DRDC-RDDC-2016-P089

DEVELOPMENT OF A BASELINE WORKLOAD MODEL FOR FUTURE SUBMARINE SONAR CREWING ANALYSIS

Wenbi Wang
Defence Research and Development Canada,
Toronto, Ontario, Canada

Abstract

A computational human performance model was developed to analyze sonar operator workload using the legacy sonar system on the Royal Canadian Navy's Victoria-Class Submarine. The paper describes key components of the model, including a task network for representing operator activities, the Visual, Auditory, Cognitive and Psychomotor (VACP) algorithm for characterizing task demand, and a discrete-event simulation for predicting operator workload. Results from a simulation experiment revealed high levels of task demand imposed by the adopted mission scenario and an uneven distribution of workload among four sonar operators under a specific contact assignment scheme. The results establish a set of performance benchmarks that will be used for evaluating workload predictions made by a future model of a new sonar system, supporting quantifiable assessment of crewing options for the future sonar system.

INTRODUCTION

The Royal Canadian Navy (RCN)'s Victoria-Class Submarines (VCS) are undergoing upgrades to meet future capability requirements of the Canadian Armed Forces. Part of the upgrade includes changes to the submarine's sonar system, including the replacement of the current Type 2040 and Type 2046 systems with a newer sonar suite. Such changes will have a significant impact on sonar operators' tasks and workflow. The RCN is concerned about whether the current sonar crew configuration needs to be adjusted to fulfil future operational requirements. A comprehensive approach was adopted to address this question that involved the use of modelling, simulation, and human-in-the-loop experimentation [1]. This study reflects one project initiative which focused on the use of computational workload models for assessing crewing options. The overall plan involves development of two sets of computational human performance models, representing sonar operation using either the legacy or the new sonar system. This paper describes the construction of the 'as-is' model based on the legacy sonar system. The purpose of this model is to establish a performance benchmark to which workload predictions for the new sonar suite can be compared, supporting RCN decisions on crew configuration for the new sonar system.

MENTAL WORKLOAD AND ITS MODELLING

As a multi-faceted concept, mental workload is considered by many researchers as a hypothetical construct to indicate the interaction between operator mental capacity and task demand [2]. Extreme levels of workload (i.e., overload and underload) are a known cause of operator performance degradation, and one mitigating solution is to adjust system manning. In this study, the modeling effort focused on overloading conditions as they reveal the upper limit of crewing requirement which the RCN is mostly concerned about. A decision was made to represent expert behaviour and performance in the model which assumes sufficient task proficiency has been obtained by operators from previous training.

A software tool, the Integrated Performance Modelling Environment (IPME), was used for model construction and workload analysis. The IPME enables an analyst to represent human behaviour using task network models and to study human performance based on discrete-event simulation [3]. The modelling process started with a task analysis in which operator activities were iteratively decomposed into elementary tasks till a level where task parameters such as completion time, error rate, and mental resource demand could be reliably estimated. A task network model was then created by connecting the elementary human tasks together based on their logical and/or temporal relationships to re-construct operator behaviour. Such a model was executed using a discrete-event simulation engine to generate performance predictions such as mission completion time, failure rate, and operator workload. The approach was successfully adopted in the past for modelling sonar operator tasks (e.g., [4, 5]).

Predictions about operator workload were made in this study based on the Visual, Auditory, Cognitive, and Psychomotor (VACP) algorithm. Originally developed by McCracken and Aldrich [6], the algorithm has been implemented in workload analysis tools like IPME and the Improved Performance Research Integration Tool (IMPRINT). The algorithm requires an analyst to specify mental resource demand (expressed in four resource components) for each human task using a set of verbal anchors. A workload component rating is provided, on a scale from 1 to 7, for each verbal anchor to indicate the intensity of mental resource requirement. For example, "register/detect" and "scan/search/monitor" are verbal anchors for describing visual tasks with two levels of demand, rated as 1.0 and 7.0 respectively. According to the VACP model, a workload rating of seven is commonly adopted as a threshold (e.g., a workload redline) for judging operator overload. An operator is deemed overloaded, thus his performance is expected to degrade, when the aggregated instantaneous VACP workload rating for a resource component exceeds seven [6].

Based on a discrete-event simulation engine, an IPME model can be executed to generate prediction of task activation for each operator along a simulation timeline. For any time point, a workload score is computed for each operator by adding VACP ratings of all tasks that the operator is engaged in at that moment. Over the entire simulation, a workload profile for each operator can be obtained which provides a basis for more detailed analysis.

THE IPME MODEL

This model of sonar operation is comprised of three modules, representing a mission scenario, a crew configuration, and a network of tasks for describing operator activities, as detailed in this section.

Firstly, a series of scenario events was created to represent a six-hour mission that reflects a medium intensity and long endurance operation for which the RCN was mostly interested in its impact on sonar crewing. The main goal of the fictitious mission was to interdict drug trafficking in the Pacific ocean. A VCS was deployed together with a Canadian Patrol Frigate (CPF) and a Maritime Patrol Aircraft (MPA) to detect, classify and track illicit drugs being transported by a cartel using two vessels. These two Contact-Of-Interests were operating in an area where there existed other Merchant Vessels (MV) and Fishing Vessels (FV) that entered and exited the Area-Of-Operation at different time points. The model uses sixteen scenario events to represent a list of contacts described in the mission. In this study, a contact refers to any entity (e.g., vessel and aircraft) that generates sonar signals and requires task processing of sonar operators. The initiation of a scenario event is determined by the contact's entry time. Each event will trigger the generation of a contact entity in the model with a unique identifier. This contact entity will then "flow through" the rest of the task network in a simulation, in other words, will be processed by assigned operators in a generic task sequence involving detection, classification, status tracking, and reporting. A contact entity remains in the network till a time when the simulation clock advances past the contact's exit time, at which point, the contact entity is destroyed and removed from the network. In the current model, the operator assignment for processing each contact is pre-defined according to Subject Matter Expert (SME) inputs.

Secondly, a team of four operators was defined in this model, representing one configuration of the sonar watch team for the legacy system. The operators were labelled as 2040 High Frequency operator (2040 HF), 2040 Low Frequency Operator (2040 LF), 2046 Operator (2046 Op), and the Sonar Controller (SC). In the associated task network, these operators were assigned to various human tasks where appropriate.

Thirdly, a multi-level task network model was created to represent operator tasks. Figure 1 is a screen capture of the top level task network, which can be further classified into the following four clusters:

(1) Simulation management: two non-human tasks were used for handling initialisation and termination of a simulation run. The 'Simulation init' task enables the reset of user-defined variables when a run starts, while the 'Simulation termination' task halts model execution at the simulation clock time dictated by the scenario (i.e., 21600 seconds (6 hours) in this case).

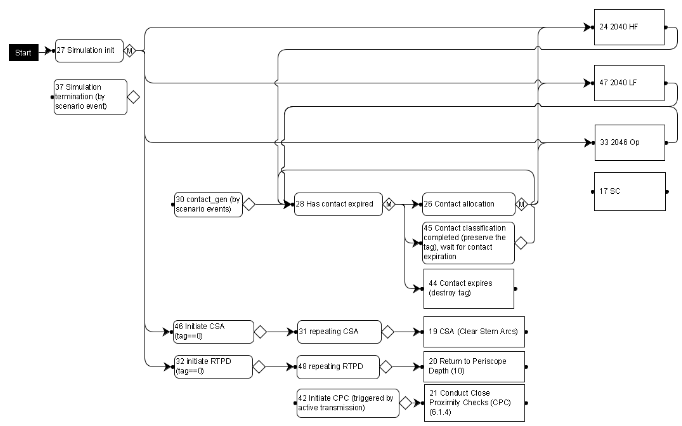

Figure 1: The top level of a task network for sonar operation.

(2) Contact management: a set of non-human tasks for controlling the generation, allocation, and status update for all contacts that are defined in the scenario. The 'Contact-gen' task is triggered by a series of scenario events each of which corresponds to a specific contact defined in the scenario. Once a contact is generated (i.e., its signature becomes detectable), it is assigned to sonar operators for further processing (e.g., classification, tracking). When the contact disappears (i.e., at its exit time), the entity is removed from the task network.

(3) Sonar operator tasks that are not associated with a specific contact, and in this model, the following three activities were represented in detail: Clear Stern Arc (CSA), Return to Periscope Depth (RTPD), and Close Proximity Check (CPC). These are common submarine manoeuvres that require the support of the sonar watch team.

(4) Four sub-networks representing tasks performed by four sonar operators respectively. As an example, the sub-network for the 2040 HF is further described in this paper.

As shown in Figure 2, this sub-network has two entry points (i.e., indicated by a triangular shape with a label 'in'). One is used to initiate continual visual and aural search and the other to process a particular contact signal. A distinction was made among three contact signal types, high bearing rate contact, active transmission, and all others, since the priority for handling each type and the detailed processing requirements are different. After a contact is classified, it is then tracked either in manual or automatic mode, until the contact signal disappears. Such tracking activities are represented in the bottom of Figure 2 for three contact types.

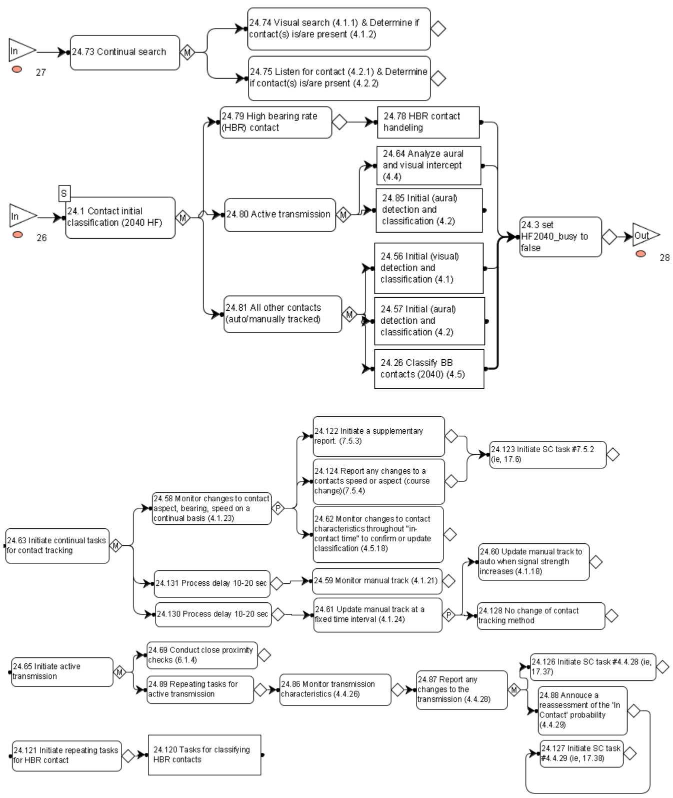

Figure 2: The sub-network model for 2040 HF.

Each operator task in this network model is represented by a set of parameters. Specifically the following task information was obtained in this study from SMEs.

Task completion time: the amount of time required for completing a task. A distinction was made in this study between tasks that are discrete or continuous in nature. For discrete tasks, the variability of their completion time was estimated for minimum, median, and maximum situations. For continuous tasks, they were treated as a repeating activity which has an active performing phase followed by an idle phase within a task cycle. In this case, the parameter task-completion-time indicates the time for completing the active phase of a repeating task.

Operator assignment: the identification of an operator for performing the task. Notably for sonar operation, a task could be assigned to different operators depending on the dynamics of a situation. For example, tasks for 2040 operators can typically be performed by either 2040 HF or 2040 LF. In this model, operator assignments were completed with assistance from SMEs based on an analysis of the scenario.

Mental resource requirement: an indication of both the type(s) of mental resource required by a task and the intensity of resource demand using the VACP verbal anchors.

RESULTS

Upon completion of the model, a simulation experiment was performed to analyze operator task activation and mental workload under the scenario condition. The data from a total of 100 simulation runs were collected and analyzed. In this paper, only results from two workload measures are presented and discussed, for the sake of brevity.

(1) Average workload

In this study, average workload is computed as a weighted mean where the duration of a workload level is treated as the weight. Figure 3 shows the average workload for four operators across four VACP components under the full scenario. A two-way ANOVA revealed a significant main effect for workload components ($F(3, 1584) = 1386.114$, $p<0.001$), operator ($F(3, 1584) = 3382.197$, $p<0.001$), and a significant interaction between these two factors ($F(9, 1584) = 326.366$, $p<0.001$). A subsequent simple effects test was performed to analyze the significant interaction.

For the 2040 HF, significant differences existed among four workload components and the order from the highest to the lowest level was visual ($M = 2.10$, $SD = 0.28$), cognitive ($M = 1.43$, $SD = 0.21$), auditory ($M = 1.06$, $SD = 0.13$), and psychomotor ($M = 0.72$, $SD = 0.44$).

For the 2040 LF, task demand from the visual component ($M = 1.82$, $SD = 0.64$) was significantly higher than the other three components. There was not significant difference between the auditory ($M = 1.25$, $SD = 0.10$) and the cognitive ($M = 1.13$, $SD = 0.59$) component. The psychomotor demand ($M = 0.86$, $SD = 0.71$) was significantly lower than the rest.

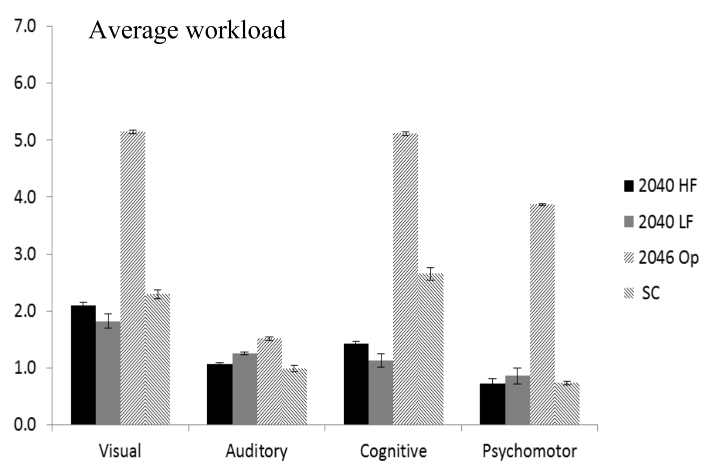

Figure 3: Average workload across four workload components.

For the 2046 Op, demand from the visual (M = 5.15, SD = 0.16) and the cognitive (M = 5.11, SD = 0.17) components was significantly higher than the auditory and psychomotor components, but the difference between them was not significant. Psychomotor demand (M = 3.87, SD = 0.07) was significantly higher than auditory demand (M = 1.51, SD = 0.13).

For the SC, significant workload differences were predicted for all four domains and the order from the highest to the lowest score is cognitive (M = 2.65, SD = 0.54), visual (M = 2.29, SD = 0.38), auditory (M = 0.99, SD = 0.29), and psychomotor (M = 0.74, SD = 0.16).

The aggregated overall workload (which is computed by adding scores from four components for each operator) predicted the highest average workload for the 2046 Op, followed by the SC, 2040 HF and 2040 LF. Among workload components, the average demand across operators indicated that visual demand is the highest, followed by cognitive, psychomotor, and auditory components in this order.

(2) Percentage of time overloaded (TO%)

TO% was calculated to measure the occurrence of task demand that exceeds an operator's mental capacity. In this study, overload is operationally defined as a condition where an operator's instantaneous workload for a domain exceeds a threshold of seven. The measure is computed by dividing the aggregated operator overloading time by the total duration of the scenario.

Figure 4 shows the percentage of time overloaded for four operators across four workload components. A two-way ANOVA revealed a significant main effect for workload components ($F(3, 1584) = 901.072$, $p < 0.001$), operator ($F(3, 1584) = 2315.259$, $p < 0.001$), and a significant interaction between these two factors ($F(9, 1584) = 295.255$, $p < 0.001$). A follow-up simple effects test revealed a pattern of results similar to the average workload measure.

For the 2040 HF, significant differences were revealed for this measure across workload components. The highest score was obtained for visual (M = 6.95%, SD = 1.79%), followed by cognitive (M = 2.74%, SD = 0.75%), auditory (M = 1.45%, SD = 0.44%), and psychomotor (M = 0.26%, SD = 0.51%) components.

For the 2040 LF, the percentage of time overloaded was significantly higher for the visual component (M = 4.71%, SD = 4.07%) than the rest of the components, which were not statistically different among themselves (auditory: M = 1.4%, SD = 0.34%; cognitive: M = 2.39%, SD = 4.78%; psychomotor: M = 1.53%, SD = 4.49%).

For the 2046 Op, significant differences were predicted for this measure across four components where the highest score was visual (M = 30.53%, SD = 2.10%), followed by cognitive (M = 23.72%, SD = 1.70%), psychomotor (M = 12.91%, SD = 1.37%), and auditory (M = 3.09%, SD = 0.42%).

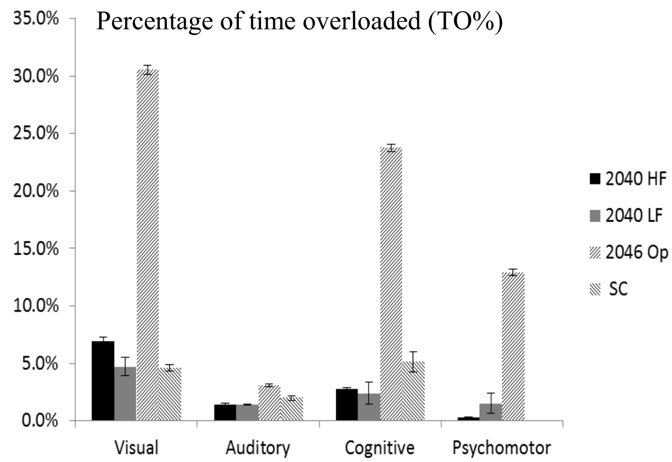

Figure 4: Percentage of time overloaded across four workload components.

For the SC, scores for the visual (M = 4.60%, SD = 1.42%) and the cognitive (M = 5.12%, SD = 4.47%) components were significantly higher than the other two components, but between them there was not statistically significant difference. The score for psychomotor (M = 0.03%, SD = 0.07%) was also significantly lower than the auditory component (M = 1.98%, SD = 0.90%).

DISCUSSION

The key predictions generated from this simulation experiment are summarized as below.

1. The adopted operational scenario is challenging. It is predicted to impose a high level of task demand and cause periods of mental overload for all operators in this 6-hour mission.

2. The contact assignment scheme used in this study will generate an uneven distribution of workload among four operators, with the 2046 Op severely over-tasked than the rest of the team.

3. For the current sonar system, major sources of workload originate from visual and cognitive activities. And this is consistent for all four operators.

Besides these general predictions, the model also revealed different patterns of task activation at the individual level that enables an analyst to trace down the key task parameters contributing to operator workload.

For operators of the Type 2040 and Type 2046 systems, their primary patterns of task activation are similar which is comprised of a constant contact search activity throughout a mission and a series of contact management tasks (e.g., classification and status tracking) after a contact is detected.

For contact search, there are two main factors that affect operator workload: the time for completing a search and the search cycle (i.e., which determines how often a search is performed). If one assumes a constant level of operator proficiency, then the search time is largely determined by the sonar system (i.e., configuration of the system for visual or aural search) whereas the search cycle is often influenced by the mission environment (e.g., search more often while sailing in hostile water). For contact management tasks, mental demand is mainly determined by the number of simultaneously tracked contacts and their types. Different types of contact are analyzed/tracked differently and the task difficulties differ. As the contact number increases, the heightened level of workload is represented in this model by interference in tracking, for example a saturation of working memory which leads to a higher level of cognitive demand. A review of the model execution trace file (which logs the initialization and termination of all tasks in a simulation run) indicated contact tracking is accountable for a majority percentage of all tasks these three operators perform and the tracking of multiple concurrent contacts is the main source of mental overload.

In contrast, the pattern of task activation for the SC is quite different. All SC's tasks represented in this model were reporting activities. Either the SC receives a report from any of the three operators or he delivers one to the Fire Control operators or the Officer-Of-The-Watch. These activities are typically triggered by a particular event, e.g., detection of a contact, completion of contact classification, or change of contact status. This model revealed that the SC's workload is mainly influenced by a contact's status update interval (which determines how often the status of a tracked contact is updated). Since detailed contact properties were not represented in this model, consequently there was no information whether the status of a traced contact was changed. The model was created with the assumption that each tracking update by the 2040 and the 2046 operators would require the SC to initiate a corresponding reporting task. As a result, workload predictions for the SC were likely inflated. However, this was considered acceptable in this study since the requirement for an SC in the future sonar watch team is robust and therefore this role is not the main focus of the current analysis.

Workload predictions generated from this study establish a set of performance benchmarks to which outputs from the to-be system model (where crewing options will be represented and manipulated) can be compared. Such a comparison is meaningful when models of the as-is and to-be systems are constructed based on the same modelling methodology, which includes using the same set of assumptions and mission scenarios. It is useful to note that the current legacy sonar system is considered by many SMEs to be very demanding in terms of mental demand. Therefore if a particular future crewing option generates workload prediction that is worse than this baseline for all operators, it is reasonable to conclude that the option is not acceptable. If the prediction is mixed, that is some operators are greater than the baseline while others are less, then a closer examination of the task assignment scheme is further required. In all cases, a detailed investigation in individual workload domains is likely needed, particularly for the visual domain, due to an expected increase of requirement for visual processing by the new sonar system.

To sum up, the goal of this study was to develop a baseline human performance model that supports quantifiable predictions on the effectiveness of crewing options for a future sonar system. This report documents the methodology that was adopted to construct the baseline model, including the creation of task networks for representing operator behaviour, the use of the VACP algorithm for characterizing task demand, and the prediction of operator workload based on discrete-event simulation. Results from the simulation experiments confirmed that operators' workload predictions were sensitive to the independent variables that were manipulated. A set of workload benchmarks were established from this study that can be used for evaluating workload prediction made by the future model of the new sonar system constructed based on the same modelling methodology.

Acknowledgements

The author thanks Dr. Jacquelyn Crebholder for coordinating the research effort, and Dr. Michael Matthews and Courtney Kersten-Kwan for conducting a comprehensive task analysis of sonar operation, the output of which was used for parameter configuration in this IPME model.

[1] Crebolder, J., Hunter, A., Hazen, M., Stoddart, M., Theriault, J., Chow, R., Wang, W., Tanner, L. and Mansfield, L. (2013), An investigation of crewing for the Canadian Navy submarine sonar system, From the TTCP HUM TP18 Crewing systems for maritime warfare workshop.

[2] Gopher, D. and Donchin, E. (1986), Workload: An examination of the concept, In K.R. Boff, L. Kaufman and J.P. Thomas (Eds.), Handbook of perception and human performance, Vol. 2: Cognitive processes and performance, pp. pp.1-49, Oxford, England: John Wiley & Sons.

[3] Alion Science and Technology (2008), Integrated performance modelling environment: User guide, Alion Science and Technology.

[4] Matthews, M., Bos, J. and Webb, R.W.G. (2003), A prototype task network model to simulate the analysis of narrow band sonar data and the effects of automation on critical operator tasks., (DRDC Toronto Contract Report CR2003-131) Defence Research and Development Canada.

[5] Matthews, M.L., Bos, J. and Bennett, J. (2004), Experimental validation of a model and simulation of generic sonar system operator tasks, (DRDC Toronto Contract Report CR2004-131) Defence Research and Development Canada.

[6] McCracken, J.H. and Aldrich, T. (1984), Analysis of selected LHX mission functions implications for operator workload and system automation goals, US Army Research Institute.