Comparing the Usability of RoboFlag Interface Alternatives

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

Abstract - *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 usability issues. The existing interface 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 both new interfaces improved the system usability over the original interface. Twenty-four participants completed the evaluation. This paper reports the design considerations, the experimental apparatus, and the usability based statistical analysis. The results indicate that both new interfaces provide improved usability over the RoboFlag version 2.0 interface.*

Keywords: Usability, User Interface Design, RoboFlag

1 Introduction

RoboFlag is a robotic team game based on "Capture the Flag" [1]. Two teams rely on a human operator to govern their movements. 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. The system includes uncertainties such as incomplete information, latency, intelligent adversaries, and neutral entities that require full, realistic solutions. Critical to a successful system is a user interface design that provides high usability and low cognitive load. Our study evaluated the RoboFlag 2.0 interface usability compared with two new interface designs. The new interface designs were developed to improve usability, decrease cognitive load, and increase situational awareness. The focus of this paper is the results pertaining strictly to usability; further information pertaining to situational awareness is available in [2].

Two previous studies have been conducted using the RoboFlag 2.0 system. Veverka and Campbell [3] conducted a study to assess users' workload and information load. The evaluation varied users' degree of control over the robotic team during a series of games. Parasuraman et al. [4] also conducted a study using RoboFlag 2.0. This study focused upon enemy engagement styles and environmental uncertainty.

This paper is organized as follows. The next section provides an overview of the design and discusses the design of the RoboFlag task environment. This is followed by an outline of the experimental evaluation. The results of the evaluation are then presented, followed by a discussion of the results and the conclusions.

2 System Description

The RoboFlag task environment consists of two teams of six robots each. The operators control the robots in each team. Their objective is to enter the opponent's territory, capture their flag, and return to the "Home Zone" as many times as possible. The game is a mix of offense and defense: secure the opponent's flag, while at the same time prevent the opponent from securing your flag [5]. Each team's playing field includes a Home Zone, a Defense Zone and an Attack Zone. The team attempts to defend their defense zone and the attack zone is where they stop the opponent from entering the defense zone.

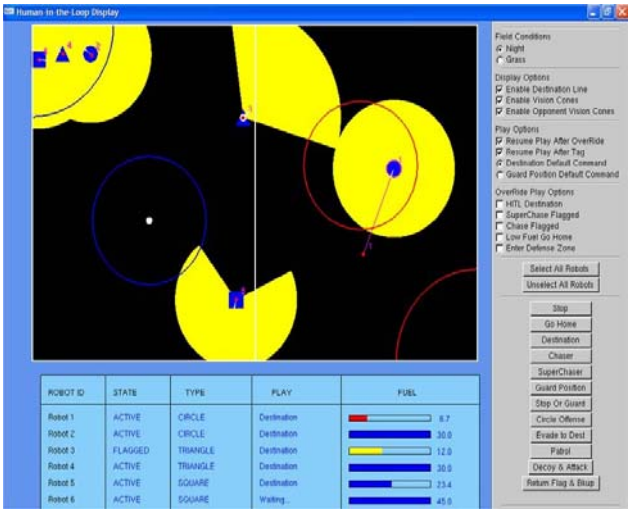

Figure 1. RoboFlag 2.0 interface.

The RoboFlag 2.0 interface (Figure 1) provides a **Home Zone** or safe haven for the robots, where the robots are placed at the beginning of the game. The Home Zone for the blue and red teams are located on the top left corner, and bottom right corner of the playing area respectively in Figure 1. The **Defense Zone** is the area where the team's flag is placed and represents the area that the team's robots are attempting to defend. The Defense zones for the two teams in Figure 1 are located in the circular regions in the center of each team's half of the playing area. The robots can travel 30m without refueling. Each robot's **Vision Cone,** represented as the yellow colored region around the robot in Figure 1, represents that particular robot's field of view. The configuration buttons

along the top half of the right hand side in Figure 1 allow the operator to set various game settings. The commands for controlling the robots are displayed in the bottom portion of the right hand side of Figure 1 as "Play Buttons" [1]. These buttons represent the types of play a robot may execute. Plays include a range of offensive capabilities and defensive capabilities. The various properties of each team robot are displayed below the playing area in Figure 1. This information includes the robots type, fuel status, status etc.

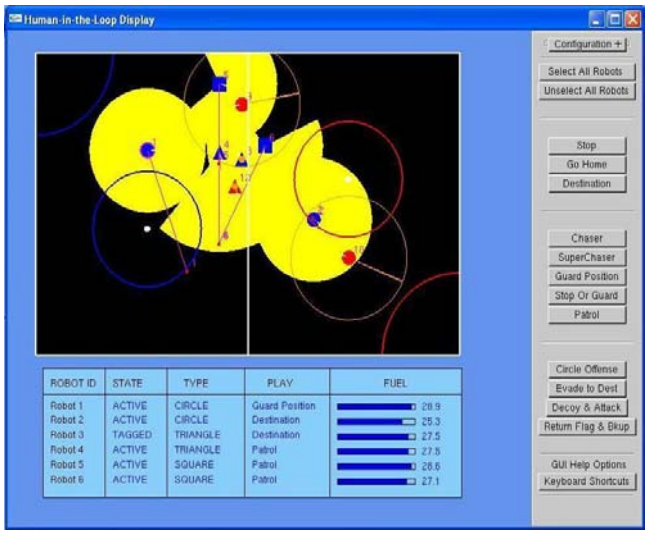

Figure 2. Improved Interface One.

A number of usability issues were identified with the RoboFlag 2.0 interface. The design team playing multiple games using the original interface identified usability issues. Therefore, Interface One was designed with the intention of improving general usability. The first issue was that all Plays (Play Buttons) were presented with no apparent order or groupings. This lack of organization might affect the user's 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, it may be an important factor affecting the user's performance. The modification for Interface One was to organize the Plays according to the strategy type they follow (shown in Figure 2 along the right hand side), that is offense or defense. There are also a few general play buttons. The 2.0 interface includes commands for modifying the game settings that are continuously displayed in the upper half of the screen as shown in Figure 1. The operator does not need continuous access to these settings; therefore, they have been reorganized into selections made available by a rollout button located at the top right of Interface One in Figure 2. This option is accessible at all times. These modifications greatly improved the screen appearance by reducing the clutter.

Situation awareness is “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” [6]. While playing the game using the 2.0 Interface, it was felt that the user’s situational awareness could be improved. Interface Two represents the redesigned Interface One that focuses on improving situational awareness (See Figure 3).

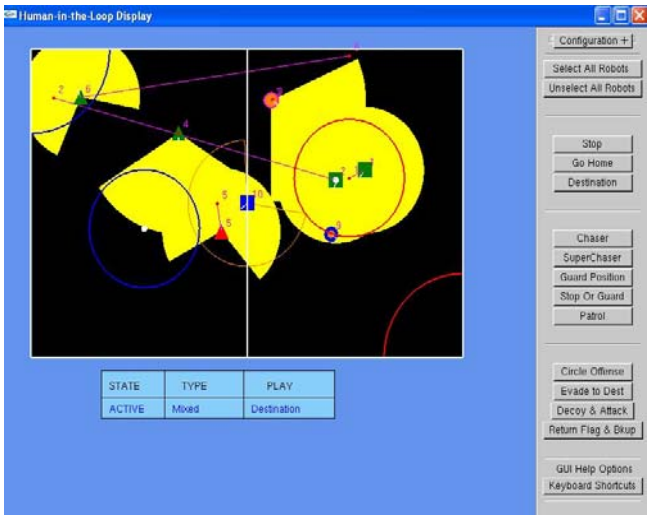

Figure 3. Improved Interface Two.

The identified factors that may affect the user’s situational awareness focused on the robot status information displayed at the bottom of the screen in the 2.0 Interface. This information is not in the operator’s focus of attention. Therefore, it is possible that the operator may not use this information during game play, which may negatively affect performance.

The redesign involved embedding as much of the robot status information into the robots as possible. The robot status information includes the robot’s state, type, fuel level, and play being executed. First, the robots’ colors were modified to represent their current fuel level. Another very important piece of information is the type of play a robot is currently executing. Therefore, the robots’ shapes were modified to represent the play type being executed. The shape of the robot changes from a circle (original shape for all robots) to a square when playing “offense” or to a triangle when playing “defense”. A circle represents a mixed strategy. Finally, when robots are selected, the status panel in the bottom of Interface Two displays the selected robots’ type and the specific play name.

3 Experimental Evaluation

A user evaluation involving 24 participants was conducted to assess the usability of the modified interfaces compared to the RoboFlag 2.0 Interface. The participants used all three interfaces. Participants were members of the Vanderbilt University community and were 62% male and 38% female. The participants’ age range was between 18 to 40 years with over 60% falling in the 25 to 30 years age group. Most participants had no prior experience using either robots or robotic games, although most were familiar with computer games.

Each participant completed a pre-experimental questionnaire and read a brief system overview. The experimenter explained the game while playing for 2-3 minutes. The participants played the game using the 2.0 Interface during a ten minutes training session. The training was followed by the completion of a questionnaire that gathered usability and situational awareness information. The participants then played two 7.5-minute games with each of the modified interfaces. Each game was followed by the same questionnaire as was used after the training session.

All users trained on the RoboFlag 2.0 interface. 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 followed by Interface One. The participants' game play was timed; therefore task completion was not recorded. Since we evaluated the interfaces based on the users' perception of the usability and their own performance, the game scores were not recorded.

The questionnaire assessed important design features that were believed to affect usability. The questionnaire consisted of twelve questions, each capturing a different aspect of system usability on a Likert scale ranging from 1 (low value) to 7 (high value). The null hypothesis was that Interface One and Interface Two do not result in any significant change in the system usability. The study also involved collection of situational awareness data using a 3-D SART questionnaire. The situational awareness results are available in [2].

4 Results

An ANOVA was conducted for each usability question across the three interfaces. The results were considered significant if $p <= 0.05$. A one-way ANOVA compared the 2.0 Interface to Interfaces One and Two. Two-way ANOVAs compared paired comparisons of the three interfaces.

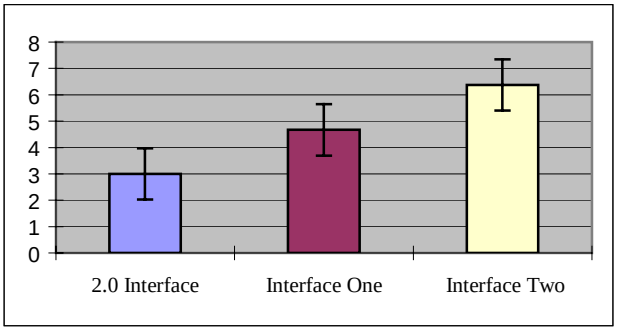

Figure 4. Mean satisfaction level by interface.

The participants rated how frustrating (1) or satisfying (7) they perceived each interface to be. The mean responses by interface are provided in Figure 4. As can be seen, Interface Two received the highest rating (mean = 6.375, std. dev. = 0.65), with the 2.0 interface (mean = 3.0, std. dev. = 0.98) receiving the lowest rating. A more satisfying interface was represented by a higher numeric value.

A one-way ANOVA found that the modified interfaces were significantly more satisfying than the 2.0 interface ($p < 0.01$, MS = 68.35, F (2,69) = 79.99). The two-way ANOVA found that Interface Two was significantly more satisfying than Interface One, ($p < 0.01$, MS = 35.02, F(1,44) = 44.99). The same comparison was significant between the 2.0 Interface and Interface One ($p < 0.01$, MS = 33.33, F(1,44) = 31.54) and between the 2.0 interface and Interface Two ($p < 0.01$, MS = 136.68, F(1,44) = 191.44). While these results indicate that both modified interfaces rated more satisfying than the original interface, Interface Two proved the most satisfying interface.

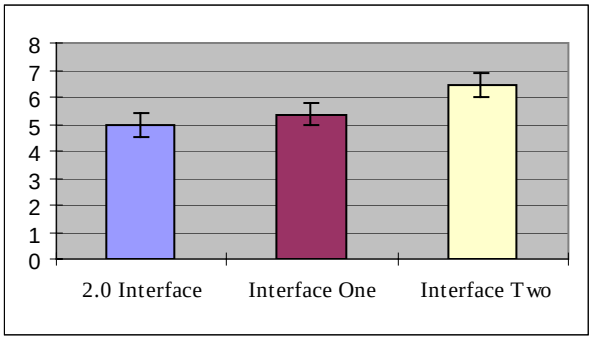

Figure 5. Mean stimulation results by interface.

The participants rated the system based on perceived dullness (1) and stimulation (7). Figure 5 shows the means responses by interface. Interface Two was rated as the most stimulating (mean = 6.42, std. dev. = 0.58). The least stimulating interface was the 2.0 interface (mean = 4.79, std. dev. = 1.14).

The one-way ANOVA found that both modified interfaces proved more stimulating ($p < 0.01$, MS = 15.88, F(2,69) = 22.02). The two-way analyses between the two modified interfaces, was significant ($p < 0.01$, MS = 9.19, F(1,44) = 20.64). The two-way ANOVA between the 2.0 Interface and Interface One provided a significant result, ($p < 0.01$, MS = 6.75, F(1,44) = 7.16). Similar analysis between the 2.0 interface and Interface Two was also significant, ($p < 0.01$, MS = 31.68, F(1,44) = 37.43). These results indicate that Interface Two was significantly more stimulating than the two other interfaces. As well, the results found that the modified interfaces significantly increased user stimulation over the original interface.

Figure 6 provides the mean results pertaining to the participants' results regarding ease of use. The participants felt that both new interfaces were easier to use (One: mean = 5.38, std. dev. 1.13; Two: mean = 5.67, std. dev. = 0.87) than the 2.0 interface (mean = 3.17, std. dev. = 1.0).

The one-way analysis was found to be significant ($p < 0.01$, MS = 44.84, F(2,69) = 44.02). The two-way analysis between the modified interfaces was insignificant, while the two-way ANOVAs between the original interface and the modified interfaces were significant. The two-way ANOVAs between the 2.0 Interface and Interface One found ($p < 0.01$, MS = 154.08, F(1,44) = 273.01) and between the 2.0 interface and Interface Two found ($p < 0.01$, MS = 31.68,

F(1,44) = 331.24). This result implies that the two modified interfaces were easier to use than the 2.0 interface.

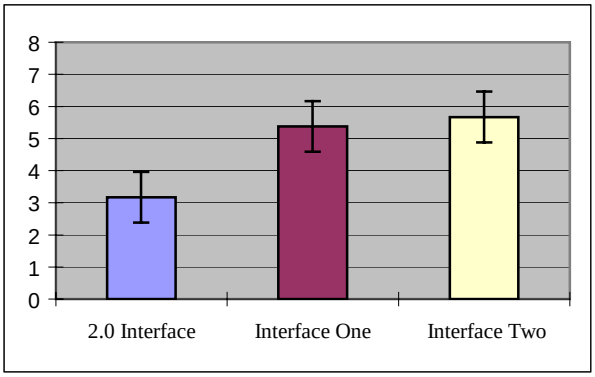

Figure 6. Mean perceived ease of use by interface.

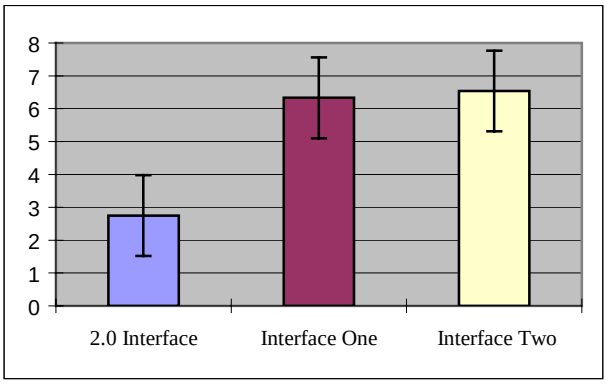

Figure 7. Mean value of determining play selection by interface.

The participants rated the difficulty of determining play selection, where easier decisions received higher ratings. As Figure 7 indicates, the participants rated play selection as easier with both modified interfaces (One: mean = 6.33, std. dev. = 0.7; Two: mean = 6.54, std. dev. = 0.66). The 2.0 interface was rated as somewhat difficult (mean = 2.75, std. dev. = 0.85).

The one-way ANOVA across all interfaces was found to be significant ($p < 0.01$, MS = 109.04, $F(2,69) = 199.09$). The two-way ANOVA comparing Interfaces One and Two was not significant. This particular result was not surprising since there was no change in the play presentation between the two interfaces. The two-way ANOVAs between the original and each of the modified interfaces resulted in significant differences. The 2.0 Interface compared to Interface One found ($p < 0.01$, MS = 154.08, $F(1,44) = 273.01$) while the 2.0 interface compared to Interface Two found ($p < 0.01$, MS = 31.68, $F(1,44) = 331.24$). Clearly, the reorganization of the buttons in the modified interfaces helped improve the participants' ability to determine play selection.

The number of plays was constant across the interfaces and mean responses as to whether the number of plays was adequate or inadequate are provided in Figure 8. The modified interfaces reorganized the buttons into meaningful groups. The modified interfaces received a higher rating of adequacy regarding the number of plays (One: mean = 6.33, std. dev. = 0.82; Two: mean = 6.63, std. dev. = 0.5) as compared to the 2.0 interface (mean = 4.0, std. dev. = 1.33).

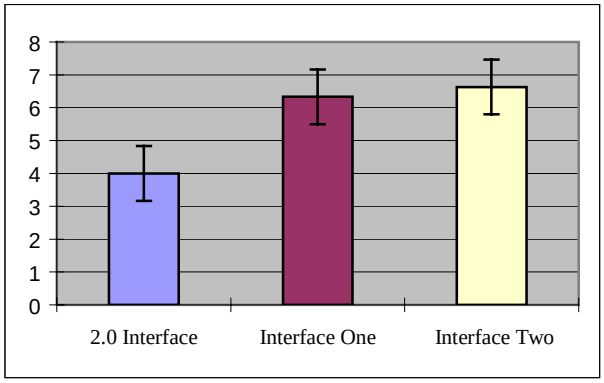

Figure 8. Mean response regarding number of plays by interface.

The adequacy rating of the number of plays for the modified interfaces was significantly different from the 2.0 Interface, as found by the one-way ANOVA ($p < 0.01$, MS = 51.35, $F(2,69) = 57.22$). Even though Interface Two received a slightly higher rating, the two-way ANOVA between the modified interfaces was not significant. The two-way ANOVA comparing the 2.0 Interface to Interface One was significant ($p < 0.01$, MS = 67.68, $F(1,44) = 63.48$). The same analysis comparing the 2.0 interface to Interface Two was also significant ($p < 0.01$, MS = 85.33, $F(1,44) = 86.98$). Therefore, the reorganization of the presented plays increased the perception that the number of plays was adequate with the modified interfaces.

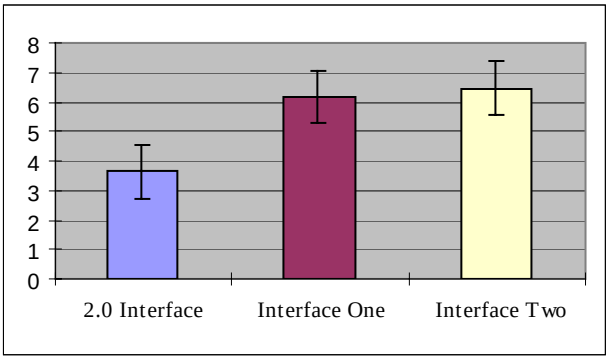

Figure 9. Mean play organization ratings by interface.

The participants rated the arrangement of the plays from illogical (1) to logical (7). Figure 9 provides a graphical view of the mean responses. Both modified interfaces were similarly rated as providing a more logical arrangement (One: mean = 6.17, std. dev. 0.92; Two: mean = 6.46, std. dev. = 0.51). Again the 2.0 interface received the lowest rating (mean = 3.63, std. dev. = 0.88).

The modified interfaces were found to be significantly more logical based upon a one-way ANOVA ($p < 0.01$, MS = 58.29, $F(2,69) = 93.72$). The two-way ANOVA between the two modified interfaces found no significant difference.

However, the remaining two-way ANOVAs were significant. The comparison between the 2.0 interface to Interface One found ($p < 0.01$, MS = 77.52, $F(1,44) = 96.76$) while the comparison of the 2.0 interface and Interface Two found ($p < 0.01$, MS = 96.33, $F(1,44) = 197.15$). Therefore, the participants found the modified interfaces' play arrangement and grouping more logical.

The participants were asked to rate the helpfulness of the robot status panel. Very little difference in the responses existed across the three interfaces. The mean responses ranged from 4.33 to 4.79 (2.0: mean = 4.33, std. dev. = 1.24; One: mean = 4.79, std. dev. = 1.18; Two: mean = 4.58, std. dev. = 1.32). The statistical analysis found no significant results for either the one-way or two-way ANOVAs.

The participants were also asked to rate the adequacy of the information displayed on the robot status panel. Again, there was little difference across the three interfaces (2.0: mean = 5.33, std. dev. = 0.96; One: mean = 5.67, std. dev. = 0.70; Two: mean = 5.42, std. dev. = 0.88). Both the one-way and the two-way ANOVAs were not significant. Thus reducing the amount of information displayed and when it is displayed seemed to have little effect.

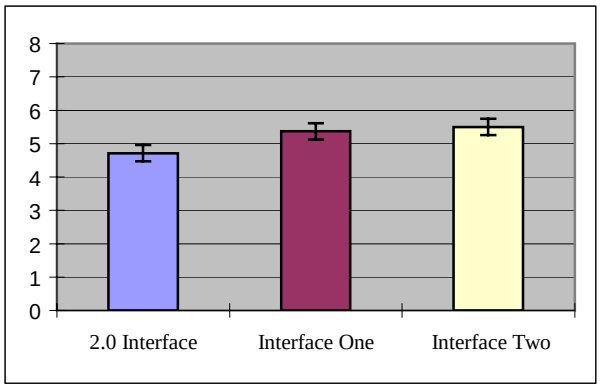

Figure 10. Mean rating regarding status panel information presentation by interface.

The arrangement of the robot status panel information was rated on a scale from illogical (1) to logical (7). The modified interfaces received the highest responses (One: mean = 5.37, std. dev. = 0.71; Two: mean = 5.5, std. dev. = 1.02), as shown in Figure 10. The 2.0 interface received the lowest rating (mean = 4.71, std. dev. = 1.0).

The one-way ANOVA found these results to be significant ($p < 0.01$, MS = 4.35, $F(2,69) = 5.12$). The two-way analysis comparing the two modified interfaces was not significant, while those comparing the original with each of the modified interfaces were significant. The comparison between the 2.0 Interface and Interface One found ($p < 0.01$, MS = 5.33, $F(1,44) = 6.8$) while the comparison between the 2.0 interface and Interface Two found ($p < 0.01$, MS =7.52, $F(1,44) = 7.08$). The result is contradictory given that the status panel was identical for the 2.0 interface and Interface One. One the other hand, Interface Two did provide a status panel with much less information over the other two interfaces. The analysis between Interfaces One and Two found no significant result, therefore, the results pertaining to this question must be considered inconclusive.

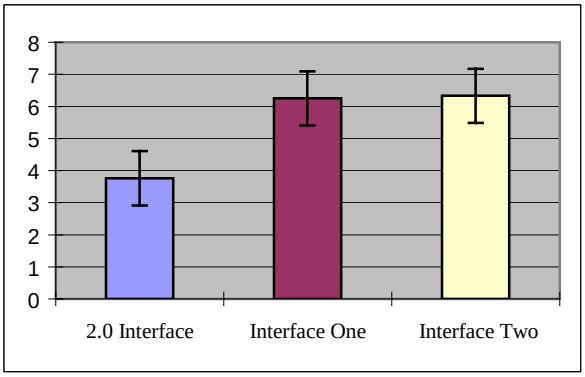

Figure 11. Mean rating for learning the system by interface.

The participants rated their ease of learning to use the system. Both of the modified interfaces received higher but similar ratings (One: mean = 6.25, std. dev. = 0.74; Two: mean = 6.33, std. dev. = 0.82), as shown in Figure 11. The 2.0 interface received the lowest rating (mean = 3.76, std. dev. = 0.88).

The one-way ANOVA found a significant difference across interfaces ($p < 0.01$, MS = 68.10, $F(2,69) = 103.36$). The two-way analysis comparing the modified interfaces to each other was not significant. The analysis comparing the 2.0 interface to Interface One was significant ($p < 0.01$, MS = 99.18, $F(1,44) = 181.21$). Similarly, the analysis comparing the 2.0 interface to Interface Two was also significant ($p < 0.01$, MS = 105.02, $F(1,44) = 160.73$). While the two modified interfaces are indistinguishable, they are easier to learn than the 2.0 interface.

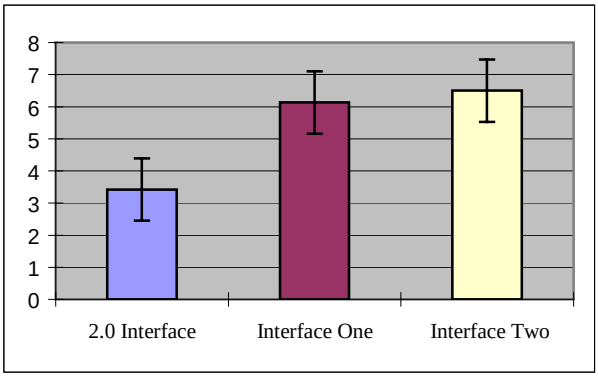

Figure 12. Mean rating for getting started by interface.

Figure 12 provides the participants mean ratings for getting started with the system. The responses for the modified interfaces were again quite similar (One: mean = 6.13, std. dev. = 0.95; Two: mean = 6.5, std. dev. = 0.51) while the 2.0 interface was rated as more difficult (mean = 3.42, std. dev. = 0.93).

The one-way ANOVA across interfaces found a significant result ($p < 0.01$, MS = 67.93, $F(2,69) = 100.89$). The comparison between the modified interfaces found that Interface Two was significantly easier to get started with than

Interface One ($p = 0.04$, $MS = 1.68$, $F(1,44) = 4.11$). There was also a significant difference when comparing the 2.0 interface and Interface One ($p < 0.01$, $MS = 88.02$, $F(1,44) = 118.26$). Additionally, the two-way analysis between the 2.0 Interface and Interface Two was significant ($p < 0.01$, $MS = 114.08$, $F(1,44) = 212.1$). While both modified interfaces are significantly easier to get started with, Interface Two was rated significantly easier to start using over the others.

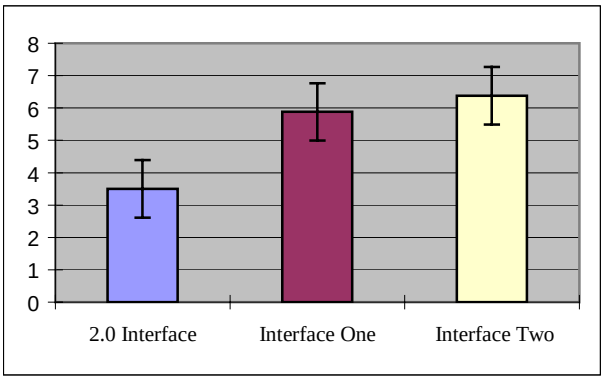

Figure 13. Mean play name and use recall by interface.

The ability to recall the names and uses of the different plays was rated as shown in Figure 13. The highest rating was provided for Interface Two (mean = 6.38, std. dev. = 0.65), while the 2.0 interface was rated the lowest (mean = 3.5, std. dev. = 1.06).

The one-way ANOVA found that both modified interfaces were significantly easier than the original interface ($p < 0.01$, $MS = 56.63$, $F(2,69) = 52.62$). The comparison between the two modified interfaces found a significant relationship ($p = 0.04$, $MS = 3$, $F(1,44) = 4.69$). The two-way comparison between the 2.0 interface and Interface One was significant ($p < 0.01$, $MS = 5.36$, $F(1,44) = 60.27$), as was the comparison between the 2.0 interface and Interface Two ($p < 0.01$, $MS = 99.18$, $F(1,44) = 143.48$). These results imply that Interface Two was the easiest interface for play name and use recall.

5 Discussion

The results obtained from this study are positive. They established that the two modified interfaces increased the RoboFlag system usability. The use of the original interface to train the participants may have skewed the results, but due to the lack of previous usability evaluations for RoboFlag, as well as time and resource constraints, the choice to use the original interface for training was made.

The modified interfaces fared better than the original interfaces on all usability questions but two. The results pertaining to the status panel usability was contradictory and therefore inconclusive. The modified interfaces created a user interaction that was more satisfying and stimulating. Interface Two was found to be significantly easier to use, more stimulating, and more satisfying.

The modified interfaces' categorization of the plays simplified the participants' play selection process. An interesting result was that the participants perceived that the adequacy of the number of plays in the modified interfaces was better, even though all interfaces contained the same number of plays. This result may be attributed to the reorganization of the play presentation that categorized the plays thus simplifying the ability to find a play for selection.

As previously stated, the actual game scores were not logged. However, the experimenters observed that the participants' performance improved with the modified interfaces, and was the best with Interface Two.

6 Conclusions and Future Work

In this work, the RoboFlag 2.0 system was evaluated to identify issues that might cause decreased usability and negatively affect game performance. Once these issues were identified, alternative design solutions that could resolve the identified issues were developed. Two new interfaces were implemented to incorporate the proposed changes. A study evaluating the modified interfaces was then conducted with 24 participants. The analysis of the usability results indicates that the modified interfaces were indeed more usable than the RoboFlag 2.0 interface. Additional analysis regarding perceived situation awareness across the three interfaces was also conducted and is reported in [2].

Acknowledgement

We thank Dave Schneider and Raja Parasuraman for their helpful comments and assistance.

References

[1] RoboFlag Website: http://roboflag.mae.cornell.edu/.

[2] S. Shankar, L. Su, Y. Jin, J. A. Adams, and R. Bodenheimer. "Enhancing RoboFlag Users' Situational Awareness"**,** *Proc of the 48th Human Factors and Ergonomics Society Annual Meeting*, Sept. 2004.

[3] J. Veverka and M. Campbell, "Experimental Study of Information Load on Operators in Semi-Autonomous Systems", *Proc. of the 2003 AIAA Guidance, Navigation and Control Conference*, 2003.

[4] R. Parasuraman, S. Galster, and C. Miller, "Human Control of Multiple Robots in the RoboFlag Simulation Environment", *Proc. of the 2003 IEEE International Conference on Systems, Man, and Cybernetics*, 2003.

[5] R. D'Andrea and M. Babish, "The RoboFlag Testbed", *Proc. of the American Controls Conference*, 2003.

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