Field Experiments using SPEAR^{TM}: A speech control system for UGVs

Siddharth R. Chhatpar*[a], Chris Blanco[a], Jeffrey Czerniak[b], Orin Hoffman[b], Amit Juneja[a], Tarun Pruthi[a], Dongqing Liu[b], Robert Karlsen[c], Jonathan Brown[a]
[a]Think-A-Move, Ltd., 23715 Mercantile Road, Suite 100, Beachwood, OH, USA 44122-5931;
[b]iRobot Corp., 63, South Avenue, Burlington, MA, USA 01803
[c]TARDEC, 6501 E. 11 Mile Road, Warren, MI, USA 48397-5000

ABSTRACT

This paper reports on a Field Experiment carried out by the Human Research and Engineering Directorate at Ft. Benning to evaluate the efficacy of using speech to control an Unmanned Ground Vehicle (UGV) concurrently with a hand-controller. The SPEAR system, developed by Think-A-Move, provides speech-control of UGVs. The system picks up user-speech in the ear canal with an in-ear microphone. This property allows it to work efficiently in high-noise environments, where traditional speech systems, employing external microphones, fail. It has been integrated with an iRobot PackBot 510 with EOD kit. The integrated system allows the hand-controller to be supplemented with speech for concurrent control. At Ft. Benning, the integrated system was tested by soldiers from the Officer Candidate School. The Experiment had dual focus: 1) Quantitative measurement of the time taken to complete each station and the cognitive load on users; 2) Qualitative evaluation of ease-of-use and ergonomics through soldier-feedback. Also of significant benefit to Think-A-Move was soldier-feedback on the speech-command vocabulary employed: What spoken commands are intuitive, and how the commands should be executed, e.g., limited-motion vs. unlimited-motion commands. Overall results from the Experiment are reported in the paper.

Keywords: SPEAR, speech control, UGV control, human-robot interaction, in-ear speech, testing and evaluation

1. INTRODUCTION

This paper describes the development of a prototype system with speech-based control supplementing hand-controller based tele-operation of a UGV. It also describes a Field Experiment at Ft. Benning to test the efficacy of the supplemental speech control system in reducing operator task-time and workload.

There are many applications with inherently noisy environments where hands-free device control is a requirement. One such rapidly growing application is the control of robots, particularly with the rise in the number of Unmanned Ground Vehicles (UGV) employed by the military [1]. Moreover, as the Army's Future Combat Systems (FCS) program—their principal modernization program—is implemented, the number of Small UGVs (SUGV) employed by the Army is going to increase significantly. On the domestic scene, robots are used for homeland defense; including bomb-disposal squads, SWAT teams, and border patrol. Besides this, potential future applications include relief operations in disaster zones, e.g., fire sites, flood-stricken areas, nuclear accident sites, etc, that are hazardous for human rescuers.

Traditionally, these robots are tele-operated through a joystick-based Operator Control Unit (OCU), which tends to be big and bulky, requiring the operator to be confined to a military vehicle or some other stationary position. Moreover, joystick control of the robot requires the operator to employ both of his hands, leaving him defenseless in case of an attack. Hence, there has been a recent push to: 1) Reduce the size and weight of the OCU, making it wearable, allowing the operator to be mobile, and 2) Find a hands-free means of controlling the robot leaving the operator free to use his hands for carrying weapons or other tactical devices. These requirements are assimilated into the concept of the *Warfighter's Associate* developed by researchers at the Space and Naval Warfare Systems Command San Diego (SPAWAR) [2]. This concept calls for a significant effort to introduce novel control techniques to provide the robot with adequate autonomy to function as a *Warfighter's Associate,* as well as novel command input schemes to allow the operator to interact with the robot as he/she would any other fellow soldier. In this context, speech based command input is both intuitive and capable of handling hundreds of discrete commands. However, UGV autonomy is still in its infancy, and it will be a few years before it matures enough for use in Theater. Meanwhile, speech can still provide great benefit to the Warfighter by augmenting joystick-based tele-operation. As a supplemental control capability, speech commands can allow simultaneous control of numerous functionalities, not otherwise possible using only a hand-controller. Speech commands also provide an efficient means of accessing menu options, which otherwise require navigating through several layers of a menu system.

Traditional speech-command based control systems use external microphones—mounted on a boom to place them close to the user's mouth, and in field environments their performance suffers because of high decibel ambient noise: typically above 90dBA with impulses of up to 110dBA [3]. The technology developed by Think-A-Move (TAM), called SPEAR^{TM}, is based on capturing user speech in the ear canal, as shown in Figure 1, and can provide excellent command recognition accuracy even in noisy environments. A patented earpiece, called the SPEAR earpiece, has been developed with a microphone pointed into the ear canal that captures the speech signal [4][5][6]. For the prototype system described here, SPEAR was integrated with iRobot's *Lightweight OCU* to provide simultaneous control of the robot through a hand-controller as well as speech commands. To test the efficacy of using speech commands to supplement joystick control, the system was subjected to a Field Experiment at Ft. Benning. The Experiment had dual focus: 1) Quantitative measurement of the time taken to complete each station and the cognitive load on users; 2) Qualitative evaluation of ease-of-use and ergonomics through soldier-feedback. Also of significant benefit to Think-A-Move was soldier-feedback on the speech-command vocabulary employed: What spoken commands are intuitive, and how the commands should be executed, e.g., limited-motion vs. unlimited-motion commands.

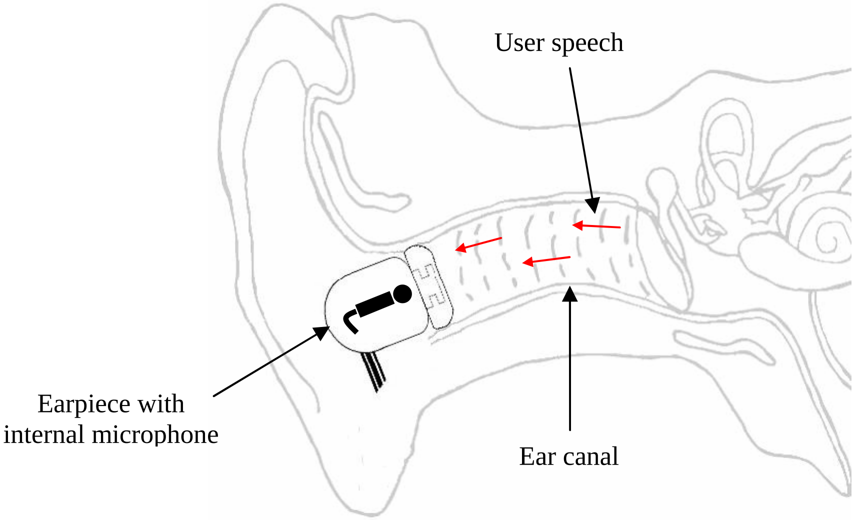

Figure 1. The earpiece fits in the ear and seals the ear canal from the environment. It contains a microphone that points into the ear canal and captures sound waves generated by user speech and transmitted internally.

The initial plan with the Field Experiment was to have 30 soldiers, each running the UGV through a testing course twice: once using manual control alone and another time using speech commands to supplement manual control. In designing the course for the Experiment, TAM wanted to focus on concurrent tasks that pose maximum difficulty in execution when only using the hand-controller. For example, navigating the menu using the hand-controller is very cumbersome and time-consuming. However, the procedure adopted by the Human Research and Engineering Directorate (HRED)—which was conducting the Experiment—in designing the course was to ensure that only those tasks that could be concurrently executed using both techniques—manual only and speech-supplemented—were to be included in the course. Therefore, some concurrent tasks had to be eliminated, and others, such as menu navigation, had to be altered to accommodate the above condition. For menu navigation, instead of using the hand-controller, buttons on the Portable Command Console were used, as described in Section 4.3.

As a result of the Field Experiment conducted at Ft. Benning, TAM gained significant knowledge enabling it to improve the implementation of speech commands for UGV control. The lessons learned ranged from hardware related – elimination of Push-To-Command button and checking for good sealing with the earpiece – to developing a more consistent structure for speech command vocabulary. The Experiment confirmed that there are substantial benefits to be gained by using speech commands simultaneously with the OCU or hand controller to potentially reduce time-to-mission-completion. TAM also collected valuable information on the performance of SPEAR in terms of speech recognition accuracy under field conditions. Based on the lessons learned from the Experiment, TAM has already undertaken several improvements to SPEAR, and has more planned for the future.

The rest of the paper is organized as follows. The anticipated benefits of using speech command control are outlined in Section 2. Section 3 describes the prototype system developed through the integration of TAM's speech control system with iRobot's Lightweight OCU to provide concurrent control using speech commands along with a hand-controller. The integrated system was used in a Field Experiment at Ft. Benning. The Experiment is described in Section 4. Section 5 provides results from the Experiments and a discussion of the lessons learned. Conclusions and a discussion of future work are provided in Section 6.

2. ANTICIPATED BENEFITS OF USING SPEECH COMMANDS FOR UGV CONTROL

This Section describes the anticipated benefits of using speech command-control. It starts with a brief background on new control input methods—e.g., video-game-style hand-controllers—to make UGVs more accessible to dismounted operators and expand their applicability to other missions besides the EOD. Against this background, the need for several new capabilities, such as hands-free and heads-up control, concurrent control and easier menu navigation, is established, and speech-based control is shown to be an ideal candidate to fulfill these requirements.

2.1 Background

Unmanned Ground Vehicles (UGV) have traditionally been tele-operated using joysticks mounted on bulky Operator Control Units (OCU). Examples include the iRobot Portable Command Controller (PCC) and the Foster-Miller Common Control Unit (CCU). While these bulky control units are adequate for EOD missions, where the operator is stationed inside an armored vehicle, they are highly impractical for Infantry missions, where the Warfighter is dismounted and mobile. Therefore, there has been a recent push to move to more practical solutions for dismounted operations, e.g., iRobot's Dismounted OCU, composed of a laptop-based control unit combined with a video game-style hand-controller. However, a dismounted operator on a mission faces several security challenges. While controlling the robot, the operator's hands are on the hand-controller and not on a weapon, hindering his ability to defend himself in case of an emergency. Being occupied with tele-operating the robot, the operator's situational awareness is diminished, making him more vulnerable to an attack. Thus, under the current setup, a dismounted operator would require a security detail to ensure his safety while he operates the robot. This limits the practical deployment of UGVs for Infantry missions since manning levels are not increased to support the operation of UGVs under these conditions. To meet this challenge, a hands-free and heads-up method for control input to UGVs needs to be devised, and a speech-based command input system provides an effective solution.

Some UGV control issues are exacerbated in moving from a large command console to a small hand-controller. One such issue is the activation of discrete commands, such as the selection of pre-configured arm poses and autonomous behaviors, through a many-layered menu tree. On the PCC, accessing a discrete command requires pushing several buttons to navigate through the menu. This is cumbersome and time-consuming. Moreover, menu navigation requires the operator to take his hand off the joystick to punch the buttons. With the hand-controller, menu navigation becomes even more cumbersome, because selecting from a list in the menu requires moving up or down the list sequentially using the joystick. Also, as a consequence, the joystick cannot be used for robot control while the menu is being accessed. On the other hand, using speech, the same discrete commands can be input without the need to navigate a menu.

Another issue is related to concurrent control – actuating multiple functions on the UGV simultaneously, e.g. panning the camera while driving the UGV. With the PCC, some concurrent control is possible depending on the dexterity of the operator, since more than one knob and/or joystick needs to be handled for controlling different functionalities simultaneously. To activate a command from the menu, the operator has to free one hand to push the appropriate buttons. In this context, concurrent control is not possible with the hand-controller, because navigating the menu requires the use of the joystick, thus eliminating its use for robot control. Concurrent control can be beneficial in many different situations as described below.

2.2 Benefits of concurrent control

Concurrent control can not only help reduce the overall time to mission completion, but also make robot control more efficient and intuitive. There are several mission scenarios where concurrent control plays a critical role; one such scenario is described here.

Mission scenario

Consider an EOD mission where the operator is driving the UGV out to an identified IED location. While it is critical for the operator to get the UGV out to the target within the shortest possible time, the operator is also tasked with scanning the surroundings on the way to the target to identify any secondary devices. This requires the operator to control the camera (pan, tilt, and zoom) while concurrently driving the robot. Based on operator preference, speech commands can be used to drive the robot while controlling the camera with the hand-controller, or vice versa. While driving, the operator encounters a pile of debris on the side of the road. Wanting to look over the debris, the operator gives speech commands to lower the flippers and raise the arm into high pose, while simultaneously using the hand-controller to position the robot in front of the pile. Similarly, the operator is able to position the arm for a good view through the window of a car parked on the side of the road as he/she drives the robot by it without having to stop.

In the absence of concurrent control, the operator would have to stop at short intervals to activate arm poses and control the camera. Thus, concurrent control contributes to significant reduction in time-to-target and overall time to mission completion.

3. INTEGRATED SYSTEM FOR CONCURRENT CONTROL

The ultimate goal of our project is to integrate Think-A-Move's speech control system with iRobot's Common OCU being developed for their Aware 2.0 software. However, at the time of this project, the Common OCU was still under development, and there was no fully developed OCU available that could support joystick control and a display interface. Therefore, TAM and iRobot undertook the development of a "Stage 1 integration" based on iRobot's prototype *Lightweight OCU*. In this section, we describe TAM's speech control system, SPEAR, and its integration with the Lightweight OCU in detail.

3.1 SPEAR for speech command-based control

SPEAR is Think-A-Move's speech command input system. Instead of picking up speech radiating from the mouth, it captures user speech in the ear canal as shown in Figure 1, and can provide excellent command recognition accuracy even in noisy environments. SPEAR consists of the earpiece, TAM's proprietary speech command recognition engine, and an API for rapid integration with robotic platforms.

The patented TAM earpiece, shown in Figure 2, has been developed with a microphone pointing into the ear canal that captures the speech signal [4][5][6]. A wired connection carries the signal from the TAM earpiece, and ends in a standard 3.5mm audio jack that can be plugged into a computer soundcard. The earpiece also contains TAM's patent-pending acoustic waveguide, which resolves a peculiar challenge faced by all in-ear sound capture devices. Generally, speech captured in the ear canal is missing sounds in the higher frequency range (above 2 kHz). This deteriorates the quality of the captured speech, and would adversely affect the accuracy rate of speech command recognition. The acoustic waveguide alters the frequency spectrum of the sound that reaches the microphone. By carefully selecting the dimensions of the acoustic waveguide, the signal in the desired frequency range can be amplified, thus restoring the quality of the captured speech.

A speech command recognition (SCR) system is used to identify the command spoken by the user. The SCR system collects the captured speech signal and outputs a recognized command. TAM has developed a proprietary SCR system with versions for both Linux and Windows platforms. The SCR engine is a Hidden Markov Model (HMM)-based system, and incorporates speech parameterization amenable to in-ear speech characteristics. TAM has trained specialized models tuned for recognition of in-ear speech, and the recognition accuracy target has been set to more than 95% even in extreme conditions, which include battlefield conditions where the operator is involved in intense physical activity like hiking and running, with loud noise in the background. The TAM SCR system is being developed for robustness to conversation and other operator speech that is not aimed at commanding the UGV. This capability is called command-spotting, and signifies that the SCR system can distinguish speech commands from conversational speech. However, at the time of the Experiment, the command-spotting capability was still under development. Therefore, a temporary push-button prototype was developed. This prototype allows the operator to activate the SCR system by pressing and holding a button whenever he/she wants to give a speech command. When the button is not pressed, the SCR system is not active, and any conversation or other operator speech will not trigger false command recognition. The push button was attached at a convenient location on the hand-controller so that the operator could use the ring finger or the pinkie to push the button. This allowed uninterrupted operation of the hand-controller, which requires the use of the thumb, index finger, and middle finger.

(a) Earpiece shell with foam tip

(b) Earpiece inserted into the ear

Figure 2. One-size-fits-all earpiece

Building on the SCR system, TAM has developed a standard C++ Application Programming Interface (API) to facilitate easy integration with a host of different systems. The API exposes all of the major components of a speech recognition system including microphone adjustment, training, and speech decoding. The API also contains a profile management module and advanced audio features such as a software microphone that allows pre-recorded speech to be decoded. This functionality allows for consistent and rapid integration with any platform. The software used in the Experiment was built upon this API.

3.2 Integration of SPEAR with iRobot PackBot EOD to provide concurrent control

The overall configuration for the Stage 1 integration is shown in Figure 3. The *Lightweight OCU* is a prototype developed by iRobot for in-house experiments. In order to support the functionality of speech commands as supplemental control inputs to a hand-controller, additional components were programmed on the robot, such as the *Teleop Listener*. As shown in Figure 3, the robot is controlled through two different OCUs concurrently: one is the *Legacy OCU* on a Portable Command Console (PCC) and the other is the Lightweight OCU running on a Panasonic Toughbook. The hand-controller is connected to the PCC, and facilitates tele-operation through the Legacy OCU. The TAM earpiece is connected to the Panasonic Toughbook, and speech control is facilitated through SPEAR integrated with the Lightweight OCU.

The integrated system in the Stage 1 prototype provides speech control supplementing the hand-controller. The supplemental control functionality available includes tele-operation, predetermined arm poses, as well as general commands, e.g., camera selection and control, including pan/tilt/zoom commands, flipper control, and illuminator on/off. Table 1 lists the speech command control functionality available with the Stage 1 integration.

The system is configured such that concurrent control is possible, i.e., the operator can control separate functionalities on the robot using the hand-controller and speech commands at the same time. For example, while driving the robot using the hand-controller, the operator can switch cameras or zoom in/out using speech commands without having to stop the robot, and vice versa.

Table 1. Functionality for the TAM-iRobot Stage 1 integration

Speech Command	Control functionality
Arm drive	Put arm in Drive mode
Arm deploy	Put arm in Drive High mode
Arm high	Put arm in High mode
Arm low	Put arm in Low mode
Arm retract	Put arm in Stow mode
Forward small	Move robot forward indefinitely with small velocity
Reverse small	Move robot reverse indefinitely with small velocity

Left large	Turn robot left indefinitely with large velocity
Left small	Turn robot left indefinitely with small velocity
Right large	Turn robot left indefinitely with large velocity
Right small	Turn robot left indefinitely with small velocity
Halt/Stop	Stop all motion on the robot; apply brakes
Brake on	Apply brakes
Brake off	Release brakes
Flippers forward	Move flippers forward
Flippers back	Move flippers back
Camera turret	Switch camera view on the PCC to the Turret camera
Camera underarm	Switch camera view on the PCC to the Underarm camera
Camera over arm	Switch camera view on the PCC to the Over-arm camera
Pan left	Pan attack camera to the left by a fixed amount
Pan right	Pan attack camera to the right by a fixed amount
Tilt up	Tilt attack camera up by a fixed amount
Tilt down	Tilt attack camera down by a fixed amount
Zoom in	Zoom in with the attack camera
Zoom out	Zoom out with the attack camera
Light on	Increase illuminator intensity on the attack camera to maximum
Light off	Reduce illuminator intensity on the attack camera to minimum

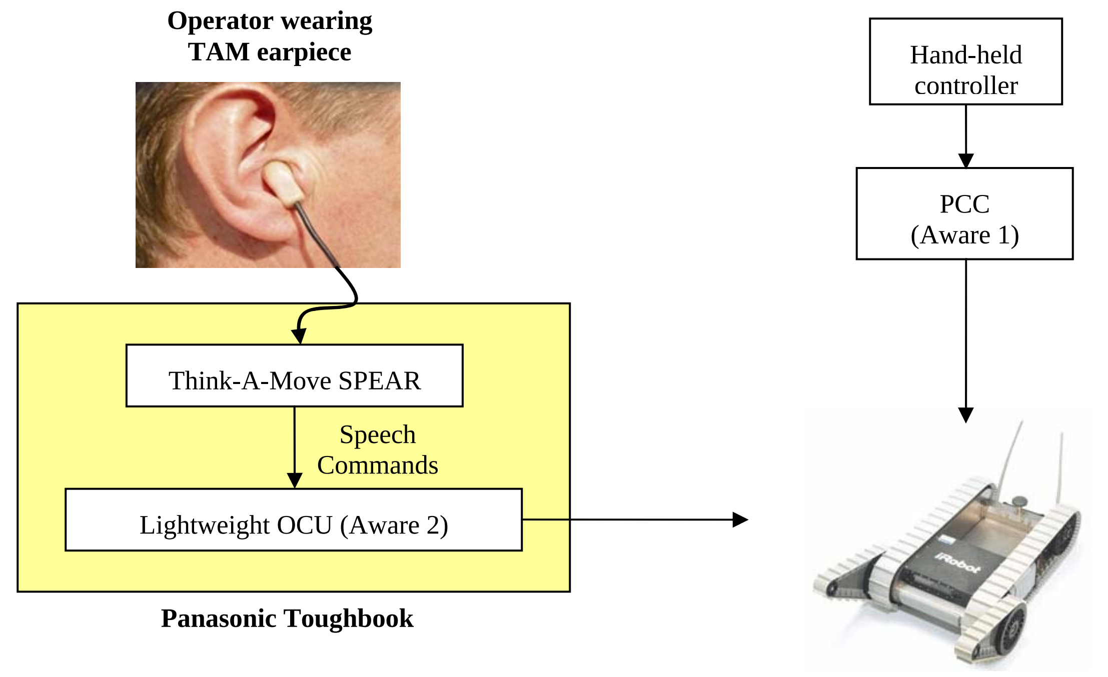

Figure 3. Overall configuration for the *Stage 1 integration*: The robot is controlled concurrently through the PCC using a hand-controller and through the Panasonic Toughbook using SPEAR integrated with iRobot's Lightweight OCU

4. FT. BENNING EXPERIMENT

The Stage 1 integrated system was utilized in a Soldier Experiment conducted by the Human Research and Engineering Directorate (HRED) at Ft. Benning. The Experiment was designed to evaluate the effectiveness of using speech as a supplemental control input method with the hand-controller. Soldiers from the Officer Candidate School (OCS) were recruited to act as test subjects. Think-A-Move and iRobot supported the Experiment with personnel and resources. In addition, the Tank-Automotive, Research Development and Engineering Command (TARDEC) provided the PackBot that was used in the Experiment. The overall plan included 2 days for set-up (August 21 & 22, 2008), and all of the following week (August 25-29, 2008) for the Experiment. The tasks involved in the set-up included:

- Designing the overall course and the individual stations,
- Determining the command set to be used to complete the tasks at each station, and
- Formulating the training procedure

Once the set-up was completed, some pilot runs were carried out. Based on these test runs, improvements were made to the course layout as well as the training procedure. The original plan was to collect data on 30 soldiers running the course. However, due to weather conditions (heavy rains and water-logging on the course), we were only able to collect data on 11 soldiers. Data were collected by HRED personnel on the time taken to complete each station and the cognitive load on users, as well as qualitative feedback from each soldier on speech control. TAM also collected feedback from the soldiers on the benefits/drawbacks of the speech control system. Summary and discussion of the results are provided in Section 5.

4.1 Motivation/Objectives

The general goal of the Experiment was to evaluate the effectiveness of using speech as a supplemental control input method with the hand-controller. To achieve this goal, the Experiment was designed with dual focus: 1) Quantitative measurement of the time taken to complete each station and the cognitive load on users, and 2) Qualitative evaluation of ease-of-use and ergonomics through soldier-feedback. For comparison, each soldier completed the course twice: once with only manual driving, and once with speech supplementing manual control.

TAM was also focused on collecting soldier-feedback on the speech-command vocabulary employed: What spoken commands are intuitive, and how the commands should be executed, e.g., limited-motion vs. unlimited-motion commands.

4.2 Course setup

The test course was divided into five stations, simulating common scenarios encountered in EOD missions, and designed to require the control of several capabilities on the UGV simultaneously. The stations are described below. The route between stations was marked with engineering tape. A driving error was assigned whenever the robot touched the engineering tape.

Station 1: Window reconnaissance

Station 1 has a wall with a high window. An IED is placed behind the wall, such that it is visible through the window. The operator needs to drive the robot up to the wall, and put the arm in *High* pose to look through the window and identify the IED. To minimize the time taken at this station, the operator is encouraged to multi-task: As the robot is being driven up to the wall, the flippers need to be moved forward and the “Arm High” pose needs to be activated concurrently. Once the robot and the arm are positioned correctly, the camera needs to be controlled to scan the view behind the window, zoom into any suspicious objects, and identify the IED. A driving error is recorded if the robot runs into the wall.

Station 2: Scan area and track moving target

This station also has a wall, but with no window. The operator needs to drive the robot to the wall, and put the arm in *High* pose to look over the wall, and to move the front flippers forward to provide stability. Similar to Station 1, the operator is encouraged to multi-task. Once the robot is in position, the operator needs to scan the horizon by panning the camera, and identify a human target. As the target starts moving across the horizon, the operator must pan the camera to track the target and report his/her activity. A driving error is recorded if the robot runs into the wall.

Station 3: Search area and identify five IEDs

At this station, five IEDs are scattered on the ground at random locations. The operator needs to drive the robot around the area and use the camera to search for the IEDs. The operator was allowed a maximum time of 5 minutes to complete this station. In case the operator failed to locate all the IEDs within the allocated time, the operator was asked to terminate the search, and the number of IEDs identified was recorded. A driving error is recorded if the robot runs over an IED.

Station 4: Bunker reconnaissance

The robot is to be driven up to the bunker, and the arm is to be put in *Low* pose to look into the bunker. To minimize the time taken at this station, the operator is encouraged to multi-task: As the robot is approaching the bunker, the flippers need to be moved forward for stability, and the "Arm Low" pose needs to be activated concurrently. Once the arm is in position, the camera is to be tilted down to look into the bunker and identify the IED. A driving error is recorded if the robot runs into the bunker covering.

Station 5: Tunnel reconnaissance

The robot is to be driven into the tunnel, and the light is to be switched on to identify an IED. A driving error is recorded if the robot touches any part of the tunnel.

4.3 Operating the robot

The operator was sitting in a tent, and did not have a direct view of the course. All the driving was executed using robot-centric camera views, shown on the PCC screen. All the commands, except the arm poses, could be actuated through the hand-controller. To activate the arm poses, the operator had to reach over to the PCC, pull up the menu, and navigate it to select the desired arm pose. Large labels were stuck next to the PCC buttons to remind the operator of the menu selections. This made it much easier for the operator to navigate the menu than under normal conditions, where the labels are not available for guidance.

During the speech-supplemented run through the course, at Stations 1, 2, 4, and 5, the soldiers were encouraged to drive the robot using the hand-controller, while simultaneously using speech to control the flippers, activate arm poses, and control camera motions. However, at Station 3, soldiers were encouraged to use speech to drive the robot and execute camera controls (pan, tilt, and zoom) using the hand-controller.

4.4 Experiment

Since each soldier runs the robot through the course twice, once driving manually, and once with speech supplementing manual control, there is a possibility that after the first run through the course, the soldier is more experienced, which helps with the second run. To eliminate bias related to this effect, alternate soldiers were required to switch the sequence in which they performed the manual and speech-supplemented driving, i.e., if Soldier 1 performs manual driving first followed by speech-based, then Soldier 2 would perform speech-based driving first followed by manual. This should minimize the "training" effect on the results of the Experiment.

To compare the mental workload on the operator under the two driving methods, a cognitive test was administered during the run through Station 3. While the Soldier was driving the UGV through Station 3, looking for IEDs, the test administrator would provide several pieces of information to the Soldier: challenge, password, time of attack, and direction of attack. The Soldier was expected to memorize this information, and at the end of Station 3, the Soldier was asked to recall the information. The administrator would score the Soldier based on how much information he/she was able to recall accurately.

A data collector walked with the robot as it was being driven through the course. The data collector recorded driving errors and time taken at each station. All the data collected through each run was tabulated for comparison between manual driving and speech-supplemented driving.

5. RESULTS & DISCUSSION

The main objective of the Field Experiment was to evaluate the efficacy of using speech commands for supplemental control input. The focus of the Experiment was not on the performance of the speech system in terms of recognition accuracy rates. However, we present a short discussion of the speech system performance below.

5.1 SPEAR performance

The performance evaluation of SPEAR was carried out based on the analysis of sound files recorded during the Experiment. These files were recorded and analyzed for 8 soldiers. The sound files provided information on the effectiveness of the earpiece seal in blocking ambient noise, the correct use of the Push-To-Command button, correct usage of speech commands, and recognition accuracy rates. The results are summarized below:

- The average recognition accuracy rate for 7 of the 8 soldiers was above 96%, and for one soldier it was 74%.
- For approximately 8% of the cases, the speech commands were wrongly spoken or the Push-To-Command button was used incorrectly—the button was pressed too late or released too early with respect to the speech command. There were also some instances when the button was pushed but no command was spoken.
- Some users had a bad seal on the earpiece, possibly from improper insertion of the earpiece. The bad seal contributed to reduced accuracy.

Think-A-Move undertook several improvements to SPEAR based on the conclusions drawn from the analysis of the sound files. These are listed below, and described in greater detail in Section 6.1.

- Eliminating the Push-To-Command button, and moving to an open microphone system that is capable of differentiating speech commands from conversation.
- A software method for checking the seal on the earpiece every time the earpiece is inserted into the ear.
- Providing a more consistent structure to the speech command vocabulary used, to minimize the load on the Operator.

5.2 Evaluation of supplemental control using speech commands

The quantitative evaluation of supplemental control using speech commands was based on time taken, IED identification, and information retention to test cognitive load. The actual numbers are presented in the Technical Report published by the Human Research and Engineering Directorate (HRED) at Ft. Benning [7].

Time taken to complete tasks

For the first four stations, the total time taken to complete the station was less when using manual control alone than when supplemented with speech commands. For the fifth station – Tunnel reconnaissance – speech supplemented control was faster than manual control. However, none of the discrepancies in time were statistically significant. In fact, for the first two stations – Window reconnaissance and Scan area and track moving target – speech supplemented control resulted in lower times for approaching the wall and employing the arm pose compared to manual control. But, this advantage was overshadowed by the higher time taken to move the camera using speech to detect the IED behind the window or track the target behind the wall. Camera control using speech was not efficient because motion commands were implemented in incremental fashion, where a “Pan left” command would turn the camera by a fixed angle to the left. Therefore, for large motions, the operator had to repeat the speech command several times, which resulted in higher task completion times. Thus, although supplemental control using speech commands had a distinct time-saving advantage, it was lost because of inefficiency in the use of speech commands for controlling camera motions.

IED identification

The Soldiers were required to identify IEDs at Stations 1, 3, 4, and 5. On average, more IEDs were identified when using manual control, 6.91, compared to speech supplemented control, 6.36. However, the difference is not statistically significant. The numbers for the speech supplemented runs were slightly lower mainly because of Station 3, *Area search*. This station required many fine adjustments in direction and speed, and the control implementation of speech commands at the time of the Experiment was not optimal for this task.

Cognitive load

For the cognitive load test at Station 3, the Soldiers were given three cognitive tasks. On average, there were more correct responses with manual control, 1.73, compared to speech supplemented control, 1.36. However, the difference is not statistically significant. Again, the non-optimal driving control implementation using speech commands resulted in some confusion for the Soldiers, and hence, slightly lower scores on the cognitive load test.

5.3 Lessons learned

Think-A-Move gained significant knowledge from the Field Experiment, enabling it to improve the implementation of speech commands for UGV control. The Experiment also helped in understanding the potential concepts of operations (CONOPS) for SPEAR in the future. It became clear that speech commands need to mimic, as closely as possible, the current robot functionality provided by either the Operator Control Unit or the hand controller. Also, there appears to be substantial benefits to be gained by using speech commands simultaneously with the OCU or hand controller to potentially reduce time to mission completion. Some of the immediate improvements applied to SPEAR are listed below.

Prior to the experiments, Think-A-Move and iRobot had implemented a number of speech commands that did not provide the same functionality as the current OCU. As an example, speech commands controlling the pan and tilt on a robot-mounted camera were interval commands, i.e. they would rotate the camera by a pre-determined number of degrees. However, this is different from the continuous motion used to pan or tilt a camera with a joystick. From the Experiment, it became clear that interval commands were not as effective as continuous commands for camera control. Therefore, subsequent to the Experiment, camera pan and tilt control implementation has been changed to a continuous motion; now, speech commands are given to start and stop camera motion. Preliminary observations suggest that this approach improves the benefits provided by speech commands.

On the other hand, a pointed need for interval commands related to turning for small corrections in direction was discovered through the Experiments. During the Experiment, driving commands – forward/reverse motion and left/right turns – were programmed to be continuous. However, the turning commands were not adequate for adjusting the direction of motion by a small amount. Even a small-velocity left or right turn command – *left small* or *right small* – overlapped on a forward/reverse command would result in large corrections by the time the turning motion could be stopped with another speech command. To avoid this issue, the need for a "veer" command was recognized, and has since been implemented successfully. *Veer left* and *veer right* produce small fixed-angle turns.

One of the positive outcomes of the Experiment was the conclusion that there is significant benefit to providing concurrent control of functionalities on the UGV with speech commands and the OCU or hand controller. Therefore, the inclusion of concurrent control in the CONOPS for UGVs should be explored further. For certain UGV maneuvers, a number of the soldiers preferred to use speech commands simultaneously with the OCU or hand controller. These UGV maneuvers included positioning the manipulator arm so that the camera could look through a window or over a wall while at the same time driving the UGV up to the window or wall.

6. CONCLUSIONS AND FUTURE WORK

In this paper, we described the development of a prototype system that allows speech command-based control to supplement hand-controller based tele-operation of a UGV. We also described Field Experiments at Ft. Benning to test the efficacy of the supplemental speech control system in reducing operator task-time and workload.

The prototype involved TAM's speech control system, SPEAR, integrated with iRobot's *Lightweight OCU* to provide simultaneous control of the robot through a hand-controller as well as speech commands. This allowed the operator to drive the robot while simultaneously commanding arm poses, controlling the pan, tilt, and zoom of the camera, moving the flippers, and changing the intensity of the illuminator on the robot. The prototype system was used in a Field Experiment at Ft. Benning. The Experiment had dual focus: 1) Quantitative measurement of the time taken to complete each station and the cognitive load on users; 2) Qualitative evaluation of ease-of-use and ergonomics through soldier-feedback. Think-A-Move was also interested in soldier-feedback on the speech-command vocabulary employed: What spoken commands are intuitive, and how the commands should be executed, e.g., limited-motion vs. unlimited-motion commands.

The initial plan was to collect data on 30 soldiers running the course. However, due to weather conditions (heavy rains and water-logging on the course), we were only able to collect data on 11 soldiers. Data were collected by HRED personnel on the time taken to complete each station and the cognitive load on users, as well as qualitative feedback from each soldier on speech control.

As a result of the Field Experiment conducted at Ft. Benning, TAM gained significant knowledge enabling it to improve the implementation of speech commands for UGV control, as well as regarding SPEAR's potential concepts of operations. It became clear that speech commands need to mimic, as closely as possible, the current robot functionality provided by either the Operator Control Unit or the hand controller. As an example, subsequent to the experiment, camera pan and tilt control was changed to a continuous motion, e.g., speech commands are given to start and stop camera pan and tilt. Preliminary observations suggest that this approach improves the benefits provided by speech commands, and TAM has adjusted its approach to the implementation of speech commands accordingly. Also, there appears to be substantial benefits to be gained by using speech commands simultaneously with the OCU or hand controller to potentially reduce time-to-mission-completion.

6.1 Future Work

In the near future, we would like to implement changes in SPEAR based on the lessons learned from the Experiment. The improvements planned and/or undertaken are described below.

- Eliminating the Push-To-Command (PTC) button: Think-A-Move has already undertaken an upgrade of SPEAR that would eliminate the PTC button. The upgrade has been planned in two stages. The first stage, where SPEAR can be operated under Continuous command mode, has already been developed. This mode allows the operator to give commands at any time without having to press a button. However, the operator is not allowed to engage in conversation while the system is active. This limitation is being addressed in the second stage, where the Command spotting mode will allow conversation while SPEAR is active. With this mode, the system will be equipped to distinguish speech commands from normal conversation.
- Ear seal check: Think-A-Move's patented earpiece uses a foam tip as interface with the user's ear canal. The foam tip is to be squeezed before the earpiece is inserted into the ear. As the foam expands, it blocks off the ear canal and the microphone in the earpiece from the external environment. This prevents ambient noise from leaking into the microphone, allowing user speech to be captured with minimal noise. However, if the earpiece is not inserted properly, it leaks more noise into the ear canal reducing the signal-to-noise ratio, and as a result, recognition accuracy. To avoid this condition, Think-A-Move is investigating a software method to check the seal on the earpiece every time it is inserted into the ear.
- Improving speech command structure: The Field Experiment highlighted the need for a more consistent structure to the command vocabulary and more intuitive spoken commands. This can help minimize the cognitive load associated with remembering the commands. For example, a number of the Soldiers participating in the Experiment recommended that using {Camera up, Camera down, Camera left, and Camera right} set of commands instead of {Tilt up, Tilt down, Pan left, and Pan right}, would make it easier to remember the commands. This was a valuable insight, and as a result Think-A-Move is actively reviewing its entire command structure to improve its consistency.

Think-A-Move is also implementing additional functionalities based on user feedback from the Experiment. One such functionality is the concept of user-defined *macros*. There are several control command sequences that are executed in the same order numerous times during a mission. One example is the lowering of the flippers before actuating a high or low arm pose to provide greater stability to the robot against tipping. This sequence of actions can be combined into a *macro* and activated with a single speech command. We intend to identify more such sequences of actions and code them as macros, and also provide the user with the facility to define macros.

Another functionality being considered is the addition of modes to the speech command structure. For EOD missions, once the robot is in place for dismantling an IED, it would be beneficial to disable speech command based control of all motion functionality to avoid any possibility of unintended motion because of false command recognition. The addition of a Safe Mode could provide this capability by disabling all motion functionality. The operator will be able to activate the Safe Mode using a speech command. The commands that are enabled in the Safe Mode can also be user-defined.

There has also been substantial interest from the user community in exploring the benefits that speech commands would provide to a dismounted Warfighter, who is using a wearable OCU to control the UGV. A speech command system, such as SPEAR, could be used to enable hands-free, heads-up control of a UGV by a dismounted Warfighter, enabling him to maintain both of his hands on his weapon. TAM is very interested in participating in future experiments to better understand the quantifiable benefits derived from hands-free control of a UGV by the dismounted Warfighter.

REFERENCES

[1] Office of the Secretary of Defense, "Unmanned Systems Roadmap 2007-2032," (2007).

[2] Everett, H., Pacis, E., Kogut, G., Farrington, N., Khurana, S., "Towards a Warfighter's Associate: Eliminating the Operator Control Unit," SPIE Proceedings 5609: Mobile Robots XVII, Philadelphia, PA, (2004).

[3] Henry, P., Mermagen, T., Letowski, T., "An Evaluation of a Spoken Language Interface," Technical Report ARL-TR-3477, Army Research Laboratory, MD, (2005).

[4] Nemirovski, G., "System and method for detecting an action of the head and generating an output in response thereto," U.S. Patent No. 6,503,197 (2003).

[5] Nemirovski, G., “Sensor pair for detecting changes within a human ear and producing a signal corresponding to thought, movement, biological function and/or speech,” U.S. Patent No. 6,647,368 (2003).

[6] Nemirovski, G., “Ear microphone apparatus and method,” U.S. Patent No. 6,671,379 (2003).

[7] Pettitt, R., Redden, E., Carstens, C., “Scalability of robotic controllers: Speech-based robotic controller evaluation,” US Army Technical Report. *To appear*

[8] Ahlberg, C., Chamberlin, D., Bushong, J., Oliveira, R., Kolpe, V., “Hearing aid ear piece having disposable compressible polymeric foam sleeve,” U.S. Patent No. 4,880,076 (1986).

[9] Oliveira, R., Chamberlin, D., Babcock, M., “Ear piece having disposable compressible polymeric foam sleeve,” U.S. Patent No. 5,002,151 (1989).

[10] Vaidyanathan, R., Chung, B., Gupta, L., Kook, H., Kota, S., West, J., “Tongue-movement communication and control concept for hands-free human-machine interfaces,” IEEE Transactions on Systems, Man, and Cybernetics, Part A, 37(4), 533—546 (2007).

[11] Westerlund, N., Dahl, M., and Claesson, I., “In-ear microphone techniques for severe noise situations,” Blekinge Institute of Technology, Research Report No. 2005:12.

[12] Rabiner L. and Schafer R., [Digital Processing of Speech Signals], Prentice Hall (1978).

[13] Schaub, A. and Straub, P., “Spectral sharpening for speech enhancement/noise reduction,” Proc. of the International Conference on Acoustics, Speech and Signal Processing, pp. 993-996 (1991).

[14] Berouti, M., Schwartz, R., Makhoul, J., “Enhancement of speech corrupted by acoustic noise,” Proc. of ICASSP, 208-211 (1979).