Using Cognitive Work Analysis to Drive Usability Evaluations in Complex Systems

Aren Hunter and Tania Randall

Defence R&D Canada – Atlantic, Dartmouth, Nova Scotia, Canada
{Aren.Hunter,tania.Randall@drdc-rddc.gc.ca}

Abstract. This paper describes how Cognitive Work Analysis (CWA) can be utilized to support a system-level usability analysis. Overall, we suggest that CWA-derived work tasks should be considered as useful in guiding the development of scenario-based usability questions. We also suggest that usability practitioners be mindful of the importance of time consistencies in developing scenarios and in the appropriate timing of questions throughout the scenario. When evaluating the results of a system level usability experiment it is useful to view the results in light of cognitive and attentional biases.

Keywords: Attention, Biases, Cognitive Work Analysis, Mental Models, System, Usability, Work Tasks.

1 Introduction

1.1 Overview

The intent of this paper is to profile the use of Cognitive Work Analysis (CWA) as a tool in performing a system level analysis. Although the primary objective of this research project was to assess the usability and functionality of an Integrated Information Display (IID), this particular paper addresses the process we undertook to perform the usability assessment. This paper is an attempt to fill the notable gap in the literature with respect to system level usability evaluations. As such, we believed that the use of CWA to inform our assessment was unique, generalizable and worth reporting. We also believed that there was value in reporting on the use of mental models and cognitive biases when evaluating a design. This paper will also shed light on the valuable aspects of CWA and the challenges in using it to define system level usability parameters. The process we describe is generalizable and valuable for researchers in various domains.

1.2 Background

Time-sensitive military missions often require operators to incorporate and process data that are distributed and presented in a variety of formats. In an attempt to understand and reduce demands on "information analysis" (p.65) [1] in submarines,

D. Harris (Ed.): EPCE/HCII 2013, Part I, LNAI 8019, pp. 39–48, 2013.

Defence R&D Canada – Atlantic (DRDC Atlantic) designed an IID to aid the warfighting capabilities of the Officer of the Watch (OOW) [2]. This IID is the focus of the following usability analysis.

As part of the IID design process a CWA was completed [2]. The CWA allowed for an analysis of the OOW's work domain. In general, CWA is used in these contexts to expose work restrictions that define decision making [3, 4]. The majority of researchers use CWA to gather information to aid the design of an interface for a complex system [5]. CWA extracts information requirements that are needed by operators to make effective decisions. In essence, the information requirements provide an explanation of what information is important in the work domain [2]. Once this is complete, the second step in display design is to determine how the information should be presented. The challenge at this point becomes translating hundreds of information requirements into meaningful graphics that support operator decision making to complete various work goals. To do this effectively the information needs to be integrated in a way that defines the limits of the work system [4] and minimizes pressure on cognitive and attentional resources. Unfortunately, CWA techniques have not been optimized to easily turn information requirements into a usable design [5].

Information from the IID was categorized in the following eight categories: Date and Time Group, Primary Ownship Status, Sound Velocity Profile, Tactical Picture, Contact Management, Schedule of Events, Alerts/Alarms, and Dynamic Information Area [6]. The layout of the categories on the IID are depicted in Figure 1. Due to issues related to intellectual property we are not able to present the readers with a fully mocked up display. However, certain isolated components of the display will be presented in this paper.

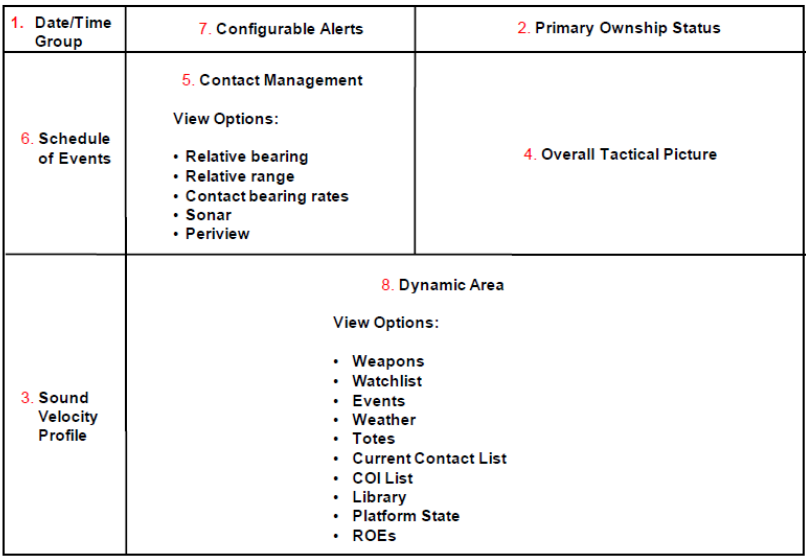

Fig. 1. Represents the layout of information in the IID and the titles assigned to each [2]

Following the initial design layout an independent team conducted usability testing to assess usability and functionality prior to implementing the completed display. For this assessment, each of the areas depicted in Figure 1 was evaluated by submariner subject matter experts (SMEs) for issues related to their functionality and usability. To simulate the "dynamic" aspects of the display a series of five IID screenshots (scenes) were developed. Each of the screenshots included time relevant changes to the IID to mimic what would happen if the IID was fully functional. The scenes were manipulated by the researcher using a button press at the appropriate time in the evaluation. For example, Figure 2 depicts the change in tactical information in area four in the display from scene three (time 12:57) to scene four (time 14:06) [6]. This was an effective way to implement some level of dynamic fidelity without having to feed real data into the display. At each new "scene" we asked the user specific questions about the content of each display area, the functionality and usability of each display area, the anticipated content changes in the elements and the expectations for change in the next scene.

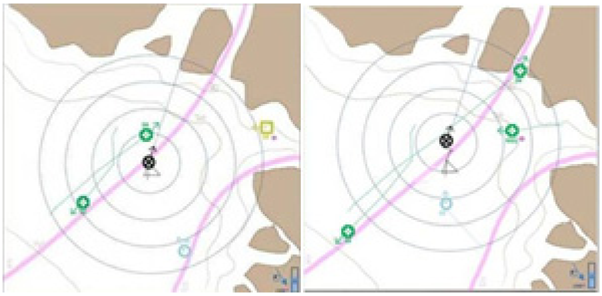

Fig. 2. Scenario for the tactical picture at 12:57 (scene 3) and 14:06 (scene 4) which depicts the movement of contacts across time periods [6]

2 System Level Evaluation

Often a newly designed display is built to replace an outdated one allowing for a baseline evaluation between the old system and the new system [7, 8, 9]. However, since the IID is a new concept there was no old display available for comparison. As such, the display was evaluated based on these three criteria: CWA derived requirements, mental models, and attention biases. In the sections to follow we will give examples of tests using each of these criteria. It should be noted that traditional usability testing is efficient when evaluating the one-to-one relationship between elements, but these techniques are not easily applied to complex integrated displays. The complexity of integrated displays requires both an evaluation of individual components (i.e., a particular gauge) and a "holistic" evaluation of the system (i.e., the integration of information) [10]. The remainder of this section will outline how we tested this new system. It also outlines what portions of the available CWA were most effective in supporting our evaluation.

Roth and Eggleston [7] indicate that complex system usability needs to be driven by a "work-centered evaluation" (p.204) to determine the value of the display in supporting work functions and work tasks. These types of evaluations require an understanding of specific work tasks and contexts, cognitive and attentional resources, task complexity, and performance expectations which matches well to the outputs of CWA [7]. This requires that scenario appropriate metrics and questions be designed for use in the usability evaluation [11]. Understanding when, and under what conditions, the display supports and overwhelms cognitive and attentional resources is also vital in determining the limitations of the display [11].

3 CWA Derived Requirements

3.1 Scenario Development

The first task, prior to beginning the evaluation, was to develop a detailed scenario with enough complexity to allow for realistic work centered decisions [9, 11]. The literature is vague on guidelines for developing these types of scenarios, but we found that scenarios or storyboards used during the CWA were sufficient enough in detail to support “work-centered” decision making. Dynamic scenarios, regardless of domain, require realistic timelines and event sequences. Our experience suggests that users are particularly sensitive to deviations in time and the progression of elements across time. As such, maintaining predictability across time are key factors in scenario realism. When inconsistent patterns, such as slight target movement (jumping too far ahead or not far enough) are present in a scenario then the SME’s mental model becomes unreliable. This becomes particularly important for scenarios that require users to maintain awareness and predict the future status of the system as is often the case with dynamic displays [12]. Another important aspect comes from the naturalistic decision making literature which suggests that realistic time constraints are key to encouraging users to make realistic decisions. When testing a new system the researcher should ensure that realistic time constraints are in place to force SMEs to make decisions that provide a reasonable solution [13]. Providing them with too much time is not realistic and does not accurately reflect the way real world decisions are made. In effect, time is one way that researchers can induce ecological validity into scenario-based decision-making.

3.2 Question Development

Once the scenario was established a series of questions were developed. Questions were required as part of the system level analysis to assess if the IID supported the level of decision making that was intended. It is important to note that question development was one of the most challenging phases involved in the system level evaluation. It was also the phase of the analysis that leveraged the most from the CWA. In order to design system relevant questions we utilized the work functions that were extracted from the CWA design work [6]. In total, ten work functions were assessed for their applicability to question design. For each of the work functions a list of high-level and low-level work tasks were also identified. From these we narrowed down the list of work functions and tasks to a set of scenario relevant functions and tasks. In the end, we had four relevant work functions, four high-level work tasks, and 22 low-level work tasks. A sample of the scenario relevant work functions and tasks are presented below in Table 1.

Table 1. Outline of work functions extracted from the CWA with high and low level work tasks [6]

Work Functions	Work Tasks	High Level Work Tasks	Low Level Work Tasks
Overall tactical picture interpretation	Work Task 1	Integrate information related to tactical picture	
	Work Task 2		Acquire information related to the tactical picture
	Work Task 3		Update understanding of the tactical picture as required
	Work Task 4		Understand relevance of tactical picture to safety covertness and mission
Contact Management	Work Task 1		Classify Contact
	Work Task 2		Establish best Target Motion Analysis (TMA) solution (range, course, speed)
	Work Task 3		Monitor bearing rates of contacts
	Work Task 4	Track Contact of Interest (COI)	

We found that low-level work tasks, in comparison to high-level work tasks, were most suitable for constructing questions with measurable outcomes. While CWA results were useful for determining what tasks need to be completed to achieve a particular work function, they provided no indication as to when these tasks need to be performed. For example - is a low-level target motion analysis (TMA) task best made at the beginning, middle, or end of the scenario? ; is it best made before or after a particular event occurs? For this reason, we had to review the availability of task related information at each point in the scenario to ensure that the questions were being posed at an optimal time in the scenario. Posing questions too early or too late would not provide an adequate representation of the system's ability to support the work task in question.

To answer these types of questions, we had an SME review the scenario and construct subtasks that could be asked as lead-up questions to the low-level work tasks. This allowed us to gauge how early in the scenario users began gathering information and what information sources on the display were most helpful in gathering this information. We also had the SME evaluate which areas of the display were most likely to support this question so that we could compare the user's extraction of data to that of the SMEs. An example of this process is provided in Table 2 below.

Table 2. Link between low-level work tasks, scenes (time), subtasks and display area

CWA Work Functions	Low Level Work Tasks	Scene	Subtask related usability questions	Relevant IID Areas
Contact Management	Classify Contact	1	Not enough detail for scenario specific questions. Is there anything currently happening that would affect your mission goals?	All
		2	Not enough details for scenario specific questions. Utilizing the information presented on the IID, please give us your interpretation of the current tactical situation, noting any changes for the previous scene. Is this information consistent with your expectations? Could you have anticipated these changes?	Tactical Picture
		3	Utilizing the information presented on the IID please give us your interpretation of the current tactical situation, noting any changes from the previous scene. Is this information consistent with your expectations? Could you have anticipated these changes? New contacts entered the scenario and these questions are based on a correct recognition of these contacts. If the user did not recognize the new contacts then point them out. Follow-up Questions What information did you use to determine there are new contacts? If you wanted to find out if the contacts have characteristics of your Contact of Interest (COI) what display information would you use? What are the propulsion characteristics of your COI? What are the weapons characteristics of your COI?	Tactical Picture Range Information Information Area

4 Mental Models

4.1 Mental Models and the IID

While CWA derives various useful information requirements there is still a need to evaluate how the information requirements should be integrated and how they support user decision making. A “mental model” is essentially what drives users to perform in certain ways and to make certain decisions, and is representative of their expectations. A display that supports the user’s mental model reduces uncertainty and aids the decision making processes of the user [14]. We found it useful to assess the user’s mental model in an attempt to understand how users use the display and what information in the display best supports their decision making.

As an example, we assessed the user’s mental model of motion. From the literature we know that mental representations of motion differ from static information [15]. As such, we believed there to be value in testing both static and dynamic forms of the gauges, especially since some of the dynamic gauges are slow moving and therefore have more static characteristics than dynamic ones. To start we isolated the static gauges from the whole display (Figure 3) and we asked users questions such as- "how do you expect this gauge to change as fuel level decreases?" "which way do you expect the dashed line to move over time?" "what do you think will happen when the line reaches the darker colour?" [16]. These questions forced the users to verbalize their mental model so we could compare their mental model to the actual movement of the gauge.

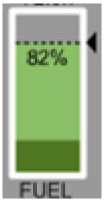

Fig. 3. Fuel gauge. Dashed line represents current fuel with 82% available

Once we obtained an understanding of the user's static mental model we used dynamic sliders and dropdown menus to simulate movement and changes to the system as they would happen in the dynamic display. The sliders and drop-down menus, presented underneath of the gauges in Figure 4, were adjusted to manipulate the lines depicted in the graphic. This allowed us to follow-up on the user's expectations and to clarify confusing elements. We found this strategy particularly helpful for graphics that were less obvious. Speed (Figure 4 right graphic), was a particularly difficult graphic for the users to delineate because it had three separate colour coded lines. The meaning of these lines in the static condition was not apparent to the users. Motion was introduced by changing the numbers in the dropdown menus and by manipulating the slider. In doing this it became clear which line represented current speed (black dashed) and which one represented planned speed (blue line). We believe that testing elements in their static and dynamic form adds value to the assessment by tapping into both the static and dynamic mental models of the user.

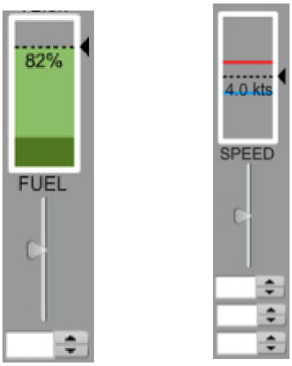

Fig. 4. Fuel and speed gauges with dropdown menus and sliders used to induce motion

As a second example the IID had a new design for ownship spatial orientation based upon common aerospace displays (Figure 5). While CWA provided information requirements, there was still a need to determine how, and what, information should have been integrated. We found it useful to assess mental models to determine the "how" for the newly integrated graphics. Properly integrated information should be less effortful to evaluate than the components that make up the integrated concept. Again, we assessed the user's mental model to determine how separate concepts may be integrated and what the spatial representation of those elements should be. In doing this we found that our user's mental model of "ownship" was from a "side-view" which made the new graphics with "look through" perspectives difficult for them to understand. The OOWs view of their submarine had an impact on how they integrated information related to the spatial orientation of ownship. As a result of this we redesigned the graphic in Figure 5 for a more simplistic integration from a "side-view" perspective. As such, we suggest that mental models that include an understanding of spatial representations be used to assess integrated concepts.

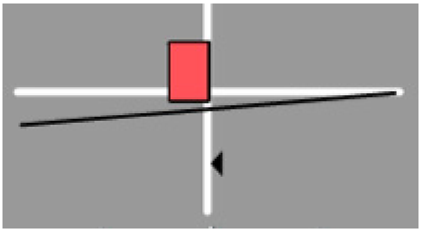

Fig. 5. Ownship attitude indicator with depth, trim, roll, pitch, rudder angle information

5 Attention

5.1 Attentional Factors in the IID

One of the main difficulties with CWA for use in display design is that it assigns equal priority to all information requirements as a way to support all possible decision making tasks. This makes it difficult to determine how information should be oriented, sized, arranged, and integrated in the display. In order to support the user and aid the design, it is imperative that priorities be assigned and accurately reflected in the display layout. While understanding the expectations of the user is important, there is also a need to understand the cognitive and attentional resources required to process the relevant information. Part of our assessment evaluated user behaviours and outcomes with respect to perceptual and cognitive biases. In doing this, we are better able to predict the potential shortfalls of the display in high-stress and high-workload conditions.

As an example, it is known that display logistics (i.e., arrangement, size, and proximity of information) directs attention and display viewing patterns [17]. Display sampling refers to the sequence of gaze patterns, which is driven by attention, to areas on the display [14, 17]. Ideally more important or vital areas of the display will be sampled more frequently than areas of low importance. With respect to the IID we found that some of the areas did not hold enough real-estate to reflect the importance of the information. Of course, this is likely a reflection of the fact that priorities were not assigned to the information extracted from the CWA. For example, we found that SMEs used Area 4 more than Area 5 (see Figure 1), yet the two areas were of equal size. Had priorities been assigned to the information areas we would have known that the tactical picture was of high importance and required more screen real-estate.

While we noted that the users gathered a lot of their information from the tactical picture we also have concerns that making the tactical picture too large would promote attentional tunneling [17]. By making one area of the display larger we run the risk of directing the user's attention to this area at the expense of other vital information in the display. To combat this bias we suggest that scanning techniques, such as scanning the display in a particular pattern, be presented to users to maximize the amount of information they retrieve from the display. This would also help combat "event rate" [17] (p.73) biases which direct attention to quickly changing areas at the expense of more static areas in the display [17, 18]. Ideally, the recommended sampling technique would mimic the OOWs current data extraction mental model to allow the user to spend more time and resources evaluating the acquired information and making decisions.

6 Conclusion

We believe that usability professionals can minimize the ambiguity of system level testing by utilizing CWA derived work tasks. In the current assessment, we found that both the work functions and high-level work tasks were too general to adequately formulate questions. We did find that the CWA derived low-level work tasks were useful in guiding the development of scenario based usability questions. However, it was necessary to break the low-level work tasks into subtasks. The subtasks allowed us to formulate time relevant specific questions with measurable outcomes. We also recommend that priorities be assigned to the CWA derived information requirements to aid the design of the system-level display. Furthermore, we found the consideration of mental models and attentional biases in our assessment valuable in identifying ways to improve design and decrease attentional load.

References

1. Zelik, D.J., Patterson, E.S., Woods, D.D.: Measuring attributes of rigor in information analysis. In: Patterson, E., Miller, J.E. (eds.) Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams, pp. 65–83. Ashgate, Aldershot (2010)
2. Chalmers, B.A.: Developing an Information Integration Display for Submarine Command and Control. Presented at Undersea Human Systems Integration Symposium 2011. American Society of Naval Engineers (2011)
3. Xiao, T., Sanderson, P.M.: Developing and evaluating the organizational constraints analysis (OCA) approach to analysing work coordination via resource allocation case studies. Presented at the Human Factors and Ergonomics Society 56th Annual Meeting, pp. 373-377 (2012)

4. Vicente, K.J.: Cognitive Work Analysis: Toward Safe, Productive, and Health Computer-Based Work (eds). Lawrence Erlbaum Associates Inc., Mahwah (1999)
5. Read, G., Salmon, P., Lenne, M.: From Work Analysis to Work Design: A review of Cognitive Work Analysis Design Applications. Presented at the Human Factors and Ergonomics Society 56th Annual Meeting, pp. 368–372 (2012)
6. Rehak, L., Karthaus, C., Lee, B., Matthews, M., Taylor, T.: Human Systems Integration (HSI) Personal Communication
7. Roth, E.M., Eggleston, R.G.: Forging new evaluation paradigms: Beyond statistical generalization. In: Patterson, E.S., Miller, J. (eds.) Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams, pp. 203–219. Ashgate Publishing (2010)
8. O'Hara, J.: A quasi-experimental model of complex human machine system validation. International Journal of Cognition, Technology, and Work 1, 37–46 (1999)
9. Redish, G.: Expanding Usability Testing to Evaluate Complex Systems. Journal of Usability Studies 2, 102–111 (2007)
10. Pfautz, J., Roth, E.: Using cognitive engineering for system design and evaluation: A visualization aid for stability and support operations. International Journal of Industrial Ergonomics 5, 389–407 (2006)
11. Patterson, E.S., Roth, E.M., Woods, D.D.: Facets of complexity in situated work. In: Patterson, E.S., Miller, J. (eds.) Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams, pp. 1–10. Ashgate Publishing (2010)
12. Rousseau, R., Tremblay, S., Lafond, D., Vachon, F., Breton, R.: Assessing temporal support for dynamic decision making in C2. Presented at the 51st Annual Meeting of the Human Factors and Ergonomics Society Baltimore, MD, pp. 1259–1262 (2007)
13. Bryant, D.J., Webb, R.D.G., McCann, C.: Synthesizing two approaches to decision making in command and control. Canadian Military Journal 4, 29–34 (2003)
14. Bellenkes, A.H., Wickens, C.D., Kramer, A.F.: Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. Aviation, Space, and Environmental Medicine 68, 569–579 (1997)
15. Freyd, J.J.: Dynamic Mental Representations. Psychological Review 94, 427–438 (1987)
16. Lamoureux, T., Pasma, D., Kersten, C.: CAE PS Personal Communication
17. Wickens, C., Hollands, J.: Engineering Psychology and Human Performance. Prentice Hall, Upper Saddle River (2000)
18. Donk, M.: Human Monitoring behavior in a multiple-instrument setting: Independent sampling, sequential sampling or arrangement-dependent sampling. Acta Psychologica 86, 31–55 (1994)