FINAL TECHNICAL REPORT
VINTHEC II - WP8 - TR 01

CONTRACT N° : G4RD-CT-2000-00249

PROJECT N° : GRD1-1999-10543

ACRONYM : VINTHEC II

TITLE : Visual INTeraction and Human Effectiveness in the Cockpit, II

PROJECT CO-ORDINATOR : National Aerospace Laboratory NLR (NL)

PARTNERS :
BAE SYSTEMS (UK)
QinetiQ Ltd. (UK)
National Defence Research Establishment FOI (SE)
Linköping University (SE)
RISØ National Laboratory (DK)
Technion - Israel Institute of Technology (IL)
Institute of Biocybernetics and Biomedical Engineering (PL)

REPORTING PERIOD : FROM 1 April 2000 TO 31 July 2003

PROJECT START DATE : 5 April 2000 DURATION : 40 months

Date of issue of this report : July 2003

Project funded by the European Community under the 'Competitive and Sustainable Growth' Programme (1998-2002)

Document control page

VINTHEC II Technical Report

Document title	Full-scale simulator experiment
Code:	VINTHEC II - WP8 - TR 01
Workpackage:	8, Final report
Workpackage leader:	NLR (G.D.R. Zon)
Contributors:	BAE SYSTEMS (C. Neary) QinetiQ (M.R. Taylor) FOI (P. Berggren) LiU (M. Jacobsson) RISØ (H.H.K. Andersen) Technion (D. Gopher) IBIB (J.K. Ober)

Document Revision log:

Issue	Date	Pages	Remarks
0.1	25-7-03		Draft - Some partner input still missing, therefore final results incomplete.
1.0	22/01/04		Final

Signed for document release:

Workpackage Leader	Co-ordinator

Project funded by the European Community under the 'Competitive and Sustainable Growth' Programme (1998-2002). The following partners take part in the ***VINTHEC II*** consortium:
National Aerospace Laboratory NLR, BAE SYSTEMS, QinetiQ Ltd., National Defence Research Establishment FOI, Linköping University, RISØ National Laboratory, Technion – Israel Institute of Technology, Institute of Biocybernetics and Biomedical Engineering.

Distribution list

Organisation	Name
Commission of the European Communities DG-XII	J. Prieur (3 copies)
National Aerospace Laboratory NLR	G.D.R. Zon E.K.S. Hanson M. Joosse H. Bohnen P.G.A.M. Jorna
BAE SYSTEMS	C. Neary
QinetiQ Ltd.	M.R. Taylor
National Defence Research Establishment FOI	E.A.I. Svensson P. Berggren J. Alfredson
Linköping University	K.A. Ohlsson S.W.A. Dekker M. Jacobsson
RISØ National Laboratory	H.H.K. Andersen R. Pedersen
Technion - Israel Institute of Technology	A.J. Grunwald D. Gopher
Institute of Biocybernetics and Biomedical Engineering	J.K. Ober W. Gryncewicz J. Dylak D. Paleczka

Table of contents

Executive publishable summary

Human error remains a stubbornly persistent problem in aviation, and is implicated in some 70% of all commercial aviation accidents. In recent years, maturation of cognitive theories (concerning, for instance, team performance and Situational Awareness (SA) has been paralleled by a revolution in physiological measurement techniques. Together, such developments have enabled researchers to non-invasively study various aspects of non-overt human performance.

Under the Fourth Framework, the VINTHEC (Visual INTeraction and Human Effectiveness in the Cockpit) consortium developed an innovative methodology for assessing individual flightcrew SA using Eye Point-of-Gaze (EPOG). VINTHEC II is an extension of this project, with the overall objective of refining and extending the VINTHEC research to the multiple-crew scenario. VINTHEC II focuses on conducting a series of part-task experiments to further develop and validate the measurement methodology. The project has conducted a motion-based flight simulator experiment to demonstrate the final methodology. A computer cognitive modelling exercise was performed as a benchmark for the experimental results.

The VINTHEC II approach turned out to be successful. The measures that were new, compared to VINTHEC I, did measure aspects of sSA. The cognitive modelling work did deliver data that are a valuable supplement to human-in-the-loop simulation data.

The total measurement battery comprises:

1. Task analysis
2. Eye Point-of-Gaze (EPOG)
3. Peripheral Arterial Tone (PAT)
4. JAZZ
5. Subjective data
    - Questionnaires
    - Communication
    - (De)briefings / auto-confrontation (incl. video)
    - Rating scales
    - Rough Performance Indicator
6. Cognitive modelling

The immediate application of this work would be systems development and training for the civil flightdeck, in which crew co-ordination, SA and interaction with automation are increasingly recognised as critical to safety. By developing a universal, standardised and validated methodology for objectively assessing shared Situational Awareness (sSA) and human interaction in complex aeronautical settings, the project promises to benefit training development, display / flightdeck systems design, and certification. The added value of such an approach is that it will result in more effective and realistic assessment methods for evaluations of complex human-machine system interactions, thereby facilitating quicker (and more cost effective) system design and evaluation techniques. Potential other applications, which the project is explicitly targeting through subject matter input to both test scenario development and technology implementation / knowledge transfer, include airframe and avionics development, flight safety training, medicine, nuclear industry and maritime bridge operations.

VINTHEC II explicitly emphasises implementation and transfer of usable results to industry. Although the immediate focus of this implementation would be aviation (systems design and safety training communities and Human Factors certification), there are clear applications of the VINTHEC II methodology to other job domains. This is because the shared SA scenario is not unique to civil flightdeck operations. Various domains share fundamental similarities in terms of, for instance, display, teamwork, human performance, and training issues. Exporting the results of VINTHEC II to such domains would be expected to provide potential safety benefits, as well as costs savings (via improved system designs, and reduced training requirements).

1 Objectives of the project

In Annex 1 to the contract the consortium wrote that: VINTHEC II intends to demonstrate the following achievements:

1. Detailed assessment methods - a universal, standardised and validated methodology for objectively assessing shared situational awareness and human interaction in complex aeronautical settings, which will reinforce the European position with respect to training development, display/flightdeck systems design, and certification. The added value of such an approach is that it will result in more effective and realistic assessment methods for evaluations of complex human-machine system interactions. This will result in quicker (and more cost effective) system design and evaluation techniques, that can help ensure better iterative design efforts;
2. Procedural guidelines- for the practical implementation of the objective assessment methodology, including specification of operational scenarios for CRM training and system evaluation purposes;
3. Implementation recommendations - for applying the assessment methodology to work domains other than the flightdeck, which is believed capable of benefiting systems design and training efforts in any number of other industrial sectors;
4. Cognitive engineering recommendations - regarding how to incorporate objective assessment methodology into the system design process;
5. Improved aviation safety- a great number of aviation accidents and incidents involve human error as a contributing factor. To understand the source of this error, knowledge about the pilots' and crews' understanding of the situation is indispensable;
6. In particular, automation is becoming more competent, and assuming the role of an additional "crewmember" on the modern flightdeck. This situation underscores the need for techniques to assess shared situational awareness and co-ordination between crewmembers (including between crewmembers and automated flightdeck systems);
7. Better (cost effective) design processes – instead of proving that new designs are in fact an improvement nowadays major studies including numerous participants are needed. The foreseen crew SA model together with the efficient EPOG measurement methodology will ensure easier and smaller experiments to effectively evaluate new designs. This will facilitate quicker (and hence, more cost effective) iterative system design efforts.

2 Scientific and technical description of the results

The relationships between the workpackages are as explained in the workplan flowchart in **Figure 2.1**. WPs 1, 2 and 3 fed their output to WP4 in which scenarios were generated. The experiences from especially WPs 1 and 2, but also aspects of scenario design for the human in the loop experiments were first pre tested in WP5 and finally evaluated in WP7. While the scenarios from WP4 were also directly fed to WP6 (which cognitively modelled the same phenomena as were the focus of research in the human-in-the-loop simulations). Eventually all results and conclusions come together in the WP8 final report and the "Engineering design guidelines and recommendations".

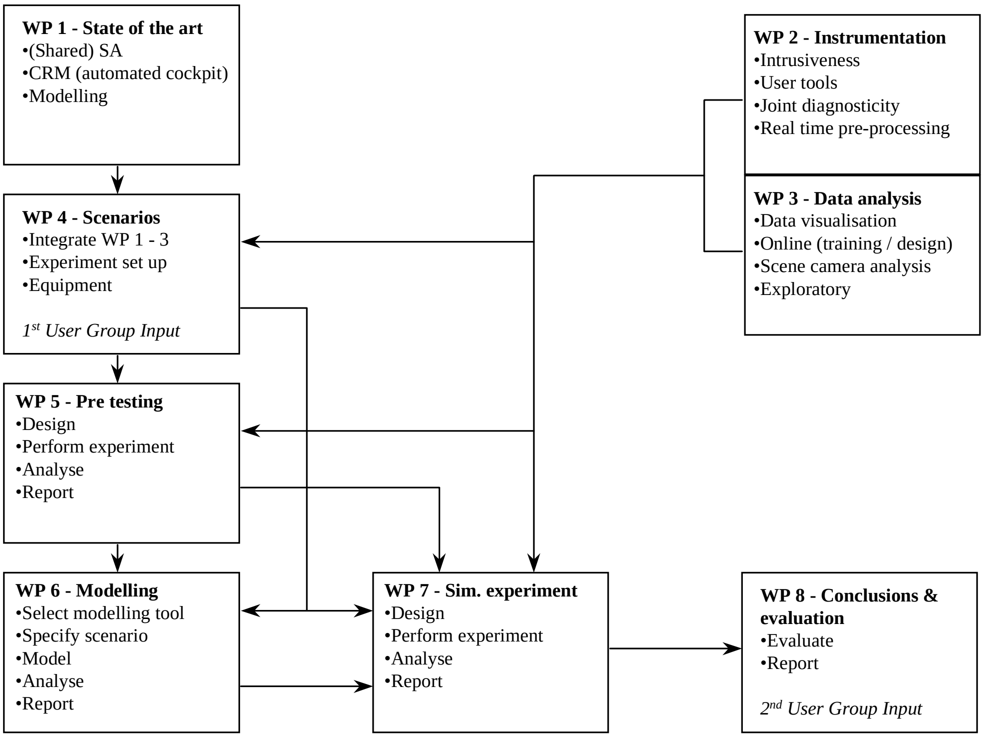


Figure 2.1. **Workplan flow chart with Work Packages (WPs) indicated.**

In the remainder of this Chapter the precise research, results and conclusions will be explained for each workpackage by the workpackage leader who was responsible for that workpackage.

2.1 WP1 State of the art

(WP leader: Linköping University)

Increasing automation on civil aircraft flight decks has produced great safety and economic benefits, but has also had unfortunate and unanticipated side effects. While crew complements have been reduced from three or four pilots/engineers to only two pilots over the past 2 decades, the ability of this reduced crew to build, maintain and share a coherent picture of what is going on (i.e. what the automation is doing with their aircraft) has not necessarily been enhanced. Characteristics of the design of automation interfaces make it both more necessary and more difficult to really form such shared situation awareness, or sSA:

- Each crew member has private access to the aircraft's computer systems through his or her own Flight Management System Control Display Unit or FMS-CDU.
- Primary Flight Displays (PFD's) have upset traditional scanning patterns: crews now have to ask themselves specific questions to know what to look for: scanning has become more model- than data-driven. This puts a high premium on the accuracy of the crew's mental model of the automation.
- Automation mode annunciations are limited to status abbreviations only: little future behaviour (especially with respect to the vertical flight path) is visible in the current representations.

VINTHEC II aims in part to experimentally test what goes into the build-up and maintenance, as well as the breakdown, of shared situation awareness (sSA) in commercial flight decks. The research has focused on using

Among the objectives for WP 1 were:

- Capture state-of-the art with respect to research into Shared Situation Awareness. SSA was defined as the extent of agreement between multiple crew members' continuously evolving assessment of the current state and trends of the process they jointly manage.

- Describe the current applicability of CRM (Crew Resource Management) training as a means of improving the attainability of sSA in automated cockpits. Assess the relevance of appropriate metrics for implementing CRM aspects in the VINTHEC II experimental scenarios

- Describe relevance of cognitive modelling to the problem at hand, which types of scenarios could profitably be managed and what techniques would be available or potentially available.

The report of WP 1 stresses how during empirical investigations of shared situation awareness:

- The phenomenon of interest must be defined, even if it is within the narrow context of study or application domain. We cannot rely on intuitive characterisations of shared situation awareness, because it leaves conclusions about its demonstration or manipulation unverifiable;
- Multiple research tactics must typically be used to produce converging evidence on the existence of, and ability to influence, a not-directly observable phenomenon. Complimentary research tactics discussed here included field observations and various forms of simulator studies.

- All of these tactics involve different kinds of trade-offs between sources of experimental uncertainty. Being explicit about the selected mix of experimental control can sponsor debate about alternative interpretations of results, rather than endowing the knowledge base with ad-hoc assertions that the phenomenon under investigation was demonstrated or influenced.

Difficulties in measurements that could validate the impact CRM courses actually have on pilot behaviour are a major concern. It is recommended that training captains be consulted to verify that certain techniques (such as eye-point-of-gaze (EPOG) tracking) could help in this validation. Two suggestions are made on how to identify the relationship between EPOG and CRM:

- Direct mapping of CRM behavioural markers onto specific EPOG points
- Comparison of good versus poor CRM behaviour

The modelling tool to be used for sSA should be able to answer questions about sSA imposed by, for example, particular cockpit configurations, training procedures, operational requirements or environmental conditions.

- Integrated modelling environments (such as MIDAS and IPME) seem to provide some practical benefits as they allow multiple applications relevant for VINTHEC II. However, MIDAS is still under development and is not yet fully documented.
- Nevertheless, to build a number of relatively simple stand-alone MIDAS, IPME and possible other applications, such as TOPAZ, could be a reasonable choice as:
- Numerical output from one model, using one tool, could feed into a second model, using another tool, and so forth.
- Different tools combine features such as: eye movement data assessment (possibly valuable pull through from VINTHEC-I in this area), cockpit systems and interface modelling with Performance Shaping Functions (PSF) for aircrew.

2.2 WP2 Instrumentation

(WP leader: RISØ)
A large portion of the first year's efforts was devoted to improving the methods by which Eye Point of Gaze (EPOG) data were collected, and analysed. The research consortium identified, on the basis of VINTHEC1 (and other related) experience, the need for improvements in the way EPOG data are collected and handled. This need was partly addressed through the work packages 2. Whereas WP2 focused on the instrumentation

issues (e.g. the more ergonomic and practical matters associated with collecting EPOG data), WP3 centred on techniques for reducing, presenting and visualising these data, so that they might be more useful for the research and operational communities. The two work packages are closely related and to some degree have common characteristics.

2.2.1 Task 2.1 Novel Methods

2.2.1.1 Objectives and hypothesis

Instrumentation work (WP2) involved developing and describing current and future/novel data measurement techniques for assessing Shared SA. The overall effort sought to demonstrate how objective (e.g. eye tracking), subjective (e.g. workload or SA self-assessment) and task performance data could be combined. It centred on identifying up-to-date methods and tools used to measure eye-point-of-gaze, describing those methods and tools, and providing guidance on their use for assessing situational awareness in a multi-crew environment.

WP2 of the VINTHEC II project consisted of 3 subtasks:

- Exploring novel methods for analysing shared situational awareness based on visual and other behavioural data.
- Development of tools to facilitate usability and efficiency of EPOG data collection.
- Exploration and documentation of EPOG data measurement techniques.

Risø have together with Technion, NLR and IBIB focused on developing and validating novel methods for the analysis of shared situational awareness. This work is presented in the VINTHEC II technical report 2-2 “Coding of and Inferences from Visual and Other Behavioural Data”. QinetiQ and BAE SYSTEMS have focused on the usability and ergonomic aspects of EPOG systems and exploration of data measurement techniques. The result of this work is presented in the VINTHEC II technical report 2-1 "EPOG Ergonomic Guidelines." NLR reviewed the deliverables together with FOI. Risoe acted as work package leader.

That VINTHEC II has been designed to collect physiological measures of information processing and response, along with the EPOG and performance measures. WP2 instrumentation development work sought to develop and define novel measurement techniques that can be applied to the assessment of Shared SA. This effort built directly on the State-of-the-Art review (WP1), which helped identify potentially promising applications. The more specific objectives of WP2 were [1] to evaluate the potential for integrated approach to SA assessment (i.e. via incorporation of EPOG and other objective measures); and [2] to conduct an ergonomic assessment of EPOG measurement techniques to help improve current measurement procedures; and to improve the usability of the techniques.

The major outcome of this part of WP2 is the prototype VINTHEC II methodology, later to be implemented in the main experiment in WP7. WP2 cross-feed data with WP3 including prototype visualisation of SSA measures. WP2 has an important link to scenario definitions (WP4), i.e. to make sure that the relevant SSA behaviour is present during the main experiment in WP7.

As the following sections detail, at a task level, WP2 addressed its stated objectives primarily through a series of part-task simulations and developmental studies.

2.2.1.2 Main results

The chief emphasis of WP2 was on the development, or refinement, of physiological assessment techniques, possibly as an adjunct to EPOG. One effort was work on Technion’s *Peripheral Arterial Tone* (PAT), measured at the fingertip. This is a clean measure of sympathetic activation, which is known to be highly sensitive to the exertion of mental effort. We believe that the addition of this measure to the existing measurement battery will improve the ability to trace changes in mental effort during the performance of flight missions. Technical steps were taken to adjust this measure for the requirements of the final Flight Simulation study at NLR.

An experiment was developed at Technion to employ the PAT. The experiment involves a single pilot performing a flight over an agricultural field. The pilot task is to keep the vehicle exactly above a vegetation lane, indicated by a field flag. During the flight, at random intervals of time, this field flag will randomly appear in a different lane. The pilot has to change his lateral position to the new lane. Since the random appearance of the flag introduces a sudden demand for pilot action, it will cause a sudden change in pilot activity and workload. Developments have also proceeded this year on development at Technion of part task team simulations in the Flight Control Laboratory.

IBIB has developed a device (JAZZ) that provides operators with real-time as well as historic data on crew members' attention mode. This might help eliminating certain actions that would otherwise disturb other operators' SA like untimely (distracting when in planning mode) verbal communication co-operator or prolonged planning mode without maintaining proper scanning of instruments. The device consists of a three-state indicator of other operators’ current attention mode, with the possibility to retrospectively display data of visual activity with planning, monitoring and exploration marked on the scale. The IBIB prototype is built upon the idea that switching between these mental modes can be recognised by a visual activity signal. According to this model, visual activity increases across the planning, monitoring and explorations modes respectively. It is the thought that this model could be utilised to give feedback to operators in terms of team-mates current mental mode and thereby support maintenance of shared situational awareness.

In a separate experiment (at the NLR) a low fidelity flight deck team scenario was investigated. This was a first evaluation of the applicability of EPOG to the multiple pilot scenario: Simultaneous EPOG readings were collected from both members of a two-man flight crew, during the performance of a series of flight-related task. Data streams were synchronised and merged, as a proof of concept of simultaneous measurement capability. Degree of team interaction (co-operation) was experimentally manipulated by task instructions. Workload (high versus low) was varied experimentally through simulator task parameters. The experiment demonstrated that such data streams could be integrated, and that the integrated data could be used to evaluate team interaction, in a meaningful way. These Results has been integrated with the WP2.2 report.

Much effort was devoted to small scale (e.g. part task) experiments conducted at Risoe. These experiments addressed both data collection (WP2, tasks 2.1 and 2.2) issues, as well as methods for analysing data (relevant to WP3). Given the interest in extending the applicability of the VINTHEC II techniques to domains beyond just the civil flight deck, these experiments were carried out in a variety of simulated settings. These include Risoe’s nuclear research reactor, the ATC simulator at Copenhagen airport, and the anaesthesiology simulator Herlev (DK) university. IBIB was instrumental in helping conduct and analyse these experiments.

With respect to data collection Risoe used two head- and eye-tracking systems for measuring subjects’ eye movements. In addition they audio- and video taped the teams to record their verbal and non-verbal communication activities. After each session the team used debriefing techniques to assess team situation awareness measures. In terms of data both non-verbal behaviour protocols, verbal protocols, interview transcriptions, EPOG data from the subjects and debriefing questionnaire results were all analysed. Final versions of the two WP2 deliverables are available. Finally, NLR visited University of Leuven (B) to receive demonstrations on developments in eye tracking technology, to exchange ideas about measurement theory and EPOG measurement methodology.

2.2.1.2.1 Measured proposed based on results from part-task experiments and developmental studies

Tracing the information gathering of subjects seems incapable by itself of indicating how subjects use the detected information, i.e. EPOG data cannot represent so-called higher cognitive processes without an analysis of a given context. For this analysis (non) verbal data are needed. It was suggested that these data be derived from observations and transcriptions of cockpit voice recordings augmented with video recordings of non-verbal communication during the WP7 experiment.

As a point of departure for the novel sSA methodology, it was proposed that a series of techniques for verbal data measurement could be used (see also Table 2.1). These comprise the transcripts of each crew-members speech, coded into pre-defined types. As such it was proposed that this coding could be analysed using three communication parameters: Number of speech acts (few/many), homogeneous communication patterns (similar/different), and temporal aspect of communication (past, present or future tense). In addition, it was proposed to measure non-verbal measures in terms of four main interaction categories or modes of interaction: Maintaining reciprocal awareness, directing attention, assigning tasks and handing over responsibility of processes in the field of work.

Table 2.1 Verbal data measurements

	Dimensions of communication		
	Number of speech acts	Homogeneous communication patterns	Temporal aspect of communication
Types of speech acts			
Command			
Observation (about flight or system status)			
Inquiry or request for information			
Response uncertainty			
Agreement			
Acknowledgement			
Repetitions (of already stated commands or inquiries).			
Social			
Directing attention			

Regarding and EPOG data, in task 2.1 and as part of a novel sSA methodology it was proposed to measure sSA in terms of whether the pilots applied a distributed or a focused attention strategy. This measure share resemblance with the in the eye-tracking methodology discussed notion of “scanning entropy.” Moreover, it was proposed to measure sSA by way of lag sequential analysis looking for “hidden” patterns in the data. Furthermore it was proposed to use sequential analysis to study of the temporal structure of sequences of events. As such it was proposed to use lag sequential analysis, which allows for calculating frequencies of transitions between dwells on AOI’s within a certain lag in a time series. Lag sequential analysis allows for answering questions like: "How many times is the a dwell on the primary flight display followed by dwell on co-pilots hands?" Finally it was proposed to us a debriefing questionnaire containing a short battery of questions (ATSA: Assessment of Team Situation Awareness) designed to elicit a subjects' estimates of his or her team-mate's situation awareness and view of task allocation in addition to the subjects' own first-order knowledge of significant system parameters and their trends. The ATSA form may be used by interrupting a simulation session and eliciting subjects’ estimates, continuing the session and then conducting a repeat interruption and elicitation of awareness judgements (like SAGAT is used). It was suggested that it instead should be applied during post-session de-briefings. When comparing responses from crews it is possible to gauge:

- The accuracy of individual crew members’ estimates of system parameters (and, by extension, the agreement between crew members).
- The ability of crew-members to correctly predict the awareness of their fellow crew-member.
- The extent to which crew members may correctly predict the workload of their colleague and their colleague’s perception of task allocation.

2.2.2 Task 2.2 Ergonomics

2.2.2.1 Objectives and hypothesis

Before a course to train novice users of EPOG equipment can be developed a clear understanding of the important subject variables is required. It may not always be possible or desirable to screen subjects, especially if the EPOG methodology is to be used satisfactorily for pilot training. Guidelines (or solutions to these kind of equipment related problems) are required so that EPOG recordings can be optimised and if necessary made on any subject regardless of their characteristics.

Usability tests to systematically investigate the ease of use of a helmet-mounted video based EPOG system (ASL 4000 series) were therefore carried out by BAE SYSTEMS at the BAE SYSTEMS Advanced Technology Centre - Sowerby. The usability tests were performed in two parts:

The first part was performed in conjunction with an experiment planned for the DIVA (Design of Human/Machine Interfaces and their Validation in Aeronautics, Brite-EuRam III) project. The aim of the DIVA project is to develop a new methodology that systematically uses state of the art human factors techniques for design and evaluation of the human-machine interface (HMI). The DIVA experiment was designed to evaluate new HMI integrated in the fixed-base Human Factors Research (HFR) Simulator at Sowerby. As part of the evaluation, a number of crews flew various flight scenarios in the HFR simulator while the EPOG of one crewmember (the pilot in the left-hand seat) was recorded. The timing and location of this experiment gave the VINTHEC 2 project the opportunity to assess, using a group of pilots, a number of factors related to the ease of use of the EPOG equipment. The EPOG usability testing used biographical data, operator and user questionnaires, video recordings and EPOG data quality. In particular, helmet comfort/intrusiveness, ease of calibration, individual differences including body (e.g., height, weight, leg and arm length, and head circumference) and eye (e.g., eye colour, eyelash length, and width and height of eye aperture) dimensions together with measures of EPOG data quality were examined.

The second part of the usability study evaluated these same measures (helmet comfort/intrusiveness, ease of calibration, body and eye characteristics, EPOG data quality) to assess the use of the EPOG system in a group of non-pilots. These subjects viewed targets presented on a number of displays within the HFR simulator. The effect of head and arm motion on the EPOG data quality was also examined.

In addition to empirical testing, this task 2.3 also relied on QinetiQ analytic review of currently available EPOG measurement systems and techniques. This review was forwarded for inclusion in the D2.1 deliverable.

The Risoe/IBIB small scale studies, as mentioned above in section 2.1, also contributed to this task 2.2 effort, in that they also addressed the ergonomic (e.g. how should data be collected? What are the issues related to subject intrusiveness and test reliability, etc?). Finally, NLR conducted small-scale experiments into integration of scene camera and EPOG trace views that will allow for overlay of the two visual streams. This will be invaluable for real-time data quality assurance, as well as qualitative assessments and real-time analysis.

2.2.2.1.1 Main results

A comparison of the two parts of this study indicates that the non-pilots were slightly more difficult and took longer to calibrate, were less willing to wear the helmet for any length of time, and found the equipment more intrusive. The non-pilots also tended to be slightly taller and heavier with longer lashes and smaller heads.

- In general, the pilots and non-pilots gave similar results:
- The closeness of fit of the helmet-mounted eye tracker is a key factor in determining measure reliability and amount of data loss

- Larger head size (presumably because of tighter helmet fit) will reduce risk of data loss.
- The perception of helmet comfort may be a good indicator of likely data loss
- Eye colour has little effect on calibration time or EPOG data quality, although aperture height (for non-pilots) can be related to data loss.

An operator experienced with EPOG equipment set-up, calibrated and collected the EPOG data. Even so, one of the pilots (after touching the visor) produced very poor quality data and the degree of data loss overall was quite high. The calibration time was kept to a minimum and re-calibration was not attempted (due to the time constraints of the DIVA experiment). It may be that frequent checks on calibration accuracy (and re-calibration) during for, example, a training session may be necessary to ensure good quality data.

It should be remembered that these results were found using one type of eye-head tracker using one type of helmet. It may be that if the system was removed completely (remote) from the head many of the difficulties in usability described in this report may be resolved. On the other hand, a remote eye-head tracker may introduce different problems, which would also need to be considered.

2.2.3 Task 2.3 EPOG Usability Testing

2.2.3.1 Objectives and hypothesis

These guidelines will take the form of procedural guidelines- for the practical implementation of the objective assessment methodology, including specification of operational scenarios for CRM training and system evaluation purposes.

This task explored and catalogued the ease of use of EPOG methods and techniques focusing on a review of eye-point-of-gaze (EPOG) relevant techniques and a review of human factors techniques relevant to measuring eye-point-of-gaze (EPOG). The work was mainly carried out under the auspices of QinetiQ and the purpose was as follows:

- To identify up-to-date methods and tools used to measure eye-point-of-gaze;
- To describe those methods and tools; and
- To provide guidance on their use for assessing situational awareness in a multi-crew environment.

2.2.3.2 Main results

The analysis of the different EPOG systems resulted in recommendations in terms of guidelines / checklists that can be used by for operators / instructors and other interested parties that want to implement VINTHEC II SSA methodologies in for example CRM training courses. To describe the tools in which methods are incorporated, a simple tabular format is used for the presentation of the guidelines Each table describes a tool in the following terms:

- The title of the tool;
- A general description of the tool;
- The broad procedure for using the tool (e.g. inputs and outputs);
- An outline of the advantages and disadvantages;
- Any potential use in investigating shared situational awareness.

The main results are presented in Table 2.2

Table 2.2 shows the main results from analysing the usability of several different EPOG system (only the advantages - including prospects for measuring SSA - and disadvantages are included)

EPOG systems	Advantages	Disadvantages	Potential use in SSA
Applied Sciences Laboratory Series 4000 Eye hardware	The system is a lightweight head-mounted system and provides good spatial resolution.	Data recording of the head-tracker can be affected by metal objects near the system. It is a relatively expensive	A possible, if expensive choice.

		system and has some limits on temporal resolution.	
EYENAL software programme	The system is easy to use after some training has been given.	The system has some software problems; there is some loss of data with eye blinking, and detailed training is required to be able to use it.	If the hardware were used, this would be the appropriate software package to accompany it
GazeTracker II software	A comprehensive system.	The equipment is intrusive. There is a risk of losing data from accidental movement of the equipment. Extensive training is required to be able to use it	Possible, but not ideal.
ISCAN miniature eye imaging system	The system is relatively inexpensive.	Without head position sensing there is no absolute direction of gaze available. Though Incorporating the head tracker (ISCAN Headhunter and magnetic tracker) overcomes the weakness of the system	Acceptable, if able to work with scene camera view, or if head sensing is also used.
ISCAN Headhunter and magnetic tracker	This system provides real time data and is a non-invasive system.	There appear to be no major shortcomings.	This appears to be a viable solution, provided the system is easy to set up.
ISCAN eye track systems	These systems are easy to set up and calibrate and are non-intrusive.	The systems risk some loss of data on large eye movements	This may be a usable system, but without a head tracker there would be no absolute EPOG data available.
Advanced System IR eye-tracker	A sophisticated, high-speed system.	The system appears to measure eye movement rather than eye point of gaze	Too complex for VINTHEC requirements.
Cyclop Multisensor IR eye-tracker	Speed and precision.	The system appears to measure eye movement rather than eye point of gaze. An updated version is designed to generate EPOG data. Such a system is not yet available commercially, but is under development	Overly complex.
Fourward Optical Technologies Inc.	The system can cope with a large visual field, is accurate, reliable,	The system is expensive and does not meet the VNTHEC II requirement for using a head-mounted	Not appropriate.

	and easy to use.	system.	
EYEGAZE (LC Technologies)	The system can operable in light and dark conditions. No attachments required on the subject so there is no obvious intrusion.	High levels of infrared light are detrimental to data collection; this suggests that bright, glare conditions of a cockpit may cause difficulties. The system relies on an unobstructed view of the subject's eye for the camera. The Vinthec II requirement for a head mounted system makes EYEGAZE unsuitable to EPOG	Not appropriate.
Microguide Inc systems	These systems are small and lightweight, incur minimal intrusion to the subject's visual field, are sensitive to small eye movements, and are relatively inexpensive.	They have a limited range of measurement and do not meet the requirements of pupil/corneal reflection methods under Vinthec II	Not appropriate.
Permobil Meditech's ober2 system	This system overcomes the problems of blinking in data collection.	Technical training required to use the system and no head tracker is incorporated into the system.	Not suitable.
SMI SensoMotoric's iView system	The full visual field is maintained for the subject. iView is an accurate system and can be configured with other systems for wider EPOG measurement and analysis.	Data recording of the head-tracker can be affected by metal objects near the system. It is a medium expensive system and has some limits on temporal resolution.	The system's ability to integrate into other systems may make it suitable for assessing shared situation awareness through EPOG assessment of multi-crew environments.
Skalar Medical's Electromagnetic Scleral Search Coil Technique	The system has a high level of accuracy.	The system does not incorporate a head tracker or a pupil/corneal method of measurement. It is invasive.	Not appropriate.
Skalar Medical's Infrared Light	The system is accurate, easy to use, requires little	The system is uncomfortable to wear and has no head tracker	Not appropriate.

Reflecting Eye Tracking System	training and provides good temporal and spatial resolution.	incorporated. It is limited to one plane and no pupil/corneal method is incorporated	
Point of Regard Tracker (PORT)	Accuracy reported as ~1o and relatively insensitive to camera shift.	The range of eye excursion is limited by the CR and some subjects will be unable to wear the equipment (e.g. because they wear spectacles or some contact lenses). This system can require expensive, bespoke processing equipment	Not appropriate.

2.2.4 Deliverables, WP2

Two reports are ready from WP2:

D2.1 EPOG Ergonomics Guidelines

D2.2 Coding of and Inferences from Visual and Other Behavioural Data

2.3 WP3 Data analysis methodology

(WP leader: QinetiQ Ltd.)

2.3.1.1 WP3 in Context

Work package three (WP3) was an extensive programme of work comprising a number of complementary activities. In many respects this WP could be regarded as the transition point in the VINTHEC II programme where theoretical and practical knowledge was taken and turned into the first stages of a user-centred methodology.

The theoretical understanding of SSA (from WP1) and the practicalities of the available EPOG equipment specifications and ergonomic guidelines (from WP2) were interpreted in WP3 in the context of user requirements. The outputs from this WP were then used subsequently in WP4 for defining the scenarios, and in WP 6 and 7 in experimental testing and data analysis. In addition, the practicalities of the methodology, required by users and addressed in this WP, were subsequently assessed in a workshop with the WP7 pilots who participated in the study.

2.3.1.2 WP3 Objectives

The top-level objective for WP3 was to identify new and enhanced ways to analyse and visualise EPOG data customised with respect to the critical SA and CRM skills, adapted for the information needs of the end users.

To address this objective a number of sub-tasks were identified:

Statement of User Needs (SUN)

The first sub-task was to identify users for the VINTHEC II methodology and to investigate their needs and obtain a Statement of User Needs (SUN). The SUN defines the requirements for the SSA measurement technique, and is necessary to guide the research toward production of a useful evaluation tool. The proposed users for the VINTHEC II derived SSA evaluation method were principally trainers, training captains and flight deck designers, with (probably) a fairly basic understanding of SA, and with limited resources to devote to the analysis of data derived from training sessions. In addition, these user groups were expected to have different evaluation requirements. The SUN sought to define the constraints that each group of users was likely to have in the use of an SSA evaluation tool. LiU performed activities in this sub-task.

Knowledge Elicitation

The second sub-task, related to the first, was to seek to identify critical skills that are required for good SSA and critical behaviours that would be indicative of the level of SSA in a team. This work complemented activities in this WP to investigate novel data analysis methods, by providing behavioural markers on which it might be possible to assess SSA. This was felt to be particularly important, as an evaluation based purely on behavioural observation would be prohibitively expensive, time consuming, and unreliable when used by non-experts. Also, it was thought that it might be possible to incorporate simple behavioural measures within an evaluation technique or, indeed, correlate novel objective measurement techniques with behaviours indicative of SSA. Finally, the identification of critical SSA behaviours in the civilian aircraft cockpit fed into the identification of appropriate experimental evaluation scenarios later in the programme. QinetiQ performed activities in this sub-task.

Exploratory Data Analysis and Feasibility Study

The third sub-task was a substantial piece of work to explore existing data gathering techniques to identify new ways in which these data might be usefully analysed and to assess the feasibility of using eye movements (rather than EPOG) to investigate SSA. Emerging techniques suggested that eye movement analysis, rather than EPOG alone, may hold vital clues as to the mental state of the individual. Alternatively, the fusion of EPOG measures with other available data streams, such as heart rate variability (HRV), verbal communications, or individual actions, may result in a more coherent measure of SSA. This sub-task investigated these areas, and activities were performed by RISØ and IBIB.

Data Visualisation

The fourth sub-task brought elements of the other sub-tasks in this WP together to develop a flexible data analysis software tool. This tool, named the DAta VIsualisation Tool (DAVIT), converts EPOG data from a variety of available systems into a standard time-based format, and combines this with data streams from a large number of other sources, and then allows manipulation and analysis in a wide variety of ways. The modular design meant that it was possible to build into the software the novel data analysis techniques that were identified, whilst the SUN provided a user requirements capture for the software, allowing design of the most appropriate interface and tools. It was hoped that the user-centred approach to the design of an EPOG methodology for the analysis of SSA, this would result in a tool that was both usable and useful. Technion led the data visualisation tool development, with programming input in WP3 from QinetiQ.

2.3.1.3 Role of contributing partners

The activities undertaken in WP3 were co-ordinated and led by QinetiQ, and six consortium member groups contributed to this large WP, as noted in the relevant sub-tasks above. In addition, the output from these sub-tasks was reviewed by FOI. Subsequent evaluations of DAVIT were undertaken by the whole consortium membership. Development of DAVIT and the IBIB novel measurement systems continued beyond WP3 and these were refined for future use in VINTHEC II.

2.3.2 Review of sub-task activities

2.3.2.1 Introduction

In this section the sub-task activities will be briefly described. The approaches taken and the outcomes of each sub-task will be discussed with regard to the objectives of the WP. More extensive details of the activities in this WP may be found in the two formal report documents and the DAVIT user manual produced as deliverables and detailed at the end of the WP3 section.

2.3.2.2 Statement of user needs

Approach: The (SUN) was prepared in discussion with various potential user groups. Structured discussions were used for this study, and the basic premises and ideas of the VINTHEC II end product were explained. Participants were invited to comment on the aspects that they deemed relevant to their practice and possible improvements to it. Discussions with practitioners were held at various locations over a number of months.

The groups of interest to the VINTHEC II project were thought to be those most closely involved in aircrew training before, or during, operations. It is in these situations that appropriate scan patterns and co-ordinating behaviours matter most, and can be studied and modified realistically.

The user groups included in this statement of user needs were:

Trainers (Flight and Simulator Instructors at Airline Flight Academy);
Training captains (European Commercial Airline);
Aircraft Designers (Saab Technologies).

Discussions included (but were not limited to) consideration of the points in pilot training at which the VINTHEC II end products could be best applied for greatest benefit; practicality of EPOG measures and the requirements of any proposed system; data analysis and data reduction needs; visualisation and accessibility requirements; and ethical issues regarding the collection and use of data.

Outcomes: Most of the instructors who were approached expressed interest in a new piece of training support hardware; as they would with any new piece of hardware that affected their work in a cockpit. It was felt that EPOG as a training aid would appear to be both useful and worthwhile.

The perceived benefits of EPOG were felt to be, primarily, for the following:

- the possibility for deeper insight into crew co-ordination breakdowns or other problems that may shape subsequent training;
- the availability of data traces of EPOG for evaluation during debriefings, which could be seen as a useful tool to support debriefings.

It was felt that the application of EPOG methodology might best be used in the recently mandated MCC (Multi Crew Co-operation) courses, taught at Airline Flight Academies and other Flight Training Organisations (FTOs). This training phase is meant to close the gap between ab initio training and initial airline training, and is used to emphasise the development of non-technical skills, such as leadership, team building, crew communication and co-ordination, applicable to working in a multi-crew environment.

The biggest limiting factors to the utility of a system for training needs were the time taken to set up the equipment, and the time taken to process the data. Time for briefing, simulator sessions, and de-brief was already well utilised, so there was little additional time available for fitting and calibration of EPOG systems or for any analysis of output. Preference was for a ‘switch on and go’ data collection mechanism with instant, non-expert, data analysis. The ability to use EPOG data video output for debriefing shortly after simulator sessions was popular, but the requirement for quick access to flight episodes and having data in a format that was usable immediately after a training session was felt to be important, both to maximise training benefit and to ensure that the methodology was used.

Finally, it was felt that the use of an EPOG measure would require adherence to a pre-defined ethical code, and data confidentiality would be an entry requirement for the use of an EPOG-based training system. Issues of litigation and use of video data of individuals being used for professional and organisational purposes were discussed, and felt to be worthy of careful consideration.

In conclusion, and in the context of the WP3 objective, the SUN study was felt to be successful and of use in guiding the programme towards production of a useful evaluation tool. The practicalities and constraints likely to be encountered were identified and a series of user requirements was generated, against which the consortium could seek to aspire with the methodology.

2.3.2.3 Knowledge elicitation

The aim of the work in this sub-task was to identify variables or behaviours that may be indicative of good and bad SSA, in an attempt to describe measures of SSA. Some of these variables were identified by the ‘state of the art’ report (WP1), and are based on empirical research and theoretical perspectives. However, it was regarded at important to elicit this information directly from experts and the potential users themselves.

Approach: The bulk of the technical work in this sub-task centred on a two-day workshop held with 40 pilots in September 2000. The majority of these pilots were from British Airways and had recently been flying Concorde but, as this aircraft was grounded at the time, the opportunity to gather data from a significant number of aircrew was taken. All pilots who took part in the workshop had either recent or current experience of glass-cockpit aircraft. The pilots were interviewed collectively (using a system of individual, but linked, laptop computers) and they were also given questionnaires to take away with them and complete after the workshop. Data from this workshop was used to support two EU programmes, VINTHEC II and ESSAI, and served as an opportunity for researchers from the two programmes to share ideas and approaches to SA, SSA, and CRM issues. For the purposes of VINTHEC II the questions set to the pilots fell into three main categories, related to obtaining and losing SSA, EPOG as an evaluation method for SSA, and cockpit design in relation to SSA. These complemented the broader VINTHEC II requirement to address questions of theory, measurement, and scenario design in relation to SSA. To elicit unconstrained, candid and valuable responses, it was necessary to ask open, unquantifiable questions, which yielded information-rich responses.

Outcomes: Thirty-two pilots completed the questionnaires and returned them within a month of the workshop. As expected, information-rich data were collected and this was reported in detail in the WP3-TR 3.1 Report. As might be expected, the critical stages of flight during which EPOG was felt to be most revealing (in terms of SSA) related to high workload periods such as take off and landing. Pilots felt that procedures related to the FMS, routing, and altitude awareness were of particular importance and further discussions revealed that SOPs on certain tasks dictated cross checking, but these procedures often became habitual and, though pilots always direct their eyes in the appropriate way, it was possible not to perceive or comprehend the information.

The observation above posed a fundamental problem with EPOG systems, i.e. that even though it was possible to measure what was being looked at, it was not a trivial task to measure what the individual comprehended. Furthermore concerns were raised that it might be difficult to evaluate habitual behaviours in a simulator that, by its nature, is a non-habitual environment. However, even with these concerns, it was noted that it would be possible to measure SSA through other means.

Responses to questions regarding how SSA was gained (briefings, audible communications, cross-checking, watching other crew members, and empathising) indicated the methods by which knowledge was shared; and these grouped nicely under the recognised SA levels of perception (communications and cross-checks), comprehension (empathy) and projection (briefings). Factors that influenced the losing of SSA were mostly poor communications and high workload.

The influence of flightdeck design and SSA were very mixed. Highly electromechanical display environments (such as Concorde) were felt both to promoted good SA and hinder good SA. The higher levels of workload in these cockpits was felt to lead to more active information processing, which in turn was felt to promote good SA. However, others felt that the high level of information processing led to a lack of spare resources, and overall less time to concentrate on new events or information. This, and specifics of cockpit display layouts, were felt to hinder cross-checking and create teamwork and SSA problems.

In the glass cockpit environment, which generally was viewed as being poor for SA, the FMS was perceived as being a source of error. In terms of shared knowledge and SSA, glass cockpits appeared to be better. The FMS was cited as being a useful source of shared information, and the ability of the pilots to cross-monitor each other's displays is much easier than in older flight decks.

In terms of using EPOG as a measure of SSA, the pilots were not, in general, EPOG experts. However, a number of eye movement behaviours were identified. Eye contact, particularly during briefings, was regarded as an important SSA behaviour, though presumably eye contact on the flight deck is limited. Perhaps more interesting was the more general comment that 'body language' was important. As well as relating to general non-verbal communication, observations from the exploratory data analysis suggested that this would also include observing hand movements and other actions. This serves as a way for one pilot to judge the tasks that the other is performing.

Overall, it was felt that the two best indicators of levels of SSA were:

Body Language – During briefings eye contact was important. During flight, actions and hand movements allowed one crew member to follow the behaviours of the other; and

Cross-checking – of power setting; throttle movement; frequency selections on radio; correct pressurisation during climb/descent; current and next headings of the aircraft; wind readout from/to way-points; charts. This would reduce under high workload, and is likely to be flight phase-dependent.

In conclusion, and with respect to the WP3 objectives, this piece of work identified a number of key skills and behaviours that were felt to reflect good SA and good SSA. The behaviours could be readily identified and measured using EPOG techniques and therefore were felt to strengthen the role of this methodology for evaluation of SSA. However, there were issues raised which questioned too simplified an approach to EPOG measures, and although these are understood and acknowledged by the experts in the VINTHEC II programme, they could be overlooked by novices or those searching for a 'quick-fix'. Overall, it was felt that understanding the differences between the SA of each crewmember, and how levels of SA interact, was key to the evaluation of SSA.

2.3.2.4 Data visualisation

Approach: From the SUN, it was apparent that a vital component of an EPOG SSA analysis methodology was an automated data analysis process. This would allow non-experts to use the software quickly, reliably, and inexpensively. As existing off-the-shelf EPOG analysis software packages were regarded as neither flexible nor tailored to the evaluation of SA, there was a clear requirement in WP3 to begin the development of a bespoke EPOG tool. The requirements of this tool were that it should be flexible, standardised, and easy to use, and must allow data input from a number of streams, and allow many different analyses. The goals within WP3 for the visualisation tool were for it to be designed with the following:

- a user-friendly graphical-interactive method - to allow potential users easy access and retrieval of information embedded in EPOG data records;

- a standardised EPOG data recording format and software - for translating the formats of the different systems/users into a standardised format; and

- a standardised graphical interface - for presenting EPOG data and for enhancing co-operation and communication between users.

Outcomes: Over the period of WP3 a beta version of the software was developed. The data specification tool, developed by QinetiQ, took raw EPOG data and combined this with data from many other sources, including audio and video data, and placed the data in a common format that could then be read by the data visualisation tool. The second module, developed by Technion, took the data streams and the display specifications defined in the first module and provided the front end by which the desired data manipulations could be made. Once data were specified, they would not be passed through the data specification tool again, and the files could be archived or passed to other users to allow identical recall at a later date. The visualisation tool was based on a modular system to allow for further development and refinement. A schematic of this version of the tool is shown in Figure 2.2 below to illustrate how the two modules work together. A detailed explanation of this version of the data visualisation tool was included in the WP3 TR-01 Report.

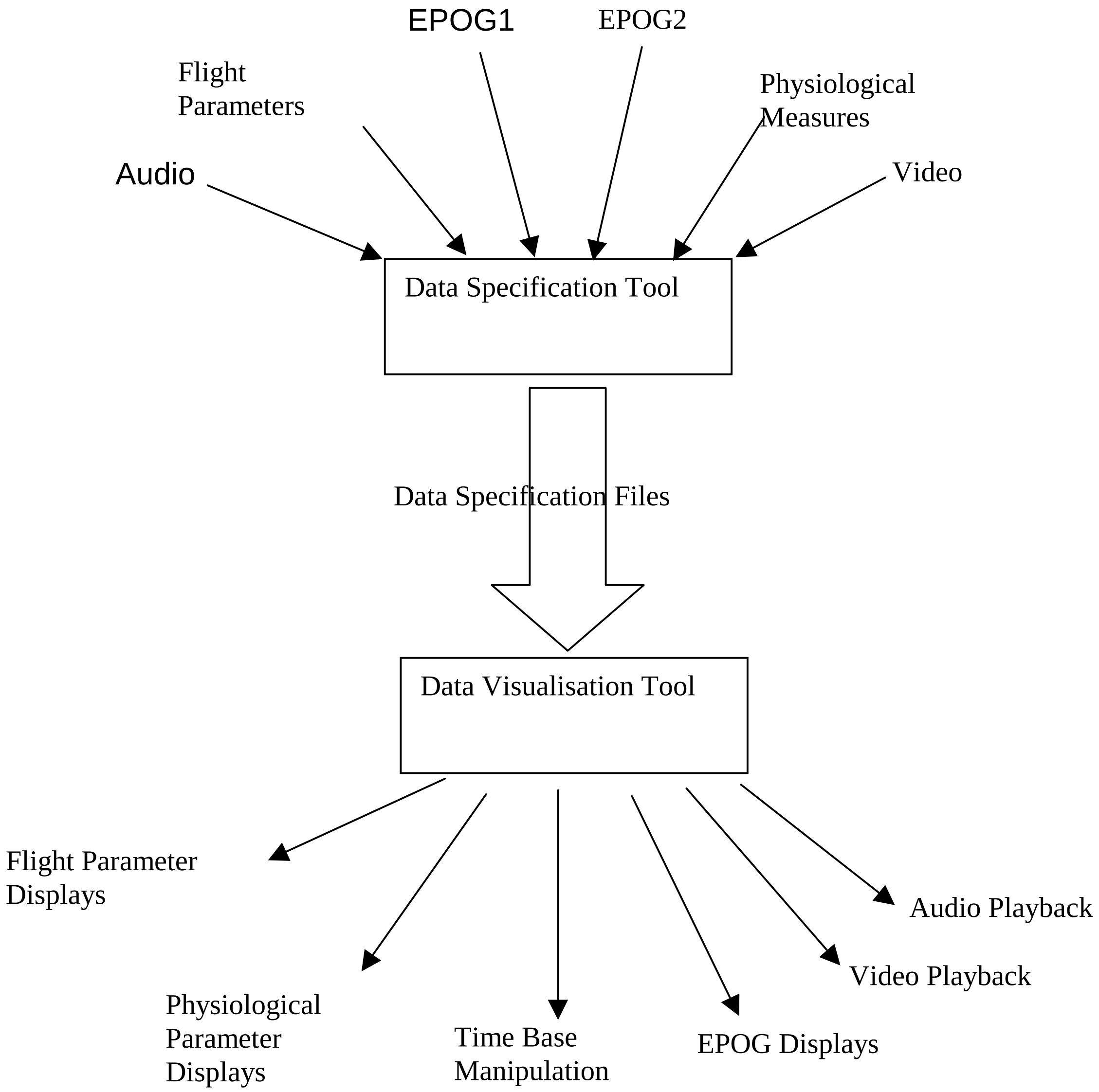


Figure 2.2: Modular Design of DAVIT the sSA Data Analysis Tool.

Following the reporting of WP3 the data visualisation tool was further developed by Technion, with input from the other VINTHEC II consortium members. At the close of the VINTHEC II programme (July 2003) the latest version of the tool, now named DAVIT (DAta VIsualisation Tool), is Version 4.0a. This version has been described in a manual by A.J. Grunwald (2003) and a brief overview of the latest version, taken from this document, is provided below.

DAVIT: Introduction and Description

The activity relating to DAVIT is the definition and development of a comprehensive method and its software package for visualising and analysing complex human response data in flight simulator experiments.

The software package, developed for processing and visualising the data consists of three programs:

the *'Grab'* tool, which extracts a synchronising sequence from digitised video recordings,
the *'Prepare'* tool, which prepares and synchronises the data for use by the visualisation tool, and
the *EPOG visualization* tool, which visualises the EPOG and human response data. This tool also allows the user to select and prepare data for further statistical processing.

The tool serves as a user-friendly, universal and intuitive environment, in which EPOG data of one or more crewmembers, flight data, event data, audio and video recordings, and human response data such as the Peripheral Arterial Tone (PAT) or Jazz, can be read from files, processed, visualized and effectively analysed, with emphasis on team situational awareness. The main purpose of the tool is to give the user fast and easy access to the data. Although some statistical processing is included, the tool is not intended for complex statistical analysis. However, the tool enables the user to readily select cross sections of the dataset, and to export this cross section in a format, which matches spreadsheets or statistical analysis programs such as SAS, Excel or Matlab.

The complete visualisation software includes three stages of operation: *A 'preparation' stage*, a *'playback stage'*, and an off-line *'statistical processing' stage*.

In the *'preparation' stage* the tool is prepared for processing the particular data set of the user. This is carried out by the 'Prepare' tool. The preparation stage deals with reading, formatting, synchronising, and preparing the data for visualisation. Output of this 'Prepare' tool is a set of intermediate binary files and ASCII configuration files, which are used by the visualisation tool. For each experiment, a set of these intermediate files is created. The user can simply start visualising the data of a particular experiment, by loading its intermediate files.

The visualisation tool carries out the *'playback stage'*. It allows a selected time frame of the data set to be visualised in actual or slow motion. The playback stage includes visualisation of the time histories of the simulation variables, the EPOG data, the video and audio signals and human response data. By manipulating a 'time-line', the user is able to review any part of the complete data set in detail and in particular, to investigate how the EPOG patterns of the pilots correlate with the simulator data, video, audio or significant events. The playback can be at normal speed, accelerated speed, slow speed or frame-by-frame, forward or backwards. It also includes an option to play back a selected time span.

The visualisation tool carries out the *off-line statistical processing stage* as well. The processing is limited to very basic scores in a specified time frame, such as the statistics of EPOG transitions between surfaces, the number of fixations, average fixation duration, surface dwell time, average blink frequency, and average blink duration. However, this stage also enables the user to select and export a cross section of the dataset to spread sheets or statistical analysis programs.

Overview of Data Visualisation: The activities in this area, both during the lifetime of WP3 and since, have led to the development of an increasingly useful visualisation tool and a truly exploitable output from the VINTHEC II programme. The complexities of data synchronisation have been non-trivial and it has been necessary to work on a number of iterations of the software to get to the version used in the WP7 data analysis. In the context of the WP3 objectives this area of technical development has been extremely satisfactory – providing a common approach to the use of EPOG and other data for users – both researcher in the consortium and, potentially, other less expert users in future.

2.3.3 Output from WP3

2.3.3.1.1 Feed-through to later workpackages

As can be readily appreciated from the foregoing details of WP3 activity, the outputs from WP3 were key to a number of the issues and activities in later workpackages. The user-centred approaches to the statement of user needs and the knowledge elicitation had direct applicability to the generation of scenarios and evaluation of behavioural markers for use in WP4, 6, and 7. In addition, the possibilities for data visualisation made real by the beta version of the data visualisation software in WP3 enabled the software tool to be further developed and made fit for purpose during the later workpackages.

2.3.3.2 Deliverables

Formal Deliverables for WP3 were:

1. Technical Report: User Needs, Knowledge Elicitation, and Software Tools. VINTHEC II-WP3-TR 3.1.
2. Technical Report: Feasibility Study of IBIB Online Prototype and Visualisation and Exploratory Data Analysis. VINTHEC II-WP3-TR3.0.

In addition, a third deliverable related to data visualisation was produced by Technion:

> DAVIT – DAta VIsualisation Tool: Description of a method and its software package for visualising and analysing complex human response data in flight simulator experiments. User Manual for Version 4.0a.

The activities conducted within WP3, or begun during this time, have formed much of the basis for exploitable products within the VINTHEC II programme. Both the initial development of JAZZ and DAVIT began in this workpackage and the exploratory work on data analysis and visualisation software and the user-based inputs from the other activities in WP3 have informed their development and the design of the methodology used in the main simulation study in WP7.

2.3.4 Review of WP3

2.3.4.1 Measures of success: Were all objectives met?

During the foregoing discussion of WP3 activities the success of each sub-item of work as been judged against the overall objective for WP3 – which was to identify new and enhanced ways to analyse and visualise EPOG data customised with respect to the critical SA and CRM skills, adapted for the information needs of the end users.

Overall, the success of WP3 depends on the perceived success of the workpackages that followed – in particular the main simulation study in WP7. Viewing each sub-item activity in isolation, each successfully fulfilled its objectives, although in most cases further work could have usefully followed. As always, the real benefits are to be gained when the parts come together (in this case making up the complete methodology) proposed as the final output for VINTHEC II. For the main exploitable areas, i.e. IBIB's JAZZ system and Technion's DAVIT this further development took place during the later stages of the programme. However, the user requirements/needs addressed in the SUN would ideally be re-evaluated more fully as a consequence of the outcome of WP7 and then further, when the methodology is in the hands of the non-VINTHEC users. Some evaluation was usefully pursued through the workshop for WP7 participants (held in November 2002), but would, ideally, be an iterative process with a wider scope of users.

2.4 WP4 Scenario definition

(WP leader: FOI)

2.4.1 Objective

The objective of Work package 4 was to generate operational scenarios or experimental flights in which practicable (i.e., reliable and non-intrusive) psychological and psychophysiological measures are integrated. The selection of measures was based upon the results from the work packages performed earlier and the aim of VINTHEC II.

Work package 4 describes the experiment facilities, equipment, and setup used in the experiment. Together with subject matter experts (instructors) from SAS Flight Academy and TFHS, Lund University School of Aviation, Sweden, operational scenarios were developed with varying degrees of difficulty. The subject matter experts were instructed to consider events in a flight that demanded interaction between the pilots.

2.4.2 Results

The work package 4 report consisted of two parts. One part integrating the theoretical aspects and practical experiences concerning shared SA, CRM, and modelling from WP1, practicable psychophysiological (ECG, Blink rate, including EPOG-measures) and psychological measures (Shared SA, Workload, Performance) from WP2, and statistical procedures and visualisation techniques from WP3.

The second part of the report describes how the scenarios were developed, what the scenarios looked like, and what limitations there were due to the NLR RFS (Research Flight Simulator). The integration of theoretical aspects and practical experiences were input as instructions when meeting with the SMEs for development of operational scenarios. The SMEs were operational pilots and flight instructors from SAS Flight Academy and TFHS, Lund University School of Aviation, Sweden. They were all very experienced, both as pilots, but also as instructors.

It was stressed that it was important the scenarios triggered difficulties with respect to shared SA, CRM, and automation and also that the measures could be obtained without loss of realism. The scenarios were divided into several phases from start to landing. The phases differed in complexity and mental load to mirror the variance of real flights. They also tried to reflect relevant crew SA behaviour.

2.4.2.1 Summary of theoretical legalisation

A thorough summary regarding psychological and psychophysiological measurement techniques of team situation awareness was performed. Communication, measures, visual attention, non-verbal interaction measures, eye point of gaze measures, and debriefing questionnaires were reviewed. Communication and its relation to team performance was discussed in WP2. The reviewed research shows a connection between different types of utterances and team performance. It was therefore recommended that the scenario was able to provoke different kind of speech acts. When it comes to visual attention, it might be an indicator of what attention mode an operator is in; exploration, planning, or monitoring. In order to measure these modes the scenario would demand that kind of activities. Non-verbal interaction measures are embedded in team work. Thus, it is necessary to incorporate an element of team work into the scenario. In terms of eye point of gaze measures it was suggested that the dwell should be focused upon. That is, fixations within an area of interest (AOI). It is suggested that the overlap of dwells on AOIs for both pilots could be used as a Team Situation Awareness measure. Therefore, AOIs should be defined a priori, as a part of the experimental design. When it comes to debriefing questionnaires a battery of questions were designed to elicit a subject’s estimates of his or her team member’s situation awareness and view of task allocation in addition to the subject’s own first-order knowledge of significant system parameters and their trends.

2.4.2.2 Pre-testing study

The pre-test study equipment and experimental setup were described as they were to be used in WP5. The pre-testing runs were to test some ideas and concepts generated in earlier work packages in an operational setting. In particular, the JAZZ system in combination with EPOG equipment was to be tested.

The scenarios that were used in the pre-testing study were not the same as was to be used in the experimental study. That was partly due to the limitations of the pre-testing simulator limitations. Scenarios that were chosen for the pre-testing study did not include emergency procedures since emergency procedures are what pilots usually come to the simulator to practice.

2.4.2.3 Summary of the scenario

A scenario with two flight legs was developed in cooperation with SMEs. The originally developed scenario was then adapted to fit within the NLR RFS. The scenario was normal according to company specific procedures. It consisted of a Fokker 100 flight from Amsterdam (EHAM) to London (EGLL) and back.

A thorough task analysis was performed with the purpose to generate events that demanded cooperation within the cockpit, concerning both hard and easy tasks in order to get varied and diverse scenarios. The task analysis was performed in close collaboration with the SMEs. The SMEs were asked to come up with with events that were plausible, valid, realistic, and not the everyday training situations that pilots are exposed to at flight simulator facilities. The cooperation with the SMEs was considered successful, resulting in a good experimental scenario. It was also regarded as useful as a modelling scenario.

After the original scenario had been adapted for the NLR RFS it consisted of three parts:

- One training session so that the pilots could freshen up their Fokker 100 skills and get acquainted with each other and the simulator.
- One experimental flight leg, going from Amsterdam to London, including some experimental probe events.
- One experimental flight leg, going back from London to Amsterdam, including some experimental probe events.

2.4.2.4 Experimental procedure

Each experimental day involved one crew. They arrived at the facility and were briefed about what was the experiment was about. Then the training session commenced. After that, the psychophysiological equipment was applied to the pilots and calibrated. They then manned the simulator. Experimental flight one (Amsterdam to London) set off. Each experimental flight leg took approximately 45 minutes. After landing it was time for lunch, so all measurement equipment was removed from the pilots. After lunch, the same procedure as before was repeated, psychophysiological equipment was applied and calibrated. Next experimental flight two (London to Amsterdam) begun. When the equipment had been removed after the flight was finished, the debriefing started.

2.4.2.5 Simulator

A thorough description of the Research Flight Simulator (RFS) at NLR was included in the work package. It is portrayed with limitations and possibilities, and also with a detailed description from a technical point of view, among other things concerning time synchronisation, motion system, hardware setup, etc.

2.4.2.6 Scenario for modelling purposes

The scenario that was developed originally was intended to be used both for simulation and for modelling purposes. This was made clear to the SMEs as a part of the instructions that they received. Later, the scenario was adjusted to better fit the different aspects of a modelling scenario and a simulation scenario. Since the scenario, and the aspects brought up in the scenario, were already adapted to both simulation and modelling purposes, the later adaptation was minor. The formalisation of ideas from conceptual thoughts into a scenario is an important step towards a formal model used in modelling. In a wider sense then, the scenario description is part of the modelling effort. As such, the scenarios developed in WP4 aided both simulation and modelling efforts, performed later.

For the modelling scenario the original scenario developed with the SMEs was used.

2.4.3 Description of the role of contributing partners

NLR – Simulation equipment, adaptation of scenarios and measures to simulation facilities

Risoe – Psychological and psychophysiological measurement techniques of Shared SA and PMWL

BAe – Experimental setup, adaptation of measures to scenarios

LiU – Description of Shared SA, CRM, modelling, and automation, review

FOI (WP Leader) – Co-ordination, integration of contributions from partners, performance measures, usability analyses, report on performed work

2.5 WP5 Pre testing

(WP leader: BAE SYSTEMS)

The main objectives of WP5 were to pre-test concepts and ideas generated in earlier workpackages in an operational setting and to indicate problem areas prior to the full-scale simulator experiment planned in WP7.
The scope of the Pre-testing was defined in collaboration with the other WP5 participants (FOI, IBIB, NLR, and Technion) and comprised the following five sub-tasks:

1. Pre-test scenarios and methodology
2. Pre-test PAT system
3. Pre-test JAZZ system
4. Pre-test Data Specification and Visualization tools
5. Review measurement approach

BAE SYSTEMS (WP5 leader) had responsibility for the selection of the measures to be tested, the integration of the necessary equipment in the BAE SYSTEMS simulator, the running of the trials, the scenario development, and reporting. Technion had responsibility for the integration and testing of the PAT system. They also further developed the EPOG Visualization Tool using EPOG data recordings from the Pre-testing trials. IBIB had responsibility for the integration and testing of the JAZZ system and as part of WP5 developed the JAZZ software and analysis methodology. The main responsibility of FOI in WP5 was to pre-test the Data Specification and Data Visualization Tools. While the role of NLR was to ensure that problems encountered in WP5 were used as “lessons learned” to the benefit of the WP7 simulator experiments to be carried out later in Amsterdam. They also co-operated with consultant’s TestUsability in carrying out a review of the measurement approach. Furthermore, FOI, NLR and Technion all contributed to the experimental design and the scenario development.

2.5.1 Pre-test scenarios and methodology

The aim of this part of WP5 was to test scenarios, equipment, measurement systems and general procedures in pilot runs in the BAE SYSTEMS fixed base simulator in Bristol, UK. A detailed description of the experimental set-up was given in the WP4 report (see VINTHEC II-WP4-TR 01 ‘”Scenario Definition”).

Operational scenarios which triggered difficulties with respect to sSA, and CRM without loss of realism were developed after a joint team from BAE SYSTEMS, NLR and Technion met in Bristol with two senior Air 2000 pilots, for a detailed evaluation and discussion of prospective flight scenarios. The BAE SYSTEMS team prepared several alternative scenarios based on the earlier work that was done in WP4 by the LiU and FOI teams. The detailed review and discussion highlighted a number of major points, which were clarified and demonstrated in the context of the specific segments of the prepared flight scenarios. The main emphasis was on a better understanding of decision alternatives, and on possible variation, that would increase team collaboration and co-ordination. With these aims in mind, and given the practical limitations of the system used, three scenarios for the Pre-testing were designed.

Scenario A: Dual Hydraulic Failure (Medium Awareness/Major Event)
Scenario A, known as the ‘medium awareness/major event condition’, included a ‘major’ failure (Yellow system begins to depressurise 20 minutes into the flight) which the pilots would have practiced many times before in simulator training and which should result in the pilots following familiar standard procedures. This failure required the crew to land as soon as possible. The scenario was intended to last no more than 30 minutes. The crew was not required to land the aircraft.

Scenario B: Fuel Temperature High (Low Awareness/Minor Event)
Scenario B, known as the ‘low awareness/minor event condition’, included a ‘minor’ failure (a gradual increase in fuel temperature over a 20 minute period shortly after the aircraft has taken off) and events (distracting ATC calls including deviating from the planned course in order to avoid nearby “traffic”), which the pilots would have been less likely to practice during simulator training. Thus, Scenario B was predicted

to demonstrate elements of crew interaction/coordination perhaps hidden when crews are tested on highly trained events.

The event was a minor problem and once noticed required the crew to begin the “to do list” for fuel temp high. At about 15 minutes into the flight an Advisory Page was automatically displayed on the lower ECAM display warning the crew that the temperature was high. If the increase in temperature was not noticed (or resolved) the Master Caution sounded at 21 minutes into the flight, with the potential for diversion. The pilots were not required to land the aircraft

There were two specific points in the Fuel Temperature High scenario when the crew might become aware of a problem with the fuel temperature. One occurs when an Advisory Page is displayed on the lower ECAM display, the other when the Master Caution sounds 6 minutes later. It was anticipated that a ‘good crew’ with high situational awareness would pick up the change on the ECAM display before the Master Caution sounds and that the EPOG pattern around about this time would reflect good crew performance.

Scenario C: Fuel Temperature High (High Awareness/Minor Event)
Scenario C, known as the ‘high awareness/minor event condition’, was a repetition of Scenario B at the end of the Pre-testing session. The reasoning behind this re-run was that the crew would now be familiar with the ‘minor event’ and it was predicted that they would therefore fly the scenario with high situational awareness.

2.5.1.1 Experimental design and procedures

The facility used was the Human Factors Research Simulator at BAE SYSTEMS, Advanced Technology Centre in Bristol, UK. The simulator itself was designed as a generic glass cockpit, solely for research purposes, and bears a partial similarity to the Airbus A320 flightdeck layout. The subjects were eight male civil airline pilots from the Air2000 airline. At the start of a session, each pilot completed a Pre-flight Questionnaire. The questionnaire assessed flight experience, and flight roles, together with questions on whether the pilots had worked with one another recently as a team and their attitude towards teamwork. After familiarisation of the simulator, the three recorded scenarios were flown, in A, B, C order.

Before a recorded scenario, the pilot flying (PF) was equipped with the helmet-mounted ASL (Applied Science Laboratories) Series 5000 Eye/Head tracker (EPOG1), while the pilot non-flying (PNF) was equipped with the headband mounted ASL 5000 Series system (EPOG2), the integrated JAZZ device and the PAT system. (Note: crew 4 only wore the PAT device.) A miniature colour scene camera was also mounted on the head to capture each pilot’s forward view. When ready, the equipment (EPOG1, EPOG2, JAZZ and PAT) was calibrated. The crew was then given a short briefing on the flight to be performed. Briefings varied according to the scenario to be recorded, although the pilot was not told which scenario would apply.

Two simulator-mounted video cameras including microphones were located above and to the side of each pilot to capture: communication between the pilots and with ATC; and video records of the trials including pilots’ interactions, some instruments and hands.

At the end of each scenario, the crew remained seated within the simulator while the EPOG calibrations were checked and while they completed the Post Flight Questionnaire. This questionnaire assessed SA (CARS), workload (NASA TLX), team behaviour and equipment comfort/intrusiveness.

One member of the Technion team had responsibility for the PAT (set-up, recording, analysis etc), while the BAE SYSTEMS team had responsibility for all other functions during the trials.

A Training Captain from the Airline Air2000 assessed the crew performance from an analysis of the video-tapes a number of weeks after the trials were completed. He was instructed to rate the crews overall (baseline) team behaviour using the first 20 minutes of recording in Scenario A. He was also asked to rate situational awareness, performance and workload for the crew, the PF and the PNF for all three scenarios making his judgement from observations of the tapes made around the time of the failures. The twelve unmarked video-tapes (one for each scenario) were assessed in random order.

2.5.1.2 Data Synchronisation Method

The Scenario Phase was encoded as a number and sent simultaneously to digital input ports on the JAZZ system, the PAT system and both eye/head trackers. Phase transitions therefore synchronised these data sets with each other. Phase One was synchronised manually with a short countdown to one of the pilots pressing the Master Caution Reset, which was logged in the simulator data; this synchronised simulator state and switch operation with the phase.

The computer recording the JAZZ data and the digitised video also recorded a subset of the simulator data, which allowed synchronisation of these data sets with the main simulator data and thus the phase, through comparison of the file system time stamps on the files.

The video-tape was not directly synchronised. However, the countdown and the button press at the start of Phase One appeared on the tape.

The computer recording the JAZZ and video data created the Synchronisation file. A second computer recorded the Event, Continuous & Switch State files. The timestamps in these three files corresponded with each other but not with those in the synchronisation file from the other computer.

2.5.1.3 Lessons learned

2.5.1.3.1 Pre-testing dataset

Data structure
The continuously recorded aircraft time histories provided an accurate image of the vehicle motions and flap and throttle settings. Since the pilots mainly flew automatically, control stick and rudder signals were not recorded. The dataset showed the autopilot response resulting from a change in heading, flight level, or autothrottle commands. Inconsistencies in vehicle motions that might result from system failures would also be revealed.

The events including start-of-run event were on the same time base as the simulator data, although the set does not specify the action itself. Only one failure event occurred, so the resulting system and pilot response could be easily reviewed by considering the time histories and discrete actions at the time of the event. This might very well be done manually. However, in the NLR full-scale simulations numerous events would occur, like ‘serious’ events such as a primary flight display failure to more ‘routine’ events such as a cabin crew briefing. The WP5 pre-flight testing clearly indicates that an efficient automated method for reviewing a large number of events would be an absolute necessity in analysing the NLR dataset.

The EPOG systems operate in parallel to the simulator on a stand-alone base. This systems output highly processed data such as fixations and fixation patterns. The EPOG Visualization Tool, under development in the framework of WP5 and to be used in the full-scale simulator experiments, would not utilise this feature, since it would compute these fixation scores within its own environment. The EPOG recording systems use their own internal time stamps. This poses a problem in synchronising the EPOG time bases with the simulator and video time bases.

Data synchronisation
Synchronisation is usually achieved by simultaneously transmitting a step or pulse to all the devices participating in the simulation. In the BAE SYSTEMS Pre-testing experiment this synchronisation was achieved by having the pilot that was flying the aircraft press the ‘Master Caution’ switch at the beginning of each scenario. This button press has been recorded on video. The button pressing has been recorded as a

time-stamped event in the simulator data and was also transmitted as a logical signal to the EPOG and JAZZ systems.

Analysis of the video of the BAE SYSTEMS pilot experiments has indicated that such a single step or pulse transmitted at the start of the simulation is not an adequate means for synchronisation. First, each video-tape must be viewed frame-by-frame to identify the exact frame of the button press. This is a time-consuming process sensitive to human error, which lacks the accuracy and robustness of an automated system. Second, the synchronisation is based on one instant of time only; if the pulse was transmitted while one of the devices was not yet switched on, or the video was not clear, synchronisation would be impossible.

Automatic data synchronisation

The BAE SYSTEMS pilot experiments stressed the necessity for using a more robust synchronisation method than the method of the single pulse or step. The problems occurring with the use of a single step or pulse for synchronisation can be overcome by continuously transmitting a unique pseudo-random telegraph wave for the entire duration of the experiment. This random signal will be transmitted by a master device and recorded simultaneously by all other devices. By correlating between the recorded random sequences of two devices over a certain time span, the time shift can be computed. A fast, accurate and robust correlation method was developed for computing the time shift between the two pseudo-random sequences..

The BAE SYSTEMS Pre-testing experiments also indicated the necessity for developing a *robust* method for *automatic* synchronisation between *video* and *simulator* data. Since the video format basically differs from digital simulator data, synchronisation is not possible by correlating between signals.

In order to overcome this problem, a novel method was developed in which the random telegraph wave is displayed on the video image through a numerical text window. This method will be employed for the first time in the NLR experiments. A blinking asterisk will be used to convey the random signal. When the asterisk appears, the telegraph wave status is ‘high’ and when it disappears, the wave status is ‘low’. In order to make synchronisation possible, the analogue video-tapes will be first converted to digital format.

Software was written to automatically extract the random sequence from the digitised videotapes by identifying the status (on or off) of the blinking asterisk. Since this random sequence now appears in digital format, the above mentioned correlation methods can be readily employed in an automated synchronisation scheme.

2.5.1.3.2 EPOG Visualization Tool

The basic EPOG data recordings of the BAE SYSTEMS experiments have been used in the development of the EPOG Visualization Tool. Although the basic software was written before the BAE SYSTEMS experiments, an actual database on which the software could be tested was lacking. The EPOG database of the BAE SYSTEMS experiments was very useful in debugging the software and testing the handling of two simultaneous EPOG recordings. However, at the time of WP5, the EPOG Visualization Tool was not yet developed to a stage that it could effectively be used in the analysis of the data.

2.5.1.3.3 Pre-test Scenarios

An important question for WP7 was whether the scenarios used in the Pre-testing did in fact manipulate shared situational awareness as predicted. An analysis of the subjective and objective data was carried out to see if this might be the case. The subjective data is split into three dimensions – situational awareness, workload and team working. For each of these dimensions, mean ratings were obtained for each scenario for the PF rating himself (PF_Own), the PF rating his co-pilot (PF_PNF), the PNF rating himself (PNF_Own) and the PNF rating his co-pilot (PNF_PF).

Both the PF and the PNF rated their own situational awareness as being lowest for Scenario B (Low Situational Awareness/Minor Event Condition). The PF and PNF also rated each other as having the lowest situational awareness in Scenario B. Interestingly, the pilots always appeared to rate their co-pilots' situational awareness as lower than their own for all three scenarios. The difference between the PF's rating of the PNF and the PNF's rating of his own situational awareness was largest for Scenario B suggesting that the mismatch between the pilots' perceived situational awareness was particularly high in this scenario.

Examination of the workload assessment questions showed that the crew rated workload the lowest for Scenario C (High Situational Awareness/Minor Event Condition), and the highest, except for the PNF, for Scenario B. In general, the pilots rated their co-pilots' workload as slightly higher than their own thus tending to estimate the workload of their fellow crewmember as higher than that estimated for themselves. As the crews were not informed that Scenario C was to be a repetition of Scenario B, the reduction in workload was likely to have occurred due to the recent handling and thus familiarity with the events.

All crew had a good attitude (Pre-flight Questionnaire) to team working suggesting team interaction during the scenarios was likely to be good. The crews in general rated the extent in which they engaged in team behaviours during a flight as quite high with the mean highest ratings for the PNF for Scenario C. For Scenario B team behaviour appeared to be reduced. This reduction in reported team working may have been due to either the higher workload during this flight interfering with normal team working behaviour or that it was more difficult to judge team working when workload was high. No consistent pattern was seen in the way the crew rated each other although the largest differences in perceived team working occurred for Scenario C. The overall team working for the crew as rated by the PF and PNF at the end of each scenario was calculated. The PF tended to rate the crew's team working as better than the PNF, while both the PF and PNF rated the crew as having the worst overall team working in Scenario B.

The Training Captain (objective measures) rated crew situational awareness and performance lowest for Scenario B agreeing with the subjective measures above. However, surprisingly workload was rated highest for Scenario A with Scenarios B and C rated as having quite similar workload levels, although for the PF the difference in workload between the scenarios was quite small.

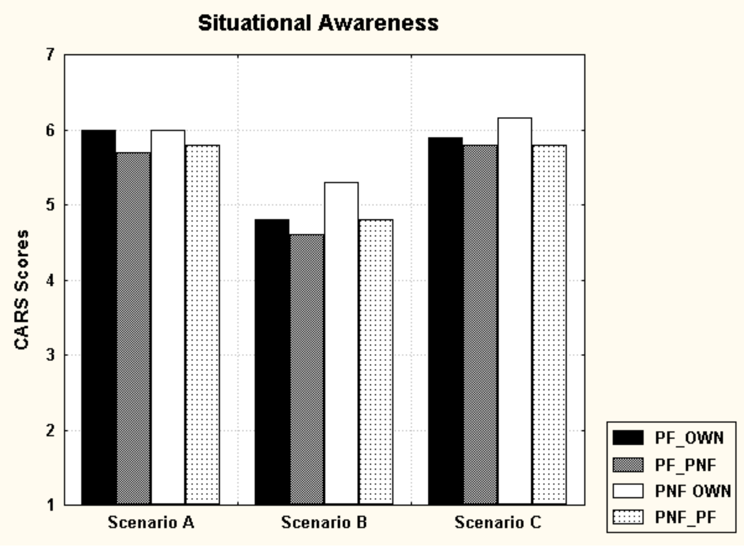


Figure 2.3. Mean crew awareness rating scale (CARS) results for each scenario. Smaller numbers indicate lower SA.

The reason for this difference in workload assessment between the subjective and objective measures is unclear, but may have been due to different interpretations of workload by the crew and Training Captain. For Scenario A, very little difference was seen between the ratings for the PF and PNF for situation awareness and workload whereas for Scenarios B and C the Training Captain tended to rate the PNF as having higher situation awareness, better performance and lower workload than the PF.

The experimenters noted during the trials that for Scenario B three of the four crew did not appear to notice the Advisory Page on the lower ECAM display and only started to deal with the fuel temperature high when alerted by the Master Caution. The crew that did pick up the problem early was, however, rated as having poor overall team working behaviour and poor performance in dealing with the fuel temperature high by the Training Captain. The Captain also commented that this particular crew demonstrated little interaction; he did, though, rate the PNF (but not the PF) as having good performance in dealing with the event.

These results suggest that the design of the scenarios in WP5 resulted in the desired manipulation of situational awareness. A clear difference between Scenario A and Scenario B in terms of situational awareness was observed indicating that minor ‘untrained’ events have a greater capacity to reveal aspects of crew coordination and situational awareness in the flightdeck. On the other hand, the relationship between team interaction and crew performance was not as clearly manipulated. While the pilots generally rated team interaction as high themselves, good team interaction did not appear to be necessary to detect minor events. These lessons learned regarding the scenarios provide valuable insight into how crews interact and deal with events, important issues for the final full-scale studies in WP7.

2.5.2 Pre-test PAT system

One of the objectives of the pilot runs at BAE SYSTEMS was to investigate whether it was in principle possible to use the PAT measurement as a successful measure of mental activity of interacting crewmembers in a simulated flight scenario. This would be the first time that the PAT would be tested outside the strictly controlled laboratory environment. The transfer from very restricted and well-controlled laboratory conditions to the active and relatively unconstrained cockpit environment was not trivial. In previous studies the subject was in a situation in which he/she was either lying down or in which his/her arm was immobilised in a special armrest. Since in actual flight scenarios considerable arm and hand motions are an integral part of pilot behaviour, the recorded PAT signal was expected to substantially differ from the type of signal previously recorded under benign laboratory conditions. As a result, the previously used software and data reduction techniques were no longer found to be effective for the signal recorded in a flight scenario and an entirely new data reduction package had to be developed.

In the framework of WP5 a relocatable PAT system was developed by Technion, based on a laptop computer with a 16-bit, four channel, Analogue-to-Digital (A/D) converter with discrete I/O channels. The A/D was implemented as a National Instruments miniature interface card. Data acquisition and visualisation software was developed and custom-tailored to the PAT task. This software was developed within the National Instruments Labview environment. It allowed the display and recording of the PAT signals of two pilots simultaneously. A fast and efficient algorithm was implemented to compute a running average of the PAT amplitude over a three-seconds interval. This average was graphically displayed as an amplitude bar and constituted a useful tool for the experimenter to obtain an online preliminary estimate of the operator mental activity.

The most effective and universal way for transferring time-stamped data between the PAT and the local host simulator, would be through an internet link. However, due to security reasons, such a connection could not be used at either BAE SYSTEMS or NLR. Transferring data from the host to the PAT system would be essential for synchronising the data files. In the framework of WP5 this difficulty was overcome by transferring data back and forth to the host through discrete data lines. This method was later also employed in the NLR full-scale simulator experiments in Amsterdam. One of these discrete I/O lines was used for synchronisation purposes.

During the implementation of all recording devices for the Pre-testing technical problems were experienced and the PAT device could not be operated correctly, resulting in a considerable loss of data. In addition, some technical problems with the EPOG device also occurred and combined data for PAT and EPOG were collected for one crew only. Data were collected for Scenario A (dual hydraulic failure), and Scenario C (repetition of Scenario B, fuel temperature malfunction). Because of the limited amount of data collected, statistical analysis was not possible. However, the observation of the data showed some changes in the behaviour of the measure during the different phases of the scenario, thus suggesting that PAT could represent a sensitive measure of workload in simulated flight missions. Note the pilot judged the device as comfortable and not intrusive.

2.5.2.1 Lessons learned

The PAT data, measured during the BAE SYSTEMS simulator experiments, differed significantly from the signals obtained previously in a benign laboratory environment, for a person at rest. Rather large fluctuations in PAT amplitude were noticed as a result of pilot arm motions. Moreover, the quality and the amplitude of the PAT signal were found to vary strongly as a result of hand motions. At some instances, the signal was found to deteriorate so strongly that the peaks and valleys of the signal, signifying the instances of maximum and minimum pressure, could no longer be detected uniquely. These involuntary variations could be attributed to slight shifts of the fingercup on the finger as a result of the hand motions, resulting in impaired contact between the LED and the photosensor, and the skin. In addition, arm and hand motions, such as lifting the arm to reach an overhead switch, would yield large pressure variations, which would not necessarily correlate with mental stress. The existing data reduction software was not designed to cope with these changes. Therefore a new approach was developed, in which the effects of artefacts resulting from hand and body motions are filtered from the data set.

The magnitude of the amplitude variations critically determines the gain setting of the PAT system that should be used in the experiment. The lesson learned from the BAE SYSTEMS experiments is that the gains chosen should not be too large. Although a large gain is desirable, since it increases the signal-to-noise ratio, it might result in a saturated signal, when the variations become very large. Therefore the gain setting should be chosen such that it is not only good at the beginning of the experiment, but it also takes into account a certain degree of reserves to cope with large variations that might occur later on.

A practical, but not less important lesson learned from the experiments at BAE SYSTEMS, is that the mobile PAT recording system should use a local power outlet plug, rather than an adapter. The use of the adapter caused a power surge, which has inactivated and damaged part of the simulation set-up. The existing data reduction software was not designed to cope with these changes.

2.5.2.2 Novel PAT data reduction software

Based on the data set measured during the BAE SYSTEMS experiments, a novel and robust data reduction technique for the PAT signal was developed by Technion as part of WP5, which is able to cope with the changes in the signal due to arm and hand motions. First, the maximum and the minimum pressures for each heartbeat are determined by identifying the ‘peaks’ and ‘valleys’ of the PAT signal. Second, the false peaks and valleys (due to hand motion artefacts) are rejected on the basis of a statistical analysis. This rejection procedure involves several stages. In the first stage, the compliance of the *magnitude* of the peak or valley of the signal is tested by considering the trend of the signal. In the second stage, the compliance of the *rate of appearance* of the peak or valley of the signal is tested. In the final stage, it is made sure that each peak can be associated with a corresponding valley, and non-associated peaks or valleys identified as artefacts, are discarded.

2.5.3 Pre-test JAZZ system

Another major concern of the pilot runs at BAE SYSTEMS was to test the use of the JAZZ system in a flight simulator prior to the final experiment in WP7. The main objectives for this part of WP5 were: to integrate and test the JAZZ system with the Eye-Point-of-Gaze head-mounted ASL system; to gather information on the implementation requirements in order to collect data in a simulator; to evaluate the intrusiveness of the JAZZ system and the acceptance by the participants; and to develop an analysis and software methodology for the JAZZ system for the full-scale studies planned for WP7.

Technical steps were taken by IBIB with the help of BAE SYSTEMS to modify the JAZZ system for use by pilots wearing the ASL eye/head tracking system within the BAE SYSTEMS research simulator. These steps included the development of hardware and software suitable for the testing environment. In particular, modification of the JAZZ hardware to enable the JAZZ system's attachment to the centre of the headband of the ASL system was carried out so that synchronised JAZZ and EPOG recordings could be made from one of the pilots (PNF) during the pre-test flight scenarios. The modifications meant that only a limited JAZZ signal was possible, as it was found the PPG-JAZZ sensor could not be applied together with EPOG headband.

During the Pre-testing trials themselves, the Post-flight Questionnaire revealed that the JAZZ system had little effect on comfort and was slightly less intrusive than either the helmet or headband mounted ASL systems.

An important issue, especially for the full-scale study, was whether the integrated JAZZ system would affect the quality of the EPOG data. In order to test if this might be the case, EPOG was recorded while three subjects performed a non-flying task (they were required to fixate in turn four targets located on scene plane 1) within the BAE SYSTEMS simulator with and without wearing the integrated JAZZ system.

An index of the quality of the EPOG data was determined for each condition by calculating the overall Root Mean Square offset of the fixations from the original targets (for full details see "EPOG Ergonomic Guidelines" VINTHEC II-WP2-TR 01). The results showed that the integration of the JAZZ and ASL system had little or no effect on the quality of the EPOG recordings. This finding suggested that the presence of the JAZZ system would not interfere with the EPOG data and that JAZZ data collection should be considered in the full-scale simulator experiments at NLR.

Although the BAE SYSTEMS Pre-testing experiment used only a limited JAZZ signal, because the PPG-JAZZ sensor could not be applied, software tools and file standards for the full range of JAZZ signals were written by IBIB for storing and analysing the JAZZ signals during WP5 in preparation for WP7.

The JAZZ software was developed with a modular structure where there is one main window – JAZZManager – from which different modules can be run. All the signals arriving from the JAZZ multi-sensor device are stored into the one binary file from which data can be later read out and visualised off-line. The process of recording the data can be carried out together with real time visualisation of the incoming signals or without it, depending on the PC processor power. Modules for on-line visualisation and recording, off-line data review and analysis and on-line analysis of pulse-oxymetry and head movements were all written and tested.

.

The concept of Visual Attention (VA) was explored and defined by IBIB as the psycho-physiological parameter expressed by the saccadic activity of the eye, *reflecting the conscious brain involvement in the process of visual exploration*s. Based on this definition a methodology for measuring VA using the JAZZ system was developed for WP7. The methodology (and associated analysis plan) involves the analysis of a number of parameters including: eye movements (number and amplitude of saccades, fixation durations, and VOR); pulse-oxymetry (heart rate, amplitude of heart activity, saturation of oxygen in the blood (SO2), and oxy-total balance); and head movements (the "energy" of accX signal, the "energy" of accY signal, and the sum of accX and accY "energies").

2.5.3.1 Lessons learned

The JAZZ system used in the Pre-testing as a Visual Attention monitor device was a prototype, incorporating the still developing model of the high-level function of the brain. Expanding the visual attention model may result in changing the definition of Visual Attention, e.g. currently IBIB places some effort in differentiating consciously and preconsciously driven saccades and incorporating the head movements into Visual Attention. In such a case, there will be a need for new algorithms and software.

As a direct result of the Pre-testing, a second version of the JAZZ system has been designed. This new system will include such features as automatic gain control, better signal-to-noise ratio and higher artefact immunity. From the point of view of the signal processing, received data will be compatible with its present form, but the sensor driving hardware and software will be redesigned. The interface, most probably, will still be based on IrDA and USB transmission, however Ericsson's BlueTooth will also be considered, as it allows a wireless PC link.

2.5.4 Pre-test Data Specification and Visualization tools

The purpose and scope of this part of WP5 was to evaluate the Data Specification Tool ("EPOG Data Gizmo", Beta version 1.0) and the Data Visualzation Tool ("EPOG Visualization Tool", Beta version 1.02), and to provide feedback to the developers of the software (Technion). The "Data Specification Tool" is a tool developed to import and correctly set up the parameters for the simulator and EPOG data. It automatically creates the necessary configuration files needed by the main program, the Visualization Tool. The Visualization Tool is the main tool for replaying a scenario and simultaneously visualising EPOG-data. It also provides some statistical summaries of the eye movements. A more thorough description of data analysis software tools, the Data Specification Tool and the Data Visualization Tool was provided in the workpackage 3 report (VINTHEC II-WP3-TR 1.0 "Technical Report: User Needs, Knowledge Elicitation and Software Tools").

A comprehensive assessment of the tools was carried out using two different data sets: (1) the test data provided together with the Visualization Tool; and (2) data from the WP5 Pre-testing study. In an attempt to thoroughly assess the software, some of the data were manipulated in order to have complete control of the input to the software packages. The Pre-testing data used for these purposes were Simulator data, EPOG data, JAZZ data and Video data. No evaluation of the data was performed since the tools at that time did not produce reliable results.

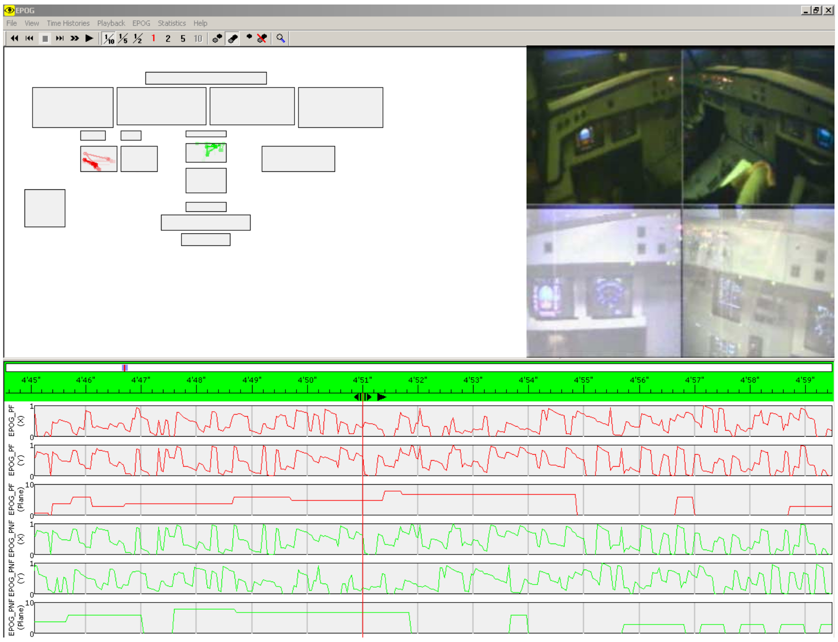


Figure 2.4. A cockpit layout can be constructed so that when the EPOG traces are presented they reflect the actual visual behaviour, including scan patterns. By using different colours for the EPOG-traces of the Pilot Flying, and the Pilot Not-Flying they are easily distinguished

2.5.4.1 Overall assessment

Generally, the tool was found to fulfil the requirements according to the purpose set. The tool could be used for visualising (and analysing) data. To be able to look at the dynamics of the situation with the use of the Visualization Tool was very helpful. How the situation, and the shared situational awareness, changes over time was easier to understand when using the tool. Also, the tool supported the detection and evaluation of detailed analyses of specific parts of the data, such as a short time sequence. The ability of easy "zooming" between details and the whole scenario in visualising and analysing is an important aspect of the tool since shared situational awareness is present in all the spectra.

A number of suggestions for future versions of the Data Specification Tool and the Visualization Tool were given by FOI. One, for example, was that it would be helpful if the window used for the Data Specification Tool could be "minimised" so that the user could easily leave the program for a quick look at another program, and come back to where he/she was before. Another was that in the configuration file, one gauge can only correspond to one EPOG-surface. If surface (scene plane) number 5 for the Pilot Flying is not defined as the same surface as surface number 5 for the Pilot Not-Flying, as in the Pre-testing study, then there is a problem with the tool. Three solutions to this problem were proposed.

The statistics part of the Visualization Tool was found to be very useful. The capability to be able to quickly shift between the visualised data and statistical descriptions of the same data is valuable. The way that a user can dynamically, very rapidly, shift between these environments, and also between "levels of detail" (by zooming into something interesting) works well. The version of the program assessed gives the options of presenting a "Scan Surface Transition Table", a "Scan Surface Transition Map", and a "Dwell & Fixation Table". The "Scan Surface Transition Table" summarises both the percentage of transitions, as well as the actual number of transitions between each surface.

A set of tests with very simple test files, however, showed that the statistical part of the program was not always correct. Several “fake” EPOG data strings were entered to try to understand the values of the different parameters, but some parts of the analyses seem to yield surprising values. For instance the “dwell time” on a surface may well be negative in some instances and the number of fixations on a surface does not always correspond to the input data strings.

2.5.4.2 Lessons learned

The problems that were encountered with the Data Specification and Visualization Tools were provided to Technion as “lessons learned” for the future development of the tools. In brief, the lessons for the Visualization Tool included: the necessity to pre-process the data before it could be entered into the Visualisation tool; the difficulty the tool had in handling large files; and the problems with the “Print Screen” button.

The Data Specification Tool also had difficulty compiling large data files, but was found to be helpful in picking up when small mistakes are made, for example if a comma (,) is used instead of a decimal point (.) or if one forgets to enter some data. Although, the Data Specification Tool, sometimes displayed a negative scale on its pre-visualisation even when only positive numbers had been entered. For example, if the altimeter was defined with values between 0 and 2000 and the scale as set to 100, a scale appeared with values from -10 to +10. However, the user could neglect this problem, since it will not be replicated in the layout in the Visualization Tool.

In conclusion, despite the problems encountered, the tool is already useful. It provides an easy-to-use solution for visualising flight data together with EPOG data from one or more operators. It is fairly straight-forward and easy to use. The “Statistics” feature summarises the EPOG data to give the user the bigger picture of the operators’ scanning behaviour. However, there are some issues that need to be corrected, and some issues that, when further developed, will probably make the tool very attractive for visualisation and evaluation.

2.5.5 Review measurement approach

The final task for WP5 was to evaluate the approach adopted in the Pre-testing study and to make recommendations, especially with regard to EPOG measurement, for the final experiment in WP7. An independent expert (TestUsability) with knowledge and experience in the field (a member of the VINTHEC consortium) made the following observations and recommendations based on studying three video-tapes from the WP5 Pre-testing study.

Composition/EPOG quality

Four sub images in one frame
The level of detail is limited, because an original video image is reduced to 25% of the size of a standard image.

If EPOG images are reduced to 25%, one must expect the level of detail that can be observed, while coding the data, to be limited.

Quality per image
It is essential that the brightness/contrast per video sub image fall in the same range. This must be realised in the original separate video images before merging. The composite image is not usable if the differences are too large.

Quality of EPOG information
It is to be expected that periods of disturbed EPOG data will occur, even if the calibration of the EPOG system has been completed successfully. An **experienced engineer** must be involved in performing the coding of video-based EPOG data. The engineer determines the quality of the data to be used for analysis and interpretation.

Pupil diameter is the result of the amount of ambient light reaching the eyes. It is recommended that light-emitting displays be tuned correctly (e.g. flight and system displays).

Parallax error (due to scene-camera offset). If the offset is not extremely small, errors will occur due to the large differences in viewing distances. An example is looking at one's own PFD, as compared to looking at the PFD of the other crewmember.

Images can, and have been, spoiled due to **reflections** in the visor of the helmet-mounted eye tracker. A well-designed, low mass "cap" can be used in most cases. If "over-head panel" activities take place, the use of a "cap" may be counter-productive (blocking of looking upwards).

In the tapes of WP-5, the cross hairs are not always discernible. Analysis requires visible cross hairs.

Quality per image in slow motion
As the coding will take place on images, which are presented in slow-motion replay, the readability must be checked for that particular mode of operation.

Generally, the readability of the cross hairs is degraded in slow motion as compared to replay in real time.

Information per video channel

Observation (WP5)
Apart from the Analogue EPOG images of the two crewmembers, there are additional cameras in operation.

The best and most informative areas should be selected.

Limited crew action was recorded in the two upper video images (fixed camera) of WP5 Pilot Testing. There is a clear need for more fixed cameras.

Recommendation
Include at least one birds-eye view camera image, taken from a position behind and above the crewmembers (facing forward).

Organisation/synchronisation

Recommendation

- Use "ID numbers" on every video-tape (visible on screen at the start).
- Each video frame should have its own time-code signal imprinted.
- If more than one recorder is used, a time code from one source should be included in each tape (or counters with a common start signal).
- If more than one image per frame is used, the mixing should be realised in real time and the time-code signal should be included on each frame.
- Video-based EPOG recorded on tape must be synchronised with Digital EPOG in the computer system.
- One has to make absolutely sure that the system for time recognition used in the analysis (hardware and software) is compatible with the format of the time code used in the measuring phase.

Audio quality

Observation (WP5)
There is a large difference between the clarity of the spoken words of the individual crewmembers (WP5).

R*ecommendation*
Use high-quality headsets and boom mikes. This is not simple, as such gear must be integrated with the helmet of the eye-tracking system.

Observation (WP5)
A large percentage of the time the pilots talked to each other, they did so in a very low voice, almost a whisper.

Recommendation
The large, dynamic range of the spoken word requires high-quality audio equipment.

2.5.5.1 Lessons learned

The review emphasised the importance of synchronisation of the video-based EPOG data with other data streams and the need for an adequate quality of recorded speech (ATC, inter-crew communication). The overall conclusion was that video-based EPOG measurements will improve the interpretation of digital EPOG data, and thus increase the value of the shared SA determination, when used for the development of pilot-centred automation in future aircraft development.

It is to be expected that the prime benefits of video-based EPOG will result from improved interpretation of digital EPOG data, and an improved understanding regarding the application of digital EPOG, video-based EPOG or both. One must realise, though, that the overall picture must be considered, i.e. the complexity of the equipment required, the effort of fitting equipment into the cockpit, the calibration effort per pilot, and the complexity of analysis of digital EPOG.

2.5.6 WP5 Conclusions

WP5 (Pre-testing) revealed a number of potential problems and issues for consideration prior to the full-scale experiments in workpackage 7. Furthermore, considerable effort was devoted to the further development of the PAT and JAZZ systems and the EPOG Visualization Tool in preparation for the final experiments. The problems that were encountered during these pilot runs were used as “lessons” that shall be avoided in the final full-scale studies.

2.6 WP6 Cognitive modelling

(WP leader: Linköping University)

The modelling work for VINTHEC II had two distinct purposes. One was to support, inform and help refine the scenario development work for WP 7. Contributions from various partners (especially Technion and FOI) were helpful in this regard. Modelling helps in identifying critical variables and supports the researcher in finding out what counts as a stimulus in the observed behaviour. The other distinct purpose was to try to generate *converging evidence*. In scientific methodology, converging evidence is judged to be critical particularly in cases of not directly observable phenomena. sSA is certainly not directly observable – its existence and nature has to be construed from other constituent performance measures. If similar or even identical factors can be shown to have similar or even identical effects on the phenomenon *using different methods*, the experimenter can be more confident that these factors indeed play a role in the phenomenon in the real world. Multiple methods can also show divergences or differences in results, which function as a challenge for current understanding and a cue for deeper probing into the nature of the phenomenon. What binds modelling even more strongly with human-in-the-loop experiments is that they both aim to demonstrate the experimenter’s understanding of the phenomenon by allowing the experimenter to explore his/her control of the phenomenon. If the experimenter can make the phenomenon happen, either in the computer or in a simulated human-in-the-loop setting, then he or she demonstrates control, and through that, greater understanding of the phenomenon.

With respect especially to the second purpose, the Integrated Performance Modelling Environment (IPME) was used to model and test the influence of Power Distance and Uncertainty Avoidance on Shared Situation Awareness in the cockpit. Shared Situation Awareness was measured by the amount and quality of the communication. Uncertainty Avoidance was implemented as influencing the arousal of pilots. Power Distance is implemented as influencing the probability of sharing information with the other crewmember. Two experiments were carried out to see if different crew characteristics influence SSA in the model. The model is a simulation of an airline crew shortly after take-off. At some point the Lateral Navigation system fails. Depending on the monitoring and cooperation of the crew they can: return to their course without anything else happening, get a traffic alert, or get a traffic alert and a resolution advisory. From the results of the experiments it could be concluded that, most of the time, high UA crews maintain better but not more SSA, but when there are many problems the SSA becomes less. Low UA crews maintain, most of the time, not so good but not less SSA. However, when problems escalate, crews maintain more SSA. The conclusions of the effects of Power Distance are: most of the time the different PD crews have the same amount of SSA, but when there are many problems the low PD crews maintain more SSA. This enables them to maintain better SSA when they get into trouble and helps them to prevent the problems getting worse.

The second part of the WP 6 report presents perspectives of the Information Processing (IP) view and the Cognitive Systems Engineering (CSE) view on cognitive modelling. Both views are described, contrasted, and to a certain extent applied to a scenario of LNAV failure, which is a sub-scenario of the larger sequences of actions and events as flown in the simulation experiments of WP 7. The Information Processing view is in this document represented by the Model Human Processor, GOMS models, and COGNET. The Cognitive Systems Engineering view was addressed by analysing aspects of COCOM modelling. The analysis and application of the combination of concepts of Cognitive Systems Engineering brought forth the most interesting and – in terms of Crew Resource Management, procedure evaluation, and

Shared Situation Awareness – useful way of cognitive modelling, in contrast to the possibly more readily programmable but artificial way of cognitive modelling in the Information Processing view.

The results from the modelling efforts in WP 6 made intuitive sense and could be quite consistent with actual performance results — to the extent that the variables could be manipulated and tested in the actual simulation, that is. Modelling efforts from WP 6, moreover, were of considerable help in shaping and focusing the discussions related to scenario design for WP 7.

2.7 WP7 Full-scale simulator experiment

(WP leader: Nationaal Lucht- en Ruimtevaartlaboratorium)

2.7.1 Objectives

The objective of WP7 was to perform an experiment in a motion-based flight simulator that provides insight about shared situational awareness, and serves as a "proof of concept" demonstration of shared SA measurement in an operationally realistic setting.

2.7.2 Approach

The, in WP4 developed, scenarios were slightly adjusted so that they could be implemented while they remained relevant and could be analysed with the measurement battery that was available for the full-scale simulator experiment.

Emphasis was put on so-called Periods of Interest (POIs). These were selected events that were introduced in the scenarios by the researchers. It was agreed that all partners would, for the data analysis that they performed, focus on these POIs thereby enabling comparison of the measures during exactly the same situations afterwards. And also enabling the possibility to evaluate the added value of each measure in relation to assessment of sSA.

2.7.2.1 The Periods of Interest

The POIs were:

- POI 1: Pre-flight preparation.

The discussion of NOTAMS (e.g. inoperative airports), weather reports and flight plan should be registered.

- POI 2: Departure crew briefing.

A complete crew briefing was held, which helped pilots build up their Situational Awareness. Essential components of briefing were:

- Maintenance log
- Navigation (route charts)
- Lining up aircraft
- Situations (e.g. ATC, weather)

- POI 3: Noisy Passenger

At 3000 ft, the purser called to inform the crew about a noisy passenger:
<gong><response by pilot A> "<first-name of pilot A>, dit is de purser. Ik wil meedelen dat wij een lastige passagier aan boord hebben. Hij maakt ruzie met iedereen om zich heen". (Notice that this text is in Dutch, since all crews were Dutch crews and the consortium wanted the flights to be as 'normal' as possible. It means "this is the purser, I want to inform you that we have a troublesome passenger on board. He is quarrelling with everyone in his vicinity".)

- POI 4: ATC directives

Give clearances to FL 80, 100, 120, 140 and finally to cruising FL.

- POI 5: Cruise.

Nothing special occurred.

- POI 6: Approach crew briefing.

The remaining POIs occurred in the second experimental flight:

- POI 7: Pre-flight preparation.

- POI 8: Departure crew briefing.

- POI 9: Passenger unwell.

After takeoff, about 3000 ft, cabin crew called flightdeck to discuss a problem with a sick passenger…. At this stage the passenger is just feeling unwell.
<gong><response by pilot A> “<first-name of pilot A>, dit is de purser. Ik wil meedelen dat een passagier behoorlijk misselijk aan het worden is.” (Notice that this text is in Dutch, since all crews were Dutch crews and the consortium wanted the flights to be as 'normal' as possible. It means "this is the purser, I want to inform you that one of our passengers is getting nauseated".)

- POI 10: Primary Flight Display (PFD) failure.

The Aircraft (A/C) generated a PFD failure on the side of the PNF at D8.0 SPL. N.B. the last event (POI 9) may trigger a lengthy discussion. This may be before SPL. If so, the POI 10 should be delayed till later.
The PFD blacked out for 20 seconds. The pilots were forced to give the controls to the pilot not-flying. This initiated a discussion that could be handled quickly. After sometime pilots could switch controls back.

- POI 11: Distracting ATC calls

A number of distracting ATC calls were given (e.g. a diversion due to a weather balloon).

- POI 12: Cruise.

Nothing special occurred. Normal behaviour was be observed. Questions about past events should be answered.

- POI 13: Passenger heart attack

<gong><response by pilot A> “<first-name of pilot A>, dit is de purser. Ik wil meedelen dat de passagier die zonet behoorlijk misselijk werd, aan het overgeven is. Hij klaagt ook over pijn in de borst. We denken dat ie een hartaanval krijgt”. (Notice that this text is in Dutch, since all crews were Dutch crews and the consortium wanted the flights to be as 'normal' as possible. It means "this is the purser, I want to inform you that the passenger who was getting nauseated, is vomiting. He is also complaining about pain in his chest. We think that he is developing a heart attack".)

Cabin crew advise medical problem has worsened and doctor wants to open medical kit… pax may be suffering a heart attack. This event should occur past half way through the flight, Just before LAM.

- POI 14: Second PFD failure on the same monitor.

Re-introduce PFD failure at LAM

- POI 15: Approach crew briefing.

- POI 16: Glide slope capture failure.

Glide slope capture failure on ILS RWY 27L. A go-around is forced. The glide slope signal is corrupted, and is not intercepted. A/C passes through glide slope and stays at 2500 ft.

Special attention was given to POIs 3, 9, 10, 12 and 16 since these were identified as the POIs in which all relevant and measure aspects of sSA assessment (like high WL and expected individual SA, need to collaborate, were present.

2.7.3 Hypotheses

A number of hypotheses were evaluated in the WP7 study. They comprise:

1. Together the applied assessment methods / measurement instruments accurately measure sSA as defined in WP1 of the VINTHEC II project.
2. There will be a strong correlation between (s)SA and SA as well as between (s)SA and workload[1]. By developing scenario's in which different combinations of expected (s)SA and required mental effort or workload are present these correlations may be proven.
3. The under 2 described relationships will turn out to be useful as the basis for a sSA measurement battery.
4. The under 2 described relationships will, as a result of the scenarios and the events that are introduced in the scenarios become measurable. That is, by introducing certain events in the scenarios the workload, SA, sSA will be influenced. The VINTHEC II measurement battery will register these changes due to introduced events.
5. The objective of VINTHEC II (to perform an experiment that provides insight about sSA, and serves as a “proof of concept” of the sSA measurement methodology) implies the hypothesis that the measurement methodology will turn out to be insightful and a useful methodology for sSA assessment. The current study will demonstrate whether this is the case for the individual measures as well as for the battery in general.

2.7.4 Results and conclusions

For the current report it would be getting too far of the subject if all results were discussed in details. Therefore we limit ourselves to a description of the tools (and in Section 2.7.5) the most important conclusions. For more details about the results the reader is referred to VINTHEC II WP7 - TR - 01 (2003).

During the experiments, in a realistic flight simulator, measures to understand how knowledge, action and communication were linked together in forming a mental picture by interactions with tools in the environment (actions), perceiving and recognising patterns (knowledge), and sharing this (communication) were recorded.

These measures were analysed by different partners and the results / conclusions reported back to the co-ordinating partner. It would be getting too far of the subject of the current document to discuss the results of all of these measures in detail. For that information the reader is referred to (VINTHEC II - WP7 - TR 01, 2003).

The measures that were applied comprised;

2.7.4.1 Eye Point-of-Gaze

Eye Point-of-Gaze (EPOG) is the point on a predefined surface where an imaginary line coming straight from the centre of the eye crosses that surface via the lens of the eye. As such this is the central point in the pilot’s field of vision. This point was measured by means of an EPOG-recorder called GazeTracker (Mooij & Associates, 1996). The optical, and at the same time electronic, equipment from the EPOG recorder was mounted on helmets that the pilots were wearing.

The duration that a pilot looks at a particular area of interest, is called a “dwell”, which was stored in a computer file. In addition to the dwell-times the scanning pattern, the amounts of fixations, the pupil

---

[1] Note that the terms mental effort, mental workload and workload all three refer to slightly different but to a large extent overlapping concepts. In scientific literature it is quite common to make a distinction like; mental *workload* = determined by a combination of mental effort (as for example measured by the RSME) and performance. However in VINTHEC II the terms were used interchangeably and all refer to the amount of perceived effort / work that was performed by a pilot or a crew.

diameter and the eye blink activity (which permits blink rate, duration and other measures to be derived) of the pilots' left eyes were recorded as indicators of mental- and visual workload. (See Harris et al, 1986 / Stern, 1994 / Wilson, 1987 and 1993.) The scanning behaviour was considered to be an indicator of the pilot's mental state and focus of attention. The common accepted assumption was made that if a pilot looks at a particular area of interest he is paying attention to that area.

The accuracy used to discriminate between fixations was 1.5 degrees visual angle from the eye and a minimum fixation duration of 150 ms. For the analysis it was therefore assumed that only if the subject looked at a point more than 1.5 degrees from the previous fixation, and for longer than 150 ms., a new fixation was made. For the determination of the eye blink rate, blinks that lasted longer then 40 ms. were considered as real blinks. Blinks were considered to be all further situations during which the pupil diameter was zero (closed eyelid).

The GazeTrackers were implemented in a way that is comparable to the way that they were implemented during the VINTHEC I project. For details the reader is referred to VINTHEC - WP8 - TR 01, Paragraph 5.2.2.2 and 5.2.2.3.

Besides the above-described digital output of the GazeTracker there was also video. Cameras mounted on top of the helmets (the same helmets that carried the other optronics) recorded the pilots' fields of vision on videotape. Overlaid on these fields of vision were crosshairs indicating the pilots' exact point of regard within these fields. This video based way of data recording was specially enabled for the VINTHEC II project.

2.7.4.2 Peripheral Arterial Tone

Peripheral arterial tone (PAT) and heart rate (HR) were measured throughout the experiment using a pneumatic plethysmography (Itamar Medical, Ltd., Israel). Details of the device have been published previously (Lavie, Schnall, Sheffy, & Shlitner, 2000; Schnall, Shlitner, Sheffy, Kedar, & Lavie, 1999). The signal was band-passed filtered (0.3-30 Hz), amplified and then stored to memory. A PC equipped with a Pentium II (166 MHz) processor was used for data acquisition. Artefacts were replaced by extrapolation. Subjects with too many missing values in succession were excluded from further analyses. The important extracted measures are changes in the amplitude of the peripheral vasoconstriction. As shown in previous studies, an increase in mental load is associated with decreased amplitude, or greater vasoconstriction of the arteries of the fingertips. Reduced mental load is accompanied by increased amplitude measured at the fingertips.

2.7.4.3 JAZZ

The JAZZ is a multi-sensor system that allows the acquisition of the eye movements with excellent spatial and temporal resolution, together with other physiological and environmental signals. The main idea behind using Jazz multi-sensor was to gather different kinds of information about pilot's interaction in the cockpit, using a single device. The physiological signals measured by Jazz include:

- eye movements in horizontal and vertical axis (1000 Hz sampling frequency),
- head acceleration in horizontal and vertical axis (250 Hz sampling frequency),
- photoplethysmographic signal in two lengths of the light wave (250 Hz sampling frequency),
- ambient illumination (250 Hz sampling frequency),
- voice recording (4000 Hz sampling frequency).

For the eye movement measurement the Jazz system utilises the Cyclops-ODS technology (infrared oculography - IRO) optimised for easy set-up and minimal intrusiveness, which is crucial for monitoring of the subject behaviour in non-laboratory environment. Unlike most IRO setups, which place the infrared emitters/sensors around, or even next to the eye, Cyclops-ODS's set of optoelectronic sensors is placed between the eyes, hiding the sensor in the "shadow" of the nose. Thanks to such a setup, limitation of the visual field is minimal, reducing the risk of interference with subject's visual exploration of the environment. The eye movement measurement performed with high temporal and spatial resolution allows precise detection of saccades — fast eye movements used to move the point of gaze around the available field of view. Statistical processing of detected saccades over selected periods of time (their quantity, amplitude, duration of preceding fixations) provides important information about pilot's visual attention during different flight phases.

Head acceleration measurement allows detection of the head movements connected with visual exploration of the cockpit environment. The photoplethysmographic signals measured by JAZZ system allow evaluating subject's heartbeat and blood oxygenation. As these signals carry the information about the vasodilatory response of the sympathetic system, this measurement can be also used to access subject's workload.

Voice recording carries information about experimenter's/subject's comments, audio signal in environment, communication in the cockpit. Ambient illumination can be use to document amount of light in the environment and differentiate between the light source (natural or artificial).

2.7.4.4 Subjective

Before the experiments started all pilots were asked to fill in a number of bibliographical questionnaires.

The questionnaires that were applies were the Rating Scale Mental Effort (RSME). This is a validated univariate scale for subjective (i.e., self-report) assessment of mental effort (Zijlstra, 1985). The RSME scale ranges from 0 to 150 point anchors. The instrument can be administered in either paper-and-pencil form, or electronic form. For ease of use, the first option was chosen. 0 means low mental effort and 150 means high effort.

Further dedicated questionnaires for performance, SA and sSA were applied.

Questionnaire were not applied during the experiments but (except for the bibliographical ones) presented after the experiments during a so-called auto-confrontation session. During this session crews were confronted with videotapes of their own behaviour that were recorded during the experiments. In a systematic all pilots were asked the same questions. This approach allowed a systematic analysis of data while pilots had plenty of opportunities to explain what, according to them, had happened.

2.7.4.5 Communication

All communication with the outside world as well as within the crews was recorded. The Dutch spoken text was translated so that all partners could understand what was discussed. During the analysis attention was paid to: number of speech acts (few/many), homogeneous communication patterns (similar/different), and temporal aspect of communication (past, present or future tense).

2.7.4.6 Rough Performance Indicator (RPI)

The consortium felt the need to have an indication of pilot performance and SA available when viewing the videotapes and examining the data for the first time. In order to provide the partners with such an indicator an experienced pilot, who was also involved in the scenario development plan, watched videotapes of all crews and rated performance SA and shared SA of all pilots on ten-point scales (where 1 = lowest and 10 = highest). He also added remarks when appropriate in order to explain the ratings.

2.7.5 Conclusions

The objective of WP7, (to perform an experiment that provides insight and serves as a "proof of concept" that sSA is measurable in a realistic setting), was met because the experiment took place and did demonstrate the applicability of the VINTHEC II sSA measurement methodology.

JAZZ, PAT and EPOG are useful for on-line assessment. Clearer definitions need to be formulated in order to determine which measure should be used and what the 'borders' for good or bad sSA are. The interesting aspect of JAZZ data is that they respond immediately to changes. As such they are great indicators of possible changes in sSA. JAZZ, EPOG and PAT all three deliver data that are perfect for - on-line - analysis of data that are recorded during a couple of minutes, according to the sliding windows principle. Therefore all three of those techniques are useful for real time data analysis and interpretation.

For JAZZ, and EPOG and PAT it is also true that a number of technical hick-ups still need to be solved before the tools can be implemented in any (not just a controlled flight simulator) environment.

Eye related data demonstrates:

- a trend towards more distributed scanning behaviour when workload increases
- a trend towards focused than distributed scanning behaviour during eventful flight segments
- that homogeneity can show that pilots have expectations of their colleague
- that mutual gaze increases with events (seeking information about others intention)
- that unfortunately peripheral-vision is not measurable while it clearly is important when assessing sSA
- EPOG and / or JAZZ recording requires calibration and may be considered intrusive
- JAZZ can be applied in real fight

Peripheral Arterial Tone (PAT):

- Not yet applicable in real flight
- Interferes with dexterity of subject
- *Other conclusions will be provided by Technion*

EPOG does not reveal ongoing cognitive processes if no context is available, communication can provide that context by:

- number of speech acts (correlates with performance, which in turn correlates with high sSA)
- speaking in future tense (correlates with performance, which in turn correlates with high sSA)
- homogeneity (structure) indicates high sSA
- communication recording is non intrusive

Communication really teaches us a great deal about the crews sSA. Unfortunately it is, with the current state of technology, not possible to apply content analysis of communication in real time. With the help of voice keys the frequency (and direction) of speech can be measured and analysed in real time. But that is just an aspect of speech analysis, interpretation like past-present-future tense can not automatically be measured and interpreted.

Communication is also one of the main factors that can be used for sSA assessment as opposed to one persons SA assessment. The fact that sSA is not just a matter of what the team has noticed but also how the team, consisting out of two individuals who are working together, has shared that knowledge springs, especially clear, forward from the communication within the team.

Auto-confrontation and task analysis enabled to put the results that were found into perspective. Like LISREL these methods are difficult to apply for real time sSA assessment but they are vital for developing sSA measurements batteries and for evaluation of existing sSA measurement batteries.

Concrete outcomes of the auto-confrontation were:

- SOP related problems have minor impact on sSA
- A clear division between tasks of PF and PNF increases sSA
- High familiarity and power distance were best sSA predictors
- Communication: clear roles, re-confirming actions have positive effect on sSA

LISREL and questionnaires are useful tools for fine-tuning the measurement battery and evaluating the relationships between the psychological constructs that sSA is build upon. They have demonstrated:

- a positive correlation between SA and sSA
- a positive correlation between SA and performance
- a negative correlation between mental effort and performance

- a positive correlation between sSA and performance
- a negative correlation between SA and mental workload
- a negative correlation between sSA and mental workload

The events that were introduced in the scenario's in order to vary the (s)SA and workload levels turned out to have had the following effects:

- noisy passenger and glide slope capture failure were low SA events
- glide slope capture failure was a low sSA event
- noisy passenger, sick passenger and glide slope capture failure were low performance events
- glide slope capture was a high workload event

A number of the individual measures that were applied correlate well with sSA and look promising. All measures that were applied contributed to sSA assessment;

- some as instruments to fine tune the measurement battery
- others as potential candidates for the battery

The current approach seems to be heading in the right direction, though not yet leading directly to the measurement battery that the consortium was looking for and that they can start exploiting right away.

The measures that were applied indeed match with the in WP1 created definition of sSA: "the extent of agreement between multiple crewmembers' continuously evolving assessments of the current state and trends of a process they jointly manage". Especially communication and eye scanning behaviour allow to monitor the "overlap" between both crew members, as well as the evolving of the crews assessment from past and present to future.

The results of the present study are of relevance to the improvement of current training for civil aircraft operations, for creating a safety culture since many of the in this report mentioned factors are trainable and for 'Human Factors certification' purposes! The results can be exported to other relevant domains that share similar characteristics (e.g. team interaction, complex system dynamics, and highly automated systems).

The current report is an attempt to bridge the gap between theory and empirical findings. Besides connecting the purpose of the project with the empirical findings from the experiment, it also gives sSA a position in the map of psychological theory. The concept of sSA in the present study and applied measurement seem promising for future applications in realistic environments.

2.7.6 Interrelation with other WPs

The WPs 1, 2, 3, 4 and 5 provided their input to WP7 with respect to respectively:

1. Theoretical / scientific background
2. Measures
3. Analysis / visualisation
4. Scenarios
5. Lessons learned from pilot studies.

Further the WP gave output to WP8 about:

1. Guidelines for others
2. Comparison with the cognitive modelling exercise from WP6.

2.7.7 Deliverables

The WP resulted in one thick main deliverable: VINTHEC II WP7 - TR - 01. It also provided data and experiences to further develop / fine tune DAVIT (DAta VIsualization Tool) and the JAZZ. As preparation for the full scale studies a number of working documents were generated. These comprise:

1. Pilot briefing guide for the RFS experiment at NLR
2. CBIS - Crew Briefing and Information System
3. Radio scripts LHR-AMS and AMS-LHR

4. PAT measures analysis
5. Task analysis - Interviews with El-Al pilots
6. Data Acquisition and Reduction of Peripheral Arterial Tone (PAT) Measurements

2.7.8 Role of all partners

All VINTHEC II partners have contributed to the full-scale simulator experiment.

- NLR has co-ordinated the work (planning and preparation, integration as well as data analysis). Of course all of this was done in close collaboration with all partners.
- BAE has provided their experience from the pilot studies to prevent the final studies from making too many misjudgements or mistakes.
- QinetiQ has together with FOI selected number of questionnaires and analysed them. Further FOI has evaluated the subjective ratings in their LISREL model.
- LiU has provided the theoretical background and watched over the relationship between experiment and scientific theory.
- RISØ has provided analyses of EPOG as well as communication data.
- Technion has performed an analysis on the PAT data and has performed a task analysis.
- IBIB has installed their JAZZ tool in the simulator. They analysed the data and thereby delivered a fresh view on eye data in relation to sSA.

2.8 WP8 Conclusions and evaluation

(WP leader: Nationaal Lucht- en Ruimtevaartlaboratorium)
The objective of WP8 was to integrate all information that resulted from the VINTHEC II project in a final report and in so-called "engineering design guides and recommendations". The first 5 workpackages have provided their output to WPs 6 and 7. Therefore most of the integration of results from all workpackages was already taken care of. However, relationship between cognitive modelling as done in WP6 and human-in-the-loop simulation as done in WP7, was not described yet. Therefore the emphasis of the current section will be on the simulation - modelling relationship rather than on the technical accomplishments and benefits of the entire project.

Further WP8 would, if applicable, indicate the validity of the sSA measures. It would provide an elaborate definition of sSA. It would provide guidelines in a separate document.

2.8.1 Simulation versus modelling

Most of the results from the project come together in the full-scale experiment (WP7). Though one very interesting comparison was not made yet. This comparison concerns the relationship between simulation and cognitive modelling.

The advantage of multiple methods to specify and understand sSA as pursued by VINTHEC II is that they can generate "converging evidence". In scientific methodology, converging evidence is judged to be critical particularly in cases of not directly observable phenomena. sSA is certainly not directly observable – its existence and nature has to be construed from other constituent performance measures. If similar or even identical factors can be shown to have similar or even identical effects on the phenomenon using different methods, the experimenter can be more confident that these factors indeed play a role in the phenomenon in the real world. Multiple methods can also show divergences or differences in results, which function as a challenge for current understanding and a cue for deeper probing into the nature of the phenomenon.

What binds modelling even more strongly with human-in-the-loop experiments is that they both aim to demonstrate the experimenter’s understanding of the phenomenon by allowing the experimenter to explore his / her control over the phenomenon. If the experimenter can make the phenomenon happen, either in the computer or in a simulated human-in-the-loop setting, then he or she demonstrates control, and through that, greater understanding of the phenomenon.

2.8.1.1 Modelling

The VINTHEC partners have gathered a great deal of modelling experience before, but also during the project run time. Especially concerning modelling sSA-aspects. During the project runtime it became clear that the consortiums' experience with simulations is so much bigger than with modelling that modelling at the same detailed level as simulation was not yet possible for the consortium. This is just a matter of time. It simply takes a great deal of time (and effort) to design, and implement, complete models of pilots as well as complete scenarios. Because of that a practicable solution had to be identified. Relevant aspects of sSA were modelled instead of complete situations. During the simulation studies care was taken to create an overlap between the modelled and the simulated sSA aspects, in order to allow comparison afterwards.

Based upon the discussions that had taken place in WP4 (scenario development) it was decided that "looking for a certain amount of time at relevant objects" and "communication" are the most useful and measurable aspects of sSA. The model experiments were designed to demonstrate that uncertainty avoidance (UA) and power distance (PD) do influence sSA. Both are reflected in combinations of "looking for a certain amount of time at relevant objects" and of course "communication".

The modelling exercises concluded that one can say that high UA crews maintain better but not more sSA, however, when they do get into trouble sSA reduces. Low UA crews generally maintain less than optimal sSA, but the amount of sSA is not reduced when many problems arise. And also that the relationship between UA and sSA is not fixed, but varies with the other circumstances.

These exercises also gave the impression that low PD crews have more and better sSA when they get into trouble, which enables them to resolve a problem quicker than medium and high PD-crews can.

The effects of PD and UA as concluded are based on the assumptions that were necessary for the implementation. These assumptions are derived from the definitions of the psychological constructs. The results give therefore an indication of how UA and PD would work in different situations that airline crews could encounter. This means that for the evaluation of modelling in general, these experiments show that one can make predictions about behaviour when different airline crews encounter the same situations.

2.8.1.2 Simulation

The full-scale simulator studies have taught (amongst others) that a solid predictor of sSA, during the environmental problems, was a high PD. It was also found that a high PD between pilots is associated with a higher sSA. According to Hofstede’s theory on PD, we should expect western societies to score low on power distance. In contrast, Asian cultures are expected to score high on power distance. Collectivist high PD cultures tend to show great concern for the group, more harmonious relationships and deference to leaders (in contrast to autonomous functioning and personal gain). This means that the tendency to “level off” the power gradient in many western airlines may not be very desirable. Thus, based on the limited number of crews studied in the VINTHEC II project, it may be concluded that a high power distance can have favourable effects on sSA.

Further discussing the found results, we feel that the discovered relation between PD and sSA should be considered with some caution. Primarily, sSA can vary throughout an aircraft operation, whilst the indices of PD are relatively stable and remain the same. This means that there can never be a strong causal relation between power distance and sSA. However, it could be that power distance stimulates the occurrence of sSA.

2.8.1.3 Conclusion on simulation - modelling relationship

Cognitive modelling of complex psychological / interpersonal phenomena has a number added benefits to human-in-the-loop simulations:

- Computer-based scenarios, once constructed / programmed, can be run a huge number of times. Individual factors (e.g. power distance or cross cockpit authority gradient) can be manipulated easily and extensively (in many cases much easier than in human-in-the-loop scenarios) to assess their role in the creation or breakdown of the phenomenon.

- Another preferable aspect of fast time simulations (modelling) is that the costs are far less. Especially ones a scenario is created and pilot behaviour is defined, it's cheaper than human-in-the-loop simulations, to make numerous manipulations.
- There are no psychological costs of failure to human participants, since there are no human participants. This allows the experimenter to explore the boundaries of the phenomenon more fully without having to worry about consequences of breakdown to participants.
- Modelling provides the researcher with a more concrete, algorithmic description of the phenomenon. Such a more concrete description can be of enormous value when understanding of the phenomenon has to lead to systems development that aims to improve the phenomenon. For example, the development of an effective decision support systems is much easier on the basis of a concrete, modelled description of the phenomenon of decision making in some context, than on mere field observations of the phenomenon.
- Cognitively modelled environments are very useful when researching behaviour in an environment, where a great deal of the variables is understood. Complex but highly procedural tasks, like flying an aircraft, are examples of such an environment. However, since humans operate aircraft there will always be exceptions to the well-understood environment. In order to identify these so-called exceptions-to-the-rule, human-in-the-loop simulations and cognitive modelling form great supplements to other.

With respect to the current experiments, and sSA assessment, the following was concluded:

Above is described that the advantage of applying multiple methods to measure the same scientific concept is that convergent and divergent results may be found. The first making the scientific result more valid the second forcing the scientist to examine the data from a different perspective which may deliver new insights and an even better understanding of the reality. The comparison between human-in-the-loop simulation and cognitive modelling delivered both kind of results:

Convergent
In both the modelled and the simulated experiment the crew was disturbed immediately after take off. This was considered to be a high workload situation. It, indeed, turned out that both methods indicated that the disturbance took place during a high work situation.

Divergent
The interesting thing is that the conclusion from the simulation study about the high PD that seems to correlate well with a high sSA, seems to be the opposite of the conclusions from the modelling study. Both studies indicate that there is a relationship between PD and sSA, but that it is probably not fixed but depended on a number of other factors.

2.8.1.3.1 Over all

Because the above described added benefits of cognitive modelling, VINTHEC II recommends to include cognitive modelling in a standard sSA-assessment-battery. There are clear benefits. In theory as well as actually established during the VINTHEC II project. These benefits guarantee having a better, more detailed way of measuring, and especially understanding, sSA.

To a lesser extent, but still important, to realise is the fact that training and experience is needed for modelling as well as for simulation. For both it is possible to work with quick and dirty approaches, but for both it is recommended to operate as realistic as possible in order to obtain results that are indeed transferable to 'the real world'.

3 List of deliverables

In the reported period, the following documents were delivered:

Table 3.1: List of deliverables 1 April 2000- 31 July 2003

Description	Title	Delivery date/status	Responsible Party
Project web site	http://www.nlr.nl/projects/VINTHEC-II/	August 2003	NLR
WP0-PR 01	Six Months Management Report (1-4-00 until 31-9-00)	November 2000	NLR
WP0-PR 02	Twelve Months Progress Report (1-4-00 until 31-3-01)	May 2001	NLR
WP0-PR 03	Eighteen Months (mid-term) Progress Report (1-4-00 to 31-9-01)	August 2001	NLR
WP0-PR 04	Twenty-four Months Progress Report (1-4-01 to 31-3-02)	December 2002	NLR
WP0-PR 05	Thirty Months Management Report	November 2002	NLR
WP0-PR 06	Thirty-six Months Progress Report (1-4-02 until 31-3-03)	May 2003	NLR
WP0-PR 07	Forty Months Progress Report (1-4-03 until 31 July 2003)	2 months after final meeting	NLR
WP1-TR 01	State of the Art	January 2001	LiU
WP2-WD 01	Eye Point-of-Gaze (EPOG) as an indicator of shared situational awareness.	July 2001	NLR
WP2-TR 01	EPOG Ergonomic Guidelines	October 2002	RISØ
WP2-TR 02	Coding of Inferences from Visual and Other behavioural Data	February 2003	RISØ
WP3-TR 01	Document describing part 1 final results of WP3	April 2003	QinetiQ
WP3-TR 02	Document describing part 2 final results of WP3	June 2003	QinetiQ
WP3-TR 03	DAVIT - DAta VIzualisation Tool *(Software Tool + manual)*	June 2003	Technion
Additional	JAZZ latest version	July 2003	IBIB
WP4-TR 01	Scenario definition	January 2003	FOI
TU-02-07	Technical Note: VINTHEC II - Scenarios and Measurement Approach	May 2002	TU (subcontractor of NLR)
WP5-TR 01	Document describing final results of WP5	June 2003	BAE
WP6-WD 01	The effects of uncertainty avoidance and Power Distance on Shared Situation Awareness in the cockpit. *An IPME simulation study.*	June 2003	NLR
WP6-TR 01	Document describing final results of WP6	July 2003	LiU
WP7-WD 01	Pilot briefing guide for the RFS experiment at NLR	March 2002	NLR
	CBIS - Crew Briefing and Information System	April 2002	NLR
	Radio scripts LHR-AMS and AMS-LHR	April 2002	NLR
WP7-WD 02	PAT measures analysis	July 2003	Technion
WP7-WD 03	Task analysis - Interviews with El-Al pilots	July 2003	Technion
WP7-WD 04	Data Acquisition and Reduction of Peripheral Arterial Tone (PAT) Measurements	July 2003	Technion
WP7-TR 01	Full-scale simulator experiment	December 2003	NLR
WP8-TR 01	Final report	December 2003	NLR
WP8-TR 02	Engineering design guidelines and recommendations	December 2003	NLR

4 Comparison of initially planned activities and work actually accomplished

4.1 Time delay

As can be seen in different progress reports the VINTHEC II project is from the beginning hindered by a delay that was never made up for. As a result most of the deliverable were published later than scheduled. The main cause for these delays is the fact that partners tended to extend their research, remained examine the data after the 'deadline' was passed. This has delivered some new insights, especially at the beginning of the project.

Care was taken that the above described delays would not affect the further progress of the project by making sure that the information that was gathered was available to all partners, on the VINTHEC II website, as soon as it was drafted. This procedure guaranteed that each workpackage could start on time and could use the input from other workpackages exactly as scheduled in Annex A to the contract.

At the end of the project the consortium feels that there is still a wealth of information available in the data that were gathered and that these data in fact deserve a great deal more attention than the consortium has given. The consortium is confident that they set their priorities right and researched the most useful issues first, but at the same time the consortium is quite curious to see what further analysis could have revealed.

4.2 Switching roles (and names) of partners

During the runtime of the project several changes in staffing of the project. Most important are:

1. The Technion staff, and also the area's of expertise had changed totally when the original staff has left the Technion. As a result a totally new plan was made for the involvement of Technion in the project. Thereby the role of Technion became broader, they were involved in more aspects of the work, since that new involvement fitted better with the expertise of the new Technion staff. No need to say that this also made it more difficult to fit al the Technion work in the work. Looking at it backward it turned out that the consortium has managed quite well to do so.
2. The company "DERA" commercialised and changed it's name into "QinetiQ"
3. The name "British Aerospace" was changed into "BAE SYSTEMS"
4. The name "FOA" was changed into "FOI"
5. After about two years Rolf Zon took over VINTHEC II co-ordination from Brian Hilburn.

4.3 Additional in deliverables

Besides the deliverables that were scheduled at the beginning of the project a number of additional deliverables were created.

In WP3 an additional report (WP3-TR-02) was published because the originally scheduled work was split in separate tasks. As a side effect of one of these tasks the JAZZ (a hardware tool invented by IBIB) was further developed. This development was also not a part of the original plan, but definitely a useful development which may result in a product that actually can be brought on the market.

Another WP3 deliverable, that also was not scheduled is DAVIT (DAta VIsualization Tool). Technion took the initiative to this development, though a number of partners have contributed experiences and ideas. This software tool was also not scheduled as a VINTHEC II deliverable. It turned out to be a very useful tool for the consortium. Even a manual (or guide) is written so that it can be brought to the market as well.

For the JAZZ as well as DAVIT the consortium still needs to decide about how to continue with these developments and if / how / under what conditions, to bring them to the market.

The Technion contributions to the project, and especially to WP7, were not originally scheduled. They turned out to be complete reports on their own. One about a task analysis and two about the Peripheral Arterial Tone (PAT) which is a measure that was introduced to the project by Technion. These three reports are published within the consortium as so-called working documents. A discussion with Technion and partners should determine whether these reports should become part of the WP7 deliverable, or that they deserve to be published as stand alone additional reports.

For more information about the JAZZ, and its successor the HAG, the reader is referred to Appendices A and B of the current document.

4.4 Expert User Groups

In Annex A to the contract it was anticipated that two Expert User Groups (EUG) one in WP4 and one in WP8 would be organised so that the consortium could get some feed back from the outside world upon their plans ideas and results. This feed back would be a kind of guarantee that the end results of VINTHEC II would also be relevant to others (outside the VINTHEC II consortium and even from other domains. The first EUG was organised. The results recorded on CD and distributed over the partners. Even though it was an interesting discussion the consortium felt that instead of a second EUG after all data were gathered and analysed, it would be more appropriate to have a meeting with the pilots who had participated as subjects in the VINTHEC II full-scale simulator experiment. Such a meeting would allow the partners to evaluate whether the conclusions that they had drawn so far, would fit with the daily practise of pilots and airlines. The consortium was pleased to hear that Brussels has accepted this request. Therefore the second EUG was replaced by a workshop where all the pilots were present and where all the partners had made notes so that they could improve their own analyses and reports before submitting them as contributions to the WP7 deliverable.

5 Management and co-ordination aspects

5.1 Contact information

The NLR offers to point persons who are interested in VINTHEC II results or follow up projects in the right direction. For any VINTHEC II related question it is therefore recommended to contact:

G.D.R. Zon (VINTHEC II Project Co-ordinator)

Nationaal Lucht- en Ruimtevaartlaboratorium (NLR)
Human Factors Department
Visitors: A. Fokkerweg 2, 1059 CM Amsterdam, The Netherlands
Mail: P.O. Box 90502, 1006 BM Amsterdam, The Netherlands

Phone: +31 (0)20 - 511 31 90
Fax: +31 (0)20 - 511 32 10
E-mail: zon@nlr.nl

Or contact other representatives from the NLR "Human Factors" department.

Additional information will be available on the VINTHEC II web site at: http://www.nlr.nl/public/hosted-sites/vinthec2/.

5.2 Intellectual Property Rights and confidentiality

In fact IPR issues are covered already in the Technical Implementation Plan (TIP) just for completeness of this final report the main IPR-related and confidentiality-related issues will be summarised here.

During the final meeting decisions were be made about the confidentiality of each VINTHEC II deliverable. It was decided that all technical deliverables will become public if non of the partners objects within one month after the final meeting. Since non of the partners has used this opportunity all VINTHEC II documents can be considered available for the general public.

VINTHEC II resulted in two deliverables that are not just documents but actually tools that may be interesting to exploit. These are the JAZZ (a hardware tool made by IBIB) and DAVIT (a software tool initiated by Technion with a great deal of input from partners). Both of these tools may be brought on the market directly, or with minor fine tuning effort. The IPR and confidentiality aspects related to these issues are not settled yet. The main developers of these products will take care of IPR issues themselves, of course in close co-operation with the VINTHEC II partners who have contributed to these products.

5.3 Roles of partners

All partners contributed their share. Even though minor misunderstandings have risen during the project runtime in general there was always the spirit that pushed the consortium towards finding compromises and via mutual understanding solving all problems.

Each partner clearly had it's own role, or speciality, which makes that the same result could not been achieved in the way if one or more partners could not have participated.

The motivation, performance and dedicating were all up to the level that one may expect.

From the perspective of a co-ordinator it was sometimes difficult to acquire all that is needed (contributions to progress reports, deliverables, presence at meetings, and sometimes even cost statements) at the desired moments. Given the fact that all partners, including the co-ordinator, do have other projects and responsibilities to pay attention to as well, this all very understandable. Eventually it has all worked out fine. From a management perspective the consortium can be satisfied.

The consortium consisting of eighth partners is a group that according to the co-ordinator is 'overseeable'. The physical distances between partners made that the desire to visit one another in order to work more efficiently together, could not always be satisfied. The project web site and mailing list are considered effective tools to stay in touch and all be well informed. The consortium has, for future projects, high expectations of internet-meeting tools.

5.4 Consortium composition

		Business activity
NLR	NL	Aerospace Research Institute
BAE	UK	Manufacturer of aerospace
QinetiQ	UK	Research Institute
FOI	SE	Defence Research Institute
LiU	SE	University
RISØ	DK	Research Institute
Technion	IL	University, Dept Aerospace Engineering
IBIB	PL	University / Laboratory

Below will follow a detailed description of each partner's role (written by that particular partner):

5.4.1 NLR

The National Aerospace Laboratory (NLR) is the central institute for aerospace research in the Netherlands. Its principal mission is to render scientific support and technical assistance on a non-profit basis to Dutch and foreign aerospace industries and organisations, civil and military aircraft operators, and governmental agencies concerned with aviation and space flight.

The NLR's "human factors" department provided most of NLR's contributions to VINTHEC II. Fields of expertise include psychology, psychophysiology, engineering and domain specific expertise on ATC, flight deck and military operation. Relevant research topics include: impact of new technology and mission and task analysis, displays and controls design; human centred automation concepts; operator performance and workload; simulation and training; and human factors cockpit evaluation.

Other NLR departments who have contributed to VINTHEC II are the "simulation and training" department, the "flight testing and safety" department, the "data and knowledge systems" department and the "instrumentation" department.

The NLR has participated in all the workpackages except for WP3 (data analysis methodology).

The NLR has co-ordinated the VINTHEC II project. This co-ordination task comprised planning, making sure that the project remained on the right track, sharing the work between partners, facilitating and reporting, intermediate between the consortium and the European Commission are a number of the major tasks of the co-ordinator. Further web sites were build, and maintained, to facilitate the project, to aid the collaboration between partners, and to make sure that all partners had always access to the latest versions of all documents, and of course to present the project to the rest of the world.

A number of small scale studies were performed to develop, or fine tune, techniques that were eventually applied in the full scale simulation studies. An example is the master thesis: A. Blechko, B.G. Hilburn and G.D.R. Zon (2001) - *Eye Point-of-Gaze (EPOG) as an indicator of shared situational awareness. Evidence from simulated flight task*. In which two sets of eye tracking equipment were running simultaneously, recording the behaviour of two subjects who were both working on the same task.

The NLR has, especially for the VINTHEC II project, bought IPME (a software environment that enables cognitive modelling). In this environment it has modelled crew behaviour during scenarios that are comparable to the scenarios that were applied in the full-scale simulator study. An example is the master thesis: J.R. Kuipers (2003) - *The effects of Uncertainty Avoidance and Power Distance on Shared Situational Awareness in the cockpit. An IPME simulation study*. In which, based upon the events that would be simulated in the human-in-the-loop studies, the effects of "uncertainty avoidance" and "power distance" were modelled.

Further the NLR has provided a great deal of expert knowledge about; measurement methodology, simulation, and daily practise of commercial flying via NLR staff but also via the network of relations that the NLR has.

The full-scale simulator experiment was co-ordinated by the NLR, in close co-operation with all the partners. The NLR has provided a simulator environment for the full-scale simulator experiment including eye-tracking equipment (the GazeTrackers), video recording and loggings of simulator output. The NLR has enabled partners to build their own measurement equipment in the NLR simulator. In order to facilitate that work the NLR has provided all partners with technological support so all partners knew how to install and connect their equipment in the simulator and what the output format of the data was like. For the full-scale simulator experiment NLR has applied (and even developed) a great deal of measurements.

5.4.2 BAE

BAE SYSTEMS, formed by the merger of British Aerospace and Marconi Electronic Systems, has become Europe’s largest producer of aerospace products, and holds a 20% stake in the European Airbus company. Employing approximately 90,000 in 7 EU Member States, BAE SYSTEMS enjoys a proud heritage in the civil aviation market, producing some of the best known and most successful aircraft. In addition BAE SYSTEMS has a long history of designing and producing high quality electronic systems through continuous innovation in microwave, television and radar technology. BAE SYSTEMS has a history of effective collaboration with European partners and its size and market share make it an influential force in determining International Standards.

It is company policy to promote research and development across a wide range of disciplines and to ensure the successful commercial application of results, especially within Europe as its largest market sector.

BAE SYSTEMS Advanced Technology Centre is the corporate research centre for BAE SYSTEMS. It employs 420 staff and carries out long term industrially relevant research for the BAE SYSTEMS business units. It provides innovation, expertise and services for the creation and development of future technology capability covering a broad spectrum. From hardware to software development, composite materials to stealth techniques, sensing systems, to synthetic environments, the ATC provides innovative solutions to customer requirements.

BAE SYSTEMS' Sowerby Research Centre (SRC) is the corporate research organisation for BAe and provides a research service to most of the operating companies. SRC currently employs around 175 highly qualified scientists and engineers, and is divided into six major Departments, of which one, the Human Factors Department, was involved in this project.

The Human Factors Department has at present 22 full-time staff who are professionally qualified in either psychology, ergonomics or, in a few cases, other disciplines such as computer science. The department has over 25 years of experience in applying human factors to the design and evaluation of human-machine systems, and of research into aspects of human performance. This experience covers many topics in the field of human factors (e.g. control/display design, vision modelling, simulation, mental workload, and human-computer interaction), and relates to a wide range of systems and equipment (e.g. aircraft cockpits and command and control systems).

SRC contributed to the project expertise in EPOG, EPOG recording equipment and use of a part-task (A320) simulator for pre-testing activities. It also contributed its knowledge of EPOG recording systems, systems design, software development, and optics and electronics.

5.4.3 QinetiQ

During the life of the VINTHEC II project this partner organisation changed its name and commercial status. Initially it was DERA (Defence Evaluation and Research Agency) and was an Agency of the UK Ministry of Defence. In July 20001 it became QinetiQ Plc, a commercial technology solutions organisation. The impact this has had on the VINTHEC II programme has been minimal – and the role and deliverables for this partner have remained unchanged.

QinetiQ is a large UK technology organisation with staff numbers of approximately 7,000. The majority of its work is conducted for the UK Ministry of Defence in the form of fundamental (far-sighted) research, applied research (related to existing or near-future military systems) and project support (research and consultancy relating to current military issues). In addition, there is an increasing emphasis on growing business in the commercial sector, and this has been the greatest change to the organisation over the life of the VINTHEC II programme.

The QinetiQ personnel who have contributed most to the VINTHEC II work programme are based in the Centre for Human Sciences (CHS). CHS is a human factors research and consultancy group and comprises approximately 330 staff, the majority are specialists in either psychology, ergonomics, or physiology. The core VINTHEC II team have experience in human factors and occupational psychology consultancy and research in military and commercial aviation, and in the wider application of human factors to other domains – such as education, airport security, military systems, and other transport environments.

Like many of the VINTHEC II partners, QinetiQ personnel are multi-disciplinary and multi-skilled. QinetiQ (then DERA) was a partner in the VINTHEC (I) programme and drew on its experience from its role (specifically data visualisation) in that programme for VINTHEC II. In VINTHEC II QinetiQ has provided specific expertise in the areas of technology review of commercial eye tracking equipment, user requirements capture for the VINTHEC II methodology, software expertise for the Visualisation tool development, led Technion, and methodological and analytical input for the use of subjective measures in the VINTHEC II methodology.

5.4.4 FOI

The main contribution of FOI to the project has been the generation of operational scenarios, selection of practicable psychological and psychophysiological measures, and integration of results. FOI specific experience in assessment of operative performance, in team performance and team processes, in data based modelling, and in development of scenarios was of importance. FOI co-operation with TFHS (Lund University School of Aviation, Sweden) was of special importance for VINTHEC II. It provides experienced instructors as well as trainees, and VINTHEC II results on training was exploited and tested in the school.

5.4.5 LiU

Not received yet.

5.4.6 RISØ

Risoe has contributed to describe the state of the art in shared Situational Awareness (sSA), and the further development of the concept. We have performed an extensive literature research on SA and sSA and this has been used in state-of-art survey deliverable and initial definition of sSA. In addition, Risoe has contributed to the development of novel methods for analysing shared situational awareness based on visual and verbal data. Risoe has tested some of these measures of subjects' awareness of crew-members via part-task experiments and development studies. This work was also contributed to the development of scenario specification by ensuring that the measure relevant behaviours would be available for analysis. Thirdly Risoe has contributed to the development of exploration and visualisation techniques related to sSA through the development of the Risoe Data Analyser and through a feasibility study of the JAZZ EPOG equipment developed by IBIB. Moreover, Risoe has tested and validated a number of the developed measures for analysing the visual and verbal data from the final flight simulator study carried out at the NLR facilities. Regarding EPOG data, they have explored measures of sSA in terms of whether the pilots applied a distributed or a focused visual attention strategy. This measure shares resemblance with the in the eye-tracking methodology discussed notion of “scanning entropy”. Moreover, they have measured sSA by way of lag sequential analysis looking for “hidden” patterns in the data. Furthermore, they measured sSA in terms of mutual eye point of gaze and dynamic visual elements. That is, elements that moves dynamically in the cockpit and therefore cannot be captured by the automatic data recordings of the eye-movement tracking system. With respect to verbal measures Risoe has explored three communication parameters: Number of speech acts (few/many), homogeneous communication patterns (similar/different), and temporal aspect of communication (past, present or future tense). Finally, Risoe has fed information about the work described above into the final report and the engineering design guidelines and recommendation report. Risoe was work package leader of WP2.

5.4.7 Technion

Not received yet.

5.4.8 IBIB

IBIB has brought to the VINTHEC II project the innovative concept of monitoring the conscious attention involvement in three different domains.

1. Real time interaction with the environment requiring maximum visual attention – the *exploration mode*.
2. Intermittent involvement of attention engaged on request from the environment when the environment does not follow the model, as well as engaging the attention to the isolated aspect of the real environment requested by the thought process – *the monitoring mode*.
3. Visual in-attention resulting from exclusive engaging of attention in the thought process, logical evaluation of the environment signals and knowledge based problem solving – *the planning mode*. In the broad sense it means also retrospection and relates to the situation when the attention is involved in processing of the information already available in the system.

The above hypothesis was experimentally evaluated by the RISØ group lead by dr. Hans Andersen which carried the feasibility study of JAZZ for systematic detection of oculomotor – saccadic behaviour which can be used for objective monitoring of the visual attention engagement.

Answering the question in which way the global measures of oculomotor behaviour systematically recognised by JAZZ hardware and analysis software can be used for mental workload studies requires further research. The JAZZ multi-system proved to be the appropriate research tool, which due to its minimal intrusiveness and good toleration by pilots can open the possibility for performing this kind of research during the real flight. The JAZZ multi-sensor system during the course of the VINTHEC II experiments, was modified and adopted to the requirements of synchronising it's operation with events generated by the computer controlling the simulator experiments.

5.5 Financial overview

Information about finances and effort that was spent, was removed for this *public* version of the document.

5.6 Man-months overview

Information about finances and effort that was spent, was removed for this *public* version of the document.

6 Results and conclusions

The overall results and conclusions of the project will be briefly summarised, for more details the reader is referred to Chapter 2 of this document. The results and conclusions were split into planned (the objectives of the project) and additional (not scheduled but still interesting outcome of the project). In general we can say that the consortium did manage to find very interesting approaches to sSA measurement. They found that, as such, the main objective of the project is accomplished. Of course the partners had hoped to create a marketable sSA measurement battery. Due to the complexity of the subject matter this turned out to be not possible within the VINTHEC II timeframe, though technically speaking it seems possible. Therefore the consortium has high hopes and expectations for the coming years.

6.1 Planned

Objective 1: Detailed assessment methods.
The consortium managed to evaluate a set of measures. For all of those measures the advantages and disadvantages with respect to sSA assessment were established. This wealth of knowledge and experience enables us to measure relevant aspects of sSA in a number of situations.

One of the technical risks that was identified at the beginning of the project were the huge individual differences between pilots. These differences were found indeed. The cognitive modelling work that was done in WP6 turned out to be an interesting tool to study large amounts of data without these variations between subjects.

Due to unforeseen difficulties and time constraints the consortium did not manage to create one tool that measures sSA in a kind of standardised sSA units. Of course such an outcome would have been fantastic. But still the consortium is able to make detailed statements about the sSA of a particular crew in a number of settings. Therefore the consortium judges the overall outcome of the work as very positive and useful.

The methodology that was applied by the consortium can be used in numerous settings ranging from training, display / flightdeck systems design, Human Factors certification, but also outside the aviation domain the methodology will be a great help when sSA assessment is relevant.

Objectives 2, 3 and 4: Procedural guidelines, implementation and cognitive engineering recommendations.
A number of workpackages have taught us a great deal about measurement, analysis and experimental design techniques (scenario development). These experiences are brought together in deliverable "VINTHEC II WP8 - TR - 02". Researchers who are planning to do research on sSA can use this guide as kind of reference to prevent them from pitfalls that the VINTHEC II consortium has encountered. Further it can provide advice on the practical implementation of the objective assessment methodology, including specification of operational scenarios for system evaluation purposes.

Objectives 5 and 6: Improved aviation safety.

Is an objective that still stands but at this moment it will be hard to prove that this particular project has contributed to that yet. There is no doubt about the fact that sSA, or less than optimal sSA, contributes to aircraft accidents. So by understanding how sSA builds up, and by being able to identify situations during which sSA decreases, one can contribute to aviation safety. The consortium is confident that (aspects of) the VINTHEC II sSA assessment methodology will be used in a broader context and that the knowledge and insight that will be gained by such research will contribute to safer situations in aviation, but in other domains as well.

Especially since automation is becoming more competent, and assuming the role of an additional “crewmember” on the modern flightdeck, the need for techniques to assess sSA and co-ordination between crewmembers (including between crewmembers and automated flightdeck systems) has increased in order to improve aviation safety. The VINTHEC II project is a valuable contribution to these improved sSA assessment techniques.

The knowledge and experience that results from VINTHEC II can be perfect input to aviation safety improvement.

Objective 7: Better (cost effective) design processes.
The added value of cognitive modelling turned out to be very promising. Even though it is obvious that the consortium needs more experience on this area of expertise in order to receive the full profit from this technique, the partners are confined that a great deal of information that normally requires numerous subjects to obtain, can now be gathered without any subject. This finding may not be interpreted as if cognitive modelling will be able to replace human-in-the-loop simulations. It merely means that both techniques can be great supplements to each other, and that emphasis with human-in-the-loop simulations may shift towards individual differences and exceptions to the rule rather than towards statistically significant data gathering analysis.

6.2 Additional

On top of the objectives that were determined in the proposal phase of the project there were also additional side effects that are worth mentioning as results.

The development of the software tool DAVIT.
This tool runs under Microsoft Windows and helps researchers to quickly gain insight is their data. It is a generic tool, for the current project Eye Point-of-Gaze-, psychophysiological and simulator data were analysed in DAVID. All data sources can be displayed in a synchronised way. Even though DAVID does not perform detailed analysis it is very helpful in acquiring a quick overview before the real analysis starts. Initially the tool was developed so that it could only be used when researching data coming from simulators that are in service at the VINTHEC II partners. Now the tool is quite flexible and can be used to analyse data coming from numerous sources and simulators.

The further development of the hardware tool JAZZ.
The first version of this hardware tool was brought into the project by IBIB as background knowledge. During the project runtime this tool was further developed and fine tuned in close co-operation with other VINTHEC II partners. Due to the VINTHEC II project the tool is now more marketable.

7 Acknowledgements

The work that was done in the project could never have been accomplished by any of the partners individually. Therefore we would like to thank all the partners for their participation in the project. We also would like to thank all individuals…
From NLR:
Adri Marsman, Arun Karwal, Bart Heesbeen, Brian Hilburn, Cees Krikke, Eamonn Hanson, Eric Kruijsen, Eric Vermeulen, Gerrit Oort, Hans Brugman, Jacco Dominicus, Jan Willem van der Laan, Martin Joosse, Paul Kuiper, Pieter Teeuwen, Rob Egter van Wissekerke;
From BAE:

Air2000 airline, Kelvin Adams (Air2000), C. Pritchard, A. Leggatt, M. Williams, B. McGuinness, K. A. Davies;
From QinetiQ:
Dr Ken Catchpole, Dr Peter Naish, Dr Paul Enterkin, Dr Dale Richards, Dr Helen Dudfield, Lucy Emery, Audrey, Caldeira-Hankey;
From FOI:
E.A.I. Svensson, J. Alfredson
From Linköping University:
Sidney Dekker, Kjell Ohlsson, Rogier Woltjer;
From RISØ:
Henning B. Andersen, Gunnar Hauland, Rud C. Pedersen;
We are very grateful to the pilots (to whom we have promised to keep their names out of any contributions) who have participated as subjects in the full-scale experiments and all really helped us a great deal!!! Further we would like to thank Heralt Mooij for his contributions and advises. Especially we would like to thank the Directorate-General for Research of the Commission of the European Communities and all the scientific officers (Doris Schröcker, Koen Liem, Jonas Forsman, and Joseph Prieur) who have guided us.

8 Abbreviations and acronyms

ATC	Air Traffic Control
BIOSIS	Bibliographic references for life sciences research
CRM	Crew Resource Management
EPOG	Eye Point-of-Gaze
EUG	Expert User Group
FOI	Swedish Defence Research Establishment - FOI (SE) (*formerly FOA*)
HF	Human Factors
IBIB	Institute of Biocybernetics & BioMedical Engineering of the Polish Academy of Sciences (PL)
INSPEC	Institution of Electrical Engineers
IPME	Integrated Performance Modelling Environment
LiU	Linköping University (SE)
LOFT	Line Oriented Flight Training
MM	Man-Month
NASA	National Aeronautics and Space Administration
NLR	National Aerospace Laboratory (NL)
NTIS	National Technical Information Service
PAT	Peripheral Arterial Tone
PD	Power Distance
SA	Situational Awareness
SME	Subject Matter Expert
sSA	Shared Situational Awareness
SUN	Statement of User Needs
TSA	Team Situation Awareness
UA	Uncertainty Avoidance
VINTHEC II	Visual INTeraction and Human Effectiveness in the Cockpit, II
WP	Work Package

9 References

Final Technical Annex 1 “Description of Work”, March 2000
“Guidelines for reporting” FP 5 GROWTH Programme, Electronic version, 1st edition – January 2000
VINTHEC consortium (1996). Technical Report: VINTHEC - WP1 - TR 01 *Human Awareness Criteria.*
VINTHEC consortium (1997). Technical Report: VINTHEC - WP3 - TR 01, *Dynamic Measures of PMWL, PP and, SA.* NLR Amsterdam, NLR-TR-97353 L.

VINTHEC consortium (1999). Technical report: VINTHEC - WP8 - TR 01, *Full-Scale Studies*. NLR Amsterdam, NLR-TR-99404.

VINTHEC II Consortium (2000). Project Plan version 9.2. *Annex 1 "Description of Work"*.

VINTHEC II Consortium (2001). Technical Report: VINTHEC II-WP1-TR 01, *State of the Art*. NLR Amsterdam, NLR-TR-2000-665.

VINTHEC II Consortium (2001). Technical Report: VINTHEC II-WP2-TR 01, *EPOG ergonomic guidelines*. NLR Amsterdam, NLR-CR-2002-508.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP2-TR 02, *Coding and inferences from visual and other behavioural data*. NLR Amsterdam, NLR-CR-2003-098.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP3-TR 01, *User Needs, Knowledge Elicitation and Software Tools*. NLR Amsterdam, NLR-CR-2003-148.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP3-TR 02, *Feasibility Study of IBIB Online Prototype and Visualisation and Exploratory Data Analysis*. NLR Amsterdam, NLR-CR-2003-429.

VINTHEC II Consortium (2002). Technical Report: VINTHEC II-WP4-TR 01, *Scenario definition*. NLR Amsterdam, NLR-CR-2002-695.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP5-TR 01, *Pre-testing study*. NLR Amsterdam, NLR-CR-2003-443.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP6-TR 01, *Modelling*. NLR Amsterdam, NLR-CR-2003-350.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP7-TR 01, *Full-scale simulator experiment*. NLR Amsterdam, NLR-CR-2004-029.

VINTHEC II Consortium (2003). Technical Report: VINTHEC II-WP8-TR 02, *Engineering Design Guidelines and Recommendations*. NLR Amsterdam, NLR-CR-2004-035.

Appendix A - *Jazz* multisensor system

The *Jazz* is the **multisensor system** that allows the acquisition of the eye movements with excellent spatial and temporal resolution, together with other physiological and environmental signals. The main idea behind using *Jazz* multisensor is to gather different kinds of information about pilot's interaction in the cockpit, using a single device. The physiological signals measured by *Jazz* include:

1. eye movements in horizontal and vertical axis (1000 Hz sampling frequency),
2. head acceleration in horizontal and vertical axis (250 Hz sampling frequency),
3. photoplethysmographic signal in two lengths of the light wave (250 Hz sampling frequency)
4. ambient illumination (250 Hz sampling frequency),
5. voice recording (4000 Hz sampling frequency).

Signals measurement technique

For the eye movement measurement the Jazz system utilises the Cyclops-ODS technology (infrared oculography - IRO) optimised for easy set-up and minimal intrusiveness, which is crucial for monitoring of the subject behaviour in non-laboratory environment. Unlike most IRO setups, which place the infrared emitters/sensors around, or even next to the eye, Cyclops-ODS's set of optoelectronic sensors is placed between the eyes, hiding the sensor in the "shadow" of the nose. Thanks to such a setup, limitation of the visual field is minimal, reducing the risk of interference with subject's visual exploration of the environment. The eye movement measurement performed with high temporal and spatial resolution allows precise detection of saccades — fast eye movements used to move the point of gaze around the available field of view. Statistical processing of detected saccades over selected periods of time (their quantity, amplitude, duration of preceding fixations) provides important information about pilot's visual attention during different flight phases.

Head acceleration measurement allows detection of the head movements connected with visual exploration of the cockpit environment.

The photoplethysmographic signals measured by Jazz system allow evaluating subject's heartbeat and blood oxygenation. As these signals carry the information about the vasodilatory response of the sympathetic system, this measurement can be also used to access subject's workload.

Voice recording carries information about experimenter's/subject's comments, audio signal in environment, communication in the cockpit.

Ambient illumination can be use to document amount of light in the environment and differentiate between the light source (natural or artificial).

Area of application and impact on European Community

The feasibility study of Jazz Multi-Sensor System carried by the RISOE group under the leadership of dr Hans Andersen found that Jazz can be used for systematic detection of specific oculomotor behaviors which can be linked with the current status of the visual attention engagement. The Jazz is a complex psycho-physiological measurement system which operation is controlled by several microprocessors. It is system which releases the experimenter from taking care of the measurement technicalities which usually requires technically oriented personnel to setup the system and check it's performance.

The Jazz multi-sensor microprocessor technology makes these setting and checking tasks invisible for the experimenter. There is no necessity on setting and checking the system operation. Only the simple calibration – scaling experiment requires experimenter's participation. The Jazz technology can be considered as a milestone in automation of psycho-physiological measurements, of which the main objective is to monitor the engagement of visual attention, or attention in general, into the process of the human interaction with the environment.

The most important aspect of visual attention engagement is the possibility to observe the process of learning which can range from simple, skill based interaction with environment to complex knowledge based solving of unexpected novelty situations. The learning, acquiring skills and real time problem solving are the essence of operator interaction with complex technical system. Just to name few of them: airliner cockpit, captain bridge on the ship, interaction of anesthesiologist team within the surgery theater, application in advanced military training as well as learning how children learn and what are the individual measures of the learning process which decide about success or failure in school.

Jazz multi-sensor system has enormous practical potential and because of it's simplicity to use, the advanced signal processing algorithms and efficient software as well as it's minimum intrusiveness. Monitoring using the Jazz system does not disturb the process of interaction between operator and controlled environment or the pupil computer aided learning.

It can be foreseen that the Jazz multi-sensor and VINTHEC II methodology, developed for the advanced studying of pilots shared situation awareness can, revolutionise in the near future our understanding of learning in general and in particular making the learning more physiological, cognitive based. What ever we are capable to do first we need to learn it.

Appendix B - Head Area of Gaze (HAG)

(*An instrument for pilots training based on knowledge of AOIs attended by pilots*.)

The EPOG methodology has been developed for precise measurements and recording all data. The system is for quantitative analysis of points of gaze. Instructor needs simple qualitative information about which AOI is attended by a pilot. Instructor has to interact during a test flight, he should not be destructed by a moving cross on a moving scene which is usual for EPOG system. Since it causes a sensory conflict a subjective feeling is that it looks awful which is very mild response for such a conflict. It might happen that an observer will experience nausea and headache. Certainly the trainer will not accept it.

We need to optimise the human factor during pilot courses. It is also important that pilots need to be equipped with less intrusive monitoring system than the EPOG instrumentation used today. The time required for setting up and calibration the EPOG system is too long. In pilots training this will cost extra money since the training on simulation is charged on number of hours that a pilot spends in the cockpit.

Moreover today EPOG system requires well trained engineers to set it up as well as to calibrate whole system before an experiment/training. For many potential users of the VINTHEC for sSA (Shared Situation Awareness) training methodology the cost of the EPOG system may be prohibitive.

The EPOG methodology can be also used in training of operators of other complex technical systems in military as well as in civil domains. It can be exploited outside pilot cockpit.

All above mentioned aspects of the EPOG system can destruct an implementation of the VINTHEC developed methodology in pilot training environments. Preliminary experiences with a presentation of the EPOG methodology to BA instructors carried at BAE (British Aerospace) systems should warn us about the complexity of this issue and pitfalls of the EPOG methodology in training.

Introduction of the EPOG system in pilot training will also require training of instructors to learn and understand advantages and limitations of the new training. Based on the today instructor experience and their limited knowledge about psychological-cognition aspects of the EPOG it is natural to expect that they will not accept it without complete understanding of its benefits. A simple demonstration with switching off the voice channel between the cockpit and instructor can be helpful in proving the usefulness of the EPOG information. It is used for knowing what pilots are doing and understanding and appreciate a new information about pilots behaviour which the EPOG system makes accessible for trainers.

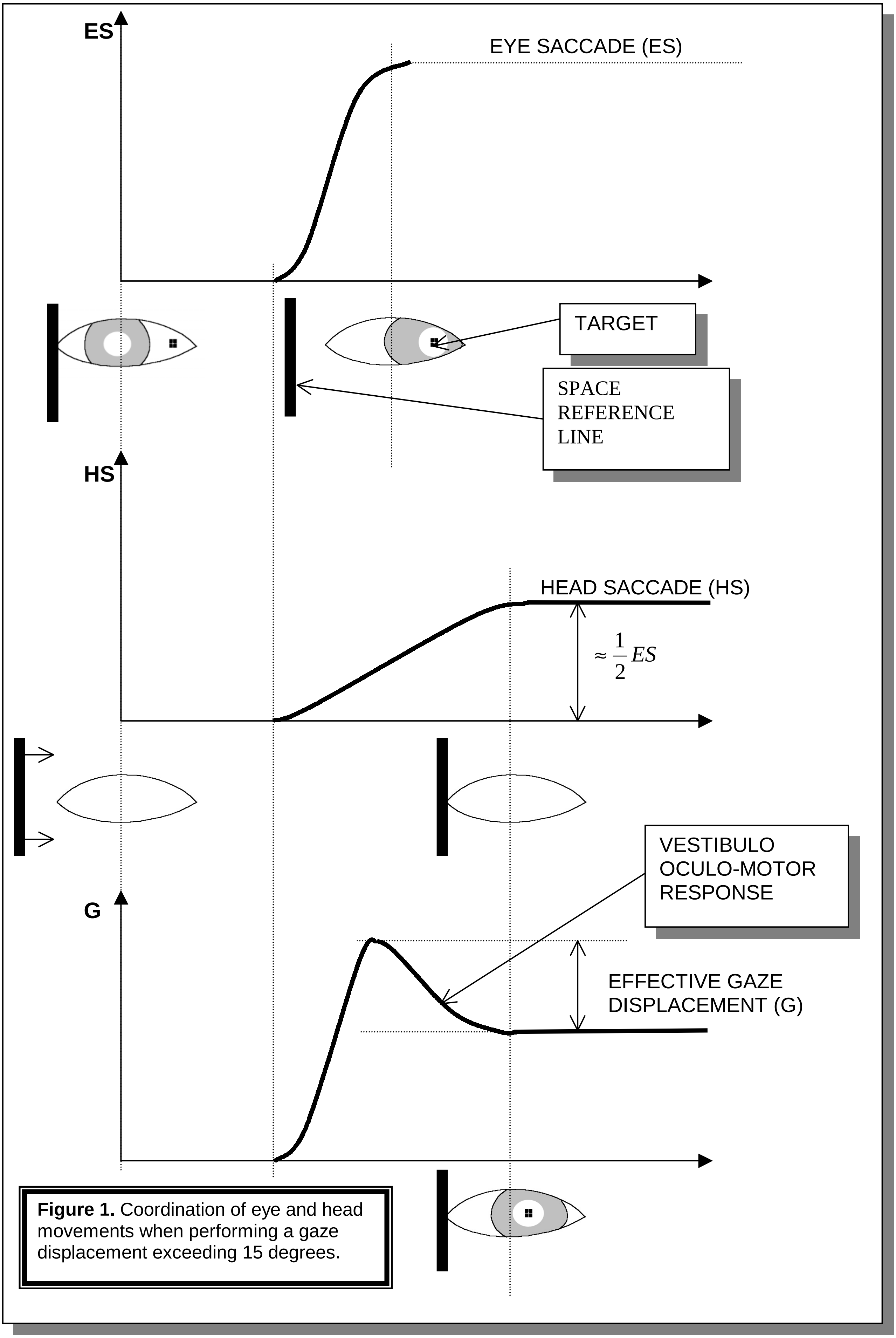


Figure 1. Coordination of eye and head movements when performing a gaze displacement exceeding 15 degrees.

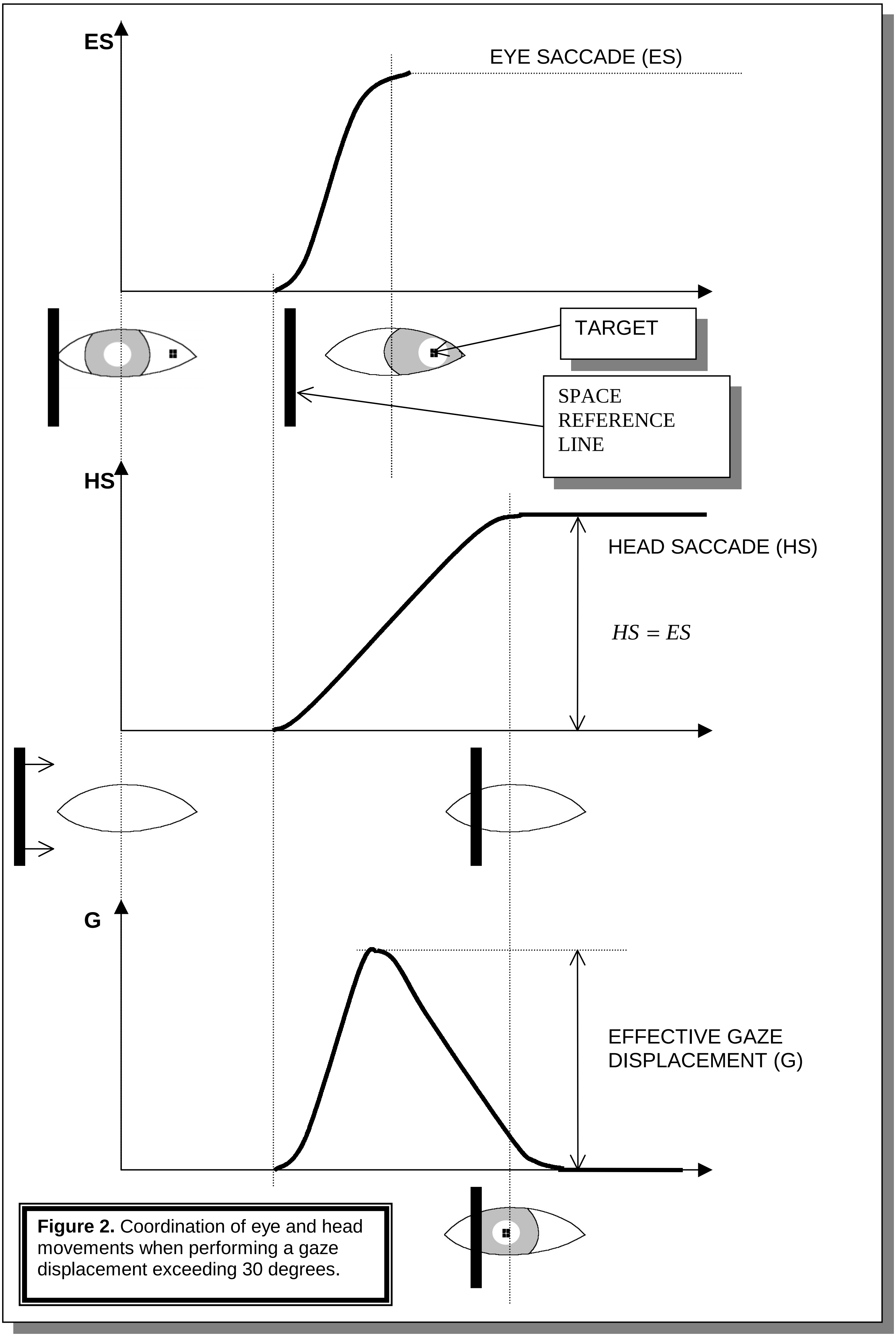


Figure 2. Coordination of eye and head movements when performing a gaze displacement exceeding 30 degrees.

HAG philosophy.

HAG is the shortening of Head Area of Gaze.

The basic idea of it is to replace high resolution points of gaze with low resolution areas of gaze. The second key idea is to substitute the measurement of eye position in relation to a head with direct monitoring of a head position in relation to the cockpit. There is a well known physiological phenomenon called orientation response which is manifested by turning the head toward the source of information. The orientation response in humans is linked with a redirection of gaze. It means that turning the head toward something is accompanied with saccadic eye movements toward some particular location.

When the extent of a gaze displacement is more then 15 degrees the eye saccade is accompanied with the rotation of the head in the same direction. There are two known strategies of such combined eye saccadic movement with a head displacement. Most of humans are assuming preferably only saccadic movements and 20% invoke head movements for all gaze displacements above 15 degrees. When the extent a gaze displacement exceeds 30 degrees all of us are using the second strategy: we rotate our head toward the area of interest (AOI).

On figure 1 there is a schematic representation of how the head and eye contribute to the gaze displacement {1}.

Gaze displacement in relation to an environment is disassembled into two component movements: the eye saccade and the head movement which can be considered as head saccade toward the AOI. The head due to its large inertia moves slower than the eye during the saccade. For simplicity of description eye and head saccade are drawn in a way that they start simultaneously which is not always true. It can happens that either eye or head starts first. When the gaze displacement is a result of appearing stimuli in a periphery the oculomotor response precedes head response. The possible explanation of it is that the visual system needs to inform head-neck motor system about their necessity to move but in the same time eyes are moved as quickly as possible toward the target.

The other possibility is when a necessity to move to displace the gaze is not triggered by an external stimulus but as a result of a volitional displacement. In this case head saccade start simultaneously with eye saccade or it can happens that in some instances head starts as the first one. One can think about using the synchrony of eye and head saccade to learn if a particular gaze redirection is caused by an external stimulus or if it is the result of attain guided visual search.

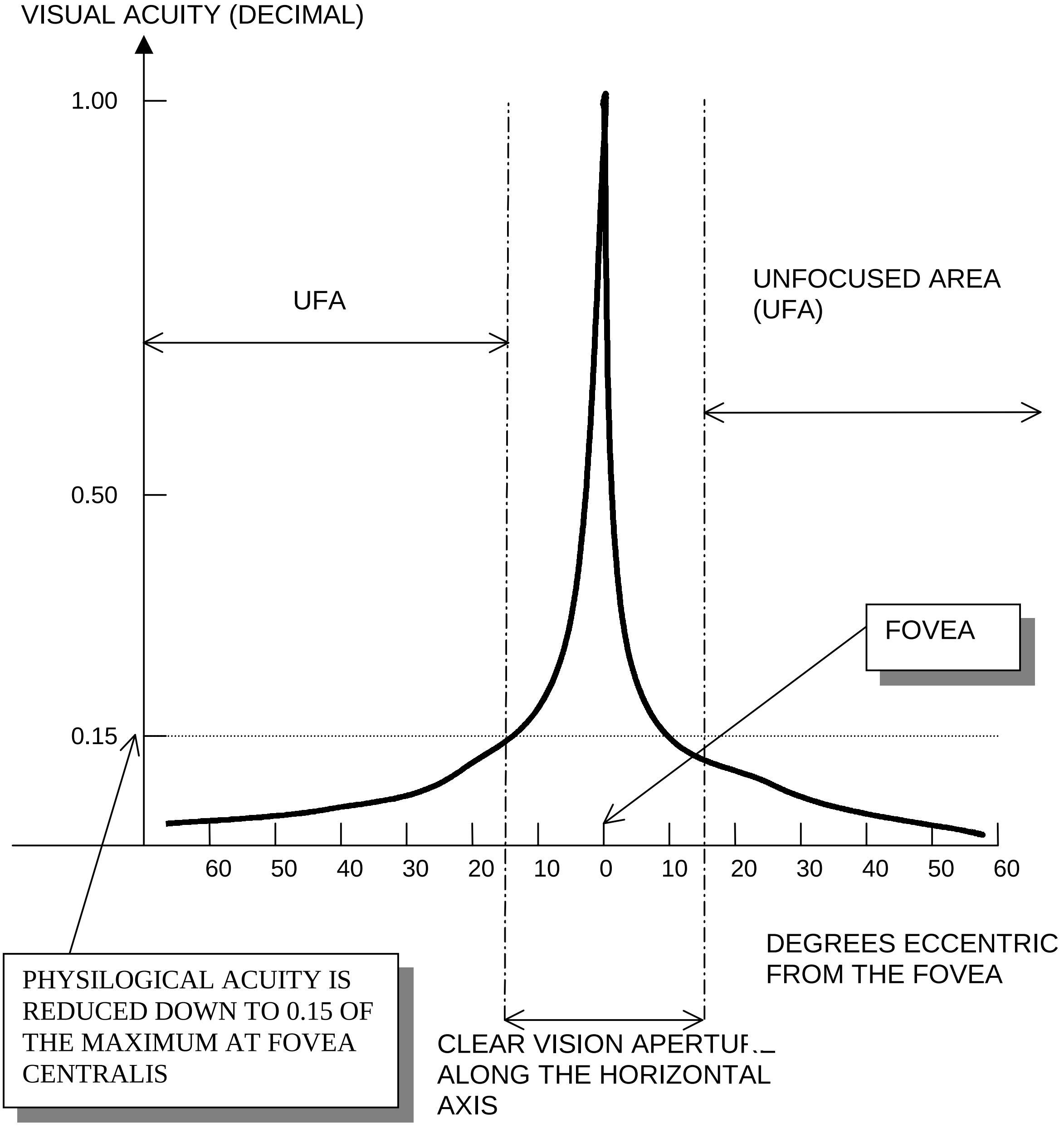
The amplitude of the head movement when performing the gaze shift is varying around value equal to 0.5 of eye saccade. Because the head displacement is only rarely related to the gaze displacement there is a necessity of monitoring independently of the extend of the eye movement in relation to the head as well as a rotation of the head in relation to the visual environment. The effective gaze displacement is the algebraic difference between eye saccade and head saccade. This decomposition does not mean in reality they can be observed independently. After the eye saccade landed the head continuous to rotate. It evokes the veestibulo-oculomotor response which is used to stabilise the image on the retina in situations when head moves. Thus when the head rotates eyes move in opposite direction with the same velocity. Measuring eye position in relation to the head allows to define the gaze displacement in relation to the outer environment. In healthy subjects as presumed pilots are the eye movements can be used to measure head movements without additional instrumentation. In fact the vestibular system, semi-circular rotational accelerometers constitute the finest motion sensing system. Such physiological opportunity has nor been exploited until now for gaze tracking systems and it is opened for research.

Anyhow measuring the eye movements in relation to the head is not an easy task from the point of view of existing instrumentation. One can think about developing an easy system for pilot training so he is tempted to develop a system which will not require measuring the eye position. It is feasible if we can only force the head to rotate the same amounts as the eye saccades. Then the rotation or head saccade will be equivalent to the gaze displacement. Such conditions are shown on figure 2.

The only problem to solve is how to force the head to move the same amount as the eye saccade. We started our experiments with information about oculomotor behaviour by patients wearing the bifocus spectacles, especially with continuously changing power (varilooks type). Due to theoretical constraints and technological limitations a useful vision is limited to the vertical stripe through which the focusing power is changing. It means that subject wearing such glasses experience some kind of limitation of visual angle. This releases the central adaptation process which forces the head to realign with a direction of gaze. In this way subject is centring spontaneously the sharp sector around the position of gaze. Basing on this phenomenon we have artificially deteriorated the vision outside the horizontal streak with the angular size of 30 degrees horizontally and 10 degrees vertically. The deterioration of a vision has been achieved by adding strong +3 dioptre lenses. For undisturbed vision we have cut an opening having the shape of the above mentioned horizontal streak.

When one would like to get an undisturbed image he needs to align the symmetric axis of a horizontal streak with the gaze direction. It forces the head to follow exactly the gaze displacement. The area outside horizontal streak is accessible for a subject but it is out of focus. The quality of the image accused through the area outside the horizontal streak is adequate for serving the orientation if a person in relation to an environment. The central vision requires the highest optical quality of projection. This can be only achieved through the area of horizontal streak. It forces the head to align with a gaze.

Applying the camera on the spectacles and equipping it with telelense view angle 45 degrees allows to monitor the gaze position with the resolution +/- 5 degrees. For the research purpose it is not applicable. For operators training it seems to be more than adequate.



PHYSILOGICAL ACUITY IS REDUCED DOWN TO 0.15 OF THE MAXIMUM AT FOVEA CENTRALIS

CLEAR VISION APERTURE ALONG THE HORIZONTAL AXIS

UNFOCUSING POWER OVERLAPS THE PHYSIOLOGICAL DECREASE OF VISUAL ACUITY

Figure 3. Visual acuity and unfocused area.

One can expect that the deterioration of vision caused by +3 dioptre at the periphery will be experienced as some sort of visual discomfort. To our astonishment we observed that subjectively speaking peripheral vision is not disturbed to the extend we have expected. In opposite shortly within few minutes we are not consciously notice the presence of the focusing for the covered area overlaps with physiological reduction of vision acuity {2}. See figure 3.

The image from the spectacles is transmitted wirelessly to the operator. Tested subject can move freely within a control room. When used in a cockpit there is no necessity to connecting the system to a simulation hardware. Subject wearing the spectacles are advised to re-establish the highest acuity within a horizontal streak by adding an extra lens as prescribed in front of HAG spectacles.

References:

1. Brennan D. H., Vision in Flight, *Aviation medicine* edited by Ernsting J. and King P., Butterworths, London 1996
2. Gauthier G. M., Semmlow J. L., Vercher J. L., Pedrono C., Obrecht G., Short-Term and long-Term Adaptative Changes in Eye-Head Movement Coordination Resulting from Reduces Peripheral Vision, *Presbyopia Research* edited by Obrecht G. and Lawrence W. S., Plenum Press, New York and London 1992