Proceedings of the
XV EUROPEAN
ANNUAL CONFERENCE ON
HUMAN DECISION MAKING
AND MANUAL CONTROL

held at the
TNO Human Factors Research Institute
Soesterberg, The Netherlands
June 10-12, 1996

organized by

TNO

TU Delft
Proceedings of the XV EUROPEAN ANNUAL CONFERENCE ON
HUMAN DECISION MAKING AND MANUAL CONTROL :
held at the TNO Human Factors Research Institute,
Soesterberg, The Netherlands, June 10-12, 1996
Peter A. Wieringa, Henk G. Stassen (Eds).
Delft: Delft University of Technology, Faculty of Mechanical Engineering
and Marine Technology
ISBN 90-370-0152-1
Subject Headings: man-machine systems / human operators / process automation / human factors
Preface

The 15th Annual Conference on Human Decision Making and Manual Control again shows a large variety of interesting presentations. Ranging from "Modelling of Individual Car-following Behaviour", via "Cognitive Support to Human Needs", to "A global Approach to solve Social Aspects of Usability and Acceptability, especially in Developing Countries", the conference offers ample possibilities for discussion and information exchange between participants.

The range of presentations more or less reflects a tendency to concentrate not only on the individual operators and not only on modelling, but on the user (in general) and on applications as well.

A strong point of these conferences is that they stimulate interdisciplinary cooperation, whereas in the beginning representatives of the technical disciplines on the one, and those from Human Factors and Ergonomics on the other side showed some sort of shyness in evaluating each other's approach. Nowadays both sides know that they cannot do without the other.

The world in the past 15 years changed faster than ever before. Therefore, I hope, the 15th "Annual Manual" will be succesfull in cooperation, especially where it concerns the problems of tomorrow.

Prof. J. Moraal
European Annual Manual Conferences

European Annual Conference on Human Decision Making and Manual Control, for short the European Annual Manual, was first held in 1981. The Conference was inspired on the idea to stimulate meetings between Ph-D students and professionals working in the field of Man-Machine Systems in the USA. The NASA -University Annual Conference on Manual Control was already more than 15 years organizing such meetings.

Since 1981 a series of conferences have been held in eight European countries, namely:

1981: The Netherlands, Delft, Delft University of Technology
1982: Federal Republic of Germany, Bonn, Forschungsinstitut für Anthropotechnik
1983: Denmark, Roskilde, Risø National Laboratories
1984: The Netherlands, Soesterberg, Institute for Perception TNO
1985: Federal Republic of Germany, Berlin, Technical University of Berlin
1986: United Kingdom, Wales, Cardiff, University of Wales, Institute of Science and Technology
1987: --
1988: France, Paris, Electricité de France
1989: Denmark, Lyngby, Technocal University of Denmark
1990: Italy, Ispra, CEC Joint Research Centre
1991: Belgium, Lieges, University of Liège
1992: France, Valenciennes, University of Valenciennes
1993: Germany, Kassel, University of Kassel
1994: Finland, Espoo, Technical Research Centre of Finland
1995: The Netherlands, Delft, Delft University of Technology
1996: The Netherlands, Delft, Delft University of Technology

It has been announced that the 1997 European Annual Manual will be held in France, Toulouse at EURISCO.

\[\text{1 This symposium has been sponsored by the RoHMI network; a network of the Human Capital and Mobility programme of the EC on the design of Robust Human Machine Interaction.}\]
Table of Contents

Preface

Session 1  Manual Control and Human Operator Modelling

1.1  J-M Pleijsant, M. Mulder, H. van de Vaart (TUDelft, AE).
How do Pilots perceive Time-To-Contact from the Ground Surface? Results of a Visual Simulation Experiment.

1.2  M. Mulder (TUDelft, AE).
Modelling Manual Control of Straight Trajectories with a Perspective Flight-Path Display.

1.3  J.H. Hogema (TNO, HFRI).
Modelling of individual Car-Following Behaviour.

Session 2: Human Operator Decision Support Systems And Interface Design

A Fuzzy Decision Support System for Magnetic Component Design.

2.2  W. Sillevis Smitt, B. Bruggeman

2.3  M. Tiemann (UKassel, IMAT).
Evaluation of a Methodology for Human-Machine Interface Development.

2.4  A. Nendjo Ella, M. Grislin, C. Kolski (UV, LAMIH / AICADA).
Towards a Global Approach to solve Social Aspects of Usability and Acceptability, especially in Developing Countries.

* Paper not presented at the Conference for personal reasons. Program committee proposed to include it into the Proceedings
Session 3: **REMOTE CONTROL**

3.1 E.F.T. Buïël (TUDelft, ME).
*The Development of a Man-Machine Interface for Space Manipulator Displacement Tasks.*

3.2 J. Herder, M. Horward, W. Sjoerdsma (TUDelft, ME).
*Mechanical Solutions versus Electronic Teleoperation; an example.*

3.3 J.B.F. van Erp, P. Padmos (TNO, HFRI).
*Real and Simulated Driving with Camera View.*

3.4 J.B.F. van Erp, B. Kappé, J.E. Korteling (TNO, HFRI).
*Visual Support in Operating Unmanned Aerial Vehicles.*

Session 4 **TASK ALLOCATION AND AUTOMATION OF COMPLEX SYSTEMS**

4.1 M. Neerincx (TNO, HFRI).

4.2 E. Cherifi, B. Riera, E. Hais (UV, LAMIH).
*Filtering Information for the Supervision of Complex Systems.*

*Experimental investigation on the effect of Task Allocation on the Human Operator.*

*Subjective evaluation of Task Allocation: An application of the Analytic Hierarchy Process.*

4.5 N. Moray (UV, LAMIH).
*A Method for Identifying Coupling between Humans and Machines.*

Session 5 **HUMAN RELIABILITY ASSESSMENT**

5.1 T.W. van der Schaaf, C.E. Shea, A.S. Nyssen (TUE / UL).
*MECCA: Medical Errors and Complications Causal Analysis.*

5.2 T.W. van der Schaaf, M. Frese, D. Heimbeck (TUE / UvA).
*Human Recovery and Error Management.*

Index
Session 1

Manual Control and Human Operator Modelling
How Do Pilots Perceive Time-to-Contact from the Ground Surface?
Results of a Visual Simulation Experiment

J.M. Pleijsant, M. Mulder, J.C. van der Vaart
Faculty of Aerospace Engineering, Delft University of Technology
The Netherlands

P.C.W. van Wieringen
Faculty of Human Movement Sciences, Free University of Amsterdam
The Netherlands

Abstract

This paper describes an experiment on the pilot's perception process during the landing manoeuvre - or flare - of an aircraft. In particular the role of the so-called Time-to-Contact, TTC, is considered. When approaching an obstacle, TTC is the time remaining to collision if no action were taken. In all time-constrained tasks like car driving or braking, subjects tend to use TTC as a cue by which their actions are triggered. Previous research on manned simulator landing tasks with only a runway outline scene visible indicated that pilots indeed use some kind of TTC strategy but that the timing of the flare was also related to the height above the ground. The present experiment was designed to examine whether addition of ground texture to a simulated visual runway scene would improve the perception of TTC, and hence the timing of the flare. The results suggest that addition of texture indeed enables a pilot to improve the landing performance due to a significantly better perception of TTC, as compared with a runway outline only.

1. Introduction, Time-to-Contact

The application of flight simulators for training pilots is still growing. It is important that the information presented by the simulator be as realistic as possible so that a pilot would perform as in real flight. Still, there are noticeable differences between simulated and real flights caused by the fact that real and visual motion are not perfectly matched in simulators. In order to achieve high fidelity, designers usually attempt to create extremely detailed visual scenes but as this requires extensive amounts of computer power, unacceptable time delays may result. If one wants to avoid or minimise time delays it is important to identify what visual information is essential to obtain a realistic moving scene. This is a challenge for both aerospace engineers and psychologists.

Gibson (1979) asserts that highly detailed scene elements provide less information on egomotion than the optic flow field created by dynamic transformations of those elements. Pilots are able to perceive information from the optic flow about the direction of heading or the aiming point (AP), and about the so-called Time-to-Contact (TTC, Lee 1980). TTC, \( \tau \), or the tau margin is the time remaining to collision if no action were taken.

When a subject approaches a wall or an object, TTC is decreasing with time. In different types of time-constrained tasks like car driving or braking, subjects tend to use TTC as a cue by which their actions are triggered. That TTC or \( \tau \) triggers the onset of the landing flare was demonstrated by flight simulation research at Delft University of Technology (Advani et al. 1993). In that experiment the timing of the flare was also related to the height above the ground surface. Height above the ground corresponds to a certain value of the angular size or optical angle \( \psi \), which
is the angle between two points sideways from the AP and the pilot's eyes. The dependency on ψ may have been caused by the low visibility of the simulated night approach, in which the only visible cue was provided by the runway outline. In order to further examine the influence of TTC on the timing of the flare the present experiment was designed to test the hypothesis that addition of ground texture to a synthetic runway scene would increase the visible optic flow field and thus would improve the perception of TTC as compared with a runway outline only.

2. Landing an Aircraft

During the final approach to a runway a pilot has to fly the aircraft along a glide path with a slant angle of about 3 [deg] (Figure 1). Generally the vertical component of the airspeed vector, the sink rate C, is too large for a smooth landing. Assuming a typical approach speed V = 60 [m/s], the glide path angle γ = -3 [deg] results in C = 3 [m/s]. At touchdown this sink rate is highly unacceptable for both passengers and undercarriage. Hence C should be reduced before touchdown, which is done by executing the flare manoeuvre. The flare is initiated by pulling the steering wheel backwards, resulting in a more positive pitch angle (Figure 1). The increase of the pitch angle coupled with an approximately proportional increase of the lift force effectively reduces the sink rate. The onset of the flare requires a precise coupling of timing and action. A flare that is too late or too weak results in a hard landing, a flare initiated too early or done too strong may result in a soft landing or no landing at all. The latter is highly undesirable since the aeroplane has a natural tendency to return to a climbing flight after such a missed touchdown. In order to avoid this and to establish an early firm contact between wheels and runway so that effective wheel braking can start immediately, pilots usually aim at a reasonably firm touchdown with a certain positive sink rate.

3. Visual Information from the Ground Surface

In most TTC experiments the existence of a TTC strategy for the onset of a collision avoiding action has been demonstrated by using approaches to a plane perpendicular to the observer's line of movement. In Figures 2 and 3 perpendicular approaches to a square object are shown. Assuming an approach along a straight line with constant velocity V, the tau margin τ as introduced by Lee (1980) equals the real TTC. It has been pointed out that τ can be described in terms of the optical angle θ by means of the following relationship:

\[ \text{TTC} = \frac{x}{V} = \frac{\sin \theta \cos \theta}{\frac{d\theta}{dt}} = \tau \]  

in which θ represents the optical angle between a random point and the aiming point (AP), and in which \( \frac{d\theta}{dt} \) is the rate of change of θ. In the following the variable τ will be used to designate the optical angle ratio and can be regarded as the perceived TTC. This τ information is directly available from the optic flow field, since the inverse ratio \( \frac{d\theta}{dt} \) represents the relative velocity of the optical image expanding across the retina. Here the relationship between TTC and optical angle information was shown by using the aiming point related angle θ. However, such a
relation can also be shown for other optical angles for instance the optical angle \( \varsigma \)
between two points from the lower side of the square (Figure 2). As a consequence, the optical angle ratio can be applied to every point of the object resulting in the same \( \tau \). Hence, TTC can be perceived from the entire optic flow field during a perpendicular approach.

Next we consider a slant approach to a square object. The optical sizes of the square using a 1 second interval are shown in Figure 4.

Figures 4 and 3 seem rather similar, but the optic flow field of the slant surface does not expand isotropically as in Figure 3. Rather, this flow field can be divided into two different areas (Lee 1974): the upper area is that part of the ground surface in between the horizon and the aiming point, whereas the lower area is covered by the remaining part of the visible ground surface, see Figure 5.

It can be seen that the perceived TTC from the upper area \( \tau_{UA} \) is larger than the real TTC at the AP, whereas \( \tau_{LA} \) from the lower area is smaller than TTC. The border line between both areas (the line parallel to the horizon and passing through the AP, therefore called the aiming line, AL) is the flow line providing the real TTC, because it is perpendicular to the line of movement.

4. Runway Outline versus Ground Texture

Before we can compare runway outline and ground texture, we should define both. A runway outline can be regarded as a visible trapezoidal shape, sharply distinguished from the surroundings. Ground texture can be defined as a spatial array of patches, lines or points varying in size, shape, posture, colour or brightness (partly adopted from Bookout and Sinacori 1993).

In conformity with the optical angle \( \theta \) as introduced in Figure 2 the optical angle \( \theta_{RW} \) (= \( \frac{1}{2} \psi \)) can be defined for a slant approach to a runway outline (Figure 6) being the angle between a point sideways from AP, the pilot's eyes and AP itself. For large values of TTC, the angle \( \theta_{RW} \) is very small. Approaching the runway and thus reducing TTC, the value of \( \theta_{RW} \) will rapidly increase as can be seen in Figure 7 where the horizontal line represents the pilot's line of motion to AP. Further, the slope of the \( \theta_{RW} \)-path corresponds to its rate-of-change \( \left( \frac{d\theta_{RW}}{dt} \right) \). Along the aiming line, TTC can only be perceived by observing the ratio:

\[
\tau_{AL} = \frac{\theta_{RW}}{\left( \frac{d\theta_{RW}}{dt} \right)} \quad (3)
\]

Next consider a ground surface containing a runway outline and texture elements. In this case all the texture elements along the aiming line provide optical information to the pilot as given by Eq. (3). This is illustrated by Figure 8 where a number of \( \theta \) paths as a function of TTC are shown. Since more optical information is available in a textured visible environment a better perception of TTC can be expected if ground texture is present compared to the case of a runway outline only.

Results of landing experiments in simulators as reported in the literature are somewhat contradictory. Addition of large spaced checkerboard texture to a simulated runway scene did not improve landing performance as compared to an outline only scene in one particular experiment (Harris et al., 1978). Rather landing performance deteriorated when texture was visible. From other experiments (Bennett et al. 1986; Warren and Riccio 1985; Wolpert et al. 1983; Zacharias 1985),...
it appeared that a terrain following task, which is slightly different from a landing approach, is best executed when the outline of a road or runway is visible. The presence of ground texture again appeared to reduce the pilot's performance.

5. Method

The experiment was done by using a Silicon Graphics Iris Indigo workstation with a Silicon Graphics 17" colour monitor. The experimental subjects were positioned at a viewing distance of 33 cm resulting in an eye-field-of-view (EFOV, Mulder 1994) of approximately 49° (azimuth) by 37° (elevation). Together with a 17 Hz image update rate the experimental design suffices the workstation simulation requirements as suggested by Batson et al. (1992).

In order to properly test the hypothesis that addition of ground texture improves the perception of TTC three different synthetic runway scenes were displayed on the monitor (Figure 9):

A. runway outline
B. ground texture
C. runway outline and ground texture

The ground texture was represented by a random line pattern consisting of about 5m long lines. The lines were randomly oriented and located across the ground surface. At the start of the simulation 1000 lines were visible. During a simulation run a continuously decreasing part of the ground surface was visually available to an experimental subject, using a constant EFOV. In order to obtain a sufficient number of texture elements remaining visible during the approach, the elements were divided around the AP according to a normal distribution. Further, a density gradient was added to the texture to enable a sufficient perception of the flat ground surface, as has been pointed out by Cutting and Millard (1984).

Assuming that in real flights the onset of the flare would mainly be based on TTC, the possible temptation of a subject to trigger the flare at a certain height corresponding to a value of the optical angle $\theta_{RW}$, should be prevented. Hence two runway widths were used, resulting in a total number of five synthetic runway scenes. The dimensions of both runways varied in width ($W = 40m$ and $W = 60m$) and in the distance between the runway threshold and the aiming point ($LT = 200m$ and $LT = 300m$).

The synthetic scenes were approached along a straight line using three different approach speeds ($V = 50m/s$, $60m/s$ and $70m/s$) and two different glide path angles ($\gamma = -2^o$ and $\gamma = -4^o$). The combination of the different values of $V$ and $\gamma$ resulted in six different sink rates $C$, ranging from $C = 1.74m/s$ to $C = 4.88m/s$. Because every sink rate was applied to every runway scene, the experiment consisted of thirty different simulation conditions, which were selected by the computer in random order.

6. The pre-programmed Flare Manoeuvre

As explained earlier, the approach is to be succeeded by the flare manoeuvre in order to obtain a smooth landing. The profile of the flare executed by transport aircraft has been shown to be approximately exponential in nature (Roskam 1979). The flare in the present experiment was pre-programmed using the following exponential equation for the height:
In this expression $H(t)$ is the eye height [m] above the ground surface and $T_f < t$ [sec]. Further, the index $f$ represents the moment of onset of the flare, and the index $opt$ represents the optimal moment of flare initiation. If flare initiation time $T_f$ is taken to be zero for a perfect or optimal flare, so $T_f = 0$, and $H_f = H_{opt}$ then:

$$H(t) = H_{opt}e^{-\frac{t}{TTC_{opt}}}. \quad (5)$$

The sink rate $C(t) = -\frac{dH(t)}{dt}$ [m/sec] during the exponential flare of Eq. (5) is then,

$$C(t) = \frac{H_{opt}}{TTC_{opt}}e^{-\frac{t}{TTC_{opt}}}. \quad (6)$$

Notice that the sink rate at flare initiation ($t=0$) is:

$$C(0) = \frac{H_{opt}}{TTC_{opt}}. \quad (7)$$

At the moment of touchdown, ($t = t_{td}$) the eye height above the ground is equal to the eye height $\Delta H$ above the wheels, so,

$$H(t_{td}) = \Delta H = H_0e^{-\frac{t_{td}}{TTC_{opt}}}, \quad (8)$$

and the touchdown sink rate $C_{td}$ is:

$$C_{td} = -\frac{dH(t_{td})}{dt} = \frac{H_0}{TTC_{opt}}e^{-\frac{t_{td}}{TTC_{opt}}}. \quad (9)$$

After some elaboration it can be shown that the sink rate $C_{td}$ at touchdown for the 'ideal' exponential landing flare can be expressed by:

$$C_{td} = -\frac{\Delta H}{TTC_{opt}}. \quad (10)$$

It follows from Eqs (6) through (10) that the flare is completely determined by the parameters $H_0$, $TTC_{opt}$ and $\Delta H$. After extensive pre-experiment testing it was decided to set $TTC_{opt}$ at 6 [sec]. Next, $\Delta H$ was set at a value of 3 [m], resulting in an 'ideal' or optimum touchdown sink rate of:
\[ C_{td} = \frac{\Delta H}{\text{TTC}_{\text{opt}}} = 0.33 \text{ [m/sec]} \].

Notice that now with \( \text{TTC}_{\text{opt}} \) and \( \Delta H \) being set, the value of \( H_0 \) follows directly from Eq. (7). In fact \( H_0 \) is set by the various initial sink rates \( C(0) \) of the experiment. A margin for acceptable touchdowns was set by:

\[ 0.05 \leq C_{td} \leq 0.90 \text{ [m/sec]} \].

This touchdown margin can be translated to a \( \text{TTC}_f \) margin: the TTC Funnel (Figure 10). In this figure the required \( \text{TTC}_f \) at flare initiation for an acceptable landing has been plotted against the approach sink rate \( C(0) \). The horizontal line represents the optimal \( C_{td} = 0.33 \text{ [m/sec]} \). The upper curve represents the minimal \( C_{td} = 0.05 \text{ [m/sec]} \) and the lower curve stands for the maximal \( C_{td} = 0.90 \text{ [m/sec]} \). From this figure it can be seen that the higher the approach sink rate, the narrower the \( \text{TTC}_f \) margin will be. Hence it can be expected that approaches with large values of \( C(0) \) (i.e. the steepest approaches) are more difficult to land than others.

Seven subjects without any previous flying experience participated in the experiment. After several training sessions, each subject completed ten replications of all thirty experimental configurations. The subjects were instructed to initiate the flare - by pressing the spacebar - in order to obtain a landing with a sink rate as close as possible to \( C_{td} = 0.33 \text{ [m/sec]} \). After each trial, subjects were provided with feedback information of the sink rate at touchdown.

7. Results

As has been explained earlier, a satisfactory perception of TTC corresponds to a touchdown within the limits set for acceptable touchdown sink rates. Hence the percentages of achieved touchdowns per display offer a first indication for the quality of the perception process for each display.

In Table 1 touchdown percentages of all simulation runs conducted by the seven subjects are shown. From this table it appears that the displays containing texture enabled subjects to achieve more touchdowns due to a possibly improved perception of TTC as compared with displays without visible texture.

<table>
<thead>
<tr>
<th>Display</th>
<th>Successful Touchdowns</th>
</tr>
</thead>
<tbody>
<tr>
<td>I small runway</td>
<td>51 %</td>
</tr>
<tr>
<td>II small runway + texture</td>
<td>57 %</td>
</tr>
<tr>
<td>III texture only</td>
<td>60 %</td>
</tr>
<tr>
<td>IV large runway</td>
<td>55 %</td>
</tr>
<tr>
<td>V large runway + texture</td>
<td>62 %</td>
</tr>
</tbody>
</table>

Table 1. Percentage of successful touchdowns
This preliminary conclusion is confirmed by extensive multivariate Analysis of Variance (ANOVA) computations, using a significance level at $\alpha = 0.05$. The application of ANOVA's for statistical analysis of the experimental data has been allowed by Kolmogorov-Smirnov goodness-of-fit tests. These tests clearly revealed the approximately normal distribution of the data, a requirement for correctly interpreting the ANOVA results.

Recalling the TTC Funnel, one may assume that the best perception of TTC would be revealed by a constant TTC strategy independent of the approach speed $V$, the glide path angle $\gamma$ and the runway size. The results presented below are illustrated by the data of one typical subject, as shown in Figure 11 through 15. These figures show typical mean values and standard deviations of $TTC_f$ and $\psi_f$ at onset of the flare as functions of the initial sink rate $C(0)$. For the two runway outline only displays (I and IV, Figure 11) $TTC_f$ at onset of the flare was significantly influenced by $V$ for all subjects. Except for two subjects, $\gamma$ did not affect $TTC_f$. Further, the runway size significantly affected $TTC_f$ for most subjects.

The three displays containing visible ground texture (II, III and V, Figure 12) predominantly enabled subjects to use constant TTC strategies. $TTC_f$ was neither affected by $V$ nor by $\gamma$ for most (five) subjects. The significant differences for the others mainly occurred at the lowest and highest values of the approach sink rate $C(0)$. In spite of the seemingly invariable TTC strategies, the runway size significantly influenced $TTC_f$ for some subjects, indicating that they were gazing at both texture and runway.

Besides $TTC_f$ the effects of the simulation variables on the optical angle $\psi_f$ at onset of the flare were also examined. For the displays without texture (I and IV, Figures 13 and 14) $\psi_f$ was not significantly influenced by $V$ for all subjects and was affected by $\gamma$ for only two subjects, suggesting a constant $\psi$ strategy. However, $\psi_f$ was significantly affected by the runway size for all subjects, which may seem rather paradoxical.

As a confirmation of the exposed TTC strategy for the textured displays (II, III and V, Figure 15), $\psi_f$ was significantly influenced by $V$ for all subjects and was influenced by $\gamma$ for some subjects. Finally, $\psi_f$ was significantly affected by the runway size as might have been expected.

8. Discussion

The results indicate that the availability of visible ground texture in a simulated runway scene yields improved perception of TTC as compared with a runway outline only scene. This enhanced perception allows a pilot to initiate the flare on the basis of TTC only.

Without texture, subjects seem to prefer constant $\psi$ strategies to trigger the flare, such that a 'small' $\psi$ strategy is used when approaching a small runway, whereas a 'large' $\psi$ strategy is used for a larger runway.

Although these results are clear cut, the question remains why texture provides improved perception of TTC compared to a runway outline only scene. The present results are rather surprising, especially regarding the previously mentioned simulator experiments (Harris et al. (1978), Bennett et al. (1986), Warren and Riccio (1985), Wolpert et al. (1983), Zacharias (1985)), the outcome of which that the outline of a runway or road improves pilot's performance if compared to ground texture.
A suitable explanation for the different results of the present and the other TTC experiments appears to be well possible by a further detailed analysis of the differences in the tasks and the displays used. Such a detailed analysis, however, is beyond the scope of the present paper.

9. Conclusions

The presence of visible ground texture in simulated landings was shown to significantly improve the perception of TTC and hence the landing performance as compared with a runway outline only scene. The presence of a visible runway appeared to have negligible effect on the perception of TTC. Hence without ground texture as a visible cue, subjects seemed to time the onset of the flare manoeuvre on the basis of the optical angle $\psi$.

Because the real TTC is only provided by the aiming line, the conclusion may be drawn that pilots mainly perceive TTC from this line. However to validate this hypothesis, a next experiment should be conducted, in which the effects of texture patterns along several lateral lines - corresponding to virtual flow planes - on the perception of TTC can be identified.

References

What Optical Cues Do Pilots Use to Initiate the Landing Flare? Results of a Piloted Simulator Experiment. Proceedings of the AIAA Flight Simulation Technologies Conference, Monterey, CA (USA), August 9-11, AIAA-93-3561-CP.


Synthetic Perspective Optical Flow: Influence on Pilot Control Tasks.

Bookout, G. and J. Sinacori (1993)


Gibson, J.J. (1979)

Harris Sr, R.L. and M.C. Waller, S. Salmirs (1978)
Lee, D.N. (1974)
Visual Information during Locomotion. In: McLeod, R.B. and H. Picks (Eds),

Lee, D.N. (1980)

Displays, Perception and Aircraft Control. Report LR-762. Delft: Delft University of
Technology, Faculty of Aerospace Engineering.

Roskam, J. (1979)
Airplane Flight Dynamics and Automatic Flight Controls, Part II. Lawrence (KA):
University of Kansas.

Visual Cue Dominance Hierarchies: Implications for Simulator Design. Society of
Automotive Engineers (SAE) Technical Paper Series, Long Beach, October, 14-17.

Eyeheight-Scaled versus Ground-Texture-Unit-Scaled Metrics for the Detection of
Loss in Altitude. In: Jensen, R.S. (Ed.), Proceedings of the Second Symposium on
Aviation Psychology. Columbus (OH).

Zacharias, G.L. (1985)
Modelling the Pilot's Use of Flight Simulator Visual Cues in a Terrain-Following
USAF ASD Contract no. F33615-81-0515.
Figure 1  Approach and flare manoeuvre

Figure 2  Perpendicular approach to a square surface
Figure 3  Expansion of square surface during a perpendicular approach; optical sizes of square with a 1 second interval

Figure 4  Expansion of square surface during a slant approach; optical sizes of square with a 1 second interval
Figure 5  Side view of the slant approach defining the optical angle $\phi$.

Figure 6  Approach to a runway with the optical angles $\psi$, $\theta_{\text{RW}}$. 
Figure 7  Optical angle $\theta_{nw}$ as a function of TTC

Figure 8  Optical angles $\theta_{tx}$ as a function of TTC
Figure 9  Synthetic runway scenes: A. runway outline, B. ground texture, C. runway outline and ground texture

Figure 10  TTC Funnel: TTC\(_i\) margin for obtaining a smooth landing
TTC\textsubscript{f} at onset of flare

Display IV: large runway

one subject

![Graph of TTC\textsubscript{f} versus C at onset of the flare. Display IV: large runway (results of one subject)]

Figure 11  \textit{TTC\textsubscript{f} versus C at onset of the flare. Display IV: large runway (results of one subject)}

TTC\textsubscript{f} at onset of flare

Display V: large runway + texture

one subject

![Graph of TTC\textsubscript{f} versus C at onset of the flare. Display V: large runway and texture (results of one subject)]

Figure 12  \textit{TTC\textsubscript{f} versus C at onset of the flare. Display V: large runway and texture (results of one subject)}
PSIf at onset of flare
Display I: small runway

one subject

\[ \psi \] versus \( C \) at onset of the flare. Display I: small runway (results of one subject)

PSIf at onset of flare
Display IV: large runway

one subject

\[ \psi \] versus \( C \) at onset of the flare. Display IV: large runway (results of one subject)
PSIf at onset of flare

Display V: large runway + texture

one subject

\[ 1.74 \, (50;2) \quad 2.09 \, (60;2) \quad 2.44 \, (70;2) \quad 3.48 \, (50;4) \quad 4.18 \, (60;4) \quad 4.88 \, (70;4) \]

Figure 15 \( \psi_r \) versus \( C \) at onset of the flare. Display V: large runway and texture
(results of one subject)
MODELLING MANUAL CONTROL OF STRAIGHT TRAJECTORIES WITH A PERSPECTIVE FLIGHT-PATH DISPLAY

Max Mulder

Delft University of Technology, Faculty of Aerospace Engineering, P.O. Box 5058, 2600 GB Delft, The Netherlands
e-mail: m.mulder@lr.tudelft.nl

Abstract. A perspective flight path display shows the flight trajectory to be flown in a synthetic three-dimensional world. The application of such a display has important consequences for a pilot because the guidance information is presented via spatial sources of information. To understand pilot manual control behaviour with a perspective display, it is essential to investigate the manner in which pilots use these optical cues. The paper describes the approach chosen and discusses results of one of the experiments.

Keywords. Perspective flight-path display, Manual control, human operator modelling

1. INTRODUCTION

A perspective flight-path display, showing the planned trajectory to the pilot in a synthetic three-dimensional world (Figure 1), is not a new concept. Since the early 1950's it has been hypothesized that such a pictorial display could mean, in many ways, an important improvement in information-transfer to the pilot. Its application was impractical, however, due to technical limitations.

Basically, two developments in technology made the pictorial display concept practical. First, rapid improvements in computer technology made sufficiently detailed real-time graphics possible. Second, the advance of new positioning systems, such as GPS (Global Positioning System) and MLS (Microwave Landing System), provided the capability of measuring the position of the aircraft with a sufficient update rate.

The application of a perspective display in the cockpit has important consequences. In a conventional cockpit the pilot mentally reconstructs the aircraft's spatial and temporal situation from a number of planar, i.e. two-dimensional, displays. With a perspective flight-path display this information is presented in a spatial format [Theunissen and Mulder, 1995].

At the Delft University of Technology, research is being conducted to investigate the implications a perspective flight-path display has on pilot's manual control behaviour. The main subject of interest is not, however, whether such a display will lead to an improved man-machine interface. Rather, the research impetus is to determine how a pilot is able to control a complex dynamic system (the aircraft) along a space-constrained trajectory, with a perspective flight-path display as primary source of information. Once this question has been answered, an attempt can be made to represent the important characteristics of the
pilot in a mathematical model. This paper describes one of the experiments which have been conducted. The intent is to show clearly the manner in which the research issues are approached. Section 2 discusses the decomposition of the problem into different subsets. Then results from a cue-inventory are presented in section 3. In order to investigate some of the hypotheses following from this cue-inventory, an experiment has been conducted. Section 4 describes the setup of this experiment while sections 5 and 6 discuss the time-domain and frequency-domain results from the experiment respectively. Finally, section 7 contains a discussion of the approach followed and conclusions.

2. TOWARDS A CONTROL-THEORETIC MODEL OF PILOT MANUAL CONTROL BEHAVIOUR WITH A PERSPECTIVE FLIGHT-PATH DISPLAY

In [Mulder, 1995] the methodology of the research project has been discussed. Below, the main points of interest are briefly repeated.

2.1. Research goal

Our goal is to understand and, ultimately, to model the main characteristics of pilot manual control behaviour with a perspective flight-path display. The particular research interests are especially the information-transfer between the display and the pilot, including the influences of display design variables and the effects of additional display symbology.

2.2. Modelling approach

The pilot's task is to follow the reference trajectory, a typical guidance task. Based on the characteristics of the reference trajectory, the guidance task can be divided into a number of phases (or sub-tasks):

1. to maintain a straight section of the trajectory,
2. to maintain a curved section of the trajectory,
3. to control a transition between a straight and a curved section of the trajectory and vice versa.

In other words, or, from a system-theoretical point of view, following the planned trajectory leads to maintaining a series of different system steady-states (or references) and controlling transitions between these steady-states. The modelling attempts are structured according to this decomposition principle.

Maintaining a certain state of the system against disturbances is simply a regulation task. To model this task, two well-known modelling methodologies can be applied: The classical approach, resulting in a multi-channel version of McRuer's crossover-model [McRuer et al., 1965], and an optimal control approach, using the Optimal Control Model (OCM) of Kleinman et al. [Kleinman et al. 1969]. The transition phase between two different steady-states can be modelled using an extension to the conventional OCM, i.e. the Optimal Control and Preview Model (OCPM), which has been described in [Mulder, 1995].

In the following, the discussion will be restricted to the regulation task of following straight sections of the tunnel trajectory.

3. OPTICAL CUES IN A STRAIGHT TUNNEL SEGMENT

3.1. General

A perspective flight-path display shows the planned trajectory in a synthetic three-dimensional world. The task of the pilot is to control the aircraft along this path. To fulfill his task, the pilot estimates the state of the aircraft with respect to the trajectory and, based on the estimated state, decides upon and activates the necessary control action(s). In order to understand the interaction between the pilot and the display it is essential to get grip of this state estimation process. This has been investigated from two different points of view.

In [Mulder, 1994] it was examined what effects a spatial display has on the control behaviour of a pilot: The man in the man-machine interface was taken to be the central element. Main questions that were addressed were the availability, the usefulness and the potential utilization of all sorts of spatial, or optical sources of information present in the real world and/or in a perspective display.

In [Mulder, 1996] the machine side was the main issue. An attempt was made to make an inventory of all spatial cues in a basic perspective flight-path display. Here, irrespective of the human operator, mathematical relations are derived that express the state of the aircraft with respect to the reference trajectory in terms of these optical cues. Obviously, most of the available cues will probably be neglected by the operator for reasons of their perceivability thresholds or simply because...
they are too far-fetched for the operator to be recognized as potential sources of information. Some of the cues, however, are so evident and can be so easily perceived from the display that it is most probable that they are used by the operator. It are these particular cues which are the main subject of investigation.

3.2. Straight tunnel sections

In this paper, the discussion will be restricted to sections of the trajectory that are straight and infinitely long. Furthermore, wind-effects are neglected. As has been discussed in [Mulder, 1995], in case of no position errors (and relatively small aircraft attitude angles) the tunnel image will be symmetric. Any deviation from the trajectory leads to a distortion of this symmetric condition (Figure 2). To minimize the discrepancy between the actual and the planned trajectory, the operator must maintain a symmetric tunnel image.

In [Mulder, 1996] it is shown that there are many optical cues in the display that are directly related to the deviation of the aircraft position and attitude from the reference trajectory. For example, the aircraft heading error, or Track-Angle-Error (TAE), can be perceived from the translation of that part of the tunnel which is located at a large distance from the viewpoint, as was observed in [Grunwald and Merhav, 1976] and [Theunissen, 1994]. In this paper the discussion will be restricted to the optical cues related to a position error.

3.3. Optical cues

The lateral and vertical position errors can be estimated using a large number of optical cues of which the most salient ones are illustrated in Figure 3. First of all, we have the relative lateral displacements \( \delta e_{ij} \) and \( \delta \eta_{ij} \) of the tunnel frames \( i \) and \( j \) located at distances \( D_i \) and \( D_j \) (with \( D_j = D_i + \Delta D \), and \( \Delta D \) the fixed distance between two successive frames). Changes in these relative lateral displacements from the zero-error condition are a function of lateral position error \( X \) only (in approximation):

\[
\delta e_{ij} = +\kappa X \left( \frac{D_j - D_i}{D_i D_j} \right) \\
\delta \eta_{ij} = -\kappa X \left( \frac{D_j - D_i}{D_i D_j} \right)
\]

with \( \kappa \) a display constant dependent on the field-of-view of the perspective projection and the size of the display screen.

The same holds for the relative vertical displace-
Fig. 3. Position error cues in a straight tunnel section

ments $\mu_{ij}$ and $\nu_{ij}$, which are a function of the vertical position error $V$ only (in approximation):

$$\delta \mu_{ij} = -\kappa V \left( \frac{D_i - D_j}{D_i D_j} \right)$$

$$\delta \nu_{ij} = +\kappa V \left( \frac{D_j - D_i}{D_i D_j} \right)$$

Note that in the formulas stated above, the deviations are for all possible pairs of tunnel frames $i$ and $j$. As one can see, however, the applicability of these cues deteriorates fast when the distances involved become larger. The term rapidly decreases the amplitudes of the displacements below threshold.

The second set of cues result from the longitudinal lines connecting the individual tunnel frames. The angles that the projections of these lines make with the horizon are also merely a function of lateral and vertical position error only:

$$\omega_1 = -\frac{V}{W} - \frac{X}{W}$$

$$\omega_2 = -\frac{V}{W} + \frac{X}{W}$$

$$\omega_3 = +\frac{V}{W} + \frac{X}{W}$$

$$\omega_4 = +\frac{V}{W} - \frac{X}{W}$$

with $W$ the (square) tunnel width.

A third cue results from the imaginary line connecting the intersections of the altitude poles with the bottom of the tunnel frames. The result is an angular cue which changes almost identically with the lateral position error:

$$\omega_5 = -2 \frac{X}{W}$$

This cue will be neglected in the following, because in the displays used in the experiment described below the altitude poles are not presented.

3.4. Discussion

The linear and angular optical cues discussed above are both a function of vertical and lateral position error only. There are two fundamental differences between these cues.

First of all, when the aircraft moves through the tunnel, the tunnel frames translate towards the perceiver, while the longitudinal lines connecting the frames appear to do not. This is an important fact, since the motion of the tunnel frames could prevent an accurate estimation of the optical displacement cues: The pilot constantly has to shift attention towards a new set of frames. The angular cues, however, only change because of a changing position error, making a shift in atten-
tion forwards and backwards into the tunnel image unnecessary.
Second, as one can see from the formulas stated above, it is clear that a lateral and a vertical position error both determine any one of the four angles. This is in contrast to the fact that a change in the relative lateral (vertical) displacements of the tunnel frame lines is only a function of the lateral (vertical) position error. In other words, the angular cues are coupled and the linear cues uncoupled with respect to the lateral and vertical position errors. This is also an important fact. The coupling of the angular cues means that any change in any of the four angles can be the result of both a vertical and a lateral position error. The linear cues, on the other hand, are uncoupled and only change within the same dimension (i.e. lateral or vertical) as the occurring position error.

3.5. Hypotheses and experiment justification

The discussion above shows the virtues and disadvantages of the two primary sets of cues to estimate a position error on a straight section of the tunnel display. To examine the usefulness of both sets and the extent to which the above mentioned characteristics influence their relative usefulness, an experiment has been conducted.

Three displays were defined, which are all abstractions of the basic tunnel-in-the-sky display:

1. Display A, showing only the longitudinal lines connecting the tunnel frames (Figure 4),
2. Display B, showing only the tunnel frames themselves (Figure 5),
3. Display C, a combination of displays A and B (Figure 6).

It is clear that in display A only the angular cues are available, while in display B only the linear displacement cues are present \(^1\). From display C, both sets of cues can be perceived. Further note that in all displays the aircraft attitude, i.e. pitch angle \(\theta\), roll angle \(\phi\) and heading angle \(\psi\) can and will be perceived in identical fashion. To analyze the usefulness of the optical cues from the three displays, two additional variables were introduced in the experiment:

I The effect of control channel:

Three control channels were applied:

\(^1\) The angular cues could be estimated from the imaginary lines connecting the tunnel frames' vertices. Results from a pilot questionnaire revealed, however, that this was not the case.

\(^2\) Since the reference heading angle of the tunnel is set zero, the aircraft heading angle equals the track-angle-error.

---

(i) **Roll:** lateral tracking only, the vertical position error $V$ was kept zero,

(ii) **Pitch:** vertical tracking only, the lateral position error $X$ was kept zero,

(iii) **Dual:** a combination of (i) and (ii), i.e. both the lateral as the vertical position errors had to be minimized simultaneously.

II The effect of forward motion:

Two situations were examined:

(i) **No (forward) motion,** in which the longitudinal position of the aircraft was fixed, resulting in a hovering task$^3$.

(ii) **(Forward) motion,** in which the longitudinal position of the aircraft was set free, resulting in a conventional tunnel tracking task.

These additional experimental variables can be used to examine the usefulness of the two sets of position estimation cues according to the following a priori hypotheses:

I The presence of forward motion has no effect on the performance$^4$ with display $A$.

II The presence of forward motion deteriorates performance with display $B$.

III The addition of an additional control channel deteriorates performance with display $A$.

IV The addition of an additional control channel deteriorates performance with display $B$, but to a significantly less degree than the performance deterioration for the same condition of display $A$.

The effects of the independent variables motion and control channel on display $C$ are expected to be a mixture of those effects on displays $A$ and $B$. Since display $C$ is a combination of displays $A$ and $B$, the relative virtues and disadvantages of the combined sets of cues could compensate for each other.

4. EXPERIMENT

4.1. Goal of the experiment

The goal of the experiment was twofold. First of all, the validity of the hypotheses stated in the former section must be examined. Second, the observed control behaviour of the subjects must be described and explained with a mathematical model.

Testing the hypotheses does not demand much ingenuity in the experimental design. One could for instance define a conventional compensatory tunnel tracking task, measure all performance variables of interest, and do a post-hoc analysis of the empirical data using statistical tests.

The efforts to describe the observed behaviour via a mathematical model, however, are not so straightforward. Boldly, one could state that making models is relatively simple, while validating them can be extremely difficult. Naturally, the model validation could be restricted to using the empirical time-domain data only. Most, if not all, of the widely-used operator modelling techniques, however, have shown their applicability especially in the frequency-domain.

The frequency-domain data, i.e. the operator frequency-response functions, can be hard to obtain. In our situation this is especially true because:

- the pilot is operating in closed-loop, which introduces a lot of subtleties in the identification procedure,
- with the type of displays discussed here, the pilot is essentially a multi-input, multi-output system.

In order to obtain the frequency-domain data, an identification method will be used that was developed in [Lunteren, 1976] and applied in [Paassen, 1994]. The application of this method has important consequences for the definition of the experiment, as will be discussed in the next section.

4.2. Identification procedure

4.2.1. Method. In [Lunteren, 1976], a non-parametric identification method is developed to estimate pilot frequency-response functions in closed-loop. Although a full discussion of this method is beyond the scope of this paper, the main concepts will be briefly addressed.

Basically, for each pilot input signal an independent reference (or, in our case, disturbance) signal, usually a sum of sinusoids, has to be introduced into the closed pilot/vehicle loop. In a well-chosen experimental setup, all variables of interest then contain much power at the frequencies of the disturbance signals and only little power at all other frequencies. The operator's frequency-response function can then be estimated by interpolation and proper manipulation of the Fourier coefficients of the FFT-ed frequency-domain data in a manner described in detail in [Paassen, 1994].
For all frequencies of the reference sinusoids, an estimation is obtained of the operator frequency-response functions. In this estimation process, it is assumed that the operator does not insert any noise into the loop. Since this is not the case, the estimations can be biased and contain uncertainties. In [Paassen, 1994] it has been shown how analytic expressions can be obtained for the bias and variance terms of the frequency-response functions.

4.2.2. Pilot Model. The control of the lateral motion of the aircraft requires the pilot to close at least three feedback loops: First, serving as the most inner loop, the aircraft’s roll attitude is fed back. Second, the aircraft’s heading angle serves as the middle loop, and finally the lateral position is the outer loop. The same can be stated for the control of the aircraft vertical motion. Thus, to use the aforementioned identification method, three independent input signals (for both channels) must be inserted to estimate the three operator feedback loops. In multi-loop tasks it is often assumed, however, that all pilot equalization is applied to get a tight and well-operating inner loop. Outer loops are then fed back using simple gains. This assumption allows us to decrease the number of input signals for each channel from three to two: The first serves for the estimation of the inner (attitude) loop, while the second will be used to estimate the combined middle/outer loops. The combined outer loop frequency-response function can then be divided into its middle-/outer-loop components using parametric operator models. This is the approach followed here. The resulting pilot model is illustrated in Figure 7.

4.2.3. Dual-axes tasks. When both the vertical as the lateral position errors have to be minimized simultaneously, the operator model consists of six input signals and two output signals. Applying the aforementioned assumptions, i.e. combining the middle and outer loops, results in a operator model with four inputs and 2 outputs. The identification method can also compute all potential cross-couplings between the two channels. In the following, however, it is assumed that in dual-axes tasks the lateral and vertical channels are controlled in parallel, neglecting all cross-coupling effects. Preliminary investigations have shown that this is a safe assumption.

4.2.4. Choice of input signals. The lateral and vertical control channels are both disturbed with two independent input signals each. The input signals were all computed as a sum of 12 sinusoids. The amplitudes of these sinusoids for all 12 frequencies were determined by the choice of the input signal spectrum. These spectra should be chosen simultaneously with the system dynamics.

4.2.5. Choice of system dynamics. The system dynamics were the linearized dynamics of a small business jet, the Cessna Citation I. Figure 8 shows the simulated aircraft dynamics for the lateral control channel and the insertion points of the two lateral disturbance signals. A well-chosen combination of the system dynamics, the disturbance signals’ spectra and the intensity of the disturbance signals, is crucial in the experiment.

4.3. Experimental setup

4.3.1. Design. The experimental design was a 3 x 3 x 2 factorial design (display, channel,
motion), resulting in 18 conditions (Table 1). Every condition was conducted 14 times: The first eight runs were used as training runs and the last six runs were the actual measurement runs. The 18 conditions were fully randomized over all 252 (18 x 14) runs.

4.3.2. Setup. Subjects were seated in a chair in a darkened, noise-free room in front of a 17 inch CRT monitor. The control manipulator was a servo-controlled hydraulic side-stick with conventional characteristics. The display update-rate was 20 Hz. The tunnel was presented as a grey wireframe on a black-and-white background. A flight-path vector was not presented.

4.3.3. Subjects. Three professional pilots collaborated in the experiment. Results of only two pilots are available at this time.

4.3.4. Task. The task is simply to fly the aircraft as accurately as possible through the tunnel. In other words: Minimize all occurring position errors, despite the effect of the disturbances acting on the vehicle.

4.3.5. Questionnaire. After completion of the experiment, all subjects received a pilot questionnaire in which they were requested to answer a number of questions. The subjective results of this questionnaire were used to increase insight into the observed, i.e. objective, overt control behaviour.

5. TIME-DOMAIN RESULTS

The performance variables of interest were all recorded during the measurement runs of the experiment. Since the pilot’s task was to fly as accurately as possible through the tunnel, the discussion will be restricted to the standard deviation of the lateral and vertical position errors. The time-domain results are examined along the hypotheses of section 3. The statistical method used was a conventional ANOVA followed by a Newman-Keuls post-hoc analysis. Although individual differences between subjects were found, the overall trends were more or less equivalent. The results will therefore be discussed using the data of one subject.

Figure 9 shows the mean and standard deviation of the six measurements of the standard deviation of the vertical position error \( V \) for all (vertical) conditions. Because in the following all results are illustrated by the same graphical method, an explanation is given here. The lower part of the figure contains a shorthand description of the conditions. First of all, the figure is divided in three parts for all displays: A, B and C. Secondly, a ‘.’ depicts the clean condition (i.e. single-axis vertical control only, no forward motion). The ‘R’
means that the lateral control channel was added, resulting in a dual-axis tracking task. The '+M' means that forward motion was activated, while the '+R', '+M' symbol depicts the situation that both an extra control channel as the forward motion were added.

From this figure it is clear that for display A the addition of forward motion has no effect on pilot performance, while the addition of an extra control channel significantly (p=0.01) deteriorates performance. These findings support hypotheses I and III.

For display B the results show that there is a significant (p=0.01) effect of the addition of forward motion on pilot performance, while the addition of an extra control channel has no effect on pilot performance. These findings support hypotheses II and IV.

The combination of displays A and B, i.e. display C, shows basically the same trends as those for display B. Furthermore, for three of the four conditions, pilot performance was superior with this display. The performance deterioration due to the addition of forward motion is substantially lower than that of display B, while the performance deterioration due to the addition of the roll control task vanishes. Tentatively then, one could state that the linear displacement cues are dominant over the angular cues, a suggestion that was supported by the results from the pilot questionnaire. The same conclusions can be drawn when the performance of the lateral control channel is examined (Figure 10). Again, the results for displays A and B support hypotheses I-IV, although the performance deterioration for display A was significant only at the p=0.05 level. Display C, however, shows a mixture of the effects of forward motion and the addition of a control channel in the sense that both effects were found significant (p=0.01).

In conclusion, one could state that the time-domain data support the a priori hypotheses of the experiment. The effects of forward motion and control channel are clear for displays A and B, while the results for display C are a mixture of the effects of its components. A final note is that the performance decrements due to the addition of an extra control channel are generally larger in the vertical channel than in the lateral channel, except for display C. This could be attributed to the fact that the vertical control channel was judged significantly more difficult to control than the lateral control channel. In order to fully explain the causes for the performance decrements one should take this task difficulty aspect into account, as will be shown in the next section.

6. FREQUENCY-DOMAIN RESULTS

The identification procedure results in estimates of the inner- and outer-loop pilot frequency-response functions. These frequency responses can be combined with the system dynamics, allowing the computation of the inner- and outer-loop crossover frequencies and phase margins. As has been stated above, in the identification of the dual tasks it was assumed that no cross-coupling effects between the two control channels exist.

Figure 11 shows the outer-loop crossover frequencies and phase margins for the vertical position tracking task. These variables were computed according to two approaches. First of all, one could
estimate the frequency-response functions for all six measurement runs for all conditions, resulting in six crossover frequency estimations and six phase margin estimations for each of the 18 conditions. The circles in the figure show the means and the vertical line segments the standard deviations of these six estimations. The second approach was to first obtain a time-domain average of the time histories over all six measurements and then, using the averaged set of time histories, estimate the crossover frequency and phase margin. This results in a single estimate of the variables of interest for all 18 conditions, depicted by a cross in the figures. Note that generally the crosses are equal or close to the estimated means.

One can see that the frequency-domain results generally follow the time-domain results\(^6\). For display A, the addition of an extra control channel decreases the crossover frequency, although the phase margin remains more or less constant. An explanation for this result will be given later since it requires knowledge of the behaviour of the pilot in the other channel. Display B shows a significant effect of forward motion while display C follows the trends of display B. The improved performance with display C over display B is reflected by significantly higher crossover frequencies and lower phase margins.

Figure 12 shows the outer-loop crossover frequencies and phase margins for the lateral position tracking task. From the lower half of this figure it is immediately apparent that the addition of an extra control channel results in a highly significant decrease in phase margin for all conditions. The crossover frequencies, on the other hand, show

\(^6\) Note that, theoretically, a higher crossover frequency generally implies a better performance and is usually accompanied by a smaller phase margin and vice versa.
that they are hardly affected by the control channel dimension. These findings are quite the opposite of those of the pitch channel analysis. An explanation for the dual-axes results is obvious: Theoretically, in a dual axes task the pilot has to divide his attention over two channels, increasing his information-processing time delay, which forces the pilot to sacrifice crossover frequency (and thus performance) to maintain sufficient phase margin (i.e. stability). For the vertical control channel this was indeed the case. The lateral control channel, however, was regarded (supported by the results of the pilot questionnaire) to be much easier to control than the difficult pitch channel dynamics. Thus, the lateral dynamics provided the pilots the opportunity to maintain a relatively high performance (i.e. crossover frequency) by simply sacrificing their phase margin. Concluding, one can state that the frequency-domain analysis can reveal some of the uncertainties that could arise from examining the time-domain data only. Instead of attributing all observed trends in the empirical data to the independent variables of the experiment, the frequency-domain data revealed that part of those trends are hidden in the experimental setup itself.

7. DISCUSSION/CONCLUSIONS

A perspective flight-path display shows the trajectory to be flown in a synthetic three-dimensional world. The aircraft's state is presented by means of a large number of spatial sources of information resulting from the perspective distortion of the geometrical shape of the tunnel. In order to control the aircraft through the tunnel, the pilot must estimate the aircraft's state using these optical cues. The experiment described was designed to examine some of the hypotheses resulting from a cue-inventory of potential optical cues for a position error. The hypotheses addressed the availability and relative usefulness of two sets of cues for a position error: Linear displacement cues caused by the relative displacements of the individual tunnel frames and angular cues resulting from the longitudinal lines connecting the tunnel frames. The results supported the a priori hypotheses very well. It is shown that performance using the angular cues only (display A), though independent of the forward motion through the tunnel, deteriorates fast when both axes have to be controlled simultaneously. This can be attributed to the fact that the vertical and lateral position errors are shown via the angular cues in a coupled manner. Performance with only the linear displacement cues (display B) showed an opposite effect. Due to the uncoupled presentation of the lateral and vertical position error, the dual axes task had only minor effect on pilot performance. The forward motion through the tunnel, on the other hand, did significantly deteriorate performance. This can be explained by the fact that the tunnel frames move towards the pilot, forcing him/her to constantly shift attention to a new set of frames. The conventional tunnel display, combining both sets of cues (display C), resulted in the best performance for all conditions. The effects of the independent variables on performance with this display showed a mixture of both aforementioned trends, but to a significantly lesser degree. Tentatively one could state that the addition of both sets of cues led to a more robust performance.

The results from the frequency-domain analysis fully support the time-domain data. Moreover, it is shown that part of the observed performance effects could be attributed to the different adaptation of the subjects to the dynamics of the system to be controlled. This allows a much better insight into the observed performance characteristics. It is expected that the future modelling efforts, including a parametric analysis of the identified operator frequency-response functions, will substantially increase this insight. The dual approach followed allows for a more substantial analysis of the observed manual control behaviour. An important disadvantage of the approach, however, is that it requires a highly abstracted design and definition of the experimental variables of interest, reducing the transfer of the results to the 'real world'.

8. ACKNOWLEDGEMENTS

The author would like to thank guru René van Paassen for his support concerning the identification procedure. Further, the author would like to thank Erik Theunissen for the use of the DELPHINS tunnel display software.

9. REFERENCES


MODELLING OF INDIVIDUAL CAR-FOLLOWING BEHAVIOUR

Jeroen H. Hogema
TNO Human Factors Research Institute
PO Box 23, 3769 ZG Soesterberg,
The Netherlands
e-mail: hogema@tm.tno.nl

1 INTRODUCTION

The still expanding traffic demand causes increasing problems in terms of traffic congestion, traffic safety, and environmental load. By use of new technologies (Advanced Transport Telematics, ATT), Dynamic Traffic Management may contribute considerably to a more efficient use of the existing infrastructure. Traffic simulation models constitute an important tool in this context, since these can be used to study the effects of various measures prior to their actual implementation. In the area of traffic simulation, several TNO institutes1 are cooperating in the project Intelligent Traffic Systems (ITS), which aims at the developing of an instrument for the assessment of ITS applications (De Vos & Van der Horst, 1996). The ITS project is sponsored by the Dutch Ministry of Transport, Public Works and Water Management. Its framework consists of both a macroscopic and a microscopic simulation model. The microscopic model, named MIXIC, has been developed in separate projects commissioned by the Transport Research Centre of the Ministry of Transport, Public Works and Water Management (Van Arem et al., 1994).

In microscopic traffic simulation models, a key component is the driver model, which describes the dynamic behaviour of individual road users. Such driver models generally consist of a lane-change model and a longitudinal driving model. Aspects of longitudinal driver behaviour are free driving (where emphasis is on speed regulation), car-following, and approaching slower vehicles (i.e. the transition of free driving to car following). More specifically, static and dynamic car-following behaviour can be distinguished. The static part describes at what steady-state distance the driver will follow a lead car under given stable circumstances, whereas the dynamic part represents how the driver reaches that steady state and how he reacts to disturbances in a car-following situation.

This paper describes a driving simulator experiment which was conducted to investigate both these aspects of car-following behaviour. The reason for using the driving simulator was to have full experimental control over the traffic environment. The aim of the study was to acquire more knowledge on car following, in order to be able to apply it in microscopic traffic simulation models. Simultaneously, several assumptions concerning the current version of the MIXIC driver model could be tested. A full description of this

---

1 Involved in the Intelligent Traffic Systems Project are: the TNO Human Factors Research Institute, TNO Institute of Infrastructure, Transport and Regional Development, the TNO Road-Vehicles Research Institute, and the TNO Institute of Applied Physics.
model is given by Van Arem et al. (1994). The present study was sponsored by the Ministry of Transport, Public Works and Water Management.

Steady-state car-following behaviour is normally described in the literature in terms of distance or time headway control. Headway is computed by dividing a vehicle's following distance by its current speed. Following distance is generally found to increase as a function of speed. Headway is often implicitly assumed not to vary with speed because it combines following distance and speed (see for instance Fuller, 1984). For modelling purposes, it would be attractive if the steady-state following behaviour could indeed be described simply by a constant headway. However, only a few studies explicitly investigated the effect of speed on headway, and the results are not always consistent. Therefore, a question remains whether steady-state car-following behaviour can simply be described on the basis of a constant time headway model.

A specific type of disturbances of a car-following situation will be discussed in this paper, i.e. lane-change manoeuvres which are carried out by a lead car. This results in a sudden increase or decrease of the following distance. Van der Horst and Bakker (1991) carried out a field study in which the reaction of drivers to the manoeuvre of Fig. 1A was registered. Two instrumented vehicles were used: one was the initial lead car and the other carried out the actual lane-change manoeuvre. By using video images recorded in both cars, it was possible to analyze the following distance during the entire manoeuvre. The subjects were not aware that their behaviour was registered. Since speed could not be varied in this study, a possible effect of this factor on behaviour during the manoeuvre could not be systematically investigated. The current simulator study is an extension of this field study in the sense that the manoeuvres of both Fig. 1A and 1B were included, and furthermore speed was varied systematically.

![Diagram of lane-change manoeuvres](image)

**Fig. 1** Lane-change manoeuvres of a car ahead of the subject.  
A decreasing the following distance  
B increasing the following distance
2 METHOD

2.1 Subjects

Eighteen male subjects participated in the experiment. They all had a driving licence for at least three years and they all drove at least 10 000 km a year. Subjects were paid for their participation.

2.2 Apparatus

The experiment was conducted in the fixed-base driving simulator of the TNO Human Factors Research Institute, which is described in detail by Van der Horst, Janssen, and Hoekstra (1991). During the experiment, the subject was seated in a fixed base mock-up of a Volvo 240 and had all normal controls (steering wheel, accelerator, brake, etc.) at his disposal. Based on the control signals, a vehicle model computed the momentaneous state of the vehicle, which had the characteristics of a Volvo 240. A vehicle with automatic gear-shift was used in this simulation study. Feedback of steering forces was given to the driver by means of an electrical torque engine, and of sound by an electronic sound generator (noise of engine, wind, and tyres). The momentaneous position (X,Y) and heading angle (FI) were transmitted via the supervisor to the Computer Generated Images system (Evans and Sutherland ESIG 2000), which computed the visual scene as seen from the position of driver. This image was projected on a screen in front of the mock-up by means of a high-resolution BARCOGRAPHICS 800 projector (visual angles: 50° horizontally, 35° vertically). The outside world used in this experiment consisted of a two-lane motorway. The experimenter was seated in a room next to the mock-up room, where he has access to the control system. Communication with the subject was possible by means of an intercom.

2.3 Procedure

The main issue investigated was how subjects react to lane-change manoeuvres carried out by cars just ahead of them. In addition, the steady-state car-following behaviour was examined. Subjects were instructed to stay in the right lane and the presence of traffic in the left lane prevented overtaking manoeuvres. In order to obtain actual car-following situations, subjects were asked to imagine they were on their way to an important appointment, for which they did not want to be late. They had to drive as they would in reality in such a situation, however, without endangering traffic safety.

There were two lane-change situations: in the first, a car from the left lane changed to the right lane ahead of the subject (Fig. 1A), and in the second, the car ahead of the subject would carry out a lane-change manoeuvre to the left lane (Fig. 1B). This was done at three separate driving speeds of the lead car: 80, 100 and 120 km/h. To prevent the occurrence of lane-changes becoming too predictable, there were also parts included that just consisted of driving at a constant speed, yielding a total of three manoeuvres (lane change to the left lane, to the right lane, and no change). Combined with three speed levels, this gives nine separate conditions for each subject. These were balanced to avoid sequence effects. Each condition was presented twice.
The runs were started on the right lane of the motorway at almost standstill, with no other vehicles present. The subject would then accelerate to what he considered a suitable speed. Three minutes after the start of the run, ten other cars merged into the motorway, five in each lane. These vehicles had the same speed as the subject, and they merged at such a distance ahead of the subject that a car-following situation with a headway of approximately 1.5 s was realized. During the simulation these other cars maintained a headway of approximately 1.5 s with respect to each other. Then the nine scenarios were executed successively, each consisting of the following steps:

- The speed of the other cars was increased or decreased to the speed required for the scenario (80, 100 or 120 km/h).
- Next, there was a certain time of driving at a constant speed to obtain a steady car-following situation. To avoid a too predictable scenario, this interval was varied pseudo-randomly using a uniform distribution between 18 and 34 s (i.e. a mean of 26 s).
- Then, the actual manoeuvre was carried out. Lane-changes were initiated by the changing vehicle turning on its indicator, followed by a lateral movement according to a cosine profile which took 6 s. The new lead car merged just behind the initial lead car, causing a fixed distance reduction of 9 m. This approach was followed because of the limited visual angle of the projection system: in order to guarantee the merging car would always be visible on the projection screen, it had to be well ahead of the subject.
- After the manoeuvre there were 30 s of driving at a constant speed to allow the subject to adjust the following distance and obtain a new, steady car-following situation.

3 RESULTS

3.1 Steady-state car-following

Following distance
The steady-state following distance was measured just before the car ahead of the subject changed lane in each of the three manoeuvres. No effect of 'manoeuvre' was expected because the scenarios were identical before the actual lane-change manoeuvre was carried out. This was confirmed by an initial ANOVA that was carried out using the factors speed, manoeuvre (3 levels: lane change to the left lane, lane change to the right lane, and no lane change), and repetition (2). The latter two factors were not significant and, therefore, the measurements were averaged over manoeuvres and repetitions.

Outliers were identified in the following manner. For each speed condition, the distribution of following distances was derived. Data points that were more than 2.5 sd's away from the mean were identified as outliers and subsequently ignored in the analysis. This process revealed 7 outliers in a total of 324 data points (i.e. 2.2% of all measurements). In the ANOVA, there was an effect of speed on following distance \([F(2,34)=54, p<0.001]\). The mean following distance as a function of speed is shown in Fig. 2.
Headway

In the analysis of steady-state time headway, the same 7 outliers were taken out. The ANOVA with speed as the only factor showed that the headway increased significantly as a function of speed \( F(2,34)=22, p<0.001 \); see Fig. 3. Post-hoc tests showed that the headways of the 80 and 100 km/h conditions differed significantly \( p<0.05 \), as well as the 100 and the 120 km/h condition \( p<0.001 \).
A further question, relevant for modelling and simulation purposes, is how the steady-state following distance can be described using regression. First, linear regression was performed using all data together, but ignoring the outliers mentioned above. The regression equation takes the form of

$$D = a_d + b_d \cdot v$$ \[1\]

where $D$ is the following distance [m], $v$ is the speed [m/s], $a_d$ is the intercept [m] and $b_d$ is the regression coefficient [s]. This can be interpreted as a time headway of $b_d$ [s] added to a constant distance margin of $a_d$ [m]. There was a statistically significant relationship between speed and distance [$p<0.001$]; the intercept was found to be $-71$ m and the regression coefficient $5.1$ s. Consequently, the regression equation yields negative (i.e. senseless) distances for speeds smaller than $14$ m/s (50 km/h). This clearly shows that the regression results should not be considered valid over a much wider range of speeds than the 80 to 120 km/h used in the experiment.

Next, it was tested if the following distance could be described better by a quadratic regression equation:

$$D = a_d + b_d \cdot v + c_d \cdot v^2$$ \[2\]

The quadratic model was significantly better than the linear model [$F(1,314)=3.8$, $p<0.05$]. The parameters found were $a_d=144$ m, $b_d=-11$ s, and $c_d=0.29$ s$^2$/m. The resulting function has a minimum of $42$ m at $v=18.8$ m/s (68 km/h). Below this minimum, decreasing speed would increase as a function of speed according to these results. Again, the results should not be extrapolated outside the 80-120 km/h range.

If Eqs. [1] and [2] are to be used over the full range of realistic speeds (i.e. from 0 to motorway speeds), $a_d$ represents the distance at 0 km/h, for instance in queues. A realistic value for $a_d$ seems to be a few meters. Regression was performed again with the intercept fixed at 3 m. The coefficient in the linear regression was now found to be $b_d = 2.6$ s; the coefficients in the quadratic regression were $b_d = -0.4$ s and $c_d = 0.1$ s$^2$/m, respectively.

The linear regression equation of headway as a function of speed was also calculated. This equation has the following form:

$$HW = a_h + b_h \cdot v$$ \[3\]

where $HW$ is the headway [s], $v$ is the speed [m/s], $a_h$ is the intercept [s] and $b_h$ is the regression coefficient [s$^2$/m]. There was a statistically significant relationship between speed and headway [$p<0.001$]. The regression coefficient was found to be $0.1$ and the intercept $0.07$, which did not deviate significantly from 0 [$p<0.8$].

### 3.2 Lane-change reducing the gap

The results presented in this section refer to the situation where a car merges from the left lane to the right lane, in the gap between the subject's vehicle and his initial lead car (see Fig. 1A).
Following distance
As in the study of Van der Horst and Bakker (1991), the following distance of the subject during these manoeuvres is described by four parameters as shown in Fig. 4. The instant at which the lane-change occurs, $t_{\text{merge}}$, is defined as the moment at which the centre of the changing car passes the lane markers between both lanes. Because the initial gap ahead of the subject is filled by the merging car, there is a discontinuity of the following distance $D$ at $t_{\text{merge}}$.
- $D_0$ is the initial following distance at the beginning of the manoeuvre.
- $D_1$ is the distance immediately before $t_{\text{merge}}$.
- and $D_2$ is the distance immediately after $t_{\text{merge}}$.
- After the reaction of the subject, the final following distance $D_3$ is reached, which is not necessarily equal to $D_0$.

The process of distance adjustment can be characterized by a first-order time constant $\tau_D$, defined as the time interval between $t_{\text{merge}}$ and the moment at which the distance reaches 63% of the range between $D_2$ and $D_3$.

From each run, the four parameters $D_i$ and the time constant $\tau_D$ were determined. Outliers were identified in the following manner. For each speed and for each of the four measurement points, the distribution of following distances was derived. Data points that were more than 2.5 sd's away from the mean were marked. If more than two of the four points describing a single manoeuvre were marked, then the entire manoeuvre was regarded as an outlier and therefore ignored in the analysis. This occurred in three out of 108 manoeuvres (2.7%).

For comparison with results from Van der Horst and Bakker (1991), the distributions of the four parameters describing following distance are indicated in Fig. 5 by the median and the 15th and 85th percentile. There is a clear reduction in following distance between $D_1$ and $D_2$, and between $D_2$ and $D_3$ a recovery is apparent for the median and the 15th percentile.
Fig. 5 Median, 15\textsuperscript{th} and 85\textsuperscript{th} percentile of following distance at the four moments describing the manoeuvre when a car merges into the subject’s lane.

Fig. 6 Mean following distance as a function of speed at the four moments describing the manoeuvre when a car merges into the subject’s lane.

To investigate the effect of speed on the following distance during the manoeuvre, an ANOVA was carried out with the factors speed, and time moments D\textsubscript{0} to D\textsubscript{3}. The averages over the two repetitions were used in this ANOVA. Both factors were statistically significant \([F(2,34)=33, p<0.001\) and \(F(3,51)=9.4, p<0.001,\) respectively]. There was also a significant interaction, see Fig. 6 \([F(6,102)=2.5, p<0.05]\). Planned compari-
sons revealed that the decrease in distance at $t_{\text{merge}}$ ($D_2$-$D_1$) was significant at all speed levels [all three $p<0.001$]. Averaged over all speeds, there was a significant difference between $D_0$ and $D_1$ [$p<0.05$]; this effect can be attributed to the 80 and 100 km/h conditions [both $p<0.05$]. This indicates anticipatory behaviour of the subjects: an adjustment of the following distance has already started before the merging vehicle has physically occupied the gap ahead. From $D_2$ to $D_3$, the only significant effect is found at 80 km/h [$p<0.01$]. When comparing $D_0$ and $D_3$, it appears that the final and the initial following distances do not differ significantly at 80 and 100 km/h [both $p<0.4$], but at 120 km/h $D_3$ remains smaller than $D_0$ [$p<0.05$].

A further ANOVA was carried out to investigate the effect of speed on the time constant $\tau_D$, with speed as the only factor. No significant effect was found [$F(2,34)=0.04$, $p<0.96$]. The mean of $\tau_D$ was 9.1 s and the sd 4.5 s. This indicates that the process of distance adjustment after the lane-change is not influenced by the speed at which it occurs.

**Headway**

The headway during the manoeuvre is described by 4 points $H_0$ to $H_3$, in the same manner as the distance measurements $D_0$ to $D_3$ defined above. The time constant $\tau_H$ is defined in the same manner as $\tau_D$. Two out of 108 manoeuvres (1.9%) were found to be outliers using the procedure described above. These outliers were ignored in the analysis.

The median and the $15^{th}$ and the $85^{th}$ percentile of the four headway measurements are given in Fig. 7. The data appear to be comparable to those of Fig. 5, although the effect of speed on time headway is smaller than the effect of speed on following distance.

![Fig.7 Median, 15th and 85th percentile of headway at the four moments describing the manoeuvre when a car merges into the subject's lane.](image)
An ANOVA was carried out on the mean headway, using the factors speed and time moments $H_0$ to $H_3$. Results showed that both factors were significant [$F(2,34) = 9.6, p < 0.001$ and $F(3,51) = 16, p < 0.001$, respectively]. There was no significant interaction [$F(6,102) = 1.9, p < 0.37$]. The means as a function of both factors are shown in Fig. 8. Planned comparisons revealed that there is a significant difference between $H_1$ and $H_2$ [$p < 0.001$]. No significant differences occurred between $H_0$ and $H_1$ or between $H_2$ and $H_3$. Averaged over all speed levels, $H_3$ remains significantly smaller than $H_0$ [$p < 0.01$].

![Fig. 8 Mean headway as a function of speed at the four moments describing the manoeuvre when a car merges into the subject's lane.](image)

The values found for $\tau_H$ equalled those of $\tau_D$ (mean $\tau_H 9.1 \text{ s, sd 4.5 s}$). Therefore, an ANOVA on $\tau_H$ with speed as the only factor revealed no significant results either.

### 3.3 Lane-change increasing the gap

This manoeuvre is depicted in Fig. 1B: the subject's lead car carries out a lane-change manoeuvre to the left lane, thus increasing the subject's following distance. The manoeuvres are described by the same parameters as used in Section 3.2. Outliers were identified with the method described in that section as well. The number of outliers was 2 out of 108 (1.9%) both for following distance and for headway.

**Following distance**

An ANOVA which was carried out on the following distance with the factors speed and time moments $D_0$ to $D_3$ revealed significant effects of both factors [$F(2,34) = 54, p < 0.001$, and $F(3,51) = 166, p < 0.001$ respectively]. There was also a significant interaction between speed and time moments [$F(6,102) = 2.7, p < 0.05$], see Fig. 9.
Planned comparisons revealed that at all speed levels the following distance increases from $D_1$ to $D_2$ and decreases from $D_2$ to $D_3$ [all $p < 0.001$]. There is no significant difference between the initial distance $D_0$ and the final distance $D_3$ [all $p > 0.43$]. From $D_0$ to $D_1$, there is a small increase at 80 and 100 km/h [$p < 0.05$].

An ANOVA on $D_0$, with speed as the only factor, revealed no significant effect. The mean was 9.5 s and the sd 3.3 s.

**Headway**

Also the headway was analysed by means of an ANOVA with the factors speed and time moments $H_0$ to $H_3$. Both factors were significant [$F(2,34) = 16$, $p < 0.001$, and $F(3,51) = 212$, $p < 0.001$, respectively]. There was no significant interaction between the factors [$p < 0.9$]. Planned comparisons showed that there was a minor increase of headway between $H_0$ and $H_1$ [$p < 0.05$]. Furthermore, headway increased from $H_1$ to $H_2$ and decreased from $H_2$ to $H_3$. The final headway $H_3$ did not deviate significantly from the initial headway $H_0$ [$p < 0.18$].
An ANOVA on $\tau_H$, with speed as the only factor, revealed no significant effect. The mean was 10.4 s and the sd 3.4 s; the means of $\tau_H$ and $\tau_D$ did not deviate significantly [$p<0.16$].

### 4 DISCUSSION AND CONCLUSIONS

Summarizing the results of this simulator study, it was found that the steady-state time headway was not constant, but increased as a function of speed. Furthermore, when a lane-change manoeuvre was carried out by a vehicle ahead, subjects appeared to compensate the changed headway fully, i.e. there was no difference between the initial and the final headway or following distance. The results also revealed an anticipatory reaction to these lane-change manoeuvres: subjects already started their reaction before the changing lead car was half way its lateral lane-change displacement.

A basic question concerning car-following behaviour remains how following distance varies as a function of speed. For modelling purposes, it would be attractive if steady-state car-following behaviour could simply be described based on a constant time headway, but the current results show that time headway increased with speed. There are not many other studies investigating time headway as a function of speed. Colbourn, Brown, and Copeman (1978) did vary speed systematically in a car-following experiment, and they reported a significant effect of speed on following distance. Unfortunately a statistical analysis of time headway was not presented, but their results seem to indicate that headway increases slightly as a function of speed. Van Winsum (1993) found in a driving simulator experiment that headway did not vary as a function of speed. Time headway varied among subjects, but appeared to be stable within subjects. However, the range of
speeds used was 40 to 70 km/h, which is rather low compared to normal motorway
speeds. In a study on driver behaviour in fog, Hogema and Van der Horst (1994) found
that headway increased with speed.

The anticipatory behaviour when a lead car changes lane was found both in this experi-
ment and in the study of Van der Horst and Bakker (1991). Their results also indicate that
drivers do not fully compensate a reduction in headway whereas the current experiment
revealed a full distance compensation in the 80 and 100 km/h condition. This discrepancy
could be caused by the difference in the stimulus: the lane-change manoeuvre itself was
executed differently. In the field study, the car coming from the adjacent lane merged
approximately in the middle of the initial gap, which approximately halved the initial
following distance (i.e. a proportional distance reduction, new headway around 0.5 s). In
the current experiment, the new lead car merged just behind the initial lead car, causing a
fixed distance reduction of 9 m (see Section 2.3). Since in the simulator experiment the
following distance is only slightly reduced, a minor reaction of the subject is sufficient to
regain the previous following condition. On the other hand, at higher speeds following
distances are larger and therefore a fixed distance reduction of 9 m becomes relatively
small, and may therefore possibly be ignored by subjects.

For the situation where the lane change of the lead car increases the following distance, it
was found that subjects adjust their following distance to its original level (in terms of
following distance as well as in terms of headway). No real-world study is available for
comparison here.

The findings of the current study have several implications for the driver model of the
microscopic traffic simulation model MIXIC. First of all, the distance controlling
mechanism should not be based just on a constant headway assumption. Therefore, the
current MIXIC approach (where the reference following distance increases as a quadratic
function of speed) should not be simplified. The regression results of this experiment were
obtained in the range of 80 to 120 km/h. These results should not be extrapolated far
outside this range because this can lead to senseless results (negative following distances
for instance). Therefore, the regression results are not suitable for direct use in MIXIC,
also because the absolute levels of the headways found in this experiment were relatively
high. The overall mean steady-state headway was 2.7 s, whereas in real traffic, drivers
are easily capable of maintaining headways of 1 s (Van der Horst, 1993).

The results of the lane-change manoeuvres of cars ahead revealed a form of anticipation:
drivers already initiated an adjustment of the following distance before the merging
vehicle has completed its manoeuvre. In the current MIXIC model, however, the reaction
is only started after the lane-change manoeuvre has been completed. It is recommended
that the anticipatory mechanism is also introduced in MIXIC; this can be expected to have
a stabilizing effect on the traffic stream.

It should also be noted that in traffic there may be a discrepancy between the headway a
driver would prefer to maintain and the actually realized headway, since the latter is partly
inflicted by the traffic situation. When a single driver attempts to maintain a gap that is
relatively large, this gap will soon be filled by another vehicle. Following distance is not
only a function of speed, but also of the traffic situation. For instance, Saad (1984)
reports that long periods of critical following specifically occur in heavy traffic conditions.
Also Hogema, Van der Horst, and Bakker (1994) found that the percentage of short
headways (< 1 s) increases as a function of traffic volume, which shows that headway is
not purely a behavioural measure but also is partly determined by the traffic situation. A question remains how this can be adequately modelled.

In conclusion, the findings of this study confirm the assumption of a non-constant time headway on which the microscopic driver model of MIXIC is based. With respect to reactions to lane-change manoeuvres of other traffic, some suggestions were derived for improving this model.

REFERENCES


Session 2

Human Operator Decision Support Systems and Interface Design
A Fuzzy Decision Support System for Magnetic Component Design.

D. Holt  U. Kaymak*  J.B. Klaassens  H.R. van Nauta Lemke

Delft University of Technology
Faculty of Electrical Engineering, Control Laboratory
P.O. Box 5031  2600 GA  Delft, the Netherlands
tel: +31-15-2783371, fax: +31-15-2786679, Email: U.Kaymak@et.tudelft.nl

Abstract
In many areas of electrical engineering, magnetic components like transformers and inductors are necessary parts of an electrical circuit. Designing such a component means finding a suitable combination of a ferrite core and a copper wire, while a great variety of requirements have to be satisfied. However, finding the optimal combination is complicated because of the large number of available sizes and types of cores and wires. A Decision Support System (DSS) can assist a designer of magnetic components in selecting the suitable component alternatives and in ranking them, hereby improving the efficiency and the results of the design procedure. This paper describes the aims and the typical problems of magnetic component design, and discusses a Decision Support System for facilitating the design procedure. As an example, the structure of a DSS is described that selects an ac-inductor of a power converter circuit.

Keywords: Decision Support Systems, magnetic components, fuzzy sets, fuzzy multiple attribute decision making, fuzzy design.

Introduction
In many areas of electrical engineering magnetic components are necessary parts of an electrical circuit. For example in power electronics these components are essential for the operation of power converters, in addition to the power semiconductors and capacitors. Magnetic components utilize their magnetic circuit to store or transform energy in order to serve applications such as transforming power, filtering and resonating. The two most commonly used magnetic components are the transformer and the inductor, which are basically composed of a core of magnetic material (ferrite) and turns of copper wire.

Designing a magnetic component means finding the optimal combination of the core and the wire, while a great variety of requirements are satisfied. However, because of the large number of available sizes and types of cores and wires, a designer uses an iterative design procedure, while making extensive use of experience, physical and heuristic knowledge and rules-of-thumb. Even if a computer is used to perform the tedious calculations, still the human design method is inefficient because of the time that is involved in the design, and may cause a sub-optimal final design.

This paper describes how a Decision Support System (DSS) can be realized that assists the designer of magnetic components by selecting and ranking a set of feasible component alternatives. The paper is organized as follows. Section 1 describes magnetic component design and its problems. Then, in section 2, the information necessary to create the DSS is discussed and the structure of an example DSS is described. Finally, section 3 presents the conclusions and the expectations from the proposed DSS.

* Author to whom all correspondence should be addressed.
1 Magnetic component design
This section explains what magnetic components are, how they are designed and it treats the shortcomings of the traditional human design method.

1.1 Magnetic components
Basically, magnetic components are constructed of a core of magnetic material and one or more windings of isolated copper wire with a certain number of turns per winding. The component is assembled by winding the wire around a coil former and then joining the core halves together, as shown in Figure 1a. and Figure 1b.

![Diagram of E-core, coil former and wire.](image)

Manufacturers offer a large number of types and sizes of cores and wires. More than 15 different ferrite core shapes can be obtained, all of them available in a range of sizes and most of them in several ferrite grades. Figure 1a. shows the example of an E-core. The available wire can be round, flat or bunched (Litz wire), each available in a range of copper cross-section areas and with different classes of isolation. Some examples of wire types and sizes are shown in Figure 1c. It is obvious that, to construct a component, a large number of combinations of cores and wires is possible. Applications like the inductor can have a varying discrete number of turns depending on the airgap width, which even increases the number of possible alternatives.

1.2 Design of magnetic components
The specific design of a magnetic component has an important influence on the overall performance and the cost of the system for which it is applied. This influence is caused by three troublesome characteristics of magnetic components: 1) usually they are relatively large and heavy items of a circuit, 2) they produce losses and heat and 3) they are not easy to wind and assemble automatically. Therefore, it is useful to find a component that has optimal performance concerning the requirements.

It has already been mentioned that a magnetic component has to be satisfy a great variety of requirements. The requirements imposed on a magnetic component are physical constraints, specifications and designer preferences. They depend strongly on the application for which the component is designed. Some examples of requirements are maximum storable energy, maximum allowable temperature rise, peak current, allowed power loss, weight and cost.

The role of decision making during the design is that a designer has to take decisions about which alternatives are not appropriate and which of the remaining alternatives are the best. The decisions are taken on the basis of how much each alternative satisfies the requirements. Because these requirements can be vague, imprecise or uncertain, a human designer makes design trade offs using his (or her) human ability to deal with vague information. Despite this ability to make fuzzy decisions, he has some shortcomings when dealing with large quantities of information, as explained in section 1.3.

1.3 Design problems and solutions
In order to manage the large number of component alternatives and the large number of requirements, a human designer uses several decision steps to come to the final design. The first step is the initial choice of a core by means of heuristic knowledge, experience or by using the catalogue suggestions. Then, the designer
starts an iterative 'trial and error' decision procedure, repeatedly choosing a different wire or a different core and calculating the resulting values of the component characteristics.

This procedure has two drawbacks. The first drawback is that the designer has limited insight in the consequences of a decision. Because it is not clear whether a choice leads to a better alternative or not, many iterations may be necessary. The second drawback is that the designer may reject many alternatives too early in the decision process, because he does not assess all possible configurations of cores and wires. Hence, finding the optimal design may not be possible.

To get around the drawbacks of this method, designers are supported with tables in books or catalogues that simplify the initial choice of the core [5] [6] [11] [12]. The first computer algorithms were implemented in the 60's [4], taking over tedious calculations. However, the critical decisions have always been taken by the designer himself, because only human knowledge and experience can handle the complex dependencies and trade-offs.

With the developments in artificial intelligence and fuzzy logic, new doors have been opened towards computer design of magnetic components [1]. In the following section a fuzzy Decision Support System is introduced that assists the designer of magnetic components by selecting a set of possible alternatives and ranking them depending on how much they satisfy the design constraints and preferences.

2 DSS for component design
This section describes a Decision Support System (DSS) for designing magnetic components. The aims of the DSS and the information necessary to realize the system are explained. The last part proposes a structure of a DSS used to design an ac-inductor for a power converter circuit.

2.1 A Decision Support System
The proposed DSS will assist the designer of magnetic components by selecting a set of appropriate alternatives and ranking them depending on how much they satisfy the design constraints and objectives. The aim of the DSS is to select a few possible, good, alternative designs by performing preparatory decisions, selections and rankings. The designer's attention can then be focused on the evaluation of these designs. In this way the quantitative abilities of computers and qualitative abilities of humans are used efficiently.

To create the DSS, the information that is necessary to design a component has to be determined. This information can be retrieved by interviewing an expert. However, because not all information will be obtained at once, there is need for a continuing personal working relationship with the expert during the construction of the DSS. Two kinds of information have to be retrieved, namely the design constraints and objectives as well as the expert knowledge of the design procedure. In the next part a characterization of the information is given, illustrated by means of the design of an ac-inductor.

2.2 Fuzzy decision criteria
This paragraph describes the first type of information necessary to design a component, namely the constraints and the objectives of the design. The constraints and objectives are defined by 1) the physical boundaries, 2) the application specifications and 3) the designer preferences. Most of these are set up by means of expert knowledge. Because the information can be vague, imprecise or uncertain, a mathematical representation based on the fuzzy set theory is introduced. The constraints and objectives together are also called fuzzy decision criteria.

Physical constraints
The physical constraints are boundaries imposed by physical properties of the used materials. The boundaries are not exact, but contain a certain range of tolerance. In general, the tolerance can only be estimated by experts who have practical experience with the materials. The three physical constraints of the example of an ac-inductor are the maximum possible value of flux density B in the core, the maximum allowable airgap width I_g and the temperature (maximum environmental temperature θ_e and maximum allowable temperature rise Δθ).

Specifications
The component specifications are the constraints imposed by the application. They form the necessary inputs to the DSS, such as current waveshape, maximum current, inductance, peak
voltage, frequency and ambient temperature. The specifications usually also include maximum (or minimum) values for the inductor properties such as mass, volume, power loss and cost.

Specifications are often provided as exact (crisp) values, hereby forming strict limiting boundaries. However, this assumption of exact specifications does not always make sense, because in practice specifications have a range of tolerance. For example, an inductor used in a filter application with a specified inductance \( L \) usually has a substantial range of tolerance around \( L \). If this range of tolerance is allowed, then the set of suitable alternatives may be larger than the case when only a single value for \( L \) is allowed. The former set might contain better alternatives than the latter, so the assumption of exact specifications actually obstructs the finding of an optimal component.

**Designer preferences**
Although the designer has to take the specifications into account, he usually has some personal preferences about several properties of the component, like mass, volume, power losses, cost or temperature rise. The preferences of the designer depend on aspects such as the application, the specifications, knowledge and experience. To help a designer making realistic assumptions for his preferences, the DSS can be made interactive: the designer is allowed to change his preferences based on results of the DSS during the design procedure.

Most of the constraints and the preferences mentioned before contain vague, imprecise or uncertain boundaries. The fuzzy set theory is an advantageous way to represent these requirements.

**Fuzzy logic and fuzzy sets theory**
Fuzzy logic has been proposed by Zadeh in the 1960's [14] as a means to model uncertainty of non-probabilistic nature and is an extension of conventional (Boolean) logic. A good example of such uncertainty is the imprecision in human behaviour and reasoning. Fuzzy logic introduces the concept of partial truth values that lie between "completely true" and "completely false". There is a strong relation between *fuzzy logic* and *fuzzy sets theory*, similarly to the relationship between Boolean logic and conventional set theory. Fuzzy sets theory is introduced as an extension of conventional sets theory, allowing partial membership in the set. A fuzzy set \( A \) is defined in the universe of discourse \( X \) via its characteristic function, usually called membership function, \( \mu_A(x) : X \rightarrow [0,1] \) that is defined as follows:

\[
\begin{align*}
\mu_A(x) &= 1 & x \text{ belongs completely to } A \\
\mu_A(x) &= (0,1) & x \text{ belongs partially to } A \\
\mu_A(x) &= 0 & x \text{ does not belong to } A
\end{align*}
\]

A fuzzy set can also be represented as a list of ordered pairs \( \{x, \mu_A(x)\} \). The value of the membership function \( \mu_A(x) \) is called *membership grade* (degree). The terms "membership function" and "fuzzy set" are sometimes used interchangeably.

**Examples**
The uncertain, vague and imprecise nature of the criteria defined before can be represented by membership functions. A membership function gives the degree to which a criterion is satisfied as a result of a certain parameter value of a component alternative. This is illustrated in Figure 2, where some examples of membership functions are shown.

![Membership functions expressing physical constraints](image1)

![Membership functions expressing preferences](image2)

![Membership functions expressing physical constraints](image3)

**Figure 2** Examples of membership functions describing the fuzzy constraints and preferences.

2.1-4
In Figure 2a the membership functions are shown expressing the grade of appropriateness of the magnetic flux density $B$ in the core and of the airgap width $l_g$. The upper boundary of $\mu(B)$ is the saturation flux density $B_{sat}$ of the core material. The lower boundary of $\mu(B)$ is determined by the maximum storable energy. The fuzzy upper boundary of $\mu(l_g)$ of a specific core is determined by the information of an expert on the maximum allowable airgap width $l_{g,\text{max}}$.

Figure 2b and Figure 2c show the membership functions expressing the designer preferences of an inductor in a space application and a consumer product application. For example, the shape of the fuzzy set $\mu(m)$ shows that the mass of the inductor for the space application should be low, while the mass of an inductor for a consumer product is not important, except for a specified maximum allowable mass $m_{\text{max}}$. From the membership functions of the cost $\mu(K)$ it can be seen that there is no limiting boundary for the cost of an inductor in space design, but in the application of a consumer product the costs should be low.

By means of these fuzzy sets the fuzzy decision criteria are described. The following part discusses the design procedure. In order to realize an efficient decision procedure for the DSS, expert knowledge about the procedure has to be retrieved.

2.3 DSS structure

In this part, firstly the retrieved knowledge for designing ac-inductors is described. Secondly, a DSS structure is proposed using this knowledge.

**Designer knowledge**

Besides the expert knowledge included in the decision criteria, additional expert information is retrieved in order to realize a more efficient decision procedure. The expert information may consist of rules, procedures, assumptions or any other kind of applicable knowledge.

For the design of an ac-inductor, the following information is retrieved during the construction of the decision procedure:

1. The equations and assumptions expressing the relations between the component parameters.

**Example:** We assume that the energy stored in the inductor is negligible, except from the energy stored in the airgap. We assume that the cooling factor $\alpha_{\text{cool}}$ is equal for all component alternatives.

2. The multi step character of the decision structure and the sequence of decision steps. **Example:** The first decision steps distinguish the set of possible alternatives using the limiting boundaries. The most efficient sequence of the steps is found to be: 1) core selection, 2) calculation of the range of number of turns for each core, 3) wire selection for each core 4) determination of the inappropriate numbers of turns.

3. Heuristic information and the rules of thumb. **Example:** The experience of the expert can be used to determine more preferable combinations of component parts or certain parameters that are needed by the decision system, such as the support and the kernel of the membership functions. For instance, round wires are easier to wind which simplifies the assembly of the magnetic component. Similarly, a very large airgap is not desirable since the fringing of the magnetic field causes eddy current losses in the copper.

Not all expert knowledge is useful for implementation in the DSS. Some rules-of-thumb, like for example the information that the temperature rise should not be too low, or the rule that the core window should be completely filled with wire, are rules that simplify only the human design method. However, for computer implementation, these simplifications reduce the set of possible alternatives unjustly.

**DSS structure**

The DSS has to be structured in such a way that the calculation effort should be reasonably small, while no alternatives are rejected unjustly. Based on the knowledge of the designer, a decision structure such as shown in Figure 3 is used.
The decision structure of an example DSS for ac-inductor design is illustrated in Figure 3. The decisions are taken in a hierarchical manner. At each level of the hierarchy, a number of alternatives is rejected, which are not within the set of best alternatives. In this way, the computational effort is reduced considerably. The initial set of alternatives is the set of all possible combinations of cores and wires. Steps a. and b. distinguish the (groups of) suitable alternatives. Step c. is a ranking step.

**a. Initial selection**

The initial selection steps reject the totally inappropriate combinations using the limiting boundaries of some requirements. The outcome is the complete set of possible alternatives satisfying the constraints to some degree.

**b. Removal of dominated alternatives**

The second decision step rejects the alternatives that are dominated by any other alternative. This means that the dominated alternatives, i.e. the alternatives that have a mass, volume, cost and power loss higher than some other alternative, are not considered further. Hence, the irrelevant alternatives are removed.

**c. Final ranking**

The final step ranks the remaining alternatives depending on how much they satisfy the requirements. This ranking is determined by a decision function that combines the membership functions of the requirements, taking into account their weights. For the decision function, often called a goal function, various types of fuzzy mathematical operators can be used. In general, the goal function should reflect the aims and the preferences of the decision maker. Fuzzy sets theory provides a variety of decision operators for modelling conjunctive, disjunctive, averaging, compensatory and other types of decision behaviour. Examples of these functions can be found in [2] [7] [8] [9] [10] [13] and [15]. At this moment a parametric generalized goal function is implemented for fuzzy decision making with unequally weighted criteria [3].

Between the three decision steps, an interactive method of preference assignment can be implemented to enable the component designer to have direct influence on the decisions. After the final ranking step, the designer's attention can be focused on a few satisfactory remaining design alternatives. The determination of the best alternative will be a trade-off between the suggestions of the decision support system and the personal preferences of the designer.

**3 Conclusions**

In this paper the design of magnetic components has been considered. In order to manage the large number of component alternatives and the large number of requirements, a human designer uses several decision steps to come to the final design. Because of the drawbacks of the human design method, a Decision Support System is proposed that assists the designer of magnetic components with the selection and ranking of the component alternatives.

The aim of the DSS is to select a few possible, good, alternative designs by performing preparatory decisions, selections and rankings. The designer's attention can now be focused on a few satisfactory design alternatives, which can be studied further. In this way the quantitative abilities of computers and qualitative abilities of humans are used efficiently.
The proposed DSS for the design of magnetic components is now being implemented. The system uses expert knowledge and heuristic information in order to improve significantly the speed and the results of the design procedure. The structure of the proposed DSS is plain and transparent. All tedious operations, comparisons and classifications are performed by the system. An interactive method of preference and weight assignment will enable the component designer to have direct influence on the decisions of the system.

References


Décision support systems and modern maritime air defence
Fuzzy identification of targets

W.W. Sillevis Smitt, B. Bruggeman

Royal Netherlands Naval College/Delft University of Technology
Department of sensor, weapon and command systems
P.O.Box 10.000, 1780 CA Den Helder, The Netherlands

1. Abstract

Naval air defence officers nowadays face an air threat that is characterised by increasing complexity, while reaction times get shorter. Therefore the development of air defence systems is of primary interest in the design of new naval combatants. Besides weapons and sensors, also the improvement of the decision systems for air defence is subject of study.

Target identification is one of the key processes in maritime air defence (MAD). In this process information from different sensors and sources is integrated in order to determine the class or type of the target. Difficulties in this process are caused by uncertainty in information and the real time character of the process. Especially information about the target trajectory is uncertain and, although of great importance, hardly used in automatic identification systems.

In this study the specific problems of target identification are addressed. A multi-level structure for target identification has been developed. Target types and classes are represented as hypotheses. A multi-hypothesis testing technique is used to compare observed features and hypotheses. In order to capture the uncertain trajectory data a fuzzy algorithm is proposed. Two methods to determine the similarity between observations and hypotheses for these uncertain data will be discussed. Finally an experiment with the proto-type simulation model of the identification system illustrates the possibilities of this fuzzy approach.

2. Keywords

Maritime air defence, decision support systems, target identification, similarity measures, fuzzy decision making, fuzzy systems.

3. Introduction

During the last ten years the international situation has changed dramatically. Warships are more and more deployed in maritime blockades and peacekeeping operations, where they have to operate close to enemy shores within reach of all kinds of anti-ship missiles (ASM's) and aircrafts. The development of new ASM's is focused on short reaction times by using higher speeds, lower altitudes and all kinds of evasive manoeuvring [BONSIG93] [HOOTON95]. Therefore the air threat is now characterised by short reaction times and a diversity of modern ASM's.

Consequently a lot of effort is put into the improvement of maritime air defence systems [RICHARD92]. The improvements concentrate on the development of new sensors, new weapons, integrating sensors within a task force and advanced command and control systems.

This trend leads to another development: the increase in information. New sensors and data link systems overwhelm the air defence officer(s) with data that plays a role in the ship's air defence.

In order to cope with the increasing amount of information in shorter reaction times, decision-support systems (DSS) are needed. A good description of a DSS is: 'it exploits intellectual and computer-related technologies in order to improve creativity in decisions that really matter'. This means that the coordination between man and computer is of great importance.

For a DSS the distinction can be made in:

- the kernel: the kernel of a system consisting of architecture/structure, communication, command & control, man-computer dialogue, information/knowledge;
- the application: the applications are (depending on the mission) grafted upon the kernel.
Important questions, with regard to the applications, are 'How to structure them?' and 'How to cope with all types of uncertainty and information?', that the interaction between application and man should be realistic and acceptable for men.

The aim of this study is to come to a general structure for a key-process of MAD: the identification process. Incidents with USS Stark (struck by two Exocet missiles) and USS Vincennes (shot an Iranian civilian aircraft) are examples of what can go wrong when the target is not correctly identified.

The correct identification is not only important for the correct assignment of "friend" or "foe", but also for improving the ship's defence tactics. Early identification can support the tuning of parameters for the ship's sensors and weapons.

In this article an identification algorithm based on fuzzy decision making and similarities is presented. Sensor observations are the input for the identification process. The output is a set of possible target identities with a measurement of belief. This measurement is called a fuzzy identity measure. In chapter 4 some characteristics of the problem of target identification are explained. In chapter 5 a multi-level structure for target identification is proposed. The particularities of each level are discussed. Special attention is given to the interpretation of target trajectory information, which is closely related to human reasoning. In chapter 6 the results of a simulation are demonstrated. Finally the conclusions and recommendations for further research are given in chapter 7.

4. The problem of target identification

The process of target identification in maritime air defence can be defined as the determination of the identity of an enemy flying object.

There are different levels of identification. Targets can be classified according to sort, class and type (figure 1).

In figure 1, identification to the level of target class gives the general identity of the aircraft or missile. The characteristics and capabilities are similar within a class, so the identity known at class level permits general conclusions on defence tactics and defence system settings. Each class contains one or more target types. A target type is the most detailed identity level in maritime air defence. All kind of information on characteristics of ASM's and aircraft is known at the type level. A target identified by type permits a suitable defence tactic and parameter setting of the ship's defence systems.

The process of target identification is one of the processes in the air defence model. The process is fed by sensor information and provides the command and control process with information that is used to eliminate the threat. It is important to realise that not only sensor information is used, but also historic knowledge and situational information. Figure 2 shows the identification process in relation to other processes and information/knowledge in the air defence model.

Three main characteristics have to be taken into account when designing an identification system.

1. Multi-type information.

The identification problem is a multi-type information problem. The information comes from various sensors and sources, on different time-scales, formats and dimensions. For instance, a threat situation can be characterised by the following information:

- Radar contact: Incoming target in position 045 degrees, 20 miles, speed around 300 knots, course 225 degrees, height 1000 ft;
- ESM-intercept: Interception of radar signal 10 Ghz, pulse width 2.5 μS, PRF 0.7 sec;
- Intelligence: A hostile ship of type A might be about 50 miles north-east. This ship is fitted with ASM's of type B.
All this information is important for the identification of the target, but the structure and scale are not uniform, so integration and fusion is not 'just like that' possible. This leads to the following important requirement for an identification process: all information has to be transformed to comparable units of the same dimensions. Only then integration and fusion of information will be possible.

2. Uncertainty.

A problem in air defence arises because the information may be incomplete, imprecise, fragmentary, contradictory or deficient in one way or another. Some information is observed by the unit’s own sensors, other information through data-link, telex or voice reports. The information is subject to atmospheric losses, interference from other sources, manipulation or deception. Uncertainty in information can be divided into three different kinds of information uncertainty [KLIR95]:

- fuzziness (lack of definite or sharp distinctions), also known as vagueness, unclearness and haziness;
- discord (disagreement in several alternatives), also known as conflict and discrepancy;
- nonspecificity (two or more alternatives not distinguishable), also known as variety, diversity and imprecision.

The uncertainty is increased, because a target in an air defence scenario will be not co-operative. Instead, the enemy tries to remain undetected and conceal its identity and intentions. In maritime air defence these "countermeasures" can be divided into different classes:

- concealment: Methods to prevent the enemy from observing deployments, capabilities intentions and movements;
- jamming: Interfering the enemy sensors by sending out high power signals;
- deception: Injection of false or misleading information.

3. Real-time process.

Another characteristic is the real-time character of the identification process. The development of there being less time for reaction means that the time factor in maritime air defence is becoming increasingly important. The time factor has three aspects. Firstly, there is often no time for collecting all the necessary information. The identification process starts immediately after detecting an air contact and conclusions as to the identity have to be made, even if only limited and uncertain data is available.

Secondly, the process is dynamic. During the process, sensor data will vary, so the conclusions on target identity might differ. In order to take these variations into account and to support the decision process it is desirable to visualise the belief in the target identities. Figure 3 shows an example of the output of the identification process versus time.

![Figure 3: Dynamic belief values](image)

It is obvious from this figure that at different times target 1, 3 and 4 are the most likely target identity. The aspect that only limited time is available can have a big impact on the decision. Whenever possible this should be taken into account in the design of an identification system.

Thirdly, some sensors will need more time for data collection and interpretation then others. This means that various time scales are used.

5. Process of target identification

As a result of the three mentioned characteristics of the identification process the requirements for the fusion techniques are high. The process should be able to handle different kinds of information and uncertainties, within a real-time environment. The last two decades a variety of representations and calculi for data fusion have been developed. Methods for handling uncertainty range from the classical Bayesian probability model [PEARL87], Shafer-Dempster evidence theory [SHAFER76], Zadeh's fuzzy set theory [YAGER77] to the linear and logarithmic opinion pool [GENEST86].

All methods have their advantages and shortcomings, but so far no single algorithm has been developed, that deals with all problems [DILLARD92].
In order to fuse different kinds of information with their uncertainties and incompleteness, a multilevel structure for the fusion and integration of data, information, and knowledge is required. Each level has its own degree of abstraction and type of uncertainty. The integration or fusion technique at each level depends on the nature of these characteristics.

In order to distinguish the levels of abstraction some terminology is introduced. Firstly the sensor observations. These are the detections of target information (e.g., the interception of an enemy radar signal). Secondly the target feature. A feature is a quantifiable signal characteristic, that can be extracted from observations and that permits the discrimination among target classes (e.g., the feature “frequency” and “pulse width (PW)” from the radar signal). Thirdly, the signature, that is described by one or more specific features (e.g., the electromagnetic (EM) “signature”, described by the “features” frequency and PW).

The following variables are introduced:

- the possible target identities (type and class) form \( k \) hypotheses \( \{h_1, h_2, \ldots, h_k\} \), \( p \) hypotheses for target classes \( \{C_1, C_2, \ldots, C_p\} \) and \( q \) hypotheses for target types \( \{T_1, T_2, \ldots, T_q\} \), with \( k = p + q \);
- the hypotheses are characterised by \( m \) features \( \{f_1, f_2, \ldots, f_m\} \), not all features apply for each hypothesis (e.g., passive homing missiles do not have radar related features);
- the knowledge about possible values of \( f_i \) for class \( C_j \) can be represented by a fuzzy set \( A_{ij} \);
- a set of features \( f \) form a signature \( \{\text{sign}_1, \text{sign}_2, \ldots, \text{sign}_n\} \), with \( n < m \);
- from the sensor observations \( s_1, s_2, \ldots, s_r \) the feature vector \( (f'_1, f'_2, \ldots, f'_m) \) is determined;

The aim of this study is to come to a general architecture for an identification system that can handle all kinds of uncertainties. Therefore the approach is based on central decision making and passing uncertainties to higher levels of decision.

To continue with the structure of the identification process: figure 4 illustrates the four levels. First a general description of the four level process will be given. Then each process is described in more detail. Special attention is given to level 2.

![Figure 4: The four-level identification process](image-url)
5.1 Level 1: Feature determination

The first process in the identification algorithm maps observations $S_i$ on the features $f_j$ and is called "feature determination". The sensor observations may come from different sensors. Sensors in maritime air defense are radar (search and track), electronic support measures (ESM), infrared sensors, electro-optical sensors. The features are quantifiable signal characteristics, which permit the discrimination of a target class or type. The mapping is not always a one-to-one relation.

There are many different ways to obtain target features from sensor observations. The appropriate technique depends on the feature and type of sensor. In the literature the following methods are mentioned:

- **State estimation** is the process of determining an estimate of the state of a target, based on observations related to its state. State estimation techniques are especially used for estimating the target's position, velocity, accelerations and future track. A large and diverse set of estimating techniques have been developed, resulting in several overview and survey articles [TAYLOR88]. Examples of estimating techniques are $\alpha$-$\beta$- and $\alpha$-$\beta$-$\gamma$-filters, Kalman filters, fuzzy target trackers [SICKING93].
- In ESM-sensors various kinds of waveform classifiers are used to obtain features like frequency, pulse-width and pulse repetition frequency from an intercepted radar signal.
- **Pattern recognition techniques** are used for imaging sensors such as infrared and electro-optical sensors.

5.2 Level 2: Signature similarity

The "signature similarity" is a mapping from a set of features $f_j$ to a value between 0 and 1. This value is a degree of belief that the signature originates from a certain target class or type (hypothesis).

In general, the signature similarity $S$ between a set of features $F^*$ and their possible values $A$ for the hypotheses $H$ can then be defined as follows:

$$S(F^*, A; H) \in [0, 1]$$

In maritime air defense several signatures can be distinguished. Each signature has its own characteristics. The fusion technique should depend on the characteristics of the signature's data and knowledge.

In practice however, trajectory information is important for air defense officers, when identifying targets. Features like target speed, height and manoeuvres are easily recognised on the tactical displays and intuitively interpreted by the members of the air defense team.

In order to incorporate this valuable information in the target identification process, a method to determine the signature similarity for this uncertain trajectory information is proposed. The fact that humans perform relatively well (as long as there is enough time) in this uncertain environment brings the fuzzy methods into scope. The use of fuzzy logic in knowledge base systems is usually referred to as approximate reasoning. The identification of trajectory signatures is therefore based on the principles of approximate reasoning.

In the next part of this chapter the implementation procedure will be discussed. First, the structure for representing the uncertain knowledge about signature features is presented. Second, the fuzzy implication in this case is discussed. Third, two methods of inferring the observed feature values $f_*$ with the knowledge are discussed. Both methods result in a fuzzy similarity measure for each hypothesis. Finally the uncertainty and lack of information are discussed.

1. Representation of knowledge

The knowledge about characteristic values or features $f_1, f_2, ..., f_n$ for target class $C_i$ can be represented in the following rule $R_1$:

$$\text{If } f_1 \text{ is } A_{1_1} \text{ and } ... \text{ and } f_n \text{ is } A_{1_n} \text{ then ident is } C_j$$

Similar rules are formulated for each target class and type.

In these rules the premise (or antecedent) is a combination of fuzzy propositions, the consequence is a singleton, representing the target class or type. When the fuzzy sets $A_{ij}$ are identified by the membership functions $\mu_{A_{ij}}(f_i)$ then the following n-dimensional fuzzy relation $R_1$ representing the combined fuzzy features of class $j$, can be constructed:

$$R_1 = I(T(A_{1_1}, A_{2_1}, ..., A_{n_1}), C_j)$$

In this relation, $T$ is a conjunction based on a general t-norm and $I$ (will be discussed later) is a fuzzy implication. $T$ represents the and connective. $I$ represents the if-then connective. Many t-norms have been developed. A discussion about the best choice in this case is beyond the scope of this study. The "minimum operator" (Zadeh's t-norm) will be used because of its simplicity and widespread use.
The multi-dimensional fuzzy relations can represent a correlation between elements of a product space. In this case this product space can be called a “fuzzy feature space”. Figure 5 shows the features for two target classes in a two dimensional fuzzy feature space.

In this way a n-dimensional fuzzy feature space, with membership functions for all features and each hypothesis can be established. This fuzzy feature space is the knowledge representation of the fuzzy trajectory signature information.

2. The fuzzy implication

The implication is a mapping from the antecedent and consequence space to \([0,1]\) (\(f_1 \times f_2 \ldots f_m \times f_n\)). In the literature many implication functions have been proposed. Two basic types of implications can be distinguished [JAGER95]:

- fuzzy implications complying with the classical implication; these I-operators are a generalisation of the classical implication (S-implications, e.g. Kleene-Dienes implication \(I(a,b)=\max(1-a,b)\));
- fuzzy implications complying with the classical conjunction (T-norms) for instance the Mamdani implication \(I(a,b)=\min(a,b)\) and Larsen \(I(a,b)=ab\).

Based on this distinction a number of combinations can be defined.

In this case an appropriate implication is type b, Mamdani-implication. This implication is easy to implement and very popular in fuzzy control.

3. Inference of the rule-base

The question is now, how to infer the class from the observed feature values \(f_1\) with the characteristic values \(A_j\) of the hypotheses. The answer is a two step process. Step 1 is the establishment of a fuzzy subspace. Step 2 is the actual inference which results in the similarities for all classes \(C_j\).

**Step 1.** Establish a subspace that contains only the features that have been observed, extracted or (intuitively) determined. This determines the structure of the antecedent space. It is important to realize that the discrimination between certain target types and classes in a subspace might be impossible. In that case the similarities for these classes are equal and the discrimination has to be done at another level. Figure 6 illustrates in a two dimensional feature space that the difference between \(C_1\) and \(C_3\) can not be distinguished, when \(f_2\) is unknown.

**Step 2.** The inference of determining the class from the observed feature values with rules can be done in various ways. As the consequences are crisp sets, the inference is reduced to the determination of the similarity between observed features and the premises. In the literature different methods are described, from the generalisation of the reasoning schemes from classical logic (generalised modus ponens), distance functions and similarity measures to various fuzzy truth values. In this application a method that fits the representation of knowledge and features is required.

In order to cope with different situations two methods for inferring the rule-base are proposed. The first method (a.) is based on distance- and similarity measures, the second method (b.) is based on t-norm implications. First of all both methods are introduced, then some considerations as to using method a. or b. are given.

The properties of both methods are discussed below.

- **Inference based on distance- and similarity measures.** In this method the similarity between features and knowledge is obtained by some kind of similarity measure or distance function. The feature values are represented by a multi-dimensional fuzzy set \(F^*\), the possible values for class \(j\) by the multi-dimensional fuzzy set \(A_j\). The similarity for class \(C_j\) is then given by the following expression:
  \[ S_{C_j} = S(F^*, A_j) \]
For the calculation of the similarity or distance between $F^*$ and $A_j$, many methods have been reported in literature. When using a distance function $d(F^*, A_j)$ the similarity equals \[ S(F^*, A_j) = \frac{1}{1 + \frac{d(F^*, A_j)}{1 - d(F^*, A_j)}}, \quad S \in (0,1) \]

Many methods for measuring the similarity (or distance) between fuzzy sets have been reported in the literature. In [SETNES95] an overview of these measures and their properties is given. Most of the similarity measures that are proposed are symmetric similarities. In this application however $F^*$ and $A_j$ are not symmetric (illustrated in IV). In [TVERSKY77] it has already been mentioned that similarity measures may not be a symmetric relation. The requirements for a similarity measure in this application are:

1. If $F^* \subseteq A_j$ then $S_{C_i}(F^*, A_j) = 1$.
2. If $F^* \cap A_j = \emptyset$ then $S_{C_i}(F^*, A_j) = 0$.
3. If $F^* \subseteq A_j$ then $S_{C_i}(F^*, A_j) \in (0,1]$.
4. If $F^* \supseteq A_j$ then $S_{C_i}(F^*, A_j) \leq 1$.

Based on the similarity measures proposed in [DUBOIS80] an a-symmetric measure for this application is developed. The measure is based on fuzzy intersection and fuzzy cardinality and meets the above mentioned requirements:

\[ S_{C_i}(F^*, A_j) = \frac{|F^* \cap A_j|}{|F^*|}, \quad \text{with} \quad |A_j| = \sum_{x \in \Sigma} \mu_{A_j}(x). \]

The fuzzy intersection $F^* \cap A_j$ can be determined using a minimum-operator and $|A_j|$ is called the cardinality, also known as the power of a fuzzy set. This similarity measure can be interpreted as the "relative fit of $F^*$ in $A_j$". It meets the intuitional judgement of a set of data against background knowledge about the possible values of the features.

b. Inference based on a T-implication. The inference of a fuzzy rule by a T-implication was proposed in [DUBOIS84] and is now extensively used in fuzzy control. The similarity can be obtained by a sup-T composition. It can be shown [JAGER95] that for the minimum and the product operator the similarity between $F^*$ and $A_j$ is given by the following expressions:

- Minimum operator:
  \[ S_{C_i}(F^*, A_j) = \text{hgt}(F^* \cap A_j) \]

- Product operator:
  \[ S_{C_i}(F^*, A_j) = \text{hgt}(F^* A_j) \]

This similarity measure gives a "degree of fulfilment" between observed features and premises.

4. Discussion about uncertainty in knowledge and feature values

In the choice between methods a. and b. the uncertainty in the feature values is crucial. Three different situations are distinguished (figure 7):

1. When no uncertainty is taken into account features can be represented by crisp values. Method b. seems to be an appropriate similarity measure. The similarity for one feature $i$ can be obtained by evaluating the membership value: $\alpha_{i,j} = \mu_{A_j}(t^*_i)$. The similarity for $F^*$ can be determined by using some kind of T-norm. For numerical feature values of $t^*_i$ the following applies:

\[ t^*_i = 1 \quad \text{and} \quad |t^*_i \cap A_j| = \text{hgt}(t^*_i \cap A_{i,j}) \]

In this case method a. and b. are similar.

Other distance measures can be developed to measure the distance between fuzzy sets and numerical values, but that is beyond the scope of this research.

2. When uncertainty originating from measurement uncertainty is taken into account, feature values can be represented by fuzzy numbers. The shape of the membership function that represents the fuzzy number can depend on the performance of sensors, environmental conditions, etc. and may vary with time. Both methods a. and b. can be used in this case. Method a. is preferred, because R-1 to R-4 are met.

3. When the feature values are more uncertain, qualified by linguistic terms and represented by fuzzy sets, they can be represented by fuzzy numbers. Method a. is now preferred, because R-1 to R-4 are met (situation 3a in fig. 7).
A specific situation arises when the uncertainty in the features is more than the fuzziness in the antecedent. Features are only roughly known. This is another kind of uncertainty. The more uncertain about the feature values relatively to the required accuracy, the smaller the similarity. This kind of similarity is, in fact, a measure of nonspecificity (situation 3b in fig. 7).

From the three situations of uncertainty in features can be concluded that method a. always applies. Methods a. and b. are equal in the case of crisp feature values. A major (practical) disadvantage of method a. is the considerably higher processing requirement.

In the next table the validity of method a. and b. is summarised.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Method A</th>
<th>Method B</th>
</tr>
</thead>
<tbody>
<tr>
<td>crisp</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>fuzzy number</td>
<td></td>
<td>+/-</td>
</tr>
<tr>
<td>fuzzy set</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>

The negative results for method b. is influenced by the choice for a Mamdani-implication. Further research is needed into the validation of method b. for other implication functions.

5.3 Level 3: Fusion of similarities

The next level in the identification process is the combination of the signature similarities for each signature and hypothesis in order to determine the combined belief in a hypothesis. This process can be seen as a multi attribute decision making problem (MADM).

There are many algorithms to fuse data from different sensors, for example, two popular algorithms are Dempster-Shafer rules of combination [BOGLER87] and the Bayesian method. In [WALTZ90] an overview is given.

In this chapter two methods that are used in connection with MADM and fuzzy logic, the averaging and distance functions will be discussed. Subsequently some considerations about weight- and confidence factors are given.

a. Average value. The average value of the signature similarities for a certain alternative. Many different averaging operators are used in the decision theory. Most of them are a special form of the generalised function:

\[ D_2(s) = \left( \frac{1}{m} \sum_{j=1}^{m} \mu_{ij} \right)^{1/s} \]

For \( s=\infty \) the general goal function becomes the "maximum" operator and for \( s=\infty \) the "minimum" operator. For \( s=1 \) the function is the arithmetic mean and for \( s=0 \) the geometric mean.

In many articles the fuzzy approach in decision making is associated with the "max-min" criterion (\( s=\infty \)). However, the right value of \( s \) depends on the decision problem and the person who takes the decision. Further study is needed to examine the value of \( s \) that represents the decisions of air defence officers.

b. Distance function. A target feature vector is a vector with a value between 0 and 1 for each feature. The best solution is represented by a reference alternative "X". The "distance" between the reference alternative and the actual values is a measurement of the similarity. The target closest to this reference is the best alternative. A general distance function to calculate the distance of alternative C and X is:

\[ D(C, X) = \left\{ \sum_{i=1}^{m} (\mu_{ij} - x_i)^{1/x} \right\}^{1/x}, \quad x \geq 1 \]

c. Weight factors. To account for the relative importance of certain features weight factors \( w_i \) between 0 and 1 can be added. These factors do not represent the confidence in a feature, but give the relative weight of a certain (group of) signature in the fused similarity. A discriminating signature could result in a high weight factor. The weight factors should comply with the following condition:

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

d. Confidence factor. To account for the confidence in features (and sensor observations) an extra confidence factor (CF) can be introduced. This factor, from 0 (no confidence) to 1 (high confidence) is accounted for, prior to the fusion of similarities. The confidence factor depends on the performance of the sensors and the likeliness of spoofing, jamming and deception by the enemy.

The operator that combines the CF with the feature similarity can be a T-norm or some other operator.

5.4 Level 4: Evaluating the result

The process of calculating the support for each hypothesis is repeated for every sample. This is one of the inputs of level 4. Another input is information about the tactical situation. The output of this process is an evaluated fuzzy identity measure for the most likely hypotheses.
In the literature several studies on this high level of decision support can be reported. In [HEUSDEN95] an overview is given of the possible use of artificial intelligence in Command and Control. In [BOASSON95] some requirements for new technologies are mentioned. In practice this level of decision making is not yet supported. Based on the practical situation some factors can be mentioned that play a role in the evaluation of ATR data:

a. *The dynamic behaviour* of the "identity measures" are a direct consequence of the real-time aspect of the identification process. Some reasons that cause this dynamic behaviour are:

- New sensor observations, for instance new EM-emissions;
- Significant changes in trend data, for instance change of target's height or detection of manoeuvres;
- Fuzzy data becomes clearer, for instance radar image recognition at a closer range;
- Change of weight factors, due to variation in sensor performance;
- Noise.

Figure 3 showed how dynamic behaviour influences the decision process. This can be taken into account by considering the change in belief values. The easiest way of doing this is to display the fuzzy identity measures versus time (as in figure 3). Automatic methods for taking this into account can range from approximate reasoning to various filtering techniques.

b. *The situational information (tactical environment)* gives important information about the possibility that a hypothesis may occur. Some factors that play a role in the tactical situation are:

- Knowledge about the presence of a weapon carrier (position [relative], capabilities and intentions);
- Enemy behaviour (surveillance activities, jamming, spoofing etc.);
- Knowledge about the environment, the proximity of land, airfields, etc.

This information plays an important role in the human interpretation of threat information. Based on this information some hypotheses can be excluded, while others have a high possibility of taking place. Several systems for taking this information into account have been proposed. In [WALTZ90] an overview is given. An early study into the possibilities of fuzzy logic for these kinds of decision making is done in [DOCKERY77].

Approximate reasoning techniques and fuzzy expert systems might provide sound basis for this kind of decision support.

c. *Equal support for different alternatives* is another factor that has to be taken into account when evaluating the results of level 3. This unspecificity between several hypotheses is caused by similar target types, uncertainty in information and lack of special information. The level 4 evaluation process should be able to deal with this kind of uncertainty.

Some considerations concerning this problem and possible approaches:

- If several target types are indistinguishable, a distinction between target classes might be possible;
- A lack of information might result in information requests to certain sensors (e.g. special attention is given to determine a distinguishing feature value);
- If the unspecified target types result in similar defence actions, no further identification is required;
- Eliminate the impossible hypotheses and concentrate on the possible;
- Take the situational information into account (pt. b.) and modify the belief values.

6. Simulations

A simulation model of the target identifier has been build using MATLAB with the "Fuzzy Toolbox". Inputs for the model are:

- Knowledge about the threat: height/distance profile, speed- and EM-information about three target types. The height pattern is defined for three fuzzy distances: far, medium and close.
- The target parameters of the missile that is to be identified.
- Noise can be added.
- Selection of identification method a. or b.

After defining the model, a missile run has to be made in order to generate a set of 'observations'.

Outputs are:

- the fuzzy feature space and the missile’s flight pattern;
• the fuzzy similarity versus time.

Figure 8 shows an example of similarity versus time.

![Graph showing fuzzy similarity over time]

From this example the following conclusions can be drawn:

- the EM (radar) signature is supporting target 2 and decreasing the similarity of target 1 and 3;
- based on the height/distance and speed information target 1 is the most likely target.

More simulations are needed in order to investigate the possibilities and limitations of the target identification system.

7. Conclusions and recommendations

In this article the problem of target identification in maritime air defence has been addressed. A general structure to deal with its specific problems has been proposed. The four-level system with the particularities of each level are described. Special attention is given to the incorporation of uncertain target trajectory data. A similarity measure for this type of information is suggested.

7.1 Conclusions

From this study the following conclusions can be drawn.

- The proposed four level identification system seems to be a flexible structure that offers the possibility to account for different levels of abstraction;

- The proposed representation of uncertain knowledge in a fuzzy feature space offers the possibility to account for uncertainty in knowledge. It has been shown that the proposed similarity measure (method a.) is a flexible and promising measure that handles both crisp and fuzzy data. Even data that is only roughly known can be incorporated in the identification process.

- At the highest level, important factors of the identification system are the situational information and dynamic behaviour. As this information is in practice of major importance for air defence officers a decision support system should be able to deal with this kind of information as well.

7.2 Recommendations

Suggestions for further study.

- Further research is needed into methods a. and b. for representing uncertainty and integrating different kinds of data. In this study special attention should be given to: how to handle “lack of information” and the use of specific fuzzy operators and membership functions. Especially the use of other fuzzy implications in combination with method b. needs further research.

- For level 2 (signature similarity) the determination of the similarity for other signatures needs further research. Method a. might be useful for other kinds of data.

- The use of fuzzy numbers to express the measurement uncertainty needs further research. The adjustment of membership functions, depending on the uncertainty at a certain moment, might be an interesting possibility to account for this kind of uncertainty.

- For level 4 further study is needed into the use of situational knowledge and dynamic behaviour. Decision making at this level is characterised by a high level of abstraction.

- An overall simulation model has to be built to test the proposed structure with (classified) data of real targets.

- The similarity measures in this study only account for “positive evidence”. It might be interesting to study the use of certainty factors defined on the interval from -1 to 1 [SHORTLIFFE92].
- The fuzzy similarity as output of the identification system is a number between 0 and 1. Further research can be done into a fuzzy (linguistic) output (e.g., "similarity class i = "very similar"). This kind of output is probably easier to comprehend for the user.

Literature


Evaluation of a Methodology for Human-Machine Interface Development

Markus Tiemann
Laboratory for Systems Engineering and Human-Machine Systems
University of Kassel

Introduction

Experiences of software development companies show that about 20% to 50% of the overall software development costs are to be spent for the development of the human-machine interface. Therefore, the process of user interface development becomes more and more an important economic factor for software developers.

The DIADEM methodology for human-machine interface development is intended to give software developers a support for this part of their development activities. Therefore, the overall objectives of the DIADEM method are

• to enhance the efficiency in developing human-machine interfaces and
• to improve the quality of the developed interfaces.

These objectives shall be met by providing the software development teams with a method to structurize, organize, and supervise their activities.

Goal of the European DIAMANTA Trial Applications Project is to evaluate the DIADEM methodology with regard to their functionality and usability.

This publication will start with a short presentation of the DIAMANTA project. Then an overview of current evaluation techniques will be given, followed by a description of the project-specific objectives and constraints. Thereafter, the suitability of the presented evaluation techniques against the background of these objectives and constraints will be assessed. The paper will conclude with a presentation of the evaluation procedure elaborated for the DIAMANTA project.

The Trial Application Project DIAMANTA

The Diamanta project is a Trial Application Project and belongs to the domain Software Technology of the ESPRIT program. It’s goal is to validate the suitability of the DIADEM Methodology for Human-Machine Interface Development.

DIADEM (Dialogue Architecture and Design Method) is a methodology developed by Thomson-CSF/RCC, France, for supporting the development of human-machine interfaces. It aims at providing a methodological approach to the software development of the interfaces with all their prototypes.

The theoretical works DIADEM is based on are besides others the Seeheim model, PAC, MAD, and GOMS. The model used in this project covers the presentation part of interactive software as well as the control part and the entry points of the application part, called application interface.

More details of the DIADEM methodology and particularly the formalisms provided can be found in [1].

For the evaluation of the DIADEM methodology it will be applied to three trial applications:
• a Tourism Multimedia System for Travel Agencies,
• a Supervisory Control of Chemical Processes,
• and a Sales Support System for Used Cars.

The processes of the trial application developments become observed, and the results will be evaluated. The peculiarity of this evaluation procedure is that it has to deal with both, the Method User Evaluation and the End User Evaluation.

Method User Evaluation means to evaluate the DIADEM methodology from the development teams' point of view, mainly by observing the process of application development. Against that, end user evaluation means to evaluate the interface developed by using DIADEM. This will chiefly be done by performing usability testings with real end users.

For both, method user evaluation and end user evaluation, objective as well as subjective data will be taken. Examples for objective and subjective measures for the method user evaluation are presented in Table 1:

Table 1: Objective and subjective measures for method user evaluation.

<table>
<thead>
<tr>
<th>Objective Measures</th>
<th>Subjective Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>time spent for certain development phases</td>
<td>user commitment during development</td>
</tr>
<tr>
<td>number of iterations</td>
<td>developers' satisfaction</td>
</tr>
<tr>
<td>number of modifications</td>
<td></td>
</tr>
<tr>
<td>lines of code</td>
<td></td>
</tr>
<tr>
<td>number of software bugs</td>
<td></td>
</tr>
<tr>
<td>complexity of formalisms</td>
<td></td>
</tr>
</tbody>
</table>

Examples for objective and subjective measures for end user evaluation, i.e. the evaluation of the developed interface, can be found in Table 2:

Table 2: Objective and subjective measures for end user evaluation.

<table>
<thead>
<tr>
<th>Objective Measures</th>
<th>Subjective Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>time measures</td>
<td>user satisfaction</td>
</tr>
<tr>
<td>number of tasks</td>
<td>user's mental workload</td>
</tr>
<tr>
<td>number of interactions</td>
<td></td>
</tr>
<tr>
<td>number of screens/dialogue boxes</td>
<td></td>
</tr>
<tr>
<td>number of button clicks</td>
<td></td>
</tr>
<tr>
<td>length of mouse moves</td>
<td></td>
</tr>
<tr>
<td>number of user errors</td>
<td></td>
</tr>
<tr>
<td>number of complete/incomplete tasks per time</td>
<td></td>
</tr>
</tbody>
</table>

The techniques suited to take these objective and subjective measures for the method user evaluation as well as for the end user evaluation will be described in the following chapter.
Overview of Current Evaluation Techniques

In general, one makes a distinction between heuristic evaluations and usability testing methods.

Table 3: Types of evaluation techniques.

<table>
<thead>
<tr>
<th>Heuristic Evaluations</th>
<th>Usability Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods depending on the Heuristics</td>
<td>Performance Measurements</td>
</tr>
<tr>
<td></td>
<td>Thinking-Aloud Method</td>
</tr>
<tr>
<td></td>
<td>Observations</td>
</tr>
<tr>
<td></td>
<td>Interviews and Questionnaires</td>
</tr>
<tr>
<td></td>
<td>Focus Groups</td>
</tr>
<tr>
<td></td>
<td>Logging Actual Use</td>
</tr>
<tr>
<td></td>
<td>User Feedback</td>
</tr>
</tbody>
</table>

To perform Heuristic Evaluations means to examine an interface or a method against specified usability principles. These principles are called the usability heuristics.

Against that, Usability Testing means to perform test sessions with real end users. There are several kinds of usability testing methods:

- *Performance Measurements* means to collect data (for example time measures or error ratings), while test users perform a given task or set of tasks.

- In *Thinking-Aloud* sessions test users speak out continuously what they are thinking while they are performing predefined tasks with the system to be tested.

- Another technique is to observe users while they are performing real tasks under true field conditions.

- In *Interviews and Questionnaires*, users are asked certain questions, and their answers are recorded.

- *Focus Group* sessions are discussions of new concepts and identification of issues while a moderator is responsible for maintaining the focus of the group on the issues of interest.

- *Logging Actual Use* means that data are continuously and automatically collected while a real user is performing real tasks.

- And finally, the *feedback* a Distributor of a system receives from the system users, for example by the way of the customer support, can be used to collect statistics about frequencies of user problems.

All these techniques have certain advantages and disadvantages. Their suitability depends on the specific demands on the evaluation procedure. The objectives and constraints of the DIAMANT A project will be presented in the following chapter.
Project Specific Objectives and Constraints

DIADEM is a methodology developed for professional human-machine interface development. It is intended for use in commercial development teams. Therefore, the analysis of the costs and benefits of the use of DIADEM is one important point.

Another point is that the three trial applications are built by three different development teams at three different sites. The effects the specific staff compositions have on the use and the suitability of the DIADEM methodology shall be elaborated.

Furthermore, the evaluation procedure shall be designed in a way that makes a comparison across all the three trial application development processes possible.

Over and above that, DIADEM shall become evaluated against theoretical and formal criteria, independent from the three applications.

A constraints of the DIAMANTA project is that people from different countries, speaking different languages, are involved. As will be shown later, especially the aspect of the different languages has a certain impact on the suitability of some evaluation techniques.

Another important constraint is the availability of data. The three trial applications are built for commercial purposes, and their development is not funded by the project. Therefore, the software developers keep certain information as company internal items.

A summarized presentation of the objectives and constraints of the DIAMANTA trial application project is given in the following table:

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs and Benefits Analysis</td>
<td>Multi-National Testing</td>
</tr>
<tr>
<td>Effects of Staff Composition</td>
<td>Availability of Data</td>
</tr>
<tr>
<td>Cross-Application Comparison</td>
<td></td>
</tr>
<tr>
<td>Application Independent Evaluation</td>
<td></td>
</tr>
</tbody>
</table>

Assessments of the Suitability of Evaluation Techniques

Following, the effects of the project-specific objectives and constraints on the suitability of the specific evaluation techniques will be assessed. Of course, these assessments are not objective. They are based on hints found in literature [2-9] and on experiences of the group of the Laboratory of Systems Engineering and Human-Machine Systems at the University of Kassel. By regarding a wide range of literature and experiences it was tried to reduce subjectivity to a minimum. Nevertheless, it shall be remarked that other evaluators with different experiences, who see the problem against another background, may come to different conclusions.

Multi-National Testing

Since partners from three countries with three different native languages are involved in the project, this must be considered for all evaluation activities that demand for verbal or written communication between the evaluator and the test user. Whereas the method evaluations with the three development teams as test users will not be affected by this circumstance, the
interface evaluation involving representative end users may require to communicate in the language of the respective country.

Against this background all techniques that demand a verbal communication are less suited. This concerns the thinking-aloud technique as well as observations and focus group sessions. Also interviews are difficult, whereas questionnaires can better be prepared in advance.

Best suited techniques in this context are those that deliver directly data that can be evaluated statistically. Those techniques are performance measurements and logging actual use. Also the feedback a system provider receives from his customers can be evaluated against the frequency of the occurrence of user problems.

Costs and Benefits Analysis

Since the DIADEM methodology for interface development is designed to be used in commercial application development processes, an important criterion for the evaluation of the methodology is the ratio between costs and benefits. Therefore, it must be analyzed what the costs and what the benefits are when using DIADEM. On the base of these analyses, it can be judged whether the benefits, when using DIADEM, justify the costs.

Direct hints on the costs and benefits of using the DIADEM methodology can be got by performance measurements and extensive data logging during the development process. Also interviews and questionnaires promise good results, since they provide the assessments of the developers against the background of their experiences.

Against that, no useful results with regard to the costs and benefits can be expected from using the thinking aloud method and focus group sessions, because these techniques focus on receiving spontaneous reactions. Observations can only be performed for the interface evaluation with test users, to observe the whole process of application development is not possible.

Staff Composition

When evaluating the development processes of the three trial applications, it must be considered that the compositions of the three development teams may be quite different. Especially in smaller companies the missions of the employees will often change. For example, there may be no explicit human factors specialist, but a person who is very familiar with human factors aspects and who will be consulted every time a usability problem occurs.

An interesting aspect for the evaluation of the DIADEM methodology is to assess how far the staff composition affects the efficiency or even the applicability of the methodology.

To investigate this point only method user evaluations can be performed, since end user evaluations can not provide any data that show relations to the development staff composition. Suited means are performance measures, interviews and questionnaires, and log data. User feedback and observations may bring up some additional information. However, observations can hardly be performed during the whole development phases. From heuristic evaluations only few, and from thinking aloud testings no results can be expected.

Availability of Data

Since the collection of the data needed to perform the evaluations must be conducted mainly within the development teams, the effort for this must be kept to a minimum as not to disturb the development processes. Further, it must be considered that certain information related to the application development processes that are not funded by the project may be company confidential and will not be provided for performing external evaluations.
Heuristic evaluations will be possible even if only few data would be available. Also performing interviews and/or questionnaires will be possible, since the questions can be selected and formulated in a way that does not impact company confidentialities. Against that, all techniques that can deliver unpredictable information are not suited to keep certain data confidential.

Cross-Application Comparisons

The three trial applications to be developed with DIADEM are of significant different types. Especially the phone interface developed by ISA will put other requirements on the development methodology than the graphical interfaces will. Therefore, it is an interesting point to compare the evaluation results by examining the three development processes and, by this, to come to a conclusion for what application developments the DIADEM method is more suited and for what application developments there are certain difficulties.

All evaluation techniques that deliver statistical data make a comparison across the three application development processes easy. These are heuristical evaluations as well as performance measurements, interviews and questionnaires, and evaluating user feedback and data logs. The other evaluation techniques provide spontaneous reactions and comments that depend to a high degree on personal and environmental circumstances and, therefore, can hardly be compared with those from other people under different circumstances.

Application Independent Evaluation

The conclusions from the comparison of the three application development process evaluations should be supplemented by an examination of the DIADEM methodology against a theoretical background. Focus of these investigations shall be, besides others, the consistency, the correctness and the efficiency of the formalisms provided by DIADEM.

The only way to receive results that are usable with regard to this aspect are heuristic evaluations and information collected from interviews or questionnaires. All other techniques only deliver data that depend to a high degree on the evaluation.

Summarized Effects of the Constraints on the Available Techniques

The following tables summarize the suitability of the specific evaluation techniques against the background of the project specific objectives and constraints. The ratings used within all tables in this chapter have the following meaning:

The use of the method against the background of the specific objective/constraint is
-- not possible
- not recommendable
0 possible
+ recommendable
++ highly recommendable

The tables show that the techniques with no significant deficiencies for the method user evaluation as well as for the end user evaluation are conducting interviews and questionnaires. Another suitable means to assess both the efficiency of the development method and the usability of the interfaces designed with it are performance measurements.

However, all techniques have certain advantages against specific criteria. Only the thinking-aloud technique shows deficiencies against all objectives and constraints of the DIAMANTA project, because it is a technique to be applied under laboratory conditions and therefore not suited for this evaluation procedure.

2.3-6
Table 5: Summarized effects of the project specific objectives and constraints on the evaluation methods with regard to the method user evaluation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic Evaluation</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Performance Meas.</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>--</td>
</tr>
<tr>
<td>Thinking Aloud</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Observation</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Interv./Questionnaire</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Focus Groups</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>--</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Logging Actual Use</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>++</td>
<td>--</td>
</tr>
<tr>
<td>User Feedback</td>
<td>++</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 6: Summarized effects of the project specific objectives and constraints on the evaluation methods with regard to the end user evaluation.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristic Evaluation</td>
<td>0</td>
<td>+</td>
<td></td>
<td>+</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>Performance Meas.</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Thinking Aloud</td>
<td>-</td>
<td>-</td>
<td></td>
<td>--</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Observation</td>
<td>-</td>
<td>++</td>
<td></td>
<td>--</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Interv./Questionnaire</td>
<td>+</td>
<td>++</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Focus Groups</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Logging Actual Use</td>
<td>+</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>User Feedback</td>
<td>++</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

Against this background, the evaluation of the three trial application development processes will mainly be based on questionnaires and heuristic evaluations. For the end user evaluation, besides heuristic evaluations chiefly performance measures will be taken and the subjective assessments of the test users will be recorded.

**Evaluation Procedure Elaborated for DIAMANTA**

The overall evaluation of the DIADEM methodology, covering (a) the evaluation of the development processes when using DIADEM as well as (b) the evaluation of the developed prototype or final interfaces, can be divided into three different types of evaluation steps: (1) continual evaluations of the development process, (2) periodical evaluations of the development process and prototypes, and (3) the final evaluation of the development process and the developed interface. These steps are shown in Figure 1.

The arrows directing from the evaluation steps back to the development process symbolize the feedback from the evaluator to the application developer sites. Certain feedback from the
The evaluator may occur after all continual and periodical evaluation steps in order to improve the measurements and data acquisition at the application developer sites, e.g., for getting more reliable comparative evaluations during the further steps.

Figure 1: Evaluation Steps in Project Life-Cycle.

All forms and questionnaires necessary for performing the evaluation steps are delivered by the evaluator in order to achieve comparable results across all three application development processes. As far as application-dependent features are concerned, they will be developed in cooperation with the application developers.

**Continual Evaluations**

The evaluations performed continually along the whole development process are performed in order to study the progress of the development activities as well as changes in objective and subjective measures. They will show the efficiency of the DIADEM methodology in dependence on the development phase and the growing experience of the members of the development team with the DIADEM formalisms.

Therefore, administrative data such as effort for performing certain activities as well as experiences and subjective assessments of the people involved in the development process are recorded and evaluated.

At each application developer site, all people involved in the development process record the data needed for the evaluation internally. One person responsible for DIAMANTA evaluation activities concludes all internal logs in the "Continual Evaluation Log" and sends it to the evaluator via electronic mail weekly.

Figure 2 shows the procedure for the continual data acquisition with the different people of the development team.
Periodical Evaluations

In addition to the continual evaluations, every time a development phase is closed or a prototype interface is finished, periodical evaluations shall be performed. These evaluations shall concern both the experience with the DIADEM methodology up to that specific development step and the developed prototype.

The prototype evaluation shall mainly deal with the investigation whether the prototype meets the requirements of the application and the usability requirements. Furthermore, the subjective satisfaction of the customer (or even end user) with the prototype and the usability of the DIADEM formalisms for communications with the customer shall be studied.

The evaluation of the development process in the frame of the periodical evaluations shall assess the advantages (and disadvantages) of using DIADEM in the specific software development phase.

The steps to be performed for the periodical evaluations are shown in Figure 3.
**Final Evaluations**

In the same way as the periodical evaluations, the final evaluations deal with both the development process and the developed interface.

The final evaluation is more comprehensive than the periodical ones, since it must involve real end users in usability testings of the developed interface. Without the end user's involvement, a decent evaluation of the interface can not be conducted. It should be regarded that end users means those users who will actually use the system, i.e., the employees of a travel agency, the clients of a used car dealer, or the operators in a chemical plant.

Ideally, it seems to be reasonable to perform the final usability testings with end users on-site by the evaluator together with the ergonomists of the application developer sites. This will allow to provide the demanded objectivity and, hence, reliability of the evaluation results. However, such participation of the evaluator in the final evaluations may be beyond the possibilities of the budget and the available time and labour-force constraints within this Trial Application.

Therefore, at least extensive test plans for the usability test sessions will be provided by the evaluator in consense with the application developers. These test plans shall ensure objective and comparable test results.

The procedure for the final evaluation is depicted in Figure 4.

![Figure 4: The final evaluation procedure.](image)

**Synthesis of Evaluations**

At the evaluator site, all data acquired within the three evaluation steps are collected and checked against their correctness and consistency. Then they become prepared for further data processing and finally evaluated.

The results of this evaluation will be synthesized for each application development process on its own, and, in a second step, across all three applications.
Conclusions

Concluding shall be noted that all described evaluation activities represent a compromise between the scientific demands on a comprehensive investigation, on the one hand, and the specific constraints of the DIAMANTA project, especially the reasonable demands of the application developers to keep certain data company confidential, on the other hand. The described procedures are promising the most comprehensive and objective evaluation of the DIADEM methodology against the background of all project-specific constraints.

Acknowledgements

In conclusion, I would like to thank Dipl.-Psych. Jens Heuer for his advice and invaluable support during the elaboration of the presented evaluation procedure.

References


Towards a global approach to solve social aspects of usability and acceptability, especially in Developing Countries

André Nendjo Ella
AICADA (Agence Internationale des Cadres Africains pour le Développement de l'Afrique)
1.Midan Ahmed Orabi El mancheya
BP: 415 Alexandria -EGYPTE
Tel: 819981/80
In doctoral research at LAMIH-CNRS 1775
e-mail anendjo@univ-valenciennes.fr

Martial Grislin and Christophe Kolski
LAMIH-CNRS 1775
Le Mont Houy BP 311
59304 Valenciennes Cedex-FRANCE
Tel: (33) 27.14.11.73
e-mail (Kolski, Grislin)@univ-valenciennes.fr

ABSTRACT. For a long time, only usability and utility were the main criteria considered in design and evaluation of the human-computer interfaces. That vision is now limited and presents a lot of inadequacies, especially in Developing Countries. Indeed, utility and usability which are technical and ergonomic dimensions, are used to the detriment of some important criteria such as social criteria, induced costs, and so on. This aspect of the things could not let indifferent AICADA, agency participating to the development of Africa.

Our document presents a way followed for the design and the evaluation of the human computer interfaces intended to the developing countries and in some case to the western countries. The way we suggest, is a new approach called PEOPLE'S CO. (Path Enclosing Operator's Participation and Logical Evaluation of Social Criteria and Others (usability, utility...). It is actually an approach of assistance based on the cultural constraints and the usability problems of HCI.

KEY WORDS. Human Computer interface (HCI), Evaluation, Participation approach, Social criteria, Developing countries.

1. INTRODUCTION

Several years after the development of the first human machine interfaces, time is now to the assessment. This last does not present a positive result referring to some global profitability objectives. Limited to the alone year 1994, the losses (fires, pollution, ...) caused by the interfaces dysfunctionning are valued at 1,43 billion of french Francs (Haehnsen, 95).

This not satisfying scenario in front of whom many designers keep the silence, constitutes the fate reserved to many interfaces in a lot of firms having invested high sums in these softwares. There is therefore an emergency to rethink our point of view in matter of interface evaluation.

In general, the evaluation consists in comparing a model of the object observed to a référence model (evaluation model) allowing to draw conclusions (Senach 90; Grislin, 95), figure 1.

For Whitefield and al. (91), the evaluation is an estimation of the close corresponding between the performances of system and the expected performances. This estimation includes a method (the procedures used to realize the estimation) and a ratio (the outcome products).

Also note that numerous methods and techniques able to contribute to evaluate the HCI usability exist in the literature. They are listed by Grislin and Kolski (see Grislin et al., 95, Grislin and Kolski, 96), figure 2.

Generally, an evaluation is done with a pre-established objective, that determines the target data to be collected. In fact, the evaluation should be much more demanding in taking into account the dimension of utility, usability, safety, cost and social acceptability (Nielsen, 93), figure 3.

Figure 1. Evaluation principle (adapted from Senach, 90 and Grislin, 95)
2. IMPORTANCE OF THE SOCIAL ACCEPTABILITY IN THE HCI DESIGN AND EVALUATION

Regarding the social requirements, one notes that the mentalities have developed a good deal, the social considerations become as important as the constraints of utility and of usability. The following examples illustrate situations where social acceptability criteria are not taken into account:

In the following example, we see how not taking into account some social requirements have led to modifications in the interfaces and thus high costs. The following example of graphic interfaces had been designed, realized and evaluated in the United States of America for an hygiene plan in Egypt, figure 4.

When the picture visible in figure 4, which shows two Egyptian women with one is undergoing a contraception, had been presented in Egypt, the Egyptian women had found it unacceptable (cultural and dignity problems). Some of them had commented it in these terms: "we don't look like the western women and why does she let its fingers be naked?"

Decision was taken to modify the interfaces with the consequences of the additional costs. The designer were obliged to revise or to eliminate the pictures including some questionable elements.

A succession of modifications gave some diagrams, that were socially acceptable, like the one of the figure 5.

This example underlines a case of frustration who makes the users do not feel useful for the factory any more; they couldn't therefore accept it. That effect of social acceptability and motivation on human machine system is presented in Figure 6.

3. TOWARD AN APPROACH CALLED PEOPLE'S CO.

With PEOPLE'S CO. (Path Enclosing Operator's Participation and Logical Evaluation of Social Criteria and Others (utility, usability, cost, reliability...), we propose an evaluation approach adapted on one hand to the developing countries where the social worths are very important, and on the other hand, to the western countries in some cases.
3.1. PEOPLE'S CO. presentation

Path (itinerary, way, path to follow in matter of HCI design and evaluation)

Enclosing (the path to follow takes into account...)

Operator's Participation (...the operators' participation and...)

Logical Evaluation (...a logical obedient to an order 1-2-3-4-5-6 of evaluation...)

'Social Criteria (...with social acceptability criteria and...)

Others (...other practical acceptability criteria, as utility, utilisability, induced cost...)

3.2. The PEOPLE'S CO. 1-2-3-4-5-6 logic

In PEOPLE'S CO., one can recognize a logical approach that follows the order 1-2-3-4-5-6. This logic leads to respect the order of the evaluation steps and enables to give an answer to the question: what must be done first in the aim to save time and to reduce or to eliminate the risks for the interface to be modified or not used after their development? Those steps are successively described below.

The proposed approach allows to do the evaluation after the stages of functional specification. The evaluation must be done first in term of social acceptability, then in term of practical acceptability. One will define thus six stages according to a well defined order (1-2-3-4-5-6), called here logical order of evaluation, figures 7 and 8.

The figure 7 presents the evaluation logical order in the present situation. The social, society and users requirements are more and more important. This new human behavior is particularly emphasized in the developing countries where the cultural constraints add to the HCI difficulties of use. So, to take into account all these aspects and moreover, to avoid the HCI modification or withdrawal because of its social non acceptability or of too high induced costs, we consider that the HCI must be evaluated first in terms of social
acceptability before its detailed design and development.

3.2.1. Evaluation-1 (evaluation of HCI in term social acceptability and human operators's participation)

The aim of this evaluation is not only to take into account the users needs but above all to seek for the social acceptability in allowing the users to take an active part in the plan. Their propositions and their social acceptability will certainly imply them in the process of final automation. Today, that way is one of the success keys for the interfaces design and evaluation. The evaluation-1 includes the presentation to the users of models and/or prototypes of the future interfaces.

3.2.2. Evaluation-2 (evaluation of HCI in term of social acceptability by the society)

The principle is the same as for the evaluation 1, and the following stage is evaluation 3.

3.2.3. Evaluation-3 (evaluation of HCI in term of induced cost, compatibility, etc)

According to Nielsen (93), we introduce here the notion of costs linked to softwares use. Indeed many interfaces are abandoned nowadays because of the entailed costs due to their use (training of the human operators by an expert, information, set-up of some work station(s), necessity of training, maintainability) become more important than the gains, not always evident, of productivity.

After this evaluation, if the conditions are favorable that is to say, for example, that the induced costs are reasonable, then one could go on to the following stage, which is the conception of the interface, which may be followed either by an a priori evaluation, or by an a posteriori evaluation. In the other case, pass on to the actions and retroactions of modifications aiming at the reduction of the induced costs (for instance predicting the costs induced by the training of the human operator so that he is a future formative agent himself instead of calling an expert who will cost ten times more; this is the most frequent problem for the Third World companies, which, in spite of their low liquidities and in a competitiveness goal, rush into automation process that forces them to make training experts to come, which annual cost is as high or even more as a significant part of the salaries and sometimes induces prejudices to the result of the company.

3.2.4. Evaluation-4 (evaluation of HCI in term of utility)

The evaluation is done a priori if the interface is not already implemented; the evaluation is done a posteriori if it exists already. If it is the evaluation a priori, one will proceed step by step to the actions and
retroactions of modifications, if problems of conformity to the specifications are found and until satisfaction; in the other case, one will go to the following stages of the interface development.

For an evaluation after the development, two case can be found: the conclusions are favorable, then according to our objectives one can consider evaluation 5 (evaluation of HCI in term of usability). In the other case, one must apply the required actions of modification until the conclusions are acceptable.

3.2.5. Evaluation-5 (evaluation of HCI in term of usability)

See evaluation 4, but this time according to the classical utilisability criteria (Cf. Ravden et al., 89 or Scapin, 90).

3.2.6. Evaluation-6 (evaluation of HCI in the context of single or multi-workstations)

This phase consists in evaluating a priorio single or multiple work stations. This ergonomic set-up is called ergonomic conception which aims at integrating the ergonomic criteria into the conception of a work station. For the static set-up this stage is based on ergonomic guide lines (Millot 88). On the other hand, the ergonomics of conception for the evaluation and the set-up of the task’s dynamic components can only be done with a simulation of the operator’s behaviour. It is based on a predictive model of the performances and of the workload.

When the (multiple) work station(s) and interface exist, one will proceed to ergonomic evaluation, called evaluation of conception which consists in measuring the performances of and the workload of the operator.

At the end of this evaluation, if the conclusions are favorable then the Human(s) Machine(s) system is globally acceptable and one can dream of a global best performance. In the other case, go on to the actions of set-up of (multiple) work station(s).

4. FIRST DISCUSSION ABOUT THE NEW METHOD AROUND AN INDUSTRIAL AIM

The exemple we describe in figure 9 is a real case that the first author has experienced as a designer. He was in charge of (1) the start-up of a human-computer interface for water management within two hydroelectrical power stations and (2) the training on the spot of the operators who must pursue the work. This happened in Gabon (Africa) from November 1991 until September 1992 within the SEEG (Société d’Energie et d’Eaux du Gabon) (Nendjo Ella 92). The decision to use a human-computer interface for water management was the consequence of an exceptionally dry year that the Gabon had never known since 1983.

This example is not only an industrial case but also illustrates, on one hand the crucial problem of induced costs for the company, on the other hand the problem of social acceptability by the users. This situation has forced the leadership of the company to abandon this human-computer interface. This interface, which is nevertheless of good quality, spends its last days in the cupboards of the company.

One of the main tasks to carry out was the water management of the two hydraulic basins of TCHIMBELE and KINGUELE (types of power stations in cascade) with some other sub-tasks like the capture of the data relative to the load curve. Our industrial objective was to obtain due to the funcionalités of the human-computer interface, an output rate (Qd) equal to zero.

The values of Qd allows to draw the curve of performance PF3, figure 10. But the expected performance (Qd= 0) were not reached for the following reasons:

• The total lack of motivation for the human operators. It was due to the absence of social acceptability: the operators saw the arrival of this automatised management like the end of their know-how and moreover as the end of their importance within the company. Already, at the beginning of the project, the power station agents let provoke a conflictual ambiance. For them, a danger could occur from this new system which came from abroad and which was strange for them (cultural shock), and they could loose their job. This behaviour was globally the same than those in the LU production plant (see above). This permanent fear would be avoided using the logic 1-2 of the PEOPLE’S CO approach.

• The lack concerning the training on the new system. The training period was too short (few days). Nevertheless, according with many authors, training quality is a factor influencing performance.

• The lack of improvements concerning the work station. Indeed, the workstation was not adapted for the situations of this type. The vibrations and noises coming from machines were very disruptive into the control room, and influenced strongly the human performance (a problem of physical ergonomics, taken into account by the logic 6 of the PEOPLE’S CO approach).

• The functional capacities of the new system had been evaluated by experts using empirical methods, on site and in real use cases. The results were good :Qd was smaller or equal to 0. We have confirmed these good results; so there was no utility problem (logic 4 of the PEOPLE’S CO approach). But the human operators obtained bad results owing to the reasons explained above.

• The leadership of the company decided of seeing again all the system. The new planning advocated to first solve the social problem in order to create a climate of confidence and motivation for the human operators (Logic 1 of the PEOPLE’S CO approach). One must send some operators abroad for a complete training period. It was also planned to improve the workstations. These new objectives led to high induced costs which were not evaluated at the beginning of the project (logic 3 of the PEOPLE’S CO approach). When the spending concerning the induced costs were estimated, they seem so high that it was judged better to give up completely the project.
Using the PEOPLE'S CO approach, the HCI development or use could not get ahead of the evaluation concerning the induced costs (Logic 3): it would be necessary to follow strictly the 1-2-3-4-5-6 approach. It would be more possible to avoid such a situation (like described above), to better implicate the human operators in HCI design and evaluation, and consequently to increase the confidence and the motivation. The third curve in figure 10 gives a theoretical view of the obtained performance using the PEOPLE'S CO approach.

Figure 9. Automatic hydraulic pool management at the hydroelectric power station of KINGUELE and TCHIMBELE in GABON (Central Africa, 1991-1992)
5. CONCLUSION AND PERSPECTIVES

With the evolution of knowledge and technology, we must take into account some other research directions as regards of human(s)-machine(s) system design and evaluation. For one decade, only usability and utility were the main criteria considered in software design and evaluation. The objective is to realize a wellusable software to enable the users to perform the tasks quickly and without any effort. If this is the aim of a good "usability" then we are tempted to ask: are there any other ways to achieve the same goal? According to this remake, it is worrying when we see that the design takes into account the human operator requirements. It is a good thing but the humans form a very complex system, whose requirements vary from an individual to another, from a culture to another.

Indeed, a lot of human computer interfaces designed, realised and evaluated in western countries and intended to the developing countries, have been abandonned or modified because the technical aspects have been mostly considered until now to the detriment of the ergonomics, social or cultural considerations. These interfaces are generally designed and evaluated according to the criteria of utility and usability. The requirements can be purely social (problem of dignity and frustration, Nielsen 93), cultural (Russo et al. 93; Boor 93, connected to the nature of the pictures or some gestures incompatible with the social norms), of ecological order (problem of pollution due to the defects of the data processings, the sinister of origin data processings led to important loss; Haehnsen, 95) and socio-economical order (for example, a complete automation would suppress some operators). To this effect we present an idea of assessment often little flattering in the matter of the interface development.

It is why, it is very important to consider explicitely the social critera during the evaluation. In the next months, we aim at going further into the PEOPLE'S CO approach.

6. REFERENCES


Acknowledgements. The authors would like to Thank Emmanuelle Le Strugeon for her strong help concerning the translation of the paper.
Session 3
Remote Control
THE DEVELOPMENT OF A MAN-MACHINE INTERFACE FOR SPACE MANIPULATOR DISPLACEMENT TASKS

Eric F.T. Buiël

Delft University of Technology, Department of Mechanical Engineering and Marine Technology, Laboratory for Measurement and Control (Man-Machine Systems Group), Meekelweg 2, 2628 CD Delft, The Netherlands. E-mail: E.Buiel@wbmt.tudelft.nl

Abstract: A space manipulator is a lightweight robotic arm mounted on a space station or a spacecraft. If the manipulator is manually controlled from a remote location (teleoperation), the human operator can only see its movements in the pictures from the cameras mounted on the manipulator and the pictures from the cameras installed in the neighbourhood of the manipulator. Typical tasks of space manipulators are transportation tasks and inspection tasks. During the execution of these displacement tasks, the human operator has to be alert not to cause a collision between the manipulator limbs and objects in the environment. This is not an easy job, because the distances between the manipulator and the objects can hardly be estimated from the available camera pictures.

Recently, at our laboratory, a conceptual man-machine interface has been developed for space manipulator displacement tasks. The new anthropomorphic European Robot Arm (ERA) served as a reference in this project. At the control-side of the interface, a force-activated control device with six degrees-of-freedom (the Spaceball) is applied to control the movements of the ERA end-effector (the hand of the manipulator). At the display-side of the interface, a single camera picture is shown: the picture from the camera, that is mounted near the ERA elbow joint. To assist the operator in deriving spatial information from the elbow camera picture, a graphical camera overlay is added: the Raindrop Overlay. In this graphical overlay, the actual distances between the manipulator and the objects in the environment are visualised by means of raindrop-shaped distance lines.

Keywords: Teleoperation, Manual Control, Collision Avoidance, Graphical Displays

1. INTRODUCTION

1.1 The manual control of a space manipulator

A space manipulator is a lightweight robotic arm mounted on a space station or a spacecraft. There, it performs inspection and maintenance tasks, e.g. the repair of a damaged satellite. Figure 1 shows an example of a space manipulator: the European Robot Arm ERA\(^1\) (van Woerkom et al, 1994; Traa, 1995). This fully symmetric manipulator with two grapples (the end-effectors) is meant for the new International Space Station (ISS; Dooling, 1995). It is approximately ten metres long and will be able to walk across the station by moving an end-effector from its actual base point to a new one; alternately with the first and the second end-effector.

Recurrent tasks of a space manipulator (e.g. the replacement of Orbital Replaceable Units containing scientific experiments) may well be automated and performed under supervisory control. This does not seem plausible for tasks that are not well defined in advance (e.g. repair tasks). Here, the inventiveness of the human operator is

---

\(^1\) The European Robot Arm is developed at Fokker Space B.V. in Leiden, The Netherlands.
required more often. Then, teleoperation (Sheridan, 1992) seems a suitable control method. With this method, the human operator controls the manipulator by hand from a remote location (e.g. a space station's manned module or a ground station on earth). Astronauts do not have to go outside their spacecraft to control the manipulator; the manipulator movements are controlled with the help of the pictures from the cameras installed in the neighbourhood of the manipulator and the pictures from the cameras mounted on the manipulator itself.

The mentioned teleoperation task is a hard job for the human operator. First, the lack of spatial information in the available camera pictures complicates manual control. Besides, task execution suffers from the manipulator dynamics: because of the lightly constructed limbs, the manipulator will be flexible. Finally, when the operator controls the manipulator on earth, time delays are introduced in the control loop. These delays are caused by the transmission of control signals from earth to space, and back again.

At the Delft University of Technology (DUT), the manual control of a space manipulator at a remote location is an object of study (Bos, 1991). The research is aimed at the development of a conceptual man-machine interface (MMI), that can diminish the three problems mentioned above as much as possible. Elements of the interface are implemented and tested in a simulator (see Figure 2). In this simulator, the movements of the European Robot Arm ERA are simulated by means of a Silicon Graphics graphical workstation (©). This computer animates simplified camera pictures of the ERA movements and additional information displays (©) in real time. Subjects control the movements with a Spaceball® control device (©): a force activated control device with six degrees-of-freedom (DOF's).

Generally spoken, the activities of a space manipulator can be subdivided in two elemental tasks: the positioning task and the displacement task. Breedveld (1995a and 1995b) has developed the DUT interface for the positioning task. This paper will focus on the development of the DUT interface for the displacement task.

1.2 The space manipulator displacement task

Transportation tasks and inspection tasks are typical displacement tasks. During the execution of these tasks, the manipulator covers large distances. Then, the human operator has to be alert not to cause a collision between the manipulator limbs and objects in the environment. This is not an easy job, because the distances between the manipulator and these objects can hardly be estimated from the available camera pictures. Even worse, sometimes a dangerous object isn’t even visible in the camera picture currently observed by the operator.

1.3 The MMI for the displacement task

If we want the human operator to avoid collisions, the MMI for the displacement task has to provide the operator all the information he needs to be able to estimate and control the actual risk of a collision for all parts of the manipulator. Two instruments can provide the necessary information: sensors measuring the current manipulator position, and (obviously) the available cameras. In the case of the European Robot Arm, a number of cameras are located on the ISS (environment cameras), and others are mounted on the ERA itself (four robot cameras; see Figure 1). Besides, the actual joint angles are measured by angle sensors, and the locations of the elements of the ISS are registered in a geometry database (the world model).

This paper will guide you through three stages in the design of the DUT interface for the displacement task. In the first stage, the information to be presented at the display-side of the interface has been selected. Here, one of the available camera pictures has been chosen to be the central camera picture on the interface console. In the second stage, the six Spaceball DOF’s have been mapped consciously to the movements of the manipulator in that picture (design of the control method). Third, a graphical overlay has been designed. This overlay emphasises the spatial information in the central camera picture, and presents additional information about the collision risk at the locations that are invisible in it.

2. THE DISPLAY–SIDE OF THE INTERFACE

2.1 Introduction

In current teleoperation testbeds (e.g. Pauly and Kraiss, 1995; Blackmon and Stark, 1995; Silva and Gonçalves, 1993; Bejczy, 1996) all of the available camera pictures and position information are often integrated in a small number of spatial information displays. Each of these displays presents a different view of the remote environment to the operator. This view can either be an existing camera picture with a spatial graphical overlay, or a synthetic spatial image of the environment (an artificial camera picture). In both cases, 3D computer graphics are used to visualise the available position information.

2.2 Two subtasks in the displacement task

The availability of multiple viewpoints of the remote site is important for planning the rough collision-free path to the desired location of the manipulator; in this planning subtask, information about the whole working environment is of major importance. But while moving along a chosen path, that is not the case anymore. For this
control subtask, the operator requires detailed information about the collision-risk at the current manipulator location only. If he is expected to extract this information from the same information displays as the ones used for the planning subtask, the operator may find difficulties, especially if the manipulator has to cover large distances. Then, multiple (artificial) camera pictures can show valid information at the same time. E.g. one picture might show the current position of the end-effector, and another one might show a hazardous object at short distance from the manipulator base. Then, it can be difficult to decide on future control actions.

Until now, relatively small attention has been paid to the operator’s information needs in the control subtask. In most of the current teleoperation testbeds, the above mentioned planning-oriented displays are applied to pre-program the desired displacement (e.g. Blackmon and Stark, 1995), or to perform the job in (semi-)supervisory control (e.g. Park, 1991). Generally, fully manual control of gross motions is only a redundant control method meant for extraordinary situations in which the normal control method is malfunctioning. Consequently, there is no proper display concept for the control subtask. For this reason the development of the DUT interface for the displacement task has mainly been focused on that part of the job. It is assumed that suitable planning-oriented displays are available, and that the operator already knows the rough collision-free path to the desired location while using the information display for the control subtask.

2.3 The information display for the control subtask

In the ideal situation, all the information the operator needs for the control subtask is visualised in one single spatial image of the current manipulator location. To ensure that this image will always show valid information, the viewpoint of the image has to move simultaneously with the manipulator movements. In the ERA case, the viewpoint for a synthetic ‘best view’ of the current manipulator environment can always be computed from the actual manipulator position and the ISS world model. For a teleoperator with moving base, Das (1989) proposed to calculate the viewpoint from which the operator can see the end-effector, and the two objects at the shortest distance from the manipulator. Then, the operator always controls the movements of the end-effector in the best view of the area with the largest danger of a collision. Unfortunately, with this method, the automatic movements of the artificial camera are tied up with the locations of objects in the actual environment of the manipulator. Then, it is difficult for the operator to predict the camera movement that will result from his intended manipulator movements. Therefore, he must check in which way he looks at the remote environment after each control action. This ‘mental displacement’ of the viewpoint will take more time in case the viewpoint change is large (experiments carried out by Kleinhans (1992) confirm this theory).

A more intuitive way to move the camera viewpoint simultaneously with the manipulator movements can be realised by using the picture from an (artificial) robot camera. If the control method is consciously designed, the operator will know exactly in which direction the ‘home limb’ of the camera will move for each elemental movement of the manipulator. Then, the mental viewpoint displacement will hardly take any time. Ultimately, the mapping between the movements of the control device and the camera movements results in a feeling of telepresence (Sheridan, 1992): the operator feels as if he controls the movements of his own eyes in the remote environment and flies along with the manipulator. In that case, he might well be able to perceive spatial information in the camera picture in a similar way as in daily life. According to Gibson (1979), the flow patterns in the retinal picture of the human eye (optic flow) form an essential cue for body motion perception in daily life. While displacing a space manipulator, the home limb of a robot camera will never move in the accompanying camera picture. The operator must perceive the manipulator motion from the resulting movements of the objects visible in the background of the picture. This continuous flow of objects might well be an analogue cue for manipulator motion perception as optic flow is for body motion perception in daily life.

In the DUT spatial information display for the control subtask, the above mentioned idea has been adopted. The picture of the elbow camera mounted on the ERA forearm (camera ☞ in Figure 1) is the central spatial image in the display. Figure 3 shows the elbow camera picture as it is animated in the experimental facility. The picture will always show the movements of the end-effector, and two other parts of the manipulator often in danger of a collision: the wrist and the forearm. Since the elbow camera is mounted near the elbow, the forearm will partially cover the operator’s view of the actual environment at any time.

Figure 3 The ERA elbow camera picture
The usage of the elbow camera picture makes two demands on the further design of the MMI for the control subtask. First, a control method has to be found, that enables the operator to predict the camera movements that will result from his intended control actions. Second, the graphical camera overlay has to provide information about the danger of a collision for the invisible parts of the manipulator: the upper arm and the backside of the forearm. In an ideal situation, the visualisation of the collision risk in the overlay suggests the control actions required to minimise this danger simultaneously. The graphical overlay has to be adapted to the applied control method to achieve this aim. Therefore, the choice of the control method preceded the development of the graphical overlay.

3. THE CONTROL-SIDE OF THE INTERFACE

3.1 Introduction

With the force-activated Spaceball, the translational and angular velocity of a spatial object in a three-dimensional workspace can be controlled intuitively. The magnitude of the force (torque) applied to the Spaceball determines the magnitude of the object's translational (angular) velocity. The direction of the applied force (torque) determines the direction of the object's translational (angular) velocity. So, if the user grasps the Spaceball as if he grasps a car's gear lever, he might feel it as if he grasps the controlled object (virtual grasping, see Figure 4).

Normally, if a Spaceball is used to control the movements of a robot, the principle of kinematic control is applied. With this method, the operator virtually grasps the end-effector. After he has defined a desired end-effector pose change, the joint velocities necessary to attain the desired pose change are automatically computed from the manipulator's inverse kinematics. The implementation of kinematic control requires the choice of a control base frame. This is the coordinate frame in which the operator specifies the desired end-effector pose changes. After the control base frame has been chosen, the mapping method must be selected. This method defines in which way the six Spaceball DOF's are mapped to changes of the end-effector pose in the control base frame.

3.2 The choice of the control base frame

The origin of the control base frame (the control origin) is imaginarily and inseparably linked to a part of the space manipulator. The orientation of the frame defines the principal movements of the end-effector. The operator can define a desired translation of the end-effector as a combination of three orthogonal translations in the directions of the frame axes. A desired change in orientation can be defined with a rotation vector. The direction of this vector defines the direction of the rotation axis; the vector length defines the rotation angle.

Position and orientation of the control base frame have to be chosen carefully. To avoid mental rotation problems, the orientation of the frame in the elbow camera picture should never change. Therefore, the orientation of the control base frame has been equated to the orientation of a frame imaginarily linked to the elbow camera (the camera frame; see Figure 5). The frame position (as defined by the control origin) determines the point that is insensitive for rotation commands. At first sight, it seems wise to place the control base frame upon the end-effector (the end-effector frame; see Figure 5). In this case, the end-effector position and orientation can be controlled separately. E.g. if the control origin is located at the end-effector tip, the operator can first move the tip to the desired location. After that, the orientation of the end-effector can be corrected without changing the tip position. Unfortunately, in the case of the ERA, the usage of an end-effector frame has a major drawback. In almost all poses of the manipulator, a movement of the end-effector in one of the principal movement directions will require rotations of all six ERA joints. Because of this, all manipulator limbs will move in different directions during the change of the end-effector pose. Then it will be difficult for the operator to predict the resulting limb and camera movements. As a result, he can hardly control the collision risk around the limbs.

To avoid the mentioned problems, the control base frame has been placed at the end of the forearm: the wrist (see Figure 5). Now, a translation in the direction of one of the frame axes requires rotations of the two shoulder joints and/or the elbow joint only (see Figure 6). Generally, an end-effector rotation defined by a rotation vector located at the control origin will still require movements of all joints. But this number of joint rotations can now be decreased if the desired orientation changes are no longer specified with rotation vectors.
3.3 The choice of the mapping method

Note that if the axes of the three wrist joints (joints IV through VI in Figure 1) would have intersected at the control origin, only rotations of these joints would be needed to rotate the end-effector in the direction of commanded rotation vector. This situation can be approximated if a change in orientation is specified with a sequence of wrist joint rotations, instead of a rotation vector. With this alternative method, each joint rotation defines a principal rotation of the end-effector: joint IV influences the pitch-rotation, joint V influences the yaw-rotation, and joint VI influences the roll-rotation. Each of the rotational Spaceball DOF’s is mapped to one of the principal rotation directions. Since the wrist joints are all located in the region of the control origin, the operator still feels as if he controls the end-effector pose in the control base frame.

The proposed semi-kinematic control method results in distinct responses on translation and rotation commands. In most cases, a displacement of the end-effector in a principal direction will result in a rotation of one single joint. Only a translation in z-direction (‘forward’) will result in simultaneous rotations of two joints (see Figure 6). These distinct responses make it very easy to predict the limb movements resulting from a specific control action. As a result, the camera movements can also be predicted easily (note that the camera will never move after a rotation command). Therefore, this control method has been implemented in the DUT interface for the displacement task.

3.3 The choice of the mapping method

Just like the six principal movements of the end-effector are defined by the control base frame, the six principal movements of the Spaceball are defined by the Spaceball frame. The origin of this frame is imaginarily linked to the centre of the Spaceball sphere. The orientation of the frame defines the mapping method: a translation or rotation of the Spaceball in the direction of one of the frame axes results in an analogous displacement of the end-effector in the control base frame.

Earlier man-machine experiments with the DUT interface for space manipulator positioning tasks have shown the benefits of the downward mapping of translations (Buiël and Breedveld, 1995). With this method, the operator must push the Spaceball downward to translate the end-effector forward in the elbow camera picture (i.e. in the z-direction of the wrist frame). So, there is a 90° rotation between the end-effector translations observed in the camera picture and the translations of the Spaceball with respect to the top of its supporting table (Figure 4 shows the Spaceball frame orientation that’s required for this method). With the downward mapping, the tabletop serves as a reference plane. Just like in the mouse parallel to the tabletop corresponds to the movements of the Spaceball parallel to the tabletop (the control reference plane) correspond to the movements of the Spaceball parallel to the tabletop (the control reference plane) correspond to the movements of the end-effector parallel to the elbow camera lens (the movement reference plane). Because of the demonstrated advantages of this mapping method, it has also been implemented in the DUT interface for the displacement task. In accordance with the downward mapping of translations, the Spaceball x- and y-rotation have been mapped to the yaw- and pitch-rotation of the end-effector resp. Finally, the Spaceball z-rotation has been mapped to the end-effector roll-rotation.

4. THE GRAPHICAL CAMERA OVERLAY

4.1 Introduction

The chosen control method ensures that the operator can predict the resulting limb movements for every displacement of the end-effector. If a limb is in danger of a collision, the graphical overlay for the elbow camera picture can now assist the operator in avoiding the collision by suggesting the limb displacement that’s needed to decrease the collision danger. An intuitive way to do this is to visualise the shortest distance between the limb and the hazardous object in the environment. At DUT, de Beurs (1995) developed a computer algorithm for computing distances between objects in the ISS world model. For each pair of objects, the two object nodes at closest distance are calculated by the algorithm in very little computing time. If both nodes are visible in the camera picture, the distance can be visualised by means of a cleverly shaped distance line (see 4.2). If they’re not visible, the control actions required to decrease the collision risk have to be visualised in a different way (see 4.3).

4.2 Visualisation of distances with distance lines

Figure 7 shows the basic idea for the graphical overlay with distance lines (de Beurs, 1995). Three dashed lines visualise the closest distances between the space manipulator and the elements of a central truss at the ISS. To the operator, it seems as if each of the lines is
Figure 7 Elbow camera picture with basic distance lines

actually present in the remote environment. Pilot experiments with this distance line overlay demonstrated the usefulness of the added distance information. But they also indicated two problems inherent to the shape of the distance lines. First, if both the environment-end and the manipulator-end of a distance line are (almost) in one line with the viewing direction of the elbow camera, the operator can hardly estimate the length of the line. Second, the danger of a collision increases at the moment the length of the distance line decreases. This is a major drawback from an ergonomic point of view: at the moment the danger grows, its display indicator becomes less eye-catching. Ultimately, at the moment a collision occurs, it isn't even visible.

Next to the shape of the basic distance line, Figure 8 shows three alternative shapes for this line that (partially) come up to the observed problems. The first option, the elastic line, solves the first problem only. The cross-section of this line increases after a decline of the line's length. Indeed, if it is observed from one of its ends, the line becomes more distinctive at the moment its length decreases. But if it is observed from the side, the line will still be flattened. The second option, the sphere, solves this problem. Here, a sphere marks the environment-end of the original distance line. The sphere radius increases while the indicated distance decreases. Since its size now grows in all directions, the distance indication will be clearly visible from any side. Note that at the moment the sphere radius equals the remaining distance, the manipulator will intersect the sphere (distance < 5 cm in Figure 8). At this moment, the maximum amplitude of the limb vibrations due to the limb flexibility roughly equals the remaining distance, and major attention is needed from the operator. To indicate this, the sphere changes color (green turns red).

In this way, the sphere provides information about the manipulator flexibility and solves both of the observed visualisation problems at the same time. But it introduces a new problem also. At the moment the sphere does not intersect the manipulator, it does not visualise the direction of the distance line. The third (and finally preferred) option, the raindrop, solves this problem. Here, the sphere merges with a dashed cone when the indicated distance exceeds the sphere radius. The top of the cone marks the manipulator-end of the original distance line. At the moment the sphere radius exceeds the distance once more, the raindrop turns into a sphere again. Figure 10 shows the resulting elbow camera picture with raindrop-shaped distance lines. Once again, the picture shows three distances to a primary truss of the ISS. At the location of the large sphere, the distance between the manipulator and the truss is almost zero.

<table>
<thead>
<tr>
<th>Distance</th>
<th>Basic line</th>
<th>Elastic line</th>
<th>Sphere</th>
<th>Raindrop</th>
</tr>
</thead>
<tbody>
<tr>
<td>35 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 8 Four optional shapes of the distance lines
4.3 Visualisation of the collision danger outside the visible area

Two parts of the ERA are invisible in the elbow camera picture: the upper arm, and the backside of the forearm. The collision danger for the backside of the forearm can be visualised by virtually transforming the forearm into a transparent glass tube. As a result, the raindrops currently located at the backside of the forearm will be visible at the back of the tube. Of course, this strategy can not be applied to visualise the collision danger around the upper arm. The distance lines located near this manipulator limb will normally be located outside the viewing volume of the elbow camera. For each of these invisible raindrops, the overlay must visualise the size of its sphere and the direction of its dashed cone in a different way.

Figure 9 shows an example of a situation in which the upper arm is in danger of a collision. Here, a hazardous object is located close to the bottom of the upper arm. The direction of the accompanying distance line is visualised in a cross section of the upper arm. Because of the cylindrical shape of the upper arm, this line will always be directed perpendicular to the surface of the upper arm. The radial position of the line (quantified by the angle $\alpha$) is an important cue for the determination of future control actions. Since the line is located at the bottom of the upper arm, the operator knows that a collision will occur if he moves the wrist forward. In the same way, if the line would have been located on the left side of the upper arm, a collision would have occurred if he had moved the wrist to the left.

From Figure 9, it can be concluded that the radial position of each invisible distance line implicitly shows the control action that’s needed to decrease the local collision risk. This important observation is utilised in the display indicator for the collision danger around the upper arm (see Figure 11). The main element of this indicator is a large circle. This circle represents the cross section of the upper arm. For each invisible distance line, a sizeable sphere marks the location of its manipulator-end (e.g. the sphere located at the bottom of the circle in Figure 11 represents the location of the hazardous object visible in Figure 9). Just like the raindrops in the elbow camera picture, the sphere radius increases while the indicated distance decreases. From the radial position of a sphere, the operator can read the control action that’s needed to decrease the local collision risk.

5. CONCLUSIONS AND FUTURE RESEARCH

In this paper, most attention has been paid to the problem of collision avoidance during gross motions of a space manipulator. The developed graphical overlay with spherical and raindrop-shaped distance indicators

---

Figure 9 Collision danger near the upper arm

Figure 10 Elbow camera picture with raindrop-shaped distance lines

Figure 11 Visualisation of collision danger near the upper arm
clearly visualises the locations currently in danger of a collision, and the control actions that are needed to decrease the danger. As a result, the overlay provides a solution for the first general problem in teleoperation tasks: the lack of spatial information in the available camera picture(s). At the same time, a solution for the second problem - the flexibility of the manipulator limbs - is provided. Raindrops and spheres implicitly show the maximum amplitude of the manipulator limb vibration caused by the flexibility of the limbs. Only the last problem - the introduction of time delays when the operator controls the manipulator on earth - has not been considered yet.

In the near future, man-machine experiments will be carried out to demonstrate the usefulness of the provided distance information. Finally, the time delay problem will be considered. To eliminate this problem, the developed spatial information display will be transformed into a setpoint display (Breedveld, 1995b). A setpoint display visualises the operator's control actions (i.e. the setpoint for the manipulator velocity) immediately after they have been carried out. E.g. the movements of a transparent (phantom) manipulator can visualise the current velocity setpoint. Since he can directly see the results of his control actions, the operator does not have to wait until actual camera pictures arrive at his location.

**ACKNOWLEDGEMENTS**

This research is supported by the Dutch Technology Foundation STW (Utrecht, The Netherlands) and Fokker Space B.V. (Leiden, The Netherlands). See http://www.mr.wbmt.tudelft.nl/~buiel on the World Wide Web for additional information.

**REFERENCES**


Mechanical solutions versus electronic teleoperation, an example

Just Herder, Marcel Horward, Wouter Sjoerdsma
Delft University of Technology, Faculty of Mechanical Engineering, Dept Measurement and Control, Wilmer group
Mekelweg 2,2628 CD Delft, The Netherlands, tel +31.15.2784713, fax +31.15.2784717, e-mail: j.l.herder@wbmt.tudelft.nl

Abstract
In laparoscopic therapy, grasping instruments are the only source of tactile information for the surgeon. Therefore, the force transmission characteristics of these instruments are important features. The force transmission function should ideally be constant throughout its range of motion and be free of hysteresis. Of three approaches to achieve this ideal, mechanical solutions, active compensation and teleoperation, the first mentioned seems most appropriate in the case of medical applications. In this paper, the design of a purely mechanical laparoscopic grasping instrument with outstanding force transmission characteristics is presented.

Introduction
Laparoscopic surgery is an operating technique based on several small incisions in the abdominal wall instead of a single large one. The advantages for the patients are such that this technique has found widespread acceptance. In comparison with conventional operating techniques however, the surgeon is severely handicapped, because laparoscopic surgery deprives the surgeon of direct vision and touch. Instead of direct vision, an image must be obtained from a monitor, and contrary to conventional surgery, in which the surgeon can manipulate and palpate tissue with his hands, all manipulations must be carried out by using laparoscopic instruments.
Unfortunately, currently available grasping instruments allow only minor feedback of grasping force due to their poor mechanical construction, the main problem being Coulomb friction, which absorbs considerable part of the grasping force information before it reaches the surgeon's hand, thus decreasing the signal to noise ratio unacceptably.

Methods
Basically, three methods exist to solve the problem of friction. (1) Mechanical solutions. By designing high efficiency mechanisms, friction may be nihilized. (2) Compensation for friction [e.g. Sorine]. As the amount of friction is known in the design phase, a motor can be connected to the operating lever which appends the moment lost through friction. This method may work well when the mechanism is moving, but is not reliable in stationary situations, since then friction is indeterminate. (3) Teleoperation by application of a master-slave servo system [Howe, Hunter, van Hemert tot Dingshof]. The problem of friction can be circumvented by measuring the grasping force and reproducing it with a motor on the operating handle (the master unit), while the position of the master is reproduced at the grasper (the slave unit). This approach allows easy adjustment of feedback gains, which may be advantageous for example in situations where operating forces are below the human sensory threshold, such as in microsurgery, and tremor can be filtered out before the master's movement is transmitted to the slave unit.
When these three methods are inspected, it seems that, theoretically, possibilities increase, but practical usefulness decreases. Therefore, and because there are separate uses for each concept, this study concentrates on the development of mechanical low friction devices.
Ideal characteristics
In order to obtain a norm, the first step is to conceive of an ideal laparoscopic grasping instrument, with ideal force transmission characteristics. With an ideal laparoscopic grasping instrument, the surgeon should feel no difference between open surgery and laparoscopic surgery. Taking into account the limited movement possibilities of the laparoscopic graspers, the reference situation in open surgery would be as illustrated in figure 1. The surgeon's fingers are shielded from the tissue by metal thimbles, and movement is limited by a frictionless pivot. With this set of 'ideal tweezers', the pinching force is transmitted directly, one to one, from the tissue to the surgeon's fingers. Although there is no direct contact, the surgeon is supplied with reliable information on the pinching force exerted.

With an ideal laparoscopic instrument, a surgeon should not be aware of the fact that the tissue to be handled is not between his or her fingers, but somewhere far away at the end of the shaft: the ideal laparoscopic instrument gives the same sensation as the set of ideal tweezers. Therefore, the ideal mechanical laparoscopic grasper has a constant force transmission function throughout its range of motion, without hysteresis (energy dissipation). In special cases, the value of this constant may be chosen deviant from unity, for instance to enable the surgeon to perceive forces normally below the force threshold of the human finger.

Mechanical deficiencies
Current laparoscopic instruments deviate from the ideal characteristics because of friction, a non-constant force transmission ratio and several other mechanical deficiencies. Efficiency measurements revealed a mechanical efficiency not exceeding 30% [Horward]. Additionally, current instruments show a wide variety of force transmission characteristics (figure 2, calculated transmission functions, hence no friction): the force transmission ratio varies up to a factor 6 over the working range. Non of the transmission functions is flat, and the variety may cause problems when interchanging instruments during an operation, as the surgeon is confronted with different force transmission functions. This may be confusing, since they have similar outward appearances and function.
Low friction design

Low friction is one of the basic and essential demands in the case of the laparoscopic grasper with force transmission capacities. Of many possible low friction pivots, the rolling link principle was selected [Kuntz]. In this technique, elements roll directly on one another, and no specific bearing elements are needed.

The demand for a constant force transmission function was satisfied by the application of a symmetrical construction. As a result of the symmetry, the rollers' angular velocity is equal and, conform the principle of virtual work, their moments are transferred one to one from the grasper to the handle. Additionally, due to the symmetrical construction, the mechanism could be fashioned with high stiffness. As a consequence, the instrument has an outstanding internal stiffness. Details on the mechanism will be available as soon as the patent negotiations are concluded.

A prototype was made, which offered the opportunity to evaluate the mechanism, and to investigate the hypotheses on force feedback in a laboratory setting.

Evaluation

By using a tensile testing machine, the mechanical efficiency and the force transmission function of the prototype were assessed. The movable jaw of the grasper was loaded by a weight, simulating constant grasping force, while the handle was driven at low speed (10mm/s approx.) so as to eliminate dynamical phenomena. Operating force and translation were recorded and from this data the mechanical efficiency and the force transmission characteristic were determined. The mechanical efficiency was calculated as the ratio of output energy and input energy, while the force transmission function was defined as the ratio of output force and input force, as a function of input movement.

From the measurement data, graphed in figure 3, the non-constancy and the energy loss of the transmission were calculated. The overall energy losses turn out to be only about 4% of the input energy. Hence, the mechanical efficiency of the described instrument amounts to approximately 96%. The force transmission characteristic shows a maximum deviation of only 3% from the calculated function.

In addition to the quantitative measurements, some explorative experiments were carried out with the prototype. A flexible hose of 3mm diameter was filled with water and connected with a syringe. With the syringe, pressure pulses were introduced in the hose, so as to simulate a pulsating artery, and objects of various stiffness were manipulated. With this simple experiment, the force transmission of the instrument was investigated qualitatively by a number of students and surgeons. The explorative experiments with the flexible hose have surprised many participants. Preliminary results are that the pulses in the hose were clearly perceivable with the prototype, whereas this sensation was absent when the conventional instruments were used.
Discussion
Of three methods to create a manually operated instrument with force reflection, the method of mechanical solutions was chosen as the starting point for the design of a laparoscopic grasper. This method will only be successful if a high mechanical efficiency mechanism with constant force transmission characteristic can be conceived of. Thorough redesign has resulted in an instrument with substantially improved mechanical qualities: an almost constant force transmission characteristic (deviation 3%) with only marginal friction (4%) was realised.

Explorative experiments have indicated that with the purely mechanical instrument presented in this paper, several physiological parameters, such as stiffness of tissue and pulsating blood pressure in arteries, can be perceived. This quality is perhaps the most essential improvement compared to current instruments. It is expected that it will facilitate the localisation of tumours (stiffness different from healthy tissue) and arteries (transmission of pulses). Furthermore, as the surgeon has feedback on grasping force, it is anticipated that applied grasping force and tissue damage will decrease.

In the near future, the following developments are foreseen. A clinically usable product is to be derived from the prototype, which will include work on ergonomics, improved simplicity, reduced cost, improved sterilisability, increased effective shaft length and reduced weight. In vitro experiments will then be carried out to compare the sensitivity threshold of the here described and conventional instruments, by using psychophysical measurement methods [Stevens, Gescheider]. Finally, in vivo experiments should give subjective judgements of surgeons, and supply data in order to verify hypotheses on decreased grasping force and tissue damage, and improved perception, control and safety.

References
Sorine M, Models of hysteresis, application to friction modelling and compensation.
REAL AND SIMULATED DRIVING WITH CAMERA VIEW

Jan B.F. van Erp & Pieter Padmos
TNO Human Factors Research Institute
Soesterberg, The Netherlands.

ABSTRACT
This paper attempts two answer two questions: firstly, is it possible to drive a vehicle properly when only camera view is available? And secondly, if this is possible, what are the human factor requirements for imaging systems which provide controllers of land vehicles with a view of the environment? To answer these questions, two experiments were conducted. The first experiment was performed at a driving circuit with an instrumented car, and the driver situated inside the vehicle itself. This may be regarded as a simulation of driving under armour. The second experiment was conducted in the TNO - HFRI fixed-based driving simulator, and may be regarded as a simulation of a second major application of camera-monitor systems: operating an unmanned ground vehicle. Objective driving performance was measured across an extended taskbattery, including driving sharp curves, performing a lane change, and estimation of speed and distance. Investigated camera-monitor parameters were: placement and aiming of the camera, field of view, image magnification, and presence of spatial orientation cues. The results show that driving with a camera-monitor system is indeed possible, be it with moderate performance degradation compared with direct view. In addition, best performance is attained with a camera mounted at the back of the car, with a wide field of view (100°), combined with magnification 1.0.

INTRODUCTION
There are two main (military) applications for camera-monitor systems in ground vehicles. Firstly, driving of tanks or other armoured vehicles on the battlefield. Traditionally, in this situation the outside view is provided exclusively by periscopes, which impose serious restrictions on the field of view (Van Erp, Padmos & Tenkink, 1994; Padmos & Van Erp, 1994). The advantages of a camera-monitor system with respect to periscopes are a larger flexibility in field of view, placement and aiming of the camera. Secondly, for Unmanned Ground Vehicles (UGV-s), when the operator is situated at a remote location, and camera images of the outside world around the vehicle are transmitted to a monitor in the control station.

Two questions arise. Firstly, is it possible to drive a vehicle properly when only camera view is available? And secondly, if this is possible, what are the human factor requirements for imaging systems which provide controllers of land vehicles with a view on the environment? To answer these questions, two experiments were conducted. The first experiment was performed at a closed driving circuit with an instrumented car, and the driver situated inside the vehicle itself. This may be regarded as a simulation of the first application mentioned: driving under armour. The second experiment was conducted in the TNO - HFRI fixed-based driving simulator, and may be regarded as a simulation of the second application mentioned: operating a UGV. Objective driving performance was measured across an extended taskbattery, including driving sharp curves, performing a lane change, and estimation of speed. Investigated camera-monitor parameters in the first and/or second experiment were: placement and aiming of the camera, field of view, image magnification, and presence of spatial orientation cues.

LITERATURE
Over the past years much research is aimed at the effects on driving performance using only a camera-monitor system. Unfortunately there is only little known about specific demands on those systems, partly because these experiments typically were conducted to develop and test a specific vehicle or a specific image system, or were not involved in real road driving but in surveillance or repair tasks indoors (for example see Mestre, Peruch, Terre & Fournier, 1995).

Below general drawbacks of camera-monitor systems and a brief overview follows of what is known to be the effects of important restraints on camera view: restrictions of field of view and image minification (for a more extended review see Tenkink, 1989 or Padmos, 1995).

General drawbacks of camera-monitor systems Compared to direct view, camera-monitor systems have important drawbacks, including declining contrast, missing binocular depth cues (for example Holzhausen, Pitrella & Wolf, 1993) black and white instead of colour
vision, and limited addressable resolution. For example, Van Erp & Padmos (1994) measured the addressable resolution of their camera-monitor system for driving under armour, which was only about one fourth of the human eye (measured value .5 arcmin¹, human eye about 2 arcmin¹).

Another problem with camera-monitor systems is accommodation micropsia. This may lead to distance overestimation and size underestimation errors when the eyes of the operator are accommodated at short viewing distances to the monitor. This phenomenon is described by Roscoe (1979, 1984) and others (Alexander, 1975; Hollins, 1976; Meehan & Day, 1997), and is relevant for the present experiment. Accommodation micropsia leads to apparent minification of objects when the eyes accommodate inwards. Of course the result of this phenomenon will be an overestimation of distance. In the present experiment overestimation of distance on the basis of this phenomenon may be expected in the camera view conditions, when subjects have to accommodate their eyes at a distance of less than 30 cm. But also at larger distances (i.e. with projection on a screen) this may also occur, Meehan (1993) reported apparent minification indoors over object distances of 3 to 9 m.

None of these factors is systematically investigated with subjects performing extended task batteries on paved roads. However, it is hypothesized that these factors may increase the known performance deteriorations with camera view.

**Effects of restricted field of view.**
The side of the road is an important cue for maintaining a correct lateral position. In this respect, most important seems to be the road section between 15 and 50 m in front of the car. Rockwell, Ernst & Rulon (1970) report lower variation in steering angle and less line crossings when the part of the road between 23 and 30 m in front of the car was visible. McLean & Hoffmann (1972) report problems when sight distance was less than 25 m; Riemersma (1987) reports higher standard deviation in lateral position when there was no sight of the first 20 m in front of the car. Tenkink (1988) found that decline in performance mainly emerges in curves. This is in accordance with the observation of McGovern (1987) of drivers not feeling comfortable turning corners.

A second impact of restricted field of view concerns speed estimation. Decrease of field of view will lead to a lower estimation of speed. Salvatore (1968) reports estimates which were between 2 and 10 km/h lower than the target speed (target speed between 40 and 100 km/h, horizontal field of view 25°). Osaka (1988, 1991) found a power function between estimated speed and field of view with an exponent of .3 (driving speed 80 km/h, field of view between 3° and 55°). Wood and Troutbeck (1992, 1994) did not find an effect varying field size from 20° to 40°, and neither from 90° to unrestricted.

Problems will also arise in curves. Most reported effect is that subjects start to turn too early (Moler & Brown, 1960; Ernst, Sanders & Ter Linden, 1967). In the latter experiment subjects had to drive a jeep on a windy circuit, with or without an infra red camera (round field of view 30° or 50°). Most errors were made in the sharp curves (arc radius less than 20 m). It may be argued that problems arise when the inside of the curve disappears out of view (see also Land & Lee, 1994; Land & Horwood, 1997). This depends on the position on the road, field of view, speed and the curve itself (Tenkink, 1989). The same effects are found when manoeuvring around obstacles or trying to position a car properly: action is taken too early (for example Ernst et al., 1967; Nagata & Kuriyama, 1984).

**Effects of image minification.**
Minification (either combined with larger field of view or not) will lead to a changing relation between the image motion and the steering moves of the driver. The reduced visual flow may lead to larger course instability because of underestimation of lateral distance and lateral speed. Indeed, Schulz-Helbach, Dongs & Rothbauer (1973) found an effect when driving in a tank simulator with minification factor of 2.6: standard deviation of the lateral position increased from .3 to .4 m.

An other effect of image minification is an overestimation of speed. Evans (1970) reports overestimation between 4 and 6 km/h with driving speed between 10 and 100 km/h with a minification factor of 2.3. Minification was accomplished by changing the viewing distance, accordingly field of view was constant. Schulz-Helbach et al. (1973) found a decrease in speed from 55 km/h to 45 km/h driving a tank simulator with minification factor 2.6.

A third effect of minification concerns the estimation of distance. Distance (and size) estimation has been the subject of various studies, both theoretically and in the field of UGVs. For example Kraft (1989) had subjects...
watch slides taken of the same scene, but with different focal lengths and taken from different distances. Of course focal length is not only equivalent to field size but, at a constant viewing distance (as in the experiment of Kraft), is also related to magnification factor. Results of the study of Kraft that are relevant to the present study are the following: distance estimations increase when field size increases (and magnification decreases), and standard error of estimations increases when actual distance increases.

The overestimation found in experiments on size and distance judgements is confirmed in field experiments with UGV's. For example McGovern (1987, 1988) and Miller (1988) report subjects driving over cones and other examples in which subjects overestimate distance. The imaging systems used in these experiments always had a magnification between 0.4 and 0.7, so there was no comparison with magnification 1.0. On the basis of these experimental observations McGovern (1988) repeats the suggestion of Roscoe (1984) that overestimation will occur until a magnification of 1.25. Despite this suggestion no follow up experiments with UGV and true geometric similarity or image magnification have been reported to test this. Earlier experiments with indirect view and magnification larger than 1.0 were rarely performed, and results are contradictory. Eernst et.al. (1967) found with magnification 1.0 an underestimation of distance, whereas with magnification 0.6 distance estimates were correct (i.e. a relative overestimation of distance at minification). Brown and McFaddon (1986) did not find consistent effects of magnification within a range of 0.75 - 1.5 on distance estimation.

EXPERIMENT 1

Experiment 1 is a field experiment, in which subjects performed an extended taskbattery on a paved driving circuit closed for other traffic. The taskbattery was run both with direct view and with camera view.

Methods

An instrumented passengers car (DODGE CARAVAN V6, see Verwey, Burry & Bakker, 1992) was used. In the direct view conditions, no restrictions were imposed. In the camera view conditions, the driver's exclusive view on the road was supplied by a video camera (black and white, JVC TK-S310 EG) mounted on the car's roof over it's longitudinal midline, and a video monitor (PHILIPS LDH 2152/00) mounted above the steering wheel (see Figure 1). Screen size of the monitor was 186x137 mm, viewing distance was fixed (by means of headrests) at 25 cm, thus leading to a field size of 50° diagonal.

Within the camera view conditions, three camera factors were varied: camera position: low (at the front of the roof) or high (at the end of the roof), field size (46° or 100° diagonal), and spatial orientation aids called monimark (absent or present). See Figure 1 for the combination of camera position and field size. The spatial orientation aids provided a horizon, a scale with distance to the front bumper, and the path of the wheels; outer sides at straight course. The monitor images with the car aligned with the right hand road markings are depicted in Figure 2. Only for wide field size the car is visible.

During the experiment, primary driving measures were recorded digitally as a function of time (20 Hz sampling frequency), and included: longitudinal speed, distance travelled, lateral distance to the right hand road marking, and steering wheel angle. Fixed speeds were maintained by means of force feedback on the accelerator pedal.

Figure 3 depicts the driving circuit and tasks. For each viewing condition, the taskbattery was run in a fixed order, consisting of (among others):

1. driving sharp curves (left and right) at a fixed speed of 20 km/h. Main instruction during the sharp curves was to follow the road marking as closely as possible, without touching or crossing it. Used dependent variables: mean lateral distance to the marking (only counted as positive; Harms, 1993), Standard deviation of lateral speed (at fixed speed equivalent to course instability, Blaauw, 1984), and mean absolute steering wheel speed (a measure for steering activity, Godthelp, 1984).

2. performing a lane change at a fixed speed of 40 km/h, changing from left lane to right lane and back, according to ISO (1975) standard. Used dependent variables used are the course instability, the steering activity, plus the standard error from midlane.

3. estimating speed, which meant pulling up to the target speed of 50 km/h and indicate this by pushing a button. In this task, the speedometer and the force feedback on the accelerator pedal were turned off. Used dependent variable is the percentage error from target speed.
Figure 1. Schematic side view of the car to scale.

Figure 2. Monitor images of experiment 1 including spatial orientation aids.

Figure 3. Driving circuit, tasks indicated with numbers and arrows.
Eight male subjects (age 23 to 38 years) participated in the experiment. They were all military driving instructors and had a driving experience of at least 150,000 km. All of them had normal vision; before participation they were tested with standard TNO-tests for visual acuity ($\geq 2.0 \text{ arcmin}^{-1}$, see Vos, Van Norren & Boogaard, 1981), and stereovision (threshold disparity $\leq 60$ arcsec). The subjects received no extra reward for their participation.

One pair of subjects was tested during a week. After arrival on site they received a general instruction on the goals of the experiment, followed by an instruction run with direct view, aimed to teach car handling and performing the task battery (an instructor always sat beside the subject). After each task in a run the subjective task difficulty was recorded, as well as various comments from the subjects on the task and on strategies they used to perform the task. The first and last experimental run per subject was always in direct view, in between these control runs were the eight camera conditions (three factors with two levels each, full factorial design, order balanced across subjects). A short instruction proceeded every new camera view condition.

Analyses of variance were conducted per performance measure and task with two statistical designs:

**CAMERA:** subject (8) $\times$ camera (2) $\times$ replica (3). The levels of camera are the mean of the two runs with direct view and the mean of the eight runs with camera view.

**CAMERA FACTORS:** subject (8) $\times$ position (2) $\times$ field of view (2) $\times$ orientation aids (2) $\times$ replica (3).

**Results**

Below, the results are presented per effect.

**Direct view vs. camera view**

In the **CAMERA** design we found an effect (means in the order direct view - camera view) in the sharp curves task on the mean lateral distance ($F(1,7)=6.9$, $\alpha=.034$; 47.1 - 58.4 cm) and the steering activity ($F(1,7)=45.5$, $\alpha=.006$; 42.2 - 45.8 °/s). Both performance measures indicate performance decrease in the conditions with camera view.

In the lane change task we found an effect of camera on the course instability ($F(1,7)=44.2$, $\alpha=.001$; 20.1 - 17.1 cm/s), and the mean absolute steering wheel speed ($F(1,7)=16.8$, $\alpha=.005$; 44.4 °/s - 47.3). In the speed estimation task we found an effect of field of view on course stability ($F(1,7)=24.7$, $\alpha=.002$; 10.2 - 13.9 cm/s). In the speed estimation task we found an effect on the estimation error ($F(1,7)=51.1$, $\alpha=.000$; +14.0 - -4.3%). These results show a beneficial effect of wide field of view in the sharp curves task, although it required higher steering activity. This beneficial effect is probably caused by the better lateral view and the presence of vehicle references. In the lane change task a favourable effect of normal field of view is present. This is due to the unfavourable effect of underestimation of lateral speed with minification 0.5 with 100° field of view, while not needing better lateral view. The effect of field of view on the speed estimation task is discussed below.

The interaction camera position $\times$ field of view was significant for the mean lateral distance in the sharp curves task ($F(1,7)=16.0$, $\alpha=.005$), see Figure 4.
overestimation of speed, not significantly dependent on position. These effects are in accordance with the idea that optic flow in the image contributes to speed perception (e.g., Van der Horst, 1991). At normal field, the main flow comes from the nearest road structures; which decreases at camera position high. At wide field the image minification decreases the flow from the road, but there is more flow from peripheral structures (immediately near the course was woodland), both at low and high camera position. These results suggest that the peripheral flow is a more effective speed cue than the flow from the road. This is in accordance with the results reported by Salvatore (1968). In his experiments subjects had to estimate their speed with only peripheral, or only foveal, view on a further empty highway. In the peripheral view conditions, estimated speed was higher and closer to the target speed. Effects of field of view as found in our experiments were also reported previously, for example see Osaka (1988).

Subjective difficulty and subjects' strategy
All statistical tests are executed on the z-scores calculated per individual. However, ratings of subjective difficulty are presented in terms of mean raw scores (scale 1 (no problem) - 5 (almost impossible to perform)).

In the sharp curves and lane change task, there is a significant effect of camera position and the spatial orientation aids show no main effect. In two tasks there is an effect of field of view. In the sharp curves (normal - wide: 2.3 - 1.9°) subjects judge the wide field conditions easier, probably because of the better lateral view and better reference. In the lane change task (normal - wide: 1.9 - 2.3°) wide field of view is judged harder, probably because of minification of the lens, while not needing better lateral view. These effects are similar to the results of the performance measures.

The strategies subjects said to have used during completion of the tasks are summarized below per task. The results show that several strategies are mentioned across different tasks. This includes using a part of the car as a reference (ie. the roof, lines on the bonnet, and the side mirrors); using 'intuition'; using a fixed place on the monitor (ie. the right corner, the lines in the middle of the monitor indicating the place where the monimarks had to be attached); and finally using the monimark (ie. the lines indicating the tracks on the straight parts, and distance markings in curves).
In the sharp curves task, it is clear that subjects use a point of reference for lateral position whenever possible. This is often a fixed point either on the camera image of the car or on the monitor. Subjects seem to prefer reference points on the car's image. Only when there are no such points (as in the normal conditions and in curves) subjects make use of fixed points on the monitor too. A similar pattern is found in the use of intuition: mainly in the normal conditions and in curves. It seems that this is a kind of compensation for the restricted view on the road. Two other points are worth noticing. Firstly the extensive use of the monimark (if present), although this seems contrary to the absence of effects in the objective measures. And secondly the use of the markings on the outside side of the curve only in the normal conditions, the conditions in which the markings can disappear out of sight in the sharp curves.

For the lane change task the use of (a part of) the car as a reference is about the same as in the eight-shaped circuit and the sharp curves: subjects use it whenever possible. Only when points of reference on the car are not visible in the monitor image, subjects use fixed points on the monitor. Worth mentioning is the fact that in the high - normal conditions the side mirror is sometimes visible on the edge of the image. This was due to variability in the adjustment of the camera angle during the one hour run. Two subjects said to have used this. Again the monimark (if present) is extensively used, and again this seems in contrast with the objective measures.

It seems that in the speed estimation task only half of the subjects made use of visual cues. Auditive cues were used far more often, which may account for the fact that there is no difference between the viewing conditions with respect to the used strategies.

Conclusions
Differences between camera view and direct view are evident for several performance measures and tasks. If present, these differences generally show that control of lateral position and course is moderately worse with camera view: lateral distance to road marking and course instability are up to 50% larger. Given the subjects' lack of experience with camera driving, with restricted field of view, different viewing geometries, and absence of stereopsis, this is not surprising and found in other experiments on RPV-control (for example Spain, 1991). These findings indicate that driving on flat roads without traffic, is indeed performable with camera view.

Spatial orientation aids show no improvements in performance and subjective difficulty. This is against expectation, because the markings were designed to provide cues for lateral position, course and distance. Probably, the limited instruction runs were not sufficient to teach proper use of the markings. Contrary to the modest effects on performance and subjective difficulty, subjects extensively used the monimark to complete most tasks. From the overview of the strategies subjects used, it is clear that (fixed) reference points on the car's image, instead of on the monitor, are preferred. This plea for placing special reference points on the vehicle itself.

Main effect of camera position is evident in only one instance. More evident are effects of field size and combined effects of position and field. However, effects are not quite consistent for each task and dependent variable. It is clear that preferable camera conditions depend for a great deal on the task characteristics. An interesting option for future research is to make use of a lens with variable field of view, adjusted to the subjects' comfort in a particular task and particular camera position.

EXPERIMENT 2
Experiment 2 was a replication of experiment 1, but conducted in the TNO - HFRI fixed-based simulator. Therefore experiment 2 may be seen as a simulation of controlling a Unmanned Ground Vehicle. Simulated direct view was by means of projecting colour images onto a cylindrical screen, simulated camera view was by means of a black and white monitor mounted above the steering wheel. Most important differences with experiment 1 are the lack of mechanical motion information, and the lack of stereoscopic viewing in the direct view conditions.

By comparison with experiment 1, spatial orientation aids is left out as camera factor, and magnification 1.0 for the wide field of view conditions is introduced.

Methods
A three channel Evans & Sutherland ESI 2000 image generator was used for the experiments. For the direct view conditions the image was projected on a cylindrical screen (radius 3.75 m) with the subject positioned in the centre. Total field size was 120° × 40°.
The image was projected by three projectors (BARCO graphics 800) with a resolution of 1024 x 1024 pixels each and a refresh rate of 60 Hz. The dynamic vehicle model was based on the characteristics of the car used in the field experiment, including automatic gear (Godthelp, Blaauw & Van der Horst, 1982). For fixed speed tasks the simulator limited the speed irrespective of subjects pushing the accelerator pedal further down. To eliminate the use of sound cues in the speed estimation task the simulated sound of the engine could be switched off.

The TNO driving simulator mockup used for the experiment is the front part of a VOLVO 200-serie, including the two front seats and the bonnet. An emblem, which was exactly the same as in the field experiment, was positioned at the front end of the bonnet. Steering wheel, accelerator pedal, and braking pedal were the same as in the original vehicle. The mockup had a two-way intercom to the control room, in which the instructor sat. The instructor had at his disposal: information about the status of the simulator, information about the behaviour of the subject, a monitor with the image presented to the subject, and a monitor with an overview of the simulator room.

For the camera view conditions a monitor (Mitsubishi colour display monitor, HL7955SBK) was placed in the mock-up, with the cylindrical screen left blank. Viewing distance, screen size, and aiming could easily be adjusted to the specific conditions. The monitor was placed directly above the steering wheel, and at a right angle with the line from eye to the image of the horizon. The subjects' head was supported by a head rest. Refresh rate of the monitor was 60 Hz, resolution 1024 x 1024 pix. Corresponding the field experiment, the colour monitor was used as a black and white monitor.

The camera factor position was varied by changing the view point of the simulator in accordance to the positions used in the field experiment. Camera factors field of view and magnification factor were varied by changing the viewing angle of the simulator, the viewing distance and the effective monitor size.

Terrain, primary measures, tasks, instructions, dependent variables, and training were equivalent to those used in experiment 1. Subjects were military driving instructors with normal vision and driving experience of at least 150,000 km.

The design used differs from the design of experiment 1. Again first and last run were with direct view, but in between were only six camera conditions. Again three camera factors: camera position (low at the front, high at the back), field of view (50°, 100° diagonal), and magnification factor: 0.5 (only for 100° field of view), and 1.0 (both for 50° and 100° field of view).

Results

The results of Experiment 2 are presented below per effect.

Direct view vs. camera view

In the CAMERA design (means in the order direct view - camera view) we found an effect in the sharp curves task of both lateral speed (F(1,7) = 8.56, α = .022; 31.66 - 39.00 cm/s), and steering activity (F(1,7)=26.03, α = .001; 47.20 - 55.06 °/s), which indicates better performance in the direct view conditions. This means that course stability and steering performance are moderately degraded (about 20%) in camera view. Mean lateral distance to the right side markings is not significantly different in direct view or camera view.

In the lane change task we found an effect on the lateral speed (F(1,7) = 16.82, α = .005; 5.83 - 4.74 cm/s), the steering activity (F(1,7)= 21.00, α = .002; 29.58 - 41.94 °/s), and on the standard error from midlane (F(1,7)= 4.41, α = .074; 13.84 - 8.82 cm). Lateral speed and error from midlane show a more stable course and a better overall performance in the camera view conditions compared to direct view. The advantage of the camera in the lane change task is the fact that it was always positioned over the longitudinal midline of the car. This results in a symmetrical camera image when the vehicle is exactly positioned in the middle of a lane. Furthermore, possible points of reference (for example the bonnet or the emblem) become more useful to indicate the exact middle of the lanes. On the basis of these advantages it may be expected that with camera view subjects can more accurately determine the lateral position of the vehicle in a lane, which will lead to a more stable course and a better overall performance. The higher score on steering activity in the camera view conditions indicates that the better performance in these conditions required more steering activity.

In the speed estimation task we found an effect on the error from target speed (F(1,7)=9.01, α = .019), overall mean for error...
from target speed in the direct view conditions is \(+3.2\%\) (which is equivalent with 51.6 km/h), and in the camera view conditions \(+19.8\%\) (equivalent with 59.9 km/h). This means a relative underestimation of speed in the camera view conditions.

The cause of the slight (absolute) underestimation in the direct view conditions is not exactly known, but was reported before (for example Van Erp (1995); Padmos & Van Erp, 1997). The possibility of an internal representation of the target speed which is not equal to 50 km/h may be an explanation. After all, subjects got only a short training which consisted of three trials with feedback. Because of this minimal training it is more meaningful to look at relative effects only.

**Camera factors**

In the camera factors design we found that mean lateral distance in the sharp curves task is somewhat larger in the high camera position \((F(1,7)=6.76, \alpha=.035; \text{low: 97.12 - high: 111.02 cm})\). It may be expected that in this position lateral distance would be underestimated, mainly because of the minification of lateral distance compared to position low. In the speed estimation there was a trend towards underestimation of speed with a high camera position \((F(1,7)=5.24, \alpha=.056; \text{16.03 - 23.63 \%})\). In this camera position the visual flow from the road is less compared to camera position low. Underestimation of speed in position high is therefore in accordance with the hypothesis that visual flow in the image contributes to speed perception (Van der Horst, 1991). The fact that this effect does not reach significance is in accordance with the small number of subjects that explicitly mentioned watching the road as a strategy to estimate speed, which indicates that flow from the road is only a weak cue.

Field size (means in the order 50° - 100°) showed some significant results. In the sharp curves task, performance on mean lateral distance \((F(1,7)=37.10, \alpha=.000; 107.23 - 77.49 \text{ cm})\) and lateral speed \((F(1,7)=13.46, \alpha=.008; 27.44 - 36.79 \text{ cm/s})\) is better in the wide field conditions. This indicates that both steering performance and determining lateral position are worse in the 0.5 conditions, up to a factor 6 for error from midlane. This indicates that both steering performance and determining lateral position are worse in the 0.5 conditions. Magnification of 0.5 will lead to underestimation of lateral distance and larger course instability. This qualitatively confirms Schulz-Helbach, Donges and Rothbauer (1973) who found that a magnification of 0.4 increased the standard deviation of lateral position on a straight road with 30 \%. The effect on steering activity is rather small.

In the speed estimation task, substantial differences are present as well, on the course instability \((F(1,7)=4.69, \alpha=.067; 4.19 - 5.42 \text{ cm/s})\), the steering activity \((F(1,7)=23.88, \alpha=.002; 36.79 - 53.77 \text{ \degree/s})\), and on the standard error from midlane \((F(1,7)=23.88, \alpha=.002; 2.98 - 18.01 \text{ cm})\). All effects show worse performance in the 0.5 conditions, up to a factor 6 for error from midlane. This indicates that both steering performance and determining lateral position are worse in the 0.5 conditions. Magnification of 0.5 will lead to underestimation of lateral distance, and to reduced lateral flow in the image compared to magnification 1.0.

In the speed estimation task, magnification has an effect on the percentage error from target speed \((F(1,7)=13.70, \alpha=.008; 26.50 - 5.76 \%)\), which means a relative speed
overestimation in the conditions with magnification 0.5. As stated before there are only few experiments in which magnification factor is systematically varied. Unfortunately, confounding of magnification and field size seems to be common in many experiments. However, Evans (1970) reported a similar effect of magnification factor. In this experiment magnification was accomplished by changing the viewing distance to the projection screen, so magnification was not confounded with field size.

Subjective difficulty score
The camera view conditions are moderately more difficult than the direct view conditions in the sharp curves task (significance level indicated as follows: ⋆⋆ means α < .01, ⋆ means .05 < α < .01, t means .10 < α < .05): (direct view - camera view: 1.47 - 1.80⋆), and the lane change task (1.75 - 2.52⋆⋆). Position reaches significance on none of the individual tasks, nor over all tasks. There is only a trend in the lane change task (low - high: 2.29 - 2.75⋆). There are no effects of field size for any of the tasks, nor over all tasks. All Three tasks show an effect of magnification on subjective difficulty: the sharp curves task (1.0 - 0.5: 1.73 - 1.94⋆), the lane change task (2.10 - 3.38⋆⋆), and the speed estimation task (1.97 - 2.25⋆). All effects indicate larger subjective difficulty in the 0.5 conditions.

Strategies
The results in the sharp curves task confirm an important finding of the field experiment: subjects use points of reference whenever possible. Again points of reference on the depicted car seem to be the most useful. Extensive use of a point of reference on the monitor is only made in the normal field of view conditions in which no part of the car is visible. These last conditions also don’t seem to allow looking further ahead, a strategy which was used by some of the subjects in the other conditions. Remarkable is the fact that in the present experiment in the camera view conditions the car was simulated by attaching a passe-partout on the monitor screen. This simulation of the visible parts of the car seems to be as useful to the subjects as the real image of the car.

In the lane change task, the results point in the same direction as the results of the sharp curves task. Subjects prefer points of reference in the image (when present). In the normal conditions, in which there are no such points, subjects try to hold the cones symmetrical, which is actual using the monitor as point of reference. In this task too, the normal conditions do not seem to invite subjects to look further ahead.

In the speed estimation task, actually the only used cue is watching the trees. This is according to the hypothesis that peripheral view contributes to speed perception. There doesn’t seem to be a difference between different levels of field size and magnification factor. Using structures from the road is hardly mentioned. Note that the sound of the engine was switched off in this task, and that changing of gears did not occur. These were two important cues in the field experiment.

Conclusions
Experiment 2 focused at two questions. The first question concerned the magnitude and character of differences in driving performance when outside view is provided by a camera-monitor system. Experiment 1 showed moderate performance degradation in driving with camera view. However, in this field experiment full mechanical motion information was available, which is only relevant in driving under armour. In operating an Unmanned Ground Vehicle such information is absent. This absence may cause additional performance loss in driving with camera view, which makes this study relevant. The second question was concerned with the effects of manipulating the following camera factors: position and aiming, field size, and magnification factor.

Simulated direct view vs simulated camera view
All tasks showed one or more differences between simulated direct view and simulated camera view. In the sharp curves task all measures show a moderate advantage of the simulated direct view conditions. However, this advantage was not evident or present in the other tasks; sometimes performance in the camera view conditions even seems to be better (for example in the lane change task). On the basis of the results of all tasks, it may be concluded that, on flat paved terrain without other traffic, driving performance in a simulated UGV is adequately compared to simulated direct view, provided that the camera factors are well-chosen.

Camera factors
Of the investigated camera factors, position appeared less critical than field size and magnification. Important characteristic of the present experiment is the fact that field size and magnification were varied independently. In
UGV's studies, both factors are usually confounded. Wide field size generally leads to better performance. Performance improvements with wide field size were substantial for taking sharp curves. Magnification appeared an important camera factor. The results showed that magnification of 0.5 of the outside world may lead to performance deterioration in taking sharp curves and lane change, but decreases speed underestimation caused by camera view. In experiment 1, where field size and magnification factor were confounded, the explanation for the relative speed overestimation in the wide field conditions was that of extended (peripheral) visual flow. The results of the present experiment seem to contradict this explanation completely: when field size and magnification are independently varied, field size seems to have no effect at all on the estimation of speed! The present experiment indicates that varying field size in the range 50° - 100° does not influence the estimation of speed.

Roscoe (1984) already suggested that there is an optimum magnification for every imaging system. Present results indicate that an optimum magnification is at least 1.0. On the basis of the results it is expected that magnification larger than 1.0 will not give substantially better performance than magnification 1.0.

FIELD VS. SIMULATOR EXPERIMENT
To draw a comparison between experiments 1 and 2, the raw data was re-analyzed using the same design for both data sets. This means that conditions with monimark present and wide field combined with magnification 1.0 were excluded, and that field of view and magnification are confounded. The results of this analysis show two interesting incongruencies, the effect of field of view in the sharp curves task, and the effect of camera view in the lane change task.

The effects of field size on mean lateral distance and lateral speed in the sharp curves task do not correspond between both experiments. Wide field results in better performance in the field experiment, but in worse performance in the simulator experiment. There are two factors involved which may each account for one of the effects. The wide field gives a better lateral view and visible reference points, which may lead to better performance on mean lateral distance and lateral speed. This seems to be a plausible explanation for the result of the field experiment. On the other hand, wide field also results in a magnification of 0.5, which causes underestimation of lateral distance (and lateral speed), and thus worse performance on mean lateral distance and lateral speed. This seems to be a plausible explanation for the result of the simulator experiment. The fact that the underestimation of lateral speed can not nullify the positive effect of better lateral view on lateral speed in the field experiment may be due to the presence of mechanical motion information. This cue for lateral speed was not present in the simulator.

Note that the suggested effects of field size and magnification are in fact demonstrated in the simulator experiment when field size and magnification factor were not confounded.

The second discrepancy between both experiments is concerned with the effect of camera view in the lane change task. In the field experiment course stability is better in the direct view conditions, in the simulator this effect is in the opposite direction. The reason for this worse performance in the field experiment direct view condition is probably the lack of binocular depth perception, while not having the advantage of a symmetrical image presentation.

GENERAL CONCLUSIONS
The driving with camera view experiments have shown that driving performance with a camera-monitor system is indeed possible without dramatic performance loss. Hitherto, the project has studied four parameters of the system: placement and aiming of the camera, additional spatial information, field size, and magnification factor. Below, main results from the previous conclusions are summarized, supplemented with miscellaneous observations.

In both experiments a modest performance degradation with camera view is found on some of the performance measures. However, the results indicate that driving with camera view on flat, paved terrain, without other traffic is possible. Of the investigated camera factors, spatial orientation aids attached to the monitor do not sort any effects, and camera position and aiming is of minor importance. Field of view and magnification factor often show substantial effects, in favour of wide field of view (100°) and magnification 1.0. Of the investigated camera conditions, best overall score is with the camera mounted at a high position at the back of the
vehicle, with field size 100° diagonal and magnification 1.0. From a comparison between the field and simulator experiment, it seems that underestimation of lateral speed may be lessened by correct mechanical motion cues.

Subjective difficulty and Strategies
The subjective difficulty scores follow the objective scores. Most important conclusion concerning the strategies used to perform the tasks is that subjects benefit from points of reference. Points in the camera image, like the bonnet, are preferred above points on the monitor itself. This pleads for a camera placement and aiming such that the vehicle is visible in the camera image. This may be completed with the installation of special markings on strategic places on the vehicle. For example markings on the front end of the vehicle which indicate the width and thus help to perceive the lateral position of the vehicle.

In the normal field size conditions, subjects hardly mentioned looking further ahead. This is remarkable, because, compared to the wide field conditions, there is no real difference with respect to the position of the horizon on the monitor.

Training camera view driving
In the present experiment, there was only a very brief familiarization period for each camera condition. It is expected that more extensive training for the preferred camera condition may improve performance considerably, enabling pupils to adapt to the particular viewing geometry, which differs considerably from direct view. Expected is that with more extensive training the performance in the best camera conditions can be as good as direct view for all tasks and measures, for driving on flat roads without traffic.

Observed practical problems
During the field experiment, three practical problems were demonstrated. Some subjects showed moderate signs of motion sickness after the first run with camera view. In the course of one or two subsequent camera runs symptoms disappeared. Apparently, the discrepancies present between visual and vestibular stimuli, which were relatively small compared to the visuo-vestibular discrepancies encountered in simulators or in remote control stations, are able to evoke motion sickness. Second problem was glare from the sun and the surrounding sky, which gave a considerable image deterioration when the elevation of the sun was lower than 15°. This notwithstanding anti-glare measures like a hood of 25 cm depth above the camera lens, and a downward pitch of the camera of 5° for the normal field and 11° for the wide field. Glare from sun reflections from a wet road decreased visibility considerably too. Technical solutions to the glare problem are a prerequisite if camera view is to be used in a variety of practice conditions. Third observed problem is lens pollution. In one instance the camera lens was covered with dew, which ruined the camera image. This illustrates that, especially for battlefield situations, the camera must be shielded by a window with an appropriate cleaning installation.

References


Nagata, M. & Kuriyama, H. (1984). Drivers’ visual behavior under conditions of negoti-

ating obstructions. SAE-Technical Paper 840550.


Roscoe, S.N. (1979). When day is done and shadows fall, we miss the airport most of all. Human Factors, 21 (6), 721-731.


Salvatore, S. (1968). The estimation of vehicu-

lar velocity as a function of visual stimula-


Spain, E.H. (1991). Objective assessments of mobility with an early unmanned ground vehicle (UGV) prototype viewing system,
presented at DRG seminar 6-8 March 1991.
Naval Ocean Systems Center - Hawaii Laboratory, Kailua Hawaii.

Road user behaviour: theory and research. 
Van Gorcum, Assen.


VISUAL SUPPORT IN OPERATING UNMANNED AERIAL VEHICLES

Jan B.F. van Erp, B. Kappé & J.E. Korteling
TNO Human Factors Research Institute
Soesterberg, The Netherlands.

ABSTRACT
Most important source of information for operators of Unmanned Aerial Vehicles in controlling airframe and camera are the outside world images from the on-board camera. However, these images are of degraded quality, i.e. limited field of view, limited spatial resolution, and low update rate, which leads to poor operator performance and loss of situational awareness.

An important operator task is controlling the camera, which requires information on airframe heading and speed, and camera heading and pitch. This information may be displayed by digital or pictorial indicators. The TNO Human Factors Research Institute developed a new principle of presentation: a view on a computer generated Environment (CQE). This view is directly coupled to the actual position and orientation of the camera, and may be depicted with arbitrary field of view, zoomfactor, and update rate. For example zoomed-out around a zoomed-in camera image to improve situational awareness, or with high update rates superposed over a slow updated camera image to improve tracking performance.

The CQE (for example a grid positioned at sea-level) provides a caricature of the optic array that is equivalent to the optic array that would have been observed on-board the airframe. We call this ecologically compatible. This is fundamentally different from a presentation by indicators which are per definition an abstraction of reality, and require transformations to create a sense of camera and airframe state. An ecological compatible display allows the use of elementary information processing mechanisms normally used for processing information on layout of the environment. Obviously, utilizing such mechanisms will result in better performance and reduced workload compared with the traditional methods of operator support. This claim is confirmed by experiments on situational awareness.

INTRODUCTION
In modern warfare, success and failure become increasingly dependent on technological developments and applications of intelligent and (semi) autonomous systems such as unmanned vehicles. In this connection, the Royal Netherlands Navy is especially interested in Maritime Unmanned Aerial Vehicles (MUAV's). Potentially, these systems may contribute to many tasks, such as intelligence, target acquisition, battle-damage assessment, communication relays, radar observation, etc.

Technological constraints
However, several technological problems have to be solved before these goals can be reached. From a human-factors point of view, one of the main problems is related to the images provided by the imaging devices, the most common MUAV payload. These imaging devices, e.g., video or infra-red cameras, have two principal limitations: a low update-rate and a limited and zoomed-in field of view. These limitations may become critical when the operator would need manual control of MUAV and/or camera movements, for instance when searching for targets or in battle damage assessment (for an overview see: Michelson & Patton, 1988; Van Brda & Passenier, 1993; Eisen en Passenier, 1991).

The low update rate of the payload image is mainly due the restricted capacity of the data link between MUAV and operator. The current state of technology allows a maximum update rate of 4 Hz, combined with low spatial resolution, limited field of view, and image compression-decompression (Bombardier, 1990), NATO, 1990).

However, even with an optimal update frequency, the camera's field of view is limited. Due to the large minimum following distance (2000 m, NATO (1990)), the camera needs to zoom-in on targets, resulting in a recorded field of view of only a few degrees. Such limited field of view does not allow the operator to develop a sense of spatial awareness, i.e., knowing the position and direction of translation of the MUAV, knowing the viewing direction of the camera with respect to the orientation of the MUAV, or knowing the position of the different sites of interest with respect to the MUAV (Van Erp, Korteling & Kappé, 1995; Van Erp, Kappé & Korteling, 1996; Korteling & Van der Borg, 1994).

Operator tasks
A MUAV typically is controlled by human operators at a remote location. This crew performs a number of basic functions which are common to most UAV missions, i.e., mission planning, navigation and platform control, payload (sensor) control, data analysis, launching and recovery, and communication. Depending on the kind of missions, mission
payload, and task-allocation strategies these functions can be divided over one or several operators. In many existing systems these functions are carried out by at least three operators (Denaro, Kalafus and Ciganer, 1989). In case of a single operator and missions utilizing imaging payloads, primary operator tasks of interest are: detection, recognition, designation, and tracking. These tasks are complicated, partly because of the numerous degrees of freedom of the system: altitude, horizontal position, and yaw of the MUAV, and pitch, yaw, and field of view of the camera. Combined with constrains on spatial resolution and update rate (due to the restricted bandwidth), and constraints on the field of view this may result in serious problems for the operator. Previous research shows that target acquisition and tracking are the most critical tasks. Reported problems concerning operation of a MUAV, and interpretation of the camera image (e.g. Korteling & Van der Borg, 1994; Korteling & Van Breda, 1994) include: - loss of situational awareness, and loss of the ability to discriminate between movements of the MUAV and movements of the camera, - degraded perception of object- and ego-motion, - disorientation because of the absence of direct visual feedback, and relative independence between the viewing direction and vehicle movements, - degraded perceptual anticipation (Poulton, 1974), due to the low quality of the payload image, - difficulties in operating and interpreting of the payload image due to restricted field, low resolution, and low update rate.

Synthetic image augmentation

One promising way of solving situational awareness problems and improving perceptual capabilities of camera operators is synthetic image augmentation. Recent studies (Chavand, Collie, Gallard, Mallem, & Stomboni, 1988; Mestre, Savoyant, Péruch & Pailhous, 1990) focused on spatial orientation problems in remote control situations. These problems arise by the fact that it is difficult for stationary operators, using a visual display as the main source of spatial information, to convert the visual transformations on the display, generated by the combined effects of platform and camera motion, into the proprioception (literally: self-perception) of translations and rotations of the platform and the rotations (viewing direction) of the camera. This lack of situational awareness may be compensated by providing augmented visual proprioceptive information concerning platform and camera motions. This would especially be helpful in missions with poor visual feedback (flying above sea, hazy weather) and/or when the camera operator has to fly a UAV by himself. A simple and well-known solution in such situations would be the presentation of an extra display on the status screen presenting indirect orientation information concerning platform and camera motion. An example of a display presenting indirect orientation information is a so called Horizontal Situation Indicator (HSI), or Combined Heading and Pitch indicator (CHPI). These indicators present in a compass like manner the viewing direction of the camera, the moving direction of the platform, and geographical north. In the present experiments such indicators are used as a starting-point.

A potential drawback of this solution is that this extra display requires the operator to combine artificial and rather abstract orientation information with straightforward, concrete environmental information. Therefore a more direct solution to this disorientation problem may be provided by generating synthetic visual cues which are stationary relative to the world or to a target. A simple version of this was used by Agin et al. (1980), who added small symbols to the payload image, one for the movements of the image centre in real time, and one for the position of the image centre in the next frame. Since self-motion is normally specified by global visual transformations, we believe that a synthetic visual grid (pattern of lines) will be more suitable for enhancing spatial situational awareness.

Computer generated environments

Such a Computer Generated Grid may be supposed to generate clear perspective optical transformations, termed optic flow or motion perspective. According to Gibson (1950, 1966, 1979) this kind of visual information is directly perceived and continuously available and utilized to control self-motion. Korteling (1994) argues that the nervous system is very efficient in combining this kind of information with other sources of environmental information. Accordingly, flow information does not pose any extra demands upon presumed limitations in information processing capacity, which are often merely due to a lack of information, rather than to a surplus. Optic flow transformations can be analyzed in a few basic components. These components independently specify UAV translations and UAV or camera rotations, whether or not in combination (e.g., Koende-
rink, 1986; Kappe & Korteling, 1995), and thereby enhance the separate perception of platform and camera attitude in an environment. When such a grid is positioned at a certain height above the ground, it may also be an aid in the perception of ground speed and altitude.

One of the most powerful advantages of this augmented synthetical information is that it can be presented with a high update frequency, e.g. 30-60 Hz without any need for extra datalink capacity or platform payload (the CGG is generated at the control station, using current MUAV status and attitude (updated at a high frequency), and a MUAV model). Hence, the most significant part of the proprioceptive motion and camera information is continuously available to the operator.

It is expected that synthetic image augmentation by a grid will reduce problems related to object- and ego-motion and thereby enhance situational awareness. The present study consists of two experiments to test the benefits of a CGE on situational awareness.

EXPERIMENT I

In a previous experiment we tested the efficiency of the CGG currently used in a tracking task, in which subjects had to track a moving ship from a moving MUAV under different update frequencies of the camera image (see Figure 1). A very significant positive effect of a CGG (consisting of parallel and perpendicular lines at sea level) on the RMS tracking error was found, in particular at lower update frequencies (factor 2 at 0.5 Hz).

Experiment I was designed to test the potential beneficial effects of this CGG on the subjects’ ability to reposition the camera at an initial position. This is an important aspect of the camera operators task, for example to relocate lost targets, or when operator wants to point at an other location and return to the original target shortly afterwards). Problems with the apprehension of spatial information (i.e., the perception of orientation and position of the aircraft and/or the sensor in space) are supposed to be reinforced by low update frequencies of the monitor image. Repositioning the camera may benefit from pictorial status information as displayed by a Horizontal Situation Indicator (HSI) presenting geographical north, viewing direction, and flying direction. Therefore, the experiment was designed such that the effects of update frequency of the monitor image, and presence of a CGE and HSI could be accounted for.

Methods

Subjects

Eight paid male subjects (age 21 - 35 years, mean 26.3) participated in the experiment. All subjects had normal or corrected to normal vision, and some experience with similar tracking tasks.

Mock-up and instrumentation

The experiment was conducted in the TNO-HFRI RPV Research Simulator. This facility (see Figure 2) is especially designed for simulating unmanned missions (Korteling & Van Breda, 1994). The subject was seated in a chair in the middle of the operator table. The chair could be adjusted to personal comfort. The only control needed was a joystick (square type, RS type 162-732) placed on the operator table at a comfortable distance for the right or left hand. Joystick deflections resulted in changes of camera rotation and/or pitch (left and right deflections: horizontal rotations, fore- and backward deflections: pitch). No deflection (2% range) resulted in a stationary camera. The monitor (Mitsubishi colour display monitor HL7955SBK) was placed at eye height at a distance of approximately 60 cm. All images and displays were generated by an EVANS & SUTHERLAND ESIG 2000 (600 × 800 pix.).

Further instrumentation consisted of computers for scenario and data storage (4 Hz sampling frequency), and for supervisory functions.
Figure 2. Side view of the TNO - HFRI RPV research simulator.

Experimental task
Subjects had to watch a computer generated image for 15 s. The image was a simulation of the camera image of a MUAV flying above an empty sea. At the beginning of each trial an image was presented which contained a target ship (fixed position, in the middle of the screen, target distance 6000 m, vertical viewing angle 11.3°). The ship disappeared two seconds after the onset of the trial. Directly after the beginning of a trial the MUAV started to move (translation and rotation), and the camera started to rotate. All movements were unknown to the subject. After 15 s the image would freeze. The subjects task was first to move the camera (by means of a joystick) in such a way that it was pointed as accurately as possible at the location at which the ship had disappeared. After confirmation by pushing a button, the subject had to indicate the horizontal angle between the initial position of the target ship and the viewing direction of the camera at the moment the image froze. The camera error was calculated as the square root of the sum of the squared pitch error and heading error. The mean camera error was calculated over the six scenarios in each condition.

Independent variables
Three factors were included in the experiment in order to test the effects of the synthetic image augmentation by the more traditional HSI, by a CGE, and update frequency. On the basis of previous experiments it is expected that update frequencies below 4 Hz are most critical, leading to the following factors and levels:
1. Augmentation by means of a head-up HSI
   - present
   - absent
2. Augmentation by means of a Computer Generated Environment
   - present
   - absent
3. Update frequency of payload image:
   - 0.5 Hz.
   - 2 Hz.
   - 4 Hz.
All independent variables were varied within subjects, thus leading to 12 conditions for each subject.

Dependent variables
Two dependent variables were calculated:
- mean camera error
- mean reported heading error.
The camera error was the difference between the position of the target and the position of the camera footprint (resulting from heading and pitch) set by the subject in order to point at the target (both in spherical coordinates). The camera error was calculated as the square root of the sum of the squared pitch error and heading error. The mean camera error was calculated over the six scenarios in each condition.

The reported (horizontal) heading error was the absolute difference (in deg.) between the actual heading of the target relative to the final heading of the camera, and the subjects' verbal report of the heading of the target. The mean reported heading error was calculated over the six scenarios.

Statistical design
Conditions were order balanced across subjects (Latin square; Wagenaar, 1967). In every condition the subjects had to perform six scenario's. To exclude effects of scenario difficulty, every subject performed the same six scenario's (order balanced) in every condition. All measures were analyzed by a 8 (subject) x 2 (HSI) x 2 (CGE) x 3 (update frequency) ANOVA with the package STATISTICA®.

Procedures
The subjects first received written instructions concerning tasks, conditions, and procedures which they had to read carefully. Subjects worked in pairs; one subject completed three conditions consecutively while the other one
rested. After a short training, one scenario in each condition, the experiment began. The instructor always explained the oncoming condition to the subject. When the subject was ready the instructor started the first scenario. The subject watched the camera image, which would freeze after 15 s (indicated by a beep). Subsequently the subject could reposition the camera in the position of the initial position of the target ship. When the subject was satisfied with the aiming of the camera he confirmed this by pushing a button. Finally, the subject had to indicate the horizontal angle between the position of the target ship and the viewing direction at the end of the 15 s scenario. After writing this down and reporting it to the instructor, the next scenario would start. One condition consisted of 6 scenario's.

Results
During the experiment, one subject indicated to have problems with the camera repositioning task, in spite of extensive additional instruction. In line with this, his mean camera repositioning error over all conditions differed more than 2 standard deviation from the overall mean camera error. Therefore this subject was excluded from further analysis.

![Figure 3. Mean camera repositioning error as effected by update frequency and image augmentation.](image)

Mean camera error data did not show an effect of the HSI; the subjects did not seem to use the information on the camera's viewing direction and the heading of the MUAV that was depicted in the HSI display ($F(1,6) = 1.62$, n.s.). The mean camera error was reduced when the visual information was augmented by means of the CGE ($F(1, 6) = 12.7$, $p = 0.012$). As can be seen in Figure 3, the subjects were more accurate in repositioning the camera when the CGE was presented in the camera image. This figure also shows the effect of update frequency ($F(2, 12) = 3.9$, $p = 0.048$), the subjects had a higher mean camera error for the lower update frequencies. Figure 3 shows that image augmentation by means of a CGE tended to interact with update frequency. Probably due to the high variance of the camera error data, this interaction just failed to reach significance ($F(2, 12) = 3.4$, $p = 0.067$), but it does show that the presence of the CGE resulted in a low mean camera error irrespective of the update frequency (Post hoc Tukey test shows significant differences between no-CGE 0.5Hz. and no-CGE 2Hz. vs. all the CGE conditions and between no-CGE 0.5Hz. and no-CGE 4Hz.).

With respect to the second dependent variable, providing a verbal indication of the heading of the ship relative to the viewing direction of the camera, the subjects reported severe difficulties. Most subjects were very uncertain about their answers and substantially overestimated the target's heading (about 10 to 20 times). In this task, all subjects were included in the analysis. In line with the reported difficulty of the task and the resulting high rate of error variance, the results did not show a significant effect on any of the independent variables.

Discussion
Experiment I was designed to test the potential beneficial effects of two forms of synthetic image augmentation on the subject's spatial awareness under different conditions of update frequency of the payload image. Upon an imposed motion pattern of the platform/camera the subjects had to reposition the camera in order to look in the direction of a previous visible target and to mention the direction of this target relative to the heading of the camera.

The presence of the horizontal situation indicator (HSI) did not show any effect on camera repositioning error. This might be explained by the fact that the interpretation of this display and the translation of the information into visuo-spatial knowledge or spatial awareness, may require more training than was provided during the experiment. This explanation is substantiated by the finding that subjects did not base their verbal reports concerning the heading of the target relative to the viewing direction of the camera on what they could read simply from the rude compass scale of the HSI. This means that subject did not use this abstract information, even not in
conditions of highly degraded visual motion information.

Improved camera repositioning performance with image augmentation indicates that, as expected, the grid aided subjects to continuously and directly perceive orientation and position of the aircraft and/or the sensor in space. In addition, the positive effect of the CGE may be explained by the fact that the subjects could use a simple heuristic to find the original position of the target. The CGE enabled subjects to monitor the movements of camera and MUAV by observing the horizontal and vertical lines that passed (the centre circle of) the display. Consequently, the subjects could find the initial position of the target by reproducing these line passages in the opposite direction. That the subjects used this heuristic, is substantiated by the finding that the verbal reports on the targets heading did not show any effect of the presence of the CGE. Overestimation of the camera rotation, and consequently the poor awareness of 'compass space', is probably caused by the fact that the subjects were viewing a zoomed-in camera image. A zoomed-in camera image increases the magnitude of translational optic flow generated by camera rotations, resulting in overestimation of camera rotation. This is cause to adjust the CGG used in experiment I.

**Improving situational awareness**

The current experiment was designed to explore the effects of an improved CGG on the operators target search performance, and to compare its performance with a more traditional method of operator support. The subjects' task was to search for target ships as fast as possible, in a predetermined order. A simulated radar image provided information on the position of the MUAV and the targets. Point of departure in the present experiment was a high quality camera image. If a positive effect of visual support can be demonstrated for such images, these effects may be even more pronounced in with degraded camera images.

![Figure 4. The Computer Generated Grid (CGG), the camera image, and the Combined Heading and Pitch Indicator (CHPI) used in experiment II.](image)

Besides the zoomed-in CGE that was presented in the previous experiment, we used a perspective correct presentation of the CGE
Expérimental task
fixed order, as fast as possible by operating a
Subjects had to locate five target ships, in a
31, 24, sd 3 years, mean 24 years) participated in
The camera from a MUAV.
subjects had normal or AU
The experiment.
subjects were
Eight higher educated, male subjects (age
method consisted of a set of head-up linear quantitative indicators on the top and right side of the camera image, indicating heading and pitch angles of the camera. The indicators provide the operator with accurate numerical information on the status of MUAV and its camera. Obviously, this type of information can only be used when the operator actively estimates target position, and adjusts the camera heading and pitch accordingly.
In the experiment, these two methods of operator support were combined in a full factorial design. Such a design includes a condition without CGG and quantitative indicators, which would make the search task impossible to accomplish. Therefore, a third type of operator support was introduced, that was always present in the display: a so called Combined Heading and Pitch indicator (CHPI, analogue to experiment I). The CHPI presented the operator a pictorial, clock-like, head-up display that depicts camera and MUAV status (see Fig. 4). The specifications of the CGG, the quantitative indicators and the CHPI are given in the method section.
Note that all three indicators (CGG, pictorial, indicators) do not directly indicate required pitch. When the MUAV is flying at a fixed altitude, the subject will have to develop a concept about the relation between distance on the radar screen, and required pitch.
Methods
Subjects
Eight higher educated, male subjects (age 21 - 31 years, mean 24, sd 3 years) participated in the experiment. All subjects had normal or corrected to normal vision. The subjects were payed for their participation, and had no experience with similar operator tasks.
Experimental task
Subjects had to locate five target ships, in a fixed order, as fast as possible by operating a camera from a MUAV. The camera image simulated the image recorded by a movable camera located underneath a MUAV. The camera was controlled by means of a joystick, controlling the MUAV was not part of the subjects’ task.
After locating the actual target, subjects had to track it for 2 s (so called target-locking). Target-locking was introduced to be sure that the subject really located the target, and to avoid accidental hits. After target locking, the screen would freeze and turn green for five seconds, after which the subject could proceed with the next target. At the beginning of each trial the camera was aimed at the first target ship, which was visible in the camera image. This means that the initial camera heading and pitch were known to the subject. After two seconds (target-locking) the image would turn green, and the subject could begin the search for the first of the five real targets.
On the basis of pilot studies the maximum search time per target was limited to 90 s. If the subject failed to locate the current target within this 90 s, the camera screen would turn red, and the computer would take over the control of the camera and point it at the current target. Again, after locking, the screen would turn green, and the subject could take over the control and begin the search for the next target.
Subjects came in pairs, and completed four scenarios in succession, each scenario consisting of five targets. The completion of four scenarios never lasted longer than 20 minutes, after which the subject could rest, while the other subject completed four scenarios.

Image and displays
A simulated radar screen displayed a radar image generated on board the mothership (see Figure 5). The image was north-up, and depicted information about the position of MUAV and target ships. Information about the location of the target ships was presented by means of numbered dots, numbers indicating the search order. The colour of the actual target ship was red, all others were yellow.
The current location of the MUAV was indicated by means of a blue dot, the position of the mothership (always in the centre) with a white cross.
Around the radar image a north-up compass scale was depicted, with markers every degree, and values every ten degrees.
A second display depicted the simulated camera image, which displayed a sea, and, when in sight, the current target ship, was depicted in a 12 cm × 9 cm window in the centre of the
camera screen (exactly one third of the total camera screen, see Figure 4).
Apart from the current target, no other ships were shown. Field of view of the camera image was 5°, zoom factor was 10.2. The horizon was simulated by a transition from light (sky) to dark blue. In the conditions without CGG, 500 dots were randomly positioned at sea-level to simulate the slight texture of an empty sea. The camera image was updated at 30 Hz.

The second display could contain (a combination of) three kinds of information on platform and camera position and movements: a CGG, linear quantitative scales, and a Combined Heading and Pitch Indicator.

The CGG consisted of a pattern of parallel and perpendicular lines. The CGG was north-oriented (north indicated with arrows at the intersection of grid lines), and positioned at sea level (Figure 4).

The distance between two parallel CGG lines was 100 m, over a total surface of 10 x 10 km. The CGG was updated with 30 Hz. The CGG was both depicted in the camera image, and around the camera image on the camera screen (remember that the camera image was only a window in the centre of the camera screen). The camera image, and thus the CGG in the camera image, was zoomed-in (factor 10.2, fixed). The CGG around the camera image was perceptually correct, which means that it was depicted at the correct size when viewing directly from the MUAV (zoomfactor 1.0).

The Linear quantitative indicators could be presented head-up along the top and right edge of the camera image. These indicators depicted respectively the camera heading and pitch (see Figure 4).

The indicators consisted of a moving line with values at every degree marker, and a fixed triangle in the centre of the scale with the accompanying value (digital, round at 1 degree). The indicators indicated the heading and pitch values of the zoomed-in camera image, correct in respect with the zoomfactor of 10.2.

The combined heading and pitch indicator (CHPI) was presented as a head-up display located in the horizontal centre of the camera screen between the camera image and the bottom edge (see Figure 4).

The CHPI was always oriented north-up, and consisted of three indicators. A partially hollow bar indicated the heading of the camera. The hollow bar was filled to a certain extent, indicating camera pitch (no filling indicated 90° pitch, completely filled indicated 0° pitch, or pointing the camera to the horizon). The third indicator was a thin line, which indicated the heading direction of the MUAV. All indicators were updated with 30 Hz.

Figure 5. The simulated camera image

Mock-up and instrumentation
The experiment was conducted in the TNO-HFRI RPV Research Simulator, described in the previous experiment. The camera screen, right in front of the subject (Mitsubishi colour display monitor HL795SDBK, 38 cm x 27 cm, 1024 pix x 1024 pix), depicted the simulated camera image, the CHPI and the CGG and/or the indicators when present. These were all generated by a SILICON GRAPHIC IRIS image generator. The camera image was depicted in a window in the centre of the camera screen (12 cm x 9 cm, 340 pix x 340 pix). The visibility (visual angle and contrast) of the target ships was such that they could always be detected when they appeared on the camera screen.

The radar screen (MAGIC VIEW 14'' DIGITAL) was placed next to the camera screen, and depicted the simulated radar image. This super VGA image was generated by a 486-based personal computer.

Parameters of MUAV system, target ships, and test scenarios
The MUAV always flew at an altitude of 600 m. In half of the trials, the MUAV flew a circle with radius 1.0 km, at a constant speed of 43.6 m/s, around a point which was located 1.5 km north and 1.5 km east of the middle of the radar image. The target ships were trawlers (70 m long), positioned at a fixed location.
The 64 scenarios which were used during the experiment were randomly generated under the following conditions: distance between the initial position of the MUAV and each of the target ships was between 1900 and 4900 m. Distance between two successive target ships was at least 1000 m separate in both the north-south and the east-west direction. The ship at which the camera was pointing at the beginning of each scenario was located 2500 m north of the initial position of the MUAV. Location of the target ship would never coincide with the intersection of two grid lines (when present).

**Independent variables**

Two factors, with two levels each: CGG (absent / present) and indicators (absent / present). Note that the CHPI was not an independent variable, but was present in all conditions. All independent variables were varied within subjects, thus leading to 4 conditions for each subject.

**Dependent variables**

Three dependent variables were used: search time, total heading movement, and total pitch movement. Search time indicates search effectiveness: the time it took to locate the target. Total heading and pitch movement indicates search efficiency: larger values for these variables mean that subjects required more camera movements to locate the target, indicating a decreased search efficiency. These effects do not depend on the speed of the movements, which is directly related to search time.

Since it is expected that different kinds of synthetic visual support could have different effects on the quality of pitch and of heading estimations, total heading and total pitch movement were analyzed separately. Thus, from the recorded data three dependent variables were calculated:

a. **search time** (s), defined as the time elapsed between the beginning of the search for a new target till the locking of the target (minus the 2 seconds it took to actually lock the target). When the maximum search time of 90 s passed without the subject locking the current target, camera control was passed onto the computer. Depending on the actual heading and pitch error, the computer located the target between 1 and 20 s. This final search time (including the first 90 s) was taken as score.

b. **total heading movement** (rad), the integral of camera heading during target search.

c. **total pitch movement** (rad), the integral of camera pitch during target search.

**Statistical design**

Each dependent measure was analyzed by a 8 (subject) × 3 (session) × 2 (CGG) × 2 (indicators) ANOVA with the statistical package STATISTICA®. Each cell consisted of five observations, one for each of the five targets in a scenario.

**Procedures**

After arrival, subjects received a brief written explanation of the general nature and procedures of the experiment. The instructor explained the controls, images, procedures, purpose and task in more detail.

The training consisted of a scenario with five targets positioned at regular distances in the north-south direction, five targets at regular distances in the east-west direction, and four random positioned targets, to teach subjects the relation between positions and distances in a certain direction and accompanying heading and pitch.

During training the instructor always sat next to the subject and explained the experimental condition, and the images (the radar and camera image, and the CHPI and the synthetic visual support, if present). This same training scenario was performed under different conditions of synthetic visual support (none, or both CGG present and indicators present).

The experiment would begin when both subjects had finished their training. The subjects were informed about the oncoming condition. During the experiment the subjects had to track five scenario's in one condition.

**Results**

Analyses were executed for each dependent variable separately, results are presented per effect.

CGG (absent/present) showed significant effects on all three dependent measures: search time \(F(1,7) = 10.14, p = .015\), total heading movement \(F(1,7) = 6.78, p = .035\), and total pitch movement \(F(1,7) = 14.46, p = .010\). All effects pointed in the same direction: lower scores, and thus better performance, in the CGG present conditions (see Figure 6). Largest effect was found on total pitch movement, in the CGG present conditions the total pitch movement to locate the target was reduced with more than 50%. There were no interactions of CGG with any of the other manipulated factors.
Figure 6. The effect of CGG on search time, total heading movement, and total pitch movement.

Indicators (absent/present) did not produce significant main effects on any of the dependent variables, nor did any of the interactions of indicators, except with session ($F(2,14) = 5.06, p = .022$). A post-hoc Tukey test revealed a decreased effect of indicators with successive sessions.

Conclusions

The current experiment was designed to test the effects of a Computer Generated Grid (CGG) on target search performance of operators of Maritime Unmanned Aerial Vehicle’s (MUAV’s), and to compare performance with a traditional method of operator support by means of linear quantitative indicators depicting camera heading and pitch. For this aim a search task was introduced, in which operators had to locate target ships on the basis of a camera and radar image. Point of departure was a simulated, high quality camera image (30 Hz update frequency, 340 pix x 340 pix) and a pictorial indicator for camera heading and pitch. The update rate of the simulated camera image is above the present state of technology, and one may expect that effects are more pronounced when update frequency is lower.

The data show positive effects of the presence of a CGG in improving search efficiency of the MUAV operator; search time, total heading and total pitch movement were reduced considerably in presence of the CGG. In contrast, the more traditional method of operator support, by means of indicators, did not show significant effects on any of the dependent variables. Using indicators requires the operator to perform mental calculations in order to determine the desired heading and pitch angles from the radar image.

The absence of an effect of quantitative indicators might have been due to the pictorial Combined Heading and Pitch Indicator (CHPI). The CHPI was introduced in order to prevent that the search task would be impossible in the conditions without CGG or quantitative indicators. However, the results suggest that the CHPI, which was always presented to the observer, may be a more powerful way of operator support than the quantitative indicators.

The success of the pictorial CHPI may be based upon the same principle as the success of the CGG. Subjects could compare the heading of the target relative to the MUAV on the radar screen with the camera heading indicated by the CHPI. Comparing two orientations is a more visual task than estimating desired camera heading and adjusting indicators accordingly.

In conclusion, the Computer Generated Grid (CGG) is a powerful method to improve the search performance of MUAV operators. Search times, and camera movements are significantly reduced in presence of the CGG. The success of the CGG may be explained by its provision of perspective and distance texture gradient information, which are supposed to improve the operators ‘situational awareness’. These elementary natural invariants can easily be picked-up and interpreted by the
visual system, without demanding substantial (visual) attentional effort. The more traditional method of operator support that was investigated (indicators depicting camera heading and pitch), showed no effect on any of the dependent variables. In contrast to the CGE, reading and interpreting quantitative indicators does demand attentional resources, since the target's heading had to be estimated and the scale values had to be adjusted accordingly. These estimations and adjustments, in turn, may have degraded the accuracy of the heading observations.

Surprisingly, there were indications that the Pictorial Combined Heading and Pitch Indicator (CHPI), which was used as a baseline condition, was a successful method of operator support. Since both heading indicator and radar image were north-up, subjects could match the orientation of the CHPI's heading indicator with the estimated heading of the target. Again, this indicates that visual information that is directly picked-up and interpreted by the visual system, without requiring higher cognitive information processing, may be preferred above more abstract information for vehicle control.

**GENERAL DISCUSSION**

In Maritime Unmanned Aerial Vehicles, the function of imaging devices, is to provide the operator with visual information for information acquisition and vehicle guidance. The quality of the visual information will be crucial to many tasks in which control of the imaging system is a prerequisite for success, such as scouting, tracking, and battle-damage assessment. However, the quality of the visual information for MUAV operators may be severely degraded, which is partly due to the need to limit the data-link between the MUAV and the control station. In the present experiments, a new method for operator support was investigated: combining the camera images with a view on a Computer Generated Environment, which is composed of parallel and perpendicular lines. This view is generated without using the data-link for images between MUAV and control station. Depiction of the view may be with arbitrary field of view, update rate, zoomfactor etc.

The kind of CGE used in Experiment I enhanced the perception of the rotations of the camera or platform. These rotations generate translations in the camera image, and the presence of a grid with a high update frequency will enhance the perception of such translations. However, there is a problem in the perception of platform and camera movements with almost vertical viewing directions (large pitch angles). Under these conditions a small change in rotation can not be distinguished from a translation of the platform, since they both generate translations in the camera image. This problem can be solved by using a second grid that is positioned at some distance above the first grid. During a translation of the vehicle, the more proximal grid will move faster (motion parallax) than the grid at sea-level. However, in case the camera or platform rotates, both grids will translate at the same speed. This results in a clear distinction between the visual effects of translation and rotation.

The camera repositioning data and the verbally reported heading data jointly indicated that a pictorial indicator which was used as a control condition, was not effectively used. In contrast, the Computer Generated Environment clearly aided subjects to aim the camera at a previously seen target. Furthermore, the CGE fully compensated for performance decline with decreasing update frequency. This absence of an effect of the quality of the outside image when the CGE was present in combination with the poor verbal heading reports, even with the aid of a CGE, demonstrates that the subjects completely relied on the CGE in performing the repositioning task by using a simple heuristic. That is, the subjects could find the initial position of the target by reproducing line passages of the grid in the opposite direction.

With regard to the verbal direction estimations, subjects seriously overestimated the camera rotations and thereby misjudged the compass direction of the target. This was true for all task conditions. This poor awareness of 'compass space', is probably caused by the fact that the subjects were viewing a zoomed-in camera image (without being told so). Normally, in everyday life, the translations generated by a rotation of the observer are directly related to the speed of rotation (Koenderink, 1986). Therefore, subjects can rely on the fact that the speed by which the visual image is translated is equivalent to the rate of rotation. However, When a camera is zoomed-in, the translation speed of the camera image depends on the speed of rotation multiplied by the zoom-factor. Therefore, zooming will increase the translation speed of information over the monitor. Since the exact zoomfactor was unknown to the subject, the magnitude of
camera rotations could not be assessed from image velocity.

However, it still may be regarded rather surprising that the subjects overestimated camera rotation in the presence of the CGE, since the subjects could easily notice that the orientation of the grid lines did not change substantially during rotation. Modest grid rotations imply that the platform/camera only turns over small angles. Therefore, the overestimation of rotations shows that the subjects primarily relied on the translation speed of the camera image, ignoring the orientation information in the grid. Probably, subjects can be trained to use the orientation of the grid lines in order to improve their spatial awareness. The current results, however, show that the subjects normally tend to utilize the translation of the camera image in order to assess information on rotation.

The above findings were cause to extend the depiction of the view on the CGE in Experiment II with a perceptually correct view (zoomfactor 1.0).

Experiment II was designed to test the potential beneficial effects of the use of the CGE on the subjects’ spatial orientation or situational awareness, i.e., the perception of orientation and position of the aircraft and/or the sensor in space, and to compare performance with a traditional method of operator support by means of linear quantitative indicators depicting camera heading and pitch. For this aim a search task was introduced, in which operators had to locate target ships on the basis of a camera and radar image. Point of departure was a simulated, high quality camera image (30 Hz update frequency, 340 pix x 340 pix) and a pictorial indicator for camera heading and pitch.

The data show positive effects of the presence of a CGG in improving search efficiency of the MUAV operator; search time, total heading and total pitch movement were reduced considerably in presence of the CGG. In contrast, the more traditional method of operator support, by means of linear quantitative indicators, did not show any effect on the dependent variables.

This result clearly indicates the superiority of the CGG in improving the operators search performance. The CGG presents the visual information on MUAV and camera attitude as the operator would have seen if he was flying there himself. The optic flow of the CGG allows an effortless perception of MUAV and camera attitude. It is supposed that the CGG improves the operators ’situational awareness’, resulting in improved search effectiveness, i.e., reduced search time, and improved search efficiency, i.e., less camera-heading and pitch movements required to find the target.

The effects of the CGG are most pronounced in controlling camera pitch. Controlling pitch was expected to be difficult, since, in contrast to heading, the desired pitch angle could not be clearly perceived from the radar image. Instead, subjects had to learn which pitch angles corresponded with the distances on the radar screen. This proved to be a difficult task, even when the altitude of the MUAV was kept constant. It is to be expected that the determination of pitch angle is even more difficult for non constant altitudes.

The more traditional method of operator support, by means of the linear quantitative indicators, did not show significant effects on any of the dependent variables. Using indicators requires the operator to perform mental calculations in order to determine the desired heading and pitch angles from the radar image.

The absence of an effect of quantitative indicators might have been due to the pictorial Combined Heading and Pitch Indicator (CHPI). The CHPI was introduced in order to prevent that the search task would be impossible in the conditions without CGG or quantitative indicators. However, the results suggest that the CHPI, which was always presented to the observer, may be a more powerful way of operator support than the quantitative indicators.

The success of the pictorial CHPI may be based upon the same principle as the success of the CGG. Subjects could compare the heading of the target relative to the MUAV on the radar screen with the camera heading indicated by the CHPI. Comparing two orientations is a more visual task than estimating desired camera heading and adjusting indicators accordingly.

In conclusion, the Computer Generated Grid (CGG) is a powerful method to improve the search performance of MUAV operators. Search times, and camera movements are significantly reduced in presence of the CGG. The success of the CGG may be explained by its provision of perspective and distance texture gradient information, which are supposed to improve the operators 'situational awareness'. These elementary natural invariants can
easily be picked-up and interpreted by the visual system, without demanding substantial (visual) attentional effort. The more traditional methods of operator support showed no effect.

References

Bombardier, Inc. (1990). The CL-227 Sentinel unmanned air vehicle system; Current system configuration. St. Laurent, Quebec: Author.


Erp, J.B.F. van & Kappé, B. Computer Generated Environment for steering a simulated unmanned aerial vehicle. TNO report in press, TNO Human Factors Research Institute, Soesterberg, The Netherlands.


Session 4

Task Allocation and Automation of Complex Systems
Mapping Cognitive Support to Human Needs

Mark Neerincx

TNO Human Factors Research Institute, Soesterberg, The Netherlands

Abstract

Research in system design has been focussing mainly on either the technology or the user of the technology. Due to this separation of interest, efficient design solutions and negative side-effects may be overseen. This article follows an alternative approach in which system design is focussing on the joint human-computer task performance by analyzing human needs for support and harmonizing methods of system design to these needs. By designing support functions which supplement users' knowledge and capacities, human involvement in human-computer systems can be improved. For this type of cognitive support, requirements were established and a design method was constructed. The analysis of specific users' needs for support is part of the design method. The design method was applied to two domains: statistics and railway traffic control. In experiments, the resulting support functions improved human task performance, especially when task load was high. This article presents the general design approach exemplified with the design and evaluation of a prototype support function for railway traffic control.

1. Introduction

Designing human-machine systems involves, among other things, modelling and testing of (i) task allocations, (ii) support functions and (iii) user interfaces. Recently, three studies were done aiming at an empirical-founded method for the allocation of tasks and the design of user interfaces which provide support. First, Neerincx & Griffioen (1996) developed a cognitive task load analysis for the allocation of a set of tasks to persons which corresponds to their capacities. Second, de Greef & Neerincx (1995) developed a design method for user interfaces with support functions which can supplement human knowledge. Third, Neerincx & de Greef (to appear) extended the aiding concept and showed that it can also supplement lack of human capacity. Each of these three studies focused on a specific part of the method aimed at. This paper will present an overview of the method for the design of cognitive support incorporating the cognitive task analysis.

Design of cognitive support can be distinguished from two “classical” design approaches: the technology- and user-centered approaches. A major approach towards system design was technology-centered. The aim of this approach was to automate some primary tasks, such as medical diagnosis. Software engineering models and methods were developed aiming at the design of complex systems. The human role was reduced in such systems, for example, only consisting of a data entry function. A serious shortcoming of the technology-driven approach is that it does not address human task performance. Consequently, human skills may be employed suboptimally and unforeseen human errors may occur.

Partly as a reaction to the technology-centered approach, the user-centered approach focused on human task performance, especially regarding the interaction with the user interface. Research on human-computer interaction has provided principles and techniques to establish a consistent and simple communication between user and computer (e.g. Eberts, 1994). However, the traditional focus on the computer user has a serious shortcoming: the computer is viewed as a tool for a user who is performing a task. However, the user may sometimes be unable to perform his or her task well. A user-friendly interaction does not help to overcome this problem.

The design of cognitive support is a third, alternative approach towards system design. It focuses on the joint execution of the primary task and tries to combine the benefits of the technology- and user-centered...
approaches. Software engineering (SE) techniques are used to model human-computer task performance in the process of system development. The content of the models differs from the “classical” design models: their scope is not limited to the machine, but incorporates human roles also. Current SE-models had to be extended to incorporate principles and techniques for the design of human-computer interaction.

2. A Design Method for Aiding

The goal of this research is to improve human involvement in human-machine systems for specialist expert tasks, like railway traffic control. Cognitive support might establish such improvement by supplementing human knowledge and capacities. However, the support should be designed with care, because support facilities can have harmful effects (e.g. Neerincx & de Greef, 1993). There is a trade-off between the costs and the benefits of such facilities and the benefits can be outweighed by the costs. The question is how to optimize this trade-off.

Cognitive support aims at supplementing human knowledge and capacities. Two general benefits can be distinguished: aiding can provide knowledge the human task performer is lacking and it can off-load the task performer who is overloaded. In general, the costs of a support facility consist of the overhead of utilizing it. Optimizing the cost-benefit trade-off means optimizing the benefits and minimizing the overhead. An optimal benefit can be established when the support function is directed at major determinants and bottle-necks of human task performance.

The costs, or overhead, can be minimized, first, by integrating support into human task performance. Using support should not be a separate activity, apart from “normal” work. Acquisition of aiding should require minimal investment. Second, the user interface should be easy-to-learn and easy-to-use, the interaction should require minimal overhead, that is, minimal knowledge and capacities.

For optimizing the costs-benefits trade-off, de Greef & Neerincx (1995) formulated three requirements about how it should operate, the content of advice and the user interface:

1. The support function should take the initiative to provide information at the right time. Consequently, the user does not need to know when and how to search for information and does not need to invest in these actions.

2. This information should consist of context-specific, procedural task knowledge. The advice is minimal: not more than necessary. However, the advice provides complete procedures to solve the problems.

3. The user interface should be a well-integrated part of the human-machine dialogue. A minimal and consistent interaction requires little knowledge and contributes to efficient task performance.

Subsequently, de Greef & Neerincx (1995) developed a design method for aiding which agrees with these requirements. System development is an iterative process consisting of generating design models and testing the result. The design method consists of three steps, the generate-and-test cycle is embedded in this method:

1. In the first step, a “plain interface” is designed. This interface is easy to learn and easy to use. Users with complete and correct task knowledge and sufficient time to perform their tasks will perform their tasks well with this interface. This might be fictive users or experts.

2. In the second step, human performance is analyzed to acquire users’ needs for support. Future users may perform a number of tasks with a prototype of the plain interface. Performance deficiencies point to lack of knowledge or capacities. For these parts of the task, step 3 of the method is entered.

3. In the third step, the aiding function is designed. An expert model and a model of joint task execution are constructed and added to the plain interface design. The result is an aiding interface.
De Greef & Neerincx (1995) discuss the modelling part of the first and the third step in detail. Here, I will go into the human performance analysis briefly. In the performance analysis, the framework of Rasmussen (1986) and Reason (1990) is used, distinguishing skill-, rule- and knowledge-based actions. Figure 1 shows a simplified model with the three levels of information processing. Cognitive task load is higher for higher levels of processing. Cognitive support provides procedural task knowledge, that is, the “short-cut” on rule-based level. High task load on knowledge-based level might be reduced by cognitive support, because problem analysis and procedure planning are minimized for the supported tasks.

Based on the framework of Figure 1, Neerincx & Griffioen (1996) developed a method to assess task load as a function of the level of information processing and the number of actions. Figure 2 shows a theoretical assessment of task load on these two dimensions (see Figure 2). The vertical axis shows the level of information processing, that is, the ratio between knowledge- and rule-based actions. The more knowledge-based actions, the higher the ratio, the higher the task load. The horizontal axis shows the number of actions. The more actions, the higher the task load. The grey area in Figure 2 represents a standard for task load in terms of the level of information processing and the number of actions. Task load is too low for job X and, therefore, this job could be extended with actions, especially knowledge-based actions. Task load is too high for job Y and, therefore, this job could be slimmed down or a support facility could be designed for job Y.
3. Aiding for Railway Traffic Control

De Greef & Neerincx (1995) applied the three steps of the method to design an aiding interface for statistical analysis software. In an evaluation the aiding function improved performance and learning considerably. The next question is whether the design can be applied for a very different task, such as railway traffic control (RTC). Neerincx & de Greef (to appear) applied the design method for aiding to the task of railway traffic control. Railway traffic controllers set switches and signals for a part of the railway network. The railway timetable can be viewed as a work plan, prescribing the procedures to follow. Next to carrying out this work plan the traffic controllers have to deal with irregularities, such as a switch getting out of order. For each irregularity, they have to find the best solution and plan a new procedure.

![Prototype plain interface for railway traffic control](image)

**Figure 3:** Prototype plain interface for railway traffic control (Neerincx & de Greef, to appear).

**Step 1: Design of the Plain Interface**

Figure 3 shows the railway-yard of Amersfoort, which can be controlled by one person. In Amersfoort trains go to and come from four different directions: Amsterdam, Utrecht, Zwolle, and Apeldoorn. It contains four platforms. In the example, a train is coming from Zwolle and going to platform 6. I will not go into much detail about this Figure (see Neerincx & de Greef, to appear). Here, it suffices to say that railway traffic control consists of two subtasks: carry out work plan and deal with irregularities.

Carry out work plan is done in the upper window. For this subtask the operator sets the train routes and, subsequently, the system sets the switches and signals in the positions according to these route settings. For the second subtask, dealing with irregularities, the operator must perform some extra tasks.

Irregularities can be disturbances (such as a switch out-of-order) or deviations from the time-table (such as a train arriving too late). These extra tasks can be done with the buttons in the window right-below.
Step 2: Human Performance Analysis

The current interface of the Netherlands Railways can be viewed as a plain interface (it does not offer cognitive support). Three control posts of the Netherlands Railways were observed (Neerincx & de Griffioen, 1996). The number of actions and the ratio between knowledge- and rule-based actions were calculated for 11 observations of one hour each. Among other things, the task load proved to be high for dealing with irregularities in busy control posts. Re-allocating tasks could not solve this problem. The question is whether aiding can be a solution for such a problem.

Step 3: Design of the Aiding Function

Based on the human performance analysis, existing documents, and interviews with experts, an expert model was formulated for dealing with a number of irregularities. This model of the controller's task prescribes the actions to carry out when an irregularity occurs. After constructing the expert model, it was specified how the human-computer cooperation operates: the computer (1) forwards the events (irregularities) to the user, (2) fetches the context-specific rules for the events based on the current state, and (3) applies the operations the user orders. When an irregularity appears, the corresponding context-specific procedure is presented in the window left-below of Figure 3.

4. Evaluation of Aiding

The section above described the design of a prototype interface which provides context-specific help for railway traffic control. Neerincx & de Greef (to appear) tested whether this kind of support can really be of help under different conditions of task load. The prototype interface was extended with a simulator and users of this prototype had to deal with a number of irregularities at the railway-yard. The same irregularities appeared when a lot of trains had to be controlled (high task load condition) and when few trains had to be controlled (low task load condition). Users worked with the aiding facility or without this facility. The hypothesis was that aiding can complement human knowledge and capacities, or, in other words, that aiding is beneficial, especially when task load is high.

With aiding, users of the railway simulator performed 91% of the actions prescribed by the expert model correctly, whereas without aiding, users performed only 46% of these actions. Furthermore, users spent 38% less time to accomplish a single action with aiding. This is important, because dealing quickly with irregularities promotes safety and minimizes delays. Aiding was especially beneficial when task load was high, that is, when there was hardly time for the knowledge-based actions: problem analysis and procedure planning (see Figure 1).

5. Conclusions and Discussion

The paper presented a brief overview of a method to design cognitive support. This method was applied to develop a prototype user interface for railway traffic control. A human performance analysis conveyed the needs for aiding and, subsequently, an aiding function was integrated into the interface. The prototype was used in an experiment to study the effect of aiding under various loading conditions. The first conclusion is that the design method can be used for process-control tasks like railway traffic control.

Based on the framework of Rasmussen (1986) and Reason (1990), it was hypothesized that presenting procedural task knowledge decreases the level of information processing, because it provides short-cuts at the rule-based level. In accordance with this hypothesis, aiding proved to be beneficial, especially in the high task load condition. The second conclusion is that aiding can supplement human knowledge and capacities.
Human needs for aiding depend on current task demands, but also on the current knowledge and capacities of the task performers. A major part of the irregularities in process control appear seldom, so that the operators will hardly acquire or keep up procedural knowledge for dealing with irregularities during their work (experience). This paper proposed to develop aiding to solve this problem. The operators can learn the short-cuts of figure 1 by training and, therefore, training can be viewed as an alternative for aiding. However, operators might forget the learned procedures, might have problems to retrieve it from memory immediately, or might not realize that the procedures are applicable in the current situation. In these three circumstances, aiding could complement the training. In this perspective, the human performance analysis provides aiding and training requirements. Aiding is harmonized to the users, because it is harmonized to the knowledge the users learned by training.

6. References


Experimental investigation on the effect of task allocation on the human operator

A.R. Paauw, Z.G. Wei and P.A. Wieringa

Man-Machine Systems Group, Laboratory for Measurement and Control
Faculty of Mechanical Engineering and Marine Technology, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands. E-mail: paauw@tudw03.wbmt.tudelft.nl

Abstract: The relation between task properties, system performance and operator mental load in supervisory control is investigated using an artificial model of a pasteurisation plant. The relation is evaluated using the degree of automation which is based on the relative weights of the manually controlled tasks. Three ways of calculating task weights are investigated, based on 1) the mental load perceived by the operators while all tasks are manually controlled, 2) the system output, and 3) task properties. Experiments were performed using the pasteurisation plant.

1 Introduction

Task allocation between human operators and automation in the control room of complex systems is an important phase in the human-machine system design process. Task allocation and task properties will influence the mental load which an operator will have to invest to operate a system and the level of performance he or she is able to achieve. To determine the effect of task allocation on system performance the combination of task weight and task allocation was used to define a Degree of Automation (DoF) [Wei, et al., 1994]. The DoF is defined as:

\[ \text{DoF} = \frac{\sum_{i=1}^{N} t_i w_i}{\sum_{i=1}^{N} w_i} = \frac{\text{sum of automated task weights}}{\text{sum of all task weights}} \]

where \( w_i \) is the weight of task \( i \) and \( t_i \) is the automation index of task \( i \) (\( t = 0 \) or 1 i.e. manual or automated respectively).

Task weights were measured in three ways, providing three different descriptions of the DoF. First, we measure the effect that a task has on system output. This is done comparing the performance of a system in which one task is not executed to the performance of the same system in which all tasks were executed automatically, the reference case. This gives an indication for the effect a task has on system performance. This weight measure is named Task Effect on System performance, or TES. The DoF based on this description of weights, \( \text{DoF}^{\text{TES}} \), is an indication of the fraction of system performance controlled by automation.

The second measure for the task weight is obtained from the workload that is imposed by a task upon the operator. This depends on task properties only and is independent of the operator! This
weight measure is named Task Demand Load, or TDL. The DofA based on TDL, DofA^{TDL}, is an indication of the fraction of the maximum TDL that is taken care of by automation.

The third measure for the task weight is obtained from the mental load that an operator perceives for each task while he controls the system. This is called Task Mental Load, or TML. The DofA based on TML, DofA^{TML}, is an indication for the mental load that is perceived by the operator in a partly automated system relative to the mental load that would have been perceived had the system not been automated at all.

For a first exploration into this relation, Wei, et al., [1996], performed experiments with an artificial system consisting of a forwardly coupled network of first-order sub-systems. They indicated that there is a relation between the measures for calculating the DofA and system performance. Another finding from these experiments is that the operator’s overall mental load may be predicted from the TML, obtained from a situation where the operator controlled a system fully manually. To verify these results with more realistic system a modified version of a pasteurisation plant [Muir, 1989] was constructed. This plant has been used extensively in man-machine studies and was reprogrammed using LabView (National Instruments Corp.).

2 Experiments

In Figure 1 the juice pasteurisation plant is schematically displayed. The plant consists of an input vat from which juice is pumped through a heat-exchanger system powered by a steam circuit. The temperature of the juice flowing out of the heat-exchanger must be maintained within an interval. If the pasteurisation temperature is too low it is recycled to the input vat. If the temperature is too high the juice is considered burnt and pumped into the waste vat. The input vat should not become empty or overfilled. To obtain a mixed set of tasks a discrete task was added to the system’s continuous tasks. This task consisted of the regular operation of a 4-position switch.

The operator can control the feed pump, steam pump, overflow valve, steam heater setting and the distribution system. All these controls can be set to automatic or manual. They were grouped

Figure 1: The Pasteurisation Plant
grouped to form three functional groups: flow system, steam circuit and distribution system. The plant now consists of three continuous tasks, requiring control in a range of values, and two discrete tasks, requiring control in a limited number of values.

Experiments were performed with 9 students from our laboratory. After operating the plant in several configurations for half an hour the subjects performed 15 minute sessions in all possible configurations of the plant, except the fully automated case. After each session the subjects rated their mental load using the Rating Scale Mental Effort (RSME) [Zijlstra, 1993]. They rated their mental load for each task individually, which provides the TML measurements and for the overall plant, or Overall Mental Load (OML).

As a measure for system performance, initially the pasteurised fraction was used, (i.e. the amount of juice that is properly pasteurised divided by the total amount available.) However, the input vat as well as the pasteurisation temperature allow for large deviations from the optimum before any product is wasted and the pasteurised fraction is affected. Since primary interest is in how well the operator performs, rather than the system itself, a new performance measure was defined. This measure consisted of a combination of several system variables. The Combined Performance (CP) measure uses a variable for every part of the system that could be controlled manually and uses the production performance. For control of the steam circuit the standard deviation of the pasteurisation temperature is used. For operating the flow system the standard deviation of the input level is used and for the distribution system the total time that the system is in an alarm state is used as the performance measure. These values are all compared to the reference case.

3 Analytical results

Before performing the experiments two methods of calculating the DofA can already be applied:

DofA based on TES

To calculate the $D_{\text{ofA}}^{\text{TES}}$ the plant was first ran fully automatic. In this way a reference performance could be obtained. Then runs were made each with one of the tasks failed. This results in the values for the $D_{\text{ofA}}^{\text{TES}}$ listed in Table 1, which are calculated with the formula:

$$D_{\text{ofA}}^{\text{TES}} = \frac{\sum t_i \cdot w_i^{\text{TES}}}{\sum w_i^{\text{TES}}}$$

For this system, with three tasks, $i$ ranges from 1 to 3, indicating the flow system, steam circuit and distribution system respectively. The weight based on TES is defined as the difference in performance with the reference case $E_0$:

$$w_i^{\text{TES}} = CP_{E_0} - CP_i$$

in which $CP_i$ is the performance when control task $i$ is failed.
The total weight can now be defined as:

$$\sum w_i^{TES} = \sum_{i=1}^{3} (C_{P_{E_0}} - C_{P_i})$$

The weights are normalised to 1 using the total sum of weights. Thus the DofA$^{TES}$ can be calculated with:

$$\text{DofA}_{TES} = \frac{\sum_t \cdot w_i^{TES}}{\sum w_i^{TES}}$$

Where $w_i^{TES}$ is the normalised task weight.

A problem occurs for the TES measurement for this system with the distribution system. If this control is 'failed', e.g. the switch is not operated, the first distribution vat will overflow and most product will be wasted resulting in a CP of 0. This causes the TES measurement for the configurations that include manual control of the distribution system to deviate considerably from the other configurations.

Table 1: DofA results

<table>
<thead>
<tr>
<th>session number</th>
<th>manual tasks</th>
<th>DofA$^{TML}$</th>
<th>DofA$^{TES}$</th>
<th>DofA$^{TDL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E0</td>
<td>none</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>E1</td>
<td>steam</td>
<td>0.64</td>
<td>0.73</td>
<td>0.68</td>
</tr>
<tr>
<td>E2</td>
<td>flow</td>
<td>0.60</td>
<td>0.73</td>
<td>0.52</td>
</tr>
<tr>
<td>E3</td>
<td>distribution</td>
<td>0.77</td>
<td>0.54</td>
<td>0.80</td>
</tr>
<tr>
<td>E4</td>
<td>steam + distribution</td>
<td>0.40</td>
<td>0.27</td>
<td>0.48</td>
</tr>
<tr>
<td>E5</td>
<td>flow + distribution</td>
<td>0.36</td>
<td>0.27</td>
<td>0.32</td>
</tr>
<tr>
<td>E6</td>
<td>flow + steam</td>
<td>0.23</td>
<td>0.46</td>
<td>0.20</td>
</tr>
<tr>
<td>E7</td>
<td>all</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**DofA based on TDL**

The TDL was calculated using a Hierarchical Task Analysis (HTA) [Kirwan, 1992]. An HTA defines what subtasks need to be executed in order to complete a task. The HTA for the pasteurisation plant is given in Figure 2, on the next page. To calculate the DofA$^{TDL}$, the amount of effort needed from the operator to control a task was set to 1. Then, based on the knowledge and experience of the experimenter, the amount of effort needed for each of the subtasks was estimated, relative to each other. The same weighting method is repeated for tasks lower in the hierarchy. To calculate the TDL for a component anywhere in the HTA the weights assigned are
simply multiplied upwards through the tree. If, for instance we would calculate the TDL for the control of the steam circuit, both steam pump and heater are on manual, we find:

\[
\text{TDL}_{\text{steam circuit}} = (0.1 + 0.7) \cdot 0.4 = 0.32
\]

The remaining values for the TDL are listed in Table 2. The advantage for estimating weights in this way is that the plant does not have to be regarded as a whole but weighting can be focused on a single task. The DofA^{TDL} is defined as:

\[
\text{DofA}^{TDL} = \frac{\sum t_i \cdot w_i^{TDL}}{\sum w_i^{TDL}}
\]

\(w_i^{TDL}\) can then be calculated by:

\[
w_i^{TDL} = TDL_i \quad i = 1, 2, 3
\]

with the values for the TDL from Table 2.

Table 2: Task weight based on TDL

<table>
<thead>
<tr>
<th>Sessions</th>
<th>flow system</th>
<th>steam circuit</th>
<th>distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>i = 1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>TDL</td>
<td>0.48</td>
<td>0.32</td>
<td>0.20</td>
</tr>
</tbody>
</table>

The value for the TDL for control of all tasks equals 1, the TDL for control of the top level "produce as much juice as possible" was defined as such. Hence the total weight of all tasks based on TDL equals one:

\[
\sum w_i^{TDL} = 1
\]
Thus the DofA\textsuperscript{TDL} can be calculated with:

\[
\text{DofA}\textsuperscript{TDL} = \sum_{i=1}^{3} t_i \cdot \text{TDL}_i
\]

where \( t_i = 1 \) for the automated systems. This results in the values listed in Table 1.

### 4 Results

With the experiment data the DofA and the system performance can be calculated. In Figure 3 the combined performance and the overall mental load are displayed in a single graph. As can be seen, the combined performance decreases and the overall mental load increases as more tasks are allocated to the operator. The flow control task is perceived to be the most demanding by the subjects, followed by the steam circuit and the distribution system.

Performance for the steam circuit and the flow system does not suffer much from the additional allocation of the control of the distribution system to the operator. However, this allocation does result in an increased mental load perceived by the operators.

**Figure 3: Performance and Overall Mental Load**

**DofA based on TML**

With the experiment data the DofA\textsuperscript{TML} can be calculated. The TML ratings provided by the subjects for the session where all controls are operated manually, or DofA=0, is used to scale the OML for the sessions with higher degrees of automation. This provides an indication for the mental load needed to control the plant in all sessions using the information from only one
session. The weight of a task based on TML equals the TML rating taken from the fully manual session. The DofA\textsuperscript{TML} can then be calculated with:

$$DofA_{TML} = \frac{\sum t_i \cdot w_i^{TML}}{\sum w_i^{TML}}$$

or the DofA\textsuperscript{TML} equals the sum of TML ratings taken from run E7 (Table 1) for the automated controls, divided by the sum of all TML ratings from E7. The results are listed in Table 1.

**DofA and OML**

If we plot the three methods of calculating the DofA against the results obtained from the OML measurements we can see how well the DofA compares to the mental load that the subjects perceived while operating the plant. In Figure 4 the DofA\textsuperscript{TES} is compared to the OML. A line has been drawn between points DofA=1 (E0) and DofA=0 (E7), to indicate the distribution of data points, assuming that the relation between DofA and OML is linear.

Weighting the distributions system with the TES results in a value for the task weight that is too high. This results in a value for the DofA\textsuperscript{TES} for the sessions where the distribution is manually controlled (E3, E4, E5, E7) that is too low. Hence two trends can be identified in Figure 4, one with and one without distribution system, indicated by a dotted line.

The datapoints E1, E2, E3 are based on measurements from single-task sessions. The task weights calculated for these sessions are combined to obtain the DofA\textsuperscript{TES} for the other, multi-task sessions. Because of this and the difficulty in calculating the TES for the distribution system the data is spread.

**Figure 4: DofA\textsuperscript{TES} compared to Overall Mental Load (OML)**

The datapoints E1, E2, E3 are based on measurements from single-task sessions. The task weights calculated for these sessions are combined to obtain the DofA\textsuperscript{TES} for the other, multi-task sessions. Because of this and the difficulty in calculating the TES for the distribution system the data is spread.
In Figure 5 the $\text{DofA}^{\text{TDL}}$ is compared to the OML. The results are considerably better than those of the $\text{DofA}^{\text{TEs}}$. This implies that with a TDL analysis a reasonably accurate prediction can be made for the distribution of the OML for each session, without carrying out any experiments. Since the TDL only indicates relative values of workload, the absolute values for the OML cannot be calculated. In case the OML for session E7 (all tasks manual) is also known, this distribution can be used to calculate the absolute values for all the sessions.

Finally, in Figure 6 the $\text{DofA}^{\text{TML}}$ is compared to the OML. The $\text{DofA}^{\text{TML}}$ also has a strong relation to the OML. This implies that from a single experiment, the session with all controls on manual ($\text{DofA}=0$), the OML for the sessions with higher degrees of automation can be estimated.
To combine the results of performance and mental load measurements, both the overall mental load and the combined performance measurement are plotted against the DofA$^{\text{TDL}}$ in Figure 7. A second order polynomial has been drawn through the data points of both OML and CP. As the DofA increases the OML decreases and the performance improves. This relation is stronger for the OML measurement. The data for the CP is more scattered.

5 Conclusion

Performance measurement is redefined as a combination of system variables rather than production performance only. This allows us to clarify the operator's role in system performance more clearly.

The DofA based on TES does not produce useful information for this system. This is caused by the control task for the distribution system. The discrete nature of this task causes the system performance to be zero when it is failed, which makes TES analysis meaningless.

The DofA based on TML can be used with reasonable accuracy to predict the OML, as was found with the artificial system used by Wei, et al., [1996]. From a single experiment where all tasks are controlled by the operator, the OML can be estimated for control of the plant at higher degrees of automation.

The DofA based on TDL also has a strong relation to the OML, as was the case with the artificial system, used by Wei [1996]. However, it is not clear if this analysis can be reproduced. It was not used before and can not be applied to the simple system.

The relation between system performance and the DofA$^{\text{TDL}}$ is less favourable than the relation between the OML and the DofA$^{\text{TDL}}$. The nature of the tasks and the amount of data points that can be obtained with only three tasks may be the reason for this.
In Figures 4 through 6 linear relationships between the DofA and the OML are suggested. A non-linear relation can be found, but this relation may only be applicable to this specific system and have no general value.

References


Filtering information for the supervision of complex systems

Elyès Cherifi, Bernard Riera and Eric Hais
University of Valenciennes, LAMIH, Le mont Houy 59300 Valenciennes,
U.R.A CNRS 1775
FRANCE
E-mail : cherifi@univ-valenciennes.fr

Keywords

Entropy, Transinformation, Coupling variables, Coupling matrix, Functional analysis, Pertinent information, Complex systems, Human operator.

1. Introduction

The automation seems one of the most important factors for the development of the big continuous industrial process. The aim of the automation is to improve the quality and the quantity of the product, this implies also to guarantee the security of the installation, the humans and the environment. However, in a complex system, the automation is not sufficient to achieve all the production and security goals. This is the reason why a supervision system is still necessary. At the moment, this activity of supervision requires the presence of one or several operators. Indeed, the human being has the main quality to know how to manage the uncertain and the unexpected situations and to create strategies adapted to the difficulties. However, with the increasing degree of complexity of the plant (large number of variables), the difficulties of the human operator to supervise a complex system is a real problem. The operator has at his disposal many information about the process, but he is not guided in his taking decision, by any analysis on these elements or their relations [Rakoto, 95].

With regard to the increasing complexity of production systems and the high level of automation, the supervision system design becomes more and more delicate and has to be human centered. Indeed, a simple report must be remembered. The Human-Being is an "information channel with limited capacities" and constitutes, in term of information treatment capacities of this Man-Machine system, an inescapable bottleneck and fully justifies, for this reason, that the design of all supervision systems has to be centered on the operator. It is no use designing a supervision system which provides all the information coming from the production system in an aleatory way because the imposing mass of data will not permit the supervision operator to treat them and therefore to work effectively [Lambert, 96]. The supervision operator has a tendency to be too often reactive to the alarms. Indeed, he waits for alarms instead of foreseeing and anticipating the anomalous situations. Face to an alarm cascade (christmas tree) often translating a lack of anticipation effort, he has difficulties to process and to act. Well, in the high level supervision tasks, some support functions to solve the problem are added, more and more. That can have, like perverse effects, for instance, a reduction of the anticipation efforts. Consequently, the supervision interfaces must use the best of the human facilities in terms of perception, anticipation and cognitive.

One of the solution to solve this problem is to display to the operator the pertinent variables of the process. For that, it is necessary to filter information by analysing the system in order to extract the pertinent ones. The pertinent information are represented by the most important variables of the system, as well as the different relationships between these variables, expressed under the form of coupling index. The hypothesis which has been done, consist in saying that a variable is pertinent for the human operator, if it is strongly coupled.
with the other ones. The information theory, in addition with a functional analysis of the process, allows by the means of entropy concept, to calculate the degree of coupling between variables. This is done by computing the transmitted information on each variable and then, we deduce a set of pertinent information of the system which are useful for the operator assistance.

In this paper, first we will present the functionalities of a supervision system. Secondly, we will detail the concept of information theory, and the way to extract pertinent variables in addition with a functional analysis of the process. At least, we will show an application of filtering information applied to a process of reprocessing nuclear waste.

2. The supervision system and its functionalities

In a very general way, the supervision system (cf. Figure 1) can be split up into two subsystems: the control/command system and the supervision tools. There are three supervision tools: the information synthesis system, the automatic supervisory system and the support systems.

The Automatic Supervisory System: according to our point of view, an automatic supervisory system is a traditional supervisory system i.e. a system which provides for the supervision operator a hierarchical list of alarms generated by simple comparison with regard to some pre-defined thresholds. The criteria of classification can be relative, for example, to the instant of detection or to the degree of dangerousness.

The Information Synthesis System: the information synthesis system manages the presentation of information coming directly from the process, via any support (synoptic, console, panel, and so on), toward the supervision operator. It is at this level that we intend to give our contribution by defining pertinent information to give to the operator.

The Operator Support Systems: a support system is characterized, on the one hand, by a contextual integration of the process states and, on the other hand, by the information generation demanding, if they must have been produced by an operator, a knowledge-based behavior. Among the possible systems bringing an useful and substantial help to the supervision operator and considering some inconveniences of the today supervisory systems, we find the filtering alarm systems, the anticipation support system, support to the diagnosis of the causes of dysfunctions, the support to the resumption of failures. Support systems are located to all the human decisional levels, i.e.: the level of detection (filtering of alarms, anticipation support system), the level of problem resolution (support to the diagnosis of the causes of dysfunctions) and finally the level of action (support to the resumption of defect).

As well as its command capacities, an industrial process supervision system has to collect, supervise and register an important quantity of process data in order to detect the possible dysfunctions and alert the supervision operator.

More precisely, the supervision system helps in two different contexts: out line and in line. Out line, the supervision system allows to establish, in deferred time, some reports and thus to analyze the production performances. In this case, some actions can be then undertaken in order to improve the Safety of Working (Reliability, Maintainability, Availability and Security) of the installations. To this topic, the archived data regarding to the dysfunctions are very rich in teachings because they permit, for instance, to define the preventive maintenance policy. It is, for these reasons, that the supervision system is, in all factories, useful because it is an important source of information. Indeed, as well as the maintenance, the automation team are interested in these collected information which are centralized in the supervision room. In line, the supervision system allows, on the one hand, to have access to the measurable information relative to the process and, on the other hand, to attract attention's operator by signalling him some important events.
3. Concept of the information theory

The information theory is presented here by the entropy and the transinformation concepts. These two concepts are used to measure the degree of the transferred information between variables. This allows to compute indexes which represent the degree of coupling between variables [Titli, 79]. We think that this approach is interesting because it is quite easy to apply it and to get useful results. The aim is not to get a complete dynamic model of the process which is more often impossible to get with regard to the complexity and the high number of variables. We intend by this analysis of the process to get information about the most important variables. Knowing them, we will give the possibility to the human operator to supervise efficiently the process. In others words, we will identify the process, not in order to design a command algorithm, but to design a "human supervision system" which will propose the right information to the human operator.

3.1 Entropy

Given a discrete variable $X$, having $N$ values with the probabilities $p_1, p_2, \ldots, p_n$, we define the entropy of the discrete variable $X$ by [Titli, 79; Richetin, 75]:

$$H(X) = -\sum_{i=1}^{n} P_i \log P_i$$

with $P_i = \Pr(X = X_i)$

The common logarithm used is expressed in base 2.

If $U = \{x_1, x_2, \ldots, x_n\}$ is a set of values of a discrete variable $X$, $n(i)$ the occurrence number of the event ($X=x_i$). The empirical probability of the event ($X=x_i$) is defined by [Titli, 79]:

$$P(i) = \frac{n(i)}{N}$$
Where \(N\) is the total number of the observation.

We define in the same way, the empirical entropy by:

\[
H(X) = -\sum_{i=1}^{p} P(i) \log P(i)
\]

We note that if the states are equiprobable, then \(H(X) = \log N\), it corresponds to the maximal value which means the maximal disorder. If contrary, we have \(p_1=1, p_2=p_3=...p_n=0\), then \(H(X)=0\). It corresponds to the minimal value of the entropy.

For two variables \(X, Y\), if \(U_1 = \{x_1, x_2, ..., x_p\}\) is a set of values of a discrete variable \(X\) and if \(U_2 = \{y_1, y_2, ..., y_q\}\) is a set of values of a discrete variable \(Y\), we define in the same way:

\[
H(X,Y) = -\sum_{i=1}^{p} \sum_{j=1}^{q} P_{ij} \log P_{ij} \text{ where } P_{ij} = P(X = x_i \land Y = y_j)
\]

If we have a partition, \(P(\Omega) = \{P_1, P_2, ..., P_n\}\) we define the entropy \(P(\Omega)\) as:

\[
H(P) = -\sum_{i=1}^{n} \Pr(P(\Omega) = P_i) \log(\Pr(P(\Omega) = P_i))
\]

3.2 Transinformation

The transmitted information between variables \(X\) et \(Y\) or transinformation is defined by [Titli, 79]:

\[
I(X:Y) = H(X) + H(Y) - H(X,Y)
\]

\(I(X:Y)\) measures the transmitted information between the variables \(X\) and \(Y\). This quantity is zero if and only if the variables \(X\) and \(Y\) are statistically independent and maximal when they are linked.

In general, we have:

\[
I(X_1; X_2; ....; X_n) = \sum_{i=1}^{n} H(X_i) - H(X_1, X_2, ..., X_n)
\]

3.3 Coupling indexes

The coupling indexes measure the links between the variables [Titli, 79; Richetin, 75]. There are two kinds of coupling indexes, the static coupling indexes and the dynamic coupling indexes. The static ones are related to the static systems and the dynamic ones are related to the dynamic systems.

The static systems are the systems which variables don’t depend on time. The dynamics systems are the systems which variables depend on time.

For the static systems, we define the static coupling index by:

\[
m_{ij} = \frac{I(X_i; X_j)}{H(X_j)}
\]

This index allows to compute the degree of links between the variables. This is expressed by a matrix called “statistical coupling matrix”. The general term of the coupling matrix is
Uly

$m_{ij} :$

\[
0 \leq m \leq 1 \quad \begin{cases} 
  m_{ij} \to 0 \text{ if the variables } X_i \text{ and } X_j \text{ are statistically independent} \\
  m_{ij} \to 1 \text{ if the variables } X_i \text{ and } X_j \text{ are statistically dependent}
\end{cases}
\]

For the dynamic systems we define the dynamic coupling index by:

\[
m'_{ij} = \frac{I(X_i;X'_j)}{H(X'_j)}
\]

We note $X'$, the variable $X$ shift with a sample period. For instance if $X=(x_1, x_2, \ldots, x_n)$, then $X'=(x_2, \ldots, x_n)$. This index allows not only to compute the degree of the link between the variables, but gives the sense of the coupling. For instance, $m'_{ij}$ shows the degree of coupling of $X_i$ to $X_j$ due to the variation of $X_j$. We can see then which variable is influenced by the other. This allows us to see the causality on the process. The interaction degree of the variable in the system, is given by computing the transinformation between the system and this variable:

\[I(S;X_i) ; \quad S = \{X_1, X_2, \ldots, X_n\}\]

4. Combination of information theory and functional analysis

Our technique consists in using this theory to the complex systems, which is represented by a very important number of variables. It is the reason why it is necessary in the first step to filter the variables, i.e. to classify and to eliminate those which are not a priori necessary to process with information theory. For that, we think that functional analysis of the process could be very useful.

The term "Functional Analysis" includes, in fact, two intrinsically different aspects:

- The external functional analysis which consists in the expression of functional need, stage specific to the Value Analysis. We do not present the concepts because the objectives of the present research concern existing continuous process.

- The methods of internal functional analysis called also methods of technical functional analysis are a necessary stage in all analysis of which goal is the understanding of system working. Two complementary definitions relative to the functional analysis exist:

  - (1) The methods of functional analysis describe the expected functions of a system and its features: they bring an important help during the fundamental stage of understanding and during the artificial description of some nominal runnings of working.

  - (2) A functional analysis defines the functions of a system without determining the components. Generally, the analysis led on the complex systems are some “structuro-functional” analysis i.e. with the purely functional analysis, a structural analysis is jointly achieved in order to define the places of some components without describing their role in the system organization [Lambert, 96]. This conjugated use permits thus to define the places of some components in the system organization and their function with regard to the goal of system.

Hence, the functional analysis is a very good complement to the information theory, because some knowledges about the process are used.

The methods of functional analysis which could be applied to physical systems are for instance: SADT and MFM [Lind, 90]. We do not describe in this paper these methods but only the expected results. Through the functional analysis, one can clearly define all the production and security functions. Functional analysis enables a description of the process and a decomposition of the systems into sub-systems. In a certain way, we have the possibility to know the sub-systems which are functionally independent. After, it is
necessary for each variable of the process to indicate the corresponding functions. Indeed, each process variable is a source of information about one or several functions. After, the idea is to choose, firstly variables which are involved in several functions and secondly variables which are very representative of one function. The information theory will be applied on this set of variables. Hence, it is possible to extract variables having a lot of interactions with most parts of the system. This sub-set of variables represent the variables that the human operator has to really supervise.

We have applied this technique on a simplified process in order to test its validity.

5. Filtering variables in the reprocessing of nuclear waste

We will present an application of this theory to the reprocessing of nuclear waste (see figure 2). It is a complex system represented by 50 variables [Leyval, 91]. Even if the number of variables is not too high, the complexity of the system is sufficient because nobody can supervise continuously 50 variables simultaneously.

This process is placed in a shop of spent fuel reprocessing, composed of two pulsed coupled columns and their components. The function of this shop consists in recovering the uranium and plutonium contained in some bars of used fuels proceeding of nuclear reactors.

All the radioactive bodies (Uranium, Plutonium, Products of fission) are dissolved in the nitric acid (H+), and the function of the extraction column consists in selectively transferring the uranium and plutonium toward an organic phase composed of tributylphosphate (TBP) then that the products of fission stays in the aqueous phase composed of nitric acid (H+). The extraction necessitates to well mash the aqueous phase and the organic phase to maximize the surface of contact between the two solvents and, so, to optimize the chemical exchanges. Then the two phases are separated by gravitation because the nitric acid has a bigger density than the tributylphosphate. Every pulsed column includes a central part where these chemical exchanges take place, and two zones enlarged to the extremities, which are some decanting zones permitting the separation of the two phases. The efficiency of chemical transfer, in the central body of each pulsed
column, is conditioned by the presence of a periodic pulsation pressure which encourages the mixture with the light phase (the Tributylphosphate noted TBP) and which allows to slow down the coming down of the heavy phase (the acid noted H+). In order to permit the circulation against current of the phases, the light phase (the TBP) is injected in the bottom of the body, and the heavy phase (the acid), in the top. The organic phase is equivalent to a continuous phase, i.e. that it occupies all the volume of the column, at the departure, and stays, after the introduction of the aqueous phase, the majority phase, in which some droplets of dispersed phase of acid circulate. The existence of the second pulsed column called "wash column" is due to the necessity of isolating the products of fission still present in the organic phase at the exit of the extraction column. Indeed, the wash column is supplied by its own aqueous phase and so, can recover the vestigial products of fission then this aqueous phase is sent toward the extraction column via the retraining loop in order to be reused for the extraction. Then, this cyclic process can continuously restart a new reprocessing cycle.

5.1 Functional analysis

We have managed a functional analysis of this process firstly in order to understand well its running and secondly to define a subset of variables on which we could apply information theory. In fact, in this process, we can distinguish three kind of variables:

- setpoint variables,
- control variables which belong to a feed-back control loop,
- production variables.

We see that setpoint variables are linked to the control variables, then it is useless to consider them in computing the coupling indexes. Even if the number of variables is low (50), we decided to filter the variables with the results of the functional analysis. We have got a new subset of 28 variables.

5.2 Information theory analysis

The information theory analysis, is applied to this system with the 28 variables remaining from the functional analysis. Therefore from a numerical process simulator, we have got a file of data representing the different evolution of the variables. Firstly, we compute the coupling matrix of the system and secondly the transmitted information between each variable and the global system. The pertinent variables are deduced by the combination of these two results. The hypothesis are:

- Dynamic system,
- Number of variables: 28,
- \(N=2042\),
- Normal operation mode.

The coupling dynamic matrix is given in figure 3. This figure shows the different links and causality between the variables. We can see some variables which are more or less coupled with the others. For instance: the variables X1, X2, X3 are more coupled together and less coupled with the other ones. If we look at the schematic of the system, these variables corresponds to the same sub-system. If we take the variable X24, we see that it is strongly coupled to the variables X21, X22, X23, X25 which correspond to the washing column and not so much with the others. In fact this variable is acting directly on this column by providing it the necessary pressure. We can see that the variables X8, X14, X15 and X26 are strongly coupled with all the variables of the system. We deduce that they should play a big part in the system, because they are involved in each part of it. The causality aspect could be deduced in looking this matrix too. For instance we can see that the variables X8, X9, X10, X11 and X12 related to the extraction column, are influenced by X4 and not inversely. This is normal, because if we look at the system, that is the pressurizer (X4) which is acting on the column by providing it the necessary pressure to make the chemical reaction in it more efficient. That is some interpretation of this matrix which represent a big
field of information we can get. The figure 4 shows the transmitted information between the global system and each variable.

Figure 4 shows the transmitted information between the global system and its variables. It represents the transmitted information between the system and its different part. The weight on the arrows represents different degrees of interaction between the system and its variables. More the degree is high, more the corresponding variable has much interaction with the system. We can see that the variables which have more interactions with the system are: X8, X14, X15, X26. The matrix coupling gives the same variables, it corresponds to those which are coupled with all the variables of the system. We deduce then these variables are important and pertinent since they are acting in every part of the system. The human operator should give one's whole mind to these variables because if some problems appear on these variables, it will have repercussions on all the others and consequently on each part of the system.
6. Conclusion

We have presented in this paper an application for the extraction of pertinent information of the reprocessing of nuclear waste using the information theory and based on a preliminary functional analysis of the process. This technique offers a good tool for the analysis of complex systems. However, the results depend firstly, on the quality of the functional analysis and secondly, on the data which must be representative of the system evolution. Through the example, we see that the information theory gives useful information about the interactions within the system using the dynamic coupling matrix. This matrix gives a lot of information about the internal organisation of the system. It shows the links between variables and its different interactions, the causality in the system and allows to detect on which part of the system the variables are acting more or less than the others. The information theory allows us to show the interaction between the system and its different parts. This is done by computing the transinformation between the system and its variables. It allows us to see the degree of the variable influence on the global system, more the degree is high more this variable should be important to be supervised. This method is really a complement to knowledge extraction from process experts and operators for the choice of information to display to human supervisors.

This paper has presented some results of what the information theory can offer to solve the problem of analysing complex systems. A lot of perspectives can be carried out from this theory for the design of supervision systems really adapted to human beings. For instance the detection of sub systems on a complex systems[Dufour, 76], and the choice of new variables aggregating several variables can be thought of doing.

References


Subjective Evaluation of Task Allocation: An Application of the Analytic Hierarchy Process

Z.G. Wei, A.R. Paauw and P.A. Wieringa

Man-Machine Systems Group, Laboratory for Measurement and Control
Faculty of Mechanical Engineering and Marine Technology, Delft University of Technology
Mekelweg 2, 2628 CD Delft, The Netherlands

Abstract: This paper describes an application of the Analytic Hierarchy Process (AHP) in subjective evaluation of task allocation between human and automation for a simulated process. The task allocation was evaluated by human subjects with respect to easy operation, human control confidence, comfortable operation, and situation awareness. The experimental results reveal that the AHP can be used to assess human mental load and to give a reasonably consistent ranking of task allocation decisions based on different criteria. However, the AHP only provides a reference for the evaluation. As the number of alternatives increases, the implementation of the AHP will become more difficult.

1. INTRODUCTION

Task allocation between human operators and automation in the control room of complex systems is one of the important phases in human-machine system design (Grote et al., 1995). The evaluation of decisions of task allocation is a necessary step in the task allocation phase. The evaluation involves many aspects, such as performance requirements, operational safety, cost and benefits, or more details, operational difficulty, human mental load, the level of automation and situation awareness, etc. Depending on system design requirements, the evaluation process employs different techniques, such as subjective evaluation and objective evaluation, workload analysis, part-task simulation as well as full scale simulation (Rouse and Cody, 1986). For the subjective evaluation, human subjects must be employed. After they have operated the system the subjects are asked to rate the allocation decisions based on the given criteria. The data from this rating can be processed statistically. However, if the population of subjects is small and the data does not satisfy the conditions for statistical analysis, one may have difficulty. In this paper, we present an application of a so-called Analytic Hierarchy Process (AHP) approach (Saaty, 1980) in subjective evaluation of task allocation decisions based on a simulated process.

2. METHOD

2.1 Experimental set-up

A simulation of a juice pasteurization plant (developed originally by Muir, 1989; and used by Huey, 1989, and Lee and Moray, 1992) was used as the experimental task. The plant employed in this study was adapted from the version used by Huey (1989) and some new subtasks were defined comparing with the old plant. Fig. 1 depicts the simulated pasteurization plant which was also the human-machine interface presented to the operator and was built on a Pentium PC using a graphic language.
The pasteurization process was to heat the raw juice to a level where bacteria in the juice could be killed. As shown in Fig. 1, the plant consisted of two loops. One was the juice circulation loop, and another one was the steam circulation loop.

In the juice circulation loop, the raw juice entered the input tank through the inflow pipe with a certain rate and temperature. From the input tank, the raw juice was pumped by the feedstock pump through the secondary heater and the primary heater to a distribution system. The secondary heater was a passive heat exchanger and was warmed by the hot pasteurized juice. The temperature, $T_d$, of the juice exiting the primary heater, must be maintained between 70 °C and 85 °C. If $T_d$ was within this range, the juice was led to the distribution tanks through a 3-way valve. If $T_d$ was below the allowed temperature, the heated juice was recycled to the input tank through the 3-way valve, unless the level of the input tank was high. If the level of the input tank was too high, the juice was diverted to the waste tank. If $T_d$ was above the allowed temperature (the heated juice was burnt), the juice was directed to the waste tank.

In the steam circulation loop, the steam was first pumped to the primary heater where the heat exchange between the juice and the steam took place, and then was sent back to the boiler. The temperature of the primary heater, $T_j$, could be controlled by adjusting the steam rate and the temperature of the steam heater. The heater temperature was controlled by changing the supplying amount of the energy entering the heater.

2.2 Experimental task

The experimental task was to pasteurize as much of the available raw juice as possible, which meant to maintain the pasteurization temperature with the allowed range, and to distribute it to an available tank. The experimental task could be broken into subtasks, or sub-subtasks, and consisted of a number of sub-subtasks.
monitoring and control subtasks. In the experimental task, the important variables to be controlled were the temperature $T_d$, and the juice flow rate. $T_d$ depended mainly on the juice flow rate and the temperature of the steam heater. The juice flow was maintained to keep the input tank neither empty nor overflow. The control of the feedstock pump speed, of the steam pump, and of the steam heater could be defined as subtasks. In the present study, three control subtasks were defined and could be allocated either to the operator or to the automation.

**Subtask 1:** Control of juice flow rate and input tank level.
This subtask included two components: feedstock pump speed control and overflow valve control. The speed of the feed pump affected the juice flow which would further affect $T_d$, and the level of the input tank. The overflow valve should be opened to deliver the juice to the waste tank if the juice level of the input tank was too high. The input tank was prevented from empty by reducing the feedstock pump speed.

**Subtask 2:** Control of the temperature of steam entering the primary heater, $T_j$.
This subtask could be further divided into the heater temperature control and the steam pump control. In the control of $T_j$, the steam heater control was a rough adjustment, and the steam pump control was a fine adjustment.

**Subtask 3:** Storage distribution control.
Three storage tanks were used to store the pasteurized products. These tanks might be emptied irregularly by tanker trucks. If one of the tanks reached its maximum capacity, the surplus product was diverted to the waste tank. In order to maintain a high production, this should be avoided. Moreover, the effort should be made to let the storage tanks not empty. Otherwise, alarm would appear, and it would be difficult to maintain the maximum capacity of other tanks. It is obvious that this subtask did not affect the pasteurization process, but it did influence the amount of the pasteurized juice.

From the above definitions, it can be found that the plant provides two types of tasks. The control of the pump speed and the steam temperature is a continuous task, while the control of the overflow valve and the distribution belongs to the discrete task. Because all the subtasks could be allocated to automation, each subtask had an automatic controller. Each controller worked under its control algorithm and acted independently with each other.

### 2.3 Operator’s task

When the operator controlled the plant, even though he performed only one subtask, he had to fulfill one of the following three objectives: (1) to maintain the temperature of the juice flowing out of the primary heater, (2) to maintain the input tank at a proper level, and (3) to distribute the pasteurized juice and to keep all storage tanks neither full nor empty. The ultimate goal was to pasteurize as much of the available raw juice as possible and to minimize the volume of the waste tank.

During the experiment, the operator was asked to assess mental load associated with the operation of the overall plant. After he completed all the experimental sessions, the operator was asked to evaluate the task allocation decisions by using pair-wise comparisons.

### 2.4 Experimental sessions

In the experiment, allocation of task was carried out on the basis of three subtasks. Thus, in total, there were 8 experimental sessions, including the fully automated session. Table 1 presents these sessions with the configurations of task allocation.
Table 1 Experimental sessions (A: an automated subtask; M: a manually controlled subtask)

<table>
<thead>
<tr>
<th>Session number</th>
<th>Task allocation between human and automation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subtask 1</td>
</tr>
<tr>
<td></td>
<td>feedstock</td>
</tr>
<tr>
<td></td>
<td>pump control</td>
</tr>
<tr>
<td>E₀</td>
<td>A</td>
</tr>
<tr>
<td>E₁</td>
<td>A</td>
</tr>
<tr>
<td>E₂</td>
<td>M</td>
</tr>
<tr>
<td>E₃</td>
<td>A</td>
</tr>
<tr>
<td>E₄</td>
<td>A</td>
</tr>
<tr>
<td>E₅</td>
<td>M</td>
</tr>
<tr>
<td>E₆</td>
<td>M</td>
</tr>
<tr>
<td>E₇</td>
<td>M</td>
</tr>
</tbody>
</table>

2.5 Subjects

Nine students (male and female) from the Faculty of Mechanical Engineering and Marine Technology of Delft University of Technology participated voluntarily as operators in the experiments. They received a fee for their participation. Before formal sessions started, the subjects performed 7 training sessions of 5 minutes each to learn to control the system and to practice to fill out mental load assessment forms. After completing the training sessions, the subject performed a 10 minute fully manually controlled session. The performance was evaluated by the experimenter based on the percentage of the pasteurized juice. If the score was too low (<80%), the subject was asked to perform more training sessions. Only when a subject was qualified to work as an operator, was he, or she, allowed to start formal sessions. Each subject performed therefore 7 formal sessions of 15 minutes each.

2.6 Performance measurement and mental load assessment

During the experiment, the following main plant variables were recorded: (1) the amount of the input raw juice; (2) the amount of the pasteurized juice; (3) the amount of the wasted juice.

The system performance was defined as the percentage of the properly pasteurized juice -- the ratio between the amount of the pasteurized juice and the available input raw juice during the experimental period.

During the experiment, the subjects were asked to assess their mental load for controlling the plant. In the experiment, a subjective rating scale, Rating Scale Mental Effort, RSME (Zijlstra, 1993), was used to assess mental load. Immediately after each session, subjects gave numbers according to the RSME that corresponded best to the mental load for controlling the whole plant.

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

In brief, the AHP is structured to encompass the basic elements of a decision. As shown in Fig. 2, the objective of a problem is placed at the highest level, then to subobjectives affecting the objective followed by criteria in the next level and so on, from more general to the more specific. For each level of the hierarchy, the AHP breaks up multiple options into a series of paired comparisons which are then recombined to produce an overall weighted ranking. The subjects are required to compare all possible combinations. The results of the comparisons are placed in a pair-wise comparison matrix which is called judgment matrix. The rows and columns of the judgment matrix are headed by the options included in the comparison. The principal eigenvector for the matrix is calculated, which gives a weighted ranking scale for each option.

For each level of Fig. 2, data in the AHP is collected using a series of the comparisons between each pair of design options at this level. Saaty (1980) and Mitta (1993) suggest a format which presents the two options to be compared on opposite ends of a 17 slot rating scale. The scale is a measure of dominance of one alternative over the other. It uses five descriptions in a pre-defined order and allows a single space between each one for comparison. The descriptions are “equal”, “weak”, “strong”, “very strong”, and “absolute”. Of course, these terms can be modified to provide easier comprehension for a particular subject group. Fig. 3 shows such a scale used in the present study to compare task allocation configurations in two experimental sessions, $E_i$ and $E_j$ (i

**Figure 2** A hierarchic structure
\(<j, i, j = 1,2, \ldots, 7\), based on the criterion of easy operation (see Table 3).

\[ E_i \text{ is easier to operate} \quad \quad E_j \text{ is easier to operate} \]

| \( E_i \) | \( \text{absolutely easier} \) | \( \text{much easier} \) | \( \text{easier slightly} \) | \( \text{same} \) | \( \text{slightly easier} \) | \( \text{much easier} \) | \( \text{absolutely easier} \) |
| \( E_j \) |

\[ \begin{align*}
\text{(absolute)} & \quad \text{(very strong)} & \text{(strong)} & \text{(weak)} & \text{(equal)} & \text{(weak)} & \text{(strong)} & \text{(very strong)} & \text{(absolute)} \\
\end{align*} \]

**Figure 3** Dominance scale used for pair-wise comparison of two task allocation schemes

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

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

\[
\begin{bmatrix}
E_1 & E_2 & \cdots & E_n \\
1 & m_{12} & \cdots & m_{1n} \\
1/m_{12} & 1 & \cdots & m_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
1/m_{1n} & 1/m_{2n} & \cdots & 1 \\
\end{bmatrix}
\]

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

- \( m_{ij} = 1 \) if \( E_i \) and \( E_j \) are of equal strength;
- \( m_{ij} = 3 \) if \( E_i \) weakly dominates \( E_j \);
- \( m_{ij} = 5 \) if \( E_i \) strongly dominates \( E_j \);
- \( m_{ij} = 7 \) if \( E_i \) very strongly dominates \( E_j \);
- \( m_{ij} = 9 \) if \( E_i \) absolutely dominates \( E_j \).

Scale values 2, 4, 6, 8 will reflect compromises between ratings of equal strength and weak dominance, weak dominance and strong dominance, and so on, respectively.

Matrix \( M \) has a reciprocal structure, i.e.:

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

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

$$C.I. = \frac{\lambda_{\text{max}} - n}{n - 1}$$  \hspace{1cm} (2)

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

After all the pair-wise comparisons are completed, an overall ranking scale can be calculated. A frequently used approach to calculate the ranking scale is the eigenvector method. In this method, the ranking scale is produced by calculating the principal eigenvector of $M$. According to the matrix theory, the principal eigenvector of a matrix, $w = [w_1 \ w_2 \ ... \ w_n]^T$, corresponds to the largest positive eigenvalue of the matrix, i.e. $\lambda_{\text{max}}$, and is determined by solving Eq. 3:

$$M \cdot w = \lambda_{\text{max}} w$$  \hspace{1cm} (3)

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

$$\sum_{i=1}^{n} w_i = 1$$  \hspace{1cm} (4)

The principal eigenvector $w$ can be considered as a vector in the space of the $n$ different options where the magnitude of the components in each direction is a measure of the strength of the respective option. The degree of dominance between two options is the ratio of their $w$ components.

For Subject $k$, a judgment matrix $M^k$ can be obtained and so does a principal eigenvector $w^k$. Thus, a matrix $W$ containing the rankings of individual subjects can be constructed from the eigenvectors $w^k$ of all subjects. Supposing that there are $s$ subjects and $n$ options to be compared in the experiment, $W$ would be an $n \times s$ matrix:

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

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

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

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

where $s$ is an $s$ vector and $s$ is the number of subjects who participate the evaluation.
The final overall ranking in which the subject contributions are taken into account is:

\[ r = W \cdot s \]  \hspace{1cm} (7)

\(r\) should be normalized if it does not sum up to 1.0. Thus, the entries of \(r = [r_1, r_2, ..., r_n]^T\) provide weights for the \(n\) options as well as the relative differences between two options. Fig. 4 summarizes the AHP procedure.

The degree of dominance between two is represented by the ratio of their weights, \(r_i\) and \(r_j\). Based on this ratio, the dominance of one over another one can be converted into a qualitative description in the terms presented in Fig. 3. Table 2 shows the conversion.

![Diagram of the AHP evaluation process]

Figure 4 Scheme of the AHP evaluation process

2.8 Evaluation criteria and procedure

The attributes, or criteria, used in the evaluation are listed in Table 3.
Table 2 Conversion of ratio value to qualitative description
(adapted from Yang and Hansman, 1995)

<table>
<thead>
<tr>
<th>( r/r_i )</th>
<th>Dominance of option i over option j</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>equal</td>
</tr>
<tr>
<td>3</td>
<td>weak dominance</td>
</tr>
<tr>
<td>5</td>
<td>strong dominance</td>
</tr>
<tr>
<td>7</td>
<td>very strong dominance</td>
</tr>
<tr>
<td>( \geq 9 )</td>
<td>absolute dominance</td>
</tr>
</tbody>
</table>

Table 3 Subjective evaluation attributes

<table>
<thead>
<tr>
<th>Number</th>
<th>Attributes</th>
<th>Descriptions</th>
<th>Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Easy operation</td>
<td>The overall system is easy to control</td>
<td>Which system is easier to control?</td>
</tr>
<tr>
<td>2</td>
<td>Human control confidence</td>
<td>Human is actually in the control of the system, rather than automation</td>
<td>In which system do you feel more that you are in control instead of automation?</td>
</tr>
<tr>
<td>3</td>
<td>Comfortable operation</td>
<td>Operators feel uncomfortable because too much automation has been used</td>
<td>With which system do you feel more comfortable to control?</td>
</tr>
<tr>
<td>4</td>
<td>Situation awareness</td>
<td>Whether the operator is aware of basic system state variables and understands them</td>
<td>For which system do you know the most about the system status (all necessary variables)?</td>
</tr>
</tbody>
</table>

Attribute 1 was chosen as an indicator for assessing the workload level. We expected that the results should be comparable with the mental load rating based on the RSME. Since the operator often complained, in a highly automated system, that he was not in the control of the system, but automation. The operator lost the confidence in controlling the plant. Attribute 2 was used to obtain the subjective opinion on this issue. Too much automation could make the human operator uncomfortable in operating a system. It was our interest, by using Attribute 3, to have a quantitative estimation on how a high level of automation affected the operator’s feeling in the operation. Situation awareness has become an important aspect in the design of a human-machine system (Endsley, 1995). In this experiment, by using Attribute 4, we tried to perform a subjective evaluation on the situation awareness associated with different levels of automation.

The hierarchic structure for the evaluation in this study was similar to Fig. 2, where \( n = 7 \) was the number of the experimental sessions. Because two of the nine subjects did not have time to complete the evaluation, the number of subjects, \( s \), was seven. Each subject was asked to carry out pair-wise comparisons after all of the sessions were performed.
3. RESULTS

3.1 System performance and mental load

The average percentage of pasteurized juice and the average mental load for each experimental session are presented in Table 4, from which one can see that there is not significant difference between the system performance for different task allocation configurations, but the mental load perceived by the operators appears very differently.

<table>
<thead>
<tr>
<th>Session number</th>
<th>Manual operation</th>
<th>System performance</th>
<th>Mental load (RSME)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>E1</td>
<td>Steam loop</td>
<td>0.92</td>
<td>0.03</td>
</tr>
<tr>
<td>E2</td>
<td>Flow loop</td>
<td>0.88</td>
<td>0.04</td>
</tr>
<tr>
<td>E3</td>
<td>Distribution</td>
<td>0.90</td>
<td>0.00</td>
</tr>
<tr>
<td>E4</td>
<td>Steam + Distribution</td>
<td>0.93</td>
<td>0.03</td>
</tr>
<tr>
<td>E5</td>
<td>Flow + Distribution</td>
<td>0.85</td>
<td>0.05</td>
</tr>
<tr>
<td>E6</td>
<td>Flow + Steam</td>
<td>0.85</td>
<td>0.09</td>
</tr>
<tr>
<td>E7</td>
<td>Fully manual</td>
<td>0.79</td>
<td>0.14</td>
</tr>
</tbody>
</table>

3.2 AHP Subjective evaluation of task allocation decisions

For easy operation, i.e. Attribute 1, the easy level was compared. Matrix $W_{A1}$ shows each subject’s evaluation on how easy to operate the plant based on a task allocation configuration. A large value means an easier operation. Each column of $W_{A1}$ represents a subject, and each row represents an option, in an order: $E_1$, $E_2$, $E_3$, $E_4$, $E_5$, $E_6$, and $E_7$.

$$W_{A1} = \begin{bmatrix}
0.3944 & 0.1742 & 0.1434 & 0.3631 & 0.2150 & 0.2335 & 0.1422 \\
0.2434 & 0.1576 & 0.2207 & 0.2747 & 0.1134 & 0.1312 & 0.1252 \\
0.1796 & 0.4916 & 0.5084 & 0.2123 & 0.5116 & 0.4650 & 0.5441 \\
0.0937 & 0.0614 & 0.0443 & 0.0390 & 0.0538 & 0.0521 & 0.0523 \\
0.0480 & 0.0678 & 0.0443 & 0.0665 & 0.0306 & 0.0766 & 0.0900 \\
0.0236 & 0.0305 & 0.0262 & 0.0285 & 0.0594 & 0.0268 & 0.0231 \\
0.0173 & 0.0169 & 0.0126 & 0.0159 & 0.0161 & 0.0148 & 0.0231
\end{bmatrix}$$

When the ranking of subjects is taken into account, we substitute $W_{A1}$ and Eq. 6 with $s = 7$ into Eq. 7 and yields the vector specifying the operational difficulty ranking for task allocation configurations which were represented by the experimental sessions as shown in Table 1:

$$r_{A1} = w_{A1} \cdot s = [0.238 \ 0.181 \ 0.416 \ 0.057 \ 0.061 \ 0.031 \ 0.017]^{T}$$
For human control confidence, i.e. Attribute 2, the subjects were asked to compare for which configuration they felt more in the control rather than automation. $W_{A2}$ shows subjects' rating on how much they felt that they were in the control. A large value indicates that the operator felt more control.

$$W_{A2} = \begin{bmatrix} 0.0487 & 0.0519 & 0.1429 & 0.0348 & 0.0452 & 0.0281 & 0.0551 \\ 0.1119 & 0.0663 & 0.1429 & 0.0941 & 0.0481 & 0.0539 & 0.0551 \\ 0.0388 & 0.0241 & 0.1429 & 0.0157 & 0.0222 & 0.0189 & 0.4958 \\ 0.0867 & 0.1157 & 0.1429 & 0.0396 & 0.0734 & 0.1269 & 0.0985 \\ 0.1606 & 0.1105 & 0.1429 & 0.1077 & 0.0967 & 0.1190 & 0.0985 \\ 0.2736 & 0.2440 & 0.1429 & 0.3489 & 0.3398 & 0.2677 & 0.0985 \\ 0.2798 & 0.3875 & 0.1429 & 0.3593 & 0.3747 & 0.3856 & 0.0985 \end{bmatrix}$$

After the rankings of all subjects being considered, the vector specifying the ranking of the human control confidence ranking for task allocation configurations is:

$$r_{A2} = w_{A2} \cdot s = [0.058 \ 0.082 \ 0.108 \ 0.098 \ 0.119 \ 0.245 \ 0.290]^T$$

For comfortable operation, the subjects were asked to compare for which configuration they felt more comfortable in the operation of the plant. $W_{A3}$ shows subjects' rating on how comfortable they felt. A large value indicates that the operator felt more comfortable.

$$W_{A3} = \begin{bmatrix} 0.2885 & 0.3417 & 0.1748 & 0.0536 & 0.1988 & 0.4103 & 0.3568 \\ 0.1184 & 0.2138 & 0.1748 & 0.1724 & 0.1224 & 0.2689 & 0.2387 \\ 0.0217 & 0.0809 & 0.5041 & 0.0166 & 0.4427 & 0.0479 & 0.0190 \\ 0.3044 & 0.1426 & 0.0546 & 0.0543 & 0.1217 & 0.1502 & 0.1123 \\ 0.1251 & 0.1232 & 0.0546 & 0.1427 & 0.0356 & 0.0675 & 0.0871 \\ 0.0729 & 0.0571 & 0.0247 & 0.2906 & 0.0644 & 0.0375 & 0.1098 \\ 0.0691 & 0.0408 & 0.0124 & 0.2698 & 0.0145 & 0.0177 & 0.0762 \end{bmatrix}$$

The vector specifying the ranking of the effect of automation on human by combining all subjects' judgments is as follows:

$$r_{A3} = w_{A3} \cdot s = [0.261 \ 0.187 \ 0.162 \ 0.134 \ 0.091 \ 0.094 \ 0.072]^T$$

For situation awareness rating, the subjects were asked to compare for which configuration they knew more about the plant status. $W_{A4}$ shows each subject's rating. A large value indicates that the operator knew more variables and had more situation awareness.

$$W_{A4} = \begin{bmatrix} 0.1405 & 0.0753 & 0.0643 & 0.1040 & 0.0651 & 0.0988 & 0.0435 \\ 0.0842 & 0.0683 & 0.0432 & 0.1336 & 0.0559 & 0.1179 & 0.0435 \\ 0.0182 & 0.0245 & 0.0141 & 0.0186 & 0.0216 & 0.0235 & 0.0435 \\ 0.0949 & 0.0953 & 0.0559 & 0.0562 & 0.0858 & 0.1621 & 0.0435 \\ 0.1065 & 0.0769 & 0.0954 & 0.2084 & 0.0984 & 0.1212 & 0.0435 \\ 0.2352 & 0.3248 & 0.3635 & 0.2396 & 0.2804 & 0.2099 & 0.3913 \\ 0.3205 & 0.3349 & 0.3635 & 0.2396 & 0.3929 & 0.2756 & 0.3913 \end{bmatrix}$$

The vector specifying the ranking of the situation awareness from a combination of all subjects' judgments is:
4. DISCUSSION

4.1 AHP evaluation and mental load

Easy operation and mental load. The relationship of the easy level in the operation (AHP) and the measured mental load (RSME) appears linear as shown in Fig. 5. The correlation coefficient between two was -0.963, p < 0.001.

![Figure 5 Relationship between easy operation and mental load](image)

Fig. 5 indicates that the easy operation level based on the AHP evaluation agreed quite well with the subjective assessment of mental load. The easiest controlled system had a lowest mental load. As the easy level increased, the mental load decreased accordingly. The only exception in the order of the two measures was in the cases of steam loop plus distribution control and feed loop plus distribution control. Both the mental load and the AHP values in these two configurations were very close. This means that the two configurations were at the same level. The results presented Fig. 5 have further demonstrated that the AHP approach could be used as a method for the mental load assessment as suggested by Vidulich and Tsang (1987).

Given vector $r_{A_4}$ in Section 3.2, one can find that $E_3$, where only distribution was manually controlled, was ranked as the easiest configuration to operate, and $E_7$, where all subtasks were manually controlled, was ranked the most difficult to control by a factor of at least (with respect to $E_6$) 1.82. $r_{A_1}$ also tells us that subjects perceived $E_3$ as being easier to control than $E_1$, $E_2$, $E_4$, $E_5$, $E_6$, and $E_7$ by factors 1.75, 2.23, 6.82, 7.30, 13.42, and 24.47, respectively. The degree of the dominance of $E_3$ over $E_i$ and qualitative interpretations can be found in Table 2.

Human control confidence and comfortable operation. These two criteria aimed at evaluating how the level of automation affected human operator’s feeling during operation. Vector $r_{A_2}$ indicates that $E_2$ was ranked as the most human control configuration, and $E_1$, where the steam loop was manually controlled, was ranked the least to control by the operators by a factor of at least (with respect to $E_2$) 1.41. Vector $r_{A_2}$ also indicates that subjects perceived $E_7$ as having more human control than $E_1$, $E_2$, $E_3$, $E_4$, $E_5$, and $E_6$ by factors 5.0, 3.54, 2.69, 2.96, 2.44, and 1.18, respectively.
Vector \( \mathbf{r}_{A3} \) indicates that \( E_1 \) was ranked as the most comfortable configuration to control because of automation, and \( E_7 \) was ranked the least comfortable to control by a factor of at least (with respect to \( E_5 \)) 1.26. Vector \( \mathbf{r}_{A3} \) also indicates that subjects perceived \( E_1 \) as being more comfortable than \( E_2, E_3, E_4, E_5, E_6 \), and \( E_7 \) by factors 1.40, 1.61, 1.95, 2.87, 2.78, and 10.07, respectively.

Comparing vector \( \mathbf{r}_{A2} \) with vector \( \mathbf{r}_{A3} \), one can find that the ranking of the human control confidence perceived by the subjects was opposite to the ranking for the operational comfort. \( E_1 \) was ranked as the most comfortable configuration, but was the least in the human control confidence. \( E_7 \) was ranked as the most in the human control confidence, but the least comfortable configuration. This implies that a compromise has to be made between these two aspects. An interesting point is that \( E_3 \) was the easiest to control, but was not the most comfortable configuration. This means that a configuration with a low workload does not necessarily make the operator feel comfortable in the operation.

**Situation awareness.** Task allocation between human and automation will directly affect operator’s awareness of the system situation. For the subjective evaluation of the situation awareness in this study, vector \( \mathbf{r}_{A4} \) indicates that \( E_7 \) had the highest situation awareness, and \( E_3 \) had the lowest situation awareness by a factor of at least (with respect to \( E_2 \)) 3.39. Vector \( \mathbf{r}_{A4} \) also indicates that subjects perceived that for \( E_7 \), the situation awareness was higher than for \( E_1, E_2, E_3, E_4, E_5, \) and \( E_6 \) by factors 3.89, 4.24, 14.39, 3.89, 3.09, and 1.14, respectively. This is understandable because in \( E_7 \) the operators performed all subtasks and monitored all necessary system variables.

### 4.2 Synthesis of the subjective evaluation

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

\[
\mathbf{M}_C = \begin{bmatrix}
A_1 & A_2 & A_3 & A_4 \\
1 & 2 & 3 & 1 \\
A_2 & 1/2 & 1 & 1/2 \\
A_3 & 1/3 & 1/2 & 1 & 1/3 \\
A_4 & 1 & 2 & 3 & 1 
\end{bmatrix}
\]

The principal eigenvector of \( \mathbf{M}_C \) is:

\[
\mathbf{r}_C = [0.351 \ 0.189 \ 0.109 \ 0.351]^T
\]

The consistent index of the comparison is C.I. = 0.0035 which indicates the paired comparison is consistent.

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

\[
\mathbf{R}_{TA} = [\mathbf{r}_{A1} \ \mathbf{r}_{A2} \ \mathbf{r}_{A3} \ \mathbf{r}_{A4}]
\]

Thus, the final rank specifying all task allocation decisions \((E_1, ..., E_7)\) by combining four criteria is:

\[
r_F = \mathbf{R}_{TA} \cdot \mathbf{r}_C = [0.153 \ 0.127 \ 0.192 \ 0.083 \ 0.086 \ 0.170 \ 0.185]^T
\]
From vector $\mathbf{r}_k$, one can find that $E_3$ with the distribution subtask being manually controlled, and $E_7$ with all subtasks being allocated to the operator have a high ranking. However, we must take into account all aspects in the evaluation before making a decision. For example, $E_3$ was ranked an extremely high score at Attribute 1, which effects a high score in the final ranking. However, at Attribute 4, $E_3$ got the lowest ranking. $E_7$ was the most difficult configuration to operate, i.e. it demanded a very high mental load, although it got the highest ranking at Attribute 2 and 4. Both $E_3$ and $E_7$ had a ranking at two extreme ends of Attributes 1 and 4 in an opposite order, and they could not be considered as an optimum task allocation. In the final ranking $E_1$ and $E_6$ had a high ranking, but they did not have an extremely high or low ranking at any attributes. Thus, they could be considered as an optimum task allocation in the control of the pasteurization plant. If the system performance as well as the mental load as shown in Table 4 were taken into account, $E_1$ might be the best task allocation decision.

4.3 Consistency of the judgments

One of the advantages of the AHP is its capability to calculate consistency in human judgments. The consistency indices for all the subjects in the evaluation were calculated by Eq. 2, and are listed in Table 5.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Consistent Index (C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subj. 1</td>
</tr>
<tr>
<td>Easy operation</td>
<td>0.318</td>
</tr>
<tr>
<td>Human control confidence</td>
<td>0.135</td>
</tr>
<tr>
<td>Comfortable operation</td>
<td>0.148</td>
</tr>
<tr>
<td>Situation awareness</td>
<td>0.171</td>
</tr>
</tbody>
</table>

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

The consistency indices of each subject are very helpful in evaluating each subject's ability to judge. So, a rank, $s$, for all subjects could be obtained. The judgment of the subjects that were least consistent could be eliminated by the individual preference. Thus, although the data from all subjects are included in the final ranking, the judgments from the subjects with the least judgment consistency will have less effect than the subjects with the high consistency. However, in the
present study, each subject was given an equal weight to avoid biased judgments although some
of the subjects had demonstrated less capabilities of providing a sound judgment.

4.4 Problems in the AHP application

The advantages and limitations of the AHP approach have been addressed by Mitta (1993). Here, we
only address the problems met in the application of the AHP for subjective task allocation evaluation.
In our application, we met with three problems. The first one was that the number of pair-wise
comparisons increased greatly as the number of options and the number of the attributes. This was a
very heavy burden to the subjects who felt very boring during paired comparisons. This made it
difficult for the subjects to have a reasonable consistency in their judgments.

The second problem deals with the combination of the individual rankings. When a final ranking for all
options was obtained by combining the ranking from each attribute, an extremely high ranking of an
option in one of the attributes might play a significant role, even though the weight for this attribute
was not so large. For example, E3 was ranked as the easiest configuration to operate, and, according to
Table 2, it dominated strongly over most of the other configurations. Therefore, at the final ranking, E3
ranked very high even though other attributes had been considered in a weight approach. When we
took into account other attributes, E3 could not be an optimum task allocation configuration. From this
example, we could conclude that the AHP ranking can only be taken as a reference, and other aspects
have to be taken into account when making a final decision.

The third problem was that for task allocation evaluation, the consistency of the subjective judgments
were difficult to reach. When the number of the options is large, or the attributes are not simply defined
and not clearly expressed, it will be more difficult to have a reasonable consistency. Thus, to carry out
the task allocation evaluation by using the AHP, it is better to reduce the options as many as possible,
and to make the evaluation criteria be easily understood.

5. CONCLUSION

The paper presents an application of the Analytic Hierarchy Process in the subjective evaluation of task
allocation decisions between human and automation. From the experimental results, the following
conclusions can be drawn:

• A subjective ranking of different task allocation decisions can be obtained using the AHP approach
  instead of a statistical approach; but,
• The AHP ranking can only be taken as a reference in the evaluation of task allocation;
• A configuration which is easy to control does not necessarily make the human operator
  comfortable;
• The AHP can be used to measure the relative level of human workload.

REFERENCES:

functions in automated work systems and their use in simultaneous engineering projects.


A METHOD FOR IDENTIFYING COUPLING BETWEEN HUMANS AND MACHINES.

Neville Moray

LAMIH-PERCOTEC, Universiti de Valenciennes, B.P.311, Le Mont Houy, 59304 VALENCEIENNES CEDEX, FRANCE Phone (+33)-27-14-14-36: Fax (+33)-27-14-12-94 Home phone: (+33)-20-55-09-53, email at university: nmoray@univ-valenciennes.fr

Abstract

In 1976 Conant introduced a new method for identifying the natural decomposition of complex systems. It is based on the analysis of the transmission of Shannon information among the subsystems and components of a large system, and can identify the "molecular structure" of complex systems, that is, the natural way in which components form "clumps" which are subsystems of the main system. The method will be briefly described.

In the past Conant has applied it to economic systems, weather systems, etc., and Conant and Moray used it to analyse the acquisition of a teleoperation skill.

Recently we have applied it to the analysis of human-machine interaction in a (somewhat) complex system, and have been able to describe the difference between poor and good operators by identifying the different ways in which the operators couple themselves to the physical system. We believe that Conant's method will prove an interesting addition to the methods of analysing human-machine system, including perhaps the identification of operator mental models of complex systems. Some preliminary results will be described.
Session 5

Human Reliability Assessment
MECCA: Medical Errors and Complications Causal Analysis

T.W. van der Schaaf
Safety Management Group
Eindhoven University of Technology
P.O. Box 513, Pav. U-8
5600 MB Eindhoven, The Netherlands

C.E. Shea
Department of Education
University of Manchester
England*

A.S. Nyssen
Department of Psychology
University of Liege
Belgium

Abstract

In February 1996 the MECCA project was started as a joint research programme between the Eindhoven Safety Management Group and a large teaching hospital. Its aims are to apply the Near Miss Reporting framework and PRISMA incident analysis approach, developed in the chemical process and steel industries, to the medical domain. A similar Ph.D. project was conducted at the University of Manchester and two teaching hospitals in the U.K.

The paper describes the MECCA project. Two preliminary applications of MECCA are briefly outlined: the risk management feasibility study in the surgical department of the Eindhoven hospital and the analysis of incidents in Accident and Emergency departments in hospitals in the Manchester area. Using a common framework, both studies showed a large number of human errors and even more organizational failures. The potential for the integration of MECCA with other risk management tools such as, process models and Failure Modes and Effects Analysis, is discussed. A third related project investigating human error in anaesthesia conducted at the University of Liege is also briefly outlined. The current and future plans for MECCA are outlined to conclude.

1. Introduction

This paper outlines the research project MECCA that began in February 1996 to investigate the possibilities for risk management in the medical domain. Risk management in the medical domain and more specifically medical accidents is a burgeoning area of research. Vincent, Ennis and Audley (1993) noted that at the beginning of their research in 1985 there was “virtually no research on the subject”.

*now with the Eindhoven Safety Management Group
In the last decade interest has steadily increased most notably by those in the fields of psychology and human factors. Cooperation and interest in the medical field is also increasing alongside these research initiatives. The value of such investigations is becoming more widely recognized among healthcare professionals.

This paper addresses the implementation of a risk management system in a medical domain. The potential for risk management in the medical domain is outlined with reference to the current situation in hospitals. The MECCA project and its general methodology is introduced with reference to the two initial projects that have successfully been completed.

Finally, conclusions are drawn and the future of the MECCA project is discussed.

2. The Role of Risk Management in Hospitals

The introduction of a risk management system in the medical domain is receiving increased attention for several important reasons. The resolution of legal issues is becoming steadily more expensive. The growing tendency of individuals to resort to expensive litigation to resolve problems has increased awareness of the need to accurately assess and analyze the quality of care provided in order to protect staff and improve patient care.

At the same time the ability to accurately represent and assess system function through the implementation of an holistic risk assessment or quality management program may aid hospitals in forecasting expenditures and identifying problem areas. The ability to demonstrate system efficiency may further translate into reduced insurance costs in some areas.

A self-evident stimulus for the management of risk in medicine is the protection of patients. Patients are increasingly viewed as consumers in the market of healthcare services. This role change is occurring in tandem with a trend toward greater patient awareness and participation in medical decisions regarding their own care. The “rights” of patients are being recognized as individuals ask for more information, including assurances of safety. The ability to provide such information will thus be an asset for a particular institution.

An important yet infrequently investigated area is staff safety. The ability to minimize risks to their personal safety will be an added benefit of a risk management system.

Finally, in a domain dedicated to helping individuals, all avenues to reduce risks to patients and improve the quality of their care should be pursued on ethical grounds.

3. Risk Management in Medicine

3.1 Current Situation in the Medical Domain

Risk management in the medical domain lags behind initiatives in a variety of other high-risk industries. From a safety based perspective the process industry, nuclear power and aviation fields have incorporated risk management strategies into the design, implementation and ongoing management of systems.

Conversely, there is no systematic or comprehensive approach to safety and the management of risk in medicine. This is not to suggest that the topic is ignored but that unlike similar
complex, dynamic and high-risk domains, medicine has infrequently looked beyond itself for recommendations.

Clinical audit is perhaps the most widespread method of medical assessment employed by hospitals. This method involves systematically looking at the procedures used for diagnosis, care and treatment, examining how associated resources are used and investigating the effect care has on the outcome and quality of life for the patient. (DoH 1995).

Clinical audit is generally performed through the retrospective analysis of patient cards. A particular medical issue is chosen for review such as the management of sprained ankles, in a selected department. One of three main topics is then addressed:
- timeliness of response,
- appropriateness of medical intervention,
- anticipated benefit of medical intervention.

Clinical teams composed of doctors and nurses then review current practice of the selected process, by reviewing past patient notes. Standards of care for the particular condition under audit are then established. A second audit may be performed at a later date to assess the impact of changes proposed following analysis of the results of the initial audit.

A second broad approach to quality management facilitating risk management is the TQM (Total Quality Management) or CQI (Continuous Quality Improvement) approach. Generally these involve the selection of a particular process for review. The process for improvement and possible indicators are determined collaboratively by a group of specialists. Each selected element is assigned a numeric value representing the threshold or point at which further evaluation should occur. "Rate based monitors" (Buckley 1994) correlate with a numeric value determined by a group of experts, as an acceptable level of quality. If more than a pre-established percentage of patients experience complications following a particular treatment, than a review of the situation should be made. Data is then collected and evaluated as required followed by the implementation of appropriate remedial actions. The effect of the changes should then be evaluated and the process repeated if necessary. Finally, the knowledge acquired through the process should be disseminated appropriately.

The execution of clinical audit and the assessment of a particular process using a TQM or CQI approach is systematic, however the selection of a process to be audited is in many cases discretionary. These methods also fail to account for the multitude of factors which may influence system performance by considering procedures in isolation. Research reviewed later in this paper strongly suggests that this type of approach is inadequate given the complex web of work processes involved in hospital function.

3.2 Research Initiatives / State of the Art

A variety of different approaches to the study of medical error are now underway. The approaches outlined below remain predominantly domain specific. They are included to give an overview of current work in the area of medical error and, in their specificity, highlight the need for a comprehensive risk management approach. The MECCA project is an attempt to develop a medical risk management system.

Human error in medicine is the focus of several projects. Bogner approaches medical error and specifically human error in medicine from a human factors perspective. Human Error in Medicine (Bogner 1994) brings together a number of different approaches concentrating on
the issue of human error (Moray; Helmreich and Schaefer; Cook and Woods) and begins to address the myriad forms human error may take in the medical domain.

From a different position, Medical Accidents (1993) edited by Vincent, Ennis and Audley, is intended to provide clinicians with a guide to investigate errors and accidents in medicine. In the view of the editors, the study of medical accidents and efforts to minimize their effects is vital as accidents “will inevitably still occur.” Increasing awareness of the lack of a systematic approach to accountability and compensation is also underlined in light of both the legal and financial consequences of medical accidents.

An in-depth study of human error in anaesthesia is being conducted at the University of Liège by A.S. Nyssen. (De Keyser and Nyssen 1993). The project explores the mechanisms by which an individual commits, persists and recovers an error in anaesthesia. Field work and observations of experts in a simulator have been incorporated into this investigation.

Extensive work in accident research from a safety management perspective has recently been applied to the medical domain. Wagenaar et al. have begun to investigate safety management in the Intensive Care Unit (ICU). (Wagenaar, Souverijn, Hudson 1993). This work is beginning to apply organizational safety management strategies learned in other industries, in this case oil exploration, to the medical domain.

An American project in transfusion medicine aims to design, develop and implement an ideal event reporting system for collecting and processing information on incidents related to error in transfusion medicine. Ultimately the project hopes to enhance human performance in this primarily human work domain. (Kaplan et al., 1996; Battles et al., 1996).

A project in the human factors department in industrial engineering at the University of Wisconsin is developing a questionnaire based self-reporting system for human error. This approach is intended to be included in a safety management system and will be tested in a radiology department of the teaching hospital.

4. MECCA Medical Errors and Complications Causal Analysis

4.1 Participants

MECCA is a four year joint project between a large teaching hospital in Eindhoven, the Netherlands and the Eindhoven University of Technology. A number individuals are involved from both institutions including hospital participants:

Bastien van der Hoeff, MSc., researcher
Gerard Op de Weegh, quality supervisor
2 stagères working part-time
3 staff from the surgical department working on the project part-time,

and University participants:

Christine Shea, Ph.D. post-doc
Wim van Vuuren, Ph.D. student
Tjerk van der Schaaf, co-ordinator.
4.2 The Methodology Behind MECCA

The aim of MECCA is to investigate the applicability of existing risk management tools and methods in the medical domain, specifically in hospitals. The method is based on the PRISMA approach to near miss management developed in the chemical process industry by van der Schaaf (1995). MECCA combines a predictive method in which possible risks are identified with a registration method of actual risks or incident causes.

The actual root causes and the predicted root causes are compared. This comparison is intended to act as a validation process in two ways. The predicted root causes may highlight different problem areas than those found in actual incident analysis. The failure of staff to report particular situations, possibly through fear of reprisal or simply because they do not recognize a routine failure as worthy of reporting, may be highlighted during the validation process. Similarly, construction of possible scenarios may be enhanced using the results of actual incidents as additional complementary inputs. (Figure 4.1).

Possible scenarios are developed incorporating the knowledge of domain experts. These situations are analyzed using the PRISMA approach (van der Schaaf 1992) to identify possible risks and their associated root causes. In tandem with the elaboration of possible risks, the collection of actual incident information is carried out. Incident reports may be collected in two ways, either through voluntary self-reports completed by individuals involved in an incident or through observation performed by the investigator. Confidential interviews are then performed to obtain more detailed information regarding the incident when necessary. The collected information is again analyzed using the PRISMA approach.

![Figure 4.1. Model of the methodological approach of MECCA.](Image)
5. Initial Projects in the Medical Domain

5.1 Risk Management in a Surgical Department

One of the first applications of the MECCA method was conducted by van der Hoeff (1995) in the surgical department of a teaching hospital in Eindhoven. Possible risks to patients were predicted using Failure Modes and Effects Analysis (FMEA). FMEA is particularly useful when systems may be decomposed into individual components. The systematic assessment of the failure modes of every component is performed to establish both their effects and causes. This information is registered on a form containing columns for:

- the failure mode or modes of every component,
- the effects of each failure mode,
- the severity of the effects,
- the causes of a failure mode,
- the frequency of occurrence of a certain cause,
- the extent to which the cause can be corrected.

It is essential that those performing the FMEA analysis have a clear understanding of the functions of all components and their inputs and outputs. The root causes of the predicted incidents were identified and classified using the PRISMA approach.

Actual critical incident information was gathered from 17 confidential interviews with various staff members involved in the relevant incidents in the surgical department. These incidents were analyzed using the PRISMA approach to identify actual root causes.

Two databases were created. The first included the root causes identified using FMEA while the second was constructed using the actual root causes. The distributions of root causes in the databases were then compared. Van der Hoeff found that human and organizational root causes predominated in both databases. (Figure 5.1). A statistical comparison using Spearman’s rank correlation coefficient was significant at the 5% level. The application of near miss analysis and the PRISMA approach in the surgical department was straightforward.

![Figure 5.1 Results of classification of root causes from actual incidents in a surgical department. (van der Hoeff 1995) (n = 95).](image)
5.2 An Application of MECCA in the Accident and Emergency Department

A Ph.D. project (Shea 1996) also incorporated the MECCA methodology. The aim of the project was the investigation of factors influencing the performance of work processes within the Accident and Emergency department (A&E). (Figure 5.2). The research compared the results of expert assessment of a normative process model of the A&E department with the results of actual incident analysis using the PRISMA method to identify factors negatively affecting system performance.

The normative process model of system function was developed to be used as an archetype of ideal system function. The process model was developed using a Systems based approach to process modelling. (In 't Veld 1985). The method is flexible enough to allow one to choose between a variety of different levels of aggregation when modelling a particular system. A system is modelled chronologically and may include feedback and feedforward loops.

The normative process model was developed to act as a measuring tool against which actual system function would be measured. Possible problematic processes in the organization and performance of work processes in the department were highlighted during a comparison between the actual system and the ideal process model. The assessment was conducted by a panel of domain experts.

Following a discussion with the panel a set of problematic processes was confirmed as causing difficulties in the performance of routine work processes in the A&E department. Three processes were identified:

---

Figure 5.2 Accident and Emergency department project methodology (Shea 1996)
- coordination with extra-departmental services performing investigations for example radiology and various laboratories,
- communication with other departments including other specialities (RMO, RSO/ orthopeadics etc.),
- admission of patients to hospital.

These processes were compared with the actual risks identified using near miss analysis. Incidents were collected through observations and self-reports by those involved in an incident. Flanagan's (1954) Critical Incident Technique was employed in follow-up interviews used to obtain a complete picture of all the activities and decisions that culminated in an incident.

The causes of an incident are drawn to illustrate their causal relationship to each other forming a causal tree. (Figure 5.3). The actual incident sits at the top of the tree while the root causes of an incident are drawn at the bottom. This method is useful in revealing the combination of technical, organizational and human factors which together combine resulting in an incident.

**Figure 5.3 Causal tree of an actual A&E incident.**

The root causes were classified using the Eindhoven Classification Model (ECM) (van der Schaaf 1995) which was adapted for the medical domain. Data gathered during the research suggested that organizational factors beyond the A&E department influenced system function within the A&E department. It became evident that the web of relations between the hospital and the A&E department was an influential and inescapable factor in the organization of work in the A&E department.
A method of identification of the causal factors of incidents that allowed differentiation between the organizational factors under the direct control of the A&E department and those under the control of other bodies, including senior hospital management and various external departments and services, appeared necessary and useful. In practice this division facilitated the accurate allocation of responsibility for problems. Problems that originated outside the A&E department could be identified and attributed to the correct department for attention and correction. The Classification /Action matrix (van der Schaaf 1992) was then consulted to develop solutions to the problems identified.

The results of the A&E project strongly suggested that organizational factors play a large part in departmental function. The development of a database of root causes using the extended organizational categories clearly showed this finding. (Figure 5.4). The problematic processes identified when assessing the normative process model also occurred at points where A&E processes overlapped with external departmental processes.

![Figure 5.4](image)

Figure 5.4 Results of classification of root causes from actual incidents in the A&E department.

6. Conclusion

The results of the preliminary MECCA projects suggest that an holistic approach to risk management in hospitals will be useful for management, staff and patients. The application of the PRISMA method in the framework of a near miss reporting system in hospital settings has revealed the diverse influences that result in incidents and the ability of the MECCA methodology to successfully cope with this diversity.

The MECCA project will continue for the next four years. The initial project in the surgical department of the Catherina Hospital in Eindhoven is currently being extended to other departments including radiology, neurology and pharmacy. Extension of the database to include information on factors influencing successful recovery is also underway. (van der
Schaaf, Frese and Heimbeck 1996). Initial contacts have been extended and now include collaboration with the University of Manchester, the University of Wisconsin, The University of Cardiff and the University of Texas Southwestern Medical Centre at Dallas.

Acknowledgment

The presentation of this paper was supported by a CEC Human Capital and Mobility Network on Human Error Prevention.

References


5.1-10


HUMAN RECOVERY AND ERROR MANAGEMENT

T.W. van der Schaaf
Safety Management Group
Eindhoven University of Technology
P.O.Box 513, Pav. U-8
5600 MB Eindhoven, The Netherlands

M. Frese
D. Heimbeck
Department of Psychology
University of Amsterdam
Roetersstraat 15
1018 WB Amsterdam, The Netherlands

Abstract

This paper highlights the positive role that human operators often play in preventing small failures and errors from developing into an actual system breakdown. The resulting 'near misses' may provide an insight into a powerful alternative to human error prevention, namely human recovery promotion and error management. Theoretical approaches to modelling error recovery are discussed and translated into empirical research questions. These are partly answered by a number of pilot studies. The main conclusions are that error recovery is much more than simple luck or coincidence, that root causes can be identified, and that these should have design implications for the technical and organisational context of the human operator's task as well as for an alternative training concept of error management.

1. Introduction

The research project described in this paper focuses on the positive role that human operators often play in preventing an ongoing sequence of usually small failures and errors from developing into an actual total system breakdown or accident. This new concept of human recovery may provide designers, managers and researchers with a powerful alternative approach to the traditional one of human error prevention in process control, namely: human recovery promotion and error training.

First, a simple incident causation model is presented in which the presence or absence of successful human recovery plays a decisive role in determining the effects of process deviations, technical failures and errors on the safety and reliability of man-machine systems (MMS's).
Then, the process of human recovery is described. It consists of three phases: detection of symptoms, localisation of their cause(s) and correction which returns the system to its normal status. A short introduction into the error management concept (Frese, 1991) is given.

The following section deals with the relationship of human error causes and the probability of recovery, and with the error detection lag. These theoretical predictions are mainly based on well known cognitive limitations and feedback-related aspects of the task situation. Implications for software design concerning criteria for error messages are mentioned (Zapf et al, 1991).

Four ways of classifying (human) recovery in actual process control situations are proposed. The first classification deals with the type of preceding failure(s), for instance technical, organisational or human failure respectively, with the error taxonomy developed by Frese and Zapf (1991). Another way to look at human recovery is to distinguish the reaction after symptom detection: ignore the deviating status, repeat a sequence of actions, or attempt fault localisation and correction. Thirdly and most importantly, the factors in the man-machine system that triggered or enabled recovery are categorized: technical factors related to process design (for instance to allow for reversibility), or interface design (e.g. to maximize observability of symptoms and effects); the organisational and management context (e.g. proper procedures, positive safety culture) and operator factors (e.g. accurate mental models). The fourth classification locates the phase in which a recovery factor primarily contributes to the recovery process: detection, localisation or correction. These theoretical approaches are subsequently translated into the specific empirical research questions on which this project is focusing.

Finally the results of recent pilot studies in the energy production and steel industry, as well as those of medical errors in a surgical ward will be presented and their implications for designing recovery into man-machine systems will be discussed.

2. Theoretical approaches

2.1 Incident causation model

In Van der Schaaf (1992), a simple incident causation model is used (see fig. 1) to define accidents, near misses and their common root causes consisting of technical, organisational and human (operator) factors. When incident development cannot be stopped by the system's predetermined barriers and lines of defence, the only distinguishing factor between an accident and a near miss effect is the presence or absence of successful 'accidental' or unplanned recovery.

Although actual accidents may also contain attempts at recovery, it is obvious that near misses as defined above are the optimal source of data to study the phenomenon of recovery as the positive counterpart of failure.
These failure factors (or root causes) have so far been modelled successfully by the Eindhoven Classification Model (ECM) of system failure (Van der Schaaf, 1991, 1992; Van Vuuren, 1995). In the pilot studies mentioned in section 4 of this paper, the ECM subcategories will not be used, only the main groups of Technical (T), Organisational and Management (O), and Human operator (H) failure factors will be referred to.

Fig. 1: The incident causation model.

2.2 Human recovery process phases

Van der Schaaf (1988) proposes that human recovery be defined as "the (unique?) feature of the human system-component to detect, localize and correct earlier component failures. These component failures may be either his or her own previous errors (or those of colleagues) or failing technical components (hardware and software)". This definition implies the following phases in the recovery process:

* **Detection:** of deviations, symptoms, etc.,
* **Localisation:** of their cause(s) (diagnosis in the strictest sense),
* **Correction:** of these deviations by timely, effective counter actions, after which these deviations are nullified and the system returns to a stable status.

In the theoretical framework of error management (Frese, 1991), these three phases in the recovery process correspond to the phases in the general error process: error detection, error explanation and error handling. The concept of error management (or error training) can be seen as an alternative to error prevention. The focus of this new approach is to avoid negative error consequences, to deal quickly with error consequences once they occur and to reduce future errors. Thus, an important distinction between the error per se and the error consequences has to be made. From this error management perspective three implications for system design can be deduced:

5.2-3
1. Systems should decrease the error detection time (error detection phase):
   More negative consequences will occur if errors are not detected early. Thus, the
   sooner the error is detected, the better for error management. Error training can
decrease the time between error occurrence and error detection by increasing
awareness of how many errors one makes.

2. Systems should facilitate one’s understanding of errors (error explanation phase):
   Although it is not absolutely necessary to know why an error occurred often good
error explanation helps to handle errors quickly. Systems can facilitate good error
explanation by being transparent and by giving helpful feedback.

3. Systems should reduce error handling time (error handling phase):
   Zapf, Frese et al (1991) have described various supports for error management in
software design. This includes memory aids (e.g. history function), backups, making
additional actions possible without loosing track of what one was doing before (e.g.
with a window technique), easy access to a known starting point (e.g. with the
ESCAPE key), undoing functions (e.g. unerase), direct correction (e.g. insert func­
tion), support for error search and correction (e.g. language checks), and support for
active exploration (e.g. tutorials).

2.3 Dependency of recovery on preceding errors

Embrey and Lucas (1988) discuss several factors affecting the probability of recovery
from error and the error detection lag. This relationship is highly relevant to understand
the role of feedback in the recovery mechanism. Their main points may be summarized as
follows:

* Causes of skill-based slips and lapses are relatively unrelated to subsequent recovery
   factors; their human recovery probability is high and the error detection lag will be
   small.
* For rule- and knowledge based mistakes the opposite holds: their recovery factors
   depend on the same preceding failure factors; probability of recovery is small and
   the error detection lag is large.

The main reasons for these predictions given by Embrey and Lucas (1988) include feed­
back related aspects and cognitive limitations: the awareness of an error possibility and the
visibility of its effects are high for slips and lapses, but low for mistakes while cognitive
limitations (e.g. confirmation-, fixation- and groupthink biases) would be small for slips
and lapses, but large for mistakes. For the present paper the main implication is that the
nature of the preceding human error(s) should be highly predictive of any subsequent
recovery.

Another feedback related aspect of errors is error messages. Zapf et al (Zapf, Frese,
Irmer & Brodbeck, 1991) emphasize that good error messages should conform to the
following criteria: they should be quite visible and salient, informative, easy to under­
stand, orient the user to further actions, and be polite and short.
2.4 Classification of human recovery aspects

The preceding sections lead to the following four ways of classifying (human) recovery aspects: according to the preceding failure(s), according to the human operator’s reaction after detecting an initial deviation or symptom; according to the type of recovery factor (or recovery root cause); and according to the phase in which this recovery factor makes its main contribution.

2.4.1 Classification based on preceding failure

Both the ECM (see section 2.1) and Embrey and Lucas (1988) provide the rationale for this taxonomy. Technical, organisational and human root causes of failures may be linked with their subsequent recoveries. Additional subcategories might include: recovery from one’s own error, or from a colleague’s (same or previous shift, when applicable); technical failure of equipment outside the central control room (CCR), of the interfaces within the CCR, of process control software, etc.

An error taxonomy developed by Frese and Zapf (1991) includes several different types of errors structured by the levels of regulation (sensorimotor level, level of flexible action patterns, and intellectual level) and the steps in the action process (goals, planning, monitoring, and feedback) according to the Action Theory (Hacker, 1986). Thus the taxonomy includes sensorimotor errors, habit errors, omission errors, recognition errors, thought errors, memory errors and judgement errors.

2.4.2 Classification according to operator reaction after symptom detection

As noted by Reason (1990) in his GEMS model people seldom go through the entire analytic process of fault diagnosis when confronted with a deviation. This was confirmed by Brinkman (1990) who collected verbal protocols during a fault finding task. He observed the following three reactions after his subjects detected an error in their reasoning process:

* Ignore the error and continue: rely on system redundancy and subsequent error recovery factors.
* Simply repeat the most recent sequence of actions: try again, without any attempts at fault localisation.
* Attempt fault localisation and optimize corrective actions: either by forward analysis (repeat the most recent action sequence and check every step) or backward analysis (trace back from symptom detection to previous actions, until the error is found).

By applying this classification, transitional probabilities between the recovery phases of section 2.2 might be established.
2.4.3 Classification according to type of recovery factor

Such a classification should be the most important one for MMS-designers. The ECM for failure root causes could serve as a basis for recovery root causes too, with the following extensions:

* **Technical design of the process:**
  aim at maximum reversibility of process reactions (Rasmussen, 1986) and 'linear interactions' plus 'loose coupling' (Perrow, 1984) of process components; these may be achieved by structural characteristics (e.g. buffers, parallel streams, equipment redundancy) and by dynamic characteristics (e.g. speed of process reactions, response delays).

* **Technical design of the man-machine interface:**
  aim at maximum observability (Rasmussen, 1986) of deviations and their effects (e.g. transparency instead of alarm inflation).

* **Organisational and management factors:**
  particularly an updated, clearly formulated and well-accepted set of operating procedures and a positive safety culture must be mentioned here (see also Van Vuuren, 1995).

* **Human operator factors:**
  optimize the cognitive capabilities (e.g. accurate mental process model) of operators through selection and (simulator-) training, but also by supporting them with software tools to test hypotheses and avoid certain biases.

2.4.4 Classification according to recovery process phase

As mentioned earlier in 2.2 this final classification aims to distinguish between detection, location and correction as the phases of impact of the recovery factors in 2.4.3.

3. Empirical research questions

Based on the proposals in section 2, the human recovery research project of the Eindhoven Safety Management Group is directed at the following empirical research questions:

1. Is recovery more than sheer luck or coincidence? If so, then the potential for recovery can be built into a MMS and managed!
2. Can recovery be classified with the same ECM root causes as for failures? If so, what is the contribution of human recovery relative to technical and organisational failure barriers? How large is the contribution of human recovery in a variety of task situations and over a variety of system effects?
3. Are recovery factors identical to failure factors in a given MMS? If so, then preventing errors and promoting recovery would focus on the same MMS aspects.
4. In which phase(s) of the recovery process do recovery factors contribute most to system performance: symptom detection, fault localisation or correction?
4. Pilot studies

Pilot studies have recently been carried out in steel making, energy production and surgery. A variety of system effects have been investigated: safety, reliability and environmental effects of system breakdown.

4.1 Safety incidents in a steel plant

In a Dutch steel plant Mulder and Van der Schaaf (1995) identified failure and recovery factors in the same set of 25 safety-related near misses. The results are given in fig. 2-4.

Fig. 2.: Distribution of 154 failure factors in 25 safety incidents in a steel plant.

Fig. 3.: Distribution of 34 recovery factors in 25 safety incidents in a steel plant.

Fig. 4.: Distribution of 34 recovery factors according to recovery phase in 25 safety incidents in a steel plant.

4.2 Medical safety incidents in a surgical ward

In a large teaching hospital in Eindhoven Van der Hoeff and Van der Schaaf (1995) found the following failure and recovery factors in the same set of 17 medical near misses with patients undergoing surgery (fig 5 and 6).

Fig. 5.: Distribution of 95 failure factors in 17 medical incidents with patients undergoing surgery.

Fig. 6.: Distribution of 22 recovery factors in 17 medical incidents with patients undergoing surgery.
4.3 Reliability and environmental incidents in an energy production plant

In a small energy producing unit of a chemical plant Zuijderwijk (1995) classified failure and recovery root causes of 23 reliability and environmental near misses (fig. 7 and 8).

![Fig. 7: Distribution of 86 failure factors in 23 reliability and environmental incidents in an energy production plant.](image1)

![Fig. 8: Distribution of 56 recovery factors in 23 reliability and environmental incidents in an energy production plant.](image2)

4.4 General discussion of the pilot studies

Figures 3, 6 and 8 show a range of 2 to 11 percent of unclassifiable causes (that is: luck or coincidence) of recovery. This must be interpreted as a positive answer to question 1: around 90 percent or more of all recovery factors are clearly technical, organisational or human in nature and therefore researchable and eventually manageable.

The same figures show human recovery root causes contributing 21 to 66 percent. Comparison with the failure factors of figures 2, 5 and 7 shows this human recovery range to vary at least as much as the human failure range (e.g. 33 to 56 percent). The human component should therefore also be taken seriously in terms of recovery possibilities (see question 2). This result can be seen as an additional argument for the implementation of error management in every work related training situation. If the role people play in preventing serious accidents is this big, they should be trained more rigorously in dealing efficiently with errors. Empirical results of studies in error training (Frese, Brodbeck, Heinbokel, Mooser, Schleiffenbaum & Thiemann, 1991; Dorman & Frese, 1994) have shown that error training in contrast to error avoidance training leads to higher performance. Error training means allowing and encouraging people to make errors in the training process and ultimately encouraging them to learn from these errors. As a result subjects who received error training have shown fewer errors and performed better, even in other areas of performance.

Zuijderwijk (1995) showed that the patterns of failure and recovery factors are clearly different. Rule- and skill-based factors dominate the operator failures, while knowledge-based insights are very important in human recovery. Similarly, ‘material defects’ are the most prominent technical failures, while ‘design’ covers all technical recovery factors (see question 3).
Finally, relating to question 4, figure 4 shows that hardly any recovery process goes through the more analytic localisation phase. Again, this could be interpreted as confirmation of Reason’s GEMS model, but there is also the possibility of an explanation in terms of time stress. If recovery is present only in the very last phase of accident development (as was the case in most of the steel plant near misses) there may simply not be enough time for a time-consuming diagnostic effort; detection and correction ‘just-in-time’ may be all one can do in such cases.

5. Implications for MMS design

In spite of the immaturity of the proposed models and classifications, and of the small number of recovery incidents gathered so far, these ideas and results are intriguing enough to formulate the following tentative implications for designing a MMS:

* Consider recovery promotion and error management as an alternative to failure prevention, especially when certain errors or failures are predictably unavoidable.
* Do not simply "design out" failure factors without considering the possible reduction of recovery factors: raising the level of automation in process control, or installing too many decision support tools for your operators, may leave them helpless under certain situations.
* Try to support all recovery phases, primarily by means of an optimal man-machine interface: detection, localisation and correction (see section 2.4.4).
* Invest in deep process knowledge of operators: reasoning beyond procedures appears to be essential for many recovery actions. Error management supported by error training seems to be an adequate way to enhance the ability to deal quickly and efficiently with errors.

Acknowledgement

The presentation of this paper was supported by a CEC Human Capital and Mobility Network on Human Error Prevention.

References


