ENHANCING ROBOFLAG USERS' SITUATIONAL AWARENESS

Sangeeta Shankar, Yi Jin, Julie A. Adams, and Bobby Bodenheimer
Department of Electrical Engineering and Computer Science
Vanderbilt University, Nashville, TN 37235
[sangeeta.shankar, yi.jin, julie.a.adams, bobby.bodenheimer]@vanderbilt.edu

The RoboFlag system was designed as a testbed to study distributed control of multiple vehicle teams with humans in the loop. This work analyzed the RoboFlag version 2.0 interface to identify existing issues with the users' Situational Awareness (SA). The existing interface for RoboFlag was modified to create two new interfaces. The first interface focused on improved usability, while the second focused on improved Situation Awareness. A user evaluation was conducted to determine if the new interfaces improved the users' SA over the original interface. Twenty-four participants completed the evaluation. This paper reports the design of the task environment, the evaluation method, and the statistical analysis. The results indicate that both new interfaces provide improved SA over the RoboFlag version 2.0 interface.

INTRODUCTION

RoboFlag is a robotic team-based game similar to "Capture the Flag" (RoboFlag, 2003). The teams rely on a human operator to govern their play. The operators specify plays that the robots then autonomously fulfill. The system is designed to allow researchers to explore basic and complex issues in several areas, including cooperative control, path planning, team strategies, team dynamics, operator interfaces, and cognitive engineering. Critical to a team's success is the operator's performance. Our hypothesis in this work is that the operator's performance should be improved with enhanced situational awareness (SA) and lowered cognitive workload. This paper provides the results pertaining to perceived situational awareness while using the RoboFlag 2.0 interface and two new interface designs. The new interfaces were designed to improve usability, decrease cognitive load, and increase SA. This work employs Endsley's 3-level SA model (Endsley 1995).

Two previous studies have been conducted using the RoboFlag 2.0 system. Veverka and Campbell (Veverka 2003) conducted a study to assess users' workload and information load. Parasuraman, Gastler, and Miller (Parasuraman 2003) conducted a study to evaluate the interface based upon the enemy engagement style and environmental uncertainty. This study gathered perceived workload and SA data but found no significant results related to overall SA across conditions.

The paper provides an overview of the RoboFlag environment, the original interface, and the redesigned interfaces. The Methodology section provides details of our evaluations, and the Results section provides the evaluation results. We provide a discussion of our results and finally a conclusion.

SYSTEM DESIGN

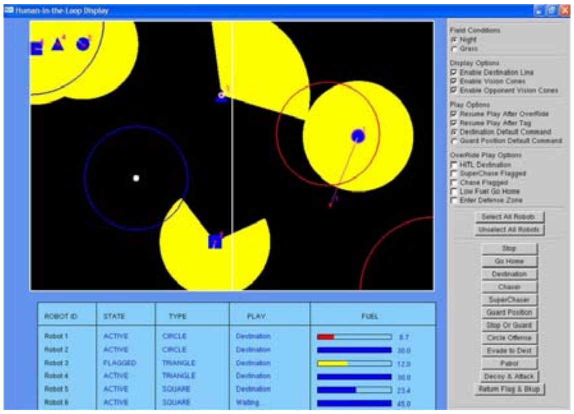

Figure 1 RoboFlag 2.0 interface.

The RoboFlag 2.0 interface (Figure 1) provides a **Home Zone** or safe haven for a team's robots. The robots are placed in the home zone at the beginning of the game. The **Defense Zone** is the area where the team's flag is placed and represents the area that the team is to defend. Robots can travel 30m before refueling, which occurs when the robot enters the Home Zone. Each robot's **Vision Cone** represents that particular robot's field of view. The game settings can be modified to particular preferences via the configuration buttons along the upper right hand side of the interface. Commands for controlling the robots are also displayed along the bottom right side of the interface. These buttons, called "Play Buttons," permit the selection of the play a robot is to execute. Plays include a range of offensive capabilities and defensive capabilities. Properties of each robot are displayed directly below the playing area.

The RoboFlag environment provides a number of autonomous plays that are intended to assist the operator with game play. The Circle Offense and the Decoy and Attack plays allow the robots to try and capture the opponent team's flag. The Go command can be used to send any selected robot back to its home zone. Examples of defensive plays include Guard position, Patrol and Chaser. The Guard Position play commands the robot to guard a specified area on the field. The Patrol play instructs the robot to guard the centerline boundary along the team's end of the field. The Chaser play instructs a robot to sense (look for) the nearest opponent robots, and chase them. Each of these plays enables the team to protect their own flag. The Stop play stops the selected robot, while the Destination play allows the operator to select a destination point that the robot then attempts to get too.

The described work sought to improve both the usability and SA of the human operator while playing the game. Therefore, an interface was designed with the intention of improving usability and decreasing cognitive load (Shankar et al. 2004), Interface One (Figure 2). A second interface, Interface Two (Figure 3), extended Interface One with the intention of improving the operator's perceived SA.

Figure 2. Improved Interface One

A number of usability and SA issues were identified with the original interface. First, there was no ordering of the Play Buttons. This lack of ordering may affect the operators' performance, as time is required to browse through the Play Buttons and select the desired play. Since the game is played under time constraints and is fast-paced, the lack of organization could be an important factor affecting the operator's performance. Interface one (Figure 2) organized the plays according to the strategy type, e.g., offense, defense, and general plays. Examples of the plays classified as offense include: Circle Offense, and Decoy and Attack. The classification of defensive plays includes commands such as: Guard position, Patrol, and Chaser. The plays classified as general include the Stop and Destination commands which can be used at any point in the game.

The second identified issue with the 2.0 interface is related to the display of configuration options that modify the game settings. These options are displayed along the upper right side of the interface. It is unlikely that the operator will modify these settings during play; therefore continuous access to these options is not required. The buttons were hidden and made available via a rollout button located at the top right of the command area in Interface One. This option is accessible at all times and reduces screen clutter. Hiding the buttons was intended to decrease users' cognitive load, thereby increasing interface usability.

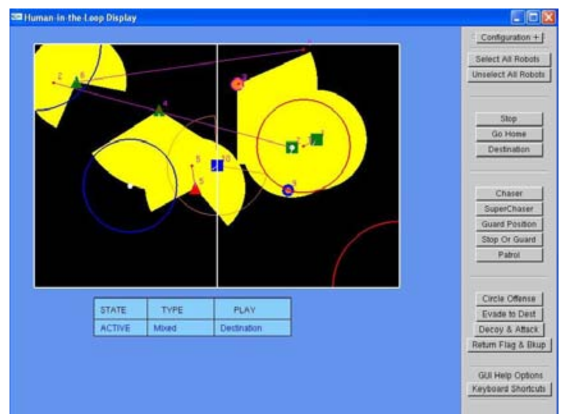

Figure 3. Improved Interface Two.

Interface Two included further modifications intended to improve SA, as shown in Figure 3. These modifications focused on the robot status information displayed at the bottom of the original interface. This information is not in the operator's direct field of view. Information not in the operator's direct view may not be used during game play and may cause adverse affects on performance. Therefore, Interface Two embedded as much of the robot status information into the representation of the robots as possible. The robots' color was changed to reflect their current fuel level. If the robot has sufficient fuel, it is colored green, when the fuel level drops below 15 the robot turns orange. Finally, when the fuel level drops below 7.5 the robot turns red. Also, the robots' shape was modified to represent the type of play being executed: defense, offense, or a general play. The status panel itself was modified to display the type and play name of selected robots rather than every robot on the team. This modification should make it easier to understand the information by reducing the amount of information displayed.

METHOD

A user evaluation involving 24 volunteer participants (15 male and 9 female) assessed the usability and SA of the modified interfaces compared to the original. The participants were members of the Vanderbilt community, with ages ranging from 18 to 40. Most participants had no prior experience using robots or robotic games, although most play computer games.

Each participant completed a pre-experimental questionnaire and then read a brief system overview. The experimenter explained the game while playing for 2-3 minutes. The participants then trained by playing a game with the original (2.0) interface for a duration of ten minutes. The training was followed by the completion of a usability questionnaire and a 3-D Situation Awareness Rating Technique (SART) questionnaire. The participants then completed two games lasting 7.5 minutes with each of the modified interfaces. Each game was followed by the 3-D SART and usability questionnaires. Order effects were controlled by first presenting half of the participants with Interface One followed by Interface Two while the remaining participants used Interface Two first. This presentation ordering was used to ensure that the use of the new interfaces did not affect their ratings. The participants' game play was timed; therefore task completion was not recorded. Since we evaluated the interfaces based on the users' perception of their own Situational Awareness, the game scores were not recorded.

The 3-D SART questionnaire was based upon that used by Parasuraman et al. (Parasuraman 2003). This questionnaire included ratings for attentional demand, attentional supply, understanding, and an overall SA assessment. The demand, supply, and understanding ratings were combined to create a single SA value using the algorithm provided by Jones (Jones 2000). A 7-point Likert scale (1 - low value; 7 - high value) was employed. Higher SA is indicated by a lower value for demand, and higher values for supply, understanding and the overall rating. The null hypothesis was that the modifications in Interface One and Interface Two do not result in a significant change in the operator's perceived SA.

An analysis of the usability results has been conducted (Shankar et al. 2004). These results indicated improved usability with both of the modified interfaces.

RESULTS

The participants provided an overall perceived SA value in addition to the three constructs of attentional demand, attentional supply, and understanding. These constructs were also employed to calculate perceived SA. The mean and standard deviations of these values for each of the three interfaces are provided in Table 1. The average attentional demand was highest for the original interface, followed by Interface One. Interface Two received the lowest demand rating. Attentional supply was rated as the lowest for the 2.0 interface, while the average rating for Interface Two was slightly higher than Interface One. The understanding component also improved for the modified interfaces over the 2.0 interface with Interface Two receiving the highest rating. In general, these results show lower demand with higher supply and understanding. This implies improved SA for Interface Two over the other interfaces.

Table 1. Demand, Supply and Understanding for the three interfaces.

	Demand		Supply		Understanding	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
2.0	5.21	1.32	3.25	1.11	2.67	1.05
One	3.75	1.30	5.13	0.85	5.46	1.14
Two	3.50	1.59	5.46	0.98	5.58	1.02

The values for these constructs were compared using one-way and two-way ANOVAs. The one-way ANOVA compared the ratings across all three interfaces, while the two-way ANOVA compared individual interface pairs to determine which was rated the best. The one-way analysis was performed for each of the three constructs across all screens. The one-way ANOVA for attentional demand found a significant result ($p < 0.01$, MS = 20.43, $F(2,69) = 10.33$). This result implies that the participants found the two modified interfaces less demanding in terms of SA. The results for attentional supply were also significant ($p < 0.01$, MS = 34.01, $F(2,69) = 34.99$). This result shows that the participants found improvement in the new interfaces for this construct. The rating for understanding increased for Interface One and Interface Two over the 2.0 interface as implied by the significant results found via the one-way ANOVA ($p < 0.01$, MS = 65.27, $F(2,69) = 56.91$).

The two-way ANOVA results for attentional demand were significant when comparing the RoboFlag 2.0 interface to Interface One ($p < 0.01$, MS = 25.52, $F(1,44) = 14.41$) and when comparing the 2.0 interface to Interface Two ($p < 0.01$, MS = 35.02, $F(1,44) = 16.83$). However, the comparison between Interface

One and Interface Two was not significant. This verifies that both new interfaces improved SA over the 2.0 interface, but does not show that Interface Two significantly improved SA over Interface One.

The two-way analysis for the attentional supply component found significant relationships between the original interface and Interface One ($p < 0.01$, MS = 42.19, $F(1,44) = 43.09$) and between the original interface and Interface Two ($p < 0.01$, MS = 58.52, $F(1,44) = 54.69$). The analysis comparing the two modified interfaces was not significant. This result indicates that the two new interfaces each improve the attentional supply but that Interface Two does not significantly improve this element over Interface One.

Similar analysis for the understanding component was significant when comparing the RoboFlag 2.0 Interface with Interface One ($p < 0.01$, MS = 93.52, $F(1,44) = 78.26$) and the 2.0 Interface with Interface Two ($p < 0.01$, MS = 102.08, $F(1,44) = 105.69$). The two-way comparison for understanding between the two modified interfaces was not significant.

Table 2 provides a comparison of the mean and standard deviation values for the overall SA values provided by the participants on the SART questionnaire and the calculated SA values. Figure 2 provides a visual comparison across the three interfaces. As the table and figure indicate, the RoboFlag v. 2.0 interface received the lowest SA ratings for both the perceived rating and the calculated values. A large increase in the perceived rating and the calculated values was found for both Interfaces One and Two. Interface Two received the highest ratings for both the perceived overall rating and the calculated value.

Table 2. Overall SA rating versus the 3-D SART calculated value.

	Overall Rating		Calculated Rating	
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**
2.0	3.21	0.88	0.71	2.05
One	5.13	0.95	6.83	2.44
Two	5.29	0.91	7.54	3.08

An interesting result, clearly shown in Figure 3, is that the participants' average overall rating was much higher than the calculated SA values for the v. 2.0 interface. The results for the remaining interfaces indicated that the overall perceived awareness was lower than the calculated SA.

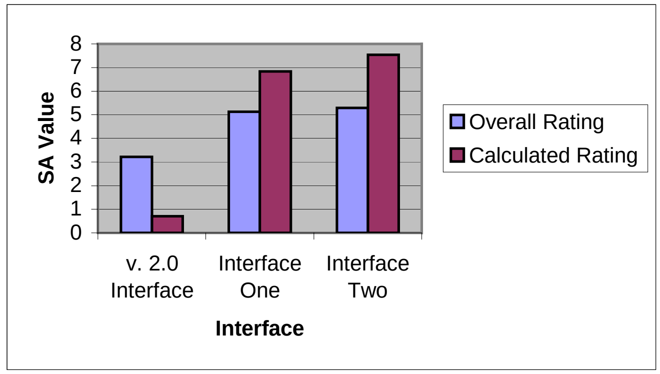

Figure 2. Comparison of the perceived overall rating versus the calculated 3-D SART values.

A one-way ANOVA was calculated to compare the results across the three interfaces for both the overall perceived ratings and the calculated SA. The analysis of the overall rating found a significant relationship, ($p < 0.01$, MS = 32.17, $F(2, 69) = 38.57$). The analysis for the calculated SA was also significant, ($p < 0.01$, MS = 338.85, $F(2, 69) = 51.70$). This result implies that the participants felt that the two new interfaces provided better SA than the RoboFlag 2.0 interface. To better understand these results, a two-way ANOVA was calculated comparing results between interfaces for both the overall perceived ratings and the calculated SA.

The two-way ANOVA comparing the overall perceived SA ratings found significant relationships for the RoboFlag 2.0 interface compared to Interface One ($p < 0.01$, MS = 44.08, $F(1,44) = 55.68$) and as compared to Interface Two ($p < 0.01$, MS = 52.08, $F(1,44) = 64.55$). The comparison between Interface One and Two was not significant. This result shows that the participants rated their overall perceived SA significantly higher for both new interfaces compared to the RoboFlag 2.0 interface. There did not appear to be a significant improvement in the overall rating between the two new interfaces.

A two-way ANOVA analysis was also conducted for the calculated SA value. The analysis found a significant relationship between the RoboFlag 2.0 interface and Interface One ($p < 0.01$, MS = 450.18, $F(1,44) = 88.59$). Similarly, the relationship between the 2.0 interface and Interface Two was significant ($p < 0.01$, MS = 560.33, $F(1,44) = 91.48$). The relationship between Interface One and Two was not significant. These results support the participants overall ratings and provide evidence that the two new interfaces greatly improved the operator's SA over the RoboFlag 2.0 interface.

DISCUSSION

The goal of this work was to provide an improved interface for the RoboFlag environment. As a result, two new interfaces were designed based upon the RoboFlag 2.0 interface. Interface One was specifically designed to improve the system usability and reduce cognitive load. Interface Two was designed based upon Interface One with the intention of improving the operators' overall SA. The use of the original interface to train the participants may have skewed the results, but due to the lack of a previous evaluations involving RoboFlag focusing solely on users' SA, as well as time and resource constraints, the choice was made to use the original interface for training.

The results obtained from this evaluation are positive and supported our hypothesis that the new interfaces would improve SA. The evaluation establishes that the two modified interfaces increase the operators' SA over the RoboFlag 2.0 interface. Both modified interfaces provided significantly improved perceived and calculated SA over the RoboFlag 2.0 interface. The modified interfaces created a user interaction that led to enhanced awareness of the situation. The modified interfaces' categorization of the plays simplified the participants' play selection process and caused the demands on the users' attentional resources to decrease. This categorization resulted in an increase in the users' overall SA.

While both modified interfaces had significantly improved SA over the RoboFlag 2.0 interface, Interface Two did not significantly improve SA over Interface One. A reason for this lack of improvement may be that usability improvements made in Interface One decreased cognitive load thereby increasing SA. Since the Interface Two design was based upon Interface One, it appears that the additional SA changes did provide a slight but not significant improvement to SA. These results also prove that in our experiment, SA correlated with usability and cognitive load. The statistical power of our experiment was insufficient to distinguish between SA improvements resulting from improved usability and further modifications.

CONCLUSIONS

Consideration of users' needs and human factors is vital in the development of better designed, easier-to-use interfaces that enhance users' SA. For our study, we assessed the original system interface to identify gaps that potentially caused reduced usability and SA or appeared to negatively affect users' game play performance. Design solutions were identified and implemented in two new interfaces. The affect on SA was evaluated by conducting a user evaluation that included 24 volunteer participants. The participants evaluated perceived SA using a 3-D SART questionnaire. The results from the user evaluation clearly demonstrated that both of the modified interfaces succeeded in increasing the users' Situational Awareness over the original RoboFlag 2.0 interface. The analysis also found that while Interface Two received better SA ratings than Interface One, it did not provide significantly improved SA. The results also established that the increase in usability and decrease in cognitive load helped improve SA

ACKNOWLEDGEMENTS

We thank Dave Schneider and Raja Parasuraman for their helpful comments and assistance. We also thank Li Su for his contributions to the project.

REFERENCES

D'Andrea, R. & Babish, M. (2003). The RoboFlag Testbed. *Proceedings of the American Controls Conference*, USA, 656-660.

Endsley, M.R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors*, 37 (1), 32-64.

Jones, D.G. (2000). Subjective Measures of Situation Awareness. In M.R. Endsley & D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement* (pp. 113-126). London: Lawrence Erlbaum Associates.

Parasuraman R., Galster S., & Miller, C. (2003). Human Control of Multiple Robots in the RoboFlag Simulation Environment. *Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics*, USA, 3232-3237.

RoboFlag (2003). RoboFlag website retrieved December 15th, 2003, from Cornell University: http://roboflag.mae.cornell.edu/.

Veverka, J. & Campbell, M. (2003). Experimental Study of Information Load on Operators in Semi-Autonomous Systems. *Proceedings of the 2003 AIAA Guidance, Navigation and Control Conference*, USA, AIAA No. 2003-5661.

Shankar, S., Su, L., Jin, Y. Adams, J. A. & Bodenheimer, R. (2004). Comparing the Usability of RoboFlag Interface Alternatives. *2004 IEEE International Conference on Systems, Man, and Cybernetics*. 2004. The Hague, Netherlands.