first european annual conference
on human decision making
and manual control

25-27 may 1981
delft university of technology
proceedings
of the
european annual manual '81
Proceedings of the

first European annual conference
on human decision making
and manual control

held at the
DELFT UNIVERSITY OF TECHNOLOGY
May 25 - 27, 1981

Conference Chairman: Henk G. Stassen
Organisation: Wim L.Th. Thijs

Man-Machine Systems Group
Laboratory for Measurement and Control
Dept. of Mech. Engineering
Delft University of Technology
Mekelweg 2
2628 CD DELFT
Netherlands
This volume contains the proceedings of the First European Annual Conference on Human Decision Making and Manual Control, for short European Annual Manual, held at the Delft University of Technology at Delft, The Netherlands from May 25 - 27, 1981. It contains the complete manuscripts of the papers presented at the meeting. The papers are ordered as presented.

The European Annual Manual can be seen as an European version of the Annual Manual as held already for 17 times in the U.S.A. It is characterized as an informal and flexible conference, where a group of people, all workers in the Man-Machine Systems area can discuss about their problems. Different from the Annual Manual, U.S.A. where aircraft and spacecraft control plays such an important role, the European Annual Manual will deal to a larger extent with topics in the process industry, cardriving, manipulators and rehabilitation. Also the supervision of complex systems will be a very important topic.

The next European Annual Manual is planned at FAT, near Bonn, Federal Republic of Germany, in the summer of 1982.

Henk G. Stassen
Wim L.Th. Thijs
# Table of Contents

## Session 1

**Performance and Workload**

- **Prediction of Visual Search Performance Based on Visual Lobe Area Measurements**
  F.K. Kraiss, A. Knaeuper *
  
- **Visual Search: Relation Between Detection Performance and Visual Field Size**
  A.F. Korn
  
- **Effect of Learning, Age and Type of Shift on the Performance of a Critical Instability Tracking Task (CIT) by Busdrivers Before Their Work (A Preliminary Report)**
  
- **A Model of the Human Helicopter Pilot During I.L.S. Approach Design of a Display for Controls Operating**
  D. Diep, J. Papon, D. Viard

## Session 2

**Displays**

- **Displaying Process Structure**
  L. Bainbridge
  
- **How People Discover Input/Output Relationships**
  R.N. Pikaar
  
- **Coping with Complexity**
  J. Raemussen, M. Lind
  
- **The Structuring of Information on Visual Display Units**
  A.J.B. van Boxtel, C. Slappendel
  
- **Search Time and Color Code Size**
  G. Derefeldt, H. Marmolin
  
- **Ergonomics of Man-Machine Communication: Visual Performance Aspects**
  P. Haubner
  
- **A Device for the Measurement of Contrast Resolution, Spatial and Temporal Resolution by Means of a VDU Screen**
  P.W. Umbach, J.W.H. Kalsbeek, D. Bosman
  
- **Prediction and State in Simple Dynamic Systems: What Is Learned?**
  P.J. Vensmans

* The presentation was made by the author typed in italics.
# TABLE OF CONTENTS

## SESSION 3

**Supervisory Control, Decision Making, Fault Management**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Solving Behavior of Pilots in Abnormal and Emergency Situations</td>
<td>142</td>
</tr>
<tr>
<td>G. Johannsen, W.B. Rouse</td>
<td></td>
</tr>
<tr>
<td>Simulation of the Pilot's Long Term Strategy During IFR Flights</td>
<td>151</td>
</tr>
<tr>
<td>D. Sculatges</td>
<td></td>
</tr>
<tr>
<td>Experimental and Theoretical Analysis of Human Monitoring and Decision Making Behavior in Failure Detection Tasks</td>
<td>165</td>
</tr>
<tr>
<td>R.C. van de Graaff, P.H. Wewerinke</td>
<td></td>
</tr>
<tr>
<td>A Monitoring and Decision Making Paradigm: Experiments and Human Operator Modeling</td>
<td>180</td>
</tr>
<tr>
<td>W. Stein</td>
<td></td>
</tr>
<tr>
<td>Field Study of the Activities of Process Controllers</td>
<td>195</td>
</tr>
<tr>
<td>J. Queinnea, G. de Terssac, F. Thon</td>
<td></td>
</tr>
<tr>
<td>Modeling the Human Operator's Supervisory Behavior</td>
<td>203</td>
</tr>
<tr>
<td>T.N. White</td>
<td></td>
</tr>
<tr>
<td>Human Problem Solving in a Process Control Task</td>
<td>218</td>
</tr>
<tr>
<td>W.B. Rouse, N.M. Morris</td>
<td></td>
</tr>
</tbody>
</table>

## SESSION 4

**Manual Control Studies**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Research Needs in Manual Control</td>
<td>228</td>
</tr>
<tr>
<td>E. Edwards</td>
<td></td>
</tr>
<tr>
<td>Motivational Factors and Performance in a Manual Tracking Task</td>
<td>233</td>
</tr>
<tr>
<td>T. Bösser, E. Helchior, M. Schütte</td>
<td></td>
</tr>
<tr>
<td>On the Stability Problem of Human Arm and Hand Movements</td>
<td>243</td>
</tr>
<tr>
<td>Controlling External Load Systems</td>
<td></td>
</tr>
<tr>
<td>W. Setzer, G. Vossius</td>
<td></td>
</tr>
<tr>
<td>Effects of Visual and Vestibular Motion Perception on Control Task Performance</td>
<td>254</td>
</tr>
<tr>
<td>R.J.A.W. Honman, J.C. van der Vaart</td>
<td></td>
</tr>
<tr>
<td>Man-Machine Interaction in Aerospace Control System</td>
<td>275</td>
</tr>
<tr>
<td>D.R. Towill</td>
<td></td>
</tr>
<tr>
<td>Computer Assisted Manual Control of Cargo Handling with Ship Cranes</td>
<td>287</td>
</tr>
<tr>
<td>B. Schmidtbauer, S. Rönbäck</td>
<td></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

## SESSION 5
### Human Behaviour in Car Driving

**SIGNIFICANT CHANGES IN DRIVER-VEHICLE RESPONSE MEASURES FOR EXTENDED DURATION SIMULATED DRIVING TASKS**  
W.W. Wierville, W.H. Muto  
298

**DRIVERS' INTERNAL REPRESENTATION AND SUPERVISORY CONTROL: A FIRST MODEL VERIFICATION IN RELATION TO DRIVING EXPERIENCE TASK DEMANDS AND DETERIORATED VISION**  
G.J. Blaauw  
315

**PSYCHO-MATHEMATICAL MODEL OF VEHICULAR GUIDANCE BASED ON FUZZY AUTOMATA THEORY**  
U. Kramer, G. Rohr  
326

**LEVELS OF STEERING CONTROL; SOME NOTES ON THE TIME-TO-LINE CROSSING CONCEPT AS RELATED TO DRIVING STRATEGY**  
H. Godthelp, H. Konings  
343

## SESSION 6
### Miscellaneous

**THE EFFECT OF ELECTROTACTILE AND VISUAL FORCE FEEDBACK ON LEARNING AND PERFORMANCE IN A TELEMANIPULATION TASK**  
J.P. Gaillard  
358

**COMPUTER AIDED TREATMENT OF PATIENTS WITH INJURIES OF THE SPINAL CORD**  
376

**COMPETENCY AND OPERATOR TRAINING REQUIREMENTS IN PROCESS INDUSTRIES**  
J. Wirstad, H. Andersson  
389

**VOICE INPUT AND THE MEDICAL RECORD**  
H.C. Price  
403

## LIST OF PARTICIPANTS  
410
session 1
performance and workload

chairman: w.b. rouse, usa
PREDICTION OF VISUAL SEARCH PERFORMANCE BASED ON VISUAL LOBE AREA MEASUREMENTS

K.F. Kraiss, A. Knaeuper
Research Institute for Human Engineering (FAT)
D-5307 Wachtberg-Werthhoven, F.R. Germany

ABSTRACT

After a review of the pertinent literature, a novel procedure to predict search performance is presented which is based on visual lobe area measurements. In addition, the procedure distinguishes between random and systematic searching strategies.

A task network model has been worked out that permits to dynamically simulate visual search with various parameters of field size, visual lobe area and eye movement characteristics.

The validity of the proposed predicting method is demonstrated by comparison with experimental data as well as by simulation runs using the above mentioned model.

INTRODUCTION

The prediction of visual search performance in various applications has attracted considerable attention during the last years (ENGEL (1976), DRURY and CLEMENT (1978), KRENDEL and WODINSKY (1970)). A review of the current status of visual models is given, e.g., by MORAWSKI, DRURY and KARWAN (1980), and BLOOMFIELD (1975).

Most of these models are based on eye movement considerations only. In some papers an attempt has been made to explain differences in search performance which may be attributed to systematic or random search strategies.

On the other hand one can find in the literature a few papers dealing with the measurement of the so called visual lobe or conspicuity area (ENGEL (1976), WIDDEL and KASTER (1981)).

In this paper it is suggested that the extend and shape of the visual lobe area is one of the main factors influencing the overall search time in a particular task.

1. VISUAL LOBE AREA MEASUREMENTS

The visual lobe area is defined as the peripheral area around the central fixation point from which specific information can be extracted in a single glance. This region is affected by various parameters as, e.g., adaptation.
level of the eye, target characteristics, background, experience, and motivation. It is small, if the target is embedded in a complex background or surrounded by irregularly positioned non-targets of high similarity, and large, if simple stimuli stand clearly in front of a homogeneous background.

In the literature two methods to measure the visual lobe area are given. ENGEL (1976) describes a tachistoscopic approach. In this case the subject is asked to fixate in the middle of the screen. Subsequently, the target appears for a short time at various distances and directions in the periphery. After a series of presentations, the target detection probability can be calculated as a function of the distance from the fixation point.

WIDDEL and KASTER (1981) suggest a more realistic procedure which makes use of eye point of regard measurements. They make the assumption that at some fixation point during the scanning process the target is detected peripherally. Subsequently, the next saccade is voluntarily directed onto the target resulting in a fixation on the target. Therefore the fixation preceding target detection may be used to calculate the extend of the visual lobe area. After various presentations the visual lobe area may again be described as target detection probability over fixation distance.

Up to now a comparison of both methods has not been performed. Therefore it cannot be stated whether both yield the same results. From the data given by ENGEL, and WIDDEL and KASTER it may, however, be concluded that the visual lobe area in many cases can be approximated by a Normal Distribution as described by the following equation

\[ p(\delta, \phi) = p(0) \cdot e^{-\frac{\delta^2}{2\sigma^2}} \cdot k(\phi) \]  (1)

with \( p(0) \) = probability of target detection at a single glance during foveal inspection
\( \sigma \) = standard deviation
\( \delta \) = fixation distance from the target
\( \phi \) = fixation direction with respect to the target

In this equation \( p(0) \) is determined by one of the two methods described above by averaging over various subjects and experimental runs. In some cases, the visual lobe area appears to have an elliptic form. For simplicity, a circular visual lobe area, i.e., \( k(\phi) = \text{const.} \), is used in this paper to describe the probability to detect target in a distance \( \delta \) at a single glance. It has, however, to be emphasized that other shapes of the visual lobe area can be used as well if more appropriate.

2. PREDICTION OF SEARCH PERFORMANCE

2.1 Probability of Target Detection

As described above, the probability \( p(\delta) \) to detect a target by a single glance follows from equation (1). In the special case of foveal inspection we have \( p(\delta) = p(0) \). If \( p(0) < 1 \), the target region must be scanned several
times to guarantee detection. The contributions \( p_i(\delta) \) of various observations may be summed up yielding a parameter value \( n \cdot p \) which is later referred to as \( p_{\text{SUM}} \).

The value of \( p_{\text{SUM}} \) is, however, not equivalent to the resulting target detection probability \( P \). Since single glances are independent of each other the probability \( P \) that the target will be detected \( x \) times during \( n \) subsequent glances must be calculated using the Binominal Distribution

\[
P(x/p,n) = \binom{n}{x} p^x \cdot (1-p)^{n-x}
\]

with \( x = \) frequency of target detection in \( n \) fixations
\( p = \) probability of target detection in a single glance.

Hence, it follows that the probability to see a target with \( n \) glances at least once may be derived from equation (2) by setting \( x = 0 \), i.e., to see the target not at all, and subtracting this from unity.

\[
P = \sum_{x=1}^{n} P(x/p,n) = 1 - P(0/p,n) = 1 - (1-p)^n
\]  

For the special case where \( p << 1 \) equation (3) can be substituted by (Poisson Distribution)

\[
P = 1 - e^{-n \cdot p} = 1 - e^{-p_{\text{SUM}}}
\]  

According to the above equation (4), a value of \( p_{\text{SUM}} = 100 \% \) then will yield a target detection rate of \( P = 63.2 \% \). From equations (3) and (4) it appears that the probability of target detection in fact is not equal to the probability \( p_{\text{SUM}} = n \cdot p \) summed up from various glances as may have been supposed at the first instance. This fact is illustrated by figure 1 where \( P \) is shown to be depending not only on \( p_{\text{SUM}} \) but also on the value of \( p(0) \).

The probability \( p_{\text{SUM}} \) summed up for various observations is proportional to the search time needed for a certain detection probability \( P \) as will be shown later.

The upmost curve in figure 1 corresponds to the case where only one glance yields 100 \% target detection probability \( (p(0)=1.0) \). As may be seen, \( p_{\text{SUM}} = 100 \% \) corresponds directly to \( P = 100 \% \) target detection (see point S in figure 1).

The lowest curve in figure 1 is described by equation (4) and corresponds to the case where very little reconnaissance is gained from a single glance \( (p<<1) \), and hence, many subsequent observations of the target area are needed \( (n \rightarrow \infty) \). The target detection score corresponding to \( p_{\text{SUM}} = 100 \% \) is here only
Figure 1  Target detection probability $P$ resulting from the probability $P_{\text{SUM}} = n \cdot p$ summed up for $n$ subsequent glances on the target.

$(1-e^{-1}) = 63.2 \%$ according to equation (4), representing a distinct decrease in performance (see point R in figure 1). Trajectories for intermediate values of $p(0)$ are also depicted in figure 1.

2.2 Visual Scanning Behaviour

The number of fixations actually contributing to target detection depends on the ratio between interfixation distance $d$ and the size of the visual lobe area which in this paper is described by the standard deviation $\sigma$ and the amplitude $p(0)$ of a Normal Distribution. If a fixation happens to be too far away from the target, e.g., $\delta > 2\sigma$ its contribution is negligible. This is illustrated in some more details by figure 2 where a target is randomly positioned in an equidistant scan pattern. All fixations which may contribute to target detection, i.e., which lie within $2\sigma$ distance off the target are marked. Each fixation contributes to target detection corresponding to its actual distance $\delta_i$. If $n$ is the number of relevant fixations and $p(\delta_i)$ the target detection probability from a single glance we get

$$P_{\text{SUM}} = \frac{n}{i=1} p(\delta_i)$$

(5)
Figure 2 Regular quadratic scan pattern (interfixation distance d). The circle comprehends 5 fixations within 2\( \sigma \) distance to a randomly positioned target. Only these fixations contribute to target detection.

From figure 2 it becomes obvious that the level of \( p_{\text{SUM}} \) resulting from a regular scan is determined by the selected interfixation distance \( d \), as well as the standard deviation \( \sigma \) and amplitude \( p(0) \) of the visual lobe area. These interrelations have been calculated for the regular scan pattern of figure 2 using equation (5) for all fixations within \( \delta < 2\sigma \). The results are presented in figure 3. It may be seen that a \( \sigma/d \) ratio of .8 is needed together with a \( p(0) \) of .28 in order to achieve a \( p_{\text{SUM}} \) of 100 \%. Other scan patterns as, e.g., a hexagonal arrangement of fixation points have also been investigated and yield only slightly different data.

Obviously, the regular scan pattern depicted in figure 2 can only be observed in extreme cases of trained systematic search while usually a random procedure must be taken into account. During random search, successive fixations are independent and hence, overlap each other that means that many subsequent observations of the target area are needed until detection. This case is represented by the lowest curve in figure 1 and by equation (4).

3.3 Estimation of Search Time

As soon as the interfixation distance \( d \) that corresponds to a \( p_{\text{SUM}} \) of 100 \% has been determined from figure 3 (see dashed line A-B), the number of fixations \( N_{\text{FIX}} \) required for the regular scan of the searched field may be calculated:
Given the required number of fixations an estimate of the search time ($t_s$) can be determined using the following equation:

$$t_s = N_{\text{FIX}} \times (\bar{t}_{\text{FIX}} + \bar{t}_{\text{SAC}})$$  \hspace{1cm} (7)

with \( \bar{t}_{\text{FIX}} \) = mean fixation time,
\[ \bar{t}_{\text{SAC}} \] = mean saccade time.
A good estimate for the time needed for a fixation including a saccade ranges between 350 ms and 400 ms. These values are backed up by many different experiments.

The search time derived above is equivalent to the time constant during random search (compare equation (4)) and gives therefore an estimate of the time needed to find 63.2% of the targets. The improvement that could possibly be reached by a trained systematic search strategy is entirely depending on the visual lobe area characteristic p(0) and may be looked up from figure 1.

3. SIMULATION OF SEARCH BEHAVIOUR

In section 1 and 2 a theoretical approach for the prediction of visual search performance was presented. Subsequently, a computer simulation model is developed that is used to confirm the prediction procedure described above.

For the model formulation a network approach seemed to be most appropriate. As a tool for network synthesis the simulation language SAINT (Systems Analysis of Integrated Networks of Tasks) was used. For details concerning the symbology and network synthesis the reader is referred to CHUBB (1981).

3.1 Network Model for Random Search

This network (figure 4) consists of eleven tasks with the following labels

START, STOP, STRTCOOR, SACDIREC, SACAMPLI, FIXCOORD, SACDURAT, FIXDURAT, TRGTDIST, DETECTPR, PRESSBUT

Main functions of each of these tasks are described below.

START : TASK 1 starts the simulation run for a predetermined time, e.g., 100 s, and if required it collects data for plotting eye movements.

STOP : TASK 2 stops the simulation run after completion if the simulation is not stopped by TASK 1 which signals the detection of target.

STRTCOOR : Target and start coordinates are drawn from a uniform distribution within the limits of the searched field. The time of this first fixation (in s) is drawn from a Beta Pert Distribution \( B(m,a,b) \), approximated to experimentally determined data from WIDDEL and KASTER (1981). The parameters of this distribution are:

\[
\begin{align*}
m &= .2 & \text{(most likely value)}, \\
a &= .02 & \text{(optimistic (smallest) value)}, \\
b &= 1.0 & \text{(pessimistic (largest) value)}. \\
\end{align*}
\]

After completion there is a deterministic branch to TASK 4.
SACDIREC : A random direction for the next saccade is selected in TASK 4. Task duration is 0 seconds. After completion a deterministic branch to TASK 5 is made.

SACAMPLI : The saccade amplitude (in deg) is also drawn from a Beta Pert Distribution which is approximated to experimental data from WIDDEL and KASTER (1981) as above. The parameters are:

\[ m = 1.5, \quad a = 1.5, \quad b = 23.0. \]

This also takes no time and a deterministic branching is performed to TASK 6.

FIXCOORD : A new fixation point is computed using the last fixation coordinates, the saccade direction determined in TASK 4 and the saccade amplitude determined in TASK 5. It is checked whether the new point is within the screen. If not tasks 4, 5, and 6 are repeated in order to determine a new fixation point. Subsequently, a branch to TASK 7 is performed.

SACDURAT : TASK 7 yields the saccade duration using the relation \( t_{\text{SAC}} = ((SA-1) \cdot 0.04) \) s (SA: saccade amplitude in degree) which is
in accordance with experimentally determined data (WIDDEL and KASTER (1981)), but which is only valid for small saccades. For large saccades (> 6 degrees) a maximum duration of .2 s is assumed followed by a deterministic branch to TASK 8.

**FIXDURAT** : TASK 8 yields the fixation duration which is drawn from the same distribution as in STRTCOOR.

**TRGTDIST** : The distance between the last fixation point and the target (determined in STRTCOOR) is computed. After task completion a deterministic branching to TASK 10 is performed.

**DETECTPR** : TASK 10 simulates the visual lobe area as described in section 1. If the target is detected branching is performed to TASK 11, otherwise to TASK 4 where a new fixation point is computed.

**PRESSBUT** : TASK 11 simulates target recognition (signaled by pressing a button) and stops the simulation run. Simultaneously, the search time is registered.

### 3.2 Network Model for Systematic Search

This network, as depicted in figure 5 consists of eight tasks with the following labels

- **START, STOP, STRTCOOR, FIXATION, SACCade, TRGTDIST, DETECTPR, PRESSBUT**
- **START** : same function as above
- **STOP** : same function as above

![Figure 5 Network model for systematic visual search](image-url)
STRTCOOR : Target coordinates are drawn from a uniform distribution as above. The start coordinates are selected by the modeler (e.g., the search begins at upper left corner of the screen).

FIXATION : The sequence of fixations is controlled in such a way that a regular scan pattern with interfixation distance d covering the whole searched field is generated (as depicted in fig. 3). The time for TASK 4 (one fixation) is derived from the experimental distribution of fixation duration (as above). A deterministic branching to TASK 5 is performed.

SACCADE : TASK 5 yields the saccade duration corresponding to the interfixation distance d followed by a deterministic branch to TASK 6.

TRGTDIST : same function as above

DETECTPR : same function as above

PRESSBUT : same function as above.

3.3 Simulation Results

Using the networks described above, the effect of various visual lobe areas on search performance can be tested analytically. In order to do so, two shapes have been selected for the visual lobe area as depicted in figure 6. In addition, an extend of 30° 25 degrees was assumed for the searched field.

As a first step an estimate is made about the time constant for random search using visual lobe area No. 1. As explained earlier a value of at least 63.2% target detection (time constant for random search) is achieved as soon as the summed up detection probability from all fixations is \( p_{SUM} = 100 \% \) all over the searched field. This is the case for a \( \sigma/d \) ratio of .8 and an interfixation distance \( d = 4.4 \text{ deg} \) as derived from figure 3 (marked point B). Using equations (6) and (7) this corresponds to a number of \( N_{FIX} = 56 \) fixations and a search time of \( t_s = 22 \text{ s} \).

When performing the same calculations for visual lobe area No. 2, the interfixation distance is \( d = \sigma/.45 = 4.4 \text{ deg} \) (figure 3, point A), i.e., the same as above, yielding an identical result for random search. Hence, the different shape of the visual lobe area should not influence random search performance at all. As indicated in figure 7 this can be justified by network simulation runs where about identical curves for random search (RANDOM 1,2) can be seen for both visual lobe areas.

As theoretically predicted a systematic search strategy doesn't improve performance for visual lobe area No. 1 where \( p(0) \) reaches only a value of .275 as compared to random search (compare also figure 1).

In contrast to this finding the results for systematic search show marked differences for the second visual lobe area characteristics. For the \( p_2(0) = 1.0 \) case a target detection rate of 86 % at point \( P_{SUM} = 100 \% \) can be observed from the simulation run. This is less than the theoretically expected 100 % (see point S in figure 7). The difference may be explained by less than perfect coverage of the searched area during the simulations.
Figure 6  Two arbitrarily selected visual lobe areas for the network simulation runs. Parameters are $p_1(0) = 0.275$, $\sigma_1 = 3.5$ deg and $p_2(0) = 1.0$; $\sigma_2 = 2.0$ deg.

Figure 7  Simulated search performance for random and systematic search strategies and the visual lobe areas of figure 6.
The deviation may also be due to the finiteness of the random test and the unsystematic coverage at the edges of the searched field. In principle however, the analytical results strongly support the theoretical findings.

In addition to the example presented above, simulations with other visual lobe areas and interfixation distances \( d \) were made. All of these experiments confirmed our theoretically expected data. The described network models may therefore be considered as a valid representation of visual search strategies.

4. COMPARISON WITH EXPERIMENTAL DATA

In order to perform a further validation theoretical search performance predictions are subsequently compared with experimental data as presented by WIDDEL and KASTER (1981), and ENGEL (1976).

In the experimental set-up used by WIDDEL and KASTER the subjects had to scan a screen with an extend of 30 x 25 deg on which the visual search material was presented consisting of a number of uniform randomly spread symbols. These symbols were identical with the exception of one, the target which had to be detected by the subjects. The target was positioned randomly within the stimulus field during each presentation. A systematic variation of the difficulty of visual perception was obtained by increasing the number of background symbols. The subjects were randomly assigned to two experimental groups who had to take part in eight sessions with fifteen presentations. One group was trained with a scan line which was moving over the screen in order to stimulate systematic search while the other group could apply a free individual scan.

The authors present experimental data for the visual lobe area in three conditions. These data were in each case approximated by a Normal Distribution resulting in the \( p(0), \sigma \) parameters given in table 1.

<table>
<thead>
<tr>
<th>Field Size (deg)</th>
<th>Number of Items</th>
<th>Visual Lobe Area Parameters</th>
<th>Interfixation Distance (deg)</th>
<th>Number of Fixations for ( P_{SUM} = 1.0 )</th>
<th>Fixation Duration (incl. Saccades) (ms)</th>
<th>Time to Search Field Once (s)</th>
<th>Detection Rate in ( t_9 ) (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>352</td>
<td>.275/3.5</td>
<td>4.4</td>
<td>56</td>
<td>393</td>
<td>22.0</td>
<td>63.2(39-83)*</td>
</tr>
<tr>
<td>x</td>
<td>467</td>
<td>.224/3.5</td>
<td>3.8</td>
<td>63</td>
<td>397</td>
<td>25.0</td>
<td>63.2(39-83)*</td>
</tr>
<tr>
<td>25</td>
<td>576</td>
<td>.184/3.5</td>
<td>3.3</td>
<td>80</td>
<td>363</td>
<td>29.0</td>
<td>63.2(39-83)*</td>
</tr>
</tbody>
</table>

* confidence interval for \( N=20 \)
Using the field size and the visual lobe areas parameters the number of fixations needed for $p_{SUM} = 100 \%$ can be calculated using figure 3 and equation (6). The time $t_s$ needed to search the field once results from a multiplication with average fixation times. Subsequently, from figure 1 predictions for the target detection rate may be derived assuming random or systematic search strategy.

For comparison the last column for table 1 gives a range of experimental data as determined in various experimental conditions. Since these data are in each case based on only 20 single measurements they should be compared with the corresponding confidence intervals which are also given in table 1.

From the data presented in table 1 it appears the search performance predictions are in line with the experimentally determined times. In addition, it is again shown that not much improvement can be expected from a systematic search strategy as long as the $p(0)$ value of the visual lobe area is considerably lower than unity.

In the paradigm used by ENGEL (1976) the subjects had to search a screen of 22.3 deg $\times$ 16.8 deg of visual angle. The screen contained a random dot pattern of 220 similar disks with a diameter of .55 deg and one dissimilar test object, the target which differs from the background disks in diameter. The minimum centre-to-centre distance was 1.5 times the diameter of the background disks in order to prevent overlapping in the stimulus pattern. The stimulus field was presented 4 s in each search experiment and the observer had to indicate the discovery of the test object by means of a light signal.

Table 2 gives a comparison of theoretically predicted and experimentally determined data in much the same way as was demonstrated for table 1.

Table 2 Comparison of theoretical and experimental results  
(averages of experimental data from ENGEL (1976))

<table>
<thead>
<tr>
<th>Field Size (deg)</th>
<th>Experimental Condition</th>
<th>Visual Lobe Area Parameters</th>
<th>Interfication Distance (deg)</th>
<th>Number of Fixations for $P_{SUM}^{*1.0}$</th>
<th>Fixation Duration (incl. Saccades) (ms)</th>
<th>Time to Search Field once (s)</th>
<th>Detection Rate in $t_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.3</td>
<td>0.08</td>
<td>1.0/2.0</td>
<td>4.4</td>
<td>30</td>
<td>330</td>
<td>9.9</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.2(48-78)*</td>
</tr>
<tr>
<td>x</td>
<td>0.10</td>
<td>1.0/2.6</td>
<td>6.3</td>
<td>20</td>
<td>340</td>
<td>6.8</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.2(48-78)*</td>
</tr>
<tr>
<td>18</td>
<td>0.13</td>
<td>1.0/4.8</td>
<td>11.7</td>
<td>9</td>
<td>330</td>
<td>2.98</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.2(48-78)*</td>
</tr>
<tr>
<td>grid</td>
<td>0.21</td>
<td>1.0/6.5</td>
<td>15.8</td>
<td>6</td>
<td>310</td>
<td>1.86</td>
<td>100.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>63.2(48-78)*</td>
</tr>
</tbody>
</table>

* confidence interval for N=48  
** extrapolated values  
D : diameter of background disks  
Do : diameter of test disk
It may be seen that the actual search performance ranges for all of the four indicated experimental situations between the values predicted for random and systematic strategy. Extreme values for systematic search, i.e., 100 % do not occur. However, the increase in performance as compared with the random approach is for some cases considerable reaching up to 83 % due to the high value of \( p(0) = 100 \) %. Thus, also for this example visual performance is correctly predicted when the proposed procedure is applied.

5. CONCLUSION

In this paper a method has been described to predict visual search performance using data on the visual lobe area and on eye movement times. These data must be made available by searching the literature for appropriate data or by special experimentation.

As soon as these parameters are known a reliable prediction of search performance can be performed in a simple straightforward way. The described method is superior to other models published in the literature in that it discriminates various strategies of visual search, i.e., random and systematic. It is shown that the systematic procedure does not in any case lead to an improvement in performance but only in cases where detection probability \( p(0) \) is high for single glances.

While in this paper only simple search situations have been considered it is expected that this method can also be applied in more complex applications where the visual lobe area changes within one search field or where various targets appear.

To support the theoretical findings a task network model has been worked out that permits to dynamically simulate visual search with various parameters of field size, visual lobe area and eye movement characteristics. This model will be used for the further analysis of search behaviour.

6. REFERENCES


VISUAL SEARCH : RELATION BETWEEN DETECTION PERFORMANCE AND VISUAL FIELD SIZE

A.F. Korn
Fraunhofer-Institut für Informations- und Datenverarbeitung
D-7500 Karlsruhe 1, FRG

Abstract

We are considering the following situation: a remotely controlled viewing device produces a visual presentation of a remote scene. This display provides the same degree of visual presence to a remote observer as if he were viewing the scene directly. In order to minimize required operator command and data link bandwidth, the field of view should be an optimum. The following experiments were performed to investigate this problem.

In a visual search experiment a single vehicle was to be found among confusing objects consisting of various components of aerial pictures. Here typical sceneries have been simulated: a forest, a meadow, a village, and an open field.

The visual search time was measured as a function of the visual field size that was varied systematically. The field could be shifted across the TV display by eye and head movements or by hand (joystick). In the first case the coordinates of the corneal reflex were used to adjust the position of the visual field concentrically around the fixation point.

Decreasing the visual field size from 24°, which is the size of the TV display, to 3° viewing angle, which is the threefold size of the vehicle, results in an increase of processing time by a factor 10. There are no significant differences between shifting the picture-window by eye, head, or hand movements.

The results can be interpreted in terms of one of the major functions performed by peripheral vision, namely the guidance of eye movements and orientation in space.

Analysis of the different scanpaths gives some hints at the size of conspicuity areas.

Introduction

In RPV (remotely piloted vehicle) reconnaissance missions TV operators have to interpret the transmitted data on-line. Viewing time is physically determined by the flight parameters and the optical parameters of the sensor. The human factors aspects of the electro-optical system design are concerned primarily with the interaction between the displayed material and the observer's visual and visual perception systems.

The detection performance depends upon (Ref.1):

- System parameters such as resolution and the size of the area to be covered.
- Scene parameters which constitute the visual conspicuity of targets in a natural scene such as size, shape, and contrast of a target or global...
and local context, i.e. the immediate and total surroundings.

- Observer-related factors such as modulation transfer function of the visual system, luminance adaption, oculomotor response (eye movements, scanning behaviour), reaction time, the amount of information which can be accepted from individual fixations (glimpses), stress placed on the observer by the physical environment or nature of the task, and the observer's training and skill.

In order to maximize the target acquisition process a very large number of picture points should be transmitted to the ground station to obtain all the information which is needed by the observer during the search process. In general, the literature indicates that display field of view sizes greater than 30 degrees but less than 60 degrees are desirable (Ref.2). A 40 degree field of view with a spatial resolution of 60 lines per degree, a 8-bit grey level resolution, and a frame rate of 25 pictures per second require a data transmission rate of 1 150 Mbit per second. Because of technical difficulties and jamming sensitivity the required high transmission rate is not desirable.

Considering the limited capabilities of the human visual system one recognizes that only a small fraction of this information can be accepted from the search scene at any individual glimpse. Therefore a satisfactory solution of the problem of bandwidth reduction without loss of relevant information can be achieved by transmitting only a small picture area around the observer's fixation point. A further reduction of bandwidth is possible by the variation of the spatial resolution in that area according to the properties of the human retina (Ref.3,4). The retina is characterized by a central region of high resolution (fovea) with gradually reduced resolution as distance from the fovea increases.

This paper is concerned almost exclusively with the experimental question how much of the field of view is needed at each fixation pause in order to get a similar detection performance as in the case of non-restricted visual field. The average length of fixation pauses are about 1/3 second each, the eyes being fixed about 90 % of the time where involuntary eye movements with small amplitudes (< 1 degree) are not considered here.

There have been several works on the determination of the size of this field of view for a single fixation (e.g. Ref.5-7). This area around the fixation point from which usable information can be extracted depends upon the task, the kind of target and background objects, and the picture quality. It is often called conspicuity area which is described as "the retinal field in which the relevant object can be discovered (without prior knowledge of location) in its background, during a brief presentation of the stimulus pattern" (Ref.5). In Ref.7 one can find the term "useful visual field of view" with a similar meaning.

The aim of the present experiment is to determine the size of conspicuity areas when subjects are searching for a vehicle in a natural scene. A special apparatus by which a visual field with adjustable size could be shifted by eye, head, or hand movements was employed. With such a device the hypothesis can be tested that the detection time in complex scenes such as aerial pictures, displayed on a TV screen, is essentially determined by properties of the visual and cognitive system and not so much by the kind of muscles involved in shifting the visual field on the display.
GENERAL METHOD

Stimuli
Static scenes were taken as slides from a 1:87 terrain model with four summer-sceneries: forest, meadow, field, village (see Fig. 4). A scene consisted of one of these sceneries and covered a ground area of 130x130 meters at a simulated altitude of 300 meters with vertical aspect angle. The slides were taken from diffusely illuminated scenes. The targets were military vehicles of 4 types: 2 different types of tanks and 2 different types of trucks. Each scene contained only one military vehicle as a target which must be detected.

Apparatus
The slides were either scenes or a neutral pause picture. They were scanned by a TV camera and presented on a standard 625-line TV monitor with 30 cm screen size. The subjects used a bite board to immobilize their head. The horizontal viewing angle was 24 degrees. The scheme of the experimental arrangement is shown in Fig. 1. The picture on the TV monitor appeared to the subject only partly within a small square whose position coincided exactly with his visual axis when eye movements are used to change the position of the window as indicated in Fig. 1.

Fig. 1: Schematic view of the apparatus to restrict visual field size and to change the position by eye movements.

Fig. 2: Device for analyzing a video signal and the generation of different patterns.

The twodimensional eye movements of the subject's left or right eye were detected by the corneal reflection method using a silicon TV camera for registration of the IR-light which is reflected from the subject's eye and an IR-mirror (Ref.8). The subject could adjust the oculometer by himself because he observes his own corneal reflex on the TV monitor. Looking at the center of the display he changes the position of the bite board until the position of the reflex and the center of the display coincide. The output of the TV camera is fed to an electronic device where the x-, y-coordinates of the corneal reflex are extracted out of the video signal. This device also contains a function generator which generates a square of adjustable size the corneal reflex being the center of the square. The scheme of this electronic device, which is called Video-Analyzer and -Generator, is shown in Fig. 2. Inputs from a TV camera, a videorecorder, a lightpen, a joystick, and from a computer are established. The outputs can be fed to an oscilloscope, a monitor, a plotter, or a computer. We did use
the x-, y-analog output signals for changing the position of the square on the monitor.

In addition to the device depicted in Fig. 2 an electronic circuit was used to perform the transformation

\[
x' = (x - x_M) \cdot A \\
y' = (y - y_M) \cdot B
\]

of the measured eye position \(x, y\) to the desired position \(x', y'\), which is the fixation point in an equally spaced grid. The coordinates \(x_M, y_M\) give the measured eye position when the eye fixes the center of the monitor and the displayed corneal reflex coincides with the centerpoint. A and B are constants for the adjustment.

As mentioned earlier, the position of the visual field could also be changed by head movements. The appropriate scheme of the experimental arrangement is shown in Fig. 3. A light source which emits a parallel light beam is closely connected with a flexible ring around the subject's head. The beam is projected on a diffusing screen. The position of the corresponding light spot is measured by an electro-optical device which is nearly identical to that shown in Figs. 1 and 2. The amplitudes of the x-, y-signals, i.e. the output of the video analyzer, are proportional to the rotation angles of the head around a vertical and a horizontal axis for the small visual angles under consideration.

Finally a change of the position of the visual field was achieved by manual control with a joystick. The direction of the hand movements corresponds exactly to the motion direction of the displayed visual field.

![Diagram of the experimental setup](image)

**Fig. 3:** Schematic view of the apparatus to restrict visual field size and to change the position by head movements.

**Fig. 4:** Open field with eye movement traces (see text).
Subjects and Procedure

The subjects were 11 untrained male school boys and college students with a Snellen visual acuity of 1 or better. Four different visual field sizes were employed: 2.25 x 3 deg., 4.5 x 6 deg., 9 x 12 deg., and non-restricted size. The first three covered 1.56 %, 6.25 %, and 25 % of the area of the entire picture. The length of the largest target (10 t truck) was about 1.5 deg., i.e. half of the smallest visual field size.

At the beginning of each session the subjects read instructions indicating the nature of the detection task. They familiarized themselves with models of the four targets. A series of 10 training scenes was shown to each subject with feedback. The subjects were instructed to finish the search process only if they were quite sure in the detection of a military vehicle whose position they were to indicate. The maximum viewing time was restricted to 100 sec. Each slide was presented to a subject four times corresponding to the four different fields of view. Memory effects were avoided by turning the slides. The investigated dependent variable was the search time for a correct detection which was measured as a function of the visual field size.

RESULTS

To analyze the eye and hand movement behaviour the scanpath has been recorded for each visual field size. In Fig.4 the record of eye movements with a x-,y-plotter is superposed on the picture of an open field. Here the visual field size of 18 x 24 deg. was not restricted. Starting point for all presentations was the center of the picture. The target in Fig.4 is a tank in the lower right corner which has been detected after 6 sec. The reason for this rather long detection time is the low contrast and the position in the picture periphery. The search pattern clearly indicates that visual conspicuities such as trees and bushes are mainly the goal of saccadic movements. The sequence of fixations or glimpses is apparently determined by the structure of the scenery.

Fig.5 shows the frequency distribution of the saccadic sizes for a non-restricted visual field size. This distribution has been obtained in previous experiments with the same scenes and targets (Ref.9). 20 % of all saccadic sizes are larger than 10 deg. The median is about 5 deg. Such distributions yield a very crude estimation of the size of conspicuity areas.

Two examples of eye movement records are shown in Fig.6 for two different visual field sizes. Here a tank must be detected in an open field. For a 2.25 x 3 deg. visual field size in Fig.6a the detection time was 100 sec. For a 4.5 x 6 deg. visual field size, the same scenery, and the same subject the detection time was 16 sec. The corresponding scanpath is shown in Fig.6b. Since peripheral information has been reduced the subject must choose a systematic scanpath. Otherwise the orientation in the scene is very hard.

A systematic search can be performed easier by head or hand movements. Fig.6c shows an example for the shift of the visual field by head movements. Here a systematic path for the 3 deg. visual field size is perceptible. Nevertheless the subject couldn't find the target in 120 sec, i.e. a tank in a village. In 6d) head movements are shown for a non-restricted search in the
Fig. 5: Distribution of the relative frequency of saccadic eye movement sizes.

same scene. The detection time was 28 sec. This record shows very clearly that head movements can be very large during a search in natural scenes in spite of the relatively small display size of 24 deg. visual angle. Fig. 7 shows the detection time as a function of the visual field size for different modes of control. The large variability of the detection time can be explained by the very different types of sceneries and differences between the detection performance of the subjects.

Detection time decreases with increasing visual field size. The average detection times for the 3 deg. visual field size and non-restricted search differ by a factor of about ten. No significant dependency upon the mode of control could be found. In other words, if a

Fig. 6: In a) the scanpath of eye movements is recorded for a 3 deg., and in b) for a 6 deg. visual field size. In c) and d) records of head movements are shown for 3 deg. visual field size and for a non-restricted search (see text).

Fig. 7: Detection time as function of the visual field size for different modes of control.
visual search is performed with a restricted visual field which can be shifted by eye, head, or hand movements, the mechanism of motor control doesn't play a significant role for the detection performance.

Discussion

The rapid increase of the difficulty of the search task for visual field sizes smaller than 8 deg. shows important peripheral characteristics such as the guidance of eye movements and orientation in space.

Concerning the different types of sceneries no significant differences of the detection time could be found for the village, meadow, and forest scenes. Only for the search within an open field scene, using the 6 deg. and 12 deg. visual field size, the detection time was significantly shorter.

Fig.7 shows that the detection time for the non-restricted search (24 deg. visual field size) is shorter than the detection time for the 12 deg. visual field size. The size of the conspicuity area is obviously at least 12 deg. Taking into account the slope of the curve a cut at a 8 deg. visual field size appears meaningful. This is five times the length of the largest target dimension (truck). For the identification the width of the truck, which is assumed to be one third of the length, should be 10 TV lines. Therefore a suitable visual field size should contain about 150 TV lines for the scenes considered in this paper.

The records of eye and head movements show a systematic search strategy for small visual field sizes. The performance of head movements was easier than the performance of eye movement because of the better proprioceptive feedback. The large amplitudes of head movements during the non-restricted search were rather unexpected.

References


Acknowledgement: This work was supported by the Ministry of Defence.
EFFECT OF LEARNING, AGE AND TYPE OF SHIFT ON THE
PERFORMANCE OF A CRITICAL INSTABILITY TRACKING
TASK (CIT) BY BUSDRIVERS BEFORE THEIR WORK
(A PRELIMINARY REPORT)

Netherlands Institute for Preventive Health Care/TNO
P.O. Box 124, 2300 AC Leiden - The Netherlands

Abstract
Performance on a critical tracking task (CIT) was measured for 6 'young' sub­j ects (26 - 32 years old) and 6 'old' subjects (53 - 58 years old) before the
beginning of their work at different times of the day. Also oral temperature
was measured for the same subjects minus 2 'old' ones, at about the same time.
These measurements were carried out in the context of a broader project, fo­cusing on the possible effect of his task on the busdriver.
The younger subjects show an overall better performance than the older ones.
Both age-groups show a better performance in the late afternoon (+ 15.40 hrs.)
than in the early morning (06.15 and 06.40 hrs.).
A continuing learning effect for at least 7 experimental days appears to be
present.
In the late afternoon, oral temperature in 'young' and 'old' subjects is higher
than in the early morning. This increase in oral temperature and in the per­formance on the CIT over the same part of the day, suggests a sensitivity of
this task to the level of activation of the subjects.
The results have implications for the use of the CIT in experimental practice.

Keywords: Age, busdriver, circadian rhythm, compensatory tracking, learning,
oral temperature, shift work, tracking-task.

Introduction
The data presented in this paper are some preliminary results of a more exten­sive project which is in progress since 1977 at the Netherlands Institute for
Preventive Health Care/TNO in Leiden - The Netherlands.
The general purpose of this project is: The development of a measuring instru­ment capable of demonstrating the (assumed) effect of the performance of a task
on a task-performer and, assuming that this would prove possible, to ascertain
to what extent the task effect changes under different task conditions.
In order to look for a possible effect of a task on a task-performer the bus­
driver task was chosen.

The more detailed information about the total project is useful because it
gives the background for the data presented in this paper.

Measurements performed in this project can be roughly divided into two cate­gories namely measurements with respect to the task-performer (physiological
and psychometric) and with respect to the bus (speed and steering-wheel move­ments). For a better interpretation of the data, also events in and around
the bus were observed during the driving task performance. Besides this, an
analysis on bus-accident data was performed.
The measurements on the individual were carried out in a kind of mobile laboratory, in a specially equipped van, during 'rest periods' before, during and after the normal task of the busdriver. During the driving the measurements on the bus were carried out (Fig. 2).

**Fig. 1.** Scheme of the total project.

**Fig. 2.** The different measurement situations in the course of an experimental day, before, during and after the service.
Type of shift was treated as an experimental variable in the project. In the work schedule of the bus company concerned, three different types of shift (Fig. 3) can be distinguished:

1) Early shift: a shift with a mean duration of 8 hrs., starting at about 6.40 hours.
2) Late shift: a shift with a mean duration of 8 hrs., starting at about 15.40 hours.
3) 'Broken' shift: a compound shift, consisting of two parts. The first part starting at about 6.15 hours and lasting for about three hours, the second part starting at about 13.15 hours and lasting for about five hours.

Another experimental variable was the 'driving situation'. The above mentioned shifts were driven as well under predominantly 'urban' as under predominantly 'rural' conditions. Additionally, a control situation under work-free circumstances was created. Measurements in this situation were carried out on the subject while they were not working, but stayed at home.

The last experimental condition was the age of the subjects. The subjects were male professional busdrivers with a task-experience of minimal two years. From those busdrivers alternatingly the oldest* and the youngest subject available was chosen. In this way a total of 8 'old' and 8 'young' subjects were obtained for the total project.

The experimental design thus includes three types of shift, two driving situations and two age-groups. For reasons of reliability the measurements were repeated once for each of these conditions (Fig. 3).

The schedule of measurements was therefore \((3 \times 2 + 1) \times 2\) for two age-groups of eight subjects each. This means that each subject was measured during 14 days. These days were preceded by a training-day (+ 8 hours), during which the subject became acquainted with the measuring situation in the mobile laboratory and had to learn a number of psychometric tasks to be performed during the experiment. The 15 experimental days were completed during the workdays of 3 successive weeks.

\[ \begin{align*}
4 & \quad 8 & \quad 12 & \quad 16 & \quad 20 & \quad 24/0 & \quad 4 \\
\text{EARLY - URBAN} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{EARLY - RURAL} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{LATE - URBAN} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{LATE - RURAL} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{"BROKEN" - URBAN} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{"BROKEN" - RURAL} & \quad & \quad & \quad & \quad & \quad & \quad \\
\text{"DAY-OFF"} & \quad & \quad & \quad & \quad & \quad & \quad \\
\end{align*} \]

Fig. 3. Experimental design for one subject.

* With the restriction that they had to be younger than 60 years. Busdrivers above 60 years old are placed into special service-schedules.
In the research strategy for this project it was assumed that the state of the subject changes under the influence of an occupational task to be performed by this subject. Further it was assumed that the effect of a task performance on a task-performer is a cumulative one.

These considerations led to the decision to measure the supposed effect of the driving task on physiological and psychometric variables not during the driving task itself but before the driving task, during some of the rest periods within the driving task and after the end of the service (viz. Fig. 2).

From the different measurements, both psychometric and physiological that were carried out in the mobile laboratory, at this moment only a part of the tracking performance data and the oral temperature values, both taken before the start of the service of the busdriver, are available.

The critical instability tracking task (CIT) was originally developed by Jex et al (1966). It is a compensatory tracking task based on a first order divergent control element. The subject has to control the unstable system as long as possible. He uses some controlling device for correction of the error induced by the controlled element. The system instability is determined by the value of \( \lambda \), i.e. frequency in radians per second. This task can be used at least in two different ways. First: the difficulty of the task increases automatically. This means that at the beginning of the task performance \( \lambda \)-value equals zero and becomes gradually larger until some point where further control is rendered impossible. The value of \( \lambda \) at the moment that control is lost is the so-called critical \( \lambda (\lambda_c) \). A higher \( \lambda_c \)-value represents therefore a better tracking performance of the subject. Second: the difficulty of the task is constant, with tracking performance measured for a fixed \( \lambda \)-value.

More specific technical information about the CIT as it has been used in this experiment is given by Soede (1980).

**Material**

For this preliminary report data from 6 'old' subjects (53 - 58 years old; mean age 55 years) and 6 'young' subjects (26 - 32 years old; mean age 28 years) are available. With respect to the oral temperature measurements, the data from 2 'old' subjects are missing.

In order to know which fixed value of \( \lambda \) had to be given to a certain subject on a certain day each experimental day was started with a so-called maximum-estimation.
Presented data with respect to the CIT are the means of four successive $\lambda_C$-values* (= $\bar{\lambda}_C$) carried out before the beginning of the first measurement period in the mobile laboratory. (In Fig. 2 before the first 'R'-block.)

The subject had the task of keeping a spot on a vertical LED-bar display on a blackline in the middle of the display frame** (Fig. 4). The subject used a joystick (force type) to perform the control task. When the spot moved upwards he had to pull the joystick towards himself, when the spot moved downwards he had to push the joystick in the opposite direction. At the moment the subject lost control the lighted spot disappeared from the display.

The maximum-estimations were carried out before the 'broken' shift (+ 06.15 hour) the early shift (+ 06.40 hour) and the late shift (+ 15.40 hour). The oral temperature measurements (in °C) are performed by means of a standard clinical disposable thermometer, about ± 3 minutes after the CIT maximum-estimations.

Statistical analysis

The mean of the 4 $\lambda_C$-values per day (= $\bar{\lambda}_C$) and oral temperature data were analyzed using analysis of variance***. The experimental variables in this analysis were:

a) 'age': old, young
b) 'shift': 'broken', early, late****
c) 'different days of the experiment' (12 days)

Additionally, with regard to the estimation of a possible effect of training on the value $\bar{\lambda}_C$ by 'old' and 'young' subjects an analysis of variance was also applied to the experimental variable 'successive period'. Each of the 4 'successive periods' consists of a 'broken' an early and a late shift, each in a different order.

The experimental variable 'driving situation' is most likely not relevant for data presented in this paper. The maximum-estimations of the $\lambda_C$ and the oral temperature measurements took place before the subject actually started his work, always at the same location, viz. the garage of the bus company. The data from the work-free days were not analyzed because the measurements were not followed by a work situation.

Results

Performance data

One of the experimental variables was 'age'. Fig. 5 and Fig. 6 give the mean values of the $\bar{\lambda}_C$ per experimental day sepa-

* For the rest of the day the subject received 85% of the highest $\lambda_C$-value ($\lambda_{C\text{ max.}}$) as a fixed task-difficulty. The predictive value of the $\bar{\lambda}_C$ for the value of the $\lambda_{C\text{ max.}}$ appears to be quite accurate.

** Technical realisation of the CIT: Department of Physics NIPG/TNO.

*** Program name: VARIAN. Orginal authors: Kwaaitaal and Roskam (University of Nijmegen, NL). Rewritten by J. Gerkema (NIPG/TNO Leiden, NL).

**** The experimental variable 'shift' is nested under the variable 'different days of the experiment'.
rately for 'young' and 'old' subjects. A higher value means a better performance. On the abscissa in Fig. 5 the successive days of the week are shown, including the weekends and the work-free days. For a clear presentation of the data, the data-points which belonging to the same experimental variable are connected.

Fig. 5. Mean values of the $\overline{\lambda}_c$ per age-group shown separately for the total duration of the experiment.

Fig. 6. Mean values of the $\overline{\lambda}_c$ per age-group.
On the abscissa in Fig. 6 the time-scale was condensed somewhat and only the data for the days on which the subjects were actually driving are presented. Statistical analyses were performed for the data without work-free days, as presented here. The abscissa of Fig. 6, and of the following figures, shows the scheme of the experiment with respect to the experimental variables, 'shifts', and the 'driving situations' on that particular day. As has been stated before, the possible influence of the driving situation is probably not relevant here.

In this experiment 'young' subjects perform better on the CIT throughout the experiment than 'old' subjects. The apparent difference is confirmed by an analysis of variance ($F(1; 90) = 92.19; p < 0.001$). Besides this difference between the two age-groups, there is also a general increase apparent in CIT-performance over the experimental period.

![Figure 7](image-url)

Fig. 7 shows the influence of a second experimental variable on $\bar{\lambda}_c$: i.e. 'shift'.

Firstly, there appears to be a general rising trend in performance during the 3 week periods. Further, the performance before the beginning of the late shift, taking into account the upward trend, is everywhere higher than before the other two shifts. Between the situations before the early and 'broken' shifts, however, there seems to be not much difference in performance. Therefore, the highly significant difference with respect to the factor 'shift' ($F(2; 90) = 12.18; p < 0.001$), could be attributed to the higher values of the late shift. No interaction between the variables 'age' and 'shift' is found ($F(2; 90) = 0.50; n.s.$). With regard to the influence from the factor 'shift' on CIT-performance, the time of day the CIT was performed was the only experimental condition that was different. Therefore, it is likely that the difference in performance is caused by the difference in time of day on which the measurement took place.
Generally there is an impression of an overall increase in performance during the 3 successive weeks. As has been stated before, the hypothesis has been tested that there is an influence from further practice on the CIT-performance. Testing has been performed by looking for possible differences in tracking performance during the 'successive periods'. Each 'successive period' consists of a 'broken', an early and a late shift.

Fig. 8 shows a general increase from period 1 through period 3, and a levelling off towards period 4. The differences between the 4 'successive periods' are highly significant ($F(3;80)=66.06; p < 0.001$). No interaction between the variables 'successive period' and 'age' is found ($F(3;80)=2.18; n.s.$).

**Temperature data**

The temperature data are presented for the same experimental variables as has been done for the performance data. Mean values of the oral temperature are given per age-group (Fig. 9).

---

**Fig. 8. Mean values of the $\lambda_c$ per 'period', shown per age-group and for all subjects.**

**Fig. 9. Mean values of the oral temperature per age-group.**
The difference between the age-groups with regard to temperature is significant but is not impressive ($F(1;72) = 4.89; p < 0.05$). What seems to be more interesting are the two rather pronounced peaks.

A different presentation of the same data (mean values of the oral temperature per shift) is shown in Fig. 10. The overall difference found for the influence of the variable 'shift' on oral temperature ($F(2;72) = 79.37; p < 0.001$) can be ascribed to the higher temperature level before the late shift, compared with the level before the two other shifts. In Fig. 10 there is no trend apparent in temperature in the course of the 3 weeks. This is confirmed by testing for a possible difference between the four 'successive periods' ($F(3;54) = 1.90; n.s.$).

Finally a combined presentation of performance and temperature data is shown in Fig. 11. It appears that both the best performance and the highest temperature were obtained before the beginning of the late shift, that is at about 15.40 hours.
Discussion

The preliminary character of the present data is stressed again. Values of some subjects and values of important variables (psychometric and physiological), that both might be necessary for a more complete interpretation of the data, are not yet available.

With these restrictions the following tentative conclusions are given:

1. The 'young' subjects in the experiment show a better performance on the CIT than the 'old' subjects.
   The results, with regard to age, point to a reduced tracking performance for the older subject. Similar results have been found by Cassell (1973) for a pursuit tracking task.

2. A learning effect, even after a day of intensive training, continues at least until the third 'successive period' (i.e. at least for 7 experimental days).
   Also the training period of two calendar weeks used by Jex (1967) points to a rather long learning effect for the CIT.

3. In the late afternoon, the performance on the CIT has shown to be better than in the early morning. Testing the possible sensitivity of the CIT to fatigue in truck drivers, O'Hanlon (1981) measured CIT-performance on different times of the day. He also found a significant influence of time of day on CIT-performance, but the worst performance occurred at midnight.

4. Concerning the physiological state of the subjects, it was found in this study that the subjects had a higher oral temperature in the late afternoon than in the early morning. Oral temperature is sometimes used (Colquhoun, 1971) as an indication of level of activation. Therefore the increase in performance on this task and the increase in oral temperature over the same part of the day suggest a possible sensitivity of the critical instability task to the level of activation of our subjects.

The above mentioned results suggest that it is possible for the experimenter using the CIT to be confronted with an influence of age, learning processes and circadian effects on CIT-performance.

Literature


A MODEL OF THE HUMAN HELICOPTER PILOT DURING I.L.S. APPROACH -
DESIGN OF A DISPLAY FOR CONTROLS OPERATING

by D.DIEP, J. PAPON, D. VIARD
ADERSA 2 avenue du 1er Mai 91120 PALAISEAU (France)

1 - INTRODUCTION

This work (1), which main axis is the analysis of the pilot's mental workload, aims to:
1 - provide a Pilot Model which could be sufficiently representative;
2 - experience a new information display related to this model;
3 - propose means of evaluation and reduction of the pilot's workload.

As far as the first point is concerned, it appeared that a strictly behaviourist model was to be left out, and that it was preferable to develop a "cognition model", able to describe the mental mechanisms which are involved. This cognition is expected to make the representation easier.

Besides, because of the failure of a category of models based on Information or Signal Detection Theories, we directed this study towards transfer function or optimal control representations.

As a methodological approach, we adopted a combination of the Control Engineer and the Psychologist viewpoints. This approach was very fruitful and many advantages are expected:
- in the field of Experimental Psychology, such a model, due to its formal properties, is able to describe the continuous interaction between the dynamics of the human operator and the controlled system;
- in the field of Automatic Control, it is possible to bring a lot by introducing psychological data and techniques in several application cases of Artificial Intelligence.

Concerning a new information display, it is underlined that one can respect the principle of compatibility between the visual display and the mental processes activated by the pilot. We thus reach a quantitative identity between some display properties and some psychological variables: duration of anticipation, action threshold, etc...

---

(1) This work was supported by the French military research services (D.R.E.T.)
At last, concerning the workload analysis, let us recall that on a historical and theoretical point of view, the workload problem can't be dissociated from a global theory of the Human Operator. The proposed modular model is fit for the use of additional tasks. According to some previous works, it may be admitted that any task demand does preferentially act upon a particular step of the operator's information processing. It may thus be sought how a visual display increases or decreases the constraints undergone by the operator.

2 - MODELLING THE HELICOPTER PILOT

2.1 - Instrument Landing

The Instrument Landing System (I.L.S.) is a two-pointer indicator which is used to land along a reference path, which appears as a wireless beam. The deviations of the pointers are proportional to the angular deviations of the aircraft with respect to the I.L.S. beam (see fig. 1): the vertical pointer indicates a deviation with respect to the landing track axis (Localizer deviation), and the horizontal one a deviation with respect to the nominal slope (Glide deviation).

![Figure 1](image)

Speed regulation and head stabilization are here supposed to be executed either automatically or manually, and in this last case the actions on the foot pedals and on the longitudinal cyclic control stick are neglected.

With these hypotheses, the I.L.S. piloting task can be considered as two entirely independent compensatory tasks (fig. 2).

![Figure 2](image)

fig. 2: simplified helicopter (+ ILS) model
2.2 - Pilot Model

This model was elaborated from:
- data recorded during flight tests;
- technical and ergonomical enquiries with test pilots;
- application of knowledge and notions from both Automatic Control and Psychology fields.

Its structure is a modular structure, and in the case of a single task performance (Localizer compensatory task), it can be summarized by figure 3 (ref. [1], [2]).

(1) The Acquisition module is the first step, at which the pilot estimates the state of the aircraft, composed of the Localizer deviation $\xi_L(t)$ and its derivative $\dot{\xi}_L(t)$ at time $t$.

(2) The pilot acts only when the state $(\xi_L, \dot{\xi}_L)$ reaches a determined action threshold. This threshold is represented by a domain of action (DA) which is limited by two straight lines in the plane $(\xi_L, \dot{\xi}_L)$. Actions are then calculated in order to minimize a criterion at time $t+\Delta t$:

$$C(t+\Delta t) = \xi_L^2(t+\Delta t) + \lambda \dot{\xi}_L^2(t+\Delta t)$$

where $\lambda$ is a ponderation coefficient.

(3) Action is delivered by the pilot during a period $\tau$.

(4) The pilot has in mind an internal model which determines the coefficient values used for his control actions: DA, $\lambda$, $\Delta t$, $\tau$, with respect to the distance to touchdown $D$.

This internal model is also used to predict the future state of the aircraft, and if needed (that is to say, if the predicted state variables differ from the actual ones), the pilot matches the gain $G$ of his internal model to the actual gain.

fig. 3 : single task Pilot Model
One can establish a comparison between the notions used in the model and corresponding notions used in Psychology [5].

<table>
<thead>
<tr>
<th>notions used in the model</th>
<th>notions used in Psychology</th>
</tr>
</thead>
<tbody>
<tr>
<td>internal model</td>
<td>mental model</td>
</tr>
<tr>
<td>criterion</td>
<td>subjective cost</td>
</tr>
<tr>
<td>adaptation</td>
<td>learning</td>
</tr>
<tr>
<td>state variable representation</td>
<td>perception &amp; processing of variables</td>
</tr>
<tr>
<td>domain of action</td>
<td>response threshold</td>
</tr>
</tbody>
</table>

A Glide Pilot Model has been elaborated according to the same structure. With regard to the Localizer Pilot Model, this model is simpler, since the dynamics of the Glide task are simpler (1 integrator instead of 3).

Concerning the multi-task model (dual axis task), a first model based on a single channel hypothesis has been realized, and it gives satisfactory results in simulation. Nevertheless, some possible interferences between the two tasks are still to be analyzed through specific experiments.

3 - SIMULATION - PRESENTATION OF A FLIGHT DIRECTOR

In order to validate some of the hypotheses used in the model and to make some performance and workload evaluations, a real-time simulation of the I.L.S. approach has been implemented in a computer.

Aircraft controls were reproduced by two sticks, and the panel was visualized on a graphic display.

This laboratory tool also permitted to test a new presentation of informations, in the shape of a Flight Director (F.D.) likely to reduce the pilot's mental workload.

The Flight Director is represented as follows: the pilot's control stick motions appear as a mobile circle on a dial (roll attitude and collective control); the controls calculated by the Pilot Model are presented on the same dial as a moving square. The action that is proposed to the pilot is to adjust the circle position onto the square position. The square comes back to the center (zero position) when it is useless to act.

Such a presentation is to be distinguished from a usual F.D., which principle is to maintain two pointers crossed at the dial center, the instructions being delivered by an autopilot.
The expected advantage on a usual F.D. is twofold:

(a) the compensatory task becomes a tracking task. With the considered aircraft dynamics and category of signals, it is known [3] that the performance of the tracking task is superior to that of the compensatory task.

(b) autopilot instructions are replaced by human-type instructions. The actions that are suggested by the new F.D. correspond to those of a human pilot, when faced to the same situation. These actions are thus expected to lighten some steps of the task performance: estimation of derivatives, instant of decision, calculation of the motions amplitude and duration, etc. Moreover, the alternation of action and non-action phases should increase the free time of the pilot: instead of acting permanently (usual F.D.), the pilot will act in an intermittent way.

4 - WORKLOAD EVALUATIONS

Some tests have been made with test pilots. The results obtained, though mainly qualitative because of the limited number of pilots, are fully encouraging.

4.1 - Subjective ratings

The actions that are proposed by the F.D. have been recognized as valuable and they facilitate the piloting task.

The F.D. instructions were considered by the pilots as reasonable and similar to their own actions, as well for the action itself (amplitude, shape, duration, instant), as for the global strategy of the I.L.S. approach (reduction of the threshold of action).

4.2 - Secondary task

We have introduced a secondary "sub-critical" task, as defined by JEX [4], which seems to be compatible with the I.L.S. task. It is realized by a pointer rotating in a dial, with unstable dynamics, which the pilot controls by the mean of a turning handle, similar to a throttle handle.
The difficulty of performing this unstable compensatory task is measured by its degree of unstability. The mental workload margin, here defined as the pilot's capacity to perform simultaneously other tasks, varies in a linear way with this degree of unstability. The results concerning 3 pilots are shown in the table below.

<table>
<thead>
<tr>
<th>pilot</th>
<th>$\lambda_c$ (rad/s)</th>
<th>task without F.D.</th>
<th>task with F.D.</th>
<th>$\Delta L$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda_x$</td>
<td>$L$</td>
<td>$\lambda_x$</td>
<td>$L'$</td>
</tr>
<tr>
<td>1</td>
<td>3.3</td>
<td>0.4</td>
<td>0.12</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>3.8</td>
<td>0.9</td>
<td>0.24</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>0.6</td>
<td>0.19</td>
<td>0.9</td>
</tr>
</tbody>
</table>

with: $\lambda_c$ = critical coefficient (unstable task only)
$\lambda_x$ = sub-critical coefficient (primary + secondary task)
$L$ = workload margin
$\Delta L = L' - L$ workload margin variation

The measurements clearly show an increase in workload margin (25% to 53%) when the F.D. is introduced.

5 - CONCLUSION

This work is mainly based on simulation results and needs to be soon validated through flight and simulator tests.

The main research axes have been defined: Pilot modelling, workload analysis and realization of a pilot aiding display. The results obtained in each axis [6] suggest a lot of perspectives.

1) Pilot modelling

In order to improve the Pilot Model, one has to analyse precisely the complex piloting task and the pilot's behaviour (intermediate variables or supplementary tasks like speed control are to be taken into account), and to use data and specific experiments in laboratory.

2) Mental workload

The first workload evaluations with a secondary task seem rather meaningful; the mental workload level of the pilot appears to be particularly high, especially during the final phase of the approach (86% to 92% of full capacity).

A limitation of this method is that it can only be used in simulation. Further studies could have such purposes as:

- an extension and a validation of measurements with a larger number of pilots;
- the use of other evaluation techniques applicable during a flight, such as subsidiary secondary tasks, psycho-physiological measurements, etc...
3) Pilot aiding display

The Flight Director which is presented here has yielded rather good results, since it reduces the pilot's measured workload. Several improvements may be foreseen:
- the experimental comparison with a usual F.D. has only been made through pilots subjective ratings, and should be deepened.
- the relations between the Pilot Model and the displayed informations are to be analysed in order to define how a model can efficiently help the pilot, and what its limits are. In other words, a problem consists in determining to what extent a cognition model is necessary, and how to define a useful model.

REFERENCES


session 2

displays

chairman: j.e. rijnsdorp, netherlands
DISPLAYING PROCESS STRUCTURE

Lisanne Bainbridge

Department of Psychology, University College London
London WC1E 6BT, England

The operator of an industrial process has several different types of knowledge about his task. A very simplified representation of their relationships is given in Figure 1. Suppose that one wants to find the displays which best convey knowledge about general process structure, as well as giving information about the actual values of the variables at a particular time, as is done by conventional instrumentation. This paper outlines some of the decisions to be taken in setting up a study of this problem. There are three basic issues to be considered, the type of display to concentrate on, the types of process structure to find displays for, and the details of doing the experiments.

1. Types of Display

Four basic types of display were discussed in this presentation.

a. Geographic/Spatial (0): showing the spatial position of parts of the process.

b. Static mimic (3): The conventional type of mimic diagram in which physical links between parts of the process are shown in a topological diagram; actual values of the variables may be shown separately, typically in a different area of the instrument panel.

c. Dynamic mimic (25): A topological diagram of the process which includes information about the actual variable values, either by conventional instruments sited within the mimic or, on VDUs, by special means such as varying the width of a pipe to symbolise the flow rate within it.

d. Conventional scale instruments. (3)

Delegates at the conference were asked to say which of these they thought was the most important to study; the number of votes is given in brackets above. Several delegates voted 'other', as the above alternatives are all within a one variable-one display approach, and do not include the important types of display which show abstractions and combinations at a higher level than the basic process variables, as discussed in the paper by Rasmussen in this conference.

The main interest in all these types of display is their effectiveness in conveying information about the underlying functional structure of the process, the causal relations between variables and the actual nature of the functions relating them. When the operator has these types of knowledge he may be able to predict future events, and work out new strategies in unfamiliar situations. Figure 2 gives some suggestions about the effectiveness of the above types of display for giving these types of information. When information is explicit, this assume that it is obvious from the display without special knowledge of this particular process, though it does require some basic skills such as the ability to interpret instruments. Two types of implicit information are suggested. In one, as on conventional instru-
ments, experience of actual readings allows the operator to learn about the general underlying structure and function of the process. In the other, as on static mimics, the display acts as a guide while the operator uses his knowledge of general principles to predict what is actually happening. The table suggests that the reason why dynamic mimics are of interest is because they make the largest number of types of information explicit, with conventional instruments second as they provide data for learning. The interest of VDU graphics is that they may have the potential for making all types of information explicit. We do not yet know whether these displays are limited only by programming ingenuity, or whether there are important aspects of process knowledge which cannot be shown in two spatial dimensions plus time.

2. Process Structures

The next step is to choose the type of process for which one will study the displays. There are two basic approaches to this. One is to use typical modules of real processes which occur in industry. Several analyses of these are available in the literature. The alternative is to use theoretically possible structures, generated according to some simple constraint. For example, Figure 3 shows some of the structures of causal relations in which there are not more than 2 inputs to each output variable. The left side shows some structures which are possible with 5 input variables; the right side shows some of the possibilities for structures which have a different number of input variables but contain a similar configuration.

I did not ask delegates to express preferences for these two approaches but I would expect that working design engineers, and those interested in displays and training aids, would be more interested in real process components, while the theoretical structures would be useful in studies of the limits to an operator's comprehension of process structure. For example, using the types of alternatives shown in Figure 3, it could be possible to do balanced experiments to separate the effects of number of variables and of particular configurations, on the difficulty of understanding a process. Many of the studies at this conference used theoretical structures, although in most cases the full potential for comparing alternatives has not been realised. There are obviously, especially when one also considers the functions which might link the variables, millions of theoretical possibilities, so the choice of fruitful constraints becomes very important. These would probably be best obtained from an analysis of typical real modules and their dimensions of variability.

3. Experimental Design

A small example will be given to illustrate how complex experiments are in this area, even when the above issues have been resolved. This experiment was done with the aim of finding the best way of laying out conventional instruments to aid inference to the function connecting them. Figure 4 shows one page from a test booklet. The booklet showed examples for both vertical and circular displays; the input and output variables were labelled in half the cases, when they were unlabelled people were asked to say which they were. In all cases people were asked to say which arithmetic function linked the numbers on the scales.

The unlabelled scales are of particular interest, as they allow one to find out what factors influence people in interpreting the displays. There are three possible answers in each case:
This could be used by those used to arithmetic.

This could be used by those used to programming.

A large number of people gave this answer, although it is the most difficult. Unfortunately the study was designed assuming that only the first two would be used, so the results cannot be analysed in the way required.

Here are two specific examples, of increasing overall difficulty:

a. \( 8, 6, 2 \) could be
   \[ 8 - 6 = 2 \]
   \[ 8 = 6 + 2 \]
   \[ + 8 (= 6) - 2 \]

b. \( 7, 2, 14 \) could be
   \[ 7 \times 2 = 14 \]
   \[ 7 = /2 \times 14 \]
   \[ /7 (= 2) \times 14 \]

The choice of answers can show whether the person is influenced by position (left, middle, right), by choosing the largest or smallest number as the answer, or by a preference for a particular function, e.g. plus rather than minus, or multiply rather than divide. (For \( \div \) and \( \times \) function is confounded with size of output, so \( - \) and \( \div \) are of great interest.)

Reactions to the task are different when the numbers are presented on scales, rather than digitally as above for ease of demonstration. Also, dynamic displays are a much richer source of information about underlying functions. However this example demonstrates that it is possible to use very simple paper-and-pencil techniques to do pilot studies of this type of experimental question.

Although this study needs to be repeated there are some trends in the results so far which can be mentioned. One is that small numbers lead to the largest number of errors. For example, \((3,2,5)\) is easy to interpret when presented digitally, but when presented on a scale it is difficult to distinguish this from \((3,2,6)\). This has the practical implication that when displays are used with a longer scale than necessary, so that only the lowest part of the range is used, operators may misunderstand the nature of the functions relating them and so make the wrong predictions in unusual situations with larger excursions. The second point is that non-mathematicians and people with a hard mathematics background give different results, which reinforces the point that an engineer should not use his own preferences in designing a display.

A problem will arise if there are no clear biases within any population,
as it will be necessary to choose some other reason for imposing a structure on an instrument layout. Inversely, it may well be the case that the results of experiments like this would be misleading if generalised to other situations. For example, in a tank display, showing input rate, level, and output rate, 'level' is the answer to the arithmetic, but most naturally appears in the centre, while in a heating display showing fuel, material and temperature the result, 'temperature', could appear more naturally on the right. This criticism might suggest that all experiments on displays which are based on theoretical rather than real processes might be misleading.
plus general 'search and process' procedures for making use of data base in relation to goals.

Figure 1: Simplified diagram of types of knowledge available to industrial process operators.
Figure 2: Suggestions about the effectiveness of different types of display for conveying different types of knowledge.

<table>
<thead>
<tr>
<th></th>
<th>static mimic</th>
<th>dynamic mimic</th>
<th>scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>actual levels at time</td>
<td>I(G to A)</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>causal relations</td>
<td>E (flows only?)</td>
<td>E (some)</td>
<td>I(A to G)</td>
</tr>
<tr>
<td>numerical constants</td>
<td>-</td>
<td>I(A to G)</td>
<td>I(A to G)</td>
</tr>
<tr>
<td>behaviour over time</td>
<td>I(G to A)</td>
<td>E (chart recorders)</td>
<td>I(A to G)</td>
</tr>
<tr>
<td>conditions on events</td>
<td>E (flows only?)</td>
<td>E (flows only?)</td>
<td>I(A to G)</td>
</tr>
<tr>
<td>goals (conditions on actions)</td>
<td>-</td>
<td>E (constant)</td>
<td>E (constant)</td>
</tr>
</tbody>
</table>

Symbols:

E Explicit
I(G to A) Implicit, operator must predict actual values from knowledge of general principles
I(A to G) Implicit, operator might infer general descriptions from actual values after considerable experience
Figure 3: Examples of possible theoretical structures with the constraint that there are not more than 2 input variables to each output. (0 input, o output)
I think that these inputs **add** to give the output.

* subtract
* multiply
* divide
* don't know

I found this problem **obvious**

* quite easy
* difficult
* impossible.

*Figure 4: Example page from test booklet*
HOW PEOPLE DISCOVER INPUT/OUTPUT RELATIONSHIPS

R.N. Pikaar
T.H. Twente
Netherlands

1. Abstract

The influence of different display formats on how people discover input/output relationships of algebraic processes is investigated. An experiment was carried out to compare a numerical- with an analog presentation of variable values. Eighteen processes were tested, of which the outputs linearly depend on one or more inputs. Subjects observed the inputs and outputs, and had to write down their assumptions on existing relationships at fixed intervals. The results indicate that there are significant differences between the numerical and the analog format in favour of the latter.

2. Introduction

The introduction of Visual Display Units in process supervision and control raises questions on the attention paid to human aspects of these technological changes. There is little knowledge on how to structure process information on Visual Display screens and on the assimilation of the displayed information by operators.

For this research an important topic is the concept of internal model or representation of a process, that is all the knowledge that one possesses of this process. Bainbridge (1974) states:

"People appear to work from an internal model of the system they are controlling."

Or Stassen (1976):

"All forms of human behavior involve some internal representation - the internal model - of the system being observed or controlled, or can be explained in such terms."

In this research project the aim is to obtain more insight in this concept of internal representation, and especially in the influence of the display-format of process information on the internal representation. As stated in the title, we would like to know how people discover input/output relationships of processes, and how the internal representation develops.

3. Experiment

3.1. Process definition

Processes are defined by the following general description: the values of output variables linearly depend on one or more values of input variables (see figure 1.). The numbers of inputs and outputs both range
from 2 to 5. The inputs are correlated filtered noise signals and the characteristics of the input signals are all equal (within some range of amplitude and frequency). The processes have no internal lags, but they are updated every 5 seconds.

Two display formats were designed. One format with numerical presentation of variable values (figure 2.) and one format with an analog presentation (figure 3). The accuracy of the values is equal for both formats. The simulation of the processes and display formats was realised by means of a personal computer with some graphical features, and a display screen in an isolated room.

Figure 1 Process definition

A process is defined by: \[ \text{Output} = S \cdot \text{Input} \]
with: \( S \) : matrix of relationships
Output, Input are vectors.

For example:

\[
\begin{bmatrix}
\text{output 1} \\
\text{output 2}
\end{bmatrix} = \begin{bmatrix}
0 & 0.9 & -0.4 \\
1.2 & -0.5 & 0.5
\end{bmatrix} \cdot \begin{bmatrix}
\text{input 1} \\
\text{input 2} \\
\text{input 3}
\end{bmatrix}
\]

3.2. Experimental design

In an experiment, data will be collected to test the following statement:
1. The number and accuracy of assumed relationships of processes depends upon the visual presentation format.

Furthermore we would like to give an answer to the following question:
2. Which parameter(s) contributes most to the perceived (in terms of performances) complexity of a process (number of inputs, number of outputs, number of relationships).

As a derivative of the data we expected to find that:
3. The number and accuracy of assumed relationships of processes depends upon individual characteristics and circumstances.

And:
4. How do subjects discover the input/output relationships as time progresses (development of internal representation of a process).
To investigate these questions the following measurands were used:
- number of relationships found.
- number of relationships quantified.
- number of non-existing relationships found or quantified.
- squared sum of differences (between the actual and the assumed
  values of the relationships).
Furthermore subjects had to fill in a questionnaire in order to dis-
cover their observation strategies.

In this experiment only student subjects participated and the research
was confined to observation of inputs and outputs; the relationships
are unknown to the subjects (see "blackbox" in figure 1.). The subjects
had to observe the input and output values of processes during several

**Figure 3. Numerical format.**

```
   1  30
   2  35
   3  25
```

**Figure 3. Analog format.**
(up to eight) intervals of time (one interval: 100 seconds). After each interval they had to write down their assumptions on the input/output relationships.
For each display format a group of seven subjects carried out this task during three sessions of about three hours each. During the first session the subjects had to learn their task. During the second and third session the subject performances on fifteen different processes were registered. Three of these processes appeared in both sessions (with only a change in the sequence of the inputs or outputs).

4. Results
Statistical analysis of the numbers of relationships was carried out by non-parametrical tests (Mann-Whitney U-test, Kendall's Coefficient of Concordance W), as these measurands did not meet the conditions of an analysis of variance.
For both groups of subjects, mean and variance of the squared sum of differences was computed for each process at each interval of time. For these results the analysis of variance was used (Hays, 1977; Siegel, 1956).

4.1. Differences in performances between second and third session.
The final responses (performances after the last interval of time) of the subjects, with respect to the three processes that were presented twice were analysed as follows:
- measurand: squared sum of differences.
- method: two-way analysis of variance.
- significance level: \( P = 5 \% \)
- data population: seven subjects per format, three processes, two values per process.
- results: -there are no differences between subject performances during the second and the third session. -there are significant differences between the performances of the individual subjects.
These results apply to both formats. Apparently subjects did not perform better during the third session than during the second session (no learning effects on task aspects).

4.2. Differences between numerical and analog format.
For all the different processes the following hypotheses were statistically analysed:
1. There are no significant differences in the numbers of relationships found, between the analog and the numerical format.
2. There are no significant differences in the numbers of relationships quantified between the analog and the numerical format.
- measurands: number of relationships found, number of relationships quantified.
- method: Mann-Whitney U-test.
- significance level: \( P = 5 \% \)
- data population: seven subjects per format; results for each process at each interval of time.
Hypothesis 1 has to be rejected. The number of relationships found is significantly larger for the analog display format during the first three to five intervals (of 100 seconds) of the information presentation. This result applies for all processes. Figure 4 illustrates these findings.

Hypothesis 2 cannot be rejected. For the numerical format the raw scores, as well as the mean numbers of relationships found and quantified differ very little. If a relationship is found, it was quantified during the same interval of time. In connection with the rejection of hypothesis 1, the subjects of the analog format seem to look for relationships first, before quantifying them. However the variances of the numbers of relationships quantified are rather large for the analog format; there are large differences between the subjects. Figure 5 illustrates these findings.

4.3. Complexity of processes

The perceived complexity (in terms of subject performances) of processes was investigated by ranking the processes for each subject by the squared sum of differences of the final responses. For each group of subjects Kendall's Coefficient of Concordance W was computed.

4.3.1. Numerical format

For the numerical format the agreement on the ranking of the processes among the subjects appeared to be significant. This significance was totally determined by the two processes that scored extremely high, and the two processes that scored extremely low. The last two processes were the only processes with five input variables. The first two processes were processes with two or three input variables and a rather small number of relationships.

The application of Mann-Whitney tests on the numbers of relationships found or quantified of processes with an equal number of inputs and outputs and different numbers of relationships, showed no significant differences between these processes.

Tests on these measurands for processes with an equal number of relationships and different numbers of inputs showed significant differences: a higher number of inputs corresponds to a lower performance of the subjects.

4.3.2. Analog format

Contrary to the numerical format the agreement on the ranking of the processes among the subjects was not significant. For processes with an equal number of inputs and outputs and different numbers of relationships, no significant differences were found.

For processes with an equal number of relationships and different numbers of inputs, differences exist, but these differences are not significant.

5. Discussion

The results of this experiment could be summarized as follows:
- The performances on the experimental task are strongly linked to the
Figure 4 Number of relationships found
Process 22, 7 relationships, 3 inputs, 3 outputs.

- = one subject analog format.
- = one subject numerical format.

Number of relationships found.

Figure 5 Number of relationships quantified.
Process 22, 7 relationships, 3 inputs, 3 outputs.

- = one subject analog format.
- = one subject numerical format.

Number of relationships quantified.
individual subject, except for the measurand "number of relationships found".

- People discover the input/output relationships of processes significantly faster and better (more complete) with an analog presentation of variable values than with a numerical presentation of variable values.

One could say that the subject is invited by the analog format to gather a picture (internal presentation) of the process as a whole before quantifying the relationships. In contrast, for the numerical format the subject is immediately invited to quantify the relationships found. While he or she does not know the process as a whole, errors are introduced.
- The perceived complexity (in terms of performances) of a process mainly depends on the number of inputs in the case of the numerical format. For the analog format no significance on this aspect was found, probably because the range of the number of inputs was not large enough to find significance (if any).

The results of the questionnaire also show great differences between the individual subjects. Therefore it is not possible to find systematic differences between both formats. There was one exception: the subjects of the numerical format expected to have made errors, in naming the correct inputs or outputs, and the direction of variable value changes. All subjects answered that they looked for relationships first, and quantifying them later on. For the numerical format this is contrary to the results found in section 4.2. This could be explained by the assumption that "later-on" for the subjects of the numerical format still means "within the same interval of time". In most cases they found and quantified one relationship per interval of time. In the analog case, far more relationships were found per interval, but not (always) quantified (in the same interval).

This experiment has demonstrated the necessity to investigate the effects of information presentation formats on Visual Display Screens, in view of the significant influence of the format on the discovery of process relationships.

This experiment can easily be extended to more complicated and dynamic processes; processes that conform more with reality in process industry.

Acknowledgement

The research took place within the measurement and instrumentation group of the department of electrical engineering of the Twente University of Technology. The author wishes to thank D. Bosman, T. M. J. Lenior and J. E. Rijnsdorp for their contributions during this doctoral research.

References


COPING WITH COMPLEXITY

Jens Rasmussen
Morten Lind
INTRODUCTION

A major topic in a discussion of safety aspects of modern industrial installations is invariably the complexity of the plant operators' work situation during abnormal plant operation. A situation is painted of a control room with thousands of instruments and indicators offering potentially important information, while hundreds of alarms assisted by a couple of screaming horns try to guide the operators' attention. The situation in highly automated plants has very picturesquely been characterised by 99% boredom and 1% horror (Bibby et al. 1975).

The conclusion of such discussions is typically one of two: Either it is argued that operators need more effective training, or it is concluded that modern information technology should be used to assist the operator in coping with complexity by using computers for analysing disturbances and presenting information by advanced displays. The present paper aims at a discussion of the potential for assisting the operator in coping with complexity by means of computerized situation analysis. Before turning to computer support, it will be helpful to discuss the concept of complexity and to analyse how people cope with complexity without the assistance of computers.

THE CONCEPT OF COMPLEXITY

What is complexity and how is it measured? In a discussion of complexity of diagnostic tasks, Rouse et al. (1980) distinguish between subjective complexity and objective complexity, which can be defined and quantified. The literature review of Rouse et al. mentions a number of attempts to quantify complexity in terms of number of items to consider during analysis, or the number of alternatives to choose from. But what does this
complexity measure describe? One may argue that objective complexity of a physical system does not exist. The complexity observed depends upon the resolution of the information seeking attention. A simple object becomes complex if observed through a microscope. Objective complexity can only be defined for a given representation of a system, not for the system itself.

For industrial plants, the complexity faced by operators is determined by the representation of the internal state of the system as presented by the information displayed. This means that the complexity perceived by the operators is determined by the technology of the interface system. During a period when instrumentation is governed by the one sensor - one indication technology, only one level of resolution of the representation is available to the operator, and this has to be the most detailed one needed in any situation. In that case the interface must be complex by the law of requisite variety (Ashby 1960). However, if the resolution of the representation and the focus can be selected to suit a given situation, complexity need not be a fact of reality. To do this is precisely what is possible by use of computer processing of the measured data. However, for operators to trust computer support they should be able to understand and monitor the processes of the computer. Therefore the computer strategies should be compatible with the way in which operators could handle the same problem, were they given the necessary time.

OPERATORS' TRICKS IN COPING WITH PRESENT COMPLEXITY

Since the human capacity for analysis and decision in a non-routine situation is notoriously limited to consideration of a very limited number of items of information, the only way to cope with the high number of information sources and of devices of elementary actions (e.g. switches and valves) found in an industrial plant, is to structure the situation and to transfer the problem to a representation at a level with less
resolution. The total data processing task then is: To structure the information at a higher level representation of the states of the system; to make a choice of intention at that level; and then to plan the sequence of detailed acts which will suit the higher level intention, see fig. 1.

Humans are very well equipped for this tripartite task in the everyday concrete environment when navigating their body or manipulating objects - being it physical objects or symbols on paper appearing as a kind of artificial objects. This capability depends on the possibility of direct perception of higher level states and values as features in the information patterns received from the environment, and on the possibility of forming integrated motor patterns which can be activated by higher level intentions or orders. Both are depending on direct operation in a time-space world where movements are controlled by signals which have no symbolic or indirect meaning, and which can be treated simultaneously by data driven transformations in a parallel processing network. This processing also depends on quantitative (analogue) representation of the time-space signals for the control of movements.

The higher level conscious decision-making is related to states, values, and intentions for acts. This depends on another human trick in coping with complexity: Common sense natural language reasoning is based on qualitative representation of (large) sets of physical variables in terms of objects and functions which are characterised by states and properties rather than by physical variables and their quantitative relationships.

In both respects, the work situation in a traditional control room is posing problems to the natural human way of processing the data. Only in special and very familiar situations can operators operate directly on the time-space aspect of the display devices - only in some tracking and feedback adjustment tasks they may operate from the "expressions of the face of the system". In most cases they have to consider the information to be symbols of the internal state of the system. Basically, this
Characteristics related to work environment:

<table>
<thead>
<tr>
<th>EVERY-DAY CONCRETE, PHYSICAL ENVIRONMENT</th>
<th>DATA INTEGRATION</th>
<th>COORDINATION OF ACTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PERCEPTUAL FEATURE FORMATION AND IDENTIFICATION</td>
<td>AUTOMATED SENSORI-MOTOR PATTERNS</td>
</tr>
<tr>
<td>CONTROL ROOM WITH ONE-SENSOR-ONE INDICATOR TECHNOLOGY</td>
<td>CONSCIOUS, FUNCTIONAL ANALYSIS OR SHORT CUTS FROM STEREOTYPE SIGNS, EXPECTATIONS AND PROCESS-FEEL</td>
<td>PROCEDURALISED ACTION SEQUENCES</td>
</tr>
</tbody>
</table>

Figure 1. Characteristics of human data processing depend on work situation.
means that a set of physical variables presented should be consulted and integrated by functional diagnosis, since traditional display techniques do not allow for efficient perceptive identification. In this case humans exercise another efficient way of coping with the most frequent situations: They notice correlations and select one or a few convenient indications as signs of internal states. Generally a very efficient trick, but disastrous when faults change the system's behaviour, since the convenient but not defining signs then lead operators into traps. The basic feature of signs is that they refer to actions and are not one to one representation of system states.

A few other tricks assist the operator in the less familiar work situations. An efficient one is not to operate on absolute data, but to base the judgements on deviations from normal or familiar situations and system states. This is of course tightly connected to the use of qualitative information, since unfamiliar situations are qualitatively most conveniently labelled by referring to known familiar situations or system states. Another efficient trick is not to start every decision by collecting all the information needed. A skilled operator who cooperates with a system has very firm expectations regarding the state of the system, and therefore only looks for signs which are suitable to confirm or disprove his expectations - and only when he has doubts. A simple input-output model of an operator is therefore not acceptable for less familiar situations.

Decisions based on signs are only effective for situations for which the necessary conventions have had the chance to evolve. For new or unfamiliar system state caused by infrequent conditions or faults, the operators' identifications and decisions must be based on observations treated as symbols, i.e. representations for concepts related to the system's internal causal structure. The operators' symbolic data processing then depends on an internal or mental model of the causal structure of the system, and again humans have a number of ingenious ways to circumvent complexity by transfer of the problem to a
representation suited to treat the present problem (Rasmussen 1979). The major tools are hierarchical aggregation/decomposition to change the resolution of the attention applied for the problem - which is very often coupled to a change in the level of abstraction used for the causal representation. Another tool is transformation into a representation for which solutions are ready from previous occasions. Hierarchical decomposition/aggregation is related to the span of attention of the operator, to the level of detail or resolution applied for data processing. A change in the level of abstraction is, however, related to the type of concepts used for representing the system and is basically independent of the level of hierarchical decomposition applied, although in practice there seems to be some correlation in the two concepts, as illustrated in fig. 2. Aggregation and abstraction hierarchies play an important role in human problem structuring and for systematic computer support, and will therefore be discussed in more detail. Fig. 3 gives illustrations of aggregation and abstraction.

AGGREGATION AND ABSTRACTION HIERARCHIES

The internal representations of the system's functional properties which are necessary for causal reasoning are available to operators in very flexible variations, and can be fitted to the problem at hand by varying the span and resolution of the model and the level of abstraction of the concepts used for modeling.

The resolution of the model is controlled by aggregation/decomposition of the elements used for representing the system. The system is considered a hierarchy of parts ranging from elementary parts and components - nuts and bolts - to the complete plant.
Figure 2. Illustration of the aggregation and the abstraction hierarchy and their typical coupling.
Figure 3. Illustration of abstraction and aggregation of the representation of an electrical power plant.
The hierarchy can be structured in many ways. However, in the context of control system design and operator decision making, the hierarchy is naturally structured by the way in which the components are connected to functional units. In order to have an orderly synthesis of overall plant function during start-up, it is necessary to establish a number of autonomous functional units at one level before they can be connected to one functional unit at the next higher level, compare Simon's watch maker (Simon, 1969). This definition of autonomous functional unit at several levels is likewise important for orderly breakdown of system function for shutdown or emergency actions. It immediately appears that a set of generic operator tasks during start-up can be defined: Coordination of functional states in a number of autonomous functional units; a network task of switching and valving to integrate into one higher unit; an adjustment of the operation of the unit to stabilize and optimize the total function of the unit.

In the abstraction hierarchy, the system's functional properties are represented by concepts which belong to several levels of abstraction, see fig. 4. The lowest level of abstraction represents only the system's physical form, its material configuration. The next higher level represents the physical processes or functions of the various components and systems in a language related to their specific electrical, chemical or mechanical properties. Above this, the functional properties are represented in more general concepts without reference to the physical process or equipment by which the functions are implemented, and so forth. At the lower levels, elements in the process description match the component configuration of the physical implementation.

When moving from one level of abstraction to the next higher level, the change in system properties represented is not merely removal of details of information on the physical or material properties. More fundamentally, information is added on higher level principles governing the co-function of the various functions or elements at the lower level. In man-made systems these higher level principles are naturally derived.
LEVELS OF ABSTRACTION

FUNCTIONAL PURPOSE

PRODUCTION FLOW MODELS, CONTROL SYSTEM OBJECTIVES ETC.

ABSTRACT FUNCTION

CAUSAL STRUCTURE, MASS, ENERGY & INFORMATION FLOW TOPOLOGY, ETC.

GENERALISED FUNCTIONS

"STANDARD" FUNCTIONS & PROCESSES, CONTROL LOOPS, HEAT-TRANSFER, ETC.

PHYSICAL FUNCTIONS

ELECTRICAL, MECHANICAL, CHEMICAL PROCESSES OF COMPONENTS AND EQUIPMENT

PHYSICAL FORM

PHYSICAL APPEARANCE AND ANATOMY, MATERIAL & FORM, LOCATIONS, ETC.

Figure 4. The abstraction hierarchy used for representation of functional properties of a technical system.
from the purpose of the system, i.e. from the reasons for the configurations at the level considered. This involves a shift in concepts and structure for representation as well as a change in the data suitable to characterise the state of the function or operation at the various levels of abstraction. For display design this means that matching the presentation to the most effective level of abstraction is not only a question of changing the format for arranging measured data (bar-graphs, curves, mimic diagrams), but the data must also be converted and integrated to match the abstract concepts. Some of these variables can be measured directly, as for instance liquid flows and levels of a mass balance for a flow representation, whereas energy flows and the levels of an energy balance must be derived by means of computations based on the measured data.

To us, a systematic use of this abstraction hierarchy seems important for formulation of the information needed by an operator to be able to identify and perform the proper control task in a given situation. At each level of abstraction, the reasons and specifications, i.e. the requirements for proper function, are formulated from above, and the means for control and potential for function, i.e. the physical capabilities and limitations, are coming up from below. In case of disturbances due to technical faults, the causes of malfunction are propagating bottom-up through the hierarchy of abstraction, at the same time as rules for proper functions are derived top-down (Polanyi 1958). Depending upon the situation, the operator's immediate task is related to one or another of the levels in the hierarchy - as will be discussed below - but in any case the task will be formulated from an identification of the discrepancy between the "top-down" proper function and the "bottom-up" actual function.

RELATIONSHIPS BETWEEN AGGREGATION AND ABSTRACTION

As already mentioned, there is generally a close correlation between the processes of abstraction and aggregation, between
the span of attention applied when considering the system and the abstractness of the concepts used for representation. A design process is well suited for illustrative purposes, in particular since the task of an operator dealing with an unfamiliar system malfunction will be to design a control algorithm to close the gap between the proper function and the actual function of the system. In case of an idealised, systematic design, i.e. a design process which is not performed by updating a previous design and not merely based on accepted standard practices, the process will be a systematic top-down realisation or materialisation of the stated overall functional purpose of the system; through a selection of a suitable production flow structure, a selection of appropriate physical processes for the production, identification of the relevant equipment, and finally selection of the components suitable for the equipment. This process is ideally an orderly change of view by concurrent change of aggregation and abstraction. For the design of control algorithms for normal operation and for plant protection, this process will generally be an iterative one. When means for realisation of a function or process have been selected at the next lower level, the implications and possible side effects at the level above must be evaluated, and causal links (for instance control loops) must be introduced to remove unwanted degrees of freedom which were added by the physical reality introduced when moving to lower abstraction levels. The aim of a design process is to coordinate and constrain the possible states and functions of a physical system to those appropriate for the purpose of the system, by means of proper system configuration and proper control links. At each functional level, reasons and requirements for the function are obtained from above, whereas support and potential for functions as well as causes of malfunction propagate from below.

The basic concepts used for describing the system at the various levels of abstraction do not depend much on the specific system considered. However, the way in which the aggregation and abstraction is coupled is very much related to the way in which proper operation is synthesized during
start-up and is, consequently, very much depending upon the specific type of system.

GUIDING OPERATORS THROUGH COMPLEXITY

Such control of the physical degrees of freedom by system reconfiguration and control action is exactly the task of the human operators in case of disturbances not met properly by the automatic control systems. Design of a relevant control strategy for fault management depends on the identification of a discrepancy between the specified or target state of operation and the actual state. This discrepancy can be formulated at each level of the abstraction/aggregation hierarchy. The one to select depends on the specific situation and the priorities of the different relevant operator tasks. The natural way to judge priorities and to select the proper level of abstraction/aggregation to formulate control strategies will be a top-down evaluation of the situation. This is partly so because the highest priority is generally related to the highest levels: First, judge overall consequences of plant production and safety to see whether the plant mode of the operation should be switched to a more safe state - for instance, standby or emergency shutdown. Next, consider whether the situation can be counteracted by reconfiguration or use of alternative resources. This is a judgement at a lower level of physical equipment and function. Finally, the question is the basic cause of the disturbance and how it can be corrected. This implies a search at the level of physical function of parts and components.

Another reason for the top-down evaluation is the simultaneous change towards more material, physical properties of the system and the narrowing down of the span of attention which enables a direct zooming-in on the discrepancy between actual state and target state. This, however, depends on the availability of information about the "actual state" of the system at each
level which can only be obtained by an evaluation of the measured data and of the actual system configuration. This bottom-up data integration must be based on functional analysis of the measured, quantitative data, not on combinatorial analysis of off-normal signals for the measured data individually, followed by a state identification from reference to stored symptom patterns for known disturbances. Since disturbances are propagating bottom-up through the hierarchy, bottom-up evaluation is advantageous to detect abnormalities and to give early warnings to announce the need for top-down identification of the proper task.

This approach immediately leads to several data processing tasks which are well suited for computers: First, storage and retrieval of technical specifications for production and safety, and of information regarding the purposes and reasons related to the various operating modes of the plant, together with the requirements and target states for each level of the hierarchy. This information can only partly be obtained by measurements on the plant (collection of data patterns defining "normal states"); much information must be made available by the system's designer. Secondly, identification of the actual state of operation at each level derived by data integration of measured values and information on systems configuration. And thirdly, presentation of information in properly formatted displays.

The way to assist operators to avoid complexity is then to make a repertoire of display formats available to him, structured in a hierarchy with a small number at the high levels of abstraction/aggregation, and a larger number at the low detailed levels, together with an orderly and structured way to seek through the hierarchy to "zoom-in" on the relevant display. The properties of the individual displays and the quality of cross references to related displays at higher and lower levels of abstraction are, however, important for the perception of complexity.
DISPLAY FORMATS

The operator actually faced with the proposed hierarchical set of displays will probably not realise the multilevel structure of the representation of the total system. In a given work situation, an operator will have only a certain part of the plant within his span of attention. This part is to him "the system" to be represented in the actual situation, and the rest of the plant is part of "the environment" of this "system". Considering only this more restricted "system" in a specific situation, generally only three levels of abstraction are relevant to the operator, viz. the process or function under consideration (the "what" level); the purpose of the process for the next higher level (the "why" level); and, finally, the level below representing the more physical properties (the "how" level, i.e. the implementation level). These relations between a functional representation and the adjacent levels are independent of the actual location in the abstraction hierarchy and are, therefore, well suited as a basis for display formats, see fig. 5.

When the focus and span of attention change in accordance with the requirements of his work situation, the typical coupling between aggregation and abstraction will lead to the effect that these three levels of abstraction - purpose, process, and implementation - as the operator sees them in a specific situation, will develop the full abstraction hierarchy of the designer (see fig. 4) by recursion as his attention shifts. This means that the system properties which are represented in the three levels and used by an operator in a specific situation will vary, and to keep the complexity of the interface low as perceived by the operator it is important to identify a consistent and uniform language to express the functional representations used for the displays. Similarly, the links used to refer operators to the display levels above and below the one in use should be standardised, and the typical operator tasks should be identified, i.e. the designer
PURPOSE
("WHY" LEVEL)

Consult to judge effects of disturbances and to prioritize goals

REQUIREMENTS

PROCESS REPRESENTATION
("WHAT" LEVEL)

DATA DISPLAY:
- Actual states and target states of process
- Critical variables and limits for supporting systems

IMPLEMENTATION
("HOW" LEVEL)

SUPPORT FUNCTIONS
- FUNCTION
- SUPPLIES
- CONDITIONS
CAPABILITIES AND LIMITS

Consult to find causes and to select alternative resources and means for action

Figure 5. Schematic illustration of the relationship between the three levels of abstraction to be considered for a control task.
must realise explicitly the types of control task he wants the operator to perform. The effect of this will be to make the concepts and structures used by the operator's higher level analysis and decision making as shown on fig. 1 more uniform and situation independent. The language used for representing the generalized causal function or process of the figure is based on a flow representation (mass, energy and/or information flow; Lind 1981) of a physical system, which in addition to its generality has the quality of being easily visualized (Rasmussen 1980). This may be used for displays which allow direct perception of system states and related operator tasks (Goodstein et al. 1980).

The typical interaction between the three levels to be considered by an operator is illustrated in fig. 6. In the example we have described a conventional power plant in a three level model. The first level describes the plant as an energy conversion system which distributes energy from the fuel to two sources, the electric grid and the environment (cooling tower etc.). The conversion process is conditioned by two support systems indicated by two critical variables related to the efficiency of the conversion process. At the next level below we have described the energy flow through the power plant with the systems supporting the energy conversion at the level above as subprocesses (energy transport systems). But the air/gas system and the steam generating systems are again supported by the "air/gas path" and the feedwater system. If the air/gas flow is not established, the air/gas system does not exist as an energy transport system. If the feed flow is not established and the levels are not proper, the steam generating system does not exist either as an energy transport system.

This example shows how the interlinkage between descriptions of different abstraction levels is established. In the representation a change from one level to the level below includes both a shift from "what" to "how", but also a shift in focus of attention such that support systems are described in terms of what they do. This example illustrates that the purpose of a process, i.e. its role in support of the level above, can in
Figure 6. Example of three levels of representing the functional properties of a power plant in terms of flow structures.
general be described by a few categories: First, the process may serve to implement part of the function at the level above, for instance a mass flow system can serve as carrier for an energy transport at the next higher level. Secondly, the process may serve to supply energy or material necessary for the function of the next level. Thirdly, the process may serve to maintain a condition within proper limits to support the function above, for instance to maintain a pressure constant in order to convert an energy balance condition to a pure transport condition; or to maintain bearing conditions of a pump. When the rôle of a process for the next higher level has been identified, the information necessary to characterise the exchange of requirements and capabilities across the boundary can be determined immediately. This exchange of requirements and capabilities also identifies the cross reference path between the levels of the display hierarchy which should be used to guide the operator through the information available to him, see fig. 5. Considering a disturbance of a process at a given level, he has to move to the level above to judge the effects and to prioritize; to explain and to find causes, or to find alternative functional capabilities and means for action, he will need to consult the level below.

In the individual displays, details are ignored which are irrelevant for the task at hand and which could only lead to an increase of the apparent task complexity. In addition to ignoring details (aggregation), processes are described in a language convenient for design of control strategies. This description involves "an element of abstraction". In this type of model, the effect of different automatic control systems is given by support systems. These simplifications allow for a considerable reduction of the number of basically different tasks which the operator has to learn and distinguish.

The linkage of processes and support systems can also be considered as an explanation of the decomposition of the overall control task. Support systems should be started up and be in proper state (target state) before the processes on the next level can be carried out, etc.
CONCLUSION

This structure of information processing would lead to the system in which the operator can consult a computer to obtain information with the degree of resolution matching his immediate need. The computer is used to store and process a large number of measured variables and other data available from the design process. It is also used to make this information available to operators at that level of detail and in terms of those higher level concepts which are necessary for system's monitoring and supervisory control. Data integration used in this way will serve to counteract the tendency to use subsets of data as stereotype signs since the operator will not have to spend mental resources for complex, but elementary functional deductions to integrate information contained in the numerous measured variables during stressed situations.

The multilevel modelling framework provides a knowledge base which can be used as a common denominator for the computations in the computer and the activities of the operator. This is necessary in order to establish an advanced dialogue between the operator and the computer during for instance diagnosis. Furthermore, the modelling framework is a basis for the specification of the functions to be performed by the computer, i.e. serve as a tool for design of the information processing system supporting the operator. This means that the framework is used by the system designer to cope with the complexity in specifying the functions to be performed by the information interface. The model framework defines the proper way of thinking of the process plant, i.e. the logic of its functional organisation.

An additional advantage with the multilevel approach in coping with complexity is that it leads to structured problem solving in diagnosis. The repertoire of strategies used in diagnosis is limited to a small number of generally applicable methods. This facilitates the transfer of diagnostic skills obtained by the operator during normal operation to diagnosis of infrequent incidents involving high risk.
In diagnosis using these models, the computer will guide the operator in a top-down search through several levels of abstraction. In response to early warnings indicating a plant disturbance, the operator/computer starts at the highest level of abstraction describing plant overall function and the systems supporting the process on that level. The search may then continue supported by the computer, in deeper levels by picking out one or several subsystems for investigation. Although the effect of plant disturbances always first appears at a low level as early warning signals, the efficiency of a top-down approach will help the operator in quickly performing a plant state identification.

The depth of the search depends on the nature of the actual disturbance and of the task of the operator. In disturbance compensation it is only necessary to identify the plant state to a level of detail where proper control actions are known to the operator/computer.

In conventional alarm systems the problem of diagnosis is left completely to the operator. The alarm patterns are situation dependent and do not include any clues to how to interpret the available data. In this way the operator has to perform a very complex inference process where measured plant data and alarms are combined with his knowledge of process functions and properties. This bottom-up approach to diagnosis excludes the explicit consideration of plant information which is known to the designer, such as the purpose of subsystems. Conventional alarm systems are situation dependent and their design requires specification of patterns which are virtually infinite in number. The approach described here is function-oriented and provides a formal method of relating different types of plant information so that it is operational to the operator and the computer.
REFERENCES


THE STRUCTURING OF INFORMATION ON VISUAL DISPLAY UNITS

Ir. A.J.B. van Boxtel*
Ms. C. Slappendel, B.Tech.*

Abstract

Computerized Visual Display Units offer great flexibility in structuring information display formats. However, there remains a lack of good criteria for choosing between alternative formats. In laboratory experiments alternative formats can be compared by using a subjects' accuracy, reaction times and subjective preferences as measures for the relative readability and clarity of information displayed. Meaningful alternative formats can only be designed following an analysis of operator tasks. In this research experiments have been based upon cases from within the process industry.

Introduction

The number of Visual Display Units (VDU's) used in various work situations has undergone a rapid increase. This increase has caused questions to be raised concerning the suitability and effectiveness of VDU's, considering the possible effects that working with VDU's may have upon users.

The introduction of VDU's into the workplace will have consequences for such aspects as the organisation structure and the workplace lay-out. These aspects may contribute to physical and psychological stress experienced by users. To minimise these effects it is also necessary to ensure that the information presented on the screen is readable and has clarity. Poor readability occurs when symbols cannot be easily distinguished. Poor clarity may result in the user having difficulty in interpreting the information. Poor readability and clarity may result in delays and human errors; the user may experience eyestrain, postural fatigue, frustration.

Information display format design using guidelines

The software of an information system can be made very flexible. By using coding methods as sizes, shapes, colours, brightnesses, directions, blink frequencies and spatial organisation, information can be arranged and distinguished on the screen. The total arrangement is referred to as the display format.

A large number of completely different display formats can be generated that, in essence, contain the same information. Choosing between the range of alternatives is a problem that many designers face. There is a lack of good criteria for basing any choice upon. This is despite the fact that guidelines exist in Ergonomics literature (eg. McCormick). Research which has produced these guidelines has concentrated upon the presentation of information in alphanumeric form.

In the process industry the current trend is towards the display of process variables in a graphical or semigraphical form. Larger structures such as line charts, bar charts, tables, flow diagrams are used, and an

* Twente University of Technology, Enschede, The Netherlands
assortment of different symbols is found in the same format. Besides the fact that inadequate guidelines exist, their application also leaves much to be desired.

The encoding of information to produce a display format requires the integrated application of coding methods. The separate guidelines which can be found in the literature foster the individual application of these coding-methods, rarely giving an integrated picture.

Research method

Typical for the process industry is the large amount of data that is required for the monitoring and control of a process or processes. This data can be structured to form large systems with hierarchies of many display formats. Information within a single format can also be structured. The structure of the whole system, which includes the dialogue; and of each display format interrelate. To date our research has concentrated upon the structuring of single formats. Graphical formats for study have been derived from case-studies chosen from within the process industry. In laboratory experiments (see also Verhagen, 1981) evaluating alternative display formats, efforts are being made to produce guidelines for the design of optimal display format structures.

The method that has been used is as follows:
- selection of a process (or information system)
- an analysis of operator tasks
- the design of alternative display formats
- experimentation

Format designs should be based on the way they are to be used. The task analysis is therefore a crucial step. However, it is very difficult to do properly. This is because the reasons why operators make certain monitoring and controlling actions are not yet fully understood.

A selection of display format alternatives to be tested may be necessary. The testing of the final selection could occur with the use of one of the two sets of apparatus available in our laboratory.
- A closed circuit camera through which (projected) slides can be reproduced on a visual display screen;
- a computerized visual display system.

The computer system has been used to display data stored in computer memory on a visual display screen. Using the data, subjects, seated in an isolated room, carry out specified tasks. The tasks are derived from the previously defined operator tasks.

The evaluation of the display formats is based on the variables: speed of reaction, accuracy of response and subjective preferences of a subject for a particular format. These criteria can be compared, to give an evaluation that reflects the readability and clarity of the different information structures.
Examples

An experiment has been based on a case-study chosen from within the steel industry. Researchers were able to gain a background knowledge of the process and observe operations in the blast furnace control room during 2 work shifts (i.e. 2 successive days, each day with a different work team).

A picture was chosen from an existing information display system for further study. This was an overview format representing the heat transfer distribution through a blast furnace wall (see Fig. 1), 72 values, each with a value that fell into one of six defined value categories, were displayed in one format. The categories of values were represented by various scales. These were:
1. a scale of squares of varying size
2. a greyscale: going from white, low value to black, high value
3. a greyscale: going from black, low value to white, high value
4. a colour scale including the primary colours
5. a blue green colour scale

Using two groups of students (the majority were following an education for process operator, at a secondary, technical level), the 5 scales were compared. An analysis of results shows that for the primary colour scale (4) reaction times were significantly longer and accuracy was significantly less than for the other scales. The non parametric test used was a two-sided Wilcoxon Matched-Pairs Signed-Rank test. The significance level chosen was $\alpha = 0.05$.

Relatively lower reaction times were in almost all cases reflected by greater accuracy. The subjective preferences that subjects gave, correspond to the relative performance (based on reaction times and accuracy) of the scales. The scales that the subjects worked best with and gave their preference to were the two greyscales.

Another experiment was set up to evaluate the application of various colour sets to one display format. Sets of 1, 2 and 3 foreground colours (upon a dark background) were applied to a flow diagram format derived from a steam generation process (see Fig. 2). The variations which were compared in laboratory experiments are as follows.
1. One colour for the structure, variable names and values.
2. One colour for the structure. Variable names and values are presented in a different colour.
3. The structure, variables names, variable values are each presented in a different colour.

Similar colours tend to be visually organised into groups. As a result the underlying structure of information may be perceived quite differently as was the case for the above example. It appears from the first analysis of data, that the colour sets used in this experiment to distinguish between the different information types (i.e. structure, names and values), have negligible influence on the task performance.

The experiment could be further extended to include the evaluation of colour usage for distinguishing between groups of variable values, or different process segments.
Conclusions

The purpose of this paper has been to show how alternative display formats can be evaluated by means of laboratory experiments. The results are specific for the tasks and display formats chosen. Whether these results can be used in other practical situations depends on how similar the task characteristics are.

The research, however, remains valuable in general, providing examples of how the phase of selecting between alternative formats could be approached in the design process in practice.

References


Verhagen, L.H.J.M.; Experiments with bar graph process supervision displays on VDU's; Applied Ergonomics, 1981 39-45
FIGURE 1: Overview format representing the heat transfer distribution through a blast furnace wall

FIGURE 2: Flow diagram format derived from a steam generation process
SEARCH TIME AND COLOR CODE SIZE
Gunilla Derefeldt and Hans Marmolin
National Defence Research Institute
Box 1165
S-581 11 LINKÖPING SWEDEN

Abstract
Search time was studied as a function of color code size and display density and for larger code sizes and higher display densities than hitherto have been used.

The results indicate that quite a large number of colors can be used for color coding, probably as many as 20 colors.

Introduction
Appropriate use of color in information displays can lead to performance improvements in visual search tasks [1,2]. As summed up by Krebs and Wolf [3] color appears to be an effective code when used:

- as a cue or alerting signal.
- as a method of grouping similar items or separating items.
- as a means of increasing symbol visibility.

The location, grouping and visibility of symbols are of course ultimately dependent on the composition of individual colors as well as the total number of colors chosen for the code.

Data on absolute identification of colors [4-6] show that 10-15 colors can be identified with nearly perfect accuracy. However, with some amount of training and careful selection of colors as many as 20-30 colors can be identified [7-9].

In search studies only small code sizes of 1-8 colors have been used [10-19]. Display density, number of items of the same color as the target symbol, subject's knowledge of the target color, and display heterogeneity have been found to be parameters of relevance.

The data by Green and Anderson [14] and Smith [16] show that for a fix display density, search time decrease as number of colors increase and thus number of items of the same color as the target decrease. Extending the investigation by Smith [16] to 10 colors Cahill and Carter [20] found the importance of this variable to hold only for intermediate and low densities. For higher densities and a larger number of colors, a decreasing discriminability between the colors being used results in increasing search time although number of items of the same color as the target decrease. The practical consequences, as discussed by Cahill and Carter [20], would be that as many as 10 colors can be used for low density displays, and probably as many as eight or nine colors for high density displays.
The present investigation was designed to test this conclusion by studying search time as a function of color code size and display density extending the study to larger code sizes and higher display densities.

Method

The effect on search time of display density, and number of colors used to code the display were tested in two experiments. In experiment 1 target color was held constant while it was randomized in experiment 2.

Displays: The display symbols were all colored squares of the size 0.8 x 1.0 cm subtending a visual angle of 30 minutes of arc. Each symbol could be identified by a letter or a figure attached to it. All symbols within each color category were combined with different letters. For the lower densities capital letters were used only, but for the higher ones also small letters and figures were used. For each display the same combinations of letters and/or figures were used within each color category. The symbols were randomly distributed across a TV-screen, having a grey background color with a mean luminance of 17.55 cd/m$^2$. Its CIE chromaticity coordinates was $x=.2776$, $y=.3220$. For the levels of density and the number of colors used, see Table 1. In each display number of items per color category was fix and equal to the ratio density/number of colors.

Table 1: Levels of density and number of colors used in Experiments 1 and 2.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Density</th>
<th>Number of colors</th>
<th>Search item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14-15</td>
<td>1 2 3 5 7 15</td>
<td>Red &quot;A&quot;</td>
</tr>
<tr>
<td></td>
<td>28-30</td>
<td>1 2 3 5 6 7 10 15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>1 4 5 8 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60-63</td>
<td>1 2 4 5 6 7 10 15</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>1 3 5 15</td>
<td>Randomized</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1 3 6 10 15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>39-40</td>
<td>1 4 8 10 13 20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>1 6 10 12 15 20</td>
<td></td>
</tr>
</tbody>
</table>

The displays described above were generated by means of a computer system with an image memory.
Colors: In the two experiments all color codes were made up of colors with fairly large distances in the NCS color space. For the CIE specifications of the colors used in experiments 1 and 2, see Table 2, and Figures 1 and 2. All symbols on the single-colored displays were red. With two colors on the display, the first two (red, green) colors from Table 2 were used. With five, six, seven colors etc. on the display, the first five, six, seven colors and so forth were used. In experiment 1 the target symbol was throughout red with the capital letter A attached to it. In experiment 2 the target color and the letter attached to it were randomized.

Table 2. Color names and CIE chromaticity coordinates of the colors used in experiments 1 and 2.

<table>
<thead>
<tr>
<th>Color name</th>
<th>CIE coordinates</th>
<th>x</th>
<th>y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. red</td>
<td></td>
<td>.6099</td>
<td>.3445</td>
<td>17.30</td>
</tr>
<tr>
<td>2. green</td>
<td></td>
<td>.2775</td>
<td>.6107</td>
<td>57.70</td>
</tr>
<tr>
<td>3. yellow, with a tinge of green</td>
<td></td>
<td>.3792</td>
<td>.5333</td>
<td>49.90</td>
</tr>
<tr>
<td>4. blue</td>
<td></td>
<td>.1450</td>
<td>.0813</td>
<td>8.92</td>
</tr>
<tr>
<td>5. violet</td>
<td></td>
<td>.2167</td>
<td>.1223</td>
<td>12.05</td>
</tr>
<tr>
<td>6. orange</td>
<td></td>
<td>.4795</td>
<td>.4526</td>
<td>35.90</td>
</tr>
<tr>
<td>7. light blue-green (turquoise)</td>
<td></td>
<td>.2023</td>
<td>.3209</td>
<td>64.70</td>
</tr>
<tr>
<td>8. rose</td>
<td></td>
<td>.3548</td>
<td>.3125</td>
<td>39.50</td>
</tr>
<tr>
<td>9. brown</td>
<td></td>
<td>.4613</td>
<td>.3109</td>
<td>4.64</td>
</tr>
<tr>
<td>10. dark green</td>
<td></td>
<td>.2459</td>
<td>.4054</td>
<td>14.19</td>
</tr>
<tr>
<td>11. light yellow</td>
<td></td>
<td>.3499</td>
<td>.4450</td>
<td>75.30</td>
</tr>
<tr>
<td>12. dark blue-green</td>
<td></td>
<td>.2205</td>
<td>.3837</td>
<td>22.20</td>
</tr>
<tr>
<td>13. dark blue</td>
<td></td>
<td>.1480</td>
<td>.0962</td>
<td>4.13</td>
</tr>
<tr>
<td>14. light yellow-green</td>
<td></td>
<td>.3300</td>
<td>.5700</td>
<td>46.50</td>
</tr>
<tr>
<td>15. olive green</td>
<td></td>
<td>.3720</td>
<td>.5274</td>
<td>23.90</td>
</tr>
</tbody>
</table>

Experiment 2

The colors 1-2 and 4-8 were the same as in Experiment 1.

<table>
<thead>
<tr>
<th>Color name</th>
<th>CIE coordinates</th>
<th>x</th>
<th>y</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. yellow</td>
<td></td>
<td>.3968</td>
<td>.5206</td>
<td>69.80</td>
</tr>
<tr>
<td>9. reddish brown</td>
<td></td>
<td>.5157</td>
<td>.3217</td>
<td>6.61</td>
</tr>
<tr>
<td>10. buff</td>
<td></td>
<td>.3615</td>
<td>.4035</td>
<td>42.70</td>
</tr>
<tr>
<td>11. yellowish pink</td>
<td></td>
<td>.4769</td>
<td>.3864</td>
<td>29.70</td>
</tr>
<tr>
<td>12. light yellow</td>
<td></td>
<td>.3192</td>
<td>.3902</td>
<td>77.30</td>
</tr>
<tr>
<td>13. light blue</td>
<td></td>
<td>.2213</td>
<td>.2459</td>
<td>43.00</td>
</tr>
<tr>
<td>14. purplish red</td>
<td></td>
<td>.4352</td>
<td>.2529</td>
<td>12.50</td>
</tr>
<tr>
<td>15. dark blue-green</td>
<td></td>
<td>.2229</td>
<td>.3841</td>
<td>22.00</td>
</tr>
<tr>
<td>16. dark blue</td>
<td></td>
<td>.1508</td>
<td>.0971</td>
<td>4.11</td>
</tr>
<tr>
<td>17. yellow-green</td>
<td></td>
<td>.3283</td>
<td>.5708</td>
<td>46.00</td>
</tr>
<tr>
<td>18. yellowish brown</td>
<td></td>
<td>.4789</td>
<td>.4044</td>
<td>6.30</td>
</tr>
<tr>
<td>19. dark olive green</td>
<td></td>
<td>.2740</td>
<td>.4883</td>
<td>3.35</td>
</tr>
<tr>
<td>20. dark yellow-green</td>
<td></td>
<td>.3996</td>
<td>.4766</td>
<td>9.31</td>
</tr>
</tbody>
</table>
Figure 1. C.I.E. chromaticity diagram showing locations of the colors used in Experiment 1. The locus of the "white" background is indicated by $\Phi$.

Figure 2. C.I.E. chromaticity diagram showing locations of the colors used in Experiment 2. The locus of the "white" background is indicated by $\Phi$. 
Procedure

The subjects were seated at a terminal in front of the TV-screen. The viewing distance was 1.2 m. The target symbol which was to be located was first separately presented on the screen. By pressing the return button, the display was presented and the recording of the search time was started. As soon as the target was located, the subject again pressed the return button, whereas the recording of the search time stopped and a square pattern with numbers was presented. The subject had to respond with the number of the area, where the target was found on the display. Displays of different densities were presented in blocks and within these, the displays were presented in random order. Each subject made 10 replications of each display. The room was illuminated by an ordinary desk-lamp containing a 60 W incandescent bulb giving an illumination of about 2 lx. The subject was adapted to the illumination for at least 10 min before the experiment.

Subjects

10 subjects from our laboratory, some of which had some previous experience of search studies took part in the study. All had normal acuity and normal color vision. In experiment 1, 6 male and 4 female, ranging in the age from 23 to 39 years, and in experiment 2, 7 male and 3 female, ranging in the age from 23 to 48 years, took part. Five of these subjects participated in both experiments.

Results

The results of experiments 1 and 2 show a general decrease of search time with increasing number of color categories.

Figure 3 and 4 show for each level of density median search time (only correct responses were included) plotted as a function of number of color categories. For all levels of density, a rapid decrease in search times can be observed as the first few colors are added to the display. As more colors are added, generally search times continue to decrease but the decrease becomes more like a flattening. No pronounced minimum was found although as many as 15 or 20 colors were used.

As can also be observed from Figures 3 and 4 search times increased with increasing display density. There is a good agreement between the results obtained from the two experiments. For each density level, search times were about equal for the same number of colors.

In order to test the hypothesis that search time mainly is dependent on the number of items displayed in the same color category as the target [14], median search times are plotted as a function of number of colors with number of items per color category held constant. Figures 5 and 6 indicate that this variable is of major importance. However, search time is also affected by number of colors and/or display density, and this effect seems to increase with increasing number of items per color category.
Figure 3. Median search time as a function of number of color categories at each level of display density (15, 30, 40, and 60). Experiment 1.

Figure 4. Median search time as a function of number of color categories at each level of display density (15, 30, 40, and 60). Experiment 2.
Figure 5. Median search time as a function of number of color categories, with number of items per color category (2, 3, 5, 10) held constant. For the 5-items curve and 10 colors, the number of items per color category is 6. For the 3-item curve and 15 colors, the number of items per color category is 4. Experiment 1.

Figure 6. Median search time as a function of number of color categories, with number of items per color category (2, 3, 5, 10) held constant. Experiment 2.
In order to analyse the results more in detail a simple search model was formulated. Search time (ST) was assumed to be composed of:

- time to scan the display for possible target colors (TS)
- time to process these colors (TC)
- time to process the letters (TL)

Thus

\[ ST = TS + TC + TL \]

If one assume TL to be fix, TC to depend on number of colors (NC) and TS on density (ND) then in average the relation between ST and ND, NC could be written

\[ ST = 0.5 \cdot S \cdot ND^m + NI \cdot C \cdot NC^n + NI \cdot L \]  

(1)

where S, C, L, m, and n are constants and NI = number of items/color category = ND/NC.

This function was fitted to the data obtained (including only correct trials) in Experiment 1 and 2 by means of a general curve-fitting algorithm. Figures 7 and 8 show that (1) could be used to describe data. The obtained values of the fitted parameters are shown in Table 3.

According to this analyses in average no time was spent on color discrimination in Experiments 1 and 2. Search was composed of time to search the display for symbols of the same color as the target and of time to identify the target by its letter. The former process was affected by display density while the latter only was dependent on number of items of the same color as the target.

Table 3. The obtained values of the fitted parameters in (1) (see text), and the RMS-error of fit.

<table>
<thead>
<tr>
<th></th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>0.49</td>
<td>0.71</td>
</tr>
<tr>
<td>m</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>n</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>L</td>
<td>0.085</td>
<td>0.080</td>
</tr>
<tr>
<td>RMS</td>
<td>0.13</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Figure 7. Obtained search time as a function of predicted search time. Experiment 1.

Figure 8. Obtained search time as a function of predicted search time. Experiment 2.
Proportion of errors are plotted against number of colors in Figures 9 and 10. Considering the large number of colors used, and that the subjects had no prior training, rather few errors were obtained. In Experiment 1 no systematic relation was obtained between number of colors and number of errors. In Experiment 2, number of errors increased systematically with number of colors, but only for high density displays with more than 10 colors. An analysis of the errors showed mainly these to be related to confusion between the colors with the smallest perceptive distances.

Discussion

The results of this study indicate that quite a large number of colors can be used for color coding, probably as many as 20 colors. This number is larger than hitherto has been reported in the literature [20]. However, it approximates the number of colors which can be identified on an absolute basis [7-9]. The increasing number of errors with increasing number of colors on the display and increasing display density were probably due to color confusion between colors of small perceptive distances. However, it is to be noted that the subjects had no prior training of the color codes.

Figure 9. Proportion of errors as a function of number of color categories at each level of display density (15, 30, 40 and 60). Experiment 1.
Figure 10. Proportion of errors as a function of number of color categories at each level of display density (15, 30, 40 and 60). Experiment 2.

References


ERGONOMICS OF MAN-MACHINE COMMUNICATION:
VISUAL PERFORMANCE ASPECTS
P. Haubner

Abstract
Within the scope of a research project on communication ergonomics some results are reported on visual performance capability taking account of optical noise on VDT-screens, caused by specular reflections. Surprisingly those anti-reflection devices were superior, which induced relatively low screen luminances.

Introduction
Work-systems increasingly apply computer-based man-machine interaction. Therefore, the communication between man and machine and especially the information display, which is an essential part of it, are important ergonomic design factors. The tremendous amount of literature available on human factors unfortunately reveals some lack of knowledge as far as information processing by man is concerned.

Hence, we are carrying out a research program to obtain data and establish guidelines for the design of man-machine communication. Restricting ourselves to the interface at first, these are the major problems:

- Quasistatic display of information - concerning largely perceptibility and coding - more or less a psychophysical problem
Dynamic man-computer dialogue - concerning above all the type of information and its spatial or temporal organisation - basically a cognitive problem.

CRT-terminals are frequently used as displays. One of the major problems arises from reflections on the screens, caused by light sources, selfluminous or non-selfluminous, such as fittings, windows, papers etc. However, specular and veiling reflections act as contrast-reducing optical noise, which may affect task visibility. Further, some equipment applied to avoid or diminish specular reflections, as for instance micro-mesh filters, lead to relatively dark screens. Such, in the visual field around the task, luminance ratios may occur higher than 3:1, a limit claimed for in literature.

Dealing with the first item mentioned above, we investigated the influence of optical noise upon visual performance and subjective appraisal. Experimental comparisons of several technologies applied to reduce specular reflections were made to answer the following questions:

- How does optical noise affect visual performance quantitatively?
- Is there any indication, that dark screens come off badly because of their high luminance ratios to other devices?
- To what extent may a loss of accuracy be compensated by speed to maintain visual performance?

Some results are presented subsequently.
Experimental investigation

An experimental work station was set up, providing an observer with five noise situations and a reference task. The noise situations were given by light reflections on terminals of the same type with different screen surfaces. The reference task was performed by using a sheet of printed white paper instead of the VDUs. The experimental set up is shown in figure 1 and is described in more detail in /1/.

Figure 1  Experimental set up and the test room
(VDU-place in our light laboratory)

Two practical cases could be simulated by an artificial window behind and a wall of mirrors in front of the observer; one most unfavorable situation of high noise and a situation of low noise, if the source of disturbance was turned off.
Observers were asked to compare a "formal text" on a copy with a corresponding text on the VDU, to find out 30% discrepancies and to mark them with a pencil. The formal text consisted of random pairs of the capital letters D O Q, which are easily confused. The visual task included 270 pairs on the screen and on the document respectively. The task is illustrated in figure 2.

Figure 2 Formal text and the noise conditions:
Left side low noise - right side high noise.

The 28 observers were skilled in VDU-work; 76% were male, 24% female and their mean age was 32 years, ranging from 20 to 52 years.
On the average, the test took 4 hours per observer and was split into 2 sessions on successive days. Time and errors per screen were recorded for every person and 10 task specific items were scaled by the method of ordinal categories \( /2/ \). The empirical data were treated by an analysis of variance. As far as the psychometric results are concerned, we used non-parametric statistics (Friedman, Wilcoxon). A probability of \( \alpha = 0.05 \) was chosen as level of significance. Further specifications of experimental conditions are given as follows:

Table 1 Short description of the surfaces

<table>
<thead>
<tr>
<th>Surface</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM:</td>
<td>Micro-mesh filter</td>
</tr>
<tr>
<td>EL:</td>
<td>Etched glass filter (low transmission)</td>
</tr>
<tr>
<td>EM:</td>
<td>Etched glass filter (medium transmission)</td>
</tr>
<tr>
<td>SP:</td>
<td>Spray-coating</td>
</tr>
<tr>
<td>UG:</td>
<td>Untreated glass</td>
</tr>
<tr>
<td>WP:</td>
<td>White diffusing paper</td>
</tr>
</tbody>
</table>

Table 2 Reflexion characteristics of the surfaces

<table>
<thead>
<tr>
<th>Surface</th>
<th>q</th>
<th>( \kappa )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM:</td>
<td>0.015</td>
<td>0.006</td>
</tr>
<tr>
<td>EL:</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>EM:</td>
<td>0.03</td>
<td>0.015</td>
</tr>
<tr>
<td>SP:</td>
<td>0.04</td>
<td>0.025</td>
</tr>
<tr>
<td>UG:</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>WP:</td>
<td>0.22</td>
<td>0.00</td>
</tr>
</tbody>
</table>

q: luminance factor in \( \text{cd/m}^2 \cdot \text{lx} \)

\( \kappa \): specular reflectance
Table 3 Photometric data (illuminance, luminance)

<table>
<thead>
<tr>
<th></th>
<th>low noise</th>
<th>high noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>work place</td>
<td>500 lx</td>
<td>750 lx</td>
</tr>
<tr>
<td>(table)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>screen</td>
<td>210 lx</td>
<td>1050 lx</td>
</tr>
<tr>
<td>window</td>
<td>---</td>
<td>6000 cd/m²</td>
</tr>
</tbody>
</table>

The characters on the screen were generated in a 5 x 7 dot-matrix, had a contrast of about 6.5:1 and were seen under a size of 27 minutes of arc. Their color was green, except for the spray-coated surface, which had an amberish appearance. The characters on the paper (reference task) had a contrast ratio of about 15:1 and the same visual angle as the VDU-characters. The work-station was designed according to ergonomic requirements found in earlier research and published in /3/.

**Experimental results**

**Performance test**

We found, that time $t$ and errors $e$ increase with high-level noise; $\Delta t/t \sim 5 \%, \Delta e/e \sim 25 \%$. Moreover, performance depends on the screen surface significantly. This is true for each of the levels as may be seen in figure 3, where the relative visual performance $P_v$ is plotted against the specular reflectance $g_s$. $P_v$ will be explained later. In the case of low noise micro-mesh and the etched surfaces are equivalent. With high noise, it is remarkable, that only micro-mesh remains unaffected and increases the performance about 16 % compared with the untreated glass tube.
Figure 3 Relative visual performance $P_v$ in relation to reflectance $g_s$ for two noise conditions

Let us return to $P_v$. We felt, that two factors of performance should be combined to a total score, i.e. speed (time) and accuracy (error). This is achieved in a formally clear and objective way by the transinformation rate $\tilde{T}$. Hence, we defined visual performance as the relative transinformation-rate:

$$P_v = \frac{\tilde{T}}{\tilde{T}_{\text{max}}} \cdot 100 \%$$
where $\tilde{T}_{\text{max}}$ is the maximum possible $\tilde{T}$ in the given task. $\tilde{T}$ was calculated for each individual observer and averaged after that. $\tilde{T}_{\text{max}}$ was approximated by theoretical and empirical considerations as follows:

Analysing the confusion frequencies we found that the perceptibility of $D$, $O$, $Q$ was about the same. Hence the set of messages contains only two events $X_1$, $X_2$ with associated probabilities of occurrence $p(X_1)$, $p(X_2)$ and corresponding responses $Y_1$, $Y_2$:

$X_1$: The compared characters are identical

$X_2$: The compared characters are not identical

Because of $p(X_1) = 5/6$ and $p(X_2) = 1/6$, the transmissible information content is 0.650 bit in the ideal errorless case. Under the particular set of conditions of the task (size, contrast, luminance) the following minimal elementary reaction times may be estimated per pair of characters to be compared:

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small head movement</td>
<td>$\sim 200$</td>
</tr>
<tr>
<td>Readaptation</td>
<td>$\sim 100$</td>
</tr>
<tr>
<td>Reaccommodation</td>
<td>$\sim 150$</td>
</tr>
<tr>
<td>Fixations by saccadic eyemovements each</td>
<td>$\sim 100$</td>
</tr>
<tr>
<td>Processes of identification each</td>
<td>$\sim 240$</td>
</tr>
<tr>
<td>Sum:</td>
<td>$\sim 790$</td>
</tr>
</tbody>
</table>

Note: $\sim$ denotes an approximation.
Furthermore, the effector- and motor system must be activated to mark the non-identical characters on the document. Thus the response time is increased by:

<table>
<thead>
<tr>
<th>Time Component</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing time in the central nervous system (CNS)</td>
<td>~250 ms</td>
</tr>
<tr>
<td>(simple yes/no-decision)</td>
<td></td>
</tr>
<tr>
<td>Transmission time CNS-muscles</td>
<td>~20 ms</td>
</tr>
<tr>
<td>Muscle-latency and activation time</td>
<td>~50 ms</td>
</tr>
<tr>
<td>Motor action time</td>
<td>~100 ms</td>
</tr>
<tr>
<td><strong>Sum:</strong></td>
<td>~420 ms</td>
</tr>
</tbody>
</table>

Thus a transmission rate of 0.61 bit/s or so can be expected theoretically for the given task taking account of the probabilities of occurrence of the events $X_1, X_2$. An empirical estimate of $\tilde{T}_{\text{max}}$ may be derived from the dispersion of $\tilde{T}$. It turns out, that transmission rate is normally distributed with a mean at 0.44 bit/s and a standard deviation at 0.08 bit/s for the white paper display. Since, as illustrated in figure 4, the observers improved in performance due to learning, we used the asymptotic part of the graph to calculate transinformation rates. By taking the 99 th percentile of the distribution as an upper limit, we got 0.63 bit/s. This value is in a fairly good agreement with the theoretical one and was taken as reference (figures 3 and 4).
Figure 4 Relationship between performance $P$ and task duration $t^\Sigma$ (cumulated times per screen).

Depending on the task, the eye has to move more or less frequently between the objects of vision. At the worst, opposed to the 3:1 rule, luminance ratios may occur as high as 30:1. In table 4 are compared the ratios $l_R$ of average document luminance and average screen luminance with the percentages of visual performance.

Table 4 Luminance ratio $l_R$ in the visual field and relative performance $P_v$

<table>
<thead>
<tr>
<th>Low noise:</th>
<th>MM</th>
<th>EL</th>
<th>EM</th>
<th>SP</th>
<th>UG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_R$</td>
<td>18.3</td>
<td>9.4</td>
<td>8.2</td>
<td>6.5</td>
<td>4.9</td>
</tr>
<tr>
<td>$P_v$ %</td>
<td>98.0</td>
<td>98.4</td>
<td>98.3</td>
<td>94.7</td>
<td>90.1</td>
</tr>
</tbody>
</table>
The values in table 4 demonstrate that the 3:1 rule does not apply without exception to VDU-workstations. On the contrary, the data reveal a positive rank correlation with Spearman's coefficient at 0.70. Specular reflectance seems to be of stronger influence than transient adaptation or readaptation.

Figure 5 shows the transinformation rate $\tilde{T}(XY)$ in relation to response equivocation $H(Y/X)$. $H(X/Y)$ is a measure of the ambiguous or irrelevant component of the response information. Figure 6 represents the time $t$ against the equivocation $H(Y/X)$. From these figures, it appears that a loss in accuracy due to confusion by errors can be compensated by speed to a certain degree only. On the average, beyond some rate of error, which differs individually, the mean performance deteriorates quickly with increasingly unfavorable visual conditions due to a rapid decrease of response speed. This holds for each of the noise levels.

Figure 6

```
Transinformation Rate $\tilde{T}(XY)$ = P

0.45

0.40

0.35

bit/s

0.02 0.03 0.04 bit

Response Equivocation $H(Y/X)$
```

Figure 6 Performance $P$ versus response equivocation $H(Y/X)$
Experimental results

Psychometric scaling

The object of this part was the subjective appraisal of the observers' performances. For that, at the end of a session, each observer was provided with a set of items, which had to be judged on a five-step rating scale. The data were treated according to the "law of categorical judgement" /4/. Since the resulting scales had about the same interval spacings, we used a common scale for the following selected items:

- Legibility of characters LC (figure 7)
- Image quality RC (figure 8) (resolution of characters)
- Brightness of screen BS (figure 9)

Figure 6  Relation between time t/screen and response equivocation H(Y/X)
o Brightness contrast BC (figure 10) between screen and document

The figures illustrate, that subjective performance is reduced by specular reflections. There are also significant differences between the screen surfaces, which can be grouped into the following categories, ranging from very good to very poor:

<table>
<thead>
<tr>
<th>Low noise</th>
<th>High noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>MM</td>
<td>MM</td>
</tr>
<tr>
<td>EL</td>
<td>EL</td>
</tr>
<tr>
<td>EM</td>
<td>EM</td>
</tr>
<tr>
<td>SP</td>
<td>SP</td>
</tr>
<tr>
<td>UG</td>
<td>UG</td>
</tr>
</tbody>
</table>

Thus, we may conclude, that the subjective results are consistent with the objective performance scores. To the observer's opinion, however, the VDU-screens were too dark on the whole and the brightness step between the VDU and the document was felt to be somewhat too high.
Figures 7-10 Results of the subjective appraisal
Conclusions:

Within the scope of ergonomics of man-machine communication, several anti-reflection devices for VDUs were evaluated, based on visual performance and on psychometric scaling. With regard to the above three questions, we may state as follows:

- Visual performance is reduced markedly by optical noise due to specular reflections—objectively and subjectively. The best surface, practically unaffected by noise, was micro-mesh.

- The results do not confirm, that luminance ratios in the visual field must not exceed 3:1 as a general rule. Moreover, relatively dark screens were superior, probably owing to their low specular reflectances. A loss in contrast due to optical noise appears to decrease performance much more than transient adaptation does.

- To maintain constant performance, an interchange of speed and accuracy is possible over a very limited range only, a finding, which is in agreement with Fitts' earlier studies on choice reaction time /5/.

Finally, it may be pointed out, that our current activities deal with further aspects of performance, as e.g. color, prolonged work and visual fatigue and on the other hand, with software ergonomics.

References

2. Haubner, P. A method of scaling judgements
No. 3, 182-184

3. Benz, C., Grob, R. Gestaltung von Bildschirmarbeitsplätzen,
P. Haubner Köln, Verlag TÜV Rheinland 1981

4. Torgerson, W.S. Theory and methods of scaling,
New York, John Wiley 1958

5. Fitts, P.M. Cognitive aspects of information
processing
J. Exp. Psychol. 71 (1966) No. 6,
850-857

Dr.-Ing. Peter Haubner
Angewandte Arbeitswissenschaft FWO 2
Forschungszentrum der Siemens AG
Paul Gossen Straße 100
D 8520 ERLangen
A device for the measurement of contrast resolution, spatial and temporal resolution by means of a VDU screen.

by: Ir. F.W. Umbach
Dr. J.W.H. Kalsbeek
Prof. Ir. D. Bosman

Ergonomics Technology Foundation
C/O Dept. of Electrical Engineering
Twente University of Technology
P.O. Box 217
7500 AE ENSCHEDE, Netherlands

Abstract

Visual fatigue as an effect of prolonged VDU reading has frequently been studied by ophthalmologic tests as used by the prescription of eye-glasses. For this purpose, however, these tests prove to be insensitive. We used besides these classical means in our experiments a newly developed electronic instrument that measures contrast resolution, spatial and temporal resolution by means of test pictures presented on a VDU. The subject responds using a keyboard; the tests are fully automatic. We tested this instrument giving subjects a task of at least two hours on the VDU and compared the results of tests before and after the task.

Introduction

The question whether there are reversible and irreversible effects of prolonged VDU reading and if so whether they can be measured by ophthalmologic tests is not yet solved. Nevertheless, the problem is of great practical importance as well for the longitudinal testing of people working at VDU's as for the development and testing of VDU's. For large scale testing it is advantageous to have investigation methods which are easily applied and do not need the intervention of specialists or even test-assistants. In order to create such measuring methods we have realized ophthalmologic tests as presentations on a VDU screen using some intrinsic properties of CRT's. Because in principle each VDU can be used for these tests, the tests can be presented (as an interrupt procedure) on the VDU at which a person is working while he is performing his task. Because the tests act upon some other properties of the human visual system than do classical ophthalmologic tests, new forms of test procedures might be realized. Because the tests are performed fully automatically, the test data can be handled also automatically.

The tests

The principle of all our tests is that the subject has to recognize characters in a background, like in the Ishahara test, which distinguish from the background by a different pictorial construction. The intensity of the character increases gradually step by step, so that that level of intensity, at which the character is recognized, can be used as a measure for the quality of the subjects ability on a certain vision property. The characters are randomly chosen by the device from an alphabet. The device compares the answer with the presented character. If the answer is correct the test stops and the following test starts. If the answer is not correct the intensity of the character is further increased until a good answer is given.

The results of each test can be handled by the device in several possible ways. They can be presented on the VDU after all test items have finished. They can also be stored on mini-cassette for later retrieval or sent to a central computer. It is also possible to secure the results against illegal use by constructing the need for a pass-word for output of the data.

The device can test the following three vision qualities:
(i) The contrast resolution.
(ii) The spatial resolution.
(iii) The temporal resolution.

The contrast resolution test

The contrast resolution test is performed by giving those pixels (elementary picture elements) which belong to the character a brightness different from the brightness of the background, as depicted in figure 1. The total brightness scale has 512 steps. So, if the background is given a brightness corresponding to 384, the character is given, in successive stages, brightnesses corresponding to 384 plus 1, 2, 3 and so on. There is an option to increase the resolution with a factor 4. The time that each step lasts is chosen such that normal response time of the subjects causes only small errors in the measurement.

The spatial resolution test

The usual way of measuring the threshold of spatial resolution of the eye is to increase the number of lines per cm in a picture, until they are no longer resolved as separate lines. The lines have a constant brightness difference from the background. Fig. 2. However, because there is a relation between brightness modulation depth and spatial resolution at threshold, the measurement can be done in a different way. Instead of changing the number of lines per cm, we can also increase, at a fixed number of lines, the brightness modulation. See ref. 1 p.333 and ref. 2 p.16.

With the aid of a VDU, the last method is technically far more easily to realize because we can make use of the property that a VDU picture is composed of horizontal lines. On a 12 inch screen we have, with no interlacing, about 15 lines per deg. With sufficient brightness for Weber's law to hold, the threshold brightness modulation then is about 1 percent. The threshold level is measured by increasing the modulation depth from zero up until the character is recognized.

The temporal resolution test

The relation between the frequency of a flickering lightsource and the brightness modulation depth needed to resolve the source as a non steady one, is
known as the "Curve of de Lange". Ref. 1 p. 40b, ref. 2 p. 19.
The frequency at which, even at 100% modulation, no
flicker is recognized is called the critical flicker
fusion frequency (CFF).
The temporal resolution of the eye normally is measu-
red by determining the CFF of a subject.
Because of the fixed 50 Hz repetition rate of T.V.
frames it is very difficult to do this kind of measu-
rement by means of a VDU.
A simple method, however, is to measure some distinct
points of the curve, namely at 25, 12.5 and 6.25 Hz.
Because these are subharmonics of the frame repeti-
tion frequency.
To make the display test-character flickering at for
instance 25 Hz we let simply the brightness be high
in one frame and low in the next frame. For flickering
at 12.5 Hz we use to successive high brightness frames
and two low brightness frames. For 6.25 Hz it is four
and four.
In this way we can measure some samples of the
"curve of de Lange".
The measurement is done by increasing at one of the
distinct frequencies, the brightness modulation from
zero percent. on.
A further refinement is the use of shaped curves to
avoid errors caused by higher harmonic components in
the 12.5 Hz and 6.25 Hz high-low signals.
The test procedure
By programming the internal u-processor a complete
test can be arranged by composing a sequence of the
described test-elements. The standard way of storing
the test-results is to write them on a build-in mini-
cassette. By command, the data on the cassette can be
output.
Besides the generation of the test-pictures, the
device has also the possibility to generate tasks
that can be used to create fatigue in test-subjects.
An example of this is a picture as in fig. 3.
Only one of the generated numbers has a value greater
than 600. The subject is asked to key in the name
(for instance "H1") of that number.
Laboratory results
Fig. 4 and fig. 5 show the preliminary results
obtained with 28 subjects (office employees working
normally with VDU).
The scores are the number of steps needed to recognise
a character. The graphs compare results obtained just
before and after a standard visual search task pre-
sented on VDU. Increased scores mean deterioration
of visual performance. Decreased scores mean better
visual performance.
The test on VDU shows a statistical deteriorating
effect whereas traditional tasks show hardly any
effect.

Fig. 1. Test picture in the form of a character "H".
The brightness of the character increases
step by step. The level of intensity at which
the character is recognized is used as a
measure. The size of the character is 4 cm.

Fig. 2. Test picture in the form of a character "7".
The character is composed of successive more
and less bright lines while all the lines of
the background have the same brightness.

Fig. 3. Example of a task to create fatigue.
Contrast, spatial and temporal resolving capacity of the human eye as measured by a newly developed test on VDU. Differences are shown between test performances just before and just after a 2 hours' continuous standard visual search task presented on VDU (Fig. 3).

N is the number of subjects out of a total of 28.

Fig. 4. Contrast, spatial and temporal resolving capacity of the human eye as measured by a newly developed test on VDU. Differences are shown between test performances just before and just after a 2 hours’ continuous standard visual search task presented on VDU (Fig. 3).

Visual acuity and phories as measured by a Screenoscoop. Differences are shown between test performances just before and just after the standard task mentioned in fig. 4. obtained with the same subjects.

Fig. 5. Visual acuity and phories as measured by a Screenoscoop. Differences are shown between test performances just before and just after the standard task mentioned in fig. 4. obtained with the same subjects.

References:
PREDICTION AND STATE IN SIMPLE DYNAMIC SYSTEMS: WHAT IS LEARNED?

P.J. Venemans
Department of Psychology
Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

ABSTRACT

In controlling slow dynamic systems, prediction appears to be an important part of the control task, closely related to some internal representation of the system under control. This internal representation should include a state estimation of the system and a functional relationship between state and control variables.

Such a functional relationship is simple for a deterministic system consisting of a double integrator; it can be inferred from two responses of the subject in each trial. First a response P, related to the state of the system at the moment of prediction and second a response Q, the prediction of a future system state.

The results indicate that the subjects depend for their prediction of the future system state (Q), more on the visible track than on the preceding response P for actual system state. The prediction showed to be more influenced by the duration of prediction time span than by visual track duration.
PREDICTION AND STATE IN SIMPLE DYNAMIC SYSTEMS: WHAT IS LEARNED?

P.J. Venemans
Department of Psychology
Tilburg University
P.O. Box 90153 5000 LE Tilburg, The Netherlands

INTRODUCTION

In discussing the task of controlling dynamic systems, Kelley (1968) states that the H.O. makes some planning based on expected differences between what the operator thinks is going to happen and what he wants to happen. Kelley describes the first of four planning phases:

"goal conception, i.e. predicting (envisioning) possible future states of a controlled variable, a. if nothing is done to affect it, and b. if certain of the available control actions are taken".

Johanssen and Rouse (1979) state that the generation of alternative plans is followed by 'imaging of consequences' before subsequently 'choosing and initiating a plan'.

These descriptions of the planning process of the H.O. in controlling dynamic systems calls for measuring of the performance of a H.O. on the subtask of predicting future states of the controlled variable. For stochastic systems experiments about such prediction performance are published (e.g. Rouse, Tainsh, Van Heusden, Van Bussel). For prediction of the track of deterministic dynamic systems, the classic work of Gottsdanker seems appropriate (Gottsdanker, 1952, 1956). His results show that the subjects' performance with linear motion was very good, but that effects of acceleration were difficult to estimate. Acceleration tended to be underestimated or ignored. Runeson (1975) got similar results. The same effect of linearization is found with extrapolation of exponential curves (Wagenaar and Sagaria, 1975).

These experiments cover only part of the subtask predicting future states, i.e. future states under condition 'if nothing is done to affect it' and contain rather an extrapolation task than just prediction. For the other subtask: predicting future states 'if some of possible control actions are taken', the H.O. has to combine the state of controlled variable at the moment of taking action with the prediction of the effect of the control action. The H.O. has to perform a task resembling the subtask in combining the state of the system with the prediction of the effect of a control action, if the effect of a recently applied control action is not yet visible. Generally he is able to observe the system state from the output and/or some status displays. However, he cannot observe the effect of a possible control action before applying that action to the system. Sometimes a predictor display can be adapted for this purpose (Kelley: trial control). Because predictor displays are hardly in use the problem of the performance of the H.O. in predicting future states, with or without possible control actions, becomes important.

For these prediction tasks, it is assumed in the literature (Pew (1974), Jagacinski and Miller (1978), Kok and van Wijk (1976)) that a human has to his disposal some form of internal representation of the system under control, the so called 'internal model'.

Jagacinski et. al. (1977, 1978) describe an experiment in which they train subjects to stop a simulated pendulum at a certain point with a single switch action. They state that for the behavior of their subjects "a more plausible explanation is not simply a set of conjunctive cues, but a functional relationship that relates system acceleration to position, velocity and applied force" (Jagacinski et. al., 1977, p. 221).
In their experiment it is not possible to measure in any direct way the use of such a functional relationship, separated from the performance in perceiving the state variables (position, velocity, etc.) to be used in such a relationship. This perceiving of the state of the system will strongly depend on display type and configuration. For other aspects of the prediction task the subjects should be sufficiently trained to become experienced with that particular system, and the control amplitudes available as system input. This experience enables the subjects to form some internal representation. A functional relationship as indicated by Jagacinski can be part of the structure of that internal representation.

For control systems consisting of a double integrator ($H(s) = \frac{K}{s^2}$) the state information consists of velocity and position, and the functional relationship which can be used for prediction, is a simple one: adding a constant distance to the predicted position if track should continue with constant velocity (zero acceleration) for a perceivable acceleration amplitude and prediction time span. From two responses in each trial, one the predicted positions $P$, if the velocity from that moment should remain constant (thus depending on the subject's estimates of velocity and position); and second the predicted position $Q$, if the acceleration remains constant throughout the trial, use of such a relationship can be inferred. If subjects use such a relationship, the errors of the two responses $P$ and $Q$ should be positively correlated, and the difference ($Q-P$) should approximate some constant related to the amplitude of acceleration used.

So we decided to test this in a couple of small experiments. If in these experiments a dependency of responses $Q$ on responses $P$ could be shown to develop with training, a third experiment was going to be done with the same subjects to measure the generalisation of this relationship to the predictions 'if certain of the possible control actions are taken'.

**METHOD**

**Apparatus and procedure**

The system dynamics ($K/s^2$) were simulated on a small analogue computer (Vidac 169), with some programmable timers, digital logic, beep generator, analogue switches and a small scratch card reader added. The cards contain the sequence of the experimental conditions for each block of trials of the experiment. The output of the simulated system was connected to the $Y$-input of an X-Y recorder (HP 7004B). The X-input was connected to an external time base at a rate of approximately 6.5 mm/sec. The track was drawn on blank recording paper. At the start of the trial the pen of the X-Y recorder was set to its starting position at the left side of the recorder. The $Y$-coordinate of this starting position was constant during both experiments. The subject was seated at normal reading distance in front of the X-Y recorder. He had two push buttons available, one to start a new trial when he was ready for it, and one to indicate that he had finished the prediction responses. After start of the trial the pen travelled to the starting position, came down and started drawing the track of that trial. In all trials the track ran from left to right, leaving free sight on the track drawn. A vertical reference line was visible through the paper at the right side of the recorder. The track ran continuously through four intervals (see figure 6). The first interval (called history, typically 24 sec) started with the start of the track, the end marked with a short lift of the pen and an auditory signal. This signal informed the subject that the second interval (response interval, 6 sec during almost all trials) had started. In this interval the subject had to mark first the predicted position $P$ and then the predicted position $Q$. With the second push button he indicated that he had completed the responses. If the button was not pushed in time, the recorder stopped at the end of the response interval.
After the 6 sec of the response interval (or at stop, triggered by
the push button) the third interval (called prediction time span,
typically 13 sec) covered the duration to the track reference line.
When the pen crossed this line a beep was given in order to stimulate the
subject to compare his predictions with the actual track. The fourth interval
covered subsequently the track from the reference line until 3 sec further.

Subjects
The subjects were 4 university students, 1 female, 3 males, age 21-24 years.
None of them had any experience with this type of experiments.
At the start of the experiment they received a written introduction and a demon-
stration of the task required. After a couple of practice trials the subjects
were instructed how to change paper after each trial.

Design
In experiment 1 the durations of the track intervals were fixed; the history
duration was 24.0 sec., the response interval was 6.0 sec., the prediction
time span being 13.0 sec.
Five values for the starting velocity in vertical direction were used: -4.4,
-2.9, -1.5, -0.3 and +0.7 mm/s for acceleration upwards (a > 0) and the same
velocities, but with changed signs for acceleration downwards (a < 0). These
10 combinations were presented in each block in random sequence. The curvature
within a block was fixed.

Two small values for the curvature of the track were used, a low one for an
acceleration |a| = 0.025 mm/s² and a high one for acceleration |a| = 0.067 mm/s².
These curvatures were easily discriminable.
These small values for the acceleration input caused the system to change slowly.

The six sessions of experiment 1 contained each two blocks with a high curvature
and two blocks with a low curvature, in a-b-b-a sequence, starting alternatively
with high or low curvature. In order to avoid warming-up effects, each session
started with two or four practice trials. Each subject took one or two sessions
each day, with a minimum separation between the sessions of two hours.

In experiment 2 the effects of shortening the duration of history or of changing
prediction time span were studied, for the same subjects of experiment 1. In each
session now either high or low curvatures were presented, with the starting
velocities and X-positions adapted to get the tracks comparable with the tracks of
experiment 1. Starting Y-positions were fixed however.
For each block a new value for either the history or the prediction time span was
chosen (see table I). In the session 7, 8, 9 history durations of 24, 12 or 6
sec were used. In sessions 10, 11, 12 history durations of 6 or 1 sec were used,
with the response interval shortened to 2.0 sec (and thus causing a temporal stop
in most trials). In sessions 13, 14, 15 history duration was set back to 24 sec
and the prediction time span set on 7, 13 or 25 sec.

Rewards and KR
In addition to the KR during each trial a performance score was calculated after
the session, based on the cumulated absolute Q-error for the session. This score
was shown to the subject and recorded on a public score-board.
With better performance the subjects earned more lottery tickets, according to
a lottery system which appeared to motivate the subjects (Van Bussel, 1980).
Extra tickets could be earned when the subjects made their predictions in time
in at least 50 % of the trials of the session. The subjects earned an extra money
bonus (f 2,50 DFl.) each time the performance score was a personal record.
TABLE I

Conditions used in blocks of experiments 1 and 2. In each block of trials the curvature and track interval durations were fixed. In each block the track accelerated in 5 trials upwards (a > 0) and in 5 trials downwards (a < 0). P-reference corrections were calculated from tangents at middle of response interval, i.e. 3 sec later. If response interval lasted less than 3 sec, corrections from tangents at end of interval were calculated.

<table>
<thead>
<tr>
<th>session history</th>
<th>curvature in mm/s²</th>
<th>response interval in sec</th>
<th>prediction time span in sec</th>
<th>P-reference corrections in mm</th>
</tr>
</thead>
<tbody>
<tr>
<td>experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 to 6</td>
<td>0.067</td>
<td>24.0</td>
<td>6.0</td>
<td>13.0</td>
</tr>
<tr>
<td>experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7,8,9</td>
<td>{0.067, 0.025}</td>
<td>{6.0, 12.0, 24.0}</td>
<td>6.0</td>
<td>13.0</td>
</tr>
<tr>
<td>10,11,12</td>
<td>{0.067, 0.025}</td>
<td>{6.0, 2.0, 1.0}</td>
<td>2.0</td>
<td>17.0</td>
</tr>
<tr>
<td>13,14,15</td>
<td>{0.067, 0.025}</td>
<td>{24.0, 6.0, 13.0, 25.0}</td>
<td>6.0</td>
<td>13.0</td>
</tr>
</tbody>
</table>

Analysis
The tracks of each trial and the response marks P and Q were fed into a microcomputer (DEC LSI-11) using a graphic X-Y tablet (with grid resolution of 0.25 mm) and recorded on floppy disks for further analysis. Errors in the P and Q prediction responses were computed and changed signs for trials with a < 0 to make the errors comparable for tracks with a > 0. A positive error indicated for a > 0 (upward curvature) a too high prediction.

For the error in the P-response a reference was calculated using the tangent to the track at the response interval (see figure 6). Since this interval lasted some time the start of the interval is taken as basis for calculations, because the instruction to the subjects requires them to use this point for the P-response, this point clearly visible at the moment of responding. When the middle of the response interval is taken, some correction should be subtracted from the P-responses, as indicated in the figures. Values are shown in table I. Such a correction being constant does not affect the variance. For each block of trials the data of all subjects are averaged.

RESULTS

Results of experiment 1 are shown in the figures 1 and 2. In figure 1a, b mean and variance are given for the P-response error, in figure 1c, d the same for the Q-response error. In figure 1e, f mean and variance for the differences between the Q and P responses in each trial are given.

The results of experiment 2 are presented in the figures 3, 4 and 5. In all figures mean and variance are given for respectively the P-response error, Q-response error and the difference in the responses Q and P.

In addition for experiment 1 figure 2 shows the correlation and regression coefficients between the Q and P response errors.
DISCUSSION

Experiment 1.

1. It was hypothesized that the results in both tasks with different values for the acceleration (0.067 and 0.025) would resemble each other. So variances and possible biases, should show the same tendencies, with greater biases and variances for the stronger curvature.

Figures 1b and 1d show the variances for the low a-value to be nearly as high as for the high a-value, for both the P and Q responses, higher for the P-responses than for the Q-responses.

For estimates of the future track (Q-responses) there should be a tendency to linearization (underestimation of the curvature), causing a negative bias. Figure 1c shows a negative bias for the Q-responses on high a-valued tracks, as expected. On low a-valued tracks, however, biases in Q-responses have about the same amplitude, but are in positive direction. This unexpected positive bias indicates an overestimation of the curvature of the a-value used in that block of trials.

For the P-responses also a positive bias is found. If the reference tangent for the P-response is calculated at the middle of the response interval (reference correction values are shown in table I), this bias show a pattern comparable to the one found for the Q-response bias. This indicates that P and Q-responses are influenced in the same way by the value of the acceleration.

2. It was expected that with increased experience in the task, the variances of both the P and Q responses would decrease. Figures 1b and 1d show that for the blocks of session 5 and 6, compared to the blocks of sessions 1 and 2, the variance reduction is limited to about 3:1, however no asymptotic behavior is reached.

Increased experience with the system used should cause a decrease in the response bias. Figures 1a and 1c do not show such a decrease.

3. When the subjects rely for their responses on some infernal model, of the form given above, their response positions should show a constant distance from each other, the prediction time span being constant. For the actual track this distance had the value +12.0 for high acceleration, and +4.5 for the low accelerations. If the tangent is calculated from the middle of the response interval these distances become 8.2 and 3.2 respectively.

The variance of the difference Q-P should be lower than the variance of Q alone.

Figure 1e shows a tendency for the average of Q-P to become constant, being 7.2 and 3.5 respectively, thus close to the distance between the tangent from the middle of the response interval and the actual track. However, figure 1e shows higher variance for Q-P than figure 1d for Q-response errors alone.

Another measure of a possible dependency between the responses P and Q is the correlation coefficient between the errors in P and Q from the same trial. Figure 2a shows these correlations for each block. The errors do not appear to be strongly correlated, although most of them are significant as indicated. For the low accelerated tracks the responses appear to be more correlated than for high accelerated tracks. Figure 2b shows regression for Q-response errors on P-response errors. It shows roughly the same pattern. An interesting result seen in the figure is the shifts in the correlations between blocks with the same conditions. This can be a transfer effect between the blocks, especially influencing the responses for tracks with low acceleration.
Conclusions experiment 1

From these results it appears that for the subjects with this dynamics \((K/s^2)\) and this display \((x=t)\) it was easier to extrapolate a parabolic curve than to make predictions for a change in acceleration which asks for linear extrapolation based on the system state. Apparently the subjects estimate the \(Q\)-responses directly from the visible track and did not use the \(P\) estimate which was made first. Therefore an indication of the use of the functional relationship mentioned above is not found. This indicates that for the subjects in this experiment either the \(P\)-response is not a measure for the state-estimates, or the state-estimates are not used in extrapolating the track. In the first case the question remains how to measure appropriately the subjects' state-estimates.

From the overestimation of the low accelerated tracks it appears that the subjects have learned some 'average' curvature, although both curvatures were easily discriminable from each other and not presented in a mixed sequence. The transfer effect found in the correlation coefficient also favors this interpretation.

The third experiment intended to measure the generalisation to predictions 'if control actions are taken', was cancelled based on these results.

Experiment 2

Except for the very short history duration \(h = 1\) sec, no main effects are found for shortening the visible track. The increase for the \(P\)-response bias with high accelerated curves is remarkable. The differences found for \(h = 6\) sec in comparing the figures 3\(^a\) and 4\(^a\), respectively 3\(^c\) and 4\(^c\), can be explained from the shortening of the response interval in the latter. The \(P\)-responses should not be influenced by shortening the history from 6 to 1 sec, because no perception of curvature is needed. Figures 3\(^a,b\) and 4\(^a,b\) confirm this. The increase in variance of the \(Q\)-response errors can be explained by a misperception of the direction of the curvature (upward or downward), causing large errors.

With increase of prediction time span the error variance strongly increases (figures 5\(^b,d\)). Surprisingly little influence is found on the error-bias of the \(Q\)-responses. The effect on the \(P\)-responses is higher than expected: \(P\)-responses should be based on linearised tracks, so a linear increase with time span was expected. Perhaps the subjects show an overestimation of the duration of the long time span used.

Conclusions experiment 2

For the dynamics \((K/s^2)\) and the display used, durations of the visible history of the tracks seem not to be important after the experience with the tracks in experiment 1, provided the direction of the curvature can be perceived. However the prediction time span appears to be important. So in case where subjects have to choose the right moment to apply a control input the subtask predicting the future track becomes easier if the subjects delay their control actions. Therefore one might expect that in some (difficult) control tasks subjects tend to delay the action, not only because of too long RT's, but also in order to reduce perceptual uncertainty.
LITERATURE


Figure 1, a,b, c,d,e,f
1a,b. Experiment 1. Mean and variance of P-response error, averaged for each block. The horizontal lines indicate the correction of the P-response reference.

1c,d. As 1a,b. Mean and variance of Q-response error.

1e,f. As 1a,b. Mean and variance of response difference Q-P.

2a. As 1a. Correlation coefficient between Q and P-response error pairs. Significance on 1% level is indicated with ★.

2b. As 1b. Regression coefficient for Q-response error on P-response error.
Figure 3a,b,c,d,e,f
Experiment 2, with history duration varied.
(a,b) (left) Mean and variance of P-response error averaged for each block of sessions 7, 8, 9. The horizontal lines indicate the correction of the P-response reference.
(c,d) (centre) As a,b. Mean and variance of Q-response error.
(e,f) (right) As a,b. Mean and variance of response difference P-Q.

Figure 4a,b,c,d,e,f
Experiment 2, with history duration 6 and 1 sec and response interval shortened to 2 sec.
(a,b) (left) Mean and variance of P-response error averaged for each block of sessions 10, 11, 12. The horizontal lines indicate the correction for the P-response reference.
(c,d) (centre) As a,b. Mean and variance of Q-response error.
(e,f) (right) As a,b. Mean and variance of response difference Q-P.
Figure 5
Experiment 2, with prediction time span varied: sessions 13, 14, 15.
(a, b) (left) Mean and variance of P-response error for each block. The horizontal lines indicate the correction for the P-response reference.
(c, d) (centre) As a, b. Mean and variance of Q-response error.
(e, f) (right) As a, b. Mean and variance of response difference Q-P.
The data point marked with † has a value of 18.8 mm.

Figure 6
Track with intervals indicated. Track has grown from left to right, typically in 46 sec. Response marks P(x) and Q(8i) are shown. The reference tangents used for calculation of P-response error were not present actually.
(a) Q-response error (shown with negative value).
(b) P-response error (negative too) with reference tangent at start of response interval.
(c) P-response with reference tangent at middle of response interval.
(b-c) P-reference correction.
session 3
supervisory control
decision making
fault management

chairman: J. Rasmussen, Denmark
PROBLEM SOLVING BEHAVIOR OF PILOTS IN ABNORMAL AND EMERGENCY SITUATIONS

Gunnar Johannsen
Research Institute for Human Engineering (FAT)
D-5307 Wachtberg-Werthhoven, F.R. Germany

William B. Rouse
University of Illinois
Urbana, Illinois 61801 U.S.A.

ABSTRACT

The planning behavior of professional aircraft pilots when dealing with abnormal and emergency situations was studied. Events simulated included engine failure, loss of hydraulic pressure, and runway closure due to weather. Both manual and automatic control were considered. Subjective assessments of workload were also collected. The combination of planning measures, workload ratings, and flight performance data is used to develop an overall perspective of problem solving behavior of pilots in abnormal and emergency situations.

INTRODUCTION

A problem solving situation arises when a human must consciously search through a set of possible solutions. This has been characterized as a mixture of planning and control (Ref. 1). Planning involves searching for a sequence of actions that will potentially lead to problem solution. Control includes execution and monitoring of this sequence of actions.

In a review paper (Ref. 2), the authors concluded that planning will become an increasingly central function for the human as automation takes over more and more of the manual tasks. It was also concluded that the human's planning process is not particularly well understood. Therefore, a series of experiments was started to increase the understanding in this area.

The design and the results of the first experiment were reported in Ref. 3. It was proposed that the human operator's planning process is hierarchical in nature, with the degree of detail in the hierarchy ranging from broad and sketchy to narrow and concise. The depth of planning, with respect to a particular task, was hypothesized as being related to the timeliness and criticality of the task as well as the anticipation of any increased difficulty in successfully performing the task. The experimental study of the planning process of professional pilots during normal, emergency, and abnormal flight scenarios was discussed. Several questionnaire techniques that were developed for the study were reviewed.

The results of the study led to some fairly interesting conclusions. First, it appeared that the online questionnaire technique for depth of planning worked fairly well. More importantly, it was found that depth of planning increased during abnormal situations when workload was relatively low. The unanticipated confounding of these two variables provided immediate and sufficient motivation for another experiment which will be discussed in this paper. Depth of planning
was also found to be strongly related to probability of increased difficulty and weakly related to criticality.

An important conceptual result was the differentiation between event-driven and time-driven planning. This idea will probably be an important ingredient in any model of the human's planning process. It is also interesting to note that the individual differences found in the study appeared to relate for the most part to event-driven rather than time-driven planning.

The next stage of the research reported in this paper is a replication of the study of Ref. 3, with modifications of the flight scenarios and procedures to reflect what has been learned from the initial study. Further, the correlations of planning activity with objective measures of performance and subjective assessments of workload have been determined.

THE TASK

Like in the first experiment, the planning process of aircraft pilots has been investigated. An HFB 320 Hansa Jet simulator at the Research Institute for Human Engineering in the Federal Republic of Germany was employed (Ref. 4). The HFB 320 is a 5 - 12 passenger, twin engine jet used for both military and commercial purposes. It normally has a two-man crew.

The flight simulator allows full maneuverability, is fixed base, and has no visual simulation of the outside view. The cockpit is an original mockup from the aircraft manufacturer. It is instrumented with conventional displays for flight, engines, and navigation as well as controls that include steering force simulation. Also, a fairly sophisticated autopilot as well as a flight director with V-bar indicators in the artificial horizon are available. However, some limitations are present as the flight instruments for the copilot, the controls in the overhead panel, and a simulation of waypoints for navigation are missing. These limitations restrict the possibilities for simulating emergency situations. Further, it was necessary to run the experiments with a second experimenter playing the combined role of the copilot and the air traffic controller.

Using this simulator, three professional HFB 320 pilots flew several 20 minute missions from cruise to touchdown (TD). In order to resolve the confounding between the influences of flight scenario and level of automation found in the first experiment, a 5 x 2 factorial experimental design for this second experiment was chosen. One factor was level of automation whose two levels were manual and autopilot (both levels included flight director and autothrottle). The second factor was scenario whose five levels included one normal flight (N), two flights with abnormal situations (A1 and A2), and two flights with emergency situations (E1 and E2).

The "normal" flight scenario N which was the basic one in this study is illustrated in Figure 1. There, eight flight phases are shown, namely: 1) Cruise, 2) Descent, 3) Holding, 4) Initial Approach, 5) Final Approach, 6) Landing, 7) Ground Roll, and 8) Cruise to Alternate. The overall mission of the N scenario lasted approximately 20 minutes whereat no cycles of the holding pattern or cruise to alternate were to be flown. No unusual events occurred.
Abnormality $A_1$ involved a temporary runway closure due to snow removal, which was announced 4.2 minutes into the flight and presented the possibilities of requiring the pilot to enter the holding pattern or to cruise to the alternate airport. Abnormality $A_2$ involved temporary Cat-III conditions due to a dense fog, which was announced 7.5 minutes into the flight and presented the same possibilities as abnormal situation $A_1$. While the possibilities of holding or cruising to an alternate were clearly shown on the map furnished to the pilot (and available in the cockpit), the abnormal situations were always resolved at the last minute and holding and cruising to an alternative always avoided. In this way, all flights were limited to 20 minutes.

Emergency $E_1$ involved the failure of the right engine at 4.2 minutes into the flight. The failure was announced by an alarm in the cockpit similar to that in real HFBs. In this way, there was no problem with the pilot missing the failure. Emergency $E_2$ involved a severe loss of hydraulic pressure due to a total loss of hydraulic fluid, i.e., even the hand pump for the gear was inoperative and an emergency landing was requested. This failure was announced at 7.5 minutes into the flight by the alarm in the cockpit.
THE EXPERIMENT

The objective of this study was to measure the planning process of pilots during the different flight conditions just described. The notion of depth of planning has been introduced as the basic concept for this purpose. By depth, we mean level of detail which can range from broad and sketchy to specific and concrete.

Depth of planning was measured by an online questionnaire technique. As the flight proceeded, the pilots received verbal queries concerning the depth of planning associated with the present and future flight phases and three selected subtasks for each flight phase. Depths ratings were made using the 10-point scale shown in Table 1. This should lessen the occasional "chattering" between, e.g., ratings of 2 and 3 as obtained in the first experiment with a 5-point scale.

The flight task of the pilots should be disturbed as little as possible by the online questionnaire. Therefore, the pilots were thoroughly familiarized with the complete questions and possible responses during the instructions before the flights. The associated text explanations for some of the possible responses in Table 1 should serve only as an aid for getting a feeling of the scale. During the experiments, the pilots only heard the short names of the flight phases and subtasks, e.g., "Final Approach". The answers were coded by the numbers 1 to 10 of Table 1 which were the only verbal responses of the pilots.

Table 1. Questionnaire for depth of planning

<table>
<thead>
<tr>
<th>Depth of Planning</th>
<th>To what extent are you planning with respect to the flight phase or subtask?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>not at all</td>
</tr>
<tr>
<td>2</td>
<td>generally aware of task</td>
</tr>
<tr>
<td>3</td>
<td>overall qualitative assessment only</td>
</tr>
<tr>
<td>4</td>
<td>specific information needs</td>
</tr>
<tr>
<td>5</td>
<td>considering specific actions</td>
</tr>
</tbody>
</table>

Depth ratings were only made for flight phases 4, 5, and 6 (Initial Approach, Final Approach, and Landing). These 3 phases with their 9 subtasks constituted a set of 12 possible queries to the pilot. All queries were randomly and independently chosen from the set of 12, with the exception that the 3 flight phases were twice as likely to be chosen as the 9 subtasks. Queries occurred every 20 seconds with air traffic and navigational information supplied in the intervals. Thus, with 20 minute flights and 3 questions per minute, there were 60 questions per flight, 8 for each flight phase and 4 for each subtask.
Dependent variables in the experiment included not only depth of planning but also workload and performance. Subjective workload was assessed with rating scales as in the first experiment.

No single measure of performance seemed appropriate for the entire flight. However, the pilot's control signals in terms of elevator, aileron, and rudder angles can be viewed as indirect measures of performance. This is similarly true for the attitude signals in terms of pitch and roll angles. Certainly, any deviation from the desired flight path has to be corrected for using one of these controls and changing the attitude of the aircraft. However, these controls vary even for flights that stay exactly on the desired flight path. Thus, a baseline is needed with which to compare measures of control activity and attitude. A good choice is to use the same measures applied to the autopilot's activities as a baseline.

As a result of this consideration and after some experimentation with data from the first experiment, the square roots of sums of variances, with respect to time-varying means, integrated over flight phases \( j \) and divided by their time duration \( T_j \) were chosen as scalar performance measures for both, control actions \( \sigma_c \) and attitude \( \sigma_A \), e.g., for control:

\[
\sigma_{C,j} = \frac{1}{T_j} \sum_{t=t_0}^{t_0+T_j} \sqrt{\sigma_1^2(t) + \sigma_2^2(t) + \sigma_3^2(t)}.
\]

The variance for a particular control \( u_i(t) \) over a time window of length \( T = 10 \) s is given by the equations:

\[
\sigma_i^2(t) = \frac{1}{T-1} \sum_{\tau=t-\frac{T}{2}}^{\tau=t+\frac{T}{2}} \{u_i(\tau) - \bar{u}_i(\tau)\}^2,
\]

\[
\bar{u}_i(t) = \frac{1}{T} \sum_{\tau=t-\frac{T}{2}}^{\tau=t+\frac{T}{2}} u_i(\tau).
\]

The sum of variances of elevator, aileron, and rudder angles was taken for control actions. The variances of pitch and roll angles were calculated correspondingly to the above equations and summed for the attitude measure.

Three subjects participated in the experiment. During the flights, depth of planning and flight performance were measured as explained in the preceding paragraphs. The treatments for the 3 subjects were the 5 flight scenarios \( (N, A_1, A_2, E_1, E_2) \) combined with the 2 levels of automation, i.e., manual (M) and autopilot (A), which are described in the earlier discussion of the task. The 5 x 2 factorial experimental design actually used is shown in Table 2 with tests \( T_1 \) through \( T_{10} \). The test \( T_{11} \) was added in order to investigate the
Table 2. Experimental design

<table>
<thead>
<tr>
<th></th>
<th></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>Nsubscript M</td>
<td>A.subscript 1A</td>
<td>E.subscript 1M</td>
<td>A.subscript 2A</td>
<td>E.subscript 2M</td>
<td>Nsubscript A</td>
<td>A.subscript 1M</td>
<td>E.subscript 1A</td>
<td>A.subscript 2M</td>
<td>E.subscript 2A</td>
<td>ME.subscript M</td>
<td></td>
</tr>
<tr>
<td>Nsubscript M</td>
<td>E.subscript 1A</td>
<td>A.subscript 1M</td>
<td>E.subscript 2A</td>
<td>A.subscript 2M</td>
<td>Nsubscript A</td>
<td>E.subscript 1M</td>
<td>A.subscript 1A</td>
<td>E.subscript 2M</td>
<td>A.subscript 2A</td>
<td>ME.subscript M</td>
<td></td>
</tr>
<tr>
<td>Nsubscript M</td>
<td>A.subscript 1A</td>
<td>E.subscript 1M</td>
<td>A.subscript 2A</td>
<td>E.subscript 2M</td>
<td>Nsubscript A</td>
<td>A.subscript 1M</td>
<td>E.subscript 1A</td>
<td>A.subscript 2M</td>
<td>E.subscript 2A</td>
<td>ME.subscript M</td>
<td></td>
</tr>
</tbody>
</table>

influence of multi-events (ME) during manual flight control (M). This scenario was a combination of the scenarios E.subscript 1, A.subscript 1, and E.subscript 2. An engine failure (E.subscript 1) occurred at 4.2 minutes into the flight, followed by a runway closure (A.subscript 1) at 5.2 minutes and a hydraulic failure (E.subscript 2) at 7.5 minutes.

The experimental procedures will now be briefly described. First, the subjects became familiarized with the special features of the flight simulator by briefing and practicing as well as with the instructions for the experiment. The instructions were written in the pilot's native language, i.e., German. Also, the pilots answered a questionnaire concerning their flight experience which was for all three pilots between 1000 and 2000 hours on the HFB 320 aircraft with almost no experience on a flight simulator. The subjects responded also to subjective scales for criticality, thereby getting acquainted with the flight phases and their subtasks and answering the question how critical these are to overall mission success. Then, the tests T.subscript 1 through T.subscript 11 were performed.

After each test, the pilots estimated their experienced workload using appropriate subjective scales, similar to those for criticality, for each of the flight phases. The whole experimental session ended for each pilot with a final interview in which he was asked to express his experiences with the experiment and to comment on some elements of his behavior observed by the experimenters.

RESULTS

Three sets of analyses of variance (ANOVA) were performed with the following independent variables:

1. Scenarios denoted by S; includes five levels (N, A.subscript 1, A.subscript 2, E.subscript 1, and E.subscript 2).
2. Flight phases denoted by P; includes cruise, descent, etc.
3. Level of automation denoted by A; includes two levels (manual and automatic).
The dependent variables considered were:

1. Depth of planning denoted by D on a scale of 1 to 10.
2. Workload denoted by W on a scale of 1 to 7.
3. Flight performance denoted by $\sigma_C$ and $\sigma_A$ (in degrees) which are the standard deviations of control activity and aircraft attitude with respect to a time-varying mean.

Initial analyses indicated that comparisons were best made for the period 4:00 to 14:40 because: 1) prior to 4:00, all flights were equivalent except for the availability of autopilot, 2) after 14:40, the abnormalities and emergencies inherently differed because the abnormalities were resolved (i.e., "runway open") while the emergencies were not.

Within this period, the flight phases or subtasks which consistently received fairly high depth ratings were: initial approach, final approach, weather minima, landing, and runway condition. Further, since the number of depth queries for weather minima and runway condition included only two each in the period of interest, only initial approach, final approach, and landing were used in the following analyses. Thus, there were three levels of independent variable P.

The ANOVA for average depth of planning indicated that S was the only significant effect ($F_{4,60} = 4.22, p < .005$), averaging 4.82, 6.54, 6.08, 5.79, and 6.64 for N, A, A', A'', E', and E'', respectively. In order to isolate the differences between abnormal and emergency scenarios, the N scenario was dropped and the ANOVA repeated, yielding a significant S x A effect ($F_{3,48} = 2.99, p < .05$). The nature of this interaction is shown in Table 3.

### Table 3. Average depth of planning as a function of scenario and level of automation

<table>
<thead>
<tr>
<th>Automation</th>
<th>Scenarios</th>
<th>E'</th>
<th>E''</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual</td>
<td>6.91</td>
<td>5.79</td>
<td></td>
</tr>
<tr>
<td>Automatic</td>
<td>5.74</td>
<td>6.67</td>
<td></td>
</tr>
</tbody>
</table>

The ANOVA for workload indicated that S ($F_{4,60} = 9.42, p < .001$), P ($F_{2,60} = 9.29, p < .001$), and A ($F_{1,60} = 4.51, p < .05$) were significant effects. More detailed analysis showed that scenarios N, A_1, and A_2 were not significantly different among themselves, averaging 4.00, but were significantly different from scenarios E_1 and E_2 which averaged 5.49.

The ANOVA for $\sigma_C$ indicated that S and P were the only significant effects ($F_{4,60} = 6.15, p < .001$ and $F_{2,60} = 13.17, p < .001$, respectively). For $\sigma_A$, ...
the significant effects were S, P, and A (F_{4,60} = 8.35, p < .001; and F_{2,60} = 6.19, p < .005; and F_{1,60} = 5.28, p < .05, respectively). Scenario E resulted in \( \sigma_C = 1.45^\circ \) and \( \sigma_A = 4.39^\circ \) while the other four scenarios resulted in similar, averaging \( \sigma_C = 0.77^\circ \) and \( \sigma_A = 2.01^\circ \). The autopilot reduced \( \sigma_C \) from 0.99° to 0.82° and \( \sigma_A \) from 2.87° to 2.10°.

Considering correlations among D, W, \( \sigma_C \), and \( \sigma_A \), it was found that \( \sigma_C \) and \( \sigma_A \) were highly correlated (r = 0.98, p < .05). Correlating workload with flight performance, the correlation of W and \( \sigma_C \) was r = 0.56 (p < .05) while the correlation of W and \( \sigma_A \) was r = 0.49 (p < .05). Depth of planning was not significantly correlated with W, \( \sigma_C \), or \( \sigma_A \).

**DISCUSSION AND CONCLUSIONS**

The significant S x A effect illustrated in Table 3 is very interesting. Why does the availability of autopilot result in decreased planning during abnormal scenarios and increased planning during emergency scenarios? While one might postulate this effect to be a by-product of the lower workload with the abnormal scenarios, the absence of a correlation between depth of planning and workload does not support this hypothesis.

This unusual effect of autopilot on depth of planning can perhaps be explained by the nature of the abnormal and emergency scenarios. The abnormalities involved changes in the environment (i.e., runway closure due to snow or fog) while the emergencies involved changes within the aircraft (i.e., engine failure or loss of hydraulic pressure). Despite these differences, the average depths of planning were remarkably similar, averaging 6.31 and 6.22 for abnormal and emergency scenarios, respectively. Yet, the autopilot did have a differential effect on planning.

One can conjecture that the key to this somewhat counterintuitive result is the effect of the autopilot on the types of event. The autopilot controls the aircraft but not the environment. Therefore, the autopilot can help to compensate for events within the aircraft but cannot directly affect events in the environment. Thus, when an engine failure or loss of hydraulic pressure occurs, the autopilot can help to compensate and thereby, free the pilot to plan the course of actions necessary to deal with the failures. As a result, the availability of autopilot during such emergencies results in increased planning.

During abnormal situations such as the runway closures used in this experiment, the pilot's main task is to hold and wait. While some planning might be associated with the possibility of diverting to an alternate airport, this possibility was not heavily stressed in the experiment. If the autopilot is available, it performs much of the "holding" task and the pilot has little left to do. As a result, planning decreases.

To summarize the conjecture offered here, during emergencies the autopilot frees the pilot to devote more time to planning; during abnormalities the
autopilot assumes a significant portion of the task and lessens the need for planning. While the notion is speculative and needs more investigation, it does serve to emphasize the possible subtle effects of automation. Another interesting result is the lack of correlation between workload and planning. While the significant correlation between workload and control activity agrees with one's intuition, the fact that an increased need for planning does not affect perceived workload is rather counterintuitive. It is quite possible that the pilots perceived workload in terms of having to do something and, since planning is an internal activity, they did not associate planning with work. Alternatively, this result can be viewed as evidence that workload is a multidimensional concept that cannot be reduced to a scalar metric.

ACKNOWLEDGEMENT

This research was partially supported by the National Aeronautics and Space Administration under Ames Grant NSG-2119. The authors gratefully acknowledge the assistance of K. Hillmann, W.-D. Huland, and W. Krüger in this work.

REFERENCES


SIMULATION OF THE PILOT'S LONG TERM STRATEGY DURING IFR FLIGHTS

By D. SOULATGES

OFFICE NATIONAL D'ETUDES ET DE RECHERCHES AEROSPATIALES,
29 Avenue de la Division Leclerc, 92320 CHATILLON (France)

Abstract

Research on the problem of modelling the human pilot behaviour was conducted at ONERA since 1973. It led to a pilot model simulating the short-term safety loop. Accordingly, control of the trajectory is simulated during particular flight-phases, while the piloting objectives as well as the aircraft state and the environment state remain constant.

An important task for the pilot that remains to be studied is the long-term safety loop. It is the object of the work presented here. It was performed in 1977 and was centered on commercial IFR flights. To simulate a complete flight, the proposed model may call for simulation programs that are specific to each piloting phase. It may also be used in order to automatically generate sufficiently risky flight phases but yet sufficiently probable to require close attention. Last it may be used to generate flight scenarios for flight simulator crew testing.

The aircraft dynamical model used here is most simplified. Yet, it includes a number of sub-systems with associated failure rates. The environment model carries, in addition to a wind vector, a model of ground équipements and régulations, as summarized on navigation Charts.

Decisions are made by calling upon a procedure list, i.e normal IFR procedures, but also a fall-back of after-failure procedures as set by the flight crew manual. When no incident occurs, the flight progresses according to schedule. The latter may be modified by application of after failure procedures.

A numerical application, dealing with to and fro flights on a french domestic trip, is presented here; these flights are simulated from brakes released to full stop. Failure rates of all systems have been homogeneously increased in order to provide an artificially augmented frequency of flights with fault, during simulation, thus inducing a wide variety of after failure situations, with possibly unexpected ones.

1 — Introduction

Most of the work carried out in the field of modelling human pilot behaviour deals with the short term piloting loop, which makes the performance on a flight phase. During each phase, the aircraft state, that of the environment and the piloting objective remain constant.

It seemed useful to build a model of the long term piloting loop making it possible to choose the piloting objectives at each phase, and to match the phases making up the complete flight.

2 — The model

2.1 — Definition of the model

A local, ground-based reference frame is assuming a constant magnetic deviation, defined by the axes:

- $\vec{OX}$, horizontal, positive towards magnetic north,
- $\vec{OY}$, horizontal, positive towards magnetic east,
- $\vec{OZ}$, vertical, positive downwards.
The origin O is located at zero altitude, at the vertical of a radio-navigation ground station.

Aircraft heading $H_A$ is an angle, in the horizontal plane, counted from a vector equipollent to $\vec{OX}$; positively from 0 to 360°, in the clockwise direction (Fig. 1).

![Fig. 1 — Magnetic heading.](image)

Wind is defined by the modulus $W$ of its velocity in the ground frame, and by the direction from which it blows $H_W$ (Fig. 2). It is assumed horizontal and constant with altitude.

![Fig. 2 — Wind vector.](image)

Trajectory is controlled through aircraft aerodynamic speed. This vector is determined by

\[
\vec{A} = \begin{cases} 
S_a & : \text{air speed, in the (xoy) plane}, \\
\frac{z}{2} & : \text{aircraft vertical speed}.
\end{cases}
\]

Aerodynamic speed components are:

$A$ : aerodynamic speed modulus,
$H_A$ : aircraft heading,
$\gamma_a$ : aerodynamic speed slope (angle between horizontal plane and vector $\vec{A}$).

The aircraft speed in the ground frame (true speed) is then obtained by the sum

\[
\vec{S}_g = \vec{A} + \vec{W}
\]
At constant air speed, aircraft trajectory is controlled by performing "standards turns", at a constant rate of 3°/s, a complete turn (360°) being performed in 2 minutes. The turning control can thus be reduced to

\[ t_s = \begin{cases} 
-1 & \text{left turn}, \\
0 & \text{neutral}, \\
+1 & \text{right turn}.
\end{cases} \]

Angle being expressed in degrees and the time in seconds, we add to the state equations:

\[ \dot{H}_A = 3 \cdot t_s. \]

2.2 — Control in the horizontal plane

The aircraft is equipped with radio-navigation receivers, and its position is expressed relative to those of the ground stations. The information is either angular (general case) or a distance (DME).

The angles used are the following (Fig. 3):

- QDM : magnetic bearing towards the station,
- QDR : magnetic bearing of the aircraft, from the station,
- Gt : relative bearing of the station, from the aircraft.

Each of the radio-navigation stations is characterized by its call sign, the position, frequency and range of its transmitter (Table I).
Table I — Ground station characteristics

<table>
<thead>
<tr>
<th>Receivers</th>
<th>Transmitters</th>
<th>Frequency</th>
<th>Position</th>
<th>Call sign</th>
<th>Range</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>RC</td>
<td>L</td>
<td>MF</td>
<td>airfield landing</td>
<td>2 letters</td>
<td>30 - 50 km</td>
<td>~20 NM</td>
</tr>
<tr>
<td>RC</td>
<td>NDB</td>
<td>MF</td>
<td>along airways</td>
<td>3 letters</td>
<td>350 km</td>
<td>~100 NM</td>
</tr>
<tr>
<td>RC</td>
<td>BS</td>
<td>MF</td>
<td>broadcast without reference to air navigation</td>
<td>very large</td>
<td></td>
<td>Gt</td>
</tr>
<tr>
<td>VOR</td>
<td>T/VOR</td>
<td>VHF</td>
<td>airfield</td>
<td>2 letters</td>
<td>visual 50 km</td>
<td>~20 NM</td>
</tr>
<tr>
<td>VOR</td>
<td>VOR/NAV</td>
<td>VHF</td>
<td>along airways</td>
<td>3 letters</td>
<td>visual 100 km</td>
<td>~50 NM</td>
</tr>
<tr>
<td>DME</td>
<td>VOR/DME</td>
<td>VHF</td>
<td>along airways</td>
<td>3 letters</td>
<td>visual 100 km</td>
<td>~50 NM</td>
</tr>
</tbody>
</table>

The aircraft distance to a DME station is directly provided to within 1 NM. The time-distance to a station of another type can be empirically evaluated from the variation in time of the received QDM (Fig. 4).

Fig. 4 — Estimation of time-distance with station across.

In (2), when \( a' = a \), the distance between aircraft and station is equal to the distance travelled between (1) and (2).

The navigation calculations require that the pilot has at his disposal, as well as radio-navigation receivers, a magnetic compass and a clock.
Basic calculations

The graphic presentation of the sum $\vec{S}_a = \vec{S}_a + \vec{W}$ constitutes the "wind triangle" (Fig. 5). The, $(\vec{S}_a, \vec{S}_g)$ angle is called drift angle, $\delta$. The $(\vec{S}_g, \vec{W})$ angle is denoted $\alpha$. As $\vec{S}_g$ is carried by the magnetic path $PM$, one can write:

$$\delta \equiv PM - HA \quad \text{(modulo 360°)}$$

Fig. 5 — Principle of the wind triangle.

Vector $\vec{W}$ has then following components:

$$\vec{W} = \begin{cases} W_a : \text{effective wind, on } \vec{S}_a \\ W_c : \text{cross wind, perpendicular to } \vec{S}_a \end{cases}$$

The pilot is led to perform his basic calculations by using the following intermediate values:

- the base factor $F_b = 60/S_a$ representing the time, expressed in minutes, to travel at speed $S_a$, the distance of 1 NM, $S_a$ being expressed in knots.

- the maximum drift angle $\Delta = F_b W$, in case of cross wind, a dimensionless value corresponding to an angle expressed in degrees.

- the head-on wind effect $I = \Delta \cos \alpha$. These values make it possible to calculate the no-wind-time and the time-for-flight.

We call no-wind-time the time necessary to travel at the speed $S_a$ (kt), a distance $D$ (NM). This time, expressed in minutes, is calculated as:

$$TF = F_b \times D$$

The time-for-flight, expressed in minutes, for travelling the same distance $D$ in the presence of wind, i.e. at the speed $S_g$, is:

$$TFA \ (\text{min}) = TF \ (\text{min}) \pm TF \ IAC \ (\text{sec}),$$

where $IAC$, corrected wind effect, appears as the correction, expressed in seconds, to apply for every minute of flight, and is calculated as a function of $I$ by an empirical law (Table II).
Table II — Wind effect correction.

| || Down wind | Head-on wind |
|---|---|---|
| 0 to 5 | || | |
| 6 to 9 | || -1 | || + 1 |
| 10 | 9 | 12 |
| 11 | 9 | 13 |
| 12 | 10 | 15 |
| 13 to 24 | || - (|| - 10)/2 | || + (|| - 10) |

Navigation along a ground station axis

By definition, it is enough, in no wind condition, to take and maintain a heading equal to the QDM of the station (Fig. 6). In the presence of wind, this strategy leads to a hunt and hare pursuit trajectory, ending straight against the wind (Fig. 7).

![Fig. 6 — Zero wind path. $H_A = QDM$](image)

![Fig. 7 — Hunt and hare pursuit trajectory $H_A = QDM$.](image)
The cross wind effect can be corrected by choosing a heading different from the QDM, allowing one to maintain a constant magnetic path towards the station. Considering realistic wind speeds, the correction will always be lower than 20°. As he ignores the value to be taken, the pilot has to approach it empirically. For this, he uses a dichotomic method, with 1-minute steps (Fig. 8). There exists a variant to this method when the station is behind.

It is often desired that the station be approached by the aircraft along a given magnetic path. One must then join back on this path before applying the drift correction: for this, the heading is altered by a value function of the time-distance to the station.

**IFR procedures**

The axis allowing navigation makes it possible to travel along airways from beacon to beacon. Procedures inside TMAs, during departures and arrivals, may require the performance of conventional patterns: these constitute the "IFR procedures". Only two of them have been retained here:

- the 40° procedure (Fig. 9)
- the holding pattern, overall time staking. (Fig. 10)
The patterns are travelled from a basic reference point, usually marked by an NDB or VOR/L beacon. Wind corrections are made distinguishing cross wind and head-on effective wind, correcting each according to its direction. This way, wind corrections are performed using the same calculation rules as in normal path navigation.

2.3 — Control in vertical plane

Two types of controls are considered:
- change of flight level (or altitude) at a given vertical speed,
- following a trajectory with given slope \( \gamma \).

It is necessary, in both cases, to determine the slope \( \gamma_a \) to be assigned, as a function of the desired value, of the air speed \( S_a \) and of the wind.

Flight level changes are performed at a vertical speed \( \dot{z} \), of 1000 ft/min in general and of 10000 ft/min for an emergency dive.

The assigned value adopted is then: \( \gamma_a = \frac{\dot{z}}{S_a} \).

In case of an imposed slope \( \gamma \), usually followed by an ILS beam of - 3° slope, the assigned value takes the head-on wind effect \( W_a \) into account:

\[ \gamma_a = \gamma (1 - \frac{W_a}{S_a}) \]

3 — Procedures

3.1 — Normal procedures

The flight is made of the three following phases:
- departure,
- cruise,
- arrival.

**Departure**

At time \( t = 0 \), the aircraft is assumed immobile, at the runway's first end, brakes "on". The aircraft heading is then equal to the true heading of the runway, and its altitude zero in QFE. The aerodynamic speed is opposed to that of the wind. The simulation starts at brakes release.
The departure sub-phases (Fig. 11) are the following:

- from $t = 0$ to $S_a = V_1$: ground running at constant acceleration, during about 10 s, with $V_1 = 90$ kt.
- from $S_a = V_1$ to $S_a = V_R$ ($V_R = 118$ kt): no change, except procedure after engine failure;
- from $S_a = V_R$ to $S_a = V_{LOF}$ ($V_{LOF} = 120$ kt): rotation; aircraft drag is assumed to increase; acceleration decreases; $\gamma_a$ increases;
- first segment: the aircraft clears the ground, undercarriage is lifted (duration 8 s); the segment is ended if:
  - the undercarriage is up (or failure with undercarriage locked down),
  - $Z_{QFE} \geq 35$ ft,
  - $V \geq V_2$ ($= 150$ kt);

According to the wind magnitude and direction, the order of occurrence of these three events is different;
- second segment: climb at constant speed ($\neq V_2$), ending at $Z_{QFE} \geq 400$ ft;
- third segment: acceleration at zero slope; ends at $S_A \geq V_{FD}$ ($= 180$ kt);
- fourth segment: climb at constant speed up to $Z_{QFE} = 1500$ ft;
- fifth segment: the aircraft joins the navigation point of origin, at $z = 0$.

The departure procedure is then finished. From now on, altitude is measured in flight levels.
Cruise Navigation on airways axis at constant speed and altitude or changes of flight level ($\gamma_a = \frac{g}{S_a}$). The conditions of execution are prescribed by the navigation charts; speed and flight level are prescribed.

Arrival

The descent starting point is reached at a given altitude and speed.

- start of descent: assignment of $\gamma_a = \gamma (1 - W_a/S_a)$ and following the localizer by navigation on its axis;
- at $ZQFE \leq 50$ ft: aircraft heading is fixed and flare initiated: $\gamma_a = \gamma_a/2$;
- at $ZQFE = 0$, touching down of wheels: $\gamma_a = \text{run way slope}$; running at constant deceleration down to $S_0 = 0$;
- end of simulation.

3.2 — Procedures after failure

The procedures applied after simulated failure will be examined:

**Single engine failure**

- at take-off before $V_1$: stopping the aircraft on the runway at constant deceleration;
- at take-off after $V_1$: procedure of continued take-off, then arrival procedure on the same airfield;
- in cruise flight or on arrival: continuation of the flight towards emergency airfield.

**Double failure (twin-engine aircraft)**

The aircraft can be stopped on the runway before $V_t$ at departure, or on arrival after touchdown (minor or significant consequence).

The aircraft is lost in all other cases ("catastrophic" consequence).

**Undercarriage failure (locked down)**

At departure: execution of take-off procedure, followed by arrival on the same field (with $z_i < 15000$ ft).

Cruise: not applicable

At arrival: normal execution (with $S_a < 150$ kt).

**Pressurization failure**

At departure: return to base after take-off procedure completed (with $z_i < 15000$ ft).

Cruise: if $z_i > 15000$ ft, emergency dive ($\gamma_a = \frac{g}{S_a}$), with $\frac{z_i}{S_a} = 10000$ ft/min; then, if $z_i < 15000$ ft, continuation of the flight at $z_i < 15000$ ft.

According to the flight phase when the failure occurs, the consequences are either "minor" or "significant" or "critical".

4 — Results

4.1 — Simulations

As an illustration, the programme was used to simulate Nantes-Bordeaux flights and back, performed with a fleet of three aircraft. The running of the flights is governed by the trajectory charts, Figures 12 and 13.
Each aircraft is subjected to the four above failures, and is supposed to perform four return journeys. The occurrence of failures is submitted to the probabilities chosen for each four above; the consequences of these failures on the running of the flights are listed according to their class of importance.

The hourly frequency of the failure rates attached to each sub-system is of the form $\lambda_1 = \lambda_0 10^C$, where $\lambda_0$ is the basic rate and C a class of probability. Realistic values of C are 5 to 6. Here, in order to allow the occurrence of failures with a reduced number of simulations, reduced values of C were retained; their results are presented on Table III.

Fig. 12 — Bordeaux arrival procedure.
Fig. 13 – Bordeaux let down procedure.
Table III — Simulation results.

<table>
<thead>
<tr>
<th>Aircraft</th>
<th>Flight n°</th>
<th>Path</th>
<th>n° of return flight</th>
<th>Events</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>N.B.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>B.N.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>N.B.</td>
<td>2</td>
<td>PR ; DU ;</td>
<td>Critical</td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>B.N.</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>N.B.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>6</td>
<td>B.N.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>7</td>
<td>N.B.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>8</td>
<td>B.N.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>N.B.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>B.N.</td>
<td>1</td>
<td>MD ; B</td>
<td>Significant</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>B.N.</td>
<td>1</td>
<td>PR ; DU</td>
<td>Critical</td>
</tr>
<tr>
<td>B</td>
<td>4</td>
<td>N.B.</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>B.N.</td>
<td>2</td>
<td>MD : → D</td>
<td>Significant</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>B.N.</td>
<td>2</td>
<td>PR ; DU</td>
<td>Critical</td>
</tr>
<tr>
<td>B</td>
<td>7</td>
<td>N.B.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>8</td>
<td>B.N.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>9</td>
<td>N.B.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>10</td>
<td>B.N.</td>
<td>4</td>
<td>PR</td>
<td>Minor</td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>B.N.</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>B.N.</td>
<td>1</td>
<td>PR ; DU</td>
<td>Critical</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>N.B.</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>B.N.</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>5</td>
<td>N.B.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>6</td>
<td>B.N.</td>
<td>3</td>
<td>MD : → B ; PR ; DU</td>
<td>Critical</td>
</tr>
<tr>
<td>C</td>
<td>7</td>
<td>B.N.</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>N.B.</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9</td>
<td>B.N.</td>
<td>4</td>
<td>PR ; DU</td>
<td>Critical</td>
</tr>
</tbody>
</table>

MG : failure left engine
MD : failure right engine
-> B : return to Bordeaux
-> N : return to Nantes

TBS : undercarriage locked down
PR : pressurisation failure
DU : emergency dive.
4.2 - Interpretation

If the actual failure rate of a sub-system is \( \lambda_i \), the probability of occurrence of the failure during a flight of duration \( \tau \) is \( p_i = \tau \lambda_i \). The probability of not having a failure during \( n \) flights is then \( f_i = (1 - p_i)^n \).

By simulating \( m \) flights with a fictitious failure rate \( \lambda'_i = \alpha \lambda_i \), the probability for not having failure \( i \) is \( f'_i = (1 - p'_i)^m \), with \( p'_i = \tau \lambda'_i \).

The equality \( f_i = f'_i \) is obtained if:

\[
\frac{n \log (1 - p_i)}{m \log (1 - p'_i)} \approx 1
\]

i.e. if \( p_i \) and \( p'_i \) are small and if \( \alpha = m/n \).

In case of double failure (engine failures), if the failure rates, a priori different, are \( \lambda_i \) and \( \lambda_j \) for the two similar sub-systems, the probability of having no failure for \( n \) flights of duration \( \tau \) will be:

\[
f = (1 - p_i)^n (1 - p_j)^n, \quad \text{with} \quad \begin{cases} p_i = \tau \lambda_i \\ p_j = \tau \lambda_j \end{cases}
\]

and the probability for not having a double failure \( s = (1 - p_i \cdot p_j)^n \).

Simulating \( m \) flights of duration \( \tau \) with fictitious failure rates \( \lambda'_i = \alpha \lambda_i \) and \( \lambda'_j = \beta \lambda_j \), the probability for not having any simple failure will be:

\[
f' = (1 - p'_i)^m (1 - p'_j)^m
\]

and the probability for not having a double failure:

\[
s' = (1 - p'_i \cdot p'_j)^m
\]

We could then choose \( \alpha = \beta = n/m \) to obtain \( f = f' \), but then \( 1 - s \approx m/n (1 - s') \), which makes more probable the occurrence of a double failure.

If we chose \( \alpha = 2 n/m \) and \( \beta = 1/2 \), then \( f = f' \) and \( s = s' \), but a difference of order of magnitude (of the order of \( n/m \)) is introduced in a parasitic manner between \( p'_i \) and \( p'_j \).

So we chose to take \( \alpha = \beta = 2 n/m \) as long as more of the \( i \) (or \( j \)) failures occurred, then to modify the failure rate on the system remaining in operation by taking, at the failure occurrence, \( \beta \) (or \( \alpha \)) = 1/2.

5 - Conclusion

The simulation programme presented here makes it possible to simulate a flight as a whole. By applying to each of the considered sub-systems an artificially augmented hourly failure rate, it is possible to generate undesirable flight phases with a reduced number of simulations. These phases comprise those occurring with simultaneous failures on several sub-systems whose probability of occurrence is a priori difficult to evaluate, and risks to be underestimated.

Using calculation expedients makes it possible to obtain statistics on the probability of occurrence of these phases, in spite of the augmentation of failure rates.

By introducing into the model all the sub-systems accounted for in the flight crew manual, we may offer a contribution to the generation of after failure scenarios to be studied by simulation during the course of aircraft certification.
EXPERIMENTAL AND THEORETICAL ANALYSIS OF HUMAN MONITORING AND DECISION MAKING BEHAVIOR IN FAILURE DETECTION TASKS

by

R.C. van de Graaff and P.H. Wewerinke *

ABSTRACT

This paper contains the results of an experimental program designed to validate a model of the human observer and decision maker formulated in terms of linear estimation and classical sequential decision theory. The experiment included a variety of monitoring tasks in which the occurrence of ramp failures had to be detected, which were superimposed upon zero-mean stochastic Gaussian processes. The independent variables were signal bandwidth, correlation among displays, number of displays, failure rate, failure type, and prior knowledge about failure type. The dependent variables were response times, display deviations at the moments of response, and false alarm rates. In addition, the experiment included measurements of heart rate, skin resistance, and eye point of regard.

A good overall agreement was found between the experimental results and the model predictions. Furthermore, the additional experimental analysis provided a useful insight in human monitoring and failure detection behavior.

INTRODUCTION

In the evolution of many man-machine systems (e.g. aerospace vehicles) during the last decades the increasing complexity of the human operator's task becomes more and more apparent. Hereby a gradual shift can be noted of the human operator's role from controller to supervisor. The last two decades considerable research effort has been devoted to the study of human control behavior. This has resulted in the development of a number of mathematical tools, of which the optimal control model has been shown to provide an adequate framework for describing the human processing of information of a dynamic system (Refs 1-11). With respect to the human operator's functioning as supervisor of an automatic system, however, until now only a few attempts have been made to investigate and model human failure detection and decision making behavior (Refs 5-10).

This paper summarizes the results of a theoretical and experimental study of the human observer and decision maker in multivariable failure detection tasks. A more detailed presentation of the model and experimental results is given in references 11 and 12, respectively. The decision tasks which were investigated in this study consisted of the detection of failures superimposed upon zero-mean stochastic Gaussian processes. Various task variables were included in the selected (16) configurations which were analyzed in terms of the following model of the human observer and decision maker. The model predictions are compared with the results of an experimental program discussed in the subsequent chapter.

* National Aerospace Laboratory NLR, The Netherlands
MODEL OF THE HUMAN OBSERVER AND DECISION MAKER

It is assumed that the human perceives information of a linear dynamic system which is described by a Gauss-Markov random sequence. Based on the learned dynamics of this system and the perceived information (i.e. noisy observations), the human makes the best estimate of the system state. This is described in standard linear estimation theoretical terms (Kalman filter, Refs 13 and 14) and is part of the well documented optimal control model (Refs 1 and 2).

Now, in the normal mode of operation, the discrepancy between perceived and expected information (the so-called innovation sequence \( n_k \)) is a zero mean Gaussian random sequence (Ref. 11) with covariance \( N_k \). It is assumed that abnormal system operation, as caused by errors in display instruments, malfunctioning of the system, excessive system disturbance levels (e.g. large windshears in aircraft operation) can be represented by a deterministic process, as such unknown to the human observer but detected on the basis of a non-zero mean innovation sequence whose statistic is sufficient to make decisions (test hypotheses) when the system is completely observable.

In terms of classical sequential decision theory (Ref. 15) a so-called generalized likelihood ratio test can be formulated. The test amounts to the comparison of the probability of a non-zero mean with the probability of a zero mean innovation sequence assuming that the human operator makes a short-term estimate of the mean of the innovation sequence on the basis of the sample mean of \( m \) past observations \( \bar{n}_k \).

It can be derived (Ref. 11) that the effect of each observation at stage \( k \) on the (log of the) likelihood ratio is given by

\[
\Delta L_k = \frac{1}{2} \bar{n}_k^2 N^{-1} \bar{n}_k
\]

under the assumption that the sample mean \( \bar{n}_k \) is constant during \( m \) observations. The accumulating effect of each observation on the total (log of the) likelihood ratio is given by the recursive expression

\[
L_k = L_{k-1} + \Delta L_k
\]

assuming that the innovation sequence is a white noise sequence (independent samples) which is exact in the normal mode of operation.

When the likelihood ratio \( L_k \) (representing the total evidence of abnormal system operation) is equal to, or larger than, a decision threshold \( T \), the decision is made that an abnormal condition has occurred. This decision threshold can conveniently be related to the accepted (or assumed) risk according to (Ref. 15) \( T = \frac{(1-P_M)}{P_F} \), with \( P_M \) the miss probability (i.e. of no response to an abnormal condition) and \( P_F \) the false alarm probability.

Referring to references 11 and 15, an expression can be derived for the average number of samples used to make the decision that the system is operating abnormally, which is, for a given sample rate, uniquely (linearly) related to the average detection time.

This implies a relationship between the average number of observations and the decision error probabilities \( (P_F \text{ and } P_M) \), for a given innovation covariance \( N \) and the non-zero mean failure state sequence which is, however, a
given task variable. Thus, the only human decision model parameters are the short-term average sample size \( m \) and the innovation covariance \( N \) which depends exclusively on the human observation noise covariance. If an optimal allocation of attention among the display variables is assumed the only remaining human decision model parameters are the short-term average sample size \( m \) and the overall level of attention \( P_0 \). The model outputs are the average failure detection time corresponding to given (or assumed) error probabilities and various attentional characteristics.

Experimental results of single-variable, (step and ramp) failure detection tasks, which are reported in reference 8, were used for a preliminary validation of the human decision model and to "calibrate" the model with respect to the short-term average sample size \( m \). The resulting value of \( m \) of four seconds yields an excellent fit to the experimental failure detection times. This can be expressed in the linear correlation coefficient between model and experimental results of 0.99 and in the ratio of the model detection times and corresponding experimental values \( t_m/t_e \). This ratio was on the average 0.98, with a standard deviation of 0.09. The value of four seconds, which lies in very well with the short-term memory span typically ranging up to five seconds for visual stimuli, will be assumed and kept constant in the following validation experiment.

The general model structure (the only assumptions are that the dynamic system is linear and that abnormal conditions can be represented by a deterministic process) accounts for the effect of a variety of task variables. These task variables were included in the selected task configurations investigated in the validation program discussed in the next section.

EXPERIMENTAL PROGRAM

In order to test validity of the model, a variety of tasks were specified: number, bandwidth and mutual correlation of the displays, and various failure characteristics.

Experimental set-up

Up to four-display tasks were considered consisting of two separate (independent) identical processes. Each process could be observed via two displays; a relatively high bandwidth variable \( y_i \) (second order process with break-frequency of 1.2 rad/s) was additionally filtered (first order filter with a time constant of four seconds). The output of this filter which is correlated with \( y_i \) \( (r = 0.5) \) was displayed as \( y_i^\prime \). This process was duplicated resulting in a four-display process \( (y_i \text{ to } y_i^\prime)^2 \), of which the displays were two by two correlated. Two failure types were investigated whether appearing on one display (display failure, \( F_i^1 \)) or on two correlated displays (system failure, \( F_i^2 \)). Also the effect of prior knowledge about failure type was investigated. The resulting 8 configurations representing these variables are summarized in table 1. The configurations were investigated for two failure rates (a slope of 0.1 standard deviation of the display position per second \( \sigma_i^1/s \) and a slope of 0.2 \( \sigma_i^2/s \) ) yielding 16 experimental conditions.

Thus, these conditions contained as independent variables and factors: signal bandwidth, number of displays, additional (correlated) display infor-
mation, failure rate, failure type, and prior knowledge about failure type. The dependent variables were: detection times, display deviations at the moment of detection, and false alarm rates. Furthermore, several measures of heart rate and skin resistance were included as indicators of the subject's psychophysiological state, as well as eye point of regard measures. Three general aviation pilots participated in the experiment. In each session the 16 configurations were presented in a random order. Twelve replications per experimental condition per subject were obtained. In contrast to the extensive training period no performance was fed back to the subjects during the formal sessions. The subjects indicated the detection of an abnormal state by pushing a button. After this response, they had to push the button a second time at the moment they were able to diagnose the failure. Diagnosis was also expressed verbally to exclude guessing.

Experimental results

Detection times

The experimental results of two subjects achieving an overall false alarm probability $P_a$ of 0.05 could be compared directly with the model predictions (the results of the third subject will be discussed separately). These model predictions of the (ensemble) mean failure detection times for all 16 failure detection tasks were obtained on the basis of the two constant model parameters: the overall level of attention $P = 0.01$ and the short-term average sample size $m = 4$ seconds. Based on previous results a value of the false alarm probability of 0.05 was assumed. Additional procedural details are discussed in reference 11. The results are summarized in figure 1. The linear correlation coefficient between the model predictions and the experimental mean failure detection times is 0.86 which reflects a very good overall predictive capability of the human decision model. The ratio of the model and experimental failure detection time $t_m/t_e$ is another measure for the agreement between the model and experimental results. On the average (over all 16 tasks) this ratio is 0.98. The standard deviation is 0.12 which is comparable with the reliability of the estimated ensemble means $(\sigma/\sqrt{N})$ of the data.

Figure 2 shows the results per configuration. For one configuration (Conf. 4) the model predictions clearly disagree with the experimental results. A very plausible explanation of this discrepancy (which was confirmed during the debriefing of the subjects) was that the subjects did not realize (use) the system failure dynamics but assumed that a system failure would appear simultaneously on both (correlated) displays. The model results based on this assumption are also shown in figure 1 and indicated with model refinement. In that case the linear correlation coefficient is 0.95; the mean ratio $t_m/t_e$ is 1.01 with a standard deviation of 0.09.

Yet, an overall validation test of the model (discussed in reference 12) indicated that, at least for one configuration, the model predictions disagree significantly ($p < 0.05$) with the experimental results.

The experimental results with respect to the specific task variables and the corresponding model predictions are summarized in table 2. The configura-
tions involved in the pair-wise comparison between the configurations which differ with respect to the specific task variable (only) are indicated. Comparing the measured and model results shows that the effect of display bandwidth, of additional (correlated) displays and of the failure rate is excellently predicted by the model. The predicted interference between uncorrelated displays (because of the human attention sharing involved) is larger than obtained experimentally. The model predictions are based on a constant level of attention. However, the heart rate measures (discussed in the following) indicate a small, but statistically significant, increase in attention with an increase in the number of displays which can easily explain the small difference in interference.

The effect of failure type is discussed before (with respect to Conf. 4). It can be seen that the model refinement and experimental results agree closely.

The experimental results of the third subject reflect a distinctly different decision strategy. He made no false alarms and his failure detection times were, on the average, 40% higher. Yet his results correlated well with the model predictions (r = 0.80).

Display deviation analysis

The analysis was performed to examine whether the differences in the detection performance over configurations could originate from relatively constant response strategies.

Referring to table 1b the following deviation measures were considered to be utilized in possible constant decision strategies: (1) the deviations from zero-level of each display ($y_1^*$): $|y_1^*|$ for $F_1$ and $F_2$ failures, $|y_2^*|$ for $F_2$ failures; (2) the average of correlated deviations: $|y_1^* + y_2^*|/2$; (3) the discrepancies between correlated deviations: $|y_1^* - y_2^*|$; (4) the discrepancy between the magnitudes of correlated deviations: $||y_1^*| - |y_2^*||$.

Figure 3 shows the results of an analysis of variance (including a multiple classification analysis (Ref. 16)) which was performed on the values of the deviation measures (1) to (4), obtained for the different configurations and failure types. This figure indicates that the deviation measures (1), (2) and (3) are varying over (groups of) configurations and therefore no candidates for relatively constant decision criteria (note the large differences (significant at level 0.05) between the values of these measures for the display failures in correlated display configurations and the corresponding values for system type failures). For deviation measure (4), however, no significance was found ($p \leq .71$) between all correlated display configurations, clearly indicating that strategy (4) can be considered as an overall detection strategy. This criterion also explains the relatively large detection times in the case of system failures, corresponding with small discrepancies. A criterion in terms of absolute values, (1) on (2), would have led to considerably better response performance.

Configurations 1, 2 and 6 form a distinct group, as in these configurations no correlated display information is present. The detection strategy with respect to these configurations can only be based on deviations from zero-level. This is represented by all four deviation measures, which are identical for these configurations (apart from a factor 2 involved in (2)).
So, in conclusion the results of the deviation analysis strongly suggest that
detection strategy is based on discrepancies between magnitudes of correlated
deviations.

Eye scanning

Eye point of regard measurements have been obtained in order to compare
the various attentional characteristics predicted by the model with the cor­
responding measures derived from the raw scanning data. In the following we
consider subsequently the experimental results in terms of the dwell fractions
on the displays, the link values between the displays, and the scan rates.

The dwell fraction on a given display is defined as the relative amount
of time spent on that display. The results obtained are presented in table 3.
The model predicts a relatively constant fraction of attention to the high
bandwidth display(s) between 0.5 and 0.6 for all correlated display tasks
(Conf's 3, 4, 5, 7, 8). As table 3 shows, this prediction could be confirmed
for subjects A, B. Subject C adopted a significantly different allocation
strategy with a relatively constant fraction of attention of 0.7 for high
bandwidth displays.

The link values between displays are defined as the relative number of
looks from a given display towards another display. The results obtained for
configurations 7 and 8 are presented in table 4. The subjects adopted similar
scanning strategies. Furthermore, a scanning preference for the correlated
displays \((y_1, y_2, y_3, y_4)\) is apparent as well as for the high bandwidth
displays \((y_1^* and y_2^*)\). These experimental results are in agreement with the
corresponding model predictions (Ref. 11).

The scan rate is defined as the average number of transitions between
displays per second. Table 5 presents the scan rate values per subject for
the configurations 3-8. The only significant effect is an increase in scan
rate when the number of displays is increased from 2 to 4.

Heart rate and skin resistance

The foregoing analysis was limited to the detection stage of the experi­
ment. A relevant discussion of the results of the physiological measurements
of heart rate and skin resistance, however, has to be extended to the period
after detection, because of the time lags involved in these quantities.
As measures for heart rate and skin resistance were considered: mean, stand­
ard deviation, root mean square, and root mean square of successive differ­
ences. Furthermore, of the skin resistance the low frequency component
(resistance level) and the relatively high frequency component (resistance
response) were obtained separately.
Table 6 shows the results of an a posteriori analysis (Scheffe's S method,
\(p \leq 0.05\), performed on the afore-mentioned measures. The table indicates that
the heart rate measures were significantly sensitive to the number of dis­
plays involved, whereas only significant differences in the skin resistance
measures were obtained systematically for configurations 5 and 8, clearly due
to the diagnosis involved in these configurations, which can be related to
task difficulty. (For configurations 5 and 8 the number of diagnosis failures
and the amount of time needed for diagnosis were significantly larger than for the other configurations.)

Table 6 further indicates that heart rate level is the most sensitive measure for discriminating among the number of displays to be monitored, as this measure is already able to discriminate during the detection stage. With respect to the skin resistance measures, table 6 indicates that resistance response quantities are more sensitive to task difficulties than resistance level quantities. Furthermore, for the skin resistance response measures it is indicated that the mean measure is the most sensitive.

CONCLUDING REMARKS

A good overall agreement was found between the experimental detection times for all 16 task configurations and the corresponding model predictions.

An analysis of display deviations suggested that the human detection strategy is based upon discrepancies between deviation magnitudes of correlated displays.

The analysis of eye-movement data indicated a relatively constant dwell fraction over the displays for all configurations. Furthermore, a clear scanning preference for correlated displays and high bandwidth displays was found. These results are in agreement with the corresponding model predictions.

The main conclusions from the physiological data were:
- heart rate level is significantly sensitive to the number of displays
- significant differences in skin resistance level and skin resistance response correspond to differences in task difficulties.

REFERENCES


11 Wewerinke, P.H.: A model of the human observer and decision maker. NLR TR 81... U, April 1981 (Forthcoming).


### TABLE 1a
Task configurations

<table>
<thead>
<tr>
<th>CONF.</th>
<th>DISPLAY</th>
<th>FAILURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$y_1$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>2</td>
<td>$y_2$</td>
<td>$F_2$</td>
</tr>
<tr>
<td>3</td>
<td>$y_1$, $y_2$</td>
<td>$F_1$</td>
</tr>
<tr>
<td>4</td>
<td>$y_1$, $y_2$</td>
<td>$F_s$</td>
</tr>
<tr>
<td>5</td>
<td>$y_1$, $y_2$</td>
<td>$F_1$ or $F_2$ or $F_s$</td>
</tr>
<tr>
<td>6</td>
<td>$y_1$, $y_3$</td>
<td>$F_1$ or $F_3$</td>
</tr>
<tr>
<td>7</td>
<td>$y_1$, $y_3$, $y_4$</td>
<td>$F_1$ or $F_{s_2}$</td>
</tr>
<tr>
<td>8</td>
<td>$y_1$, $y_3$, $y_4$</td>
<td>$F_{s_2}$</td>
</tr>
</tbody>
</table>

### TABLE 1b
Display configuration

![Diagram of display configuration](image-url)
TABLE 2
Specific effect of task variables on model and experimental failure detection times

<table>
<thead>
<tr>
<th>EFFECT OF CONFIGURATIONS</th>
<th>RATIO $t_i/t_j$</th>
<th>measured</th>
<th>model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>m</td>
<td>s</td>
</tr>
<tr>
<td>bandwidth (low/high)</td>
<td>1, 2, 5, 8</td>
<td>1.01</td>
<td>0.12</td>
</tr>
<tr>
<td>additional display info.</td>
<td>1, 3, 6, 7</td>
<td>0.76</td>
<td>0.06</td>
</tr>
<tr>
<td>interference</td>
<td>1, 3, 5, 6, 7, 8</td>
<td>1.11</td>
<td>0.05</td>
</tr>
<tr>
<td>failure type</td>
<td>3, 4, 5, 8</td>
<td>1.43</td>
<td>0.33</td>
</tr>
<tr>
<td>(system/display)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>failure rate</td>
<td>all</td>
<td>0.67</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3
Dwell fractions on high bandwidth display(s)

<table>
<thead>
<tr>
<th>CONF.</th>
<th>SUBJECT</th>
<th>DWELL FRACTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>3, 4, 5</td>
<td>A</td>
<td>.52(.07)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.57(.07)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.71(.08)</td>
</tr>
<tr>
<td>7, 8</td>
<td>A</td>
<td>.52(.06)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>.57(.04)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>.71(.07)</td>
</tr>
</tbody>
</table>

(...)= standard deviation
TABLE 4
Link values between displays
(Confs 7, 8; 3 subjects)

\[ \begin{array}{cccc}
  & y_1 & \rightarrow & y_2 \\
 15 & \rightarrow & .14 & \rightarrow \\
 12 & \rightarrow & .12 \\
 17 & \rightarrow & .16 \\
 & y_3 & \rightarrow & y_4 \\
 & .04 \\
 & \end{array} \]

(.,)= standard deviation

TABLE 5
Scan rates for two- and four-display configurations

<table>
<thead>
<tr>
<th>CONF.</th>
<th>SUBJET</th>
<th>SCAN RATE (1/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,4,5,6</td>
<td>A</td>
<td>1.69(.22)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1.87(.17)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.98(.21)</td>
</tr>
<tr>
<td>7,8</td>
<td>A</td>
<td>2.78(.20)</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>2.97(.20)</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1.99(.22)</td>
</tr>
</tbody>
</table>

(.,)= standard deviation
TABLE 6
Significant differences between configurations for various scores of heart rate, skin resistance response and skin resistance level (Scheffé's S method, p ≤ .05)

<table>
<thead>
<tr>
<th>PHYSIOLOGICAL PARAMETER</th>
<th>SCORE</th>
<th>PERIOD 1</th>
<th>PERIOD 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>M</td>
<td>(1,2)<em>(3,4,5,6)</em>(7,8) n.s.</td>
<td>(1,2)<em>(3,4,5,6)</em>(7,8) n.s.</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>n.s.</td>
<td>(1,2)<em>(3,4,5,6)</em>(7,8) n.s.</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>n.s.</td>
<td>(1,2)<em>(3,4,5,6)</em>(7,8) n.s.</td>
</tr>
<tr>
<td>Skin Resistance Response</td>
<td>M</td>
<td>(5,8)*(1,2,3,4,6,7) n.s.</td>
<td>(5,8)<em>(1,2,4,6)</em>(3,7) n.s.</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>n.s.</td>
<td>(5,8)*(1,2,3,4,6,7) n.s.</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>n.s.</td>
<td>(5,8)*(1,2,3,4,6,7) n.s.</td>
</tr>
<tr>
<td>Skin Resistance Level</td>
<td>M</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>RMS</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>n.s.</td>
<td>(5,8)*(1,2,3,4,6,7) n.s.</td>
</tr>
<tr>
<td></td>
<td>RMSSD</td>
<td>n.s.</td>
<td>(5,8)*(1,2,3,4,6,7) n.s.</td>
</tr>
</tbody>
</table>

M : mean
RMS : root mean square
SP : standard deviation
RMSSD: root mean square successive differences
n.s.: no significance

PERIOD 1: 0 < t < t detection
PERIOD 2: t detection < t < t diagnosis
Fig. 1 Overall comparison of model predictions and experimental results
Fig. 2 Model predictions and experimental detection times per configuration
Fig. 3 Various deviation measures at the moment of detection for all monitoring tasks.
A MONITORING AND DECISION MAKING PARADIGM:  
EXPERIMENTS AND HUMAN OPERATOR MODELING  

Willi Stein  
Research Institute for Human Engineering (FAT)  
D-5307 Wachtberg-Werthoven, F.R. Germany

ABSTRACT

A paradigm is developed for human decision making in multiple process monitoring situations. The tasks imply the monitoring of a given number of stationary Gauss-Markov processes by means of tolerance-band displays. The process related events to be detected consist of the time intervals exceeding the tolerance bands, so that a yes/no response process is requested. An extended coverage of the paradigm is provided by a wide range variation of a system of experimental factors: number of processes monitored (1), bandwidth (2), and event probability (3) of the processes, field of view required (4), and couplings between the processes (5). The decision making performance is related to the visual instrument sampling behavior, recorded by an oculometric device. The human behavior is modeled on two parallel ways: A descriptive approach is based on multilinear regression functions. A normative approach uses a combination of a continuous and a discrete processing stage, based on optimal theoretic principles and decision theory. The extensively investigated system of factors and the fidelity of modeling yielded favour applications within process control, vehicle guidance and other man-machine areas.

INTRODUCTION

As the complexity and automation of man-machine systems increase, the human operator must supervise more processes of greater variety. Examples of such multi-task situations include the monitoring of industrial plants, power systems and aircraft piloting tasks. The observation of continuous or approximately continuous processes and the generation of well selected discrete actions are attributes of many monitoring and decision making tasks.

Numerous theoretical and experimental approaches to the various fields of monitoring and decision making have been undertaken (ref. 1, 2, and Proceedings of Annual Conference on Manual Control, since 1971). Nevertheless this field of research is supposed to be at an early stage. The categorization of tasks, models, and results seems to be rather unsatisfactory. Some stimulating systematization is imposed by the human operator models applied. The study reported in this paper is closely related to the approaches of Senders (ref. 3, 1, 2), Levison (ref. 4, 5), and Wewerinke (ref. 6, 7). The research aims to an extensive and well validated representation of human monitoring and decision making behavior. Hence, the study of the operator's sensory limitations and attention allocation combined with the analysis of the visual instrument sampling behavior are indispensable. The experimental situation has been derived and generalized from the pilot's task of deciding whether or not he is within the landing window. This type of monitoring task involves tolerance-band displays and approximates many situations where display vari-
ables have to be related to explicitly indicated reference values. Hence, these tasks are more frequent and in closer relation to many practical problems than continuous manual control.

**EXPERIMENTAL SITUATION**

An experimental situation was developed to facilitate the study of human monitoring and decision making performance in a specific multiple process task. As figure 1 illustrates, the task involves the simultaneous monitoring of a given number \( m \) of dynamic processes \( y_i(t) \) and the binary decision \( u(t) \in \{ D^0, D^1 \} \) whether or not all display variables are within the corresponding tolerance bands. The response state \( D^1 \) indicates the push button depressed. The stationary Gauss-Markov processes \( y_i(t), i=1, \ldots, m, \) have second order Butterworth characteristics with the natural frequencies \( \omega_{0i} \) and a constant damping ratio of \( 1/\sqrt{2} \). The inputs \( w_i(t) \) to the system are zero-mean Gaussian white noise processes of unit variance. Hence, a multivariable system model with \( m \) inputs, \( m \) outputs, and \( 2m \) states is involved. The addition of static cross-couplings is provided for situations requiring correlated display variables. The appropriate state space representation shows that there is no feed-back of human response \( u(t) \):

\[
\begin{align*}
\dot{x}(t) &= Ax(t) + Ew(t), \\
y(t) &= Cx(t),
\end{align*}
\]

(1)

(2)

where \( x(t) \): state vector, \( A \): system matrix, \( w(t) \): input noise vector, \( E \): noise input matrix, \( y(t) \): display vector, \( C \): display matrix.

**Dependent Variables and Measures**

Figure 2 illustrates the time history of a task with \( m \) display processes. The monitoring state of each display variable \( y_i(t) \) is represented by the binary function \( h^i(t) \in \{ H^0, H^1 \} \), where \( H^1 \) indicates the tolerance band exceeded. The superscripts \( o \) and \( i \) relate to normal and abnormal operating conditions of the system monitored, respectively. The monitoring states \( H^0_i, H^1_i \) of the \( m \) displays are combined to the binary function

\[
h_{\text{tot}}(t) = \left\{ \begin{array}{ll} H^0 & \text{if } (H^0_1 \land \ldots \land H^0_m) \\
H^1 & \text{if } (H^1_1 \lor \ldots \lor H^1_m) \end{array} \right.
\]

(3)

Function \( h_{\text{tot}}(t) \) is identical with the required response

\[
u_c(t) = \left\{ \begin{array}{ll} D^0 & \text{if } h_{\text{tot}} = H^0 \\
D^1 & \text{if } h_{\text{tot}} = H^1 \end{array} \right.
\]

(4)

where response state \( D^1 \) indicates the push button to be depressed. The error function \( e_c(t) = u(t) - u_c(t) \) characterizes the difference between the performed response \( u(t) \) and the required response \( u_c(t) \) containing the error.
types false alarm \( \equiv (H^0D^1) \) and missed signal \( \equiv (H^1D^0) \). Furthermore, a probability may be assigned to any state of the binary functions. In particular, the event probability \( P(H^1) \) of the total task is defined as probability of monitoring state \( h_{tot}(t) = H^1 \). An analogous event probability \( P(H^1) \) may be formulated for any display variable \( y_i(t) \). The joint probabilities of the incorrect combinations of states (i.e., \( H^0D^1 \) and \( H^1D^0 \)) result in the false alarm probability \( P_{fa} = P(H^0D^1) \) and the missed signal probability \( P_{ms} = P(H^1D^0) \). The task performance of the human operator is reflected by the decision error \( P_e = P_{fa} + P_{ms} \) whereas the error ratio \( R_e = P_{fa}/P_{ms} \) characterizes some aspects of decision strategy applied. The corresponding measures \( P_{fa}^1, P_{ms}^1, P_e^1, \) and \( R_e^1 \) for any display variable \( y_i(t) \) may be derived analogously.

Discretely valued processes in continuous time are somewhat characteristic of human decision making. The ordinary and joint probabilities defined before may be interpreted as time fractions of the experiments considered. The binary functions \( h_i(t) \) and \( u(t) \) for instance are called indicator variables and may be regarded as alternating renewal processes. Again, the class of renewal processes is part of the point processes. The latter are also used to represent the visual instrument sampling behavior. The process scheme underlying the experimental situation starts from Gaussian white noise, leads to Gaussian-Markov processes and results in alternating renewal processes finally. The statistical theory for mathematical treatment of these processes and derivation of required measures is available (ref. 8, 9, 10, 11).

**System of Experimental Factors**

As Table 1 shows, the five experimental variables investigated constitute a system of factors reflecting characteristics of many monitoring and decision making tasks. The system of factors is intended as a step towards systematization and categorization of tasks. In general, the multiplicity of real situations must be reduced in order to perform systematic and extensive investigations. The factors of the system may be viewed as wide-ranging key-factors concerning the presented information and the field of displays. The number of processes monitored, \( N_a \), may range from a single one in a laboratory task to some hundred in nuclear power plants for instance. However, continuous attention may be directed to a few display variables only. During landing approaches about 60 to 90 per cent of pilot’s attention may be directed to 3 or 4 instruments under given conditions (ref. 12, 13). The number of processes monitored has been varied from 1 to 4 in this study.

The bandwidth of processes monitored plays a part as experimental factor in several monitoring models, especially in the context of the required visual sampling and the sampling theorem (ref. 1, 2). The bandwidth to be considered ranges from small fractions of a rad/s (e.g., control of large ships and process control) to some rad/s during maneuvers of fast response aircrafts. In this study the cut off frequency or natural frequency \( \omega_{01} \) is used as experimental factor being proportional to the bandwidth. In case of a power-spectral density with irregular shape an equivalent bandwidth may be defined (ref. 14, 15). The cut off frequency \( \omega_{01} = 1.5 \) seems to be a sufficient upper limit of the range investigated. The lower limit \( \omega_{01} = 0.25 \) is somewhat arbitrary, but substantially lower frequencies would require prolonged experimental runs (ref. 14).
Another factor to be considered is the probability of events requiring human response. In general, rare events and frequent events have to be distinguished and the frequency of manual interventions decreases in many situations with increasing automation. The event probability \( P(H_i) \) has been defined as a fraction of time in this situation, however, the frequency of manual interventions is nearly proportional to the product of event probability and bandwidth. Interventions may be requested by system failures or by intolerable deviations from a desired state. The latter are more frequent than system failures in general. The event probability \( P(H_i) \) of display variable \( y_i(t) \) depends on the ratio \( \sigma_i/Y_i \), where \( \sigma_i \) and \( Y_i \) denote the standard deviation and the breadth of the tolerance band, respectively. The event probability is varied from 0.125 to 0.5 in this study. The upper range of variation from 0.5 to 1.0 seems to be dispensable, since the decision error \( P_e \) yields a symmetry within both ranges.

Each man-machine task may be characterized by a specific field of view and a thereby connected allocation of attention. The field of view is determined by the task related displays to be fixated foveally in this situation. The big field runs up to a square of 34 times 34 degrees of visual arc, so that only one display may be read at a glance and fractions of attention have to be allocated to the different displays. The small field of view approximately involves 5 degrees of visual arc, so that all displays may be read at a glance. Hence, the analysis of visual instrument sampling has been reduced to the experiments involving the big field of view.

A certain degree of couplings may exist between different display variables in any real situation including a multivariable system. In this study the correlation coefficient \( \rho_{ij} \) implies frequency - independent couplings between the display variables \( y_i(t) \) and \( y_j(t) \). The coefficients 0.0 and 1.0 denote totally dependent and totally redundant display variables, respectively. The demands of a monitoring task are assumed to be lowered by process couplings, since a correlated display variable yields a certain amount of information concerning other display variables to be monitored. The coefficients investigated are 0.0, 0.5, and \( 1/\sqrt{2} \), so that the range of couplings found in real situations has been definitely exceeded.

**EXPERIMENTAL RESULTS**

An extensive investigation is required, since the system of experimental factors is part of the models to be validated. A systematic investigation of all factors, levels, and combinations described before would require some thousand hours of experimental work, so that a complete design is definitely unfeasible. In order to reduce the experimental expenditure, the appropriate techniques of experimental design have to be applied (ref. 16). The experimental program has been partitioned into homogeneous and inhomogeneous conditions, whereby the different display variables of a homogeneous task have identical parameters. Oculometric records are restricted to inhomogeneous conditions and the big field of view. The training requirements, the duration of a run and the number of subjects and replications are reduced to a minimum expenditure determined by pilot studies. However, the experimental design is complicated, since 4 of the 5 factors are highly interacting.
Figure 3 illustrates monitoring and decision making results as a function of bandwidth $\omega_0$. The decision error $P_e = P_{fa} + P_{ms}$ as well as its two components are proportional to the bandwidth of the display variables. The error ratio $R = P_{fa}/P_{ms}$ is approximately constant here and reflects a stable decision strategy within the considered range of frequency. The conditions kept constantly within this part of experiments are the event probability $P(H^1) = 0.5$, the $N_a = 4$ uncorrelated and homogeneous processes monitored, and the small field of view with about 5 degrees of visual arc.

Another aspect of human behavior is characterized by the Pearson variability $V = \sigma/m$, where $\sigma$ and $m$ are standard deviation and mean of the sample considered, respectively. $V_{tot}$ and $V_{intra}$ denote the total and the intraindividual variability. Both measures show increasing decision stability in case of increasing bandwidth. The intraindividual variability of the decision error $P_e$ falls on 5 per cent and indicates a very high stability, whereas the stability of the error components $P_{fa}$ and $P_{ms}$ is less high. The variability of $P_{fa}$ and $P_{ms}$ is supposed to reflect both performance and strategy variations, whereas the variability of $P_e$ reflects performance variations only. Variability aspects of human decision performance have not been investigated extensively yet.

Figure 4 shows the decision error $P_e$ as a function of the event probability $P(H^1)$, whereby the bandwidth $\omega_0$ is taken as an additional parameter. These experiments further involve $N_a = 4$ uncorrelated and homogeneous display variables in combination with the small field of view. The decision error increases monotonously within the lower half of the $P(H^1)$-interval and decreases monotonously within the upper half, so that the peak of $P_e$ at the event probability $P(H^1) = 0.5$ is the centre of symmetry of these curves. The factors $\omega_0$ and $P(H^1)$ have to be assumed interacting since the slope of $P_e$ increases monotonously as a function of the bandwidth. The symmetry may be interpreted on the following way. The two system states $H^0$ and $H^1$ as well as the two states of the response button are mutually exclusive. Hence, the more likely state may be regarded as normal condition and the less likely state may be regarded as abnormal condition to be detected. This somewhat optimal behavior of the human operator seems to be reflected by the symmetry.

The number of processes $N_a$ and the field of view $G_a$ have a strong influence on the monitoring and decision making performance. Both factors are closely connected with the allocation of attention. The experimental results illustrate both factors monotonously increasing the decision error $P_e$. An extension of the field of view yields a reduction of the effective attention due to increasing scan losses. The latter are caused by the time intervals for the eye movement and by associated inhibitory periods (ref. 13). A saccadic eye movement of 34 degrees of visual arc lasts approximately 80 ms, and a value of 0.2 s may be adapted as an estimate of a typical inhibitory period. The scan losses depend on the scan rate (i.e., glances per second) and may amount 30 to 50 per cent of available attention.

Within the situation investigated, the monitoring and decision making behavior is not affected by any dependence of the processes monitored. It has been assumed that the decision error $P_e$ would sink in case of correlated display variables. These somewhat surprising findings have been supported by model results. There may be found only a weak effect when the correlation coefficient nears 1.0, but this degree of dependence exceeds that of real systems by far.
The case of $p_{ij} = 1.0$ implies a reduction of the number of display variables. A different study shows that correlation of display variables is playing a part in tasks, where additive bias failures have to be detected in stationary Gauss-Markov processes (ref. 17). The possible influence of correlated display variables may depend on specific task properties.

Figure 5 illustrates results derived from the experiments having inhomogeneous task conditions, whereby the bandwidth $\omega_i$ and the event probability $P(H_i)$ are varied and the $N=4$ uncorrelated processes within the big field of view are the stabilized parameters. The inhomogeneous case seems to be closer to real systems, since the different display variables $y_i(t)$ of a task don't have identical parameters in general. The inhomogeneous case looks like the homogeneous in respect of the factorial relationships but shows considerably increased decision errors.

The analysis of the visual instrument sampling is applied to experiments involving the big field of view and the inhomogeneous conditions. Aperiodic sampling is assumed, so that successive observations will appear to be independent (ref. 1). Figure 6 shows the dwell fractions $P^a$ of the display variables $y_i(t)$ as a function of the respective bandwidth $\omega_i$ and the event probability $P(H_i)$. $P^a$ seems to be monotonously increased by both experimental factors, whereby the scan losses and the inhibitory intervals are not reflected explicitly.

MODEL APPROACHES

Both approaches of human operator modeling intended must account for the full range of experimental factors investigated. The first model approach is based on multiple regression functions. As the experimental results have shown, most of the relationships between the independent variables $x_i$ (i.e., $N_a$, $ms$, $o_i$, $P(H_i)$, and $G$) and the dependent variables $y_k$ (i.e., $P_{fa}$, $ms$, $P_e$, $P_i$, and $P_i^a$) are monotonous. Within a limited range, the relationships can be approximated by linear functions. Therefore a descriptive modeling approach based on regression functions has been selected. Starting with the simple linear model $y = \beta_0 + \beta_1 x + \epsilon$, any relationship between a singular $x_i$ and a singular $y_k$ of figure 7 can be matched (ref. 16).

Furthermore, a representation of all factors given is required. Since the analysis of variance has revealed a considerable degree of interactions between all experimental factors, the obvious approach by means of multiple linear regression does not fit. The type of interaction occurring is shown in figure 4. As figure 7 illustrates, these bunch-like curves can be represented by a bilinear regression model, consisting of a dual linear regression model and an augmenting interaction term. In general, the representation of $m$ factors is resulting in multilinear regression models. A multilinear equation is a special case of a nonlinear equation and may be viewed as a polynomial in $m$ variables (ref. 11, 16).

In order to estimate the parameters of the multilinear model, an appropriate procedure has to be formulated. Since the linear model $y = XB + \epsilon$ is a general model, it can be used to fit any relationship that is linear in the unknown
parameters $\beta$ (ref. 16). Hence, a multilinear equation may be inserted into the general model by appropriately transforming the independent variables. The parameters $\beta$ are determined by a least squares estimator (ref. 16).

The second model approach aims at a simulation model with dynamic characteristics, consisting of a system model and a human operator model (ref. 18, 19). The optimal monitoring and decision making model (OMDM) herein applied is illustrated by figure 8. Starting with the optimal control model (OCM), the primary version of the OMDM has been developed by Levison and Tanner (ref. 5, 4, 18, 19). The extended version of the optimal monitoring and decision making model has been developed by Wewerinke (ref. 6, 7). As the OMDM in figure 8 shows, the perceptual submodel as well as the estimator/predictor of the optimal control model are retained, whereas the control law of the OCM is replaced by an optimal Bayesian decision rule (ref. 20, 1).

The human operator is regarded by the OMDM as maximizing the expected utility on the basis of the perceived display variables $y_i(t)$. Therefore, the binary human response signal $u(t)$ may be formulated by the Bayesian decision rule

$$u(t) = \begin{cases} D^1 \text{ if } \frac{P(H^1|y_p(t_1))}{P(H^0|y_p(t_1))} \geq K_u \\ D^0 \text{ else} \end{cases}$$

where $t_1 \leq t$ denotes the time interval of the perceived variables $y_p(t_1)$ and $K_u$ denotes the decision threshold defined as

$$K_u = \frac{U_{01} - U_{10}}{u_{11} - u_{01}}$$

Therein $U_i$ is the utility choosing $D^i$, when the monitoring state $H^i$ is given. The decision threshold $K$ is the only free parameter of the decision rule. $K$ may be considered a parameter of decision strategy, since it primarily affects the error ratio $R = P_{fa}/P_{ms}$. Other parameters of the decision rule, e.g., the breadth of the tolerance band and the signal variances, have been fixed by the experimental situation and by the processing stage preceding the decision rule.

The perceptual submodel and the information processing stage of the OMDM are adapted from the optimal control model. Hence, a mathematical deviation and an extensive description of these stages may be found (ref. 1, 15, 18, 19), so that a short explanation of the parameters applied in this study may be sufficient. The perceptual submodel involves the time delay $T$ and the observation noise components $v_i(t)$ consisting of Gaussian white noise processes. Human attention allocation is assumed to be represented by

$$f_0 + \sum_i f_i = 1$$

where $f_0$ denotes scan losses due to eye movements and $f_i$ relates to the fraction of attention devoted to the $i$-th display variable. The various $\sigma^2$ of the $i$-th display variable, the respective observation noise variance $V_{yi}$, and the atten-
tion index $f_i$ constitute the relationship

$$\frac{V_{y_i}}{\sigma_{y_i}^2} = \frac{P_0}{f_i}$$

(8)

where $P_0$ denotes the observation noise ratio corresponding to full attention within a single variable monitoring task. Typically, a value of $-20$ dB is associated with $P_0$ in manual control.

Certainly, the time delay $\tau$ is a free parameter of the perceptual submodel, but it is not varied within the present matchings due to its relative stability supposed. A value of $0.2$ s has been adopted as an estimate of a typical time delay $\tau$. The observation noise and/or the attention index are the only free parameters affecting the decision performance which have been varied in this study.

Figure 9 shows typical experimental results and corresponding model results. The situation considered involves a task with $N_a = 2$ uncorrelated and homogeneous display variables $y_1(t)$ and $y_2(t)$ to be monitored. The display variables have the event probability $P(H^*) = 0.5$ and a second order Butterworth characteristic with the natural frequency $\omega_0$. Hence, the decision error $P_e$ from the small field of view is compared with that from the big field of view. In case of the small field of view, the observation noise ratios $P_1 = P_2 = -20$ dB can be associated with the two display variables so that the corresponding observation noise ratio for full attention is $P_0 = -23$ dB. As the observation noise ratios $P_1 = P_2 = -17$ dB of the other curve show, the big field of view yields additional scan losses of $3$ dB. In all of these cases position and rate perception is assumed not interfering.

CONCLUSION

The field of human decision making in multiple process monitoring situations has been studied by means of a paradigm, both theoretically and experimentally. The two model approaches applied reflect the full range of experimental conditions investigated. The multilinear regression model turns out to be a capable descriptive instrument of the monitoring and decision making tasks. Approximately 70 to 90 per cent of the total variance have been represented by the experimental factors and their interactions. Possible model applications are favoured due to its simplicity and appropriateness for both decision performance and visual sampling behaviour. The optimal monitoring and decision making model applied, too, is in far-reaching accordance with the experimental results.

REFERENCES


FIGURES AND TABLES

**MACHINE SYSTEM**

- **$w_i(t)$**: input noise process
- **$x_i(t)$**: display variable
- **$u(t)$**: response process
- **$A_i$**: tolerance-band display
- **$B$**: binary response button

**FIGURE 1**: Monitoring and decision making situation

**FIGURE 2**: Example time history of a task
TABLE 1
System of experimental factors

<table>
<thead>
<tr>
<th>FACTORS</th>
<th>LEVELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of processes monitored</td>
<td>( N_a ) = ( 1 ), ( 2 ), ( 4 )</td>
</tr>
<tr>
<td>bandwidth (cut-off frequency) of ( i )-th process</td>
<td>( \omega_{bi} ) = ( 0.25 ), ( 0.5 ), ( 1.0 ), ( 1.5 )</td>
</tr>
<tr>
<td>event probability of ( i )-th process</td>
<td>( P(H^i_0) ) = ( 0.25 ), ( 0.25 ) j, ( 0.5 )</td>
</tr>
<tr>
<td>field of view (degrees of visual arc)</td>
<td>( G_a ) = ( 30 ) deg., ( 34 ) deg.</td>
</tr>
<tr>
<td>correlation coefficients of ( i )-th and ( j )-th process</td>
<td>( P_{ij} ) = ( 0.8 ), ( 0.5 ), ( 1/42 )</td>
</tr>
</tbody>
</table>

\( \omega \) = bandwidth
\( \bar{P}_f \) = false alarm probability
\( \bar{P}_m \) = missed signal probability
\( \bar{P} = \bar{P}_f + \bar{P}_m \) = decision error

\( V_{\text{intra-individual}} \) = intra-individual variability
\( V_{\text{total}} \) = total variability

FIGURE 3: Monitoring and decision making performance
FIGURE 4: Monitoring and decision making performance

Independent variables:
- $R_l$: i-th display (display variable $x(i)$)
- $\omega_0$: bandwidth (cut off frequency)
- $P(H_i)$: event probability of $x(i)$

Experimental conditions:
- $N_u = 4$: uncorrelated display variables
- Field of view: 5 degrees

FIGURE 5: Monitoring and decision making performance

Dependent variables:
- $P^m$: mixed signal probability of $x(i)$

Experimental conditions:
- $N_u = 4$: uncorrelated display variables
- Field of view: 34 degrees
**FIGURE 6:** Small probability in visual instrument sampling

**Simple Linear Regression Model:**

\[ y_k(x_i) = \beta_{k0} + \beta_{ki} x_i + e_k \]

**Bilinear Regression Model:**

\[ y_k(x_i, x_j) = \beta_{k0} + \beta_{ki} x_i + \beta_{kj} x_j + \beta_{kij} x_i x_j + e_k \]

**Multilinear Regression Model:**

\[ y_k(x_1, x_2, ..., x_n) = \beta_{k0} + \beta_{k1} x_1 + ... + \beta_{km} x_m + \beta_{klm} \prod_{i=1}^{m} x_i + e_k \]

**FIGURE 7:** Regression models
**FIGURE 8**: Optimal monitoring and decision making model (OMDM)
FIGURE 9: Experimental results and model results
FIELD STUDY OF THE ACTIVITIES OF PROCESS CONTROLLERS

Y. QUEINNEC, G. DE TERSSAC and P. THON

Laboratoire de Psychophysiologie - ERA 700
Université Paul Sabatier - 31062 Toulouse Cédex (France)

INTRODUCTION

The efficiency with which a "Man-Machine" system works depends, schematically, on three conditions (a) the characteristics of the technological equipment (b) the degree of knowledge and training of the human operator (H.O.) and finally (c) the fact that the demands of the job be compatible with the human capacities of the operator.

Although the development of process control has not brought about the disappearance of the H.O., who remains an important element in the Man-Machine system, the tasks which fall upon him and the nature of his work are deeply modified, in that normal running is automatically controlled. In this case, the efficiency of human intervention depends on the ability to act at the right moment, that is to say to quickly change from a state of passive observation to one of active involvement in the process (1). In numerous situations the role of the H.O. consists of anticipating malfunctioning in order to reduce its frequency of occurrence. At any given time therefore the response of the H.O. will depend on:

(a) his capacity to move quickly
(b) his idea as to the state of the system
(c) his knowledge of the causes which may have brought about the disturbance in the system. The latter condition implies that the H.O. has at his disposal, at all times, the data concerning the development of the technical parameters i.e. the changes which occurred a certain time before the incident.

Furthermore, over the last several years there has been a considerable growth of shift work notably in automatic process control situations. At the same time chronobiological research has proven the existence of cyclic variations affecting the functioning of the different systems of living creatures, including man (2,3). This circadian rhythmicity seems to persist, in a lot of cases, in shift workers at least when they are subjected to rotations every 2 or 3 days (4). Field studies, have, until now, mainly concerned either the analysis of various physiological (temperature, heart rate, blood pressure, urinary excretions etc...) and psychophysiological factors (reaction time, flicker, clock test, memorisation etc...), or the recording of various operator performances. Very rarely have researchers (5,6) shown interest in the real behaviour of the operators during the working period, they were satisfied to evaluate the direct or indirect consequences of the activity. However, in automated situations where the operator is not involved in the normal running of the process, it could be thought that he would adopt strategies compatible with his functional state at the time (6). So the psychophysiological capacities of the H.O. make up a set of conditions which can change, to a varying degree, the reliability and the security of the Man-Machine system.

The present report aims at studying the actual functional state of the human operator in control of continuous processes.
PRESENTATION OF THE STUDIED SITUATION METHODS

The company chosen is a chemical factory situated in south-west France. It employs about 2 000 people.

The role of the studied plant is to produce a synthetic gas (ammonia) from natural gas. It is a work station which, permanently manned, uses the shift work system with shift changes at 4.00 am, mid-day and 8.00 pm on a short rotation pattern \([2 - 2 - 2(±1)]\) 3.

After 3 weeks of work, the fourth week includes a resting period and on call days. Five teams of eight members each are responsible for the running of the installation, in particular, two (or three) of the members of each team have to follow the process and the correct running of annex installations from within a control room (fig. 1A).

FIG. 1:  
A - Scale diagram of the control room. The numbers correspond to the different information zones (see text).

B - Schematic representation of the different panels (2 to 16), of the plant schema (17) and of the display unit (1).

Audio warnings and lights indicate any malfunctioning. During normal running as well as their supervision duties, the operators must regularly take readings and keep a log book.

In the control room there is a display of various information: firstly an illuminated plant schema and also on a wall divided into panels grouping together 292 sources of information; finally there is a control desk with a diode display unit (and its keyboard) which is linked up to
78 temperature probes. Numerous controls, 192 in all, are spread over the instrument panels and the control desk.

Analysis of operator activity was carried out in two different teams over two observation periods (spring-summer 1979 and 1980). The method which was used (7) consists of observing and noting, on recording charts, elements of spontaneous operator behaviour. Two analysis charts were used. The first concerns motor activity and the speech behaviour of the operators (30 minutes observation every 2 hours) and the second deals with changes in the direction of the head and eye axis of one of the operators (30 minutes every 2 hours) who was randomly chosen each time. These recordings involved taking note, chronologically of each movement of the line of sight in the direction of one of the instrumented zones of the control room (fig. 1B). The latter consist of: the control desk display unit (Zone 1), the plant schema (Zone 17) and also the various wall panels (Zones 2 to 16).

RESULTS

1) Circadian variation in the activity of the operators

Various activities, not directly linked with work, such as changes in posture, locomotor activity and conversation show that the behaviour of the operators is subject to a high degree of circadian rhythmicity with minimum activity in the middle of the night and high activity at the beginning of the afternoon (P < 0.01) (Figs. 2 and 3).

![Chart showing hourly variation of the frequency of position changes. The broken lines indicate shift changes. The dispersion index represents the standard deviation.](chart.png)
FIG. 3: Hourly variation of the time spent, by the operators, in the sitting position (see fig. 2).

Furthermore, during one certain task carried out in free time (copying on to a chart 78 temperatures detected by probes in the various installations the values can be displayed, when desired, in the control room) a large increase can also be observed in the time taken over it at the end of the night or the beginning of the morning.

These variations in operator activity can be explained by neither variations which might arise in production (difference between the 3 shifts non-significant), not by a difference in the incidents (no significant variation detectable) not finally by the presence or absence of maintenance teams or members of the management.

2) Analysis of the supervising activity of a team of 2 operators

2.1. Change of line of sight frequency

The hourly frequency of line of sight changes varies considerably over 24 hours (P < 0.001). The average number of glances in the direction of any of the zones in the control room passes from 0.54 per minute at 3.00 a.m. to 1.59 per minute at 5 pm (fig. 4). It can however be noted that the start of the night shift shows relatively high values.
2.2. Qualitative activity variations

Whatever the shift considered, operators do not look at the various zones evenly. Certain zones (e.g., 5, 7, 10, and 13) are frequently watched whereas others (3, 11, and 17) are almost never looked at. However, taking into account the overall reduction in the operators' activity at night and their high level of activity in the afternoon, the data should be interpreted in a relative manner. Furthermore, the high value of a given zone could be linked to a variation in the frequency of incidents in that zone; it is for this reason that we estimated the values in percentage and, also, that we took into account the regularity (or non-regularity) of glances at each zone during all our observations. We thus calculated a relative glance frequency which allowed us to compare the various shifts (fig. 5). It was seen that, on considering all the zones, the order of importance accorded by the operators is almost identical over 24 hours. For the 6 zones considered the most important however, there are large differences concerning glance frequencies. Zone 5, for example, which is the most watched at night is in fourth and fifth position in the morning and the afternoon (P < 0.01).

3) Comparison between the different teams

Comparison of the results obtained in a team of 2 operators (team I) with those of a team of 3 operators (team II) brings to light a certain number of results which, not only confirm the previous results but also slightly modify them.
In as far as the overall supervision activity is concerned, the presence of three operators tends to reduce the amplitude of circadian rhythmicity (fig. 6). Whereas in team I the difference between the maximum and the minimum was from 1 to 3, in team II this difference was reduced from 1 to 1.7. This phenomenon results notably from a rise in the nocturnal activity for the whole team.

The order of importance of the various zones in the control room confirms, in both teams, that certain information is important whereas other elements are of less pertinence. Similarly it is noticed that some zones which are considered important at night (e.g. zone 7) are not considered so during the day - this is true for all the teams (fig. 7). We should point out however that a difference can exist between the 2 teams concerning the relative importance of various parameters in the installation. For example, Zone 5 is looked at much more by team II than by team I, on the other hand zone 1 is rarely surveyed, apart from during temperature recordings, by team II.

CONCLUSIONS

From the data collected throughout this study some questions can be posed as to the reliability conditions of a Man-Machine system considering that the overall reliability of the system is the product of the reliability of automation and that of human operator. We have observed that: a) The H.O. is not stable, but he presents, depending on the time, large differences in the amount of information he perceives.
FIG. 6: Distribution of glances (in per cent) during the three shifts in team I (2.C) and team II (3.C).

FIG. 7: See Fig. 6 (Team II).
b) The H.O.'s use different working procedures, in team I at night glances are concentrated towards 3 zones. There is more dispersion of looks in the afternoon.

c) The mental representation of the system which the H.O.'s build up changes with the different positions occupied and the relationships which occur within a working team.

As mentioned at the beginning of this study, in automated situations the operator must pass rapidly from a state of passive observation to one of active intervention in the process. Part of his working capacity then must be rapidly mobilised to track down the causes of the incident, come up with a solution, make a decision and carry it out. It is however clear that, notably at night, the H.O. does not have at his disposition either full knowledge of the situation or the possibility to act quickly. He should therefore be helped to pass from passive observation to active intervention and also to arrive quickly at his diagnosis. The aid should take the form of the rearrangement of the indications and the controls and by the development of procedures which allow various solutions to be tested and the consequences of the choices to be evaluated.

From the point of view of organisation, it seems that the present tendency to reduce the number of staff at night is in fact prejudicial for both the operators and the process. Indeed, it does not allow a H.O. to be temporarily released from his activity in order to be more able of fulfilling the requirements of any situation. From this point of view it seems that the presence of an additional team member would allow, through task alternation, nocturnal activity to be maintained at a relatively high level.

- (This study has been supported by the D.G.R.S.T. (79-7-1107).

REFERENCES

(1) OCHANINE (D) et VENDA (V) - De l'accroissement de l'efficacité du travail des opérateurs dans les systèmes automatisés. Questions de Psychologie, 1962, (3) (en russe).

(2) KLEITMAN (N) - Sleep and wakefulness. 1963, Univ. Chicago Press, Chicago.

(3) REINBERG (A) - Des rythmes biologiques à la chronobiologie. 1979, Gauthier Villars, Paris.

(4) RUTENFRANZ (J) - Analysis of psychophysiological load due to shift (night) work. Proceed. Analysis and Health evaluation of work place, Portoroz (Yougoslavia), 1980, 621-643.


(6) CHRISTOL (J), DOREL (M), QUEINNEC (Y) et DE TERSSAC (G) - Horaires de travail et régulation des conduites. Le Travail Humain, 1979, 42, (2), 211-231.

(7) QUEINNEC (Y) et DE TERSSAC (G) - Variation temporellee du comportement des opérateurs : le cas processus à feu continu. Le Travail Humain, 1980, 43, in press.
MODELLING THE HUMAN OPERATOR'S SUPERVISORY BEHAVIOR

Ted N. White*
Delft University of Technology, Dept. of Mech. Eng.
Lab. for Measurement & Control
Man-Machine Systems Group
Delft, The Netherlands

ABSTRACT

The investigation reported here deals with the description of the behavior of the human operator performing the supervision of a setpoint controlled process under steady state condition. The description will be given in terms of a mathematical model.

In verifying the model structure, certain task variables have been defined. Then, to each of these task variables, the particular influence of the variable on the parameters of the human operator model has been postulated. The data analysis is based on statistical and stochastic procedures.

It is shown that the model structure is applicable to the description of the human supervisory behavior. However, the parameter estimation procedures are complicated, time consuming and thus expensive.

1. INTRODUCTION

The proper functioning of technological systems, diverging from very simple to very complex, needs the ability of human intervention in most cases. More specifically, it can range from a direct manual control to a supervisory control mode.

Because of the ever increasing automatization in industry, new demands have been put on nowadays operators. They are no longer functioning as direct, manual controller; their tasks became more and more the supervision of multi-variable, slowly responding, and complex systems. As a consequence, this task is composed of monitoring the displayed system outputs, controlling the setpoints of the automatic controllers and performing the fault management.

Due to sudden process changes or disturbances, the output variables might exceed the boundaries which are specified by the management or by safety regulations. Based on the perceived deviations, the supervisor decides to change one or more setpoints. Moreover, when the information needed to supervise the plant is insufficient, the operator might decide to ask for additional information which may lead to a setpoint correction. This steady state control aspect of the supervisory behavior is in general important in process control, that is in the control of continuous processes. It is called tuning. The shut down and start-up procedure, also an important control mode, is mostly executed by special teams or automatic controller systems.

In nuclear power plants, however, tuning is far less important than fault management. These plants are highly automated, and the main task of the operator is to search for malfunctions, fault diagnosis and fault management.

In this investigation the attention is focussed on the tuning task of the supervisor.

2. MODELLING HUMAN CONTROL BEHAVIOR

The behavior of an operator performing a supervisory task, or any other manual/cognitive task, can be described in different ways.

For many years, scientists try not only to describe the behavior of the subjects

*This research is partly sponsored by the Netherlands Organization for the Advancement of Pure Research (ZWO).
of interests after the experiments, but, they also try to make predictions upon the behavior to be expected, thus before the experiments. Among the many theories which are applied in order to predict human operator behavior during manual control tasks, the cybernetic models are well-known. In this philosophy the human operator is regarded as a system, a central processor, with inputs and outputs. So, human operators can be described as input - output relations. This idea is not new; back in the twenties, the behaviorists already described human beings as data processors with an unknown central processor, the so-called "blackbox", which transferred stimuli into certain responses. By the application of a wide variety of stimuli and by measuring all responses, the input-output relationships could be determined. On the basis of system theory, from the input-output relations a model, that is a model structure and model parameters, can be derived. Consequently the "black box" is whitened.

Although in manual control situations cybernetic models can be applied successfully to predict human behavior [ACMC 1964 – 1981], up to now it is still not possible to develop such prediction models for supervisory control situations. The major differences between direct manual control of vehicles and the discrete characteristics of supervisory control in process industry can be illustrated by Table 1. [Beaverstock, Stassen, Williamson, 1977]

<table>
<thead>
<tr>
<th>Data Processing Aspects</th>
<th>Vehicle Control</th>
<th>Process Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Direct</td>
<td>Indicative</td>
</tr>
<tr>
<td>Dimension</td>
<td>About 10</td>
<td>Between 100 and 1500</td>
</tr>
<tr>
<td>Time constants</td>
<td>From sec. to min.</td>
<td>From min. to hours</td>
</tr>
<tr>
<td>Abstractness</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Origin disturbances</td>
<td>Known</td>
<td>Unknown</td>
</tr>
<tr>
<td>System properties</td>
<td>Linear</td>
<td>Non-Linear</td>
</tr>
<tr>
<td>Constant</td>
<td>Time-varying</td>
<td></td>
</tr>
<tr>
<td>PRESENTATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration</td>
<td>Integrated displays</td>
<td>Non-integrated displays</td>
</tr>
<tr>
<td>Parallel/series</td>
<td>Mostly parallels</td>
<td>Parallel and/or series</td>
</tr>
<tr>
<td>Number of displays</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Time window</td>
<td>Instantaneous and predictive</td>
<td>Instantaneous and history (trend)</td>
</tr>
<tr>
<td>Modes</td>
<td>Visual, auditive, vestibular</td>
<td>Visual, auditive</td>
</tr>
<tr>
<td>Overview</td>
<td>Good</td>
<td>Poor</td>
</tr>
<tr>
<td>MANIPULATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct feedback</td>
<td>Proprioceptive</td>
<td>No proprioceptive</td>
</tr>
<tr>
<td>Response</td>
<td>Immediate</td>
<td>Next shift</td>
</tr>
<tr>
<td>Strategies</td>
<td>Error correction</td>
<td>Boundary control</td>
</tr>
<tr>
<td>Accuracy</td>
<td>More or less consistent</td>
<td>Variable</td>
</tr>
<tr>
<td>Quantitative</td>
<td>More or less qualitative</td>
<td></td>
</tr>
<tr>
<td>CONFIGURATION</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical size</td>
<td>Small</td>
<td>Large</td>
</tr>
<tr>
<td>Personnel training</td>
<td>Highly and consistently</td>
<td>Less rigorously</td>
</tr>
<tr>
<td>Human factors</td>
<td>Well developed</td>
<td>Very little developed</td>
</tr>
</tbody>
</table>

Table 1: Comparison between vehicle control and process control
After Beaverstock, Stassen and Williamson [1977]
3. MODELING HUMAN SUPERVISORY BEHAVIOR

3.1 Definition of the system state and model description

The human operator's supervisory behavior in a stationary control mode, where a dynamic, slowly responding, and complex system is setpoint controlled is characterized by periods of waiting and of observing output variables followed by periods of executing control actions.

The process to be supervised is described by a linear, dynamic system with state $x(k)$. The system is disturbed by a noise $w(k)$. It has a controlled input $u(k)$ whereas the output $y(k)$ consists of a linear combination of the state variables $x(k)$; the system can be observed by the human operator via a display. This observation can be a continuous observation or an additional sampling of a particular system output.

System theory defines the state of a discrete system as "some compact representation of the past activity of the system complete enough to allow us to predict, on the basis of the inputs, exactly what the outputs will be and also to update the state itself" [Padulo, Arbib, 1974]. "So the future values of the state as well as the outputs are only determined by the initial values of the state and the following inputs" [Kok, Van Wijk, 1978]. Therefore, the input-output relation of a dynamic system can be described by two equations: One equation yielding the progressive change in time of the system state due to the inputs, and a second equation associating the outputs of the system to the actual value of the system state, thus:

$$x(k + 1) = Ax(k) + Bu(k) + Gw(k);$$
$$y(k) = Cx(k).$$

The matrices $A$, $B$ and $G$ can be regarded as system parameters, and the matrix $C$ is the display parameter.

In order to describe human operator's supervisory behavior, the model suggested by Kok and Van Wijk [1978] will be applied. This model is based on the philosophy that the human supervisor acts as some kind of sub-optimal controller, taking into account the inherent limitation that his observation of the output process variables can be inaccurate. This limitation is modelled by a noise $v_0(k)$, called the observation noise. It is assumed that the human supervisor possesses exact knowledge of the generated control signals $u(k)$. Hence, the model consists of three subsystems: an observer, a controller and a decision making element, DME, the last one acting on each of the two other subsystems (Fig. 1).

![Fig. 1: Structure of the observer/controller/decision model](image-url)
The structure of the observer is depicted in Fig. 2; it is modeled as a Kalman filter, thus it contains a replica of the system including the display. Based on the observed output variables, on the knowledge of the dynamics of the system, on the knowledge of the statistical properties of the noise acting upon the system, and on the setpoint changes which have been made already the observer estimates the state of the system $\hat{x}(k)$ and the associated variance of the estimation error $\psi_y$. In this part of the model, the noise vectors $v_O(k)$ with the covariance matrix $\psi_{v_O}$ is assumed to be white and normally distributed. As a consequence the matrix $\psi_{v_O}$ is a diagonal matrix. The quantity $\psi_{v_O}$ models the rate of uncertainty of the supervisor about the perceived outputs of the system. The quantity is a parameter of the observer subsystem.

The subsystem, the controller, can be regarded as a variant of the optimal control law. It is supposed that the human operator will control deviations, from outside the limits, back to the nominal value in a certain period. Hence, when the period, in which the output variable will reach the nominal value, is determined, then the amplitude of the setpoint correction is also fixed; a short period requires a larger amplitude than in a longer period. This trade-off between period and amplitude is modeled by the parameter "p" in the single step control law which will be used as controller.

From this control law: $u(k) = u_{nom}(p) - L(p)\hat{x}(k)$, where $u_{nom}$ denotes the nominal input value, $L$ denotes the gain factor and $u(k)$ the magnitude of the corrective action, all the quantities are known except $p$, the model parameter, while the gain factor, $L$, is dependent of $p$. It is supposed that the human operator bases the quantity $u(k)$ on the estimation of the actual system state, $\hat{x}(k)$, and is supposed not to take into account effects of the system noise during the period in which the action is kept constant. With the single step control law it will be possible to reach in $p$ time steps an output vector $y(k + p)$ which is equal to the nominal output vector, $y_{nom}$, as long as the system noise is zero within that period. Because the system noise is probably not zero a "subgoal" will be reached.

The DME provides two functions: Firstly it models the moments on which an additional sample has to be taken and secondly it models the instants in time on which
a setpoint correction has to be performed. The inputs of the DME are the system state estimate \( \hat{x}(k) \) and its variance \( V_x \); the first function makes use of \( \hat{x}(k) \) and \( V_x \), the last one uses \( \hat{x}(k) \) only.

The additional observation sampling is modeled according to a hyperbolic decision rule (Fig. 3). Whenever the combination of the absolute value of the estimation of the deviation \( |\hat{y}_i(k) - y_{nom,i}| \) with regard to the nominal value, \( y_{nom,i} \), of an output, \( y_i(k) \), and the uncertainty about that particular estimation, \( \sigma_{y_i}(k) \), exceeds a criterion value, a sample will be taken. The quantities \( \hat{y}_i(k) \) and \( \sigma_{y_i}(k) \) are derived as a linear function of the state estimate, \( \hat{x}_i(k) \), and the associated variance, \( V_x \), respectively. The modeling parameters are the curvature of the hyperbolic line, \( CE_i \), and the asymptotic value or minimal accepted uncertainty, \( \sigma_{y_i \min} \), where \( i \) denotes the \( i \)-th output variable.

The controller decisions are modeled as a straight line (Fig. 4). Whenever, the absolute value of the deviation between the estimated output and the nominal value, \( |\hat{y}_i - y_{nom,i}| \), exceeds the line \( q_i \), a corrective action will be made. The upper or lower tolerance limit, \( q_i \), is thus also model parameter.

\[
\sigma_{y_i} \uparrow \quad \frac{CE_i}{\sigma_{y_i} - (\sigma_{y_i})_{\min}} \quad |\hat{y}_i - y_{nom,i}| > |\hat{y}_i - y_{nom,i}|
\]

\[
\sigma_{y_i} \uparrow \quad \text{no control action} \quad |\hat{y}_i - y_{nom,i}| \quad \text{control action} \quad q_i \quad |\hat{y}_i - y_{nom,i}|
\]

Figure 3: The hyperbolic decision-rule

Figure 4: The linear decision-rule

### 3.2 Separation Theorem / Model hypotheses

As pointed out already in the preceding section, the Observer/Controller/Decision-model, OCD-model, is based on the conceptual framework of the Optimal Control Model of Baron, et. al [1969], Kleinman, et. al [1970, 1971]. The basic assumption of this model is that a well-trained and highly motivated human operator will act in a near optimal way, subject to his internal constraints which limit his performance, and subject to the extend to which he understands the task objectives. This model can appropriately be used to predict manual control behavior ACMC, [1964 - 1980]. The OCD-model is an extension of the Optimal Control Model, just by the introduction of a DME. It is also different by the fact that a single step control law as a modification of the optimal control law is proposed, instead of the optimal, continuous control law. Finally it differs by the fact that most of the internal constraints which were taken into account in the OCM are assumed to be neglectable. This assumption seems sensible due to the large time constants involved in the steady-state supervisory control mode; hence aspects like reaction times of 0.1 sec and neuromuscular dynamics with time constants of 100 to 200 msec, important in the OCM, can be neglected in the OCD-model.
In the case of the Optimal Control Model the well-known separation theorem is valuable as a basic tool. No arguments can be found, to proof that the separation principle can also be applied in the case of the OCD-model, since with the introduction of the DME, the model becomes non-linear, whereas linearity is one of the necessary conditions in order to be allowed to apply the separation principle. The importance of the separation theorem, however, makes that it may be fruitful to study to what extent the DME influences this important principle; therefore the following hypotheses have been stated:

1. Variations in the intensity of the system disturbances are changing the parameter values of the observer; they are not influencing the parameters of the controller. Changes of the parameter values of the decision making element for observation actions will also take place.
2. Variations in the display structure are also changing the parameter values of the observer, and the decision making element for the observation actions; they are neither changing the parameter values of the controller, nor the parameters of the decision making element for the controller actions.
3. Variations in the task to be performed are changing the parameter values of the controller and the decision making element for the controller actions; they are not changing the parameter values of the observer and the decision making element for the observation actions.
4. Variations in the dynamics of the system to be supervised are changing all the parameter values in the model.

In Table II the postulated relations between the task variables and the parameters of the model are represented, where a plus, +, denotes a dependency and where a zero, 0, denotes independency.

Table II: Postulated relations between task variables and model parameters.

<table>
<thead>
<tr>
<th>Parameters of:</th>
<th>Observer part</th>
<th>Observer Decision part</th>
<th>Controller Decision part</th>
<th>Controller Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>System Disturbances</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Display Structure</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Task Requirements</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>System Dynamics</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Since the parameters themselves are difficult to interpret, and since they cannot be identified easily, a smooth way of verification of these hypotheses cannot be found directly. Therefore, it is worthwhile to transform the hypotheses to directly measurable quantities. Since quantities like the number of observation actions, the number of controller actions and the amplitude of the control actions can be measured, and since they are directly related to the structure of the system to be supervised, one can reformulate the hypotheses as follows:

1. Variations in the intensity of the system disturbances will be reflected in the number of observation actions, and will be reflected in an additional way in the number of controller actions. However, it will not be reflected in the mean amplitude of the controller actions.
2. Variations of the display structure will only be reflected in the number of observation actions; they will neither influence the number of controller actions nor the mean of the amplitudes of the controller actions.
3. Variations in the task requirements will be reflected in the number of controller actions; they will not be reflected in the number of observation actions (Table III and Fig. 5).

Two remarks have to be made in relation to these hypotheses:

In the first place the task variable "system dynamics" will not be taken into account.
account here, because the verification of hypothesis on the influence of the system dynamics as a task variable will increase enormously the number of experiments.

In the second place the supplementation "in an additional way" in the first hypothesis needs some explanation. To verify the separation theorem, it has to be shown that optimization of the filtering and the control problem, separately, will lead to the same results as optimization of the combined problem. Referring to the model first-discussed, and taking into account the influences of the task variables just-mentioned, a consequence of the separation theorem is that the super-position principle should hold. Hence, the task variable system disturbances is thought to contribute, in an additional way, a number of controller actions which is independent of the number of controller actions attributed by the task variable task requirements.

Table III: Postulated relation between task variables and direct measurable quantities.

<table>
<thead>
<tr>
<th>Quantity:</th>
<th>Number of Observation actions</th>
<th>Number of Controller actions</th>
<th>Mean Amplitude of Controller actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Disturbances</td>
<td>+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Display Structure</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Task Requirements</td>
<td>0</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 5: Postulated relation between task variables and direct measurable quantities

4. EXPERIMENTAL DESIGN

4.1 Description of the research vehicle

In trying to verify the hypotheses stated in the last section, one needs a research vehicle on which experiments can be performed. For this reason a setpoint controlled process simulation has been implemented on a PDP 11/34 computer. The process, a Utility Plant, is considered to be a linear system and it consists of a Boiler, a Back Pressure Turbine and a Condensing Turbine. The Boiler produces high pressure steam which is pressure controlled. The Back Pressure Turbine consumes H.P.-steam and produces low pressure steam and electrical power; this low pressure steam is flow controlled. The Condensing Turbine consumes H.P.-steam and produces only electrical power. The total amount of electrical power produced is power controlled (Fig. 6).
4.2 Description of the task of the supervisor

The task of the human operator is to keep the three output variables, i.e. the pressure of the H.P.-steam, the flow of L.P.-steam and electrical power which are corrupted by system disturbances, within specified boundaries. These boundaries are symmetrically located around a certain nominal value. The outputs are displayed digitally; the boiler output can be observed continuously whereas the two other outputs can be sampled only on request. The task of the operator is to sample as few as possible, and to correct the set-points as few as possible. The control should be performed only with small corrective actions.

The operators will be master students of the Lab. for Measurement & Control of the department of Mechanical Engineering.

4.3 The experimental set-up

Three task variables have been chosen for the experimental set-up, i.e. a variation in system disturbances, in display structure and in task requirements.

Variations in the system disturbances: Each output variable is disturbed by a first order noise with a low intensity, denoted as n, or with a high intensity, denoted as N.

Variations in the display structure: Each deviation between output and a specified boundary can be indicated in one of the two following ways. By the local alarm structure, denoted as D, a direct indication is given which output is out of the boundaries and in what direction the variable is changing. By the central alarm structure, denoted as D, it is only indicated whether one of the output variables is deviated from one of the specified limits. The human operator has to investigate which output is deviating and in what direction the output is changing; he even has to make sure that there is only one output deviating from the specified limits. More information can be asked for by means of taking additional samples.

Variation in the task requirements: The lower and upper limits of each output variable are specified. For the case that lower and upper limit are close to each other the task is difficult (B); for the case they are far away from each other, the task is rather easily (b).
In the experimental set-up each condition consists of a combination of one of the two levels of the three task variables involved. Therefore, eight conditions can be formed. In the first verification phase four conditions were considered, they were: n.d.b; n.D.b; N.d.b; N.D.b. In the second verification phase the conditions were: n.d.B; n.D.B; N.d.B; N.D.B.

5. TESTING THE HYPOTHESES

In trying to verify the hypotheses, one needs to be sure to base the verification on a robust set of data of well-trained operators. The criteria in order to conclude whether the data is originating from well-trained subjects, are twofold. At first, the number of samples taken and the number of corrective actions performed should approach a constant level for each condition. Secondly, the computed absolute value of the deviation integrals, based on the time that the signal is outside the specified boundaries, should be as close as possible to zero.

For testing the hypotheses on the basis of the first four conditions the data of nine trials were considered. The necessary conditions in order to apply parametric statistics were checked. Thereafter, an investigation whether the collected data were statistically seen, influenced by possible learning effects was performed. No learning effects could be detected. As an indication, more than 100 (catch) trials preceded the actual 36 trials. Fig. 7 shows the 99% confidence interval of a T-Student distribution of the number of samples taken and the number control action performed, for each of the four conditions. The figure shows that a change in the display structure effects only the number of samples taken and not the number of controller actions executed; Table IV reveals the exact results.

Table IV: Results of a T-test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of cases</th>
<th>Mean</th>
<th>SD</th>
<th>2-tail prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampl group n,d,b</td>
<td>9</td>
<td>0.7</td>
<td>2.0</td>
<td>0.000</td>
</tr>
<tr>
<td>group n,D,b</td>
<td>9</td>
<td>21.7</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Contr group n,d,b</td>
<td>9</td>
<td>77.2</td>
<td>3.6</td>
<td>0.353</td>
</tr>
<tr>
<td>group n,D,b</td>
<td>9</td>
<td>79.8</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td>Sampl group N,d,b</td>
<td>9</td>
<td>14.8</td>
<td>6.5</td>
<td>0.000</td>
</tr>
<tr>
<td>group N,D,b</td>
<td>9</td>
<td>47.0</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>Contr group N,d,b</td>
<td>9</td>
<td>115.9</td>
<td>16.6</td>
<td>0.242</td>
</tr>
<tr>
<td>group N,D,b</td>
<td>9</td>
<td>108.2</td>
<td>9.1</td>
<td></td>
</tr>
</tbody>
</table>

Furthermore it can be seen, that variation in the intensity of the system disturbances is reflected in the number of observation actions, whereas it is also additive in the number of controller actions; Table V shows the exact results.
As long as the task requirements have not been changed in the four conditions, the mean amplitude of the controller actions was supposed to remain constant. For verification the results of the T-test showed no significant difference between the four conditions (Fig. 8). So the hypotheses with respect to the task variables, display structure, and system disturbances, can be accepted completely.

The third task variable the "task requirement", is supposed to have only a effect on the number of controller actions and not on the number of observation actions. By comparing Fig. 9 and Fig. 7, it can be seen that only in one case the results of a variation in the task requirements deviates from the postulated effects; in condition n,D,b more samples are taken than in condition n,D,b.

Table V: Results of a T-test

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of cases</th>
<th>Mean</th>
<th>SD</th>
<th>2-tail prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampl group n,d,b</td>
<td>9</td>
<td>0.7</td>
<td>2.0</td>
<td>0.000</td>
</tr>
<tr>
<td>group N,d,b</td>
<td>9</td>
<td>14.8</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>Contr group n,d,b</td>
<td>9</td>
<td>77.2</td>
<td>3.6</td>
<td>0.000</td>
</tr>
<tr>
<td>group N,d,b</td>
<td>9</td>
<td>115.9</td>
<td>17.6</td>
<td></td>
</tr>
<tr>
<td>Sampl group n,D,b</td>
<td>9</td>
<td>21.7</td>
<td>3.1</td>
<td>0.000</td>
</tr>
<tr>
<td>group N,D,b</td>
<td>9</td>
<td>47.0</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>Contr group n,D,b</td>
<td>9</td>
<td>79.9</td>
<td>7.4</td>
<td>0.000</td>
</tr>
<tr>
<td>group N,D,b</td>
<td>9</td>
<td>108.2</td>
<td>9.1</td>
<td></td>
</tr>
</tbody>
</table>
Results obtained with a less trained subject were similar with respect to the task variables, display structure and system disturbances in the conditions: n,d,b; n,D,b; N,d,b; N,D,b. The results were slightly less significant (p<.02). The results with respect to the task variable task requirements do not agree with the hypotheses; even the results in the conditions n,d,B; n,D,B; N,d,B; N,D,B where the task variables display structure system disturbances were varied systematically, were rather poor, although the trends were in the predicted direction.
In trying to understand why the highly significant results with the well-trained subject were not replicated by the less-trained subject, and in particular for the case of the task variable task requirement, an additional investigation was made. Due to the fact that the second subject sampled about three times more than the well-trained subject in each condition and due to the fact that he also controlled about 1.3 times more than the first subject, the idea of a roof or ceiling effect, spoiling the expected results, was probed. In a new experimental set-up the well-trained subject was excluded from any alarm indication when he was supervising the Utility Plant. So the task variable display structure got a third level. Here it should be noted that no alarm indication in the other two levels, local alarm and central alarm, carries at least the information that none of the outputs is out of the boundaries. In the new, experimental situation, he even did not get this information, so the subject had to sample much more than before, independent of whether the system disturbances were high or low, and whether the boundaries were small or large. The number of samples became about equal to the number of samples of the second subject. The results, obtained from more than 90 trials, indicated very strongly that a roof effect could have been the cause of the differences between the results obtained with the well trained subject and those with the less trained subject.

6. PARAMETER IDENTIFICATION

In order to verify the hypotheses with respect to the parameters of the model, no analytical method for parameter estimation could be derived. Therefore, certain iteration procedures have been used. During the parameter identification, many difficulties showed up. One difficulty appeared in the problem of not converging to a minimum of one of the accepted criteria, another difficulty appeared with the large computing time of 45 seconds CPU of a large scientific off-line computer system for only one iteration step. Therefore, much effort has been put in reducing the number of iterations. The main parameter of the Observer part, $\Psi_0$, has been chosen constant; the parameter value was chosen 1% of the variance of the associated observer output variable according to results obtained with the Optimal Control Model [Levison, et.al., 1969]. This parameter will be identified in the near future.

The parameters for the observation part of the DME, the curvature parameter $CE_i$, and the asymptotic $\sigma_{y\text{ min}}i$ of this curve, had to be identified for each of the outputs to be sampled. Thereafter, an adaptive random search method became too expensive. For this reason a graphical method was proposed. Suppose the human operator sampled the i-th output variable at time: $t_1$, $t_2$, $t_n$. By applying the observer part, the absolute value of the deviation between the estimation of the i-th output variable, $\hat{y}_i$, and the nominal value of the i-th output variable, $y_{\text{nom}i}$, and the associated uncertainty about this estimation $\sigma_{yi}$, can be derived, respectively:

$$|\hat{y}_i - y_{\text{nom}i}|, \ldots, |\hat{y}_n - y_{\text{nom}n}|, \ldots, |\hat{y}_m - y_{\text{nom}m}|$$

$\sigma_{yi}, \ldots, \sigma_{ym}$. The k-th sample is located on the hyperbolic decision line:

$$|y_{ik} - y_{\text{nom}i}| = \frac{CE_{ik}}{\sigma_{ik}}$$

For the k-th sample, the quantities $\hat{y}_{ik}, y_{\text{nom}i}$ and $\sigma_{yi}$ are constants; so the relation of these quantities with the parameters $CE_{ik}$ and $\sigma_{y\text{ min}ik}$ is known. This relation can be given as the straight line,

$$CE_{ik} = |y_{ik} - y_{\text{nom}i}| * \sigma_{y\text{ nom}ik} + |y_{ik} - y_{\text{nom}i}| * \sigma_{yik},$$

and can be plotted in a $\sigma_{y\text{ min}i} - CE_i$ plane. This relation can be given for all samples of the i-th output variable.

Fig. 10 shows the envelop lines of all the n "decision lines". In the area of convergence of the envelop lines, the optimal parameter set is supposed to be located. This idea can be explained when 2 samples are taken; the crossing of
the two lines, Fig. 11, shows the only solution for two samples to be located at the same hyperbolic line with the associated asymptote. In this case convergence is narrowed down to one point, and the variance is reduced to zero.

Figure 10: Envelop lines in a CE-$\sigma_{y\min}$ plane

Figure 11: Two decision-lines

By applying this method, very good results were obtained, i.e. the number of samples, taken by the model based on the acquired parameter set, agreed with the number of operator samples taken in the same condition. This result however, does not guarantee that the instants in time of operator sampling and the instants in time of sampling of the model agree with each other. Therefore, the distribution of the sampling intervals of the model, given certain parameter values of CE$_1$, and $\sigma_{y\min}$, is compared with the distribution of sampling intervals of the human operator. The distributions have the shape of binomial distributions. A close resemblance between the two distributions of operator sampling and model sampling for the two to be sampled outputs is illustrated in Fig. 12 and Fig. 13.

Figure 12, Figure 13: Distribution of sampling intervals of two outputs
The parameter, $p$, of the controller subsystem, determines the number of time steps, the setpoint correction will remain constant and determines the amplitude of that correction; that parameter $p$, have to be identified. In this investigation different parameter values have been used, whereas $p = 22$ turned out to be the lowest value before instability of the H.P.-steam control loop occurs. With $p = 22$ also the best control performances were obtained. In order to relate these results with other control situations. Table VI shows the absolute value of the deviation integrals of the three output variables of the supervisor system for several control situations.

**Table VI: Deviation scores of system outputs**

<table>
<thead>
<tr>
<th>Deviation score of:</th>
<th>Boiler output</th>
<th>Back Pressure output</th>
<th>Condensing output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situation:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non operator intervention</td>
<td>400</td>
<td>106</td>
<td>50</td>
</tr>
<tr>
<td>Optimal operator control</td>
<td>0.0</td>
<td>0.01</td>
<td>0.001</td>
</tr>
<tr>
<td>OCD-model control;</td>
<td>137</td>
<td>73</td>
<td>58</td>
</tr>
<tr>
<td>Single step control law ($p=22$)</td>
<td>0.35</td>
<td>10</td>
<td>9.5</td>
</tr>
<tr>
<td>OCD-model control;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal (continuous) control law</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results are clear: The optimal (continuous) control law deviation score agrees mostly with the deviation score of the human operator. However, the characteristics of the way of controlling of the model do not agree at all with the discrete action pattern of the human operator. The number of controller actions, initiated by the OCD model ($p=22$), agrees with the number of control action executed by the human operator; also, the discrete action pattern of the single step control law agrees with human operator control actions, but the overall performance of the model expressed by the deviation scores, is still somewhat troublesome.

7. CONCLUSIONS AND FINAL COMMENTS

Whether the separation theorem is valid or not for the OCD-model, cannot be answered right now. However, it can be said that the results obtained with a highly trained and motivated subject showed the postulated independency of observation and control. It can also be said that most of the model parameters can be identified, although the computing time is still alarming. Therefore, the research will be continued with highly trained subjects; it will deal with changes in the parameter choice of the controller part or even accepting another control law and it will deal with optimizing the parameter identification procedures.

REFERENCES


HUMAN PROBLEM SOLVING IN A PROCESS CONTROL TASK

William B. Rouse and Nancy M. Morris
University of Illinois
Urbana, Illinois 61801 U.S.A.

Abstract

Human problem solving in dynamic environments is considered. Possible conflicts between the goals of stabilization, optimization, and diagnosis are discussed. To study the nature of these conflicts, a process control task called PLANT (Production Levels and Network Troubleshooting) is introduced. Results of a preliminary investigation with PLANT are presented. Experimental variables studied include the size of the processing network (i.e., number of tanks) and the failure rate of the components (i.e., valves).

Introduction

The task of controlling a technical system (i.e., aircraft, ship, process plant, etc.) involves coordinating three goals. First, the system must be stabilized in the sense of maintaining the state of the system within some allowable range. Second, if possible, system performance should be optimized in order to maximize production, minimize energy consumption, maximize safety, etc. Finally, if anything unusual occurs, the source of the unexpected events must be diagnosed and appropriate actions taken.

Unfortunately, these three goals cannot be pursued independently. For example, achieving optimal performance may require that one operate the system on the edge of instability. Further, attention devoted to diagnosis can result in degradations with respect to stabilization and optimization. Thus, beyond performing the particular tasks associated with stabilization, optimization, and diagnosis, one must also achieve an appropriate balance among these three goals.

This task of coordinating three possibly conflicting goals is complicated by the fact that most technical systems are dynamic in that their states and even their structures (i.e., via mode changes and/or failures) evolve in time. As a consequence, the appropriate balance among the goals of stabilization, optimization, and diagnosis is likely to be time-varying. For example, stabilization is likely to be the most important goal during transition phases of system operation and when the state of the system is very close to the limits of the allowable range of operation. Otherwise, unless a failure has been detected, optimization is likely to be the dominant goal. Of course, this presupposes that failures evidence themselves in a yes/no manner. It is more likely that the onset of a failure will produce initially ambiguous symptoms and, as a result, the precedence of diagnosis over optimization will not be immediately clear.

Thus, it would seem that the control of a technical system could be viewed as requiring a two-step process of decision making. First, one must determine which goal is most important at the moment. Then, one must
determine the action or actions most appropriate for achieving this goal. Unfortunately, this perspective is too simplistic. Actions are not chosen in isolation, but instead serve as elements of a sequence of actions aimed at achieving a balance among the goals of stabilization, optimization, and diagnosis. The choice of a sequence of actions so as to coordinate these three goals is the essence of system control.

In this paper, the perspective of system control outlined above will serve as a framework within which the human controller of technical systems will be considered. The discussion will begin with a description of a fairly general system control task which was developed to allow study of human operators as they coordinated possibly conflicting goals and chose control actions accordingly. Next, an experimental study which employed this task will be discussed. Finally, the results of this study will be considered in terms of the framework espoused in this paper.

A Process Control Task

Based on the point of view advocated in the Introduction, as well as the wealth of literature available concerning human operators in process control [references 1, 2, and 3], a process control task called PLANT was designed. The acronym PLANT stands for Production Levels and Network Troubleshooting task. PLANT is a computer-generated network of tanks interconnected by valves and displayed on a CRT. The purpose of PLANT is to process fluid so as to produce an unspecified product.

An example PLANT problem with nine tanks is shown in Figure 1. The number in square brackets to the left of each tank number indicates the current level of the fluid in the tank. The displayed level may include the actual level plus noise, although a zero noise component was employed for the experiment reported in this paper. The symbols to the right of each tank represent valves which may be opened to other tanks. In this example, valves are open between tank numbers 1 and 4, numbers 2 and 5, etc. The operator may close any of the valves already open and/or open any new valves, with a limit of three per tank regardless of the number of tanks in the network. A further restriction is that valves may only connect adjacent columns. Thus, in this example, valves from tank numbers 1, 2, and 3 to tank numbers 4, 5, and 6 could be opened or closed but valves from tank numbers 1, 2, and 3 to tank numbers 7, 8, and 9 do not exist.

The network of tanks receives input to the left most column of tanks and produces output from the right most column of tanks. The amount of fluid produced by each output tank is equal to the amount of input to each input tank, a variable which is controlled by the operator. All other flow in the network is in the direction of from high to low. With each iteration, ten percent of the difference between two connected tanks will flow from the tank with the higher level to the tank with the lower level. ("Iteration" refers to updating the state of the network which is not done in real time, but instead occurs after each decision.)
Figure 1. Example PLANT Network

The operator’s responsibilities when using PLANT include both the rate and quality of production. The rate of production is influenced by both the amount of fluid input to the network and the number of valves open between tanks. As noted above, production per output tank is equal to the amount of fluid input to each input tank. Within the network, the rate of flow increases as more valves are opened. From this perspective, one would like to have maximum input with all valves open.

However, the production network has some limitations. The level in each tank starts at some intermediate level (e.g., 50) and, in order to process fluid appropriately, must be maintained within some minimum and maximum values (e.g., 0 and 100, respectively). If the level of any tank is not within the range defined by the minimum and maximum, the quality of the product produced by that tank is unacceptable. Further, since the quality of the overall output of the network depends on each tank producing acceptable output, all of the network’s production is unacceptable during those periods when any tank’s production is unacceptable.

Since the quality of production depends on every tank providing its contribution, production can also be affected by failures. Failures in PLANT are such that all of the valves associated with a particular tank close, although this is not indicated on the display. As a result, the tank in question does not produce any fluid and consequently the
production of the whole network is unacceptable. The symptoms of a failure are a backing up of fluid and hence, an increased fluid level in the tank whose valves have closed. If the network is in steady-state operation, failures are fairly easy to detect; however, during transition phases (e.g., when configuring to increase production) detection can be more difficult. Failures are diagnosed by making flow readings. When a zero flow through a supposedly open valve is found, the failure is automatically corrected.

A production run with PLANT begins with a display similar to that shown in Figure 1, except that there may be more tanks (e.g., 16 or 25). The levels of each tank begin at an intermediate level, only one valve is open per tank, and the input per tank is set at \( i = 1 \) where \( i \) denotes "input". At the top of the display, "time = 1; current prod. = 0.0; total prod. = 0.0; your action = " will be seen.

The operator is allowed one action per unit time or iteration. For this example, the operator might type "i10" in order to increase production, or something like "i1-5" or "i2-6" in order to configure the network to allow even a larger input, say "i20". After each action, time is incremented, current production calculated, and total production cumulated. The results are then shown at the top of the display and the next action chosen by the operator.

As time goes on, a failure may occur and evidence itself by a slowly increasing level in one tank. To diagnose the failure, the operator would type something like "f6-9" and, as noted earlier, if the flow reading was zero, the failure of the valves associated with tank number 6 would be automatically corrected. A non-zero flow reading would reflect a false alarm and a waste of a unit of time, although an excellent control strategy is not necessarily devoid of false alarms.

As with many tasks, the operator using PLANT has responsibility for more than just opening/closing valves, adjusting input levels, and making flow readings. The operator is assumed to have other duties that require him to be away from the production network. The operator chooses when to be away by typing "sN" where "s" denotes "skip" and \( N \) is an integer specifying the number of iterations to be skipped.

The operator's overall goal is to maximize acceptable production (i.e., total production minus all production during out of tolerance or failure situations) per observation. An alternative goal is maximum acceptable production subject to the number of observations not exceeding some maximum. However, only the former measure of production was employed in the experiment reported in this paper.

A typical strategy for approaching a production run with PLANT is as follows. Since the network starts with low input and few valves open, one has to configure the network for increased production. This initial phase is somewhat analogous to "taking off" in an aircraft or "getting under way" in a ship. One then enters into a steady-state or "cruise" phase where skip commands predominate with the length of the skip inversely related to failure rate of the network. When a failure is detected, a reconfiguring phase may be necessary, especially if the failure has had several iterations to develop. After reconfiguring, steady-state production is resumed.
While the above strategy is representative, there are considerable variations possible. The following experiment illustrates the extent to which variables such as network size and failure rate affect humans' strategies and production results.

An Experiment

Six members (one female) of the experimenters' research group at the University of Illinois served as subjects. All were students or former students in engineering or research staff members.

At the beginning of the experiment, each subject was given a 5-page booklet of instructions for operating PLANT. The instructions included both a description of PLANT dynamics and a discussion of tradeoffs that could influence production (e.g., number of valves opened). The subjects were also informed that production divided by the number of observations of the system (i.e., production per observation) was the measure of system performance that they should attempt to maximize. A hard copy of these instructions was made available to the subjects throughout the experiment.

Two parameters of PLANT were manipulated as independent variables: size of production network and probability of a failure per iteration. Network size refers to the number of tanks in the system; PLANT problems consisted of either 3x3, 4x4, or 5x5 networks of tanks (hereafter referred to as sizes 3, 4, and 5 respectively). The probability of failure per iteration was either .01, .10, or .20.

Subjects served in six 45-minute sessions. A single session consisted of three production runs (one of each network size) of 100 iterations each. Within a session, the probability of failure per iteration remained constant across production runs. At the beginning of each session, subjects were explicitly informed of the failure rate for that session.

The first three sessions, which served as training sessions, were the same for all subjects. Within each of these sessions, the order of occurrence (according to size) of production runs was 3, 4, and 5. Across training sessions, the order of probability of failure was .01, .10, and .20. The last three sessions were experimental sessions. Probability of failure across experimental sessions was counterbalanced between subjects; network size was counterbalanced both between subjects and across sessions within subjects.

Recorded measures of the subjects' performance included total production, acceptable production, unacceptable production, number of observations, time required to execute each command, and the sequence of commands used. Prior to conducting the experiment, a number of hypotheses and expectations as to experimental results existed.

First, it was expected that production per observation would be affected both by failure rate and network size. More specifically, production per observation was expected to decrease with increasing failure rate (due to the greater likelihood of unacceptable production), and increase with network size (since total output relates to both level of output and number of output tanks). Total production (i.e., both
acceptable and unacceptable) should be unaffected by failure rate, and should increase with network size.

The number of observations made should be directly related to failure rate and network size, for the following reasons. As failure rate increases, operators should be less willing to "leave" the system for a long period of time, and thus should check on the system more often. As network size increases, reconfiguring the system should require more observations due to the presence of more valves, so total number of observations should increase.

Detection of failures was expected to deteriorate as failure rate increased and network size increased, simply because of network complexity and a generally less stable system. The number of iterations required to correct a failure should provide direct evidence of this. Furthermore, failures occurring during a period of transition (i.e., while the operator was reconfiguring the system) should go uncorrected for a longer period of time than failures occurring when the system was more stable.

Results

Unless otherwise noted, all analyses referred to in the following discussion were analyses of variance for two independent variables and repeated measures. Multiple comparisons were made using Newman-Keuls tests.

An analysis using production per observation (the overall measure of system performance) as the dependent variable revealed a significant main effect both of probability of failure \( F(2,10) = 66.76, p < 0.001 \) and network size \( F(2,10) = 7.17, p < 0.025 \). For failure rates of .01, .10, and .20, the means were 56.71, 22.95, and 15.08 respectively. Multiple comparisons indicated that the significant difference was between .01 and the other two rates; there was no difference in performance with .10 and .20 probabilities of failure. The respective means for network sizes 3, 4, and 5 were 42.67, 26.95, and 25.22. Although multiple comparisons failed to indicate the locus of significance, the apparent difference was between size 3 and the other two sizes.

An analysis with number of observations as the dependent variable revealed a significant effect of probability of failure \( F(2,10) = 22.51, p < .001 \); as with production per observation, the significant difference was between .01 and the other two failure rates. Means were 40.0, 59.61, and 65.0 for failure rates of .01, .10 and .20. There was no significant difference in number of observations relative to network size. There was also no significant effect of network size or failure rate upon number of valves opened (frequency of "o" commands).

When total production was used as the dependent variable, there was a significant effect of both probability of failure \( F(2,10) = 5.47, p < .05 \) and network size \( F(2,10) = 29.86, p < .001 \). Mean total production for failure rates of .01, .10, and .20 was 2369, 1953, and 2036 respectively; production associated with network size was 2441, 2047, and 1869 for 3, 4, and 5 networks. Multiple comparisons failed to identify the significant difference associated with failure rate; the difference in production related to network size was between size 3 and the other two sizes.
The number of observations required to correct a failure was used as the dependent variable in the next analysis. This analysis indicated no significant effect of either probability of failure or network size on failure detection. However, when frequency of "false alarms" (i.e., flow readings in the absence of failures) was used as the dependent variable, there was a significant effect of failure rate \( (F(2,10) = 6.49, p < .025) \). Mean frequencies were 0.27, 2.4, and 2.3 for failure rates of .01, .10, and .20. As with production per observation and number of observations, the significance was between .01 and the other two probabilities of failure.

Whether or not when the failure occurred had an effect upon failure correction was the subject of the next analysis. A failure was said to have occurred during a transition period if the failure occurred immediately after a valve had been opened. The mean number of observations required to correct a failure occurring during a transition period was 5.079; failures occurring in a non-transition period were corrected in a mean of 2.083 observations. This difference was significant according to a one-tailed t-test for matched groups \( (t = 2.02, p < .05) \).

The effects of failure rate and network size upon length of time between observations were investigated with an analysis involving the number of time units specified in "skip" (s) commands. The main effect of probability of failure was significant \( (F(2,10) = 18.31, p < .001) \); for failure rates of .01, .10, and .20, respective means were 3.92, 3.55, and 3.40. The significant difference was once again between .01 and the other two failure rates: .10 and .20 failure rates had no significantly different effect upon performance. The main effect of network size was not significant. Using frequency of "s" (skip) commands as the dependent variable, there was again a significant effect of failure rate \( (F(2,10) = 11.37, p < .01) \), with that difference being between .01 and the other two rates. Respective means were 9.61, 15.61, and 13.11 for probabilities of failure of .01, .10, and .20.

Discussion

To summarize the results described above, the probability of a failure occurring had a definite influence upon almost every performance measure. As expected, production per observation declined as failure rate increased. The number of observations increased as predicted, with a corresponding decrease in the average length of time operators "looked away". That total production would be affected by failure rate was unexpected, since total production was controlled solely by the subject's choice of level of input; however, as noted above, increased probability of failure was associated with choosing a lower level of input and consequently increased output.

Detection and correction of failures was not greatly affected by failure rate, contrary to expectations. When one considers the significant increase in false alarms associated with increased probability of failure, however, this finding might be explained as an increased tendency to check for failures rather than an indication of ability to detect failures. Further analysis is required to resolve this question. The finding that failures occurring during transition were undetected for
a longer period of time than were failures occurring in non-transition periods was consistent with expectations.

It should be noted that in every case in which there was a significant effect of failure rate, the significant difference was between .01 and the other two rates (.10, .20). Performance under .10 and .20 probabilities of failure did not differ significantly. Lest one be tempted to conclude that there is no difference in performance if the failure rate is "substantial", it should be noted that an examination of the actual failures revealed that there was virtually no difference between the actual frequencies of failures under these conditions. (The actual failure rate for the .10 condition was .03; when the probability of failure was .20, the actual rate was .12.) The reasons for this drastic deviation from the designated .20 failure rate are not completely clear. Since only one failure could occur at a time, one might expect that the actual occurrence of failures would be somewhat less than prescribed; however, this explanation is not a sufficient account of the difference between the actual .12 failure rate and the desired one of .20. It is hoped that this problem can be resolved in future experiments.

The effects of network size were somewhat different from those expected. Recall that an increase in production associated with larger network sizes was predicted. The observed effect was in the opposite direction; production decreased as the size of the network increased. There was no effect of network size evident in number of observations made; the expected result was an increase in observations with increasing size of the network. There was also no increase in the number of valves opened. If these results are taken at face value, it seems that subjects opened a smaller percentage of valves as network size increased, in order to keep the number of observations as small as possible. Whereas this statement does seem to be descriptive of a large portion of the performance observed, it should be pointed out that there was a great deal of variability between subjects; some subjects chose to open a large portion of the valves in the system, and others opened none at all. This variability precludes making broad generalizations, but rather indicates the need to investigate various strategies.

Conclusions

While many of the initial hypotheses were supported, there were some surprises. As expected, performance was best for low failure rates. However, it was not expected that the smallest problem sizes would produce the best performance. For some reason, subjects did not take advantage of the production possibilities for larger networks, especially for the higher failure rates. It may be that failure diagnosis was the dominant goal, although failures that occurred during reconfiguration phases resulted in substantial increases in detection time. Thus, diagnosis was not the only goal. In general, it seems reasonable to conclude that subjects in this experiment were rational but conservative.

For further studies with PLANT, several changes are envisioned. First, non-zero noise levels will be used for the level indicators to provide a more realistic environment. Further, the effects of exceeding tolerances or failures may be changed so as to affect only those output tanks reached by the out of tolerance or failed tank. Finally, the
diagnostic process may be changed to include product quality tests rather than flow readings.

The short-term goal of this work is to develop an understanding of human performance in the PLANT environment. A model of human performance will hopefully result from these efforts. The long-term goal of this research is to develop computer aiding and training schemes to help the human operator in environments similar to PLANT.

Acknowledgement

This research is supported by the National Aeronautics and Space Administration under Ames Grant NSG-2119.

References


session 4
manual control studies

chairman: g. johannsen, frg
CURRENT RESEARCH NEEDS IN MANUAL CONTROL

Elwyn Edwards

1. ABSTRACT

Advances in control technology inevitably lead to changes in the types of problems which are most urgently in need of attention at any particular time. Furthermore, deployment of research resources in a given direction easily leads to a situation in which self-perpetuating programmes are generated. Such factors tend to produce disparities between the genuine problems in control, and the problem-solving activities which are taking place: we have a large lag in the system. Some examples are given of current problems in manual control, particularly in relation to aircraft guidance, and examples are given of ways in which research might be directed towards dealing with such problems.

2. INTRODUCTION

This paper is concerned to review the current research needs in relation to the role of the human operator (HO) in the control of automated systems.

It has been stressed on many previous occasions (see, e.g., Edwards 1976) that the introduction of automation does not necessarily ease the load upon the HO, although the nature of the human contribution is likely to alter considerably. Consequently, the important practical questions in the 1980s are different from those of former decades.

In Table 1, a comparison is made between "traditional" and "contemporary" states of several control system parameters. These parameters are inter-dependent; the taxonomy is purely one of convenience. Principally, attention is directed towards aircraft systems, but many observations might be of equal relevance in the control of such systems as chemical processes or of power generation.
Table 1

Some typical differences between traditional and contemporary control systems in relation to several system parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Traditional</th>
<th>Contemporary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator Involvement</td>
<td>Continuous</td>
<td>Occasional</td>
</tr>
<tr>
<td>Manual Rehearsal</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Control Technology</td>
<td>Analogue</td>
<td>Digital</td>
</tr>
<tr>
<td>Failure Diagnosis</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Integrity Testing</td>
<td>Elementary</td>
<td>Elaborate</td>
</tr>
<tr>
<td>Mental Models</td>
<td>Manageable</td>
<td>Intricate</td>
</tr>
<tr>
<td>System Language</td>
<td>Clear</td>
<td>Obscure</td>
</tr>
</tbody>
</table>

3. SYSTEM PARAMETERS

3.1 Operator Involvement

Manual tracking requires, in most cases, continuous attention to the task by the human operator. The introduction of a simple control system might reduce the direct involvement to a relatively infrequent adjustment procedure for much of the time combined, perhaps, with direct manual control during certain phases of an operation. The introduction of a sophisticated automatic system, however, might reduce the inputs from the HO to no more than two or three switch selections over the course of a complete period of duty. The problems here, in terms of the maintenance of vigilance, are fairly self-evident, and introduce new dimensions into the tasks of operator selection and training.

3.2 Manual Rehearsal

Closely allied to the question of involvement above, is that of the amount of rehearsal of the manual control task which the HO will undergo under normal operating circumstances. This is likely to be extremely low in the case of highly automated systems with the results that manual reversion, resulting from an unexpected system failure, is likely to be performed by an operator who is largely out of practice, and who is obliged to enter suddenly into the direct control loop.
3.3 Control Technology

The introduction of digital control technology brings in its wake several system changes of relevance to the HO. Displays, for example, may tend to contain more digital information, and those preserving a more traditional analogue format may well be constructed using digital technology. Pointers will then move in a series of finite jumps, and the displays will be susceptible to different types of malfunction than will their analogue ancestors.

Perhaps of more consequence to the HO is the comparative ease with which systems may be modified when under software control and when activated via digital data highways. Gone are the nightmares of re-cabling associated with traditional retrofits! There are two ways in which this ease of modification may effect the operator. First, in the relatively long-term, quite substantial changes in display formats, control procedures etc, may be introduced as a result of operating experience. Secondly, in the short-term, data in the system may be updated regularly by means of disks, tapes or other storage media such that the HO is frequently faced with an untested component in the system.

3.4 Failure Diagnosis

The escalation in the number of warning signals associated with complex vehicles has been well documented. (See, e.g., Veitengruber et al 1977). The result of this is that swift and accurate interpretation of the root cause of the problem and its rectification may become extremely difficult tasks. In many cases the HO will require a good deal more diagnostic assistance than that which is currently available if adequate degrees of safety and efficiency are to be maintained.

3.5 Integrity Testing

At various stages of design, development, and certification it is necessary to examine the integrity of a system in terms both of normal operation and under conditions of component failure or system misoperation. Similarly, the testing of system integrity forms part of the regular duties of the system operator. Whilst such a process may be relatively trivial in the case of a simple control system, it may quickly assume very large proportions with sophisticated automation. Much of the testing may itself be automated, but it is important to take due care of the possible human behaviours which may occur, particularly under conditions of stress.

Integrity testing serves as an example of a viable application of a microcomputer simulation of part or whole of an automatic system. This testing may well form part of a larger programme of simulation during system development.
3.6 Mental Models

Both tactical and strategic planning of control behaviour necessitate the use of a model of system performance. Such a mental model is developed both through formal training and through experience of system operation. It is obvious that a viable model of a highly complex system will be relatively difficult to acquire. This basic problem may be exacerbated by an interface between the HO and the remaining parts of the system which, whilst designed to facilitate certain normal control interactions, provides an effective camouflage in front of the operator in the event of unusual conditions. Training captains, for example, may complain that pilots overlook the existence of "a real airplane behind the panel".

3.7 System Language

It is inevitable that complex systems require complex explanations, but many current difficulties could be alleviated. Little attention has yet been paid to the development of languages designed specifically for the needs of system users. All too often, languages are derived from the conceptual framework of system designers, and contain characteristics originating from the history and other irrelevant aspects of the system technology. This is equally true of both vocabulary and syntax. By this means a second interface is created which places a camouflage between the HO and the rest of his system.

4. FUTURE RESEARCH NEEDS

"There is no easy answer to the question, under what conditions should one automate, and how much? Very few empirical studies have been performed to help solve this problem." (Meister, 1976) Whilst the problem may have no single answer, either easy or difficult, it does direct attention to the very large set of issues relating to human control performance in automated systems. The above analysis attempts to identify some of the characteristic features of such automated systems, and thereby to indicate some current research needs. Previous studies have been directed at the same general goal (See, e.g., Boehm-Davis et al 1981).

The following list comprises examples of current outstanding research areas:-

- Selection techniques for HOs as system monitors and managers
- Training techniques for this function
- Effects on performance of low total direct manual control experience
- Effects on performance of low recent direct manual control experience
Transfer effects between systems with differing display formats
Methods of checking system integrity during operations
Design of alerting systems to facilitate diagnosis and appropriate action
Design of user-oriented control system languages
Method of inducing mental models of system operation.

5. REFERENCES


MOTIVATIONAL FACTORS AND PERFORMANCE IN A MANUAL TRACKING TASK

Tom Bösser, Elke Melchior & Martin Schütte
Psychologisches Institut der Universität Münster
Schlaunstr. 2 D-4400 Münster

Abstract: Optimal performance in manual tracking tasks is attained only for short periods of time. We assume performance in long-term manual-control tasks to be constrained by motivation rather than the limitations of the perceptual-motor-system. This study aims at the identification of conditions limiting performance and a detailed description of performance-changes effected. A tracking task judged to be hard by the subjects affected feeling of comfort negatively, but reduced errors in an IQ-independent performance test following tracking. This supports the hypothesis that performance is a vector-valued rather than a scalar quantity. An experimental condition omitting visual feedback was introduced. Results indicate that the use of visual feedback decreases within a session, which may be a strategy to reduce workload.

Traditionally the study of manual control aimed at identifying limiting conditions of human performance, in order to present a basis for specifying requirements of technical systems like high-performance vehicles. These extreme levels of performance are attained, however, only for short periods of time, for example in critical situations during driving an automobile or in the landing-phase of aeroplane flying. Most of the time, performance on tasks of this type is well below the possible maximum. We have shown that a drivers transfer characteristic in car-following can be modified easily by verbal instructions affecting the subjects motivation (BÖSSER 1980). Readiness to switch to high-level performance at all times is most important in order to give the human controller a safety-margin in unusual situations. Constant high vigilance seems to be expensive for the resources available to the human controller, such that optimum performance can not be sustained for long periods of time, due to high cost for the individual. This view implies decision-making of the individual in order to choose between possible alternative actions. We assume the individual to optimize in allocating resources, this assumption, however, offers no explanation, as long as the criteria for optimization are not known. TVERSKY & KAHNEMANN (1981) illustrate that the human is not a rational decision-maker according to the assumption of normative decision-theory.

Because there is no absolute scaling for the cost of human resources in terms of time, effort, energy etc., this scaling has to be done on an empirical basis. This requires the dimensionality of the decision-space to be determined, which correspond to NAVON & GOPHER's (1979) 'resources'. If there is no independent basis for scaling of the axis, and scaling has to be based on preference-data, the properties of the 'value-functions' which TVERSKY & KAHNEMANN (1981) have shown to be required in a model of human decision-making, are implied by this scaling-operation. The human decision-maker is
not the rational decision-maker of normative decision-theory, so that his
decisions may not always correspond to external optimality-criteria. Empirical
investigations must be based on minimal assumptions. Two strategies may be
employed for the study of resource-allocation. NAVON & GOPHER (1979) consider
mainly parallel tasks, but the same arguments apply for sequential tasks and
performance-decrements due to fatigue. In almost all instances human behavior
can be regarded as time-sharing of various tasks. Resources have to be
allocated here such as to optimize performance and minimize cost over the
interval subject to planning. If heavy requirements on task B are expected
in the near future, for example, there might be a tendency to reduce effort
spent on task A. Independent resources may be expected to be deployed
differently in different tasks. This hypotheses is supported by the fact
that diversified sequential scheduling of tasks is assumed to be less tiring
(SCHMIDTKE 1973).

We study the relation between performance-characteristics in tracking-
experiments and motivational constraints. In the experiments reported here
pursuit-tracking and a similar experimental condition omitting feedback were
used. The last condition was introduced in order to indicate the amount of
open-loop-control employed by the subjects.

Apparatus: The subjects were seated alone in a quiet, dimly lit room. Vertical
moving lines on a display-oscilloscope were presented to the subject as
track-input and man-machine-output, a one-axis lowfriction lever served as
input to a zero-order system. All control- and data-collection was done by
digital computer (Fig. 0).

Input-Signal: A pseudo-random signal was generated from a 7-stage, 3-level
shift register at a rate of 10 Hz and a period of 218.6 sec. (DAVIES 1970).
This was passed through a second-order digital filter. Fig. 1 shows the
spectrum of the resulting signal.

Subjects: 26 male psychology students (age 19-27) took part in the experi­
ment in order to fulfill a course-requirement.

Procedure: The subjects were informed of the length of the experiment be­
fore the session. Good performance in this experiment, they were told,
qualified for taking part in well-paid experiments later. Before and after
the session two paper-and-pencil tests were administered: Test d2 (BRICKEN-
KAMP 1967), a version of the Bourdin-Test, measuring performance related to
vigilance, and the ZERSSEN (1975) 'Befindlichkeits-Skala', having high
validity for changes in subjective feeling of comfort.

A session consisted of three tracking-tasks of 10 or 30 minutes each, com­
 pensatory, pursuit and 'no-feedback'- conditions, with an interval of 1
minute.

Data-Analysis: The time series sampled at an interval of 0.1 sec were
analyzed in the frequency-domain with FFT-procedures (BENDAT & PIERSOL 1971).

Results: The task was judged hard and strenuous by all subjects, scores on
the ZERSSEN-Befindlichkeits-Skala increase for all subjects (p < .001),
indicating that the task induces a feeling of discomfort. As expected total
performance on test d2 (number of correct items) increases, reflecting the
effect of practice. The percentage of error, normally not correlated with
total performance, decreases in almost all subjects considerably \((p < .001)\).

Fig. 2 shows the spectra of control-output of two subjects for pursuit-tracking and 'no-feedback'-condition, which were identical, except that in the latter case the system-output was not displayed. The difference in level between spectra for pursuit- and 'no-feedback'-condition reflects the fact that due to the lack of feedback, there is no basis to adjust the amplification. Coherence between input-signal and control outputs (Fig. 3) demonstrates good control-performance up to frequencies of 0.5 Hz, but also a considerable coherence in the 'no-feedback'-condition.

Total RMS-error for pursuit-tracking is significantly lower \((p < 0.02)\) in the experimental sessions \((n = 7)\) lasting 10 min. as compared to sessions lasting 30 min. \((n = 11)\).

While Fig. 3 shows that in the 'no-feedback'-condition some of the variance in the input-signal is reproduced, even without visual feedback, it does not follow that this is a part of the variance common to output of pursuit-tracking and output of 'no-feedback'-experiment. Because the same input-signal was used in both cases, it makes sense to calculate the coherence between the two output-signals, which is shown in Fig. 4 for the same two subjects. It can be seen that proportion of the control-output in pursuit-tracking can be reproduced without visual feedback.

Development of the error during a session of 30 min. in pursuit-tracking and 'no-feedback'-condition is presented in Fig. 5 for one representative subject. Performance in pursuit-tracking does not indicate any effect of fatigue, and the learning-phase must be assumed very brief at the beginning of the session. It seems difficult to attribute the decrement of error in the 'no-feedback'-condition to learning, as the experiment does not offer any feedback for this learning, there must be some way, however, in which the motor-output can be equated to the input-signal without using visual feedback.

The probability-density-distributions in Fig.6 show the range being reproduced incorrectly in the 'no-feedback'-condition, also an asymmetry in the distributions, indicating individual nonlinear strategies.

Discussion of results
A highly disagreeable, difficult and tiring task, which was judged as such by all subjects, would be expected to impair performance in a general way, especially performance on a test designed to measure vigilance. Instead it was shown that in a different task \((d2\)-test) a specific aspect of performance (percentage of error) was affected in an opposite direction. We conclude from this that different tasks do use different resources, as suggested by NAVON & GOPHER (1979).

In contrast to results on tracking-tasks of similar duration, e.g. compensatory-tracking in HAIDER (1977, p. 55) we did not observe a performance-decrement as indicated by the error. This may be due to specific, so far unknown aspects of the experimental situation. Pursuit-tracking may differ from compensatory-tracking in a relevant aspect of the task. The results for the 'no-feedback'-condition suggest that subjects can learn to use
'open-loop' control to a larger extent. This hypothesis has been put forward to account for performance changes in similar motor-control-tasks (Pew 1966), but the hypothesis that open-loop control reduces workload and demand on resources remains speculative. Further investigations will be directed towards a more detailed analysis of nonlinear control-strategies and quantification of the motivational variables involved.


Brickenkamp, R., Test d2, Aufmerksamkeitsbelastungstest, Göttingen 1967


This work was supported by Deutsche Forschungsgemeinschaft (Bo 642/2).
Fig. 1  Power-Spectral-Density of input-signal for all experiments.
Fig. 2 Spectra (Syy) of control-output (lever-position) for pursuit-tracking and no-feedback experiment, subjects 14 and 29.
Fig. 3 Coherence $\gamma^2_{xy}$ input - control-output (lever-position) for pursuit-tracking and no-feedback experiment, subjects 14 and 29.
Fig. 4 Coherence $\gamma^2_{xy}$ between outputs of pursuit-tracking and no-feedback experiment. The same input-signals were used in all experiments.

- subject 14 (10 min.)
- subject 29 (30 min.)
Fig. 5  RMS-error (averaged for 25.6 secs.) in pursuit-tracking and no-feedback experiment, subject 29
Fig. 6  Probability-densities of input-signal and control-output-signals for subject 14 and 29 (pursuit-tracking and 'no-feedback'-experiment)
ON THE STABILITY PROBLEM OF HUMAN ARM AND HAND MOVEMENTS CONTROLLING EXTERNAL LOAD SYSTEMS

W. Setzer, G. Vossius

Institut für Biokybernetik und biomedizinische Technik, Universität Karlsruhe, W. Germany

Abstract

Different loads, controlled by a human operator effect different dynamic behaviour of the entire controlled Arm-Hand-Load (AHL) system. Thus, in terms of stability and optimal control, a human operator requires information about the AHL system to provide the performance of a given task. Additionaly adaptive mechanisms must be present, to determine the properties of the closed loop system. It is shown, that the receptors of the neuromuscular system, muscle spindle and Golgi tendon organs might have the functional role of an internal parallel model to the controlled AHL system. A parallel model of this kind is shown to be feasible to realize identification and compensatory mechanisms involved in the human motor control system.

Introduction

Application of control theory is necessarily based on a minimum of information about the dynamic nature of the controlled system. External loads influence more or less the dynamical properties of the physical Arm-Hand-Load system, such that they produce various control situations for a human operator in a closed man-machine system. Those loads, generally mechanical devices with linear or nonlinear, stable or unstable characteristcs are a priori unknown to the operator. There-
fore identification of the unknown system properties must take place to provide effective arbitrary control. Moreover, experimental results indicate (McRuer, 1980) that the human operator is able to determine the closed loop properties through compensatory muscle control. Both identification and compensatory mechanisms are supposed to be functional properties of the so called "inner loop", where the sensors of the neuromuscular system are mainly muscle spindles and Golgi tendon organs (Granit, 1970).

A simplified Model for AHL

Subsequently, a simplified model of limb and load in operation is described. The limb is characterized by its mass $m_A$ and elastic and frictional components $b_A, a_A$. Elasticity and friction are considered to be determined by the actual set-point of neural muscle-innervation (Jex, 1970; Johannsen, 1977). The external load system is considered by an additional mass $m_F$, spring $a_F$ and friction $b_F$. Since muscle force input $F$ and displacement $x$ are common to limb and load (Fig.1), one obtains:

$$F = (m_A + m_F)\ddot{x} + (b_A + b_F)\dot{x} + (a_A + a_F)x \quad (1)$$

One can see from Eq. 1 that the dynamic properties of AHL system might be determined by the parameters $a_A, b_F$ which are controlled from higher nervous centres through muscle innervation.

The Sensors of the Motor System

The most important sensors of the motorsystem are muscle spindles and Golgi tendon organs. Muscle spindles are located in parallel to extrafusal muscle fibres (Granit, 1970) and respond to static and dynamic stretch of corresponding muscles. An additional motorinput of the muscle spindles
provides a shift of receptor sensitivity with respect to the actual length of the muscle. The spindle output is related to length and changes in length of extrafusal muscle fibres. A linear sensor output \( y \) is assumed:

\[
y_m = ax + bx
\]

(2)

where the parameters \( a \), \( b \) are variable and depend on the motor innervation of the muscle spindle. Golgi tendon organs are sensitive receptors for muscle tension (Houk, 1980). They are found in common tendons of in series muscle fibres. In the simplified model of the AHL system, the tendon organs sense the netforce which is applied to the skeleton and the external load. From Figure 1 one obtains the output signal of the Golgi organs:
\[ y_g = (m_A + m_E) \ddot{x} + b_E \dot{x} + a_E x \quad (3a) \]
\[ = P - b_A \dot{x} - a_A x \quad (3b) \]

A third kind of sensor, joint and skin receptors produce output signals that correspond to the acceleration of the limb (Dickhaus, 1976):

\[ y_b = m^* \ddot{x} \quad (4) \]

Input for all sensors is the statevector of the controlled system \( x=(x,\dot{x},\ddot{x})^T \) and the output an equivalent force vector \( y=(y_g,y_m,y_b)^T \), where the Golgi output includes the unknown system parameters (Eq. 2) and the spindle response is based on the actual value of its own parameters, adjusted by innervation of the muscle spindle.

Fig. 2. Identification with a parallel model
Identification and Compensation

For further considerations the muscle spindle and the joint receptors will be defined as a parallel model to the controlled AHL system (Fig. 2). An error function $e$ can be assumed:

$$
e = y_g - y_m - y_b$$

$$= (m - m^*)x + (b_E - b^*)x + (a_E - a^*)x$$

where $m = m_A + m_E$

The error includes the differences in parameters of system and model. Minimizing a quadratic cost function (Weber, 1971):

$$f(e) = e^2$$

by an adaptive adjustment of the model parameter:

$$m^* = -K e x$$

$$b^* = -K e x$$

$$a^* = -K e x$$

results for $e \to 0$ and joined integration, the unknown parameters $m$, $a_E$, $b_E$.

Compensation is conceivable through the same mechanism. The Minimization of a second cost function where the error is defined by:
\[ e = F - y_g - y_m \]  \hfill (8)  
\[ = (b_A - b^*)x + (a_A - a^*)x \]

will be used to describe compensatory operations. In this case the spindle parameters remain constant while the muscle parameters \( a_A, b_A \) (Eq. 1) are adapted to the \( a^*, b^* \) parameters by additional innervation of extrafusal muscle fibres.

**Experimental Results**

Experiments were performed to prove the existence of a parallel receptor model as outlined above. A manipulator was built where a computer controlled torque motor with additional drive, moves a joystick on a bar. The controller input is the operators handforce, applied on the stick which displaces the device in the direction of force applied. The plant of the manipulator is determined by the software of the computer. Tracking experiments were performed. Sinussoidal inputs were used in a domain of \( 0.2 \ldots 1.5 \text{ Hz} \). The maximum displacement on the

![Graph showing experimental results](image)

**Fig. 3.** Experiments with a linear load system
bar was 10 cm. Both, target and displacement were displayed on a screen.

Results of two typical experiments will be discussed:

1. Tracking experiments with different dynamic manipulator systems, where the state variables of the system were used for the simulation of the parameter identification along Eq. 5-7. First the identification of an unstable linear manipulator system

\[ F_H = m_L \ddot{x} - b_L \dot{x} + a_L x \]  

controlled by the human operator was simulated. During the experiment the parameter \( b_L \) was changed step wise from 0 \( ... \) \( 6 \) Ns/m. The mass of the load system was \( m_L = 2 \) kg and the elastic parameter was \( a_L = .5 \) N/cm. Figure 3 shows a typical record of target \( w \), displacement \( x \), hand force \( F_H \) and the identified parameter \( b_L \). The arrows in Figure 3 indicate the step wise change of \( b_L \). Identification with the proposed scheme is prompt and results in the correct parameter value.

A nonlinear load system was used next:

\[ F_H = 3 \sqrt[3]{x \cdot m_L} \]  

The nonlinearity is a cubic parabola. Identification with the proposed scheme results in parameters related to the magnitude of the input \( F_H \). An example for three different magnitudes \( F_H \) is shown in Fig. 4. Corresponding to the nonlinear characteristic the scheme identifies a big mass while small forces are applied and a small mass when the force is increased. This re-
result was confirmed upon inquiry of the testpersons.

2. Experiments were performed to prove the compensatory abilities of the parallel receptor model. In a first run, unstable manipulator systems were used. The testpersons were able to stabilize the system and moved the joystick in accordance with the tracking task. In a second run the joystick was fixed, and the system, as in the first run, was simulated on the computer. Input was still the applied handforce, the analytically computed output was displayed on the screen.

Fig. 4. Identification of a nonlinear system

The difference of the two runs is that the sensors, muscle spindle and Golgi organ are available for compensation in the first case but not in the second. Only the applied handforce in the fixed stick will be measured by the tendon organs.
In both runs, target and displacement are sensed by the optic system. So far input force and output displacement are still available for compensatory commands in the second run.

Fig. 5 illustrates the most important results comparing the task performance with free moving and fixed joystick. In the first run all load systems, stable, unstable or nonlinear were maintained by the testpersons in accordance to the tracking task (dotted line Fig. 5). In the fixed joystick mode the testpersons were able to control only stable load systems (Fig. 5). Furthermore, oscillations of the load system were suppressed with free a moving stick, but not if the joystick was fixed (Fig. 6).

Conclusion

A new significance of the functional role of muscle spindle, Golgi tendon organs, skin and joint receptors was shown. In terms of system identification and compensation, the receptors
Fig. 6. Supressing of oscillations of the system are assumed to represent a parallel model to the controlled AHL system. Identification is achieved by tuning the models parameter thus, that a cost function becomes minimal. Linear load systems will be identified correctly, nonlinear systems affect model parameters which depend on the magnitude of the input force (Setzer, 1980).

Fixing the joystick, the testpersons were no longer able to perform compensatory tasks. Unstable systems became uncontrollable even if the same system was controlled perfectly with a free moving stick. Hence one concludes that compensation of AHL system through adjustment of muscle parameters in cooperation with a parallel model is more likely than a neuronal control algorithm.

References


This Report is supported by Grant of DFG (Vo 72/22).
EFFECTS OF VISUAL AND VESTIBULAR MOTION PERCEPTION ON CONTROL TASK PERFORMANCE

R.J.A.W. Hosman *)
J.C. van der Vaart *)

Abstract

The influence of visual and vestibular motion perception on pilot's behaviour in a control task has aroused many discussions during the last decades which have not yet come to an end. This influence is of direct relevance to the modelling of pilot control behaviour and to flight simulation.

Results of experiments in this field as reported in the literature appeared to be somewhat different from the experience gained in the research flight simulator of the Department of Aerospace Engineering of the Delft University of Technology. The aim of the experiment described in the present Paper was to obtain a data base on pilot's behaviour using central and peripheral visual and motion cues.

In a following task (or compensatory tracking task) and in a disturbance task (both roll tasks) using a double integrator as the controlled element, all possible combinations of central visual, peripheral visual and vestibular motion cues were presented to the subjects.

The results show significant influence of the peripheral visual and vestibular cues on subject's performance and dynamic behaviour in both control tasks.

1. INTRODUCTION

Theories of control behaviour of a pilot or a human controller are generally based on the concept of a controller as a processor of information. In these theories as well as in quite a number of mathematical models of human control behaviour three main features are usually distinguished: observation, decision making and output generation.

The perception of position and motion by way of visual, vestibular, tactile and proprioceptive cues can be considered as a first stage in the observation process. Research in the field of control behaviour at the Aerospace Department of Delft University is mainly concerned with motion perception.

An example of earlier work are the experiments (reported in Ref. 1) on vestibular thresholds of motion perception. In these tests, subjects were passive and were only required to give verbal responses to selected motion stimuli in a typical flight deck situation.

Of course motion perception by a pilot in a situation where he actively controls an aeroplane is quite different from that of a passive observer. The work reported in the present Paper was undertaken to gather more insight into the motion perception in a control situation.

In this area especially there is a considerable lack of knowledge. Very little is known in fact, about exactly what are the most vital senses for an aircraft control task and on how the information processing takes places.

*) Delft University of Technology, Department of Aerospace Engineering.
Knowledge of all these aspects is essential for formulating the mathematical models that are not only to be tools for aircraft development but also for motion simulation. It is conceivable that much of the present motion versus non-motion controversy in the field of flight simulation partly springs from a general lack of knowledge about the role of motion perception in the control of aircraft. It is obvious that there is a certain degree of overlapping (redundancy) in sensory information. Furthermore it is clear that the amount and quality of sensory information needed depend on the nature and difficulty of the control task, although quantitative data are again either scarce or lacking altogether. An ideal tool in the quest for answers to the questions posed above is a modern moving base flight simulator in which it is possible to simulate a wide variety of system characteristics, to delete or alter different sensory cues and to vary systematically the nature and difficulty of a control task. In this way the information processing task of the pilot can be influenced and studied.

In the literature a number of experiments on the influence of motion in control tasks is reported, see Refs. 2 through 7. The results of these experiments are not always consistent, however. Differences in the reported results can be attributed to differences in the dynamic characteristics of systems to be controlled, in the characteristics of the motion simulation systems and in the experimental set-ups. Experiments on the influence of peripheral visual cues on pilot control behaviour are reported in Refs. 8 through 10. It is reported there that, under certain circumstances, peripheral visual cues can partially substitute for vestibular motion cues. Because no work has been reported covering systematic variation of available sensory cues, it was decided to do an experiment in the Aerospace Department's moving base flight simulator, in which a simple roll control task and the various 'sensory displays' - i.e. central CRT-display, peripheral field display and simulator cockpit motion - were systematically varied. Two different control tasks (see Section 4) were used, the controlled element characteristics being the same in both cases (see Section 3).

Although the basic physical quantity to be controlled in a roll control task is the roll angle, the quantities perceived through the different displays will differ due to the different characteristics of the sensory organs involved and the particular way in which the sensory information is processed. For instance, it is usually assumed that a human controller is able to derive, from a central visual display, some measure of rate of change (roll rate) from the basic roll angle information. A far better impression of roll rate may be obtained, however, from the peripheral visual field, which is known to yield mainly velocity information (Ref. 11). Furthermore, the otoliths in the vestibular system are also sensitive to the simulator roll angle, due to the gravitational influences, while the semicircular canal organs are sensitive to roll acceleration.

2. INSTRUMENTATION AND DATA REDUCTION

All measurements with motion were carried out in the Delft University of Technology Department of Aerospace Engineering flight simulator (see Fig. 1). The three degrees of freedom motion system of this simulator has unique
high fidelity motion characteristics, making the simulator a very suitable tool for the present experiments. The application, in this motion system, of so called 'hydrostatic' bearings in the hydraulic servo actuators results in very smooth and almost rumble free simulator motions, see Ref. 12. Under normal operating conditions motion noise is well below the thresholds of motion perception, as determined by the tests reported in Ref. 1. For the present roll axis control task, a central (foveal) CRT display (simulating an artificial horizon) was installed in the instrument panel in front of the subject's seat in the simulator cockpit, as shown in Fig. 2. Peripheral visual motion cues were provided by two T.V. monitors mounted against the side windows of the simulator cockpit (Figs. 2 and 3). These monitors displayed a moveable checkerboard pattern. The relative positions of the displays and test subject's eye position are given in Fig. 3, the technical details of which are described in Ref. 13.

Subjects used a spring centered side-arm controller to control the dynamics of the system (see Section 3). Details of the side-arm controller can be found in Ref. 14. The dynamics of the controlled system were simulated on a hybrid computer that also generated the quasi-random disturbances acting on the system (see Section 4), as well as the signals controlling the displays and the simulator motion system. The computer algorithms driving the visual displays and the motion system were implemented such that time delays between these systems were smaller than 0.01 second.

During test runs measurements were taken at a rate of 25 per second. In the case of the 'disturbance task' the roll angle \( \varphi \), the roll rate \( \dot{\varphi} \) and the control stick deflection \( \delta_a \) were recorded. In the so-called 'following task' (see Section 4) the roll angle error \( e_{\varphi} \) was recorded in addition to \( \varphi \), \( \dot{\varphi} \) and \( \delta_a \).

Shortly after the end of a test run, data analysis was completed by a digital program yielding standard deviations of the recorded variables. Simultaneously Bode plots of the human operator transfer functions were obtained by using a Fast Fourier Transform (F.F.T.) routine, Ref. 15.

All combinations of display configurations used in the experiment, are shown in Table 1. Due to limited availability of the flight simulator, all non-motion conditions were run in a similar fixed base experimental set-up in an acoustically isolated room. An analysis of variance revealed no significant differences when a number of these non-motion conditions were replicated in the flight simulator.

3. ROLL CONTROL TASK DYNAMICS

The roll control task was chosen because it was felt that any influence of the peripheral displays would be more dramatic in comparison with the other possible modes of the flight simulator i.e. pitch and heave. The dynamics of the controlled system were those of a double integrator having the transfer function

\[
H(s) = \frac{K}{s^2}
\]

(1)

where \( K \) was set to a value of 4.
The dynamics of eq. (1) are roughly similar to the roll control of a slowly responding aircraft, such as a medium to large sized jet transport flying at low speeds. There is, however, a minor difference between the motion to be sensed in an aircraft and in the simulator. When the simulator in the present experimental set-up is made to roll by a control stick deflection, the subject senses, in addition to the rotational roll acceleration, a lateral force component due to the simulator tilt. Due to the particular dynamics of an aircraft and its larger number of degrees of freedom, this lateral force component is virtually absent in actual flight.

4. DISTURBANCE AND FOLLOWING TASK

It is known that human control behaviour is also influenced by the manner in which disturbances act on the controlled loop, see Ref. 6. Therefore two distinct control tasks were used in the present experiments.

In the first one, the disturbance task, the disturbing signal was made to act on the controlled system, as shown in Fig. 4a. In this situation, which is quite comparable to the case in which a pilot stabilizes an aircraft in rough air, the roll angle, or attitude, perceived through the peripheral display by the cockpit motion exactly corresponds to the roll attitude presented on the central display. All 'sensory displays' therefore yield attitude relative to the outside world.

In the second control task, the following, or tracking task, the displayed signal on the central display, $e_\phi$, is the difference between the disturbance signal, i, and the roll angle, $\phi$, of the controlled system, see Fig. 4b. The peripheral display and the motion system, however, correspond with the roll angle $\phi$ of the controlled system. The motion of the system in this case depends on the subject's control action only and there is no direct simple relationship between roll angle error $e_\phi$ as presented by the central display and roll angle and roll rate as presented by the motion system and peripheral displays.

If only one of the controlled variables $\phi$ or $e_\phi$ is presented on the central display, a well trained subject is able to discriminate between a disturbance and a following task, even though the task goal, which is to keep the displayed variable as small as possible, is the same. The addition of peripheral visual and motion cues serves to amplify the difference between the two tasks. In the disturbance task, the task goal can be achieved by keeping the motion of the controlled system as small as possible. This can be achieved by avoiding high roll rates, as would be caused by quick and large control stick deflections. This is contrary to the situation in the following task, where the subject is free to induce large changes in roll attitude and roll rate in order to minimize the displayed error magnitude.

The disturbing signal used in all tasks was a quasi-random one, consisting of the sum of 10 sinusoids whose frequency, amplitude and phase are given in Table 2. The standard deviation of the disturbing signal was $\sigma_i = 1.875$ degrees.
5. SUBJECTS AND TEST PROCEDURE

Three subjects, all university staff members and qualified jet transport pilots, volunteered for the experiments. Extensive training was done until stable performance, as expressed by roll angle or roll angle error standard deviation, was reached.

As the non-motion test runs were performed outside the simulator in a separate experimental set-up, the actual experiments were run in two parts. Within each part, control tasks and display configurations were presented in random order. The duration of a single test run (one particular task under one configuration) was 110 seconds. Measurements were taken only during the last 82 seconds of a run. Five replications were performed, resulting in a total of $4 \times 3 \times 5 = 60$ test runs for the following task and $7 \times 3 \times 5 = 105$ for the disturbance task.

Each series of test runs presented to a subject (6 or 7 runs) lasted approximately 20 to 25 minutes. For the purpose of training alone, 420 test runs were completed among the three subjects before starting the main test program.

6. RESULTS

Two distinct aspects of the control task are considered here: Task performance and control behaviour. Task performance is expressed by the standard deviation of the controlled variables. Control behaviour is assessed by using the computed human controller Bode plots.

Performance

Disturbance task

From Fig. 5 an impression can be obtained of the performance as expressed by the relevant standard deviation $\sigma_{\dot{\varphi}}$ as a function of display configuration. Adding the peripheral displays to the central display (Configuration 2) is seen to have a beneficial effect on the performance of the disturbance task, but the influence of motion is seen to be most dramatic (Configurations 4, 5, 6 and 7). Quite remarkable is the performance for the case of motion alone (Configuration 7). Once motion is present little appears to be gained by adding peripheral displays (Configurations 4 and 5). Addition of the central display in the case of motion (Configurations 5 and 6) gives a small but significant improvement. The standard deviations of the angular rate $\Phi$ and the control output $\delta_{c}$ also demonstrate the considerable influence of motion. In summary it can be observed that addition of the peripheral displays in the disturbance task improves the performance of the subjects just as motion does, the influence of motion begin stronger. No further improvement can be obtained by adding the peripheral displays once motion is present.

Following task

A similar decrease of $\sigma_{\varepsilon_{\varphi}}$ - i.e. improved performance - is seen for the following task due to the addition of either the peripheral displays, motion or both, see Fig. 5. In this task the peripheral visual and motion cues are not in correspondence with the centrally displayed error signal $\varepsilon_{\varphi}$. However, the same trend of decreasing standard deviations is found for both $\varphi$ and $\Phi$ although these variables are not the directly controlled ones, see Fig. 5.
The standard deviations for $\delta$ follow a similar trend. From these data it appears that the influence of the peripheral displays is stronger than in the case of the disturbance task whereas the influence of motion is slightly less.

Analysis of variance on the measured variables of the 105 test runs for the disturbance task and the 60 test runs for the following task are summarized in Table 3. These analyses show that the changes in performance, as expressed by $\sigma_{\delta}$ and $\sigma_{\phi}$, due to changes of the display configurations, are significant for both tasks. This also holds for the standard deviations of $\dot{\phi}$ and $\delta$ for the disturbance task and $\phi$, $\dot{\phi}$ and $\delta$ for the following task. In addition a significant influence of the subjects and the interaction between subjects and configurations is demonstrated for these variables. This indicates that subjects, while obtaining approximately equal performance, used different control strategies and reacted differently on the changes of the display configurations.

The mean effects of adding peripheral displays or motion can be summarized by the following relative decreases in standard deviations of the controlled variables $\sigma_{\phi}$ and $\sigma_{\phi}$.

<table>
<thead>
<tr>
<th>Task</th>
<th>Peripheral display</th>
<th>Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disturbance task</td>
<td>12%</td>
<td>59%</td>
</tr>
<tr>
<td>Following task</td>
<td>16%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Control behaviour

Bode plots of the transfer function $H(\omega)$ relating the subject's input $\phi$ or $e^\phi$ to the subject's output $\delta$ were calculated for all combinations of display configurations and control tasks tested.

Disturbance task

The bode plots of the transfer function $H(\omega)$ for the disturbance task are presented in Figs. 6 and 7. Due to the addition of peripheral visual and motion cues the modulus of the transfer function is seen to increase at low frequencies. As could be expected from the performance data, the influence of motion on the transfer function is the strongest. Of all the configurations without central display (Configurations 3, 6 and 7), Configuration 3 (peripheral display only) shows a decrease of the modulus at the low frequencies. This can be explained by the fact that in this configuration, subjects could hardly derive any roll attitude information from these displays, especially at the low frequencies. In the Configurations 6 and 7 however, the subjects can perceive the side force due to the bank angle $\phi$. From these data it follows that the side force is a good substitute for the central visual display. In Fig. 8 the crossover frequency and the phase margin $\phi_m$ have been plotted as a function of the seven configurations. As could be expected, the crossover frequency is increased when the peripheral
displays and motion are added to the central display. The phase margin remains approximately constant.

**Following task**

For the following task, see Fig. 9 the changes due to addition of peripheral visual and motion cues are opposite to the ones in the disturbance task. The modulus decreases especially at the low frequencies, the phase angle increases at low frequencies and decreases at high frequencies. In Fig. 8 the crossover frequency $\omega_c$ and the phase margin $\phi_m$ have been presented as a function of the four display configurations. In this case $\omega_c$ is hardly influenced by the addition of the peripheral displays or motion. However, the phase margin is seen to increase. Finally, Figs. 10 and 11 may serve to stress the differences in the changes of the subject's control behaviour. In these Figures the open loop transfer functions $H_p(\omega)H_o(\omega)$ for two display configurations are plotted for both the disturbance and the following task.

Summarizing the results concerning the control behaviour in both tasks it can be concluded that the performance improvement due to the addition of peripheral displays and/or motion, coincides with changes in the subject's transfer function in both tasks. For the disturbance task the performance improvement can easily be explained by the increase of $\omega_c$ at nearly constant $\phi_m$. For the following task, however, the performance improvement is seen to be accompanied by an increase in phase margin, the crossover frequency remaining nearly constant.

**7. DISCUSSION AND CONCLUSIONS**

That motion as well as peripheral visual cues should, in general, have a considerable influence on subject's control performance and control behaviour could be expected considering the results reported in the literature (Refs. 2 through 10). In the present experiment, however, the disturbance signal was small, resulting in very low values of the standard deviation of the roll angle and roll angle rate ($\sigma_\phi = 1-3$ degrees, $\sigma_\dot{\phi} = 1.5 - 5$ degrees/sec). In spite of these low values, a considerable influence, especially of motion, was found on performance and control behaviour.

Ref. 6 describes disturbance and following tasks using configurations similar to no's 1 and 4 (central display only and central display with motion). The results are quite comparable to the present ones: a performance improvement and a considerable increase of the crossover frequency for the disturbance task. The following task of Ref. 6 showed a slight improvement of the performance together with a large increase of the phase lead at low frequencies due to the addition of motion.

Another experiment with a following task only, see Ref. 9, showed the same trend as the present one although the controlled system of Ref. 9 was a much more difficult one to control.

The remarkable difference between the changes of control behaviour for both control tasks brought about by the addition of peripheral visual and motion cues can probably be explained in terms of a difference in the subjective cost function that the subjects tried to minimize. If the subject tries to maintain the roll angle $\phi$ on the central display in the disturbance task as
small as possible, the peripheral displayed roll rate $\dot{\phi}$ and the simulator motion will also be small. In the case of the following task, however, relatively large roll angles and roll rates will occur, when the subject minimizes the centrally displayed roll angle error. From the present experiment it turns out that for the following task, the subjects developed a control strategy that resulted not only in a decrease in roll angle error $e^p$ but also in an accurate control of the roll angle $\phi$ and roll angle rate $\dot{\phi}$, if peripheral visual and motion cues were present. Apparently, subjects tended to keep the roll angle and roll angle rate at relatively low values. This means that they somehow included these variables in their subjective cost function.

Summarizing the results of the present experiments it can be concluded that:

1. Performance is improved significantly in both disturbance and following task by adding peripheral visual cues and/or motion cues.

2. Control behaviour as expressed by human operator transfer functions is influenced by adding peripheral visual cues and/or motion cues in both disturbance and following task.

3. In the disturbance task the increase in performance due to the addition of peripheral visual and/or motion cues is readily interpreted by the increase in the crossover frequency.

4. In the following or tracking task the influence of peripheral and/or motion cues on the human operator transfer function shows a trend opposite to the one in the disturbance task. As a consequence improvements in performance cannot readily be interpreted in terms of human operator transfer functions.

Although improvements in performance and changes in dynamic behaviour were definitely demonstrated as a result of the addition of peripheral visual cues and motion cues, it is not clear what exactly are the causes for these improvements and changes.

Motion perception may have been improved in either of two ways. Firstly it may be that by addition of peripheral visual and/or motion cues, redundant information is made available to the subject thus improving the accuracy of the motion perception process. Another possibility is that due to differences in the dynamic characteristics of the vestibular system and the peripheral visual system on the one hand, and those of the foveal visual system on the other, a subject receives additional information that enables him to improve motion perception.

Further research into the motion perception process in particular into the separate aspects of central visual, peripheral visual and vestibular motion perception and their interactions is called for.

8. REFERENCES


Table 1: Display configurations

<table>
<thead>
<tr>
<th>Configuration no/code</th>
<th>Central display</th>
<th>Peripheral display</th>
<th>Motion task</th>
<th>Disturbance Following</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 C</td>
<td>x</td>
<td>x</td>
<td></td>
<td>D, F</td>
</tr>
<tr>
<td>2 CP</td>
<td>x</td>
<td>x</td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>3 P</td>
<td></td>
<td></td>
<td></td>
<td>D, F</td>
</tr>
<tr>
<td>4 CM</td>
<td>x</td>
<td></td>
<td>x</td>
<td>D, F</td>
</tr>
<tr>
<td>5 CPM</td>
<td>x</td>
<td>x</td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>6 PM</td>
<td></td>
<td></td>
<td></td>
<td>D</td>
</tr>
<tr>
<td>7 H</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Frequency, amplitude and phase of the sinusoids used to generate the quasi-random disturbance signal

<table>
<thead>
<tr>
<th>Frequency ω (rad/sec)</th>
<th>Amplitude (degrees)</th>
<th>Phase (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.153</td>
<td>1.106</td>
<td>4</td>
</tr>
<tr>
<td>0.230</td>
<td>1.099</td>
<td>151</td>
</tr>
<tr>
<td>0.383</td>
<td>1.083</td>
<td>43</td>
</tr>
<tr>
<td>0.537</td>
<td>0.957</td>
<td>122</td>
</tr>
<tr>
<td>0.997</td>
<td>0.957</td>
<td>324</td>
</tr>
<tr>
<td>1.457</td>
<td>0.862</td>
<td>184</td>
</tr>
<tr>
<td>2.378</td>
<td>0.646</td>
<td>281</td>
</tr>
<tr>
<td>4.065</td>
<td>0.428</td>
<td>194</td>
</tr>
<tr>
<td>7.440</td>
<td>0.247</td>
<td>162</td>
</tr>
<tr>
<td>13.576</td>
<td>0.136</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3: Results of the analysis of variance on the standard deviation of the measured variables

**Disturbance task**

<table>
<thead>
<tr>
<th></th>
<th>σ_φ</th>
<th>σ_σ</th>
<th>σ_σ_e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Configurations</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>2 Subjects</td>
<td></td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>3 Interaction subjects-configurations</td>
<td></td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>4 Replications</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

**Following task**

<table>
<thead>
<tr>
<th></th>
<th>σ_φ</th>
<th>σ_φ</th>
<th>σ_σ</th>
<th>σ_σ_e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Configurations</td>
<td>****</td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>2 Subjects</td>
<td></td>
<td>****</td>
<td>****</td>
<td>****</td>
</tr>
<tr>
<td>3 Interaction subjects-configurations</td>
<td></td>
<td></td>
<td>**</td>
<td>****</td>
</tr>
<tr>
<td>4 Replications</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td>****</td>
</tr>
</tbody>
</table>

α < 0.01 ****
α < 0.05 ***
α < 0.1 **
α < 0.25 *
Fig. 1: The flight simulator cab with peripheral displays.
Fig. 2: Simulator cockpit with central C.R.T. display, peripheral display and side arm controller.
Fig. 3. Positions of displays relative to the test subject's eye position. Central display image.
Fig. 4: Block diagram of controller and controlled element for the disturbance task and the following task.
Disturbance task

Following task

Display configurations

Fig. 5: Standard deviations of controlled variables and control deflections as a function of display configuration.
Fig. 6: Bode plots of the transfer function $H_p(\omega)$ of the test subjects in the disturbance task.
Display configurations 1, 2, 4, and 5.
Fig. 7: Bode plots of the transfer function $H_p(\omega)$ of the test subjects in the disturbance task. Display configurations 1, 3, 6 and 7.
Disturbance task

Following task

Display configurations

Fig. 8: Phase margins and crossover frequencies of $H_p(s) \cdot H_c(s)$ as a function of display configuration.
Fig. 9: Bode plots of the transfer function $H_p(\omega)$ of the test subjects in the following task.
Fig. 10: Typical examples of measured transfer characteristics
or controller and controlled element $H_P(\omega).H_C(\omega)$.
Display configuration 1 (C).
Fig. 11: Typical examples of measured transfer characteristics of controller and controlled element, $H_p(\omega).H_c(\omega)$. Display configuration 5 (CPM).
Acknowledgement
Thanks to the kind permission Prof. Towill obtained from the Institution of Electronic and Radio Engineers we could reproduce the paper as printed below, which appeared before in "The Radio and Electronic Engineer", Vol. 50, No. 9, pp 447-458, September 1980.

Man-machine interaction in aerospace control systems


Based on a paper presented at an IERE Colloquium on Design of Control Systems for Advanced Aircraft having Relaxed or Negative Stability held in London on 16th October 1979.

SUMMARY
The Multi-Role Combat Aircraft Tornado is probably the most spectacular example to date of a complex man-machine system in which design and manufacture are widely dispersed. Even sub-assemblies of this system are extremely complex relative to the technology of just a few years ago. This makes life extremely difficult for an engineer involved in the procurement of the sub-assembly, especially as the overall objective can soon be submerged in a mass of detail.

Starting with a review of desirable dynamic properties of the machine, as seen by the man, typical human operator transfer functions which have found use in closed-loop control system design are presented. Some ways in which microelectronics can reduce pilot workload are then discussed with the object of conveying to the design engineer some of the reasoning behind the performance specification he is trying to meet. Examples include helicopter stability augmentation systems and spacecraft terminal phase control.

1 Introduction
Optimizing the performance of man-machine systems is an inter-disciplinary activity requiring contributions from behavioural scientists, plant designers ('plant' is the generic term used to describe the object being controlled), servomechanism experts, electronic engineers, and systems specialists. Thus the development of a complex man-machine system, such as are met in aerospace applications, requires a large team of engineers and scientists in split locations and separate companies. The Tornado aircraft is the most obvious example.

Each group within the team will work towards an itemized performance specification appropriate to the individual hardware or software product being manufactured. Every such product will, in general, be well engineered through the customary evolutionary process which is the hallmark of good practice, and which represents the commercial expertise of industrial companies.

However, the bringing together of all these products and skills to optimize the performance of the man-machine system, i.e. the 'systems engineering' aspect is less well understood. Thus it is not difficult in times of galloping technological obsolescence for designers of individual products to lose sight of the overall system objectives. This in turn will render less effective the highly desirable interchange of ideas between design groups within the team especially with regard to performance specification where the human operator is acting within a closed loop system.

Even worse would be a situation in which a potential hazard went unrecognized. This is particularly true with the increasing use of digital microelectronics, where it is now being recognized that software reliability must be improved. Until we are satisfied in this respect, we should make haste slowly in replacing operationally satisfactory analogue equipment.

2 Contribution of the Present Paper
The purpose of this paper is to bring together the constituent components of man-machine systems in such a way as to provide a firm foundation for the 'systems engineering' aspect of design. As a consequence, the meaning of performance specifications commonly used for each sub-system will become apparent.

In particular, it is intended to highlight the increasing contribution of microelectronics to man-machine system design. For example, Table 1 illustrates the volumetric compression in commercially-developed aircraft flight control system (a.f.c.s.) electronics achieved over a fifteen-year span. It can be seen that volume needed is reduced by approximately 50% every two years. Additionally, of course, it may be argued that the improvement is even more spectacular, because present
Microminiaturization of the automatic flight control system (a.f.c.s.) due to advances in electronic technology

<table>
<thead>
<tr>
<th>Year</th>
<th>Vehicle</th>
<th>Approximate volume (cm³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>U.S. Navy A6 aircraft autopilot</td>
<td>950</td>
</tr>
<tr>
<td>1962</td>
<td>Missile autopilot</td>
<td>290</td>
</tr>
<tr>
<td>1968</td>
<td>Boeing 747 autopilot</td>
<td>40</td>
</tr>
<tr>
<td>1973</td>
<td>U.S.A.F. B1 aircraft autopilot</td>
<td>10</td>
</tr>
</tbody>
</table>

day a.f.c.s. are capable of performing a wider range of functions than earlier models.

Wherever possible, the usual systems engineering tools such as block diagrams, flow diagrams, and transfer function techniques will be used. Although most of the readily available published literature on man-machine systems deals with aerospace applications, the subject is of much wider relevance. So we find the same principles applied to skill development, shipsteering, land vehicle steering, and 'bumpless' train shunting. The opportunity will also be taken to show how microelectronics can be used to reduce human Operator workload, via such techniques as plant augmentation, display quickening, providing a model reference describing the current plant dynamics, ‘observers’ for state estimation, and fast models for prediction in terminal-phase docking operations.

3 Man-Machine System Concepts

Much of early work on the development of the ‘servomechanism’ approach to modelling human Operator performance in man-machine systems was undertaken during, and shortly after World War II.

The result is the block diagram shown in Fig. 1 in which the human operator is embedded into a closed-loop system, and reacting to visual, vestibular (perceived motions, such as via the ‘seat-of-pants’ feeling), and other inputs.

One example of a vestibular sensor transfer function is the horizontal pair of semi-circular canals located in the ear which permit the human operator to perceive angular velocity in a blind situation. A suitable model is

\[
\frac{\text{sensor output}}{\text{input angular velocity}} = \frac{10s}{(1 + 10s)(1 + s/10)}
\]

From this sensor output it is clear that the human operator has an essentially flat response in sensing angular motions in the range 0.10 to 10 rad/s.

4 Display Configurations

The display shown in Fig. 1 can be the perceived visual field, or a 'blind' display. In the former case, the choice of display is made by the operator, whereas in the latter case it is at the discretion of the system designer. As we shall see later, it is possible to relate the display chosen by the human operator to skill development in task performance.

We may describe the operation of the systems shown in Fig. 2 in the following ways:

‘Compensation’

In compensatory tracking the visually displayed effects of the human operator are not distinguishable from the system output. So the operator can only determine the effect of control motion if the input conditions are zero.

‘Pursuit’

Past experience provides the operator with information about what to expect in a future input, but he must operate in a closed loop fashion with feedback about his responses. The operator can distinguish the effect of his
actions from changes in input stimuli.

'Pre-Cognitive'
The operator has effectively complete information about the input's future. A stimulus can then trigger off a 'dictionary' of practised, properly sequenced responses on the part of the operator. It is essentially 'open-loop' control, i.e. no feedback.

5 Skill Development Viewed as a Choice of Display
For the well-known example of learning to drive an automobile, skill development may be explained via the sequencing of perceived displays in the following way. Phase 1. Novice starts by relating fixed object on car to a guideline on the road, so may aim to keep a constant angle between a point on the body and the curb. This is compensatory as the operator attends to error only.

Phase 2. Operator becomes aware of the separate characteristics of the curb and the car. He achieves a pursuit display organization of his visual field and hence makes use of the regularities and predictable features of the road.

Phase 3. On achieving complete familiarity with the 'plant', the operator samples entire visual field, and steers vehicle with deft, discrete, movements. This is 'open-loop', or pre-cognitive behaviour in which the operator has thoroughly learned his inputs, and by experience seems to provide exactly the right output. Examples are a helmsman meeting a turning ship, and a driver recovering from a skid. The time interval between these short-term bursts of 'open-loop' control depends on environmental conditions.

In apparent contradiction, we note that most published research into man-machine system performance utilizing the servomechanism approach is based on compensatory control. This has been found to be the most relevant configuration for the control of complex multi-loop systems, especially as met in aerospace applications. Examples are instrument-aided landing systems, and terrain following via head-up displays.

6 Human Operator Variability
Unlike a servomechanism, in which the response to a given stimulus is generally repeatable, human operator response will not always be the same, although in a well-designed system the skilled operator will perform with reasonable consistency. Figure 3 shows a set of step responses obtained during simulator studies with the human operator as part of an eighth-order closed-loop system. Although there is some variability in response, it does not depart significantly from the idealized third-order Butterworth response. This was chosen by the system designer so that little or no overshoot would result, which is thought to be a highly desirable characteristic in terrain-following systems.

The design concept used is of interest. Using a simple model \( (2s/5) \) to represent the human operator inside a closed-loop control system, there are then five system zeros and eight system poles, so that we may write

\[
\frac{h(t)}{h_D(t)} = \left[ 1 + \sum_{i=5}^{8} b_is^i \right] \left[ 1 + \sum_{i=8}^{16} a_is^i \right] (2)
\]

If the system is now compensated such that the denominator can be factorized into

\[
1 + \sum_{i=8}^{16} a_is^i = \left[ 1 + \sum_{i=8}^{16} b_is^i \right] \times (1 + 2(s/\omega_0) + 2(s/\omega_0)^2 + (s/\omega_0)^3) (3)
\]

then five system poles will cancel out the five system zeros, leaving the desired third-order Butterworth design to describe the system response to command signals. The \( \omega_0 \) parameter is set by the designer, based on the desired speed of response. For 'optimum' performance to be achieved in practice, \( \omega_0 \) must be compatible with human operator capability. From the results shown in Fig. 3, the value of \( \omega_0 \) chosen is clearly satisfactory.

7 'Quickening' the Display
The cancellation design is also of interest from two further points of view. Firstly, the design is mechanized by a technique known as 'quickening', as shown in Fig. 4. The sensor and synthesized data which are feedback are mixed prior to the display. The human operator therefore reacts to a single stimulus, just as he would in an unquickened system, except that the signal displayed is the 'optimum' needed to achieve the desired closed-loop performance. In contrast, where plant augmentation is used, as will be described later, similar signals are feedback, not to the operator display, but are coupled to the operator output to form one or more minor loops around the plant.

The second point of interest is that to achieve the desired denominator in equation (3), eight independent signals must be available within the system so as to
shape the $a_i$ coefficients to the requisite values. In practice, the designer has to cope with the situation in which it is unlikely that sufficient sensor generated signals are available, nor are they guaranteed to be independent of each other. To overcome this difficulty, the displayed feedback signal is synthesized from those signals which are available. For reasons of noise filtering, integration techniques are to be preferred to those requiring explicit differentiation.

The particular design technique used to achieve the desired feedback shaping has since become known as the 'H-equivalent' method.\textsuperscript{18} It implies the design and implementation of 'observers' to estimate inaccessible signals. This is an important application of electronics within man-machine systems, since either analogue or digital hardware may be used to generate the estimates. A simple example will be detailed in the section on stability augmentation systems (s.a.s.). The effective damping ratio of the s.a.s. will be increased by estimating pitch angular velocity from pitch angle, rather than incorporating a rate gyroscope to measure velocity directly.

8 Desirable Plant Transfer Functions
A standard method of improving man-machine system performance is to 'shape' the plant transfer function under control so that the human operator is better able to cope. In order to achieve this aim, we must have available 'preferred' transfer functions to which the plant should be tailored. Quite apart from the improved performance to be expected when parameters are at nominal value, correct feedback of sensor signals will reduce the effects of parameter changes,\textsuperscript{19} thus leading to improvement throughout the operating envelope.

The best-known set of target transfer functions are probably those derived for the short-period pitch motion of fixed wing aircraft, and reproduced in closed contour form in Fig. 5.\textsuperscript{20, 21} The axes are respectively $\omega_n$ (undamped natural pulsation) and $\zeta$ (damping ratio) for the quadratic lag due to the aircraft which appears in the transfer function relating stick deflection to pitch rate.\textsuperscript{22} These contours are based on pilot opinion, which in turn is related to the Cooper rating scale of Table 2.

The Cooper scale was derived in 1957\textsuperscript{8} in order that pilot ratings might attempt to define quality relative to mission and intended use. In practice it is found that using this scale pilots are able to discriminate adequately enough and repeatably for the system designer to have confidence in these opinions guiding him to an optimum solution.

9 Plant Augmentation via Instrumentation
Historically, plant augmentation has been achieved wherever possible by directly sensing 'states', such as

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|l|l|}
\hline
Adjective rating & Numerical rating & Description & Primary mission accomplished & Can be landed \\
\hline
Normal operation & Satisfactory & 1 & Excellent, includes optimum & Yes & Yes \\
& & 2 & Good, pleasant to fly & Yes & Yes \\
& & 3 & Satisfactory, but with some mildly unpleasant characteristics & Yes & Yes \\
Emergency operation & Unsatisfactory & 4 & Acceptable, but with unpleasant characteristics & Yes & Yes \\
& & 5 & Unacceptable for normal operation & Doubtful & Yes \\
& & 6 & Acceptable for emergency condition only\textsuperscript{+} & Doubtful & Yes \\
No operation & Unsatisfactory & 7 & Unacceptable even for emergency condition\textsuperscript{+} & No & Doubtful \\
& & 8 & Unacceptable—dangerous & No & No \\
& & 9 & Unacceptable—uncontrollable & No & No \\
Unprintable & & 10 & 'Motions possibly violent enough to prevent pilot escape' & & \\
\hline
\end{tabular}
\caption{Original Cooper pilot rating scale\textsuperscript{8}}
\end{table}

\textsuperscript{+} Failure of a stability augmenter.
velocity, and then feeding back to a summing junction a 'mixed' signal combining all the information necessary to shape the plant transfer function. This is analogous to the case of a position servomechanism in which artificial damping is provided by tachometer feedback, although for engineering reasons this is usually feedback separately to form a rate loop within the main position loop.

Figure 6 shows the principle applied to an aircraft stability augmentation system. By suitable choice of gains within the forward and return paths of the system, the augmented plant transfer function can be shaped to include a short period quadratic lag with acceptable values of \( \zeta_s \) and \( \omega_n \). This is not to imply that the plant is necessarily second order. For example, fixed wing aircraft exhibit a long period (phugoid) oscillation in the pitch plane. Typically this can have a period of 50 seconds, and can be divergent, with \( \zeta_n \) negative. However, it is found that pilots can cope with a mildly divergent phugoid mode provided the short period mode has satisfactory characteristics. So for stability augmentation systems, the designer will usually concentrate on short period mode compensation.

10 Plant Augmentation via State Estimation

Let us now consider how plant augmentation can be achieved when all desirable sensor signals are not available. Modern control theory refers to this problem as 'state estimation', and embraces two very different approaches. It is commonly linked with the Kalman filter, but it is also known as the 'observer' problem, associated with Luenburger, in which we seek to reconstruct missing signals using only integrators. However, modern control theory simply provides a rational solution to this problem, rather than opening up new possibilities for system design. Thus it is found that for many years 'classical' trained control system designers have been constructing observers to estimate inaccessible signals needed in configuring a satisfactory system, admittedly on an ad hoc basis.

In terms of the evolution of system design techniques, the \( H \)-equivalent method is a useful classical technique, which integrates well with existing methods, and is worthy of better publicity. However, for the purpose of this paper we shall use a pragmatic approach in which we concentrate on devising integral-based compensation to yield the desired equivalent signal at a particular point in the loop. An algorithm which generalizes this approach is available.

Figure 7(a) shows a simple second-order system in which the basic plant is a double integrator. Velocity feedback is essential to achieve adequate damping. The augmented plant transfer function is thus

\[
\frac{\theta_0}{\theta_1}(s) = \frac{1}{1 + T_s s + s^2/K}. \tag{4}
\]

If we are prohibited from measuring velocity directly, then we require an 'observer', which supplied with the only available signals, \( \theta_0 \) and \( u \), will generate the desired feedback signal \([1 + T_s \theta_0]\), so that equivalence is maintained. Figure 7(b) shows how this will be achieved. Both \( \theta_0 \) and \( u \) are passed through filters with the same denominator, which is set by the designer. The filter numerators are different, and must be related to the needs of estimating velocity.

Adding the observer will result in the augmented plant now being of third order, but reducing to second order by choosing the parameters such that pole-zero cancellation is possible. Thus we aim to achieve

\[
\left[ \frac{\theta_0}{\theta_1} \right](s) = \frac{1 + T_s}{(1 + T_s)(1 + T_s s^2/K)}. \tag{5}
\]

From Fig. 7(b), the augmented plant transfer function denominator in terms of the observer parameters is
Equating coefficients in (5) and (6) yields the solution

\[ \gamma_0 = a_0 = 1 \]
\[ \gamma_1 = a_1 = (T_r + T) \]
\[ \beta_0 = T T_r K \]
\[ \beta_1 = 0. \]

The 'observer' is summarized in block diagram form in Fig. 8(a). To avoid the use of differentiators, standard analogue instrumentation may be used, as shown in Fig. 8(b), in which the device is clearly recognized as a 'computer'. The aerospace industry has traditionally referred to such devices as 'computers', thereby highlighting present and future microprocessor applications. For example, this observer could be readily mechanized from a basic low-pass digital filter design based on the Motorola M6800.28

It should be noted that the 'observer' augmented system is only the exact equivalent to the rate-feedback compensated system for the case where all system parameters are at their nominal values. So if sufficient sensor data can be made available of the right quality at the right price, it is preferable to use the instrumentation augmented system. However, an additional useful application of the observer principle is the condition monitoring of sensors in order to increase system integrity.

11 Using an Adaptive Model Display

If the augmented plant transfer function still varies significantly over the flight envelope despite the best use of feedback compensation, microelectronic technology can be used to reduce pilot workload by continuously evaluating and updating the plant dynamics as shown in Fig. 9. Environmental data, and various inputs and outputs are monitored and fed into an on-board mini-computer. Once the transfer function has been evaluated, the characteristics can be displayed using whatever mode is thought to be most appropriate. In Fig. 9, the step response has been estimated, since this gives the pilot immediate visual cues on current vehicle damping ratio and response time, information which can assist him in taking the most effective control actions.

In this instance the mini-computer is performing the task of identification. Many techniques exist, the choice resting on the signal-to-noise ratio, non-linearities present, and the speed with which updating is required. It is also helpful, especially during quiescent periods of operation, to inject special command signals into the vehicle so as to provide adequate responses for mathematical analysis by the micro-computer. A short pulse pattern of zero mean may be acceptable, and the dynamics established via the fast Fourier transform. Otherwise a low-level p.n.s. (pseudo-noise sequence) may be generated, and the vehicle cross-correlation function estimated. Under certain circumstances the cross-correlation function approximates to the vehicle impulse response. By integration with respect to delay time, the vehicle step response may then be estimated and displayed without the need for violent manoeuvres.

It has been suggested that updating the vehicle dynamic model in the manner shown in Fig. 9 can be of considerable assistance to the pilot in detecting, and compensating for, a failure in s.a.s. The extent to which it is of help depends on the rate at which the failure occurs, and the time taken by the adaptive loop to update the display sufficiently for the pilot to identify the cause. Depending on circumstances, the pilot will have somewhere between two and ten seconds to take appropriate action, so that useful information must be provided within this timescale. Once the cause has been
identified by the pilot the action taken is likely to be automatic and in accordance with pre-conditioning attained during training. The adaptive loop will, however, help to reduce the experimentation time with the controls needed by the pilot to identify the cause of failure.

12 Human Operator Quasi-linear Models for Closed-loop Control

In previously referring to the eighth-order system designed by Tipton, mention was made of the approximate transfer function \((2.5/s)\) to describe the human operator behaviour within the closed loop. Much research has been undertaken on the derivation of such models. Figure 10 is a block diagram explanation of how the model is derived from experimental data. The plant is assumed to be a linear transfer function \(Y_c(s)\). Using offline computation, the best-fit linear transfer function is determined which minimizes the mean-square error between actual pilot response and model response. To emphasize the unavoidable curve-fit 'remnant', of Fig. 10, the model is sometimes referred to as a describing function, rather than the normal transfer function.

The generalized human operator transfer function (h.o.t.f.) relevant to one and two-dimensional compensatory control tasks is

\[
Y_{Ho}(s) = \frac{K_{Ho}(1 + T_L s)e^{-r\phi}}{(1 + T_I s)} \tag{7}
\]

where \(\tau\) is the effective time delay including transport lag and high frequency neuromuscular lags, \(K_{Ho}\) is the operator gain, and \((1 + T_L s)/(1 + T_I s)\) is the operator's equalization characteristic. \(K_{Ho}, T_I,\) and \(T_L\) apparently can be adjusted by the operator, and to some extent the variation can be used to explain differences between conflicting published transfer functions.

Some results are given in Table 3 which relates the h.o.t.f. to the plant dynamics in the region of the loop cross-over frequency. The operator appears to select the most suitable form of equation (7) for the task. So lag-lead, pure lead, pure lag, or pure gain may be chosen for the equalization characteristic as needed to achieve good low frequency response and an adequate phase margin (typically between 60° and 110°).

As a consequence, irrespective of the detailed high- and low-frequency dynamics, if the man-machine system is controllable, the compensated loop transfer function can be approximated in the cross-over frequency region by the simple transfer function

\[
Y_L(j\omega) = Y_p(j\omega)Y_{Ho}(j\omega)
\]

where

\[
Y_L(j\omega) = \left[ \frac{\omega_c e^{-j\phi_m}}{j\omega} \right]
\]

because the human operator makes the necessary adjustment. Typical values for a study in which the plant transfer function is a double integrator and the command signal covers the range 0 to 2.5 rad/s are:

- \(\phi_m = 35°\)
- \(\tau_c = 0.267\ s\).

<table>
<thead>
<tr>
<th>Approximate vehicle transfer function in region of crossover frequency, (\omega_c)</th>
<th>Equalizer type chosen by pilot from ([(1 + T_L s)/(1 + T_I s)])</th>
<th>Resulting approximate pilot transfer function</th>
<th>Location of equalizer break frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K)</td>
<td>Lag-lead</td>
<td>(\frac{K_{Ho}e^{-r\phi}}{1 + T_I s})</td>
<td>(\frac{1}{T_I} \leq \omega_c)</td>
</tr>
<tr>
<td>(\frac{K}{s})</td>
<td>High-frequency lead</td>
<td>(K_{Ho}e^{-\phi})</td>
<td>—</td>
</tr>
<tr>
<td>(\frac{K}{s^2})</td>
<td>Low-frequency lead</td>
<td>(K_{Ho}(1 + T_L s)e^{-\phi})</td>
<td>(\frac{1}{T_L} \leq \omega_c)</td>
</tr>
<tr>
<td>(\frac{K}{s(1 + T_L s)})</td>
<td>If (T &gt; \tau), use mid-frequency lead</td>
<td>(K_{Ho}(1 + T_L s)e^{-\phi})</td>
<td>(\frac{1}{T_L} \approx \frac{1}{T})</td>
</tr>
<tr>
<td>(\frac{K}{1 + 2\xi(\frac{s}{\omega_c}) + \xi^2(\frac{s}{\omega_p})^2})</td>
<td></td>
<td>(K_{Ho}e^{-\phi})</td>
<td>—</td>
</tr>
<tr>
<td>(\frac{K}{1 + 2\xi(\frac{s}{\omega_c}) + \xi^2(\frac{s}{\omega_p})^2})</td>
<td>If (\omega_c &lt; 1/\tau), use low-frequency lead</td>
<td>(K_{Ho}(1 + T_L s)e^{-\phi})</td>
<td>(\frac{1}{T_L} \leq \omega_c)</td>
</tr>
<tr>
<td></td>
<td>If (\omega_c &gt; 1/\tau), use lag-lead</td>
<td>(K_{Ho}e^{-\phi})</td>
<td>(\frac{1}{T_L} \leq \omega_c)</td>
</tr>
</tbody>
</table>
Note that $\tau_c$ allows for 90 lag from one of the integrators, the neuromuscular and reaction time delays, plus the lead developed from the equalization characteristic, giving 55 total lag at crossover frequency. The relatively low phase margin indicates that this is not an easy task for the operator, who, as shown in Table 3 is expected to provide the limit of his lead capability.

13 The Significance of the 'Remnant'

In Fig. 10 the 'remnant' is conceptually shown as the curve fit error between the output from the best linear transfer function which can be obtained by curve fitting, and the actual output from the human operator, in the mean square sense. The actual transfer function needed to minimize the error is itself a measure of efficiency of design of the man-machine system. In general, the simpler the transfer function needed to yield an adequate curve fit, the more satisfactory will be the design. Thus pilot ratings of the design can be cross-checked via modelling of simulator performance.

Table 4 shows the best curve fit transfer functions determined during the experiments leading to the construction of the $\zeta$ versus $\omega_n$ contours of Fig. 5. The values of $\zeta$ and $\omega_n$ quoted should be regarded as typical, rather than exact, since the original research indicated regions, rather than spot points, for which the transfer functions are appropriate. It can be seen that in the region rated by the pilots as satisfactory, the transfer function modelling of human operator response is just a gain plus a pure time delay. This is regarded as a general indication that the man-machine system has been optimally designed, and provides a standard against which a proposed configuration can be judged.

Thus the transfer function $(2.5/s)$ chosen for the 'quickened' system of Fig. 4 should be seen as a suitable approximation to the desired human operator behaviour in the region of cross-over frequency. Also $\zeta_{ho}(s) \approx (2.5/s)$ is a more convenient form of model for use with the particular design technique chosen, compared with a pure time delay.

Once the 'best' curve fit transfer function has been determined following simulator experiments and/or trials on a particular man-machine system configuration, the 'remnant' can now be examined. If the remnant is small, then the human operator is performing the task in a linear fashion. On the other hand, if the remnant is large, then the human operator is performing as a non-linear controller, as illustrated in Fig. 11. It is found that the large remnant also correlates with pilots' opinions. They rate systems which require them to work continuously in a non-linear mode as poor, so that the remnant can also be used to monitor the effectiveness of a particular configuration. The use of the word 'continuously' is quite deliberate: in stabilizing a divergent phugoid mode, a pilot will accept the need for occasional pulsed non-linear behaviour to dampen the oscillation.

14 Helicopter Dynamics

The helicopter is a vehicle which, without augmentation, is extremely demanding of the pilot. For example, the relationship between stick position and helicopter lateral position involves four integrations. This is because a stick movement changes the angle of attack of the rotor blades, which in turn results in a movement about the centre of mass of the vehicle, the second integral of which is roll angle. Two further integrations result because the lateral thrust of the helicopter is proportional to the roll angle, and lateral acceleration needs integrating twice to determine lateral position.

Clearly, from the 'plant' alone, 360 phase lag will occur at cross-over frequency, so that observing lateral position alone, the pilot needs to provide at least 180 of

<table>
<thead>
<tr>
<th>$\zeta$</th>
<th>0.75</th>
<th>0.60</th>
<th>0.25</th>
<th>0.15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_n$</td>
<td>2.5 r/s</td>
<td>3.5 r/s</td>
<td>3.5 r/s</td>
<td>1.5 r/s</td>
</tr>
<tr>
<td>Pilot gain</td>
<td>0.16 in</td>
<td>0.25 in</td>
<td>0.50 in</td>
<td>0.625 in</td>
</tr>
<tr>
<td>Pilot dynamics</td>
<td>$(1 + 0.67s)e^{-0.2s}$</td>
<td>$(1 + 0.77s)e^{-0.2s}$</td>
<td>$(1 + 0.25s)$</td>
<td>$(1 + 0.5s)^2$</td>
</tr>
<tr>
<td>Pilot opinion</td>
<td>Acceptable</td>
<td>Satisfactory</td>
<td>Poor</td>
<td>Unacceptable</td>
</tr>
<tr>
<td>Remnant</td>
<td>Moderately linear</td>
<td>Highly linear</td>
<td>Non-linearity becoming evident</td>
<td>Much</td>
</tr>
</tbody>
</table>

Table 4

Human operator transfer functions associated with typical combinations of short period damping ratio and undamped natural pulsatance of Fig. 5.
(a) Non-linear response by human operator (discontinuous response).

(b) Linear response by human operator (smooth continuous response).

Fig. 11. Examples of linear and non-linear human operator response.

Fig. 12. Helicopter instrumentation for 'unaided' control.

Fig. 13. Hierarchial model of pilot controlling an 'unaided' helicopter. (After Ref. 36.)

phase lead even to achieve critical stability. Unfortunately, the desire to avoid extra weight associated with the introduction of more complex flight control systems is extremely strong. This is due to the unique capabilities of the helicopter being achieved through a 50% higher power to weight ratio compared to the nearest fixed wing equivalent. The consequence is that a large number of helicopters flying today do not have any form of autostabilization system. Virtually continuous attention to the piloting task is therefore required, leaving little spare capacity for other mission requirements.

Nevertheless, helicopters continue to fly, despite the heavy workloads demanded of the pilots. There are two main systems engineering reasons why this is so. Firstly, as shown in Fig. 12, in a well instrumented vehicle the pilot can monitor $\phi$, $\dot{y}$, and $y$ in addition to the lateral position signal $y$. Monitoring $\phi$ will give 270° of phase lead, monitoring $\dot{y}$ will give 180° of phase lead, and monitoring $y$ will give 90° of phase lead. The pilot still has the problem of monitoring four signals (per channel) and has to decide in what proportion to 'mix' the signals in order to determine the best pattern of stick deflection. Note that in the 'quickened' display referred to previously, not only would the monitoring and mixing have been done on behalf of the pilot, but the mixing would have been done in a manner considered optimal for closed-loop performance.

Secondly, as shown in Fig. 1, the pilot can make use of his inherent vestibular feedback. Thus in this example, by sensing roll angular velocity $\phi$, the pilot will have 270° of phase lead available without reference to his instruments. However, signal mixing and further monitoring problems still remain to ensure a heavy workload.

The result of not providing a stability augmentation system in a helicopter is to require the pilot to operate in a highly non-linear manner. This behaviour has been studied in depth, and it is possible to construct the hierarchial model shown in Fig. 13, and which is found to correlate reasonably with simulator results. It shows how the pilot decides to operate either in bang–bang or simple tracking modes depending on the magnitude and sign of $\phi$, $\dot{\phi}$, $\dot{y}$, and $y$.

One proposal for reducing pilot workload in this situation is to provide supplementary information to the
Fig. 14. Assisting the human operator with fixed-time ahead prediction and display. (After Ref. 37.)

For example, it is suggested that displaying not only the current value of position, but additionally the expected value \( T \) seconds ahead will improve control quality. Figure 14 shows the principle in block diagram form. The future value is predicted using the Taylor series expansion

\[
x_p(t) = x_c(t + \tau) = \sum_{n=0}^{\infty} \frac{\tau^n}{n!} \frac{d^n x_c(t)}{dt^n}.
\]

Simulator results with a human operator controlling a v.t.o.l. third-order model (including a double integrator) showed considerable improvement in control, with reasonable values for the expansion being \( m = 2 \): \( \tau = 0.7 \) s. For these experiments the disturbances applied to the simulation were in frequency range \( 0.50 \rightarrow 0.5 \) Hz.

This technique is extremely simple, and is very modest in computational needs. Other methods of prediction, such as the Kalman filter, could obviously be used as an alternative to the Taylor series expansion. If the derivative terms are not available as sensor data, then these alternatives would have to be considered, pushing up the computational requirements. Note that two signals per channel are feedback, whereas in 'quickening' only the optimum mixed signal is displayed.

15 Helicopter Stabilization

Once it is decided to provide helicopter stability augmentation, the various levels of assistance can be readily identified, as shown in Fig. 15.\(^{38}\) For s.a.s., roll velocity is measured, or estimated, and fed back to turn the first integrator into a first-order lag, thereby reducing the phase lag at crossover frequency. An a.f.c.s. measures or estimates roll angle and roll velocity, feeding back to form a second-order lag whose damping ratio and natural frequency could be tuned to suit Fig. 5. Hover augmentation (h.a.s.), which would be extremely useful for air/sea rescue work, additionally feeds back yaw velocity. Finally, the most sophisticated system would feed back yaw position relative to a suitable datum.

The change in aircraft response with successive feedback closures shows that ultimately the pilot task is reduced to providing long-term adjustment of the aircraft position. It is found in practice that it is sufficiently sophisticated to provide h.a.s., thus reducing the pilot's task to a yaw velocity control function. For operation in a hostile environment, some form of inertial velocity sensing is desirable, as an aircraft could easily be detected if it relied on radio or Doppler techniques. One solution involving the further use of microelectronics is to implement a sensor-computer combination which will produce signals proportional to Earth-referenced velocities.\(^{38}\)

<table>
<thead>
<tr>
<th>Degree of Augmentation</th>
<th>Block Diagram</th>
<th>Stick Movement</th>
<th>Vehicle Roll Angle</th>
<th>Vehicle Lateral Displacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free Vehicle</td>
<td>S ( k_x )</td>
<td>( \delta ) S</td>
<td>( b ) S</td>
<td>( k_y ) S</td>
</tr>
<tr>
<td>Stability Augmentation System (SAS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic Flight Control System (AFCS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hover Augmentation System (HAS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H.A.S. Plus Position Feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 15. Various stages in improving helicopter flying controls. (After Ref. 38.)
16 Terminal Phase Control

As our final example on the use of microelectronics to improve man-machine system performance we consider the case shown in Fig. 16 where a 'fast model' of the vehicle dynamics provides long-term predictions of future behaviour. This is an important application for docking situations, especially in systems with large response times, since early action must be taken to achieve the desired goal. The 'fast model' is subject to inputs from the exploratory controls and the equations of motion, scaled faster by a factor of 100:1, or even 1000:1, are then solved to provide predictions of the future behaviour of the vehicle as a result of the present control action. The display can be quite sophisticated, as shown in Fig. 17, which gives the expected orientation and trajectory for a spacecraft.

Fig. 16. ‘Fast-model’ predictor for spacecraft landing trajectory optimization. (After Ref. 39.)

Fig. 17. Spacecraft predictor display shows position and attitude of a tumbling vehicle with respect to a desired command path. (After Ref. 39.)

It is possible to engineer the control stick and exploratory controls so that the pilot can ‘freeze’ his control action once he is satisfied with the outcome. This can be done by providing a freeze button on the stick, which he does not press until ready. Simulator studies show that using the 'fast-model' technique permits spacecraft pilots to correct rapidly changing thrust disturbances, not only more precisely, but with substantially less fuel consumption compared to results obtained with alternative displays.

17 Conclusions

Man-machine systems can be described in part using conventional control engineering terminology. By representing both human operator performance and desired plant dynamics in transfer function form, the author hopes to have drawn attention to guidelines which considerably assist in the design and understanding of complex systems. Recent advances in microelectronics offer immense opportunities to reduce pilot workload through a variety of techniques including the use of ‘observers’, ‘quickened displays’, ‘prediction displays’, ‘plant dynamic models’, and ‘fast models’. Originally implemented with analogue devices, many of these techniques would prove fruitful for microprocessor application once the reliability (including software reliability) has been proved.

18 References


-Manuscript first received by the Institution on 30th January 1980 and in final form on 28th April 1980

<Paper No. 1953/ACS 80>
COMPUTER ASSISTED MANUAL CONTROL OF CARGO HANDLING WITH SHIP CRANES

Bengt Schmidtbauer
Stefan Rönnbäck

University of Luleå, Sweden

Abstract

Handling large pieces of cargo, e.g. containers, with ship cranes is a difficult manual control task. If vertical axis control is added by introducing a torsional pendulosity, the difficulties begin to exceed what can be adequately handled by dock workers in many ports where ships are routed.

To alleviate the manual control operation, dynamical load compensation can be inserted between operator commands and actuator motion, leading to apparent load dynamics that are easier to handle. The implementation is done in a microcomputer, giving a high flexibility and possible sophistication to the compensation functions used.

The current work, dealing mainly with controlling the rotational motion, starts with a simulation study, to be followed by a small scale laboratory hardware implementation and a full scale prototype system.

For the simulation part a graphics processor with a 13" CRT, showing a visual real-time display of simulated cargo motion, is used together with a test operator with appropriate controls. Software for animated picture generation, simulation of load dynamics and implementation of above mentioned compensation functions has been produced for a Z80 - microcomputer.

The simulation set-up is currently used as an aid to find the best load compensation. In the longer run it could have another application as a crane operator training facility.

1. Introduction

Cargo transported on ships nowadays comes mainly in standardized units with weight and volume of considerable size, e.g. 20' and 40' containers. To move these pieces around human muscle has been replaced by hydraulic actuators, even for the accurate final positioning operation. The crane operator thus has to be in full control of all degrees of freedom of cargo motion, including vertical axis rotation.

The simultaneous manual control of 4 degrees of freedom (pitch and roll oscillations are disregarded), 3 of these being slow virtually undamped oscillations, is a formidable task, even for highly trained operators.
Ship cranes are usually handled by local dockers in the various ports along the ship's route. Their training on specific ship crane operations is often very limited. Thus it is particularly desirable in this case to add microcomputer intelligence to the manual control loops to improve cargo handling efficiency and ease the task of the crane operators.

The work presented here is carried out together with Swedish ship crane manufacturer Hägglunds. In the initial phase we are concentrating on vertical axis rotational control, since this is a novel area in crane operations requiring particular attention.

2. Load dynamics

To be able to exert a vertical axis torque, the load is suspended from two wires at a distance $a$ [m] from each other (Fig. 1). The wire length is $l$ [m] and the radius-of-inertia of the load (container) is $r$ [m]. With these assumptions the natural frequency of vertical axis oscillations is deduced from straightforward Newtonian dynamics as ("torsional pendulum"):

$$\omega_0 = \frac{a}{2r} \sqrt{\frac{g}{l}} \ [\text{rad/sec}] \quad \ldots(1)$$

Fig. 1 Ship crane with CRD and container
The wire suspension can be rotated with respect to the container using the cargo rotation device (CRD), the rotation angle $\Psi$ [rad] being the primary variable for vertical axis control (Fig. 2).

Neglecting energy dissipation and assuming $1 \gg a$, the equation for the angular motion $\varphi$ [rad] of the container can be expressed as:

$$\ddot{\varphi} = \omega_0^2 \cdot \sin(\Psi - \varphi) \quad \ldots (2)$$

The maximum angular acceleration is thus equal to $\omega_0^2$ [rad/s$^2$] and occurs when the wires are twisted 90° in either direction.

Fig. 2 Rotation angle definition

3. Time-optimal control

Even though this study is aimed against manual control, it is essential that the optimal control problem is analysed, to serve as one of the standards of comparison for the manual control results. A natural formulation of the optimal control problem is one of time-optimal control, i.e. to move the system to the desired target state in the least possible time under the inherent physical constraints of the process.

Case 1: No constraints on CRD velocity ($\dot{\Psi}$)

In this case the process (2) becomes a standard double integral plant, with the control magnitude constrained between $\pm \omega_0^2$, the solutions being well-known (Ref. 1). For simplicity a normalized time variable is introduced as:

$$\tau = \omega_0 t \quad \ldots (3)$$

Taking derivatives with respect to $\tau$, equation (2) is replaced by:

$$\frac{d^2 \varphi}{d\tau^2} = \sin(\Psi - \varphi) \quad \ldots (4)$$

the control then being constrained between $\pm 1$. 
The time-optimal (bang-bang) control trajectories to take the system from the origin through a cargo rotation of \( \phi_1 \) radians with zero terminal derivatives \( (\dot{\phi}, \ddot{\phi}) \) are shown in Fig. 3. The idealized control trajectory \( \Psi(\tau) \) is a solid curve whereas a more practical case with limited CRD actuation speed is indicated with dotted lines.

With the resulting parabolic arcs for \( \phi(\tau) \), the relationship between deflection angle and the corresponding minimal time duration is:

\[
\tau_1 = 2\sqrt{\phi_1} \quad \ldots(5)
\]

**Case 2: Limited CRD velocity; small deflection angles**

Here it is assumed that \( |\phi| \ll |\Psi| \ll \pi/2 \) and the maximum actuator speed:

\[
\left. \frac{d\Psi}{dt} \right|_{\text{max}} = \psi_{\text{max}} / \omega_0 = b \quad \ldots(6)
\]

The control signal is now CRD speed, the plant (4) approximately becoming triple integral:

\[
\frac{d^3\phi}{dt^3} = \frac{d\Psi}{dt} \quad \ldots(7)
\]

The corresponding time-optimal trajectories are shown in Fig. 4. In this case the function \( \phi(\tau) \) is piecewise cubic, the expression for the minimal time duration \( \tau_1 \) for an angular rotation \( \phi_1 \) becoming:

\[
\tau_1 = \frac{3}{32}\phi_1 \frac{1}{b} \quad \ldots(8)
\]
In the range not covered by case 1 and 2 the time-optimal trajectories can be computed numerically. Using a two-dimensional Newton-Raphson iteration to calculate the time-optimal switching instants, the resulting trajectories $\psi(t)$ turn out like in Fig. 5 for the case of normalized time $\tau^1_1 = 4$. For normalized CRD speed $b$ larger than around 1.5 a double saturation occurs like in "case 1" whereas for lower values of $b$ the trajectories are similar to "case 2".

![Fig. 5 Simulated time-optimal trajectories ($\tau^1_1 = 4$)](image)

The resulting relationship between $\tau^1_1$, $\phi_1$ and $b$ is shown in Fig. 6 (logarithmic scales).

![Fig. 6 Minimal time as a function of slew angle and peak actuator speed (normalised values)](image)
4. Approaches to computer-assisted manual control

The most difficult part of manual cargo rotation control is the fact that the torque changes with the rotation angle leading to the undamped oscillator type of dynamics. A first step in the compensation approach is to decouple torque control from cargo rotations, leading to an effective control signal $u_0$ according to (cf equation (4)):

$$u_0 = \psi - \varphi$$  \hspace{1cm} ... (9)

The CRD rotation command $\psi = u_0 \star \varphi$ thus has to be computed as the sum of the control command $u_0$ and the rotation angle $\varphi$, the latter obtained from measurement or state estimation techniques. From the operator's point of view the plant has been changed from harmonic oscillator to approximately double integral (if $b$ is large).

However, a double integral plant with additional lag due to CRD speed restrictions is still difficult to handle, and needs a high degree of anticipatory (lead or derivative) action to be stable and well behaved. One way to change plant dynamics from double integral to single integral with lag is to introduce artificial viscous damping to the plant dynamics. With $c$ [s] being the viscous damping constant, the primary control signal has to be:

$$u_c = \psi - \varphi + c \dot{\varphi}$$  \hspace{1cm} ... (10)

with the corresponding CRD rotation command $\psi = u_c \star \varphi - c \dot{\varphi}$.

The different versions of apparent plant dynamics are illustrated by their respective impulse responses in Fig. 7.

---

**Fig. 7** Impulse responses for different apparent load dynamics
A. Uncompensated CRD speed control
B. Decoupled with fast CRD
C. Decoupled with fast CRD and damping
5. Simulation model

Based on the principles of the previous section, different computer assisted manual control schemes have been evaluated with a number of test operators. Plant dynamics compensation functions and real-time graphics were implemented on a Z80-microcomputer, with a graphics processor and a 13" CRT for visual presentation (Fig. 8).

Fig. 8 Computer and graphics display equipment

Plant model (cf. section 2)

For the test runs the following numerical values were chosen:

- Wire length \( l = 15 \) m
- Wire distance \( a = 1 \) m
- Inertia radius \( r = 1.73 \) m (20' container)

Substituting into equation (1) we get:

\[ \omega_0 = 0.233 \text{ rad/sec} \]
Manual control task

The control task used in the simulations is to rotate the container by 90° ($\varphi_1 = \pi/2$ rad). A top view of the container and the CRD is presented on the display tube, according to Fig. 9. The operator exerts (continuous) control by rotating a potentiometer. In the uncompensated case the potentiometer signal is directly turned into a corresponding CRD speed command. With computer assistance the potentiometer signal is a setpoint for the commands $u_0$ or $u_c$ in eqns. (9) or (10). The corresponding CRD rotation angle $\Psi(t)$ is generated with a software local feedback loop with the appropriate saturation characteristics ($|\Psi| < \Psi_{\text{max}} = b\omega_0$).

A manually controlled 90° rotation is considered to be accomplished when the orientation has been within ± 0.03 rad of the target during 5 seconds.

Control parameters

Five different cases of control are investigated in the simulation study.

Case 0: Direct control of CRD speed with no compensation (low peak speed; see A1 below).

In all cases of computer assistance a decoupling of container rotation is carried out (see section 4).

The differences in the compensation are with respect to:

A. Peak CRD speed

A1 Low speed actuator: $\Psi_{\text{max}} = \omega_0 b = 3 \text{ rpm} = 0.314 \text{ rad/s}$
A2 High "-" : $\Psi_{\text{max}} = \omega_0 b = 30 \text{ rpm} = 3.14 \text{ rad/s}$

... (13)

B. Amount of artificial damping

B1 No artificial damping $c = 0$ [s]
B2 High artificial damping $c = 21$ [s]

... (14)

The computer assisted control alternatives thus can be labelled as:

Case 1: Low speed/low damping
Case 2: Low speed/high damping
Case 3: High speed/low damping
Case 4: High speed/high damping
Time-optimal control

With $\omega_1 = 90^\circ$ and $b$ given by A1 and A2 above (equal to 1.35 or 13.5 for low and high speed actuator respectively), the corresponding minimum time durations can be obtained from Fig. 6 as:

A1: $t_1 = \frac{T_{11}}{\omega_0} = 14.7$ secs
A2: $t_1 = \frac{T_{11}}{\omega_0} = 11.2$ secs

...(15)

In the case (B2) with artificial damping the minimal time is increased, since the cargo rotation speed saturates at:

$\dot{\phi}_{\max} = \frac{\pi}{2c} = 0.075$ rad/s

...(16)

Taking the transient behaviour at the beginning and end of the trajectory into account, the effective minimal time in the damped case is approximately:

$t_{1d} = 22$ secs

...(17)

6. Simulation results

In the simulations 25 different test operators were used, 5 on each of the test cases described in the previous section. Each operator had to carry out 25 runs (90° rotations).

The ensemble mean values for the required time in the successive trials are plotted in Fig. 10. The corresponding theoretical time-optimal control limits are also indicated. Based on the results we can make the following observations:

Need for compensation

The uncompensated non-decoupled dynamics in case 0 were extremely hard to control manually, even after the in this case very long training session of 25 trials. A high degree of attention is necessary to even maintain stability. Results are very poor. Learning is slow and improves drastically in the cases with compensation.

Actuator speed

The increased lag in the slow actuator cases (1 and 2) leads to a decreased manual control performance in comparison with the corresponding cases with a fast actuator (3 and 4). For a practical design this effect has to be traded against the increased cost of a fast actuating mechanism.
Artificial damping

Artificial damping has little influence on the mean values but has a strong effect in decreasing fluctuations, both between different operators and between different trials. Less attention is required. Thus for predictable and dependable manual control a fair amount of damping appears to be desirable.

7. Areas of further study

For a practical implementation of the computer assisted control schemes discussed in this paper a sensing mechanism for cargo rotation (ϕ) is needed. We are currently looking into a single-axis gyro implementation but also into the possibility of reconstructing the rotation angle with state estimation techniques using other information sources.

In a complete ship-crane manual control system, the cargo rotation controls have to be treated in an integrated framework together with translational control.

8. References

session 5
human behaviour in car driving

chairman: h.g. stassen, netherlands
SIGNIFICANT CHANGES IN DRIVER-VEHICLE RESPONSE MEASURES FOR EXTENDED DURATION SIMULATED DRIVING TASKS

Walter W. Wierwille and William H. Muto
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061, U.S.A.

ABSTRACT

As a first step in understanding the effects of fatigue on driving, it is necessary to examine how driver-vehicle measures change with time on task. Data of this type are difficult to obtain in a real-world environment because of lack of experimental control. As part of a larger study on fatigue, the simulator at Virginia Polytechnic Institute and State University was used in extended duration runs in an effort to examine trends in driver-vehicle response measures, physiological measures, and subjective measures. The simulator possesses four degrees of platform motion and six degrees of display motion. It handles realistically and has been demonstrated as valid.

A variety of measures were taken, including number of body movements per unit time, heart beat interval standard deviation, vehicle lateral standard deviation, steering standard deviation, large steering reversals per unit time, small steering reversals per unit time, steering reversal mean amplitude, yaw standard deviation, and yaw reversals. All of these measures, taken for consecutive 10 minute intervals, show statistically significant trends over a two and one half hour driving period. Furthermore, the data show a high level of monotonicity when averaged over twelve subjects, indicating that the changes are pronounced and well behaved.

Opinion ratings as a function of driving time, taken from a larger pool of subjects, indicated that subjects rated themselves significantly more fatigued as run length was increased from one half hour to one hour. When driving time was increased to two and one half hours, subjective ratings of fatigue showed a trend toward an increase, but the trend was not statistically significant.

The results of this study indicate that asymptotic performance in simulated driving is not reached within a two and one-half hour driving period. Therefore, it is very important for researchers to use equal amounts of practice and time on task when taking data in driving situations. Drivers cannot be assumed to have practiced "to asymptote". In addition, because of differences in subjective and performance-related measures, both types of measures should be taken in assessing the effects of time on driving tasks.

INTRODUCTION

Driver fatigue has been a topic of great concern throughout the world. In the United States, it has been estimated that "... 20 to 25 percent of all accidents on high-speed, long distance roads, and 40 to 50 percent of the fatalities, are ... [attributable] to driver fatigue" (Reference 1, page 304).
Investigators in Japan (Reference 2) have come to similar conclusions, indicating that up to 40% of the casualties due to traffic accidents have been related to fatigue. The problem is not limited to cars. The Bureau of Motor Carrier Safety recently investigated 286 commercial vehicle accidents and came to the conclusion that 111 were caused by a driver falling asleep or driver inattention due to fatigue (Reference 3). These 111 accidents caused 142 deaths, 281 injuries, and nearly five million dollars in property damage.

Fatigue is a general term used to describe manifestations of extended time on task and insufficient rest. Numerous technical definitions have been set forth, none of which is universally accepted. Because fatigue cannot be measured directly, and because the definitions give researchers little guidance, operational definitions have been used (Reference 4). Typically, operational definitions fall into the categories of performance measures, physiological measures, and opinion measures.

Because of the importance of fatigue particularly in transportation, numerous research studies have been conducted on fatigue and on related topics. Harris and Mackie (Reference 3) performed an extensive study involving direct observation of truck and bus drivers. They concluded that increases in performance errors and decreases in levels of physiological arousal occurred within the first ten hours of driving. However, open road testing was used, and it was difficult to control extraneous variables that are associated with experiments of this type. Platt (Reference 4) conducted a study of two drivers on the U.S. interstate highway system, with each subject alternating as driver and observer every one and one-half hours for a distance of 1200 miles (1931 km). Noticeable changes occurred in performance as measured using the Greenshields Drivometer. Safford and Rockwell (Reference 5) also studied drivers on the road over a 24 hour driving period. They concluded that velocity variation, steering reversals, and accelerator reversals generally increased over time. However, individual subjects showed decreases in some of the measures over time, causing some concern about reliability and generalizability. Other studies, have introduced techniques other than direct performance as a means of investigating fatigue (References 6 through 11).

Simulators have occasionally been used to study fatigue. The have the advantage of inherent safety, but the disadvantage of requiring proof of validity. Sussman, Sugarman, and Knight (Reference 12), using a simulator designed by Wierwille, performed a study in which subjects drove for four hours. Several conclusions were drawn from the study: the number of steering corrections and the lateral position tracking accuracy decreased with time; larger errors occurred in recovering from emergencies after the extended driving time; and EEG alpha wave bursts increased with time. Dureman and Boden (Reference 13) used an "outside in" simulator for an experiment involving four hours of driving. Subjects showed performance decrements in several measures over time. One group of subjects received mild electric shocks, and their decrements were somewhat less than for those subjects who did not receive shocks.

In concluding this very brief (and only representative) literature review, mention should be made of the voluminous work on vigilance that has been done in the last two decades. While generally of a less applied nature, vigilance clearly is an aspect of extended driver performance. The excellent review by Davies and Tune (Reference 14) covers the general ideas and results.
EXPERIMENT

The literature on driver fatigue indicates that certain performance measures, physiological measures, and subjective measures change over time. The changes are reported primarily as differences between fresh drivers and the same drivers after an extended period of time on task. There is generally no information regarding the trend (curve) shape over time. Furthermore, the previous studies conflict with one another to a certain extent. In some cases steering reversals or corrections per unit time have been found to decrease while in other cases they have been found to increase. Therefore, it was felt that as an initial step in undertaking work on fatigue, a more careful examination should be made of the performance, physiological, and opinion measures believed to be sensitive to driver time on task. Emphasis was placed on determining the shapes of the curves showing any changes over time. Important questions needed to be answered. Do drivers rapidly deteriorate in performance after a prescribed length of time, or are the trends in performance gradual? Do changes in physiological and subjective measures correspond to the performance changes, or are they characteristically different? Questions of this type have not been fully answered in previous research on driver fatigue, and they seem to require answers before more complex issues can be examined.

For safety and control over extraneous variables, the driving simulator at Virginia Polytechnic Institute and State University (VPI & SU) was used for the experiment. This simulator has been empirically demonstrated as valid in an absolute sense over as many as five performance-related measures in highway driving tasks (Reference 15). This validity was checked by examining correspondence between actual and simulated road-tests (interstate highway) for fresh subjects. While there has been no validity check in regard to fatigued subject comparisons, it is quite likely that the same curve shapes would be obtained, except possibly for some distortions in the independent variable, time.

The VPI & SU simulator has been previously described in the literature (References 15 through 17). Therefore, only a brief description will be presented here, including a description of new equipment for simulating other vehicles in the highway scene.

The simulator is composed of five major subsystems: the computer-generated roadway image system, the "video bench" system for inserting images of "other" vehicles in the driver's scene, the motion and vibration system, the sound system, and the vehicle dynamics simulation system.

The computer-generated display system presents a roadway image of a two-lane concrete highway with a dashed centerline. Unlike most other display systems, this one uses large numbers of lines to create a realistic "solid roadway" instead of an outline of a roadway. Field markings on each side of the road and additional horizontal lines give the road the appearance of being embedded in a horizontal plane. The system takes a vector input of vehicle forward velocity, lateral position, yaw, roll, pitch, and inverse radius of curvature. The image generator uses a combination of digital, hybrid, and high-speed analog techniques to produce the image. The final image is produced on a 23 in (58 cm) monochrome monitor, which is viewed through a 24 in (61 cm) by 19 in (48 cm) fresnel lens. The image, as viewed by the subject, is 48" wide and 38" high, and appears at a distance of 33 ft (10.1 m).
The video bench "shoots" models of vehicles (matchbox size) using a monochrome video camera. The resulting image is insert/keyed onto the computer generated roadway, producing a realistic scene. The vehicle model is mounted on a pedestal which can be rotated under servo control, thereby producing variable aspect angle. The camera is fitted with a 10:1 zoom lens, also under servo control, thereby producing variable apparent distance. Finally, lateral position on the highway is simulated by servo-controlling the rotation of a mirror in the optical path between the camera and the model. An interface-mapping processor is used to create and maintain correspondence between the roadway and the correct position of the model.

The simulator's motion system uses closed-loop servo-controlled hydraulic actuators. Roll, yaw, lateral translation, and longitudinal translation are obtainable. A unique platform is used to achieve these four motions (Reference 16). Only the roll axis makes use of conventional gimbals. The other three axes are achieved by a technique similar to the "hydraulic stilt" technique now widely used in aircraft simulators. Engine-drivetrain vibration is simulated on the platform by means of a motor with an eccentric mass. The velocity of the motor is controlled by the simulated engine speed. This motor and eccentric also contain external magnets, which are used to induce voltage waveforms in a coil. The voltage output of this coil is shaped, filtered, and amplified to produce engine sound. The other portions of the sound system produce tire screech (for braking), tire squeal (for large steering inputs at high speed), and rolling resistance-road sound. All of these sounds are modulated and controlled by the dynamics.

The vehicle dynamics are simulated using analog and hybrid computers. These dynamics take steering, braking, and accelerator signals, as well as gusts and curvature as inputs. The outputs consist of vehicle velocity, lateral position, yaw, roll, and pitch. Conditioned signals are also provided for the motion servos, the speedometer, the video bench processor, the sound system, and the FM recording system (used for data processing).

Overall the simulator produces a realistic, properly handling, vehicle simulation of a 1978 mid-size American made sedan (rear wheel drive).

Thirty-six subjects, ranging in age from 18 to 31, were used in this experiment. All had 20/25 visual acuity or better, with correction where needed. Equal numbers of males and females were used. Subjects were paid for their participation. On the first day of the experiment, subjects drove the simulator for thirty minutes in a practice session. Both normal and "emergency" situations were simulated. The emergencies were in three classes: recovery from step side gusts, evasive or braking maneuvers due to lead car stopping on the highway, and lane changes required by light signals. After completion, subjects were instructed to return on the next day, to get a full night's sleep, and to avoid use of stimulants until after the experiment was concluded. On the second day of the experiment, subjects were again exposed to their corresponding simulator conditions, this time for twenty minutes. Subjects then left simulator for a two hour recess. When they returned, they began the extended driving task (with no emergencies), during which time the data were taken. Subjects drove for 30 min., 60 min., or 150 min. Each subject was exposed to only one of the three classes of emergencies in practice and only one of the three durations. Thus, the experiment consisted of nine cells with four subjects (two males and two females) in each cell. Following the runs,
each subject estimated his/her level of fatigue using a rating scale sheet. The sheet contained two seven-point scales, one for rating his/her level of fatigue at the beginning of the extended duration run and another for rating level of fatigue at the end of the extended duration runs. The scales had anchor points of "extremely tired" at one end (-3 rating) and "extremely alert" at the other end (+3 rating).

Instructions during the data run stressed driving normally, in the right lane, and remaining alert for possible emergencies (although none occurred during the extended driving task). The instructed speed was 55 mph (89 km/hr). In all simulator runs, there was a lead car in the visual scene at an apparent distance of 165 ft (50.3 m). Subjects were instructed to maintain that distance. If they exceeded the instructed speed, they gained on the lead car; and if they were below the instructed speed, the lead car gained on them (increased its apparent distance).

The simulator was instrumented to obtain measurements of nine performance and physiological measures:

1. **BDYMOVES** — number of torso movements
2. **HRTINTSD** — heart beat interval standard deviation (sec)
3. **LATDEVSD** — vehicle lateral position standard deviation (cm)
4. **STEERSD** — steering wheel standard deviation (deg).
5. **STREVRG** — number of large (>2°) steering reversals
6. **STREVSML** — number of small (0.5° to 2.0°) steering reversals
7. **STRREVMM** — mean steering reversal amplitude (deg)
8. **YAWDEVSD** — vehicle yaw angle standard deviation (deg)
9. **YAWREVII** — number of yaw reversals >2°.

The BDYMOVES measure was obtained by conditioning signals from two sensors embedded in the back cushion of the driver's seat, one at the lumbar region and one 10 inches (25.4 cm) above, in the thoracic region. The purpose of this measure was to detect any straightening or stretching motions of the subjects. The HRTINTSD measure was obtained using an ear plethysmograph system. Subsequently, the interval between consecutive pulses was computed.

Data were recorded on a multichannel FM tape recorder and later analyzed using a PDP-11-55 digital computer with an LPS-11 data conversion system. Signals were sampled at a 100Hz rate. All measures were computed and printed for each 10 min segment of each run. By using a 10 min. segment, measurement stability could be maintained while still allowing trends due to time on task to be observed.

**RESULTS**

Although all performance and physiological data were thoroughly analyzed, only the results associated with the 150 min. runs will be presented here. These 150 min. data runs actually possess the same trends as the shorter runs, but of course, also contain the additional information beyond the first 60 min. of the experiment. The opinion data, however, are reported for all 36 subjects participating in the experiment.
The 150 min. performance and physiological data were first analyzed using a multivariate analysis of variance (MANOVA). The analysis included "experiment", which refers to the type of emergency used in the practice session; and subjects, nested within experiment. All statistical tests on the performance and physiological data were analyzed using the computer-packaged Statistical Analysis System (Reference 18). Table 1 shows that the type of emergency simulated in practice did not significantly alter the performance of subjects in the extended duration runs (Wilk’s criterion). However, the table does show that significant changes in the measures did take place over time during the extended runs ($p < 0.01$, Roy's Union Intersection Test). (Separate tests had to be used because the Roy's test table did not encompass the necessary range). Accordingly, a univariate analysis of variance on each of the nine individual measures was performed.

Table 1.
MANOVA Summary for performance and physiological measures.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>p</th>
<th>$\gamma_H$</th>
<th>$\gamma_E$</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment (E)</td>
<td>9</td>
<td>2</td>
<td>9</td>
<td>.000017</td>
</tr>
<tr>
<td>Subjects/E (S/E)</td>
<td></td>
<td>(Error term for E)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Roy's Union Intersection Test)

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>s</th>
<th>m</th>
<th>n</th>
<th>Max</th>
<th>$\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (T)</td>
<td>9</td>
<td>2.0</td>
<td>58.0</td>
<td>2.3134</td>
<td>.6982**</td>
</tr>
<tr>
<td>Ext</td>
<td>9</td>
<td>9.0</td>
<td>58.0</td>
<td>.6561</td>
<td>.3962</td>
</tr>
<tr>
<td>T x S/E</td>
<td></td>
<td>(Error term for T, Ext)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**$p < .01$**

Because presentation of the individual analyses is beyond the scope of this paper, a summary has been prepared and is presented in Table 2. It shows that all nine measures exhibit statistically significant changes as a function of time.

To obtain a better understanding of the trends, each measure has been plotted over time, along with corresponding 90% confidence limits. The plots are presented in Figures 1 through 9. Also two regression analyses were performed on each of the nine measures. The first analysis was performed on group means for each time segment, and the second was performed on individual driver measures for each time segment. The results appear in Table 3, which shows the group mean regression equation, and the $R^2$ (proportion of variance accounted for) for group mean data and for individual driver data. Eight of the nine measures have positive slopes, STREVSML being the exception, with a negative slope.
Figure 1. Mean and 90% confidence limits for frequency of body movements (BDYMVES) vs time.

Figure 2. Mean and 90% confidence limits for heart beat interval standard deviation (HRTINTSD) vs time.
Figure 3. Mean and 90% confidence limits for lateral standard deviation (LATDEVSD) vs time.

Figure 4. Mean and 90% confidence limits for steering angle standard deviation (STEERSD) vs time.
Figure 5. Mean and 90% confidence limits for steering reversals greater than or equal to 2° (STREVLRG) vs time.

Figure 6. Mean and 90% confidence limits for steering reversals between 0.5° and 2.0° (STRREVSM) vs time.
Figure 7. Mean and 90% confidence limits for steering reversal mean amplitude (STRREVMN) vs time.

Figure 8. Mean and 90% confidence limits for yaw angle standard deviation (YAWDEVSD) vs time.
Figure 9. Mean and 90% confidence limits for frequency of yaw reversals greater than or equal to 2° (YAWREVII) vs time.

Figure 10. Mean (post-hoc) opinion ratings of fatigue for "before" and "after" extended driving runs, by duration.
Table 2.

Summary of univariate ANOVAs for effects of time on the performance and physiological measures.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>df</th>
<th>MS</th>
<th>( F )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDYMoves</td>
<td>14,126</td>
<td>32.5341</td>
<td>1.99</td>
<td>2x10^{-2}</td>
</tr>
<tr>
<td>HRTINTSD</td>
<td>14,126</td>
<td>126</td>
<td>6.20</td>
<td>3x10^{-9}</td>
</tr>
<tr>
<td>LATDEVSD</td>
<td>14,126</td>
<td>156.5832</td>
<td>8.83</td>
<td>3x10^{-13}</td>
</tr>
<tr>
<td>STEERSD</td>
<td>14,126</td>
<td>126</td>
<td>6.96</td>
<td>2x10^{-10}</td>
</tr>
<tr>
<td>STREVLRG</td>
<td>14,126</td>
<td>9423.4698</td>
<td>4.93</td>
<td>3x10^{-7}</td>
</tr>
<tr>
<td>STREVSMNL</td>
<td>14,126</td>
<td>15473.1667</td>
<td>9.17</td>
<td>1x10^{-13}</td>
</tr>
<tr>
<td>STRREVMN</td>
<td>14,126</td>
<td>1.4320</td>
<td>6.07</td>
<td>5x10^{-9}</td>
</tr>
<tr>
<td>YAWDEVSD</td>
<td>14,126</td>
<td>0.0206</td>
<td>7.75</td>
<td>1x10^{-11}</td>
</tr>
<tr>
<td>YAWREVI</td>
<td>14,126</td>
<td>157.2460</td>
<td>4.37</td>
<td>3x10^{-6}</td>
</tr>
</tbody>
</table>

The opinion ratings of fatigue were analyzed in a two-way ANOVA with "duration" as one factor and "test" (post-hoc ratings of fatigue for before and after the extended duration run) as the other. As shown in Table 4, significant effects were found due to test, \((p < 0.0001)\) and due to duration x test interaction \((p < 0.001)\). A Newman Keuls analysis of the duration x test interaction revealed that: (a) mean subjective rating scores for all duration conditions were significantly lower (more fatigued) for the "after" ratings than for the "before" ratings \((p < 0.01)\); (b) mean "before" ratings were significantly higher (more alert) for the medium (60 min) and long (150 min) duration conditions than for the short duration (30 min) condition \((p < 0.01)\); (c) mean "after" rating scores were significantly lower for medium and long duration conditions than for the short duration condition \((p < 0.01)\); and (d) no significant differences were found among corresponding "before" and "after" ratings for the medium and long duration conditions. The mean "before" and "after" rating values are plotted in Figure 10.

CONCLUSIONS

The results of this experiment show surprisingly well-behaved trends in the measures taken as a function of time. Both physiological and performance related variables exhibit smooth, predominantly linear trends over time.
Table 3.
Regression analysis summary of trends over time.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Equation - Linear Estimate</th>
<th>$R^2$ (Group Means)</th>
<th>$R^2$ (Individual Data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDYMOVES</td>
<td>$Y = 1.04 + .35 x$</td>
<td>.92</td>
<td>.080</td>
</tr>
<tr>
<td>HRTINTSD</td>
<td>$Y = .073 + .003 x$</td>
<td>.96</td>
<td>.148</td>
</tr>
<tr>
<td>LATDEVSD</td>
<td>$Y = 29.1 + .78 x$</td>
<td>.93</td>
<td>.134</td>
</tr>
<tr>
<td>STEERSD</td>
<td>$Y = 2.77 + .06 x$</td>
<td>.97</td>
<td>.187</td>
</tr>
<tr>
<td>STREVLRG</td>
<td>$Y = 98.5 + 6.0 x$</td>
<td>.91</td>
<td>.135</td>
</tr>
<tr>
<td>STREVSML</td>
<td>$Y = 470.2 - 7.6 x$</td>
<td>.90</td>
<td>.110</td>
</tr>
<tr>
<td>STRREVMN</td>
<td>$Y = 2.12 + .075 x$</td>
<td>.94</td>
<td>.153</td>
</tr>
<tr>
<td>YAWDEVSD</td>
<td>$Y = .316 + .009 x$</td>
<td>.94</td>
<td>.165</td>
</tr>
<tr>
<td>YAWREVII</td>
<td>$Y = .47 + .79 x$</td>
<td>.95</td>
<td>.123</td>
</tr>
</tbody>
</table>

Table 4.
Analysis of variance summary for opinion ratings of fatigue.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration (D)</td>
<td>2</td>
<td>.0495</td>
<td>.03</td>
<td>&gt; .10</td>
</tr>
<tr>
<td>Test (T)</td>
<td>1</td>
<td>119.2604</td>
<td>163.64</td>
<td>.0001</td>
</tr>
<tr>
<td>Subject/D (S/D)</td>
<td>45</td>
<td>1.9087</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D x T</td>
<td>2</td>
<td>6.0964</td>
<td>8.36</td>
<td>.0008</td>
</tr>
<tr>
<td>T x S/D</td>
<td>45</td>
<td>.7288</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Significant increases were found in both heart beat interval standard deviation and in number of body movements, two measures which had been purport ed as indicators of changes in the driver's state of alertness. The results on heart beat standard deviation agree with the results of several previous studies, which showed greater variability in heart rate with fatigue (References 19 and 20); and the results on body movements tend to agree with the observations of O'Hanlon and Kelly (Reference 21). However, the data obtained on body movements in the present experiment appear much more uniform and predictable than those which they observed.

The measures which reflect vehicle control show significant changes over time also. Lane keeping performance becomes increasingly more erratic with the passage of time. This tendency has been seen in earlier investigations (References 1 and 12 for example), but again, the uniformity of the trend has not been previously observed.

Similar changes were observed in yaw angle standard deviation and in yaw reversals greater than 2°. Here again the trends are linear and increasing, indicating gradual increases in laxity. The increases in the yaw variability and numbers of reversals appear unprecedented in the driving fatigue literature.

Steering behavior shows significant well-behaved trends also. However, steering behavior is somewhat more complex than it may at first appear. Note that small steering reversals (0.5° to 2.0°) tend to decrease with time, while large steering reversals (greater than 2°) as well as steering standard deviation and steering reversal mean amplitude tend to increase with time. These results suggest a gradual shift in the strategy drivers use as time passes. Fresh drivers use small, precise corrections where necessary. This apparently allows them to maintain tighter control over the vehicle, as exhibited by corresponding smaller lane and yaw deviation. However, as time passes, drivers cease to maintain tight control, and instead, use coarser corrections. This causes a reduction in the number of small steering reversals and an increase in the number of large ones. The results of this study serve as an explanation of the conflicting results obtained in earlier studies which used steering reversals as a means of assessing fatigue (References 1, 3, 5, 7 and 12). Quite clearly, the size of the reversals is very important. Small ones decrease, and large ones increase. The results obtained by other investigators have shown either increases or decreases with time or task, depending on the size of the reversals used and to what degree large and small reversals are combined in a conglomerate measure. Any future studies involving driving fatigue should make use of this important finding by classifying the sizes of steering reversals.

Before leaving the area of physiological and performance measure results, a word of caution is in order. Table 3, showing the linear regression equations and the $R^2$ values reveals an important, but somewhat discouraging finding. The $R^2$ values obtained for regressions on group means (using 12 subjects to compute each mean) show very high values (.90 to .97), indicating that group means can be accurately modeled by linear trends over time. In contrast, $R^2$ values obtained for regressions employing individual data are much smaller (0.80 to .187), indicating the regression equations account for a relatively small proportion of the total variances. Therefore, while the group means are well-behaved over time, the data for individual subjects are not so well behaved. This makes prediction of individual behavior as a function of time impossible on the basis of the present data.
The results of the drivers' post-hoc opinion ratings of fatigue revealed that they were rather reliable, when the data are compared with the physiological and performance data. The mean ratings for "before" and "after" extended duration runs indicate that subjects judged themselves to be less alert after having driven for a prolonged period than they were at the beginning of the period. Furthermore, subjects who drove either 60 min. or 150 min. judged themselves to be less alert after the runs than subjects who drove for only 30 min. The lack of significant differences for ratings among the medium (60 min.) and long (150 min.) runs indicates an asymptotic driving time effect on subjective ratings. This asymptotic effect, which was exhibited after 60 min. of continuous driving, indicates either that (a) subjects were maximally "fatigued" after 60 min., (b) ability to judge any further changes in alertness was diminished, or (c) subjective changes are sufficiently nonlinear so that much longer driving times are needed to elicit "fatigued" ratings. Although the trend is in the right direction, the subjective ratings demonstrate asymptotic behavior, whereas the physiological and performance measures do not.

One other interesting note in regard to the opinion ratings involves the significance of the duration x time interaction. Apparently, subjects had a tendency to judge themselves more alert before the extended duration runs and less alert after the runs, as a function of run length (Figure 10). Assuming no inherent biases were present in subject sampling or experimental conditions, the most plausible explanation for the differences in "before" ratings is that subjects tended to anchor their initial ratings according to their perceived magnitude of change in alertness over the driving period. In other words, subjects judged themselves to be more alert in the "before" situation in an effort to increase the dynamic range of the changes in alertness. It is apparent that care should be taken when attempting to assess relative or absolute levels of alertness based on such post-hoc judgements.

Finally, this experiment demonstrates conclusively that physiological and performance measures in simulated driving do not reach asymptote within 2½ hours of driving. Instead, linear, well-behaved trends occur. The ramifications of this finding are far-reaching. Clearly, any differences in time on task can be expected to result in measures that are confounded with time. If the results of this experiment are generalizable to other driving experiments — and there is no reason to believe they are not — then the idea that subjects have "practiced to asymptote" is fallacious. The best that an experimenter can do is insure that each subject gets an exactly equal amount of practice prior to data taking.

ACKNOWLEDGMENT

This research was sponsored by the General Motors Corporation and by Virginia Polytechnic Institute and State University.
REFERENCES


DRIVERS' INTERNAL REPRESENTATION AND SUPERVISORY CONTROL:
A FIRST MODEL VERIFICATION IN RELATION TO DRIVING EXPERIENCE
TASK DEMANDS AND DETERIORATED VISION

Gérard J. Blaauw
Institute for Perception TNO, Soesterberg, The Netherlands

ABSTRACT

Interim results are presented of an ongoing research project in which the behaviour of car drivers is described by a supervisory control model with two main elements, 'observation/prediction' and 'control', and an intermediate internal representation. This representation is a function of both the driver's cognitive and perceptual anticipation of relevant features of the driving environment and the visual feedback information from that environment. The inevitable development of uncertainty within this representation acts to initiate the acquisition of new visual information. Control actions are initiated in accordance with the internal representation of the overall performance. After a short description of the model some hypotheses are given on the effects of the level of driving experience (indicative for the possibility to integrate several tasks and to behave in a supervisory control mode), task demands (tasks to be performed at the same time) and a deterioration of vision by comparing night and day driving. Two experiments were performed on the road to verify the structure of the model by the hypotheses on the three variables. Driver's monitoring behaviour, driver's control actions and overall performance were taken as critical variables. The monitoring behaviour was studied explicitly by asking drivers in one of both experiments voluntarily to occlude their visual information by looking downwards within the car, instead of at the road, for as long as possible.

1. INTRODUCTION

The features of manual and supervisory control can be used to describe the behaviour of car drivers in relation to their level of driving experience. Novice drivers behave in a strictly manual control mode and need full attention in performing and integrating the separate tasks (lateral vehicle control, longitudinal vehicle control, the detection and evaluation of other traffic participants, etc.) of driving. They do not as easily co-ordinate lateral and longitudinal control at the same time as experienced drivers (Blaauw et al., 1977) and perform each task in a strict controlled way. The integration of these tasks requires a substantial amount of driving experience (Kimball et al., 1971; Safren et al., 1970) and is the result of a decrease in attention in response to the individual task demands by performing each task more automatically. Then, driving behaviour can be said to be shifted towards a more supervisory control mode. Ultimately, supervisory control of very experienced drivers may be described by two aspects: The monitoring of individual tasks which are performed on a hierarchical lower mental level automatically, and the detection of failures in the traffic system, like unsafe behaviour of other traffic participants.

The manual and supervisory control mode can both thought to be based
on an internal representation drivers make of the overall task. This internal representation consists of estimates of the relevant aspects of the overall task with their associate variances, reflecting the driver's uncertainties about these aspects, and forms the intermediate between an element 'observation/prediction' and an element 'control' within a hypothetical model of driver behaviour (Fig. 1). This model is based on the concepts of the optimal control model (Baron and Kleinman, 1969; Kleinman, Baron and Levison, 1971) and was proposed earlier to describe human operator control of slowly responding systems (Kok and Van Wijk, 1978; White, 1980). In the application presented here, the structure of the model is adapted for fast responding systems (like automobiles).

The element 'observation/prediction' (Fig. 1) transforms the input information and estimates a representation of relevant system states with their associated uncertainties. In the original model the prediction was only implemented for the compensation of the inherent time delays of the human operator and the system to be controlled. It should be noted, however, that the present 'prediction' has to cover a much wider range in time and has to supply instantaneous and future values of the system states over a time interval \((k, k+K)\) in order to allow for perceptual anticipation in relation to driving experience. Cognitive anticipation is thought to be present in the internal model describing driver's knowledge of real system dynamics, lead variables (road to be followed, for example) and disturbances. A quantitative approach, however, to use the future values in the anticipation processes is not yet available, but the objective TLC-concept (time-to-line crossing) described by Godthelp and Konings (1981) and its subjective counterpart (as an element of the internal representation of the driver) are considered to be useful for a quantitative description of the future values.

The element 'control' forms a second essential part in the model which transforms the estimated internal representation via an optimization cri-
terion into the output vector reflecting driver's actions like steering wheel movements and acceleration activities. In the optimization criterion drivers evaluate the relevant aspects of driving and interpret specific task demands. Via knowledge of the car dynamics control actions are subsequently initiated in accordance with the desired change in overall performance due to the internal criterion.

Both these elements are active during the continuous process of car driving. In addition to these elements, however, an element 'decision' is incorporated as a discrete representation for supervisory activities. During supervisory control the estimated internal representation and the associated uncertainties are considered continuously which may result in discrete actions with respect to new observations and corrections. These actions stipulate the time moments on which the observations and corrections are made; the corresponding control amplitudes are set by the optimization criterion in the 'control' element. Drivers decide to make new observations on a specific variable when the corresponding internal uncertainty increases to too high a level or when the estimated value becomes critical in relation to the tolerated bounds. As a consequence, the internal representation is updated and allows non-visual control subsequently for a short time. New control actions are initiated when the estimated values deviate too much from the desired ones.

The structure of the proposed model is intuitively attractive for the description of supervisory control in relation to the level of driving experience and task demands, but needs some verification. Therefore, the specific effects of these experimental variables and its interactions under deteriorated visual circumstances, on driver's monitoring, driver's control and overall driver-car performance were predicted by using the structure of the model. Real data were subsequently gathered for these experimental variables with an instrumented car on the road in order to verify the predicted effects experimentally.

2. EXPERIMENTAL VARIABLES AND MODEL HYPOTHESES

From an experimental point of view in testing the model structure, experimental variables have to be selected which only affect one single element ('observation/prediction' or 'control') of the model. In contrast to the level of driving experience and the task demands only the deterioration of vision seems to suit this requirement completely.

Deteriorated vision refers to conditions of reduced visual circumstances under which the element 'observation/prediction' will result in estimates (the internal representation) with larger uncertainties. Consequently, drivers need more observations in maintaining a sufficiently accurate estimate as a function of time. Control actions, however, are based on these internal estimates and are not affected due to the external deteriorated vision. Overall system performance is also unaffected unless the observations cannot establish a sufficiently accurate internal representation. During the experiments the deterioration of vision was obtained by comparing night and day driving. At night, especially the estimation of future values of the relevant states (perceptual anticipation) is more difficult due to reduced preview.

An increase in driving experience results in a better internal model in the element 'observation/prediction' (cognitive anticipation) and enables a more accurate internal estimate of the actual states. Consequently, experienced drivers need less observation than inexperienced ones. At the other hand, the element 'control' is also influenced via an increasing
knowledge of system dynamics, resulting in more efficient control actions in relation to the desired change in overall performance, and via the internal criterion for the weighting of the estimated variables. It is expected that experienced drivers know better the maximum tolerated values of the output variables and will give less weight to small variations in these variables. Both reasons will probably result in fewer control actions for the experienced drivers.

Task demands can be varied experimentally by selecting different number of tasks to be performed at the same time, or by changing the tolerated boundaries in each individual task. The primary task of lateral vehicle control was extended by longitudinal vehicle control as another substantial part of the overall driving task. Previous research (Blaauw, 1980), with a between-subjects design, showed no differences in lateral and longitudinal performance in relation to the instructed tolerances for the lateral position or velocity respectively. A significant effect, however, was found in lateral control as the result of an instructed limitation for the boundaries for longitudinal control. Therefore, task demands in the present experiment were varied in accordance with these findings and only longitudinal control was instructed on two levels: a free condition (velocity 80 - 120 km/h) and a condition to drive with a constant velocity (100 km/h). The task demands will affect both elements of the model. From the previous research it is known that the 'control' element is affected differently for both groups of driving experience. The constant velocity condition will result in larger deviations in lateral position (worse performance) for the inexperienced drivers in relation to the free velocity condition, while the experienced drivers will obtain a better performance. Because of the more restricted boundaries the driver requires a more accurate estimate for his internal representation and, consequently, more observations on the speedometer can be expected during the constant velocity condition. The observations on the variables for lateral vehicle control will decrease or remain on the same level when the constant velocity condition is introduced.

3. EXPERIMENTAL VERIFICATION

The verification of the model structure is related to the specific effects of the three experimental variables on driver's monitoring, driver's control and overall driver-car performance during real driving. Although driver's control and overall performance can easily be measured during normal driving, some problems arise for the analysis of critical monitoring behaviour in relation to the internal uncertainties. In most normal situations the driver has more visual information than he needs to perform satisfactorily; drivers have considerable spare visual capacity and eye fixations may be merely seen as reference points for the organisation of peripherally acquired information (Rockwell, 1972). Consequently, well-known techniques like the recording of eye movements during normal driving are not likely to achieve the desired monitoring data and a modified technique has to be chosen. In accordance with Senders et al. (1966), Farber and Gallagher (1972), and Triggs and Caple (1978) a visual occlusion technique was used in a separate, second experiment. Drivers were asked to occlude voluntarily visual information by looking downwards in the automobile, instead of at the road or speedometer, for as long as possible. Only when their internal uncertainty on a specific variable became too high, they were allowed to look again for as short as possible in order to update their internal representation.

Two experiments were performed. The first experiment was focussed on
driver's control and overall performance during normal driving with the three selected experimental variables. Driver's monitoring was studied in the second experiment (with identical experimental variables) with the special occlusion technique. Both experiments were similar in all other details.

Subjects
Forty-eight male subjects (Ss) took part in each experiment, 24 experienced drivers and 24 inexperienced drivers. The experienced drivers had their license for at least three years and a driving experience of at least 30,000 km; they had a car of their own. The inexperienced drivers had either followed a driver training course or had just passed their driving test; they had no car. All Ss were between 18 and 29 years of age. They were paid for their services.

Procedure
Each driver had to drive in an instrumented vehicle, a Volvo 145 Express, during a one-hour session on a straight section of a four-lane highway with divided traffic lanes, having a constant road geometry and lane width (3.60 m). Ss did not receive any training with the car before the start of the experimental runs. The initial position was in the emergency lane, where Ss were instructed how to handle the car. The Ss then, accelerated to the desired speed and changed to the correct lane. In the subsequent one-hour period the driver's observation and control behaviour was measured regularly, during six intervals in which interacting traffic was absent. There were no disturbances such as sidewind gusts. Halfway during the highway the Ss had to return to the initial position.

During all runs two experimenters (Es) were present in the instrumented vehicle. One E took care of the apparatus while the other E instructed the S. The latter E was a driver-training instructor to legalize the runs with the inexperienced drivers.

The experimental conditions (night and day, inexperienced and experienced drivers, free and constant velocity) were varied between the 48 Ss and, consequently, each condition was performed by a separate group of six Ss. Each S drove his specific condition in the one-hour session. From the group of six Ss each day, three Ss drove during daytime (10.30 a.m. - 2.30 p.m.), while the remaining three Ss did the runs at night (8.00 p.m. - midnight). There was no public lighting during the latter runs; the car had dipped headlamps.

Before the run Ss were instructed about the task demands for longitudinal vehicle control (free velocity or to drive with the constant velocity of 100 km/h), about the required voluntarily occlusion technique (in the second experiment), and about some additional procedures. With respect to the occlusion technique drivers were instructed to occlude voluntarily their visual information during the measurement intervals by looking downwards within the car, instead of at the road, for as long as possible. Peripheral vision was occluded as well in that situation. When drivers needed to update their internal representation on a specific variable, i.e. when their uncertainty became too high, they were allowed to look at the road or the speedometer, for as short as possible. The experimenter supervised driver's observation behaviour and could interrupt because of traffic safety reasons. Ss were urged, however, to observe in a manner such that any interference by the experimenter would be unnecessary. Additional instructions were given in order to drive in the right lane most of the time, to allow overtaking for short periods only, and to drive at normal highway speed (80 - 120 km/h) during the free velocity condition.
Data analysis

During each run steering wheel angle, lateral position (distance between driver's eyes and the right lane marking), yaw rate, position of the accelerator and velocity were sampled over 6 periods of 32 s each, with a sampling frequency of 4 Hz. The samples were recorded on floppy-disc for subsequent computer-processing. Mean values and standard deviations were taken as dependent variables.

Driver's monitoring behaviour in the second experiment was recorded on videotape by means of a dashboard-mounted television camera looking at the head and eyes of the driver. During the night condition infrared illumination was applied in order to obtain a good video picture with the infrared-sensitive camera. The monitoring behaviour was analysed off-line quantitatively by means of a special computer-controlled video plotting device. The beginning and the end of each period as well as the type of information (downwards within the car, road ahead, speedometer) were taken as dependent variables.

All dependent variables were analysed by Analysis of Variance (ANOVA) and subsequent Newman-Keuls tests, in order to investigate whether any main effects of the experimental conditions or interactions may have occurred by chance or not (Winer, 1962). The experimental conditions for the two experiments are summarized once in their hierarchical order:

- 2 levels of vision deterioration (night and day)
- 2 groups of driving experience
- 2 task demands with respect to longitudinal vehicle control
- 6 Ss for each level of vision deterioration, group of driving experience and task demand
- 6 intervals during each run.

4. RESULTS

The analysis of driver's monitoring, driver's control and overall driver-car performance of both experiments is ongoing at this time, and therefore only intermediate results with respect to mean values and standard deviations of the several dependent variables are given in relation to the three experimental variables and the model hypotheses. Driver's control actions and the overall performance are derived from the first experiment, while the monitoring behaviour is the result of the second experiment. Because of the absence of any major learning effects over the six measurement intervals, the results are given as mean values over the intervals. Significant differences between individual drivers are ignored.

4.1 Deterioration of vision: night versus day driving

Table I presents the hypotheses and experimental results for the effect of the deterioration of vision achieved during night driving, in relation to driving during daytime. The results are focussed on driver's monitoring behaviour, driver's control actions and overall performance. The hypotheses of Table I were already given in paragraph 2. The experimental results are shown for the free velocity condition.

The actual monitoring behaviour confirms an increase of observation time during the night condition for both groups of experience. The increase in the overall percentage of time is about equal for the inexperienced and experienced drivers. With respect to the mean observation time it is no-
Table I  Experimental verification of the effect of deteriorated vision (night versus day) on the proposed structure of the model. The results are given for the free velocity condition (S.D. = standard deviation).

<table>
<thead>
<tr>
<th>MONITORING</th>
<th>HYPOTHESES</th>
<th>EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>percentage of time on road ahead:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp. drivers 35.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp. drivers 29.6</td>
</tr>
<tr>
<td></td>
<td>at night larger internal uncertainties resulting in more observation time for all drivers.</td>
<td>at night significant (p &lt; 5%) more time on road ahead; not different for both groups of drivers.</td>
</tr>
<tr>
<td></td>
<td>due to a worse internal model the inexperienced drivers will need more observation time than the experienced drivers.</td>
<td>mean observation time on road ahead:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>day</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp. drivers 2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp. drivers 2.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>at night the experienced drivers tent to look at the road for longer periods of time.</td>
</tr>
</tbody>
</table>

| CONTROL ACTIONS | no effects. |
| CONTROL ACTIONS | S.D. steering wheel movements: |
| no effects. | day  | night  |
| inexp. drivers 1.5 | 1.5° |
| exp. drivers 1.6 | 1.1° |
| significant (p < 5%) differences for the exp. drivers. |

| PERFORMANCE | - no effects, unless (maximum) monitoring results in an insufficient internal representation. |
| S.D. lateral position: | day  | night  |
| inexp. drivers 16.5 | 21.0 cm |
| exp. drivers 21.2 | 12.7 cm |
| significant (p < 5%) differences for the exp. drivers. |
| mean lateral position: | day  | night  |
| inexp. drivers 179.5 | 217.1 cm |
| exp. drivers 173.1 | 216.3 cm |
| at night both groups drive significantly (p < 0.1%) more to the left within the lane. |
| mean velocity: | day  | night  |
| inexp. drivers 109.6 | 104.3 km/h |
| exp. drivers 110.5 | 106.3 km/h |
| at night both groups drive significantly (p < 5%) slower. |
| S.D. velocity: | no significant differences. |

Ticled that the inexperienced drivers tend to have longer observations at night, while the observation times do not change for the experienced drivers between night and day conditions. With respect to driver's control actions and overall driver-car performance the inexperienced drivers show no differences in the standard deviations of the steering-wheel movements and the
Table II  Experimental verification of the effect of driving experience on the proposed structure of the model. The results are mainly given for the free velocity condition (S.D. = standard deviation).

<table>
<thead>
<tr>
<th>HYPOTHESES</th>
<th>EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MONITORING</strong></td>
<td>- percentage of time on road ahead:</td>
</tr>
<tr>
<td>- inexperienced drivers have a worse internal model and need therefore more observation time than the experienced drivers.</td>
<td>- inex.</td>
</tr>
<tr>
<td></td>
<td>night</td>
</tr>
<tr>
<td>- inexperienced drivers spent significant (p ≤ 5%) more time on road ahead.</td>
<td></td>
</tr>
<tr>
<td>- mean observation time on road ahead:</td>
<td>- inex.</td>
</tr>
<tr>
<td></td>
<td>night</td>
</tr>
<tr>
<td>- inexperienced drivers have significant (p ≤ 1%) longer periods of time on the road.</td>
<td></td>
</tr>
<tr>
<td><strong>CONTROL ACTIONS</strong></td>
<td>- S.D. steering wheel movements:</td>
</tr>
<tr>
<td>- inexperienced drivers have a worse knowledge of system dynamics and maximum acceptable boundaries: larger, inefficient actions for the inexperienced drivers.</td>
<td>- inex.</td>
</tr>
<tr>
<td></td>
<td>night</td>
</tr>
<tr>
<td>- inexperienced drivers show significant (p ≤ 5%) larger movements at night; no differences during daytime. For the constant velocity condition significant larger movements for the inexperienced drivers for both night and day.</td>
<td></td>
</tr>
<tr>
<td><strong>PERFORMANCE</strong></td>
<td>- S.D. lateral position:</td>
</tr>
<tr>
<td>- two reverse effects in lateral position: 1) inexperienced drivers show larger variations due to restricted knowledge of system dynamics; 2) inexperienced drivers show smaller variations due to small accepted tolerances in position variation.</td>
<td>- inex.</td>
</tr>
<tr>
<td></td>
<td>night</td>
</tr>
<tr>
<td>- significant (p ≤ 5%) larger variations for the inexperienced drivers at night; no differences during daytime. For the constant velocity condition significant larger variations for the inexperienced drivers for both night and day.</td>
<td></td>
</tr>
<tr>
<td>- mean lateral position: no significant differences.</td>
<td></td>
</tr>
<tr>
<td>- mean velocity:</td>
<td>- inex.</td>
</tr>
<tr>
<td></td>
<td>night</td>
</tr>
<tr>
<td>- no significant differences.</td>
<td></td>
</tr>
<tr>
<td>- S.D. velocity: no significant differences.</td>
<td></td>
</tr>
</tbody>
</table>
lateral position of the car; these results confirm the proposed hypotheses. The experienced drivers, however, show a significant decrease in both standard deviations during nighttime. With respect to the overall performance some other significant changes are obtained: At night all drivers shift about 40 cm to the left in their lane and they slow down. The exact reason for the shift to the left is not known, but it is suggested that drivers indicate some adaptation mechanism at night because 1) smaller distances to a reference line result in higher perspective sensitivities for the detection of lateral position variations and 2) a position more to the left in the lane gives more preview due to the asymmetrical European headlamp beam.

4.2 Level of driving experience

The effects of the level of driving experience on the three aspects of the model structure are given in Table II. The results are mainly given for the free velocity condition. The standard deviations of the steering wheel movements and the lateral position on the road, however, showed some significant interactions with the task demands for velocity control; these results are also given in Table II.

The monitoring behaviour confirms a higher percentage of time on the road ahead and longer mean observation periods for the inexperienced drivers during night and day. The mean observation period at night tends to increase to a larger extent for the inexperienced drivers than for the experienced drivers in relation to the daytime. Driver's control actions are larger for the inexperienced drivers during all conditions with an exception for the most easy condition (daytime, free velocity) resulting in comparable steering wheel variations for both groups of driving experience. Overall performance shows similar differences: Comparable position variations during daytime and the free velocity condition, and larger deviations for the inexperienced drivers in all other conditions than the inexperienced drivers. During daytime and the free velocity condition it is obvious that the inexperienced drivers accept smaller tolerated boundaries for the lateral position variations than the experienced drivers and reach comparable deviations. At night and during the constant velocity condition lateral vehicle control deteriorates for the inexperienced drivers, while the experienced drivers show an inverse effect and perform with smaller variations.

4.3 Task demands: free versus constant (100 km/h) velocity

Table III summarizes the results of the differences in both task demands for the monitoring behaviour, control actions and overall performance of lateral vehicle control; the corresponding results for longitudinal vehicle control are presented in Table IV.

Due to the additional constant velocity condition the monitoring behaviour shows shorter mean observation time on the road than during the free velocity condition; a trend is present in the total percentage of time spent on the road. From the results of Table IV it is shown that the shorter observation times during the constant velocity condition are adjusted to about the level for the free velocity condition when the mean observation times on the speedometer are added. Table IV confirms that both groups of drivers observe their speedometer to a larger extent during the condition to drive with a constant velocity. The overall percentage of
Table III  Experimental verification of the effect of task demands (free versus constant velocity) on the proposed structure of the model for lateral vehicle control (S.D. = standard deviation).

<table>
<thead>
<tr>
<th>MONITORING</th>
<th>EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>HYPOTHESES</td>
<td>EXPERIMENTAL RESULTS</td>
</tr>
<tr>
<td>- in the constant velocity condition it is necessary to monitor the speedometer additionally and, consequently, the monitoring on lateral control variables will decrease or, at best, remain on the same level.</td>
<td>- percentage of time on road ahead:</td>
</tr>
<tr>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td>inexp.: day 35.2 31.7 %</td>
</tr>
<tr>
<td></td>
<td>night 38.7 37.0 %</td>
</tr>
<tr>
<td></td>
<td>exp.: day 29.6 24.3 %</td>
</tr>
<tr>
<td></td>
<td>night 34.9 31.4 %</td>
</tr>
<tr>
<td></td>
<td>trend to shorter time.</td>
</tr>
<tr>
<td></td>
<td>mean observation time on road ahead:</td>
</tr>
<tr>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td>inexp.: day 2.5 1.6 s</td>
</tr>
<tr>
<td></td>
<td>night 3.1 2.7 s</td>
</tr>
<tr>
<td></td>
<td>exp.: day 2.0 1.4 s</td>
</tr>
<tr>
<td></td>
<td>night 2.1 1.9 s</td>
</tr>
<tr>
<td></td>
<td>in the constant velocity condition significant (p&lt;5%) shorter observations.</td>
</tr>
</tbody>
</table>

| CONTROL ACTIONS |
| S.D. steering wheel movements: |
| free constant |
| inexp.: day 1.5 1.7 |
| night 1.5 1.8 |
| exp.: day 1.6 1.3 |
| night 1.1 1.3 |
| no significant differences. |

| PERFORMANCE |
| S.D. lateral position: |
| free constant |
| inexp.: day 16.5 22.3 cm |
| night 21.0 28.5 cm |
| exp.: day 21.2 11.3 cm |
| night 12.7 15.5 cm |
| in the constant velocity condition show the inexperienced drivers significant (p<5%) larger variations during night and daytime; the experienced drivers show significant (p<5%) smaller variations during daytime; at night comparable variations. |
| mean lateral position: |
| no significant differences. |

| mean velocity/S.D. velocity: |
| see Table IV. |
Table IV Experimental verification of the effect of task demands (free versus constant) velocity on the proposed structure of the model for longitudinal vehicle control (S.D. = standard deviation).

<table>
<thead>
<tr>
<th>MONITORING</th>
<th>HYPOTHESES</th>
<th>EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- in the constant velocity condition more observation time on the speedometer; inexperienced drivers even need more observation time than the experienced drivers.</td>
<td>- percentage of time on speedometer:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp.: day 0 8.6 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0 4.9 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp.: day 0 3.0 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0 4.7 %</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in the constant velocity condition significant (p ≤ 0.1%) higher percentage of time on speedometer; inexperienced drivers tend (p ≤ 8%) to look more than the experienced drivers during daytime.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mean observation time on speedometer:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp.: day 0 0.8 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0 0.5 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp.: day 0 0.5 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0 0.5 s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in the constant velocity condition significant (p ≤ 0.1%) longer observations on the speedometer.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D. accelerator:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no significant differences.</td>
</tr>
<tr>
<td>CONTROL ACTIONS</td>
<td>- in the constant velocity condition more accelerator activities due to higher weightings to velocity variations.</td>
<td>- mean velocity:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp.: day 109.6 102.4 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 104.3 99.7 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp.: day 110.5 107.8 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 106.3 102.0 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in the constant velocity condition significant (p ≤ 5%) lower velocity.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D. velocity:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>free constant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>inexp.: day 1.0 0.8 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0.9 0.8 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>exp.: day 1.4 0.7 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>night 0.8 0.8 km/h</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no significant differences.</td>
</tr>
</tbody>
</table>
time and the mean observation time on the speedometer both increase. There is a tendency that the inexperienced drivers even need a higher percentage of time on the speedometer than the experienced driver; the mean observation time is not differently for both groups of experience.

With respect to the control actions and performance in lateral vehicle control some effects are postulated based on previous research (Blaauw, 1980). Although no effects are shown in the control actions of the driver, the differences in overall performance appear to be present roughly as hypothesized: Experienced drivers show smaller variations, and inexperienced drivers larger variations when the constant velocity condition is introduced in relation to the free velocity condition. There is, however, an exception for the experienced drivers at night.

With respect to the control actions and performance in longitudinal control no confirmation is found for the hypothesized effects (Table IV), although the mean velocity approaches the desired one better during the constant velocity condition than the free velocity condition. The variations in the velocity are not smaller when drivers are instructed to drive with a constant velocity. The result confirms earlier data (Blaauw, 1980) about the absence of a decrease in variations of lateral position or velocity when drivers, in a between-subjects design, are instructed to drive as straight as possible or with a constant velocity, respectively. There is, however, again an effect from the constant velocity condition towards lateral vehicle control. Significant differences do occur, at the other hand, as a result of changes in instructed boundaries in a within-subjects experimental design (Blaauw et al., 1977).

5. DISCUSSION

In general the results of this experimental verification confirm most of the hypotheses based on the theoretical structure of the model with respect to driver's monitoring, driver's control actions and overall driver-car performance. Consequently, the model with the internal representation as intermediate between the elements 'observation/prediction' and 'control', offers a promising structure for the description and prediction of the effects of experimental variables. The results in this paper already give valuable insight into the effects of the level of driving experience, task demands and deterioration of vision. More detailed theoretical analysis is needed, however, to study the hypotheses which were not confirmed experimentally.

The verification of the structure of the model, however, is based on interim results concerning mean values and standard deviations of the relevant variables. It is planned to extend the analyses towards a critical study of the decline in the estimated variables of driver's internal representation, with the decline in the associated variances (uncertainties), as a function of time after the last observation and correction in order to predict the moments in time for new observations and corrections. In this approach it is thought to establish a link towards the series of uncertainty models, as proposed for example by Ceder (1977) for car-following situations. As mentioned before, it is planned to implement the time-to-line crossing concept (Godthelp and Konings, 1981) in these extended analyses.

In addition to this experimental verification on the road and the subsequent data analyses it is also suggested to implement the structure of the model in a mathematical computer model for the driver and, in combination with a mathematical computer model for the system to be controlled, to verify the validity of the computer model for the three mentioned experi-
mental variables in relation to the corresponding results of the experiments on the road.

REFERENCES


PSYCHO-MATHEMATICAL MODEL OF VEHICULAR GUIDANCE
BASED ON FUZZY AUTOMATA THEORY

U. Kramer
G. Rohr

Institute of Automotive Engineering
Technische Universität Berlin

ABSTRACT

Within the system driver - vehicle - environment the visual-motor channel represents the most important component as to mastering the driving tasks. This channel can be characterized by perceptual complexity and behavioural dynamics. This report aims to present a conceptual framework considering both aspects complexity and dynamics in like manner. Thereby - first of all - two approaches are unavoidable: 1. description of reaction dynamics by use of quantitative-mathematical methods; 2. description of information processing complexity with the aid of formal logic (i.e. by a set of verifiable statements). Establishing fuzzy systems to describe the dynamics, it is possible to avoid too great discrepancies between these two levels of description. For this state of the visual-motor system the experimental design will be presented. Furthermore, it will be shown that the chosen mathematical approach, based on fuzzy sets theory, eases an advanced formalization of the vehicular guidance model.

I. INTRODUCTION

In the last years many publications dealt with modelling the human reaction sequences during manual control tasks. These works focussed mainly the description of reaction dynamics by use of differential and difference equations respectively. External disturbances as well as inadequately modelled system components were described by corresponding MARKOVian processes (e.g. ref. /1/). In these models the perception of environmental stimuli initiating reactions is usually reduced to the observation of certain values which are especially comfortable in mathematical handling. This procedure is sufficient as long as the mathematical description of stimuli corresponds to the test person's observation, for example reading off instruments. By trying to apply these methods to vehicular guidance the problem arises that the observation of relevant stimuli according to typical driving tasks is a complex process. Would this process be described with the same precision as the reaction dynamics it could be stated that significance and precision of
information obtained by such a model become mutually excluding features: mathematical apparatus and analyzed object are incompatible (ref. /2/).

Both aspects dynamics and complexity are taken for essential features of vehicular guidance and, therefore, have to be expressed by the model. By describing and explaining complex phenomena (e.g. in Psychology) usually systems of verifiable statements (verbal models) are employed. These verbal models, however, are hardly understandable even by simple dynamic processes, and they are so far less appropriate than formal models.

To take full advantage of both model types a vehicular guidance model is proposed which contains not only formal description of reaction dynamics but also verbal description of perceptual processes. Because of the finite number of logically connected expressions it is advisable to make use of finite automata describing dynamics. Uncertainties are modelled by fuzzy sets, the according formal model is called fuzzy automaton (ref. /3/).

This research covers two objectives: 1. description of visual motor system within vehicular guidance with the aid of partly logical expressions, partly fuzzy dynamic model (psycho-mathematical model) including experimental design; 2. development of a partial formalization of perception system, compatible with the fuzzy dynamic reaction model, and demonstration of its expressiveness by means of a field study taken from literature.

II. MODEL OF VISUAL MOTOR SYSTEM

1. Structure

Accomplishing typical driving tasks (lane control; speed adaptation; reaction to obstacles, traffic signs, other road users, etc.) is essentially a process of visual motor coordination. Our research is, therefore, restricted to the visual-motor system. First of all, it must be found which parameters of the ambient visual field are selected as references by the driver. This transformation of environmental stimuli to an internal representation includes a preprocessing as well as a feature extraction. The internal representation has to be classified by a pattern recognition process, i.e. by comparing with already stored internal prototypes to serve as basis of a decision strategy. The decision strategy for appropriate action is interpreted as a dynamic process since it depends on classified internal representation of environmental stimuli and on the preceding behaviour.

Based on the results of different authors (ref. /4/) it can be assumed that the visual system has attributes of a spatial frequency analyzer. It can also be supposed that frequency analyzes in the horizontal and vertical ranges are dominant
within the human visual system (refs (5,6/).

This kind of processing has the advantage of a parallel coverage of relevant information in the visual environment. This way, several orientation tasks can be mastered in a parallel manner and herewith partially automatized, and capacity for the search of singularly occurring events remains.

As mentioned before, the vehicular guidance is characterized by a number of those tasks. The driver has to control continuously the lane keeping as well as his speed which requires a permanent orientation to lane direction and actual speed. Moreover, he needs additional capacity for apprehension of and reaction to singularly occurring events like obstacles, traffic signs, etc.

This is mostly guaranteed if preprocessing and feature extraction proceed in the before-mentioned manner of parallel filtering. Thereby, we postulate the following assumptions (ref. /7/):

1. lane control according to low spatial frequency components
2. speed control according to high spatial frequency components

of the visual field.

Consequently, the pattern recognition processes have to be assumed working parallel referred to several orienting stimuli. Their interconnected classification results can be interpreted as input variables of the fuzzy dynamic decision unit. The connection can be modelled by fuzzy relations being composed of fuzzy constituents. As will be shown later, pattern processing stages can be mapped also into fuzzy relations.

2. Formal Model of Decision Strategy

The conversion of analyzed visual information to action sequences is performed by a series of decision processes. In principle, these processes contain an assignment from input to output variables, and represent a special case of general input-output-systems \( \Sigma \{X, Y, S\} \) with input set \( X \), output set \( Y \), and relation \( S \subseteq X \times Y \). It is a characteristic of decisions to choose one element from a set of a finite number of alternatives on the basis of actual events as well as certain a-priori-information. Thereby, the actual events are the input variables, the a-priori-information (e.g. instruction, experience) can be interpreted as global states, and the decisions generate the action sequences as output signals. An appropriate model of decision behaviour means to regard \( X \) and \( Y \) respectively as finite sets. The relation \( S \) has to reflect the ambiguity of the assignment from input to output variables. One feasible description of \( S \) is the introduction of \( A \) as a fuzzy set on \( X \times Y \).

Formally a fuzzy set \( A \) is given by the elements \( u \) of a basis set \( U \) (universe) and a function \( \mu_A(u) \) which indicates the
degree of membership of \( u \) in \( \mathcal{A} \) with values between 0 and 1:

\[
\mathcal{A} = \{(u, \mu_{\mathcal{A}}(u)) \mid u \in U, \mu_{\mathcal{A}}(u) : \mathcal{A} \rightarrow [0,1]\}
\]

whereby \( \mu_{\mathcal{A}}(u) \) means that \( u \) is not an element of the set \( \mathcal{A} \),
while \( \mu_{\mathcal{A}}(u) = 1 \) indicates that \( u \) is definitely an element of
the set \( \mathcal{A} \). Conjunctions of the fuzzy sets \( \mathcal{A}, \mathcal{B} \subseteq U \) are realized
by forming

\[
\begin{align*}
\mathcal{A} \cup \mathcal{B} & : \leftrightarrow \mu_{\mathcal{A} \cup \mathcal{B}}(u) = \max \{\mu_{\mathcal{A}}(u), \mu_{\mathcal{B}}(u)\} \\
\mathcal{A} \cap \mathcal{B} & : \leftrightarrow \mu_{\mathcal{A} \cap \mathcal{B}}(u) = \min \{\mu_{\mathcal{A}}(u), \mu_{\mathcal{B}}(u)\} \\
\mathcal{A}' & : \leftrightarrow \mu_{\mathcal{A}'}(u) = 1 - \mu_{\mathcal{A}}(u)
\end{align*}
\]

for \( U, \mathfrak{n} \) and the complement \( ' \). Regarding these operations the
set \( \mathcal{F}(U) \) of the fuzzy subsets of \( U \) forms a distributive
lattice with pseudo-complement.

The decision process

\[
\Sigma \{X, Y, \mathcal{S}\} \text{ with } \mu_{\mathcal{S}}(x_i, y_j)
\]

will be interpreted in the following manner: it represents the possibility by which the deciding person thinks the reaction

\[
y_j \in Y \ (j = 1, \ldots, n)
\]

appropriate to the according occurring event

\[
x_i \in X \ (i = 1, \ldots, n).
\]

In general, input and output variables are also fuzzy subsets \( \mathcal{A} \)
and \( \mathcal{B} \) of the universes \( X \) and \( Y \) respectively with corresponding
membership functions. Conjunctions between input and output
variables by means of \( \mathcal{S} \) can be formed by the composition

\[
\mathcal{B} = \mathcal{A} \circ \mathcal{S}
\]

By introducing the fuzzy relation \( \mathcal{S} \) as matrix

\[
\mathcal{S} = \begin{bmatrix}
\mu_{\mathcal{S}}(x_1, y_1) & \ldots & \mu_{\mathcal{S}}(x_1, y_n) \\
\vdots & \ddots & \vdots \\
\mu_{\mathcal{S}}(x_m, y_1) & \ldots & \mu_{\mathcal{S}}(x_m, y_n)
\end{bmatrix}
\]

and the fuzzy signals \( \mathcal{A} \) and \( \mathcal{B} \) as vectors

\[
x^T = | \mu_{\mathcal{A}}(x_1) \ldots \mu_{\mathcal{A}}(x_m) |, \ y^T \text{ acc.}
\]

eq (1) can be represented similar to a usual matrix product

\[
y = \mathcal{S} \times x
\]

where max- and sup-operation respectively has to be applied for
addition, and min-operation for multiplication.

By representing the decision process as a fuzzy dynamic system
the fuzzy relation \( \mathcal{R} \subseteq X \times Y \times Y \) will be
\[ B(t+1) = \{A(t) \times B(t)\} \circ R_1(A(t),B(t),B(t+1)) \quad (3) \]

indicating the connection between the fuzzy subsets \(A(t) \in X\), \(B(t) \in Y\) and \(B(t+1) \in Y\) \((t: \text{discrete time})\) (ref. /8/).

As a special case a feedback structure with fuzzy output intersection can be derived with

\[
\begin{align*}
R_1 &= \begin{bmatrix}
\mu_{B_1}(x_1(t),y_1(t+1)) & \cdots & \mu_{B_1}(x_1(t),y_n(t+1)) \\
\vdots & \ddots & \vdots \\
\mu_{B_1}(x_m(t),y_1(t+1)) & \cdots & \mu_{B_1}(x_m(t),y_n(t+1))
\end{bmatrix} \\
R_2 &= \begin{bmatrix}
\mu_{B_2}(y_1(t),y_1(t+1)) & \cdots & \mu_{B_2}(y_1(t),y_n(t+1)) \\
\vdots & \ddots & \vdots \\
\mu_{B_2}(y_n(t),y_1(t+1)) & \cdots & \mu_{B_2}(y_n(t),y_n(t+1))
\end{bmatrix}
\end{align*}
\]

and \(x(t), y(t), y(t+1)\) in a corresponding vector notation of the fuzzy sets \(A(t)\), \(B(t)\), and \(B(t+1)\), we have

\[ y(t+1) = R_1 x(t) \cap R_2 y(t) \quad (4) \]

which can be extended directly to:

\[ y(t+1) = \{R_1 x(t) \cap R_2 x(t-1)\} \cap \{S_1 y(t) \cap S_2 y(t-1)\} \]

The driver's motor activities (turning the steering wheel, pedal application) will be regarded as output variables of the fuzzy dynamic system, while the contents of the ambient visual field, coded in an appropriate manner, can be regarded as (possibly fuzzy) input variables. The membership values in the fuzzy matrices, for example \(R_1\) and \(R_2\) in eq. (4), can be determined in experiments; it seems suitable to express these membership values in each column by characteristic values of given distribution functions (e.g. bias and dispersion). Changing test conditions these characteristic values will be regarded as random variables and submitted to statistical examinations, according to a part of the subsequent frame of hypotheses.

### 3. Verbal Model of Pattern Processing

The complex of model assumptions on the verbal level can be split up into two groups. The first group (hypotheses 1 and 2) is related to the general thesis of parallel pattern processing, especially to filter processes derived therefrom during lane keeping and speed control. The second group (hypotheses 3 and 4) shall clarify the requirements which have to be met by a formal perceptual model of pattern processing in addition to the general thesis. This second group mainly concerns the transformation of actual patterns into an internal representation and its relation to reaction behaviour as well as foveal fixation.
Hypothesis 1 (general):
Certain spatial frequency components of the ambient visual field contribute differently to the various tasks of vehicular guidance (lane keeping, speed control), whereby the components are processed in a parallel manner and taken from differently preferred areas of the retina (foveal, parafoveal).

Experimental specification:
a) If a visual field is given with only lower spatial frequencies lane keeping can be controlled easily but speed estimation is very difficult.
b) If a visual field is given with dominating higher spatial frequencies speed can be estimated quite accurate but the ability for lane keeping turns down.
c) Both conditions a) and b) lead to speed reduction by the driver in contrary to a visual field containing both spatial frequency components.
d) If the high spatial frequency components are located in the peripheral visual area speed control is easier than these frequency components being in the frontal visual area.

Hypothesis 2 (general):
Sudden changes in the range of high spatial frequencies lead to a change in speed estimation and affect adaptive reactions according to the modified speed estimation.

Experimental specification:
If speed has to be controlled on a constant level, pedal application is related to sudden changes of the high spatial frequency components in the parafoveal visual area, i.e. the increasing spatial frequencies in the given visual field, the decreasing driven speed, and vice versa.

Hypothesis 3 (general):
The more the static pattern of the visual field is overlaid by high frequency disturbances, the vaguer is the assignment of image elements to the category "street". The so estimated assignments, expressible by membership functions, depend on the filter characteristics of pattern preprocessing.

Experimental specification:
If a course of certain width and curvature with a specific degree of disturbance by overlaying high frequency components is exposed as reference pattern, and if this course has to be estimated in pared comparison according to a course of different width and identical curvature being "equal" or "unequal" respectively, the certainty of this judgement - explicitly demanded - depends on the degree of disturbance. The distribution of the membership function over the image pattern as well as filter function of the preprocessing can be determined by these gained values.

Hypothesis 4 (general):
The information of that image area fixated foveally determines
the decision for the steering behaviour. This point from the overall pattern is elaborated thereby which yields maximum information. This happens with the aid of internally represented patterns, acquired by experience, and in permanent comparison with those. Then the steering movement has a fuzzy relation to the foveal fixation.

Experimental specification:

a) If a course is presented as an undisturbed pattern, the steering angle has a fuzzy relation to foveal fixation whereby the horizontal coordinate determines the amplitude and direction of steering movement and the vertical coordinate the latency between fixation and motor reaction.

b) For the area, foveally fixated at different courses, one pattern of internally stored representation can be found which will produce just this area of foveal fixation with maximum information resulting from comparison of internally stored and actual pattern.

c) The less unique the fixation behaviour is at undisturbed course pattern (frequent saccades with small amplitude), the more uncertain the steerings will become (angular changes with small amplitude).

4. Formal Approach to Pattern Processing

The stimuli of the ambient visual field to be processed by the driver for lane keeping are objects $o^j_t$, generally embedded in a space-time-continuum:

$$o^j_t = f^j(x,y,z,t)$$  \hspace{1cm} (5)

These objects are mapped onto retina

$$0^j_0 = T \{o^j_t\} = g^j(\xi,\eta,t)$$  \hspace{1cm} (6)

and are treated subsequently. The stages of the internal process are outlined already. The interface within the internal processing from which our formal approach is proceeding is assumed being between preprocessing and feature extraction.

It is supposed that there are information assigning certain sub-areas of the two-dimensional overall pattern belonging to the category "street" or not. According to a low spatial frequency filter - not yet determined experimentally - the coordinates $(\xi,\eta)$ at an instant $t$ with a membership value $\mu_{Q^t}(\xi,\eta,t)$ are assigned to the object $Q^t$ ("street"), i.e. $Q^t$ is regarded as a fuzzy subset of the overall pattern.

$$Q^t = \{(\xi,\eta,t), \mu_{Q^t}(\xi,\eta,t) | (\xi,\eta,t) \in X \times Y \times T; \mu_{Q^t}(\xi,\eta,t): Q^t + <0,1>\}$$  \hspace{1cm} (7)

In case of discrete sets $X, Y, T$ the preprocessed pattern $Q^t$ of the object street $o^j_t$ at $t$ can be described as a fuzzy matrix:
The evaluation of entries of \( O_t \) with values between 0 and 1 depends on "signal/noise-ratio" of original pattern and on filter process performed by the preprocessing stage.

At first, the subsequent processing system has to extract those features from the membership distribution in matrix \( O_t \) which affect directly the drivers' steering movements. Furthermore, it can not be assumed that the parallel filtering processes are able to induce the isolation of the individual object \( O_j \). That is why it is necessary to identify the significance of these objects, realized by the stages succeeding the preprocessing (essentially by pattern recognition). Finally, a further requirement is the possibility to determine the foveal fixation with the aid of internally represented object patterns. That means that a range within the preprocessed pattern can be derived which is distinguished by certain characteristics (as minimum uncertainty or maximum information according to lane control) and, therefore, it is the primary point of foveal fixation.

Since the local difference between target and foveal fixation defines the precision of any motor reaction in relation to this target (ref. /9/), the extraction of primary foveal fixation has also to be attached to the feature extraction. In this context, it should not be disregarded that a decomposition of pattern processing into a directed chain consisting of preprocessing, feature extraction, and pattern recognition, is a poor abstraction. Especially the direction of foveal fixation depends also on the result of pattern recognition.

For pattern recognition it is assumed that \( O_t \) has to be correlated with a reference matrix \( R_t \) which represents a pattern of membership values derived from vehicle width. The correlation of \( O_t \) and \( R_t \) is performed horizontally, combining the \( i \)-th row vector \( O^T_{i,t} \) of \( O_t \) with \( i \)-th row vector \( r^T_{i,t} \) in the following manner:

\[
\begin{bmatrix}
\mu_{O_{t}^{(n_1,\eta_1,t)}} & \cdots & \mu_{O_{t}^{(n_n,\eta_1,t)}} \\
\vdots & \ddots & \vdots \\
\mu_{O_{t}^{(n_1,\eta_m,t)}} & \cdots & \mu_{O_{t}^{(n_n,\eta_m,t)}}
\end{bmatrix}
\]

\[
O_t = \begin{bmatrix}
\mu_{O_{t}^{(n_1,\eta_1,t)}} & \cdots & \mu_{O_{t}^{(n_n,\eta_1,t)}} \\
\vdots & \ddots & \vdots \\
\mu_{O_{t}^{(n_1,\eta_m,t)}} & \cdots & \mu_{O_{t}^{(n_n,\eta_m,t)}}
\end{bmatrix}
\]

The correlation vector

\[
c^T_{i,t} = \begin{bmatrix}
\cdots & r_{i2,t} & r_{i1,t} & r_{in,t} & \cdots \\
\cdots & r_{i3,t} & r_{i2,t} & r_{i1,t} & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & r_{i1,t} & r_{in,t} & r_{i(n-1),t} & \cdots 
\end{bmatrix}
\]

with the correlation vector (\( n \) even)

\[
c^T_{i,t} = \begin{bmatrix}
c_{i(-n/2)} & \cdots & c_{i(-1)} & c_{i(0)} & c_{i(1)} & \cdots & c_{i(n/2)}
\end{bmatrix}
\]
The result is a $m \times (n + 1)$-matrix
\[ C^T_t = \begin{bmatrix} c_{1,t} & \cdots & c_{m,t} \end{bmatrix} \]
containing the information relevant for reaction behaviour and foveal fixation. The realization of such correlative conjunctions by neural networks is held possible by several authors (ref./10/). Also, these conjunctions can be extended to fuzzy operations. Without that if the entries of $C_t$ are normalized by its maximum element $C_t^-$ can be interpreted as a fuzzy matrix. That row of $C_t^T$ which exhibits the most significant progress (sharpest maximum) of correlation determines - if present at all - the vertical range of foveal fixation. Its horizontal range is determined by the maximum element (or elements) of $C_t^T_t$. This significant row of $C_t$ is selected for the succeeding pattern recognition process.

According to BREMERMANN (refs /11,12/) the pattern recognition process is operating with several internally represented prototypes (one prototype for one class of objects respectively) and by comparing those with the instantaneously gained pattern. The comparison happens by deforming each prototype as far as maximum coincidence with the actual pattern is reached. Suppose the identity of one prototype and the actual pattern, the membership of the analyzed pattern to belong to the class represented by this prototype is unique. If there is no prototype of such kind, the deformation and the distortion remaining at maximum coincidence act as a measure of membership with which the actual pattern belongs to each of all classes represented by the prototypes; i.e. the actual pattern is a fuzzy subset of the universe the prototypes constitute. By this fuzzy classification procedure good results were achieved in identification of handwritten letters.

While in the case of letter recognition several discrete, uniquely defined prototypes are suitable, in the case of lane control the straight-ahead course is the only prototype distinguishable in an unique manner as by the fact that no motor reaction is necessary. Deviations from this course are compensated by turning the steering wheel. Hence a pattern recognition model is indicated which is satisfied with one internally represented prototype, namely the pattern of straight-ahead course. In our opinion, the deformation degree of prototype, necessary for a maximum coincidence, corresponds directly to the steering wheel angle. On the other hand, the remaining distortion of coincidence represents the measure of nonuniqueness or uncertainty of the decision unit input (section II.2) and so far even a measure of uncertainty of the reactions themselves. The deformation can be determined by an operation like correlation which is a similarity measure just as deformation degree. This operation can principally be applied to the total actual pattern so that feature extraction would become redundant. The disadvantage of this procedure is that with it foveal fixation can not be derived from the pattern processing at all.
In contrary, we start from an already present prototype $P_t$ acquired by the above-mentioned preprocessing like the actual pattern. Classifying this actual pattern after feature extraction it remains only to compare that row of $C_t$ with the corresponding row of $P_t$ (showing the same structure as $C_t$) which has been recognized at feature extraction to be significant.

$$f_{i,t}^T = \begin{bmatrix} 
\cdots p_{i(-n/2+1)} & p_{i(-n/2)} & p_{i(n/2)} & \cdots \\
\cdots p_{i(-n/2+2)} & p_{i(-n/2+1)} & p_{i(-n/2)} & \cdots \\
\vdots & \vdots & \vdots & \vdots \\
\cdots p_{i(-n/2)} & p_{i(n/2)} & p_{i(n/2-1)} & \cdots 
\end{bmatrix} \quad (10)$$

Normalization of the elements of correlation vector

$$f_{i,t}^T = | f_{i(-n/2)} f_{i(-n/2+1)} \cdots f_i(0) \cdots f_{i(n/2)} |$$

by its maximum element yields the fuzzy input vector of the fuzzy dynamic system modelling the decision strategies:

$$x(t) = f_{i,t}$$

$$y(t+1) = R_1 x(t) \cap R_2 y(t) \quad (11)$$

The uncertainty of $y(t+1)$, caused by the remaining distortion of $f_{i,t}^T$, can be expressed by use of fuzziness stated in the pertinent literature (e.g. entropy of membership functions, fuzziness index; see ref. /3/).

In the following, the expressiveness of our suggested perceptual model will be demonstrated by analyzing and explaining observations in field studies, drawn from the literature (ref. /13/). SHINAR and his co-workers ascertained that certain roads showed a clearly higher accident rate than other roads, differing merely in their appearance but almost not in their geometric characteristics curvature and width ("high accident curves" HAC; "low accident curves" LAC). They reported especially that the HACs were underestimated in their curvature and overestimated in their width by the drivers. Further, it was found that HAC negotiation goes along with greater behavioural uncertainty of drivers (by means of their eye movement behaviour: frequent saccades with low amplitudes) than LAC negotiation. The authors tested several corrections of HACs, painting the road surface in different ways, only one of which succeeded, i.e. it affected an estimation and reaction behaviour of drivers like LACs. The correction consisted in painting the road surface in such a way that the inner curvature angle appeared more obtuse than that of uncorrected HAC.

Fig. 1 shows the courses of LAC, HAC, and corrected HAC according to SHINAR. To the right of each course the corresponding matrix of preprocessed spatial-discretized pattern is
represented to which the correlation operations of eq. 8 will be applied. Thereby, a matrix with row vectors

\[ r_{i,t}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \]

is chosen as reference pattern. The resulting correlation matrices are placed together in Fig. 2. Comparing the correlation matrix of HAC with that of LAC it can be seen that the LAC matrix contains a relatively sharp maximum in the second row, whereas there are three rows (1st, 2nd and 3rd) in the HAC matrix with three elements having the maximum value 1.

From the viewpoint of our model assumptions, in this case, the fixation behaviour should become more unsteady than at LAC. Just this effect is reported by SHINAR et al. Through the correction of HAC the structure of the corresponding correlation matrix approximates that of the LAC-matrix tallying with the observation of a firmer fixation behaviour. Finally, the pattern recognition, applied to significant rows of the LAC-, HAC-, and corrected HAC-correlation matrices according to eq. (10), generates the fuzzy input vector of the decision system. The resulting curves shown in Fig. 3 could explain the improved skill at negotiating corrected HAC in accordance with SHINAR's results.

III. NOTES ON EXPERIMENTAL METHOD

The method to be outlined in the following refers exclusively to the psycho-mathematical model described in sections II.2 and II.3, especially to its verbal component. As mentioned before the formal model of reaction dynamics produces among others - the dependent variables necessary for testing hypotheses. The following components of the visual ambient field have to be controlled as independent variables:

a) Systematical variation of the portion and local distribution for several frequency components within the visual field is to prove hypotheses 1, 2 and 4 (dynamic test);

b) Systematical variation of course curvature and width is to prove hypotheses 1 and 4 (dynamic test);

c) Systematical variation of course curvature and width as well as the "signal/noise-ratio" within the visual field is to prove hypothesis 3 (static test).

Such variations are obviously not producible in field experiments, and so it is advisable to use a general purpose driving simulator.

The simulator system SISY, meeting the mentioned requirements, has been developed and partly put into operation.

The dependent variables are provided by acquiring the time history of eye movements, steering wheel angle, and position of pedals. The relation between these dependent variables and the binary-coded display images is established in the fuzzy dynamic decision system. In case of variation c) a static decision
model replaces the dynamic system. Both models represent primarily an "accumulative memory" of the dynamic reaction behaviour. Because of their data reducing features they also produce characteristic values used for statistical tests as dependent variables. Identifying the reaction behaviour by fuzzy systems the membership functions (see section II.2) have to be estimated. One possible procedure is to increment the DIRAC measure of the discrete interval, assigned to the observed input and output variables. Normalizing the resulting histogram by its maximum value we get the values of membership function (ref. /3/).

This procedure resembles the estimation of empirical probabilities but the normalization related to the enclosed histogram volume has to be performed thereby. This would result in a stochastic (probabilistic) system. The fuzzy system we use is a generalization of the stochastic approach in several respects:

1. Weaker restrictions referring to the membership functions regarded as a fuzzy measure.
2. Extended possibilities for accumulating observations to membership functions (with the disadvantage that no statements about consistency of these estimations exist).
3. The uncertainty modelled by the membership function is always the fuzzyness of a classification, i.e. the fuzzyness of the assignment of an element to some class. This meets the subjective character of uncertainties occurring by sensory-motor tasks rather than the probabilistic approach.

IV. REFERENCES

/1/ JOHANNSEN, G.; et al.: Der Mensch im Regelkreis. Oldenbourg Verlag, München, 1977


/7/ KRAMER, U.; ROHR, G.: Sensumotorische Regelprozesse bei der Fahrzeugführung. Forschungsbericht Nr. 343, Institut
für Fahrzeugtechnik, Technische Universität Berlin, 1981


FIGURE 1: Original and preprocessed patterns

FIGURE 2: Normalized correlation matrices of actual patterns and prototype (PT: straight course)
FIGURE 3: Deformed prototypes for significant correlation matrix rows
LEVELS OF STEERING CONTROL; SOME NOTES
ON THE TIME-TO-LINE CROSSING CONCEPT
AS RELATED TO DRIVING STRATEGY

Hans Godthelp and Hein Konings
Institute for Perception TNO, Soesterberg, The Netherlands

ABSTRACT

A description is given of a research project on 'levels of steering control', in which the distinction between open and closed loop control strategies in driving is discussed. In most driving control models emphasis is put on a closed loop, or error control, strategy. Even when an open loop strategy is taken into account, an error control mode is always assumed to function in parallel. The preliminary thoughts laid down in this research note are based on the idea that it may be useful to develop a driving model in which the open and closed loop strategies act serially rather than in parallel. As far as open loop control is concerned, a distinction is made between an active and a passive control mode. In the active mode the driver generates an open loop steering action which may be either of a precognitive or a pursuit nature. In the passive mode the driver may either hold the steering wheel in a fixed position (fixed control) or he may release it (free control). The 'time-to-line crossing' concept is discussed as a variable which may be useful for describing the efficiency of open loop control strategies and which may also serve as a tool in modelling driving strategy.

1. INTRODUCTION

An extensive description of the project on 'levels of steering control' was given in Godthelp (1980). Roughly the aim of the project is to describe the functioning of high levels of control i.e. of control strategies in which error correcting mechanisms play only a minor role. Many of the models which consider driver's behavior in steering tasks are based on the assumption that the driver acts like an error correcting mechanism who is continuously paying attention to this specific task element. Nevertheless it is also well understood that under many circumstances driving does not require continuous error control. Therefore instead of closing the loop, drivers may temporarily behave in an open loop mode. In the present study the functioning of open and closed loop strategies is subjected to further analysis.

In order to consider the different strategies which may be adapted by the driver while operating in an open loop mode, it is important first to consider precisely what is meant by the term 'open loop control'. From a straightforward servo theoretical point of view, a system operates in an open loop whenever information about the error between desired and actual system behaviour is either unavailable or is not acted upon. From a psychomotor point of view, in the driver-vehicle system such error information is presented principally through the visual channel, and represents information about such features as lane position, relationship to other traffic elements, velocity, etc. In the event of any interruption to, or occlusion of, this channel, one would thus expect the driver to perform in an open loop mode. This is true, however, only to the extent that the driver relies on immediate visual information. In reality this is seldom the case, as the driver
has, in addition to feedback about vehicle performance through other non-
visual modalities, such as proprioceptive information, a fairly developed
'internal model' of the expected performance of the vehicle in relation to
the roadway over some time/distance into the near future. In this sense
then the driver's performance can not usually be thought of as purely open
loop even when he is not visually attending to the present roadway envi-
ronment.

In addition to the information processing aspects of open loop con-
trol, it is important also to consider some of the modes in which motor
control can be executed during open loop driving. Basically, these may be
divided into an active and a passive mode. Acting on his internal repre-
sentation about the vehicle's input-output characteristics the driver may
actively generate steering actions which are intended to guide the auto-
mobile, in a sort of pursuit fashion, along an intended trajectory, inde-
dependently of where it happens to be within the lane at the present moment.
Acting in conjunction with this type of control is the driver's invocation
of his store of various motor programmes which are available to him to aid
in the precognitive execution of certain well-learned manoeuvres. On the
other hand, in the intervals between the actively generated steering ac-
tions, the driver may find himself to be in the passive control mode in
which case he essentially does nothing during certain periods of time.
Subdividing this last category further, the driver may employ either fixed
control, by holding the steering wheel fixed and allowing the automobile
to continue along its present course, or free control, by releasing the
steering wheel and relying on the self-centering properties of the steering
system.

Given that the driver as the controller of the vehicle in relation to
the roadway may act in either an open or a closed loop fashion, it is also
important to ponder the fundamental question of whether these open and
closed loop control strategies act together simultaneously, in a mutually
parallel fashion, or whether they are adopted consecutively in time, in
an analogously serial fashion. Reviewing some of the recent literature we
see that both types of models have been proposed. Allen and McRuer (1977)
and Donges (1978) have both presented models in which an active pursuit
(with preview) open loop control acts in parallel with compensatory closed
loop error control. In the models of Carson and Wierwille (1978) and of
Baxter and Harrison (1979) a nonlinear element in the form of a threshold
or hysteresis loop, respectively, has been introduced into the standard
error-correcting closed servo loop. The implication of these nonlinear
elements is that during the system 'dead-time' the driver adopts an open
passive fixed control strategy, whereas beyond these levels closed loop
control predominates. Since the system can only be in one of these states
at any one time, the open and closed loop strategies thus switch back and
forth successively, or in series. From an information processing, or
psychomotor, point of view it is our opinion that some sort of comparable
serial alternation between open and closed loop strategies should be a
preferable strategy, in that during the passive open loop period the driv-
er should then be able to devote his attention to other task aspects.

The preliminary thoughts laid down in this research note are based on
the idea that it should be useful to extend our knowledge about the func-
tioning of open loop control strategies. From this knowledge predictions
could then be made of how open and closed strategies could interact and how
those strategies could be distributed over time. Results of experiments
in which the above given approach is tested on its validity will be pres-
ented later. In the present paper the 'time-to-line crossing' concept is
discussed as a variable which may be useful in describing the time efficiency of both active and passive open loop control strategies. This proposed measure will be illustrated by a theoretical and an experimental example.

2. THE TIME-TO-LINE CROSSING CONCEPT

From the traffic conflicts literature we are acquainted with a variable called the: 'time measured until collision between two vehicles' or in short 'time-to-collision' (TTC). Roughly, TTC is the time required for two vehicles to collide if they continue at their present speed and at the same path (Hayward, 1972). TTC is calculated by measuring speed and path of two vehicles, as well as their distance from the potential point of collision. By definition, the above given TTC concept starts from the assumption that the driver behaves in a sort of passive open loop control mode, since no action is assumed to be taken to avert the collision. However, from the vehicle control point of view the definition seems incomplete. Instead of assuming a constant speed and path angle, it is probably better to assume that the driver's control action remains fixed, which would lead to a calculation in which the longitudinal deceleration and the lateral yaw remain constant. In this way the TTC definition is more closely related to the assumption that the driver uses a fixed, open loop control strategy.

It seems furthermore worthwhile to note that it is not necessary to use the TTC concept only for conditions in which two vehicles are involved. One of the vehicles may be replaced by an obstacle or, as in the present study, by the roadway delineation. In that condition TTC describes the Time-to-Line Crossing, which will be denoted as TLC. It is suggested that TLC may be an important cue for drivers whether to use a passive, fixed, open loop control strategy and/or whether to switch over to closed loop control. By calculating TLC after an active open loop control action it may also give an idea about the efficiency over time of this type of open loop control. The following paragraph shows a theoretical example of the latter aspect.

3. A THEORETICAL APPROACH: Curve entrance.

The entrance to a curve is often referred to as a condition in which the driver uses an active open loop control strategy. During the (straight) approach the driver will use preview to estimate road curvature ($C_r$) and to plan a steering action ($\delta_0$). The actual steering wheel movement may start at a period ($t_2 - t_1$) before the curve begins. Donges (1978) conceived the period ($t_2 - t_1$) as an anticipation time. Fig. 1 gives a schematic description of this process. In particular, in cases of small steering wheel movements and/or with high speed driving, the steering action from $t_1$ to $t_3$ will be completely open loop, i.e. based on the perceived road curvature and 'weighting' of the steering characteristics of the car.

Steering wheel error at the end of the active open loop period ($t_3$) can be considered as a variable indicating the efficiency of the open loop steering movement. Godthelp (1980) gave laboratory results of how steering wheel movement amplitude, frequency and steering force do affect the errors in movement reproduction under visually open loop conditions. Further experiments using a curved roadway imagery are presently under way.
Fig. 1 Schematic description of the curve entrance process, assuming an active open loop steering action from $t_1$ to $t_3$ and fixed open loop control from $t_3$ to $t_4$.

In addition to this experimental work theoretical calculations can be made about the consequences of making a steering amplitude error at the end of the active, open loop period ($t_3$). Actually such an error will lead to a vehicle motion with an incorrect curvature ($C_r$). The time from $t_3$ to line crossing at $t_4$, assuming fixed open loop control after $t_3$, gives an impression of the consequences of this original steering error. It can be shown that this TLC at $t_3$ can be represented by a relatively simple formula (see also Figs. 1 and 2):

$$
TLC = \frac{G_1 (1+KU^2) \arccos \left( 1 - \frac{W_r - W_{ver}}{G_1 (1+KU^2)} \delta_{se} \right)}{U \delta_{se}}
$$

where:
- $TLC =$ time to line crossing at $t_3$ in seconds
- $G =$ steering ratio
- $l =$ wheelbase in meters
- $K =$ understeer/oversteer coefficient
- $U =$ vehicle speed in meters per second
- $W_r =$ lane width in meters
- $W_v =$ vehicle width in meters
- $W_{ver} =$ effective vehicle width, rightward (see Fig. 2)
- $\delta_{se} =$ steering angle error at $t_3$ in radians
- $l_f =$ distance between center of gravity and car front in meters
- $\psi =$ heading angle in radians
Fig. 2 Effective vehicle width at the moment of line crossing.

In deriving this TLC formula the following assumptions have been made about the vehicle motion at $t_3$: a) the car is in a steady curvature, b) the car is in the center of its lane and c) the heading angle is zero. Furthermore, it is assumed that longitudinal speed is and remains constant after $t_3$. An effective vehicle width was involved in the formula in order to correct for the heading angle at the moment of line crossing (see Fig. 2).

The formula shows how TLC depends on the properties of the driver-vehicle-road system. Figs. 3-5 give some illustrations. In these examples the following vehicle properties were chosen as constants: $G = 19.8$; $l = 2.62$ m and $l_u = 2.24$ m.

It is remarkable that road curvature as such is not represented in the TLC formula. However, Fig. 3 shows how TLC is affected by the driver's error in generating the correct steering wheel angle. When we combine these data with those of Godthelp (1980), who illustrated how the error of open loop steering wheel movements increases with steering movement amplitude, it can still be expected that TLC will be shorter for larger curvature.

Fig. 3 also gives an impression of the influences of speed. The curves in Fig. 3 apply to an understeered car. The TLC formula also shows a strong relation with the yaw rate sensitivity of the car which we know plays a role in vehicle handling. Fig. 4 gives an illustration of the influence of the understeer/oversteer coefficient. For cars with understeer characteristics the largest speed effects can be found in the lower speed range. With high speed TLC becomes more and more independent of speed. The curves for the oversteered car show a completely different picture. Here TLC becomes zero at the critical speed.

Lane width also strongly influences TLC. Fig. 5 shows this effect for different levels of vehicle width. From the figure it can be seen that TLC levels drop fastest when the difference between lane and vehicle width becomes less than 0.50 m.

For the examples in Figs. 4 and 5 the steering angle error is normalised at a value of $\delta = 1^\circ$. The TLC curves for this condition can be considered as normalised, theoretical curves which may be compared with each other.
Fig. 3 The relation between TLC and steering wheel error for different levels of speed, as predicted by the TLC formula.

Fig. 4 The relation between TLC and vehicle speed for different levels of the understeer/oversteer coefficient, as predicted by the TLC formula.

Fig. 5 The relation between TLC and road width for different levels of vehicle width, as predicted by the TLC formula.
The present TLC formula gives the time-to-line crossing after an active open loop control action and assuming fixed open loop control after that action. As such TLC represents a time which is decisive for the efficiency of the active open loop action. As a consequence TLC might serve as a predicting variable regarding the interplay between open and closed loop strategies. Field experiments are presently planned to verify the practical implications of these predictions. A further discussion of this point will be given in part 5.

4. AN EXPERIMENTAL EXAMPLE: Straight lane keeping.

In the previous paragraph theoretical values were given for TLC at one particular moment i.e. at the end of an active open loop steering action. In the present paragraph an illustration will be given of how TLC can be used as a continuous variable in a lane keeping situation. To perform this analysis experimental data were used from an experiment which was presented earlier by Blaauw et al. (1977). In this experiment driving performance on a straight motorway section was analyzed for experienced as well as inexperienced subjects. Driving strategy was varied by means of specific instructions for lateral and longitudinal vehicle control. In the original paper the results about steering behaviour and lane keeping performance were presented in terms of means, standard deviations and spectral density functions. Major questions to be answered in the present analysis are: a) can TLC be calculated as a continuous variable describing the time-to-line crossing at each particular moment of an experimental run, and 2) can TLC add information to 'conventional' variables such as those described in the original paper?

4.1 Method

Subjects
Six male subjects (Ss) took part in the experiment, three experienced drivers and three very inexperienced drivers. The experienced drivers had their license for at least three years and a driving experience of at least 30,000 km. The inexperienced drivers followed a driver training course and had a driving experience of about 35 hours. They were going to have their driving test within two weeks and received training in a similar car to that used in the experiment. All Ss were between 21 and 28 years of age. They were paid for their services.

Procedure
Ss had to drive an instrumented car on a straight section of a four-lane motorway with divided traffic lanes, having a constant road geometry and lanewidth (3.60 m) for about 4,000 m. The runs were carried out during long gaps in the normal traffic flow. Ss drove the experimental section in one direction, returning to the starting point after each run. During all runs two experimenters (Es) were present in the car. One E took care of the apparatus, while the other instructed the S before each run. During the runs of the inexperienced drivers a driver-training instructor acted as the latter E. A written instruction was presented to Ss before the first run, in which they were told to drive in the right-hand lane without overtaking. Driving strategy was manipulated by means of specific instructions.
There were two instructions for lateral vehicle control:
1) Instructing Ss to drive as straight as possible, informing them of the
fact that in particular their performance in straight driving was going to
be recorded and that therefore they should fully concentrate on this task.
This instruction aimed at stimulating Ss to behave in a tightly closed loop
mode, i.e. with continuous error correcting. Runs with this instruction
will be referred to as 'CL' (closed loop).
2) Instructing Ss to keep lane but to look around and to mention details
about the environment. This instruction was chosen in order to provoke a
condition in which Ss switch between open and closed loop control strate­
gies. Because of the open loop character, runs made with this instruction
will be referred to as 'OL'.

The instructions for lateral vehicle control were combined with one
of the following instructions for longitudinal control:
a) Forced speed - to maintain a constant speed of either 80 or 100 km/h.
b) Free speed - to drive with a self chosen or 'comfortable' speed.

Instructions for lateral and longitudinal control combined resulted in
six conditions (2 x 3). Each S drove each condition three times in a ran­
domized sequence during daytime. These runs were completed in two two-hour
sessions on different days. During return to the starting point no specific
instructions were given.

**Apparatus**
An instrumented car, a Volvo 145 Express, was used. Steering wheel angle,
lateral position and speed were recorded continuously. The steering wheel
angle was recorded with a potentiometer activated by the steering axis. The
lateral position of the car was measured by a transducer, scanning the road
luminance by a fast rotating prism in front of a photodetector. The detec­
tion of the right edgeline was the critical signal in measuring the lateral
position relative to this line. A tachometer mounted on the outward axis
of the gear-box measured the speed. All variables were recorded on magnetic
tape for computer-processing. Main characteristics of the instrumented car
were: steering ratio $G = 19.8$
wheeldiameter $l = 2.62$ m
distance between c.g. and front axis $a = 1.59$ m
understeer coefficient $K = 20.10^{-4}$ s$^2$/m$^2$
mass $m = 1850$ kg
moment of inertia around z-axis $I = 3180$ kgm$^2$

**Data analysis - TLC calculation**
Each signal was sampled over 128 sec with a sampling frequency of 4 Hz
(sampling period $\Delta t = 0.25$ sec). Because of the place of the lateral position
transducer at the rear of the car this signal was converted so that
it accounts for the lateral position of the centre of gravity of the car.
The TLC for each sample moment $i$ was derived from the lateral position signal
by way of solving $t$ from the following formulae, which consider line
crossing with the right-hand edgeline and with the road centerline res­
pectively:

$$\frac{1}{2} y_i t^2 + \dot{y}_i t + y_i = \frac{W}{2}$$  \hspace{1cm} (2)

$$\frac{1}{2} \dot{y}_i t^2 + \ddot{y}_i t + y_i = W_r - \frac{W}{2}$$  \hspace{1cm} (3)
with: \( y_i \) = lateral position at sample moment \( i \) m
\( \dot{y}_i \) = lateral speed at sample moment \( i \) m/s
\( \ddot{y}_i \) = lateral acceleration at sample moment \( i \) m/s\(^2\)

\( W_v \) = vehicle width 1.73 m
\( W_l \) = lane width 3.60 m

The smallest positive solution for \( t \) gives TLC at the sample moment \( i \). The formulae are based on the assumption that the motion of the car has a constant lateral acceleration after the sample moment \( i \). Actually this assumption means that the driver behaves in a fixed open loop mode after \( i \). In this mode longitudinal speed and yaw rate are assumed to remain constant. For small heading angle errors a vehicle motion with these characteristics can be considered to have a constant lateral acceleration.

The lateral speed and acceleration to be used in the aforementioned formulae were derived from the lateral position signal:

\[
\dot{y}_i = \frac{y_{i+1} - y_{i-1}}{2 \Delta t} \quad \text{and} \quad \ddot{y}_i = \frac{\dot{y}_{i+1} - \dot{y}_{i-1}}{2 \Delta t}
\]

Because these derivatives are strongly affected by a noise component in the lateral position signal a filter procedure was used to 'clean' this latter signal. A low pass Butterworth filter was adopted for this procedure. This filter was such that its power gain was 0.5 at 0.30 Hz. Together with this frequency and the sampling time of 0.25 sec the order of the filter (here chosen to be 10) determines the length of the interval around 0.30 Hz in which the power gain decreases from 0.99 (0.24 Hz) to 0.01 (0.37 Hz). By filtering the data in two different directions, the phase shift is zero. After the filtering the first and the last 20 samples of each run were ignored in the further analysis in order to eliminate the effect of the filter on the beginning and the end.

Fig. 6 gives an example of a time history of the lateral position signal and the corresponding TLC (see also the next part 4.2). Frequency distributions, mean values and standard deviations were calculated for lateral position and lateral speed. The upper part of Fig. 7 gives an example. The lower part of this figure presents frequency distributions for TLC. Specific characteristics were derived from the cumulative distribution for each particular run. The '15% TLC value' gives a representative TLC value for a particular run which indicates that 15% of the time TLC was shorter than this value, either to the left (centerline) or to the right (edgeline). Furthermore, 'the percentage of time with TLC shorter than 5 sec' was derived from these cumulative distributions.

It can be seen in Fig. 6 that each part of the TLC curves has a minimum value. Frequency distributions were also made for these minima. From these frequency distributions the median of the minimum TLC values was derived, which gives a representative minimum TLC as used by the driver in a particular run. Fig. 6 also shows how TLC predictions switch from leftward to rightward v.v. Mean times between these switches were calculated for each run. Differences between conditions were tested on their statistical significance by means of analysis of variance.
Fig. 6 Example of a time history for the lateral position signal and TLC. The dots in the lateral position history mark the width of the car.

Fig. 7 Mean frequency distributions for lateral position, lateral speed and TLC as averaged over the runs of the experienced Ss. driving with free speed and CL instruction. For the lateral position the mean of each separate distribution was shifted to the overall mean for this condition before averaging.
4.2 Results

One of the main results of the present analysis has already been shown in Fig. 6. This figure gives an illustration of how the lateral position signal can be transformed to TLC. From the figure it can be seen 1) how line crossings as predicted by the TLC concept switch from leftward to rightward and 2) how TLC curves can differ as to their shape, minimum, length, etc. The point marked with A shows an example of neglecting an increasing error leading to a very small TLC value.

Fig. 8 gives the results for the lateral position signal which show that the instruction for lateral control strongly affected performance. In the OL condition mean lateral position is more to the center of the lane \( p < 0.01 \), whereas standard deviations are larger as compared to the CL condition \( p < 0.01 \). Further to this main effects, a slight interaction \( p < 0.05 \) for the mean lateral position indicates that experienced drivers react in a different way to an additional speed instruction as compared to inexperienced drivers.

The 15\% value for TLC is given in Fig. 9. The left and right part of this figure consider TLC towards the left (centerline) and the right (edge-line) respectively. A remarkable difference between left and right can be noted. The main effect of lateral instruction as found in mean and standard deviation of lateral position is reflected in the TLC to the centerline. However, no effect exists to the edgeline. Obviously the combination of a more rightward position and a smaller standard deviation in the CL condition counterbalance in such a way that TLC levels to the right are about equal for the CL and OL condition. Fig. 10 shows that the same effect is found for the percentage of time with TLC shorter than 5 sec. For this latter variable Fig. 11 shows the interaction between lateral instruction and TLC direction when averaged over driving experience and longitudinal instruction. From the Figs. 10 and 11 it can also be seen that particularly for the CL condition the percentage of time with TLC shorter than 5 sec towards the centerline is very small. Fig. 12 gives an impression of the TLC minima. It appears that the minima to the edgeline are somewhat shorter than those to the centerline and that this effect is strongest for the CL conditions.

Finally the mean time between switches in TLC (from leftward to rightward v.v.) proved to be slightly but significantly \( p < 0.05 \) smaller for experienced Ss than for inexperienced Ss (2.88 s versus 2.96 s).

As stated earlier the lateral position signal was filtered because of the noise in the measurements. The use of the filter may have had some influence on the absolute values of the TLC data, and so we accept them with some reserve. However, it is obvious that differences as they are described above will only be found if they actually exist.

5. DISCUSSION

The aim of the present study was to analyse the time-to-line crossing (TLC) concept as a variable which may be useful in describing open loop control strategies. Theoretical predictions were made about the influence of some road, vehicle and driver related properties on the efficiency of active open loop control actions. Actually the TLC formula which was presented in par. 3 offers the opportunity to predict the effect of various parameters on the distribution over time of open and closed loop strategies. Experiments on curve entrance and negotiating will be carried out to verify
Fig. 8 Mean and standard deviation of the lateral position as a function of speed instruction and for the two groups of driving experience (exp., inexp.) and lateral instruction (OL, CL).

Fig. 9 Mean 15% values TLC for the different conditions. The left and right part of the figure refer to TLC predictions towards the left (centerline) and right (edgeline) respectively.

Fig. 10 Mean percentage of time with TLC shorter than 5 sec for the different conditions. The left and right part of the figure refer to TLC predictions towards the left (centerline) and right (edgeline) respectively.
the usefulness of predictions based on this theoretical TLC concept. Further attention will also be given to the fact that TLC predicts a line crossing, of which it can be assumed that the driver will not accept it. Before the moment of line crossing the driver will switch to an error control mode. Therefore the various factors which determine the minimum accepted TLC value have also to be analysed. From such an analysis it should become clear how an effective, passive open loop time can be derived from TLC.

The analysis of the lane keeping experiment gives an impression of how TLC can be used as a continuous variable. Although this analysis should be considered as preliminary, the results suggest that TLC offers the opportunity to characterize lane keeping with a variable which considers time as well as lane position. Fig. 6 gave an example which illustrated how TLC can serve as a sort of prediction-display which represents a driver's spare time at every moment of driving. On the one side this TLC concept may give insight in the time which is available for the driver to spend on tasks not strictly related to vehicle control. On the other hand TLC can provide information about the probability of lane exceedance at each particular moment. As such TLC can serve as an addition to variables presented by Allen and O'Hanlon (1979). Further experiments will be done in order to analyse how TLC is affected by characteristics of the road, the vehicle and the driver.

Finally, TLC should be considered as to its usefulness in modelling driver behaviour. A further analysis will be made about the relation between TLC, minimum accepted TLC values and a driver's decision to switch from an open to a closed loop driving strategy v.v. This analysis should give insight into questions such as whether and how TLC can serve as a
tool in models like those developed by Carson and Wierwille (1978) and Baxter and Harrison (1979), i.e. in models which consider open and closed loop strategies act serially. Until now these models primarily account for straight road driving. The methods as given in the present paper may be used to develop a 'serial' strategy model of driving which also accounts for curved road sections. Such a model should involve an active open loop mode in addition to a passive open loop mode. Further attention will also be given to the question of whether TLC can make predictions about a driver's observation strategy. Driving experiments in which voluntary occlusion times are measured have been done (Blaauw, 1981) and are planned. The hypothesis that the length of occlusion periods will be related to TLC will be tested on its validity. An additional feature which describes driver's increasing uncertainty during the occlusion period will be combined with TLC in the latter analysis.

REFERENCES


session 6
miscellaneous

chairman: J. Wirstad, Sweden
THE EFFECT OF ELECTROTACTILE AND VISUAL FORCE FEEDBACK
ON LEARNING AND PERFORMANCE IN A TELEMANIPULATION TASK

J.P. Gaillard

One of the best known condition for learning is the necessity for the
subject to be informed at each instant of the result of the task he is exerting.

In teleoperation this means that the flow of information to the
operator has a great importance on his performance. Therefore, the two
basic questions we able deal with could be resumed as follows :

1. What kind of information does an operator need to optimise his
control in telemanipulation ?

2. What is the maximum quantity of information an operator will accept ?

The first question is generally labeled as a qualitative one while the
second one is labeled as quantitative. Evidently, we all try to find
answers to one or both questions through a specific system and search for
original solutions to find an adequate coupling between operator and
manipulator system. Human motion is primiraly governed by vision and to act,
man must centrally process the visual information. The way, or more exactly,
the hypothetical activity underlying this central process has retained our
attention in the teleoperator control. At first view a telemanipulation task
requires the operator to control by sight the task he is executing, and
therefore increasing information using a visual display could have an adverse
effect on both learning and performance. It could either give the information
necessary to a good control or overcharge the operator's information
processing capacity and distract him from the actual control task.

The main idea is not an original one, and dates back as early as the
1960's to feedback information to the operator through sensory channels
other than vision to resolve the eventual contradictory effect. A tactile
display could be one way to do so.

Another general argument in favor of tactile feedback is the performance
in pure reaction time experiments compared to other sensory channels,
mechanical and electrical stimulation of the skin provide shorter reaction
times than visual stimulation. One could argue that faster transmitting
of a stimulus would facilitate the generation of a movement. However, both
movement generation and learning require feedback of information that must
be processed centrally. This information, either is coded, stored in memory,
and retrieved when necessary. The way in which such information is
centrally processed in motor learning and motor control is yet to be
clarified.

The two experiments reported here focus on this question and represent
a first study on teleoperator control.

In their review on "the structure of motor problems" Keele and Summers
(1976) have pointed out that "temporary provision of artificial feedback
may be useful" for skill learning. In the closed loop theory of motor learning Adams (1972) obtained results that suggested feedback such as vision of the movement, sound on the track, and spring tension on the handle play a role throughout all stages of learning.

A recall process would be responsible for selection and initiation of movement, whereas recognition would be responsible for the evaluation of sensory feedback and the detection of movement errors.

The basic process every one agrees about presently is some formal comparison processes between efference specification (Adams, 1971) or schema of motor behavior (Keele, 1975) and knowledge of results for the development of the recall process; between sensory feedback such as visual, auditive, proprioceptive feedback, and knowledge of results for the development of the recognition process.

In both theories (closed loop and open loop) and in both activities (recall and recognition) knowledge of results should be provided to learn a skilled movement. The question of what kind of knowledge of results, which sensory modality and what quantity of information are required remains to be answered to optimize both training and performance.

One could argue that the way in which informations are displayed to the operator will facilitate or complicate the comparison between informations issued from the results of the movements and informations issued from efference specification and sensory feedback. Then recall and recognition processes would be more or less efficient according the way of providing informations to the operator to access his knowledge of the results.

**Experiment 1**

A first experiment was conducted to test the above-mentioned fundamental contents in a telemanipulation task. The main purposes was to test the validity of an information feedback increasing the knowledge of the results of an operator operating with the "SPARTACUS" telethesis. The second objective of this experiment was to study the efficiency of an electrotactile modality to return information to the operator and to compare it with the visual modality.

**METHOD**

1. **Material.** The experimental system was composed of a manipulator MA23 with seven degrees of freedom. The manipulator was controlled by a head movement transducers with two degrees of freedom, a roller transducer and a bar-switch (Photos 1, 2).

   Manipulator and transducers were connected to a SOLAR 16/65 computer, programmed to allow the operator to command the manipulator using a specific language linking the degrees of freedom of the transducers to specific translations or rotations of the gripper in space. Initially the system and its program were studied for a medical manipulator as a part of the SPARTACUS
project as an aid to high-level quadriplegics. Then the same concepts were used to control an industrial telemanipulator, but through other control transducers and adequated control language.

To output the information necessary to the objet's manipulation (3 positions, 3 orientations and 1 of opening the terminal device) some phases should be organised and then grouped in a strategy (Figure 1).

Photo 1 - Head movement transducer

Photo 2 - Roller transducer and bar-switch
Fig. 1- Functional diagram of a control strategy with selection of three phases (from Guittet and al.)

A phase-selection signal permits the operator to use his continuous signals in three different ways.

i) Position control of the gripper in cartesian coordinates,

ii) Position control of gripper yaw, pitch and roll and yaw

iii) Gripper opening and closing associated with forward/backward movement.

Table 1 resume the strategy used for this experiment.

<table>
<thead>
<tr>
<th>Phase</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left/right</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward/backward</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Up/Down</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gripper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yaw</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pitch</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roll</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open/close</td>
<td>o</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head movement in frontal plane</td>
<td>Δ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head movement in logitudinal plane</td>
<td>☐</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roller transducer</td>
<td>o</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 - Strategy used in experiment 1

2. Task. The task was a peg-in-hole insertion task in which a cube with a center hole had to be fitted on an orientable peg fixed on the table (Figure 2).

The angle $\alpha$ between the bracket and the vertical was varied between $30^\circ$, $45^\circ$ and $55^\circ$. First the operator had to approach the cube to the peg from its initial position 40 cm away from it. This was the "displacement task". Secondly, he had to position the hole over the extremity of the peg to prepare the cube insertion - this was the "adjustment task". Thirdly, he had to insert the cube on the peg: "insertion task".

* For phase 1 and 2 we used a position command and in phase 3 the forward/backward movement was a velocity control.
The displacement task was operated in phase 3 with a velocity control whereas adjustment and insertion tasks were effectuated in phases 1 and 2 with a position control. When the insertion was difficult the operator had to go to phase 2 and correct the orientation of the cube and then return to phase 1 to insert the cube. The transition from phase 1 to phase 2 and back was time consuming.

3. Sensory feedback. Two types of sensory feedback were used to inform the operator about the forces exerted by the gripper on the peg. First one was a visual display of 5 LED's with each diode successively representing a higher level of force. One diode lit meant the sum of forces was low, but not negligible, and 5 diodes lit that gripper was exerting a strong force. The sum of the forces was obtained from the summation of the currents of the three torque motors approximately corresponding to the three cartesian coordinates.

The visual display was mounted on a bracket attached to the manipulator. The distance between the gripper and the display was about 20 cms but the latter was clearly visible in any position of the manipulator.

The second form of feedback was an electrotactile stimulation situated on the operator's arm with the following characteristics:

- rectangular pulse train
- frequency modulation $1 \rightarrow 50$ Hz
- pulse duration 100 $\mu$s
- pulse intensity $= 5$ mA
  on the skin
These characteristics were tested in a pre-experimental procedure and confirm others observations (Kato and al, 1970; Prior and Lyman, 1977). The frequency range provided five distinguishable levels for the operators. The frequency varied proportionally with the sum of the forces with a low frequency stimulation, of 1 or 2 Hz, corresponding to weak forces and a high frequency of 40 or 50 Hz to strong forces.

4. Measurements. The task execution times were the dependent variable, obtained by measuring the time \( T \), to complete both the displacement and the adjustment task, \( T_1 \) and \( T_2 \) for the insertion task. The total execution time was their sum \( T_1 + T_2 = T_T \).

The experimenter started a chronometer when the operator coupled the clutch at the starting position of the cube and an intermediate time was recorded when the cube was adjusted, and the operator changed phase to start the insertion. Finally, he stopped the chronometer when the cube was inserted. Then the insertion time was obtained by substracting the displacement/adjustment time from the total time \( T_T - T_1 = T_2 \).

5. Procedure

Four experimental conditions were tested successively:

a) Tasks without sensory feedback of the sum of the forces \( V - 1 \).

b) Tasks with a visual feedback of the sum of the forces \( V_+ \).

c) Tasks with an electrotactile feedback of the sum of the forces \( E_+ \).

d) Tasks without feedback of the sum of the forces \( V - 2 \).

In each condition operator had to execute 30 trials composed with \( 3 \times 10 \) trials for each angle \( \alpha \) of the peg bracket. The order for the three different angles \( \alpha \) were obtained by a circular permutation. Varying the peg bracket angle at each trial allowed to avoid a strong learning effect of a specific orientation, thereby forcing the operator to exercise a control. Operators had to use the sensory feedback when angle was varied.

Two subjects were trained operators and had to complete a total of 120 trials corresponding to \( 40 \) trials for each angle \( \alpha \) and 30 trials for each experimental condition. The experimental design was obtained by the following formula: \( S_2 \times A_3 \times C_4 \times T_{10} \) (S for subjects, A for angle \( \alpha \), C for experimental conditions, and T for trials).

An operator completed the 120 trials in about 4 hours distributed over 4 sessions. One for each experimental condition. Before each session the operator had a 10 minutes warm-up.

6. Experimental predictions

Predictions were made on the learning rate for each type of task—displacement, adjustment and insertion. If we hypothetise that information feedback relevant to the sum of forces is a way to improve the knowledge of the result (K.R.) when forces are controlled in telemanipulation, then
learning and performance should be better with its use. Increasing knowledge of the result on the operators activity should give a more efficient association between the values of the K.R. and the motor response specification (in the closed loop theory) or value from the schema of movement (in the open loop theory). Then recall process should be better and movement generation more efficient. The execution time and learning rate would decrease more rapidly than if the sum forces were not relevant for the task or the quantity of K.R. were less.

On the other hand, increasing K.R. by the summed force feedback should give a more efficient comparison between K.R. values and sensory informations values issued when the movement is produced. The recognition process should then be better and movement corrections more efficient. We assumed that the displacement and adjustment tasks did not involve contact between the cube held by the gripper and the peg, so no forces were exerted by the manipulator other than lifting the cube. The sensory feedback was of no use to the operator. However, the insertion task involved forces on the manipulator and sensory feedback could be useful.

We predict that learning and performance would be more efficient for the insertion task than for the displacement/adjustment one because pertinent quantity of K.R. would be greater in the former case.

A second study was to observe a possible difference between visual display and electrotactile stimulation, but no specific prediction were developed other than the following one. If electrotactile is not informative for the operator, then the learning effect on the insertion task could be stopped when it is feedback to operators.

Prediction was an interaction between the execution time of the two sub-tasks.

7. Results. Data reported in table have been calculated with the angles a taken together and represents an average on 30 trials for each sensory condition. D.A. is the displacement/adjustment time on T1 and I the insertion time on T2.

<table>
<thead>
<tr>
<th>Sensory condition</th>
<th>Subject</th>
<th>S1</th>
<th>S2</th>
<th>S1 + S2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D/A</td>
<td>I</td>
<td>T</td>
<td>D/A</td>
</tr>
<tr>
<td>V - 1</td>
<td>16,3</td>
<td>32</td>
<td>48,3</td>
<td>18,8</td>
</tr>
<tr>
<td>V+</td>
<td>13</td>
<td>21,8</td>
<td>34,8</td>
<td>13,8</td>
</tr>
<tr>
<td>E+</td>
<td>13,8</td>
<td>19,6</td>
<td>33,4</td>
<td>14,6</td>
</tr>
<tr>
<td>V - 2</td>
<td>12,7</td>
<td>16,4</td>
<td>29,1</td>
<td>13,1</td>
</tr>
</tbody>
</table>

Table 2
Mean time T1 and T2 for operators S1 and S2 as a function of the sensory condition of feedback.
A variance analysis has been computed according with the design S2 × C4 × E10 for the two times T1 and T2 separately.

- **T1 analysis.** The variance analysis did not show a significant difference of the displacement/adjustment time between the two operators $F(1,9) = 1.17; p > .10$.

However, one found a significant effect on the sensory conditions. $F(3,27) = 4.59; p < .01$.

Interaction between operators and sensory conditions were not significant. $F(3,27) = 0.21; p > .10$.

This lack of interaction indicated that the two operators have no different behavior in response of the sensory feedback. Then we grouped S1 and S2 to analyse the effect on sensory conditions. The orthogonal decomposition of the three degrees of freedom on the factor sensory condition was as follows:

<table>
<thead>
<tr>
<th>Comparison</th>
<th>% variance</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-1 / V+</td>
<td>62.8</td>
<td>2.86*</td>
</tr>
<tr>
<td>V-1,V+ / E+ , V-2</td>
<td>30.8</td>
<td>1.40</td>
</tr>
<tr>
<td>E+ / V-2</td>
<td>7.2</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

* None of the $F$ statistics approach the level of significance $p = .05$.

Table 3

Variance analysis of the sensory conditions factor for T1

To test a learning effect within each block of trials under the different conditions of sensory feedback we compared the five first trials for each angle $\alpha$ to the five last ones - the t-test showed no significant learning effect between trial $t(78) = 0.28; p > .10$.

- **T2 analysis.** The same variance analysis was computed for the time to complete insertion. Significant differences were found between operators: for subject 1 $m = 67.3$ sec. and for subject 2, $m = 38.9$ sec. $F(1,9) = 26.23; p < .01$.

As for T1 there was a significant difference between the various sensory conditions: $F(3,27) = 8.90; p < .01$ but their interaction was not significant $F(3,27) < 1; p > .10$. We computed both S1 and S2 to analyse the sensory conditions. The orthogonal decomposition on the three degrees of freedom was as follows:


<table>
<thead>
<tr>
<th>Comparison</th>
<th>% variance</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>V-1 / V_</td>
<td>27.6</td>
<td>2.45</td>
</tr>
<tr>
<td>V-1, V_ / E_+V-2</td>
<td>71</td>
<td>6.34*</td>
</tr>
<tr>
<td>E_+ / V-2</td>
<td>12</td>
<td>&lt; 1</td>
</tr>
</tbody>
</table>

* p < .01

Table 4
Variance analysis on the sensory conditions factor for T2

The t-test on a learning effect within each sensory condition showed no learning effect from trial to trial (t(78) = 1.53; p > .10).

The main result on task time as a function of sensory conditions is represented in figure 3.

Figure 3
Task time in seconds as a function of sensory modality
8. Discussion. In agreement with the predictions we obtained a learning effect with force feedback for insertion task and not for the displacement adjustment task. Electrotactile stimulation provided to the operator K.R. that could be associated to motor response specifications and sensory informations.

In fact it was difficult to make a clear distinction between learning and the role of the sensory aids. The comparison between tasks executed in which sensory aids are not relevant for the operator and one where it could be of use gave us a first estimation of the utility of such information. However, with this procedure no clear distinction could be made between electrotactile and visual stimulation no could the improvement be quantified in terms of theoretical quantity information.

The first result that could be interpreted concerns the decreasing insertion time for the trials performed with sensory aids including electrotactile stimulation and revealed by the variance analysis (table 4). If sensory aids were not used by the operator this learning effect should be due to a general learning factor or learning task. However, for the displacement/adjustment task where force feedback was not relevant, the learning effect occurs on the first trials but not for subsequent trials with electrotactile stimulation. Then we can hypothetize that this task time revealed a gross learning effect which, in other terms, could be a basic learning effect.

The difference in learning between the two tasks in favor of the insertion task could be explained by the effective use of sensory aids, and in particular the electrotactile stimulation. Electrotactile stimulation was pertinent for this task and effectively treated by operators.

The second result concerns the learning effect observed between the different blocks and not trials to trials. This result was observed in a preliminary experiment using electrotactile feedback and has no clear interpretation for the moment. The standard deviation in these experiments gives us the best explanation for this effect. It showed that trial to trial operators had a great variability in completion time indicating that it was not possible to observe finer effects with this procedure. So we developed a mixed sensory condition paradigm trial to trial to compare visual and electrotactile feedback starting from the idea that the latter one provides one kind of information which could be centrally processed.

Experiment 2

The second experiment was designed to test performances in a telemanipulation task with visual or electrotactile feedback on the sum forces as in the first experiment. Its purpose was to make a first comparison between the two sensory modalities when manipulations and to evaluate the theoretical information quantity processed by the operator in both sensory modalities.

1. Material. The manipulator was the same as the one in the previous experiment but the transducer was different. This transducer was a "syntaxer" with six simultaneously controllable degrees of freedom. But only three of them were used for this experiment. The three translations were operated with a velocity control by three micro-displacements of the handle.
These micro-displacements required a certain amount of forces to be exerted on the syntaxer.

2. Task. The task consisted of the displacement of a cursor along an orientable axis that could be rigidly fixed with three different inclinations and twelve horizontal orientations (Photo 4). The cursor was linked to the axis by a self-alignises bearing to avoid variation of sliding friction.

The task required the operator to precisely control manipulator movements in the space and to guide cursor held by the gripper in a direction strictly parallel to the orientation of the axis. If the gripper was not guided in parallel with the axis, an increasing effort was generated and the cursor could not be conducted to the limit switch at the end stop. The sliding distance remained the same all over the experiment.
3. Sensory feedback. The sensory feedback were the same as those in the previous experiment - visual and electrotactile feedback were alternatively used to inform the operator of the summest forces exerted by the manipulator on the axis.

4. Measures. The execution time was measured automatically at each trial, when the cursor begin to move a micro-switch was released, starting a chronometer and when the cursor load moved about 40 cms on the axis a record micro-switch was activated stopping the chronometer.

Another way to finish a trial was by the triggering of an automatic safety stop when sum of the forces exerted a certain level. Then the syntaxer was automatically uncoupled to the manipulator and the trial was considered as a failure.

5. Procedure

Three conditions of sensory feedback were permuted on 30 trials. 10 trials were run with visual, 10 with electrotactile and 10 without force feedback. The experimental design was $S_3 \times M_3 \times T_{10}$ with $S$ for subjects, $M$ for the sensory modality and $T$ for the number of trials. Each subject found the apparatus and had to slide the cursor on the axis as rapidly as possible without error. Of the 30 trials, 21 requirest the control of three degrees of freedom and 9 requirest only two degrees of freedom. The orientation of the axis was changed for each trial and there were never two identical orientations in the 30 trials. 2 subjects were the same as in the first experiment and the third was an experienced operator who had never participated in an ergonomical experiment. Before each trial the experimenter fixed the axis with a new orientation and positioned the gripper on the cursor at the starting position. Then the subject grasped the syntaxer and started the trial controlling the movement only by sight or with the addition of sensory feedback in 20 trials out of 30. After each trial he was informed on the execution time task.
6. Experimental predictions

Two complementary effects on the operators were hypothetised to exist, permitting one to make predictions on the task time and the number of errors:

1. A visual dispersion of information, where, for practical reasons, the visual display was mounted at a distance of 20 cm from the gripper.

So if the operator had to control both the course of the cursor on the axis and the visual display, the visual load might exceed the visual information acquisition capacity.

2. A visual dominance effect. Attention will be focused on visual information when competing with other sensory information (Rock and Victor 1964).

These two hypothetised effects were derived from the general assumption: treatment time increases with the quantity of information and therefore execution time, will increase with uncertainty in the system.

The following predictions were made if we predict a visual dispersion effect task time will be longer under visual feedback and if we predict a visual dominance effect error number will be higher under electrotactile feedback.

In other words when the sensory conditions are homogenous (with visual display) information about summer force will be correctly processed by operators and attention focused on this channel allowing a low number of errors but task time will be longer because of the spatial dispersion and the large quantity of information processed. The sensory conditions are heterogenous (with electrotactile feedback) attention will be focused on vision to control the course of the cursor and the forces feedback will not be well processed. The number of error will be greater and task time will be shorter because a small quantity of information is processed.

7. Results

The main results were the following for the three subjects taken together we noticed a high correlation between task time and error number.
Figure 4
Task time and error number (between parentheses) for the 2 and 3 degrees of freedom tasks.

A variance analysis on task time showed a significant interaction between subjects and the sensory conditions for both 2 and 3 degrees of freedom. \( F(4,8) = 6.44; p < .05 \) and \( F(4,24) = 4.5; p < .01 \) respectively.

The results for the three subjects are summarized in table 5.

<table>
<thead>
<tr>
<th>Sensory condition</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean task time</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>20.07</td>
<td>6.66</td>
<td>5.29</td>
<td>10.09</td>
<td>17.23</td>
<td>17.27</td>
<td>16.65</td>
<td>17.05</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>12.42</td>
<td>10.87</td>
<td>6.98</td>
<td>23.99</td>
<td>10.35</td>
<td>9.03</td>
<td>14.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>24.46</td>
<td>15.50</td>
<td>15.38</td>
<td>18.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Error number</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>18</td>
<td>8</td>
<td>2</td>
<td>28</td>
<td></td>
</tr>
</tbody>
</table>

Table 5
Task time and error number for the 3 subjects
8. Discussion. The basic results for discussion are execution time and error number under the condition without force feedback. For the 2 degrees of freedom task very few error were registered (1 error for 27 trials) indicating the task was easier to control without force feedback. So electrotactile and visual display had no necessity for the operator and were not relevant to task performe. However, for the 3 degrees of freedom task forces feedback was necessary (28 errors for 63 trials) and a comparison between electrotactile and visual modalities could be made. As predicted, increasing the theoretical information quantity tends to increase the task time and reduce the error number but no statistically significant differences were found on the modality factor. Interaction between subject and modality suggests that different operators had different strategies to resolve modality competition. The behavior of subject 1 seems different than that of subjects 2 and 3. However, in all cases the visual dominance prediction is confirmed in this experiment where under vision feedback there was no error for the 3 operators and 4 errors with the electrotactile display. This result could be interpreted as an effect of information quantity processed by the operators and suggests that under the electrotactile modality operators process less information than under the visual modality.

Another result concerns the lack of a visual dispersion effect on task time. So task time seems to be dependent of cognitive factors such as information acquisition and treatment capacity.

In conclusion the results obtained in these two experiments could be summarized as follows. The first one concerns the effect of electrotactile knowledge of the results on learning telemanipulation task. This kind of stimulation provides to the operator informations about the results that could be associated to motor responses and sensory information.

The second one concerns performances in controlling telemanipulation task. Increasing quantity of information in the system augmented task time. It means that if the information returned to the operator is effectively processed the time to treat this information and make motor decision is increased. The electrotactile stimulation of the skin provides information effectively processed by the operators, although the quantity of information centrally processed could be less than with an equivalent visual stimulation.

At first we suggest that an electrotactile display informing operator on forces exerted by a manipulator could replace other ways to give a K.R. and especially a visual feedback on a force feedback on the control handle. The result is certainly better than without any information about forces. But if electrotactile feedback is partly informative for the operators of force, it is probably not as informative as a visual display that or force feedback directly on the handle. One suggestion is that electrotactile stimulation is processed as a discrete all-or-nothing information and visual stimulation as an analogical one with a better discrimination.

In other terms an electrotactile stimulation would provide less information than a visual one. From a cognitive point of view this means that electrotactile information bypasses the sensory memory register and has a direct acces to short term memory. Presumably, the limited information
treatment capacity and the retention delay in short term memory do not allow the operator to completely treat all the information provided. Therefore, some information will be lost before it can be treated and transferred to long term memory.

The apparatus and the method we developed to study electrotactile feedback in telemanipulation ergonomics has provided an approach to study such cognitive factors as attention, memory and decision making. Most likely, these factors are of much importance if we want to use sophisticated systems. Of course these two experiments are not sufficient to answer the many remaining specific questions about finer cognitive processes.

Other experiments should be made to test the "local model" theory about processes involved in telemanipulation. However, we hope that this first approach is sufficiently informative about the possibility to use electrotactile feedback and opens the way for more psychological studies.

Acknowledgments

I wish to thank H.H. Kwee and N. Quétin for their help and comments during the preparation of this paper. This research was supported by the direction générale des équipements techniques contract N° 78606.
REFERENCES


Computer Aided Treatment of Patients with Injuries of the Spinal Cord

by


Abstract

The rehabilitation process of spinal cord injury patients is carried out by a multidisciplinary team. Such a step assures that all data needed for the treatment are available; however, it also creates the problem of the exchange of accurate and up-dated information to all members of the team. In order to solve this problem a study of the entire rehabilitation process has been made; that is the dynamics of the treatment itself have been modelled, and the information streams in relation to the organisation structure have been traced.

In applying the model developed, a digital computing facility has to be introduced in the rehabilitation center. In order to be aware of the consequences of such an introduction a study of the information flows and of the organisation has to be executed, since it is important to know how strongly automation of the data processing could act on the functioning of the individual staff members. Finally, the consequences as far as they can be overlooked at this moment will be discussed. A first experiment with the introduction of a demonstration system in the clinical setting will be elucidated.

1. INTRODUCTION

The World Health Organisation defines rehabilitation as the multidisciplinary process to bring, and to keep, handicapped patients in a state of health in which they can function in the most optimal way in their social and physical environment. Hence, in dealing with spinal cord injury patients, the rehabilitation process is carried out by a multidisciplinary team consisting of members from the medical occupational and physio-therapeutical, the social and psychological, and the technical disciplines. Such a set up assumes that all data needed for the treatment will become available to all team members, correctly and in time. It thus requires an accurate exchange of up-dated information to all team members. Because the multidisciplinary aspect in rehabilitation medicine is such a basic point, a detailed study was set up with the following goals in mind.

- To achieve a better understanding of the dynamics of the entire rehabilitation process.
- To study the possibilities of direct digital data processing of the treatment data.
- To propose, if worthwhile, a data acquisition system in order to automate the data processing within the rehabilitation center. This proposal should include the consequences in relation to the organization, the data transfer lines and the different tasks of the personnel.
At this stage the first two goals have been realised, the last one is under study and a great deal of the work has been done.

2. THE REHABILITATION PROCESS

The study of the rehabilitation process of spinal cord lesions was executed at the rehabilitation center "De Hoogstraat" at Leersum, The Netherlands. The study started in 1973 and can be divided into five main phases, which are listed as follows:

- Recording of the data concerning the treatment of the patient and his/her state of health.
- Description of the state of health of the patient, and data reduction
- Development of a prognosis model of the future state of health of the patient.
- Analysis of the rehabilitation process dynamics on the basis of the treatment (input) and the state of health (output).
- Prediction of an optimal treatment plan.

In the period of 1973-1975 58 patients were observed on a weekly basis; 74 treatment variables as well as 148 state of health variables were recorded. The average period of treatment of the patient was about 43 weeks. The data collection was carried out by the treatment team members. The data can be classified as input and output variables concerning the medical aspects (84 and 39 variables), the Activities of Daily Living (41 and 19), the psycho-social variables (6 and 7) and the support of assistive devices and home adaptations (3 and 19), respectively. The medical variables were related to aspects like urinary tract inflammation, pressure sore, and para-osteoarthropathy; the ADL variables concerned activities like dressing, eating, personal hygiene and wheelchair riding; the psycho-social variables described the relation with the partner, the family and the team members, emotional stress, etc.; and finally the support variables represented the progress in the support of wheel chairs,

Figure 1: Block diagram of the rehabilitation process of a patient with a spinal cord lesion.
orthoses and home adaptations. The input variables were quantified on the basis of the duration of the treatment and the kind of treatment. To be able to compare the completely different output variables with each other, all variables were scored at a scale from 0 to 5 in the order of a decreasing restriction that a particular variable may have on the functioning of the patient, i.e. 0 is impossible, 1 is very restrictive, 2 is restrictive, 3 is lightly restrictive, 4 is may become restrictive and 5 is not restrictive at all.

The data reduction of the output variables was achieved by the following procedure. By using the knowledge and experiences of the clinical staff a number of variables could be skipped because they did not provide additional information. Furthermore a number of variables did not change at all, and so they did not give any dynamic information about the treatment process. From the 148 variables only 67 remained. As a next step a covariance analysis was carried out, resulting into 12 major independent variables, i.e. pressure sore, micturition, paraosteo-arthopathy, dysregulation, spasticity, communication and personal hygiene, wheelchair activities, ability to stand in an upright posture, period the patient is out of his bed, acceptance at home, mood, and the relation with the members of the treatment team. In this list of variables the support variables were left out, because they turned out to be just a result of the medical, the ADL and the psycho-social variables. The final result was checked by means of the Delphi ranking method. The variables were judged by 20 specialists on the basis of the following three aspects: Direct danger for life, clinical treatment, and post-clinical function. Although a reduction of the twelve variables just mentioned to only four, being the medical, ADL, psycho-social and support variables, was found to be too rough for practical use, Fig. 2 shows as an example what rehabilitation medicine actually means: Living with your handicap (no increase or decrease in the medical state), learning to use what is left (a strong increase in ADL performance), becoming a psycho-socially stable person (only small variations in the psycho-social state), and receiving the necessary support in assistive devices etc (increase in the support variable).

The figure also shows very clearly that the support variable may be considered to be a result of the progress of the other variables. Mostly a time at least
10 weeks is found, mainly due to the procedures necessary to obtain financial support.

From Fig. 2 one learns that the discharge of the patient from the rehabilitation center is mainly determined by the moment the home adaptations and assistive devices are available, since in general most of the other variables have reached a steady state value at least 10 weeks earlier. It was therefore worthwhile to investigate whether it is possible to predict the medical, ADL, psycho-social and support variables over a 10 weeks period, so that possible supplies can be arranged in time. Two prediction methods have been studied, i.e. one based on a description of the output variables by a Semi-Markov Series, and one on the state variable description.

The dynamics of the Semi-Markov Series are given by: The probability that a variable makes a change from a certain value i directly to another value j at any time, and by the probability that the variable has a value i during (n-1) weeks before a transition is made to the value j on week n. In using this description it is assumed that the future changes of a variable describing the state of a patient depend only on the present state of health and on the amount of time the patient has been in it. By counting the transitions in the measured data, the probabilities can be easily found. Hence, given these probabilities based on a population of patients, and given the data point over a certain period of a particular patient, a prognosis can be made. Fig. 3 shows an example, where

![Figure 3: The prognosis over a 10 weeks period of medical, ADL, psycho-social and supply variables based on a 10 weeks record of a particular patient.](image)

for four variables based on the data measured over a 10 weeks period a prognosis for the next 10 weeks has been obtained, whereas Fig. 4 shows a prognosis over 28 weeks based on the previous 10 or 20 weeks, respectively.

A more sophisticated method to predict the future value of the variables is based on a state variable description. Since input and output variables have been measured the dynamics of the entire rehabilitation process can be determined. Hence with the model $y(k+1) = Ay(k) + Bu(k)$, where $y(k)$ is the state variable as well as the output
vector, \( u(k) \) is the input treatment vector, \( A \) is the system matrix and \( B \) is the input matrix, the state variable can be predicted given a certain treatment plan.

Figure 4: Prognosis of an ADL variable over a 28 weeks period based on the data measured over the previous 10 or 20 weeks.

The matrix \( A \) and \( B \) can directly be found from the input/output data measured. Fig. 5 gives an example of such a prediction. Of course, if the treatment plan is a standard one, this method can help in overseeing the consequences of a particular treatment for a given patient. However, often it is the treatment plan itself that should be defined. Thus the problem should then be formulated as an optimization problem. Given the rehabilitation process dynamics, and

Figure 5: Four predictions of state variables of a particular patient on the basis of the initial state at the second week of treatment and a given treatment plan.

given the inherent limitations in treatment possibilities, one has to optimize a cost function concerning the benefits of the treatment on the one side and the cost of the treatment on the other. In this way one can conclude to an optimal
treatment plan; this final step is presently under study.

Although not all details of the study can be given here, a number of interesting conclusions has been reached.
- The introduction of a cybernetic modelling approach in the rehabilitation center yielded that the team members became to realize that a systematic recording by means of checklists at regular times increased their understanding of what the team was doing actually. Moreover it opened the possibility to evaluate the rehabilitation process. One also became aware of the efficiency of uniformly quantified information exchange in a multidisciplinary team.
- The 153 variables originally recorded in order to describe the patient's state of health could be reduced to only 65 variables without any loss of information; out of these measurements twelve state variables could be determined for a sufficient description of the patient's state of health.
- The medical, ADL, and psycho-social variables show that at least 10 weeks earlier a patient could be discharged from the rehabilitation center, if in time the supply of assistive devices was started. The prediction models were such promising that it was believed that on the basis of this model the treatment period could indeed be decreased by ten weeks, whereas the quality of care will remain the same.
- The description of the rehabilitation process dynamics was felt to be so attractive and accurate that a proposal of an optimal treatment can be a worthwhile contribution as an input to the treatment team, in this way aiding in making the final treatment plan decision.
- The goals of the treatment (Fig. 1) seem to be dependent on the results obtained, that means the goals during the entire rehabilitation process may change due to new circumstances. Besides that, patient and team members should have equal expectations about what may be reached. Therefore a goal check list has been developed which has to be filled out by patient and team separately, once every three months. In discussing the differences between the goals of the team members, including the patient, one harmonizes the joint goal of the treatment team. The results of the introduction of these check list were so successful that presently the method is fully accepted as a standard procedure.

3. THE DATA PROCESSING IN THE REHABILITATION CENTER: HUMAN FACTOR ASPECTS.

The overall impression of the approach followed raised high expectations, so the introduction of an on-line data processing system was considered. However, the introduction of such a system means much more than just a kind of automation of data handling. From the viewpoint of man-machine communication it yields a number of different aspects, among which the most important ones are:
- With the introduction of a data processing system one starts to control the information flows along carefully planned pathways, probably different from the existing ones. This might lead to situations where people are not or badly informed, whereas others are informed in different ways.
- The information pathways are directly related to the organization structure of a clinical center. Automation thus may have its influence on the existing organization structure, that is the formal as well as the informal structure.
- If the information flows are changing, and as a result hereof the organization has to be adapted, tasks of different staff members may be influenced. Hence, the consequences of such an introduction on the personnel's tasks and
functioning should be carefully taken under consideration.

- The introduction of computer systems in a non-computer minded environment may lead to a number of interesting problems. Firstly, there is the problem of making the staff familiar with the new possibilities of the system, so that not old methods are used with new equipment, but that methods and equipment are used in an optimal way. Then, secondly, one has to train the fully naive staff in the use and operation of the system.

- Protection of privacy is a very important item in data processing in health care. Often it is wrongly believed that privacy protection will degrade when computers are used.

The five above mentioned items have been studied extensively during the last few years, although the entire project certainly will request another couple of years. To start with, a detailed study of the information structure as well as the organization structure at the rehabilitation center has been carried out, in this covering the first three items. Parallel hereto a demonstration of the possibilities of an information processing system has been worked out in order to get some feeling about the last two items. In discussing the information and organization structure the emphasis of this paragraph will be on the methodology of the study, because the results are so closely related to the particular setting at the rehabilitation center that in understanding the results a detailed description of the rehabilitation center is required.

Considering only the processing of the treatment data of the spinal cord lesion patients it is necessary to oversee the influence of modern information systems on the information exchange and the organization. This has been achieved by applying the ISAC method developed by the Department of Information Processing and Computer Science of the Royal Institute of Technology and the University of Stockholm, Sweden. The ISAC method stands for the Information System Work and Analysis of Changes. Due to the, in general, high complexity of the problems studied, the method distinguishes five interactive phases. Each phase, individually, leads to a set of documentation, i.e. graphs and descriptions. Starting with a global review graph of the entire system one zooms in to a set of graphs of the different subsystems. The five interactive phases are: (1) Change Analysis; (2) Activity Studies; (3) Information Analysis; (4) Data System Design; and (5) Equipment Adaptation.

The Change Analysis deals with the study of possible changes (improvements) in the activities of the organization. These possible changes are based on the insight gained about the organization, the observed problems and the problems to be expected in future.

The major purpose of the Activity Studies is the definition of the information subsystems, that is the system of information processing and not the actual treatment process. The results consist of a graph as well as a list of information subsystems, a cost-benefit analysis, the required financial means, a time schedule and a definition of the interaction between the subsystems. This phase will result in a final decision to continue or to stop the project.

The Information Analysis phase deals with a detailed description of all information subsystems. Three succeeding analyses have to be executed. During the Precedence Analysis one tries, starting from the output information sets, to find which preceding information sets are the roots of the outputs. This process is continued until the input information sets of the subsystem have been reached. The next step is the Component Analysis, during which an exact description of the information sets is obtained. Finally, a Process Analysis is executed, that is a description is given in which ways and under which restrictions the information sets are transformed into other information sets.
The last two phases, Data System Design and Equipment Adaptation refer to the realization of the information processing system. The DSD deals with the design of the system, that is the design of the structure of the system, the design of the required components and the data processing method - automatic or manual or a combination; batch or on-line or in-line etc -. During this phase the component and process descriptions are translated into data structures and procedures. The final step of the ISAC method is the adaptation of the data system design to the computer configuration chosen.

Figure 6: The activity graph and information graph of the treatment process of spinal cord injuries.
In the manner as described before the entire information and organization structure of the rehabilitation center can be depicted by graphs. As an example the activity graph and the information precedence graph for the treatment process of patients with a spinal cord lesion will be discussed now.

The activity graph: Before one is able to start the treatment of a patient an examination has been executed, moreover data from the hospital have been collected, hence at the beginning of the treatment a certain amount of data is available. With this information, the patient and the necessary means, one can start the treatment. During this process additional data will be collected on the basis of which the treatment can be evaluated, so that a decision can be made whether the patient will be discharged or whether the treatment has to be continued.

The information precedence graph: In order to be able to execute the rehabilitation process a perfect information exchange between the members of the treatment team is necessary. The information precedence graph illustrates how an information set is reconstructed from preceding sets. For instance, in order to supply a particular assistive device one needs a prognosis as well as data about the treatment process itself. The treatment data is constructed from the rehabilitation records, treatment instructions, and collected treatment data. A summary in which also the prognosis is included will result in the updated rehabilitation records set. The instructions for the different disciplines can be derived after the treatment goals as well as the treatment plan are known. Combined with the information gathered during the first examination as well as the rehabilitation data recorded one obtains updated rehabilitation record. Finally the set discharge information is obtained by a further evaluation of the treatment data.

Although the method discussed is a time consuming procedure, because the treatment of the bodily handicapped patient is such a complex and multidisciplinary process, the method certainly leads to a very clear insight in the structure of information and organization of the clinic. There is almost no need to elucidate how important the results are with reference to the final design of human tasks, to the introduction of an information processing system and to the man-computer communication.

4. A PRELIMINARY MAN-COMPUTER-COMMUNICATION STUDY.

In order to get more insight in the direct problems related to the introduction of a computer system in this particular situation, it was decided in an early stage to set up a simulation so that practical experiences in a clinical setting could be gained. The most important reasons can be stated as follows:

- It was necessary in an early stage to test the dialogue design, when working with naive users.
- In literature it is emphasized that it is very important that from the beginning the final users of a system are involved in the project. Their feedback is very essential, and their involvement is crucial for a successful introduction of a computer system.

For a dialogue between man and computer several possibilities exist. Several authors distinguish between a number of different techniques, ranging from fully Computer initiated to fully Operator-initiated; or as seen from the operator ranging from passive to active dialogues. For choosing a particular kind of dialogue the following criteria are given: Level and kind of education of the users of a system; frequency the system is used and kind of computer application.

For the case of the rehabilitation center the so-called menu-selection seemed
to be an appropriate choice. This is a passive dialogue, where the user has to choose from lists of possibilities presented on the screen, in order to find the information requested. This menu-selection is an easy-to-learn dialogue style, but it will become boring for experienced users. Hence, for them possibilities to reach the information, bypasses, have to be provided. Important other aspects in the design of a dialogue for naive users are: The system has to be foolproof, the quantity of input-data should be minimised, and the application of abbreviations for data input is very desirable, if and only if, strict rules are given to which an abbreviation is formed. Important ergonomic aspects are: The fact whether the information is presented line by line or page by page, the use of capitals and lower case letters; the use of special features such as reverse video; the use of special-purpose terminals as often is advised; the quantity of information per page and the standardization of the lay-out on the pages.

Based on the literature reviewed a small demonstration test system has been realised on the university's IBM 370 system with a Beehive-terminal. This terminal had a reverse video facility, only one programmable function key could be applied. The dialogue was developed in CPS, an interactive subset of PL/I. The structure of the test system is shown in Fig. 7. Every block in this figure represents a page on the screen. On the first page the user has to identify him- or herself

Figure 7: General lay-out of the demonstration system.
(Privacy protection), then one has to state whether one wants to follow the bypass or the menu. The menu consists of: A list of patients, a list of files of every patient, and a list of chapters of a particular file. As an example the goal check list as mentioned in paragraph 2 was worked out. Then, finally one had to indicate whether one wanted to fill out or just to observe the goal check list. The reason for distinguishing two versions of the goal check list in the menu is the fact the observation of the list as shown in Fig. 8 is easy to interpret, whereas during the fill out phase the positioning of an X, P or T (Fig. 8) by a cursor takes more time and leads to more mistakes than just typing in a code from 1 through 5. Hence in the observe version the data are printed in columns, and in the fill out version the data typed in code. Besides this menu selection method the system was set up in such a way that via the bypass one also could reach the data. In this interactive program one could use abbreviations.

As mentioned before; an important aspect in computerizing the data handling is privacy protection. Rules have to be formulated as answers to the questions:

- Who is allowed to enter a new patient in the data base, and who is allowed to enter a particular data about a patient? How can one prevent that people who are not authorised to do so, will change data?
- Who is allowed to read the information, and how can we prevent possible misuse of the system?

<table>
<thead>
<tr>
<th>COAL CHECK LIST</th>
<th>DATE: 10-09-1980</th>
</tr>
</thead>
<tbody>
<tr>
<td>PATIENT NAME</td>
<td>BAKKER</td>
</tr>
<tr>
<td>STATUS XR</td>
<td>0015</td>
</tr>
<tr>
<td>ADMISSION DATE</td>
<td>05-22-1980</td>
</tr>
<tr>
<td>ANAMNESIS</td>
<td>TETRAPLEGIC</td>
</tr>
</tbody>
</table>

**QUESTIONS**

<table>
<thead>
<tr>
<th>ADL</th>
<th>1 2 3 4 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>INDEPEND. IN/OUT OF BED</td>
</tr>
<tr>
<td>14</td>
<td>INDEPEND. FROM FLOOR TO CHAIR</td>
</tr>
<tr>
<td>15</td>
<td>PICK UP ARMREST FROM GROUND</td>
</tr>
<tr>
<td>16</td>
<td>GET UP FROM GROUND</td>
</tr>
<tr>
<td>17</td>
<td>USE TOILET WITHOUT HELP</td>
</tr>
<tr>
<td>18</td>
<td>USE TOILET WITH HELP</td>
</tr>
<tr>
<td>19</td>
<td>INDEPEND. DRESSING UPPER BODY</td>
</tr>
<tr>
<td>20</td>
<td>INDEPEND. UNDRESSING UPPER BODY</td>
</tr>
<tr>
<td>21</td>
<td>INDEPEND. DRESSING LOWER BODY</td>
</tr>
<tr>
<td>22</td>
<td>INDEPEND. UNDRESSING LOWER BODY</td>
</tr>
<tr>
<td>23</td>
<td>INDEPEND. WASHING UPPER BODY</td>
</tr>
<tr>
<td>24</td>
<td>INDEPEND. WASHING LOWER BODY</td>
</tr>
</tbody>
</table>

**CONTINUE?**

line in reverse video

*Figure 8: The display format of the goal check list.*
The first question implies that the system must recognize who is on the staff of the center, both questions imply that the system must identify who is who. A safe method is to make use of pass-words, a frequently used method in banking operations.

21 subjects, equally distributed over the different disciplines of the clinical staff tested the system voluntarily. Their opinions about the system were collected by watching the subjects during the training sessions, and by questioning them by means of check lists. Two classes of opinions were gathered, i.e. one class about the use of computers in a rehabilitation center in general, and another about the use of this particular system. The first check list had to be filled out before the training session, the second one at the end of the test period.

The inquiries showed that the group, although representative with reference to the distribution over the disciplines, was certainly not representative with reference to their attitude to computers. The majority judged positively about automation of the data handling, whereas it was known that a large group was more reluctant. The results can be summarized as follows:

- Speed of data transport should be in the order of 9600 baud; the system response time should certainly be in the order of a fraction of a second and in the case it takes longer it should be indicated by a message like "Your data are now being processed"; the reliability of the system should be high, and an indication whether the system works or not is essential for the use with naive subjects.
- The idea of working with pass-words appeared to be working well, although flexibility due to shifting staff members is very important; menu selection is a very attractive method for infrequent users and beginners, or as reference possibility, good bypass possibilities are certainly required for the experienced user; although only one special function key was available, the application of more function keys such as, a Begin-, Stop-, Correction-, Next Page- and a Break- key seems to be valuable; the dialogue should be user-friendly.

5. ACKNOWLEDGEMENT

The authors would like to acknowledge all members of the treatment team for spinal cord lesions of the Rehabilitation Center "De Hoogstraat" for their very enthusiastic assistance in collecting the necessary data, and their stimulative contributions in the form of workshops and discussions. Also the 58 patients, who were subject of this study on the one side, and data collectors on the other, should be acknowledged. Without their positive attitude this study could not have been done.

6. LITERATURE


3. Hoogendoorn, R.:
   The efficiency of the Multidisciplinary Treatment of Bodily Handicapped
   (In Dutch), DUT, 1979, 76+159+58 p., N-167/R-546.

4. Lundeberg, M; G. Goldkuhl; A. Nilsson:
   of Techn. / Univ. of Stockholm, 1980.

5. Martin, J.:
   Design of man-computer dialogues. Prentice Hall, Englewood Cliffs,

6. Hebditch, D.:
   Design of dialogues for interactive commercial applications, In:
   Proc. Infotech.Conf. on man-computer communication, Amsterdam, 1978,
   22 p.

7. Rouse, S.H.; W.B. Rouse:
   Design of a model-based on line management information system for
   interlibrary loan networks. Information processing and management,

8. Stewart, T.R.M.:
   Communicating with dialogues. In: Proc. Conf. on ergonomics and the
   design of computer dialogues. University of Technology, Loughborough,
   1979, pp. 77-95.


10. Stassen, H.G.; A. van Lunteren; R. Hoogendoorn; G.J. van der Kolk;
    P. Balk; G. Morsink; J.C. Schuurman:
    A computer model as an aid in the treatment of patients with injuries
    of the spinal cord. In: Proc. Int. Conf. on Cybernetics and Society,
Abstract

The paper presents a method for operator competence development. Competence is defined as the ability of the operator to meet adequately all situations and states of the plant including normal operation, disturbances, transients and accidents. A set of closely related requirements and actions aiming for the acquisition and maintenance of operator competence is called the "competency system". There are the following parts in this system:

- Job description and job requirements.
- Recruitment requirements.
- Job training content development.
- Job training programme development including courses, training means and follow-up procedures.
- Requirements on plant training organization including training management and instructors.
- Registration, storing, processing and reporting of competency data.
- Recurrent checks and revisions of the "competency system".

A description is given of the procedures to develop a "competency system" of a plant under the following headings:

- Specification of some steering factors for a particular "competency system" solutions, including present operator recruitment and job training etc.
- Specification of job requirements through job analysis and descriptions.
- Specification of knowledge and skill requirements for operating the plant.
- Specification and implementation of a job training programme including training resources and a training organization of the plant.

The development of competency system for Swedish nuclear power plants was organized in work groups with participation from the Nuclear Power Inspectorate, the utilities and Ergonområd AB.
Background

Operator licensing for nuclear power plant personnel has been enforced by safety authorities in several countries, e.g. in the U.S.A. since 1954 and in FR Germany since 1974. In Sweden the Atomic Law has regulated nuclear power since 1959. This law gives the complete responsibility for safety to the reactor owner. It means that the reactor owner is also responsible for the personnel who operate the plant.

In 1976-77 the Swedish Nuclear Power Inspectorate carried out a feasibility study on the regulation of operator competence. In July 1978 a project was given to Ergonområd AB to develop principles for operator competence under the condition of the Atomic Law and under the condition that there are utilities with different operational organizations, operator recruitment and training traditions.

A method including principles and procedures for specifying operator competence was developed and has been implemented in ten Swedish nuclear power blocks since mid-1980. The work is done on behalf of the Nuclear Power Inspectorate which is the governmental authority for nuclear safety. The work is carried out in close cooperation with the utilities.

The method has a wider applicability than to nuclear power plant operators and can be used for safety and availability purposes in other processes and in other operator jobs.

Aim of the paper

This paper is a technical presentation of a method for the acquisition and maintenance of operator competence taking the above mentioned conditions concerning law and utilities into consideration. Some principles behind the method, procedures and an organization for the work involving the method are presented.

Operator competence

Competence in this context is defined as the ability of the operator to face adequately all situations and states of the plant including normal operation, disturbances, transients and accidents. The overwhelming proportion of these situations involve one or several processes and technical systems. The operator and his shift colleagues have to face this. But it is important to realize that operator competence is not limited to the technical side of the plant. It also engages professional contacts with other specialists from inside or outside the plant, e.g. maintenance people, fuel specialists, radiological specialists and guard of the plant. There is also an operator interface with documentation, regulations etc.

What builds up operator ability? The answer is, of course, complex. But a set of knowledge and skill belongs to the most important factors. This knowledge and skill is a necessary condition for success in the job. Thus, it is an important step to analyse and define them.

There are factors other than knowledge and skill which may play an important role in adequate operator behaviour, e.g. personality factors, control room factors, work procedures and work organization factors but these factors are not taken into consideration in relation to operator
competence. Some of the factors ought to be considered in separate studies and developments like control room design and man-machine communication. Other factors are difficult to take up because of lack of interest, opposition or large resources needed. Operator personality is probably such a factor. It may be very important under certain conditions, e.g. in stressful situations. But it is difficult to define relevant personality dimensions. It may cost quite a lot to develop adequate tests or measuring devices and there is probably a strong resistance among operators to them.

The purpose of the competency work was plant safety. But contrary to the distinct technical safety features, e.g. barriers in the plant, consequence reduction systems, redundancy and diversification of systems, it is not possible to separate between what is safety related and what is not safety related in the operation of the plant. It is a well documented experience from nuclear power plants that safety incidents or accidents often start with something which is not safety related. Through interrelations between systems a non-safety related incident may develop into a safety incident or an accident. Because of this, no distinct separation was made between safety and availability related operator competence.

A system concerning competence

A method for operator competence consists of a set of interrelated concepts which are called "the competency system". Part of the work has been allocated to the development of concepts and a structure for this system (1). The main features of it are described in figure 1.

![Diagram](image)

**Figure 1. Competency and training for nuclear power operators.**
A fundamental principle in the present method is that competency requirements of the operator should be based on what the operators actually do in their jobs. If one knows what the operators do in their jobs and knows the more important job circumstances like timing, task frequency and load one can also derive knowledge and skills needed for the operator to be able to carry out his tasks successfully. Thus, a starting point is the job analysis and job description. This is input for considerations on recruitment requirements and on job training requirements. The job training requirements are met through a job training programme, which contains all courses and other types of training which are needed to fulfill the training requirements. A training organization with sufficient resources to realize the training requirements is also an important part of the "system". Thus, the "competency system" contains the following parts.

- Job descriptions and job requirements.
- Recruitment requirements.
- Job training content development.
- Job training programme development including courses, training means and follow-up procedures.
- Requirements on plant training organization including training management and instructors.
- Registration, storing, processing and reporting of competency data.
- Recurrent checks and revisions of the competency procedures.

**Competency specification procedures**

There are four main procedure steps in applying the competency method to particular operator jobs.

1. Specification of steering factors for a particular "competency system" solution.
2. Specification of job requirements through job analysis and job descriptions.
3. Specification of knowledge and skills requirements for operating the plant.
4. Specification and implementation of operator job training programme(s) including training resources and the training organization of the plant (or other organizations involved in operator training).

**Competency steering factors**

Some important factors which will influence the particular competency system solution must be identified at the very beginning of the development. To a large extent these factors belong to what can be called company policy and inherited factors. These factors can be identified through describing certain present state conditions within the company. They should be considered in the competency work either through accepting them
as steering factors or through changing them when they are hindering efficient operator competence planning and development.

Information about these factors are found within the utility or the plant to which the analysed jobs belong. Data can be collected in interviews with plant or site management in the following factors:

1. Present operator recruitment; requirements, principles and procedures, estimated need for future operator recruitment.
2. Present operator training; basic operator training programme, retraining programme, courses, training aids like simulators.
3. Present training organization; who takes care of training within the organization and how, availability of instructors and training time.
4. Present operator situation; number of operators, operational staff organization.
5. Present operator competency; competency levels and categories, amount of on-the-job experience.
6. Objectives for operator competency; safety, availability and/or job satisfaction objectives, safety authority's, management's and operator's view of operator competency.

Procedures for the specification of job requirements

The job analysis technique used has been reported elsewhere (2). It is illustrated in Figure 2 which says that the body of knowledge needed to carry out a job can be defined through a limited number of tasks. These knowledge loaded tasks are called typical tasks. Together with some job situation demands like timing and precision a representative set of typical tasks forms the job requirements.

Figure 1. Typical tasks of a job and the body of knowledge.
The procedure to generate typical tasks is based on a system analytical approach. Starting in the analysis of the power generating system, which is described in a so called state diagram (see Figure 3) containing distinct plant states based primarily on the situation in the reactor and in the turbine of the plant.

Figure 3. States and procedures for a BWR nuclear power plant (see text below).

There are eleven distinct states:

a Reactor after refuelling
b Cold subcritical plant
c Heated subcritical plant
d Hot critical, reactor power 5% turbine not running
e Turbine at nominal speed
f Normal operation
g Disturbed operation
h Emergency operation
i Hot, tripped, subcritical reactor
k Hot, critical reactor, power 5%
l Hot, subcritical reactor
State transitions can be produced through manual or automation process control or through disturbances in the process or in the control system. The operator tasks are subunits to state transitions or in activities aiming for the preservation of some state:

**Transition Procedure**

1. General plant preparation
2. Preparation for start up (heating of reactor using residue heat)
3. Start nuclear heating and increase power to 5%
   - Start aux. feedwatersystem to control waterlevel in reactor tank
   - Heat steam pipes and continue nuclear heating using control rods
   - Dump steam to condenser
4. At 5% power, switch from aux. feedwater system to feedwater system
   - Bring turbine to nominal speed
5. Synchronization and loading of generator
   - Increase generator power to 20%
6. Decrease power to 5%
7. Shut down to hot subcritical reactor
8. Cooling by dumping steam to condenser
9. Cooling by dumping into containment
10. Incident causing disturbed operation
11. Return to normal operation after disturbed situation
12. Cooling of subcritical reactor
13. Start up of hot reactor
14. or 10.17 Incident causing emergency situation
15. or 12 SCRAM or manual shutdown
16. Refuelling
17. Change of control rod pattern
18. Increase or decrease of power level
19. Change of shift
20. Maintaining state b (residue heat cooling)
21. Maintaining state i to be able to perform transition 14 later

The state diagram is the basis for task generation, which preferably can be done through interviews with plant operators, supervisors, operator instructors and system engineers. Operators are an important source of task information when there is plenty of operational experience. If operational experiences are lacking, process engineers and control system engineers become the most important information.
The job analysis is conducted in four phases.

**Phase 1 - descriptions of the main system:** The aim of the description is to identify all interaction surfaces between the operator and the main system. The main system is not limited to the technical system for direct power production - which is the object in, e.g. the state diagram - but comprises the complete plant. Thus, the main system is considered an organizational system in which the technical process of power generation is a subsystem. To find the interaction surfaces of the three nuclear power operator jobs the main system was described in operational terms, in technical terms and as an organization.

Information about the goals of the plant were collected, too. Only qualitative goals were collected in the present work but as a general rule quantitative as well as qualitative goals should, if possible, be formulated. For example, quantitative goals like "80% plant availability" can be used to generate more precise operator requirements which can be useful, e.g. when training requirements are derived.

The work in this phase started with the localization of the analysed jobs. It is important to get a clear picture of where the jobs are situated within the organization of the plant. The analyst also has to get an overview of the general content of the jobs. Does a job include operation, maintenance, planning, supervision or other activities? The answer to this question indicates the type of system descriptions needed for the derivation of all interaction surfaces between the operators and the main system.

**Phase 2 - task generation:** The aim of the second phase is to generate operator tasks. This can be done in several ways. The most convenient way is through interviews with operators, supervisors and other personnel who cooperate with the operators. A matrix found, such as Figure 4, can be used to guide the interview. Along one of the axis of the matrix there is the state diagram (or mission profile) of the main system. Along the other there is a number of possible information surfaces like systems, documents and other personnel.

![Figure 4. A first matrix used for operator task generation.](image-url)
Phase 3 - job structure generation: The aim of this phase is to formulate operator tasks which are even, i.e. have the same or nearly the same degree of resolution. A rule of thumb is that the task statements should tell what is done in the task. A statement which tells how the task is done is too precise and means that the job content cannot be described with a reasonable number of tasks. The typical tasks will also be organized according to what they will be used for. The state diagram made them easy to communicate with the operational staff and training planners of the utility. All the collected task statements should be evaluated by an experienced operator. It is important to reformulate statements which can be misunderstood.

The outcome of this phase is a set of preliminary typical tasks, which will represent all operator functions or main activities of the analysed job.

Phase 4 - performance requirements generation: The aim of the fourth phase is to evaluate the relevance of each preliminary typical task and to formulate performance requirements. This can be done through contributions from the interviews mentioned above in phase 2 - task generation. Each typical task is judged on relevancy and operator performance. Also in this phase badly formulated tasks can be reformulated and a few new tasks can be added to the set of preliminary tasks.

The typical tasks generated in the job analysis are the main result of the job analysis. The set of typical tasks in a job also represents the knowledge and skill content of the job as was illustrated previously in Figure 2. If the operator knows these typical tasks he knows the job which means that he fulfills the competency requirements.

The job analyses performed in nuclear power plants resulted in around 150 unique typical tasks each for the turbine operator and the reactor operator. The shift supervisor had around 100 unique typical tasks. There is a certain overlap in content between the jobs. The reactor operator must know some of the tasks of the turbine operator. The shift supervisor must know all the tasks of both the reactor operator and the turbine operator. As deputy supervisor the reactor operator must know an extensive part of the tasks of the shift supervisor. A sample of typical tasks for the reactor operator is found in Figure 5.

Specification of knowledge and skill requirements

The typical tasks generated in the job analysis describe what the operator must be able to do to carry out the job successfully in accordance with demands to run the station safely and with high availability. The second step of the competency method concerns the transformation of the typical task into knowledge content terms. The principles and procedures for this transformation has been reported previously (3).

Knowledge terminology

The aim is to express the typical task in terms which are relevant for the planning of personnel recruitment, operator training and follow-ups connected with these activities. A set of knowledge related terms was generated for analysing the typical tasks on knowledge and skill content.
for nuclear power plant operators. It is presented below and the terms are related to nuclear power plants. But it is likely that they can also be used in other process industries with minor modifications.

A. Knowledge categories: 13 knowledge categories were formulated. They were judged by training and operator experts to be relevant for the nuclear operator jobs. A relatively precise definition was given to each category:

1. Knowledge on plant layout.
2. Component knowledge.
4. System knowledge.
5. Process knowledge.
6. Reactor core knowledge.
7. Knowledge on localizing and identifying disturbances.
8. Knowledge on normal operation and measures at disturbances.
9. Knowledge on measures at plant fire, serious accidents and sabotage.
10. Organizational knowledge.
11. Administrative knowledge.
12. Knowledge on safety regulations.
13. Knowledge on supervision.
B. Knowledge object: 5 object categories were formulated. All objects within the categories are defined and are listed in the plant documentation (with exception for Actions):

(1) **Technical systems** according to, e.g. the System List of the plant.
(2) **Organizational units or persons** according to, e.g. the Organizational Chart of the plant.
(3) **Documents** according to, e.g. the Document List of the plant.
(4) **Disturbances** according to disturbance lists of the plant.
(5) **Actions** according to, e.g. the List of Typical Tasks of the jobs considered.

C. Knowledge depth: The depth of the knowledge or skill is defined in three levels:

3. **Thorough knowledge** or skill means learning to the extent that the material can be activated without use of instruction, advice or any other aid.
2. **Knowledge or skill** means that the material can be activated with use of instruction, advice etc.
1. **Orientation** means familiarity with the material normally without demand on performance.

Each one of the typical tasks from the job description are then analysed with regard to its knowledge content in terms of significant knowledge categories, knowledge objects and knowledge depth.

The analysis concerning nuclear power plant operators was performed by personnel of each utility representing the operational staff and training specialists. Rules and advice for the analysis were formulated by Ergonområd.

The outcome of the knowledge and skill analysis will be a list of typical tasks and its related knowledge content. Together they define the competency requirements of the job according to the present method. There are various ways to summarize and present these competency requirements.

The work on nuclear power operators resulted in competency requirements generated by the utilities and accepted by the safety authority. The requirements are being used by operator training planners and instructors within the utilities.

Within the training organization of the utilities the requirements are used in the evaluation and revision of the present operator training. New courses in the basic operation will, for instance, be added to the basic operator training due to new requirements. The requirements will also be used for design of a number of follow-up measures in the operator training.
Specification and implementation of an operator job training programme

The competency system concerning recruitment requirements and operator training is being implemented in all Swedish nuclear blocs.

Recruitment requirements have been expressed in terms of mathematics, physics, chemistry, technology and techniques. A High School education specially made for process technicians and operational personnel has been decided upon as the minimum basic education before entering job training for nuclear power plant operators.

The operator training offered by the utility should be based on this recruitment requirement. The operator training is divided into three categories:
A. Basic operator training.
B. Retraining.
C. Continued operator training.

The content of the basic operator training was specified in terms of typical tasks, knowledge categories, knowledge objects and knowledge depths. Quite an extensive part of the training is carried out in simulators, full scope simulators as well as more limited simulators.

The retraining is especially important in tasks, knowledge and skills which are seldom practiced on-the-job, e.g. in fault localization and identification and actions in disturbances and accidents. The need for retraining can be found through knowledge tests and questionnaires to operators. The retraining can be carried out, e.g. once a year.

It is important to realize that the content of a job in a large plant is never static. There are always new things, technically and organizationally, concerning regulations or new operational experiences which have to be taught to the operator. Therefore, there is also a need for updating of the operator in these new aspects. This training is called continued on-the-job training. It should be given with certain time intervals, e.g. a year.

The competency system also regulates how the utility shall follow up the individual competency. Different tests should be given to a student which will make it possible to demonstrate to the student himself and to the utility that the demanded knowledge and skills have been acquired.

The competency system developed for the Swedish nuclear power production also has some administrative procedures which made it possible for the Nuclear Power Inspectorate to fulfill its role as a safety authority.

The implementation of this competency system in the utilities has started from July 1, 1980. Each utility is responsible for the development of recruitment procedures, an operator training programme with courses and follow-up procedures and for a training organization including instructors, training aids and other resources needed to carry out the programme. It can be mentioned in this context that together the utilities are running a school for operator training which houses two so called full-scope training simulators.
Work organization for competency development

The present method for operator competency work is adapted to be used by the utilities and operators. The work on the development and implementation of the competency system in the Swedish nuclear power stations has been carried out in close cooperation between Ergonområd AB, the Nuclear Power Inspectorate which is the safety authority and the utilities. There has been a working group at Ergonområd and a working group in each one of the utilities. Every part of the system has been thoroughly worked through in the working groups before it is accepted by the Inspectorate and is sent back to the utilities as a regulation. The work organization is presented in Figure 6.

Figure 6. Work organization for the competency system.
These presented principles and procedures for competency in nuclear power plant operator jobs have a first order importance for system safety.

However, the applicability is not limited to safety in nuclear power. The concept of operator competency, consisting a number of interrelated factors as recruitment requirement, operator training and related follow-up procedures has a more general applicability.

The job analysis method can be applied more generally to operator jobs especially where there is a significant demand for safety or availability.

The principles for knowledge and skills analysis can be transferred without extensive modifications to other process operator jobs.

References


Abstract

Communications at the work face. The ability to provide Voice Input from the doctor patient interaction to a doctor's personal microcomputer as part of a clinical distributed network could solve many of the problems in the Health Services. It could also provide clinical knowhow as well as a more efficient and coordinated epidemiological and statistical service for dangerous transmissible diseases - research could be carried out during routine clinical care.

The paper discusses the needs, the options and the possibilities of such a functional approach to the application of computers to clinical care rather than hotel management.
Background, Health Communications

The health defence system evolved by the developed countries over the last century requires fast accurate communications between Patient, Doctor and Public Health Services. Information concerning dangerous social diseases, especially infectious diseases is passed to regional and national authorities so that the Public Health and Preventive Services can be alerted. Information both local and central is the essential basis of health care. Locally, at the clinical level, the doctor collects, stores and recalls vital data about the health of his patients to enable him to investigate, interfere biologically and control the disease process.

Time, medical science and a changing population have interfered with this method of healthcare monitoring. Extensive use of antibiotics over the last 30 years have removed infectious diseases, the main killers of the past. This practice of medicine and the improved standard of living have given rise to increased longevity, resulting in increased disease in middle and old age, cancers and cardiovascular episodes. These diseases are not officially notifiable and are only recognised after death by the authorities. In spite of the high cost of their morbidity, the monitoring of their course leaves much to be desired.

Retrospective studies on epidemics show that such proportions need not have been reached had the epidemic diseases been correctly reported and monitored. Too often incorrect data is passed to the authorities who are responsible for the nation's health and even when it is accurate; it frequently arrives too late. How are these vital data collected, collated, stored and distributed?

The Patient's Medical Record

The patient's medical record is the source document where all the activities of the doctor are recorded, it is the basic document in patient care. Fifty years ago any competent doctor could retain in his own memory all the data necessary for his clinical practice, but these days are long past. The complexity and divisiveness in modern medicine when practised en masse have thrown a tremendous burden on the traditional record. It has now become overloaded with laboratory printouts resulting in a very bulky, disorganised and unwieldy dossier, time consuming to read, expensive to compile, store and recall. It is frequently mislaid and difficult to study because the written section is often illegible. The doctor is the only person who can generate and record the clinical data, but because doctors and time are in short supply, many records are incomplete.

In order to maintain and extend the present clinical standards it is absolutely essential that a new type of record should be generated which mirrors the full consultation and:

(a) contains all the essential words and phrases used by the doctor;

(b) can be compiled without requiring any new skill or activity from the doctor. If possible he should have less to do and so have more time to have a relaxed relationship with the patient to assess his veracity.
(c) should supply information readily whenever required for the clinical care of the patient or for research.

It is also vital that the record is compiled in such a way that the logic of any decisions taken by the first doctor to see the patient can be obvious not only to the same doctor seeing the patient again but also to any other doctor who may act for him - this unfortunately is not so.

The Nature of the Computer Problem

Since the patient's medical record is the essential document in clinical and administrative care, a simple and easy computerised record should be developed to be so beneficial in patient care as to help overcome the doctors' prejudices against computers in general. The existing computer systems usually based on large main frame networks are expensive and suit only the dedicated doctors who have designed and implemented them. Unfortunately these doctors constitute less than 1% of the profession.

The patient's record can be an essay or a carefully designed keyword summary of the findings that form the basis of the clinical decision, delineating the logic of the diagnostic information pathway and the associated executive action pathway commencing when the patient is seen for the first consultation, until finally discharged.

For many reasons, most of which are now historical, the medical record has been avoided as far as possible in the various projects to introduce computers into the medical field. The result is that today computers are used largely as expensive desk calculators in three main areas: statistics; various laboratory services and by administration for hotel functions. Laboratory printouts are often the only computer facilities that appear in the clinical work area. If computers do appear for other functions in these areas the data is in inputted away from the actual clinical activity itself. This increases the liability of error and reduces confidentiality.

The sort of Luddite attitude of the majority of doctors to computers is understandable as we all resist change. This new technology has not been exploited to provide the doctor with what he needs, more time, nor has it given him the opportunity to really understand this new method of communication.

The doctor needs communications above all else. He needs to communicate with himself, that is, he needs to be able to remind himself of previous clinical situations. He needs to be able to communicate with his colleagues in the diagnosis and treatment of difficult patients. He also needs to know the availability and scope of test and care facilities so that he can advise and direct the patient as required and finally obtain the results of any investigations. This information must be readily available at all times for the same doctor or others who are called upon to treat the patient at an emergency or in any unusual situation.

But time is the essence of this activity. Can communications for the doctor be improved and time saved by a more realistic application of computers? This is the vital question that we have to answer.
Preliminary Investigations to define the problem

The desirability of designing and implementing a computer medical record has been highlighted by a series of studies we have carried out since 1969.

The Keyword Medical Record

The work carried out in a special study between 1969 and 1974 on a medical record suitable for general practice and outpatients in the Community Health Information Project (CHIP) (1) proved the technical feasibility and acceptability by the doctors of a simple functional keyword summary for both clinical areas with the patient record interchangeable via an IBM computer (360/156) using Remote Job Entry (RJE) in the batch mode. The key words entered on special cards by the doctors either in General Practice surgeries or outpatient clinics were used to create a punch tape which was transmitted to the central bureau computer via tele-typewriters in each consultation area; when the record was required it could be printed out and amended or brought up to date by the doctor concerned.

Automated Computer History Taking

Some very time consuming tasks in medical care are the primary and subsequent history taking procedures especially in chronic conditions now much more frequent than in the past. Pilot studies using our hospital Sigma 6 computer in this section of the medical record, the history taking or interrogation, to provide a keyword summary for inclusion in the case folder (2) (3) have been successful. The patient sits in front of a VDU in a small office and presses the YES/NO/SOMETIMES/PLEASE REPEAT keys as a result of seeing the question on the screen. The use of this technique has proved to be acceptable to patients and doctors (4). The use of the computer does not involve any function such as writing or typing. This application is now operating routinely in the Department of Medicine in Charing Cross Hospital using a microcomputer MICKIE, Motorola 6800 modified by Computer Workshops, firstly for patient follow-up in outpatients and secondly for an occupational and environmental program for inpatients. Each provides a printed summary of keywords for the patient's case folder.

Speech Recognition Experiment

An experiment was carried out during the early part of 1977 in conjunction with the Man/Computer Interface Department of the National Physical Laboratory (5) (6). This entailed the interrogation of bronchitic patients using the original Medical Research Council (MRC) bronchitis questionnaire programmed for speech recognition, developed by the late Dr. Christopher Evans from the National Physical Laboratory using the same answering words. The total number of transactions made by the fifteen patients was 796. Of these 74% were successful. The recognition performance (63% correct, 6% errors, 31% rejects) was not as good as previously obtained in laboratory trial performance figures (78% correct, 2% errors, 20% rejects) the differences, mainly the increase in rejects at the expense of correct recognitions, can be attributed to the particular group of patients used.
These were mainly old patients, a number had speech difficulties due either to age, shortness of breath or dialect. The range of tonal variation was considerable and fifteen patients included a fairly broad scan of individual dialects.

**Principal Objectives of the Present Feasibility Project**

This is to develop a medical record without (a) distorting or changing the doctor’s traditional clinical methods used at the consultation (b) forcing the doctor to carry out any additional procedures and (c) having to enhance the existing main frame hospital computer.

To avoid the known difficulties of persuading doctors to use standard input devices, it is proposed to test the feasibility of employing a Special Input Device (S.I.D.) and a PDP 1100/3 microcomputer, capable of voice recognition, to be used by the clinician during the course of his consultation, examination and testing of the patient as well as during the subsequent treatment and monitoring of the patient’s progress.

The voice input record will be validated at the time it is generated by the doctor on a VDU and a hard copy record made. The medical activities will also be recorded on a tape recorder and computer simultaneously - the two microphones working in parallel. The tape recordings will be transcribed in the usual manner by a medical secretary for comparison with the computer record. Traditional pen and paper case notes will also be compiled. In this manner there will be three records for evaluation.

Patients' and doctors' opinions will be recorded on questionnaires before the patient is discharged. These will be analysed and the acceptability of the system will be assessed at the end of the experiment.

The keyword summary record compiled with voice input to the computer will be assessed for its completeness and value in clinical care. The time factor and ease of compilation will be compared with the other two existing methods by the team.

**Immediate Benefits**

If this feasibility study proves successful the benefits to be expected are as follows:-

(a) each doctor can compile and hold his own records in the manner best suited to his work constrained only by the high level programming language BASIC. These records will contain the key functional words that specify the biological imbalance in the patient, the provisional diagnosis or symptom complex, the results or verification by physical examination and test procedures;

(b) these record procedures will be carried out on the doctor's personal computer in the clinical area

(c) all clinical data will be stored on floppy discs of a size which can be stored in the doctor's briefcase and their confidentiality ensured.
What have we achieved?

If the voice input study is successful it will be possible for every clinician to create, collate, store and use all the records of the patients he treats. Such an easily available record will enable him to expedite his clinical activities whilst at the same time carry out ongoing research.

The next step

In order to be able to exploit this information system all doctors will have to become part of a distributed network in some form or other.

At first the information that he has stored can be fed into a main computer data bank, at times suitable to both doctor and computer department using only the patient's number. Perhaps this central function could be taken over by a mini driver. The data could also be stored on microfiche fall back system.

Thus every clinician in group practices could be linked to the local hospital where similar groups of doctors in the various specialities are grouped together for functional purposes. In each case the confidential record would only be available by agreement with the individual doctor, in the accepted form and content.

In such a manner the whole of the key clinical data generated in a town or city could be merged if required. The personal data recorded by the doctor on his floppy disks would remain in his care unless released by him with appropriate safeguards (as it is today with manual records) for research purposes.
References

Hammersmith & Kensington Community Health Information 

(2) Evans C.R. Price H.C. Wilson J (1973) 
Computer Interrogation of Patients in a London Hospital. 

Some preliminary experiments in the use of a programmable 
videotape recorder as an automated history taking device 

(4) Computer Evaluation Unit Charing Cross Hospital. Patient 
History Taking by Computer. Report No. 1/73.

An Investigation of Man-Computer Interaction. 
Using an automatic Speech Recognition System. 

1978. 
An approach to the automatic recognition of speech.
<table>
<thead>
<tr>
<th>Name</th>
<th>Institution/Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Per-Anders Akersten</td>
<td>Ergonormrad AB 10032 S-65110 Karlstad, Sweden</td>
</tr>
<tr>
<td>Dr. Ing. Arthur B. Aune</td>
<td>SINTEF, Division of Automatic Control, Trondheim - NTH, Norway</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
<tr>
<td>Dr. L. Bainbridge</td>
<td>Univ. College London, Dept. of Psychology, 70a Gower Street, London WC1E 6BT, United Kingdom</td>
</tr>
</tbody>
</table>

---

IR. G. J. Blaauw  
Instituut voor Zintuig-Fysiologie TNO  
Postbus 23  
3769 ZG Soesterberg  
Netherlands

D. Bosman  
THF AFD. EL.  
Postbus 217  
7500 AE Enschede  
Netherlands

Tom Bosser  
Psychol. Instituut  
Univ. Münster  
Schlaunstrasse 2  
D-4400 Münster  
F.R. of Germany

Dr. H. H. Carls  
Fraunhofer Inst. IIIB  
Sebastian Kneipp Str. 12/14  
D-7500 Karlsruhe 1  
F.R. of Germany

M. Diani  
Centre d'Études Sociologiques Paris  
82 Rue Cardinet  
75017 Paris  
France

J. P. Gaillard  
Laboratoire Spartacus  
Universite Paris XII  
91011 Evry Cedex  
France

P. J. Gelderloos  
Dept. of Mechanical Engr.  
Lab. for Measurem. and Control  
Delft Univ. of Technology  
Meikleweg 2  
2628 CD Delft  
Netherlands

A. Van Dijk  
Dept. of Mechanical Engr.  
Lab. for Measurem. and Control  
Delft Univ. of Technology  
Meikleweg 2  
2628 CD Delft  
Netherlands

R. C. Van de Graaff  
National Aerospace Lab. (NLR)  
Anthony Fokkerweg 2  
1059 CM Amsterdam  
Netherlands

Gosta Harlin  
Univ. of Umea  
Industrial Electronics  
95186 Umea  
Sweden

Hans J. Bloem  
MIPG/THO  
Postbus 124  
2300 AC Leiden  
Netherlands

A. Den Boer  
Consultant Process Control Applications  
Foxboro  
Postbus 113  
3100 AC Schiedam  
Netherlands

H. Diandri  
Centre d'Etudes SocioLOGiques Paris  
91120 Palaiseau  
France

J. P. Gaillard  
Laboratoire Spartacus  
Universite Paris XII  
91011 Evry Cedex  
France

P. J. Gelderloos  
Dept. of Mechanical Engr.  
Lab. for Measurem. and Control  
Delft Univ. of Technology  
Meikleweg 2  
2628 CD Delft  
Netherlands

A. Van Dijk  
Dept. of Mechanical Engr.  
Lab. for Measurem. and Control  
Delft Univ. of Technology  
Meikleweg 2  
2628 CD Delft  
Netherlands

R. C. Van de Graaff  
National Aerospace Lab. (NLR)  
Anthony Fokkerweg 2  
1059 CM Amsterdam  
Netherlands

Gosta Harlin  
Univ. of Umea  
Industrial Electronics  
95186 Umea  
Sweden
PAFICIPANTS EUROPEAN ANNUAL MANUAL 1981

PROF. DR. TOWILL
DYN ANAL. GROUP
UNIST
KING EDWARD VII AVE.
CARDIFF CF1 3NU
UNITED KINGDOM

DR. D.A. TYLDESLEY
VAKGROEP PSYCHOLOGIE, IFLO
DE BOELELAAN 1081
AMSTERDAM
NETHERLANDS

F.W. UMBACH
THI AFQ. EL.
POSTBUS 217
7500 AE ENSCHEDE
NETHERLANDS

J.C. V.O. VAART
DELFT UNIV. OF TECHNOLOGY
KLUYVERWEG 1
2269 HS DELFT
NETHERLANDS

IR. P.J. VENEMANS
TILBURG UNIV.
POST BOX 90153
5000 LE TILBURG
NETHERLANDS

A. VILLEMEUR
E.D.F. - D.E.R.
1 AVENUE DU GENERAL DE GAULLE
92141 CLAMART CEDEX
FRANCE

PROF. DR. MED. C. VOSSIUS
UNIVERSITAT KARLSRUHE
INSTITUT FUR BIOKYBERNETIK UND
BIOMEDIZINISCHE TECHNIK
KAISERSTRASZE 12
7500 KARLSRUHE
F.R. OF GERMANY

H. VRIEND
DEPT. OF MECHANICAL ENGR.
LAB. FOR MEASUREMENT AND CONTROL
DELFT UNIV. OF TECHNOLOGY
MEKELWEG 2
2628 CD DELFT
NETHERLANDS

J.H.J.M. VULKER
ROYAL SHELL LABORATORY
BADHUISWEG 3
1031 CM AMSTERDAM
NETHERLANDS

P.H. WEWERINK
NATIONAL AEROSPACE LAB (MLR)
ANTHONY FOKKENWEG 2
1099 CM AMSTERDAM
NETHERLANDS

TED N. WHITE
DELFT UNIV. OF TECHN.
MEKELWEG 2
2628 CD DELFT
NETHERLANDS

J.H. J. M. VULKER
ROYAL SHELL LABORATORY
BADHUISWEG 3
1031 CM AMSTERDAM
NETHERLANDS

PROF. DR. ENG. H. P. WILLUMEIT
INST. FUR FAHRZEUGTECHNIK
TECHN. UNIV. BERLIN
STRASSE DES 17. JUNI 135
1000 BERLIN 12
F.R. OF GERMANY

J. WIRSTAD
ERGONOMRAD AB
BOX 10032
S-650 10 KARLSTAD
SWEDEN

JI ZHONGYUAN
DELFT UNIV. OF TECHNOLOGY
LAB. FOR MEASUREMENT AND CONTROL
DEPT. OF MECHANICAL ENGR.
MEKELWEG 2
2628 CD DELFT
NETHERLANDS