

A Multi-modal Interface for Natural Operator Teaming with Autonomous Robots (MINOTAUR)

Stephanie Kane, Kevin McGurgan^(✉), Martin Voshell, Camille Monnier, Stan German, and Andrey Ost

Charles River Analytics Inc, 625 Mt. Auburn Street, Cambridge, MA 02138, USA
kmcgurgan@cra.com

Abstract. Dismounted squads face logistical problems, including the management of physical burdens in complex operating environments. Autonomous unmanned ground vehicles (UGVs) can help transport equipment and supplies, but require active remote control or teleoperation, even for mundane tasks such as long-distance travel. This requires heads down attention, causing fatigue and reducing situational awareness. To address these needs, we designed and prototyped a Multi-modal Interface for Natural Operator Teaming with Autonomous Robots (MINOTAUR). The MINOTAUR human-robot interface (HRI) provides observability and directability of UGV behavior through a multi-modal interface that leverages gesture input, touch/physical input through a watch-based operator control unit (OCU), and voice input. MINOTAUR's multi-modal approach enables operators to leverage the strengths of each modality, while the OCU enables quick control inputs through lightweight interactions and at-a-glance information status summaries. This paper describes the requirements and use case analysis that informed MINOTAUR designs and provides detailed descriptions of design concepts.

Keywords: Human factors · Human-robot interaction · Watch-based interface · Multi-modal interface

1 Introduction

Dismounted squads often face logistical problems, such as the management of physical burdens in complex operating environments. Autonomous unmanned ground vehicles (UGVs) can help transport more equipment and supplies than can be carried by hand or in backpacks. However, these platforms often require active remote control or teleoperation, even for mundane tasks such as long-distance travel. This requires heads down attention from operators, which causes fatigue and reduces situational awareness, making it difficult to maneuver nimbly or watch out for threats. Poorly designed human-robot interfaces (HRIs), which integrate new autonomous capabilities at the expense of good HRI design, further limit the operational benefits of current systems. As a result, users require extensive training for interfaces that do not directly address their needs and only allow them to use a fraction of the available operational capabilities. A successful system should enable UGVs to reliably and autonomously follow a

dismounted operator and free the Warfighter from control tasks, improving situational awareness and reducing cognitive burden.

To address these needs of dismounted squads, we designed and prototyped a *Multi-modal Interface for Natural Operator Teaming with Autonomous Robots* (MINOTAUR) HRI. The MINOTAUR HRI consists of a UGV (including hardware and software) and a lightweight, wearable operator control unit (OCU), similar in form factor to a wristwatch. MINOTAUR interface designs were informed by a requirements analysis, which identified a broad set of operationally relevant use cases, such as a lead/follow arrangement, changing operational environments, and a range of UGV health and status problems.

MINOTAUR provides observability and directability of UGV behavior through a multi-modal interface that leverages gesture, touch/physical input through a watch-based OCU, and voice input. This approach enables operators to flexibly and opportunistically choose operationally appropriate input modalities and to provide redundant commands across modalities (e.g., a “stop” command simultaneously issued verbally and with a gesture), which promotes robustness in challenging environments and improves command accuracy. This approach also enables operators to leverage the strengths of each modality to provide additional information on base commands, such as giving a verbal command to go to a particular location while providing directional input with a pointing gesture. To minimize the amount of “head down” time, the MINOTAUR watch-based OCU enables quick control inputs through lightweight interactions as well as at-a-glance information status summaries. This enables operators to quickly understand and modify UGV behavior while maintaining focus on the mission at hand.

This paper describes the operational problems faced by MINOTAUR dismounted squad users, as well as the requirements and use case analysis that informed MINOTAUR interface designs. It also provides detailed descriptions of select interface design concepts.

2 Requirements and Use Case Analysis

To inform MINOTAUR design activities, we performed a work domain analysis of an envisioned small team equipped with UGVs conducting operational field maneuvers. This analysis was based on literature reviews and knowledge elicitation interviews with a Marine Corps subject matter expert. As part the analysis, we identified a set of initial support themes to inform design activities, developed a formal abstraction hierarchy [1–3] of a subset of squad and team leader operations, and defined a notional operational scenario.

2.1 Initial Support Themes

Based on the results of our work domain analysis, we defined an initial set of support themes for envisioned squad-based operations. These themes capture a broad set of support needs for human-robot interaction in squad-based contexts, and provide a basis for future analyses (e.g., scenario development, requirements definition), as well as early design activities.

Theme 1: Squad Dynamics. Generally, a platoon consists of three squads, and within each square are three teams. Each team is made up of a team leader, an M249 Squad Automatic Weapon (SAW) gunner, an A-gunner, and a general rifleman. Within a squad, each one of these teams will function as the lead team, scouring the area, while the second team would then serve as the support, with the third team providing overwatch and security for the rear. In any case, each team will require a specific item or resource related to their specific objectives. For example, if a team will be responsible for entering a building, the first team would conduct recon and get to the building first. The second team would then bring the specialized item to assist in breaking down the door. The third team would then provide security. This sort of functional division of labor demonstrates the need for synchronized coordination.

Theme 2: Mission Objectives. A platoon is always going to have a specific mission objective. These objectives could be related to patrol (e.g., securing a perimeter or going out and seeking a target), providing security, conducting search and rescue, or supplying other platoons. Each one of these missions, while functionally similar, will potentially have very different operational tempos and require significantly different forms of support. Generally, a squad would not be sent out if they will be in contact in an over-match situation (e.g., against a platoon or two platoons); however they must constantly monitor and adapt to the high potential for surprise.

Theme 3: Squad Communication. Coordination and communications depend heavily on the type of mission, the time of day, and the terrain/surroundings the platoon is facing. Communications are primarily conducted through hand-signals and voice communications over digital radios (reliable), as well as verbal commands (e.g., shouting “contact right”). At night, squads will employ night vision goggles (NVGs) that will let them use hand signals, and communicate quietly through radios. Similarly, the definitions of smoke colors would be determined when creating the mission.

For the squads and teams, hand signal communication is heavily dependent on how far away squad members are from one another, with radio comms being the typical fallback. While the most common hand signals are relatively simple, they depend largely on the types of formations used by the team/squad. The types of hand signals used are dependent on the environment – different signals are used in mobile/urban locations than forest or jungle locations.

Theme 4: Managing Spatial Proximity. For a squad, spacing depends largely on the terrain. In an open area, spacing could span 100–200 yards. However, if the terrain is difficult, spacing would likely be closer together (e.g., 75 yards), though the squad and team leader remain mindful of not being too close. Within a team, formation and spacing depends primarily on the mission, terrain, and time of day. In a team of four people, individuals would likely be operating within a 25-yard radius in a smaller area and a 75-yard radius in an open area. Both squad and team spatial organization depend largely on the level of danger of the mission and the potential for receiving enemy fire.

2.2 Envisioned Mission Scenario

Based on the previously described analyses of squad-based operations, we developed a general scenario to explore OCU use cases, HRI, and performance contexts within the setting of this scenario. This highlights coordination opportunities and challenges for an envisioned human-robot team performing a reconnaissance mission.

Mission Phase 1: Moving Out from the Assembly Area. A platoon receives a mission objective to conduct reconnaissance. The squads will leave from a secure location based on an order within the squad detailing which team would go first and how they would set the perimeter for each team to go out. Once a secure perimeter is established, the other teams with the squad take cover and set security for the team initially leaving the safe line of the assembly area.

After the first team heads out, the team leader determines how best to organize the rifleman, SAW gunner, and A-gunner. As the team progresses toward their assigned area, they move at a pace determined by the team leader. Per their training, each team member takes a few steps (3–4, depending on terrain) before looking back at each other to ensure team members are staying in contact within a particular distance. This relative distance is critical to how the team dictates how and where to give signals. The team leader, at the back of this formation, gives a series of hand signals that are passed to the rifleman (who is unable to see the team leader). Each team will depart the assembly area in a similar manner, and everyone will look back to their squad leader for orders.

For squad-level movements, the squad leader positions himself in the middle of the team leaders. For example, in a wedge position, in which the first team is in the front with a team on either side, the squad leader positions himself in the middle of the formation. The team leaders organize themselves toward the inner parts of their team formations, and frequently look back to receive instructions from the squad leader.

Mission Phase 2: Moving Through Terrain. Although the squad's radios are relatively reliable, visibility decreases as the squad moves through a hilly, heavily wooded area. In these circumstances, it is largely the responsibility of the rifleman to make a good path for the team.

As the squad and teams use different formations to better deal with difficult terrain or guard against potential ambushes, team leaders constantly optimize their position for awareness of the squad leader. The squad and team leaders constantly gauge the location and distance of teams and team members. The squad leader must maintain awareness of all team members and choose the most appropriate formation (e.g., Wedge, Skirmishes Right, or Skirmishes Left). As they move through the terrain, the squad leader relies entirely on the team leaders to communicate positions and key information. The squad is able to use provided intelligence to rapidly locate, assault, envelop, and overcome the enemy.

Mission Phase 3: Additional Support and Return. After the squad has identified, positioned, and engaged with enemy forces, the squad leader decides to drop smoke to identify the need for additional supplies and show their position to indicate where they want additional fire support (either from aircraft or ship-based). The lead team has

identified a building for destruction, and the smoke allows the squad to mark their position and provide exact coordinates where they require additional fire support (or extraction). The use of different colored smoke denotes different mission needs, and the definitions of smoke colors are determined during mission planning. Upon successful destruction of the objective, the team identifies the primary path from the objective area and returns safely.

3 OCU Interface Design Concepts

Based on the envisioned scenario, we designed interface concepts that focused on general UGV control, and more specifically on modifying the UGV's following behavior. As part of the MINOTAUR effort, we explored a broad range of control inputs, display devices, and interaction methods. The MINOTAUR multimodal Operator Control Unit (OCU) consists of a watch-based visual display, and accepts physical touch inputs, gesture-based inputs, and voice inputs. This multi-modal approach will enable the operator to flexibly and opportunistically employ input methods based on specific operational needs. It also enables redundant and orthogonal commands across multiple modalities.

We utilized an iterative, work-centered approach grounded in established Cognitive System Engineering (CSE) methods [4] to develop OCU interface concepts that minimize learning time and are well suited for the warfighter in envisioned squad-based operational contexts. The OCU provides quick control inputs through lightweight interactions as well as at-a-glance information status summaries to minimize the amount of operator heads down time. It will also increase the *observability* and *directability* [5, 6] of UGV functions to enable rapid and robust interactions with the robot in dynamic operational environments. Increased observability provides the operator with insight into the current and future activities of automated processes. Observability techniques also include support for operator understanding of the limitations of automation (e.g., speed constraints, connectivity problems). Increased directability enables the operator to efficiently and purposefully direct and re-direct resources, activities, and priorities as situations change and escalate.

3.1 Natural Multimodal Interface Design Concepts

Across the MINOTAUR multimodal interface toolkit, we designed and prototyped interface concepts to enable operators to provide control inputs over multiple modalities, including touch, gesture, and voice inputs [7]. A key advantage of multimodal information display and input methods is their ability to improve the amount of information that can be conveyed to and provided by the operator, as well as the likelihood that the operator will perceive and respond to conveyed information. We will purposefully leverage channels and rendering methods that will be perceptually compelling and successful in squad-based operation contexts. One important aspect of multimodal information design is the consideration of priority of perceptual channels, as information in some channels is harder to ignore, particularly when information conflicts.

The MINOTAUR multi-modal interface approach will enable the operator to *flexibly and opportunistically* determine which interface modality or modalities to employ. For example, the operator could use voice commands in relatively safe operating conditions, and gestures when voice commands are dangerous due to nearby enemies. This approach also enables the operator to *redundantly* provide commands across multiple modalities to improve transmission of information. For example, the operator could simultaneously issue a stop gesture and voice stop command. This promotes robustness in challenging environments (e.g., degraded sound quality or visibility), and improves the accuracy of commands.

Finally, our approach enables orthogonal commands across multiple modalities, such as verbally directing the robot to go to a location and pointing a direction to convey additional, more specific spatial information. For example, MINOTAUR's multi-modal interface approach enables the operator to *orthogonally* provide commands across multiple modalities, such as verbally directing the robot to go to a location and pointing a direction to convey additional spatial information. This approach leverages the strengths of each modality, and enables the operator to provide additional information on base commands. This approach leverages the strengths of each modality, and enables the operator to provide additional information on base commands.

3.2 Watch-Based Control Interface

Initial MINOTAUR design activities focused on the Operator Control Unit (OCU), which enables the operator to interact with the UGV through a watch-based form factor. We designed a Tracker View, which provides observability and directability of current UGV behavior, a Command Status display, which provides observability of the UGV's processing of received commands, and a Command Log, which provides observability of the history of commands provided to the UGV.

Follow Modes. An initial focus of MINOTAUR OCU design efforts was a Tracker View, which provides observability of the UGV's current following behavior and enables the operator to modify the manner in which the robot follows the human operator through lightweight interactions. We explored a range of interaction methods for modifying robot behavior, including toggling and drag-and-drop concepts.

Figure 1 below shows a workflow for changing the robot's follow mode from loose to exact using the Tracker View. In this figure, the pink lines show the operator's path through the interface. The pink circles indicate the location of user interactions or gestures, such as finger presses and drag-and-drop locations. The blue circle with white outline (located at the bottom of the screens) represents the robot and the white circle with the black X-shape (located at the top of the screens) represents the human team leader. The unbroken white and green line represents exact following and the dashed, curved green and white line represents loose following. The first panel of Fig. 1 shows the primary Tracker view display, which provides a high level summary of the robot's current leader/follower mode behavior (e.g., exact/loose) through a vertical orientation to visually reinforce the specified leader/follower configuration. Additional support information on the execution of the behavior, such as distance and speed, appears in the

bottom right hand corner. The operators can set distance and speed as constraints or they can be derived from the vehicle itself. User-defined constraints are displayed with a lock icon preceding the data value. This allows the operator to maintain a higher level of control over the robot’s task execution.

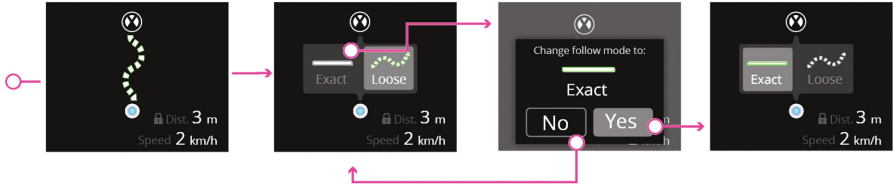


Fig. 1. A notional series of interface concepts for changing the robot’s follow mode from loose to exact using the Tracker View. The vertical orientation of operator and robot symbols visually reinforces the specified leader/follower configuration. Saliency cueing visually guides the operator through the task of switching the robot’s follow mode, while a two-step confirmation prevents operator errors.

Within the Tracker View, the operator can toggle between the robot’s two follow modes – exact and loose. In exact mode, the robot follows the operator’s exact route. In loose mode, the robot autonomously navigates its own route. The second panel of Fig. 1 depicts the updated touch-based toggle control. In this toggle control, the operator can see visual representations of each mode, with supporting text describing the available modes. This toggle also provides salient cueing to the operator of the robot’s current follow mode. Providing this additional context directly within the toggle control itself reinforces the operator’s mental model of the current state of the robot and accelerates learning of controls and available options for new users who may be unfamiliar with the interface. To prevent accidental touch-based inputs, this view employs a two-step confirmation approach. The third panel displays a dialog box that shows the operators new along with “yes” and “no” buttons. Finally, the last panel of Fig. 1 displays the updated control for switching from the new exact mode to the original loose mode.

Figure 2 below shows a parallel transition from *exact* follow mode to *loose* follow mode. In the first panel, the robot is currently following the operator and must maintain a distance of 3 meters (denoted by the lock icon). The second panel depicts the updated control depicting the current mode (*exact* follow mode) along with the *loose* follow mode option. As the operator selects the *loose* option, a confirmation dialog pops up to

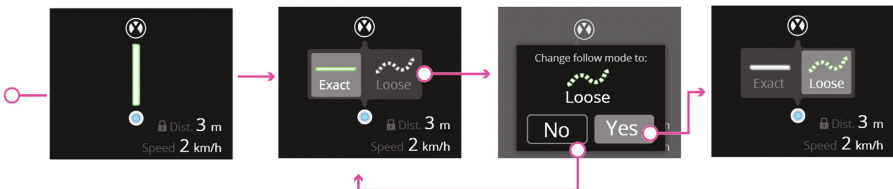


Fig. 2. A notional series of interface concepts for changing the robot’s follow mode from exact to loose using the Tracker View.

confirm the change, as seen in the third panel. Finally, the fourth panel depicts the updated control for switching from the new *loose* mode to the original *exact* mode.

Command Status. The Command Status Region appears below the Tracker View and provides observability of the processing status of commands issued to the UGV, including receipt of commands, processing of commands, and acceptance/execution of commands. Figure 3 below shows a series of screens that reflect the transition from exact follow mode to loose follow mode. At the top left of the figure, both the Tracker View and Command Status region show the UGV in “following exact mode.” The second display shows that a “follow loosely” command was received. The smaller size and less salient color of text and “Next:” text helps the operator understand that the command has been received, but is not yet being executed. In the next screen, a “refreshing” symbol appears on the right side of the Command Status region to indicate that the UGV is processing the new command. In the tracker view, the solid line between the leader and robot icons is semi-transparent to indicate that a new command is being processed. In the next screen, the follow loosely command is shown in larger white text at the top of the region to indicate that the command has been accepted and is now being executed, while the previous command is shown in smaller, darker text. The “following exact” symbol in the Tracker View (solid white line) has also been replaced with a “following exact” symbol (curved dashed line). Eventually, the text for the previous command disappears, leaving only the current command (as shown in the final screen in Fig. 3).



Fig. 3. A series of screens illustrating the Command Status Region, which appears below the Tracker View and provides observability of command status, including receipt of the command, processing, and acceptance/execution. This display region uses variable salience and integrates with the Tracker View to promote operator understanding of command status.

Command Log. The Command Log display provides observability of the history of commands provided by the operator to the UGV. The operator accesses this display by touching the Command Status Region, and enhances operator awareness of robot functioning in context. It can also provide the operator with insight into the effectiveness of

different command modalities (e.g., the operator can see that no gesture inputs have been accepted over the course of a mission).

Figure 4 shows the command log display, which shows operator commands ordered by recency, with the most recent command appearing at the top of the display. Each command in the log occupies a row within the display. An icon to the left of each command indicates the modality with which it was provided (e.g., speech bubble icon for a voice command, eye icon for a gesture command). The time since the command was given is shown for commands issued a short time ago (e.g., “1 min ago”, “10 min ago) and a timestamp is shown for commands issued longer ago. All past commands are displayed with dark grey text. The word “now” appears next to commands that are currently being executed, which are shown in white text, and a refreshing symbol appears next to commands that are being processed, which appears in light grey text. Finally, failed commands are shown in red text. Saliency mapping throughout the Command Log helps the operator quickly understand the various command statuses, and quickly identify commands that failed or that are pending.

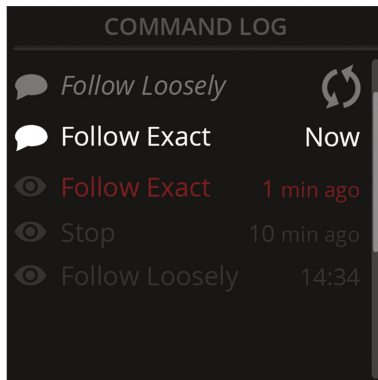


Fig. 4. The Command Log display, which provides observability of the history of commands provided by the operator to the UGV, and enhances operator awareness of robot functioning in context. Relevant command properties (e.g., time since command was given, current command, pending commands, failed commands) are provided, and saliency mapping helps the operator to quickly understand current, past and future UGV functioning.

4 Conclusion and Future Work

This paper described *Multi-modal Interface for Natural Operator Teaming with Autonomous Robots* (MINOTAUR), a human-robot interface for dismounted squad operations. This effort built upon analyses of squad-based operations to develop OCU concepts that improve observability and directability of UGV functions through lightweight interactions and at-a-glance information summaries. Examples of OCU concepts were presented, including a Tracker View, Command Status display, and Command Log.

This work sets the stage for continued development of OCU display concepts to accommodate additional squad-based use cases, such as waypoint-based navigation. Another focus area for follow-on efforts is “progressive enhancement” of UGV commands (i.e., commanding the UGV to go to a location and later modifying that command so the UGV chooses its own route and goes to the location more quickly). This would allow the operator to update commands based on the current operational context without needing to repeat commands unnecessarily. Future efforts will also explore robust UGV health and status displays. For example, status displays for the various command modalities will explore ways to provide operator cues that enable graceful degradation when one or more modalities are unavailable. By alerting the operator to failures as they occur, health and status displays will also enable the proactive management of UGV issues. The key challenge of these displays will be the balance between showing critical information and alerts while minimizing operator heads down time. Because we developed a broad set of display concepts, future efforts will also focus on user testing to refine our OCU designs.

References

1. Vicente, K.J.: *Cognitive Work Analysis: Towards Safe, Productive, and Healthy Computer-Based Work*. Lawrence Erlbaum Associates, Mahwah (1999). Foster, I., Kesselman, C.: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (1999)
2. Rasmussen, J., Pejtersen, A., Goodstein, L.P.: *Cognitive Systems Engineering*. Wiley, New York (1994)
3. Woods, D.D., Roth, E.M.: Cognitive engineering: human problem solving with tools. *Hum. Factors* **30**(4), 415–430 (1988)
4. Hollnagel, E., Woods, D.D.: *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. CRC Press, Boca Raton (2005)
5. Woods, D.D., Hollnagel, E.: *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. CRC Press, Boca Raton (2006)
6. Kilgore, R., Voshell, M.: Increasing the transparency of unmanned systems: applications of ecological interface design. In: Shumaker, R., Lackey, S. (eds.) *Virtual, Augmented and Mixed Reality. Applications of Virtual and Augmented Reality*. Springer, Orlando (2014)
7. Oviatt, S.: Multimodal interfaces. In: Jacko, J., Sears, A. (eds.) *Handbook of Human–Computer Interaction*. Lawrence Erlbaum, Mahwah (2002)