31</td>
</tr>
<tr>
<td>4</td>
<td>10.50</td>
<td>7.89</td>
</tr>
<tr>
<td>4</td>
<td>23.83</td>
<td>21.64</td>
</tr>
<tr>
<td>4</td>
<td>44.17</td>
<td>40.48</td>
</tr>
<tr>
<td>6</td>
<td>35.50</td>
<td>33.86</td>
</tr>
<tr>
<td>6</td>
<td>36.17</td>
<td>32.68</td>
</tr>
<tr>
<td>6</td>
<td>46.33</td>
<td>38.14</td>
</tr>
<tr>
<td>6</td>
<td>50.00</td>
<td>47.20</td>
</tr>
<tr>
<td>8</td>
<td>27.83</td>
<td>30.28</td>
</tr>
<tr>
<td>8</td>
<td>47.00</td>
<td>43.53</td>
</tr>
<tr>
<td>8</td>
<td>56.83</td>
<td>56.53</td>
</tr>
<tr>
<td>10</td>
<td>43.67</td>
<td>46.86</td>
</tr>
<tr>
<td>10</td>
<td>50.33</td>
<td>54.49</td>
</tr>
<tr>
<td>10</td>
<td>62.67</td>
<td>62.24</td>
</tr>
</tbody>
</table>
5.3.9 Task Demand Load, System Performance, and Human Perception

The relation between the system OTDL and system performance has been presented in Figure 5.8. When the OTDL is smaller than 20, the performance shows little variation. This implies that there is an OTDL region in which system performance is almost constant during normal operation. Within this range, human operators may not perceive a high workload. According to Fitts list (Fitts, 1951), automation has more advantages than the human operator in the case that a set-point control task, as used in this study, has to be executed. The operator realizes an optimal performance as long as the OTDL falls within a certain region. When the OTDL exceeds 20, the performance will decline. This indicates that when the OTDL passes a critical point, system performance can be improved by automation.

The relation between the OTDL and the OML is presented in Figure 5.13. As expected, the trend is that the OML increases with the OTDL ($R^2 = 0.96$). Moreover, Figure 5.7 and Figure 5.9 show that for a system configuration with a certain number of manually controlled cells, both the OTDL and the OML are directly proportional to the sum of the manually controlled cell index numbers. It demonstrates that human subjective perception is very sensitive to task complexity. This result is consistent with the experimental results reported by Bi and Salvendy (1994). However, there are other factors besides the task complexity which may affect human perception of workload, for example, the lay-out of the manually controlled cells in the human-machine interface (see in Figure 5.4). We investigated this factor as follows.

When TML was assessed for the fully manually controlled system (Table 5.3), the difference between two TMLs, one for a cell on the far left of a row on the display and one for a cell on the far right of the row above, seemed larger than the difference between two TMLs for cells in the middle of a row (for example: Cells 4 and 5, Cells 8 and 9 compared to Cells 2 and 3, 6 and 7, 10 and 11). This suggests the effect of the position of the cell interface on the display on mental load perception.

![Figure 5.13 Relation between average overall mental load and overall task demand load.](image-url)
However, we could not proof the significance of the differences between cell pairs across rows and within rows. Thus, we conclude that the effect of the position of a cell interface on the display is minor.

5.3.10 Relation between OML and System Performance

In previous sections, only the relations between system performance and task demand load, as well performance and DofA are presented. The relation between system performance and human perceived mental load is worth checking. This section discusses whether a hypothetical relationship (Figure 3.2) between system performance and the mental load is satisfied.

The hypothetical relation between performance and workload is known as an “inverted-U” (Sheridan, 1992) and is illustrated in Figure 3.2. For convenience, the “inverted-U” hypothesis is re-plotted in Figure 5.14 (Bi and Salvendy, 1994). As it can be seen from Figure 5.14, for extremely low levels of mental workload, in Region 1, the performance is poor. A main reason is that because the operator has only to do a little, the operator may become inattentive and bored. Boredom can lead to an operator’s monitoring performance decline, such as: Missing tasks and instructions. Another explanation may come from the unwillingness or laziness of the operator to invest enough mental load for achieving high levels of performance. The operator may continue to perform tasks at a low level of performance, using his improved knowledge just to reduce his mental load (Stassen et al., 1990). Within a reasonable level of mental load, acceptable performance may be achieved, as shown in Region 2. However, when mental load enters into Region 3, performance will greatly decline. As it can be seen from Region 2, performance may remain at an acceptable level over a considerable range of mental load variation. It means that task mental load is certainly not the same as task performance. Therefore, task performance is not suitable to be regarded as a legitimate measure of mental load (Sheridan, 1992; Yeh and Wickens, 1987).

![Figure 5.14 Hypothetical relationship between performance and mental load (see, for example, Bi and Salvendy, 1994; Sheridan, 1992).](image-url)
Figure 5.15 System performance vs. overall mental load.

By noticing that human performance may be measured by system performance (Sheridan, 1992), system performance in our experiments is used as the performance addressed in Figure 5.14. Based on the experimental data, the relation between mental load and system performance, as shown in Figure 5.15, illustrates the hypothetical relationship between system performance and OML. In the case that the OML is smaller than 35, the system performance shows very small variations. It suggests that the mental load falls into Region 2 of Figure 5.14. When $35 < \text{OML} < 50$, system performance shows variations, and it becomes poorer, but not too much. It indicates that the mental load has reached the critical section of Region 2. However, as soon as $\text{OML} \geq 50$, system performance declines drastically and becomes very poor. This means that the mental load has passed the critical point and has fallen into Region 3. For our experimental system, the upper critical range of OML for an acceptable performance may be between 45 to 55, where system performance is going down drastically.

Although a pretty good system performance has been achieved when OML is at extreme low level, it does not mean that the experimental results are conflicting with Region 1 in Figure 5.14. As mentioned before, for the set-point keeping task, automation can lead to better performance than that the human operator can achieve. Thus, the more cells which invoke strong mental load are automated, the better performance can be achieved. As shown in Table 5.4, when OML is at a very low level, system performance has small variations, approaching to, or even better than the performance when all cells were automated. However, almost all of the subjects complained that controlling a system with a mental load below 15 was very boring. Based on system performance and mental load in this experiment, we cannot observe the relation in Region 1. We conclude that this experimental system is not suitable to conduct a study for Region 1.
5.3.11 RSME and NASA-TLX

During the experiment, the RSME was used to assess mental load. Nevertheless, the NASA-TLX was applied for some of the sessions to verify the outcomes of the RSME. Veltman and Gaillard (1993) have tested the RSME and compared it with the NASA-TLX. They find that the RSME is better for Dutch subjects and is simpler to use in the experiment. Analysis of the data obtained by using these two scales shows that the assessment of mental load can be achieved equally well from the RSME as well as from the NASA-TLX in a consistent way. Figure 5.16 shows the normalized overall mental load assessed according to the RSME and the NASA-TLX, averaged over all subjects.

5.3.12 Further Research

The experimental system used in this study is a simple linear system. It has only forward connections. If a backward connection (i.e. feedback) is employed, the task complexity to control a cell will increase. This needs to be studied further.

In analyzing the task effect on system performance, only the effect of a single control task has been investigated. The effect of the control tasks that occurred simultaneously was not considered.

The computation of DofA presented here is related to process tuning. As more automation is introduced, the operators will have to deal with discrete tasks (see Ch. 1) as well. The question arises whether the definition of DofA can be used for the computation of DofA for systems that contain discrete tasks. Further research is needed to quantify DofA for industrial systems.
The effect of a change of DofA due to a failure in the system's automation on the mental load, system performance, and human performance is worth investigating. The following chapters will be devoted to these aspects.

5.4 CONCLUDING SUMMARY

In this chapter, a quantitative measure for the level of automation, DofA which is defined as a function of the number and the nature of control tasks to be executed by operators, has been applied to an artificial linear system. This is a preliminary step towards developing an index to quantify the level of automation in supervisory control. In this study, the relationship between different measures of the DofA is discussed based on the experimental results. This relationship is limited in the experimental system. The effects of the change of the DofA on the human operators are studied. From this study, the relation between performance, mental load and the DofA is drawn. The application of the DofA measurement in real world will depend on further studies using more complex systems with different task natures (discrete vs. continuous), task criticality, and task uncertainty. Moreover, the monitoring tasks created by automation can have strong influence on the operator.

Based on the experimental study, the following conclusions may be drawn:

- When the operator controls a partly-automated system, the overall mental load perceived by the operator may be predicted from the mental load obtained from a specific situation as long as the overall mental load is within a certain range. In the specific situation, all tasks are manually performed. The mental load obtained from this situation includes the overall mental load and task mental load for each task component.

- System performance may level off when the DofA is above a certain level. This suggests that the human operator can remain in the control loop without degrading the system performance.

- The difference between the DofA based on task demand load and the DofA based on task mental load is small. This suggests that task demand load can reflect the level of task mental load. Thus, within a range that the humans are willing to spend, the level of mental load may be estimated by task demand load which is obtained from engineering parameters.

Although the experimental task was simple and consisted of controlling set-points of identical subsystems with only forward connections, the research may have the following implications for the design of human-machine systems. In adaptive automation, or in dynamic task allocation, when human mental load exceeds a certain level, tasks should be reallocated between human and machine to reduce human mental load. However, it is difficult to measure the mental load on line. Task demand load can predict the mental load as shown in this study. So, the task
demand load can be used as a trigger in adaptive automation. Moreover, the operator's mental load in controlling a partly automated system could be predicted from the mental load obtained from a specific situation. Thus, the overall mental load to trigger a reallocation could be computed from the task mental load in this specific situation.

The situation awareness and operational skill of the operator can be corrupted by full automation (Endsley and Kiris, 1995). So, it is important to keep the operator in the control loop and to let him have a reasonable level of workload. As shown in this study, the human operator can be kept in the control loop, that is to say in this simplified situation, without degrading the system performance. In this way, the operator will not lose awareness of system situations and operational skills. If automation fails, the operator may more effectively take over the control of the automated task than in the case that the operator has a low awareness.
Chapter 6

Determination of Degree of Automation of a Pasteurization Plant

The quantitative measure, as described in Ch. 4, for the degree of automation in supervisory control is implemented in a simulated pasteurization plant. The consequence of the change of the degree of automation is experimentally investigated, and is compared with the experiment in the artificial linear system (Ch. 5). Similar conclusions to those from the linear system have been drawn. In addition, a mathematical tool - Analytic Hierarchy Process - is used to carry out the subjective evaluation of task allocation decisions. The results derived by this tool may be taken as a reference for the task allocation evaluation.

6.1 INTRODUCTION

The quantitative measure of the degree of automation in supervisory control, DofA, introduced in Ch. 4 has been applied to an artificial linear system as described in Ch. 5. The relationship between the tasks in this system is simple although they are coupled. In order to demonstrate the possibility of the implementation of the DofA measure in the real world, the DofA is applied on a simulated pasteurization plant which has been used by other researchers for different purposes (Hucy, 1989; Lee and Moray, 1992, 1994; Muir, 1989). Realistic thermodynamics and heat transfer relationships govern the dynamics of the pasteurization process. The plant dynamics incorporate many of the properties of an actual process control plant. The operator of the simulated plant can achieve his/her goals in a variety of ways. According to Lee and Moray (1992), this simulated plant has a reasonable level of complexity. This will impose cognitive and motivational demands on the subjects similar to those experienced in actual work situations. Thus, the results of the experiments from this plant may be more generalizable to actual work domains than those described in Ch. 5.

This chapter presents the results from a series of experiments. Similar to the experiment presented in Ch. 5, the task weights based on the TES, the TDL, and the TML are analyzed and assessed. Thus, DofA's for the cases with different task allocation configurations can be
computed. Consequently, the effects of a change of degree of automation on system performance and human perception of the task workload will be analyzed. Besides, the relationship between the demanded workload and the mental load, and the relationship between the system performance and the mental load will be analyzed and discussed from the experimental results. In addition, for different task allocation decisions, subjective evaluation results using Analytic Hierarchy Process (Saaty, 1980) based on a set of criteria are presented.

6.2 METHOD

6.2.1 Experimental Set-Up

A simulation of a juice pasteurization plant (developed originally by Muir and Moray, 1987; and modified by Huey, 1989; Lee and Moray, 1992) was used as the experimental task. Figure 6.1 depicts the modified pasteurization plant. The pasteurization process is to heat the raw juice to a level where bacteria in the juice will be killed. Although the pasteurization plant represents a somewhat simplified and abstracted version process of the control plant, it is not a replication of any real process control plant. It is a microworld, designed to capture some of the complexities of the actual work domain, but in a way that allows a greater degree of experimental work. It makes the seemingly simple system a challenging control problem (Lee and Moray, 1994).

The plant was run in iteration, rather than in real time. Each iteration had different time lengths. The shortest iteration time in our experiment was set to 35 ms. However, some iterations were 3 to 4 times longer than this time because of calling subroutines and I/O operations. For a fully automated plant, the average running rate was 21.4 iterations/s.

![Figure 6.1 Scheme of the pasteurization plant (after Paauw, 1996).](image-url)
As shown in Figure 6.1, the plant consists of two loops. One is the juice circulation loop, and the another one is the steam circulation loop. These two loops contain three subsystems which will be described in the following sections.

6.2.1.1 Flow control subsystem

The flow control subsystem consists of an input tank, a feedstock pump and an overflow valve. In the juice circulation loop, the untreated juice enters the input tank through the inflow pipe. The raw juice in the input tank is pumped by the feedstock pump through the secondary heater and the primary heater to a distribution system. The secondary heater is a passive heat exchanger between the cold juice and the hot, pasteurized juice. The temperature of the juice leaving the primary heater, $T_d$, must be maintained between 70 °C and 85 °C. The juice leaving the primary heater is distributed by a 3-way valve, according to $T_d$, to each of the following lines:

- If $T_d$ is below 70 °C, the juice is recycled to the input tank. If the level of the input tank is maximum, the juice is redirected to the waste tank.
- If $T_d$ is within the required range, the juice is led to the secondary exchanger and consequently to the distribution tanks.
- If $T_d$ is above 85 °C, meaning that the heated juice is considered to be burnt, the juice is directed to the waste tank.

In order to induce control demands for the operator, both the temperature and the flow of the input juice are perturbed every 200 iterations. The temperature of the juice entering the plant varies between 5 and 15 °C. The input flow varies between 0.205 and 0.285 m$^3$/iteration. The temperature and the flow are varied according to a uniform distribution.

The input tank, with a maximum value of 23 m$^3$, is subject to overflow. The overflow valve can be controlled both manually and automatically. In automatic control, the valve is opened when the input tank reaches its upper limitation. When the level of the input tank becomes too high during manual control, and the overflow valve is not opened by the operator, the input flow will be cut down from average 0.24 to 0.008 m$^3$/iteration until the input tank is half full. The juice surplus is directed to the waste tank.

The input tank may also become empty. This is prevented by reducing the feedstock pump speed automatically. If the juice volume in the input tank is less than 2.5 m$^3$, the pump speed is automatically reduced with its maximum rate of change until the speed is below 0.15 m$^3$/iteration. This results in the input tank being quickly filled to about 5 m$^3$. Due to the sharp drop of the pump speed and the unchanged steam temperature, the temperature, $T_d$, increases sharply so that nearly all pasteurized juice is burnt during this period and therefore the performance is degraded.
The feedstock pump is operated at a speed between 0 and 0.4 m³/iteration with a maximum rate of change of 0.01 m³/(iteration)². It can be operated automatically or manually. If it is automatically controlled, it follows the amount of juice flowing into the input tank and keeps the level of the input tank constant. When the level of the tank approaches to 17 m³ (the input tank will start to overflow) or to 6 m³ (the input tank becomes empty), the automatic controller adjusts the pump speed with reference half the input tank volume.

When the pump is manually controlled, the pump speed can be set by a control slide steered by using a mouse or by typing a value using the keyboard. Based on the input flow of the input tank, a reference value, called flow set-point, for the pump speed is estimated with an accuracy of ±0.05 m³/iteration and is displayed below the control slide. The flow set-point can be used to adjust the pump speed in accordance to a new input flow. The warning will appear when the flow rate deviates more than 0.02 m³/iteration from the flow set-point. Consequently, the steam temperature and flow in the steam loop should be adjusted for a new flow set-point value.

6.2.1.2 Steam control subsystem

The steam control subsystem consists of a steam heater in which the steam is generated by heating water with gas, and of a steam pump. In the steam circulation loop, the steam pump sends the steam to the primary heater where the heat exchange between the juice and the steam takes place. The temperature in the primary heater can be controlled by adjusting the temperature of the steam heater and the steam flow rate.

The temperature of the steam heater is controlled by changing the amount of the gas supply; it depends on the actual steam flow and the temperature of the steam returning from the primary heat exchanger. The gas supply can be set from 31% to 150% of the normal working condition which is considered to be 100%. The supply can be changed immediately, but the steam temperature cannot change instantaneously. In order to increase the operational difficulty, the gas supply is periodically disturbed. The gas supply is dropped every 105s to 30% of the normal condition and this lasts for 4.2s. This causes the pasteurization temperature to drop below 70 °C in whatever state the plant is, and consequently the juice is diverted to the input tank.

Although the flow rate of the steam is affected by the speed of the steam pump, the control of the pump speed does not show very much effect on the pasteurization temperature. A change in the steam flow rate not only affects the amount of the heat transferred in the primary heat-exchanger, but also causes a compensatory effect in the steam heater. Thus, a step change in the steam pump speed causes only a temporary change in the pasteurization. The pump is operated between 0.02 m³/iteration and 0.25 m³/iteration with a rate of change of 0.1 m³/(iteration)².

Both the steam heater and the steam pump can be controlled manually or automatically. In automatic control, PI controllers are used to keep $T_d$ at its optimum of 77.7 °C.
6.2.1.3 Production distribution subsystem

In this modified pasteurization plant, we added a new production distribution subsystem for the juice flow loop. This subsystem consists of three distribution tanks and a distribution switch. The operator is required to distribute the pasteurized juice among three tanks by pointing on the distribution switch to different tanks. It is assumed that these tanks may be emptied irregularly. If one of the tanks reaches its maximum capacity, the surplus product is diverted to the waste tank and the alarm, both visually and auditory, is activated. In order to maintain a high production, this should be avoided. Moreover, the distribution tanks should not become empty. Otherwise, the alarm will be activated. It is obvious that this subtask does not affect the pasteurization process, but it does influence the amount of juice in the waste tank. Hence, it further influences the rate of production. This subtask can be executed by the operator or by automation. The automatic controller simply switches to the tank with the least juice continuously and keeps the levels in all three tanks almost the same. This subsystem requires only discrete operation, and has 4 distinct states: Flow to tanks 1, 2, 3, or to the waste tank.

Figure 6.2 Human-Machine Interface of the pasteurization plant. T₁₉: inflow temperature; Frecycle: the flow rate of the recycle juice; Input vat: juice volume in input tank and its temperature are displayed; Overflow: overflow valve and the flow rate; FlowSP: the SP value for the rate of the flow pump; T₂: juice temperature after the secondary heater; T₃: juice temperature after pasteurization; Fwaste: the flow rate to the waste tank; Vwaste: the juice volume in the waste tank; Steam temperature and steam flow rate are also displayed.
It should be noted that both the feedstock pump and the steam pump respond in a non-linear way to a change in the pump rate (for details, Huey, 1989; Lee and McKay, 1994).

When the pasteurization plant starts operation, all plant disturbance variables and initial conditions are read in from a file to ensure that each experiment will have an equal difficulty, regardless of automation setting, test subjects, etc.

The plant was built on a Pentium PC with a 17 inch display using a graphic language, LabVIEW (National Instruments, 1993). Figure 6.2 shows the human-machine interface presented to the operator.

6.2.2 Experimental Task

The experimental task is to pasteurize as much of the available raw juice as possible, which means that the tasks to maintain the pasteurization temperature within the allowed range, and to distribute it to an available tank have to be realized. The task can be split into subtasks and task components. That is to say that the experimental task consists of a number of monitoring and control subtasks which can be allocated to the operator or to automatic controllers. The monitoring subtasks are considered to be a part of the control task since it is difficult to distinguish between these task components.

In order to accomplish the experimental task, the two important variables, \( T_d \) and the juice flow rate, are to be controlled. \( T_d \) should be maintained in allowed range, which depends mainly on the juice flow rate, the steam flow rate, and the temperature of the steam heater. The juice flow is maintained at a certain level in order to keep the input tank neither empty nor in overflow. In the present study, three subtasks are defined for the operation of the plant.

**Subtask 1:** Control of the pasteurization temperature.

This subtask deals with three control components: Feedstock pump control, steam heater control, and steam pump control. When the steam heater temperature and the steam pump rate are set constant, a change in the juice flow rate will affect \( T_d \). The operator can control all three control components to maintain \( T_d \). In the control of \( T_d \), the steam heater control can be taken as a rough tuning. However, a change in the steam flow does not cause a big change of \( T_d \), since it is counteracted by the steam temperature. Thus, the steam pump control is taken as a fine tuning.

In the experiment, any or all of the three control components can be allocated to the operator or to the automatic controllers.

**Subtask 2:** Control of the juice flow rate.

This subtask fulfills two goals: To follow the flow set-point request and to avoid the input tank from overflowing or from becoming empty. The operation includes two control components:
Feedstock pump control and overflow valve control. The speed of the feedstock pump affects the juice flow that consequently will affect \( T_d \), and the level of the input tank. When the pump speed increases, \( T_d \) decreases if the steam temperature and the steam flow rate are constant. As the pump speed decreases, the level of the input tank may increase, causing overflow, and \( T_d \) may exceed 85 °C. If the set-point value for the pump speed has been followed, the input tank will not overflow. The overflow valve control will only be operated when the input tank is going to overflow.

**Subtask 3: Distribution of product.**

This subtask includes a discrete control component — control of a distribution switch. Flow to three storage tanks should be distributed to avoid overflowing or becoming empty. The discrete control task can also be allocated either to the operator or to the automatic controller.

In summary, three subtasks are realized by five control components:

- The Overflow Valve Control, OVC;
- the Feedstock Pump Control, FPC;
- the Steam Pump Control, SPC;
- the Steam Heater Control, SHC, and
- the Distribution Switch Control, DSC.

In this experiment, the allocation of the task between the operator and automatic controllers was carried out on the basis of *three task groups*:

1. Flow control subsystem: FPC and OVC;
2. Steam control subsystem: SPC and SHC;
3. Production distribution subsystem: DSC.

From the discussions above, it can be seen that FPC is coupled with SPC and with SHC by thermodynamic and heat transfer process taking place in the primary heater. The plant provides both continuous and discrete tasks (see Ch. 1). FPC, SPC, and SHC are continuous tasks, whereas OVC and DSC are discrete tasks.

### 6.2.3 Operator’s Task

When the plant is in the manual control mode, the operator has to fulfill the following three objectives: (1) Maintain the temperature of the juice flowing out of the primary heater; (2) maintain the level of the input tank at a proper level; and (3) distribute the pasteurized juice and keep all storage tanks neither full nor empty. The ultimate goal is to pasteurize as much of the available raw juice as possible and thus to minimize the volume of the waste tank.

During the experiment, the operator was asked to fill out the RSME form for each subtask and
its components, as well as for the operation of the entire plant. After all of the experimental sessions were completed, the operator was asked to evaluate, based on a set of criteria, the task allocation configurations by using pare-wise comparison required by the Analytic Hierarchy Process (Saaty, 1980). In Section 6.2.8., we will discuss the Analytic Hierarchy Process as well as the evaluation criteria.

6.2.4 Subjects

Nine students (male and female with an average age of 24.0 years) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology participated voluntarily as operators in the experiments. They received a fixed fee for their participation. Before formal sessions started, the subjects performed 6 training sessions of 650 iterations or approximation 5 minutes each to learn to control the system and to practice filling out mental load assessment forms (Appendix 3). After completing the training sessions, the subjects performed a test session which was in fully manual mode and lasted 5 minutes. The performance of this session was evaluated. The evaluation was based on the percentage of the pasteurized juice relative to the total raw juice that entered the input tank during the session. If the score of a subject was below 90%, the subject was asked to perform more training sessions and some operational guidance was given until the required score was achieved. Only when a subject was qualified to work as an operator, he/she was allowed to start formal experimental sessions. Immediately following each formal session, the subject was asked to rate mental load for performing each subtask and its components. Each subject performed 7 sessions of 2000 iterations or approximation 15 minutes each.

### Table 6.1 Experimental sessions (A: An automated subtask; M: A manual subtask)

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Task Allocation between Human and Automatic Controller</th>
<th>Operational Descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FPC</td>
<td>OVC</td>
</tr>
<tr>
<td>E₀</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>E₁</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>E₂</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>E₃</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>E₄</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>E₅</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>E₆</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>E₇</td>
<td>M</td>
<td>M</td>
</tr>
</tbody>
</table>
6.2.5 Experimental Sessions

As said before, the simulated plant has three subtasks that consist in total of five control components. These task components can be allocated to the operator and to an automatic controller (Table 6.1). In total, 8 experimental sessions are designed, including the fully automated session (E₀).

6.2.6 Performance Measurement

During the experiment, the following plant variables were recorded:
- The amount of the input raw juice;
- the amount of the wasted juice;
- the level of the input tank;
- the pasteurization temperature, \( T_d \);
- the number of alarms due to a coming overflow or an empty in a distribution tank.

In this study, System PerFormance, SPF, and Operator PerFormance, OPF, are used to measure the performance. The definitions of the SPF and the OPF are as follows.

*System PerFormance* is defined as the percentage of pasteurized juice relative to the total raw juice that entered the plant during the experimental session. The juice could be wasted in three situations: (1) Overflow occurred in the input tank or the overflow valve was opened, (2) pasteurization temperature exceeded 85 \(^\circ\)C, and (3) one of the distribution tanks was overfilled.

*Operator PerFormance* reflects the performance of each subsystem in the plant control. It takes into account every part of the plant that can be manually controlled by the operator. Thus, except for the percentage of the pasteurization, three variables from each subsystem are firstly weighted and are next combined to describe a measure of the operator performance. The OPF for a session has been defined as (Paauw, 1996):

\[
OPF^k = SPF^k \frac{SD_{T_d}^0}{SD_{T_d}^k} \frac{SD_{IT}^0}{SD_{IT}^k} \frac{ND_{AM}^k}{ND_{AM}^0},
\]

(6.1)

where:
- \( k = 1, \ldots, 7 \), the experimental session number;
- \( SD_{T_d}^0 \) and \( SD_{T_d}^k \): the standard deviation of pasteurization temperature in Sessions E₀ and E_k;
- \( SD_{IT}^0 \) and \( SD_{IT}^k \): the standard deviation of input tank level in Sessions E₀ and E_k;
- \( ND_{AM}^0 \) and \( ND_{AM}^k \): the number of data points recorded without an alarm on any of the distribution tanks in Sessions E₀ and E_k.
Equation (6.1) shows that the operator performance is defined as a combination of the system performance, the input tank level variation, the pasteurization temperature variation, and the alarm frequency of distribution tanks. All items are normalized to the value for session \(k = 0\).

6.2.7 Estimation of Task Weight

6.2.7.1 Task effect on system performance
The simulation technique suggested in Ch. 4 is used to estimate the task effect on system performance. The system performance for a reference case, where all control subtasks are executed correctly and promptly, is compared to a case where one specific control subtask is not executed or partly executed (for details, see Ch. 5). For the pasteurization plant, if a subtask is not performed, the whole plant will have a very low production. To measure the task effect on system performance, TES, each continuous subtask is partly performed. A fixed value is set to FPC, SPC, and SHC, respectively. If the DSC was not operated, all products would be sent to the waste tank. The plant will have a very poor system performance. So, a discrete time series with a time interval of 14s is applied to perform DSC. The system performance from each simulation with a subtask partly controlled is measured and is compared to the reference system performance. Eq. (5.8) was used to calculate \(w^{TES}\).

6.2.7.2 Analysis of task demand load
As mentioned in previous chapters, the TDL may be estimated by task analytic methods, expert opinions, and system engineering parameters. In Ch. 5, the TDL could be estimated by using a concept of task complexity because the system was linear and the relation between tasks could analytically be analyzed. However, since the pump response is nonlinear (Lee and Moray, 1994), the pasteurization plant has non-linear dynamics. Hence, it is difficult to find a mathematical model to calculate task demand load for all subtasks. In this case, the task analysis approach combined with expert opinion in the calculation of the TDL may be easier than the approach used in Ch. 5. Hence, for the pasteurization plant, the Hierarchic Task Analysis, HTA, (Kirwan and Ainsworth, 1992) and expert opinions were employed in order to analyze the task demand load for each subtask. Of course, this approach could also be applied to the linear system presented in Ch. 5.

The hierarchic structure based on the HTA approach for the pasteurization plant is plotted in Figure 6.3. In the HTA of the plant, we suppose that the goal — *produce as much juice as possible* — requires that the operator invest the maximum effort of 1, i.e. the maximum TDL is set to 1. Based on the knowledge of the plant design and the experience in the operation of the plant, we estimate the relative amount of the TDL for each subtask and control of each component when it is performed to achieve the goal.
For the second level of the HTA structure, three operational subtasks are: Controlling the pasteurization temperature, controlling the juice flow rate, and distributing the products. According to the operational experience of the experimenters, these three subtasks are assigned a TDL value of 0.4, 0.4 and 0.2, respectively.

As discussed before, Subtask 1 includes three control components. The temperature in the steam heater affects $T_d$ strongly and the steam heater is therefore frequently operated. Considering the importance of the steam heater in controlling $T_d$, the frequency of the operation in the SHC, and the disturbance added to the steam heater, we assign a TDL value of 0.7 to the SHC. Since the SPC acts as a fine tuning and is less important in adjusting $T_d$ and is less frequently used, we assign a TDL value of 0.1 to the SPC. The feedstock pump has an obvious influence on $T_d$. However, since it is seldom used as an active means for controlling the temperature, we assign a TDL value of 0.2 to the FPC within Subtask 1.

Subtask 2 includes two task components: follow flow set-point and avoid high/low input tank levels. In order to avoid the extremes in the input tank level, the level control is considered to be more important than following the flow set-point because the former can degrade the production. However, follow the flow set-point requires almost continuous adjustment and needs a lot of attention of the operator, and therefore, it is assigned a TDL value of 0.6. Since avoiding extremes in the input tank level only requires actions occasionally and is easier to be operated, it is assigned a TDL value of 0.4.

Two operations, or subtask components, may be performed to avoid the extreme levels in the input tank: FPC and OVC. Since the FPC can indirectly affect $T_d$ and the input tank level, it is far more frequently used, and requires much more attention than the OVC does. Thus, we assign a TDL value of 0.9 to the FPC and 0.1 to the OVC.
As mentioned before, the task allocation in this experiment is performed among three control task groups. Thus, the TDL in the case that the operator operates one or more control task groups, can be computed by multiplying the assigned values vertically through the hierarchical tree. For the same level, the TDL is added up. If, for example, we compute the TDL for the control of the steam control subsystem where both the steam pump and the steam heater are manually controlled, it follows:

\[ TDL_{\text{steam subsystem}} = 0.4 \cdot (0.1 + 0.7) = 0.32. \]  
(6.2)

The TDL when the overflow valve and the feedstock pump are manually controlled is:

\[ TDL_{\text{flow subsystem}} = 0.4 \cdot 0.2 + 0.4 = 0.48. \]  
(6.3)

*Overall Task Demand Load, OTDL, is determined in similar fashion. For instance, if only the flow control subsystem is controlled by the operator, the OTDL is calculated by Eq. (6.3). When all subtasks are manually controlled, as assumed above, the OTDL is 1. The task weight based on the TDL, \( w^{\text{TDL}} \), can be calculated as follows:

\[ w_{i}^{\text{TDL}} = \frac{TDL_{i}}{OTDL}, \]  
(6.4)

where \( i \) indicates a subtask, \( i = 1, 2, \) or, 3. Obviously, \( \Sigma w_{i}^{\text{TDL}} (i: 1, 2, 3) \) will be 1.

**6.2.7.3 Task weight based on task mental load**

In this experiment, the subjective rating scale, Rating Scale Mental Effort, RSME, (Zijlstra, 1993), was used to assess the mental load experienced by human subjects. Immediately following each session, subjects were asked to give numbers according to the rating scale that corresponded best to the experienced mental load for controlling each manual subtask and for controlling the entire plant. Again, the Overall Mental Load, OML, is introduced to indicate the total mental load experienced by the operator in operating the entire plant.

The \( w^{\text{TML}} \) is calculated by the TML that is experienced by the operator while controlling a task. We use \( TML_{i} (i = 1, 2, 3) \) to stand for the TML of a task group obtained from the fully manually controlled plant. When \( TML_{i} \) is used to calculate the \( w^{\text{TML}} \) for a task group, Eq. (5.14) can be employed. Furthermore, to measure the OML experienced by the operator when he/she controls a fully manually (FM) controlled plant, the parameter OML\(^{\text{FM}}\) is used.

**6.2.8 Subjective Evaluation of Task Allocation based on Analytic Hierarchy Process**

As described in Ch. 3, the evaluation of the decisions on task allocation between human and machine is an important step in system design. It involves many aspects and may employ
different techniques. In this study, we performed a subjective quantitative evaluation of task allocation on the basis of a given set of criteria. In order to collect the data given by the subjects, we employed a mathematical tool, so-called the Analytic Hierarchy Process, AHP, (Saaty, 1980). The following sections will briefly describe the AHP approach combined with the experimental conditions, and will address the criteria used in the evaluation.

6.2.8.1 Analytic Hierarchy Process

The AHP was originally developed by Saaty (1980) as a decision making aid, and is a theory of measurement for dealing with quantifiable criteria. It has found wide usage in a broad range of areas, including market prediction, project evaluation, and medical decision making etc. (Vargas, 1990). It has been used as a tool for subjective workload assessment (Vidulich and Tsang, 1987). It has been demonstrated that the AHP provides a means to evaluate multiple design options using weighted ranking scales, and that it can be used as an tool for obtaining subjective preferences in Human Factors studies (Mitta, 1993; Yang and Hansman, 1995). For task allocation between human and machine, Papantoniopoulos (1990) has developed an AHP cognitive task allocation model. In comparing the AHP to the traditional psychophysical methods for generating measurement scales, the AHP has the following advantages: (1) It possesses the ability to quantify consistency in human judgments; (2) it has the ability to provide useful empirical results in the event of a small sample of subjects and when the likelihood of obtaining meaningful statistical results may be restricted; and (3) it requires no statistical assumptions regarding the distribution of human judgments (Mitta, 1993).

In brief, the AHP is structured to encompass the basic elements of a decision. As shown in Figure 6.4, the objective of a problem is placed at the highest level, then moving to the subobjectives affecting the objective followed by criteria in the next level and so on, from more general to the more particular.

Figure 6.4 A hierarchic structure.
For each level of the hierarchy, the AHP breaks up multiple options into a series of paired comparisons which are then recombined to produce an overall weighted ranking. The subjects are required to compare all possible combinations. The results of comparisons are placed in a pair-wise comparison matrix which is called judgment matrix (Figure 6.5). The rows and columns of the judgment matrix are headed by the options included in the comparison. The principal eigenvector for the matrix is calculated, which gives a weighted ranking scale for each option.

For each level of Figure 6.4, data in the AHP are collected using a series of comparisons between each pair of design options at this level. Saaty (1983) and Mitta (1993) suggest a format which presents the two options to be compared on opposite ends of a 17 slot rating scale (Figure 6.5). The scale is a measure of dominance of one alternative over the other.

Figure 6.5 summarizes the AHP approach for the subjective evaluation conducted in this study. For a detailed description of this approach and the definition of Consistency Index, C.I., the reader is referred to Appendix 4. It should be stressed that if C.I. is smaller than 0.1 the judgments in pair-wise comparisons are considered consistent (Mitta, 1993).

**Figure 6.5** Scheme of the AHP evaluation process. \( n \): the number of options to be evaluated; \( s \): the number of subjects.
Table 6.2 Subjective evaluation attributes

<table>
<thead>
<tr>
<th>Number</th>
<th>Attributes</th>
<th>Descriptions</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>Easy operation</td>
<td>The overall system is easy to control</td>
<td>Which system is easy to control?</td>
</tr>
<tr>
<td>A₂</td>
<td>Human control confidence</td>
<td>Human is actually in the control of the system, rather than automation</td>
<td>In which system do you feel that you are in control instead of automation?</td>
</tr>
<tr>
<td>A₃</td>
<td>Comfortable operation</td>
<td>Operators feel uncomfortable because too much automation has been used</td>
<td>Which system do you feel comfortable to control?</td>
</tr>
<tr>
<td>A₄</td>
<td>Situation awareness</td>
<td>Whether the operator is aware of basic system state variables and understands them</td>
<td>For which system do you know the most about the system status (all necessary variables)?</td>
</tr>
</tbody>
</table>

6.2.8.2 Criteria for subjective evaluation of task allocation

The attributes, or criteria, used in the evaluation of task allocation of the current study are listed in Table 6.2. Attribute 1, A₁, is chosen as an indicator for assessing the workload level. We expect that the results will be comparable with the mental load rating based on the RSME. The operator often complains, in a highly automated system, that he is not in control of the system, but automatic systems are in the control of the system. He/she loses the confidence in the control of the plant. A₂ is used to obtain the subjective opinion on this issue. Too much automation may make the human operator uncomfortable in operating a system. It is our interest, by using A₃, to obtain a quantitative estimation to what extent a high level of automation will affect the operator’s feeling in the operation. Situation awareness becomes an important issue in the design of a human-machine system (Endsley, 1995a, b). In this experiment, by using A₄, we try to perform a subjective evaluation on the situation awareness associated with different levels of automation.

6.2.8.3 Evaluation procedure

The hierarchic structure for the evaluation in this study is similar to Figure 6.4, where \( n = 7 \) is the number of the experimental sessions. Because two of the nine subjects did not have time to complete the evaluation, the number of subjects, \( s \), is 7. Each subject was asked to carry out pair-wise comparisons after all of the sessions were performed. The comparison was done by using pencil and paper. A sample of pair-wise comparison tables is presented in Appendix 5.
6.3 RESULTS AND DISCUSSION

6.3.1 Task Weights Based on Task Effect on System Performance

A task weight based on TES, $w^{\text{TES}}$, can be calculated using Eq. (5.8). To obtain the task effect on system performance each subtask is controlled manually by setting the control input at a median value. The operator performance as defined in Eq. (6.1) is used to calculate the task effect on system performance. Table 6.3 shows the reference performance, TES performance, and the task weights based on the TES.

6.3.2 Task Weights Based on Task Demand Load

Table 6.4 presents TDL values of all subtasks and the associated weights based on Eq. (6.4). Steam control and flow control can be divided into different control components. The task weights for these components are presented in Table 6.5 for use in the following chapters.

Table 6.3 Task weight from TES, $w^{\text{TES}}$, (single task effect on performance) based on Eq. (5.8)

<table>
<thead>
<tr>
<th>Subtask</th>
<th>TES Performance (OPF)</th>
<th>Reference Performance (OPF)</th>
<th>Task Weight ($w^{\text{TES}}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>steam control</td>
<td>0.38</td>
<td>0.82</td>
<td>0.34</td>
</tr>
<tr>
<td>flow control</td>
<td>0.37</td>
<td>0.82</td>
<td>0.35</td>
</tr>
<tr>
<td>distribution</td>
<td>0.43</td>
<td>0.82</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Table 6.4 TDL and $w^{\text{TDL}}$, Eq. (6.4); TML in fully manual control and $w^{\text{TML}}$, Eq. (5.14)

<table>
<thead>
<tr>
<th>Subtask</th>
<th>TDL</th>
<th>$w^{\text{TDL}}$</th>
<th>$T_M^{\text{FM}}$</th>
<th>$w^{\text{TML}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>steam control</td>
<td>0.32</td>
<td>0.32</td>
<td>77.5</td>
<td>23.9</td>
</tr>
<tr>
<td>flow control</td>
<td>0.48</td>
<td>0.48</td>
<td>86.6</td>
<td>15.4</td>
</tr>
<tr>
<td>distribution</td>
<td>0.20</td>
<td>0.20</td>
<td>50.4</td>
<td>22.6</td>
</tr>
</tbody>
</table>

Table 6.5 TDL and $w^{\text{TDL}}$, TML in fully manual control and $w^{\text{TML}}$ for task components

<table>
<thead>
<tr>
<th>Task Component</th>
<th>TDL</th>
<th>$w^{\text{TDL}}$</th>
<th>$T_M^{\text{FM}}$</th>
<th>$w^{\text{TML}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>SPC</td>
<td>0.04</td>
<td>0.04</td>
<td>31.4</td>
<td>25.6</td>
</tr>
<tr>
<td>SHC</td>
<td>0.28</td>
<td>0.28</td>
<td>46.1</td>
<td>22.3</td>
</tr>
<tr>
<td>OVC</td>
<td>0.016</td>
<td>0.02</td>
<td>62.3</td>
<td>17.8</td>
</tr>
<tr>
<td>FPC</td>
<td>0.46</td>
<td>0.46</td>
<td>24.3</td>
<td>13.1</td>
</tr>
</tbody>
</table>
Figure 6.6 The relation between performance and OTDL. Larger values represent better performance.

The relationships between system performance, operator performance and the OTDL are plotted in Figure 6.6. The performance is averaged over all subjects. From Figure 6.6, one can observe that there is no significant difference among the system performance (t-test: p > 0.05) of each experimental session even though the OTDLs are different. The SPF cannot reflect the effect of OTDL on the performance of the operators. However, the OPF was different across all sessions. The relation indicates that when the OTDL increased the OPF decreased.

6.3.3 Mental Load and Task Weights

Average OMLs across all subjects for all experimental sessions are plotted in Figure 6.7. When all subtasks were manually controlled, each subject gave an RSME rating for each subtask, and for the entire system (OMLFM). For each subtask, or task components, the average TML and its standard deviation over all subjects are listed in Table 6.4 and Table 6.5. The \( w_{\text{TML}} \) calculated from Eq. (5.14) is shown in Table 6.4 and Table 6.5 as well.

Figure 6.7 OML deviation for each session.
6.3.4 Computation of DofA

The DofA was computed for different system configurations using the definition in Eq. (4.1), and for different task weight measures presented in the preceding section. The DofAs are presented in Table 6.6.

From Table 6.6, one can find that the numerical value for the DofA obtained on the basis of the measure $w^{\text{TDL}}$ ($\text{DofA}^{\text{TDL}}$) and the value for the DofA obtained on the basis of the measure $w^{\text{TML}}$ ($\text{DofA}^{\text{TML}}$) are very close. This suggests that mental load may be predicted by task demand load that is obtained from the expert opinions in the HTA structure. The $\text{DofA}^{\text{TES}}$ is again less related to the $\text{DofA}^{\text{TDL}}$ and $\text{DofA}^{\text{TML}}$. Similar findings have been obtained in Ch. 5, and the similar explanation can be found in Section 5.3.4. The main reason is that the $\text{DofA}^{\text{TES}}$ does not consider the dependency among tasks because the task weights are derived from effects of single subtasks on system performance.

6.3.5 Relation between Performance and DofA

Relationships between the performance and the three measures for DofA are shown in Table 6.7. Here a second order polynomial is used to fit the data. Figure 6.8 shows the relationship between the performance and the $\text{DofA}^{\text{TDL}}$. Similar trend lines can be found for the $\text{DofA}^{\text{TES}}$ and the $\text{DofA}^{\text{TML}}$. One may conclude from these results, as in Ch. 5, that when the DofA increases above a certain level, for example, DofA $\geq 0.7$ (Figure 6.8), the improvement in performance becomes small. This again proves that an operator can achieve a system performance as good as a completely automated system. Advantages for keeping the operator in control have been discussed in Ch. 5.

<table>
<thead>
<tr>
<th>Session</th>
<th>Manual Subtask</th>
<th>$\text{DofA}^{\text{TES}}$</th>
<th>$\text{DofA}^{\text{TDL}}$</th>
<th>$\text{DofA}^{\text{TML}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_0$</td>
<td>none</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$E_1$</td>
<td>steam control subsystem</td>
<td>0.66</td>
<td>0.68</td>
<td>0.64</td>
</tr>
<tr>
<td>$E_2$</td>
<td>flow control subsystem</td>
<td>0.65</td>
<td>0.52</td>
<td>0.60</td>
</tr>
<tr>
<td>$E_3$</td>
<td>distribution subsystem</td>
<td>0.69</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>$E_4$</td>
<td>steam + distribution control</td>
<td>0.35</td>
<td>0.48</td>
<td>0.40</td>
</tr>
<tr>
<td>$E_5$</td>
<td>flow + distribution control</td>
<td>0.34</td>
<td>0.32</td>
<td>0.36</td>
</tr>
<tr>
<td>$E_6$</td>
<td>flow + steam control</td>
<td>0.31</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>$E_7$</td>
<td>all subsystems</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 6.6 Calculated DofA based on TES, TDL and TML, Eq. (4.1)
Table 6.7 Relation between operator performance, overall mental load and DofAs. OPF = a(DofA)^2 + b*DofA + c: Proposed equation for the performance, and y = b*DofA + c for overall mental load. R^2 is the regression coefficient.

<table>
<thead>
<tr>
<th>Degree of Automation</th>
<th>Operator Performance</th>
<th>Overall Mental Load</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>TES Based</td>
<td>-0.73</td>
<td>1.43</td>
</tr>
<tr>
<td>TDL Based</td>
<td>-0.87</td>
<td>1.63</td>
</tr>
<tr>
<td>TML Based</td>
<td>-0.82</td>
<td>1.54</td>
</tr>
</tbody>
</table>

Figure 6.8 Relations between performance, overall mental load and DofA.

6.3.6 Relation between Overall Mental Load and DofA

The average OML of a session (Figure 6.7) is also mathematically related to the three measures for the DofA. Table 6.7 presents the relationship between the OML and the DofA using a first-order polynomial. A fitting of the data to DofA_{TDL} and DofA_{TML} with R^2 > 0.95 further demonstrates that the relationship between the OML and the DofA can be adequately described by a straight line. The relationship between the OML and the DofA_{TDL} is plotted in Figure 6.8 as well. Similar trends have been found for DofA_{TES} and DofA_{TML}.

The regression coefficients for the chosen relationship between OML and DofA are higher than that for the relationship between performance and DofA. The explanation for this result may be: (1) The relation between the performance and the DofA may have an order higher than the second order although a third order polynomial did not improve the regression very much; and (2) the measurement of the OPF includes discrete numbers/counts, such as the number of alarms. This indicates that the OPF measure may be incomplete.

The relation between OML and DofA can be described by a linear relation. This is in agreement with the relation obtained in Ch. 5. Although the data fit is not as good as for
OML, the relation between OPF and DofA reflects, at least, the trend that as the DofA increases, the OPF will increase to a certain level. This is consistent with the conclusion obtained in Ch. 5.

6.3.7 Prediction of Mental Load

As discussed in Ch. 5, the OML perceived by the operators during the control of the entire system is related to how many and which control tasks are manually performed. This is also true for this experiment as shown in Figure 6.7. In this study, the operators were also asked to assess $TML_i^{FM}$ from controlling each subtask in a fully manually controlled plant (as listed in Table 6.5). Based on these measurements and the $OML^{FM}$ as shown in Figure 6.7, Eq. (5.17) is employed to predict the OML for controlling a partly-automated plant ($OML^{PA}$). Table 6.8 presents the predicted $OML^{PA}$ and the assessed $OML^{PA}$.

The $t$-test shows the predicted $OML^{PA}$ is not significantly different from the assessed $OML^{PA}$ ($p < 0.05$). The correlation coefficient between two column is 0.99. However, the relative error between them is larger than that in Ch. 5. This is because there exists non-linear relation between subtasks in the pasteurization plant. Moreover, one can observe that when the DSC was manually performed the predicted error became larger. The error was measured when the DSC was the only subtask to be controlled manually. According to Pauw (1996), most subjects complained about the tediousness of controlling only the distribution subsystem. They also complained that this subtask was demanding when it had to be manually performed together with other subtasks. This may have caused the lower rating for the mental load for the DSC when it was manually performed alone and the higher rating when it was performed together with other subtasks. In Ch. 5 we have noted that when the mental load is low (< 30), RSME becomes less sensitive and that the dependency between task components has an effect on the perception. In this study, this is also the case when the distribution task was manually performed, and when it was combined with other subtasks.

Table 6.8 Comparison between predicted and assessed overall mental load for a partly-automated pasteurization plant

<table>
<thead>
<tr>
<th>Session</th>
<th>Manual Subtask</th>
<th>Assessed OML$^{PA}$</th>
<th>Predicted OML$^{PA}$</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>steam control subsystem</td>
<td>35.9</td>
<td>19.1</td>
<td>33.4</td>
</tr>
<tr>
<td>E2</td>
<td>flow control subsystem</td>
<td>44.0</td>
<td>18.9</td>
<td>37.3</td>
</tr>
<tr>
<td>E3</td>
<td>distribution subsystem</td>
<td>15.3</td>
<td>7.6</td>
<td>21.7</td>
</tr>
<tr>
<td>E4</td>
<td>steam + distribution control</td>
<td>66.3</td>
<td>11.3</td>
<td>55.1</td>
</tr>
<tr>
<td>E5</td>
<td>flow + distribution control</td>
<td>70.3</td>
<td>9.3</td>
<td>59.0</td>
</tr>
<tr>
<td>E6</td>
<td>flow + steam control</td>
<td>81.0</td>
<td>16.1</td>
<td>70.6</td>
</tr>
</tbody>
</table>
6.3.8 Task Demand Load, Performance, and Mental Load

The relation between the performance and the system OTDL has been presented in Figure 6.6. It is concluded from this figure that there is an OTDL region where the performance is almost constant during normal operation. Within this range, human operators may not have perceived a high workload. When the OTDL is larger than a certain level, say 0.5, the performance declines. This indicates again that when the OTDL exceeds a critical point, the performance may be greatly affected by automation.

The relation between the OML and the OTDL is presented in Figure 6.9. As expected, it demonstrates that operator perception of mental load is consistent with our analysis for the TDL estimation which was made before the experimental results were obtained.

Figure 6.10 shows the relationship between performance and overall mental load. Since Figure 6.9 shows a straight line, Figure 6.10 is almost the same as Figure 6.6. The system performance does not show much variation when the OML increases. However, when the mental load increases, the operator performance presents a decreasing trend although this is not consistent. A closer observation reveals that when the OML increases from 15 to 60 the OPF does not change very much except for Session E2. However, when the OML exceeds 70, the OPF decreases greatly. This indicates that the mental load has exceeded the critical point for maintaining an acceptable operator performance. This observation tells us that the relation between the OPF and the OML is in agreement with the hypothetical relation (Figure 5.14) in Region 2 and in Region 3. Hence, similar comments as in Section 5.4.7 can be made for this relation. Note that the present plant is not suitable to conduct a study for low mental load situations.

Figure 6.9 The relation between the OML and the OTDL. $E_i$ ($i = 1, 2, \ldots, 7$) is the experimental session index.
Figure 6.10 The relation between the performance (SPF and OPF) and the OML. \( E_i \) (i = 1, 2, ..., 7) is the experimental session index.

### 6.3.9 AHP Subjective Evaluation of Task Allocation Decisions

#### 6.3.9.1 AHP judgments and ranking

Following the steps in Figure 6.5 and considering the ranking of subjects, we can obtain the vector that specifies the easy operation ranking for task allocation configurations as shown in Table 6.1:

\[ r_A = \begin{bmatrix} 0.24 & 0.18 & 0.43 & 0.06 & 0.06 & 0.03 & 0.02 \end{bmatrix}^T, \]

where the first element of the vector represents the easy operation level of experimental session \( E_1 \), the second element for \( E_2 \), and successively, the last element represents the level of \( E_7 \). From this vector, one can find that the rank of the easy operation level is: \( E_3, E_1, E_2, E_4, E_5, E_6, \) and \( E_7 \). Session \( E_3 \) is ranked the easiest task allocation configuration and Session \( E_7 \) is ranked the most difficult configuration.

The vector specifying the ranking of human control confidence for task allocation configurations is:

\[ r_X = \begin{bmatrix} 0.06 & 0.08 & 0.11 & 0.10 & 0.12 & 0.25 & 0.29 \end{bmatrix}^T. \]

The order of the human control confidence is: \( E_7, E_6, E_5, E_3, E_4, E_2, \) and \( E_1 \). Session \( E_7 \) is ranked the highest and Session \( E_1 \) is ranked the lowest.

For comfortable operation, the vector specifying the ranking of the effect of automation on human behavior by combining all subject's judgments is:
\[ \mathbf{r}_{A} = \begin{bmatrix} 0.26 & 0.19 & 0.16 & 0.13 & 0.09 & 0.09 & 0.07 \end{bmatrix}^T.\]

From this vector, one can find that the rank of the comfortable operation is: E_1, E_2, E_3, E_4, E_5, E_6, and E_7. Session E_1 is ranked the highest and Session E_7 is ranked the lowest.

The vector specifying the ranking of the situation awareness from a combination of all subject's judgments is:

\[ \mathbf{r}_{A_s} = \begin{bmatrix} 0.09 & 0.08 & 0.02 & 0.09 & 0.11 & 0.29 & 0.33 \end{bmatrix}^T.\]

The rank of situation awareness is: E_7, E_6, E_5, E_1, E_4, E_2, and E_3. Session E_7 is ranked the highest and Session E_4 is ranked the lowest.

### 6.3.9.2 Consistency of the judgments

One of the advantages of the AHP is its capability to calculate consistency in human judgments. The consistency indices for six subjects (one subject who always achieved very poor consistency is not included) in the evaluation are calculated by Eq. (6.6), and are listed in Table 6.9.

From the consistency indices, we can conclude that the most consistent comparison takes place in the evaluation of the situation awareness since the average C.I. is approaching 0.1. In other evaluations, the average C.I. is around 0.2 which indicates that the consistency in the paired comparisons is lower. The possible reasons are: (1) The comparison capabilities of some subjects are poor; (2) the questionnaire are not completely understood by the subjects; and (3) there are too many options, i.e. too many experimental sessions to be compared. After all of the sessions completed, the subjects could not accurately remember what they had experienced.

### Table 6.9 Consistency Indices

<table>
<thead>
<tr>
<th>Number</th>
<th>Attributes</th>
<th>Consistency Index (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Subj. 1</td>
</tr>
<tr>
<td>A_1</td>
<td>Easy operation</td>
<td>0.32</td>
</tr>
<tr>
<td>A_2</td>
<td>Human control confidence</td>
<td>0.11</td>
</tr>
<tr>
<td>A_3</td>
<td>Comfortable operation</td>
<td>0.12</td>
</tr>
<tr>
<td>A_4</td>
<td>Situation awareness</td>
<td>0.17</td>
</tr>
</tbody>
</table>
6.3.9.3 AHP evaluation and mental load

Easy operation and mental load. The relationship of the easy level in the operation (AHP) and the measured mental load (RSME) appears to be linear as shown in Figure 6.11. The correlation coefficient between two variables is -0.963, \( p < 0.001 \).

Figure 6.11 indicates that the easiest controlled system has the lowest mental load. As the easy operation level increases, the mental load decreases. Since the OMLs for the experimental sessions \( E_3 \) and \( E_4 \) are very close, the easy operation levels for these two sessions are also very close. The results presented in Figure 6.11 demonstrates that the AHP approach can be used as a method for mental load assessment as suggested by Vidulich and Tsang (1987).

Given Vector \( r_{A1} \), one can find that experimental session \( E_3 \), where only the distribution is manually controlled, is ranked as the easiest configuration to operate, and Session \( E_7 \), where all subtasks are manually controlled, is ranked the most difficult to control by a factor of at least 1.8 with respect to Session \( E_6 \).

Human control confidence and comfortable operation. These two criteria aim at evaluating to what extent the level of automation affects human operator’s feeling during operation. Vector \( r_{A2} \) indicates that Session \( E_7 \) is ranked in the situation where a maximum manual control is required, and Session \( E_1 \), where the steam control is controlled manually, is ranked the least to control by the operators by a factor of at least 1.4 with respect to Session \( E_2 \).

Vector \( r_{A3} \) indicates that Session \( E_1 \) is ranked as the most comfortable configuration to control because of the introduction of an automatic control system, and Session \( E_7 \) is ranked the least comfortable to control by a factor of at least 1.2 with respect to Session \( E_5 \).

![Figure 6.11](image-url) Relationship between overall mental load (RSME) and easy operation level (AHP).
Comparing Vector $\mathbf{r}_{A_2}$ with Vector $\mathbf{r}_{A_3}$, one can find that the ranking of the human control confidence perceived by the subjects is opposite to the ranking of the operational comfort. Session $E_1$ is ranked as the most comfortable configuration, but is the least in the human control confidence. Session $E_7$ is ranked that the human has the most control confidence, but it is ranked the least comfortable configuration. This implies that a compromise has to be made between these two aspects. An interesting point is that Session $E_3$ is the easiest to control, but it is not the most comfortable configuration. This means that a configuration with a low workload does not necessarily make the operator feel comfortable in the operation.

Situation awareness. Allocation of tasks between human and machine will directly affect operator's awareness of the system state. For the subjective evaluation of the situation awareness in this study, Vector $\mathbf{r}_{A_4}$ indicates that Session $E_7$ has the highest situation awareness, and $E_3$ has the lowest situation awareness by a factor of at least 3.4 with respect to Session $E_2$. This is understandable because in Session $E_7$ the operators perform all subtasks and monitor all necessary system variables.

### 6.3.9.4 Synthesis of the subjective evaluation

All criteria, or attributes ($A_1,..., A_4$), used in the evaluation can also be ranked based on the AHP approach so that a final ranking of all task allocation decisions may be obtained. For the four attributes used in the study, the judgment matrix, given by the experimenter, may be as follows:

$$
M_C = \begin{bmatrix}
A_1 & A_2 & A_3 & A_4 \\
A_1 & 1 & 2 & 3 & 1 \\
A_2 & 1/2 & 1 & 2 & 1/2 \\
A_3 & 1/3 & 1/2 & 1 & 1/3 \\
A_4 & 1 & 2 & 3 & 1
\end{bmatrix}
$$

The elements in $M_C$ indicate to what extent an attribute is more (or less) important than another (see Figure 6.5).

The principal eigenvector of $M_C$ is:

$$
\mathbf{r}_c = [0.35 \ 0.19 \ 0.11 \ 0.35]^T.
$$

The consistency index, C.I., of the comparison is 0.0035 ($< 0.1$) which indicates that the paired comparison is consistent.

The rankings of task allocation configurations in the four attributes can construct a new matrix with the following form:

$$
\mathbf{R}_{TA} = \begin{bmatrix}
\mathbf{r}_{A_1} & \mathbf{r}_{A_2} & \mathbf{r}_{A_3} & \mathbf{r}_{A_4}
\end{bmatrix}
$$
Table 6.10 Final rank of task allocation configurations

<table>
<thead>
<tr>
<th>Rank</th>
<th>Session</th>
<th>Manual Subtask</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>E₃</td>
<td>distribution subsystem</td>
</tr>
<tr>
<td>6</td>
<td>E₇</td>
<td>all subsystems</td>
</tr>
<tr>
<td>5</td>
<td>E₆</td>
<td>flow + steam control</td>
</tr>
<tr>
<td>4</td>
<td>E₁</td>
<td>steam control subsystem</td>
</tr>
<tr>
<td>3</td>
<td>E₂</td>
<td>flow control subsystem</td>
</tr>
<tr>
<td>2</td>
<td>E₅</td>
<td>flow + distribution control</td>
</tr>
<tr>
<td>1</td>
<td>E₄</td>
<td>steam + distribution control</td>
</tr>
</tbody>
</table>

Thus, the final rank specifying all task allocation decisions (Sessions E₁, ..., E₇) by combining four criteria is:

\[
r_F = R_{TA} \cdot R_C = [0.15 \ 0.13 \ 0.19 \ 0.08 \ 0.11 \ 0.17 \ 0.19]^T
\]

According to Vector \(r_F\), the rank of all the task allocation configurations is listed in Table 6.10.

From Table 6.10, one can find that Session E₅ and Session E₇ have a high ranking. However, we must take into account all aspects in the evaluation before making a decision about task allocation. For example, Session E₃ is given a high score at A₁. This high score has made E₃ obtain a high score in the final ranking. However, for A₄, E₃ gets the lowest ranking. Session E₇ is the most difficult configuration to be operated, i.e. it demands a high mental load, although it gets the highest ranking at A₂ and A₄. Both Session E₃ and Session E₇ have a ranking at two extreme ends of A₁ and A₄ in an opposite order. So, they cannot be considered as an optimum task allocation. In the final ranking Session E₁ and Session E₆ also rank high. However, they do not have an extremely high or low ranking at any attributes. Comprising these factors, we can consider these two sessions to be the optimum task allocation in the control of the pasteurization plant. When the performance and the mental load are considered as the most important factors, Session E₁ might be the best task allocation decision.

6.3.9.5 Problems in the AHP application

Three problems are encountered in the application of the AHP for subjective task allocation evaluation. The first one is that the number of pair-wise comparisons increases with the number of options and the number of attributes. This is a heavy burden to the subjects who felt bored doing pair-wise comparisons. This makes it difficult for the subjects to have a reasonable consistency in their judgments.
The second problem deals with the combination of the individual rankings. When a final ranking for all options is obtained by combining the ranking from each attribute, an extremely high ranking of an option in one of the attributes might play a significant role, even though the weight for this attribute is not so large. For example, the experimental session E₃ is ranked as the easiest configuration to operate, and it dominates strongly over most of the other configurations (see Appendix 4 for the dominance degree). Therefore, in the final ranking, Session E₃ ranks high even though other attributes have been considered by using a weighted approach. When we take into account other attributes, Session E₃ can not be an optimum task allocation configuration. Thus, we may conclude that the AHP ranking can only be used as a reference, and at the same time, other aspects have to be taken into account when making task allocation decisions.

The third problem is with the consistency of the subjective judgments. When the number of the options is large, or the attributes are not clearly expressed, it will be more difficult to have a reasonable consistency. Thus, to carry out the task allocation evaluation by using the AHP, it is better to reduce the options, and to make the evaluation criteria easy to be understood.

6.4 SUMMARY AND CONCLUDING REMARKS

This chapter has studied the quantitative measure for the degree of automation, the DofA, for a simulated pasteurization plant. It studies quantitatively the effect of the change of the DofA on human performance and mental load. The study further demonstrates that the concept of a quantitative measure for the degree of automation is useful to investigate the effect of automation on system performance as well as human performance.

Comparing the results of Ch. 6 with those of Ch. 5, we find that similar findings have been obtained as discussed in earlier sections. Firstly, although the number of data points in this study is smaller than that in Ch. 5, the relationships between performance, mental load and DofA reveal that as the DofA increases, both the systems have an increasing performance and a decreasing mental load for the operators. However, the performance will improve only slightly beyond a certain level of DofA.

Secondly, the overall mental load for controlling a partly-automated system in both the systems could be predicted by the task mental load and the overall mental load obtained from a situation that all subtasks are manually performed as long as the mental load does not exceed a maximum.

Thirdly, DofA^{TDL} and DofA^{TML} in both the implementations are closely related because these two are measured by the relationship between a task and its workload. The TML that is perceived by the human operator is created by the TDL that is meaningful if and only if the human operator is involved. However, the DofA^{TES} differs from these in that it is measured by
the relationship between a task and its effect on the performance that is independent from the human operator.

Finally, the relationship between the mental load and the performance from both experiments is in agreement with the inverted-U hypothesis (Bi and Salvendy, 1994; Sheridan, 1992).

An important difference between the pasteurization plant and the simple liner system is the introduction of discrete tasks, such as distribution switch control and overflow valve control. When the TES for these tasks is measured, these tasks must be partly performed. Otherwise, it will cause a poor system performance. Furthermore, other continuous tasks must also be partly performed in order to measure the TES associated with them. This demonstrates that all tasks in the pasteurization plant are more critical than the tasks in the linear system.

As discussed in Ch. 4 and Ch. 5, the approaches to decide the task weight in the DofA measure are system dependent. In the pasteurization plant, expert opinion was used to estimate the TDL based on the Hierarchic Task Analysis, whereas in Ch. 5 the system parameter, i.e. task complexity, was used to measure the TDL. All of the approaches are used according to the system to be studied.

The immediate purpose, the limitations, and the heuristic contribution of the DofA measure have been addressed in Ch. 4 and Ch. 5. What we need to do in the future is to develop an index which may be used in human-machine systems design. This needs more validation, especially in real systems.

The Analytic Hierarchy Process has been employed to evaluate the task allocation. A subjective ranking of different task allocation decisions can be obtained using the AHP approach instead of a statistical approach. However, AHP ranking can only be taken as a reference in the evaluation of task allocation. AHP may also be used to measure the relative level of human mental load. Moreover, we find that a task allocation configuration which is easy to control does not necessarily make the human operator feel comfortable to operate the system in such a configuration.
Chapter 7

Effect of Sudden Loss of Automated Tasks on Human Operators

The 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. It is found that the mental load perceived by the operators becomes higher due to the sudden drop of the DofA. The mental load arrives at a peak shortly after the DofA drops. The mental load will decrease to a low level a long period after the DofA drops. This low level of mental load may be perceived if the operators start to operate the system at that low value of DofA. The experimental data show that right after the DofA drops, system performance as well as 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.

7.1 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? What is the transition behavior as the DofA changes from a high level to a low level?

To the best of the author’s knowledge, there are no reports, or experiments, that have explicitly investigated this issue. Although some researchers, such as Huey (1989), Gluckman et al. (1991), Debernard et al. (1992), and Parasuraman et al. (1996), have investigated the effect of adaptive automation, or dynamic task allocation, 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. It is our intention to study the behavior of human performance and human 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. This transition is illustrated in Figure 3.1 as a hypothetical relationship between the mental load and the degree of automation. Figure 7.1 displays such a relationship again. The solid curve in the upper part of the figure presents the relationship. The mental load increases as the DofA decreases. According to the studies conducted in Chs 5 and 6, the mental load is a linear function of the DofA.

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 7.1 presents the relationship between the performance and the DofA. In Chs 5 and 6, we have demonstrated that the performance is a second-order polynomial of the degree of automation.

Figure 7.1 Hypothetical relationships between mental load, performance and degree of automation.
Based on the results of the experiments on the linear system (Ch. 5) and based on the results obtained with a simulated pasteurisation plant (Ch. 6), two experiments were conducted that included a sudden drop of the DofA. The following hypotheses are tested:

1. The operator perceives a much higher mental load shortly after the degree of automation of the system suddenly drops (from Point $A_m$ to Point $C_m$ in Figure 7.1).
2. The operator perceives a higher mental load (Point $C_m$) shortly after the DofA drops than when the system is operated for a longer time at that low DofA (Point $D_m$). 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 Point $D_m$ lies on the hypothetical relationship between mental load and DofA.
3. The performance reduces much when the DofA suddenly drops (from Point $A_p$ to Point $C_p$).
4. The performance will recover from the low level directly after the DofA drops (Point $C_p$) to the higher level (Point $D_p$) when the system is operated for a longer time at that low DofA. The recovered performance (Point $D_p$) can be achieved if the system is operated from the beginning at that low DofA. This means that Point $D_p$ lies on the hypothetical relationship between performance and DofA.

Hypotheses 1 and 2 are about the relation between the mental load and the DofA. Hypotheses 3 and 4 deal with the relation between the performance and the DofA.

This chapter presents the experimental results. In Section 7.2, the experiment with the linear system is described, and in the Section 7.3, the experiment with the pasteurization plant is presented. Then, following a general discussion based on two experiments, conclusions are drawn from the observations from both experiments.

### 7.2 EXPERIMENT ONE

The experiment was performed using the experimental system described in Ch. 5. Only a brief description on the experimental method will be given in this section. The controlled parameters are the number and the location of the automated cells, which result in different degrees of automation in the operation.

#### 7.2.1 Method

##### 7.2.1.1 Experimental set-up

The experimental configuration has been depicted in Fig. 5.1. The experimental task was to control the system to the requested Set-Point, SP. In this study, we still used a completely
forward connection (Ch. 5). The only difference was that the average time interval between two consecutive SP requests was 30 seconds instead of 40 seconds. Thus, for a system with the same task allocation but a higher frequency of the SP request, the system performance might be poorer and the operator mental load might be higher (Wei et al. 1995b).

7.2.1.2 Operator’s task

The operator’s control task was to generate an input for the appropriate cell(s), by means of a mouse or keyboard, to bring the cell’s output to the SP requested, and to maintain the other cells at their current set-points.

7.2.1.3 Participants

Six students (all male) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology 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.

7.2.1.4 Experimental sessions

The experimental sessions were designed according to the DofAs computed in Ch. 5. In this experiment, 6 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.

Table 7.1 Experimental sessions (☐: An automated cell; ■: A manually controlled cell)

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Task Allocation Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A = B</td>
</tr>
<tr>
<td>2</td>
<td>A = B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>B</td>
</tr>
</tbody>
</table>
A session might include 2 task allocation configurations ("A" and "B" as shown in Table 7.1) 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 7.1 presents all formal 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 the operator 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.

7.2.1.5 Protocol

Figure 7.2 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\textsuperscript{1}. 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\textsuperscript{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\textsuperscript{III}. The meanings of SPF\textsubscript{1}, SPF\textsubscript{2}, and oml\textsubscript{1} etc. will be addressed later.

![Figure 7.2 Time intervals and task allocation schemes for a session. A, B: Task allocation configuration. I, II, and III: Time interval. OML\textsuperscript{1}: OML during Interval I; oml\textsubscript{1}: OML during the last 3 minutes of Interval I. SPF\textsubscript{1}: SPF during the last 3 minutes of Interval I.](image-url)
7.2.1.6 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 Ch. 5. SPF was calculated for three time intervals. OML for each time interval was rated based on the RSME (Zijlstra, 1993) using the form in Appendix 3.

It is well-known, as discussed in Ch. 3, that the subjective mental load rating instrument cannot 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 level during the 3 minutes after the DofA changes.

In Chs 5 and 6, we have demonstrated that the relationship between mental load and DofA is linear. In Section 7.2.2.2 and Section 7.3.2.2, we will proof that the time effect on the rated mental load can be ignored. Consider that during the first 3 minutes of Interval II the operators operate 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:

\[
\frac{3 \cdot oml_1^{II} + 3 \cdot oml_2^{II}}{6} = OML^{II}.
\]

(7.1)

Since \( oml_1^{II} = OML^{I} \), we have:

\[
 oml_2^{II} = 2 \cdot OML^{II} - OML^{I}.
\]

(7.2)

7.2.2 Results and Discussion

We perform a repeated measures analysis of variance, or ANOVA, (Norusis/SPSS Inc., 1993) for the experimental data because the data from the subjects are not independent under different conditions (different time intervals, or DofAs). The ANOVA results are based on multivariate tests of significance using SPSS software and are reported in a form of \( F(df_1, df_2) \) with \( F \)'s significant level, \( p \). The \( F \) is the ratio between the mean square of the hypothetical data sets and the mean square of the error. \( df_1 \) is the degree of freedom for the hypothetical data sets, and equal to: The number of conditions - 1. \( df_2 \) is the degree of freedom for the error, and equals to: The number of cases - \( df_1 \). In the analysis, if the probability, \( p \), of obtaining the observed \( F \) value is smaller than 0.05, we reject the null hypothesis. The significance test is also performed by using averaged tests of significance in SPSS. Similar conclusions from two tests are obtained.
Table 7.2 Overall mental load and system performance in Sessions 1 and 2 (□: An automated cell; □: A manually controlled cell)

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Time Interval Measured (min.)</th>
<th>Task Allocation</th>
<th>DefA (TDL based)</th>
<th>OML mean</th>
<th>OML SD</th>
<th>SPF mean</th>
<th>SPF SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-6</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.85</td>
<td>35.50</td>
<td>16.99</td>
<td>9.27</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.85</td>
<td>35.83</td>
<td>17.44</td>
<td>9.12</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.85</td>
<td>38.33</td>
<td>18.51</td>
<td>9.25</td>
<td>0.06</td>
</tr>
<tr>
<td>2</td>
<td>0-6</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.17</td>
<td>73.00</td>
<td>15.87</td>
<td>8.36</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.17</td>
<td>79.67</td>
<td>20.14</td>
<td>8.31</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>□□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□ □□□□□□□□□□□□□□</td>
<td>0.17</td>
<td>82.50</td>
<td>20.18</td>
<td>8.15</td>
<td>0.43</td>
</tr>
</tbody>
</table>

7.2.2.1 Data for performance and mental load
In Sessions 1 and 2, the reallocation was not introduced during Interval II. In Table 7.2, the OML and the SPF for these two sessions are presented for the entire time interval of all three intervals. 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. Table 7.3 presents the average values of the SPF across six subjects during the last 3 minutes of three intervals, i.e., SPF1, SPF2, and SPF3, as indicated in Figure 7.2. For the OML on the last 3 minutes, Table 7.3 lists the average values of oml1, oml2, and oml3. As discussed before, we assume that the oml1 = OML1 and oml3 = OML3. The oml2 was calculated using Eq. (7.2). 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, rather than 1 as in Ch. 5, is used to prevent the SPF from becoming a negative value. Thus, the maximum performance will not exceed 10.

7.2.2.2 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 7.4 shows the ANOVA results in analyzing the significance for the effect of time on SPF and OML. From Table 7.3, and Table 7.4, 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 7.3 Overall mental load and system performance during the last 3 minutes in Sessions 1 to 6 (square: An automated cell; circle: A manually controlled cell)

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Time Interval Measured (min.)</th>
<th>Task Allocation</th>
<th>DofA (TDL based)</th>
<th>omi₁₂</th>
<th>SPF₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3-6</td>
<td></td>
<td>0.85</td>
<td>35.50</td>
<td>16.99</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.85</td>
<td>36.17</td>
<td>20.43</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.85</td>
<td>38.33</td>
<td>18.51</td>
</tr>
<tr>
<td>2</td>
<td>3-6</td>
<td></td>
<td>0.17</td>
<td>73.00</td>
<td>15.87</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.17</td>
<td>86.33</td>
<td>28.76</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.17</td>
<td>82.50</td>
<td>20.18</td>
</tr>
<tr>
<td>3</td>
<td>3-6</td>
<td></td>
<td>0.85</td>
<td>29.67</td>
<td>15.92</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.17</td>
<td>98.00</td>
<td>27.72</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.17</td>
<td>78.33</td>
<td>17.73</td>
</tr>
<tr>
<td>4</td>
<td>3-6</td>
<td></td>
<td>0.67</td>
<td>69.50</td>
<td>23.83</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.16</td>
<td>108.17</td>
<td>12.23</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.16</td>
<td>91.33</td>
<td>19.21</td>
</tr>
<tr>
<td>5</td>
<td>3-6</td>
<td></td>
<td>0.53</td>
<td>71.33</td>
<td>16.97</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.000</td>
<td>99.00</td>
<td>22.38</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.000</td>
<td>98.83</td>
<td>14.87</td>
</tr>
<tr>
<td>6</td>
<td>3-6</td>
<td></td>
<td>0.70</td>
<td>48.50</td>
<td>16.18</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td></td>
<td>0.28</td>
<td>103.83</td>
<td>37.99</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td></td>
<td>0.28</td>
<td>83.33</td>
<td>15.51</td>
</tr>
</tbody>
</table>
Table 7.4 ANOVA for OML and SPF in three time intervals of Sessions 1 and 2

<table>
<thead>
<tr>
<th>Session</th>
<th>Variables</th>
<th>ANOVA Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F(2, 4)</td>
</tr>
<tr>
<td>1</td>
<td>Overall mental Load</td>
<td>1.57</td>
</tr>
<tr>
<td></td>
<td>System performance</td>
<td>0.37</td>
</tr>
<tr>
<td>2</td>
<td>Overall mental Load</td>
<td>7.25</td>
</tr>
<tr>
<td></td>
<td>System performance</td>
<td>0.88</td>
</tr>
</tbody>
</table>

7.2.2.3 Mental load and DofA

The objective of the present study is to investigate the effect of a sudden drop in DofA on mental load. 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$^1$ and OML$^III$ are illustrated in Figure 7.3. The ANOVA showed that $oml^2_{II}$ was significantly higher than OML$^1$ for all sessions, $F(1, 5) > 15.5, p < 0.01$. We conclude that Hypothesis 1 is confirmed by the experiment.

Figure 7.3 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 7.3, 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. $oml_2^{III}$. The ANOVA test showed that $oml_2^{II}$ was significantly higher than $oml_2^{III}$, $F(1, 5) > 5.5$, $p < 0.05$, 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 7.1. If so, we can conclude that Hypothesis 2 is also confirmed by the experiment. This is discussed as follows.

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 7.4, i.e. the relationship between OML$^1$ and DofA. According to the study in Ch. 5, a linear polynomial is fitted to the 5 data points as presented in Figure 7.4.

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 7.3. The calculated OML, noted as OML$^{CL}$, and the OML$^{III}$ for Sessions 3 to 6 are presented in Figure 7.5.

The $t$-test shows that the measured OML$^{III}$ in Sessions 3 to 6 is 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 decline 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 $D_m$ in Figure 7.1. So, Hypothesis 2 is proved by the experiment.

![Figure 7.4 Relationship between overall mental load (OML$^1$) and DofA.](image-url)
Figure 7.5 Comparison between the calculated OML_{CL} and the measured OML^{III}. ×: An OML^I point in Figure 7.4, when DofA = 0.168, OML^I = 73.0.

7.2.2.4 System performance and DofA

In Ch. 5, we have demonstrated that system performance reduces when DofA declines. In this experiment, we intend to investigate 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 analyzed based on the SPF in the last 3 minutes of all three time intervals. These performances are compared in Figure 7.6.

Figure 7.6 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.
Figure 7.7 Relationship between performance ($\text{SPF}_2^I$) and DofA.

The $\text{SPF}_2^I$ is significantly better than $\text{SPF}_2^\text{II}$ as expected ($p < 0.01$), and this confirms Hypothesis 3. $\text{SPF}_2^\text{III}$ in Sessions 3, 4, and 5 is larger than $\text{SPF}_2^\text{II}$ (not significantly, $F(1, 5) < 6.0$, $p > 0.05$). $\text{SPF}_2^\text{II}$ and $\text{SPF}_2^\text{III}$ 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.

The relationship between SPF and DofA in the case that the experimental system was operated without dynamic task allocation is plotted in Figure 7.7, i.e. the relationship between $\text{SPF}_2^I$ and DofA. A second-order polynomial, $\text{SPF}_{\text{CL}}$, is fitted to the 5 data points as presented in Figure 7.7.

Based on this function, the SPF could be calculated for low DofA values which are the values during Interval III as shown in Table 7.3. The calculated SPF, noted as $\text{SPF}_{\text{CL}}$, and the $\text{SPF}_2^\text{III}$ for Sessions 3 to 6 are presented in Figure 7.8.

Figure 7.8 Comparison between the calculated $\text{SPF}_{\text{CL}}$ and the measured $\text{SPF}_2^\text{III}$. $\times$: An SPF$_2^I$ point in Figure 7.7, when DofA = 0.17, SPF$_2^I$ = 7.60.
The t-test shows that the measured SPF^{III}_2 in Sessions 3 to 6 is 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 7.1. So, Hypothesis 4 is proved by the experiment.

### 7.3 EXPERIMENT TWO

This experiment was conducted using the pasteurization plant (Ch. 6). Only a brief description on the experimental method as well as the differences in the system conditions will be given in this section.

#### 7.3.1 Method

**7.3.1.1 Experimental system**

The pasteurization plant is shown in Figure 6.1. The experimental task was to pasteurise as much of the available raw juice as possible. In this study, the experimental task was broken up into 5 control subtasks which could be automated. The automation of the subtasks was subject to failure. The human operator had to respond to this failure and had to execute the subtasks manually. The subtasks were the controls of the feed-stock pump speed, the steam pump, the steam heater, the overflow valve, and the distribution switch. Another difference in the experimental condition compared to the experiment conducted in Ch. 6 was that the disturbance frequency for the steam heater was 50% higher in this experiment. In this way, the experimental time could be reduced.

**7.3.1.2 Operator's task**

Since in this study we want to investigate the effect of a drop in the DofA on mental load and performance, the operators should control at least one subtask. The goals that the operator must achieve were: (1) To maintain the temperature of the juice flowing out the primary heater within an allowed range, (2) to maintain the level of the input tank at a proper level, and (3) to distribute the pasteurized juice effectively.

**7.3.1.3 Experimental sessions**

The experimental sessions were designed according to the DofA calculated in Ch. 6. The experiment consists of 10 sessions. Each session had 3 time intervals of 6 minutes each, and included a transition from a high DofA to a low DofA. The time schedule for the drop in DofA and for the mental load measurement is the same as presented in Figure 7.2.
Table 7.5 Experimental Sessions in Experiment 2 (AC: An automatically controlled subtask; MC: A manually controlled subtask)

<table>
<thead>
<tr>
<th>Session</th>
<th>Task allocation</th>
<th>Subtask 1 (feed pump)</th>
<th>Subtask 2 (overflow valve)</th>
<th>Subtask 3 (steam heater)</th>
<th>Subtask 4 (steam pump)</th>
<th>Subtask 5 (distribution)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A = B</td>
<td>AC</td>
<td>AC</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
</tr>
<tr>
<td>2</td>
<td>A = B</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
<td>MC</td>
</tr>
<tr>
<td>3</td>
<td>A → B</td>
<td>AC → MC</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>MC → MC</td>
</tr>
<tr>
<td>4</td>
<td>A → B</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
<td>AC → MC</td>
<td>MC → MC</td>
</tr>
<tr>
<td>5</td>
<td>A → B</td>
<td>AC → MC</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>AC → AC</td>
</tr>
<tr>
<td>6</td>
<td>A → B</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
<td>AC → MC</td>
<td>AC → AC</td>
</tr>
<tr>
<td>7</td>
<td>A → B</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>AC → AC</td>
</tr>
<tr>
<td>8</td>
<td>A → B</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
<td>MC → MC</td>
<td>MC → MC</td>
</tr>
<tr>
<td>9</td>
<td>A → B</td>
<td>AC → MC</td>
<td>AC → AC</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
</tr>
<tr>
<td>10</td>
<td>A → B</td>
<td>MC → MC</td>
<td>MC → MC</td>
<td>AC → MC</td>
<td>AC → AC</td>
<td>AC → MC</td>
</tr>
</tbody>
</table>

Table 7.5 summarizes the experimental sessions and task allocation configurations. Again, in order to check the time effects on operator mental load and performance, Sessions 1 and 2 did not change task allocation during all time intervals. Sessions 3 to 10 changed task allocation during Interval II. The change of task allocation will make each session drop at least 60% of the initial value of DofA. Subjects were informed, in advance, about the dynamic task allocation, the moment when the allocation happened, and the cells whose automated control might fail.

7.3.1.4 Participants

Seven students (male and female) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology participated voluntarily as operators. They were paid an hourly rate for their participation plus a bonus contingent on performance. Before formal sessions started, the subjects performed 10 training sessions of 6 minutes each to learn to control the system. Each subject performed 10 formal experimental sessions that lasted 18 minutes each.

7.3.1.5 Measurements

The plant variables recorded in this experiment were the same as those in the experiment conducted in Ch. 6. The overall mental load was subjectively assessed based on the RSME (Zijlstra, 1993). System performance was defined as the percentage of pasteurization—the ratio between the amount of the pasteurised juice and the available input raw juice during the measurement period. Operator Performance, OPF, took into account several variables recorded during the experiment besides the percentage of the pasteurization. All of these
variables were weighted and then, were combined to give a measure for the operator performance (for details, see Ch. 6).

7.3.2 Results and Discussion

7.3.2.1 Data for performance and mental load

In Sessions 1 and 2, the reallocation was not introduced during Interval II. In Table 7.6, the OML and the SPF for these two sessions are presented for the entire time interval of all three intervals.

Again, we focus our analysis on the last 3 minutes of each time interval. The average values of the OML, the SPF and the OPF across seven subjects during the last 3 minutes of three intervals are presented in Table 7.7. Note that a larger value indicates a better performance. Eq. (7.2) is used to extract \( \text{oml}_2 \) from OML.

7.3.2.2 Effect of time on mental load and system performance

For the pasteurization plant, it is also necessary to check the time effect on operator’s perception of mental load and performance. For this reason, Session 1 and Session 2 can be used since the task allocation configuration was not changed. Table 7.8 shows the ANOVA results in analyzing the significance for the effect of time on the OML during three time intervals.

From Table 7.7 and Table 7.8, we observe that as the time went on, the mental load perceived by the operators for both sessions raised, but not significantly. The system performance and the operator performance did not decrease significantly either. This allows us to neglect the effect of time on mental load and system performance.

Table 7.6 Overall mental load, system performance, and operator performance from Session 1 and Session 2

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Time Interval Measured (min.)</th>
<th>DofA (TDL based)</th>
<th>OML</th>
<th>SPF</th>
<th>OPF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>mean</td>
<td>SD</td>
<td>mean</td>
</tr>
<tr>
<td>1</td>
<td>0–6</td>
<td>0.48</td>
<td>40.80</td>
<td>14.45</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>6–12</td>
<td>0.48</td>
<td>46.20</td>
<td>11.65</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>12–18</td>
<td>0.48</td>
<td>45.60</td>
<td>11.62</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>0–6</td>
<td>0.00</td>
<td>71.20</td>
<td>23.06</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>6–12</td>
<td>0.00</td>
<td>77.00</td>
<td>20.32</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>12–18</td>
<td>0.00</td>
<td>78.80</td>
<td>21.71</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Table 7.7 Overall mental load, system and operator performance during the last 3 minutes from Session 1 to Session 10

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Time Interval Measured (min.)</th>
<th>DoIA (TDL based)</th>
<th>omI₂</th>
<th>SPF₂</th>
<th>OPF₂</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>SD</td>
<td>mean</td>
<td>SD</td>
</tr>
<tr>
<td>1</td>
<td>3-6</td>
<td>0.48</td>
<td>40.80</td>
<td>14.45</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.48</td>
<td>53.43</td>
<td>10.70</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.48</td>
<td>45.60</td>
<td>11.62</td>
<td>0.90</td>
</tr>
<tr>
<td>2</td>
<td>3-6</td>
<td>0.00</td>
<td>71.20</td>
<td>23.06</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.00</td>
<td>85.86</td>
<td>18.57</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.00</td>
<td>78.80</td>
<td>21.71</td>
<td>0.91</td>
</tr>
<tr>
<td>3</td>
<td>3-6</td>
<td>0.48</td>
<td>46.71</td>
<td>14.82</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.00</td>
<td>83.85</td>
<td>35.25</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.00</td>
<td>75.29</td>
<td>21.51</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>3-6</td>
<td>0.32</td>
<td>38.33</td>
<td>16.63</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.00</td>
<td>116.67</td>
<td>64.91</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.00</td>
<td>80.33</td>
<td>22.04</td>
<td>0.88</td>
</tr>
<tr>
<td>5</td>
<td>3-6</td>
<td>0.68</td>
<td>20.43</td>
<td>10.89</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.20</td>
<td>69.00</td>
<td>32.44</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.20</td>
<td>49.57</td>
<td>18.65</td>
<td>0.89</td>
</tr>
<tr>
<td>6</td>
<td>3-6</td>
<td>0.52</td>
<td>26.29</td>
<td>17.60</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.20</td>
<td>86.29</td>
<td>30.07</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.20</td>
<td>56.14</td>
<td>18.74</td>
<td>0.90</td>
</tr>
<tr>
<td>7</td>
<td>3-6</td>
<td>0.80</td>
<td>12.14</td>
<td>8.41</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.20</td>
<td>72.43</td>
<td>50.62</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.20</td>
<td>57.57</td>
<td>22.33</td>
<td>0.89</td>
</tr>
<tr>
<td>8</td>
<td>3-6</td>
<td>0.60</td>
<td>17.86</td>
<td>12.18</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.00</td>
<td>123.86</td>
<td>49.09</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.00</td>
<td>82.7</td>
<td>26.6</td>
<td>0.86</td>
</tr>
<tr>
<td>9</td>
<td>3-6</td>
<td>0.68</td>
<td>27.17</td>
<td>18.47</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.016</td>
<td>107.50</td>
<td>18.57</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.016</td>
<td>78.8</td>
<td>20.9</td>
<td>0.89</td>
</tr>
<tr>
<td>10</td>
<td>3-6</td>
<td>0.52</td>
<td>24.71</td>
<td>11.26</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>9-12</td>
<td>0.04</td>
<td>114.71</td>
<td>34.89</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>15-18</td>
<td>0.04</td>
<td>77.29</td>
<td>24.99</td>
<td>0.89</td>
</tr>
</tbody>
</table>
Table 7.8 ANOVA tests for OML, SPF, and OPF in three time intervals

<table>
<thead>
<tr>
<th>Session</th>
<th>Variables</th>
<th>ANOVA Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$F(2,5)$</td>
</tr>
<tr>
<td>1</td>
<td>Overall mental Load</td>
<td>5.04</td>
</tr>
<tr>
<td></td>
<td>System performance</td>
<td>6.14</td>
</tr>
<tr>
<td></td>
<td>Operator performance</td>
<td>0.24</td>
</tr>
<tr>
<td>2</td>
<td>Overall mental Load</td>
<td>4.76</td>
</tr>
<tr>
<td></td>
<td>System performance</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Operator performance</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Figure 7.9 Variation of mental load after a sudden drop of DofA.
7.3.2.3 Mental load and DofA

As mentioned above, the experiment based on the pasteurization plant is used to test the hypotheses about the relationship between OML and DofA when the latter is suddenly dropped. From Table 7.7, one can see that the DofA in each session dropped at least 60% of the initial value and that the absolute drop of the DofA was larger than 0.3. Figure 7.9 illustrates the variation of the OML for Sessions 3 to 10 (the standard deviation is shown in Table 7.7). The ANOVA shows that \( oml_{II} \) was significantly larger than OML, \( F(1, 6) > 8.5, p < 0.05 \). This result confirms Hypothesis 1.

Moreover, the ANOVA shows that \( oml_{II} \) was significantly higher than OML, \( F(1, 6) > 9.0, p < 0.05 \); except for Session 3 and Session 7 in which \( oml_{II} \) was not significantly higher than OML, \( F(1, 6) < 0.8, p > 0.1 \). The similar observation has been made in Experiment 1 where \( oml_{II} \) was significantly higher than OML in most of the sessions.

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 7.10, i.e. the relationship between OML and the DofA. According to the study in Ch. 6, a linear polynomial is fitted to the 6 data points as presented in Figure 7.10. 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 7.7. The calculated OML, noted as OML\(_{CL}\), and the OML for Sessions 3 to 10 are presented in Figure 7.11.

The \( t \)-test shows that the measured OML in Sessions 3 to 10 is not significantly different from the calculated OML at the specific values of the DofA, \( p > 0.05 \). We can conclude that the OML, a longer period after the DofA drops to a lower value, can decline 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\(_{m}\) in Figure 7.1. So, Hypothesis 2 is also proved by Experiment two.

![Figure 7.10 Relationship between overall mental load (OML) and DofA.](image-url)
Figure 7.11 Comparison between the calculated OML\textsubscript{CL} and the measured OML\textsuperscript{II}. #: An OML\textsuperscript{I} point in Figure 7.10, when DoF\textalpha = 0, OML\textsuperscript{I} = 71.2.

7.3.2.4 Performance and DoF\textalpha

In this experiment, we want to confirm Hypotheses 3 and 4 addressed in Section 7.1. In Ch. 6, we have demonstrated that the system performance does not significantly change for different DoF\textalpha s. In this experiment, the SPF\textsubscript{s} for Interval I and Interval II, i.e. SPF\textsuperscript{I} and SPF\textsuperscript{II}, are also not significantly different, \( F(1, 6) < 1.9, p > 0.1 \). Although the difference is very small, the trend can be observed that the SPF is reduced due to a drop in DoF\textalpha. The OPF\textsuperscript{I} is significantly larger than OPF\textsuperscript{II}, \( F(1, 6) > 8.8, p < 0.05 \). This is expected since the DoF\textalpha dropped during Interval II. A closer observation of the SPF and the OPF during the last 3 minutes of Interval I and Interval II reveals a similar result, for SPF\textsuperscript{I} and SPF\textsuperscript{II}, \( F(1, 6) < 2.0, p > 0.1 \), and for OPF\textsuperscript{I} and OPF\textsuperscript{II}, \( F(1, 6) > 7.5, p < 0.05 \), except for Session 6. In Session 6, OPF\textsuperscript{I} and OPF\textsuperscript{II} has no significant difference, \( F(1, 6) = 0.40, p = 0.55 \), because the drop in DoF\textalpha was too small to create a significant difference. Nevertheless, we may conclude that the reduction in performance due to a drop in DoF\textalpha confirms Hypothesis 3.

For Hypothesis 4, the data cannot give an obvious confirmation. The SPF\textsuperscript{II} and SPF\textsuperscript{III} are not significantly different, \( F(1, 6) < 3.3, p > 0.1 \). We can only observe the trend that SPF\textsuperscript{III} tends to be better than SPF\textsuperscript{II} in most of the sessions. For the OPF, a similar result can be found. The OPF\textsuperscript{II}, and OPF\textsuperscript{III} are not significantly different, \( F(1, 6) < 5.2, p > 0.05 \), and only in most of the sessions OPF\textsuperscript{III} tends to be better than OPF\textsuperscript{II}. Based on this analysis, we conclude that Hypothesis 4 cannot be confirmed in this experiment.

7.4 GENERAL DISCUSSION

Many researches 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 hypothesized 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 findings in our study reveal that while the operators managed the fault, they could maintain their performance constantly. 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 drops, the operators face another system: They have to perform more tasks. After they operate this other system for a longer period of time, the operators accumulate experience so that they can improve the performance. Shortly after the DofA drops, the operators have to change suddenly their attention and to move mentally to perform more tasks. So, the operators perceive the highest mental load right after the transition. After having gained more experience, the operators can invest less mental effort and therefore they perceive lower mental load which is still higher than that they perceive before the reallocation. The mental load increases probably because more manual tasks have 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. In practice, in order to maintain their operational skills and awareness of the system states, the operators often override the automation to perform scheduled or unscheduled manual interventions even though automation does not fail (Hockey and Maule, 1995).

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 as discussed in Ch. 3.

This study suggests again that the quantitative measure for the degree of automation could be used to investigate the effect on operator performance of a change of the degree of automation. The relationships between performance, mental load and DofA have been studied in Chs 5 and 6. However, it is a problem whether the amount of change in mental load and performance can directly be predicted by the amount of change in DofA. During this study, we observe that the amount of performance reduction and the amount of mental load increase do not simply depend on the amount of the DofA reduction. They may be affected by many factors, such as: System states, operator states, the automated tasks, and the DofA before the DofA drop. This needs to be studied in the future.

A change in DofA will affect many aspects in the system design, such as: Information presentation, interface design, and alarming. The ways that the operators deal with the change in DofA will depend on situation awareness, operator’s skill, and operator’s motivation.

7.5 CONCLUSIONS

This chapter presents an implementation of the quantitative measure of the degree of automation in a linear system and in a simulated process plant. The effort has been put into investigating the behavior of performance and human mental load when the degree of automation suddenly drops. Based on the experimental 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.

These conclusions indicate that the period shortly after automation failure is very important and should get more attention. A comprehensive support system may be helpful to support the human operators during this period.
Chapter 8

Human Monitoring Behavior of Automated Systems

The human operator monitoring behavior of the automated subsystems of a simulated pasteurization plant is experimentally investigated. The automatic controllers of the subsystems are subject to failure during the operation. The operator monitoring behavior is measured by the eye fixation time and the number of monitoring samples. The experimental data show that operator performance is related to monitoring behavior. More observations are given to highly critical automated subsystems. It takes a certain time for the operator to rebuild his trust in automation after the automation recovers from a failure. The rebuilding of the trust is affected by the nature of the automated tasks. A task with high task demand load and high task criticality may have a high monitoring task demand load. The high monitoring mental load will contribute to the overall mental load.

8.1 INTRODUCTION

To supervise an automated process, the human operators monitor not only the process, but also the status of the automation itself. The monitoring of the whole process is necessary for the operators to stay aware of the situation of the process (Endsley, 1995a, b) and to carry out fault management (Moray and Rotenberg, 1989; Parasuraman et al., 1996). An increased level of automation may result in that the operator’s mental load is mainly determined by the monitoring task. It is not surprising if such situations exist where the requirement for monitoring automation creates too much workload and makes the operator feel that automation kills us with kindness (Sheridan, 1992).

Many researches have investigated the human monitoring of automation in relation to fault management (Moray and Rotenberg, 1989; Parasuraman et al., 1996), human vigilance of automation failure (Molloy and Parasuraman, 1996), human trust in automation (Lee and Moray, 1992), human use of automation (Riley, 1994) and situation awareness (Endsley, 1995; Wickens, 1994). Although the operator now experiences a high mental load mainly from monitoring due to the increased automation level, we have not seen reports on the assessment of
mental load for monitoring automated systems. In Ch. 4, we have addressed that the monitoring tasks for monitoring automation can, or should, be included in the quantitative measure for the level of automation. Moreover, in Chs 5 and 6, we take into account the mental load for controlling and monitoring a task together, and neglect the mental load created by the monitoring automated subsystems. In fact, automating a task can create new monitoring task, and consequently, will impose monitoring workload on the operator. This motivates us to investigate how the overall mental load in operating a system is affected by monitoring an automated subsystem. This is the purpose of the study presented in this chapter.

Another interesting problem is that the monitoring behavior of automated systems is associated with the human trust in automation. It is hypothesized that the operators will pay more attention to those subsystems that failed in the past and that were recovered, e.g. automated subsystems that the operator does not trust very well. The third question deals with the link between monitoring behavior and operator performance.

To investigate the above questions and to find to what extent the operators monitor the automated subtasks under different conditions, we conducted an experiment with the pasteurization plant (as presented in Ch. 6) by using an eye tracking instrument to track the eye movements and to measure eye-fixation times of the operators. This chapter presents the experimental results and the analysis.

The following hypotheses are tested in the experiment:
1. Operator performance is affected by the monitoring process.
2. If the operator is informed that automation of a subsystem may fail, he will observe the error prone subsystems more frequently. The operator will spend more time monitoring the automation.
3. The operator pays more attention to an automated subsystem that recovers from a failure than before the failure occurs.
4. Highly critical subtasks are monitored more frequently.
5. The monitoring of automated subtasks contributes to the overall mental load.

8.2 METHOD

The experiment was conducted based on the pasteurization plant as used in Ch. 6. A brief description on the experimental method as well as the differences in system conditions are given in this section.
8.2.1 Experimental System

The pasteurization plant used in this experiment was the same as the plant shown by Figures 6.1 and 6.2. The experimental task was to pasteurize as much of the available raw juice as possible, and to distribute it to an available tank. In this study, the experimental task was broken into 3 control subtasks associated with 3 subsystems. The automatic controller for each automated subsystem was subject to failure. If automation failed, the human operator had to control the appropriate subsystem manually. The three subsystems were: A flow control loop, a steam control loop, and a production distribution control. In order to shorten the experimentation time, in the experimental condition compared to the experiment conducted in Ch. 6, the disturbance frequency for the steam was 1.5 times higher in this set of experiments.

8.2.2 Operator’s Task

The goals that the operator must achieve were three fold: (1) To maintain the temperature of the juice flowing out of the primary heater within an allowed range (70 °C to 85 °C), (2) to maintain the level of the input tank at a proper level, and (3) to distribute the pasteurized juice efficiently.

8.2.3 Experimental Sessions

The experiment consisted of five sessions. For each session, the allocation of the three subtasks differed. Table 8.1 presents the experimental sessions with the task allocations. The fifth column of Table 8.1 indicates that the subjects were informed that automation could fail or not. The last column indicates whether automation really failed or not. Except for the last column, Table 8.1 was presented to the subjects before the experiment. Thus, a subject only knew about the possibility that a failure could happen and that the failure might occur after first six minutes. However, the subject did not know whether the failure would actually occur and at what exact instant the failure would occur.

Each session lasted 18 minutes. Figure 8.1 shows the time schedule. As can be seen, each session had 3 time intervals of 6 minutes each. First, the operator controlled the plant with a subsystem automated for about 6 minutes. Then, the automation for the automated subsystem could fail. The failure was indicated by the appearance of a manual control slide and switch on the display at the instant of the failure. The instant that the failure occurred depended on different sessions, and was always after 6 minutes. When automation failed, the operator took over the control of that subsystem. After the operator controlled the plant for about 6 minutes, the automation recovered from its failure status. The operator would control the plant for another 6 minutes before the end of the session. At the end of the session, the operator rated the OML.
Table 8.1 Experimental sessions and automation failure (A: Automated subtask; M: Manually controlled subtask)

<table>
<thead>
<tr>
<th>Session</th>
<th>Subtask 1</th>
<th>Subtask 2</th>
<th>Subtask 3</th>
<th>Automation Failure</th>
<th>Failure Scheduled</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>A</td>
<td>M</td>
<td>Excluded</td>
<td>No</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>A</td>
<td>M</td>
<td>Possible</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>A</td>
<td>M</td>
<td>Possible</td>
<td>No</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>M</td>
<td>M</td>
<td>Possible</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>M</td>
<td>M</td>
<td>Possible</td>
<td>No</td>
</tr>
</tbody>
</table>

Figure 8.1 Time intervals for a session.

8.2.4 Participants

Because the costs of hiring the eye tracking instrument were considerable to perform all 5 sessions, only two subjects were invited. The two students (male) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology participated voluntarily in the experiment. They were paid a fee for their participation plus a bonus contingent on performance. They participated in the experiments conducted in Chs 6 and 7, respectively. They were chosen according to their experience and their performance during the previous experiments as well as to their eye movement characteristics. Since the eye tracking instrument measures the movement of the left eye, an inactive left eye would affect the measurement too much. On the basis of previous experience and performance, it was concluded that Subject 1 understood the plant operation better. Subject 1 operated the plant 2 months ago during previous experiments, whereas Subject 2 operated the plant 4 months ago. The subjects performed 5 training sessions with the eye tracking measurement sensor mounted on their heads. The training session lasted 6 minutes each.
8.2.5 Measurements

8.2.5.1 Performance and plant variable measurement

In the experiment, the logged plant variables were the same as in Chs 6 and 7. System performance and operator performance were defined in the same way as in Chs 6 and 7.

8.2.5.2 Eye movement measurements

To measure which part of the pasteurization plant the operator was monitoring, the Gaze-Tracker (Mooij, 1995) as shown in Figure 8.2 was used. This instrument measured the operator's left eye pupil position. The following variables were recorded into the data file: (1) The X-Y ordinate on the screen indicating the operator's eye focus point, (2) fixation time duration, (3) the absolute time, and (4) the pupil diameter.

The sampling time of the Gaze-Tracker was 20ms. However, to perform data reduction, the measured data were processed such that only when the fixation time was longer than 150ms the eye movement variables were stored in the data file.

The total fixation times recorded for each subject and for each session are different due to data reduction. In order to compare the difference in monitoring time between two subjects, we use the percentage of fixation time that the subject monitored a subsystem in the total fixation time. This is noted as \( P_f \).

Figure 8.2 Gaze-Tracker for measuring the eye movement.
When a subject has a look at a subsystem, we define that the subject has made a sample in monitoring the subsystem. Each sample may have different fixation time ($\geq 150\text{ms}$). In order to observe the frequency at which a subject takes samples of a subsystem, we define the monitoring frequency as follows:

$$F_m = \frac{\text{total number of samples}}{\text{total fixation time}}.$$ 

For convenience of the measurement, the interface of the pasteurization plant (as shown in Ch. 6) was projected on the back of a flat screen with a size $51.6 \times 43.5\ \text{cm}$ (width $\times$ height) so that the image of the plant was enlarged by a factor of 2.06 horizontally and 2.12 vertically compared to the size used in the experiments in Ch. 6.

### 8.2.5.3 Calibration

The position measurement was calibrated for each subject before each experiment. After the Gaze-Tracker equipment was mounted on the head of the subject and was adjusted, calibration was carried out as follows. The 4 corners of the flat screen were used as references. The operator was asked to look at each corner successively while the output of the Gaze-Tracker was recorded. The original point was calibrated with the lower left corner. The calibration was checked by looking at the upper left corner, the upper right corner, and the lower right corner. In this way the coordinates for a subject were determined.

### 8.3 RESULTS AND DISCUSSION

#### 8.3.1 Results

The overall system performance and the operator performance for two subjects are presented in Figure 8.3 and Figure 8.4. As can be seen, Subject 1 always achieved better results than Subject 2.

Table 8.2 shows the percentage of fixation time, $P_{fn}$, that the operator monitored a subsystem or an interesting plant variable during 18 minutes operation of the plant compared to the total fixation time recorded. Table 8.2 also shows the operator's monitoring frequency, $F_m$. 


Figure 8.3 System performance from two subjects. A larger value indicates better performance.

In this study, one way Analysis of variance, ANOVA, (Norusis/SPSS Inc., 1993) is used to perform statistical tests. The ANOVA result is described in a form of \( F(df_1, df_2) \) with \( F \)'s significant level, \( p \). \( F \) is the ratio between the within-groups and between-groups variance estimates. \( df_1 \) is the degree of freedom associated with the between-groups, and \( df_2 \) is the degree of freedom associated with the within-groups. In the analysis, if the probability, \( p \), of obtaining the observed \( F \) value is smaller than 0.05, we reject the null hypothesis.

Figure 8.4 Operator performance from two subjects. A larger value indicates better performance.
Table 8.2 Percentage of fixation time, $P_{ft}$, in monitoring subsystems and variables and monitoring frequency, $F_m$. Note: Total percentage is the sum of $P_{ft}$ that a subject monitored all subsystems and interesting variables. It is smaller than 100% because the subject might monitor other unrelated parts. Total fixation time is the sum of the fixation times of all monitoring samples that were recorded.

<table>
<thead>
<tr>
<th>Monitored Subsystems or Variables</th>
<th>Subj.</th>
<th>Session Number</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>input tank</td>
<td></td>
<td></td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
</tr>
<tr>
<td>1</td>
<td>18.4</td>
<td>0.30</td>
<td>11.2</td>
<td>0.19</td>
<td>14.0</td>
<td>0.25</td>
<td>1.4</td>
<td>0.03</td>
<td>0.2</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>28.6</td>
<td>0.52</td>
<td>26.0</td>
<td>0.51</td>
<td>28.5</td>
<td>0.51</td>
<td>21.2</td>
<td>0.41</td>
<td>16.8</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flow loop</td>
<td></td>
<td></td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8.9</td>
<td>0.27</td>
<td>9.4</td>
<td>0.31</td>
<td>10.4</td>
<td>0.33</td>
<td>15.2</td>
<td>0.46</td>
<td>15.3</td>
<td>0.49</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.4</td>
<td>0.01</td>
<td>1.3</td>
<td>0.05</td>
<td>2.1</td>
<td>0.06</td>
<td>9.9</td>
<td>0.31</td>
<td>14.0</td>
<td>0.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>steam loop</td>
<td></td>
<td></td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>36.4</td>
<td>0.49</td>
<td>38.2</td>
<td>0.55</td>
<td>34.4</td>
<td>0.53</td>
<td>32.9</td>
<td>0.50</td>
<td>40.9</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>32.9</td>
<td>0.51</td>
<td>30.1</td>
<td>0.49</td>
<td>33.8</td>
<td>0.54</td>
<td>39.1</td>
<td>0.43</td>
<td>39.3</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>temperature curve</td>
<td></td>
<td></td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>94.4</td>
<td></td>
<td>96.7</td>
<td></td>
<td>96.4</td>
<td></td>
<td>94.2</td>
<td></td>
<td>94.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>96.6</td>
<td></td>
<td>96.4</td>
<td></td>
<td>98.3</td>
<td></td>
<td>98.1</td>
<td></td>
<td>97.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>distribution tanks</td>
<td></td>
<td></td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td>$P_{ft}$</td>
<td>$F_m$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1040.8</td>
<td></td>
<td>976.48</td>
<td></td>
<td>969.05</td>
<td></td>
<td>965.29</td>
<td></td>
<td>949.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>885.40</td>
<td></td>
<td>889.12</td>
<td></td>
<td>897.86</td>
<td></td>
<td>932.81</td>
<td></td>
<td>1011.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8.3.2 Operator’s Performance and Interaction with the Plant

According to Moray (1996), the difference between poor and good operators in supervisory control can be identified by analyzing the ways in which the operators interact with the plant. We use the operator’s monitoring behavior of subsystems as an indication to show in which way the operator interacted with the plant. We assume: (1) When the operator has a control action on one of the subsystems, he has to focus his eye on the subsystem; (2) the operator interacts with a subsystem after a fixation time longer than 150 ms.

As can be seen in Table 8.2, during the 18 minutes operation, the total fixation time recorded for Subject 1 from Session 1 to Session 4, except for Session 5, was longer than that for Subject 2. However, the total percentages of fixation times in monitoring all subsystems and
interesting variables are not significantly different between two subjects, $F(1,8) < 0.006$, $p > 0.1$. A close observation on how the two subjects distributed their attentions to the interesting areas shows that they had different approaches in monitoring or interacting with the plant.

Figure 8.5 plots the average fixation times on five interesting variables and subsystems across Sessions 1 and 3, where the steam subsystem was automated and automation did not fail. We observe that Subject 1 paid significantly more attention to the temperature curve in terms of fixation time, $F(1,2) = 96.73$, $p < 0.01$. The same conclusion can be drawn in terms of $P_n$ and in terms of $F_m$, $F(1,2) = 54.58$, $p < 0.05$. The two subjects did not pay significantly different attention to the distribution subsystem in terms of fixation time, $P_n$ and $F_m$, $F(1,2) < 6.5$, $p > 0.1$. To the input tank, there was also no significant difference between two subjects, $F(1,2) < 18.1$, $p > 0.05$. However, we observe from Table 8.2 that Subject 1 had a higher $P_n$ and a higher $F_m$ in monitoring the input tank than Subject 2. To the steam control loop which was automated, both the subjects did not pay much attention. Subject 2 spent significantly more time on the flow control loop in terms of the fixation time, $F(1,2) = 90.37$, $p < 0.05$, and in terms of $P_{fl}$, $F(1,2) = 4871.90$, $p < 0.002$. The frequency, $F_m$, for Subject 2 to monitor the flow control loop is also significantly higher, $F(1,2) = 11881.00$, $p < 0.0001$.

In Session 4 and Session 5, where the flow control loop was automated, both subjects did not pay much attention to the input tank. As shown in Figure 8.6, Subject 2 paid significantly more attention to the steam control loop in terms of the fixation time, $F(1,2) = 22.26$, $p < 0.05$, and in terms of $P_{fl}$, $F(1,2) = 39.94$, $p < 0.05$. However, as can be seen in Table 8.2, the frequencies that both the subjects monitored the steam control loop, were not significantly different, $F(1,2) = 9.8$, $p > 0.1$. For other subsystems and variables, there were not significant differences between the two subjects (fixation time: $F(1,2) < 1.5$, $p > 0.3$; percentage of the total fixation time: $F(1,2) < 2.70$, $p > 0.2$; and monitoring frequency: $F(1,2) < 13.9$, $p > 0.06$).

![Figure 8.5](image-url)  
**Figure 8.5** Comparison of the average fixation times between two subjects for different subsystems. The fixation time is averaged across Session 1 and Session 3.
Figure 8.6 Comparison of the average fixation times between two subjects for different subsystems. The fixation times are averaged across Session 4 and Session 5.

From the analysis above, one can find the difference between two subjects in controlling the plant. Subject 2, who achieved poor performance, focused mainly on the manually controlled subtasks. He did not pay enough attention to the automated subsystem and its related variables. This indicates that Subject 2 was busy taking care of the tasks allocated to him, and had less interaction with the automated subsystems. However, Subject 1, who achieved good performance, monitored all the possible variables that may affect the performance, rather than that he focused only on the manually controlled subtasks. We conclude that two subjects have different strategies for the plant operation. Their strategies affect the way in which they monitor, or interact with, the plant. Due to the difference in monitoring the plant, they achieve different performance. This confirms Hypothesis 1.

8.3.3 Operator’s Monitoring of Automation

The approach that the operator monitors the automated subsystems when he is informed that the automation may fail is another topic that we want to investigate. The three hypotheses about the monitoring of the automated subsystems are Hypotheses 2, 3, and 4. The hypotheses are related to each other. We assume that the operator will lose his trust in automation due to the failure of automation. Because he distrusts the failed automation, even when the automation recovers from the failure, the operator still gives more attention to the automation. The first two hypotheses will be affected by the third. If automation for a highly critical subsystem fails, the operator’s trust in automation will take longer time to rebuild. During the period that the operator regains confidence, the operator will pay more attention to this subtask and its automation.

8.3.3.1 Monitoring behavior before the occurrence of automation failure

Table 8.1 shows that in Session 1 the steam control loop was automated and the automation
would not fail, something that the subjects knew. For Session 2 and Session 3, the subjects were only informed that the automation might fail, but the failure did not take place in Session 3. As can be seen in Table 8.2, even though Subject 1 knew that automation might fail he paid less attention to the automation in Session 3 compared to Session 1. Subject 2 paid more attention to the automation in Session 3 and had a more than 50% increment in the fixation time and in the monitoring frequency. Since in Session 2 the automation failed during Interval II, we need to check the monitoring behavior in Interval I, i.e. before automation failure for these three sessions. The fixation times in Interval I are plotted in Figure 8.7. The fixation times on other subsystems and interesting variables are listed in Table 8.3, which show where the operator was monitoring if he paid less attention to the automated subsystem.

As can be seen in Table 8.3 and Figure 8.7, during Interval I, Subject 1 showed a shorter fixation time on the steam loop automation from Session 1 to Session 3. However, Subject 2 had a fourfold longer fixation time in Session 2 and a doubled fixation time in Session 3 compared to that in Session 1. These results show that Subject 1 worried less about the automation failure in the steam control loop than Subject 2.

As can be seen in Table 8.3, Subject 1 moved his attention to other subsystems or other places. An interview after the experiment indicated that Subject 1 had a good understanding about the effect of the automation failure in the steam control loop. Even though the automation in this loop failed, it would take a long time to affect the temperature of the output juice. The operator had enough time to correct the steam control. Since Subject 2 worried more about the automation failure than Subject 1, he spent more time to monitor the steam loop. Due to the different results of the two subjects, we conclude that the experimental data will not confirm the Hypothesis 2, that is the operator will spend more time on monitoring the automated subsystem that he does not trust.

![Fixation time graph](image)

**Figure 8.7** Fixation time for monitoring the steam loop automation before automation failure (Interval I) and after automation recovered from its possible failure (Interval III). Note that in Sessions 1 and 3 there was no failure.
Table 8.3 Fixation time on subsystems and interesting variables during Interval I, Interval II and Interval III. The unit of fixation time is second.

<table>
<thead>
<tr>
<th>Subsystems or Variables</th>
<th>Subject</th>
<th>Session Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>II</td>
</tr>
<tr>
<td>input tank</td>
<td>1</td>
<td>59.8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>48.4</td>
</tr>
<tr>
<td>flow loop</td>
<td>1</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>129</td>
</tr>
<tr>
<td>steam loop</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>temperature cube</td>
<td>1</td>
<td>34.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>distribution tanks</td>
<td>1</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>123</td>
</tr>
</tbody>
</table>

In Sessions 4 and 5, the flow control loop was automated, whereas the steam control loop was manually controlled. The fixation times on the flow control loop from two subjects are plotted in Figure 8.8. As can be seen in Figure 8.7, Figure 8.8, and Table 8.3, both the subjects gave significantly more attention to the automation of the flow control than that to the steam control before the automation failure during all time intervals. These observations demonstrate that the subjects considered that the flow control was more critical than the steam control. The subjects worried about the automation failure in the flow control loop. The subjects knew that the automation in both the sessions might fail. The fixation times (averaged across two subjects) between two sessions are not significantly different, $F(1,1) = 0.6, p > 0.5$. This is also true for the number of monitoring samples, $F(1,1) = 1.56, p > 0.4$.

It is understandable that during Interval II the subjects had longer fixation time to the automated subsystems when automation failed (Sessions 2 and 4). During Interval II of Sessions 3 and 5, as can be seen in Table 8.3, both subjects slightly increased the fixation time to the automated subsystems compared to Interval I even though the failure did occur. During Interval III, the fixation time decreased again. However, during Interval II of Session 1, both subjects did not show more attention to the automated steam control loop compared to Intervals I and III. The above observation shows that the subjects had been affected by the possible failure in Sessions 3 and 5.

It should be stressed that all of the above observations are highly speculative since we have only two subjects. A comprehensive conclusion will depend on more experiments.
Figure 8.8 Fixation time for monitoring flow loop automation before automation failure (Interval I) and after it recovered from its possible failure (Interval III). Note that there was no failure in Session 5.

As can be seen in Table 8.2, in Sessions 4 and 5, the subjects used more frequent samples, rather than longer fixations to monitor the automated flow loop that was perceived with a high criticality. This observation is in agreement with the observation achieved by Moray and Rotenberg (1989). They find that during fault management the operators have more frequent samples and not longer fixation times, on the failed subsystems. For our study, we can conclude that the operators took more samples, instead of longer fixations, to automated, highly error prone subsystems. This proofs the Hypothesis 4 that highly critical subtasks are monitored more frequently.

8.3.3.2 Monitoring behavior after automation recovered

Figure 8.7 and Figure 8.8 present the fixation times after the automation recovered from its failure (i.e. during Interval III) for Sessions 2 and 4, respectively. In Session 2 (Figure 8.7), both the subjects showed longer monitoring times to the automation after it had recovered from its failure than before the failure occurred (Subject 1 increased more than 4 times and Subject 2 increased more than 2 times in fixation time). The number of samples increased 2 times and 15 times, respectively. A detail analysis of the duration when these samples were done reveals that the monitoring took place within 120s directly after automation recovered. This shows that even though automation had recovered from its failure, it took a period for the operators to rebuild, or recover, the trust in the automation. This confirms Hypothesis 3, that the operator pays more attention to an automated subsystem that recovers from a failure than before the failure occurs. This finding is in agreement with the earlier studies conducted by Lee and Moray (1992) and Riley (1994).

As shown in Figure 8.8, the fixation times (average across two subjects) in monitoring the flow loop during Intervals I and III in Session 4 are not significantly different, $F(1,1) = 72.83,$
\( p > 0.05 \); and so are the numbers of samples, \( F(1,1) = 1.31, p > 0.1 \). This is, probably, because this highly critical subsystem is always monitored. No matter whether automation worked or failed, the subjects continuously kept their attention to that subsystem.

In Sessions 3 and 5, the automation failure did not occur. The total fixation times in monitoring the automated subsystems during Interval I and Interval III are not significantly different as shown in Table 8.3. For the numbers of samples, the similar conclusion can be drawn.

8.3.3.3 Remarks

The analysis of the operator monitoring behavior reveals the following remarks concerning the hypotheses:

The experimental data do not show that the operator gives more monitoring attention to the automated subsystems when the operator knows that automation might fail (Session 1 and 3). This may be because the automated subsystem in these sessions is not critical. We may assume that human trust in automation is affected by the nature of the subsystems that are automated, such as task criticality. We need more experiments to confirm this assumption.

The human monitoring behavior of the automated subsystems that recover from an automation failure depends on the monitoring behavior before the automation failure. The operator gives less attention to an automated subsystem with low criticality than to an automated subsystem with high criticality. An automated subsystem with low criticality is not frequently monitored when the automated system is working in normal. Thus, after the automated subsystem recovers from a failure the operators will pay more attention to the automated subsystem. For a highly critical subsystem, the operators have paid much attention to the automated system before the failure. When the automated subsystem recovers from the failure, the operators will pay the same amount of attention to the automated subsystem. So, the process that the operator rebuilds trust in automation is affected by the nature of the subsystems that are automated.

The finding that the operators still pay attention to monitoring the automated subsystems with low criticality for a certain period after the automated systems recover from a failure can be supported by an interesting finding achieved by Parasuraman et al. (1996). Parasuraman et al. investigated the effects of adaptive task allocation on the monitoring of automated systems, and studied the operator performance in detecting the automation failure. They found that the performance if the automated subsystem is reallocated to the operators is better than that without reallocation. They further found that the performance is also higher in the post-allocation phase when the task is returned to automated control. This means that the performance can be kept higher for a certain period after the task is performed by automation.
In their experiment, Parasuraman et al. do not measure how the operators monitor the automation, before, during and after the task is dynamically allocated to the operators. They only measure the monitoring performance (detection rate of automation failures). Why is the performance kept higher for a certain period after the task is returned to the automatic controller? This is probably because the operators need a certain period to rebuild the trust in automation. During this period, the operators still keep their monitoring on the automated subsystems. So, the detection performance can be kept higher for a certain time. However, the performance will diminish with time as less attention is paid by the operators due to the recovered trust in automation.

### 8.3.4 Mental Load due to the Monitoring of Automation

In Ch. 4, we have addressed the monitoring workload of a task. The monitoring workload is split up into two aspects: The mental load when the task is performed manually and the mental load when the task is performed automatically.

As discussed before, the mental load associated with the monitoring tasks cannot be directly measured. Because of the limited number of subjects in this study, the mental load data collected by the subjective assessment does not show a significant difference between different sessions. We use the mental load data collected in Ch. 6 as a reference to demonstrate how the mental load is related to the monitoring of the automated tasks.

Recalling the experiment conducted in Ch. 6 it reveals that the subtask flow control, has a higher task demand load and evokes a high mental load when only that subtask is manually performed. The steam control subtask demands less workload and evokes a low mental load. The experiment in the current study shows that the flow control subtask attracts significantly more attention when it is automated than the steam control subtask. Figure 8.9 shows the task demand load (the original values from Ch. 6), the overall mental load (the subsystem is manually controlled, see Ch. 6), the percentage of the fixation time of the monitoring in the total recorded fixation time (averaged across two subjects), and the monitoring frequency (averaged) on these subsystems when they are automated.

The fact that more attention was paid to the flow control subtask explains why in Ch. 6 the task demand load of flow control subtask, TDL$^{FC}$, is given a higher value than the task demand load of steam control loop, TDL$^{SC}$. However, the overall mental load when only the flow control is on manual mode, OML$^{FC}$, is not much higher than the overall mental load when only the steam control is on manual mode, OML$^{SC}$. This can be summarized as follows: Although there is a large difference between the task demand loads of two subtasks, i.e. TDL$^{FC} = 1.5$TDL$^{SC}$, the overall mental loads assessed by the operators for the two subtasks are not significantly different, F(1,8) = 7.44, p > 0.01.
Figure 8.9 Comparison of task demand load, overall mental load and the number of monitoring samples on automated subsystems.

Why is $OML^{SC}$ not significantly different from $OML^{FC}$?

In Chs 5 and 6 we have demonstrated that the overall mental load in operating a system can be predicted from the task mental load, TML, of individual tasks. So, we may assume: In the case that the flow control is on manual and no failure occurs:

$$OML^{FC} = TML^{SC} + TML_{m}^{SC} + TML_{m}^{DS},$$  \hspace{1cm} (8.1)

where $TML^{FC}$ represents the TML for controlling the flow control subsystem manually; $TML_{m}^{SC}$ stands for the TML for monitoring the automated steam control subsystem, and $TML_{m}^{DS}$ for monitoring the automated distribution subsystem.

In the case that the steam control is on manual:

$$OML^{SC} = TML^{SC} + TML_{m}^{FC} + TML_{m}^{DS},$$  \hspace{1cm} (8.2)

where $TML^{SC}$ represents the TML for controlling the steam control subsystem manually, and $TML_{m}^{FC}$ stands for the TML for monitoring the automated flow control subsystem.

We assume that automatic controllers for the flow control and the steam control do not affect $TML_{m}^{DS}$. Subtracting Eq. (8.1) from Eq. (8.2), we obtain:

$$TML_{m}^{FC} - TML_{m}^{SC} = (OML^{SC} - OML^{FC}) + (TML^{FC} - TML^{SC}).$$  \hspace{1cm} (8.3)

We assume that TML can be predicted by TDL. Since $TDL^{FC} > TDL^{SC}$, we have $TML^{FC} > TML^{SC}$. Considering $OML^{FC} = OML^{SC}$, we can conclude from Eq. (8.3):
\[ TML_{m}^{FC} > TML_{m}^{SC}. \] (8.4)

Eq. (8.4) indicates that the automated flow loop demands more monitoring workload from the operators than the automated steam loop. Hence, due to the high value for the \( TML_{m}^{FC} \), the OML\(^{SC} \) approaches the OML\(^{FC} \). Therefore, we conclude that the mental load due to monitoring the automated subtasks contributes to the overall mental load. This confirms Hypothesis 5, the monitoring of the automated subtasks contributes to the overall mental load.

It should be stressed that the above analysis is only valid if the task demand load of monitoring an automated task is estimated by the monitoring actions, i.e. by the monitoring fixation time and the number of monitoring samples. We have mentioned before that the operator pays more attention to the automation of subsystems with high criticality. So, the monitoring task demand load may be qualitatively and indirectly estimated by the nature of tasks to be monitored. That is to say: A task with a high task demand load as well as a high task criticality may also have a high monitoring task demand load no matter whether it is automated or not.

From the discussion above, questions may arise. Whether the monitoring task demand load can be estimated by task criticality or task complexity which includes such as the time request, the amount of information, and the frequency of the information change etc.? If so, whether the approaches to quantify the task complexity discussed in the previous chapters can be used to estimate indirectly the monitoring task demand load? Further research is needed to answer these questions.

### 8.4 SUMMARY

This chapter mainly concerns with the operator monitoring behavior of automated subsystems. Based on the analysis of the experimental data, we summarize the study as follows:

1. Operator performance is related to the operator monitoring behavior of the plant. A well-performing operator focuses not only on the manually controlled tasks, but also supervises thoroughly the automated subsystems.

2. Operators increase the number of monitoring samples, instead of the fixation time, of the automated, highly critical subsystems compared to the automated subsystems with a lower criticality.

3. The human trust in automation may be affected by the nature of the task that is automated. There is a time delay for the operators to rebuild the trust in the automated systems.
4. The monitoring of the automated tasks will contribute to the overall mental load. Although it is difficult to measure the mental load resulting from monitoring, a task with a high task demand load as well as a high task criticality may have a high monitoring task demand load no matter whether it is automated or not.

The experimental study described in this chapter is a preliminary study. It presents possible directions for the further research.
Chapter 9

Concluding Summary

The study included in this thesis is aimed at investigating the influence of various levels of automation in supervisory control on the performance of human operators. The study focuses on confirming the hypothetical relations between the mental load, the performance and the level of automation as shown in Figure 1.1. To start with the study, task allocation approaches between human and machine are extensively examined. Quantitative measures for the level of automation, degree of automation (DofA), are derived. By our definitions, a change of the allocation of tasks for whatever reason affects the level of automation. The validation and implementation of the DofA measures are done using two experimental systems: An artificial linear system and a simulated pasteurization plant. Human monitoring behavior of the automated subsystems of the pasteurization plant is investigated by measuring eye movements. In the following sections more specific conclusions from the present study are drawn and the implications of the study are discussed.

9.1 CONCLUSIONS FROM THE PRESENT STUDY

In Ch. 4, quantitative measures for the level of automation in human supervisory control are proposed. An important aspect for a quantitative measure of the degree of automation is that the measure should consider not only the number, but also the characteristics of the automated tasks.

The study that has been completed on the DofA measure in this thesis is a preliminary step to develop a measurement approach for the degree of automation during the design phase of supervisory control (e.g. DofATDL) or during the system’s evaluation (e.g. DofATML and DofATES).

The DofA measure was applied to a linear system and to a simulated process (Chs 5 and 6). The hypotheses (Figure 1.1) for the relation between performance and degree of automation, and the relation between mental load and degree of automation have been confirmed (as shown in Table 5.5, Figure 5.10, Table 6.7, and Figure 6.8). For the relation between performance and DofA, the following conclusions can be drawn:
• When the degree of automation increases, the system performance may level off. This suggests that the human operator can remain in the control loop without degrading the system performance.
• The relationship between performance and DofA can be expressed as a second order polynomial. The performance increases as the DofA increases within a certain range.

In the study on the relation between mental load and DofA, the following conclusions can be drawn:

• The relationship between mental load and DofA can be expressed as a first order polynomial. The mental load decreases as the DofA increases.
• When the operator controls a partly-automated system, the overall mental load that he/she perceives may be predicted from the mental load obtained from a specific situation, namely, the situation in which all tasks are manually performed. The mental load obtained from the specific situation includes the overall mental load for controlling the entire system and the mental load for each task component.
• Mental load may be predicted by task demand load as long as the task demand load is within the range that the human is willing to spend.

Based on the results obtained in Chs 5 and 6, the further effort has been directed to investigate human performance and mental load during a transition from a normal situation to an abnormal situation (Ch. 7). In the latter situation, a sudden loss of some automated subsystems occurs and the degree of automation drops largely. When the automatic control system for a subsystem fails, the human operator has to take over the control of the subsystem. The investigation is carried out on the basis of the linear system as well as of the pasteurization plant. From this investigation, the following conclusions can be made:

• A much higher mental load is perceived by the operators just after the degree of automation suddenly drops. The mental load level will decrease to a lower level after the system is operated at that low level of automation for a longer period. This low level of mental load is at the same level as may have been perceived by the operators if the system would have been operated at that low level of automation from the beginning.
• Within a short period after the degree of automation drops, the human performance degrades to a low level. After the system is operated for a longer period at that low level of automation, the performance will increase to a higher level. This level is the same for a system that would have been operated at that low level of automation from the beginning.

Human operator’s monitoring behavior of an automated task is affected by the nature of the task (Ch. 8). For human monitoring behavior of the automated tasks, the specific conclusions that have been drawn are:
The performance of operator supervisory control is related to the operator monitoring behavior. A well-performing operator focuses not only on the manual tasks, but also supervises the automated tasks.

The operator gives more frequent monitoring samples, rather than longer fixation time, on the automatic control systems of highly critical tasks, compared to the automatic systems of tasks with a low task criticality.

The monitoring of the automated, highly critical tasks will contribute more to the overall mental load compared to the monitoring of tasks with a low criticality.

Human trust in automation will be affected by the nature of the automated task.

9.2 IMPLICATIONS OF THIS STUDY

Human performance and mental load will directly affect the system performance and the operational safety which are the ultimate requirements of the system design. This is why the study focuses on the effect of different levels of automation on human/system performance and human mental load. Besides the level of automation, there exist other factors that can affect the performance and the mental load. From the present study, one may find that the nature of a task plays an important role. The nature of the task has to be taken into account when the task is allocated to humans and to automatic control systems (or machine, computer). It is also the case when the study on human trust in automation, human use of automation, human-centered automation, and adaptive automation is conducted.

The following two interesting questions have not been addressed in this study: (1) What level will the mental load (Point $C_m$ in Figure 3.1) reach, and what level will the performance (Point $C_p$ in Figure 7.1) reach when the degree of automation drops suddenly? (2) What else will affect the level of Points C in Figures 3.1 and 7.1? It is of no doubt that the system designers are interested in these questions. For instance, if the level of Point $C_m$ can be estimated as a function of the magnitude of the drop and the level of the DoFAs, the designers can take measures such that Point $C_m$ will fall below a certain high value to avoid overloading the operators in case something goes wrong (loss of automated subsystems). Although the present study did not stress the above questions, we do have the following observations (implicated in Figures 7.3, 7.6, and 7.9): (1) When the degree of automation drops suddenly, Point C depends on the mental load level (or performance level) before the drop happens; (2) Point C is also affected by the system state just before the drop in the degree of automation occurs.

The Analytic Hierarchy Process (AHP) can be used to evaluate the decisions on task allocation between human and machine (Ch. 6). The thesis demonstrates the simple use of the AHP for task allocation evaluation. In practice, subjectively optimum task allocation decisions can easily be evaluated using the AHP approach. The application of the AHP approach in the subjective
evaluation of task allocation is promising, especially when the number of options for pair-wise comparison is not too large, the number of subjects is small, and the criteria for the subjective evaluation can clearly be defined. In these situations, a high consistency of the subjective judgments can be achieved. However, it should be stressed that in human-machine systems design, the AHP ranking of task allocation options can only provide a subjective reference for the evaluation. At the same time, other aspects for the evaluation, such as mental load level, human/system performance, operational safety, costs and benefits, should be considered together.

In many situations, the mental load in monitoring the automated subtasks is often neglected due to the difficulties in the measurement. However, the study in this thesis (Ch. 8) shows that the human operator monitoring behaviors to the automated subtasks are affected by the nature of these subtasks. This implies that task demand load or mental load resulting from monitoring the automated subtasks can be estimated by the nature of the subtasks. When too many automated subtasks must be monitored by the operator, it is easy for the operator to be overloaded. It is necessary to take measures to reduce the operator monitoring workload due to the increasing level of automation. Moreover, the human operator monitoring behaviors of the automated subsystems, i.e. eye-tracking on the automated subsystems and interesting variables in the system, reflect the system model that is formed in the operator’s head and reflects human-system interactions during the operation. Thus, the human performance is related to the human monitoring behaviors. However, according to this study, the monitoring behaviors are affected by the human trust in automation, the nature of the automated subsystems, and the human understanding of the system. If an operator overtrusted an automated subsystem, he/she would pay less attention to the automated subsystem. When something is wrong with the subsystem, the operator cannot find the problem immediately. So, the operator’s performance will be reduced. It is essential that the operator has an adequate internal representation of the system and has developed a correct monitoring strategy so that the operator will not lose the awareness of the system states.

As has been addressed in Ch. 3, the estimation, or the prediction, of human mental load is an important issue in the human-machine systems design and many efforts have been made to explore approaches to measure the mental load. The present study reveals that the overall mental load for a partly automated system can be estimated from a specific situation as long as the mental load is within a limited range (Chs 5 and 6). Based on this finding, one can predict whether an operator will be overloaded when he/she controls a partly automated system, especially in the situations of dynamic task allocation and of automatic system failures. Moreover, the estimated overall mental load may be used as a trigger for reallocating tasks between human and machine.
9.3 FUTURE RESEARCH

The recommendations for further research based on this thesis are as follows.

As has been discussed in Ch. 2, dynamic task allocation, or adaptive automation, is becoming a popular topic in supervisory control. In adaptive automation, the following questions should be answered: (1) When the reallocation of tasks should take place, namely, under what condition the dynamic task allocation should be triggered? and (2) who, computer or human, will make decision for the reallocation. As an implementation, the quantitative DofA measure may be employed as a trigger to initiate the reallocation of some tasks between human and machine. When the DofA is lower than a certain value, some tasks should be reallocated to the automatic controllers. When the DofA is higher than expected, some tasks should be reallocated to human operators. In such a way, the reallocation decision may be made by human as well as by the computer. Thus, the implementation of the quantitative DofA measure in adaptive automation is an interesting topic for further research.

Operational safety associated with a high degree of automation has become more important. A possible application of the quantitative DofA measure is to investigate in what way a high level of automation affects the safety related to human errors. As has been addressed in Ch. 1, an increasing level of automation may cause many new problems for the human operator. Hence, it is necessary to investigate the relation between human errors and the degree of automation. A preliminary step to evaluate the safety with respect to human reliability has been proposed by Macwan, Wei and Wieringa (1994). This research is worth being continued.

In this thesis, the transition behavior of human performance and mental load due to a sudden loss of automated subsystems in process control has been investigated. In the study, the operators took over the control of some tasks from automatic controllers to perform the tasks manually. However, the operator’s role is now focused on supervisory control as well as on decision making. It is necessary to investigate transition behavior of human performance and mental load when operators intervene and override automation in performing decision making tasks.

In this study, an indirect analysis about the relationship between operator monitoring behavior of the automated subsystems and mental load has been performed. The study just reveals a possible relation between mental load, monitoring behavior, and task nature. It is far from having an applicable approach to estimate the monitoring mental load. Further research is needed to find a way to assess the monitoring mental load. In the implementation of the DofA measure, we did not take into account the mental load in monitoring the automated subsystems. Thus, the estimation of the monitoring mental load will, in turn, contribute to the DofA measure proposed in this thesis.
Lastly, it is recommended to carry out more research on human-centered automation in process industry. It is not because this topic has been fashionable recently. It is because human operators are still necessary for the process industry in foreseeable future. Although technology can completely operate a process plant, human operators are still necessary to supervise the operation, and to deal with the uncertainties and beyond the design situations. Human operators are essential back-up elements in case automated systems fail, and they can compensate for the lack of entire system robustness.
References


References


References


References


Appendix 1

The original Fitts List

Humans appear to surpass present-day machines with respect to the following:

1. Ability to detect small amounts of visual or acoustic energy.
2. Ability to perceive patterns of light or sound.
3. Ability to improvise and use flexible procedures.
4. Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time.
5. Ability to reason inductively.
6. Ability to exercise judgment.

Present-day machines appear to surpass humans with respect to the following:

1. Ability to respond quickly to control signals, and to apply great force smoothly and precisely.
2. Ability to perform repetitive, routine tasks.
3. Ability to store information briefly and then to erase it completely
4. Ability to reason deductively, including computational ability.
5. Ability to handle highly complex operations, i.e. to do many different things at once.

(From Fitts, 1951)
Appendix 2

RSME Rating Form Used in Chapter 5

Rating Scale Mental Effort (RSME)

Instructions: Choose a point on the scale that represents the magnitude of the mental effort in the task you just performed and write the number related to the point chosen to Table 1 (right).

The RSME measures an operator's mental effort invested when he/she performs a task. Mental effort may be defined as the total amount of controlled cognitive process in which an operator is engaged. Simply to say, mental effort is related to attention. The more attention a task demands, the more effort will be required. In the experiment, you will assess how much effort you have invested during task performance, i.e. how much cognitive costs and how hard you have to work to perform one or all tasks.

Table 1. Workload Assessment Record

Subject number:_____

Session number:_____

<table>
<thead>
<tr>
<th>Task (Cell) Number</th>
<th>RSME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Rating Scale Mental Effort (Zijlstra, 1993)

Note: For automated cells, you do not need to assess mental load.
Appendix 3

RSME Rating Form Used in Chapters 6 and 7

Rating Scale Mental Effort (RSME, Zijlstra, 1993)

De RSME meet de mentale belasting van een operator die hij of zij heeft in een bepaalde taak. Mentale belasting kan worden gedefinieerd als de totale hoeveelheid bewuste cognitive processen die een operator op een bepaald moment uitvoert.

Simpel gezegd: mentale belasting is gerelateerd aan aandacht. Hoe meer aandacht een taak vraagt, hoe meer inspanning de operator moet leveren.

Na de experimenten wordt gevraagd hoeveel inspanning je gedurende het uitvoeren van de taak ofwel hoeveel cognitieve capaciteit (beslissen, berekenen, kijken, e.d.) je hebt geïnvesteerd.

Instructions:

Kruis op de schaal aan op de plaats die correspondeert met de mentale inspanning zoals jij die ervaren hebt voor de taak die hieronder genoemd wordt. Vul ook je nummer en het sessienummer in. Als het een trainingssessie betreft het rondje hieronder zwartmaken.

<table>
<thead>
<tr>
<th>Operator nummer:</th>
<th>Sessie nummer:</th>
<th>O Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taak:</td>
<td></td>
<td>voor elk onderdeel een ander formulier invullen</td>
</tr>
<tr>
<td>O feedpump</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O overflow valve</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O steam pump</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O steam heater</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O distribution system</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O het gehele systeem</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

150 ---- 140 ---- 130 ---- 120 ---- 110 ---- 100 ---- 90 ---- 80 ---- 70 ---- 60 ---- 50 ---- 40 ---- 30 ---- 20 ---- 10 ---- 0

114 ontzettend inspannend
102 heel erg inspannend
86 erg inspannend
71 behoorlijk inspannend
57 tamelijk inspannend
37 enigszins inspannend
26 een beetje inspannend
13 nauwelijks inspannend
2 helemaal niet inspannend
Appendix 4

Analytic Hierarchy Process and Evaluation Results

A4.1 Analytic Hierarchy Process

The Analytic Hierarchy Process, AHP, was originally developed by Saaty (1980) as decision making aid, and it is a theory of measurement for dealing with quantifiable criteria. The AHP has found wide usage in a broad range of areas, including market prediction, project evaluation, as well as medical decision making etc. (Vargas, 1990). In comparing the AHP approach to the traditional psychophysical methods for generating measurement scales, the AHP has the following advantages: (1) It possesses the ability to readily quantify consistency in human judgments; (2) it has the ability to provide useful empirical results in the event of a small sample of subjects and when the likelihood of obtaining meaningful statistical results may be restricted; and (3) the AHP requires no statistical assumptions regarding the distribution of human judgments (Mitta, 1993).

In brief, the AHP is structured to encompass the basic elements of a decision. As shown in Figure 6.4, the objective of a problem is placed at the highest level, then moving to the subobjectives affecting the objective followed by criteria in the next level and so on, from more general to the more particular. For each level of the hierarchy, the AHP breaks up multiple options into a series of paired comparisons which are then recombined to produce an overall weighted ranking. The subjects are required to compare all possible combinations. The results of comparisons are placed in a pair-wise comparison matrix which is called judgment matrix (Figure 6.5). The rows and columns of the judgment matrix are headed by the options included in the comparison. The principal eigenvector for the matrix is calculated, which gives a weighted ranking scale for each option.

For each level of Figure 6.4, data in the AHP are collected using a series of comparisons between each pair of design options at this level. Saaty (1980) and Mitta (1993) suggest a format which presents the two options to be compared on opposite ends of a 17 slot rating scale. The scale is a measure of dominance of one alternative over the other. It uses five descriptions in a pre-defined order and allows a single space between each one for compromise.
Figure A4.1 Dominance scale used for pair-wise comparison of two task allocation schemes.

The descriptions are "equal", "weak", "strong", "very strong", and "absolute". These terms can be modified to provide easier comprehension for a particular subject group. Figure A4.1 shows such a scale used in the present study to compare task allocation schemes in two experimental sessions, \( E_i \) and \( E_j \) (\( i < j, i, j = 1,2, ..., 7 \)), based on the criterion of easy operation (Table 6.2).

The scale allows the subjects to indicate their judgments regarding the degree of dominance of one task allocation configuration over the other. The subjects indicate not only that one alternative dominates over a second, but also the degree by which it dominates. Given \( n \) options to be compared, each subject must make \( n(n-1)/2 \) comparisons. In the present experiment, as shown in Table 6.1, the subjects operated 7 sessions (task allocation configurations) so that for one criterion each subject made 21 comparisons which were carried out after the subjects completed all of the sessions.

The dominance measures from the pair-wise comparisons given by a subject are placed in a judgment matrix \( M \) of the following form:

\[
M = \begin{bmatrix}
E_1 & E_2 & \ldots & E_n \\
E_1 & 1 & m_{12} & \ldots & m_{1n} \\
E_2 & 1/m_{21} & 1 & \ldots & m_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
E_n & 1/m_{n1} & 1/m_{n2} & \ldots & 1
\end{bmatrix}
\]

(A4.1)

Each \( m_{ij} \) entry of \( M \) reflects the factor by which option \( E_i \) dominates option \( E_j \) in a nine-point scale as specified as follows:

\[
\begin{align*}
m_{ij} &= 1 \text{ if } E_i \text{ and } E_j \text{ are of equal strength;} \\
m_{ij} &= 3 \text{ if } E_i \text{ weakly dominates } E_j; \\
m_{ij} &= 5 \text{ if } E_i \text{ strongly dominates } E_j; \\
m_{ij} &= 7 \text{ if } E_i \text{ very strongly dominates } E_j; \\
m_{ij} &= 9 \text{ if } E_i \text{ absolutely dominates } E_j.
\end{align*}
\]
Scale values 2, 4, 6, 8 will reflect compromises between ratings of equal strength and weak dominance, weak dominance and strong dominance, strong dominance and very strong dominance, very strong dominance and absolutely dominance, respectively.

Matrix M has a reciprocal structure, i.e.:

- \( m_{ji} = 1/ m_{ij}, \) for \( m_{ij} \neq 0; \)
- \( m_{ij} = 1, \) for \( i = j \) and \( i, j = 1, 2, \ldots, n. \)

This reciprocal format results from the AHP axiom of ration scale: If \( E_i \) is \( x \) times more dominant than \( E_j, \) then, \( E_j \) is \( 1/x \) the dominance of \( E_i. \) Furthermore, if \( E_{j-1} \) is \( y \) times more dominant than \( E_i, \) then, \( E_{j-1} \) would be expected to be \( xy \) times more dominant than \( E_j. \) This reflects the consistency of the human judgment. If the human judgment is perfect in each comparison, there exists: \( m_{ik}=m_{ij}m_{jk} \) for all values of \( i, j, \) and \( k, \) and \( M \) is referred to as a consistent matrix. In fact, this is difficult to achieve. The principal eigenvalue of \( M \) is used to measure the consistency. A consistency index suggested by Saaty (1980) is:

\[
C.I. = \frac{\lambda_{\text{max}} - n}{n - 1}. 
\]

(A4.2)

where \( n \) is the number of options, and \( \lambda_{\text{max}} \) is the principal eigenvalue of \( M. \) This index indicates that if C.I. is less than 0.1, the judgments are considered consistent (for an explanation, see Mitta, 1993).

After all the pair-wise comparisons are completed, an overall ranking scale can be calculated. A frequently used approach to calculate the ranking scale is the eigenvector method. In this method, the ranking scale is produced by calculating the principal eigenvector of \( M. \) According to the matrix theory, the principal eigenvector of a matrix, \( w = [w_1 \ w_2 \ \ldots \ w_n]^T, \) corresponds to the largest positive eigenvalue of the matrix, i.e. \( \lambda_{\text{max}}, \) and is determined by solving the equation:

\[
Mw = \lambda_{\text{max}}w. 
\]

(A4.3)

In the AHP approach, \( w \) should be typically normalized such that its components sum up to 1:

\[
\sum_{i=1}^{n} w_i = 1 
\]

(A4.4)

The principal eigenvector \( w \) can be considered as a vector in the space of the \( n \) different options where the magnitude of the components in each direction is a measure of the strength of the respective option. The degree of dominance between two options is the ratio of their \( w \) components.
For Subject $k$, a judgment matrix $M^k$ can be obtained and so does a principal eigenvector $w^k$. Thus, a matrix $W$ containing the rankings of each individual subject can be constructed from the eigenvectors $w^k$ of all subjects. Supposing that there are $s$ subjects and $n$ options to be compared in the experiment, $W$ would be an $n \times s$ matrix:

$$W = (w^1, w^2, \ldots, w^s)$$  \hspace{1cm} (A4.5)

Moreover, the AHP allows for results of each subject to contribute equivalently or differently depending on their skill, experience, judgment ability, and so on. The experimenter can rate all of the subjects and give a ranking for each subject in the same way as the subjects rate an option by the AHP approach. However, according to Yang and Hansman (1995), in most cases all subjects should be given equal considerations in the final analysis; otherwise, the outcome could be easily be biased toward a certain result.

Suppose that the principal eigenvector which reflects the ranking of subjects is $s = [s_1, s_2, \ldots, s_s]^T$ and that all subjects are given the equal consideration, after normalizing, there is:

$$s = \left[ \frac{1}{s} \frac{1}{s} \ldots \frac{1}{s} \right]^T,$$  \hspace{1cm} (A4.6)

where $s$ is an $s$ vector and $s$ is the number of subjects who participate the evaluation.

The final overall ranking in which the subject contributions are taken into account is:

$$r = W \cdot s.$$  \hspace{1cm} (A4.7)

Vector $r$ should be normalized if it does not sum up to 1.0. Thus, the entries of $r = [r_1, r_2, \ldots, r_n]^T$ provide weights for the $n$ options as well as the relative differences between two options.

**Table A4.1** Conversion of Ratio Value to Qualitative Description
(adapted from Yang and Hansman, 1995)

<table>
<thead>
<tr>
<th>$r_i/r_j$</th>
<th>Dominance of Alternative $i$ over Alternative $j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>equal</td>
</tr>
<tr>
<td>3</td>
<td>weak dominance</td>
</tr>
<tr>
<td>5</td>
<td>strong dominance</td>
</tr>
<tr>
<td>7</td>
<td>very strong dominance</td>
</tr>
<tr>
<td>≥ 9</td>
<td>absolute dominance</td>
</tr>
</tbody>
</table>
The degree of dominance between two is represented by the ratio of their weights, \( r_i \) and \( r_j \). Based on this ratio, the dominance of one over another one can be converted into a qualitative description in the terms presented in Figure A4.1. Table A4.1 shows the conversion.

### A4.2 AHP Subjective Evaluation of Task Allocation Decisions

For *easy operation*, i.e. *Attribute 1* in Table 6.2, the easy level was compared. Matrix \( W_{A_1} \) shows each subject’s evaluation on how *easy* to operate the plant based on a task allocation configuration. A large value means an easier operation. Each column of \( W_{A_1} \) represents a subject, and each row represents an option, in an order: \( E_1, E_2, E_3, E_4, E_5, E_6, \) and \( E_7 \).

\[
W_{A_1} = \begin{bmatrix}
0.3944 & 0.1742 & 0.1434 & 0.3631 & 0.2150 & 0.2335 & 0.1422 \\
0.2434 & 0.1576 & 0.2207 & 0.2747 & 0.1134 & 0.1312 & 0.1252 \\
0.1796 & 0.4916 & 0.5084 & 0.2123 & 0.5116 & 0.4650 & 0.5441 \\
0.0937 & 0.0614 & 0.0443 & 0.0390 & 0.0538 & 0.0521 & 0.0523 \\
0.0480 & 0.0678 & 0.0443 & 0.0665 & 0.0306 & 0.0766 & 0.0900 \\
0.0236 & 0.0305 & 0.0262 & 0.0285 & 0.0594 & 0.0268 & 0.0231 \\
0.0173 & 0.0169 & 0.0126 & 0.0159 & 0.0161 & 0.0148 & 0.0231 \\
\end{bmatrix}
\]

When the ranking of subjects is taken into account, we substitute \( W_{A_1} \) and Eq. (A4.6) with \( s = 7 \) into Eq. A4.7 and yields the vector specifying the operational difficulty ranking for task allocation configurations which are represented by the experimental sessions as shown in Table 6.1.

\[
r_{A_1} = W_{A_1} \cdot s = [0.238 \quad 0.181 \quad 0.416 \quad 0.057 \quad 0.061 \quad 0.031 \quad 0.017]^T
\]

For *human control confidence*, i.e. *Attribute 2*, the subjects were asked to compare for which configuration they felt more in the control rather than automation. \( W_{A_2} \) shows subject’s rating on how *much* they felt that they were in the control. A large value indicates that the operator felt more control.

\[
W_{A_2} = \begin{bmatrix}
0.0487 & 0.0519 & 0.1429 & 0.0348 & 0.0452 & 0.0281 & 0.0551 \\
0.1119 & 0.0663 & 0.1429 & 0.0941 & 0.0481 & 0.0539 & 0.0551 \\
0.0388 & 0.0241 & 0.1429 & 0.0157 & 0.0222 & 0.0189 & 0.4958 \\
0.0867 & 0.1157 & 0.1429 & 0.0396 & 0.0734 & 0.1269 & 0.0985 \\
0.1606 & 0.1105 & 0.1429 & 0.1077 & 0.0967 & 0.1190 & 0.0985 \\
0.2736 & 0.2440 & 0.1429 & 0.3489 & 0.3398 & 0.2677 & 0.0985 \\
0.2798 & 0.3875 & 0.1429 & 0.3593 & 0.3747 & 0.3856 & 0.0985 \\
\end{bmatrix}
\]

After the rankings of all subjects being considered, the vector specifying the ranking of the human control confidence ranking for task allocation configurations is:
\[ \mathbf{r}_{A_2} = \mathbf{W}_{A_2} \cdot \mathbf{s} = \begin{bmatrix} 0.058 & 0.082 & 0.108 & 0.098 & 0.119 & 0.245 & 0.290 \end{bmatrix}^T \]

For *comfortable operation*, the subjects were asked to compare for which configuration they felt *more comfortable* in the operation of the plant. \( \mathbf{W}_{A_3} \) shows subject’s rating on how *comfortable* they felt. A large value indicates that the operator felt more comfortable.

\[
\mathbf{W}_{A_3} = \begin{bmatrix}
0.2885 & 0.3417 & 0.1748 & 0.0536 & 0.1988 & 0.4103 & 0.3568 \\
0.1184 & 0.2138 & 0.1748 & 0.1724 & 0.1224 & 0.2689 & 0.2387 \\
0.0217 & 0.0809 & 0.5041 & 0.0166 & 0.4427 & 0.0479 & 0.0190 \\
0.3044 & 0.1426 & 0.0546 & 0.0543 & 0.1217 & 0.1502 & 0.1123 \\
0.1251 & 0.1232 & 0.0546 & 0.1427 & 0.0356 & 0.0675 & 0.0871 \\
0.0729 & 0.0571 & 0.0247 & 0.2906 & 0.0644 & 0.0375 & 0.1098 \\
0.0691 & 0.0408 & 0.0124 & 0.2698 & 0.0145 & 0.0177 & 0.0762 
\end{bmatrix}
\]

The vector specifying the ranking of the effect of automation on human by combining all subject’s judgments is as follows:

\[ \mathbf{r}_{A_3} = \mathbf{W}_{A_3} \cdot \mathbf{s} = \begin{bmatrix} 0.261 & 0.187 & 0.162 & 0.134 & 0.091 & 0.094 & 0.072 \end{bmatrix}^T \]

For *situation awareness* rating, the subjects were asked to compare for which configuration they knew *more about* the plant status. \( \mathbf{W}_{A_4} \) shows each subject’s rating. A large value indicates that the operator knew more variables and had more situation awareness.

\[
\mathbf{W}_{A_4} = \begin{bmatrix}
0.1405 & 0.0753 & 0.0643 & 0.1040 & 0.0651 & 0.0988 & 0.0435 \\
0.0842 & 0.0683 & 0.0432 & 0.1336 & 0.0559 & 0.1179 & 0.0435 \\
0.0182 & 0.0245 & 0.0141 & 0.0186 & 0.0216 & 0.0235 & 0.0435 \\
0.0949 & 0.0953 & 0.0559 & 0.0562 & 0.0858 & 0.1621 & 0.0435 \\
0.1065 & 0.0769 & 0.0954 & 0.2084 & 0.0984 & 0.1212 & 0.0435 \\
0.2352 & 0.3248 & 0.3635 & 0.2396 & 0.2804 & 0.2009 & 0.3913 \\
0.3205 & 0.3349 & 0.3635 & 0.2396 & 0.3929 & 0.2756 & 0.3913 
\end{bmatrix}
\]

The vector specifying the ranking of the situation awareness from a combination of all subject’s judgments is:

\[ \mathbf{r}_{A_4} = \mathbf{W}_{A_4} \cdot \mathbf{s} = \begin{bmatrix} 0.085 & 0.078 & 0.023 & 0.085 & 0.107 & 0.291 & 0.331 \end{bmatrix}^T \]

Base on \( \mathbf{r}_{A_1}, \mathbf{r}_{A_2}, \mathbf{r}_{A_3}, \) and \( \mathbf{r}_{A_4} \), the subjective ranking for different task allocation configurations can be discussed. After a rank for the four attributes is obtained, the synthesis of the subjective evaluation can be performed. The results have been presented in Ch. 6.
# Appendix 5

## A Sample of Pair-Wise Comparison Tables

**Easy Operation:** Which system is easier to control? (PAIR-WISE COMPARISON: E₁/Eᵢ)

Operator number: 

<table>
<thead>
<tr>
<th>E₁</th>
<th>E₂</th>
<th>E₃</th>
<th>E₄</th>
<th>E₅</th>
<th>E₆</th>
<th>E₇</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absolutely easier</td>
<td>much easier</td>
<td>easier</td>
<td>slightly easier</td>
<td>same</td>
<td>slightly easier</td>
<td>easier</td>
</tr>
<tr>
<td>E₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E₁</th>
<th>E₂</th>
<th>E₃</th>
<th>E₄</th>
<th>E₅</th>
<th>E₆</th>
<th>E₇</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absolutely easier</td>
<td>much easier</td>
<td>easier</td>
<td>slightly easier</td>
<td>same</td>
<td>slightly easier</td>
<td>easier</td>
</tr>
<tr>
<td>E₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E₁</th>
<th>E₂</th>
<th>E₃</th>
<th>E₄</th>
<th>E₅</th>
<th>E₆</th>
<th>E₇</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absolutely easier</td>
<td>much easier</td>
<td>easier</td>
<td>slightly easier</td>
<td>same</td>
<td>slightly easier</td>
<td>easier</td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E₄</th>
<th>E₅</th>
<th>E₆</th>
<th>E₇</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absolutely easier</td>
<td>much easier</td>
<td>easier</td>
<td>slightly easier</td>
</tr>
<tr>
<td>E₄</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₅</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₆</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₅</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₆</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>E₅</th>
<th>E₆</th>
<th>E₇</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>absolutely easier</td>
<td>much easier</td>
<td>easier</td>
</tr>
<tr>
<td>E₆</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₆</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₇</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₇</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E₇</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Appendix 6  Experimental Results (Chapter 7)

Table A6.1  Overall mental load and system performance for the linear system ( homosex: An automated cell; □: A manually controlled cell)

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Interval Measured (min)</th>
<th>Task Allocation Configuration</th>
<th>DofA (TDL based)</th>
<th>OML</th>
<th>SPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-6 (Interval I)</td>
<td></td>
<td>0.85</td>
<td>35.50</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>6-12 (Interval II)</td>
<td></td>
<td>0.85</td>
<td>35.83</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>12-18 (Interval III)</td>
<td></td>
<td>0.85</td>
<td>38.33</td>
<td>0.75</td>
</tr>
<tr>
<td>2</td>
<td>0-6</td>
<td></td>
<td>0.17</td>
<td>73.00</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td></td>
<td>0.17</td>
<td>79.67</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td></td>
<td>0.17</td>
<td>82.50</td>
<td>1.85</td>
</tr>
<tr>
<td>3</td>
<td>0-6</td>
<td></td>
<td>0.85</td>
<td>29.67</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td></td>
<td>0.85-0.17</td>
<td>63.83</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td></td>
<td>0.17</td>
<td>78.33</td>
<td>1.49</td>
</tr>
<tr>
<td>4</td>
<td>0-6</td>
<td></td>
<td>0.67</td>
<td>69.50</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td></td>
<td>0.67-0.16</td>
<td>88.83</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td></td>
<td>0.16</td>
<td>91.33</td>
<td>1.86</td>
</tr>
<tr>
<td>5</td>
<td>0-6</td>
<td></td>
<td>0.53</td>
<td>71.33</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td></td>
<td>0.53-0.00</td>
<td>85.17</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td></td>
<td>0.00</td>
<td>98.83</td>
<td>2.46</td>
</tr>
<tr>
<td>6</td>
<td>0-6</td>
<td></td>
<td>0.70</td>
<td>48.50</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td></td>
<td>0.70-0.28</td>
<td>76.17</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td></td>
<td>0.28</td>
<td>83.33</td>
<td>1.53</td>
</tr>
</tbody>
</table>
Table A6.2 Overall mental load, system performance, and operator performance for the pasteurization plant

<table>
<thead>
<tr>
<th>Session Number</th>
<th>Interval Measured (min)</th>
<th>DofA (TDL based)</th>
<th>OML</th>
<th>SPF</th>
<th>OPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-6 (Interval I)</td>
<td>0.48</td>
<td>40.80</td>
<td>14.45</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>6-12 (Interval II)</td>
<td>0.48</td>
<td>46.20</td>
<td>11.65</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>12-18 (Interval III)</td>
<td>0.48</td>
<td>45.60</td>
<td>11.62</td>
<td>0.92</td>
</tr>
<tr>
<td>2</td>
<td>0-6</td>
<td>0.00</td>
<td>71.20</td>
<td>23.06</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.00</td>
<td>77.00</td>
<td>20.32</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.00</td>
<td>78.80</td>
<td>21.71</td>
<td>0.88</td>
</tr>
<tr>
<td>3</td>
<td>0-6</td>
<td>0.48</td>
<td>46.71</td>
<td>14.82</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.00</td>
<td>65.29</td>
<td>22.09</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.00</td>
<td>75.29</td>
<td>21.51</td>
<td>0.87</td>
</tr>
<tr>
<td>4</td>
<td>0-6</td>
<td>0.32</td>
<td>38.33</td>
<td>16.63</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.00</td>
<td>77.50</td>
<td>20.58</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.00</td>
<td>80.33</td>
<td>22.04</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>0-6</td>
<td>0.68</td>
<td>20.43</td>
<td>10.89</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.20</td>
<td>44.71</td>
<td>20.31</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.20</td>
<td>49.57</td>
<td>18.65</td>
<td>0.92</td>
</tr>
<tr>
<td>6</td>
<td>0-6</td>
<td>0.52</td>
<td>26.29</td>
<td>17.60</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.20</td>
<td>56.29</td>
<td>17.43</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.20</td>
<td>56.14</td>
<td>18.74</td>
<td>0.91</td>
</tr>
<tr>
<td>7</td>
<td>0-6</td>
<td>0.80</td>
<td>12.14</td>
<td>8.41</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.20</td>
<td>42.29</td>
<td>28.54</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.20</td>
<td>57.57</td>
<td>22.33</td>
<td>0.88</td>
</tr>
<tr>
<td>8</td>
<td>0-6</td>
<td>0.60</td>
<td>17.86</td>
<td>12.18</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.00</td>
<td>70.86</td>
<td>26.34</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.00</td>
<td>82.7</td>
<td>26.6</td>
<td>0.84</td>
</tr>
<tr>
<td>9</td>
<td>0-6</td>
<td>0.68</td>
<td>27.17</td>
<td>18.47</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.16</td>
<td>67.33</td>
<td>18.18</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.16</td>
<td>78.8</td>
<td>20.9</td>
<td>0.86</td>
</tr>
<tr>
<td>10</td>
<td>0-6</td>
<td>0.52</td>
<td>24.71</td>
<td>11.26</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>6-12</td>
<td>0.04</td>
<td>69.71</td>
<td>22.55</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>12-18</td>
<td>0.04</td>
<td>77.29</td>
<td>24.99</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Appendix 7

Experiment Results (Chapter 8)

Table A7.1 Fixation time, $F_t$, in monitoring subsystems and interesting variables, and the number of samples, $N_s$, in human monitoring behavior of automated systems

<table>
<thead>
<tr>
<th>Monitored Subsystems or Variables</th>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$F_t$</td>
<td>$N_s$</td>
<td>$F_t$</td>
<td>$N_s$</td>
<td>$F_t$</td>
</tr>
<tr>
<td>input tank</td>
<td>1</td>
<td>204.6</td>
<td>431</td>
<td>160.5</td>
<td>385</td>
<td>221.2</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>162.9</td>
<td>264</td>
<td>99.9</td>
<td>165</td>
<td>125.4</td>
</tr>
<tr>
<td>flow loop</td>
<td>1</td>
<td>297.3</td>
<td>530</td>
<td>253.7</td>
<td>494</td>
<td>276.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>392.9</td>
<td>695</td>
<td>375.0</td>
<td>645</td>
<td>40.5</td>
</tr>
<tr>
<td>steam loop</td>
<td>1</td>
<td>9.1</td>
<td>37</td>
<td>65.6</td>
<td>106</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4.5</td>
<td>20</td>
<td>102.6</td>
<td>133</td>
<td>7.7</td>
</tr>
<tr>
<td>temperature curve</td>
<td>1</td>
<td>92.3</td>
<td>276</td>
<td>91.6</td>
<td>298</td>
<td>100.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>3.6</td>
<td>12</td>
<td>11.8</td>
<td>40</td>
<td>18.8</td>
</tr>
<tr>
<td>distribution tanks</td>
<td>1</td>
<td>379.2</td>
<td>511</td>
<td>372.6</td>
<td>536</td>
<td>333.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>291.4</td>
<td>452</td>
<td>268.0</td>
<td>432</td>
<td>303.8</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>982.5</td>
<td>1354</td>
<td>943.93</td>
<td>1819</td>
<td>933.7</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>855.2</td>
<td>1443</td>
<td>857.4</td>
<td>1415</td>
<td>858.2</td>
</tr>
</tbody>
</table>
# List of Abbreviations and Symbols

## 1. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>Automatic Controller</td>
</tr>
<tr>
<td>AHP</td>
<td>Analytical Hierarchy Process</td>
</tr>
<tr>
<td>ANOVA</td>
<td>ANalysis Of VAriance</td>
</tr>
<tr>
<td>CH</td>
<td>Cooper-Harper scale</td>
</tr>
<tr>
<td>C.I.</td>
<td>Consistency Index</td>
</tr>
<tr>
<td>DoF</td>
<td>Degree of Automation</td>
</tr>
<tr>
<td>DSC</td>
<td>Distribution Switch Control</td>
</tr>
<tr>
<td>FPC</td>
<td>Feedstock Pump Control</td>
</tr>
<tr>
<td>HO</td>
<td>Human Operator</td>
</tr>
<tr>
<td>HOS</td>
<td>Human Operator Simulation</td>
</tr>
<tr>
<td>HTA</td>
<td>Hierarchical Task Analysis</td>
</tr>
<tr>
<td>HUMIF</td>
<td>Human Machine InterFace</td>
</tr>
<tr>
<td>KBB</td>
<td>Knowledge-Based Behavior</td>
</tr>
<tr>
<td>MABA-MABA</td>
<td>Men Are Better At ... Machines Are Better At</td>
</tr>
<tr>
<td>MCH</td>
<td>Modified Cooper-Harper scale</td>
</tr>
<tr>
<td>NASA-TLX</td>
<td>NASA - Task Load indEcx</td>
</tr>
<tr>
<td>NPP</td>
<td>Nuclear Power Plant</td>
</tr>
<tr>
<td>OML</td>
<td>Overall Mental Load</td>
</tr>
<tr>
<td>OPF</td>
<td>Operator PerFormance</td>
</tr>
<tr>
<td>OTDL</td>
<td>Overall Task Demand Load</td>
</tr>
<tr>
<td>OVC</td>
<td>Overflow Valve Control</td>
</tr>
<tr>
<td>RBB</td>
<td>Rule-Based Behavior</td>
</tr>
<tr>
<td>RSME</td>
<td>Rating Scale Mental Effort</td>
</tr>
<tr>
<td>SAINT</td>
<td>System Analysis of Integrated Networks of Tasks</td>
</tr>
<tr>
<td>SBB</td>
<td>Skill-Based Behavior</td>
</tr>
<tr>
<td>SCADA</td>
<td>Supervisory Control And Data Acquisition</td>
</tr>
</tbody>
</table>
2. Symbols

$A_i$ attribute for subjective evaluation

$DofA_{TDL}$ degree of automation based on task demand load

$DofA_{TES}$ degree of automation based on task effect on system performance

$DofA_{TML}$ degree of automation based on task mental load

$df_1, df_2$ degree of freedom of the hypothetical data sets and the error

$d_i$ difference between reference performance and TES performance

$e_i$ error between set-point and output of a cell

$err_i$ absolute value for the difference between set-point value and inputs for cell $i$

$F$ ration between mean square of hypothetical data sets and mean square of error

$F_m$ monitoring frequency

$F_{recycle}$ flow rate of the recycle juice

$F_{waste}$ flow rate of the wasted juice to the waste tank

$K$ connection matrix

$K_i$ integral coefficient

$k_{ij}$ coupling coefficient between two cells

$K_p$ proportional coefficient

$M$ total number of sampling periods during an experimental session
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ME_i$</td>
<td>absolute difference between actual output and set-point for cell $i$</td>
</tr>
<tr>
<td>$N$</td>
<td>total number of cells</td>
</tr>
<tr>
<td>$ND_{AM}$</td>
<td>number of data points without an alarm on distribution tanks</td>
</tr>
<tr>
<td>$OML_{CL}$</td>
<td>calculated overall mental load</td>
</tr>
<tr>
<td>$OML_{FM}$</td>
<td>overall mental load for a fully manually controlled system</td>
</tr>
<tr>
<td>$OML^K$</td>
<td>overall mental load during $K^{th}$ time interval</td>
</tr>
<tr>
<td>$oml_i^K$</td>
<td>overall mental load during the $i^{th}$ phase of $K^{th}$ time interval</td>
</tr>
<tr>
<td>$OML_{PA}$</td>
<td>overall mental load for a partly-automated system</td>
</tr>
<tr>
<td>$OPF_{i}^{K}$</td>
<td>operator performance during the $i^{th}$ phase of $K^{th}$ time interval</td>
</tr>
<tr>
<td>$p$</td>
<td>significant level</td>
</tr>
<tr>
<td>$P_f$</td>
<td>percentage of fixation time</td>
</tr>
<tr>
<td>$R^2$</td>
<td>regression coefficient</td>
</tr>
<tr>
<td>$r_{A_i}$</td>
<td>ranking vector</td>
</tr>
<tr>
<td>$r_F$</td>
<td>final ranking vector</td>
</tr>
<tr>
<td>$SD_{Tid}$</td>
<td>standard deviation of the pasteurization temperature</td>
</tr>
<tr>
<td>$SD_{IT}$</td>
<td>standard deviation of the input tank level</td>
</tr>
<tr>
<td>$SE$</td>
<td>system error</td>
</tr>
<tr>
<td>$SPF_{CL}$</td>
<td>calculated system performance</td>
</tr>
<tr>
<td>$SPF^K_i$</td>
<td>system performance during the $i^{th}$ phase of $K^{th}$ time interval</td>
</tr>
<tr>
<td>$T_a$</td>
<td>time available</td>
</tr>
<tr>
<td>$T_c$</td>
<td>juice temperature after secondary heater</td>
</tr>
<tr>
<td>$T_d$</td>
<td>juice temperature after pasteurization</td>
</tr>
<tr>
<td>$TDL_i$</td>
<td>task demand load from task $i$</td>
</tr>
<tr>
<td>$TDL_{COMP}^i$</td>
<td>task demand load for compensatory actions</td>
</tr>
<tr>
<td>$TDL^{SP}_i$</td>
<td>task demand load for set-point following actions</td>
</tr>
<tr>
<td>$t_i$</td>
<td>task allocation indicator, 1 if task $i$ is automated and 0 if task $i$ is manual</td>
</tr>
<tr>
<td>$T_{in}$</td>
<td>inflow temperature</td>
</tr>
<tr>
<td>$T_{r}$</td>
<td>time Required</td>
</tr>
<tr>
<td>$TML_{i}^{FM}$</td>
<td>task mental load for task $i$ when the system is fully manually controlled</td>
</tr>
<tr>
<td>$T_s$</td>
<td>sampling period</td>
</tr>
<tr>
<td>$\tau$</td>
<td>time constant</td>
</tr>
<tr>
<td>$u$</td>
<td>control input vector</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(W_{satc})</td>
<td>juice volume in the waste tank</td>
</tr>
<tr>
<td>(w_i)</td>
<td>task weight for task (i)</td>
</tr>
<tr>
<td>(w_i^c)</td>
<td>control task weight for a manually controlled task</td>
</tr>
<tr>
<td>(w_{ma}^m)</td>
<td>monitoring task weight in the case that the control task is automated</td>
</tr>
<tr>
<td>(w_{mm}^m)</td>
<td>monitoring task weight in the case that the control task is manual</td>
</tr>
<tr>
<td>(w_{TDL}^i)</td>
<td>task weight based on task demand load</td>
</tr>
<tr>
<td>(w_{TES}^i)</td>
<td>task weight based on task effect on system performance</td>
</tr>
<tr>
<td>(w_{TM}^i)</td>
<td>task weight based on task mental load</td>
</tr>
</tbody>
</table>
Summary

Mental Load and Performance at Different Automation Levels

Zhi-Gang WEI

The research, presented in this thesis, concentrates on the influence of various levels of automation on the performance and mental load of human operators.

The level of automation in human-machine systems increases continuously due to technological developments and for financial and economic reasons. The tasks that human operators perform have been changed to monitoring and managing the process, and controlling the product quality. Thus, the operator’s daily task is not directly to operate the process. The operator risks to lose his operational skill and his involvement in operating the process. The question can be asked whether the operator is capable to adequately solve the problems in production and whether he can make the process stable after it is disturbed. It is possible to reduce the level of automation and to increase the operator involvement in the operation so that the operator can keep his operational skill and can stay aware of the process state.

To start with the study, the thesis examines extensively task allocation approaches, the measures for evaluating task allocation decisions, and the qualitative scales for the level of automation. Then, a quantitative measure for the level of automation, degree of automation (DofA), is derived. The DofA has been defined as a function of the number and the nature of tasks to be automated. It is assumed that if part of the automated system fails the DofA reduces. The study is a preliminary step to develop a measurement for the degree of automation in the design of operator supervisory tasks or during system evaluation.

The validation and implementation of the DofA measure has been carried out based on two experimental systems: an artificial linear system and a simulated pasteurization plant. Between performance, mental load and the DofA, the following relations are obtained:
- The relationship between performance and DofA can be expressed as a second order polynomial.
- The performance increases as the DofA increases, but levels off at large DofA values.
- The relationship between mental load and DofA can be expressed as a first order polynomial. The mental load decreases as the DofA increases.
- When the operator controls a partly-automated system the overall mental load that he/she perceives may be predicted from the mental load measurement obtained for each task in a situation where all tasks are manually performed.

A mathematical tool, Analytic Hierarchy Process, was implemented to subjectively evaluate the decisions on task allocation. The implementation indicated that this tool can be easily used under certain conditions. The evaluation results should be used together with the results obtaining from other evaluation approaches.

Human performance and mental load during a transition from a normal situation to an abnormal situation was investigated. An abnormal situation was simulated by a sudden loss of some automated subsystems which causes the degree of automation to drop largely. The investigation reveals:

- A much higher mental load is perceived by the operators directly after the degree of automation drops. The mental load level will decrease to a lower level after the system is operated at that low level of automation for a longer period. The low mental load level is the same for a system that would have been operated at that low level of automation from the beginning.

- The human performance degrades to a low level just after the degree of automation drops. After the system is operated for a longer period at that low level of automation, the performance will increase to a higher level. This high level performance is the same for the system that would have been operated at that low level of automation from the beginning.

Finally, operator monitoring behavior of an automated task was studied using eye-tracking measurements. From this study, the following is concluded:

- A well-performing operator focuses not only on the manually performed tasks, but also supervises the automated tasks.

- The operator takes more frequent visual samples, rather than longer fixation time, on the automatic control systems of highly critical tasks, compared to the automatic systems of tasks with a low task criticality.

- The monitoring of automated, highly critical tasks will contribute more to the overall mental load than monitoring tasks with a low criticality.
Samenvatting (Summary in Dutch)

Mentale Belasting en Operator Prestaties bij Verschillende Graden van Automatisering

Zhi-Gang WEI

Het onderzoek, gepresenteerd in dit proefschrift, concentreert zich op de invloed van de mate van automatisering op de prestatie en de mentale belasting van operators.

De automatiseringsgraad van mens-machine systemen neemt continu toe als gevolg van technologische ontwikkelingen en financiële en economische argumenten. De taken die aan de operators worden toegewezen verschuiven naar het zogenaamde monitoren en het administreren van het proces en de kwaliteit van het produkt. Doordat de operator bij veel van dergelijke taken niet meer direct met de procesbesturing bezig ontstaat de kans dat minder ervaring wordt opgebouwd en de betrokkenheid afneemt. De vraag wordt dan ook gesteld of de operator nog wel in staat geacht mag worden problemen in de produktie adequaat te verhelpen of het proces te stabiliseren na een verstoring. Wellicht zou door een minder ver doorgevoerde graad van automatisering de betrokkenheid met het proces verhoogd kunnen worden waardoor de operator beter in staat geacht mag worden het proces ook in minder vaak voorkomende situaties meester te blijven.

In dit proefschrift wordt allereerst ingegaan op taakallocatie technieken en verschillende beschrijvingen van het niveau van automatisering. Daarna wordt de graad van automatisering (Degree of automation=DofA) gedefinieerd. Daarbij wordt uitgegaan van het idee dat het uitvallen van een deel van de automatisering inhoudt dat de graad van automatisering afneemt. De graad van automatisering is gedefinieerd als een functie van het aantal taken en de aard van de taken. Deze studie is een eerste stap om te komen tot een maat voor de graad van automatisering die gebruikt kan worden in het ontwerp van operator supervisie taken of voor evaluatie daarvan.

De validatie en implementatie van de DofA maat is gedaan met behulp van twee experimentele systemen: een artificieel, lineair systeem en een gesimuleerd pasteurisatie proces. Tussen prestatie, mentale belasting en DofA werden de volgende relaties gevonden:
Samenvatting

- De relatie tussen prestatie en DofA kan worden beschreven door een tweede orde polynoom.
- De prestatie neemt toe als the DofA toeneemt maar vlak: af bij grote waarden voor de DofA.
- De relatie tussen de mentale belasting en de DofA kan worden beschreven door een eerste orde polynoom. De mentale belasting neemt af bij toenemende waarden voor de DofA.
- Ingeval de operator een gedeeltelijk geautomatiseerd proces bestuurt kan de door de operator ervaren mentale belasting voor het gehele systeem worden voorspeld op basis van mentale belasting metingen van deeltaken in een volledig handmatig bestuurd systeem.

Een mathematische methode, het Analytisch Hiërarchisch Proces, is gebruikt om subjectief de taak allocatie te evalueren. De implementatie liet zien dat deze methode onder bepaalde omstandigheden gemakkelijk gebruikt kan worden. De resultaten zijn vergeleken met andere methoden.

De menselijke prestaties en mentale belasting gedurende een transitie van een normale situatie naar een abnormale situatie is ook onderzocht. Een abnormale situatie werd gesimuleerd door een plotseling verlies van een automatisch systeem wat de graad van automatisering plotseling deed afnemen. De resultaten lieten zien:
- Een veel hogere mentale belasting werd ervaren door de operators direct na de afname van de graad van automatisering. De mentale belasting neemt af tot een lagere waarde nadat het systeem voor een langere tijd bij die lagere automatiseringgraad is bewaakt. Het lagere niveau van mentale belasting komt overeen met het niveau dat zou zijn bereikt als het systeem van het begin af aan op dat niveau zou zijn bewaakt.
- De operator prestaties nemen af tot een laag niveau nadat de graad van automatisering is afgenomen. De prestaties nemen weer toe nadat het systeem voor langere tijd bij de lagere graad van automatisering is bewaakt. Dit hogere niveau komt overeen met het niveau voor de operator prestaties in het geval dat het systeem van het begin af aan bij die automatiseringgraad zou zijn bewaakt.

Tenslotte is het operator monitorendag van een geautomatiseerde taak bestudeerd door gebruik te maken van oogbewegingsmetingen. Uit dit onderzoek zijn de volgende conclusies getrokken:
- Een goed presterende operator focuseert niet alleen op de handmatige taken maar superviseert ook de geautomatiseerde taken,
- De operator neemt vaker visuele samples in plaats van langere fixatie tijden indien de geautomatiseerde taken kritische taken zijn vergeleken met geautomatiseerde taken met een lage kritische waarde,
- Het monitoren van geautomatiseerde, hoog kritische taken draagt meer tot de mentale belasting bij dan van laag kritische taken.
Curriculum Vitae

Zhi-Gang Wei was born on October 6, 1958 in Hubei Province of China. From 1975 to 1977, he was assigned to a forest plant to work as an electrician. From 1978 to 1982, he studied in the Department of Marine Engineering, Huazhong University of Science and Technology, China. In 1982, he received his Degree of Bachelor of Engineering in Ship Automation from Huazhong University of Science and Technology. From 1982 to 1990, he was an assistant lecturer, and then a lecturer in the Department of Computer and Automation, Wuhan Transportation University, China, where he completed his postgraduate courses in Electrical Engineering.

From November 1990 to March 1993, he was a visiting scholar and then a research fellow at Process Dynamics and Control Group of Kramers Laboratory, Faculty of Applied Physics, Delft University of Technology, The Netherlands.

From April 1993 to 1997, he worked as a research assistant towards a Ph.D. degree at Man-Machine Systems and Control Group, Laboratory for Measurement and Control, Faculty of Mechanical Engineering and Marine Technology, Delft University of Technology. The work presented in this thesis was performed during this period.