ACE project

Adaptive Cockpit Environment

Technical Report

Workload assessment in the cockpit

Patrick Ehlert
Leon Rothkrantz

Knowledge Based Systems group
Delft University of Technology
The Netherlands

WP 7 Report

Reference: ACE/NLR/WP7

Internal reference: DKS03-02 / ACE 03

Date: March 2003

The ACE Consortium

NLR (Amsterdam, The Netherlands), Military Academy (Brno, Czech Republic), Delft University of Technology (Delft, The Netherlands)

For the Dutch Ministry of Defence

DOCUMENT IDENTIFICATION

Results of the Literature Study		**ACE/NLR/WP6 & 10**	
written by:		reviewed by:	
Name	Organisation	Name	Organisation
P.A.M. Ehlert and L.J.M. Rothkrantz	Delft University of Technology	L.J.M. Rothkrantz	Delft University of Technology

DOCUMENT CHANGES

Issue date	**Version**	**Comments**
March 2003	1.0	

STATUS OF THE DOCUMENT

Internal	
Restricted	
Public	X

WORKGROUP
KNOWLEDGE BASED SYSTEMS
DELFT UNIVERSITY OF TECHNOLOGY

Ir. P.A.M. Ehlert
Q.M. Mouthaan
B. van der Poel
Drs. Dr. L.J.M. Rothkrantz

Preface

This report is part of the Adaptive Cockpit Environment (ACE) project, which is a joint project of the Military Academy of Brno in the Czech Republic, the National Aerospace Laboratory in the Netherlands and the Knowledge Based Systems group of Delft University of Technology in The Netherlands. The goal of the ACE project is "*the definition and evaluation of a prototype adaptable interface technique to identify the specific automation requirements and practical utility of this innovation in a military cockpit*" [NLR 2000]. The idea is that certain features in the pilot-plane interface can be adapted or automated depending on the workload and status of the fighter pilot. Physiological measurements are taken to determine this workload and information from the aircraft system is used to determine the current situation. The high-level information supplied by the workload and situation assessment modules is then used by a logical component to determine the content, format and modality of the display in the cockpit. This way the survivability of the pilot and plane, and the effectiveness of the flight or mission will be improved. In this report we will discuss methods for investigating and using the workload of a pilot within the proposed ACE system.

Summary

To help a pilot deal with information processing and decision-making, avoid information overload, and optimize flight performance, a crew-assistant system or intelligent pilot-vehicle interface (PVI) has been proposed. To correctly assess the amount of information that a pilot can handle we need to know his (mental) workload. Therefore, we need to design a workload assessment module as part of the PVI system.

We have devised a model to estimate the workload of a pilot in advance. The total workload of a pilot consists of his physical and mental workload. Manual workload can be estimated by counting the number of operations that the pilot has to perform and multiplying each operation with a subjective complexity factor. The estimated mental workload of a pilot will be a function of time and performed tasks. It will depend on the number of tasks performed and the difficulty of the tasks. There is a certain baseline that will slowly rise as the pilot becomes more fatigued. Also there is a certain limit above which the pilot's performance will degrade. In addition, a scheduling and coordination task is always present with a task load depending on the number of concurrent tasks. After finishing a task the workload of that task will slowly degrade (relaxation time).

From literature we know that physiological measures such as heart rate, hart rate variability, perspiration, and blink rate can be used as indicators of mental workload. We have designed a system that can use certain physiological measures to assess workload in real-time. However, most physiological measures can be indicators of physical work or stress as well as workload, making the system somewhat unreliable. For this real-time workload assessment method we propose to gather data with a gaze tracker. Our proposed workload assessment method uses a combination of number of fixations, fixation time, and blink rate. In addition, we record the physical actions done by the pilot, such as pressed buttons and stick and throttle movements.

We propose to "fuzzify" the measures of both workload assessment models. This makes a comparison of results between the two workload assessment methods possible. Experiments need to be performed in order to assess our two models.

Table of Contents

Chapter 1: Theoretical background

In this chapter we will address briefly the idea behind an intelligent pilot-vehicle interface and the role of a workload assessment module in such a system (section 1.1). Then we will explain relevant terms such as situation awareness (section 1.2) and mental and physical workload (section 1.3). In section 1.4 we will talk about human information processing and its influence on workload. We conclude with section 1.5 where we will provide a short overview of existing cognitive task models that can be used as a starting point for designing a workload assessment module.

1.1 Introduction

Ever since the first airplane was built by the Wright brothers the capabilities of aircraft have continuously been improved. For example, the maximum speed of the average military fighter plane has gone from approximately 100 Mph in 1920 to over 1500 Mph currently. These high speeds are responsible for the little time available to fighter pilots to process information and make decisions. In addition, the improved weapons range in military aircraft (missiles can be fired from 20 km away) reduces the pilot's decision time even more.

Where early planes only had a few meters, modern aircraft have several hundreds of meters or information displays, providing the pilot with a wealth of information. With the growth of information sources, the complexity of the available information has increased also. Research has shown that flight performance is correlated to the complexity of the presented information [Svensson et al 1997]. Performance is constant up to a certain complexity level (in Svensson's research the number of displayed items) and deteriorates when the information complexity increases beyond this point.

Combining the increase in aircraft capabilities and information complexity, it is clear that pilots run the risk of information overload. To help a pilot deal with information processing and decision-making, avoid information overload, and optimize flight performance, a crew-assistant system or intelligent pilot-vehicle interface (PVI) has been proposed [Mulgund and Zacharias 1996], [Abeloos et al 2000], [NLR 2000]. The idea is that the assistant-system presents relevant information to the pilot at the right moment and in the appropriate format, depending on the situation, the status of the aircraft, and the workload of the pilot. The intelligent interface should help to improve the pilot's situation awareness and reduce his workload. As a result the survivability of the pilot and plane, and the effectiveness of the flight or mission will be improved.

To correctly assess the amount of information that a pilot can handle we need to know his (mental) workload. Therefore, it is suggested that a good intelligent PVI should have a workload assessment module. The job of such a module is to monitor the pilot and detect situations of cognitive overload or underload (boredom). Taking the pilot's current cognitive load into account, the PVI is able to provide the pilot with the correct amount and format of information.

1.2 Situation awareness

Having a high level of situation awareness (SA), sometimes also called situation*al* awareness, is seen as one of the most critical aspects for achieving successful performance in aviation [Endsley 1999a]. Several definitions of SA exists, but the one from Mica Endsley [1995a p.36] seems to be the most cited:

> *"Situation awareness is the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future."*

Endsley refers to "perception of elements" as level 1 SA, "comprehension of the situation" as level 2 SA, and "projection of future status" as level 3 SA. Having a good SA means that the pilot is able to perceive and understand his situation and can predict the future situation. Basically a pilot builds SA by integrating the different sources of information to one mental situation model of the world. Pilots usually sample the available information sources briefly, following a pattern that is learned from experience. The information in the mental model is updated constantly by continuously sampling and processing new information. To some extent, more experience leads to faster information processing. Information will be processed quicker when the contents of the information is as expected, but people are more likely to make mistakes if it is not. Important aspects that affect a person's situation awareness are among others; workload (overload/underload situations), system design (how is the information presented), system complexity (how complex is the information) and the level of automation (whether the pilot is in or out the control loop).

Research has suggested that many human errors in (military) aviation are caused by lack of SA. Most errors in aviation can be attributed to failures in level 1 SA. Not observing the available information (e.g. due to high workload and/or distractions) has been determined to be the largest single causal factor for these level 1 SA failures [Endsley 1995b]. By understanding why SA problems occur, it is possible to design an interface or system to prevent many of these problems. Designing such a system would require [Endsley 1999b]:

- *An analysis of SA requirements*; e.g. through expert elicitation, observing operator performance, questionnaires.
- *Using design guidelines for enhancing SA*; e.g. always provide goal-oriented information, remove extraneous information, support parallel processing.
- *Evaluation during the design process*; make sure the system does indeed improve SA.

Evaluation of situation awareness is usually done by observing test subjects and/or asking questions about their knowledge, actions and assumptions. Examination is done either during a specific test scenario or after performing the scenario. An example of the first type of evaluation is SAGAT, the Situation Awareness Global Assessment Technique, where the scenario is temporarily frozen to allow the subjects to answer the questions [Endsley 2000]. An example of the latter is SART, the Situation Awareness Rating Technique [Taylor 1990]. An interesting prototype system for on-line automatic situation assessment was created by researchers from Charles River Analytics [Zacharias et al 1996]. Their system called SAMPLE uses two pilot models; the first model has perfect information and the second has to process the available information and make the decisions. SAMPLE allows investigating the way that SA is achieved, how it evolves, and draws up some SA measures by comparing the perfect-information pilot model with the actual-deciding pilot.

1.3 Workload

Many different definitions of workload exist and we will not go into a discussion about them here. We will merely provide the user with our (intuitive) notion of workload.

> *Workload is the effort, both mental and physical, that is necessary to perform one or more tasks.*

With mental workload we denote the cognitive processing that is being done. Physical workload consists of the actual actions and movements that are being performed in the physical world. A person's workload influences his degree of alertness, also known as level of arousal. People need a certain level of arousal to perform tasks and task performance improves with an increase of the level of arousal, but up to a certain point. Above this point, higher arousal levels lead to a decrease of performance [Campbell and Bagshaw 1999, p.129-130]. The relation between performance and level of arousal is described by a curve, known as the Yerkes-Dodson curve, which is shown in Figure 1.

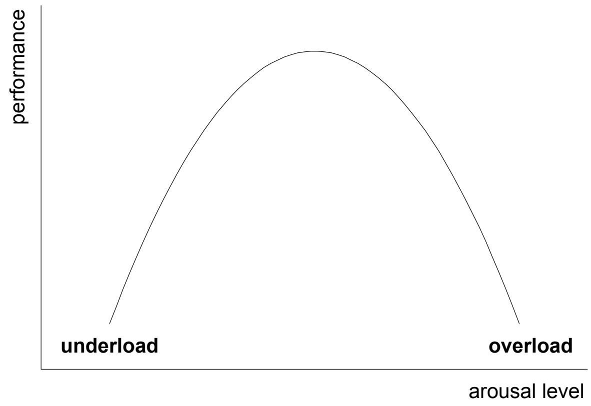

Figure 1: The Yerkes-Dodson curve showing the relation between performance and level of arousal

As can be seen in the figure above, to perform a task well a person's level of arousal should not be too low or too high. Since workload and arousal level are related, a person's workload should also not be too low or high. In situations of stress or high workload, people tend to focus on one particular aspect (information source) of their job and ignore others. For example, a combat pilot with high workload will most likely focus primarily on negative and threatening information. The normal scanning behavior of the pilot is disrupted, reducing the pilot's SA and performance. This phenomenon is known as fixation or "cognitive tunnel vision"[1]. Strong fixations can lead to a resistance to consider alternative hypotheses. Therefore it is important that the workload of a pilot does not become too high.

1.3.1 Mental workload

Since automation has replaced most of the necessary physical action of flying an aircraft, the mental workload of a pilot very much determines the total workload of a pilot. Mental workload, also called cognitive workload, has been defined as the "*degree of processing capacity that is expended during task performance*" [Eggemeier 1988]. Many theoretical and empirical studies on mental workload have been described in literature. These studies present a large number of behavioral, physiological and subjective workload indicators. However, since different researchers use different definitions of mental workload, the devised workload assessment methods are often inconsistent with each other, making the validity and reliability of these methods difficult to assess. On the positive side, the large number of studies has resulted in a number of variables that have been found to influence the level of workload, which includes:

- *Available time to perform an action*; less time means higher workload.
- *Rate of change of workload*; the more it changes, the higher the workload.
- *Use of long-term memory*; using long-term memory causes a higher workload than using short-term memory.
- *Number of items or parameters that must be remembered/controlled;* the average person can only remember about 7 items at a time.
- *The consequences and risks of making an error*.
- *Requiring a response;* workload is higher if a person is required to respond verbally to certain events.
- *Degree of training or expertise*.
- *Emotional state of the subject*.

[1] Other synonyms such as cognitive lockup, functional fixedness, or opportunistic control can also be found in the literature.

Researchers have made the observation that an objective description of influencing variables is not enough to explain the perceived complexity of the tasks and the errors that are made by a test subject. Objective measures do not always correlate with the perceived workload. Therefore, when assessing workload, subjective complexity is often used. Subjective complexity is that aspect of complexity that stems from the characteristics of the operator (expertise, stress, fatigue, and so on).

1.3.2 Physical workload

Physical workload consists of the actions that one performs in the physical world, for example pushing a button. It is not that it is straining for the pilot to push buttons or move the throttle, but physical workload originates from the number of actions that have to be performed within a particular timeframe. For example, when a military pilot is under attack he has to perform a lot of actions within seconds, which causes a high physical workload. By separating physical and mental workload our future intelligent interface can determine whether it should automate some actions (lower the physical workload) or change the modality or method of information presentation (lower the mental workload). According to Padfield [1996], the root mean square of the control activity can be used as a control workload metric, where control activity is given by the power spectral density function of the control column deflection.

1.3.3 Workload measurements

Whereas physical workload can be measured objectively by recording the visible actions that are performed by a subject, a person's mental workload is much harder to measure. There are three often-used methods for measuring (mental) workload.

Subjective ratings

In aviation, subjective ratings are often used to evaluate and validate design decisions. Subjects who are performing a task are asked how difficult they find it to perform the given task, either during or after task execution. The advantage of subjective ratings is that they are very easy to use. A disadvantage is that they can disturb the task performance or (when done after the task) are less reliable since test subjects may forget to mention things or over- or underestimate the task at a later stage. Another disadvantage of evaluation after task execution is that the rating is taken over a large period, making the outcome rather static. Subjective measures are often used in combination with physiological measures.

Examples of subjective rating methods are the Cooper-Harper Scale, the Modified Cooper-Harper Scale (MCH), the BedFord Rating Scale (BFRS), the Subjective Workload Assessment Technique (SWAT) and the NASA Task Load IndeX (NASA-TLX). NASA-TLX seems to be the most popular method. It uses six dimensions to asses mental workload; mental demand, physical demand, temporal demand, performance, effort, and frustration level. After comparing these six dimensions, weights between 0 and 5 are applied to each dimension. Then each dimension is rated using a scale from 0 to 100. The overall workload score is found by multiplying each dimension rating with its weight and dividing the sum by 15 [Hart and Staveland 1988].

Secondary tasks

Secondary tasks can be given to a test subject in order to influence the performance of the subject's primary task. The degradation in the secondary task performance indicates that the subject was required to give more attention (higher workload) to the primary task. In other words, secondary tasks provide a measure for the subjects spare processing capacity. A large drawback of secondary tasks is the intrusiveness of the method.

Physiological measures

Physiological measures, sometimes also referred to as biocybernetic measures, consist of measuring a person's physiological characteristics, such as heart rate, skin resistance, EEG levels, eye blink rate, pupil diameter etc. Changes in these characteristics may be caused by changes in workload. The advantage of using physiological measures is that they are (usually) not as intrusive as secondary tasks and require little or no extra effort for the test subject. However, the drawback is that the translation of the recorded physiological data to workload level is not always reliable. This is caused by the lack of good theoretical frameworks for interpreting physiological measures. In addition, physiological data can be unreliable due to false sensor data (noise). Several physiological measures are discussed in more detail in Chapter 3.

Time and motion studies

Time and motion studies are used to evaluate an operator's workload by recording his actual behavior along a time axis. The performed tasks of the operator are identified and placed on the time axis. Using predetermined workload ratings for a particular task (e.g. subjective ratings), the development of the operator's workload in time can be derived.

1.4 Human information processing

Human information processing consists of the complete process of perceiving signals (hearing, seeing, feeling) to the interpretation of these signals in the brain. A well-known model of human information processing is that of Wickens [1992]. Wickens model consists of a series of stages or mental operations that occur between stimuli and responses:

1. Receiving physical stimuli from the environment through eyes, ears and other senses. Each sensory system has its own short-term memory store that prolongs the stimuli for a short period of time.
2. The stimulus is perceived or recognized (perceptual encoding). Recognition consists of comparing short-term sensory memory to physical codes stored in long-term memory and by integrating the stimuli into meaningful events.
3. After the physical stimulus has been recognized, the subject has to decide what to do (response selection). Information can be stored for later use, it can be integrated with other information, or it might require a certain response.
4. If it is decided that a response is needed, then the subject needs to coordinate and execute the response.

Although information can be processed from multiple sensors at the same time, decision-making limited. Therefore it is necessary to share attention between the information sources [Campbell and Bagshaw 1999]. The capacity of the entire information processing system is in the order of 10 bits per second. If this capacity is exceeded, the probability that people will make mistakes increases. Important information might be discarded, or information belonging to different problems or input are mixed and confused.

1.4.1 Working memory

The basic idea of humans as limited-capacity information processors goes back at least as far as the 1950s and has gradually evolved into the concept of "working memory". The processing of information, that is necessary to reach a good SA, is done in a person's working memory. Perceived information, comprehension and understanding (level 1 to 3 SA) all occur for the most part in the working memory. Working memory limits reduce a person's ability to achieve good SA.
Several important characteristics of working memory have been identified, such as:

- *Maximum number of stored items;* the maximum number of items in working memory appears to be seven (plus or minus 2). This number applies to learning, decision-making, judgment and estimation.
- *Basic processing speed;* there is a relation between simple processing speed and working memory capacity.
- *Displacement/ interference;* when new items enter working memory, other stored items tend to become harder to access. The cognitive system becomes less efficient and slows down.
- *Decay;* items in working memory decay over time. That is, the longer it has been since an item was needed in working memory, the less likely it is that it is currently available.

- *Age;* working memory capacity in adults generally declines with age.

To avoid working memory limits experienced operators may use system knowledge stored in long-term memory instead. This internal representation of the system is called a person's mental model of the system. The mental model contains among others knowledge of relevant elements of the system, rules for interpreting these elements, and mechanisms to project the future status of the system (Endsley's the three levels of SA). Note that long-term memory may be distorted and can be difficult to retrieve information from.

Although working memory plays a central role in the theory and concept of information processing, there currently is no good formal theory that integrates the known working memory characteristics and allows us to predict its performance.

1.5 Cognitive task models

A way to gain insight into the mental workload of a person is to use cognitive task models. Unfortunately, there is no good context-independent theory on human cognition that can be used in different situations. However, a number of practical cognitive task models are available that work well in limited and very specific domains.

1.5.1 Skill, rule and knowledge-based behavior

Rasmussen [1986] has developed a framework for cognitive task analysis in the domain of supervisory control of industrial installations. Rasmussen identified eight stages of decision-making (see also Figure 2).

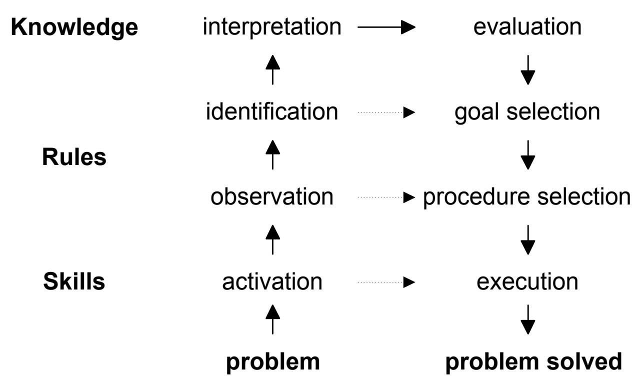

Figure 2: Abstraction of Rasmussen's stepladder model of decision-making including the skill, rules and knowledge level of information processing.

Solving a problem requires following all eight stages, but through experience or training short-cuts can be developed that allow bypassing some steps (the dotted arrows in Figure 2). These short-cuts represent other (lower) levels of information processing. Rasmussen identifies the following three levels of behavior:

- *Skill-based level* consists of highly practiced actions that are performed automatically without consciously thinking about them. In Figure 2, skill-based behavior is represented by the short-cut between activation and execution.
- *Rule-based level* which is procedure or goal oriented. On this level the executor has some experience with the task at hand. Most of the time a predetermined set of rules can be applied that have proven to be successful previously. The process of choosing a rule is done more or less consciously, but once a rule is chosen the appropriate actions are carried out automatically. In Figure 2, rule-based behavior is represented by the short-cuts between observation and procedure selection, and between identification and goal selection.

- *Knowledge-based level* behavior is applied in new situations that require planning of actions. The executor sets local goals, initiates actions to achieve them and observes if the actions are successful. If needed the plan is adjusted and new subgoals are devised to better achieve the goal. Knowledge-based behavior follows all eight stages of decision-making.

Knowledge-based tasks demand more attention and working memory than rule-based tasks. Rule-based tasks in turn require more effort than skill-based tasks. At which of the three levels a task is performed depends on the person's experience with the task and the complexity of the task. The complexity of a task is determined by the learnability of the actions, which in its turn depends on the size of the problem space (number of possibilities) and the situational variability (differences in each possible situation).

1.5.2 Multiple limited resources models

Multiple resources models work under the assumption that people have several different capacities, which can be regarded as limited resources. Tasks that are performed concurrently will interfere with each other if they have to share resources. Also, there is a limit to the amount of a type of resource that can be used. An example resource model is that of Wickens which is shown below in Figure 3.

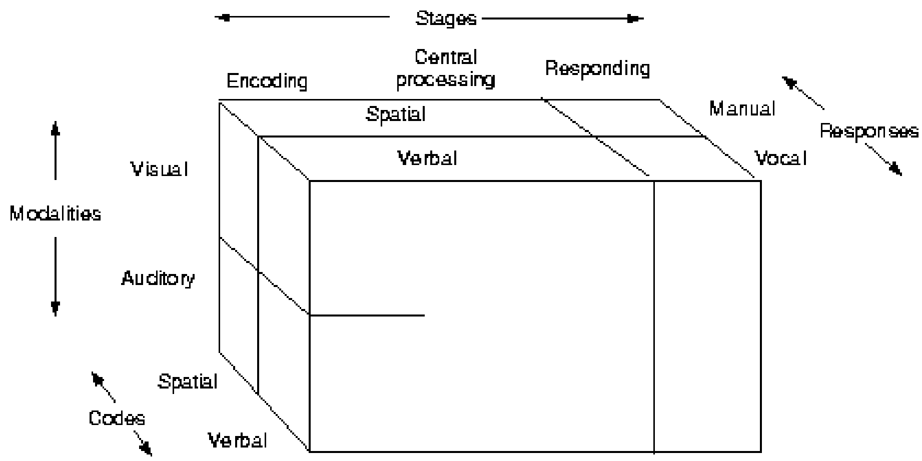

Figure 3: The multiple resource model of Wickens [1992]

Although multiple resource models are good in modeling task execution through different modalities (e.g. verbal or visual) is does not explain all phenomena associated with multiple task performance such as task switching, confusion and cooperation.

1.5.3 GOMS-based models

A well-known and much-used model for measuring cognitive load is GOMS, which stands for Goals, Operator, Methods and Selection rules [Card et al 1983]. GOMS performs very well for simple and well-practiced task, such as using a computer. A GOMS model describes the "methods" that a computer user needs to accomplish his or her goals. A method is a series of steps consisting of operators that the user performs. If there are more methods to accomplish a goal, then the GOMS-model includes selection rules that choose the appropriate methods depending on the context. This way the GOMS framework can be used to produce a variety of models with different levels of detail. Many GOMS-derived models have been designed. Here we will briefly address two of them: Soar for expert knowledge-based behavior and EPIC for expert skill-based behavior.

Soar

Soar is a theory of cognition as well as a cognitive architecture that can be used as a platform to test other cognition theories. Soar extends the GOMS-model to include planning and problem solving strategies. Just like a production system, Soar tries to solve problems by trying to go from the current state to a defined goal state [Newell 1990]. The Soar system works by specifying a “goal context” that contains a goal, a problem space, the current state, and operators to go to other states. The perception/motor interface adds new data from the external world to Soar’s internal representation (one or more goal contexts) in the working memory. The contents of the working memory trigger motor actions and/or associations in the long-term memory (LTM). The LTM contains mappings from one goal context to another. When Soar reaches an impasse (i.e. not enough knowledge is available in the current goal contexts), new associations can be generated in the LTM by using previous experiences. This way Soar is able to learn [Lehman et al 1996].

The Executive Process-Interactive Control (EPIC) model

The Executive Process-Interactive Control (EPIC) framework is a framework for human information processing suited for modeling multiple task performance [Meyer and Kieras 1996]. It couples basic information processing to perceptual-motor mechanisms using a (GOMS-structured) production system.

GOMS seems to be well suited for modeling tasks where procedural task knowledge is involved. However, obtaining the knowledge required to create a GOMS-based task analysis is very difficult and infeasible with complex tasks. John and Kieras [1994] summarized in a report the predictive capabilities of GOMS and other similar models. They showed that their predictions of duration of task execution (performance times) are generally well correlated with the duration observed in experimental situations. However GOMS-based models can only be used with a certain class of cognitive processes in a certain type of environment. Since GOMS-based models need a description of the task that is under investigation, very complex tasks or tasks in unfamiliar situations cannot be modeled.

1.5.4 Cognitive task load model

Neerincx [Neerincx and Van Besouw 2001] developed a practical theory for mental load and cognitive support in process-controls tasks. In their (nameless) model, cognitive task load is a function of the percentage time occupied on the task, the level of information processing (skill-based, rule-based or knowledge based) and the number of task-set switches (switching between different goals).

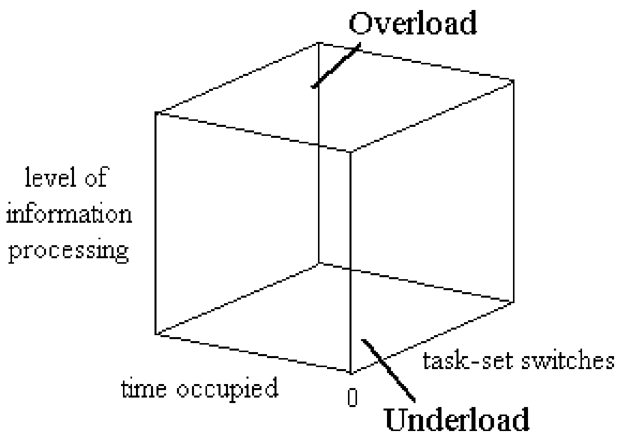

Figure 4: The 3 dimensional model of cognitive task load [Neerincx 2003]

Note that not all “parts” of the cube will occur in practice. The three axis very much influence each other, e.g. when the level of information processing rises (to knowledge-based processing) then the time occupied will also increase.

Chapter 2: Predicting a pilot's workload through cognitive task modeling

In this chapter we present a model for estimating a pilot's workload using cognitive task modeling. After a short introduction in section 2.1, we present our model in section 2.2. In section 2.3 we argue that we can use this model to get an estimation of a pilot's workload during a mission as long as we can recognize what the pilot is doing. In section 2.4 we compare our workload assessment model to some general guidelines. We conclude this chapter with a summary and example in section 2.5.

2.1 Introduction

The total workload of a pilot consists of his physical and mental workload. Physical workload is the number of manual operations (actions) the pilot has to perform (e.g. pushing a button). Mental or cognitive workload is the mental effort needed by the pilot to perform his task. Very simply stated the total workload is the result of the following formula:

$$\Sigma \text{ (operations to do)} * \text{(level of complexity)} \qquad \textbf{Equation 2-1}$$

Manual workload can be estimated in a relatively easy manner; we simply count the number of operations that the pilot has to perform and multiply each operation with a subjective complexity factor. Cognitive workload is much harder to estimate since we generally do not know exactly what the pilot is thinking about. However, if we perform a task analysis and look at the mission plan of the pilot, we can try to estimate the workload of the pilot, based on the actions (both manual and cognitive) that the "average" pilot should perform at each moment during this mission. Flying a military plane is a very procedural job, so the tasks that a pilot has to perform are very much regulated. By assessing the complexity of all tasks, multiplying them with the estimated complexity, we get an indication of what the pilot's total workload should be.

It is not possible to draw up a complete model that accurately describes a pilot's workload and changes in workload levels over time, in all its detail. For this we still lack the necessary knowledge and theories about (cognitive) workload. Instead, we present here a simplified workload model that we expect to be sufficient enough for our purposes, which is to provide an estimate of the total workload of a pilot that can be used in an intelligent interface. Since workload can be influenced by many factors, we will restrict our model and assume that the pilot is well trained, well rested and not under the influence of alcohol or drugs.

2.2 The workload model

The predicted workload of a pilot will be a function of time and performed tasks. The workload associated with time-critical tasks should be higher than that of non-time-critical tasks. Also, workload should be higher with a larger number of concurrent tasks. In addition, the difficulty of a task has to be taken into account.

2.2.1 Baseline and task load

We will assume that pilots are always busy with something (unless the pilot is unconscious) and will always have a certain baseline workload level, which is associated with the always-present tasks during a flight or mission (e.g. situation assessment, controlling the plane etc.). The exact level of this baseline workload will be pilot dependent. A very experienced pilot will have a lower baseline than a rookie pilot. Further, we assume that performing other tasks that are necessary for the mission requires a certain level of energy, time and capacity, expressed as task load. This level of task load depends on the type and difficulty of the task (skill, rule and knowledge-based processing).

A simple approach is to assign a fixed task load for each possible task in advance. Skill-based tasks will have a lower task load than rule or knowledge-based tasks. For simplicity the pre-determined load will be spread equally over the estimated time that the task will take. The total workload of a pilot at a certain time will be the sum of all task loads associated with each task that is (or should be) performed at that particular moment.

As a first-order estimate, we will set the task loads based on the number of subtasks (actions) that make up the task. Subtasks can be either cognitive or physical. Initially, we will give each cognitive task the same load. All physical tasks also get the same load, but this load will be lower than that of cognitive subtasks. As a second-order estimate, we can try to determine the task load of a particular task by using a workload evaluation method.

2.2.2 Multiple tasks; concurrent and sequential tasks

Summing up the load of all currently performed tasks to obtain the total task load is not enough. Some tasks will "interfere" with the execution of other tasks and can cause additional load on the pilot. Interference will be minimal if tasks require different processing skills (e.g. spatial and linguistic). Also, we have to make a distinction between tasks that can be handled concurrently and tasks that must be handled sequentially. Tasks that are handled in parallel are more demanding to the pilot than the sum of their parts due to the necessary scheduling and task coordination. Therefore, we expand the model to include a separate scheduling and coordination task, which will have a load that is dependent on the number of concurrent tasks and the method of processing.

2.2.3 Stress and relaxation

We expect that stress will also play a role in the workload of a pilot. For example, if a pilot is attacked and an enemy missile is approaching, some tasks have to be suspended and others have to be performed at a higher pace. In our model, the higher pace results in higher workload since the task load is divided over a smaller time frame. However, we expect that this effect is not enough to accurately predict the pilot's workload, so to accommodate we will provide the tasks that are associated with handling threats or equipment failures with a significant higher task load than other tasks. In addition, we will also raise the workload level of the scheduling and coordination task in case of threats and/or failures.

After a very demanding or stressful task has been completed, there will be a relaxation time. A pilot cannot continuously perform high-stress tasks. Even in normal cases people should only be occupied between 70 and 80 percent of the total time available. Therefore, the workload of a completed task will degrade slowly over time. This means that when a pilot has just finished dealing with a threat or failure and encounters a new threat shortly after, his workload will be higher than before since he did not have time to completely "recover" from the previous encounter. This is also known as the "carry-over effect".

2.2.4 Workload threshold, physical condition and fatigue.

People perform better when there is a certain level of arousal or tension, but will collapse under too much pression (see also section 1.3). Every pilot will have a certain workload maximum at which the pilot will perform his tasks most efficiently. Below this maximum the pilot will not perform optimally, but usually performance will be acceptable. However, above the maximum, pilot performance will degrade rapidly. Therefore, in the ideal case a pilot's workload should be kept below or at this threshold.

The workload threshold of a pilot will be dependent on the individual and the current state of the particular pilot; his physical condition and mental well-being among others. If a pilot is tired it will be more straining for him to perform multiple tasks. To reflect fatigue during a mission, the threshold workload level of the pilot will slowly go down during the mission. If we look at the pre-determined task loads we can add a certain additional weight to each task load reflecting the pilot's physical and mental condition. This also allows us to express individual pilot differences. However, these individual settings will have to be set in advance before a mission. Note that we ignore differences physical conditions that may occur as a result of the time of day (e.g. night time). The exact value of the workload threshold has to be determined in experiments.

2.3 Adapting to unexpected events

In the previous section, we assumed that we can give an estimate of a pilot's workload based on a given flight or mission plan and the actions that should be performed at each stage during the flight. This requires that the provided flight or mission plan is described in sufficient detail. Also, the workload model will fail if an unexpected event will arise that is not known beforehand. The problem is that we do not know what actions a pilot should perform once such an unexpected event arises. Therefore, we need an identification mechanism that recognizes events and determines the current situation. Once we get an interpretation of the situation, we can simply look up the associated tasks for that situation and dynamically adjust the workload model to accommodate this new event. Further, we will assume that unexpected events that are not in the mission plan will have a relatively higher task load.

2.4 Criteria for workload assessment methods

O'Donnell and Eggemeier [1986] have drawn up a number of criteria to which any good workload assessment method ideally should comply. These criteria are:

- *Sensitivity:* the index should be sensitive to changes in task difficulty or resource demand.
- *Diagnosticity:* an index should indicate the cause of a change.
- *Selectivity:* the index should only be sensitive to differences in capacity demand and not to changes in other factors such as physical load, emotional stress, alcohol use, etc. that are unrelated to workload.
- *Obstruviness:* the method for obtaining the index should not interfere, contaminate or otherwise disrupt the performance of the primary task of the test subject.
- *Bandwidth and reliability:* the index should be reliable and be able to track workload changes over time. Also, the index must be derived within a certain amount of time.

Note that our workload assessment method described in this chapter meets all of these criteria.

2.5 Summary and example

To summarize, the workload of a pilot consists of the sum of all weighted task loads. There is a certain baseline that will slowly rise as the pilot becomes more fatigued. Also there is a certain limit above which the pilot's performance will degrade. In addition, we have a scheduling and coordination task with a task load depending on the number of concurrent tasks. After finishing a task the workload of that task will slowly degrade (relaxation time). In formula:

$$W_{pilot}(t) = \Sigma\,(currenttask_i * weight_{task(i)} * weight_{pilot(i,p)}) + \Sigma\,(previoustask_{j,t} * weight_{task(j)} * weight_{pilot(i,p)} * timefactor) + baseload + taskcoordination \quad \textbf{Equation 2-2}$$

In Figure 5 below, we give an example of the workload and its development in time of a pilot that is attacked by a missile. Before the attack, the pilot is busy navigating. Once the missile is detected, the pilot stops his navigation task and starts the evade-incoming-missile task. Since this task his very dangerous for the pilot, the task has a very high task load. The pilot decides to change his radar mode which causes additional task load (and a slightly higher scheduling and coordination taskload). After changing the radar mode, the pilot decides to eject some chaffs and flares. Since these two tasks are manual task, their task load is not very high. We also see that when a task ends, its associated workload slowly goes down.

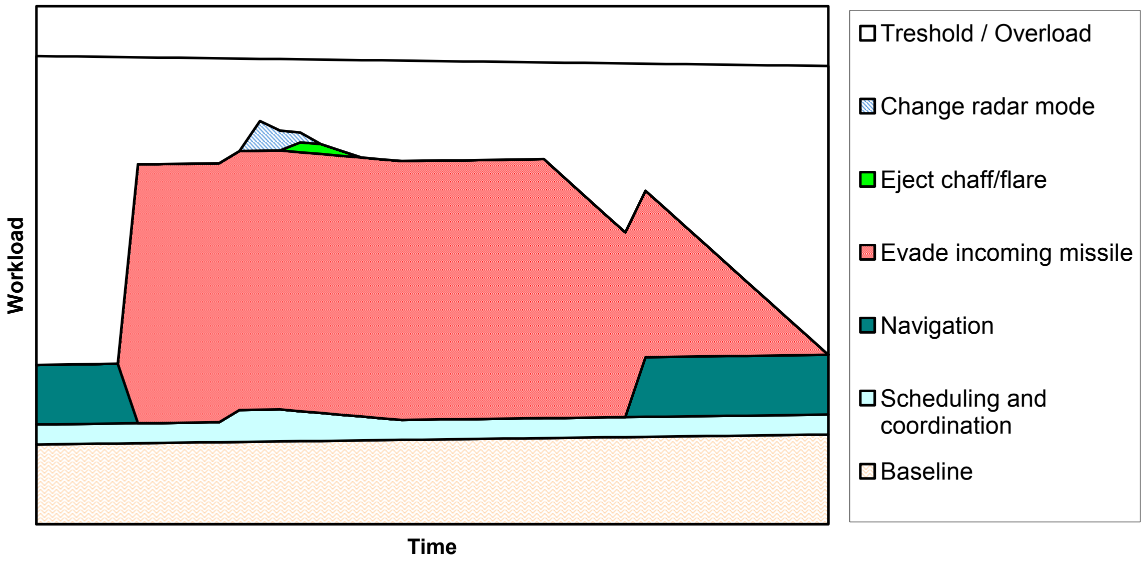

Figure 5: Example of the estimated workload of a military pilot during a missile attack

Chapter 3: Real-time physiological measures for pilot workload estimation

This chapter describes our model for estimating a pilot's workload in real-time using physiological measures. In the first section we discuss the idea shortly. Then in section 3.2 we provide some information about a popular method for gathering physiological data, which is the gaze tracker. Section 3.3 mentions other types of physiological data. Finally in section 3.4 we present our model.

3.1 Introduction

The workload assessment model presented in the previous chapter will only give us a rough approximation of a pilot's workload. The accuracy of the workload model is mostly determined by the accuracy of recognizing the current situation and pilot tasks. Another approach is to determine a pilot's workload in real-time using physiological measures. For example, we can use a gaze tracker to determine what the pilot is looking at, so we know what he is interested in and what information he is dealing with.

From literature we know that physiological measures such as heart rate, hart rate variability, perspiration, and blink rate can be used as indicators of mental workload. However, there are some difficulties. Most physiological measures can be indicators of physical work or stress as well. In addition, these measures are often person dependent and measures can contain a lot of noise. Current theoretical models are not sophisticated enough to accurately interpret physiological measures in all cases. Byrne and Parasuraman [1996] suggested that workload measures may be especially difficult to use in underload situations since other effects than task difficulty can predominate in this case. These difficulties force us to be very careful when interpreting physiological data and consider their reliability. Just like the pre-determined workload, the real-time workload will not be very precise (probably even less so), but a large advantage of a real-time workload estimate is that it will still provide relevant data in unforeseen and unknown situations.

3.2 Gaze tracker

A gaze tracker is a device that is mounted either on a person's head or is placed remotely in front of the person (e.g. on the desktop). The gaze tracker sends out infra-red light and captures reflections of this light from both the corneas and the retina of the person's eye.

3.2.1 Human perception

People use their eyes with very little conscious effort. Most of the time, we automatically look at the object we are working on. To see an object clearly, it is necessary to move your eyes so that the object appears on the fovea, a small area at the center of the retina. The fovea covers approximately one degree of visual arc. Because of this, a person's eye position provides a rather good indication (to within the one-degree width of the fovea) of a person's focus point of attention on a display [Jacob 1991]. Note that stationary objects just outside the fovea will hardly be noticed, but moving objects can be seen in this peripheral view.

Eye movements

The most common way of moving the eyes is a sudden saccade. Saccades are rapid and often irregular eye movements. They generally take between 30-120 ms and are often followed by a fixation on an object. Fixations are 200-600 ms periods of relative stability of the eye during which an object can be viewed. However, during a fixation the eye does not remain entirely still. It makes small, jittery motions, generally covering less than one degree. This means that the overall picture of eye movements for a person sitting in front of a stationary information display (e.g. a cockpit) is a collection of steady (but slightly jittery) fixations connected by sudden, rapid saccades. While the eyes are rarely entirely still (they seldom remain in one fixation for long) people generally think they are looking steadily at an object and are not aware of the small motions during fixations. Smooth eye motions, less sudden than saccades, occur only in response to a moving object in the visual field (pursuit movements). Other eye movements that can occur are nystagmus (reflexive or involuntary movements of the eyes), vergence, and torsional rotation, but these are less relevant to detect for our purposes.

Queuing theorem

Operator object scanning behavior follows the queuing theorem [Carbonell 1968] which states that the pilot will scan the instruments in such a way that the total "expected costs" will be minimized. The expected costs are defined as the true costs of missing an event multiplied by the probability that the event will be missed [Wickens 1992]. The probability of missing an event is related to the event frequency and increases with the amount of time since the channel has last been sampled. This way the most important events (glances at objects) will be scheduled first in order of importance, and will also be depending on the last time the object was checked. Some conclusions have been made which can be summarized as follows [Heiligers 2002]:

- People incline to form a model of the events that occur on each channel. This implies that if an instrument has been changing slowly for quite some time, a pilot will not respond immediately if a sudden change occurs.
- Instruments with a higher event frequency are sampled more often than instruments that remain approximately constant.
- Due to an imperfect memory, people tend to oversample channels with a low event rate because they forget the status of that particular channel.
- Sampling becomes somewhat more optimal when a preview is available of the scheduled events that are likely to occur in the future.

3.2.2 Gaze tracker data

Input from a gaze tracker consists of a continuous stream of raw data that usually specifies the x and y coordinate of the point of gaze. Generally, eye position is recorded every 1/60 of a second or more. The recorded data is first filtered for noise and sometimes the data is compensated for head movements (when this is measured). Then the meaningful events must be recognized. Most likely this will be saccades and/or fixations. In his paper, Robert Jacob [1991] describes a fixation recognition algorithm that detects sequences of 100 ms during which the eye position changes no more than 0.5 degrees. Further eye positions within approximately one degree are assumed to represent continuations of the same fixation. Note that there is no standard technique for determining fixations and that a minor change in the determination method can lead to very different results. Different definitions of fixations in the literature and different gaze tracking equipment have lead to several distinct algorithms of fixations, which makes a comparison of results very difficult. Besides fixations, other measurements that can be taken with a gaze tracker include; the point of gaze (e.g. a specific object), fixation duration, scan pattern (and its randomness), pupil diameter and blink rate.

Point of gaze

When a fixation occurs, we can determine the point of gaze and thus the object of interest of the pilot. To associate the detected eye position with a visible object on a display or console, the nearest neighbor algorithm can be used.

Fixation duration

Fixation duration (also called dwell time) can be used as a measure of difficulty of information extraction and interpretation. Svensson has found that the durations and frequencies of eye fixations change as a function of information load [Svensson et al 1997]. He noticed that an increase of the information load (which was displayed on instruments head-down) generally results in longer fixation times head-down. The fixations times head-down are correlated linear with the number of displayed items. The fixation times head-up first drop (with 6 items or less displayed) and then remain constant (with more than 6 items displayed).

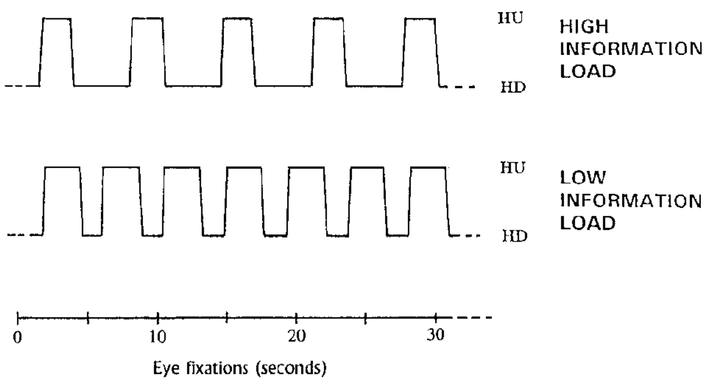

Figure 6: Eye fixations head-up and head-down over time under high and low information load [Svensson 1997].

Scan pattern

As a person's workload increases, the randomness of his scan pattern will decrease. Also note that pilot fatigue may result in a limited scan pattern; where the pilot only checks the most important instruments and neglects others.

Blink rate

Brookings et al [1996] have found that blink rate is very sensitive to the level of workload in visually demanding environments. A clear distinction between blink rates at low, medium, high and overload levels was found.

Pupil diameter

It has been known that the size of a person's pupil can be influenced by:

- Workload and stress
- The amount of light
- Alcohol level
- Emotion (e.g. sexual interest)
- Age

Therefore, only under very strict circumstances (e.g. fixed uniform lighting) can pupil size be used as an indicator of workload.

3.3 Other physiological measures

In this section we will address other possible physiological measures that can be used to estimate a pilot's workload.

3.3.1 Heart rate

A popular physiological measure is heart rate. Heart rate has been found to vary as a function of the mental load caused by an operator's task. Svensson has used heart rate has an indicator of a pilot's psychophysiological activation during a mission. He used sortie (mission) means, running means and variance and found that heart rate (sortie means and variability) increases as a function of the subjective task difficulty and complexity of the information [Svensson et al 1997]. Svensson also reports that the diagnostic value of heart rate (running means) is affected by the pilot's skill level.

Heart rate variability (HRV) is a measure of the oscillation of the interval between consecutive heartbeats. Heart-rate variability has been used as a measure of mental effort and has been reported to decrease with increases in mental demand. Rowe et al [1998] suggested that HRV reflects the mental effort of a user while interacting with an interface. In their study subjects had to respond to moving objects on a screen under various conditions. They found that the HRV decreases with an increase of the number of moving objects. They suggest that the subjects tried to maintain a certain performance level and increase their mental effort. When the number of objects increases further, HRV shows a decrease. Rowe suggested that users have to react to more objects than they can handle and as a result settle with less performance and reduce their mental effort.

3.3.2 Speech

Berthold and Jameson have drawn up a list of results mentioned in the literature concerning the effects of cognitive load on features of speech [Berthold and Jameson 1999]. They divided the symptoms into two categories; symptoms involving output quality (sentence fragments, false starts, syntax errors, etc.) and symptoms involving output rate (articulation rate, speech rate, silent and filed pauses, etc.). Berthold and Jameson also mention that some of these symptoms may be very difficult to detect by an automated system. For workload assessment in the cockpit, one of the problems is that pilots do not always talk, and if they do, their comments are usually not long enough to gather sufficient data to analyze. Another big problem is that in the cockpit there is a lot of background noise that distorts the recorded speech. We do not recommend using speech for workload assessment in airplanes.

3.4 The workload assessment method

For our online and real-time workload assessment method we propose to gather data with a gaze tracker. In addition, we record the physical actions done by the pilot, such as pressed buttons and stick and throttle movements. This makes a distinction between mental and physical workload possible. For each pilot, we first need to determine a “baseline” workload in order to compensate for individual differences in physiological response to task load.

3.4.1 Mental workload

Our proposed workload assessment method uses a combination of number of fixations, fixation time, and blink rate. At the moment point of gaze and pupil diameter are not used. We expect the latter to be too unreliable. Point of gaze might be added in a later stage in order to record the information that was already sampled by the pilot.

The mental workload of a pilot is derived from the sum of the current average number of fixations, fixation time and blink rate. These averages have to be calculated over a certain (moving) time interval which will be set to 30 seconds. Each measure will be multiplied by a certain weight, so that each measure has a fair contribution to the total workload. The exact weights have to be determined in experiments. Fixation length is used as an indication of difficulties in information extraction and processing caused by mental load. A larger average fixation length will indicate a higher workload. An increase in the number of fixations indicates less efficient search for information which may be the result of a higher workload. A decrease in the average blink rate will also indicate a higher workload. Normally, we will regard a higher number of fixations as a higher workload. However, in overload situations (tunnel vision) the number of fixations will decrease. Therefore, we will have to set a certain limit below which the number of fixations will be regarded as high mental load (instead of low mental load).

We will assume that if a pilot is interested in a particular object, he will look at this object and focus (fixate) on it. We further assume that objects in a person’s peripheral view are not noticed.

3.4.2 Physical workload

The physical workload of the pilot is a weighted sum of all the detected physical actions of the pilot. A large component of the physical workload will be the stick movement done by the pilot. For this, we look at the root mean square of the differences in control activity as proposed by Padfield [1996]. Throttle movements and pressed buttons add to the workload for a short time period after which their contribution degrades. Again, the exact weights of each activity has to be determined in experiments.

3.4.3 Total workload

The total workload of the pilot will be the sum of his mental and physical workload. If a pilot's workload becomes very high, the separate values for physical workload and mental workload can give an indication what type of adjustment can be done by our intelligent cockpit environment.

3.5 Final remarks

The described method of workload assessment using physiological measures has been derived based on results described in literature. We are planning to perform some experiments to indicate whether good results can be achieved with our proposed assessment method. It is very well possible that some of the used physiological variables found in literature do not (sufficiently) apply in our particular application. It is also unclear whether the used method of data gathering (e.g. fixation algorithm) will give accurately enough results for this method.

Chapter 4: Workload assessment and adaptive automation

In this chapter we will discuss some of the implications and possible problems of using workload assessment in an adaptive automation system. After a short introduction in section 4.1, we will present some guidelines automation in section 4.2. Section 4.3 discusses when to use either of our devised workload assessment methods. We close this chapter with some potential pitfalls that may arise when using workload assessment (section 4.4).

4.1 Introduction

Traditionally automation is static. A machine always performs a particular task in the same way. It has been known that traditional automation can lead to loss of situation awarareness, manual skill degradation, impaired decision making and monitoring efficiency. What is more, in aviation traditional automation has lead to a number of independent and non-cooperative systems that can contradict each other. Adaptive automation provides a dynamic method of automation. Adaptive automation can take over a task depending on specific circumstances, for example the number of errors made by the operator of the system. The goal of adaptive automation is to keep an operator in the control loop and "divide" the work between the system and the operator. Scheduling of tasks is automatically regulated by taking the operator status (capabilities, workload,) into account. Not only can an adaptive system decide to take over and automate certain tasks, it can also decide on a particular method of automation or automation feedback. For example, auditory messages may be given if the system detects that the operator is busy scanning other parts of the system. In order to be effective, the system and the operator must be aware of each others capabilities and performance.

Workload assessment can play a part in adaptive automation in two ways. First, workload assessment can provide information about the effectiveness and practicality of any adaptive system. Second, real-time physiological measures can be used as a "trigger" for adaptations made by the adaptive systems. However, since physiological measures are not always reliable (see our discussion in the previous chapter), it is best to use them as a "supporting" input measure, part of a set of multiple input variables, such as critical event detection, operator preferences, and performance assessment. Also, some input triggers can be defined more clearly a priori (e.g. critical event detection).

4.2 Guidelines for automation

Not much is known about using workload assessment techniques in a closed-loop adaptive environment. Many physiological measures are not sensitive to small variations in workload. This is probably the reason that the emphasis in adaptive automation research has been to prevent operator overload, which can be detected more easily.
In his PhD thesis, Mark Neerincx gives some guidelines for the "harmonization of the task to human capacities" [Neerincx 1995]:

1. The total number of actions that are performed during a certain period should have an upper and lower limit.
2. The ratio, between rule- and knowledge-based actions in cognitive tasks should have an upper and lower limit (skill-based actions are barely cognitive demanding).
3. Avoid long-term periods during which only one sort of action is performed continuously (e.g. monitoring). This may degrade performance rapidly.
4. Avoid momentary overloading that can result from several knowledge-based actions in rapid succession in a short time or (almost) simultaneously performed rule- or knowledge-based actions. Overloading may lead to an accumulation of errors and rapid performance degradation.

Neerincx also remarks that there are no fixed upper and lower limits for point 1 and 2. These will be dependent on the type of task and should be determined during cognitive task load analysis.

4.3 Model-based workload versus real-time determined workload

We expect that both the cognitive task workload assessment method described in chapter 3 and the real-time assessment using physiological measures described in chapter 4, will not be very precise. Also the appropriateness of any quantitative value for each workload assessment method is not clear. Therefore, we propose to "fuzzify" the measures. Instead of using absolute values, we use a category consisting of low, medium, high and overload workload. This makes a comparison of results between the two workload assessment methods possible.

If the predicted cognitive task workload is different from the physiological workload estimation then there are three possible explanations:

1. It is possible that the pilot has encountered a *situation in which we did not foresee* (e.g. illness of the pilot can influence his workload) or there is an unexpected situation that cannot be identified. In this case we have no choice but use the measured physiological workload.
2. *The predicted workload model is wrong*. Besides unforeseen situations, the estimated workload uses several theories of cognition that may prove to be incorrect in particular situations. Again, we have no choice but to use the measured physiological workload.
3. *The physiological workload model is wrong*. Sensors may fail and real-time monitoring can be influenced by a lot of factors. Therefore, it is very well possible that the used sensors will provide very noisy and/or erroneous data. However, by using multiple measures we are able to cross-reference the data and possibly detect erroneous measurements. If we are able to detect that the difference between the measured and estimated workload is caused by faulty measured data then we can continue to use the predicted workload, and ignore the faulty readings.
4. *Both the estimated and the measured workload model are wrong*. In this case we are lost.

Experiments are needed to assess the reliability of both workload assessment methods and determine the situations that either one might fail.

4.4 Potential problems

Some researchers have stated their concern that fully adaptive systems might be too unpredictable for human operators. Nevertheless, most researchers feel that adaptive automation in many cases can be a large improvement above static automation because it centralizes the role of the human operator. A potential problem of using workload assessment in a closed-loop adaptive environment can be learned responses. In literature there are reports of conditioned psychophysiological responses. It is possible that operators adjust their physiology after repeated use of an adaptive system in order to provoke a certain response of the system. Although, this should not necessarily be a problem, it can have unwanted side effects. Also, once operators become accustomed to an adaptive system, their interaction strategy may change, causing a change in cognitive task load. This makes workload assessment based on cognitive task analysis difficult. In addition, one should note that physiological feedback from novice users in a system will most likely be very different from that of an expert user of a system.

Bibliography

Abeloos, A.L.M., Mulder, M., and Van Paassen, M.M. (2000) "The applicability of an adaptive human-machine interface in the cockpit", in *Proceedings of the 19th European Annual Conference on Human Decision Making and Manual Control (EAM 2000)*, June 26-28, Ispra, Italy.

Berthold, A. and Jameson, A. (1999) "Interpreting symptoms of cognitive load in speech input", in *Proceedings of the 7th Int. Conference on User Modeling* (UM99), Kay, J. editor, pp. 235-244, Vienna, Austria.

Brookings, J.B., Wilson, G.F. and Swain, C.R. (1996) "Psychophysiological responses to changes in workload during simulated air traffic control", in *Biological Psychology, Vol.42*, No.3, pp. 361-377, Elsevier Science.

Byrne, E.A. and Parasuraman, R. (1996) "Psychophysiology and adaptive automation", in *Biological Psychology, Vol.42*, No.3, pp. 249-268, Elsevier Science.

Campbell, R.D. and Bagshaw, M. (1999) "*Human performance and limitiations in aviation*", 2nd Edition, Blackwell Science Ltd.

Carbonell, J.R., Ward, J.L. and Senders, J.W. (1968) "*A queuing model of visual sampling: experimental validation*" in IEEE Transactions on Man-Machine Systems, MMS-9.

Card, S. K., Moran, T.P. and Newell, A. (1983) "*The psychology of human-computer interactio*n", Lawrence Erlbaum Associates.

Endsley, M.R. (1995a) "Toward a theory of situation awareness in dynamic systems", in Human Factors, Vol. 37, No. 1, pp 32-64.

Endsley, M.R. (1995b) "A taxonomy of situation awareness errors", in Human factors in aviation operations, pp. 287-292, Fuller, R. Johnston, N. and McDonald, N. eds. , Ashgate Publishing Ltd.

Endsley, M.R. (1999a) "Situation awareness in aviation systems", in Handbook of Aviation Human Factors, Garland, D.J., Wise J.A., Hopkin, V.D. eds, Lawrence Erlbaum Associates, Mahwah, NJ USA.

Endsley, M.R. (1999b) "Situation awareness and human error: designing to suppport human performance", in Proceedings of the High Consequence Systems Surety Conference, Alburquerque, NM USA.

Endsley, M.R. (2000) "Direct measurement of situation awareness: validity and use of SAGAT", in Situation Awareness Analysis and Measurement, Endley, M.R. and Garland, D.J. eds, Lawrence Erlbaum Associates, Mahwah, NJ USA.

Eggemeier, F.T. (1988) "Properties of workload assessment techniques", in Human mental workload, Hancock, P.A. and Meshkati, N. eds., pp. 41-62, North-Holland, Amsterdam, The Netherlands.

Hart, S.G. and Staveland L.E. (1988) "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research", in *Human mental workload*, Hancock, P.A. and Meshkati, N. eds., pp. 139-183, North-Holland, Amsterdam, The Netherlands.

Heiligers, M.M. (2002) "*Prediction of pilot scanning workload and control workload based on a computer simulation of a hirozntal flight through turbulent air*", MSc. thesis, June 2002, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands.

Jacob, R.J.K. (1991) "The use of eye movements in human-computer interaction techniques: what you look at is what you get", in *ACM Transactions on Information Systems Vol. 9*, No.3, pages 152-169.

John, B.E. and Kieras, D.E. (1994) "*The GOMS family of analysis techniques: tools for design and evaluation*", CMU Computer Science Technical Report No. CMU-CS-94-181. Also Human-Computer Interaction Institute Technical Report No. CMU-HCII-94-106.

Kieras, D.E. and Polson, P.G. (1985) “An approach to the formal analysis of user complexity.” in *International Journal of Man-Machine Studies, Vol. 22*, pp. 365-394.

Lehman, J.F., Laird, J.E., and Rosenbloom, P.S. (1996) “A gentle introduction to Soar, an architecture for human cognition”, in *Invitation to Cognitive Science, Vol. 4*., Sternberg , S. and Scarborough, D. (Eds), Cambridge, MA, MIT Press.

Meyer, D.E., and Kieras, D.E. (1996) “*EPIC: adaptive executive control of human multiple-task performanc*e”, paper presented at the meeting of the Psychonomic Society, Chicago, IL. .

Mulgund, S.S. and Zacharias, G.L. (1996) “A situation-driven adaptive pilot/vehicle interface”, in *Proceedings of the Human Interaction with Complex Systems Symposium*, Dayton, OH, August 1996.

Neerincx, M.A. (2003) “Cognitive task load analysis: allocating tasks and designing support”, in Handbook of Cognitive task design, Hollnagel, E. (Ed.), Lawrence Erlbaum Associates, Inc. (forthcoming)

Neerincx, M.A. and Van Besouw, N.J.P. (2001) “Cognitive task load: a function of time occupied, level of information processing and task-set switches”, in Engineering psychology and cognitive ergonomics, Vol. 6, Industrial ergonomics, HCI, and applied cognitive psychology, pp. 247-254.

Neerincx, M.A. (1995) “*Harmonizing tasks to human knowledge and capacities*”, PhD Thesis, Rijksunversiteit Groningen, The Netherlands.

Newell, A. (1990) “*Unified theories of cognition*”, Harvard University Press, Cambridge, MA.

NLR (2000) “*Adaptive Cockpit Environment*”, memorandum VE-2000-002, Version 1.1, Nationaal Lucht-en Ruimtevaartlaboratorium, The Netherlands.

O”Donnell, R.D. and Eggemeier, F.T. (1986) “Workload assessment methodology”, in Handbook of perception and performance, Vol.2: Cognitive processes and performance, Boff, K. et al. (Eds), Wiley, New York.

Padfield, G.D. (1996) “*Helicopter flight dynamics: the theory and application of flying qualities and simulation modeling*”, Blackwell publishers.

Rasmussen, J. (1986) “Information processing and human-machine interaction: an approach to cognitive engineering”, Elsevier Science Publishers.

Rowe, D.W., Sibert, J. and Irwin, D. (1998) “Heart rate variability: indicator of user state as an aid to human-computer interaction”, in Proceedings of the Conference on Human Factors and Computing Systems (CHI’98), pp 480-487, Los Angeles CA, USA.

Svensson, E., Angelborg-Thanderz, M., Sjoberg, L. and Olsson, S. (1997) “Information complexity-mental workload and performance in combat aircraft” in Ergonomics, Vol. 40, No. 3, pages 362-380, March 1997, Taylor and Francis Ltd.

Taylor, R.M. (1990) “Situational awareness rating technique (SART): the development of a tool for aircrew systems design”, in *Situational awareness in aerospace operations* (AGARD-CP-478), pp. 3-1 / 3-17, Neuilly Sur Seine.

Wickens, C.D. (1992) “*Engineering psychology and human performance*”, Harper Collins Publishers Inc.

Zacharias, G.L., Miao, A.X., Illgen, C., Yara, J.M. (1996) “SAMPLE: situation awarenss model for pilot-in-the-loop evalution”, in *Proceedings of the 1st Annual Symposium on Situational Awareness in the Tactical Air Environment*, June 1996, Patuxent River MD, USA

Zijlstra, F.R.H. (1993) “*Efficiency in work behaviour: a design approach for modern tools*”. Delft University Press, The Netherlands.