Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR TP 97032

Context simulation: an interactive methodology for user centred system design and future operator behaviour validation

P.G.A.M. Jorna

DOCUMENT CONTROL SHEET

	ORIGINATOR'S REF. TP 97032 U		SECURITY CLASS. Unclassified

ORIGINATOR
National Aerospace Laboratory NLR, Amsterdam, The Netherlands

TITLE
Context simulation: an interactive methodology for user centred system design and future operator behaviour validation

PUBLISHED IN
Cognitive Engineering in the Aviation domain (Sarter and Amalberti, eds.)
Lawrence Erlbaum inc. Publishers

AUTHORS	DATE	pp	ref
P.G.A.M. Jorna	971210	38	11

DESCRIPTORS

Air traffic control
Air traffic controllers (personnel
Aircraft communication
Cockpit simulators
Cognition
Cognitive psychology
Computerized simulation
Flight control
Flight management systems
Flight safety
Ground-air-ground communication
Man machine systems
Man-computer interface
Pilot performance
User requirements
Voice communication

ABSTRACT

Aviation is facing the challenges caused by its own success, namely solving the congestion in major air spaces and making air travel even more efficient. Different options and solutions have been proposed to increase the capacity and throughput of air travel and most, if not all of such 'Air Traffic Management' proposals include the development and implementation of advanced technologies and software tools. Lesson learnt from the past indicated that the application of technologies can and will change the role of the human operator. The predicitve capabilities for identifying human factors problems were, and still are, limited as most knowledge is based on existing systems. Therefore, not all of these changes for the human operator were predicted. Some of the unanticipated consequences were new types of human errors, issues like complacency, over reliance in automation and safety problems associated with maintaining situation (traffic/ground) and system (modes) awareness.

As predictions have proven not to be perfect, it is necessary to gain the required new knowledge about future systems by means of simulation technology. This report describes a methodology to perform experiments that aim at identifying and documenting such "human factors problems' during the design and validation process. An essential element of the methodology is the inclusion of 'context' in order to support cognitive engineering principles and to be able to measure the (cognitive) interactions of the human operator with future systems, with a high level of objectivity and validity.

The core of the concept is an integration of modern simulation techniques for both the flight deck and Air Traffic Control with extended objective measurements techniques for evaluating and documenting the responses of human operators.

Summary

Aviation is facing the challenges caused by its own success, namely solving the congestion in major air spaces and making air travel even more efficient. Different options and solutions have been proposed to increase the capacity and throughput of air travel and most, if not all of such 'Air Traffic Management' proposals include the development and implementation of advanced technologies and software tools. Lesson learnt from the past indicated that the application of technologies can and will change the role of the human operator. The predictive capabilities for identifying human factors problems were, and still are, limited as most knowledge is based on existing systems. Therefore, not all of these changes for the human operator were predicted. Some of the unanticipated consequences were new types of human errors, issues like complacency, over reliance in automation and safety problems associated with maintaining situation (traffic/ ground) and system (modes) awareness.

As predictions have proven not to be perfect, it is necessary to gain the required new knowledge about future systems by means of simulation technology. This report describes a methodology to perform experiments that aim at identifying and documenting such 'Human Factors problems' during the design and validation process. An essential element of the methodology is the inclusion of 'context' in order to support cognitive engineering principles and to be able to measure the (cognitive) interactions of the human operator with future systems, with a high level of objectivity and validity.

The core of the concept is an integration of modern simulation techniques for both the flight deck and Air Traffic Control with extended objective measurements techniques for evaluating and documenting the responses of human operators.

Contents

(38 pages in total)

1. Introduction

1.1 Lessons learnt in Aviation

The advent of highly automated aircraft with ‘glass cockpits’ has extended the capabilities of the aircraft but also changed the nature and type of the tasks that have to be performed by the crew. Flight control assistance and Flight Management Systems (FMS) have changed the pilots role from a manual controller and navigator to a systems monitor and information manager. The psycho-motor skills of the pilot are now complemented by information and resource management, task scheduling skills and handling the programming of on-boards computers. The continuing expansion of air travel, necessitates the use of even more advanced technologies in order to accommodate the expected levels of traffic while at the same time maintaining safety. Technologies like the digital data link play a key role in realising such an advancement.

New, or retrofitted, cockpit equipment and human machine interfaces have to be used in a complex operational context that is characterised by a so-called theoretical ‘time line’. This time line represents a mission profile specifying which tasks should be performed by whom, when and in what order. During the design process, time lines are used to assist cockpit design and define requirements for the input and display devices. The analysis is however, theoretical and is burdened in reality by variations in equipment use and extended response times. The FMS interface, as an example, proved to be quite difficult for computer illiterate line pilots and it took more time to reprogram as expected, leading to an advice not to use it anymore in busy terminal area’s. The actual time lines that are observed in reality, will furthermore depend on crew procedures, company policies and the occurrence of unplanned interruptions by factors like ATC, system messages, fellow crew members and cabin staff.

These trends in technology development have been accompanied by now well documented problems in Pilot- Aircraft interaction and apparently ‘distanced’ the crew from direct control over the aircraft control surfaces to a more indirect control by means of ‘instructing’ and monitoring the automation. Maintaining ‘mode’ or ‘system’ awareness under such conditions can be compromised. These trends (an extension on Billings by Jorna 1996) are summarised in figure 1.

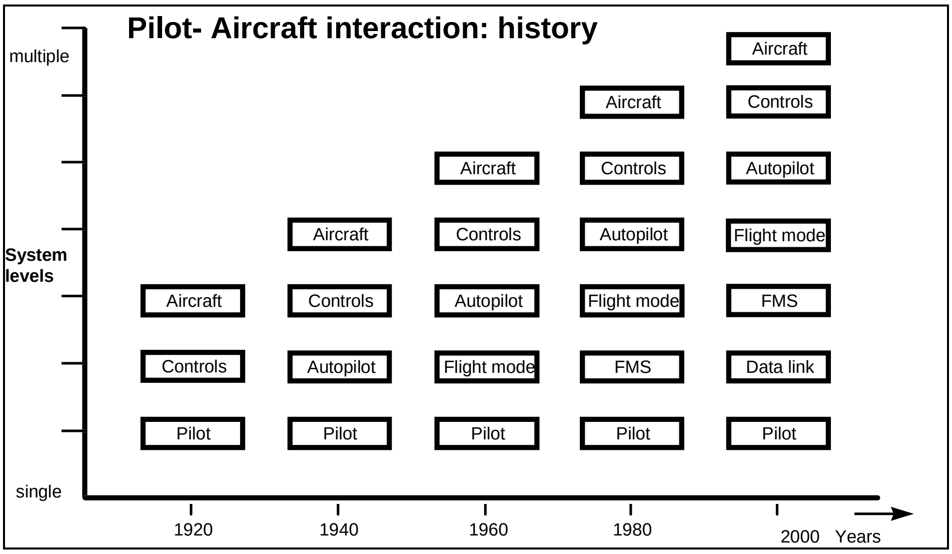

Fig. 1 Trends in 'distancing' the pilot from the aircraft including future data link ‘gating’, an auto load function of the FMS.

The working conditions for crews have also changed and are quite varying in their particular task loading during the flights. Cruise flight is relatively boring as compared to the sometimes hectic terminal area operations. Maintaining high alertness is always difficult under low task load conditions, so details can be missed, while during high task load conditions, stress can result in unexpected errors when handling tasks under time pressure. The human machine interface should support the crew adequately under all conditions, including handling multiple tasks under distracting conditions. An alert or annunciation that is perceived as adequate, or even annoying with respect to tone and sound level under low task loading, can go totally unnoticed under high task loading.

The development of the digital data link allows potential benefits for the future by creating a more efficient Air- Ground information exchange, both for the ground and the air. Technology wise, many opportunities and options now exist for a radical change in aviation operations. Data link can provide direct access to aircraft systems like the FMS, allowing in principle, fully ground controlled air traffic management. Crews in that case, monitor the progress and economics of the flight, evaluate ATC proposals and provide their consent to the up-linked clearances. Such an integration of data link with aircraft avionics is potentially able to reduce human typing errors, but can also introduce new errors. Under time pressure, pilots could revert to a strategy of 'accept first, think later'. Other ATM solutions focus on ‘Free flight’ type of solutions emphasing on-board equipment that allows aircraft to observe other aircraft in their vicinity on a cockpit traffic display. The consequences of such technologies for the operators in future systems are not yet fully understood and theoretical benefits should not be realised at the expense of (again) discovering unanticipated problems after ‘fielding’ such technology.

Designing the human operator out of the system has been a strategy for quite some time under the argument that human error is the basic safety problem. Opponents to the idea that human limitations are responsible for the majority of errors, argue that it is not the human itself but the quality of the total system concept, the quality of information provided and the effectiveness of the human machine interface that determines proneness to errors. The automation level itself is therefore not a threat to safety, but rather the transfer of information and the 'awareness' of system states and environment as mentioned earlier. Such issues, in psychological terms often described as 'Feedback' and 'Task involvement' issues, are key targets to be dealt with in new designs. The nature of the associated tasks in modern and future aircraft is and will be predominantly ‘cognitive’ and pilot-aircraft interaction is dominated by planning ahead, scanning displays, performing communication and using input devices for instructing the aircraft.

1.2 Who owns the problems?

The principal users of airspace are the airlines. They should generate a profit in order to be able to provide resources for the airplane manufacturers and the ground systems. The 'user charges' or landing fees are not negligible. Investments in cockpit technology for future Air Traffic Management systems are weighted for their economical revenues and will not be easily adopted when they only apply to part of the world. Preferred routing is not only a means for reducing costs, but should also allow business strategies like f.i. adopting an 'on time' image or a 'budget' image of an airline. Any investment in cockpit technology should be therefore be paid by an increase in revenues/safety or a reduction in costs.

Improving existing aircraft, by retrofitting equipment etc. is often performed after the aviation authorities have issued a mandatory requirement to do so. In case of the complex problems associated with the ‘Human Factors’ of design and operational issues, like ‘mode control panels’ or ‘use of colour’ as an example, certification rules do not yet seem to cover all implications of certain modifications and therefore leave options for different solutions. The manufacturer (or the proposing airline) has to prove that the solution is valid and such a process is intensive and quite expensive to perform. Economical criteria will therefore influence overall strategy. Similarly, some options on improved cockpit equipment, offered voluntarily by the manufacturers, will not sell in sufficient numbers, as the customers will signal that there is no formal requirement of the authorities to implement them. The issue of which party is responsible for research on Human factors is therefore unclear and will be at least fragmented over the many partners involved in the aviation community.

In the development process of new aircraft, a similar apparent paradox has occurred. New technology applied to these aircraft will normally require the re-training of crews. The airlines nowadays do not seem to accept such a dependency. An often cited airline requirement is the 'zero transition' requirement. Pilots with a type rating on one version of an aircraft should be able to fly the update with none (or very limited) training. 'Trainability' of cockpit concepts and modifications is therefore an important driver for the design. Designing flight decks that deal with both new developments and incorporate lessons learned, is complicated and hampered by these constraints and the net result could be a commercial design strategy of 'no change'. In the end, that strategy will be counter productive as the market clearly needs improved and more reliable levels of performance for human- machine systems.

2. The future revisited

2.1 Cognitive Engineering and user centred system development

Integration of components is a well known issue in designing technical systems , but integrating the human proved even more difficult sometimes. A classic Human Factors topic is the problem of subject variability. Pilots not only differ in qualification or experience (and personality), but also show variations in their behaviour and strategies over time. Their experience is increased or factors like fatigue, family crises etc. create distractions. Pilots are also known to be creative for inventing new ways for accomplishing their work by using equipment differently as anticipated by the original design. Flight management system functions can also be used to depict turbulence area's on the displays during oceanic flights. The turbulence or problem area is depicted by entering a scratch pad route on the Control Display Unit (CDU), and should therefore never be activated accidentally. Weather radars and Traffic Collision Avoidance Systems (TCAS) are nice to create some situational awareness, but were never intended for actual navigation.

The flexibility of human behaviour has both positive and negative effects and its underlying priciples need to be assessed and understood by realistic 'man-in the loop ' testing before fielding. Sample sizes of subject pilots or crews should be large enough to draw valid conclusions for the total range of users involved and not only for test pilots or airline 'aces'. The measurements taken should be fine grained enough for assessing both system and human performance implications of the technology and the scenario's should invite the operator to modify and try different strategies. After sufficient understanding has been obtained, adequate and efficient procedures can be developed to optimise the benefits and minimise possible risks. Procedures are therefore an integrated part of the design process, but many airlines adopt their own.

Ideally, testing of so-called 'pre-operational systems' is performed with the system as a whole, but costs will be enormous if all the technology has to be developed first. It is therefore critical to find or develop effective ways for testing the components of an envisaged system on their 'feasibility of fit'. As no one can afford the actual design of a total system and discard it after unsatisfactory validation, a stepwise approach is needed to discover flaws and wrong assumptions as early in the design process as possible. The risk of a technology driven approach, is adressing specific issues in isolation, like cockpit lay-out, avionics, ATC tools or specific procedures. Human- Machine interaction problems will not be raised or overlooked, because the total context of work is not yet known or ill-considered. A basic requirement for any systematic approach is therefore the ability for pointing at possible unanticipated interactions between the partners in a system that could prove critical to system effectiveness and safety in a later stage.

2.2 The T.E.S.T. approach

Ad hoc research on ‘ human factors’ safety issues that were discovered after fielding of the systems, often showed unanticipated interactions and relationships between factors like equipment design, the particular work environments and the responses of the human operators concerned. In many cases, additional effort was required to make the system work, like more extensive training or equipment upgrades. A mediating factor in this respect seems to be the particular focus of the engineering process.

Industries providing products for flight decks, have to sell systems or ‘boxes’ and engineering is therefore often geared towards providing many attractive functions into one device. In practice, not all of these functions are used. In contrast with such a system engineering perspective, human engineering is much more focussing on tasks. Executing a task can involve more than one system. The ‘technical’ and ‘human’ perspective can lead to different results as will be illustrated by a simple example. Computer displays have a function for controlling the amount of light emitted. An electrical engineer will perfectly design such a function with the electronics available. A pilot, wanting to ‘turn up’ the flight display, can be baffled initially, by the display suddenly going black. Looking more closely at the small print, it will be discovered that the function was actually a ‘DIM’ function. The pilot in this case, expected a brightness control but was confronted with ‘more dim= less light’. Such discrepancies in perspective occur regularly and should be identified and resolved especially for the more safety critical situations.

The so-called ‘validation process’ of cockpit and ATC equipment will have to deal with this interplay between multiple systems and multiple users differing in background, expectations etc. The relevant human performance ‘shaping’ factors can be summarised by the ‘T.E.S.T.’ approach (Jorna 1993).

The T.E.S.T. acronym lists the variables and possible mutual influences (interactions) that have to be addressed or controlled in design and validation.

T .= **Task** parameters that influence difficulty and limit human performance.

E. = **Environment**al factors that complicate task execution or limit the operators ability.

S. = **Subject** characteristics that influence individual performance, acceptance or availability.

T. = **Training** and practice requirements.

<u>Task:</u> In the design process it is ‘task analysis’ that should provide a definition of actions and duties to be performed. Experience has learned that such an analysis is either not available or not carried out at the most effective level of detail. The human factors researcher confronted

with an existing system to be evaluated or assessed, is often required to perform an 'in-field' approximation of such analysis. In the course of up front validation, it is practically difficult but essential to incorporate all possible task levels for measuring the effects of interactions between tasks on operator performance.

Environment: During short and long haul flight, interruptions of on-going activities occur that can distract the crew and leave tasks unattended or unfinished, especially when the display formats do not indicate such omissions. The working conditions experienced during cruise flight are generally not very loading, leading to vigilance and alertness problems, this in contrast with the hectic terminal area operations were crews are loaded with (too) many tasks. Additional environmental factors like noise levels, humidity, extreme exposures to time zones, or G-forces for military pilots can all affect the mental fitness level of the crew.

Subjects: A classical pitfall in design or demonstrations is the use of highly skilled subjects like test pilots or very experienced instructors. The effect will be twofold: if there is a negative transfer of old working habits to new designs, than the potential of the new design will be underestimated. Alternatively, a test pilot will not be fatigued, jet-lagged, bored, or otherwise impaired as with people who have to operate under normal daily life circumstances. Another bottleneck is that most task analysis methods will specify tasks as they are, meaning independently from the kind of operator, while cultural differences exists.

Training: Long term exposure to a new design is typically not performed. Training changes the locus of the human limitations from conscious information processing, like cognition or knowledge based performance, to the limits of particular sensory or response capabilities that are associated with practised, skill-based and more ‘automatic’ ways of performing , like the reliance on routine planning, data entry or use of input devices (Jones 1980, Wickens 1992).

Often the wording ‘tasks’ and ‘skills’ are used interchangeably, but there is a distinct difference. Performing the same task like ‘hammering a nail in a piece of wood’ under different circumstances can involve totally different skills. Imagine hammering in the open air (no problem for most people) as compared to hammering ‘under water’ by a diver (wood suddenly floats and it is a bit dark). In addition to such a factor, time restrictions also play a role in determining the required level of skill. When landing a general aviation aircraft, completing a ‘circuit’ and performing down wind checks with a slow airplane requires different skills, or levels of skills, as compared to a fast(er) airplane. If the circuit cannot be extended for noise abatement reasons, time pressure will be imposed on all the checks and communications required. Planning and anticipation are suddenly even more critical as they are normally.

As a rule of thumb, a ‘skill’ can therefore only be defined if:

1) the task to be executed is known;
2) the working environment and context, including other tasks, is known and
3) the timing pattern required is known.

Cognitive engineering is a ‘hybrid’ discipline that addresses new types of tasks or systems as evolved in aviation, not in isolation, but in their entire context to predict and assess strong and weak points of cockpit designs. In designing such systems, exploiting the human capabilities is emphasised instead of designing around the human limitations.

2.3 Context simulation

As a result of these developments, the NLR developed an integrated methodology for the design, evaluation and validation of crew interfaces, work procedures and flight deck/ATC design by adding simulations that consider the possible operational contexts. The simulation scenario includes a multi-task environment combined with operationally relevant environmental factors and uses a wide(r) range of subject pilots. These simulations are quite elaborate and are therefore performed after a selection process of identifying likely candidate solutions for the Human machine Interface concerned. This process involves rapid prototyping by the so-called NADDES system (NLR Avionics Display and Design Evaluation System) and evaluations are performed in ‘laboratory type’ of experiments (workstations etc.). In both stages of development the ‘Context’ factors as specified by T.E.S.T. are included, but they differ in scope and realism. Distractions are, as an example, generated in the laboratory experiments by using a standardised additional or dual task added to the work to be performed during the workstation experiment, while in the ‘full context simulation’ events are used that can occur naturally under operational circumstances. Both context environments do, however, include a range of variations in task loading and working conditions and allow possible unanticipated interactions to occur.

The subject pilots used are not ‘aces’ but regular line pilots from different cultural backgrounds with a range of experience from high to low. Extensive, computer supported familiarisation and training is provided with the human- machine interfaces before starting the measurements.

The importance of context for understanding pilot behaviour will be illustrated by a number of case studies, discussing the results of research performed by the methodology discussed. There are however some problems to be resolved first before the methodology can be applied. These will be discussed to illustrate the technology requirements for ‘context simulation’.

2.4 Scientific standards

A major disadvantage of using interactive simulations is that the crew will be able to influence the actual route(s) flown, decide on the initiation of communications etc. It is therefore difficult to compare the results and observations across crews, as the actual flights can develop quite differently, leading to different working circumstances and/or traffic encountered. One crew strategy can make life easy, but another can result in difficult situations depending on its effectiveness. Next to similar experimental conditions, science requires sufficient numbers of repeated measurements, in order to assess the relevance of the results with statistical testing. The solution to this problem was the definition of so-called ‘Events’ that can be inserted in the scenario at will by the experiment leader. The simplest example is an engine failure that can be initiated by an instructor in most of the flight simulators used for training purposes. More complex is to assure that the working context for initiating such event should also be comparable between subjects crews. An instructor would be required to monitor all details consistently, to adhere to more scientific standards. This issue was resolved by automatic monitoring of selected flight parameters and status of equipment. Monitoring this data, would create a situation where it is possible to initiate an ATC event like a data link up link, precisely at the moment that the Pilot Non-Flying is at a certain page of the Control and Display Unit of the Flight Management System. This test condition can be arranged for all crews involved. Also, events can now be triggered under different levels of workload by considering a particular combination or sequence of events, representing a IF this, AND that, initiate ‘event’, ELSE do not trigger.

An example of such a manipulative scenario is presented in figure 2.

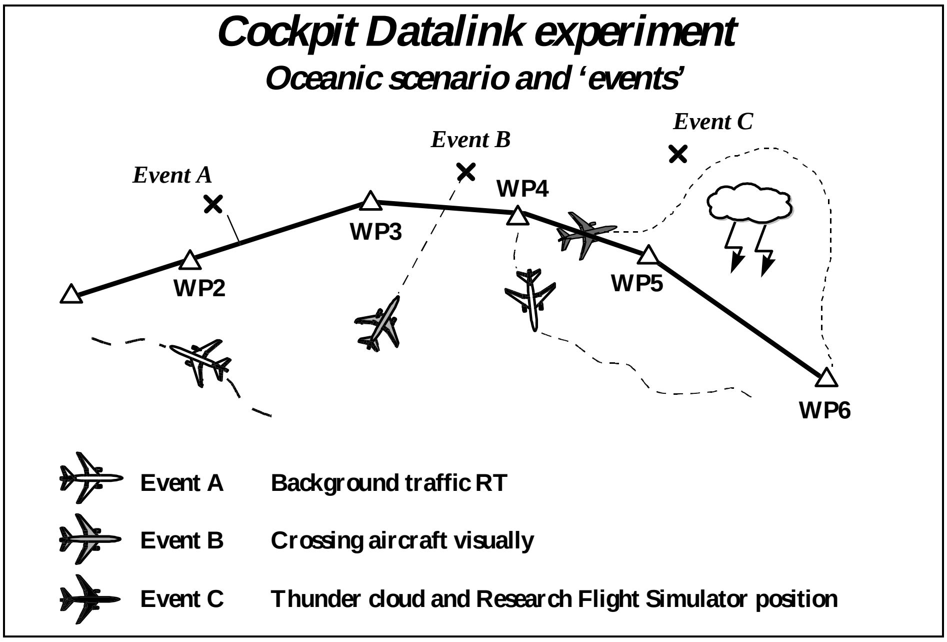

Fig. 2 Examples of ‘events’ that can be initiated under full experimental control

2.5 Simulation and scenario management

The simulation facilities need to be equipped with more extensive and accurate control over the scenario. This comprises factors like:

- Control over presentation of the events, both in number and sequence.
- Monitoring of many system status parameters, like FMS page on top, or mode control panel settings etc., to allow triggering of events in relation to a predetermined initial condition.
- Recording of parameters with accurate time coding and inclusion of event codes for later identification and automatic computer processing of data.
- Extensive and accessible storage capacity.

The so-called '*Experiment Manager facility*' developed for this purpose, is a dedicated computer that contains a list of scripted events for the simulation computers to execute. These are initiated either automatically or manually when the required conditions happen to occur during that simulation. It therefore monitors a set of prescribed simulation parameters to allow adequate triggering of the required events. The experiment manager has a 'manual' mode in which the experiment leader can work interactively whenever required. As an example, an event like 'Engine failure' can be scripted to occur twice in the scenario, once during communication with ATC, and once during a relaxed period in the flight. In this way it is possible to still allow the pilot to perform his duties under realistic circumstances, while also maintaining rigorous experimental control over the number and initial conditions of the events under investigation.

3. The Human-Machine Co-operation 'ATM Test bed'

3.1 Embedding Aircraft in Air Traffic Management

The future Air traffic control environment needs to be updated in order to resolve present day delays and congestion's and to accommodate the projected traffic loads for the future. The solutions envisaged all depend on an improved exchange of data between aircraft(s) and/or ground equipment. The establishment of digital datalinks is a key technology for realising such options. However, the working conditions of the crews and controllers can change drastically and the interactions that will occur between human-machine teams in the air and on the ground are still largely unknown.

Experiments need to involve both parties in dynamic scenario's that allow these interactions to actually occur and study them for effectiveness and safety issues. For this reason, the simulation of experimental aircraft cockpits was expanded by connecting similar experimental controller working positions containing advanced tools for Air Traffic Management. The 'experiment manager facility' will in that case have to control events and monitor the status of two sophisticated simulations in full context. In addition, a full battery of measurements concerning human perfornace has been integrated.
The resulting 'ATM testbed' is illustrated in figure 3.

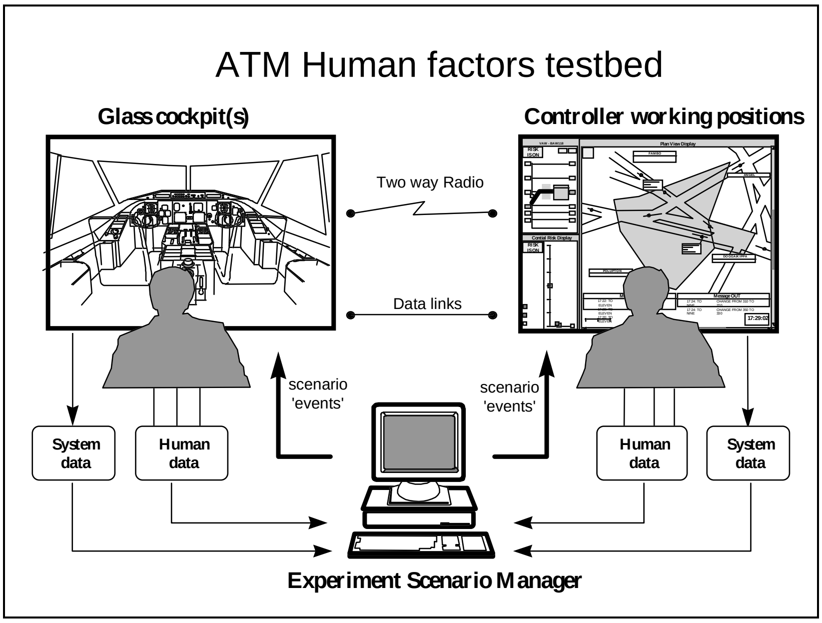

Fig. 3 An ATM testbed for designing, simulating and studying human interactions, work procedures and safety issues in experimental ATM environments.

3.2 Human data measurements

The following data is obtained to evaluate human performance, workload and effort, situational, meaning traffic awareness, systems awareness and user preferences:

- sampling of visual data on displays by ‘Point of gaze’ head mounted eye trackers which are calibrated to the particular simulator in order to depict active use of the displayed information. The system provides the following information in real time, with a sampling rate of 50 Hz:
  - Point of gaze, expressed in X and Y co-ordinates relative to the viewing plane;
  - Fixation dwell time, in msec;
  - Millisecond-accurate time stamp (i.e., starting time of fixation), to permit referencing to simulation events;
  - Transitions between display elements and other displays;
  - Pupil diameter, which can be converted into millimeters;
  - Surface identification, for translating pre-defined planar co-ordinates into viewing surfaces (e.g., separate dials of a simulated cockpit display).
- analysing changes in heart rate to assure that the information ‘looked at’ is also actually processed by the crew or controller in order to make sure that information is also ‘seen’.
- calculating heart rate variability to monitor the mental state and effort excerted during processing of the information.
- analysing vocal communications within the crew or between controllers as well as communication outside. So-called voice key’s (electronics that indicate both onset and duration of speech) are combined with ‘press to talk’ switches to discriminate between communications inside/outside.
- recording respiration to control for breathholds influencing heart rate and control over the occurrence of murmured speech not detected by voice keys.
- a battery of questionnaires depending on the goals of the study, but standard work load ratings are always used.

4. Case studies

4.1 Back driven throttles versus non-backdriven throttles

Two different design-philosophies with respect to the Human-Machine Interface of Auto Thrust Systems (ATS) were compared (Folkerts & Jorna, 1994). With the so-called moving or backdriven thrust lever concept, the positions of the thrust levers are servo-controlled. Hence, the thrust levers provide the pilot with positional information regarding actions taken by the auto thrust system. In contrast, the so-called non-moving thrust levers are not servo-controlled by the auto thrust system. The intent information can be provided by a modified engine display with an alternative visual display element reflecting the 'commanded thrust' levels. The primary flight display (PFD) includes a 'speed trend' vector on the speed tape to assist the pilot in anticipating and assessing speed changes.

Moving thrust levers could serve the following purposes:

- They could support 'attention getting' as movements can be detected by peripheral vision.
- Thrust levers can provide tactile feedback that could be valuable during head-up operations like approach and landing.
- Representative lever positions for thrust levels can increase redundancy and thrust awareness.
- Thrust levers provide 'lead time' in advance of the actual engine response to a desired thrust setting. This particular informational aspect represents a 'commanded thrust' level indication the moment the levers are set in a new position by either ATS or pilot. The engines response itself is delayed due to the spool-up time required.

4.2 Selected flight events and aircraft model

The experimental events were derived from accident histories and discussions with pilots and aircraft systems experts. The following classes of events were identified: Auto flight problems, engine troubles, environmental influences and some special cases like system anomalies, software bugs and runway obstructions.

Table 1 Selected 'Event's inserted during the flight

Type of 'Event'	Description
1.Capture of wrong FL	the AP captures FL 190, in stead of ATC cleared FL 180, early thrust increase
2. Failure to capture FL	while descending to FL100, the AP does not capture FL 100, no increased thrust.
3. Failure to reduce thrust	after initiating the descent, the ATS does not reduce thrust.
4. Failure to capture speed	after reducing speed to the selected speed, the ATS remains idle and speed drops
5. Unexpected idle	during the approach an unexpected idle occurs.
6. Engine failure slow	one engine thrust reduces slowly to idle level.
7. Engine failure fast	instantaneous loss of one engine.

The aircraft model had the following characteristics:

- Two engines, so an engine failure will cause a 50% loss in thrust that requires a significant increase of thrust for the remaining engine.
- Engines located close to the centre-line of the aircraft to minimise yaw during engine failure.
- The sound of the engines was suppressed, to reflect low noise, clean aircraft.
- A stabilisation of attitude by so-called 'control wheel steering'. This functionality provides for attitude hold and rate command response to control inputs. The net result is that a speed change does not affect the required control wheel or stick forces.

The pilots were instructed to press a button on the control column to indicate that some 'problem' was detected. Solving the problem itself was not part of the study and was simply 'reset' by the experiment leader, who served as pilot not flying (PNF). With this 'standard crew member' flight procedures could be standardised and observations could easily be made.
An important factor to obtain results with some realism, is to control the level of 'expectancy' of the subject pilot. To prevent excessive pilot readiness for detecting problems, some routine flights without any problems were included. The pilot was informed about the existence of these normal flights but could not predict their occurrence. Events were presented randomly across flights.

4.3 Detection of aircraft malfunctions

Eight recently licensed civil pilots (no prior ATS experience or preference) flew both thrust lever configurations under automatic and manual flight conditions and tried to detect problems and malfunctions by using all sources of information available to them. Within the pilots, quite varying strategies in information sampling were observed, some using all displays, others concentrating on the PFD.

Table 2 and 3 summarise the overall detection results by means of movements of the levers.
No events went undetected in case of missing the throttle responses. When flying with moving thrust levers, pilots reported the engine display and the PFD less often as a source of information than when flying with non-moving thrust levers. So, the movement of the thrust levers reduces the dependence on the engine display and the PFD as a source of information.

Table 2 Percentage of the 'events' detected by means of the thrust levers

Descriptions of the events	Moving
Capture of a wrong flight level	55%
Failure to capture the selected flight level	63%
Failure to reduce thrust	67%
Failure to capture the selected speed	50%
Unexpected idle	47%
Engine failure slow	-
Engine failure fast	11%

Table 3 Trade-off between thrust levers and alternative sources of information

	Engine Display	Thrust Levers	PFD	Stall	Total
Moving	11 (11%)	46 (47%)	41 (41%)	1 (1%)	99 (100%)
Non-moving	31 (36%)		51 (59%)	3 (3%)	87 (100%)

4.4 Response times

Overall, the pilots detected an event more quickly when flying with moving thrust levers, as illustrated in figure 4. Three particular events discriminated significantly between the non-moving and moving thrust lever conditions:

- Failure to capture the selected flight level (event 2)
- Failure to reduce thrust (event 3)
- Failure to capture the selected speed (event 4)

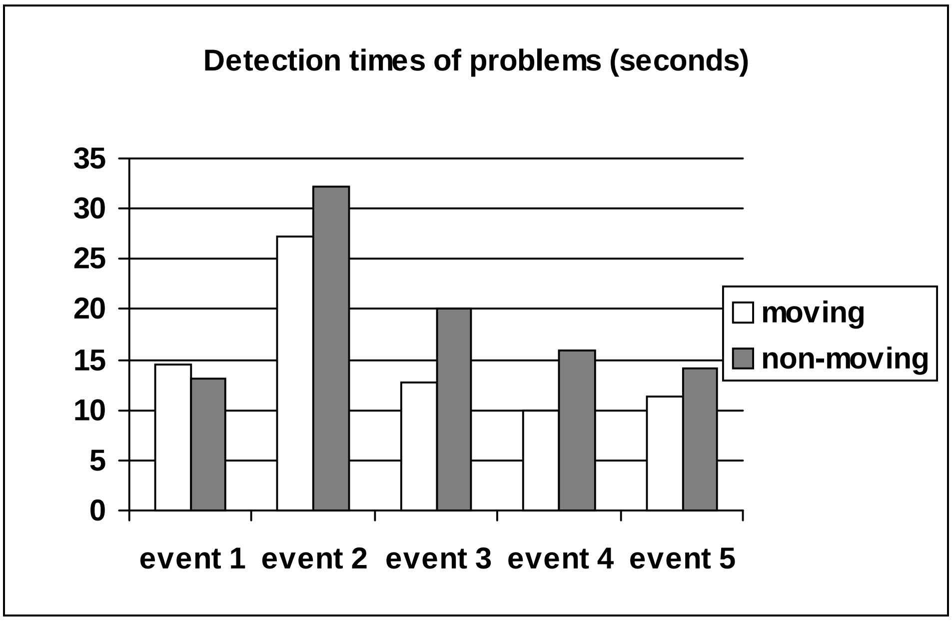

Fig. 4 Detection times of events during the flight. (engine failures excluded, for event codes see table 1)

These events seem to share a common characteristic: the aircraft or its systems 'fail' to perform some action or manoeuvre. The pilots expect the aircraft to reduce or increase thrust, so the thrust levers should move, but this movement does not occur. It is the 'missing movement' that informs the pilot of potential trouble. This finding is particularly interesting, as it indicates that the pilot's state of system awareness is supported by feedback from the system. *Missing information on the right time* represents new information. Thrust levers seem to have a particular value as providing a 'feedback' function on the operation of cockpit automation like the ATS or AP.

Expectations of pilots seem crucial to the net effect of movement of the thrust levers, as unanticipated events are clearly not consistently detected through these motions. Particular examples in this respect were the simulated engine problems. It was an unexpected result that the thrust levers apparently did not support the pilot as a primary information source on engine problems, either slow and insidious, or more violent, like the loss of half the power. An 'alert' or 'attention getting' function of the thrust levers could not be demonstrated convincingly for the descent phase of the flight as used in this study.

The general 'detection-time' required for a malfunction could be considered as rather long as values of 10-15 seconds or even more, are quite longer than the general human response times of 500-1000 msec in laboratory experiments. The reported numbers represent relative values and do not strictly represent sole pilot response time, as they include other factors. A contributing factor to the long detection times could be related to the strategy of the pilot. It was noted that 'false alarms' hardly occurred, which opens the possibility that pilots did not strictly adhere to the instructions of 'responding as quickly as possible', but performed some analysis on the origin of the problem before they decided to press the button. In that case, they knew for sure that there was a problem and did not initiate a false alarm. This phenomenon is well known in experimental psychology as the 'speed/accuracy trade off'. Accepting a cautious strategy is well respected in aviation, leaving the response times as still operationally realistic.

4.4.1 Discussion

The use of moving thrust levers as part of the human- machine interface was found to have limited but consistent advantages. Detection times are reduced especially when pilots expect a certain behaviour of the aircraft and it does not occur. The study clearly demonstrated the importance of pilot expectations and system awareness on the timely identification of potential problems. Another aspect is the observation of varying individual strategies in information sampling between pilots, even after recent training. These impact on the use of different information sources on the flight deck and the related response times.

4.5 Cockpit data link communication with air traffic control

4.5.1 Experimental design

The introduction of digital datalinks by means of UHF or mode-S, enables computer to computer communications between aircraft and ground systems. A well known example is the familiar ACARS unit that is used for communicating with the airline. In the study to be discussed, the possible application for communicating with Air Traffic Control was investigated (Gent, van et. al., 1994). A realistic scenario was designed around Oceanic routes, starting above the Atlantic, crossing England and finishing with a full stop landing at Schiphol airport. Three flight segments with different working conditions and time pressures were therefore included: Oceanic, Cruise flight and Descent. According to the concept of 'context simulation' several events were scripted including crossing and overtaking aircraft (also represented in the visual of the simulator), thunder clouds to initiate route changes and all communications were scripted as 'events' like uplinks, downlinks etc. This set up proved to be very efficient in realising fine grained analysis of crew behaviour and strategies.

Three cockpit data link devices and respective Human-Machine Interfaces were compared for effectiveness and acceptability: an Interactive Display Unit (IDU), the Control and Display Unit (CDU) and a Multi Function Display (MFD). All crews flew with all interfaces in separate flights.

Human data recorded included: head tracking to determine 'Head down time' and 'Head front time', meaning scanning either outside or on the primary flight instruments; also heart rate, respiration, voice keys, questionnaires and extensive system data including all buttons and switches were recorded. Some of the results will now be discussed.

4.5.2 Results

Crew procedures prescribe a task division between the crew members, with the 'Pilot Flying' handling the aircraft primarily and the 'Pilot Non Flying' handling the communications. In the case of data link this implies that the PNF will go head down while the PF can remain 'head-up'or head-front' for scanning and checking. The measurements of head tracking proved otherwise as illustrated in figure 5.

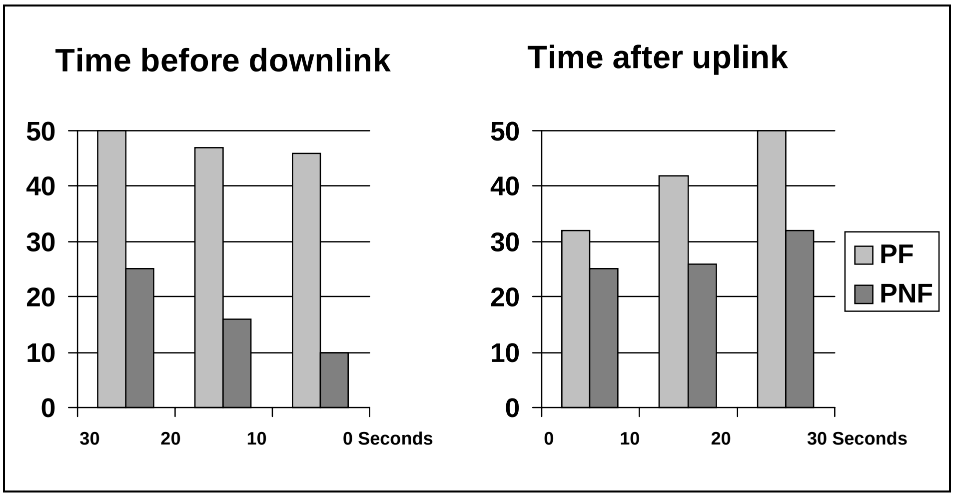

Fig. 5 Percentage of time spent 'head-up' or 'head front' for both crew members in case of composing a downlink message and receiving an uplink from Air Traffic Control.

During the downlink, the PNF is head down as expected in order to reach the input devices (IDU, CDU or MFD). The PF essentially remains head up in accordance with crew procedures. When the moment of sending is coming nearer, a trend towards head down can be observed. However, when receiving an uplink from ATC, both crewmembers go head down immediately, although the PF somewhat less than the PNF. In cognitive engineering terms this can be easily understood as humans are simply curious and want to be informed directly about anything of interest for executing their tasks.

As a result of this study, synthetic speech applications were investigated in a phase II study, to improve head-up times especially during the more critical flight phases (Gent van, 1996).

Data link transmissions take time and cause a delay factor in the communication as compared to the RT (Radio Telephony) baseline. Concerns are even greater as the delay times are variable as a function of the particular medium used in the Aeronautical Telecommunications Network (ATN). These variations prevent the crews and controllers to anticipate the delays and can disrupt task scheduling capabilities. From a cognitive engineering point of view, having one type of consistent delay would be better handled by humans than ever varying delays as they cause concern about the integrity of the system and require more feedback information concerning the actual status of the transmissions. Examples of such messages are: ‘sending’, ‘sent’ and ‘received by ground system’. With such information the crew has some insight if they are waiting for a slow controller or if there was a medium problem. In this study, the actual delays observed in accepting uplinks could be studied and replicated easily by using the ‘event coding’ of context simulation. The results are depicted in figure 6.

At first look, datalink communication seems quite slower as compared with voice, however when the data is corrected for particular types of messages requiring extensive routine device interactions, the differences are reduced considerably. So for important communications, datalink in itself is not critically slower as compared to voice, provided that the communication interface does not require extensive manipulations to create an answer.

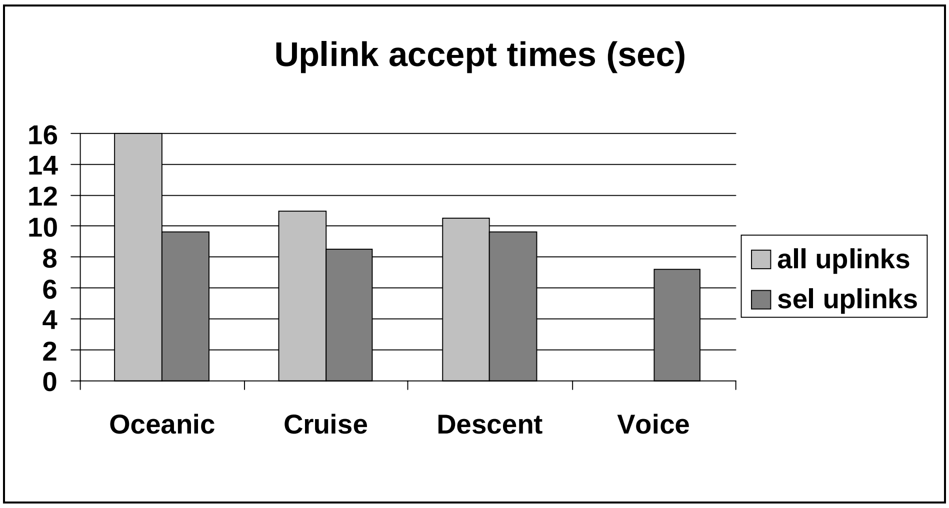

Fig. 6 Acceptance times for crews to uplinks with clearances etc. from Air Traffic Control. Also results for a subset of uplinks, excluding meteo and free text are included.

The quality of the Human Machine Interface proved to be important in more than one respect. Next to the time required for message and requests composition, the support provided to the operator proved the be a determinator for the way the device is actually being used. Quite unexpectedly, it was found that the worst device seemed to be used in a ‘single shot mode’, meaning that when you start the task, you will finish it in one go without allowing interruptions. The number of disruptions found was related to flight phase as illustrated in figure 7.

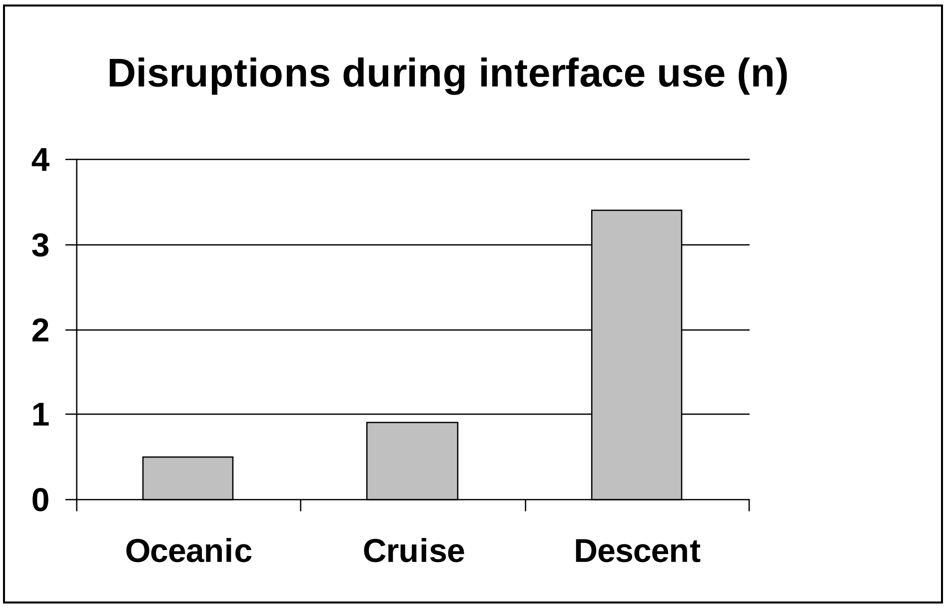

Fig. 7 Average number of procedure disruptions during use of the Human machine Interfaces for data linking. High time pressure increases the occurrences.

The IDU, CDU and MFD proved to different with respect to this factor. Apparently, when a user interface provides some feedback on 'where you are in the process', it is more easy and tempting to interrupt the task as you can find your way back. Supportive interfaces should allow for such a strategy as pilots will become more flexible in their task scheduling. In that case it is critical that the interface indicates 'unfinished business' very clearly!
The results are depicted in figure 8.

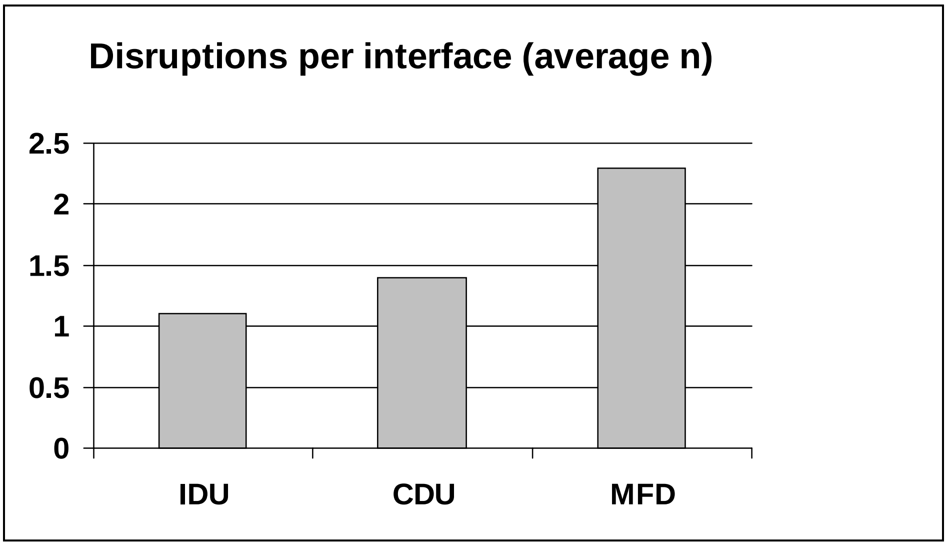

Fig. 8 Average number of task procedure disruptions during use of the Human machine Interfaces for data linking.

4.6 Pilot heart rate responses to Data link communication

The purpose of this study was to investigate the feasibility of applying a particular ‘heart rate analysis’ method to the assessment of human machine interface effectiveness in existing and advanced versions of the so-called ‘Glass cockpits’ (Jorna, 1997). The methodology proposed is: ‘Event- related Heart Rate’(EHR) as an indicator of the dynamic characteristics of human information processing during instrument and display scanning and the execution of discrete tasks. Its use is only possible within a context simulation with extended coding of ‘events’.

4.7 Experimental design

The simulation study investigated the crew interactions with an improved cockpit data link interface with two means for exchanging the message to the crew (van Gent 1996). In the first case, data link messages were displayed at a multi-function display, so the pilots(s) could read the message. The second display format was auditory, as the message was presented by means of synthetic speech. Both formats were tested under three levels of automation or ‘gating’ facilities. Up links from ATC could be stored in the FMS directly for execution after pilot consent. The pilot is not required to re-type all instructions. A second level was an extension of FMS gating with the automatic setting of the mode control panel (MCP) entries, called ‘MCP gating’. Both were compared with a full manual condition that served as a baseline, resulting in three levels of automation.

Event- related Heart Rate (EHR) is a basic technique. In the scenario, certain stimulus events are presented to which a pilot response is anticipated. After the experiment, segments of heart rate are ‘cut’ and *averaged* together. The reasoning is that all fluctuations that are not related to the specific event, will average out. The result is a net response of that subject linked to a particular event.

Comparisons can be made between type of events, for example different types of alerting systems, on the effectiveness of attracting the ‘pilots attention’. It is however, required to base that comparison on the same number of events, or on a sufficient sample size per event. The exact number required has to be determined by practical experience, or standard samples. The advantage of this method is that the response of the pilot is close in time to the event of interest and therefore only the fast responding para-sympathetic nervous branch will be involved for the first 10-20 seconds after the event.

4.8 Technology requirements

Heart rate should be recorded on digital equipment without tapes that have stretching problems and together with respiration to control for sighing, deep breaths etc. Especially breath holding will initiate a reflex that reduces heart rate. Artefact checking and correction facilities should be implemented. There must be an accurate linking with respect to real time accuracy with the stimulus or mission events. These events should be stored preferably in the same data file to allow adequate inspection and selection of relevant heart rate segments. Software should be able to locate the selected events and cut a specified section of the recording. This section includes some time before the event and some time after the event in order to study the initial values and the change in response as a function of the event. The exact time required depends on the goals of the study.

Heart rate is an a-synchronous signal that has to be interpolated in order to enable the cut and paste function accurately over a certain time segment. Averaging software for segments is required to obtain a global heart rate profile that can typify a pilots response to the event. Comparisons can only be made if the number of events in certain conditions is similar, and their presentation order is counterbalanced in the scenario, according to normal scientific standards of the behavioural sciences.
Graphing of combined profiles requires software that can determine the required scale to fit both, or more, profiles. That feature is typically not present in commercial software and is therefore included in own developments.

4.8.1 Data recordings and 'events'

The following events in the data link experiment were controlled and recorded by means of the Experiment Manager facility:

event code	description
232	'Roger'
235	''WILCO''
237	Moderate turbulence on
238	Light turbulence
239	No turbulence
240	Destination, runway change ATC
243	Radio communication other traffic
245	Traffic modification messages
246	No auto land
247	Auto throttle disconnect
248	Auto pilot disconnect
249	Comm. Light switched off
250	Data link light switched off
251	ACARS message
252	Urgent ATC up link
253	Aural alert from ATC up link
254	No aural alert from ATC up link

4.8.2 Results

The analysis performed used the event 253 'Aural alert from ATC up link' as the event for selecting, cutting and averaging of heart rate segments. Normally, that event will first attract the attention of the pilot, to be followed by the task elements: process the presented information; decide on the action required in co-ordination with the fellow crew member and finish with the initiation and execution of the response by means of a down link message to ATC. In the case of data link, all responses are made through selecting the standard reply 'WILCO' from a menu presented on a display.

The averaged segments are standard wise depicted with a running indication of the standard deviation at both the upper and lower level of the particular individual heart rate averages over time. Wide spread varying heart rate values, as a consequence of between or within- subject variability, will lead to a relative wide range of these indications, while a distinct, and consistent response to the event will lead to a restriction of this standard deviation band, as all numbers are more closely in range of each other.

4.8.3 Manual flight and synthetic speech for ATC

The results as obtained for a pilot in the simulation condition: Manual flight and synthetic speech for ATC message content are depicted in figure 9.

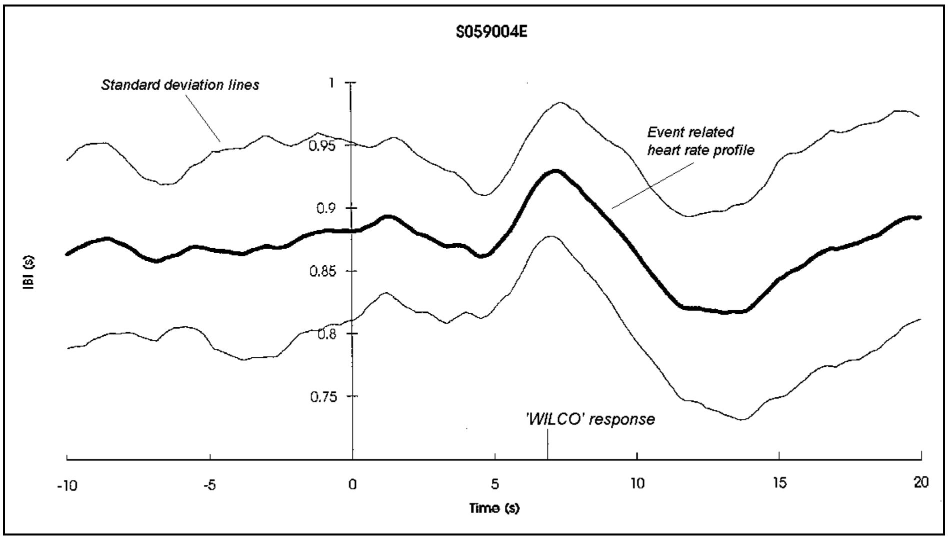

Fig. 9 Event related heart rate response, averaged over multiple events (n=26) for a pilot not flying while handling communications tasks with Air Traffic Control through data link. The centred event was an ATC up link (time zero) displayed by synthetic speech*, which required a 'WILCO' response later on. The occurrence of that response, as determined by independent measures of the time of selecting the 'WILCO' button, is indicated.*

The results show a distinct pattern of a linked response to this event as reflected in the averaged heart rate profile. Before the event, variations in heart rate were indeed averaged out, confirming uneventful conditions just prior to the triggering event for this analysis. It is, however, possible that other events occur just prior to selected events, depending on the separation in time between events in the scenario. The experiment manager can assure sufficient separation of events at will, and depending on the scenario.
After the occurrence of the up link, heart rate slows down shortly , a response known from laboratory study experience, to be often associated with information uptake, and it increases subsequently when the pilot is mentally working to absorb the information (note the plotting convention in inter beat intervals, meaning heart rate goes down when the intervals increase!).
The sequence is completed with a distinct decrease in heart rate as the 'WILCO' response is, according to the independently recorded, button activation's, selected and executed. The distinct decrease of the profile can be associated with the required concentration of the pilot but can be strengthened as a result of possible breath holding associated with such concentration. Similar event related averaging techniques can be applied to respiratory signals when required.

4.8.4 Manual flight and displayed text for ATC

The same pilot performed identical tasks under a different regime of experimental conditions. The ATC message could be spoken by means of synthetic speech like before, or could be displayed as text on the navigation display. In this case reading is required as opposed to listening to the message. The results for this case are depicted in figure 10.

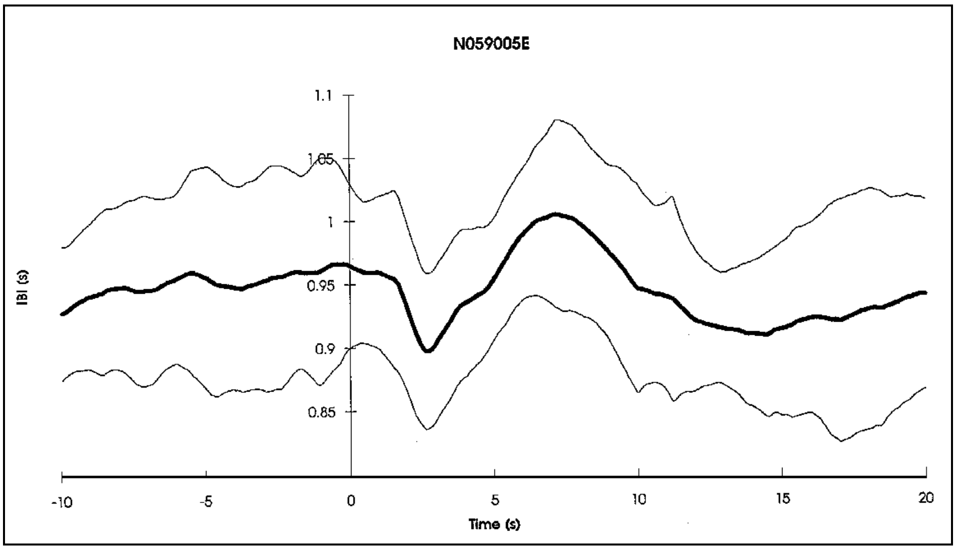

Fig. 10 Event related heart rate response, averaged over multiple events (n=26) for a pilot not flying for handling communications tasks with Air Traffic Control through data link. The centred event was an ATC up link (time zero) displayed as text *on the navigation display, which required a 'WILCO' response later on.*

In the case of reading from the display, a marked difference in response can be observed. After the alert, a short period of heart rate acceleration was observed, followed by a quite more distinct increase in heart rate as compared with the synthetic speech condition, apparently associated with ‘working though the text’. After absorbing the information again, heart rate decelerates markedly during the selection and execution of the ‘‘WILCO’’ response.

4.9 Response to Automation failures

To illustrate the response of a pilot to an automation failure in the case of FMS gating through advanced Data link, an ‘Auto pilot disconnect’ event was selected for this analysis.

These events naturally, have a low frequency of occurrence to prevent the pilots of being prepared for such failures. In the selected case, the auto pilot disconnect occurred directly after handling an up link from ATC as controlled by the Experiment manager facility.

The results obtained are depicted in figure 11 and demonstrate the occurrence of the apparently, typical ‘‘WILCO’’ response in heart rate also for this pilot, just before the surprising event of an auto pilot disconnect. The profile illustrates that even subsequent events with short intervals between them, can still cause distinct responses to be observed. Also, only three events in the average were sufficient to replicate the earlier results for the ‘WILCO’ obtained with 26 events in the earlier average (also different subject pilot).

The intuitive expectation of a pilots response to an auto pilot disconnect event would be a distinct rise in heart rate. The data however, shows a different, more particular response. The heart acceleration following the completion of the communication task is followed by a decrease in heart rate, again apparently associated with the intake of information to determine the problem. No ‘panic’ related response is observed. Note the effect of the event on the standard deviation of the averages obtained. It is almost reduced to zero at the precise moment the pilot takes corrective actions. After the event, heart rate seems to stabilise to more normal levels.

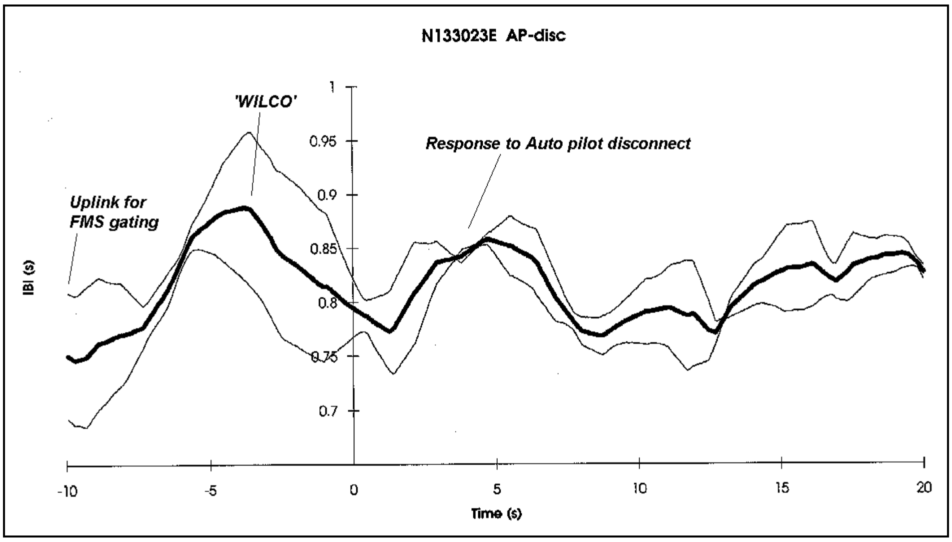

Fig. 11 Event related heart rate response, averaged over multiple events (n=3) for a pilot confronted with a auto pilot disconnect after accepting the FMS gating of an ATC up link. The disconnect was initiated automatically 10 seconds after the up links.

4.10 Discussion

The aim of this study was to demonstrate the feasibility of using Event related Heart Rate measures to complement and assist studies into human-machine interaction in automated cockpits. These cockpits are particularly suited for this application as the cockpit devices require many discrete tasks to be completed by the crew. The technical requirements for this technique are formidable with respect to the integration of both experimental control over simulation events and the associated data recordings with accurate synchronising provisions over multiple data sources. The application of the techniques could be demonstrated and resulted in the detection of distinctive heart rate responses to particular events scheduled in the scenario. In this respect, this laboratory technique could be transferred successfully to the complex environment of dynamic full mission simulation. With the EHR technique, a wealth of additional information seems to be accessible from complex simulations, improving the cost effectiveness of such studies.

4.11 Controller interactions with experimental software tools

Early attempts to introduce automated functions in ATC have resulted in mixed success. Electronic flight strips and automatic conflict resolution are examples that turmoiled both the research and ATC community for a long time.

Major issues were the reliability of computer advice and the incompatibility with human information processing strategies and working procedures. More effective and economic ATC performance is however, necessary for the future and poses a challenge to (human factors based) automation design. In collaboration with European partners (Jorna & Nijhuis 1996) and in collaboration with NASA Ames, the applicability of advanced software tools for ATM purposes was investigated. The development of such tools used to be driven by a ‘technology push’ often based on the assumption that tasks had to be removed from the controller in order to reduce the workload and increase the ‘reliability’of the provided services. However, without an effective Human- Machine Interface, no success can be attained and some developments could not be compatible with human characteristics from a cognitive engineering point of view.

4.12 Experimental design

The present study (Hilburn et al 1995, 1996) investigated the human use of possible tools in a future ATC scenario with present (low) and future (high) traffic loading. Arrival traffic approaching Schiphol was displayed on a Plan View Display (PVD).

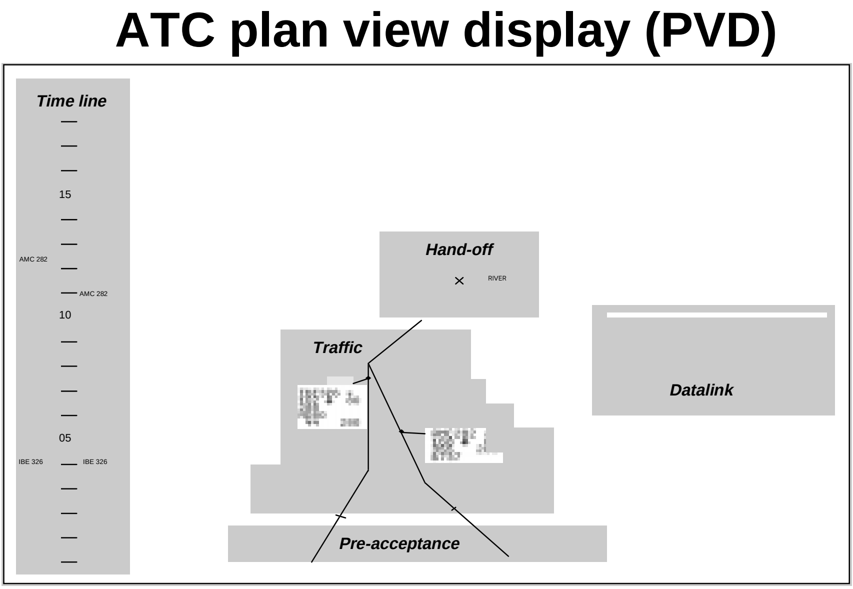

Fig. 12 Display lay out of the Plan View Display with Arrival scheduler at the left, aircraft hand-off area and data link communication status panel. The shaded area’s are used by the point of gaze equipment to provide area related data on eye scans, duration’s, transitions etc.

It contained :

- *Timeline window*– the controller must monitor this area for scheduling information, if he/she is to ensure that the arrival sequence is as desired, and that ETA (expected time of arrival) and STA (scheduled time of arrival) agree;
- *Traffic area*– the region of the screen in which controlled aircraft appear, including both the aircraft location plots, and the flight labels that display all relevant flight parameters;

- *Data link Status Panel*– displays all recently-uplinked messages, together with elapsed time since transmission, and whether the clearance has yet been acknowledged by the aircraft;
- *Hand-off region*– general area in which the PVD displays the plots and flight labels of aircraft around the time that they are handed off to Amsterdam approach (APP) control;
- *Pre-acceptance region*– the general PVD region that displays aircraft before they are accepted from the previous sector. Viewing this region provides the controller an indication of impending traffic load changes.

Three levels of assistance with the aim of reducing workload were available: conflict detection by the machine; conflict resolution by the machine and a ‘manual’ baseline condition representing full controller work.

4.13 Results

The controller was equipped with ‘point of gaze’ measurement facilities. The results of these measurements are depicted in figure 13.

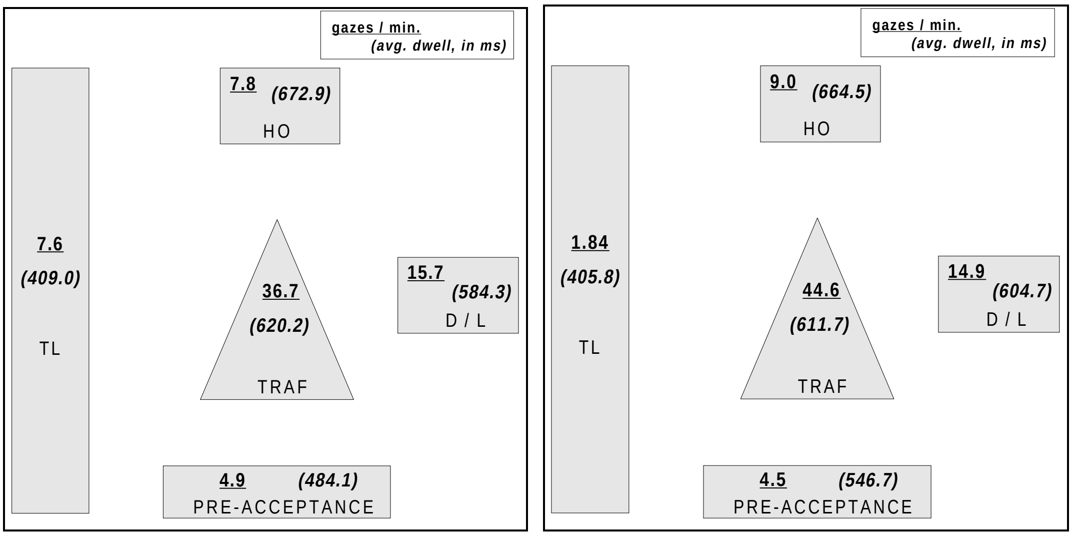

Fig. 13 Dwell time and fixation frequencies, for low (a) and high (b) traffic load

Comparing the results in figure 13 gives an indication of how traffic load influenced the visual scanning of the task-relevant surfaces. It appears that average dwell times were slightly influenced by differences in traffic load. Surfaces 1,2, and 4 (i.e., the timeline, traffic and hand-off regions) showed a decrease of 0.8% to 1.3% from low- to high-traffic conditions, whereas increased dwell times were seen for the data link (3.4%) and pre-acceptance (12.9%) regions.

Fixation frequency, however, seemed much more sensitive to the effects of traffic load. The net change from low to high traffic conditions ranged from -6.3% (for the pre-acceptance region) to -75.7% (for the timeline). The actual scanning could change dramatically as illustrated in figures 14 and 15.

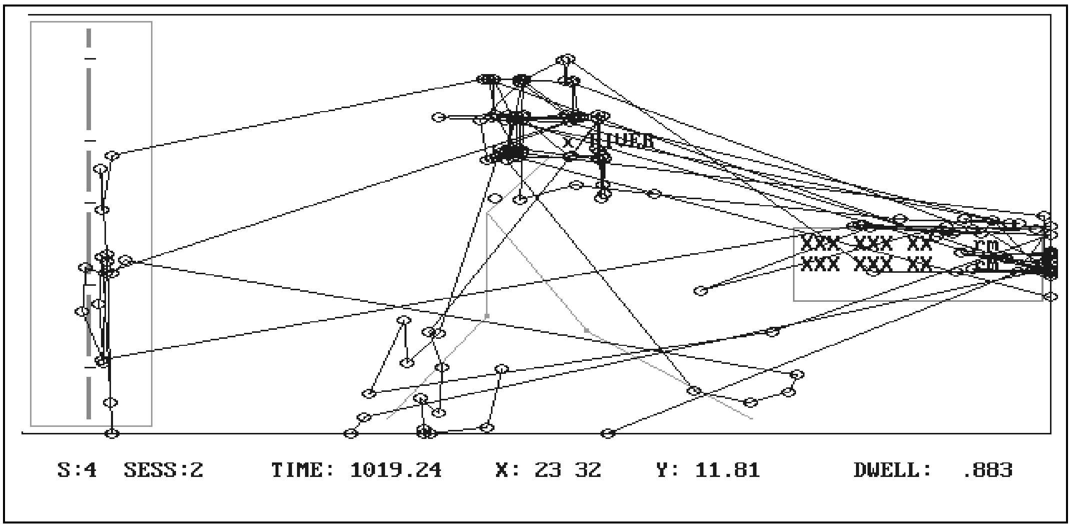

Fig. 14 Sample 120-second eye scan history trace, low traffic

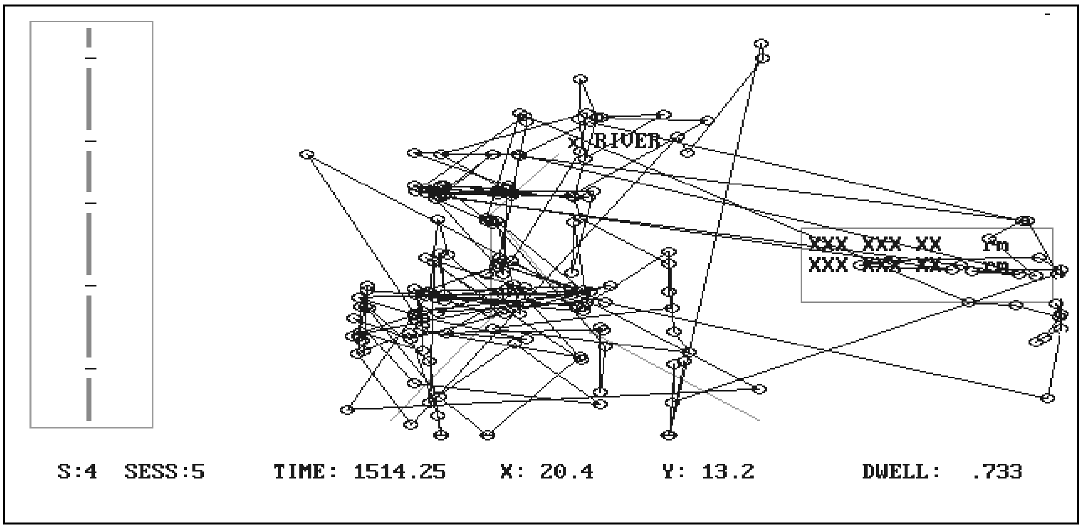

Fig. 15 Sample 120-second eye scan history trace, high traffic

4.14 Automation assistance and Workload

Increases in mental workload have not only been associated with measures of heart rate but also with pupillary dilation. This measure is also available from the 'point of gaze' equipment. Compared to either the light or focal length responses, however, the effect of mental activity on pupil diameter is quite small. Across subjects, pupil diameter was found to range from 2.72 mm to 7.20 mm, a surprisingly broad range given that the typical extremes of pupil diameter are believed to be two and eight millimetres. Pupil diameter was greater under high traffic load. Average diameters of 4.63 mm and 5.31 mm were obtained under low and high traffic, respectively. it is worth noting that this pupillary response occurred in tandem with an increase in the number of on-screen elements and, therefore, the number of active screen pixels. That is, pupil diameter was seen to increase significantly with traffic load, even though higher traffic load coincided, in this study, with higher levels of background illumination. Apparently the tools seem to reduce workload as indicated by this measure.

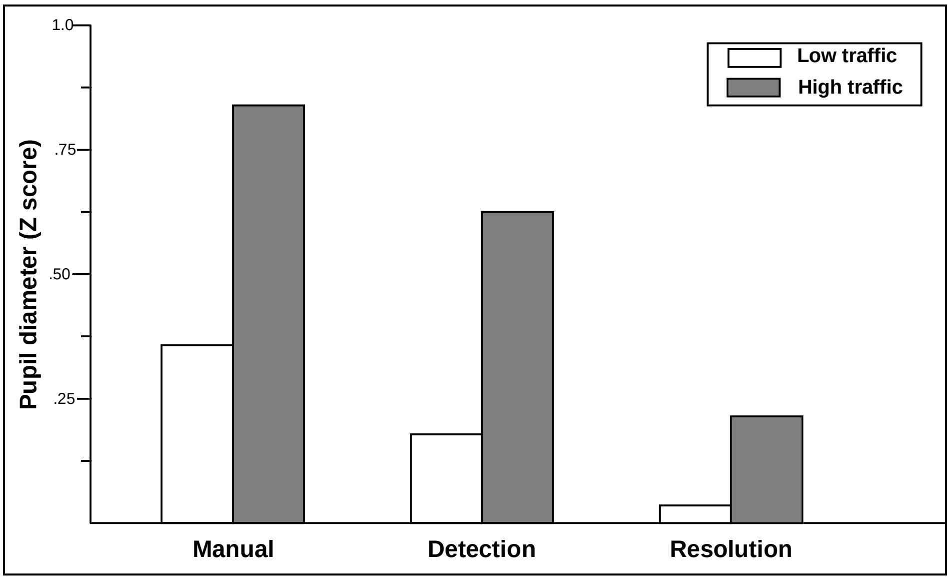

Fig. 16 Pupil diameter, by traffic load and automation condition

Heart rate variability is a measure often used for the assessment of mental workload (see Jorna 1992, for a review). The variations in heart rate rhythm tend to be reduced under mentally taxing conditions. The reduction in variability is plotted in figure 17, in an inverted value to serve as a more readily understandable indicator for the mental workload.

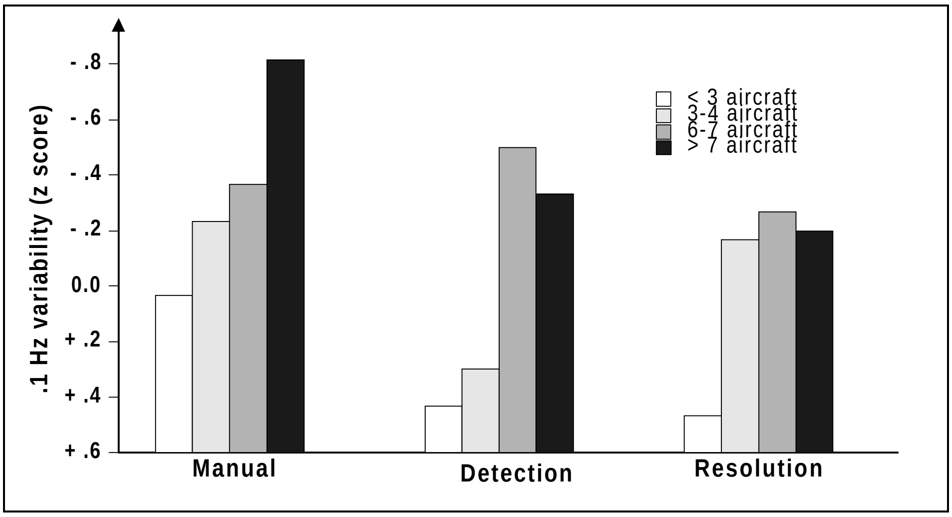

Fig. 17 Heart rate variability (.1 Hz), by traffic load and automation level (high values plotted to indicate high effort levels).

Clearly, HRV was sensitive to levels of traffic load. This effect appeared across all levels of automation. Further, HRV patterns across automation levels clearly suggest that, especially under high traffic conditions, mental workload was reduced by automation. This pattern is depicted in figure 17 for all three levels of automation, across the same four levels of traffic load.

4.15 Subjective ratings

Controllers' subjective ratings underscored the effectiveness of the traffic load manipulation in terms of perceived workload. The ratings provided regularly, show a trend contrary to the results of the more 'objective' data by indicating that workload is perceived as 'higher' when more automation assistance is added to the simulation. Other observations revealed that the controllers tend to revert to their old controlling methods under highly loading conditions. Training scenario's should emphasise the use of tools under these conditions more effectively to prevent 'high traffic panic'.

Objective and subjective performance and workload measures applied in such fine grained studies often dissociate. Performance levels increased and objective workload indices decreased when providing assistance from tools for conflict detection and resolution. Subjective ratings, however, indicated an increase representing a possible bias to *'more tools must be more workload'*. These observations underscore the importance of using objective measurements and not solely rely on 'users opinions'. These can be influenced by many factors, but practice and extensive familiarisation with advanced technologies seems crucial for user acceptance.

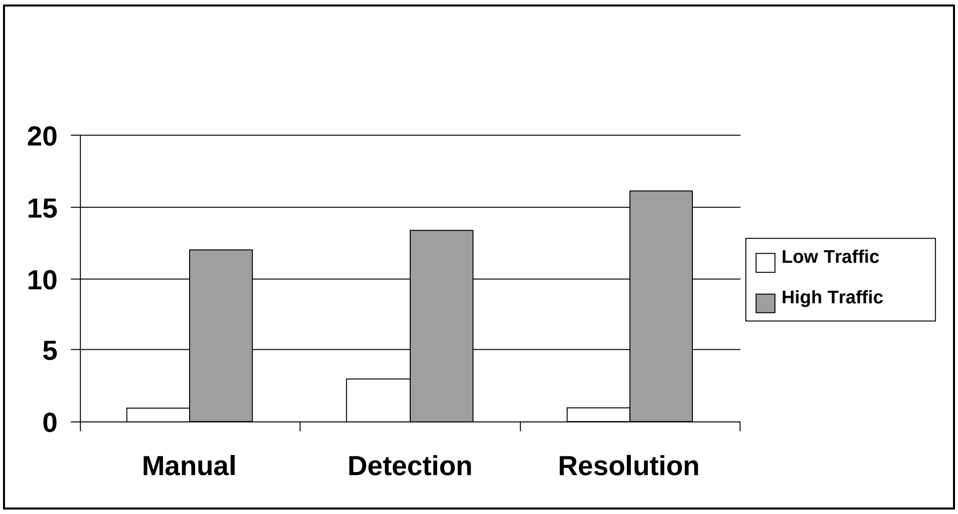

Fig. 18 Subjective ratings on influence of automation support on perceived work load

5. Challenging issues

The results of ‘context simulation’ studies provide valuable contributions to both the understanding of human behaviour in advanced systems and strategies employed for the re-design of Human -Machine Interfaces according to the lessons learned. Some of the major ones are:

- Expectations determine the main portion of information transfer and these expectations can only occur if the pre-operational context is adequately represented.
- Humans actively use and seek for information confirming their actions. The deviations or ‘incorrect’ actions by the machine are far less effectively detected and realised. Alerting should therefore understand what the context is in order to support the operator.
- Human machine interfaces will be used differently as anticipated in the design phase. Fine grained measurements are essential to understand why and what to do about it.
- Ineffective tools will be dropped immediately when time pressures increase. The opposite should be achieved by applying cognitive engineering principles.
- Humans are curious and will go for information irrespective of what the procedures prescribe.
- Subjective ratings are never sufficient for validating a design or procedure.
- Extensive training and familiarisation is essential for research into application and development of new technologies
- Context simulations and its associated methodologies proved to a powerful tool for studying and understanding operator behaviour.

The most challenging forthcoming issues seem to be associated not only with technology development and applications but also with issues concerning ‘training and transitioning’ from old to new technologies. A particular interesting challenge will be the realisation of so-called ‘Free Flight’ ATM scenario’s where the tasks are shared differently between the aircraft and ground control as today. Many opponents to such a concept, state that it will be impossible. Others argue that it is the only solution to the present congestion of the aviation system while assuring the principle of free enterprise for all businesses involved. Context simulations with the Human -Machine Co-operation ATM testbed as realised as a result of the studies discussed, seems to be the right way to go in order to provide more objective data on what can be done and what cannot be done.

6. 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. 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.
8. Jorna, P.G.A.M. (1993). The Human component of System validation. In: NATO ASI series: 'verification and Validation of Complex Systems: Human Factors Issues. Berlin Springer verlag.
9. Jorna P.G.A.M. (1996) Issues and Trends in Cockpit Automation: the challenge for enhanced Human performance. Second conference on Automation and Human Performance, Florida.
10. 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.
11. Jorna P.G.A.M. (1997). Pilot performance in automated cockpits: event related heart rate responses to datalink applications. In: proceedings of the ninth international conference on aviation psychology, Columbus, Ohio, USA.