Nationaal Lucht- en Ruimtevaartlaboratorium

National Aerospace Laboratory NLR

NLR-TP-97031 U

Human machine interactions with future flight deck and air traffic control systems: validation results and the role of objective and subjective measurements

P.G.A.M. Jorna

DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF. NLR TP 97031 U		SECURITY CLASS. Unclassified

ORIGINATOR
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

TITLE
Human machine interactions with future flight deck and air traffic control systems: validation results and the role of objective and subjective measurements

PRESENTED AT
the First international Conference on Engineering Psychology & Cognitive Ergonomics, Stratford-upon-Avon, England, 23-25 October 1996.

AUTHORS P.G.A.M. Jorna	DATE 980305	pp 30	ref 14

DESCRIPTORS
Aircraft accidents
Air traffic control
Decision making
Display devices
Flight safety
Heart rate
Human factors engineering
Human performance
Man machine systems
Mental performance
Operator perfomance
Pilot performance
Psychological factors
Stress (psychology)
Workloads (phychophysiology)

ABSTRACT
The safety level in the aviation system is extremely high and accidents are so rare that most crashes reach the headlines. Air travel is expanding cconsistently and this expansion of the number of flights will necessitate a major improvement in safety levels in order to assure that aircraft accidents will remain rare events. New technologies like digital data links and advanced software assistance to the human operators are investigated to accommodate the growth in travel and at the same time improve safety to the required levels. The occurrence and Prevention of human errors is therefore a main and high priority issue in the design and validation of new technology applications.
This lessons learnt with so-called 'automation' in aviation will be reviewed briefly as well a s the challenges that lay ahead for the industry. Studies and experiments with applications of new technology for data linking. Air traffic management and flight deck automation, will be discussed. The emphasis is on the quality of the human interactions with such future technologies as observed and measured during realistic simulations of possible operational applications. Human operator behaviour can be studied and documented by means of more advanced measurement equipment that enables objective performance and workload measurements when working with these systems. The role and importance of contrasting subjective and objective measurement techniques for decision making on the 'way forward' with the design and validation process will be illustrated and discussed. Finally, some fallacies as well as implications for future work will be highlighted.

Summary

The safety level in the aviation system is extremely high and accidents are so rare that most crashes reach the headlines. Air travel is expanding consistently and this expansion of the number of flights will necessitate a major improvement in safety levels in order to assure that aircraft accidents will remain rare events. New technologies like digital data links and advanced software assistance to the human operators are investigated to accommodate the growth in travel and at the same time improve safety to the required levels. The occurrence and prevention of human errors is therefore a main and high priority issue in the design and validation of new technology applications.

The lessons learnt with so-called ‘automation’ in aviation will be reviewed briefly as well as the challenges that lay ahead for the industry. Studies and experiments with applications of new technology for data linking, Air traffic management and flight deck automation, will be discussed. The emphasis is on the quality of the human interactions with such future technologies as observed and measured during realistic simulations of possible operational applications. Human operator behaviour can be studied and documented by means of more advanced measurement equipment that enables objective performance and workload measurements when working with these systems. The role and importance of contrasting subjective and objective measurement techniques for decision making on the ‘way forward’ with the design and validation process will be illustrated and discussed. Finally, some fallacies as well as implications for future work will be highlighted.

Contents

21 Figures

(30 pages in total)

1. Introduction

The present day aircraft are equipped with technologies that allow around the clock operations under almost any weather conditions. The safety level is extremely high and accidents are apparently so rare that most crashes reach the headlines. Air travel is however, still expanding consistently and this expansion of the number of flights could lead to an increase in the absolute number of crashes occurring, making it an unfortunate less rare event. New technologies like digital datalinks and advanced software assistance to the human operators are investigated to accommodate the growth in travel and improve safety to such levels that an airplane crash will remain a relative rare event. The occurrence and prevention of human errors is a main issue in the design and validation of new technology applications.

The existing problems with so-called ‘automated aircraft’ will be reviewed briefly as well as the challenges that lay ahead for the industry. Studies and experiments with applications of new technology for datalinking, Air traffic management and flight deck automation, will be discussed. The emphasis is on human interactions with future technologies as observed during realistic simulations of the possible operational applications and studying human operator behaviour with objective performance and workload measurements when working with these systems. The role and importance of contrasting subjective and objective measurement techniques for human performance and workload will be illustrated and discussed. Finally, implications for future work and some fallacies experienced by the industry will be highlighted.

2. Present state of the art

2.1 Automated aircraft: benefits and problems

The attainable accuracy and dependency of modern aircraft is quite superior to that of older technologies, but solving one problem often creates new and unexpected flaws in the modified or new design. Cockpit automation has provided a clear illustration of this day to day phenomenon. The flight crew have to interact with multiple avionics systems that evolved in an evolutionary way, mostly implying the addition of new technologies to existing ones. Sometimes, under unexpected operational states or circumstances these systems can operate against each other, or compensate for some malfunction leaving the crew unaware of the situation or not informed about possible consequences when the automation reached the edges of its so-called performance envelope. A summary of benefits and problems is provided in figure 1.

Automated Aircraft

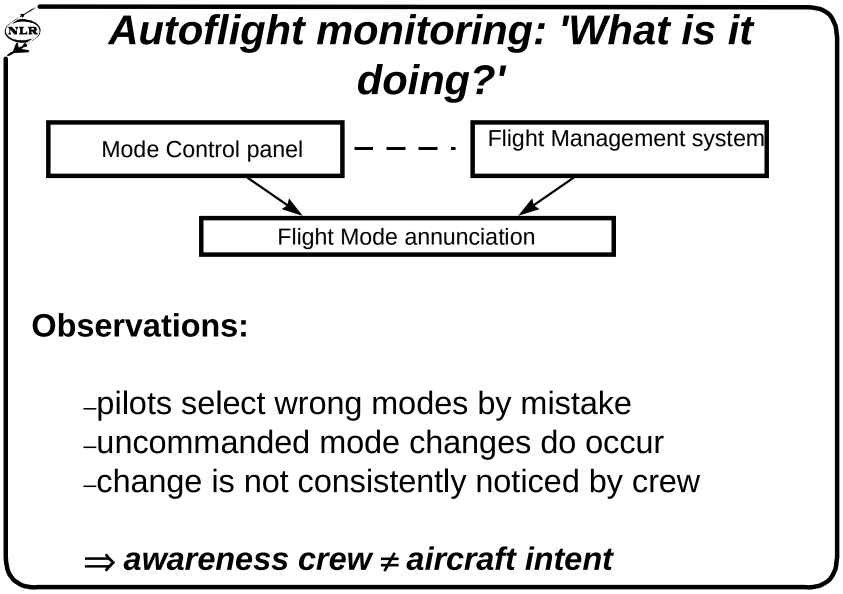
Benefits	**Problems**
Capability	Avoidable accidents
Accuracy	Complexity
Efficiency	Increased coupling
Fuel	(in)observability
Crew size	**Pilot-Aircraft interaction**

Fig. 1 Summary of benefits and problems with modern advanced aircraft

The most disturbing problem is the occurrence of accidents with aircraft that had no major malfunction and that still were in flyable condition. Confusion of perceived state of the aircraft and actual state apparently do not match consistently, indicating the nowadays well known, problems with pilot-aircraft interaction.

2.2 Pilot- aircraft interactions

It is widely reported that the dominant factor in present day aircraft accidents is consistently related to some 'human factor' in either the design or the operation and maintenance of aircraft's and their related systems. Flight crews have to operate rather complex automated equipment, often under time pressure as initiated by re-routing through ATC etc. Working with this automation implies the management of many possible ‘modes of operation’. The main problems with such modes are summarised in figure 2. including the main avionics and displays involved.



Autoflight monitoring: 'What is it doing?'

Observations:

- pilots select wrong modes by mistake
- uncommanded mode changes do occur
- change is not consistently noticed by crew

⇒ ***awareness crew ≠ aircraft intent***

Fig. 2 Summary of main problems when managing multiple modes in aircraft systems

For example, some 70% of the civil transport accidents are generally attributed to human error and of these accidents the majority were of the so called 'Controlled Flight Into Terrain' (CFIT) type. In that case a fully serviceable, intact aircraft is flown into the ground with devastating consequences for passengers, crew and the crash environment or people on the ground who are impacted. Clearly, no crew will intentionally or voluntarily perform such manoeuvre, so other factors must play a crucial role. One of them is that the level of ground support is different at many locations leaving the crews without accurate radar coverage. Approaches are flown (non-precision) by following strict and often complex procedures using data from a beacon indicating direction and distance without an independent reference with respect to terrain ahead or to check the assumed state and (vertical) position of the aircraft. The use of the Global Positioning System based on satellites are an improvement but the crew still has to analyse the maps to decide if any obstacles are to be expected.

Time pressure can hurry the crews and increase the likelihood of human errors when ‘instructing’ the aircraft. An illustration of such an event was the case were a pilot intended to select a flight mode that would control the ‘angle of descent’ but instead happened to accidentally select the ‘vertical speed mode’. All modes needs a particular setting of a value to ‘chase’ and the value was set with the idea in mind of setting a value for the angle but the input arrived at the wrong mode. The system received a vertical speed to be maintained. Pilot awareness and system intent were now inconsistent. As display space is always limited in an aircraft, only a limited number of digits is often displayed, making it difficult to detect the error (value times 100 etc.). The result was an aircraft descending (just as the pilot would expect), but at a too high rate. The aircraft subsequently hit a mountain in stead of passing over it. These type of events, can occur with all types of aircraft and are illustrated in figure 3.

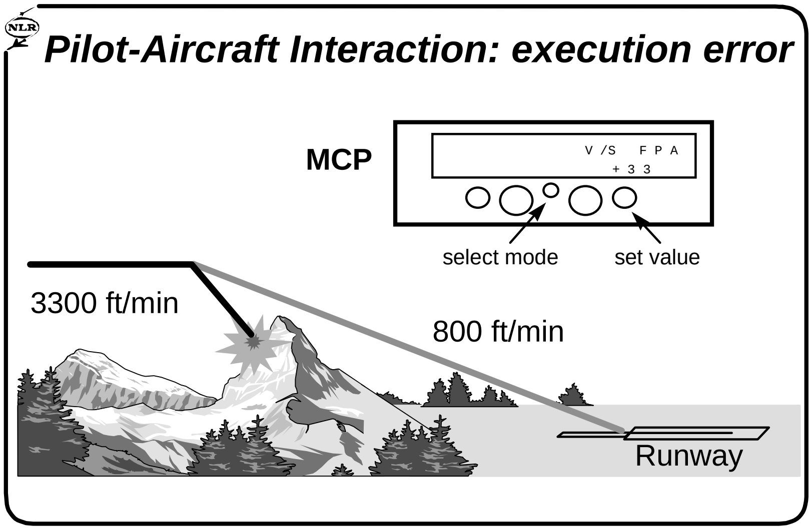


Fig. 3 Illustration of a Controlled Flight Into Terrain (CFIT) accident by unexpectently selecting a wrong mode during approach

The lessons learnt sofar of many of such pilot-aircraft (mis) interactions, indicate that improved and future systems on the flight deck and ground equipment should improve the ‘observability’ of complex software based systems as well as create systems that are more robust to human errors (Folkerts & Jorna 1994). Providing clear feedback is mandatory but not sufficient. There should also be a back-up in the sense that the aircraft is more informative about its overall ‘intent’. This design task is however formidable, especially when considering the increased future density of air traffic.

3. Challenges for the future

3.1 Accommodating more air traffic

Reading the papers, it becomes obvious that the predicted growth in air traffic (approaching more than twice as many) has actually begun to materialise and it is still expected to continue for many years. Airports are congested to full capacity at peak hours and the growth potential in the market is hard to follow as constructing new runways takes a lot of time and fighting the environmental problems like noise reductions etc. requires new technologies that also take time to develop and procure. The modern aircraft clearly are less noisy and technologies are still improving, so a main problem will be to accommodate the traffic as the absolute number of flights will increase. Building very large aircraft is not a fail safe solution as they will mainly fit the long haul routes and flying with half empty aircraft on short routes is unaffordable.

The progression in Air Traffic Control is generally considered to have fallen behind the state of the art of modern aircraft and major improvements are needed. However, the investments will be high and there is no room for failure. As a result there are multiple perspectives on how to proceed towards establishing so-called Air Traffic Management. These perspectives have a high impact on considerations concerning the Human Factors and cognitive activities required of its participants.

3.2 The ground versus airborne perspective

The renewal of present day ATC procedures and technologies brings many issues to resolve and organize (for a review see Jorna 1994). A first issue is that there will be (and are) actually 'two types of ATC existing today and in the future namely a '*High Tech-*' and a *'Low Tech'* ATC environment. Both are served by the same aircraft. The need for future updates will not only be challenging for the Western world, but will also place other sectors under economical pressure for increasing investments in facilities to accommodate air travel. A common, globally accepted ATM concept has not been established yet.

The most prominent technology that will both change the flight deck as well as the ATC system is the digital data link. This link will provide aircraft better access to information available in ground computers and ATC can receive better data on aircraft positions and intents. Also the congested radio frequencies can be freed from routine communications.

More aircraft means either a reduced separation or the exploitation of more airspace. The present fixed route structures lead to traffic jams just as on the roads. Reduced separations can be realized when aircraft fly very accurately as planned and on time within say 5 seconds. This way of thinking lead to the so-called 4D concept where the aircraft should be at a certain location within a narrow specified time window, so another aircraft can be allowed to pass right behind.

Amongst the most successful scenario's are the European PHARE (Program for Harmonized ATM Research in EUROCONTROL) concept of efficient (re-) negotiations with ATC and accurate adherence to the contracted trajectories by means of an airborne based 4D FMS. The United States CTAS (Center TRACON Automation System) environment allows the ground system itself to provide 4D type of guidance to non-equipped aircraft and seems particularly suited for a transitory phase.

Recently, however, the so called *'Free Flight' concept*, took another perspective on using the airspace differently. It proposes some level of aircraft responsibility for en-route separation while flying free, or preferred routes. It has gained considerable visibility in the media and bounded numerous advocates to its (quick) development. The role of ATC is now intentionally reduced and involves a role of providing 'arbitration' only when needed.
'Free flight ' thinking will have a major impact on ATM thinking as it is very attractive form an economical point of view. The airlines earn money with flying and will invest in on-board technologies. The alternative is to wait for all the ground systems to be updated. Using Global Positioning Systems (GPS) and digital maps allow aircraft to find their way and if robust, is tempting to be adopted as it can be fairly quick and at relative low cost. The issue however is ‘will it be safe’ and what is the role of operators involved and how will tasks be shared, monitored and executed by computers and humans. Considering the issues, more efficient and timely validation methodologies should be developed and practiced to allow such systems to operate.

4. Working with new technologies

4.1 Cockpit datalink

The advent of data link capabilities resulted in the issue of what kind of human-machine interfaces would be acceptable (certifiable) for handling ATC communications at the flight deck. An existing application is the so-called ACARS unit (Aircraft Communications and Reporting System) located in the aft part of the pedestal, so in a sub optimal location considering positioning with respect to the crew members. This unit was compared with possible alternatives like using the Multi-function displays and/ or the Control and Display Unit (CDU) of the FMS. These systems have a favourable position but share other functionality's in a single device.

These three systems were compared during a realistic simulation with the NLR research simulator with Glass cockpit instrumentation and representative operational conditions using normal line-crews (Van Gent et al.1994). In these studies, the following measurements were included: Head tracking for both crew members, heart rate and respiration, communications analysis with the ground and within crew, logging of systems inputs and outputs etc. and subjective assessments of usability and acceptability per flight phase (for a review of the methodologies involved see Jorna 1997).

The principles of crew resource management dictate that the Pilot Non Flying (PNF) is handling the communications, in order to allow the Pilot Flying (PF) to concentrate on primary flight instruments and the outside world. Translated into expected head tracking data, this means that the PF should be head-up most, if not all of the time. The following data was however obtained as illustrated in figure 4.

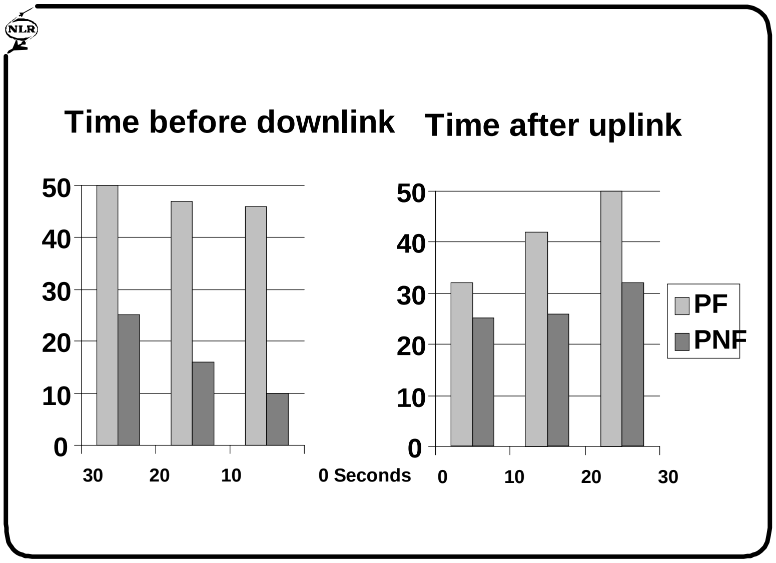


Fig. 4 Head-up percentage of time just prior to a downlink and after the occurrence of an uplink that has to be handled by the Pilot Non Flying (PNF) as opposed to the Pilot Flying (PF)

In case the PNF prepares and sends a message, his colleague does take notice, but maintains attention primarily according to the requirements and normal procedures. However, in case of an up-link, the Pilot flying goes head down significantly while this behaviour is not according to best rules of practice. From a cognitive point of view, this behaviour is quite understandable as any uplink means significant information and everybody will be curious about what is going to take place. As a result of this work synthetic speech applications for auditory presentations to the PF were re-visited to ameliorate this type of behaviour, especially during the more critical flight phases (Van Gent 1996).

In determining the ‘Good versus Bad’ characteristics of user-interfaces, task completion times seem to be a natural criterion to be considered. Analyses of the logging of ‘button interactions’ combined with head tracking data, revealed a phenomenon that could be described as ‘interruptions’ in task execution or breaches in procedures for handling certain tasks. The number of occurrences per time period are depicted in figure 5. for each of the human machine interfaces investigated in the experiment.

The expected result was that the ACARS Interactive Display Unit (IDU) would produce slowest task completion times due to location and touch control characteristics. This only proved to be partially true. What was observed is that the user has a tendency to, whatever happens, finish the job. The reason was that in case of a non-supportive interface, the user will loose track when interrupted. Interfaces that provide feedback and status of work information, however, actually allow tasks to be interrupted and that was what happened. The lesson learnt is that a new design will also elicit new human behaviours that, although unexpected beforehand, prove to make sense from a cognitive standpoint.

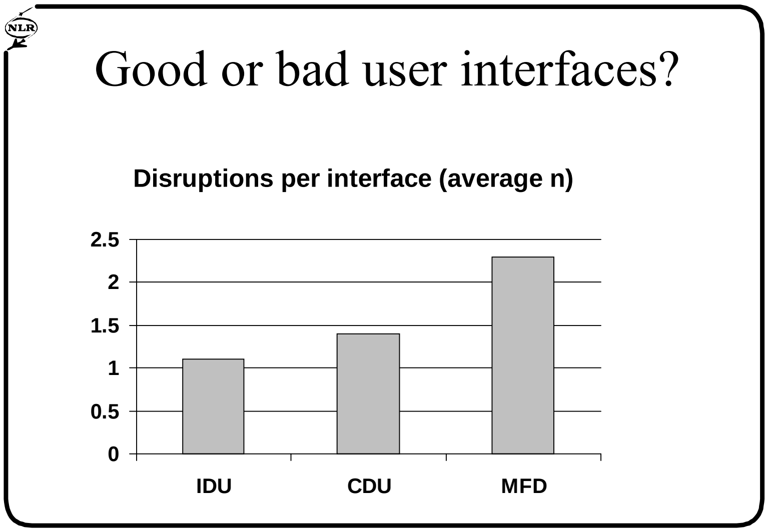
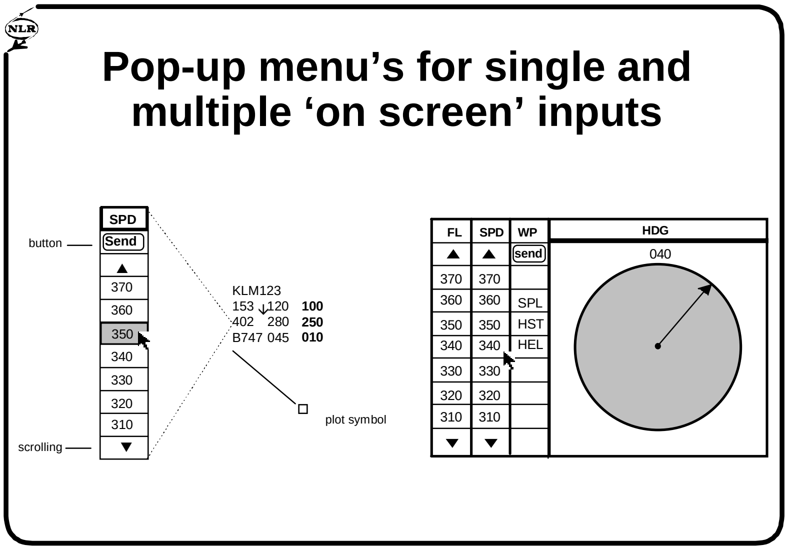


Fig. 5 Number of interruptions pilots allowed during work or ‘breaches’ of anticipated procedures

4.2 ATC datalink

Datalink communication can involve computer-computer interfacing as well as Air- Ground human operator communications. Pilot can request route changes, ask for information etc. while the controller can detect possible conflicts in the future and provide the aircraft with instructions. Finding and implementing a solution in the present day systems often requires vocal communications with the pilots as well as inputting data (the instructions) into the ATC computers in order to display the overall status to the controller. A datalink user interface can accommodate both these functions at the same time. In an experiment different versions were developed and implemented for testing (Hooijer 1996). The datalink can be implemented as a separate communications window on the controllers display or as an integrated part of the so-called radar plot symbol that is associated with a particular aircraft. Examples of such selectable pop-up menu's are depicted in figure 6.

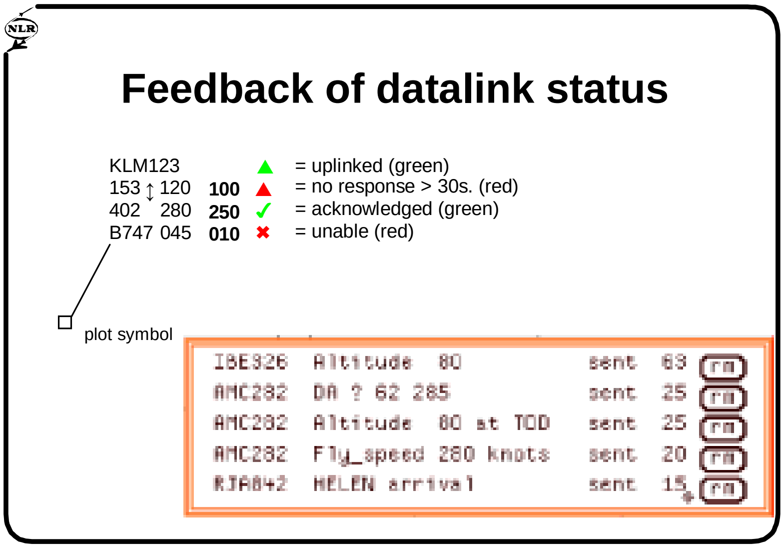


Pop-up menu's for single and multiple 'on screen' inputs

Fig. 6   Examples of datalink user interfaces that allow direct selections and transmissions of solutions of problems and instructions to the aircraft

Similarly, the feedback of actual status of the negotiations with the aircraft need to be provided as datalinks have a time delay in transmitting the data, depending on the particular medium used i.e. radio frequencies, radar signals etc.

The feedback can be provided through different means. Integrated with the radar plot data block or as an separate communications window. Examples are shown in figure 7.



Feedback of datalink status

KLM123 🟩 = uplinked (green)
153 ↕ 120 **100** 🟥 = no response > 30s. (red)
402 280 **250** ✓ = acknowledged (green)
B747 045 **010** ✖ = unable (red)

```
IBE326  Altitude  80            sent  63
AMC282  DA ? 62 285             sent  25
AMC282  Altitude  80 at TOD     sent  25
AMC282  Fly_speed 280 knots     sent  20
RJA842  HELEN arrival           sent  15
```

Fig. 7 Examples of means for providing feedback on the datalink status for communicating and receiving confirmations from many different aircraft

From these options three combinations were designed that will be designated as user interface combinations ‘A’, ‘B‘ and ‘C’. The mapping of the particular features that were combined is as follows:

Condition	Input method	Feedback
A	pop-up menus	DSP
B	pop-up menus	label symbols
C	combined pop-up menu	label symbols

The conditions A, B, and C were intended to represent an increasing level of integration of task elements and feedback onto the screen of the Air traffic controller. The higher the level of integration, the lower the number of on-screen search actions and required subsequent integration of data and information. Also, these options have different disadvantages. As an example the data link status window will change in content (ir-) regularly, thereby attracting the controllers attention at a time that such information would not strictly be required for mental processing. Alternatively, the pop-up or better, pop down menu’s of the radar data block can obscure some of the other traffic data, although at a moment selected by the controller who decides to take an action.

The experiment used the following measurement techniques: Eye point of gaze measurements, head tracking, pupil size, heart rate and respiration, heart rate variability , logging of system inputs and responses and extensive use of subjective ratings. The subjects in this experiment were, like in the cockpit data link study, normal professionals in this case regular controllers.
The results showed the following:

Heart rate variability normally decreases when working under working conditions that are cognitively loading or stressful (for a review see Jorna 1992). So a user interface that is more easy to work with could result in a relative increase as compared to more cumbersome interfaces. However, no very explicit results were expected as Heart rate variability is especially sensitive to more extreme overall working conditions that are associated with particular distinctive mental states as associated with mental work and stress. The results obtained in this experiment proved very promising as indicated in figure 8.

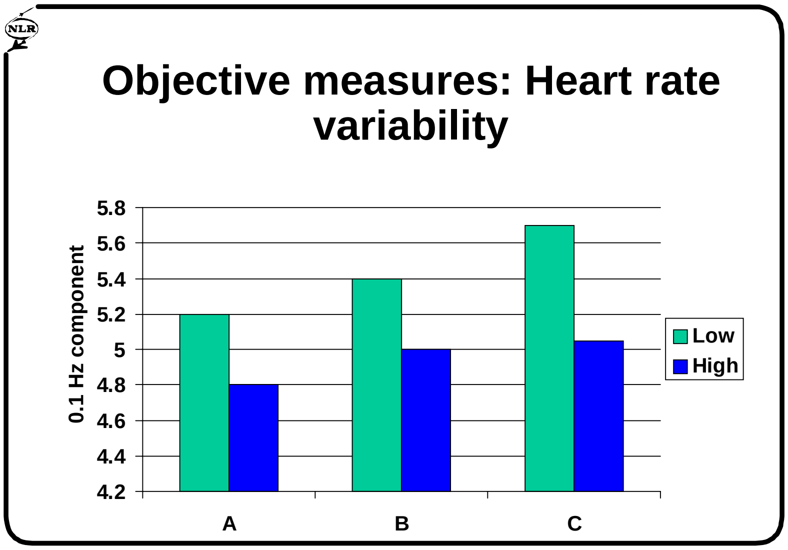


Fig. 8 Heart rate variability increase as a function of controller data link interface and traffic density (Low or High)

The impact of traffic density on heart rate variability is quite consistent and a little bit to our surprise, also quite distinct for type of user interface. Also, as an experiment, the pupil size was calculated and analysed. The results of this initial analysis are depicted in figure 9.

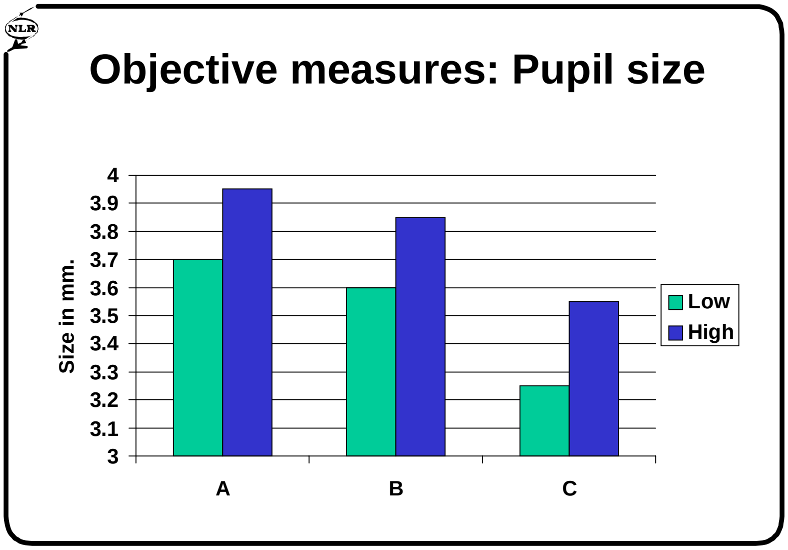


Fig. 9 Pupil size as a function of controller data link interface and traffic density (Low or High)

The size of the pupil(s) normally is modified by physical factors like the amount of light present, but it can also be influenced by the required visual information sampling and processing of visual data. In that case, its size will increase as a function of the amount of visual information processing.

The results indicated that accurate measurements of small differences in size can be realized. Similarly to the heart rate variability data, the pupil decreased in size as a function of level of integration in the controller data link interface while it increased markedly with an increase in traffic density (difference between Low or High traffic samples). Note that more traffic implies more radarplots on the screen which should tempt the pupil to downsize as a function of amount of light in the display.

An example of the subjective ratings provided by the controllers is summarized in figure 10.

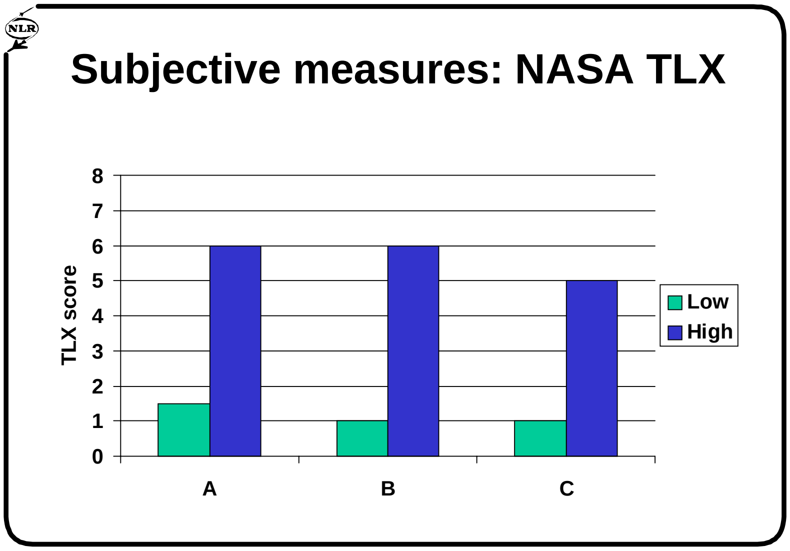


Fig. 10 Ratings provided after the experimental sessions as a function of controller data link interface and traffic density (Low or High)

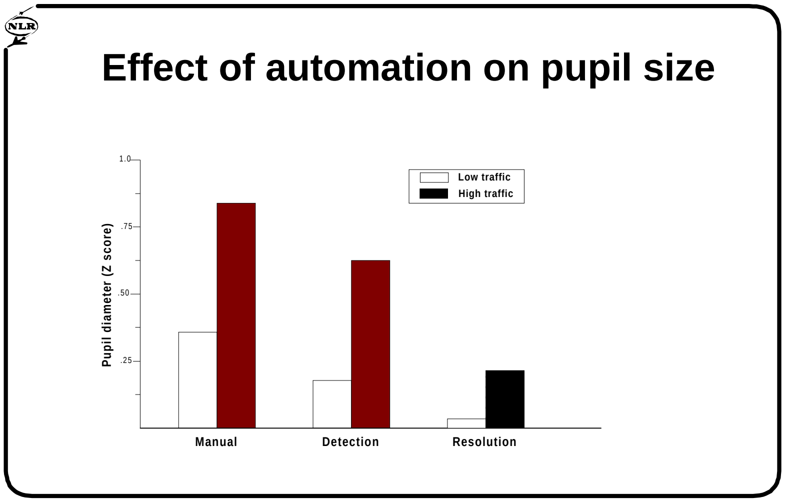
The subjective ratings displayed a marked sensitivity to the amount of traffic present, but revealed a quite less spectacular difference between user interfaces. So, in this experiment the objective results all pointed to the possibilities of designing effective controller datalink interfaces, but the subjective data did not reflect this to the same extent.

4.3 Tools for the air traffic controller

The mental processing capabilities of the air traffic controller are generally considered to represent the major bottleneck in expanding the amount of air traffic. One reason is the communication process that is of a serial nature and has to address each plane individually. Also, the controller has to build an overview over the traffic streams in order to be able to predict and anticipate possible separation issues. In case of a possible overload, the traditional procedure is to subdivide more sectors so more controller teams can share the work. The disadvantage is of course that the communication requirements also increase dramatically, thereby limiting the overall effectiveness.

Alternatively, the use of datalinking allow the controller to issue messages more quickly and both individual aircraft and groups of aircraft can be addressed if relevant. Also, clearances with multiple parameters or complete route instructions can be issued. Software tools can provide assistance in conflict detection and resolution of aircraft route or altitude infringements. The effectiveness of such possible assistance was investigated by means of simulations and extensive objective and subjective measurements (Hilburn et al. 95, Hilburn et al. 96).

The results of a comparison of a stepwise increase in the level of assistance provided with a ‘manual’ baseline are depicted in figure 11 for pupil size measurements.



Effect of automation on pupil size

Fig. 11 Pupil decrease as a function of higher levels of automation assistance

The data obtained in this study seem very promising for extended applications of these measures. Also, the results for Heart rate variability revealed an almost identical trend with heart rate variability increasing systematically when more assistance is provided to the controller.

An additional technique applied was the principle of ‘dual tasking’ but with the purpose of acquiring an objective measure of situation awareness, in this case defined as awareness of communication and aircraft status. Incidentally, aircraft would fail to acknowledge their up links and the controller had to detect these occurrences. In case of a reduced overall task load, more options are present to perform this particular task more timely. The results are depicted in figure 12.

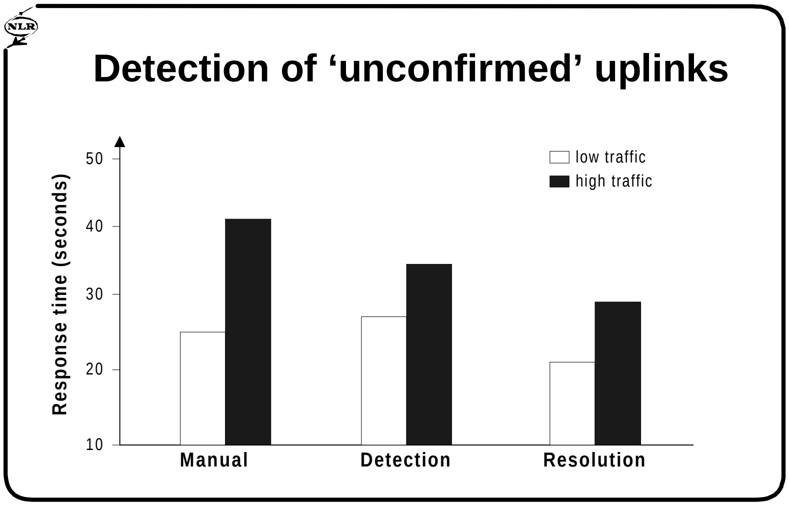


Fig. 12 Controller response times to status information on communications and aircraft responses

Apparently, the tools do allow the controller to scan the display more effectively, resulting in better overall performance. So, overall the objective measures clearly indicate the potential of the tools in helping the controller. But how do the controllers rate them subjectively? That data is depicted in figure 13.

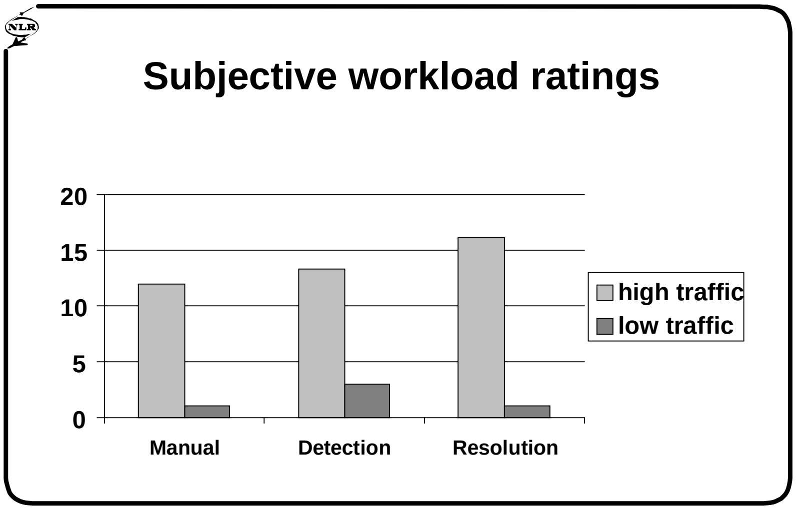


Fig. 13 Controller estimates of workload effects as a function of more software tools

Surprisingly, the controllers rate the effects of the tools quite contrary to the picture provided by the objective measurements. Possibly, the addition of extra functionality is experienced as more work to be handled.

A mediating factor in these ratings could be the particular strategy employed by the controllers in using these tools. Controllers have very particular strategies in handling their traffic and these ‘controller methods’ could influence adaptation to the new controller working position. An illustration can be provided by analyzing the eye scanning during high and low traffic density samples.

ATC plan view display (PVD)

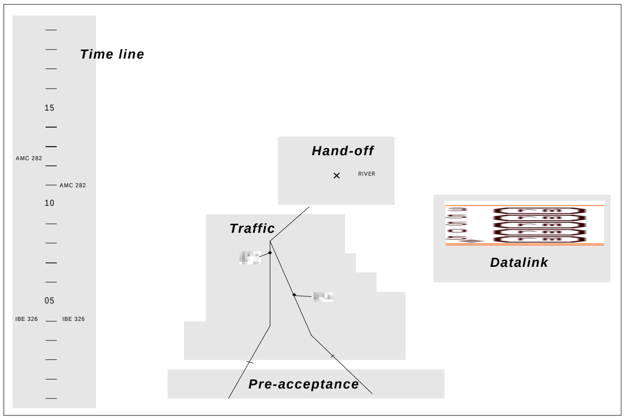


Fig. 14 Display lay out of the Plan View Display with Arrival scheduler at the left, aircraft hand-off area and data link communication status panel. The area’s are used by the point of gaze equipment to provide area related data on eye scans, duration’s, transitions etc.

Scanning the PVD: low traffic

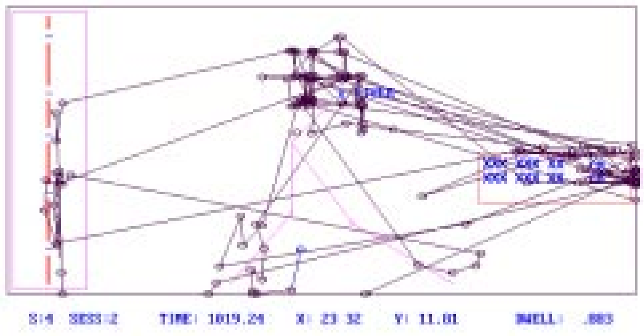
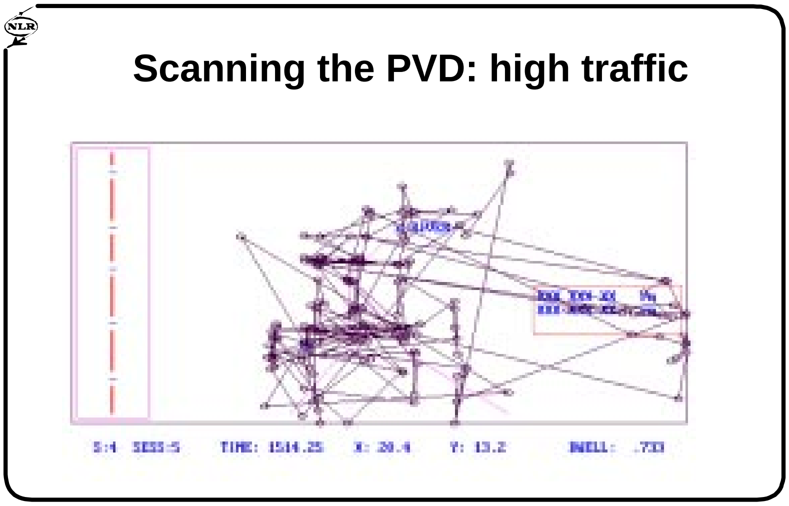


Fig. 15 Sample (120 seconds) of Point of gaze transitions



Scanning the PVD: high traffic

Fig. 16 Sample (120 seconds) of Point of gaze transitions

The results indicate that a tool as the scheduler for Arrivals by means of a time line is used especially under the low traffic conditions, but the moment traffic builds up, controllers seem to drop the tool and revert to the classic ‘on screen’ controlling methods. The paradox that occurs is that tools with technology designed to ease the job of the controller are being discarded especially in the situations where they were anticipated to benefit the most.

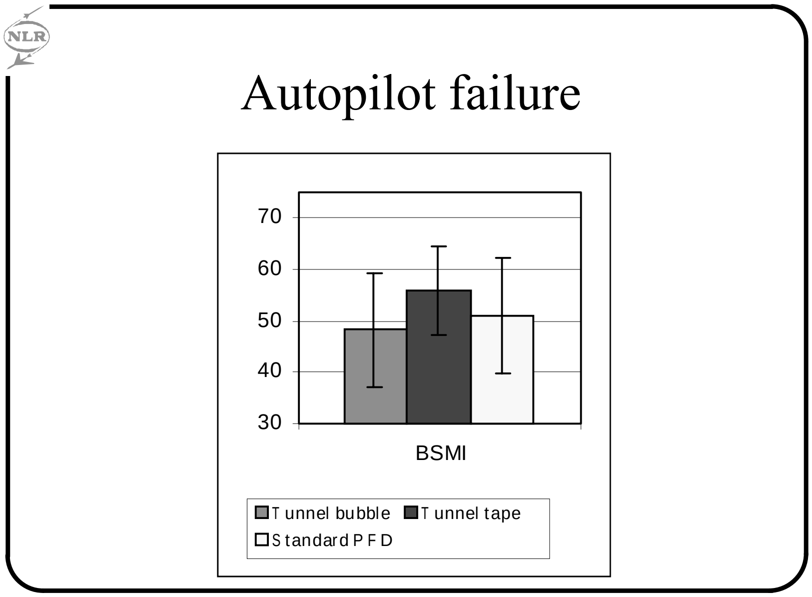
4.4 Tools for the Flight crew

In flying under anticipated Air Traffic Management scenario’s as described by EUROCONTROL programs like PHARE (Program for Harmonized Air traffic management Research in EUROCONTROL) so-called 4D scenario’s are envisaged (for more information see Jorna & Nijhuis 1996). Aircraft negotiate a preferred route that is defined in the normal three dimensions, but now precise timing is linked with certain locations and altitudes of the aircraft in the airspace. The plane now has to fly through a virtual Tube to maintain the conflict free route depicting both lateral and vertical navigation information. Within this tube they have to stay in a ‘Time box’ that moves along the tube as required to maintain conflict free.

A possible ideal display for the flight crew is the well known’ High way in the sky’ representation, but adapted to ATM purposes by designing it as a virtual ‘Tunnel in the sky’ concept to guide the crews ‘in time’ and ‘on time’.

In preparation for a validation experiment pilots (Huisman & Karwal 1995) were asked to comment on the display candidates involving a Tunnel display and a modified baseline display comprising a normal Primary Flight Display (PFD) with additional indicators for timing aspects integrated at vertical scales depicting values for altitude and speed (so-called speed and altitude tapes). As a result a hybrid display version was designed comprising the tunnel display combined with speed and altitude tapes. This version could possibly ease the transition from one display type to the other.

The experiment (Huisman & Flohr 1997) was performed in a moving base research flight simulator as mentioned in the earlier research. Twelve line pilots participated and extensive ‘Human Factors measurements’ were collected as discussed earlier. The following (preliminary) results were obtained for a flight involving a failure of the auto pilot which necessitated the pilots to revert to ‘hands on’ manual control. The performance data obtained revealed a very consistent superior performance for the Tunnel display in general. However, the following subjective ratings were noted as depicted in figure 17.



Autopilot failure

Fig. 17 Pilot ratings (n=12) on a Rating scale for Mental Effort (RSME or BSMI in Dutch)

The results indicate that the tunnel display was favoured as compared with the traditional display. Surprisingly, the hybrid display or ‘Tunnel with tapes’ was rated less favourably while it was proposed originally by members of the pilot community. The rating values themselves can be denoted as being low to average on the RSME scale.

In order to investigate the nature of this apparent change in opinion, the total group was split in two sub groups. One group consisted of pilots who clearly expressed their preference for the Tunnel (n=9) and a second group (n=3) that clearly did not! This last group was however quite small, indicating the potential for pilot acceptance of the tunnel display.

When the data was recalculated and explored, the following results were observed as depicted in figure 18.

Auto pilot failure (tunnel group)

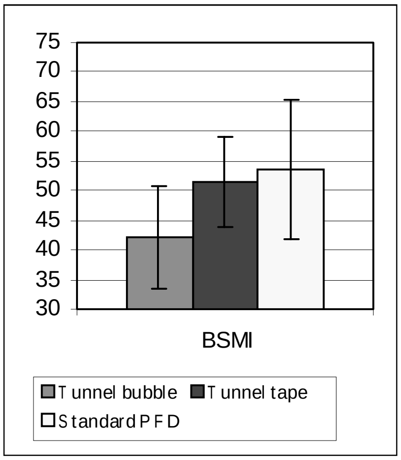


Fig. 18 Pilot ratings for crews favouring the tunnel display (n=9) on a Rating scale for Mental Effort (RSME or BSMI in Dutch)

The tunnel display for the ‘tunnel’ ‘group’ was rated as requiring less effort. The traditional display was rated worst, but the addition of ‘tapes’ to the tunnel apparently initiated a cost in its effectiveness with respect to estimated mental effort. This result seems quite in accordance with the data from Wickens (this volume) investigating the effects of conformal and non-conformal symbologies on Head up type of displays.

Auto pilot failure (PFD group)

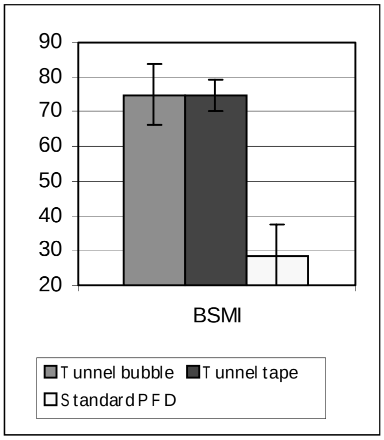


Fig. 19 Pilot ratings for crews favouring the traditional display (n=3) on a Rating scale for Mental Effort (RSME or BSMI in Dutch)

The pilots who preferred the traditional display, did this quite strongly as the results depicted in figure 19 illustrate. The traditional display was very easy for them while the tunnel display scored fairly high on costs of mental effort.

To explore the underlying reasons for the quite different ratings obtained from (some) individuals, other detailed questionnaires were explored. The results are depicted in figure 20.

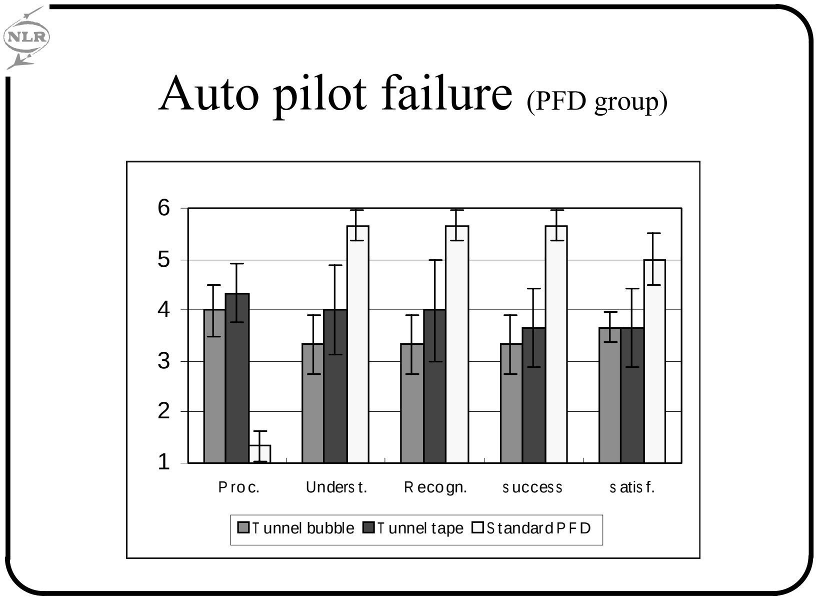
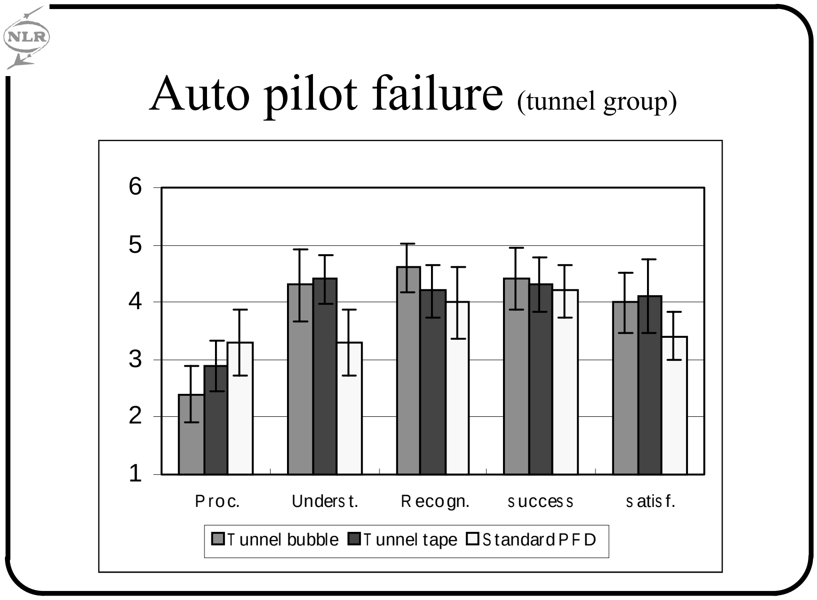


Fig. 20 Pilot ratings for crews favouring the traditional display (n=3) on a rating scales for Mental Processing- Understanding- Recognition- Success in task and Satisfaction with their performance

Note that for this group some extreme scores were obtained. The mental processing of the traditional display was rated extreme low, indicating high familiarity as confirmed by the scores on ‘Understanding’ and ‘Recognition’. Also success was rated higher although satisfaction achieved was expressed firmly but less explicit as compared to the other aspects rated.

If the results for the tunnel group are contrasted with the PFD group, a more balanced distribution in ratings is observed as depicted in figure 21.



Auto pilot failure (tunnel group)

Fig. 21 Pilot ratings Pilot ratings for crews favouring the traditional display (n=9) on rating scales for Mental Processing- Understanding- Recognition- Success in task and Satisfaction with their performance

5. Lessons to be learned

5.1 Unexpected user behaviours

If experiments allow the use of fine-grained objective measurements at both the performance and workload level, *unexpected user behaviours* can be identified and explored to understand and learn their cognitive origin and nature. The cockpit data link study revealed unwanted head down behaviour for the Pilot Flying and the task completion without allowing interruptions seemed to be pursued especially for the lesser optimal display. Supporting the user by providing more access and feedback will allow a different kind of user interaction as known beforehand. Understanding why this occurs is of crucial importance for improved designs.

The air traffic controllers were found to drop the use of new tools especially when the traffic load was high. Apparently, the tool did not help them or the training was insufficient to overcome the old habits'. More validation studies are clearly required before fielding such equipment, even when high investments in developing its technologies have been made.

Similarly, the pilots that did not favour the tunnel display as the majority of their colleague's. could also be influenced by these 'old habits'. Preference is always personal, but its dimensions have to be understood intensely in order to facilitate transitions to new technology and optimize training. An individualised approach seems to be indicated by the data obtained, imposing specific requirements to training design and tools required.

5.2 Objective versus subjective measurements

In several of our studies, objective measurements based on physiological and other data suggested a reduction in workload in most if not all of the measures taken. Subjective appreciation's however, pointed in the other direction and sometimes quite explicitly. The ATC datalink studies and the Automation assistance studies revealed this pattern of 'dissociation'. The major implication seems to be that the inclusion of objective measurements is essential to gain a full understanding of the processes involved in working with new interfaces, including adaptations and transitioning problems to be expected. Furthermore it is sometimes hard to estimate workload for complex systems and its components. Perhaps that is the reason for a possible bias in the direction of 'More tools or equipment must mean more work'!

6. Misconceptions and Fallacies

From these, and many other, studies it became apparent that certain assumptions and perhaps common sense believes are in fact not true and need to be reconsidered carefully.

- ***Automation will always be beneficial*****:** the data obtained in experiments employing fine grained performance and workload measurements indicate that many ‘tools’ will not be used as predicted or even at all, especially under high task loading conditions.
- ***Training will compensate for poor design*****:** this is an ‘oldie’ but still relevant. Without adequate testing and human validation studies, designs will be maintained that will provide difficulties during transitioning form old to new technologies. Even certain groups of users could feel, or actually be, excluded from the future work force. This aspect is detrimental for the individuals but also for a system that has to cope with decreasing numbers of potential candidates for operators.
- ***Function allocation can be best based on machine capabilities:*** this has been a trend for a long time because of obvious reasons. It works out better if designs appreciate the cognitive strategies that people employ generally.
- ***Users are the best designers*****:** this is most often not true as users will be used to particular equipment and procedures and gain pride in having mastered them. The transition to new technology will often require persuasion and re-training will prove ineffective some of the time.
- ***Human modelling is valid for future behaviour*****:** these models are especially appreciated by designers who do not want, or can not afford, the assumed burden of taking complex and extensive measurements. However, since these models are based on known behaviour patterns they will not include or cover unexpected behaviours and will not be valid for ‘certifying’ new systems.
- ***Ergonomics are checked after the design:*** in many cases basic technologies have been developed and the behavioural scientist is tasked to incorporate ergonomics. The task structure and working philosophies or sub task sequences will not fit anyway, leaving a sub optimal design for the user to master.
- ***Design is delayed by Human factors and increases costs*****:** if planned form the start, Human Factors can help and even take the lead in defining the design strategies for technology development. In all experiences known to me where engineers and physicists were teamed with behavioural scientists, clearly some adaptation problems occurred with respect to language and strategies pursued, but eventually everybody felt more secure about the quality of the products delivered. Overall costs of the total product development cycle and its introduction will be lowered instead of increased considering the early prevention of ‘after fielding’ problems like modifications or mid-life updates, excessive training requirements or even user resistance.

- ***Human Factors is a work area, not a behavioural science:*** with the increased recognition of the importance of ‘Human Factors’ for overall performance, acceptance and endurance of a system as a whole, an apparent market was opened with many contenders. It is often not realised that working with a display and computers is not sufficient to qualify as ‘Human Factors’ research. Due to the inherent individual differences between humans, adequate scientific standards are simply essential to gain interpretable results. Also the use of questionnaires is not sufficient and when not complemented by high quality objective measurements, such a strategy should be qualified as an *‘easy way out’*.

7. References

1. Folkerts, H. H. and Jorna, P.G.A.M. (1994). Pilot performance in automated cockpits: a comparison between moving and fixed throttle levers. NLR TP 94005 L.
2. Gent, van R.N.H.W., Bohnen, H.G.B. and Jorna, P.G.A.M. (1994). Flight simulator evaluation of baseline crew performance with three data link interfaces. NLR CR 94304 L.
3. Gent, van R.N.H.W., (1996) Human Factors issues with Airborne Data link: towards increased crew acceptance for both en-route and terminal flight operations. NLR technical publication, TP 95666.
4. Hilburn, B., Jorna, P.G.A.M. & Parasuraman, R (1995). The effect of advanced ATC strategic decision aiding automation on mental workload and monitoring performance: an empirical investigation in simulated Dutch airspace. In *Proceedings of the Eighth International Symposium on Aviation Psychology*, April 1995, Columbus, Ohio: Ohio State University.
5. Hilburn B.and Jorna P.G.A.M., Byrne E.A. & Parasuraman R. (1996). The Effect of Adaptive Air Traffic Control (ATC) Decision Aiding on Controller Mental Workload. Second Automation Technology and Human Performance Conference proceedings, March 7 - 9, 1996, Florida, USA.
6. Hilburn, B. (1996). The Impact of ATC Decision Aiding Automation on Controller Workload and Human-Machine System Performance. Thesis for Ph'd, available as NLR TP 96999.
7. Hooijer, J.S., (1996). Evaluation of a label oriented HMI for tactical datalink communication in ATC. NLR Technical Publication. TP 96676 L.
8. Huisman H. and Karwal A. (1995), 'Design and initial evaluation of two 4D PFD format prototypes'. NLR CR 95229 L.
9. Huisman, H. and Flohr, E. (1997) The design and validation of a Tunnel in the sky display for aircraft operating in advanced Air Traffic Control environments. NLR Technical publication TP 97xxx, in press.
10. Jorna, P.G.A.M. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. Biological psychology, 34, 237-257.
11. Jorna P.G.A.M. (1994). Impact of ATM on future flight deck design. In: proceedings of the European association of Aviation Psychology (EAAP) symposium Dublin, Ireland.
12. Jorna P.G.A.M. and Nijhuis, H.B. (1996) Ground Human Machine Interfaces (GHMI) for ATC automation technology: a European strategy towards coherent development and human factors validation. Second conference on Automation and Human Performance, Florida.

13. Jorna P.G.A.M. (1997). Context simulation: an interactive methodology for user centred system design and future operator behaviour validation. In: Cognitive Engineering in the Aviation Domain. Sarter & Amalberti (eds.) Lawrence Erlbaum Associates.
NLR TP 97032 L
14. PHARE GHMI (1995). Specification of the advanced systems GHMI Interface for PD/1. EUROCONTROL, Brussels, Belgium.