

Quantifying performance between tele-operated and supervised autonomous fire control

Eugene Pursel, Jaclyn Baron
Naval Surface Warfare Center Dahlgren Division
Dahlgren, VA
eugene.pursel1@navy.mil, jaclyn.baron@navy.mil

Benjamin Wheeler
Naval Surface Warfare Center Dahlgren Division
Dahlgren, VA
benjamin.wheeler@navy.mil

ABSTRACT

Effective weaponization of direct-fire remote and unmanned systems involves a complex interaction between warfighter and weapon system. To better understand the feasibility of such systems, a tele-operated and supervised autonomous (SA) remote weapon station simulator was used to conduct a series of warfighter experiments to validate the SA fire control methodology as a future system concept. Simulation and analysis are used to investigate the capabilities and limitations of SA, and benefits of SA as compared to tele-operation as a method of weapon system control. This paper presents the experiment design, modelling and simulation approach, and results of this study conducted with both civilians, and operational warfighters. Results indicate a normalization, and overall increase in measured operator performance in combat scenarios utilizing an 'ideal' SA combat system as compared to 'ideal' manual operation, however also indicate that peak performance of adept manual operators may exceed that of semi-automated performance.

ABOUT THE AUTHORS

Eugene Pursel is a retired Marine and currently a research scientist with the Testing, Experimentation, Assessment Modeling and Simulation group of the Naval Surface Warfare Center Dahlgren Division's Missile Systems and Integration Branch. He holds a Master of Science in Modeling, Virtual Environments, and Simulation from the Naval Postgraduate School and a Bachelors of Science in Computer Science from Penn State. Ray's current focus is on using virtual environments to prototype developmental and research systems.

Jaclyn Baron received her M.S. and Ph.D. from Clemson University in Industrial Engineering with a concentration in Human Factors. Currently, Dr. Baron is employed by the Naval Surface Warfare Center, Dahlgren Division in Dahlgren, Virginia. Prior to coming to Dahlgren, Dr. Baron worked on a wide variety of projects exploring the impact of human factors on performance in a wide variety of contexts, to include aviation, medical, entertainment and petrochemical. During her time in Dahlgren in the Human Systems Integration department, she has worked to provide support to several Marine Corps and Navy programs, including serving as HSI lead for the Supervised Autonomous Fires Technology (SAF-T) program.

Benjamin Wheeler received a Masters in Computer Science from North Carolina State University, and received a bachelor of arts in Mathematics and Computer Science from Thiel College. He is currently a research scientist at the Naval Surface Warfare Center Dahlgren Division working in the Autonomous and Unmanned Systems Integration Branch. He is also a project officer for the Office of Naval Research, Expeditionary Maneuver Warfare & Combating Terrorism Department (Code 30) Maneuver Thrust in the Autonomy group. Current projects include work in autonomous control of unmanned ground systems, and weaponization of unmanned and robotic platforms.

Quantifying Performance Differences between Tele-operated and Supervised Autonomous Fire Control

Eugene Pursel, Jaclyn Baron
Naval Surface Warfare Center Dahlgren Division
Dahlgren, VA
eugene.pursel1@navy.mil, jaclyn.baron@navy.mil

Benjamin Wheeler
Naval Surface Warfare Center Dahlgren Division
Dahlgren, VA
benjamin.wheeler@navy.mil

INTRODUCTION

This paper details the experiment and results of a study designed to evaluate the proposed system concept of the Supervised Autonomous Fires Technology (SAF-T) program, as an effective methodology of automated remote weapon station (RWS) fire control. Specifically, this study quantifies the difference between tele-operated (TO) and supervised autonomous (SA) fire control, as measured by performance in execution of a virtual combat scenario.

The purpose of the SAF-T project is to weaponize unmanned systems and increase the effectiveness of remote weapons stations by reducing the operator's need for time-critical inputs to and from the weapon system. A virtual prototype of the SAF-T was created within simulation to validate the system concept, provide a framework for a series of warfighter experiments to compare system performance versus traditional fire-control methods, and to establish key performance parameters in support of detailed system design.

In September of 2013, 38 government civilian participants executed a set of simulated combat scenarios using both TO and SA systems. Qualitative and quantitative metrics were captured through measurement of performance using the simulated systems, as well as post-experiment debriefing and a survey. In May 2014, the same experimental procedure was used with a group of 15 Marines for the purpose of verifying the civilian results.

This paper presents the experiment design, modelling and simulation approach, and results of this study conducted with both government civilians and operational warfighters. Results indicate a normalization, and overall increase in measured operator performance in combat scenarios utilizing an 'ideal' SA combat system as compared to 'ideal' TO operation, however also indicate that peak performance of adept manual operators may exceed that of semi-automated performance.

BACKGROUND

This experiment was performed in support of the SAF-T program, utilizing a virtual prototyping environment created for this purpose using Unity3D and the Robotic Operating System (ROS). The background and goals of the SAF-T program and of these tools are briefly described in this section.

Unity3D

Unity3D is a video game engine developed by Unity Technologies. It provides freeware and Pro developer versions, and uses a convenient content-oriented editor for the creation and scripting of 3D games. More information about Unity3D may be found at (Unity Technologies, 2013).

Unity3D was used in this work to create the virtual prototyping environment used for experimentation.

Robotic Operating System

The Robotic Operating System (ROS) is an open source software framework for the development of robotic applications. It is a meta-operating system which runs in Linux, and implements a publisher-subscriber architecture that enables effective collaboration and code-reuse of basic software elements common to robotic applications. The ROS architecture is well described at (Quigley, 2009), and additional information may be found (<http://www.ros.org>).

ROS was used for implementation of the SAF-T virtual prototype, as described in (Wheeler, Fazzari, & Wilson, 2013).

Supervised Autonomous Fires Technology

This work was performed in support of the SAF-T program. This program seeks to overcome current limitations of RWS such as limited situational awareness, and command latency by allocating operator tasks like targeting, tracking, and fire control to the software systems on the platform. Proper allocation of these tasks is expected to allow warfighters to surpass traditional limits of human reaction time, increasing system effectiveness over manually operated systems.

Effective weaponization of direct-fire remote and unmanned systems involves a complex interaction between warfighter and weapon system. Current systems rely on tele-operation for fire control, provide limited situational awareness, and suffer reduced operational effectiveness during remote (wireless) operation due to command latency.

Previous studies have shown that when command latency within the sensor-shoot loop exceeds 100 milliseconds, shooter performance is severely degraded (Beigbeder, 2004). Difficulties in command, control, and communications within the physical and electromagnetic environments associated with modern conflict necessitate a paradigm shift in weapon command and control for unmanned ground and surface vehicles.

SAF-T will enable the required paradigm shift in weapon control through the creation of a semi-autonomous fire control system for RWS as described by the Department of Defense (DoD) Directive 3000.09, "Autonomy in Weapon Systems" (Department of Defense, 2012).

Prior to full-scale prototyping and experimentation, a virtual prototype was created within simulation to validate the system concept, provide a framework for a future series of warfighter experiments to compare system performance versus traditional fire-control methods, and to establish key performance parameters (KPPs) in support of detailed system design.

This work is the result of studies performed in support of this system concept. Beginning in fiscal year 2014, initial experimentation began on creation of detection, decision, and engagement systems to support the overall system concept as implemented in the virtual prototype.

Related Works

The SAF-T program proposes to implement a weapon-control concept that is well established and proven successful in other areas of the military, but applied to small-arms, and within a more complex targeting and engagement environment.

The application of technologies designed in support of the video-game industry to serious applications has grown in popularity as the accessibility of tools and technologies for complex 3D modeling and simulation continues to increase. Within the DoD, many examples of the application of modeling and simulation exist, and are under continued development. Simulation tools designed specifically for DoD applications include Virtual Battle Space by Bohemia Interactive and OneSAF (<http://www.onesaf.net>). Additionally, numerous organizations have leveraged commercial game engines such as Unreal Engine (e.g. Army Games Studio at Redstone Arsenal to create combat system simulators and pre-deployment training).

MATERIALS

Virtual Prototyping

To quantify the performance differences between TO and SA fire RWS fire control system, this work implemented a virtual prototype of each system. These virtual prototypes enabled experimental subjects to exercise each system within a simulated combat environment, while key performance statistics were measured. A detailed description SAF-T simulation environment and architecture may be found in (Wheeler et al., 2013), and is summarized here.

The TO system was created to emulate the performance of the CROWS II RWS by Konigsberg, and was implemented entirely within Unity3D. Similar to a 'first person shooter' video game, the TO system enabled users to tele-operate the RWS using control interfaces representative of the CROWS II, using the gun-camera video feed as seen in Figure 1 (left). A second instance of the Unity3D simulation was created to run as networked instance to provide an experimenter interface, enabling video recording and data collection capability.

The SA system was implemented through a combination of Unity3D and ROS. Unity3D was used to create simulated sensor feeds that fed a prototype system implementation within ROS. ROS was used to implement the SA system control logic, and communicate with a graphical user interface (GUI) implemented within C++. The GUI, as seen in Figure 1 (right), presented the user with a set of potential targets, and enabled the user to engage these targets via selection on the interface.

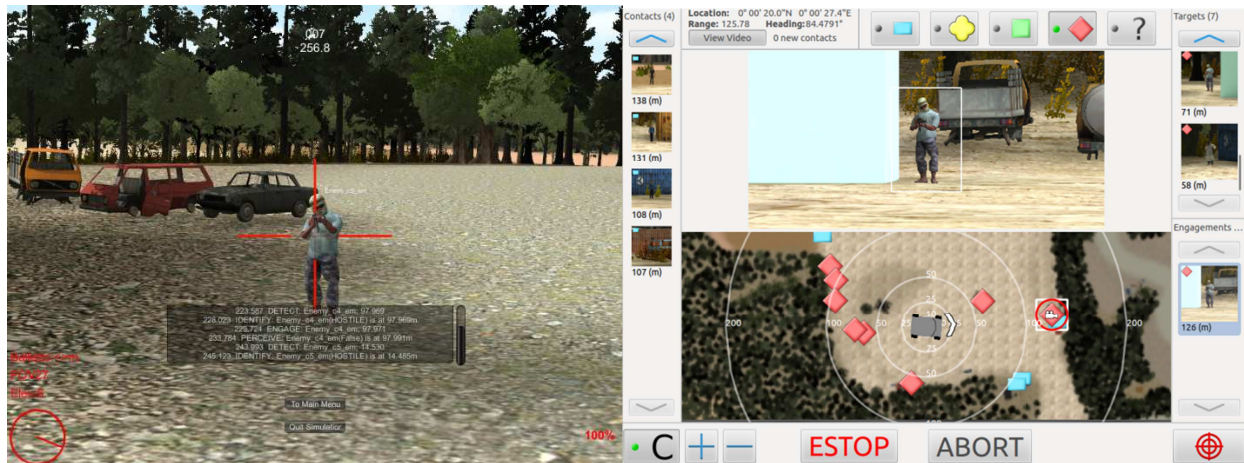


Figure 1: The left image shows the example gun-camera view from the TO remote weapon station. The right image shows the SA remote weapon station GUI interface.

Simulated Combat Scenarios

Scenarios were developed with the goal of simulating measurable responses between use of the TO and SA fire control systems. Target characters were generally placed in plain view and did not react to gunfire to provide a repeatable scenario between participants.

As both the present TO RWS system and the proposed SAF-T system have a wide array of concepts of employment, a great deal of consideration was given to which common facet might be the used for the experiment scenarios. Given limited access to the participants (about one hour) and the expected variety of military experience of within the civilian subject population, a condensed mounted patrol scenario was created. While the scenario is not operationally realistic, it enabled the team to uniformly present a dense number of targets to each participant with controlled complexity without requiring an in-depth knowledge of military procedure and protocol.

In total, three types of scenarios were created, a training, scenario A, and scenario B, each run with TO and SA to create six unique scenario configurations; comprised of two training scenarios, two SAF-T scenarios, and two TO scenarios. Each non-training scenario type (A or B) was similar in the number and intensity of targets, but different in route and location. The two different experiment scenarios were modeled on the McKenna Military Operations in Urban Terrain (MOUT) Facility at Fort Benning, GA. Participants would execute one scenario with the TO system, and one with the SA system, randomly assigned as described in the Methods section. One training scenario was created, modeled on a fictitious Middle Eastern village. The intent in having two different terrains was to limit participants' exposure to the experiment setting. All scenarios were in clear weather and daylight conditions. During the experiment, participants would practice with each system using the training scenario appropriate for each system (TO or SA).

Experiment Equipment

To perform the experiment, virtual prototypes of the TO and SA systems were each run on a dedicated set of computers and associated hardware.

The TO system was composed of two computers and two monitors, in addition to a Common Remotely Operated Weapon Station II (CROWSII) joystick and a surround sound system as illustrated in Figure 2 (left). The computers had 3.5Ghz Intel Core i7 processors, twelve gigabytes of RAM and 4GB nVidia 680GTX video cards running the Microsoft Windows 7 operating system. They were both running a networked simulation built on the Unity3D game engine, and are further described in terms of the TO participant workstation and the observer workstation (used by the experimenter).

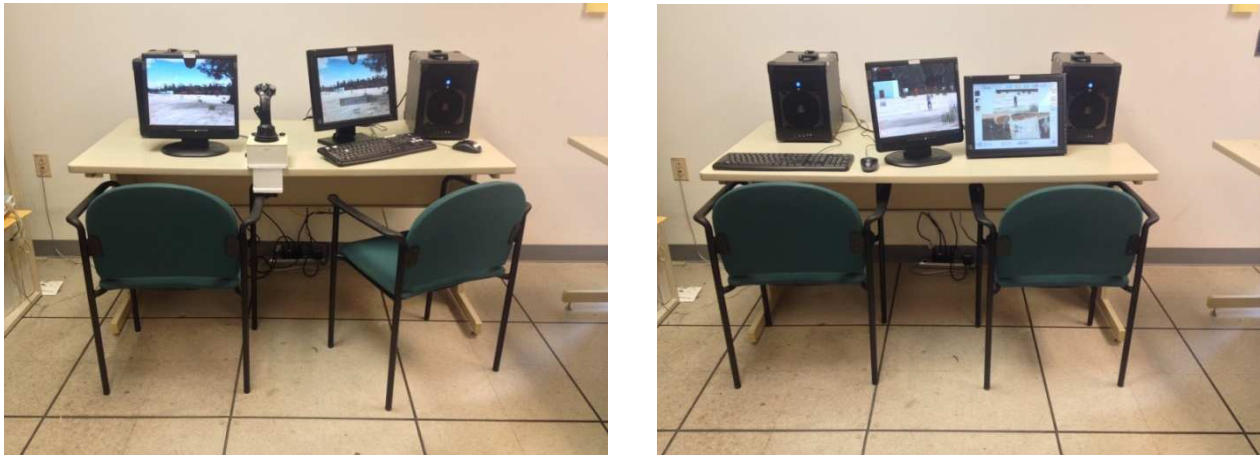


Figure 2: The left picture shows the TO experiment setup, and the right picture shows the SA system setup. Note the TO system uses a joystick, and the subject has one screen as the gun camera. The SA system uses a touchscreen, and the subject has the touchscreen as well as a gun camera view.

The TO participant workstation was a basic CROWSII simulator, and was used by the participant during the experiment. The participant's view was a 15-inch display on which they were presented with a representation of the CROWSII gun camera with heads-up display. Their input was limited to manipulations of the CROWSII joystick and several of its buttons. The TO participant workstation was the primary recorder of collected data for the TO system.

The observer workstation mirrored the participant's view, giving the experimenter the participant's view overlaid with extra context information to aid recording notes and for later reference. The experimenter could also choose, start, pause, and end the simulation scenario from the workstation. During the experiment, the experimenter used this interface to record the participants' declarations of "Contact," "Hostile," or "Friendly," and "Engage" by pressing 4, 5, 8, or 6 on the keypad, respectively. When the participant asked to stop the vehicle, the experiment could halt the vehicle and then, later, continue on the route with the participant's command. These key presses were recorded as well as screen captures of the observers view at 10Hz and reports of when every potential target came into and went out of the system line of sight.

The SAF-T system, illustrated in Figure 2 (right), had the same hardware as the TO system, but one of the systems was using the Ubuntu 12.0.4 LTS operating system to run ROS and the SAF-T prototype system software. ROS was used as the primary communications backbone between the Unity3D simulation and SAF-T system. This system is further described in terms of the SAF-T participant workstation, and the simulation workstation.

The SAF-T participant workstation ran Ubuntu 12.04, and was connected to a 15-inch touchscreen monitor, and was used by the participant during the SA experiment. The SAF-T GUI on the touchscreen was the only means by which

the participant could interact with the system. This workstation recorded all of the button presses made by the participant.

The simulation system's primary purpose was to run the simulation and present the participant with a gun camera view. This was provided by a 15-inch monitor with the same heads-up display as is used on the TO participants workstation. This gun camera view presented the gun camera feed for the participant to use as reference, but the participant had no control over the gun camera except as the gun moved in response to GUI inputs. The experimenter used the keyboard on this system to select, start, pause and end the scenario and this system recorded all shot data and the characters' entry into and exit from the vehicle's line of sight.

METHODS AND PROCEDURES

Two experiments were run; one with government civilians and one with Marines. The same test setup and procedure was used for each subject group. In September of 2013, 38 government civilians participated in the experiment, run at the University of Mary Washington Dahlgren Campus in Dahlgren, VA. A classroom was converted into the experimental space, with two separate stations for the TO and SAF-T conditions. Participants were briefed and debriefed in a separate conference room.

In May 2014, the same experimental procedure was used with a group of 15 Marines for the purpose of verifying the civilian results, run at Camp Pendleton, Oceanside, CA.

A total of 38 participants executed two simulated combat scenarios utilizing the SAF-T virtual prototyping environment, in addition to the training scenarios, each lasting approximately 10 minutes. In one scenario participants were asked to use a simulation of an existing RWS combat system, and in the other asked to use a simulation of an automated RWS combat system representing the future concept under development by the SAF-T program.

Experiment Recruitment and Participants

Government civilian employees currently employed at the Naval Surface Warfare Center Dahlgren Division were recruited for participation through base-wide distribution emails.

Participation for the Government civilians was limited to individuals who had not been involved in the development of the SAF-T system. All participation was voluntary, the informed consent process was followed, and the program paid for the participant's salary in support of the experiment. The average age of participants was 36 years old, with a standard deviation of 10 years, ranging from 23 to 61 years old. The subject pool included 31 males and 7 females.

A total of 15 active duty United State Marines participated in this experiment, and participation was organized by the Office of Naval Research Code 30 Demonstration and Assessment Team, in coordination with the United States Marine Corps. The average age of participants was 26 years old, with a standard deviation of 2.3 years, ranging from 22 to 31 years old. The subject pool included 15 males.

Experiment Protocol

When a participant arrived for the experiment, they were escorted to a conference room where they were briefed according to the experimental script, to include the purpose of the experiment and the informed consent process, their right to withdraw at any time, that their participation is always voluntary, that data is coded with a participant number, and any Personally identifiable information (PII) data is required to be secured under lock and key. The participant then read through the informed consent document and signed the document when they were comfortable with the information, indicating their willingness to participate. The participant was then given the demographic form to complete.

The participant was then directed to the experimental area. An experimental checklist provided a randomized presentation of scenario with conditions. This document was consulted to determine if the experimenter should first

present the SAF-T or TO system to the participant. The participant completed the first training session, and the experimenter answered all questions the participant had prior to beginning the scenario. The participant then completed either the TO or SAF-T first full length condition for either scenario A or B, as determined by the experimental checklist. Once the participant had completed the first scenario, they would then be moved to the experimental setup associated with the other system, and complete the alternate scenario. For example, if the participant had completed scenario A on the TO system, they would then move to scenario B on the SAF-T system.

A debriefing was provided to the participant by the technical team to describe the SAF-T system, the purpose of the study, and the benefit of their participation to the developers of the program. They were thanked for their participation, and escorted to the exit.

Data Collection

Three basic types of data were collected during the experiment scenarios: 1) TO and SAF-T simulation events, 2) SAF-T GUI events, and 3) TO screen shots. Data were collected slightly differently for each condition, and no data logging was used for training events.

Simulation events were any occurrence within the scenario, to include the entering and exiting of contacts within camera view, the firing of weapons, or movement of contacts. During TO scenarios, simulation events were captured and recorded on the participant's workstation and experimenter button presses were captured on the observer's workstation and relayed to the participant's workstation to be recorded. During the SAF-T scenarios, simulation events were captured and recorded on the simulation workstation.

The participants' interactions with the GUI were captured and recorded on the SAF-T workstation. These events included selecting a target from the map or from the contact queue, classifying a target, engaging a target, and canceling an engagement. Zooming in and out of the map was also captured. As these events were in a format different from the simulation events, they were mapped to equivalent simulation events and the two logs were combined into a single event log in post-processing.

On the TO observer's workstation, an image of the observer's screen was captured and stored at up to 10Hz. This rate occasionally decreased slightly when heavy loads were placed on the observer's workstation by the simulation. These screen shots were used in post-processing to help analyze the data by providing a means to clarify confusing or conflicting data.

The Measures of Performance (MOPs) and Measures of Effectiveness (MOEs) for this experiment are defined in Table 1. They were the guiding principles in developing the metrics and, thus, the scenarios' content and the data collection goals for the experiment.

Table 1. Experimental MOPs and MOEs

#	MOP	Definition
1.	Time to Detect	Time (seconds) as measured from moment target enters system perception range until detection.
2.	Time to Identify	Time (seconds) as measured between system detection and identification as target. Identification is defined as the authoritative user decision to classify detection as a valid target.
3.	False Target ID %	Measured as the inverse fraction of valid system target identifications over total target identifications.
4.	Percentage of Hits	The 'hits' of shots /total shots, per discrete engagement, and across all engagements within scenario.
5.	Time to Engage	Time as measured from the decision to begin engagement, until the first hit on the target.
6.	Total Time to Engage	Total time for a discrete engagement measured from time of detect to the completion of engagement, defined as time of last shot fired at the threat.
7.	Finish Time	Time (seconds) as measured from first Detection to last Event

#	MOE	Definition
1.	Engagement Effectiveness	Compound metric measuring effectiveness of shots fired per engagement.
2.	Scenario Effectiveness	Compound metric combining time, Engagement Effectiveness and Mission Accomplishment measured over each scenario.

We described the participants' Scenario Effectiveness operating each system in weighted terms of Time, Accuracy, and Mission Accomplishment.

The ability to capture the data was a fundamental design requirement for our simulation. Rather than try to capture simulation information through some passive external means, data collection components internal to the simulation were created to which other components reported events. The data collection component received those events, made some additional calculations such as miss distances and time durations, and recorded the information in a data file. The SAF-T GUI recorded each participant's button presses in a log.

Data Analysis

Statistical analyses were performed using the JMP 9.0.0 statistical analysis software. For a priori and ad hoc Power, Effect Size, and Sample Size calculations, G*Power 3.1.7 for Windows was used.

An a priori sample size calculation was made to estimate the number of participants required for significant findings. The assumed effect size for this calculation was 0.8 (moderate) and the error probability was set at 0.05. The result estimated that 35 participants would be required. This sample size was achieved with the government civilian test where 38 subjects participated.

A post hoc power calculation was made using the scenario effectiveness statistic, resulting in an effect size of 2.1 (high). Our sample size and this high effect size yielded a Power of 1.0, and we can conclude the civilian experiment had more than enough participants for significant Scenario Effectiveness results.

Another a priori sample size calculation was conducted using the effect size from the civilian experiment. The result was an estimate of 8 participants would be required to have results with 0.95 power assuming the same effect size.

To perform a set of statistical tests, we determined a goodness of fit for the collected data. We first fit the distributions of data to the Normal distribution for ease of analysis. For those that were not a good fit to the Normal distribution, the Gamma distribution was considered. For those that were a poor fit to either distribution, Non-Distribution was assumed. To determine if a distribution of data could be assumed to be Normal, the Shapiro-Wilk W test was used. For this test, the null hypothesis was that the data were from the Normal distribution. To determine if a distribution of data could be assumed to be Gamma, the Cramer-von Mises W test was used. For this test, the null hypothesis was that the data were from the Gamma distribution.

To compare the means of two sets of data, usually to compare a measure between conditions, a Student's t-Test was used for Normally distributed data. For data that were not Normal, the non-parametric Wilcoxon Signed Rank test was used. For these tests, the null hypothesis was that the difference in the means or medians was 0.

To determine if mutual relationships exist between pairs of statistics, correlation tests were conducted. There were two types of data, continuous and ordinal, and different tests were used for each type, as described below.

Percent Hit was the only measure where both TO and SAF-T distribution might be considered Normal. For all others, a Non-Parametric method, Spearman's Rank-Order Correlation (ρ), was used. A significance calculation was also made where the null hypothesis is the correlation equals 0 (no significant correlation).

For ordinal or categorical data fields such as video game experience (High/Medium/Low/None) and TO experience (Yes/No), numerical values had to be assigned to each category before conducting the Spearman's ρ in JMP. For example: High, Medium, Low, and None were mapped to values 3, 2, 1, and 0. The magnitude of the values have no effect on the outcome; only the order was relevant.

RESULTS

The overall results for the experiment are list in Table 2, Table 3, and Table 4. Table 2 shows the civilian experiment results, comparing TO to SA, and finds significant differences in Scenario Effectiveness, Time to Detect, False Target Identification, Percent Hit, and Percent Kill.

Table 2 and Table 3 compare civilian results to warfighter results.

Table 2: Comparison of SA to TO fire control for government civilians

Measure	SA Fire Control (SAF-T)			TO Fire Control			H ₀ rejected
	Mean	Std Dev	N	Mean	Std Dev	N	Y/N
Time to Detect (s)	4.98	1.79	36	17.02	6.04	38	Yes
Time To Identify (s)	1.62	0.47	36	1.45	1.06	38	No
False Target ID %	0.03%		36	2.87%		38	Yes
Percent Hit	78.0%	9.49%	36	53.10%	20.20%	38	Yes
Time to Engage (s)	1.87	0.38	36	2.47	1.15	38	No
Total Time to Engage (s)	3.7568	0.81	36	3.9241	1.59	38	No
Finish Time (s)	415.60		36	466.43	97.72	38	Yes
Engagement Effectiveness	97.21%	3.69%	36	80.64%	8.22%	38	Yes
Scenario Effectiveness	92.34%	4.30%	36	81.37%	7.27%	38	Yes

Table 3: Comparison of Civilian to Marine Participants for SA Fire Control

Measure	Civilian			Marine			t-Test (df=51)			H ₀ rejected
	Mean	Std Dev	N	Mean	Std Dev	N	t	Std err Diff	p-Value	Y/N
Time to Detect (s)	4.98	1.79	36	3.40	1.43	15	1.99	0.52	0.0637	No
Time To Identify (s)	1.62	0.47	36	3.62	2.75	15	4.28	0.47	<0.0001	Yes
False Target ID %	0.30%		36	1.20%		15	338.59	0.003	<0.0001	Yes
Percent Hit	78.13%	9.49%	36	86.53%	6.38%	15	3.14	0.03	0.0029	Yes
Time to Engage (s)	1.87	0.38	36	1.65	0.22	15	2.06	0.11	0.044	Yes
Total Time to Engage (s)	3.76	0.81	36	2.26	0.43	15	6.78	0.22	<0.0001	Yes
Finish Time (s)	415.60		36	410.06	4.64	15	1.18	4.68	0.2419	No
Engagement Effectiveness	97.21%	3.69%	36	96.31%	0.02	15	0.89	0.01	0.3772	No
Scenario Effectiveness	92.34%	4.30%	36	95.37%		15	1.10	0.01	0.2749	No

Table 4: Comparison of Civilian to Marine Participants for SA Fire Control

Measure	Civilian			Marine			t-Test (df=53)			H ₀ rejected
	Mean	Std Dev	N	Mean	Std Dev	N	t	Std err Diff	p-Value	Y/N
Time to Detect (s)	17.02	6.04	38	12.55	3.40	15	2.69	1.66	0.0096	Yes
Time To Identify (s)	1.45	1.06	38	1.09	1.14	15	1.09	0.33	0.2823	No
False Target ID %	2.87		38	5.80%	3.59%	15	3.31	0.01	0.0017	Yes
Percent Hit	53.10%	19.55%	38	63.55%	12.87%	15	1.90	0.06	0.0621	No
Time to Engage (s)	2.47	1.15	38	2.27	1.38	15	0.53	0.37	0.5989	No
Total Time to Engage (s)	3.92	1.59	38	1.96	0.63	15	4.63	0.42	<0.0001	Yes
Finish Time (s)	466.43	97.72	38	465.58	93.43	15	0.03	29.44	0.9772	No
Engagement Effectiveness	80.64%	8.22%	38	86.56%	5.11%	15	2.59	0.02	0.125	Yes
Scenario Effectiveness	81.37%	7.27%	38	0.7762	4.42%	15	12.27	0.01	<.0001	Yes

DISCUSSION AND FUTURE WORK

Foremost in considering our results, a few constraints must be kept in mind. Mainly, our experiment simulation scenarios were two of many possible scenarios and are not indicative of every, or even many, of the proposed applications of the Supervised Autonomous Fires – Technology or existing Tele-Operated systems. We do not assume that the results from this experiment directly represent those other scenarios or applications. Secondly, our population sample may not accurately represent the population of military TO systems operators. Another experiment run with those operators is necessary to remove this last caveat.

Overall, our participants performed the experiment tasks as measured by our MOPs and MOEs better with the theoretical SAF-T system than with the TO system. There are several caveats to that statement as raised above and in the Experiment Findings paragraphs. Still, these initial findings are very encouraging regarding the continued development of the SAF-T system.

An interesting conclusion can be made given the finding that there is no correlation between TO Scenario Effectiveness or prior RWS experience to SAF-T Scenario Effectiveness. This implies that in our scenarios, the skills required for effective TO performance are not necessarily required for SAF-T performance. In general, it seems that the task of identifying and persecuting targets at close range in open terrain is significantly easier with the SAF-T system for almost everyone, regardless of experience (real or video game) or age.

There were a few complications in the manner by which we captured data that may have had an impact on our statistics. Specifically, the times to engage, detect and identify seem to have been affected by the way in which we captured those times. The button-press method of recording utterances was an improvement over getting no data. However, the way this method's weaknesses were exacerbated by some participants' styles of employment was unexpected.

The experiment with Marine participants was run as a means to validate the civilian results, given it was impractical to recruit enough Marine participants to yield statistically significant results. In spite of this, evidence shows the Marine participants did not perform drastically differently from the civilian participants. This result implies that the use of civilian participants is appropriate for the evaluation of some aspects of these systems in subsequent experimentation. More analysis and additional data collection could help clarify the differences between the Marines' and civilians' performance and the benefits of using each group.

ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research Code 30, Expeditionary Warfare and Combating Terrorism Department, within the Fires Technology Investment Area, and by the Marine Corps Warfighting Laboratory.

REFERENCES

- Department of Defense Directive, DoDD 3000.09. Autonomy in Weapons Systems, dated November 21, 2012.
- Duchowski, A. (2007). *Eye Tracking Methodology: Theory and Practice, 2nd Edition*. New York, NY: Springer.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.
- M. Quigley, B. G. (2009). "ROS: an open-source Robotic Operating System". *Proceedings Open-Source Software Workshop of the International Conference on Robotics and Automation (ICRA)*.
- Unity Technologies. (2013, March 15). *Unity3D*. Retrieved from Unity3D: <http://www.unity3d.com>
- Wheeler B., Fazzari, K., Wilson, G. (2013), "Playing games with the development of future combat systems". MODSIM World 2013.