Automated assistance for human factors analysis in complex systems

A. GREGORIADES & A. G. SUTCLIFFE

Centre for HCI Design, School of Informatics
University of Manchester, P.O.Box 88, Manchester, M60 1QD, UK
ags@manchester.ac.uk

Keywords: Socio-technical systems design; Human performance; Workload assessment; Functional allocation.

This paper describes a tool and method for scenario-based workload assessment and performance validation in complex socio-technical systems design, such as command and control rooms of military vessels. We assess workload in terms of communication and the task load that each agent is able to handle. The method employs subjective task and communication estimates used to calculate the workload of human operators, using static and scenario-based analyses. This enables the identifications of bottlenecks to be addressed by the designer with the appropriate allocation of function between humans and smart technology. This task is supported by the functional allocation advisor tool. A case study is presented that demonstrates the use of the tool for the design of the command and control room (C2) of a military vessel.

1. Introduction

In complex socio-technical systems such as military command and control rooms, human agents are faced with a variety of threats that impose periods of high and low workload. An early workload assessment of air traffic controllers (Stein 1985) identified the need for workload prediction to assist planning and staffing of air traffic control rooms. According to Rouse (Rouse et al. 1993), workload is defined as the demand placed upon people which may be a behavioural response to events, communication and interaction between the humans and technology. Hart and Wickens (1990) portray workload as the cost of accomplishing task requirements for the human element of socio-technical systems. High levels of workload degrade the operator's concentration, information processing and decision making, leading to increased errors which might have catastrophic effects (Reason 1990, Leveson 1995). However, low workload can also be pathological since it causes boredom and can result in a deterioration of performance (Hollnagel & Bye 2000).

The need for lean manning of future C2 systems in order to cut costs will increase workload unless alleviated by more automation, although many have warned about the dangers of excessive automation (Reason 2000, Hollnagel 1998, Perrow 1999). In order to overcome this problem, designers need to identify the most appropriate functional allocation scheme, although little advice apart from general principles exists to help solving this problem (Sheridan 2000). According to Woods et al. (2002), ineffective functional allocation can increase information overloading in C2 systems.

Workload is characterised by multifaceted definitions (Jex 1988, Xie & Salvendry 2000), including the consideration of task and human characteristics, and the task environment. Megaw (2005) identified three main approaches in a comprehensive review of workload assessment. Physiological approaches, such as measurement of heart rate or EEG, have the merit of objective measures but can be difficult to achieve in field settings. Experimental approaches use standard task or interference task and record completion times and errors; however, experimental approaches can be criticised for being artificial. Finally subjective rating measures have the merit of being cheap and quick to administer as well as being reasonably accurate. According to Hill et al. (1987), workload assessment can be classified into two main categories, the analytical and empirical techniques. The latter requires the participation of a human operator to carry out the task(s) while observing performance changes (times, errors). Empirical approaches can be further analysed into four groups (Meshkati et al. 1995): primary and secondary task measures, subjective task measures and physiological measures. In the primary task measures technique (O'Donnell & Eggemeier 1986) the investigator assesses the performance of a primary task undertaken. Secondary task measures (Knowles, 1963) employ two tasks, one primary and one secondary. This enables the estimation of resources required for the primary task. If the secondary task is performed well then the primary task is considered easy (low workload). Subjective estimates acquire operators' ratings of demands, stress, loads and task complexity, for example the widely used NASA TLX inventory (Hart & Staveland 1988) or SWAT (Reid & Nygren 1988), and provide metrics to calculate the mental and physical workload (e.g. Wierwille 1988, Moray 1979).

Physiological measures assess changes in physiological functions during task performance, e.g. cardiac functions (Kramer 1991), brain activity (Cooper et al. 1980), eye fixation (O'Donnell & Eggemeier 1986), pupil diameter (Kahneman 1973), respiration (Backs & Ryan 1992), blood pressure (Mulder 1980), hormone levels (Wilson & Eggemeier 1991), electrodermal activity (Heino et al 1990) and muscle activity (Van Boxtel & Jessurun 1993). Physical workload measures have been assessed using physiological measures (Strasser 1977). Analytical techniques simulate aspects of human behaviour and performance to estimate workload using a variety of mathematical models based on information theory, manual control and queuing theory; and simulations and task analysis using fuzzy logic (Ntuen et al. 1996) and discrete event simulation (Laughery 1999, Wetteland et al. 2000) to model workload in task sequences.

Despite the consensus concerning the importance of workload in complex socio-technical systems, and development of measurement techniques (Wierwille 1988), no method exists that integrates assessment of multi-tasking and multi-agent communications. Assessment of communication is implicit in Multiple Resource Theory (Wickens 2002) of multitasking performance; however, this does not consider the effect and demands of different types of discourse acts (Grosz & Sidner 1986, Bunt 1989). This paper presents a method and tools for assessing the level of workload of prospective system designs, providing comparisons of the relative merits of alternative designs. The method introduces communication and task dependencies between human and machine agents as a measure of workload. The method is supported by a set of tools that conduct the analysis, and advise the analyst on whether system-level requirements should be automated, partially automated or specified as manual procedures.

The following section explains the system modelling approach used. Subsequent sections introduce the HUCRE (Human Centred Requirements Engineering) workload and performance analysis method. The supporting tools are described, followed by a case study. The paper concludes with a brief discussion.

2 System Modelling

In our approach we augmented the OODA cycle's four basic steps: observation, orientation, decision and action, with additional steps for recognition, interpretation and analysis. Operational scenarios are converted into task sequences that correspond to the above cycle. This generic structure is subsequently specialised into specific task sequences that describe the domain. Figure 1 presents the generic form of the augmented model complemented with the cognitive resources required for each step and the cognitive pathologies that could result from each step (Sutcliffe 2000).

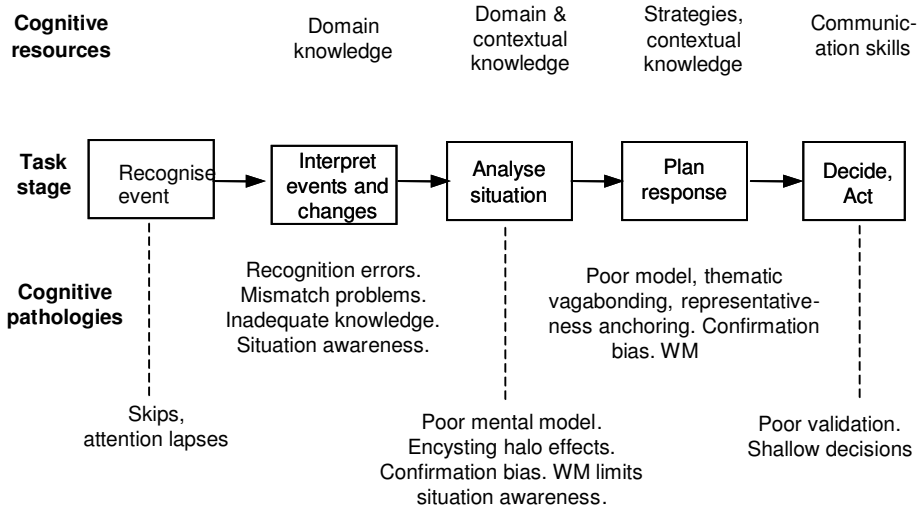

Figure 1: Extended OODA decision cycle, the cognitive pathologies are embedded as part of the advice set in the functional allocation advisor tool

The first problem we needed to address was how to model the system design. We used the i* agent-oriented modelling framework, (Yu 1994, Liu et al. 2003). The i* framework models socio-technical systems as networks of agents (human or machine), goals, system qualities called soft goals, tasks and resources. Agents, goals and tasks are connected by dependency relationships, i.e. a goal depends on an agent to achieve it. Other relationships are means-ends (tasks are the means to achieve goals) and decomposition so lower-level views can be decomposed from higher-level agents (as sub-systems) and goals. The i* framework facilitates analysis of socio-technical system designs by tracing dependency links to ensure that goals are achieved by tasks carrying out tasks with the necessary resources. Agents, tasks and goals have properties so the model can be inspected to check whether the agent has the necessary capability (i.e. training) to carry out a particular task, has been allocated responsibility for a goal, etc. However, i* does not model information flows and inter agent communication explicitly.

Since we argue that communication should be considered during workload analysis, we augmented this methodology with communication dependencies between agents, expressed as discourse acts. Modelling how communication is integrated with action is a key concern in command and control rooms. The extension is to show communication channels between agents, typed by discourse acts, e.g. Inform, Command, Report, Propose, Check, etc. (further details are given in Sutcliffe 2000, 2001). Other additions we have made to i* are to add the attributes' duration and complexity of tasks, and discourse acts to describe human computer interaction (Inputs, Commands, Simple output messages, Complex messages, Simulations, Decision support/situation awareness displays, Computer suggestions, Computer commands for human action). Inter-agent interactions and communications are categorised into inflows and outflows, and assigned a discourse act type which has an expected duration and complexity, based on weights that represent the dependency between the originator of the message and the receiver in terms of implication for action and decision making, following the theories of common ground (Clark 1996) and discourse action (Winograd & Flores 1986). Weightings are specialised for different domains from the generic defaults. The discourse acts have different degrees of force or obligation which they impose on the receiver. Requests, checks, informs and reports generally only impose the obligation to provide or process information; whereas disagrees, corrections and warnings necessitate reasoning; proposals require the receiver to consider the proposition and then agree or disagree. Commands have the highest degree of force because they necessitate reasoning and action; furthermore, the receiver has less choice, and hence both agents have a higher mutual dependency. In contrast, Proposals and Coordinates have less force but require more exchanges to arrive at an outcome so they have longer durations and couple agents during the exchanges. Discourse act weights express a load depending on the information processing implications of each act. While we do not assert that communication load is same as workload, we do argue that it needs to be considered when designing roles, procedures and socio-technical systems. Too many high dependency discourse acts focused on one agent are likely to be a source of system failure. When communication time is calculated we use a time estimate for generating or receiving the act itself plus the complex weight that represents the estimated cognitive processing time.

Figure 2 depicts the generic model of dependencies between agent-task-technology and provides an example of the interactions between the Principal Weapons Officer (PWO) and Air Picture Compiler (APC) agents during the planning task, using tactical plot technology.

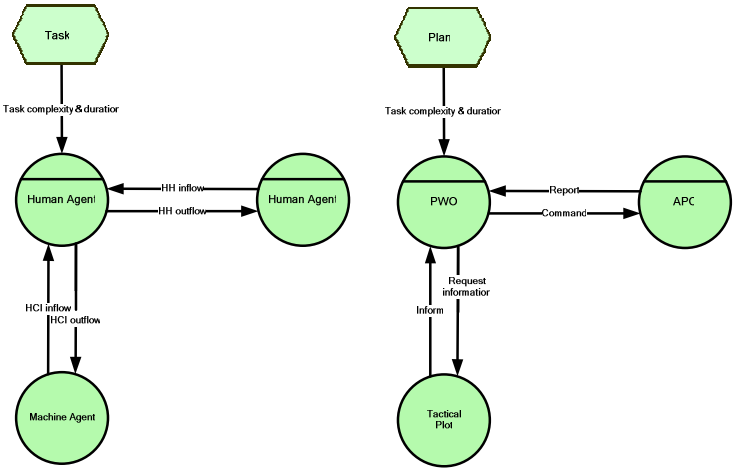

Figure 2: System modelling constructs; the left hand general schema is specialised to the right hand domain model

3 Workload Analsyis Method

The method builds upon Coupling Analysis (Sutcliffe 2000, 2001, Sutcliffe & Minocha 1999) which investigated the dependencies between agents, the tasks they were responsible for, and communication. However, coupling analysis was a laborious manual process which required automated support.

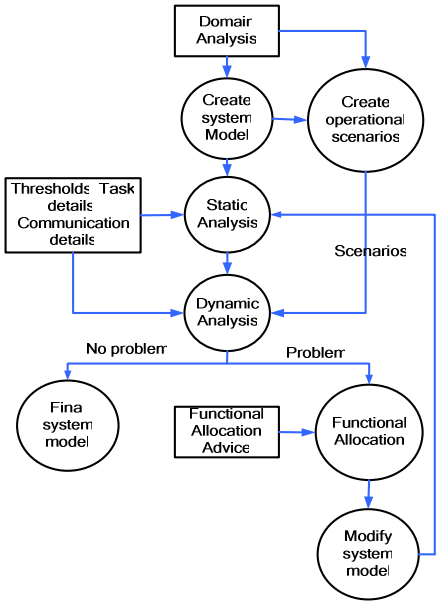

Figure 3. The HUCRE method illustrated as a data flow diagram of method steps and resources provided

The method steps are summarised in figure 3. First a model is constructed in an extended i* notation that adds communication dependencies between agents, expressed as discourse acts with an associated weighting that expresses the dependency imposed on the receiver in terms of responding to the communication. This model enables a static workload analysis which checks that human agents have sufficient time to perform the tasks they are responsible for in the time available, e.g. shift, working day. The support tool multiplies estimated task completion times by the expected iterations per day, summing the results and comparing the sum with the agent's available hours. The static analysis pinpoints agents who may become overloaded if they have been allocated too many tasks or have to perform too many iterations of the same task. The pattern of message flows inbound/outbound to each agent is investigated to determine the agent's ability to respond to the demands of others.

Static analysis is followed by a dynamic analysis which diagnoses potentially excessive workloads when human operators are carrying out two or more concurrent scenarios. The estimated task completion and communication time are compared to an ideal time for scenario completion. This analysis identifies areas in the system model where operational problems may occur, e.g. human agents have been allocated too many tasks, or do not have the time to carry out the allocated tasks. Other problems are identified, such as missing communication channels, and information resource dependencies necessary to carry out a task.

This leads to the third step, which investigates the functional allocation of system requirements. The conventional response to work overload is to increase automation; but this comes with its own penalty of impoverished situation awareness and subsequent errors resulting from leaving the human agent out of the loop (Hollnagel & Bye 2000). To help the designer consider these trade-offs, the tool provides advice on the level of automation and generic requirements customised to tasks in the application domain.

The method and supporting tool helps system engineers, who are invariably not experts in human factors, to diagnose workload problems in such systems, and then reason about requirements to solve potential overloading, as well as validating that a potential system design can fulfil operational requirements for a given set of scenarios. A case study explains the method in more depth by describing two software tools that support the method, with a scenario of use. The scenario of use was itself used as a test task for the evaluation sessions.

4 Tool Support and Scenario of Use

Two software assistant tools were developed to support the HUCRE method. The Workload Analyser, programmed in JAVA, helps to diagnose agents and tasks where overloading will occur when the system is responding to events in one or more scenarios. The Functional Allocation Advisor, programmed in Visual Basic, provides advice on

functional allocation based on the literature (Bailey 1996, Cook et al. 1999, Cook & Corbridge 2000, Wickens & Holland 2000) and generic requirements for each task. The system architecture is illustrated in figure 4.

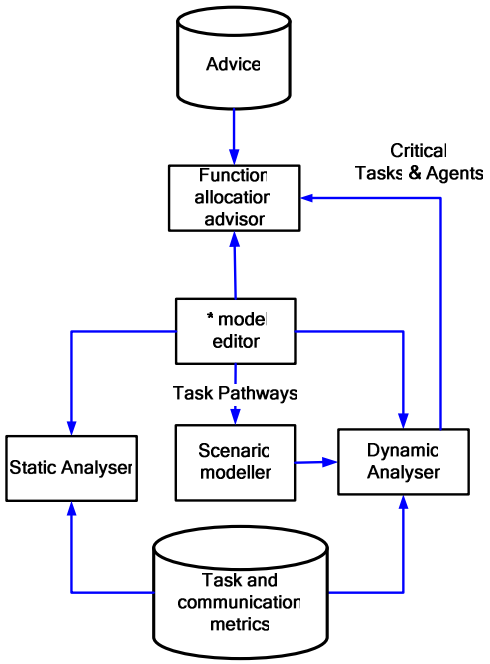

Figure 4. System architecture of the HUCRE method and support tools

The static and dynamic Workload Analysis tools read values of task completion times and complexity weights from the system database to calculate agent workloads, using the model created in the i* model editor. The dynamic analysis tool also takes input from the scenario modeller with which the user creates scenarios as task sequences. The dynamic analysis algorithm partitions task completion times into different intervals for each scenario so that loading in concurrently running scenarios can be compared by a time interval analysis. The granularity of analysis is under user control. The identities of problem agents and tasks are passed to the Functional Allocation Advisor which then highlights them in the i* model display. The user can access the requirements advice database by pointing to the highlighted task and advice cues in pop-up menus.

5 Case Study: Combat Control System

The domain is the combat sub-system of a naval offshore patrol vessel. The system is composed of 12 operator roles (Radar/sonar operators, tactical picture compilers, weapons directors, and the principal weapons officer who is in command) and computerised equipment (radars, sonars, situation displays, electronic jamming devices, etc.) in the ship's combat control system. The domain knowledge necessary for the development of the system model was elicited from domain experts from BAE Systems. The combat control configuration was a slimmed down version of the Type 23 Frigate investigated by Stanton

et al. (in press). The crew complement differed between the two studies primarily in the merging of the PWO (Principal Weapons Officer) and AWO (Air War Officer) in our case study, which our analysis demonstrated was a problematic decision.

The system is modelled using the extended version of i*. Dependency relationships are shown between tasks and agents to denote which agents are responsible for carrying out particular tasks, and the discourse acts required to fulfil the command and control collaborative task. Several goals are not shown, to simplify the diagram, since tasks fulfil the requirements indicated by goals (see figure 5).

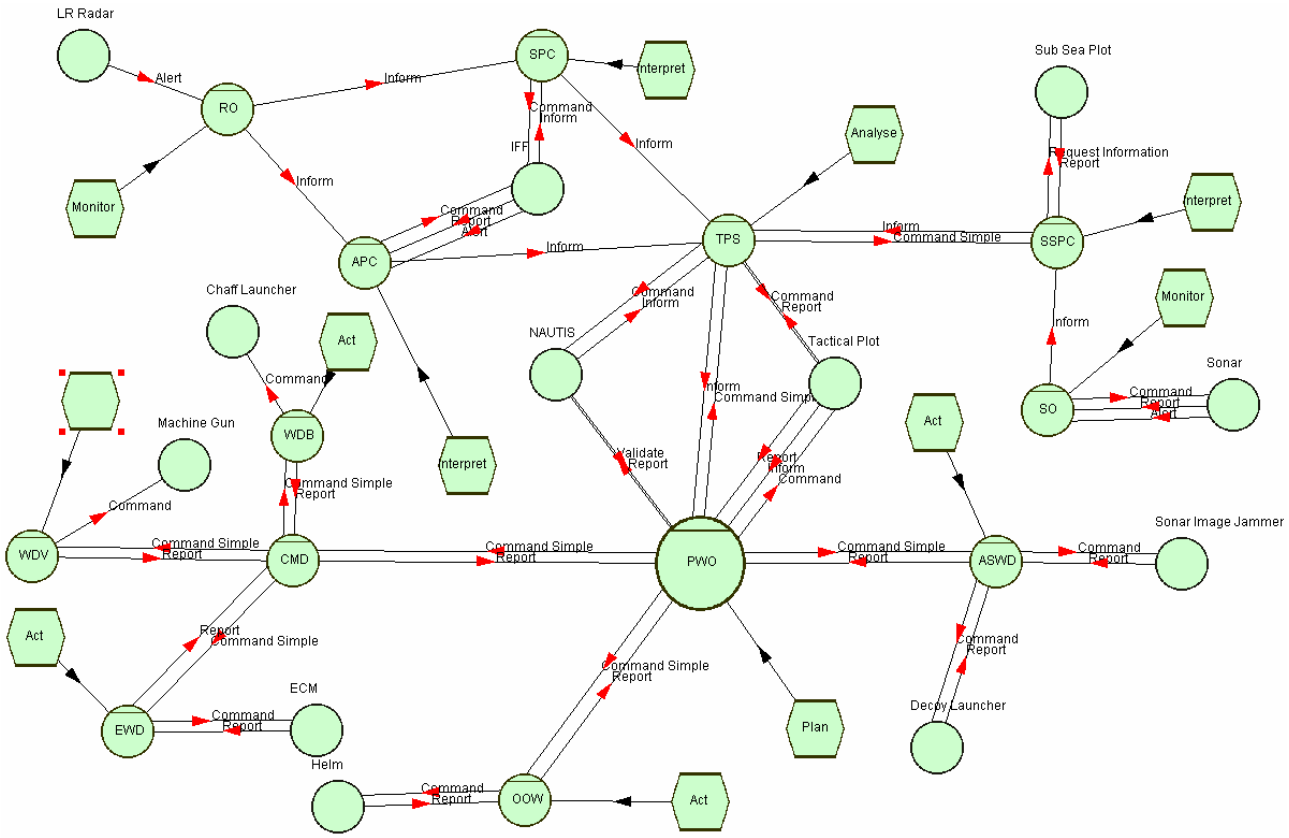

Figure 5: The system model of the command and control room

The system was tested with three scenarios (dynamic analysis) which go through several cycles of generic task sequence, from recognising and interpreting a threat, through analysing the situation, to planning a response and taking action. The scenarios are taken from naval training exercises similar to those reported by Stanton et al. (in press) when they investigated the Type 23 Frigate simulation at HMS Dryad.

Scenarios

The first air attack scenario commences with an external threat of the missile being launched from a hostile aircraft which is detected by Radar Operators (RO) who inform Air and Surface Picture Compilers (APC/SPC) who interpret significant events (i.e. friendly unit or potential threat). They notify the Tactical Picture Compiler (TPC) who integrates input from several operators and passes the information on to the Principal Weapons Officer (PWO), who is responsible for analysing the situation and planning appropriate action, in consultation with the Captain. Once the PWO has decided how to respond, orders are passed down the line of command to the Weapons Directors: Blind (WDB) who is in charge of missiles, Visual (WDV) who commands the ship's guns, Anti-submarine (ASWD), or Electronics (EWD) who controls electronic counter-measures.

The missile attack scenario consists of several cycles in the command and control task sequence, commencing when the enemy aircraft launches a missile, which is detected by the ship's radar. The TPC reports the incoming missile threat to the PWO, who decides to jam the missile's radar using electronic counter-measures (PWO→EWD). The missile keeps homing, so in the second cycle PWO decides to deploy chaff to confuse the missile's radar, and when that does not work, he decides in the third cycle to destroy it with the ship's defensive missile.

In the second scenario an unknown (potentially a terrorist) boat is detected by the ship's radar heading towards the ship. SPC identifies the direction of the boat and its speed in order to track its path. The threat is reported to TPC who assesses the threat and reports to PWO. PWO decides to notify friendly units (PWO→CMY Communications Yeoman) and tries to radio the boat (PWO→CMY, RO). WDB reports unsuccessful radio contact. PWO decides to fire warning shots to the incoming boat (PWO→WDV). The boat is still heading towards the vessel so PWO decides to destroy the boat using 76mm gun. WDV reports successful destruction of the terrorist threat.

A third scenario which we have not elaborated describes a submarine attack foiled by manoeuvring the ship and jamming the submarine sonar.

The first problem is the conversion of scenarios, which are narrative stories, into a format that can be executed by the system. This is achieved by decomposing scenarios into a number of tasks using the generic C2 task cycle described earlier; see figure 6.

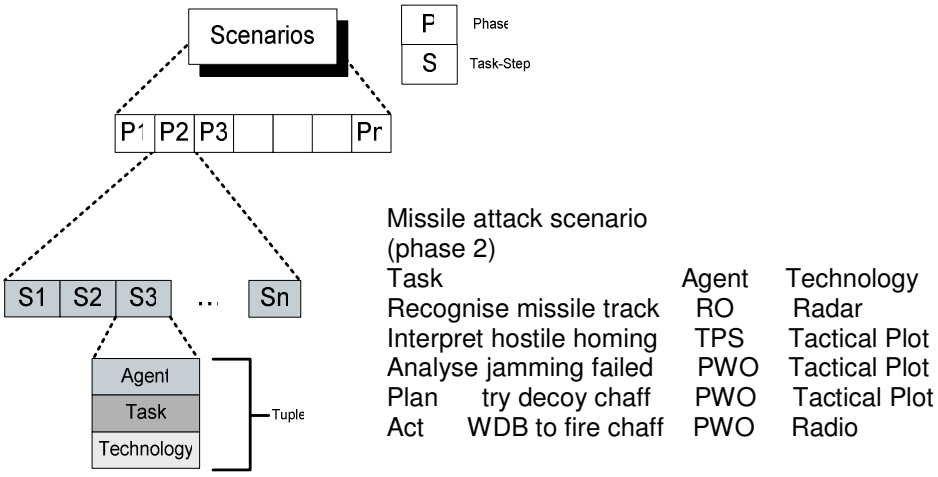

Figure. 6: Scenario structure

Communication is modelled using discourse acts that match tasks in the C2 cycle. The dependencies from RO to TPC to PWO inbound are modelled with Inform-Report acts, whereas the outbound order from PWO to WDB is a Command act. The tool automatically links acts to task dependencies specified in the C2 template, although other discourse acts can be added, e.g. Liaison and Command acts between PWO and the Captain. Scenarios can be constructed interactively by pointing to tasks on the system model; the tool then automatically constructs a scenario task sequence by tracing the human and machine agents involved with each task, as illustrated in figure 7. The system provides task sequence templates, such as the extended OODA cycle used in this example, to aid scenario construction.

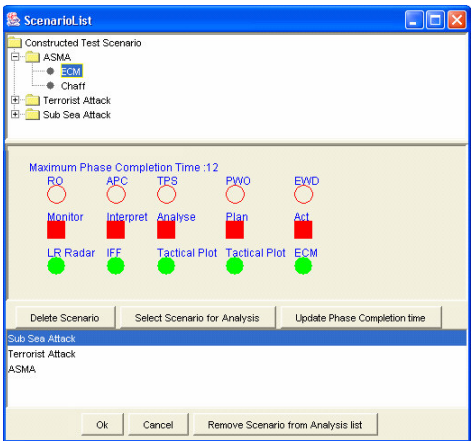

Figure 7. Scenario sequence editor showing the tuples of task, human operator and machine agents

5.1 Workload Analysis

5.1.1 Static Analysis. Static workload analysis answers the question "can the agent carry out the tasks allocated within the time available in a shift or working day?". The tool traces

responsibility relationships from agents to tasks then asks the analyst to enter the estimated number of iterations for each task in the shift. The system database contains estimated task completion times and calculates the total time per task, summed for each agent, then compares this time with the available hours. Any agent allocated more work than is feasible in the available time is highlighted (see figure 8). The analyst has to either reduce the task iterations or reallocate tasks to different agents. If the analyst desires, task times can be weighted by complexity, derived from subjective estimates (e.g. NASA TLX inventory: Hart & Staveland 1988) or expert judgement, to give an adjusted estimate of an ideal completion time. The innovation beyond standard workload analysis tools is to add an estimate of communication time to workload estimates, e.g. the time to prepare and send a message (discourse act type *x*) multiplied by the expected iterations, plus time to receive and understand a message. A static workload algorithm is shown in figure 9.

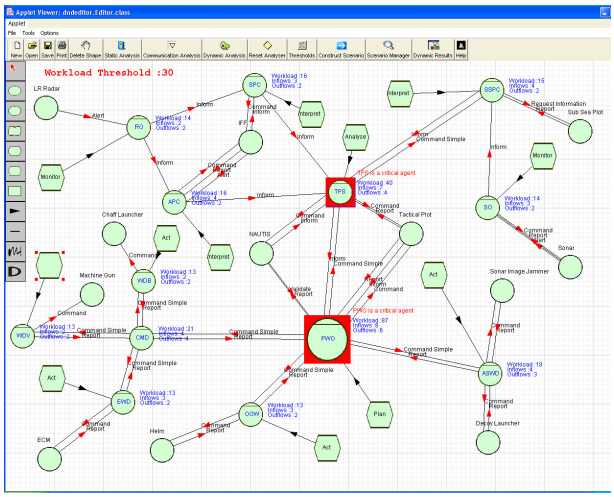

Figure 8. Screen shot of the static workload analysis tool, showing the system model in i* notation with the PWO and TPC agents highlighted to warn of excessive workloads from their task assignments. The tool shows the two agents overloaded when “available hours” is set to 30

For each agent ***in System Model do*** *{*

Identify all connected neighbour agents

For all neitgbour agent connections

$$Cl = \sum_{i=1}^{n} Cc_i \cdot Ct_i$$

For all tasks assigned to agent (i)

$$Tl = \sum_{i=1}^{n} Tc_i \cdot Td_i$$

$$Workload_{ai} = Cl + Tl$$

IF Workload>Threshold => THEN

Highlight critical agent

}

Cl=Communication load
Tl=Task load
Cc=Communication complexity
Ct= Communication time
Tc=Task complexity
Td=Task Duration

Figure 9. Static analysis control algorithm

5.1.2 Scenario-based Dynamic Analysis. The second, dynamic, analysis answers the question "can the system respond to one or more concurrently running scenarios?" to diagnose breakpoints when agents may become overloaded during operational scenarios. This provides a more realistic test of workloads because static analysis ignores the demands of concurrent operation. Scenarios are modelled as sets of discrete activities (tasks) which have known durations. The tool carries out an interval sequence analysis, using the algorithm depicted in figure 10, to assess how many concurrently running tasks will be undertaken by each agent in any one time interval and hence pinpoints intervals when operators may become overloaded. Scenarios are assigned desired completion times by a domain expert. The tool estimates completion times by summing all task durations in the scenario sequence and reports whether the desired time will be achieved. The analyst can set threshold values for desired scenario completion times and agent workloads. The system model can be interactively tested by picking one or more scenarios and placing them in a timeline template (see figure 11) to set their start time, i.e. all start simultaneously or with varying degrees of delay.

For Agent 1..N Do

For Interval 1..T Do

Determine the number of concurrent scenarios

For Scenarios 1..M in Interval (t) Doc

Determine communication from neighbour agents

For all neighbour Agents

$$Cl = \sum_{i=1}^{n} C_{HCI} c_i \cdot C_{HCI} t_i + \sum_{i=1}^{n} C_{HH} c_i \cdot C_{HH} t_i$$

Determine Tasks assigned to Agent (n) in Interval (t)

$$Tl = \sum_{i=1}^{n} Tc_i \cdot Td_i$$

Workload=Cl+Tl

End Scenarios

End Intervalss

IF Workload>Threshold THEN

Highlight critical agent

End Agent

}

Cl=Communication load

Tl=Task load

Cc=Communication complexity, HC human computer, HH human - human

Figure 10 Dynamic Analysis control algorithm

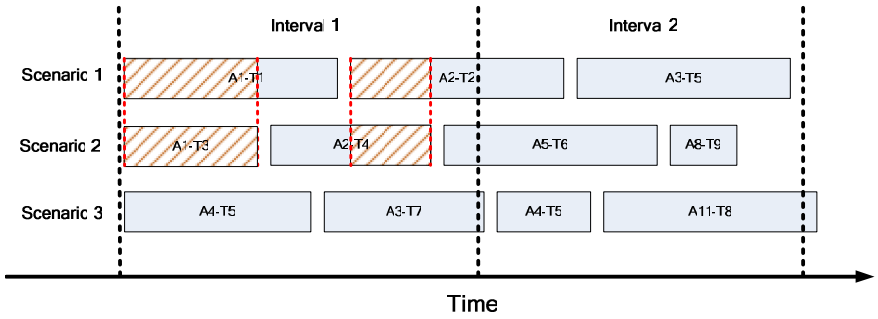

Figure 11: Multitasking example

Interval durations are under user control. The system indicates which agents (human and machine) will experience overload and may be exposed to multi-tasking conflicts. The human factors literature (Hollnagel 1993, 1998, Rasmussen 1986, Reason 1990, 2000) advises that generally people should not perform two or more tasks concurrently, unless the tasks are performed by highly trained skilled operators. Even though military personnel are highly trained there is a limit to concurrent task operation (Rasmussen 1986). For instance, it may be possible to carry out two cognitively simple tasks, e.g. monitoring and interpreting, semi-concurrently by "time slicing" between them. In contrast, knowledge-intensive tasks such as analysis and planning have to be consciously executed and are

difficult to combine with other concurrent tasks. The tool embeds a set of rules expressing permissible concurrent combinations of task types, and the safe degree of overlap between finishing one task and starting another. If concurrently running scenarios imply two simultaneous tasks that infringe the permissible limits, the tool warns the analyst of the multi-tasking problem.

These rules are expressed through a multi-tasking tolerance matrix that defines the level of acceptable multi-tasking among the various tasks in the system model (figure 12). Each cell in the matrix contains the maximum acceptable overlap time for each pair of tasks. In order to assess the level of multi-tasking, the workload analyser represents each task by a series of time intervals. Time intervals can be set to finer or coarser grained analysis (e.g. 10, 30 seconds, 1, 2 minutes, etc). The tool segments each task into a sequence of intervals that account for its duration, so a 96 second task with the interval analysis granularity set at 10 seconds would consume 9 complete and one partially complete interval. Multi-tasking identifies the intersections among all tasks performed by the same agent. Therefore in the example in figure 11, agents 1 and 2 are multi-tasking (dashed areas). Agent 1 is performing tasks 1 and 3 starting at interval 0, and agent 2 is multi-tasking with tasks 2 and 4. Given that the duration of task 1 = 11sec and task 3 = 7sec, the corresponding overlap sets are A1T1 = {1...11} and A1T3 = {1...7}. Hence multi-tasking level = $A1T1 \cap A1T3$ = {1...7} = 7. If this number is greater than the number assigned in the tolerance matrix for the two tasks then multi-tasking is critical and the agent is highlighted.

S=

	T1	T2	T3	T4	T5
T1	S11	S12	S13	S14	S15
T2	S21	S22	S23	S24	S25
T3	S31	S32	S33	S34	S35
T4	S41	S42	S43	S44	S45
T5	S51	S52	S53	S54	S55

Where $S_{ij} = S_{ji}$ for all i, j=1, 2...n.

Figure 12: Multi-tasking tolerance matrix with minimal tolerable task overlap times

5.1.3 Results. In the system being tested the analyst has selected three scenarios (missile, submarine and terrorist attack) to run concurrently. The tool indicates that three agents, RO (Radar Operator), TPC (Tactical Picture Compiler) and PWO (Principal Weapons Officer), will be exposed to excessive workloads and errors inherent in attempting the two concurrent, complex tasks, analysis and planning (see figure 13).

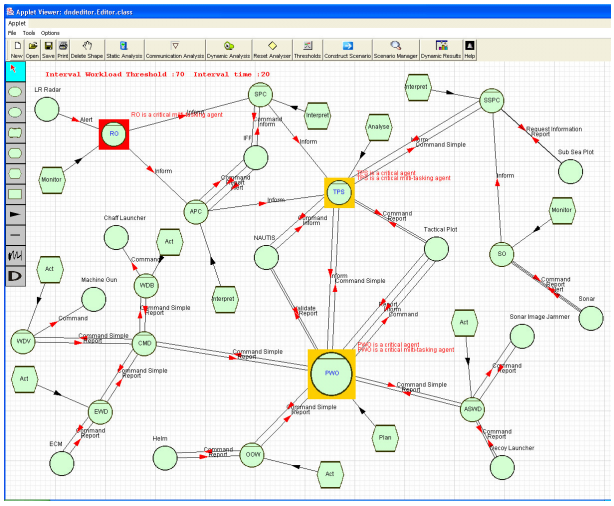

Figure 13. Dynamic workload analysis, highlighting agents (RO, TPC and PWO) with excessive workloads and multi-tasking problems

The tool also displays the workload by agent for each scenario time interval so the analyst can observe when the problem will arise. The selected scenarios are displayed in a timeline sequence, overlaid with the expected completion times (see figure 14). The results display can be configured to show the workload of all agents for each scenario in time intervals, or workload for single agents in each time interval. PWO will be overloaded in intervals 0 and 1 (0-20, 20-40 secs) when three concurrent planning tasks are necessary to deal with each threat; however, in intervals 2-5 only one planning task is necessary so the workload is acceptable. Similarly TPC has to analyse three different threats.

The display indicates that TPC and PWO have problems with overall workload as well as multi-tasking conflicts, while RO has a workload problem. In this case problems arise with the assignment of analysis and planning to PWO, who has to deal with three concurrent threats: one from a missile, one from a submarine, and an unconventional terrorist threat. The tool indicates that even if two of these scenarios happened concurrently, the system would probably fail because PWO could not deal with the threats in time.

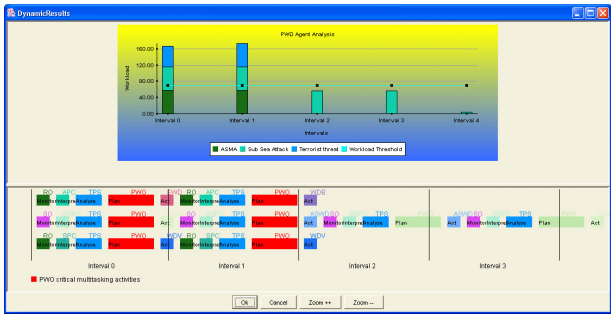

Figure 14. Workload summary for the PWO agent in 20s time intervals with three concurrent scenarios, starting at the same time

Having identified the overload problem, the next question is how to revise the system to reduce or eliminate the overload. A systems engineer would have to wait for a human factors expert to make suggestions; alternatively, we provide an advisor tool, which can be used to provide the answer with a modest amount of human factors training.

5.2 Functional Allocation Advisor

The Functional Allocation Advisor contains four sets of advice linked to specific tasks. The advice is intended to guide the system designer when making decisions about which functions to automate, or not, and suggests reusable generalised requirements which should be appropriate for the task. The generic requirements are derived from a list of functional requirements for abstract generalised tasks (Sutcliffe 2002) which have been specialised for naval command and control systems. The first set advises on functional allocation problems, i.e. the extent of automation which is desirable for the task, whether it can be safely assigned to one or more agents, problems to beware of if the task is divided between several agents, and the potential dangerous effects of multi-tasking on performance and errors. The analyst has to chose whether to use the advice or not, and then interpret it in a domain specific context. The advice is accessed by pointing to the task implicated in the workload analysis and selecting the appropriate advice set (see figure 15). The advice tool can also be yoked to the workload analysis so advice windows pop up for each highlighted agent.

The tool displays the scenario narrative, a sub-set of the i* model that is relevant to the selected scenario, showing the task sequence dependencies and agents, with selection controls to access advice and requirements appropriate to the task. The tool can be used in stand-alone mode or loosely coupled with the workload analysis, in which case the overloaded agents and task are indicated. The view can be switched between different scenarios and corresponding i* models; however, the advice is tailored only for the tasks/domain rather than for each individual scenario.

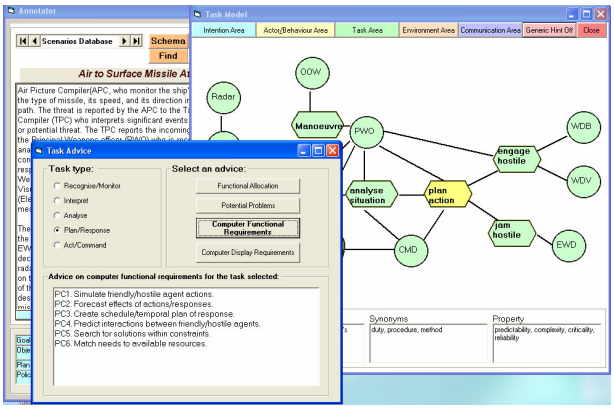

Figure 15. Functional Allocation Advisor showing functional requirements for the planning task

The second advice set indicates potential cognitive problems human operators may face when carrying out a task without the appropriate equipment or if they are overloaded. This advice requires some human factors knowledge for its interpretation. For example, recognition and interpretation tasks tend to be prone to slips and lapses in attention; whereas analysis and planning tend to suffer from inadequate mental models, poor situation awareness, inadequate hypothesis formation, failure to test assumptions, etc. The tool provides background knowledge to help the analyst interpret functional allocation advice. Generic requirements for task support are provided in the next advice set, followed by information display requirements to support human decisions in the final set.

After the system designer has used the Workload Analysis tool to locate the planning task as the cause of overload problems for the PWO agent, he/she accesses functional allocation advice which indicates that planning tasks are best allocated to one agent but may be sub-divided if a coordination task is added to integrate sub-plans. The tool advises that automated planning is possible only for well known deterministic domains, but may be necessary when response times are very short, although it should be implemented with an over-ride option.

The functional requirements to support the planning task are illustrated in figure 15. These include creating simulations and forecasts of agents' actions in the domain, suggesting response plans and predicting interactions between friendly and hostile agents. Other requirements are constraint-based searches for pre-set responses to a current threat (e.g. manoeuvre ship, use decoys, electronic jamming strategies). These requirements are selected with the system suggesting planned responses to known types of threat; so some of the planning burden for air and sub-sea threats could be alleviated, leaving PWO more time to concentrate on the unpredictable terrorist threat. The computer display requirements leave planning in human hands but advise on the information to support human decision making such as plotting movement of friendly and hostile agents, with checklists and agendas for tactical responses. The analyst has to decide which functional allocation advice to follow and pick requirements for an automated or semi-automated system depending on his/her interpretation of the domain.

5.3 System Redesign

Having identified that the main concern was for the PWO, APC and RO agents, the system designer would make appropriate changes. The designer decides to change the level of automation by replacing the NAUTIS component with an advanced DSS that would relieve PWO's and TPC's workload. The decision support system helps the analysis and planning tasks by applying a "rules of engagement" analysis to incoming threats. The system assesses the situation and suggests the more appropriate action(s) based on knowledge of past situations using case-based reasoning. Additionally the system provides the user with a checklist of actions that should have been performed, with current status, in order to support the situation awareness.

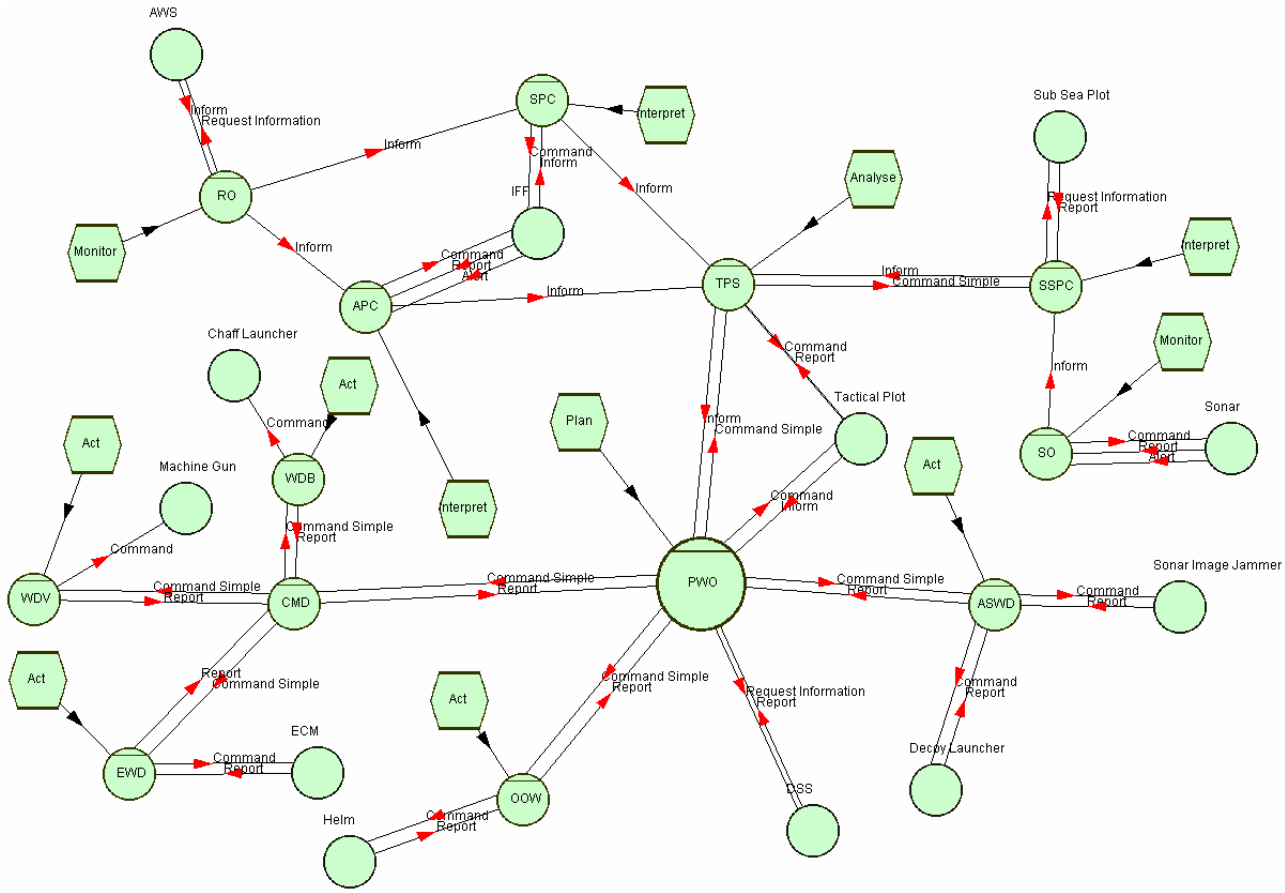

Figure 16: The system model of the command and control room

To overcome RO's overloading an Advanced Warning System (AWS) is introduced to prioritises incoming threats based on their severity. This will enable RO to focus his/her attention on dealing with the critical threats rather than being overloaded with multiple potential threats. The design of the new system is depicted in figure 16.

The new design was evaluated (see figure 17 for the results of the dynamic analysis) to confirm that all critical agents had been relieved from the excessive workload.

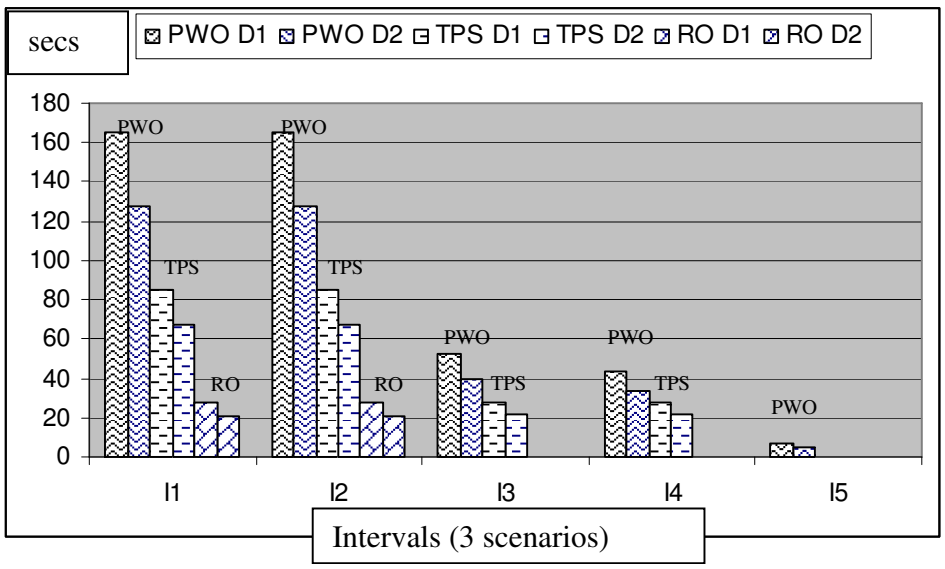

Figure 17. Comparison of estimated workloads in the original (1) and revised (2) designs for all intervals with three concurrent scenarios

Introduction of the decision support system led to a dramatic drop in PWO's workload by reducing the time taken to complete the planning and decision tasks with improved information display and filtering options. PWO's communication load also decreased due to more effective human computer communication. TPC's workload also decreased for the analysis task, due to improved decision support although the improvement was not as dramatic. The RO monitoring and interpretation tasks benefited from the advanced warning system which prioritised the incoming threats, therefore reducing complexity and completion time. The overall reduction in the static loading analysis was 20% for PWO, 21% for TPC and 17% for the radar Operator.

6. Discussion

A wide variety or workload assessment techniques exist ranging from single task workload assessment (Brookings et al. 1996) and subject assessment questionnaires (Hart & Staveland 1998) to measures of the psycho-physiological responses of individuals to varying workloads, and qualitative techniques for measuring individual and team workloads based on the operator's perception (Entin et al. 1997); see also the review by Xie & Salvendry (2000). Apart from general simulation environments such as MicroSAINT (Chubb et al. 1987) and control theory based mathematical simulations (e.g. PORCRU, Baron & Corker 1989) most techniques have no automated support. This limits their scalability when analysing complex systems which have to respond to a wide variety of scenarios and operational conditions. Jung and Jung (2001) use an analytical hierarchy process (AHP) to rank tangible and intangible factors that influence workload for a single task; however, their approach did not consider multi-tasking and multiple operational scenarios.

The method and tool presented in this paper advance workload analysis by performing a scenario pathway analysis to assess operational times against a desired standard. A further contribution is automatically diagnosing breakpoints in the system where failure might occur if the system had to respond to two or more concurrent scenarios. The tool provides a general, tailorable technology, so a variety of subjective and objective workload measures (e.g. see Xie & Salvendry 2000) could be incorporated. While workload analysis methods have been developed for multi-tasking (e.g. Time Line Analysis & Prediction (TLAP) Parks & Boucek 1989), and Workload Index (North & Riley 1989), the calculations can be time consuming and the communication component of workload is not considered. The discrete event simulations used by Wetteland et al. (2000) and Laughery (1999) for human performance assessment with the MicroSaint modelling environment is similar to our use of scenarios and uses a multi-task model, but does not consider inter-operator communication The influence of human communication on system performance has been analysed in the domain specific context of air traffic control (Manning et al. 2002); however, we believe the method we have proposed is the first to provide a general framework for explicitly analysing communication as a component of workload, as well as being diagnostic in its role in system failure.

Analysis of inter-operator communication develops the concept of the management component of workload (Xie & Salvendry 2000) for coordination while also considering communicative acts as explicit tasks. Although Manning et al. (2002) do consider communication events in combination with task measures to predict subjective workload, they concentrate on identifying correlations between the total communication load and subjective estimates of workload. Communication as interaction between roles, i.e. synchronous meetings, asynchronous messages, has been modelled as a workload component in software engineering teams (Hanakawa et al. 2002); however, this study did not consider the impact of different types of communicative acts. The importance of modelling communication as a component of workload and system performance is illustrated by Rasker et al.'s (2000) study on intra-team communication and performance in C2 domains, which demonstrated than unrestricted speech communication enables better performance compared to restricted, text message communications.

The HUCRE scenario analysis method could complement more detailed task analytic approaches such as the Distributed Situation Analysis (Stanton et al in press) which provides a set of task-knowledge maps that can be activated to investigate dependencies in a scenario event sequence. Although their analysis focuses on discovering the key focal knowledge objects that are shared among C2 teams for situation awareness, analysing the implication of different and possibly concurrent pathways through their system models might provide further insight into possible failure in situation awareness when operators are distracted by conflicting but concurrently active sets of information. Even though workload was not the prime purpose of their analysis it is interesting to note that they comment on the possible excessive workload of the PWO and AWO roles, which agrees with our findings.

Several improvements could be made to our tool; for instance, error and the concept of ineffective workload could be modelled (Xie & Salvendry 2000) so scenarios which contain failures and mistakes could be analysed. The idealised task and communication act times we use could be modified by taking the properties of the operators and the operational environment into account, using cognitive resource models (Sarno & Wickens 1992). Another improvement could be to model our scenarios in more detail so the steps could be assessed by a computation version of Multiple Resource Theory (Wickens 2002) to expand the task interference matrix in the HUCRE tool. In related work we have developed Bayesian Belief Networks (BBNs) that enable causal reasoning and the likelihood of failure in scenarios using the same task-operator-machine agent paradigm described in this paper. The BBN embeds theories of human error (Reason 1990, Hollnagel 1993) and accounts for the influences of the operational environment. We intend to integrate the BBN analysis to improve workload predictions in the next generation of the tool.

The second contribution of the HUCRE tool is to provide advice on the partitioning of system activities among manual procedures, collaborative decision support, and full automation. This advice, drawn from the human factors engineering literature (Bailey 1983, Wickens & Holland 2000), could be augmented with trade-offs between performance, situation awareness and system adaptability for different automation paradigms (Endsley & Kaber 1999). Although this advice is normally provided by human factors experts, the limited availability of such experts in many development teams suggests that it should be incorporated into system analysis tools. The evaluation of the method and supporting tools demonstrated that researchers with only limited human factors training could understand the advice and use the tool effectively.

The third contribution is to provide generalised, reusable requirements to introduce knowledge reuse into domains where similar problems arise in different applications. This follows the initiative of generating generic requirements in an aerospace domain (Lam et al. 1997) and reuse of generic requirements associated with types of events in use cases (Sutcliffe et al. 1998, Sutcliffe 2002). Reuse of generic requirements has also been demonstrated in naval and air traffic control applications using a scenario-based approach (Mavin & Maiden 2003).

The evaluation, although demonstrating that the tool and method were usable and effective, did indicate some improvement were necessary. In particular, the cognitive problems advice was difficult for our subjects to understand, so we will review whether further training can make this advice more useful. Some of our subjects commented that generic requirements would be improved if examples were included to promote understanding. This is more difficult since providing specific examples may be inappropriate for particular applications and may bias users' views with low-level solution details. Further testing with domain experts is necessary to improve the quality of the advice.

In the future we will augment the functional allocation advisor by incorporating BBNs to model the divergent influences on functional allocation decision. Validation of the

communication and task model times and weighting against experimental data is a further aim; however, we believe the value of the tool lies in exploratory "what if" assessment of different socio-technical designs rather than in absolute performance predictions *per se*. Since the tool provides an extensible architecture, more sophisticated and better validated workload models can be incorporated in future versions.

9. Acknowledgements

This research was supported by the EPSRC systems integration programme project SIMP (Systems Integration for Major Projects); partners UMIST, QMW and City University, coordinator BAE Systems. Our thanks to Leroy Osbourne, BAE Systems, for his help with the scenarios; and to reviewers for their help in improving this paper.

References

Backs, R. W., & Ryan, A. M. (1992). Multimodal measures of mental workload during dual-task performance: Energetic demands of cognitive processes. *Proceedings: Human Factors Society 36th Annual Meeting*, (pp. 1413-1417). Santa Monica CA: Human Factors Society.

Bailey, R. W. (1982). *Human performance engineering: A guide for system designers*. Englewood Cliffs NJ: Prentice Hall.

Baron, S., & Corker K. (1989). Engineering based approaches to human performance modelling. In G. R. McMillan, D. Bevis, E. Salas, M. H. Strub, R. Sutton, & Breda L.V. (Eds.), *Applications of human performance modelling to system design* (pp. 203-218). New York: Plenum Press.

Brookings, J., Wilson, G., & Swain, C. (1996). Psychological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*, 361-377.

Bunt H.C. (1989). Information dialogues as communicative actions in relation to partner modelling and information processing. In M. M. Taylor, E. Neel, & D. G. Bouwhuis (Eds.), *The structure of multi-modal discourse*. Amsterdam: North Holland.

Chubb, G. P., Laughery, K. R., & Pritsker, A. B. (1987). Simulating manned systems. In G. Salvendry (Ed.), *Handbook of human factors*. New York: J. Wiley.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cook, C., & Corbridge, C. (2000). Function allocation: Optimising the automation boundary. *Proceedings: IEE One-day Seminar on Systems Dependency on Humans, London 16 February 2000*, (pp. 3/1-3/6). London: IEE.

Cook, C., Corbridge, C., Morgan, C., & Turpin, E. (1999). Investigating methods of dynamic function allocation for naval command and control. *Proceedings: International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres*. London: IEE.

Cooper, R., Osselton, J. W., & Shaw, J. C. (1980). *EEG technology*. London: Butterworths.

Endsley, M. R., & Kaber D.B. (1999). Level of automation effects on performance, situation awareness, and workload in a dynamic control task. *Ergonomics*, *42*(3), 462-492.

Entin, E. E., MacMillan, J., & Serfaty, D. (1997). Analysis of organizational processes in

adaptive command and control architectures. *Proceedings: IEEE International Conference on Systems, Man, and Cybernetics*,. Los Alamitos CA: IEEE Computer Society Press.

Grosz, B. J., & Sidner C.L., (1., . , Vol 2(3), pp 175-204. (1986). Attention, intentions, and the structure of discourse. *Journal of the ACL*, *2*(3), 175-204.

Hanakawa, N., Matsumoto, K., & Torji, K. (2002). A communication workload estimation model based on the relationships among shared work for software development projects. *Proceedings: APSEC, 9th Asia Pacific Software Engineering Conference*,. Los Alamitos CA: IEEE Computer Society Press.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.

Hart, S. G., & Wickens, C. D. (1990). *Manprint: An approach to systems integration*. New York: Van Nostrand Reinhold.

Heino, A., Van der Molen, H. H., & Wilde, G. J. S. (1990). *Risk-homeostatic processes in car following behaviour: Electrodermal responses and verbal risk estimates as indicators of the perceived level of risk during a car-driving task (Report VK 90-22)*. Haren, Netherlands: Traffic Research Centre, University of Groningen.

Hill S G, Plamondon B D, Wierwille W W, Lysaght R J, Dick A O, Bittner A C Jr (1987). Analytic techniques for the assessment of operator workload. In Proceedings of the 31st Annual Meeting of the Human Factors Society (Santa Monica, California: The Human Factors Society) pp 368-372

Hollnagel, E. (1993). *Human reliability analysis: Context and control*. London: Academic Press.

Hollnagel, E. (1998). *Cognitive Reliability and Error Analysis Method: CREAM*. Oxford: Elsevier.

Hollnagel, E., & Bye, A. (2000). Principles for modelling function allocation. *International Journal of Human-Computer Studies*, *52*(2), 253-265.

Jex, H. R. (1988). Measuring mental workload: Problems, progress and promises. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload* (pp. 5-39). Amsterdam: North Holland.

Jung, H. S., & Jung H. (2001). Establishment of overall workload assessment technique for various tasks and workplaces. *International Journal of Industrial Ergonomics*, *28*, 341-353.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs NJ: Prentice Hall.

Knowles, W. B. (1963). Operator loading tasks. *Human Factors*, *5*, 155-161.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 279-328). Washington DC: Taylor and Francis.

Lam, W., McDermid, J. A., & Vickers, A. J. (1997). Ten steps towards systematic requirements reuse. *Proceedings: ISRE '97: 3rd IEEE International Symposium on Requirements Engineering, Annapolis MD 6-10 January 1997*, (pp. 6-15). Los Alamitos CA: IEEE Computer Society Press.

Laughery, R. (1999). Using discrete event simulation to model human performance in complex systems. *Proceedings: 31st Conference on Winter Simulation: Simulation, a*

Bridge to the Future, Phoenix, (pp. 815-820). New York: ACM Press.

Leveson, N. (1995). *Safeware: System safety and computers*. Reading MA: Addison Wesley.

Liu, L., Yu, E., & Mylopoulos, J. (2003). Security and privacy requirements analysis within a social setting. *Proceedings: IEEE Joint International Conference on Requirements Engineering*, (pp. 151-161). Los Alamitos CA: IEEE Computer Society Press.

Manning, C., Mills, S., Fox, C., Pfleiderer, E., & Mogilka, H. (2002). *Using air traffic control taskload measures and communication events to predict subjective workload Technical Report*. Washington DC: US Department of Transportation.

Mavin, M., & Maiden, N. (2003). Determining socio-technical systems requirements: Experiences with generating and walking through scenarios. *Proceedings: 11th International Requirements Engineering Conference*, (pp. 213-222). Los Alamitos CA: IEEE Computer Society Press.

Megaw, E. (2005). The definition and measurement of mental workload. In J. Wilson, & N. Corlett (Eds.), *Evaluation of human work* (3rd ed.) (pp. 525-551). London: Taylor and Francis.

Meshkati, N., Hancock, P., Rahimi, M., & Dawes, S. (1995). Techniques in mental workload assessment. In J. Wilson, & N. Corlett (Eds.), *Evaluation of human work* (2nd ed.). London: Taylor and Francis.

Moray, N. (1979). *Mental workload: Its theory and measurements*. New York: Plenum.

Mulder, G. (1980). *The heart of mental effort. PhD thesis*. Groningen: University of Groningen.

North R.A., & Riley, V. A. (1989). W/INDEX: A predictive model of operator workload. In G. R. McMillan, D. Bevis, E. Salas, M. H. Strub, R. Sutton, & L. V. Breda (Eds.), *Applications of human performance modelling to system design* (pp. 81-90). New York: Plenum Press.

Ntuen, C. A., Park, E., Strickland, D., & Watson, A. R. (1996). A fuzzy model for workload assessment in complex task situations. *Proceedings: Third Annual Symposium on Human Interaction with Complex Systems HICS '96, Dayton OH August 1996*, (pp. 101-107). Los Alamitos CA: IEEE Computer Society Press.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, 2* (pp. ch 42). New York: J. Wiley.

Parks, D. L., & Boucek, G. P. (1989). TLAP: A method for Time Line Analysis & workload Prediction. In G. R. McMillan, D. Bevis, E. Salas, M. H. Strub, R. Sutton, & L. V. Breda (Eds.), *Applications of human performance modelling to system design* (pp. 47-64). New York: Plenum Press.

Perrow, C. (1984). *Normal accidents: Living with high-risk technology*. New York: Basic Books.

Rasker, P. C., Post, W. M., & Schraagen, J. M. C. (2000). Effects of two types of intra-team feedback on developing a shared mental model in Command and Control teams. *Ergonomics*, *43*(8), 1167-1189.

Rasmussen, J. (1986). *Information processing in human computer interaction: An approach to cognitive engineering*. Amsterdam: North Holland.

Reason, J. (1990). *Human error*. Cambridge: Cambridge University Press.

Reason, J. (2000). *Managing the risks of organizational accidents*. Aldershot: Ashgate.

Reid G.B., & Nygren, T. E. (1988). The subjective workload analysis technique. In P. A. Hancock, & M. Meshkati (Eds.), *Human mental workload*. Amsterdam: North Holland.

Rouse, W. B., Edwards, S. L., & Hammer, J. M. (1993). Modelling the dynamics of mental workload and human performance in complex systems. *IEEE Transactions on Systems, Man and Cybernetics*, *23*(6), 1662-1671.

Sheridan, T. B. (2000). Function allocation: Algorithm, alchemy or apostasy? *International Journal of Human-Computer Studies*, *52*(2), 203-216.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M. S., Linsell, M., Dymott, R., & Green, D. (in press). Distributed situation awareness in dynamic systems: Theoretical development and application of an ergonomics methodology. *Ergonomics*.

Stein, E. (1985). *Air traffic controller workload: An examination of workload probe. Technical report DOT/FAA/CT-TN84/24*. Washington DC: US Department of Transportation.

Strasser, H. (1977). Physiological measures of workload correlations between physiological parameters and operational pereformance. *Proceedings: Methods to Assess Workload, Cologne*, (pp. AGARD-CP-216\(A8-1-1A8-8)).

Sutcliffe, A. G., Maiden, N. A. M., Minocha, S., & Manuel, D. (1998). Supporting scenario-based requirements engineering. *IEEE Transactions on Software Engineering*, *24*(12), 1072-1088.

Sutcliffe, A. G., & Minocha, S. (1999). Linking business modelling to socio-technical system design. In M. Jarke, & A. Oberweis, (Eds). In *Proceedings: CAiSE99, Advanced Information System Engineering*, (pp. 73-87). Berlin: Springer-Verlag.

Van Boxtel, A., & Jessurun, M. (1993). Amplitude and bilateral coherency of facial and jaw-elevator EMG activity as an index of effort during a two-choice serial reaction task. *Psychophysiology*, *30*, 589-604.

Wetteland, C. R., Miller, J. L., French, J., OaposBrien, K., & Spooner, D. J. (2000). The human simulation: Resolving manning issues onboard DD21. *Proceedings: Simulation Conference, Orlando FL December 2000*, (pp. 1402-1406). Los Alamitos CA: IEEE Computer Society Press.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, *3*(2), 150-177.

Wickens, C. D., & Holland, J. G. (2000). *Engineering, psychology and human performance*. London: Prentice Hall.

Wierwille, W. (1988). Important remaining issues in mental workload estimation. In P. A. Hancock, & N. Meshkati (Eds.), *Human mental workload*. Amsterdam: North Holland.

Wilson, G. F., & Eggemeier, F. T. (1991). Psychophysiological assessment of workload in multi-task environments. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329-360). London: Taylor and Francis.

Winograd, T., & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Reading MA: Addison Wesley.

Woods, D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? A cognitive system diagnosis. *Cognitive Technology and Work*, *4*, 22-36.

Xie, B., & Salvendry G. (2000). Review and reappraisal of modelling and predicting mental

workload in single and multi-task environments. *Work and Stress*, *14*(1), 74-99.
Yu, E. (1994). *Modelling strategic relationships for process reengineering: Technical Report DKBS-TR-94-6*. Toronto: University of Toronto.