MODELING THE EFFECTIVENESS OF TOOLS TO ASSIST SONAR OPERATORS

Michael L. Matthews, Jeff Bos
Human*systems*® Incorporated
Guelph, Ontario, Canada
Jacquelyn M. Crebolder
Defence Research & Development Canada Atlantic
Dartmouth, Nova Scotia, Canada
Sharon McFadden
Defence Research & Development Canada Toronto
Toronto, Ontario, Canada

The task of building the underwater picture from sonar data is made complex by high volumes of noise and multiple data that arrive from a variety of acoustic sources detected at great distances by modern, sonar equipment. Typically, acoustic sources from ships have a complex spectrum consisting of several base frequency components and related harmonics. The task for operators is to analyse the data to determine if there is a pattern that represents the signature of a known source, thereby leading to identification of a vessel. Since the task can be highly labour intensive automated decision aids may be of value to the operator. This project addresses how to predict and optimise the impact of new technologies in system re-design by using a modeling/simulation approach to operator-system functionality. A generic sonar analysis process was simulated and the effectiveness of a decision aid evaluated. The improvement in performance predicted by the aid was then validated experimentally.

INTRODUCTION

This paper addresses the issue of how to evaluate the effectiveness of new tools to assist sonar operators whose task is to build an underwater picture to guide tactical decisions in the Canadian Navy. The underwater picture contains information related to the identity, position, course and speed of surface and subsurface vessels detected by acoustic sensors.

The specific goal of the project[1] has been to assess the effectiveness of operator aids to enhance the process of identifying vessel signatures from sonar patterns, using a task network model of the sonar domain. A simulation approach to finding practical and effective solutions to improving human-system effectiveness was chosen for a variety of reasons. First, the operational system is complex and not readily adaptable to "bench testing" new software and hardware components for evaluation purposes. Second, operational systems are normally deployed and not available for R&D purposes. Third, trained operators, who might be used to evaluate system improvements, are few in number and also not readily available.

The task domain

Sonar data received by acoustic sensors arises from a variety of biological and mechanical sources and are affected in transmission by a variety of environmental and oceanic factors. As a result, the pattern of sonar data received from a large vessel may contain acoustic frequencies associated with many sources such as engines, shafts, propellers, generators, pumps, switches and other electromechanical devices. Each source will likely generate not only a base frequency (e.g. 60Hz) but also harmonic components at higher frequencies.

The task for the sonar operator is to distinguish among background noise the patterns of frequencies that represent likely sources and then to analyse these patterns to determine if they match the signature profile of known vessels. Information is typically provided to operators in the form of a frequency (x axis) by time (y axis) by intensity (z axis) display that updates at regular intervals with the most recent data displayed at the bottom. The display normally has a background level of "noise" that represents random signals arising from the underwater environment and detected by the system. Signals from sonar sources of a given frequency will appear on the display as vertical lines whose length corresponds to their duration and luminance to their signal strength.

The signature of a single vessel may comprise anywhere between 25 and 100 lines depending upon the number of acoustic sources active, the distance to the sensor and a variety of oceanic variables. Further, the signature may not be the same fixed pattern, but oceanic conditions may cause the base frequency to appear shifted to a different frequency and may also influence the ratio of the intensity of the harmonics to the base frequency.

Thus, for the most part, there is *no simple unique visual pattern* that represents a vessel and hence a visual identification by pattern recognition alone is not feasible. This is particularly the case when several vessels are picked up by sensors and their overlapping patterns are present on the display. Thus, the task becomes one of serial analysis of the lines aided by a set of tools, such as a variable harmonic cursor that allows lines that are harmonically related to be determined more readily.

Complicating the process further is the fact that no single display can represent 360 degrees of coverage of the ocean.

[1] This project was funded by a contract to Humansystems Incorporated® by DRDC Atlantic and DRDC Toronto and the current paper is based upon the final report by Matthews, Bos and Bennett (2004).

Therefore, individual sectors, or beams, (radiating from the ship responsible for the picture building) of the environment are filtered and each allocated to a given window that can be brought up on the display. Typically, the ocean may be divided into anywhere from 20 to 100 such beams, although for present purposes we have assumed full coverage with 44 beams.

In order to build a complete picture of the underwater environment an operator must successively step through and analyse each window associated with a beam, a process that may take hours under many operational circumstances.

A further complication is that the ship building the picture normally travels in a task group (TG) comprising five or six vessels in reasonably close proximity. Each of these vessels will generate its own acoustic pattern that will also "flood" the operator's display with lines of data. Under some circumstances these lines can number in the thousands. Further, the pattern of these lines will change over time depending upon the geospatial relationship between the vessels emitting the sounds and the vessel doing the detection, their speeds as well as the intervening oceanographic conditions.

Essentially, there are two critical picture compilation tasks to be done. The primary task is to build the underwater picture by identifying vessels of interest and their associated signatures. As part of this primary task, each line detected on a beam must be identified as belonging to one of three categories: "*known*" - a part of a target signature that is unambiguously recognised based upon information held in a knowledge database; "*unknown*" – a line that cannot be definitely associated with any known signature and "*possible*" – a line for which there is some, but inconclusive evidence, that it belongs to a known signature. The operator is required to tag each and every line on the current beam into one of these three categories, enter the information into a log and then move on to examine the data from the next beam. For simplicity, we will refer to this task subsequently as *search/id*.

The secondary critical task is to identify and eliminate from further analysis *the known vessel signatures* arising from the TG, thereby enhancing the ability to do the first task and the efficiency with which it can be done. This process of elimination is often referred to as "*sanitizing*" the display.

The sanitization task is essentially a top-down process, whereby the operator uses known information about target signatures of TG ships to direct the search for locating on which beams the individual lines can be found. The process ends with the operator entering into the log the locations and line components of all of the updated signatures. In practice, the operator looks for evidence of the individual sound sources that comprise each vessel's signature, such as acoustic lines associated with engines, shafts, propellers, generators and other electro-mechanical devices. Once all of the individual sources are found the vessel as a whole, and all of its associated lines, are then fully identified.

Because of the continuous and high work load demands associated with each of these tasks, in a typical operational environment, one operator will be assigned to build the picture (search/id) and a second to sanitize the display. When the latter operator has finished this process, she/he is then available to help out the operator building the picture.

Under some operational conditions, it would not be unusual for the sanitization process to take hours to complete, and could require the almost constant attention of one operator. The process itself does not place heavy intellectual demands on the operator, but is often referred to as "brain dead" and is known to lead to boredom and data entry errors, particularly with extended time on task.

This is a somewhat inefficient use of personnel resources, and hence the motivation for the present project was to identify possible operator aids that could assist in the sanitization process, thereby freeing up operator resources to deal with the more tactically critical task of underwater picture building. The efficacy of such aids was then to be evaluated using modelling and simulation.

The baseline model developed was also seen as a way of addressing the potential operational impact of future system re-design options, including issues such as: increasing the beam resolution and number of beams, changing the number of displays and their size, color coding of information, re-assignment of operator tasks and personnel redeployment.

METHOD

Building the model

Using existing task analyses of navy sonar systems (Matthews, M.L., Webb, R.D.G and Woods, H., 2001) and with the assistance of a subject matter expert, who was an experienced Navy sonar trainer, critical tasks were identified that comprise the processes of the detection and identification of ships from their radiated acoustic patterns. These tasks then formed the basis for creating decision-action diagrams to represent the sequential operations performed and decisions made.

The Integrated Performance and Modeling Environment (IPME) software (Copyright Micro Analysis And Design) was used to create the task network model. For each task, estimates were generated of the time to complete (means and variances) tasks, probability of success, consequences of, and tasks influenced by, failure and operator workload on visual, auditory, cognitive and psychomotor dimensions (VACP). VACP is an attentional demand algorithm based upon the task loading for an operator within the four separate channels and estimates the demands on human processing resources. To achieve a VACP rating, each operator task is rated with respect to the weighted task demand that appears appropriate for the specific task requirements for each of the four independent channels. Tasks were assigned workload ratings on a seven point scale by comparing them to normative values of the IPME workload scale. (Aldrich, Craddock, and McCracken 1984, Bierbaum, Szabo, and Aldrich, 1987).

Parameters of the model

The model simulated a two-operator environment in which one operator detected sonar lines of interest, analysed them and attempted to make identification from the observed pattern, while the other operator sanitized the array. Both operators entered their analysis of each line into a handwritten log that contained a number of fields of information relevant to the properties of the line.

System Hardware

In order to approximate the realities of existing sonar systems the hardware constraints were set as follows: (i) two high-resolution monitors; (ii) processed, narrow band, sonar data were represented on 44 beams; (iii) data were represented as frequency information over time and (iv) aural presentation of sonar data was available to operators to enable further analysis

The underwater model and sonar contacts

The starting assumption was that a number of sonar sources, each of which has a variety of sound frequency components that arrive at the sensor, are presented on a display, or can be heard through headphones or speakers. Thus, the sonar database comprised a number of signal representations that corresponded to sources that, when processed by the simulated operators, should result in identifications of non-mechanical, unknown, possible, or known.

Contact spatial and temporal dynamics

To simplify the simulation, the baseline model did not represent the complexities of TG movement through the ocean, therefore sonar sources other than from the TG were represented on a single beam only. While this may be unrealistic of many operational conditions it does faithfully represent the task of detecting and identifying sonar contacts that are represented on a single beam.

In order to simulate the temporal parameters of acoustic data, the data representations of the source targets were defined as having a finite, temporal lifespan that entered into the simulated underwater environment at varying points in time. In this way, the information available to the operator changed over time and, if the data were not processed before they expired, contacts would be missed or misclassified. Once being available for processing, each target had a life span of between 200 and 600 seconds that was randomly assigned by a model function.

Sonar data types

Four characteristics of sonar data were modelled as follows; their probability of occurrence is shown in parentheses: (i) source is a true target (.22); (ii) source is noise (.22); (iii) sonar data require the operator to wait for additional screen updates (i.e. the signal is too brief to allow analysis to occur) (.33); (iv) sonar data scroll off the display before the operator has time to make an identification. (.22)

Target identification characteristics

To capture the essence of the task, and based upon Subject Matter Expert (SME) input as to what would be representative, 20 data lines on each beam (associated with a single target) were required to be present before the target could be identified. If such lines were not present, then the target would either be missed or classified as unknown. This approach was chosen to generate identification times that were consistent with the wide range of actual identification latencies that occur under operational conditions.

Taskgroup Data

These data result from the sensor array picking up, on a continuous basis, acoustic signals generated by noise sources on all five ships that comprised the TG. The actual number of lines that might be generated by the entire TG under operational conditions could range from the hundreds to the thousands. Based upon SME input, two values were chosen for the numbers of lines per ship to be processed, representing a low and moderately high TG “noise”. For the low load condition, there were 25 lines per ship and each ship was represented on five different beams; therefore, for the entire five ship TG there was a total of 625 lines. For the high condition there were 100 lines per ship, for a total of 2500 TG wide.

The operator model

As indicated above, there were two functionally separate elements of the operator model – the basic search and analysis for contacts of interest and the sanitization of the array of the known lines arising from the TG. Both operators did the basic search/id and one operator was additionally assigned to sanitize the array at regular intervals throughout the watch. In the standard search/id process the operator searched through the beams to detect sonar signals of interest, identified the source and logged the relevant data for each line detected. One operator sequentially search up the beams from 1-44 and the other from 44-1.

At the start of the watch (i.e. when the simulation commenced) one operator "sanitized the array" while the other operator searched. Once this sanitization process was complete, the operator also searched for contacts. The sanitization process was required to recur several times during the simulated watch and the recurrence interval was set at 20 minutes, based upon SME input for the conditions that were being simulated.

Modeling operator tools.

Two labor-intensive, error-prone tasks of the operational environment were considered to be prime candidates for automation or operator assistance, namely the sanitization process and the logging of data. The decision was made to concentrate initially on the former process by considering the kind of tool that would help the operator to perform this process more efficiently.

To review, the current sanitization process is a top-down serial search for expected acoustic sources that comprise the signature lines of each TG ship on expected beams followed by an update of the log to reflect the actual current line data. This process is repeated for all ships until all their lines are accounted for. Obviously, this task involves a lot of back and forth checking between the log and the display, and requires the mental translation of written frequency values and beam numbers in the log into where to look, and what to look for, on the actual display. It seems feasible that a tool could be created that would use the information in the log to create a visual template, or pattern of lines, that when overlaid on the display would provide instantaneous feedback as to whether those lines were actually present on the beam.

The operator would then use this template to make a judgment of whether the signature is present or not and, if present, a smart data entry process captures the required information for the log. If the signature is not found, the operator analyses the line pattern found and then updates the database if the pattern is found to be at variance from the predicted identity.

Thus, from the operator's perspective, the visual analysis component of the sanitization task is greatly simplified and the human proficiency in visual pattern matching of complex features is capitalized upon.

Accordingly, a sub-network of the overall model was developed to analyse and simulate how such a template aid might work; the model was then run and debugged.

Model execution

Once the overall model was found to be error free it was run with 30000 data updates generated at a rate of 1 per second, thereby simulating 8.3 hours of real time operating conditions, during which time approximately three thousand lines of sonar data were entered for analysis. Ten model runs were executed for each of the conditions in which the sanitization aid was available (assisted condition) or not (baseline condition). Further, two conditions of TG load were simulated, in which each TG ship was represented by either 25 or 100 lines that required analysis and identification. These conditions were chosen to represent a range of operational conditions from moderate to high intensity based upon SME input.

RESULTS

Data collected from the model fell into two main categories, system performance and estimates of operator workload. For present purposes system performance data will be presented in terms of the number of times the sanitization process was completed and the total number of contacts identified and logged.

System performance

The magnitude of the performance gain in the sanitization process resulting from the template assistance resulted in approximately 4.3 times as many sanitizations being completed in the low load condition, and 8 times as many in the high load condition.

Given that the assisted condition produced more efficient sanitization performance, we should expect to see some impact when the resulting residual spare capacity of the sanitizing operator is redirected to the search/id task. This was confirmed by the identification data which showed that for both the 25 line ($t=13.38$, $df=18$, $p<.01$) and 100 line ($t=14.47$, $df=18$, $p<.01$) conditions there was a significant increase in the number of contacts logged when the task of TG sanitization was template assisted.

Operator Workload

Mean workload scores for each operator on each of the workload dimensions for each of the experimental conditions were computed based upon the individual 35225 values computed by the IPME model during each of the 10 runs.

Statistical analysis of the data was conducted using two-factor (baseline/assisted and number of lines per TG ship) analysis of variance (ANOVA). Where necessary, supplemental comparisons were made using t-tests.

For both TG conditions, there was a small trend for lower workload ratings in the template-assisted condition for visual, cognitive and psychomotor components, and a reverse trend for the auditory workload component. The effect of the number of TG lines to be analysed was not consistent. For all workload measures, except the cognitive (where the reverse was true), workload was slightly higher in the 100 line condition. However, these effects were quite small, typically of the order of less than .1 on the 7-point workload scale, but were statistically significant because of the small variance between simulation runs. Significant interactions for visual and cognitive workload scores reflected a larger effect of the template assistance under the 100 line condition, compared with the 25 line condition. This difference was in the opposite direction, however, for the psychomotor scores.

DISCUSSION

The results of the simulation show that the impact of adopting a smart, visual template to assist the sanitize process resulted in a significant effect on the speed with which this process can be executed. As a result, more sanitization cycles could be performed in a watch and there is additional residual capacity created, such that the "operator" performing the sanitization task is able to work on the search/id task thereby improving the overall rate of identifications for that task also. The results showed minimal impact on mean predicted operator workload across a test run (i.e. watch period) resulting from the use of the template for the sanitization task.

While the results of the simulation clearly pointed to the potential value of the aid, there was some concern of their validity from two perspectives. First, while the model function parameters for known sonar tasks were based upon recommendations from an experienced SME, the values used to estimate task performance with the template were derived from an analysis and estimation from published human factors data for similar task contexts, together with input derived from the experience of the human factors team. Second, the values used for the workload for each task were based upon the IPME scale values, and there may be some issues with how well such values generalize to other task situations.

Accordingly, it was decided to assess the validity of the modelling and simulation results by collecting human performance data with human operators performing the sanitization task under baseline and assisted conditions. The critical data would comprise both performance effectiveness (in terms of throughput or sanitization rate) as well as perceived workload for each task component. Consequently, a simple, computer-based simulation of the sonar sanitization task was developed that would allow us to train non-Navy personnel to the required standard of proficiency and then conduct the validation study with such personnel.

VALIDATION EXPERIMENT

This experiment attempted to replicate the tasks of TG signature identification and data entry. A simulated, sonar, frequency-time-intensity display was developed on a PC platform and a database of ship signature patterns was created. Each signature pattern comprised a representation of five different ship sound sources, each with five frequency components. The patterns of frequencies for each source were harmonically related, but different multipliers were used. Three variations in the basic signature pattern were produced by changing the base frequency. By doing this we hoped to create a sufficiently complex data set that would force analysis of the line components rather than memorization of unique patterns. There could be from 1-3 ship signature patterns on a beam and these patterns were always accompanied by a variable number of "noise" lines, indistinguishable from the actual signature lines. The total number of lines on a display was always 125, the ratio of target signature lines to noise lines therefore ranged from 25/100 to 75/50. This display "load" was comparable to the 25 line TG condition of the network model.

Twenty adult volunteers were trained to criterion in the signature identification process and assigned to either a baseline or template assisted condition. Subsequently, the subjects conducted two, separate, two-hour sessions in which they searched through 44 beam displays for potential target signatures in one of two ways. In the *baseline condition* (which simulated existing practice) subjects analysed signatures by first analysing component noise sources for each signature then identifying the ship source that was most appropriate. They did this by selecting a harmonic cursor (from a set) that could be used to identify frequency relationships among groups of five lines. Thus, to identify a specific ship signature, they would first need to analyse the lines for each of five potential noise sources. Once identified, they entered the frequencies of each of the signature components into a paper log. In the *template assisted condition*, subjects could select from an array of templates that represented a complete ship signature pattern to overlay the beam. In this way, an entire ship's signature could be determined by the selection of the appropriate template. For both conditions, a signature match was visually apparent by an increase in luminance for the component lines when the template was completely in alignment with the signature components. The log data entry for the template assisted condition was streamlined so that the operator simply indicated the template number that matched the pattern. Every 10 minutes during the trial a pop-up window containing a workload rating scale (1-7) was presented, with a check box for subjects to indicate the current sub-task they were performing and the associated workload on overall, visual, cognitive and psychomotor dimensions.

RESULTS

The major data of interest for both conditions were the time required to identify ships, the frequency of analysing the entire 44 beam array and operator workload ratings. By analysing the identification times as a function of the number of ships per beam, we showed that the slope of this function decreased from 126 sec/ship in the baseline condition to 17.6 in the assisted condition. The frequency of sanitization was computed by subtracting out the time required for data logging and the results showed an increase in frequency of the sanitization process in the template assistance condition by a factor of 4.6, which compared closely to the model prediction of 4.3. By extrapolating the human performance data to the equivalent high load condition of the model, we found that, the model predicted a performance gain of a factor of 8, whereas the human performance data suggested a gain of 7.4.

With respect to workload, visual and cognitive ratings provided by the subjects were consistently lower than the model norms for workload used by IPME. However, the pattern for reduced workload under the assisted condition predicted by the model was again found in the human ratings for specific tasks. For example, workload was lower in the assisted condition for the tasks of searching for lines and log entry. However, as might be expected, visual workload ratings for the task of template matching were higher in the whole ship template condition.

CONCLUSIONS

We were able to conclude generally that the modeling approach provided a reliable and valid method for estimating the effectiveness of this particular decision aid on system performance and operator workload. However, some revision of the IPME scale values for cognitive and perceptual workload is suggested by the results obtained. Finally, the creation of this comprehensive baseline model will allow quantifiable estimates to be made on the effectiveness of other future system re-design options.

REFERENCES

Aldrich, T.B., Craddock, W., and McCracken, J.H., (1984) "A computer analysis to predict crew workload during LHX scout-attack missions Volume 1." MDA903-81-C-0504/ASI479-054-I-84(B), US & Army Research Institute Field Unit, Fort Rucker, AL.

Bierbaum, C.R., Szabo, S.M. and Aldrich, T. (1987), Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model. Technical Report Number ASI609-302-87, Anacapa Sciences, Inc, Fort Rucker, Alabama.

Matthews, M.L., Bos, J and Bennett, J. (2004) Experimental Validation of a Model and Simulation of Generic Sonar System Operator Tasks. *DRDC Toronto Report # CR-2004-131*

Matthews, M.L., Webb, R.D.G and Woods, H. (2001) Function Analysis of TIAPS, Update of CANTASS Functionality and Human Factors Review of OMI Design for Sonar Combat Systems. DCIEM CR2001-045.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Canadian Navy for providing access to subject matter experts and in particular to Chief Petty Officer Riggs whose valuable knowledge and enthusiastic co-operation made the development of this model possible.