DRDC No. CR-2012-083

HUMAN – ROBOT INTERACTION LITERATURE REVIEW

by:

Jordan Bray-Miners, Chris Ste-Croix, and Andrew Morton

Human*systems*® Incorporated
111 Farquhar Street
Guelph, ON N1H 3N4

Project Manager:

Andrew Morton
(519) 836-5911

PWGSC Contract No. W7711-088136/001/TOR
Task Authorization No. 4500892320; Call-up No. 8136-10

On Behalf of

Department of National Defence

As represented by

DRDC Toronto
1133 Sheppard Avenue West
Toronto, ON, M3M 3B9

Project Scientific Authority:

Giovanni Fusina & Ming Hou
(416) 635-2000

March 2012

Author

Andrew Morton
Human*systems*® Incorporated

Approved by

Ming Hou
Scientific Authority

Approved for release by

J. V. Baranski
Chair, Knowledge and Information Management Committee

Abstract

In the field of unmanned vehicle (UV) systems, researchers have been striving to inverse the human-robot ratio such that one operator can control multiple robots. This goal has not yet been accomplished for military applications, despite ongoing research. Research suggests that the human-robot interaction (HRI) that takes place while an operator is in control of one or more UVs needs to be improved before the ratio can be inversed. This literature review included 53 references to provide an overview of current HRI research dealing with the operation of UVs and to identify the key human factors (HF) issues when conducting research within this area.

The literature identified three key factors in HRI research related to operating UVs for military applications: operator capacity (that is, the number and type of UVs that a human operator controls or supervises), automation, and interface design. Within the literature HRI is most often measured through the three common metrics of situation awareness (SA), workload, and task performance. In general, research shows that increasing operator capacity increases workload and decreases SA, while the corresponding impact on performance has been shown to be inconsistent. Automation and multimodal interfaces have been shown to alleviate some of the increased workload and decreased SA as operator capacity is increased, however, there is a complex interaction between the three variables. The literature suggests that adaptive automation and adaptive interfaces are promising solutions to accommodate for this complex interaction, but further research and empirical studies are necessary before they can be implemented into military operations. Three additional characteristics of military applications also need to be investigated further: one operator in control of mixed UV platforms (i.e. UAVs and UGVs), operators controlling UVs in a mobile environment, and team coordination between multiple operators each in control of multiple UVs.

To help further research in this area, the new Human-Robot Interaction laboratory being built at DRDC—Toronto should consider investigating HF issues in the design of a multimodal adaptive interface for mixed UV military operations. In particular, due to gaps in the literature and the need for more detailed research in certain areas, studies should look at the interactions between operator capacity, adaptive automation, automation reliability, adaptive interfaces, mobile environments, and team coordination.

Résumé

Dans le domaine des systèmes de véhicules sans pilote (UV), les chercheurs travaillent sans cesse à inverser le ratio humain-robot, de sorte qu'un seul opérateur puisse commander plusieurs robots. Cet objectif n'est pas encore atteint en ce qui concerne les systèmes militaires, et ce, en dépit de recherches continues. Selon les recherches, il faut améliorer l'interaction humain-robot (IHR) qui a lieu quand l'opérateur est aux commandes d'un ou de plusieurs UV avant de pouvoir inverser le ratio. Cette analyse documentaire porte sur 53 ouvrages et vise à donner un aperçu des recherches sur l'IHR concernant l'utilisation des UV qui sont actuellement en cours et à cerner les principaux enjeux liés aux facteurs humains (FH).

Dans les ouvrages étudiés, on a constaté que les recherches sur l'IHR portant sur le fonctionnement des UV militaires font ressortir trois principaux facteurs : la capacité de l'opérateur (c'est-à-dire, le nombre et le type de véhicules qu'un opérateur humain contrôle ou supervise), l'automatisation et la conception de l'interface. Dans les ouvrages analysés, l'IHR est la plupart du temps mesurée au moyen des trois paramètres communs que sont la connaissance de la situation (CS), la charge de travail et le rendement à l'exécution des tâches. En général, les recherches démontrent que l'amélioration de la capacité de l'opérateur fait augmenter la charge de travail et diminuer la CS. Cependant, l'incidence sur le rendement s'est avérée inégale. Il a été démontré que l'automatisation et les interfaces multimodales aident à atténuer quelque peu l'augmentation de la charge de travail et la diminution de la CS quand on accroît la capacité de l'opérateur. Toutefois, il existe une interaction complexe entre les trois variables. Les documents indiquent que l'automatisation adaptive et les interfaces adaptatives constituent des solutions prometteuses qui permettraient de faciliter cette interaction complexe, mais il faudra poursuivre les recherches et effectuer d'autres études empiriques avant de pouvoir les intégrer aux opérations militaires. On doit également poursuivre l'étude de trois autres caractéristiques des applications militaires : un seul opérateur aux commandes de diverses plateformes UV (p. ex. UAV et UGV), des opérateurs commandant des UV dans un environnement mobile et la coordination d'équipe en présence de multiples opérateurs commandant chacun plusieurs UV.

Pour contribuer aux recherches effectuées dans ce domaine, le nouveau laboratoire de l'interaction humain-robot, qui est en cours de construction à RDDC—Toronto devrait envisager d'étudier les questions liées aux FH dans la conception d'une interface adaptative multimodale pour les opérations militaires utilisant divers types d'UV. En particulier, comme il existe des lacunes dans la documentation et comme des recherches plus approfondies s'imposent dans certains domaines, les études devraient porter sur les interactions entre la capacité de l'opérateur, l'automatisation adaptative, la fiabilité de l'automatisation, les interfaces adaptatives, les environnements mobiles et la coordination d'équipe.

Executive Summary

HUMAN – ROBOT INTERACTION LITERATURE REVIEW

Jordan Bray-Miners, Chris Ste-Croix, and Andrew Morton, Human*systems*® Incorporated; DRDC Toronto CR2012-083; Defence R&S Canada – Toronto; March 2012.

Mandate This literature review was conducted for DRDC—Toronto as part of the Human-Robot Interaction Laboratory – Phase 1 project. The Literature search and review was conducted by Human*systems*® Incorporated from January 2012 to March 2012.

Background Unmanned vehicles (UVs) are increasingly being used for military operations because they have the potential for being a force multiplier, and for preventing exposure of soldiers to many operational dangers. UV applications have not yet been optimised, and in most situations multiple personnel are required to operate a single UV, due to the complexity of the monitoring and control requirements. Researchers have been striving to inverse the ratio such that a single operator can control multiple UVs during military operations. To accomplish this goal, improvements need to be made in the human-robot interaction (HRI) that takes place while an operator is in control of a UV.

Purpose This literature review aims to provide an overview of current HRI research in reference to UV operation. The goal was to identify the key human factors (HF) issues that researchers need to consider when studying this area, and to recommend new research avenues for The Human-Robot Interaction Laboratory.

Methods A set of HRI-related keywords was developed, and used to search relevant databases. Articles were systematically evaluated based on their quality and breadth of coverage, as well as their relevance to the purpose of the literature review. Fifty-three book chapters, journal articles and technical reports were selected to be included in the review.

Results This review shows that operator performance while in control of UVs mainly depends on three factors: operator capacity, automation implementation, and interface design. The combination of the three factors that optimise performance depends on the type of UV(s) and the mission task. During the studies reviewed HRI is most commonly measure by situation awareness (SA), workload and specific task performance metrics. In general, research shows that increasing operator capacity increases workload and decreases SA, while the corresponding impact on performance has been shown to be inconsistent. Automation and multimodal interfaces have been shown to alleviate some of the increased workload and decreased SA as operator capacity is increased, however, there is a complex interaction between the three variables. To accommodate for the complex interaction between variables, the review suggests that automation and interface design should dynamically adapt to the operator and mission status during UV operations. Based on the results of the literature review three potential areas research for military applications were found: one operator in control of mixed UV platforms (i.e. UAVs and UGVs), operators controlling UVs in a mobile environment, and team coordination between multiple operators each in control of multiple UVs.

Conclusions HRI has not been sufficiently optimised to implement UV teams with a single operator in control of multiple UVs during military applications. Consequently, the research gaps in the interaction between, operator capacity, automation, and interface design provide an opportunity for

further research in the following areas: team coordination, operation of non-simulated UVs, and mixed UAV and UGV systems.

Recommendations **T**HRIL should focus on studies that look at operating non-simulated UVs during simulated military missions, within a team scenario. These studies should ideally help develop a multimodal adaptive interface and implement an adaptive automation scheme that could be used to operate multiple, mixed platform UVs.

Sommaire

HUMAN – ROBOT INTERACTION LITERATURE REVIEW

Jordan Bray-Miners, Chris Ste-Croix, and Andrew Morton, Humansystems® Incorporated; DRDC Toronto CR2012-083; Defence R&S Canada – Toronto; March 2012.

Mandat : Cette analyse documentaire a été effectuée pour le compte de RDDC-Toronto dans le cadre du projet du Laboratoire sur l'interaction humain-robot – Phase 1. La recherche et l'analyse des documents ont été effectuées par la société Humansystems® Incorporated, de janvier à mars 2012.

Contexte : Les véhicules sans pilote (UV) sont de plus en plus utilisés dans le cadre des opérations militaires parce qu'ils peuvent constituer un multiplicateur de force, et ils évitent aux militaires de s'exposer à de nombreux dangers opérationnels. Les applications UV n'ont pas encore été optimisées et, dans la plupart des situations, il faut plusieurs personnes pour faire fonctionner un seul UV, en raison de la complexité de la surveillance et des contrôles à effectuer. Les chercheurs s'efforcent sans cesse d'inverser le ratio humain-robot de sorte qu'un seul opérateur puisse commander plusieurs robots dans le cadre d'opérations militaires. Pour atteindre cet objectif, il faut améliorer l'interaction humain-robot (IHR) qui se produit lorsqu'un opérateur commande un UV.

But : Cette analyse documentaire vise à présenter un aperçu des recherches en cours portant sur l'IHR liée à l'utilisation des UV. Il s'agissait de cerner les principaux enjeux relatifs aux facteurs humains (FH) que les chercheurs doivent examiner lorsqu'ils étudient ce domaine, et de recommander de nouvelles pistes de recherche pour le Laboratoire sur l'interaction humain-robot.

Méthodes : On a élaboré un ensemble de mots-clés relatifs à l'IHR et utilisé ceux-ci pour faire des recherches dans les bases de données pertinentes. On a évalué systématiquement des articles en fonction de leur qualité et de leur portée, ainsi que de leur pertinence par rapport à l'objectif visé par l'analyse documentaire. Cinquante-trois chapitres de livres, articles de revues et rapports techniques ont été choisis pour la tenue de cette analyse.

Résultats : L'analyse démontre que le rendement de l'opérateur qui commande un UV dépend principalement de trois facteurs : la capacité de l'opérateur, la mise en œuvre de l'automatisation, et la conception de l'interface. La combinaison des trois facteurs assurant l'optimisation du rendement est fonction du type d'UV(s) et de la tâche à accomplir. Dans les études analysées, l'IHR est la plupart du temps mesurée en fonction des paramètres de la connaissance de la situation (CS), de la charge de travail et du rendement à l'exécution d'une tâche précise. En général, les recherches indiquent que l'augmentation de la capacité de l'opérateur accroît la charge de travail et diminue la CS, alors que l'incidence sur le rendement s'avère inégale. Il a été démontré que l'automatisation et les interfaces multimodales atténuent l'augmentation de la charge de travail et la perte de CS quand on accroît la capacité de l'opérateur. Cependant, il existe une interaction complexe entre les trois variables. Pour faciliter l'interaction complexe entre les variables, l'analyse indique que l'automatisation et la conception de l'interface devraient s'adapter de façon dynamique à la situation de l'opérateur et de la mission pendant l'utilisation des UV. D'après les résultats de l'analyse documentaire, il existe trois domaines de recherche potentiels relatifs aux applications militaires : un opérateur commandant diverses plateformes UV (p. ex. UAV et UGV), des opérateurs commandant

des UV dans un environnement mobile et la coordination d'équipe entre de multiples opérateurs commandant chacun plusieurs UV.

Conclusions : L'IHR n'est pas suffisamment optimisée pour permettre la création d'équipes d'UV, avec un seul opérateur aux commandes de plusieurs UV dans le cadre d'opérations militaires. Par conséquent, en raison des lacunes constatées dans les recherches sur l'interaction entre la capacité de l'opérateur, l'automatisation, et la conception de l'interface, on aura l'occasion de poursuivre les recherches dans les domaines suivants : coordination d'équipe, utilisation d'UV non simulés et d'un assortiment de systèmes d'UAV et d'UGV.

Recommandations : Le LIHR devrait se concentrer sur les études qui portent sur l'utilisation d'UV non simulés pendant les missions militaires simulées, dans le cadre d'un scénario d'équipe. Ces études devraient idéalement aider à l'élaboration d'une interface adaptative multimodale et à la mise en œuvre d'un schème d'automatisation adaptative qui pourrait permettre de commander plusieurs plateformes UV de types différents.

Table of Contents

ABSTRACT .... **I**

RÉSUME .... **II**

EXECUTIVE SUMMARY .... **II**

SOMMAIRE .... **V**

TABLE OF CONTENTS .... **VIII**

LIST OF TABLES .... **X**

1. PROJECT OVERVIEW **1**

1.1 BACKGROUND 1
1.1.1 Unmanned Aerial Vehicles *1*
1.1.2 Unmanned Ground Vehicles *2*
1.1.3 Human Factors Issues of UVs *3*
1.2 OBJECTIVE AND SCOPE 4
1.3 WORK ITEMS 4

2. METHODS **5**

2.1 KEYWORDS 5
2.2 DATABASES 5
2.3 SEARCH STRATEGY 6
2.4 SELECTION CRITERIA 7
2.5 SEARCH RESULTS 8
2.6 STRUCTURE OF THE REPORT 8
2.7 LIMITATIONS 8

3. RESULTS **9**

3.1 SITUATION AWARENESS 9
3.2 WORKLOAD 10
3.3 AUTOMATION 11
3.3.1 Level of Automation *13*
3.3.2 Imperfect Automation *15*
3.3.3 Adaptive Automation *16*
3.3.4 Adaptive Automation for Mixed UV Operations *18*
3.3.5 Level of Automation and Operator Capacity *19*
3.4 OPERATOR CAPACITY 23
3.4.1 Operator Capacity and Task Performance *23*
3.4.2 Operator Capacity and Automation Reliability *25*
3.4.3 Operator Capacity and Interface Design *26*
3.4.4 Operator Capacity Model Development *28*
3.5 INTERFACE DESIGN 29
3.5.1 Visual Display Aids and Control *30*
3.5.2 Audio Aids and Control *36*
3.5.3 Tactile and Haptic Display and Control *39*
3.5.4 Multimodal Display and Control *40*
3.5.5 LOA, Operator Capacity, and Information Perception *43*
3.5.6 Adaptive Interfaces *45*
3.5.7 Interface Design in Mobile Environments *46*

4. CONCLUSION AND RECOMMENDATIONS **49**
4.1 OVERVIEW OF CONCLUSIONS 49
4.1.1 Operator Capacity *49*
4.1.2 Automation *49*
4.1.3 Interface Design *49*
4.1.4 Need for Further Research *49*
4.2 AVENUES OF NEW RESEARCH 50
4.2.1 Adaptive Interfaces *52*
4.2.2 Mobile Environments *52*
4.2.3 Interaction between Automation Reliability and Interface Design *52*
4.2.4 Additional Avenues of New Research *52*
4.3 RECOMMENDED RESEARCH AREAS AND METHODS 53
4.3.1 Overview of Lab Equipment and Requirements *54*
4.3.2 Measurement of Situation Awareness *54*
4.3.3 Measurement of Workload *54*
4.3.4 Measurement of Task Performance *54*
5. REFERENCES **57**
ANNEX A **A-1**

List of Tables

TABLE 1: KEYWORDS 5
TABLE 2: PRIMARY DATABASES FOR SCIENTIFIC/ACADEMIC SEARCH 6
TABLE 3: ENDSLEY AND KABER (1999) LOA TAXONOMY 13
TABLE 4 - OVERVIEW OF RESEARCH REVIEWED 51
TABLE 5 - UNIQUE CHARACTERISTICS OF THRIL 53
TABLE 6 - SUMMARY OF EMPIRICAL STUDIES A-1

This page intentionally left blank.

1. Project Overview

1.1 Background

Human-Robot Interaction (HRI) is a rapidly developing field with very strong applications in the military environments. Advances in robot technologies will soon allow robots to perform advanced reconnaissance tasks, logistics supply, and battlefield casualty evacuations, among others. These advances, however, are dependent upon the success of HRI and human-systems integration research which will help to understand the capabilities, uses and misuses of robot technologies.

One of the robot technologies that has been used extensively in recent military operations is that of unmanned vehicles (UVs). With the potential of being a force multiplier and preventing soldiers from certain dangers, UVs have the potential to save lives and money for armed forces across the globe. Thus, the focus of this literature review is on UVs, specifically unmanned aerial vehicles (UAVs) and unmanned ground vehicles (UGVs), being the robot in the HRI.

1.1.1 Unmanned Aerial Vehicles

The United States (US) Department of Defense (DOD) defines a UAV as:

"A *powerful aerial vehicle that does not carry a human operator uses aerodynamic forces to provide lift, can fly autonomously or be piloted remotely, can be expendable or recoverable, and can carry a lethal or non-lethal payload. Ballistic or semi-ballistic vehicles, cruise missiles and artillery projectiles are not considered Unmanned Aerial Vehicles."* (Wheatley, 2004, pp.3)

The development of UAVs primarily started after World War II (WWII) but the concept was started long before that with the US Navy developing a seaplane that could operate without a pilot onboard (Global Security, 2012). This idea was developed throughout the 1920s and 1930s and the US Navy used a small plywood UAV in the Pacific during WWII to attack heavily defended targets. The US Army expanded this idea by flying explosive laden B-17 aircraft into targets after the pilot bailed out as the plane reached a specific altitude. The B-17 was then piloted via a radio control from another B-17 (Global Security). The first reported use of a UAV for reconnaissance was in the 1950's when the US Army placed cameras, which were later replaced with television systems, onto the UAV.

UAVs have since been used in most of the major wars since WWII with the US Air Force having the ability to launch four drones, from an already airborne plane, like missiles on a pre-programmed flight pattern. From the mid-1960s to the end of the Vietnam War more than 3,000 missions that were conducted over North Vietnam and China (Global Security, 2012). UAVs were continued to be developed in operations such as Desert Storm, where the Pioneer UAV system provided intelligence and fire support to commanders, to Operation Iraqi Freedom where each brigade was assigned 4 teams of 22 soldiers to operate 450 UAV that were reported to be in Iraq (Global Security, 2012).

Historically, UAVs were classified into 3 categories: Close Range, Short Range, and Endurance; however, with technology developments that have increased UAV capabilities and the proliferation of UAVs for a broad range of applications, UAVS are classified into 5 categories:

- High Altitude Long Endurance (HALE) ex. Global Hawk;
- Medium Altitude Long Endurance (MALE) ex. Predator;
- Tactical UAVs (TUAV) ex. Hunter;
- Mini-UAVs; and

- Micro-UAVs.

Canada's first involvement with UAV development began in 1959 with the Canadair project which was a missile with a camera attached (CL-89) to it such that it could be used as a reconnaissance aircraft (Wheatley, 2004). Canada followed the lead of the US and other nations by converting existing drones into reconnaissance aircrafts. With the help of the British, West Germany, and US governments the CL-89 took flight in 1964 on a pre-programmed course while taking photographs, or infrared images at certain instances along the flight route (Wheatley, 2004). The CL-89 was used later by West Germany, France, Italy, and as late as the British during the Persian Gulf War. Even though the CL-89 proved to be useful the Canadian government chose not to adopt it. The CL-89 project was followed by the development of the CL-289 and even though this once again proved to be useful for the French and German militaries, where it accounted for 37 percent of all UAV missions flown during Operation Allied Force (NATO bombing of Yugoslavia in 1999), the Canadian government chose not to adopt it. While foreign involvement in UAV development and use has proved to be successful, the Canadian Forces has found a new interest in acquiring UAVs as the military transforms into joint, network-enabled fighting force (Wheatley, 2004).

Since 2000, Canada has participated in at least 5 UAV test programs and a number of evaluations including several studies within the Soldier Information Requirements Technology Demonstration (SIREQ TD). In partnership with the US, the Global Hawk HALE program, which entailed several overflights Exercise Robust Ram tested UAVs at the brigade level in terms of connectivity of a family of UAVs as well as their command and control capabilitites. During Exercise Robust Ram additional UAVs, such as, a tactical UAV were also tested and paved the way for the TUAV to be used during the G-8 Summit in 2002 where it provided real-time information for security forces. The fourth program was a Pacific surveillance and reconnaissance experiment using a MALE UAV, and lastly an Atlantic surveillance and reconnaissance experiment adding the use of mini-UAV (Wheatley, 2004).

During the Afghanistan mission Canadian Sperwer UAVs have conducted numerous intelligence gathering missions that were found to be an intrinsic part of the mission while fighting in urban areas such as Kabul. Currently, Canada is using two unmanned systems that carry out intelligence, surveillance, target acquisition, and reconnaissance missions: the Sperwer and the Skylark with plans on procuring and leasing a family of UAVs in the coming years (Carryer, 2008).

1.1.2 Unmanned Ground Vehicles

Similar to the research and development of UAVs, UGVs have undergone considerable development over the last few decades. An UGV is a powered, mobile, ground conveyance that as its name asserts does not carry an operator. The first major UGV development effort in the US was the Shakey which was developed in the late 1960s to serve as a testbed for Defense Advanced Research Projects Agency (DARPA) founded artificial intelligence. The SHAKEY system was a wheeled platform with a steerable TV camera that could accept English sentence commands. The program was determined to unsuccessful due to the lack of autonomy in the system (Gage, 1995).

The SHAKEY program was reinvented in the early 1980s as the Autonomous Land Vehicle (ALV) program. The ALV was an eight-wheeled all-terrain vehicle that was only able to achieve speeds of about 3 km/hr in the beginning and was further advanced to achieve up to 70 km/hr on the highway. This program led to the initiation of the Reconnaissance, Surveillance, and Target Acquisition (RSTA) application for UGVs. This capability was appealing to battlefield commanders as they had the ability to have a direct sensing capability without endangering human personnel. Two RSTA projects developed the Ground Surveillance Robot and the Advanced Teleoperator Technology TeleOperator Dune Buggy.

The success of these programs led to the development of a United States Marine Corps (USMC) managed program called the Ground/Air TeleRobotics Systems (GATERS). The goal of the GATERS program was to develop a TeleOperated Vehicle to support the test and evaluation of UGV product concepts by prospective military users. The TeleOperated Vehicle was a Humvee that had up to three control stations housed in a shelter mounted on the back of the Humvee. The TeleOperated Vehicle was equipped with a RSTA package, as well as, 50-caliber machine gun that could be manually controlled with a joystick in response to feedback provided by the on-board cameras.

In 1990, with many concurrent UGV developments underway in the US, the DoD consolidated all efforts into the Unmanned Ground Vehicles Joint Program Office (UGV JPO) as the central agency responsible for the development and fielding of DoD UGV systems. One of the programs that was part of the initial fielding of UGVs in military operations was the Vehicle Teleoperation Capability (VTC). The VTC was incorporated into M60 tank chassis (Panther) and used for route proofing in Bosnia. A smaller version of the Panther, the miniflail, was also deployed to Bosnia to provide a smaller version for proofing route of anti-personnel mines. Along with route proofing, the miniflail was also used for the evacuation of wounded soldiers from minefields. Along with route proofing and casualty evacuations, UGVs began neutralizing ordnances remotely in the late 1990s (Gage, 1995).

Current programs are focussing on developing lightweight, man portable mobile robots for operations in urban areas, such as tunnel, sewer, and bunker reconnaissance missions. Not only are current programs investigating smaller more compact robots but they are also developing what is known as marsupial robots. A marsupial robot is a larger UGV that has the ability to carry one more smaller UGVs either attached to or inside. Marsupial robot efforts also involve the integration and launch of both a UAV and UGV from a single robot. This concept of an operator not only controlling multiple UVs but also a mix of UGVs and UAVs is one focus area of this literature review.

1.1.3 Human Factors Issues of UVs

In 2000, the United States of America (USA) Congress mandated that by 2010 a third of all Army aircraft be unmanned , and by 2015, that a third of all ground combat vehicles be unmanned (Jentsch, Evans, & Ososki, 2010). As of 2010, the US DoD had an inventory of 10,767 manned aircraft and 7,494 unmanned aircraft clearly indicating a transformation program to have many more UVs in military operations. This poses a large personnel requirement on armed forces to have qualified personnel control these UVs.

In most situations a UAV is operated by more than one personnel with a pilot and a person responsible for the payload. In larger UVs that have a multitude of sensors even more personnel are required for each UV. This places a large personnel burden on the military. Therefore, the goal of many nations is to invert the operator / robot ratio from having multiple operators control a single robot to having a single operator control multiple robots. Having a single operator control not only a single robot but multiple robots poses a number of human factors issues.

In the majority of cases the operators of UVs are out of sight of the UV creating a need to provide the operator with enough situational awareness of the environment that the UV is operating in to make informed decisions. The operator maintains their situational awareness through some sort of an interface with the robot. Different types and styles of interfaces directly impact the level of situational awareness the operator has, as well as, the workload that is assumed by the operator. The reliability / automation of the system also have a significant impact on the operators' workload and performance, especially if the operator controls more than one UV. This is just a small portion of human factors issues that impact HRI. Theses human factors issues (situational awareness, workload, information perception, and automation) will be discussed in greater detail within this literature review.

1.2 Objective and Scope

The Human-Systems Integration Section at Defence Research and Development Canada (DRDC) Toronto plans to build The Human-Robot Interaction Laboratory (THRIL) capable of studying human-robot interactions and the integration of humans with robot systems. The vision for this laboratory is to establish a facility that will allow human factors experimentation relating to the various facets of human operators controlling robots, with particular focus on teams of robots

The literature review will review current HRI research focusing on several key areas of human factors in preparation for the establishment of THRIL. The key areas of human factors include situational awareness (SA), workload, information perception, automation, and operator capacity. Within each of these human factors areas topics investigated include, advanced interface design for multiple robots and HRI experiments involving single robot operation, multiple robot operations, and mixed ground /aerial operations. Based on this review, recommendations for research avenues in HRI research will be made.

1.3 Work Items

The following work items were undertaken:

- A search of the literature to identify relevant journal articles, reports, books, etc., pertaining to current human-robot interaction research.
- Approximately 45 articles were selected from those identified in the search and were reviewed.
- A DRDC contractor report documenting the results of the literature review and recommendations for different research avenues in human-robot integration research for THRIL.

2. Methods

2.1 Keywords

A set of keywords were developed by the project team for the literature search based on our experience with the pertinent technological, scientific, and military domains. These keywords were chosen because they focused the search topics directly related to UAVs, UGVs, advanced control interfaces, human-robot interaction, and UAV and UGV interaction. The following keywords (Table 1) were used in combination to search accessible databases.

Table 1: Keywords

Core Concept	Primary Keywords	Secondary Keywords
Advanced Interface Design	Human factors, robot control, design requirements, situational awareness, workload, display characteristics, information clutter	
UAV	human factors, ergonomics, military, multiple UAVs, tasks, coordinated control, planning, human system integration, co-operative control, multiple control, performance, situational awareness, supervisory control, workload, fan-out	Metrics, out of the loop control
UGV	human factors, ergonomics, military, multiple UGVs, tasks, coordinated control, planning, human system integration, co-operative control, multiple control, performance, situational awareness, supervisory control, workload, fan-out	Metrics, out of the loop control
Human robot interaction	Situational awareness, workload, cognitive psychology, spatial perception, metrics, automation	online

2.2 Databases

The following were primary databases that are the most relevant for searching the scientific/ academic literature.

Table 2: Primary Databases for Scientific/Academic Search

Database	Description
IEEE – Institute of Electrical and Electronics Engineers, Inc	The IEEE, a non-profit organization, is the world's leading professional association for the advancement of technology. The IEEE publishes nearly a third of the world's technical literature in electrical engineering, computer science and electronics. This includes about 130 journals, transactions and magazines and over 400 conference proceedings published annually. IEEE journals are consistently among the most highly cited in electrical and electronics engineering, telecommunications and other technical fields. (IEEE, 2007)
NTIS – National Technical Information Service	NTIS is an agency of the U.S. Department of Commerce's Technology Administration. It is the official source for government sponsored U.S. and worldwide scientific, technical, engineering, and business related information. The database contains almost three million titles, including 370,000 technical reports from U.S. government research. The information in the database is gathered from U.S. government agencies and government agencies of countries around the world. (NTIS, 2007)
STINET – Scientific and Technical Information Network	STINET provides access to citations of unclassified unlimited documents that have been entered into DTIC's Technical Reports Collection, as well as the electronic full-text of many of these documents. Public STINET also provides access to the Air University Library Index to Military Periodicals, Staff College Automated Military Periodical Index, DoD Index to Specifications and Standards, and Research and Development Descriptive Summaries. (STINET, 2007)
Psyc Info	The PsycINFO database is a collection of electronically stored bibliographic references, often with abstracts or summaries, to psychological literature from the 1800s to the present. The available literature includes material published in 50 countries, but is all presented in English. Books and chapters published worldwide are also covered in the database, as well as technical reports and dissertations from the last several decades.
DRDC Research Reports	DRDC Defence Research Reports is a database of scientific and technical research produced by and for the Defence Research & Development Canada. It is available online at pubs.drdc-rddc.gc.ca/pubdocs/pcow1_e.html.

In addition, Google Scholar and the World Wide Web (WWW) was searched with the keywords

2.3 Search Strategy

The project team systematically searched the databases using the keywords specified. The core concept keywords were the most important words used in the search, as they represented the broad concepts to be investigated. As necessary, the primary keywords were used in order to ensure sampling of literature from several different areas within the core concept. For example, when searching with the "UAV" core concept, primary keywords such as "human factors" and "coordinated control' may or may not emerge. The purpose of the primary keywords was to ensure that research and information related to several aspects of human factors issues with UAVs/UGVs was explored. If an unmanageable number of hits results from a search with three words, additional modifiers (from the keyword list) were used to focus the results. When a keyword yielded too few searches, less narrow concepts were used until the precise level of analyses has been reached.

If necessary, searches were refined and/ or revised and continued using secondary level keywords.

When keywords were searched and a relevant article was identified, the following information was documented in a spreadsheet:

- Database searched (e.g., Psych Info);

- Keyword combination (e.g., Situational Awareness AND performance);
- Topic Categorization;
- Reference Citation;
- Abstract;
- Articles downloaded (Yes/No);
- Articles/books that require purchase (Yes/No), and;
- Rating of relevance and impact.

2.4 Selection Criteria

The research team developed some preliminary criteria by which to evaluate the articles found during the search process. First, relevance was defined as how closely the article relates to the research objectives outlined in the Statement of Work. Specifically, relevance was assigned the following 3 point scale:

- 1: Primary Focus of Article
- 2: Mentioned – but not the focus
- 3: No specific mention of any of these things, but still of some relevant.

If an article was found to be directly relevant to the topics identified in the statement of work and encompassed the core concepts and primary keywords identified it was given a value of '2'. If the articles was relevant but the focus of the article was not on the core concept or a primary keyword it was given a '1'. Articles that also were found to be directly relevant to the topics identified in the statement of work and encompassed the core concepts and primary keywords identified but were too narrowly focussed they were also given a value of '1'. Following the initial literature search, a more refined search was suggested to make certain all relevant literature was obtained. This included re-examining all the articles that were rated '2' in terms of relevance (52 in total) and identifying any relevant references and additional relevant keywords that were not included in our original keyword list. The following additional keywords were identified:

- Automation;
- Supervisory Control;
- Fan-out, and;
- Co-operative control.

These keywords were paired with the appropriate keywords and searched in all of the above mentioned databases.

Once titles and abstracts were ranked according to relevance, the research team obtained as many of the primary articles as possible. Overall, the references comprised books, journal articles, and technical reports from the behavioural sciences, military, and related domains.

2.5 Search Results

Of the selected articles, 8 articles were purchased and 38 articles were downloaded giving a total of 46 relevant articles. Each article was then briefly reviewed and classified according to article theme (i.e. advanced interface design, UAV/UGV control), source of the article (i.e. keyword search, from the TA, or from another article), and whether the article included theoretical or empirical research.

Of the thousands of articles identified by the search process a total of 80 articles were considered for the report but due to the down selection method they were not found to be as significant as the selected articles.

2.6 Structure of the Report

Section 1 of this report provides the background for the current project, and presents the scope of the work and the deliverables. Section 2 describes the method used to initiate the review, including how we found and chose articles to include for this review. Section 3 begins by describing two relevant metrics that are used to evaluate UV performance: situation awareness and workload. The section continues by providing results from articles across a number of key HRI areas including: automation, operator capacity and interface design. Finally, in Section 4 the report concludes with a section highlighting the conclusions of the research, as well as, lab specifications, and test capabilities of the future HRI laboratory. This section will also indicate new directions for future research for the HRI laboratory.

2.7 Limitations

This report has a key limitation that is important to note. The major limitation of this report involves the breadth of this review and the relatively limited number of primary articles reviewed (n=46). Given the breadth of information available, it would have been impossible to review the number of source articles necessary to provide a detailed picture of the current research in all of the specific areas. One of the challenges of such a wide-ranging review is that it is challenging to get adequate coverage of all possible research questions that relate to the human factors issues surrounding human-robot interaction. Many of the human factors issues / themes identified as influencing human-robot interaction (i.e. situational awareness and workload) would each be worthy of their own review. What we have attempted to do was to select the best available and most relevant information to show a broad scope of the current research.

3. Results

3.1 Situation Awareness

Endsley (1995b) defined situation awareness (SA) as “the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future” (p. 36). Endsley defined three levels of SA: perception (Level 1 SA), comprehension (Level 2 SA), and projection (Level 3 SA). Developing and maintaining situation awareness is a requirement for optimizing human-robot interaction (HRI) during UV operations; optimising HRI is a prerequisite for optimising performance. HRI primarily occurs through the robot’s interface, which is composed of its displays and controls. The level of SA that can be developed and maintained during UV operation is limited by system design, interface design, automation implementation, sensor technology, operator attentional resources, as well as task and environment challenges (Riley, Murphy, & Endsley, 2006).

Operator SA is one of the main metrics used to evaluate the performance of a UV system. In the context of UVs, the levels of SA correspond to the operator being aware of: the UV’s current location and status (Level 1 SA), the impact of elements in the environment on the system (Level 2 SA), and the impact of the UV’s current and future behaviour on the team’s goals or mission objectives (Riley, Strater, Chappell, Conners, & Endsley, 2010).

Many techniques have been developed to measure SA; each technique requires a set of metrics that are tailored to the specific task that is being evaluated. The goal of the metrics is to determine the level of SA with minimal impact on the “normal” behaviour of the participant, thus reducing the risk of biased data. The metrics used can be embedded performance parameters that are representative of the defined levels of SA, or they can be obtained as an external subjective measure (Endsley, 1995a). SA models and measurement techniques have primarily been focused on an individual, however, modeling and measuring team SA has been discussed. Salmon et al. (2008) reviewed literature that describes team SA as being more complex than combining SA from individual team members. In general team SA models need to include two additional components: shared SA between team members and combined SA of the whole team (Salmon et al., 2008).

Operator capacity is a term used to describe the number of UVs under control by a single human operator. The research goal is to increase operator capacity to be greater than one; however, studies have shown that this can cause a decrease in performance. The decrease in performance has been linked to a decrease in SA (Nehme, Crandall, & Cummings, 2008). Researchers have attempted to develop models that can predict the operator capacity that should be used for a specific mission task in order to achieve optimal performance, but most have been deemed unsuccessful. Research has suggested that the lack of success is due to the models’ inability to account for SA (Nehme et al., 2008).

Nehme et al. present a discrete-event simulation (DES) model that does account for SA, and thereby shows improvements in its ability to predict optimal operator capacity. SA is reflected in the model as a wait time that represents the interaction delay between the operator and UV when the operator loses SA. This wait time is incorporated through an operator model that limits the operator to only attending to one event at a time, and events that need operator attention while the operator is busy are placed in a queue. The arrival rates of events are a factor of the rate at which vehicles need attention and the operator loss of SA. SA is empirically represented in the model as a variable that is a function of operator utilization. The function curve is concave up such that low and high levels of utilization result in high penalties due to loss of SA. The study concludes that the inclusion of SA into a model used for predicting optimal operator capacity is effective (Nehme et al., 2008).

This type of model can be used to predict the optimal operator capacity as well as provide guidance for methods that could be explored to reduce the loss of SA when the UV team is implemented. Traditionally, advancements in automation and interface design have been the main avenues to accomplish the increase in SA and they will be discussed in more detail in the sections to follow.

3.2 Workload

Workload is another common metric used to evaluate how effectively a human-computer team completes a task (Crandall & Cummings, 2007). When used as a metric, workload represents the demand that a task requires for completion. This can be the combination of objective requirements of the task, condition it is being performed in, as well as the skill, behaviour and perception of the person required to complete the task (Hart & Staveland, 1988). Research has reported a negative correlation between workload and SA during the operation of UVs, such that as workload increases SA decreases (Sterling & Perala, 2007). There is a coincident negative correlation with workload and performance, such that as workload increases performance decreases.

There are several variables that affect perceived workload during a particular task and therefore similarly to SA, methods for obtaining workload ratings are tailored to each application. Subjective ratings are the most common method for collecting workload data; the NASA-TLX (Task Load Index) is the framework most often used. NASA-TLX requires participants to rate the task performed in the following categories: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The ratings are then used to determine overall workload as well as conclusions in regards to what category was deemed to be most influential on overall workload (Hart & Staveland, 1988).

Another implementation of workload during UV research is to classify, rather than subjectively rate, a particular task. For example, researchers might require a test participant to complete a task under two conditions (high workload and low workload), and attempt to draw conclusions as to how the participant performs in each condition. In this case the high and low workload tasks have been dictated by the researcher not perceived by the test participant. Strategies for setting workload conditions vary depending on the experimental design and test environment. The classifications of high and low thresholds are up to the researcher's discretion. For example, a surveillance task high workload conditions can be accomplished by increasing the difficulty to find targets by: increasing the number of non-enemy targets, decreasing visibility, etc. Another example would be for a system monitoring task, by increasing the difficulty to detect what is being monitored can be accomplished by: increasing radio chatter, providing distractions, etc.

An example study that used workload and SA as subjective measures to make conclusions about UV operations is described below. Sterling and Perala (2007) also introduce the concept of mixed UV operations which involve operating more than one type of UV simultaneously. In particular they wanted to determine the optimal combination of type of military unit being supported and UV types being used (Sterling & Perala, 2007).

Sterling and Perala (2007) tested 12 participants in a simulated military reconnaissance environment, where they were responsible for planning the route and dynamic attributes, such as speed, altitude, radius of surveillance, etc. The authors manipulated the type of military unit being supported in the simulation (non-line-of-sight [NLOS], Infantry, Recon, mounted combat systems [MCS], and multiple) as well as the type of UV (UAV only, UAV+UGV, UAV+ unmanned ground sensor [UGS], and UAV+UGV+UGS). During the test trials the participants were evaluated on subjective workload (NASA-TLX), stress (both physical and mental), and situation awareness. The result from unit and UV type combinations were compared to baseline results that the authors collected on the first day of the study, and represent the workload, stress, and situation awareness of driving to work. It is important to note that the authors did not have the statistical power in the experimental design to include a statistical analysis of their results

Results show that the following unit types resulted in higher workloads than the baseline: Infantry, recon, and multiple. Supporting an infantry unit resulted in the highest perceived workload. The following UV types resulted in higher workload results than the baseline: UAV+UGS and UAV+UGV+UGS. Operating all three assets resulted in the highest perceived workload. The authors also looked at the interaction between type of unit and type of UV and suggest that the type of asset and unit each contribute to the perceived workload of the operator (Sterling & Perala, 2007).

Physical stress levels were not found to exceed the baseline condition when testing military unit type; however, mental stress was greater for the following unit types: Infantry, recon, and multiple. Supporting infantry units resulted in the highest mental workload. Physical stress was only higher than the baseline condition for the UAV+UGS condition. The following UV types resulted in higher mental workload than baseline: UAV+UGV, UAV+UGS, and UAV+UGV+UGS. The combination of all three UV types resulted in the highest mental workload (Sterling & Perala, 2007).

Situation awareness showed the trend that when workload and stress were high, SA was low. SA results were lower than the baseline condition when participants were supporting recon and multiple unit types, with recon resulting in the lowest SA score. SA results were only lower than the baseline condition, with respect to UV type, when the participants were in the UAV+UGS condition (Sterling & Perala, 2007).

The experimental design of this study did not allow for strong statistical significance to be reported, however the researchers still concluded that workload and physiological stress was highest when UV teams were supporting infantry units. The researchers hypothesized this is because dismounted infantry units are the most vulnerable out of the units tested. Workload and mental stress were also highest when operators were required to use all three UVs, due to attention sharing (Sterling & Perala, 2007).

In order to optimize HRI, and subsequently increase UV performance during military missions, it is necessary to maximize SA and minimise WL. The future direction of research is to increase operator capacity of UV systems to be greater than one, but currently the increases in performance realized due to increases in WL and decreases in SA. The effectiveness of automation and interface design are two areas of research that will be reviewed in correlation with research looking at increasing operator capacity, to provide a literature representation of the factors involved in optimising HRI.

3.3 Automation

The operation of a UV includes: controlling the movement of robot, route planning, monitoring system status, monitoring and interpreting sensor data, and communicating relative information about the system to others. It is challenging for soldiers to perform all of the operations manually, which is why currently the UV to operator ratio is typically greater than 1:1. Technology can be used to perform these tasks autonomously, causing the operator to enter into more of a supervisory role (Cosenzo, Parasuraman, & de Visser, 2010). Automated UV systems can benefit military operations by potentially reducing manpower, lowering life-cycle costs and minimizing human exposure to hazardous environments (Liu, Wasson, & Vincenzi, 2009; de Visser et al., 2008). Liu et al. state that although automated systems can be advantageous in some situations, fully automated UV systems are not as applicable for complex, dynamic, military operations as they may be in other fields. Humans need to be kept in-the-loop because of their ability to adapt to new situations and apply judgment to their decisions. As support for human involvement, de Visser et al.'s (2008) summary of previous research suggests that possible disadvantages of automation for military applications are lower SA, unbalanced workload, decision biases, over reliance and complacency, and inappropriate trust. Given the disadvantages, the important factors of automation that need to be considered, in order to optimize the HRI for military applications, are the level of automation (LOA) implemented and the reliability of that automation; optimizing the HRI equates to maximizing SA and operator performance while minimizing workload.

Endsley and Kaber (1999) developed a ten level taxonomy of LOA that could be applied to general HRI; it is summarised in Table 3. Intermediate levels of automation are mainly used for UAV operation (Liu et al., 2009). Liu et al. go on to describe two levels of intermediate automation, management by consent (MBC) and management by exception (MBE). MBC limits the system from proceeding until the human approves the recommended action. MBE allows the system to choose the recommended action and proceed unless the human intervenes. Therefore, MBE is considered to be a higher level of automation than MBC.

Table 3: Endsley and Kaber (1999) LOA taxonomy

Level	Description
1 – Manual Control	Human performs all tasks.
2 – Action Support	System assists the operator with the performance of the selected action.
3 – Batch Processing	Human generates and selects the operation to be performed and the system automatically carries it out.
4 – Shared Control	Human and system generate the options, the human still selects the operation and shares carrying out the operation with the system.
5 – Decision Support	The system generates a list of decision options that the human can select from or they can generate their own list. The selection is then carried out by the system.
6 – Blended Decision Making	The system generates a list of option, makes a selection and carries out the task if the human consents. If they do not consent they can either choose from the list provided by the system or from their own list.
7 – Rigid System	The system generates a list of options, which the human must choose from. The system then carries out the selection.
8 – Automated Decision Making	The system generates a list, which can be augmented by alternatives suggested by the human, and selects the best option that it will carry out.
9 – Supervisory Control	The human supervises, and intervenes if necessary, as the system generates a list, selects the best option, and carries out the option.
10 – Full Automation	The human is out-of-loop as the system carries out all actions.

Research has focused on the impact of the level of automation on UV performance in terms of maximizing HRI. The studies described in the following sections look not only at the independent impact of the level of automation on HRI, but also at the dependent variables such as imperfect automation, adaptive automation and operator capacity.

3.3.1 Level of Automation

This literature search aims, to provide a base knowledge of the important factors involved with how level of automation can affect UV operations. While it is difficult to conclude an individual level of automation that is best, the research does show the general benefits of automation for UV operations.

Endsley and Kaber (1999) tested the impact of their taxonomy through a cognitive control task where 30 subjects performed simulation trials with varying levels of autonomy. The simulation task was not specific to UV operation, but it was designed to incorporate the following features of dynamic control tasks: collision avoidance, location and selection of objects, and task processing. The researchers were particularly interested in the operator's ability to revert to manual control following automation failure, for the purpose of designing a human-robot interface that will have a smooth transition when failures do occur. The participants were evaluated on task performance, workload (NASA-TLX), and SA (SAGAT). During the simulation participants were required to select targets (boxes) and collapse them before they reached an endpoint on the screen or before they collided with each other. The method used to accomplish the aforementioned task was dependent on the level of automation.

Results indicate that LOA has a significant impact on task performance, as measured by number of targets collapsed, expired and collided. LOAs that incorporated computer aiding or computer assumption of the implementation aspect of the task showed significant improvement in performance. LOAs that incorporated a joint human-computer generation of options showed a decrease in some aspects of performance, when compared to LOAs that incorporate human generation of options and LOAs that incorporated computer generated options. There was no difference in task performance between LOAs that incorporated human-computer selection, human selection or computer selection (Endsley & Kaber, 1999).

The results also indicate that time to recover from automation failure was significantly affected by LOA. Time to recover was higher when failures occurred during level 3 and level 8, and was lowest during level 2. In addition to recover time, automation failures caused declines in task performance, but only for some LOA. Task performance was significantly lower after automation failures for five LOAs: level 3, level 6, level 8, level 9, and level 10 (Endsley & Kaber, 1999).

The above results suggest that the decrease in performance observed in LOAs that involved joint human-computer actions could be attributed to the human participant's shift away from system monitoring or task selection when they are conducting the joint action. The participants' ability to recover from failure improved when the operator was incorporated in task implementation. In contrast, time to recover from failures was lessened when the LOA incorporated advanced queuing of targets. In conclusion LOAs that allocated the option generation and implementation roles of a task between the participant and/or automated system have a significant impact on performance (Endsley and Kaber, 1999).

SA results show that LOA had no significant effect on level 1 SA; however, the data trended towards increased level 1 SA when LOA was increased from level 1 to level 8. The results did show that level 2 SA was significantly different under different LOAs. Higher level 2 SA was found for the following LOAs: level 6, level 8, level 9 and level 10. There was a significant difference in level 3 SA under different LOAs. The data shows an increase from LOA 1 to LOA 4, after which level 3 SA drops off with the exception of level 7 LOA, where level 3 SA peaked.

SA results suggest that improved level 2 SA was due to operators not being required to select decisions on their own. In general lower SA in lower LOAs was due to the participant being required to make decision selections and monitor the system while generating strategies. The authors were unable to provide an explanation for the trend in level 3 SA results (Endsley and Kaber, 1999). These results demonstrate the non-uniform effect of LOA on different levels of SA; the LOA best suited for creating and maintaining one level of SA may not be well suited to another level of SA.

Workload results show that LOA had a significant effect, and the following LOAs resulted in significantly lower workload: level 6, level 8, level 9, and level 10. This trend is an inverse correlation with SA results, such that for those LOAs with higher SA resulted in lower WL. A trend in the WL results showed operators' perceived success to increase as LOA increased. Workload results show that improved WL levels were found in LOAs that did not require the participant to make decision selections on their own (Endsley and Kaber, 1999).

In conclusion this study suggests that human performance is benefited most when the implementation portion of a task is automated. However, performance recovery is much lower if automation fails and the human is excluded from the implementation of the task. When the operator and system are jointly required to generate the task option list, the performance is lower than if just the operator or just the system generates the list. When the decision making portion of the task is automated the operator's work load was lowered and situation awareness was raised; however, task performance was only slightly better.

While the previous study tested the full spectrum of automation levels for their overall task, several other studies have tested the validity of automating a specific component of the UV task, such as the flight path, target recognition, etc. Typically, a few automated conditions will be compared to a baseline of no automation. Comparing the empirical metrics from these studies is challenging because each author uses a slightly different mission task or application of automation. The studies do allow for conclusions to be drawn about the general benefits of automation on UV operation, and what the effects are of automation reliability, adaptive automation, and operator capacity.

Wickens and Dixon (2006) conducted a complex UV study that looked at multiple variables that would impact UAV operator performance, with one of the variables being automating flight control. The study presents eight experiments: the first two were developed with the background of multiple resource theory where the goal was to offload workload by automation and auditory cueing. The difference between the first and second experiment is that the second included an increase in operator capacity as well as model development. The results from auditory cueing are described in more detail in section 3.5.2 and the results from increasing operator capacity and model development are described in section 3.3.5. All of the experiments were simulated UAV operations, seven of which were done in a Hunter/Shadow simulator and one which was on a general UAV simulator platform. The Hunter/Shadow simulator had a primary mission task of tracking the UAV waypoints and reporting on "command targets" (CT). There were two secondary tasks, one in which the participant was required to search for "targets of opportunity" (TOO) and another in which they were required to detect and respond to system failures (SF). The participants in these studies were primarily student pilots.

In the first experiment Wickens and Dixon (2006) investigated the ability of a 100% reliable autopilot to mitigate the workload of operating a single UAV, when compared to a baseline condition with no autopilot. The authors predicted that automation would improve performance, based on multiple resource theory.

Performance improvements were observations with respect to: the number of times instructions needed to be repeated, target monitoring, and system monitoring. Automation resulted in a lower average number of times that mission instructions needed to be repeated, indicating better parallel processing than the baseline condition. There were significantly more targets detected and system failures detected in the automation condition when compared to the baseline condition. The researchers found that the autopilot was successful in mitigating workload by removing the pilot's task of monitoring and controlling the heading trajectory. This action reduced overall demands and in turn improved time-sharing amongst the mission tasks (Dixon & Wickens, 2003; Wickens & Dixon, 2006; Wickens, Dixon, & Ambinder, 2006).

3.3.2 Imperfect Automation

As previously mentioned, the level of automation is not the only variable to consider when looking to optimize HRI. Endsley and Kaber (1999) looked at effects due to automation failures because failures will inevitably occur in real applications, including UV operation. Wickens and Dixon (2006) stress the importance of the automation supervision task as the automation may be imperfect. The source of such imperfection will differ depending on the type of UV operation.

Wickens and Dixon's experiments three to eight (Levinthal & Wickens, 2005; Wickens & Dixon, 2006) looked at automating secondary tasks such as system monitoring and target recognition. They also introduced the variable of imperfect automation to each of these tasks. For system monitoring and target recognition, they investigated the effects of altering the threshold between imperfect detection and imperfect diagnostics. Following signal detection theory, a threshold that leads to imperfect detection is classified as a miss prone (MP) system whereas a threshold that leads to imperfect diagnosis is classified as a false alarm prone (FAP) system.

The results showed that when perfect, automation was beneficial to all three tasks, which supports the findings from the first two experiments. For all three tasks, when automation became unreliable the benefits of automation were degraded, as expected. An interesting finding was that unreliability in the secondary tasks (system monitoring and target recognition) was more detrimental than unreliability in the primary task (path monitoring). They suggest that this was because participants treated path monitoring as the primary task and it was therefore treated as being more critical in mission success (Wickens & Dixon, 2006).

When the type of automation was taken into account, performance degradation as a result of FAP automation was due to participants not believing an alarm when it went off and therefore not leaving their current task to respond to it. Furthermore, since automation misses are unlikely in a FAP automation threshold, performance was also degraded because participants became complacent with the alarm system and would not see targets or system alarms if not given an automated cue. It was anticipated that MP automation would degrade performance because the participant would be required to devote more time to double checking the task that is being automated. This hypothesis was not consistently supported by the results because degraded performance was only found when the automated task was cognitively or perceptually demanding (Wickens & Dixon, 2006).

3.3.3 Adaptive Automation

Adaptive automation (AA) is an emerging concept that attempts to consider the complex human-robot interactions that take place while operating a UV. AA has been defined by Kaber, Wright, and Sheik-Nainar (2006) as "dynamic allocation of system functions to a human operator and/or automatic controller over time based on operator states and task contextual information for the purpose of optimizing system performance." (p. 528). They follow their definition by outlining the AA goals to reduce operator workload and increase SA by allowing a better match between the given task demands and operator cognitive resources. De Visser et al. (2008) add to the broader definition by describing two similar forms of automation, adaptable and adaptive, to be "a system that is flexible and responsive to user needs, environment demands and context." (p. 2). They differentiated between the two by describing adaptable automation as a system that could accept a high level set of commands, and then interpret and execute them. An adaptive system would change state based on a set of criteria, such as operator workload reaches a certain level.

In correlation with the definitions of AA previously discussed, the following articles provide support that AA will be beneficial to UV operations. Squire, Trafton, and Parasuraman (2006) review addition previous work and suggest that a UV operator is able to recognize situations when automation might not be the best choice and therefore would prefer manual control. Consequently, they suggest that the operator would benefit from an adaptive automation interface. De Visser et al. (2008) summarised work that has already shown AA that allows for improved supervisory control of multiple UVs; their review included studies that looked at operating a UAV and UGV. Kaber et al. (2006) summarised early AA research and suggest that dynamic function allocations (DFAs) associated with AA can cause a temporary decrease in performance when the system switches between automation modes. This could be due to either the fact that the operator was not ready for the mode shift or they may be in the middle of a task. They went on to suggest that sensory cues could be used to help prepare the operator for mode switching. HRI research will help improve the benefits of AA and overall UV performance.

To investigate the effect of different levels of automation and performance degradation due to mode switching, Kaber et al. (2006) performed a study that implemented two AA programs, one which was predominantly manual control and the other which was predominantly supervisory control. In addition they had fully manual and fully automated control conditions. They used 32 participants in a simulated under-water mine disposal interface. The metrics used to evaluate the participants were performance, workload (NASA-TLK) and SA (SAGAT). The performance measures include: time-to-task completion (TTC) and number of task performance errors.

The study results supported previous research that suggested performance degradation after a state change. The results showed a significant difference in SA before and after automation state changes, with SA always lower after the state change (Kaber et al., 2006). The proof of different before and after state changes allowed for conclusions regarding how different levels of automation affect the difference in performance.

In this study, the type of automation had a significant effect on TTC, SA, and workload. In general, TTC decreased as the amount of automation increased, and fully automatic control resulted in significantly lower TTC than in all other modes. Although not statistically significant, AA that incorporated primarily supervisory control resulted in lower TTC than fully manual control. In contrast, supervisory AA resulted in significantly lower SA than all other control modes, and there was no significant difference between the remaining three modes. Workload ratings revealed that fully automatic control resulted in a significantly lower perceived workload than the other three modes. AA that incorporated predominantly supervisory control was not perceived to be different that manual control; however, AA that incorporated predominantly manual control was perceived to be worse than both (Kaber et al., 2006).

The results suggest that an AA model could be developed to strategically allocate tasks between manual and supervisory control. The developed AA would ideally be able to maintain SA levels equivalent to those typically found under manual control conditions, while resulting in better operator performance and no additional perceived workload (Kaber et al., 2006). This result was supported by another experimental study by Squire et al. (2006) that went beyond the work of Kaber and colleagues to hypothesize that different combinations of automation and operator capacity would result in different switching costs.

Squire et al.'s study was conducted in a simulated UV environment, where the objective was to use the robots to capture a flag. During the trial the operator could choose between one of two strategies for each robot: offense or defence. Participants needed to adjust their strategy throughout the trial depending on the situation that was occurring on the field of play. The simulation program had three levels of automation (waypoint, play, and super-play) and three methods of UV selection (individual, group, and all). In addition there was an adaptive automation condition where the operator was able to select the automation level and UV selection method. In the waypoint automation condition, the operator selected some number of robots and selected a point on the display for the robots to move to. In the play automation condition the participant selected some number of robots and then selected from a list of predefined movements for the UVs to perform. In the super-play automation condition the participant selected multiple robots to perform a mix of plays. This study tested five different control combinations that were a mix of the following level of automation and number of UVs that could be controlled: waypoint-individual, waypoint-selectable, play-individual, play-selectable, and selectable-selectable. Participants were evaluated on overall mission completion time, number of commands, and the time taken to switch between defensive and offensive strategies (Squire et al., 2006).

The results showed that mission completion times were significantly lower in the conditions with rigid automation than those with rigid manual control. Although not statistically significant the data trended towards the adaptive automation having lower completion times than the other automation conditions. The control combination had a significant effect on the number of commands issued by the participant. The data shows that the combinations involving play or selectable automation were lower than the combinations with waypoint automation. Time to switch between strategies was significantly different between all three automation conditions, with the order from fastest to slowest being: waypoint, selectable, and play (Squire et al., 2006).

Squire et al. (2006) conclude that their research supports AA as a means of improving operator performance. They were also able to support the hypothesis that switching costs are dependent on automation level. This suggests that AA is an import aspect of automation design that needs to be considered when optimizing HRI.

3.3.4 Adaptive Automation for Mixed UV Operations

The benefits of AA have already been discussed, and in addition to the control of multiple UVs, an important area of research in AA concerns mixed UAV and UGV operations. Parasuraman, Cosenzo, and De Visser (2009) conducted two experiments that examined AA in a simulated reconnaissance mission that involved the supervision of both UAVs and UGVs.

The mission was composed of four subtasks: 1) UAV target identification 2) UGV route planning 3) communications and 4) change detection. The first experiment was predominantly used to determine the parameters of the change detection system and to classify appropriate high and low task loads. The second experiment compared three automation levels: manual, static automation, and adaptive automation. Automation was provided to the system via automatic target recognition (ATR) for the UAV task. The difference between static and adaptive automation was that in the static condition the ATR was turned on for all participants halfway through the mission, but in the adaptive condition ATR was initiated only if the operators change detection performance dropped below a threshold by the halfway point. High and low task loads were incorporated through the difficulty of the communication task. Participants had to monitor communications for their own call signs and simply identify when they heard one of their own signs. Difficulty of this task was determined by the ratio of their own signs mixed with other call signs. A lower ratio of own signs to other signs represented a high workload and vice versa for low workload (Parasuraman et al., 2009).

The second experiment used sixteen participants to test the three previously described automation levels. The metrics used for evaluation were task performance, SA and WL (NASA-TLX). The task performance metrics were: UAV target acquisition accuracy and reaction time (RT), UGV route planning RT, communication RT and percent missed for own call acknowledgment, and change detection accuracy and RT. SA was assessed based on a series of verbal questions designed to measure perception and comprehension (Parasuraman et al., 2009).

Change detection accuracy was calculated as the percent of imbedded icon changes detected by the operator. Changes would occur at random throughout the mission; participants were warned prior to the study and asked to identify when they thought a change occurred. The results were separated into pre-automation and post-automation. As expected, there was no difference in the change detection accuracy pre-automation, however, there was significant difference in automation condition, post-automation. Change detection accuracy was significantly higher in both the automation conditions than in manual control and adaptive automation was significantly higher than static. Change detection accuracy was also significantly higher in the low workload condition than high workload (Parasuraman et al., 2009).

The researchers made an interesting conclusion when they compared change detection accuracy within the adaptive automation group. There were three participants who did not have the ATR turn on because their accuracy was above the threshold at the halfway point. Therefore, their accuracy was significantly higher than the remainder of the group pre-automation; however, when the automation did turn on for the rest of the group it raised the groups' accuracy so that it was no longer significantly different from the three participants. This means that the automation was capable of raising the user performance to an acceptable level (Parasuraman et al., 2009).

The study results showed that there was no effect of automation condition on UAV target acquisition accuracy and RT, or UGV route planning RT. Automation condition did have a significant effect on communication accuracy and RT, such that significantly higher accuracy was found in both the automation conditions compared to the manual condition, and furthermore higher accuracy was found in the adaptive automation condition over the static automation condition. Adaptive automation also resulted in a significantly lower communication RT than both static and manual conditions and there was no significant difference between static automation and manual control. Unlike the automation condition, communication workload condition had a significant effect on UGV route planning RT, and communication accuracy and RT. UGV route planning RT and communication RT were both significantly lower, and communication accuracy was significantly higher when communication workload was low, compared to high (Parasuraman et al., 2009).

Parasuraman et al. (2009) found that situation awareness was higher in both the automation conditions compared to manual control, however there was no significant difference between adaptive and static automation. Workload results were significantly different for all three control conditions, the order from lowest to highest was: adaptive, static, and manual. With the combination of subjective and empirical results the authors concluded that automation enhanced SA, lowered perceived workload, and increased change detection accuracy. Furthermore, adaptive automation had significantly higher change detection accuracy and lower perceived workload than static automation. AA also had higher SA than static automation although not to statistically significant levels (Parasuraman et al, 2009).

The adaptive automation results reviewed in the previous sections indicate that the there are multiple variables that contribute to the best level of automation that should be implemented for UV control. Operator capacity is another topic discussed frequently in literature and there is a close connection between the optimal level of automation and operator capacity. The following section looks at studies that specifically incorporated level of automation and operator capacity into the main effects of their investigation.

3.3.5 Level of Automation and Operator Capacity

The interaction between level of automation and operator capacity depends on the specific mission task and UV being operated. This once again makes it challenging to compare studies, due to their different test conditions, and make a concrete conclusion about the best level of automation for each level of operator capacity. The studies described in this section manipulated both level of automation and operator capacity in such a way that it is possible to identify that there is a significant interaction between the two parameters and what that interaction means for UV operations. Wickens and Dixon (2006) found automation while operating two UAVs to be somewhat successful in primary mission completion and secondary system monitoring, but unsuccessful in the secondary target recognition task. This finding is in contrast to their earlier study that found automation to be successful in all tasks while operating one UAV. Ruff, Narayanan, and Draper (2002) found an interaction between higher management levels of automation and operator capacity, but only for subjective data. This shows that for the participant there appears to be an optimal combination of LOA and OC, which correlates with concepts discussed in adaptive automation literature. Liu et al. (2009) highlighted the importance of applying the appropriate automation management strategy in order to achieve adequate performance during high-level cognitive tasks, but they were unable to find an interaction between higher management levels of automation and operator capacity. However, there were significant results of both LOA and operator capacity. Riley and Strater (2006) conducted a study that used real UGVs in a lab and were able to find both an interaction between LOA and operator capacity, and significant performance, SA, and workload differences for different LOA conditions.

The series of experiments by Wickens and Dixon (2006) introduced in section 3.3.1 are relevant in this capacity as well. In their second experiment, Wickens and Dixon added a two UAV condition and evaluated operating one or two UAVs in either the baseline configuration or in 100% reliable autopilot, with and without auditory aids. They used the results from this study to not only analyse performance differences but also evaluate the validity of applying a single channel, single resource or multiple resource model into a general performance model. In contrast to experiment one, the autopilot was somewhat successful in mitigating the workload of the primary mission completion and secondary system monitoring during dual UAV operation; however, there was no benefit seen with the secondary TOO surveillance task. The results suggest that even with 100% reliable autopilot, an operator cannot perform effective en route surveillance while operating two UAVs.

A performance model for UV operations may provide a method for determining the operator capacity and level of automation that would optimize performance. The modelling analysis from the second experiment of Wickens and Dixon (2006) revealed that a single channel model that including a cost for switching between tasks was suitable for dual UAV operation in the baseline condition. In the autopilot condition, a single channel model without a cost for switching between tasks was adequate. However there were indications that a single channel model does not account for the effects of implementing auditory aids to assist with task completion. . When the authors looked at the implementation of a single resource or multiple resource models, they found that single resource model was adequate when the operator was under high load while operating two UAVs in either the baseline or autopilot condition, without auditory aids. Consequently when load was reduced and auditory aids were implemented there is a need for a model that incorporates a combination of single resource and multiple resource theories. In general they conclude that this type of modelling consideration would be essential for developing a performance model for UAV operations (Dixon, Wickens, & Chang, 2005; Wickens & Dixon, 2006).

Ruff et al. (2002) investigated management strategy in a simulated UAV study that was a 2x3x3 mixed subject design. Twelve participants were placed in two groups 100% fidelity and 95% fidelity, where fidelity equated to automation reliability. The other two independent variables were management type (manual, MBC, and MBE) and operator capacity (one, two, and four UAVs). Each participant completed nine target recognition and identification trials, one with each management level in combination with each capacity level. Their performance was measured by the following metrics: mission efficiency, percentage of correct rejection and incorrect decision aids, event management, and number of UAV hit points sustained. Subjective measures collected include:

NASA-TLX workload ratings, Subjective WORkload Dominance (SWORD) ratings, self-rating of SA on a seven-point Likert scale, SA-SWORD score, and a trust rating on a 100-point Likert scale.

Performance results showed that automation reliability only had a significant main effect on mission efficiency, where mission efficiency was lower for the 95% fidelity condition than 100%fidelity. Subjective data also showed a significant interaction with fidelity and management type for the post trial SA scores, trust in automation scores, and post study workload scores. This indicates that there would also be a preferred management type for different levels of automation reliability. There was also found to be an interaction between fidelity and operator capacity for post-trial situation awareness and trust in automation. This indicates that there would be a different level of operator capacity that would maximise SA with different automation reliability levels (Ruff et al., 2002). These results further support the literature on automation reliability and adaptive automation.

The operator capacity results of this study show that as the number of UAVs controlled increased, the percentage of events managed decreased. There was no main effect found between operator capacity and mission efficiency, however there was one found for number of UAV hit points sustained. When operators were in control of one UAV they had significantly lower hits than when operating two UAVs (Ruff et al., 2002).

Ruff et al. (2002) concluded that MBC had advantages over MBE and there were interactions between type of management, workload, SA, and trust. Performance data showed that MBC had better overall mission efficiency, percent correct detection of decision aid faults, and number of hit points than manual control and MBE. With respect to event management, MBE and MBC resulted in a higher percentage of managed events than manual control. The authors concluded that this makes sense because the system was identifying events to the operator. The level of automation and operator capacity interaction was evident because the drop in performance as more UVs are added was different depending on the control condition, with manual control dropping by the greatest amount.

Subjective data showed that management type had a significant interaction with operator capacity for post-trial workload and situation awareness data, as well as the trust in automation data. For example, manual control resulted in the lowest workload in the single UAV condition but in the four UAV condition manual control resulted in the highest workload. Situation awareness and trust in automation results demonstrate the interaction because different levels of automation resulted in different sensitivities to change in operator capacity. The significant interactions found by Ruff et al. (2002) indicate that for each level of operator capacity there would be a different type of management that would lower workload while keeping a high level of situation awareness and therefore provide the optimal performance.

Liu et al. (2009) also investigated the interaction between MBE and MBC automation, and operator capacity. They found similar LOA and operator capacity results, but they were unable to find the same interactions as Ruff et al. In both studies the authors found that MBC had performance advantages over MBE and as operator capacity increases more tasks can be performed but there might be a cost of efficiency or accuracy. It is important to note that Liu et al. did not include any subjective ratings in their analysis and therefore could only make conclusions based on performance results.

Liu et al. (2009) implemented a 2 x 3 between subject design, where the independent variables were management type (MBC and MBE) and operator capacity (1, 2, and 4 UAVs). Each of the 60 participants experienced one of the six conditions in a simulated environment and the metrics used were task performance and workload. The UAVs flew a predetermined flight path while capturing target images. The operator's primary task was to monitor the images to determine the accuracy of the Automatic Target Recognizer (ATR) and they were assessed on their response time, queue time, processing time, target selection accuracy, manual accepts/rejections, automatic accept/rejections, and image hold counts. The operator's secondary tasks were to operate the Mission Mode Indicator (MMI) and monitor for Unidentified Aircrafts (UAs). They were assessed on MMI event occurrences and response times, as well as UA occurrences and response times.

The results showed that type of management only had a significant effect on the image processing time of the primary task, where MBC resulted in a significantly shorter image processing time than MBE. Operator capacity had a significant effect on image processing time and task accuracy for both the primary and secondary tasks. No significant difference was found between the one and two UAV conditions for all of the results; however, the two UAV condition had significantly lower processing times and higher accuracy than the four UAV condition for both primary and secondary tasks. The one UAV condition had significantly lower processing times than the four UAV condition for both primary and secondary tasks but only had a significantly higher accuracy for the secondary task. The lack of significance for most of the automation results and the interaction results was not anticipated and the authors suggested that the metrics used in this study were not sensitive enough to detect the performance differences, and the trial durations were relative short (Liu et al., 2009).

The studies above all dealt with simulated UAV experiments, however there are additional factors worth studying that are involved when operating real UVs, but they have not been given as much attention in literature their economical resource requirements. One study that explored the interaction between the level of automation and operator capacity with real UGVs was done by Riley and Sarter (2006). studied This study required participants to operate real UGVs in a lab environment, and the results showed significant performance and workload advantages to semi-autonomous control, over manual and fully automatic. There was also a significant interaction found between level of automation and operator capacity, such that benefits to the addition of a second UGV were only seen under certain levels of automation.

The study by Riley and Sarter (2006) required twenty participants to navigate UGVs through a maze under one of four randomly assigned control modes: 1) manual control of one robot, 2) manual control of two robots, 3) manual control of one robot and supervisory control of the other in semi-autonomous mode (participant was required to give directional decision at various intersections), and 4) manual control of one robot while supervising the other in fully autonomous mode. The metrics used to evaluate the participants were SA, workload, and navigation performance (Riley and Sarter, 2006).

Riley and Strater (2006) found that navigation performance was significantly better for control condition three, when compared to the control condition two and four. There was no significant difference between control condition two and four. This result indicates a significant difference caused by the level of automation, such that semi-autonomous control of the second robot was better than both manual and fully automatic control. Although not statistically significant the data trended towards condition one being different than condition three, but not different than two or four. This result indicates an interaction between level of automation and operator capacity, because the addition of the second UGV only improved performance when in semi-automatic control.

Overall perceived workload was significantly lower for condition one compared to conditions two and four, which once again represents an effect of level of automation. When the second UGV is under semi-autonomous control the overall workload is lower than both manual and fully autonomous. When the individual workload components (mental, physical, and effort) are analysed the interaction between level of automation and operator capacity can be observed. Mental workload was only higher with the addition of the second UGV in semi-autonomous control. Physical workload was only lowered with the addition of a second UGV when it was operated with either semi-autonomy or full autonomy. Effort was only higher with the addition of the second UGV when it was operated in manual control.

Overall situation awareness was not significantly different between the different control conditions; however, there was significant difference for Level 2 and Level 3 questions. Manual control resulted in higher SA than control conditions that involved automation, suggesting an interaction between level of automation and operator capacity. For the semi-automatic condition this suggests that keeping track of the UGV motion in combination with making decisions based upon past automated control actions was a difficult task to cope with.SA was only higher with the addition of a second robot when it was controlled manually (Riley & Sarter, 2006).

The results described above combined with the previous sections highlight that the level of automation, reliability of automation, and operator capacity are important variables to consider when optimizing HRI for UV operations. Moreover, the type of UV and mission task will also have an effect on the desired level of automation, and thus need to be carefully considered. The complex interaction of variables has raised important discussions about the use of adaptive automation to provide improvements over fixed levels of automation, but there are concerns and challenges with effective implementation of AA. The type of research that has been discussed is critical in achieving the future goal of increasing operator capacity such that the ratio is greater than 1:1.

3.4 Operator Capacity

The impact of operator capacity is greater than just its interaction with the level of automation, and therefore it has received considerable attention in the literature. Operator capacity refers to the number and type of UVs that a human operator controls or supervises. As discussed in previous sections, the future direction of UV systems is to have one human operating/supervising multiple UVs. Lif, Hedstrom, and Svenmarck (2007) summarise a NATO report, stating that when determining the optimal operator-to-UV ratio, the dependent variables are: task and coordination demands, level of automation, information perception, and the operator's working memory, responsibility, and decision making. They state that two reasons the operator-to-vehicle ratio is difficult to improve is because a vehicle can only be left unattended for so long before performance degrades, and there are cost associated with switching between vehicles and/or tasks.

This literature search focused the effects of operator capacity on UV operator performance, and how operator capacity interacts with other variables such as automation and interface design. Because UV types and mission tasks vary so extensively, directly comparing studies to determine an optimal operator capacity for all UV operations is difficult. This review provides an overview of studies that have examined the performance effects of operator capacity and the impact of automation and interface design. The final section discusses the merits of a set of metrics that has been developed to consider how the variables that affect the appropriate operator capacity for UV operations interact.

3.4.1 Operator Capacity and Task Performance

Previous sections highlighted the validity of implementing AA for the control of UVs and Lif et al. (2007) used AA with real UGVs to determine the effect of operator capacity on task performance and the strategies that were implemented by the participants. They were able to detect performance benefits when operator capacity was increased. Their strategy observations provide relevant insight into proper use of AA that is important for overall system design.

The study used 12 participants in a lab to operate multiple live UGVs in a simulated Military Operation in Urban Terrain (MOUT) environment. The operator's objective was to navigate the

UGVs to a predetermined inspection point as quickly as possible under one of three operator capacity conditions: one, two, or three UGVs. The metrics used in the study were number of inspection points arrived at, time spent in different control modes, instantaneous performance (IP), and subjective results from an interview. There were three modes of operation that the participant was allowed to choose from throughout their trial: manual control, autonomous mode (continue with current speed and heading), or camera mode when the UGV is standing still (the operator is only able to rotate the camera). IP was calculated at a frequency of 1 Hz to determine the efficiency of the operators control strategy; it was calculated by dividing the actual distance travelled by the distance that could have been travelled at maximum speed (Lif et al., 2007).

Results show that there was a significant increase in inspections reached with multiple UGVs than with one, while there was no significant difference between the two and three UGV conditions. In contrast, there was significantly lower IP when operating multiple UGVs than one and again there was no significant difference between the two and three UGV conditions. It was concluded that IP could be used as an indicator for how many inspections points would be reached, based on the significant correlation observed between average IP and average number of inspection points per UGV (Lif et al., 2007).

The subjective results showed that subjects generally divided their task into three sub control tasks: 1) manual control and transition to autonomous mode, 2) utilization of the UGVs and 3) navigation. Their main strategy for operating multiple UGVs was to use all three vehicles simultaneously, and in some occasions abandon the third UGV when mental workload was high. In those high workload cases the participants would operate one UGV manually, and supervise the other in autonomous mode. It was also concluded that performance needed improvement for this application until an interface is designed that reduces switching costs. These results support other studies that have highlighted the benefit of implementing adaptive automation, and the potential cost of task switching (Lif et al., 2007).

It was concluded that performance increases when an operator is given a second UGV to operate; however, there are no benefits to the addition of a third UGV. Higher performance for this type of task could only be achieved with improved automation from what was used in this study (Lif et al., 2007).

Optimizing operator capacity is an area of interest in industries other than the Military; however, the performance benefits have not always been found. For example, Adams (2009) conducted a study that investigated the effects of operator capacity on task performance for the application of indoor materials handling. The results showed a decrease in performance as operator capacity increased. The study used 12 participants, repeated in two sessions, to complete a simplified UGV transportation tasks. Participants were required to move UGVs (one, two, or four) to a goal location while avoiding obstacles. The metrics used to evaluate the participant during the tasks were: workload (NASA-TLX), number of task completions, task completion times, number of errors, and number of commands issued per task.

Results showed that there was a significantly higher perceived workload for the four robot tasks than the one and two robot tasks, and there was no significant difference in the workload between the one and two robot tasks. This suggests that the four robot task required significantly more cognitive capabilities than the one and two robot tasks. When the results were compared between the two sessions, the findings suggest that level of experience only had a significant effect on workload for the two robot task (Adams, 2009).

Number of task completions was found to be significantly higher for the one robot task than the four robot task; however, there was no significant difference found between the other two task combinations. When the task completion results were compared between sessions there was no significant effect of experience (Adams, 2009).

Task completion times were found to be significantly different for all three task conditions. The order of fastest to slowest times was: one robot, two robot then four robot task. Experience also had a significant effect on task completion times because completion times were significantly lower in the second session (Adams, 2009).

Total number of commands per task and total number of errors committed per task were found to be significantly higher during the four robot task than the one and two robot tasks; however, there was no significant difference between the one and two robot tasks. Experience did not have a significant effect on total number of commands or total number of errors committed per task because the results were not significantly different between the two sessions (Adams, 2009).

Adams (2009) makes the same overall conclusion as Lif et al. (2007) that if operator capacity is to increase then automation and interface design need to be improved. This conclusion was supported by Adams' (2009) results that show for the four robot task, perceived workload, task completion times, and number of commands issued increased as number of tasks completed decreased.

3.4.2 Operator Capacity and Automation Reliability

Automation reliability has been identified as an important aspect of automation that needs to be considered for optimising HRI. Therefore the interaction between automation reliability and operator capacity is also of interest. Automation reliability has been discussed in detail in section 3.3.2, in particular the series of experiments by Wickens and Dixon (2006). One of their experiments investigated the interaction between operator capacity, automation level, and automation reliability. They found that increased operator capacity decreased performance, and there was a significant interaction between operator capacity and automation reliability (Levinthal & Wickens, 2005). This section examines that study in more detail.

Levinthal and Wickens (2005) required 42 participants to complete two trials, one while operating two UAVs and another while operating four UAVs. The automation reliability was randomly altered between the subjects among four possible settings: no automation, FAP, MP, and high reliability automation. The operator was responsible for two tasks: management of the UAVs and detection of tanks. Automation was varied via ATR for the detection of tanks. The metrics used to evaluate the participants were: tank detection response accuracy, response time, compliance, complacency, reliance, and UAV idle time (Levinthal & Wickens, 2005).

Levinthal and Wickens (2005) concluded that increased operator capacity significantly lowered the operator's tank detection response accuracy. They were unable to find a significant effect of automation or automation reliability on tank detection response accuracy. Operator capacity, automation, and automation reliability appeared to have an interaction on tank detection response accuracy. No automation showed the smallest decrease in tank detection response accuracy when operator capacity was increased compared to the three automated conditions. High reliability automation showed the largest decrease in accuracy when operator capacity was increased and MP automation showed the smallest decrease of the three automated conditions.

Overall response times were significantly higher when operators were in the four UAV condition and the data showed a trend towards an interaction of operator capacity, automation level and automation reliability. The baseline condition with no automation was more sensitive to an increase in operator capacity than the three automated conditions. Within the automated conditions, FAP was the most sensitive to a change in operator capacity, while MP and high reliability automation showed little effect of change in operator capacity (Levinthal & Wickens, 2005).

Compliance was measured as the response time after an alert sounded, and therefore the no automation condition is omitted from this section of the analyses. Operator capacity was shown to not have a significant effect on compliance; however, the data did show a trend of compliance increasing as operator capacity increased. An interaction between operator capacity and compliance was also visible from the data. High reliability automation showed the largest increase in compliance when operator capacity was increased, followed by FAP, and then MP (Levinthal & Wickens, 2005).

Complacency was measured as the time it took the operator to detect a tank when the ATR failed to identify it. Therefore, high reliability automation was omitted from this analysis. There was no significant effect of operator capacity on complacency; however, the data did trend towards an increase in complacency as operator capacity increased. Complacency in the no automation condition appeared to be more sensitive to an increase in operator capacity than the other two automated conditions (Levinthal & Wickens, 2005).

Reliance was measured as the proportion of the response time after an alarm had sounded and therefore the no automation condition was omitted from this analysis. Reliance was significantly higher when operator capacity was increased and the sensitivity of reliance did not appear to be influenced by the automation reliability (Levinthal & Wickens, 2005).

UAV performance was measured based on UAV idle time and the results from this analysis show the same trend as reliance. Idle time was significantly higher when operator capacity was increased and did not appear to be sensitive to the presence of automation or the reliability of automation (Levinthal & Wickens, 2005).

The results show that increasing operator capacity from two to four caused a decline in performance of both the primary control task and secondary surveillance task. The level of decline in performance appeared to be dependent on the automation condition, where different conditions were more sensitive to an increase in operator capacity. This supports a significant interaction between operator capacity and automation. The control interface used in the study only allowed for the operation of one UAV at a time and the authors suggest that an interface that provides better simultaneous control would be beneficial (Levinthal & Wickens, 2005). The operator capacity performance results and interface criticisms are similar to those that have been previously discussed, and reinforce that achieving higher levels of performance by increasing operator capacity is dependent on an interface design that addresses the specific mission requirements and implements an appropriate automation scheme.

3.4.3 Operator Capacity and Interface Design

A more detailed discussion about the factors of interface design is covered in section 3.5, but this section discusses two articles of interest. The studies presented in these two articles investigated the effects of increasing operator capacity; however, they also altered interface design. The intention was to provide empirical proof that there is an interaction between interface design and operator capacity. Both of the studies found similar declines in performance as operator capacity increased, as did the studies previously described. Both studies also found significant interactions between interface design and operator capacity (Chadwick, 2006; Baber, Morin, Parekh, Cahillane, and Houghton, 2011). The specific interface design details and related results will be discussed in section 3.5.5.

Chadwick (2006) conducted a two part study that used 24 participants to operate simulated UGVs in two separate experiments. The experiments were similar in design in that the author manipulated operator capacity (one, two, or four UGVs) within subjects for both. The difference between the two experiments was the way the interface that the operator was using to control the UGVs was manipulated. The metrics used to evaluate performance were: monitoring performance, response time, and target localization. The combined results due to manipulating operator capacity are explained below.

The results showed that monitoring performance of the UGV state, target localization accuracy was significantly lower when operator capacity was increased. Alternatively, response time was significantly higher when operator capacity was increased. This result was true for two different categories of response time: target response (either valid or invalid), and response to robot navigation errors (simulated by the robot being stuck in a loop) (Chadwick, 2006).

The study concludes that increasing operator capacity increases the resource demands on an operator. Performance degradation can be mainly attributed to the operator's ability to respond to queries presented by the UGV, and as operator capacity increases the number of queries increases. The attention limitations of the operator can be minimized by the interface design or by implementing the correct type of automation (Chadwick, 2006). The second study that incorporated interface design into their test procedure did not notice the same performance degradation as operator capacity increased, however, they were able to conclude that there was a significant interaction between interface design and operator capacity (Baber et al., 2011).

Baber et al. (2011) split their study into two simulated UV experiments. The goal of the first experiment was to determine the impact of multimodal control on UV performance, by varying the control modality (speech, gamepad, multimodal) and type of UV (UAV or UGV). The goal of the second experiment was to determine if the benefits of multimodal control hold true when operator capacity is increased, by varying the control modality (speech, gamepad, multimodal) and operator capacity (one, three, five). The participants were required to control UV sensors and classify targets, while responding to system warnings when necessary. The same 16 participants were used in both experiments and the metrics used to evaluate them were: number of commands issued, distance to target when command was issued, time to respond to warnings, and subjective workload (NASA-TLX). The operator capacity results of experiment two are discussed below.

The study showed that there was a significant effect of operator capacity on the total number of commands issued by the operator, there was also a significant interaction between operator capacity and modality. The number of commands decreased as operator capacity increased but this decrease was not the equivalent across all modalities. Increasing operator capacity did not have a significant effect on the distance from the target when the commands were issued (Baber et al., 2011). A more detailed discussion of the modal results will be discussed in 3.5.5. Time to respond to warnings were not significantly affected by operator capacity but there was an interaction between operator capacity and control modality. This shows that control modality has different effect on time to respond to warnings depending on the participant's operator capacity condition. In contrast to the operator's ability to respond to warnings in all levels of operator capacity, a significant difference in subjective workload was observed. The subjective workload ratings increased as operator capacity increased (Baber et al., 2011), which is a similar result observed by Adams (2009).

The two articles described above both concluded that operator capacity had a significant interaction with interface design. This result is consistent with articles in previous sections that concluded that interface design had an impact on performance results. The two articles were unable to show the same trend in performance data, but as previously stated, the difference in experimental design makes it difficult to directly compare results between studies. .

3.4.4 Operator Capacity Model Development

This literature review has identified that many factors are involved when determining the operator capacity that will maximize UV performance. One major factor is the mission task that needs to be performed. Research has focused on developing a model that could be used to predict the operator capacity, while accounting for automation (Crandall & Cummings, 2007).

Crandall and Cummings (2007) presented their model as a set of metric classes for human-robot teams and start by identifying three attributes that a set of metrics must have: 1) identify both the human and robot's limits, 2) be able to predict how changes in environment, mission and team members will affect performance, and 3) contain key performance parameters (KPPs) that will indicate the overall effectiveness of the team. To derive a set of metrics with these capabilities, the authors broke down general human-robot teams into relevant subparts and developed measures for each subpart. The measures were combined to predict overall team effectiveness. The capability of the set of metrics was then tested through a user study.

To break down the control loop of human robot teams, the authors specifically looked at the supervisory control condition. They started by defining the two controls loops present with an operator capacity of one UV: human-robot loop and robot environment loop. Within the human robot loop, the human processes information provided by the UV via the human-robot interface and provides an action for the robot to take, once again using the human-robot interface. Within the robot-environment loop, the robot combines the input data from the human with its own sensor data in order to decide how to act. When operator capacity is greater than one, the two loops form the base of the framework; however, now the human needs to divide attention between multiple human-robot loops. Therefore the human must first select which human-robot loop to be a part of before they can receive information and make a control decision via the human robot interface. New links are also presented between the robots that represent their ability to communicate with each other (Crandall & Cummings, 2007).

The functions of the three classes of metrics are: 1) measures the effectiveness of the human-robot loop, 2) measures the robot's autonomous capabilities, and 3) measures the efficiency of human attention allocation. The derived metrics that accomplish these functions are called: 1) interaction efficiency (IE), 2) neglect efficiency (NE), and 3) attention allocation efficiency (AAE). The three metrics can be estimated by the following parameters: 1) interaction time (IT), 2) neglect time (NT), and 3) wait time (WT) (Crandall & Cummings, 2007).

The user study conducted to assess the defined metrics implemented a simulated UGV task where the 12 participants were tasked with removing as many objects as possible from a maze. There was an 8 minute time limit set to each trial and the participants were supposed to have their entire robot team out of the maze when time was up. Each participant ran at least one trial with each of the four operator capacity conditions (two, four, six, and eight). The participants were evaluated based on a performance score, which was calculated as the difference between objects collected and robots left in the maze when time had expired (Crandall & Cummings, 2007).

The results from the user study show that as operator capacity increases, the number of objects collected also increases. The number of objects collected was significantly lower in the two robot condition, compared to the other three conditions. The four robot condition was significantly lower than the six and eight robot conditions and there was no statistical difference between six and eight.

When the authors looked at the number of robots left in the maze they found that there was a significant difference between the two and four robot groups and the six and eight robot groups. They used the combination of these results to conclude that the highest performance for this user study was seen when operator capacity was between four and six. This represents the observed fan out (FO) for this study (Crandall & Cummings, 2007).

In addition to calculating FO, the user study was used calculate the three parameters that will be used in model. IT, NT and WT were calculated for each trial and an average was given for each of the operator capacity conditions. IT decreased as operator capacity increased, which represents the limits of the human. The largest decrease was between the two and four robot conditions, followed by the difference between four and six robot conditions and there was almost no difference between six and eight. NT increased as operator capacity increased, which represent the robot's limits of how much time the operator can pay attention to it. The largest difference was again found between the two and four robot conditions. The difference steadily declined as operator capacity increased. WT showed a similar pattern to NT, such that it increased as operator capacity increased; however, the difference between conditions was relatively consistent. The resultant IT, NT, and WT values were substituted into FO equations previously used in literature and compared against the FO results concluded from the user study. The authors observed that none of the equations extracted from the literature were successful in predicting the true FO observed, or predicting the robot team effectiveness.

One of the study goals was to find a metric that would represent KPPs, and it was observed that a fraction of IT could accomplish that goal. Furthermore, the authors concluded that minimizing that KPP would result in an increased overall performance. In summary, the metric classes introduced (IE, NE, AAE) were successful in measuring human and robot limits while providing KPPs. However the metrics were not successful in predicting the FO or team effectiveness (Crandall & Cummings, 2007).

The performance results from studies investigating the effect of operator capacity on UV operations were inconsistent, but they were able to identify the design variables that impact performance when operator capacity is increased: automation level and reliability, and interface design. Due to the complex interaction between these variables, developing a model to predict the optimal operator capacity for UV operations is challenging. Further research in optimizing HRI needs to improve the implementation of automation and the human-computer interface.

3.5 Interface Design

As noted in the previous sections, interface design has been a common point of criticism for declines in performance when automation is unreliable and operator capacity is increased. Information perception of the operator through the interface design of a UV system is essential for optimal performance.

Information perception is an area of research that has been given much attention in the research literature. In reference to the operation of UVs, it mainly concerns the design of the human-robot interface. The interface can be broken down into two categories: display and control. The three main modalities of display and control are visual, tactile, and audio. The metrics of SA, workload, and performance are commonly used to evaluate effectiveness of each modality. Researchers have investigated each modality separately and in combination (multimodal).

Display interfaces represent information from the robotic system to the operator. The general objective of a display is to optimize the information being relayed to the operator and thus to optimize the operator's subsequent task performance.. This has traditionally been done by manipulating display characteristics such as display size, colour, and shape. Audio and tactile displays, in combination with the main visual display, have been explored in literature for the purpose of increasing the effectiveness of the communication link between the robotic system and operator (Maza, Caballero, Molina, Pena, & Ollero, 2010).

Controls represent the information flow from the operator to the robotic system. Examples of intermediate devices that can be included in a control system are: mouse, keyboard, touchpad, joystick, steering wheel, etc. Recent research has focused on removing intermediate devices and instead implementing more direct interface methods. One method is a touchscreen on the visual display, which would allow operators to directly engage with the interface. Another method is through a different communication channel, such as speech recognition or body movement (Jansen & van Erp, 2010; Maza et al., 2010). Advancements in control technology are focused on optimizing the flow of information from the operator to the system.

A third information flow that has received attention in the literature is the robotic system learning the operator's state. Important characteristics of operator state include body position and orientation as well as physiological measures (Maza et al., 2010; Hou & Kobierski, 2006, Hou, Zhu, Zhou, & Arrabito, 2011).The system can use its knowledge of the operator state to optimize how it relays information to the operator, which in turn can reduce the perceived workload of the operator and increase overall task performance (Maza et al., 2010; Hou & Kobierski, 2006, Hou et al., 2011)..

3.5.1 Visual Display Aids and Control

The more traditional method for enhancing information perception through a human-computer interface is by modifying the visual display. Different modification methods have been used with the common goal of increase SA while minimising workload and thus optimising performance. In correlation with previously discussed criticism of current interface design, Cavett, Coker, Jimenez, and Yaacoubi (2007) state that a new interface design is needed and that the objective of the interface should be to reduce cognitive workload, increase situation awareness, and increase operator performance, while in control of multiple UAVs. Two example display design methodologies proposed to accomplish this goal are: one which mimics how humans navigate in the world and another that uses two information relay strategies, synchronous and asynchronous (Cavett et al., 2007; Chappell, 2007). Empirical studies to determine the performance benefits of different interface designs have been able to show that different display techniques can improve UV operator performance. Gunn et al. (2005) tested cognitive (symbolic representation) and sensory (change in physical attributes) warning display for UAV target recognition, and found performance improvements with the sensory display format. A method of synthetic overlays that used picturing-in-picture (PIP), found that some form of PIP resulted in better performance results than no PIP, and furthermore 33% PIP was preferred by the participants (Calhoun, Ruff, Lefebvre, Draper, & Ayala, 2007). Two methods for providing visual cues (no cognitive countermeasure and with cognitive countermeasure) was tested and performance benefits were seen in the cognitive countermeasure group (Dehais, Causse, & Trembley, 2011).

Chappell (2007) based their control concept on the basic human ability to move within the world through a process of seeing what is around them, determining which direction they need to go, and moving in that direction. The human control concept led them to develop a UAV navigation system of point-and-click, which could be used in three dimensions. The interface for the control system consisted of three main displays: forward view, similar to the view from a cockpit, overhead map view, and a sensor view. Through these views the operator was able to provide navigation points in three dimensions. The traditional horizontal navigation points were given in the map view, while the forward view can be used to specify the navigation point in the lateral, longitudinal and vertical positions. The sensor view allows the operator to view their surroundings and promote a higher situation awareness. The three views were also necessary in order to effectively provide the user with instant feedback about the UAVs new commanded position and status of the UAV as it travels to that position. The authors state that the interface they developed helped with UAV control issues due to excessive control delay by simplifying the control process and providing important cues of the duration and variability delays as they occur.

The control method discussed by Chappell (2007) is highly dependent on effective visual displays and visual cues of important information. Two methods for presenting information to the human operator through the human-computer interface are synchronously and asynchronously (Cavett et al., 2007). Two display designs were presented for multiple UAV operations; the design process started by outlining the human-computer hierarchy and subsequent flow control system diagram (Cavett et al., 2007).

The human-computer value hierarchy provides a breakdown of the value that has been given to different components of the human performance and system performance. The total performance of the human-computer system is 50% dependent on human performance and 50% dependent on computer system performance. The authors note that for the interface design, focus should only be given to maximizing human performance, because system performance is dependent on the technology of the system. The components of human performance and their assigned weightings are: user accuracy (50%), user processing time (30%), user satisfaction (10%), and training time (10%) (Cavett et al., 2007).

Cavett et al.'s (2007) system diagram represents the information flow and relationships between the human-computer interface and the rest of the system. The additional components of the system are the user/operator, UAV, command centre, and computer hardware. The primary function of the human-computer interface is to act as the link between the operator and the UAV. Through this link, the user can navigate the UAV, control and receive data from the UAV payloads and monitor the mission and UAV health status (Cavett et al., 2007).

The authors used the two aforementioned design components two develop two different interface design alternatives. The designs were based on two different formats to relay information to the operator: synchronous and asynchronous. During the synchronous format all information about all the UAVs under the operator's control is made continuously available. During the asynchronous format, information about the UAVs is made available through alerts and pop-ups when the system deemed it necessary to have the operator's attention (Cavett et al., 2007).

Cavett et al. (2007) describe an iterative evaluation procedure that could be implemented to assess and refine the interface design. Each interface would be evaluated during a series of simulated UAV mission trials, where the participants would be evaluated on the follow performance metrics: skill development time, number of errors, and completion time. Subjective usability, SA, and workload ratings will also be used to analyse each interface.

Asynchronous display strategy requires important information to be cued to the operator, which can be done by either sensory or cognitive formats (Gunn et al., 2005). Sensory display formats use a change in physical attributes to represent information. Cognitive display formats use a symbolic representation to represent information. Gunn et al. describe cognitive displays to be more robust than sensory displays, however, they also state that links have been drawn between cognitive displays and higher workload. As previously discussed this could lead to reduced SA and task performance.

Gunn et al. (2005) conducted a simulated UAV target recognition study with sixteen participants to examine the effects of display format (sensory or cognitive), scan rates (slow and fast), cue modality (no cueing, visual, spatial audio, and haptic), and target location (between 60° and 150° or -60° and -150°, from centre of display). The participants were evaluated on threat warning detection rates, target acquisition time, and perceived workload (NASA-TLX). The display format, visual cueing, and target location results will be discussed in the following paragraphs. Spatial audio results will be discussed in section 3.5.2 and haptic results in section 3.5.3.

Display format results from Gunn et al. (2005) show that significantly more threat warnings were detected and fewer false alarms occurred in the sensory display format than the cognitive format. The sensory format also led to significantly lower target acquisition times than the cognitive format. Perceived workload was significantly higher in the cognitive format condition than the sensory condition, which was expected based on the literature summarised by Gunn et al.

Visual cueing and target location results showed that there was no significant effect on the amount of threat warnings detected by the operator, however there was a significant effect on target acquisition times. Acquisition times were lower in the visual cued condition than the no cueing condition and when targets where located closer to the centre line of the operator field of view. Also found was a significant interaction between target location and visual cueing, such that acquisition times decreased more in the visual cued condition as target location moved closer to the centre line. In other words, acquisition times were more sensitive to a change in target location in the cued condition compared to no cueing. There was no significant effect of visual cueing or target location, on perceived workload (Gunn et al., 2005).

It has been hypothesised that the combination of computer generated data and live video imagery will benefit UAV missions (Calhoun et al., 2007). Calhoun et al. summarised other work that has shown various methods and applications of synthetic overlays to improve situation awareness and performance by improving limiting factors such as: narrow field-of-view, data transmission issues, and challenging operating environment. One method of synthetic overlays that was outlined by the authors was "picture-in-picture" (PIP), where the synthetic display has an inset box that displays the real time video data. The two displays need to be tethered such that they are showing the same geo-referenced data.

Calhoun et al. (2007) conducted a simulated UAV study to determine if the benefits of PIP were dependent on the size of the overlaid display and accuracy of the computer generated landmarks. Twelve participants were evaluated in a 3x2x2 within-subject study, where there were three PIP conditions (none, 50%, 33%), two levels of registration error (low, high), and repetition (1-2). During the trial they were not responsible for any navigation of the UAV, but they manually manipulated the camera viewpoint to locate a specific landmark marked by a synthetic flag. Subjective situation awareness, workload, and task difficulty ratings were recorded after each trial and the operator performance was measured based on landmark designation time.

Performance results showed that designation time was significantly longer in the no PIP condition, compared to two other conditions. Although 33% PIP did result in lower average designation time than 50%, there was no statistical significance. There was a significant interaction between registration error and PIP condition such that the no PIP condition was more sensitive to higher errors, and therefore resulted in higher designation times (Calhoun et al., 2007).

Post-trial subjective results showed that PIP condition had a significant effect on situation awareness, workload, and task difficulty ratings. In all three of the rating scales, participants rated the no PIP condition to be worse than the two PIP conditions. Worse equates to higher perceived workload and task difficulty, and lower SA. There was no significant difference found between the two PIP conditions (Calhoun et al., 2007).

Post study subjective results showed that the participants rated the PIP conditions significantly different in terms of speed to find landmarks and workload required to find landmarks. In both cases they rated the 33% PIP condition to be the best. Furthermore, when asked what PIP condition they preferred, all of the participants chose PIP-on in the low registration error condition and all but one chose PIP-on in the high registration error condition; suggesting the benefits of PIP are not dependent on registration accuracy. There was no significant difference between the two PIP-on conditions (Calhoun et al., 2007). In conclusion PIP was an effective display aid that could be included in future UV interface design.

Another method for presenting important information to the operator asynchronously is through cognitive counter measures (Dehais et al., 2011), defined as the temporary replacement of the main visual display with a visual cue. The objective is to avoid attentional tunnelling which can lead to UGV operators missing cues for secondary tasks.

To test the validity of cognitive countermeasures, a study was conducted with 23 participants in a lab environment, where their objective was to perform target localization and identification tasks with real highly automated UGVs. While the participants were engaged in the target identification task, the authors induced a low battery event that would automatically route the UGV back to base. For the no countermeasure condition, three visual cues were presented on the display when the event was induced. For the countermeasure condition, the main operating view was temporarily replaced by a visual cue. The only variable introduced in the study was the countermeasure condition (yes and no), and participants were placed in one of two groups. During each trial participants were evaluated on decision making, eye movement, and heart rate (Dehais et al., 2011).

The decision making results showed that only 33.33% of participants in the no countermeasure group let the robot go back to base when the low battery event occurred. The remaining participants continued with the identification task and declared they did not notice the cues that appeared on the display. All of the participants in the countermeasure group noticed when the low battery event occurred and only one participant did not allow the robot to return back to base (Dehais et al., 2011).

Psychophysiological results showed that the no countermeasure group had a significantly higher average heart rate when the low battery event was induced, compared to the countermeasure group. The no countermeasure group also spent a significantly higher percentage of time with their eyes focused on the main video display, compared to the countermeasure group. This indicates that the countermeasure was successful in reducing attentional tunnelling. This result was confirmed by the number of scanned areas of interest (AOIs) and gaze switching rate. For both metrics there was significantly higher result in the countermeasure group compared to the no countermeasure group (Dehais et al., 2011).

The articles summarized above provided design strategies and examples of empirical studies that have shown improvements in UV performance through display interface design. Research has shown that different views might be desirable and one method for accomplishing this is through mixed UGV and UAV operations.

3.5.1.1 Visual Display for Mixed, UGV and UAV Operations

The concept of mixed UGV and UAV operations has already been discussed The following section reviews articles that have investigated mixed operations for the purpose of determining that additional UAV views can act as a display aid for the UGV operator. Optimal interface design for mixed operations is a complex problem that includes the compound factors of the mission task, the type of UVs that will be used, and the operator capacity (Sterling & Perala, 2007). Chadwick (2005) took a more basic look at the validity of using a UAV view to aid UGV operations, for varying numbers of

UGV teams. Another study went beyond just determining the validity of mixed operations to identify that time lag and frame rate are important variables to understand (Chen, Durlach, Sloan, and Bowens, 2008). Chen (2010) developed a simulated MOUT reconnaissance testbed to investigate the use of different UAV view types to assist with UGV operations.

Chadwick (2005) also looked at the benefits of mixed UGV and UAV operations. The study required 61 participants to manually operate miniature UGVs in a lab simulated urban search and rescue mission. The authors conducted the study in a 1:12 scaled urban terrain model, consisting of a maze of passageways filled with obstacles. Each participant was randomly placed in one of four conditions: 1) one UGV, 2) one UGV with UAV view, 3) two UGVs, and 4) two UGVs with UAV view. The UAV views were provided by monochrome video cameras that were stationary and each covered one third of the terrain. The participants were evaluated on number of targets found, target localization error, the number of robot faults, and workload (NASA-TLX). A robot fault was deemed to be any time that the UGV required participant intervention to continue (i.e., when a roll-over occurred).

The results from Chadwick (2005) showed that there were only two statistically significant main effects. One effect was the number of UGVs under control when looking at the number of targets found, such that the single UGV conditions resulted in more targets being found than the double UGV conditions. The second effect was the number of UGVs under control when looking at the number of faults per robot, such that the double UGV conditions resulted in fewer faults than the single UGV conditions. Although not statistically significant, the trend in the data showed that the localization error was lower and workload was higher in the double robot conditions.

The study showed no statistically significant differences in any of the results with respect to the addition of UAV views. Although not statistically significant, the trend in the results show that the addition of the UAV views increased the number of targets found in the double UGV condition, decreased the localization error, and decreased the total number of robot faults in the double UGV condition.

In conclusion, the addition of a UAV view to assist with the mission task was more beneficial than the addition of a second UGV. Similarly to other articles discussed, the benefit of adding views to assist the participant is dependent on improvements with interface design and automation (Chadwick, 2005)

In contrast to Chadwick's findings, Chen et al. (2008) found that increased workload and decreased performance under simulated mixed UV conditions. They hypothesised that time lag and low frame rates would cause significant performance differences. This result was not found; however, subjective ratings from the participants showed that time lag was less ideal than low frame rate (Chen et al., 2008).

The study looked at simulated mixed UGV and UAV parallel route reconnaissance missions. The researchers not only compared the difference between operating each platform independently and in parallel, but they also investigated the effects of display time lag and frame rate. A 250 ms time lag was induced to half of the participants and the other half had a low (5 Hz) frame rate induced. Participants were required to complete four trials, one while using each platform (semiautonomous UGV, semiautonomous UAV and, manual UGV) and one while all three platforms in parallel. The trials were evaluated based on task completion time, target detection, workload (NASA-TLX), simulator sickness, and usability (Chen et al., 2008).

The study results showed that task completion time was significantly affected by platform condition. The mixed condition resulted in a significant smaller percentage of participants who completed the mission in the allotted time when compared to the other three single platform conditions. There was no significant difference found between the three single platform conditions (Chen et al., 2008).

The total elapsed time required to complete the mission was significant less for the manual UGV condition when compared to the other three conditions. When the remaining three conditions were compared, the UAV condition was significantly less than the mixed condition, but not different than the semiautonomous UGV condition. Furthermore there was no difference between the mixed and semiautonomous UGV conditions (Chen et al., 2008).

The study results showed a significant impact of platform condition on number of targets lased, with the manual UGV condition being the lowest. Manual UGV control also resulted in the lowest target detection per minute value, which was used to normalize the data since manual UGV resulted in the shortest mission completion times. There was no significant difference found between the platform conditions when looking at the number of friendly lases (Chen et al., 2008).

The results showed that perceived workload was significantly affected by platform condition. The mixed platform condition resulted in significantly higher ratings than the three single platform conditions and there was no difference between the ratings of the three single platform conditions. There was no significant difference found between the platform conditions when the participants were asked to rate their simulator sickness; however, the trend in data showed the mixed condition to be the highest. The authors concluded that the simulation used for all four conditions rarely resulted in severe symptoms.

There was no significant effect of frame rate or time lag found, when looking at target detection. The authors anticipated that there would be a significant effect of time lag and frame rate on target detection, and they explain that the lack of significance could be due to a limiting analytical procedure (Chen et al., 2008).

The authors concluded that under these test conditions there was no benefit of mixed UV operations. Furthermore, improved performance will likely not be achievable until automation implementation and interface design are improved. The study suggested that participants did not take advantage of the additional views available to them because the workload of operating additional UVs was too much for them to handle. Even though there was no significant difference found between time lag and frame rate conditions, the participants perceived a significant difference, evident from the subjective usability questionnaire. In general, the time lag group ratings were less ideal than frame rate groups (Chen et al., 2008).

When the authors investigated the strategies employed by the participants in the mixed condition, they found two interesting results in relation to control and task division. First, they found that 75% of the participants employed a main strategy of sending the UAV out in front of the other platforms. In general participants would search the route with the platform they were most comfortable with and simply steer the rest of the vehicles to the endpoint with very little additional searching. Second, participants used the UAV more predominantly to lase targets. The authors found little evidence that participants employed a strategy that coordinated movements or task completions between platforms (Chen et al., 2008).

Chen (2010) conducted another mixed UV military reconnaissance study that investigated the benefit of providing a UGV operator with additional UV views. This study was conducted on a simulated MOUT testbed, and combined a simulated UGV view with simulated UAV views. The mission task was to find and navigate to a primary target (SUV) and then find and navigate to five secondary targets (enemy soldiers). The participants were evaluated on mission performance, map marking accuracy, landmark location test score, and perceived workload (NASA-TLX). The study set-up varied lighting conditions (night and day) between subjects and primary target type (stationary or moving), and UAV view type (none, micro aerial vehicle (MAV), UAV fixed view, UAV orbiting view). The MAV view simulated a UAV flying at low altitudes and therefore did not provide a complete bird's eye view of the entire MOUT area. The two UAVs were from higher altitudes and did provide a complete bird's eye view of the MOUT area.

The mission performance results from Chen (2010) show that there were no main effects of primary target type or lighting on target search times, but there was a significant effect of UAV view type on the target search times. The no UAV view condition resulted in the lowest search times when compared to the other three conditions. MAV view condition was significantly lower than both of the other two UAV view conditions, and there was no difference found between the two high altitude UAV conditions. Map marking accuracy was significantly affected by both UAV type and primary target type. No UAV view resulted in significantly lower marking accuracy than the MAV, and both UAV view conditions. MAV view condition was significantly lower than both of the other two UAV view conditions, and there was no difference found between the two high altitude UAV conditions.

Workload was significantly affected by both UAV view type and primary target type. The orbiting high altitude UAV view resulted in a higher perceived workload compared to the MAV and fixed UAV view conditions. The data shows that workload was higher when the primary target was in the moving condition.

Increased performance observed under the high altitude UAV condition and the lack of increased workload under the same condition indicates that mixed operations could be beneficial to some military operations (Chen, 2010). The inconsistent results found in visual display research demonstrate again that the many factors involved in UV studies make finding unified results a challenge. Improvements in testing technology, automation, and interface design capabilities will likely improve the eventual benefits seen by adding visual stimulus to an interface display.

3.5.2 Audio Aids and Control

The projected benefit of using additional modalities other than visual is based on multiple resource theory, such that offloading information to other resource channels is expected to reduce workload, increase SA, and ultimately increase UV operator performance (Kaber et al., 2006; Wickens and Dixon, 2006). One component of the experiments previously described by Wickens and Dixon (2006) that has not yet been discussed is their use of auditory cueing. They were able to conclude that auditory cueing provided performance benefits no matter what the operator capacity (Dixon & Wickens, 2003). The challenge of using auditory cueing in military environments has been investigated, and the effects of two types of auditory signals: discrete and continuous. Continuous cue signals produced better performance results than discrete cues (Donmez, Cummings, & Graham, 2009; Graham & Cummings, 2007). Spatial audio cuing has been proposed to not only alert the UV operator when something needed their attention put also provide them with a location on the visual display that they needed to focus on (Maza et al., 2010).

Section 3.3.1 discussed several experiments by Wickens and Dixon (2006), now experiment 1 and 2 will be discussed in more detail. The authors conducted a study in a simulated UAV environment, where the operators were primarily responsible for keeping the UAV on track through a series of waypoints while they reported on command targets (CTs). There were two secondary tasks, one in which the participant was required to search for "targets of opportunity" (TOO) and another in which they were required to detect and respond to system failures (SF). In each experiment the operators tested three conditions: baseline, auditory offload, and flight automation. In the baseline condition the participants were required to manual fly the UAV, while searching for targets and monitoring the system status. In the auditory offload condition the mental workload of monitoring the system status was alleviated by auditory cueing, and auditory mission instructions. In the automation condition the participant was no longer responsible for manually flying the UAV; instead they took a supervisory roll which only required them to enter coordinates. The difference between the two experiments was that in experiment 1 the participants were required to operate one UAV and in experiment two they were required to operate two UAVs. During the trials participants were evaluated on: flight course tracking error, the number of times they needed mission instructions repeated, TOO detection rates, TOO response time, SF detection rates, and SF response time (Dixon & Wickens, 2003).

Experiment 1 results show that auditory offloading had no impact on tracking error when compared to the baseline condition. The authors predicted that auditory offloading would improve performance, based on multiple resource theory. Although there was no evidence in the tracking performance, the authors observed performance improvement with respect to: the number of times instructions needed to be repeated, target monitoring, and system monitoring. Auditory offloading resulted in a lower average number of times that mission instructions needed to be repeated, indicating better parallel processing than the baseline condition. There were significantly more targets detected and system failures detected in the auditory offloading conditions when compared to the baseline condition. The authors suggest that, because participants were able to respond to system failures more quickly, they were in turn able to spend more time searching for targets (Dixon & Wickens, 2003).

Experiment 2 results show that once again auditory offload had no impact on tracking error, and the addition of a second UAV also had no impact. Auditory offloading allowed for similar performance improvements as experiment 1 with respect to number of times instructions needed to be repeated and system monitoring. Therefore, auditory offloading will provide performance benefits for these two tasks regardless of the operator capacity. In contrast to experiment 1, there was no benefit to auditory offloading with regards to target monitoring (Dixon & Wickens, 2003). The authors were unable to provide an explanation for this finding. To expand on previous research that has shown the benefits to auditory cueing, two studies investigated the type of auditory signal that should be used for certain UV events. They were both conducted in the same lab and had similar results, such that continuous cues were beneficial in alerting both UAV course deviations and UAV late arrivals (Donmez et al., 2009; Graham & Cummings, 2007).

In the studies, the two cues types (continuous and discrete) were used for two events during the UAV task: course deviations and late target arrivals. The tests were conducted in a simulated UAV testbed and the objective was to ensure that the correct targets were engaged and that the UAVs returned to base within the allotted time. There were four cue conditions that were a combination of course deviation cues (discrete or continuous) and late arrival cues (discrete or continuous). Participants performed two trials, one with each operator capacity condition (single and multiple), with one of the cue combinations. During the trials, participants were evaluated on the number of missed course deviations and late arrivals, reaction time to correct course deviations and late arrivals, number of responses to the secondary task of monitoring radio chatter, and perceived workload (NASA-TLX) (Donmez et al., 2009; Graham & Cummings, 2007).

The results showed that there was not enough missed course deviations and late arrivals for this parameter to be included in the results. Continuous course deviation cues resulted in significantly faster course deviation reaction times than discrete course deviation cues. Although not statistically significant, the data trended towards the multiple UAV condition resulting in higher course deviation reaction times than the single UAV condition (Donmez et al., 2009; Graham & Cummings, 2007).

The late arrival response time results show that there was no significant main effect of late arrival cue type. There was, however, a significant main effect of operator capacity, interaction between operator capacity and course deviation alert type, and interaction between course deviation and late arrival alert types. The interaction between course deviation alert type and late arrival alert type is represented but the fact that the only combination of alert types that had a significant effect on late arrival response time was continuous course deviation alerts and discrete late arrival alerts. This combination resulted in significantly lower response to late arrivals. The main effect of operator capacity and its interaction with course deviation alert type are represented by the fact that in the single UAV condition there was no effect of course deviation cue type on reaction time for late arrives but in the multiple UAV condition, continuous course deviation alerts lead to significantly longer reaction times when compared to discrete course deviation alerts. Therefore, the authors suggest that continuous course deviation alerts only interfered with late arrival response when the operator was controlling multiple UAVs (Donmez et al., 2009; Graham & Cummings, 2007).

In conclusion, continuous audio signal type was beneficial to UAV performance for both single and multiple UAV conditions. The authors suggest that discrete alerts can easily be missed in noisy military environments, where continuous cues will not turn off until the problem is attended to. Furthermore, continuous cues did not have an impact on monitoring radio chatter, which is a common scenario in military UAV environments. This was supported by the subjected workload results that showed that the participants did not find a significant effect of any cue type. One concern raised with continuous alerts is that they might be fatiguing, but further investigation is needed to determine the impact of fatiguing on performance (Donmez et al., 2009; Graham & Cummings, 2007).

Maza et al. (2010) proposed that other audio signal types could be used to assist the UV operator with understanding the information being presented. Two examples are spatial audio cueing and speech synthesis. Spatial audio cueing is suggested to provide benefits over traditional audio cueing because response times will be shortened. With spatial audio cueing the participant is given an indication of where on the visual display their attention is needed and thus time is not wasted scanning for the warning. Speech synthesis is suggested to assist by relieving the requirement that the UV operator must read information from the visual display. These two signal types were tested in a simulated UAV environment, as part of a study looking at several cue modalities (Maza et al., 2010)

Different combinations of modal conditions (visual, audio, and tactile) were tested in a simulated UAV control interface. The operators were not performing a UAV task, but were required to respond to queries. They were evaluated on reaction time and accuracy of their responses. The initial experiments were done to determine the best control type, mouse or touchscreen. The results Showed that touchscreen was better for this specific task and therefore it was as the base control type for the subsequent tests that altered cue type (speech synthesis, 3D audio, tactile, and 3D audio plus tactile). Tactile results will be discussed in section 3.5.3 and multimodal results will be discussed in section 3.5.4 (Maza et al., 2010).

The speech synthesis and 3D audio experiments demonstrated that audio cueing provides performance improvements over no cueing (decreased response time). The speech synthesis condition would simply alert the operator when they needed to attend to a task, and the operator then had to find the message on the interface display. This type of cue was beneficial because the operator was no longer required to continuously scan the visual display for messages; they only needed to start scanning when the audio alert was heard. The 3D audio experiment provided spatial audio information such that not only was the operator alerted when a message needed attention but the location of the message was communicated based on where the audio alert was generated from. This increased operator performance from the speech synthesis condition because it took away the requirement that the operator had to scan the entire display for the message. The authors found that when the operators heard the spatial audio message they would immediately focus their attention in the correct direction (Maza et al., 2010).

The resource offloading from visual to audio has proven to be beneficial for UV operators and there appears to be specific audio signal types that benefit specific tasks. Resource offloading can be done in multiple ways and needs to be customized to the required task. When audio and visual resources are being occupied, tactile and haptic displays could provide addition resource offloading.

3.5.3 Tactile and Haptic Display and Control

Tactile and haptic force feedback have important differences relevant to display and control. Tactile feedback is a static vibration that is used as a stimulus to assist the performance of a task. Haptic feedback is the combination of force and vibrational feedback that is used to represent the forces experienced by the robot in its environment. Tactile and haptic displays have been discussed in literature to assist during teleoperation and the choice of which method to use is dependent on the specific task at hand (Jones & Sarter, 2008).

Tactile displays can be either vibrotactile or electrotactile, where vibrotactile sensors mechanically stimulate the skin and electrotactile sensors pass current through the skin and stimulate nerve fibres. Electrotactile sensors have been criticised on their application because of issues that arise when skin contact is not maintained and a comfortable stimulation range is user dependent. Vibrotactile displays have been discussed as being more applicable in industry applications such as cue types for UV operators (Jones & Sarter, 2008).

One common application of tactile displays is in situations where an operator needs assistance with spatial orientation and navigation when stable reference frames are absent. For this application, torso-based tactile vibrotactile systems have been used to successfully present an aircraft operator with the information required to control the vehicle and maintain spatial awareness (Jones & Sarter, 2008). In addition, cues were used to alert the operator of important information within their environment. Similarly, torso based vibrotactile systems have been successful in presenting navigation cues to someone walking or operating a vehicle in an unknown environment. In virtual environments vibrotactile displays have used to relay important information about the virtual environment such as object collisions (Jones & Sarter, 2008).

Jones and Sarter (2008) discuss how tactile displays have proven to be beneficial to multimodal display design. They suggest that such displays are predominant in applications where visual and audio channels are under high demand. Tactile displays allow information to be presented to an operator that would otherwise be lost or delayed. In addition to the relay of information, tactile cues can be used to attract attention to a task that needs intervention (Jones & Sarter, 2008; Donmez, Graham, & Cummings, 2008).

A study by Maza et al. (2010), also described in section 3.5.2 , looked at the application of multimodal interface designs for UAVs by testing different combinations of modal conditions (visual, audio, and tactile) in a simulated UAV control interface. 3D audio results have been discussed in section 3.5.2 and multimodal results will be discussed in section 3.5.4 .

The tactile experiment used three vibrating remotes, one located on the left, right, and centre of the body. The remote that vibrated was dependent on the location of the warning that needed to be responded to. This approach employs a similar theory to spatial audio cueing, and when the results from tactile cueing were compared to spatial audio cueing the authors found that the result were equivalent. Both cueing modalities showed improvements over the no cueing condition. The results suggest that spatial cueing allowed the operators to not pay attention until there was a cue, and when the cue was initiated the operator would immediately direct their attention to the location of the warning. There was no need to scan the visual display (Maza et al., 2010).

A study conducted by Gunn et al. (2005), described in section 3.5.1 also looked at different modes of cueing for a simulated UAV target acquisition task. They describe a haptic cueing method that used force feedback in the operator control stick to guide them towards the target on the display interface. Their results showed that haptic cueing had a significant effect on target acquisition time and had a significant interaction with target location. Haptic cueing reduced acquisition times when compared to the no cueing condition. Also found was a significant interaction between haptic cueing and target location, such that acquisition times decreased when targets were closer to the centre line, at a greater rate in the haptic cueing condition when compared to the no cueing condition. The authors were unable to find a significant effect of haptic cueing on target detection rates and perceived workload.

Donmez et al. (2008) looked at the application of two haptic displays: continuous and threshold. Thirteen participants were tested in a simulated multiple UAV supervisory control mission. Haptic feedback was used to cue two events during the mission, late arrivals and course deviations. The feedback for each event was presented by a pressure vest and vibrating wrist band, respectively. The only variable that was manipulated between trials was the feedback type (continuous or threshold). The metrics used to evaluate the participants were: course deviation reaction time, late arrival reaction time, audio call sign recognition, and perceived workload.

The result showed that continuous feedback resulted in significantly faster reaction times during course deviations and significantly slower reaction times during late arrivals; however, there was no significant difference in the audio call sign recognition rate or perceived workload. The results suggest that continuous haptic feedback is more suitable for continuous events and threshold feedback is more suitable for discrete events. The authors note that they are unable to determine the general validity of including haptic display in an interface design because they did not test a baseline condition (no cueing) (Donmez et al., 2008).

In conclusion, tactile and haptic display aids have proven to provide performance benefits to UV operators. Similarly to audio displays, the signal type that will maximise performance is dependent on the mission task (Donmez et al., 2008; Gunn et al., 2005; Maza et al., 2010). For future interface design it also important to understand what the effects are of combining different cue modalities.

3.5.4 Multimodal Display and Control

Multimodal display aids are characterized by the combination of two or more of the modes previously described. The same theory behind using a single modal cue to alleviate visual resources when under high demand to improve performance is used to predict that multimodal cueing will further improve performance when two resource channels are under high demand. Multimodal cueing is also a strategy that is used to present redundant information, which will improve the performance advantages already discussed when using modal cueing (Jones & Sarter, 2008).

One study described a human-computer interface that used both visual aids and audio cueing. The design implemented a 3D virtual display with a touchscreen controller and audio cueing. Twelve participants completed a simulated UAV mission that involved two tasks, mission planning and mission supervision. Each participant completed a test trial under each level of automation: manual, semiautomatic, and automatic. There were no performance results used in the analysis, and conclusions were only made based on subjective SA, workload, and general opinion questionnaires (Crescenzio, Miranda, Persiani, & Bombardi, 2009).

Opinion questionnaires suggest that the 3D display was appropriate for this task. The following reasons were given as to why the 3D display was successful: the most relevant information was displayed in a unique way, permitted a good perception of the current vehicle and mission state with limited head movement or combining information from multiple displays, and it had good interaction and navigation features that could be used to change views. The authors also suggest that the touchscreen was a good control method for scenarios that require high level commands and they found that the audio cueing was more successful in higher levels of automation to help maintain operator vigilance (Crescenzio et al., 2009).

Studies previously discussed commented on the benefit of modal cueing for all levels of operator capacity. The interaction between level of automation and cue modality is also of interest; one study aimed to minimize switching costs by using cues. The study was conducted in a simulated underwater UV environment and was previously described in section 3.3.3. Their study implemented two AA programs, one which was predominantly manual control and the other which was predominantly supervisory control. To mitigate negative performance effects due to change in automation level they used three cue types: auditory, visual and bimodal. Thirty-two participants were evaluated on performance, workload (NASA-TLK), and SA (SAGAT). The performance measures include: time-to-task completion (TTC) and number of task performance errors (Kaber et al., 2006).

The analysis of the cue type was broken down into two parts. The first analysis compared each of the cue types versus no cueing, and the second analysis compared each cue type and included any interactions between mission difficulty and cue type. The study showed that there was a significant effect of cueing on TTC. Bimodal cueing resulted in significantly lower TTC than no-cueing; however, there was no significant difference between the single mode cuing and no cueing conditions. When comparing the cue types they found that bimodal resulted in significantly lower TTC than visual, and there was no significant difference between bimodal and auditory, or auditory and visual. There was no significant interaction found between cue type and mission difficulty, suggesting that the difference in cue type effects will be present regardless of task load (Kaber et al, 2006).

Level 1 SA results from the study were similar to TTC such that bimodal cueing resulted in significantly higher SA than no cueing, and there was no difference between either of the modal cueing conditions and no cueing. Bimodal cueing resulted in significantly higher SA than both of the modal cue types, and there was found to be no significant difference in SA between the two modal types. Level 2 and 3 SA results did not show a significant difference between all the cue types. It is suggest that the Level 2 and 3 SA results could be due to the fact that the time for operators to mentally transition from one control mode to another was greater than the time between the cue and the control shift (Kaber et al., 2006).

Workload results revealed no significant difference of bimodal or either modal cueing type when compared to the no cueing condition. The authors reported a marginally significant difference in workload when comparing the different cueing types with each other. The data trended towards higher perceived workload for auditory cues in comparison with visual cues (Kaber er al., 2006).

In conclusion multimodal cueing was effective at improving performance when AA is being implemented; SA was increased and subsequently performance was increased with no cost to perceived workload, when compared to both the no cueing condition and the single mode cueing conditions (Kaber et al, 2006). Another study observed improved performance due to redundant multimodal cueing, when compared to both no cueing and individual modal cueing (Maza et al., 2010).

This study examined the application of multimodal interface designs for UAVs by testing different combinations of modal conditions (visual, audio, and tactile) in a simulated UAV control interface; 3D audio results have been discussed in section 3.5.2 and tactile results have been discussed in section 3.5.3 (Maza et al., 2010).

The multimodal experiment combined the 3D audio and tactile cue modalities that were previously discussed. When used in a multimodal design, they provide redundant information about the location of the warning that the operator needed to attend to. The authors found an improved performance over the single modality conditions and contribute it to the effectiveness of providing redundant information to the operator. The authors conclude that multimodal display techniques will improve operator performance of UAVs by helping with high information loads and communicating with operators in a variety of environments (Maza et al., 2010).

Multimodal interfaces have been discussed as being beneficial in multiple studies that incorporated different types of simulation environments. This suggests that resource offloading is a successful way to increase performance while operating UVs. The interaction between target modality and cue modality has been raised as a point of interest for military UV applications (Ferris & Sarter, 2008).

The study was conducted in a simulation of military battlefield operations, and incorporated a briefing style based on U.S. Army protocol that emphasised a hierarchy of mission objectives. Two main factors were manipulated within the test design, target modality (visual, auditory, or tactile) and cuing (uncued or cued). The cueing modality (visual, auditory, or tactile) and spatial relationship of the cue to the target were also varied (ipsilateral or contralateral). The ipsilateral condition was when cues were presented on the same side of the body as the target and the contralateral condition was the opposite. The performance of the participants was evaluated by target response times. It is important to note that due to expected high variability the authors used a more relaxed critical p value of 0.1 for their statistical analysis (Ferris & Sarter, 2008).

The authors hypothesised that there would be performance differences between the cued (ipsilateral and contralateral combined) and uncued conditions within each target modality, but this result was not found. There was a significant difference between the uncued and one of the cued conditions (ipsilateral or contralateral) for each target modality. This confirms the general hypothesis that cues can aid performance during UAV operations. The three significant cued results compared to the uncued condition were: 1) auditory, ipsilateral cueing decreased response time of visual targets, 2) tactile, contralateral cueing decreased response time of auditory targets, and 3) visual, ipsilateral cueing increased response time for tactile targets (Ferris & Sarter, 2008).

The results showed that there was a significant difference in the in the spatial relationship between targets and cue. The authors report that previous work has shown ipsilateral cueing (visual, auditory, and tactile) to have benefits. In the current study, ipsilateral cuing only resulted in a significantly lower response, when compared to contralateral cueing, for visual targets with auditory cues. In contrast the following contralateral target-cue pairs were significantly lower, than the ipsilateral pair: visual target with tactile cues, auditory target with tactile cues and tactile target with visual cues. To explain their unexpected results the authors discuss the concept of cross-modal inhibition of return (IOR). IOR is when the attention of the operator is inhibited from a location on the interface that had already been attended to. This is done in order to avoid wasting attentional resources and was a factor in this study because of the high attentional task demands (Ferris & Sarter, 2008).

The benefit of resource offloading has been discussed for modal and multimodal display designs. The specific cue scheme that will optimise performance is dependent on the mission task, and as seen in the later section the target modality. The performance improvements that have been reported due to cueing is important for future development of interface design that will help increase operator capacity without performance degradation.

3.5.5 LOA, Operator Capacity, and Information Perception

Operator capacity and level of automation have been shown to be limited by the interface design. Information perception of the operator is dependent on the interface design. Several studies have used the limitations in their interface to explain unexpected performance declines when a high level of operator capacity was used. The testing of different interface designs in combination with varying operator capacities will allow for further insight to measures that can be taken to further improve UV performance. Two studies are reviewed in the section below, one that looked at the interaction between display aids, operator capacity, and interface designs and the second that looked at the interaction between control modality, operator capacity, and UV type (Chadwick, 2006; Baber et al., 2011)

The study conducted by Chadwick (2006), previously described in section 3.4.3, looked at the interaction between operator capacity and interface design. In their two experiments they varied the number of UGVs being operated (one, two, or four) as well as the interface design. In the first experiment, the interface was varied between two different conditions (no display aids or display aids). In the second experiment the interface was varied between two different designs (A or B). Design A had four separate mini displays for each UGV plus a larger display that was designated to the UGV that was currently being operated. Design B only had four equally sized displays for each of the UGVs. The metrics used to evaluate the participants were monitoring, responding, and detecting performances.

This study showed that operator capacity had a significant effect on monitoring, responding, and detecting. Display aids had a significant effect on responding and there appears to be an interaction between operator capacity and display aids with respect to responding. The display aid condition resulted in significantly lower response time to valid targets. With respect to the interaction, the condition with no display aids appears to be more sensitive to a change in operator capacity than when display aids are present. The authors suggest that the addition of an overhead map that shows the current UGV that is being operated in the centre was what led to the reduction in response time for valid targets (Chadwick, 2006).

Quick video playback (QVP) was one of the display aids used in experiment 1. The intention of the playback was to assist in situations when an operator would switch to a UGV and not be able to understand the context of the environment it was in. This scenario would occur if the operator was unable to watch the approach of the UGV due to other attentional demands. Being able to use the QVP would mitigate the performance effects when this scenario would occur. The results did not have statistical significance but the authors did suggest that target localization was improved in the low context situations described above (Chadwick, 2006).

Interface design A and B had the same results with respect to response time and error detection; however, design A was more sensitive to change in operator capacity than design B, which is evident from the slopes of the fitted curves (Chadwick, 2006).

Baber at al. (2011) also looked at the interaction between information perception and operator capacity, only they altered the control modality. Their study was also split into two simulated UV experiments, where the same 16 participants were used in both. In experiment 1 the authors varied the control modality (speech, gamepad, multimodal) and type of UV (UAV or UGV). In experiment 2 the operators varied the control modality (speech, gamepad, multimodal) and operator capacity (one, three, and five). The metrics used to evaluate the participants were: number of commands issued, distance to target when command was issued, time to respond to warnings, and subjective workload (NASA-TLX).

The result from experiment 1 showed that there was a significant effect of control modality on the number of commands issued by the participant. Speech control resulted in the lowest number of commands and the gamepad resulted in the highest. Distance from the target when the command was issued was only significantly affected by the type of UV being operated, where participants tended to be closer to the target while operating UAVs. Time to respond to warnings was significantly affected by control modality where speech control resulted in the highest response time and multimodal control resulted in the lowest. No significant effect of control modality on workload was found (Barber et al., 2011).

When the authors looked at the multimodal trial in detail they found that speech controls tended to be used for target recognition and response to warnings, while the gamepad controls were used for all other sensor control. This division of tasks allowed the participants to have better performance in the secondary task of respond to warnings (Barber et al., 2011).

The results from experiment 2 supported the core findings from experiment 1 because they again found a main effect of control modality on number of commands issued and time to respond to warnings. The authors also found that there was an effect of modality on the distance from the target when the command was issued, where speech controls resulted in the shortest distance (Barber at al., 2011). This effect was not found in experiment 1.

The results from experiment 2 also showed a significant interaction of control modality and operator capacity for the number of commands issued. When the participants were operating five UVs, they used the lowest number of speech and gamepad controls. The largest number of gamepad controls was experienced in the one UV condition whereas the largest number of speech controls were experienced in the three UV condition. A similar trend was found in the time to respond to errors results, although not statistically significant. Response time increased as operator capacity increased when in the gamepad control condition. This was not true for speech control. Response time decreased from one to two UVs and increased (but still lower than one UV) from three to five UVs (Baber et al., 2011).

The results from both experiments show that using speech as a control modality on its own can be costly to the performance of UV operation. Speech can be combined with gamepad control in a multimodal interface to improve the performance observed from both control methods independently. The authors found an interesting trend in that the ratio of speech controls to gamepad controls in the multimodal condition was relatively similar across all operator capacity conditions. This implies that the participants employed a similar strategy no matter the number of UVs they were being asked to operate. Furthermore, they found the trend to be using speech control for symbolic tasks (classifying targets and responding to warnings) and gamepad control for spatial tasks (controlling sensors) (Baber et al., 2011).

In conclusion, the modality combination of speech and gamepad control provided performance advantages over either modality when used on its own. Furthermore, is the studies showed an interaction between operator capacity and control modality (Baber et al., 2011). It was shown earlier that performance benefits from visual display aids were also dependent on operator capacity (Chadwick, 2006). This finding suggests that interface design is dependent on the desired operator capacity (Baber et al., 2011; Chadwick, 2006).

3.5.6 Adaptive Interfaces

The automation, operator capacity, and interface design results previously described have demonstrated that there are many factors that need be considered when designing an interface to optimize performance. The use of a rigid interface for a complex application such as military operations has been criticised and recent work has discussed the efficacy of adaptive interfaces (De Visser et al., 2008; Hou & Kobierski, 2006; Hou et al., 2011). An important aspect to an adaptive interface is that the interface needs to know something about operator state as well as mission state. For operator state this can be accomplished through a variety of physiological measures and spatial orientation such as: heart rate, eye tracking, electromyography, body motion tracking, etc. (Hou & Kobierski, 2006; Hou et al., 2011; Maza et al., 2010). For mission state this is accomplished through an intelligent aspect of a computer system that is capable of processing complex information about the goals, environment, and current status of the system components (Hou & Kobierski, 2006; Hou et al., 2011)

De Visser et al. (2008) proposed a four step design methodology that looked at both the level of automation and information perception. Three use cases were then used to demonstrate how the design guidelines could be implemented.

The four steps in the design methodology are: 1) collect data to determine the system requirements, 2) implement a model to determine the level of automation required and strategy for how that automation is going to be implemented, 3) design an interface that maintains a high level of SA and minimizes workload, and 4) validate the interface design (De Visser et al., 2008).

Use case one references a study by one of the authors that introduces the concept of Adaptive Delegation Interfaces (ADIs). Similar to previous AA concepts discussed in section 3.3.3. ADIs adapted to the user and mission state. The delegation aspect is in reference to the user being in supervisory control of the UV. They present an ADI that has three mission modes: wizard, compose, and execution. The interface has different tasks while in each mode, and they are typically executed in sequential order. The first mode, mission wizard, allows the user to select high level mission goals that are processed through a plan generator. The second mode, mission compose, presents the structure of the plan previously generated and allows the user to select and assign task within the plan. The final mode, mission execution, is only activated once the plan is put into place and from this point on the user is able to monitor and modify parts of the mission if necessary (De Visser et al., 2008).

Use case two focuses on the development of an adaptive interface for commanders, who would be supervising the UV operator. Commanders' tasks in this scenario can be broken down into three mission phases: monitoring current video feed, reviewing past information, and re-tasking the vehicle in flight. The authors developed three mission modes to facilitate UV management in each of the mission phases: monitor mission mode, review mission mode, and change mission mode. The adaptive interface will switch modes based on: mission type, critical events and commander preference (De Visser et al., 2008).

Use case three focuses on the development of an adaptive interface for multiple UV control of collaborative operators. The Automated Mission Scheduler (AMS) is composed of two main components. One component is responsible for automatically creating mission plans, accounting for mission objectives, environment, resources, client requests, and user inputs. The second component is the operator interface, which supports decision making, problem solving, communication and negotiation between the collaborative operators (De Visser et al., 2008).

In summary the design methodology proposed was derived for supervisory control of UV systems and through the three use cases a few key challenges were recognised: identification of display nodes, proper application of AA, and the development of a method for giving high level objectives and tasks to automation for plan generation (De Visser et al., 2008). Delegating a level of decision making to automation through an adaptive interface has also been proposed through a hierarchy of automated agents (Hou et al., 2011).

A communication flow multiple adaptive intelligent agents (AIAs) form an intelligent adaptive interface (IAI). Within the interface there is a hierarchy of agents (junior, working, and senior) that individually can plan and react in order to achieve goals, obtain and store information, sense and act, model the environment, coordinate with each other, and resolve conflicts. Together within their framework, the IAI is capable of presenting the right information, presenting the action sequence that is proposed by the AIAs or perform the correct action autonomously. The goal of the multiple AIA based architecture is to effectively implement AA to increase SA, reduce workload, and subsequently increase UAV performance in military applications (Hou et al., 2011).

To achieve the aforementioned goal, AIAs will alleviate the requirement that the UAV operator focus on the detailed operation of the system and instead they will serve in a supervisory role. The conceptual design framework for an IAI is presented to contain five senior agents, each with a specific function: 1) managing, 2) modelling, 3) sensing, 4) tasking, and 5) interacting. Senior agents each have multiple working agents and amongst all the working agents, four must be responsible for the following key functions: 1) behaviour, 2) perception, 3) cognition, and 4) inference. Working agents each have their own set of junior agents that each has an individual task they are required to perform (Hou et al., 2011).

The communication flow of the system would be that a junior agent would obtain information from the operator about a specific characteristic (e.g. eye gaze). A working agent would combine information from multiple junior agents to determine a component of the operator state (e.g. where they are looking on the display). The components of operator state provided by multiple working agents is combined by a senior agent to determine the overall operator state and decide the necessary action to be taken by the operator (Hou et al., 2011).

This framework and design methodology has been successful in improving the human-computer interaction of UV interface design by facilitating a natural communication flow between the operator and interface. Further empirical studies are needed to advance the design framework to increase usability and UV operator performance (Hou et al., 2011).

3.5.7 Interface Design in Mobile Environments

Researchers have shown that an important consideration for some military applications is that the operator will be located in a mobile environment. Haas and Stachowiak (2007) discuss the idea that in the future soldiers will be required to operate UVs from a mobile environment, and this could decrease visual capabilities, and thus increase the need for cues in other modalities. Tactile and audio cues may also be affected by mobile environments due to the soldier’s body vibrating while travelling in rough terrain and external noise masking spatial audio cues. The authors conducted a field study with 12 participants who performed a simulated UV target search in a High-Mobility Multipurpose Wheeled Vehicle (HMMWV). The authors varied cue type (spatial audio, tactile or multimodal), display signal azimuth (ranged from -15° to +15°), and HMMWV movement (stopped at idle, traveling over gravel, and traveling over cross-country terrain). The metrics used to evaluate the participants were response time and subjective workload and participant age was included as an independent variable in the analysis.

Performance results showed that the only statistically significant main effects were movement and age. The cross-country mobility condition resulted in highest response times over the other two conditions, for all the age groups. Participants in their 20s had significantly shorter response times than those in their 40s (Haas & Stachowiak, 2007).

The results also found a significant interaction between movement and display modality as well as movement display modality and age. Therefore, there was no one modality that provided the shortest response time for all age groups during all mobility conditions. During the idle condition the only significantly lower response time was found while participants in their 30s were using tactile cueing. During the gravel condition the only significant lower response time was found while participants in their 40s were using multimodal cueing (tactile plus audio). During the cross-country condition each age group had one cue condition create different response times than the others. For participants in their 20s audio cueing resulted in higher response times than both tactile and multimodal. For participants in their 30s lowest response times were found with multimodal cueing. Participants in their 40s had higher response times when multimodal cueing was used, compared to both tactile and audio modal cueing (Haas & Stachowiak, 2007).

Workload results showed that there was a statistically significant main effect of cue type and mobility condition. Multimodal cuing resulted in significantly lower perceived workload than both of the tactile and audio modal cue types. The authors suggest that the multimodal cue combination tested in this study was effective in an environment with audio and tactile distractors. The cross-country condition resulted in significantly higher perceived workload than both of the other mobility conditions, and there was no difference between the idle and gravel terrain conditions. The authors suggest that this result was due to the high levels of environmental noise and vibration (Haas & Stachowiak, 2007).

The authors discuss that their results might be lacking statistical power because there were only three participants in each of the 30s and 40s age groups. In conclusion, the combination of subjective ratings and performance results indicate that a single display condition does not provide more performance improvements over another across all mobile conditions. Future research is needed to gain an understanding of the role of each modality during mobile conditions before conclusions can be made about design guidelines for interface designs (Haas & Stachowiak, 2007).

The factors that influence information perception while operating UVs in a military environment include: mission task, display and control modality, automation implementation and desired operator capacity. The concept of using an adaptive interface has been raised to accommodate for the complex interaction of these factors has been discussed and example design framework has been reviewed.

This page intentionally left blank.

4. Conclusion and Recommendations

4.1 Overview of Conclusions

This report identified three main factors in HRI research related to operating UVs for military applications: operator capacity, automation, and interface design. The common goal of the literature reviewed was to investigate the effect that each of the aforementioned factors had on HRI with three commonly used metrics: situation awareness, workload, and task performance. A tabular summary of the empirical studies reviewed in this paper can be seen in Annex A.

4.1.1 Operator Capacity

In general, research shows that increasing operator capacity also increases workload and reduces SA. The impact of an increased workload and decreased SA on overall task performance while operating multiple UVs was inconsistent across the literature, however, it was common for researchers to suggest that higher performance could be reached with improved automation and interface design.

4.1.2 Automation

Automation has been shown to improve HRI for some tasks and to increase the achievable performance when operator capacity is increased; however, if the level of automation is increased to a certain point, HRI performance will begin to decrease primarily due to a loss in situation awareness. Automation reliability is another factor to consider when designing UV systems, as decreases in performance effectiveness have been shown to occur as reliability decreases. Adaptive automation has successfully mitigated some of the effects of the complex interaction between variables that determine the optimal implementation of automation, but it has its own set of challenges that need to be considered in optimizing the HRI, such as function allocation and mode switching.

4.1.3 Interface Design

The majority of HRI occurs via the interface, and therefore interface design is an important factor to consider when attempting to increase operator capacity. The general goal of an interface is to optimize the information flow between the operator and computer, and the common modalities used to accomplish this are: visual, audio, tactile, and haptic. Visual displays have traditionally been the core component of a UV interface, but research has shown that offloading information to additional resources via one or more of the other modalities has increased SA, decreased workload, and increased performance. Research has shown that the optimal design depends on the operator capacity, level of automation, and type of UV, and therefore it is difficult to identify one particular interface design best suited for all UV applications. Researchers have proposed that adaptive interfaces are a promising solution to accommodate for this complex interaction, but further conceptual research and empirical studies are necessary to design and implement an adaptive interface that optimizes HRI.

4.1.4 Need for Further Research

Further research is needed to fully understand the complex interaction between operator capacity, automation, and interface design in the context of HRI. The human-robot ratio has yet to be effectively improved such that a single operator can control multiple UVs during military applications. This is primarily because this interaction is a complicated, multidimensional design problem that requires a high level of research resources in order to solve. Despite the wide range of ongoing research projects within this field, there is still a need for more, as is evident from the inconclusive findings and failure to successfully increase operator capacity for military applications. Furthermore, there are still many areas yet to be investigated, due to gaps in the current literature, and availability of new technology.

4.2 Avenues of New Research

The literature reviewed in this report identified key areas of interest necessary for future HRI research, but due to the defined limitations of the review the report does not represent every research area within the industry. Table 4 presents a matrix of the topics of studies reviewed in this report, and possible gaps within the literature. The categories represent key test parameters, where the row is the primary test parameter and column is the secondary study parameter. The matrix cells have been populated with the reference(s) that included each particular combination of primary and secondary test parameter in their research. Shaded cells represent combinations of primary and secondary test parameters that were not found in the literature reviewed for this report, while blank but not shaded cells represent combinations that were found in reverse order.

The matrix shows two main clusters that represent research concentration areas. One cluster, in the top left corner of the matrix, represents research that has focused on the interactions between automation and operator capacity. The second cluster, in the bottom right corner, represents research that has focused on interface design. The bottom left corner and top right corner are mirror images of each other and represent the interactions between interface design, automation, and operator capacity, areas not well covered in the literature reviewed. The most significant areas of research that were lacking coverage are adaptive interfaces, mobile environments, and the interaction between automation reliability and interface design.

Table 4 - Overview of Research Reviewed

Primary Test Parameter \ Secondary Test Parameter	Level of Automation	Automation Reliability	Adaptive Automation	Level of Operator Capacity	Operator Capacity Model Development	Mixed UVs	Interface Design - Visual	Interface Design - Audio	Interface Design – Tactile & Haptic	Interface Design - Multimodal	Adaptive Interface	Mobile Environments
Level of Automation	Endsley & Kaber, 1999; Dixon & Wickens, 2003	Endsley & Kaber, 1999; Ruff et al., 2002		Ruff et al., 2002; Liu et al., 2009; Riley & Sarter, 2006, Wickens et al., 2006	Dixon et al., 2005							
Automation Reliability		Wickens & Dixon, 2006		Wickens & Dixon, 2006								
Adaptive Automation	Kaber et al., 2006; Squire et al., 2006		Hou & Kobierski, 2006; Hou et al., 2011	Squire et al., 2006		Parasuraman et al., 2009	Kaber et al., 2006	Kaber et al., 2006		Kaber et al., 2006		
Level of Operator Capacity		Levinthal & Wickens, 2005	Lif et al., 2007	Adams, 2009; Chadwick, 2006; Crandall & Cummings, 2007		Baber et al., 2011; Chadwick, 2005	Chadwick, 2006	Baber et al., 2011		Baber et al., 2011		
Operator Capacity Model Development	Ruff et al., 2002; Liu et al., 2009; Riley & Sarter, 2006	Ruff et al., 2002			Nehme et al., 2008; Crandall & Cummings, 2007							
Mixed	Chen et al., 2008			Sterling & Perala, 2007; Chen et al., 2008				Baber et al., 2011		Baber et al., 2011		
Interface Design - Visual						Chadwick, 2005; Chen et al, 2008; Chen, 2010	Gunn et al., 2005; Calhoun, et al., 2007; Dehais et al., 2011; Chappell, 2007; Cavett et al., 2007	Gunn et al., 2005	Gunn et al., 2005			
Interface Design - Audio				Dixon & Wickens, 2003; Donmez et al., 2009; Graham & Cummings, 2007; Baber, 2011	Dixon & Wickens, 2003			Maza et al., 2010, Donmez et al., 2009; Graham & Cummings, 2007				
Interface Design – Tactile and Haptic									Maza et al., 2010; Gunn et al., 2005; Donmez et al., 2008			
Interface Design - Multimodal	De Crescenzio, et al., 2009			Baber et al., 2011			Ferris & Sarter, 2008	Ferris & Sarter, 2008	Ferris & Sarter, 2008	Maza, et al., 2010		
Adaptive Interface	De Viser et at., 2008		Hou & Kobierski, 2006; Hou et al., 2011									
Mobile Environments							Haas & Stachowiak, 2007	Haas & Stachowiak, 2007	Haas & Stachowiak, 2007	Haas & Stachowiak, 2007		

Note: Shaded cells represented combinations that were not found in the literature.

4.2.1 Adaptive Interfaces

As discussed in the Results section of this report, adaptive automation and adaptive interfaces have been proposed as a promising means of decreasing or eliminating performance detriments associated with increasing operator capacity. As shown in Table 4, this literature review suggests that future research is needed to better understand the interactions between automation reliability, adaptive automation, adaptive interfaces, and mobile environments. Developments in technology have led to the design of intelligent systems capable of analysing operator state through a number of physiological and physical orientation measures. Further, development of such intelligent computer systems could improve HRI and the future application of adaptive automation and adaptive interfaces for UV military operations, and help reach the goal of increasing the human-robot ratio above 1:1.

4.2.2 Mobile Environments

Mobile environments are an important area of research because they are a physical characteristic that would be encountered during some military applications which are not being represented in most simulator research studies. This literature review suggests that there is a need to expand on current studies that have investigated the effects of operating UVs in mobile environments, and ideally all studies that have concluded that their UV system design can improve performance should confirm their results in a mobile environment. This confirmation would bring the system design one step closer to being implemented in military applications.

4.2.3 Interaction between Automation Reliability and Interface Design

This literature review highlighted that, realistically, automation will not be 100% reliable when implemented in military applications. The general effects of automation reliability have received a significant amount of attention in literature, and the general conclusion is that it will have negative effects on UV system performance. Despite the current research efforts, more detailed research in the area of automation reliability, specifically with respect to the interaction between automation reliability and interface design, is needed. As discussed earlier, adaptive interfaces and adaptive automation are important areas of future research, and therefore their interaction with automation reliability should be incorporated into future research studies.

4.2.4 Additional Avenues of New Research

As noted throughout this review, most of the studies on UV design and implementation have been conducted with a simulation testbed, where the UVs and mission task are both represented by the simulation. Fewer studies implemented a "real" component to their studies, where either just the UVs or both the UVs and mission task are physically represented. In situations when just the UV is physically represented, the operator is in control of a real UV, however, the environment and mission task is still represented through a simulation. Implementing a "real" component into HRI research is important because simulations are unable to capture all of the system behaviour that would be present during military applications. Before UVs can be used in military applications, they first must be tested in physical environments that better represent the challenges of military situations. Team dynamics will likely be an important HF issue to consider, as it is realistic to assume that military applications will involve teams of operators, each with a team of UVs.

Team coordination should also be considered in the design and implementation of the interface, because there might be an optimal design for a single operator that is not effective in a team coordination environment. Related specifically to SA, team SA is more complex than combining SA from individual team members. In general team SA models need to include two additional components: shared SA between team members and combined SA of the whole team.

In conclusion, areas of research that are not strongly represented in the literature are interactions between automation reliability, adaptive automation, adaptive interfaces, and mobile environments. Furthermore, improved HRI depends on further research in the following areas: team coordination, operation of "real" UVs, mixed UAV and UGV systems, and multimodal interfaces. The combination of exploring new research avenues and expanding current research in the key areas identified will increase the potential for inversing the human-robot ratio for UV systems that will be implemented into military applications.

4.3 Recommended Research Areas and Methods

THRIL has the potential to facilitate research in areas not well represented in the current scientific literature. Consequently, the lab would benefit from a long term research goal of investigating human factors issues in the design of a multimodal adaptive interface for mixed UV military operations. As part of this overall research goal, the lab should undertake studies that look at the interactions between operator capacity, adaptive automation, automation reliability, adaptive interfaces, and mobile environments, as these research areas are critical to inversing the human-robot ratio for UV systems. Several unique characteristics of this lab also present secondary research avenues that can be explored within the lab's overall research plan. An overview of the characteristics and example points of interest within each characteristic can be seen in Table 5.

Table 5 - Unique Characteristics of THRIL

Characteristic	Research Opportunities
Military Environment	• The type of task (ie route reconnaissance) • The type of military role (ie infantry operator, UV specialist, commander, etc.) • • The military hierarchy (ie officer supervising two operators each with a UV team trying to accomplish a common goal)
Mixed UV Platforms	• Operating combinations of UAVs and UGVs • Operating multiple UVs of one platform with the use of sensor data from another platform not under your control (ie. UAV view to assist with the control of UGVs)
Operating Real UVs	• Test the impact real UV operations as opposed to strictly simulated UVs • Test with a mix of real and simulated UVs and environments
Number of UVs Available	• Extensions of operator capacity research • What combinations of UV platforms promote the best team success • Communication between team members • What strategies are used in team coordination

4.3.1 Overview of Lab Equipment and Requirements

The general lab equipment that will be available at THRIL is:

- Multiple UGVs and UAVs
- Optical localization system
- Central control station

The fundamental requirements in order to conduct the recommended research are:

- Basic interface design capability that is able to accommodate multiple modalities, an adaptive implementation scheme, and simulation of mobile environment.
- Basic automation capability that is able to manipulate LOA, level of reliability, and accommodate an adaptive implementation scheme.
- Software infrastructure that accommodates communication with multiple UVs, mixed UVs, teams of operators, and simulated missions tasks.
- Ability to measure SA, workload, and task performance in experimentally appropriate ways.

4.3.2 Measurement of Situation Awareness

In the literature reviewed for this report, situation awareness was measured via both direct measurements and indirectly through performance measurements. Performance measures can either be external to the required task, such as a research-forced display change that a participant needs to observe, or imbedded, such as a researchers' interpretation of a low success rate for a given task. In either case, performance measures are used as indicators of the different levels of situation awareness and SA scores are calculated for each participant. Direct measures of SA are typically either a subjective rating or questionnaire responses, which allow the researcher to calculate SA scores. They can be implemented either post-trial or during random pauses while the participant is in middle of their respective trial. Given that the method used to measure SA is study dependent, the method developed for THRIL should be designed specifically for each individual study conducted within the lab. A few well know SA evaluation methods are: Situation Awareness Rating Technique (SART), Situational Awareness Rating Scales (SARS), Situation Awareness – Subjective Workload Dominance (SA-SWORD), and Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1996).

4.3.3 Measurement of Workload

In the literature reviewed, workload was most commonly measured using the NASA-TLX framework. Within the NASA-TLX framework subjective ratings are provided post-trial for six categories: mental demand, physical demand, temporal demand, performance, effort, and frustration level. These scores are multiplied by the respective weighting of each category to calculate a total workload score. The weighting are determined by a set of fifteen pair-wise comparisons were the participant is required to select which of the two categories was the source of more workload (Hart & Staveland, 1988).

4.3.4 Measurement of Task Performance

Performance measures varied significantly with the literature reviewed, however a few common categories did present themselves, namely reaction time, success rate, accuracy, and performance efficiency. These categories were used for both primary and secondary tasks. THRIL should be able to implement any number of performance measures depending on the study parameters and mission task.

In conclusion THRIL should be able to accommodate research that will support areas that have not received much attention in the literature including: multimodal adaptive interfaces, adaptive automation, mixed UV operation, military scenarios. During studies investigating these research areas, the metrics used to evaluate participants within the lab will be consistent with those commonly found in literature. The combination of: 1) implementing methods that are common practice within the industry, 2) addressing gaps within the current research, and 3) using the unique characteristics of THRIL, will allow for novel HRI research that will ultimately help reach the goal of improving the human-robot ratio such that one operator can control multiple UVs.

This page intentionally left blank.

5. References

Adams, J.A. (2009). Multiple robot/single human interaction: effects on perceived workload. *Behaviour & Information Technology, 28(2),* 183-198.

Baber, C., Morin, C., Parekh, M., Cahillane, M., & Houghton, R.J. (2011). Multimodal control of sensors on multiple simulated unmanned vehicles. *Ergonomics, 54(9)*, 792-805.

Calhoun, G., Ruff, H., Lefebvre, A., Draper, M., & Ayala, A. (2007) “Picture-in-Picture” Augmentation of UAV Workstation Video Display. *Proceedings of the Human Factors and Ergonomics Society, 51(2),* 70-74.

Carryer, A. (2008) A History of Unmanned Aviation in Canada. Retrieved from http://www.uavs.ca/outreach/HistoryUAVs.pdf

Cavett, D., Coker, M., Jimenez, R., & Yaacoubi, B. (2007). *Human-Computer Interface for Control of Unmanned Aerial Vehicles*. Paper presented at the System and Engineering Design Symposium, 2007, Charlottesville, VA.

Chadwick, R.A. (2005). The Impacts of Multiple Robots and Display Views: An Urban Search and Rescue Simulation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 49(3),* 387-391.

Chadwick, R.A. (2006). Operating Multiple Semi-Autonomous Robots: Monitoring, Responding, Detecting. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(3)*, 329-333.

Chappell, S.L. (2007). UAV Control with Pictures and Pointing. *Proceedings of the American Institute of Aeronautics and Astronautics 2007 Conference and Exhibit*. Rohnert Park, CA: American Institute of Aeronautics and Astronautics.

Chen, J.Y.C. (2010). UAV-guided navigation for ground robot tele-operation in a military reconnaissance environment. *Ergonomics, 53(8),* 940-950.

Chen, J.Y.C., Durlach, P.J., Sloan, J.A., & Bowens, L.D. (2008) Human-Robot Interaction in the Context of Simulated Route Reconnaissance Missions. *Military Psychology, 20(3)*, 135-149.

Cosenzo, K., Parasuraman, R., & de Visser E. (2010). Automation Strategies for Facilitating Human Interaction with Military Unmanned Vehicles. In M.J. Barnes, & F.G Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 103-124). Farnham: Ashgate Publishing.

Crandall, J.W., & Cummings, M.L. (2007). *Developing performance metrics for the supervisory control of multiple robots*. Paper presented at the Proceedings of the ACM/IEEE international conference on Human-robot interaction.

Crescenzio, F.D., Miranda, G., Persiani, F., & Bombardi, T. (2009). A First Implementation of an Advanced 3D Interface to Control and Supervise UAV (Uninhabited Aerial Vehicles) Missions. *Presence: Teleoperators and Virtual Environments, 18(3)*, 171-184.

Cummings, M.L., Nehme, C.E., Crandall, J., & Mitchell, P. (2007). Predicting Operator Capacity for Supervisory Control of Multiple UAVs. In J.S. Chahl, L.C. Jain, A. Mizutani & M. Sato-Ilic (Eds.), *Studies in Computational Intelligence: Vol. 70. Innovations in Intelligent Machines (1)* (pp 11-37): Springer Berlin/Heidelberg.

Dehais, F., Causse, M., & Tremblay, S. (2011). Mitigation of Conflicts with Automation: Use of Cognitive Countermeasures. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(5),* 448-460.

de Visser, E., Cohen, M.S., LeGoullon, M., Sert, O., Freedy, A., Freedy, E., et al. (2008). A Design Methodology for Controlling, Monitoring, and Allocating Unmanned Vehicles. *3rd International Conference on Human Centered Processes. Delft: Proceeding of Supervisory Control in Critical Systems Management Workshop.*

Dixon, S.R., & Wickens, C.D. (2003). Control of Multiple UAVs: A Workload Analysis. *Proceedings of the 12th International Symposium on Aviation Psychology*, Dayton, OH.

Dixon, S.R., Wickens, C.D., & Chang, D. (2005). Mission Control of Multiple Unmanned Aerial Vehicles: A Workload Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 47(3),* 479-487.

Donmez, B., Cummings, M.L., & Graham, H.D. (2009). Auditory Decision Aiding in Supervisory Control of Multiple Unmanned Aerial Vehicles. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *51(5),* 718-729.

Donmez, B., Graham, H.D., & Cummings, M.L. (2008). *Assessing the Impact of Haptic Peripheral Displays for UAV Operators* (No. HAL2008-02). Cambridge, MA: Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Humans and Automation Laboratory

Endsley, M.R. (1995a). Measurement of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1)*, 65-84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37(1)*, 32-64.

Endsley, M.R. (1996). Situation Awareness Measurement in Test and Evaluation. In T.G. O'Brien & S.G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 159-180). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Endsley, M.R., & Kaber, D.B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42(3)*, 462-492.

Ferris, T.K., & Sarter, N.B. (2008). Cross-Modal Links Among Vision, Audition, and Touch in Complex Environments. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50(1)*, 17-26.

Gage, D.W. (1995). UGV History 101: A Brief History of Unmanned Ground Vehicle Development Efforts. *Unmanned Systems Magazine, (13)3*, 9-16.

Global Security. (n.d.). Unmanned Arial Vehicles (UAVs) Retrieved January 12, 2012, from http://www.globalsecurity.org/intell/systems/uav.htm

Graham, H.D., & Cummings, M.L. (2007). *Assessing the Impact of Auditory Peripheral Displays for UAV Operations* (No. HAL2008-02). Cambridge, MA: Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Humans and Automation Laboratory

Gunn, D.V., Warm, J.S., Nelson, W.T., Bolia, R.S., Schumsky, D.A., & Corcoran, K.J. (2005). Target Acquisition With UAVs: Vigilance Displays and Advanced Cuing Interface. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *47(3)*, 488-497.

Haas, E., & Stachowiak, C. (2007). *Multimodal Display to Enhance Human Robot Interaction On-the-Move*. Paper presented at the Proceedings of the 2007 workshop on Performance Metrics for Intelligent Systems.

Hart, S.G, & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P.A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload*, North-Holland, Amsterdam: North Holland Press.

Hou, M., & Kobierski, R.D. (2006). Operational Analysis and Performance Modeling for the Control of Multiple Uninhabited Aerial Vehicles from an Airborne Platform. In N.J. Cooke, H.L. Pedersen & O. Conneor (Eds.), *Human Factors of Remotely Operated Vehicles* (Vol. 7, pp.267-282): Emerald Group Publishing Limited.

Hou, M., Zhu, H., Zhou, M., & Arrabito, G.R. (2011). Optimizing Operator-Agent Interaction in Intelligent Adaptive Interface Design: A Conceptual Framework. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, 41(2)*, 161-178.

Jansen, C., & van Erp J.B. (2010). Telepresence Control of Unmanned Systems. In M.J. Barnes, & F.G Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 251-270). Farnham: Ashgate Publishing.

Jentsch, F., Evans, W., and Ososki, S. (2010). Model World: Military HRI Research Conducted Using a Scale MOUT Facility. In M.J. Barnes, & F.G Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 419-431). Farnham: Ashgate Publishing.

Jones, L.A., & Sarter, N.B. (2008). Tactile Displays: Guidance for Their Design and Application. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50(1)*, 90-111.

Kaber, D.B., Wright, M.C., & Sheik-Nainar, M.A. (2006) Investigation of multi-modal interface features for adaptive automation of a human-robot system. *International Journal of Human-Computer Studies, 64(6),* 527-540.

Levinthal, B., & Wickens, C.D. (2005). *Supervising Two Versus Four UAVs With Imperfect Automation: A Simulation Experiment* (No. AHFD-05-24/MAAD-05-7). Savoy: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Human Factors Division.

Lif, P., Hedstrom, J., & Svenmarck, P. (2007). Operating Multiple Semi-autonomous UGVs: Navigation, Strategies, and Instantaneous Performance. In D. Harris (Eds.), *Engineering Psychology and Cognitive Ergonomics* (pp. 731-740). Springer.

Liu, D., Wasson, R., & Vincenzi, D.A. (2009). Effects of Systems Automation Management STartegies and Milti-mission Operator-to-vehicle Ratio on Operator Performance in UAV Systems. *Journal of Intelligent & Robotic Systems, 54(5),* 795-810.

Maza, I., Caballero, F., Molina, R., Pena, N., & Ollero, A. (2010). Multimodal Interface Technologies for UAV Ground Control Stations. *Journal of Intelligent & Robotic Systems, 57(1),* 371-391.

Nehme, C., Crandall, J.W., & Cummings, M.L. (2008). *Using Discrete-Event Simulation to Model Situational Awareness of Unmanned-Vehicle Operators*. Paper presented at the Virginia Modeling, Analysis and Simulation Center Capstone Conference, Norfolk, VA.

Parasuraman, R., Cosenzo, K.A., & De Visser, E. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change, Detection, Situation Awareness and Mental Workload. *Military Psychology, 21(2)*, 270-297.

Riley, J.M., Murphy, R.R., & Endsley, M.R. (2006). Situation Awareness in the Control of Unmanned Ground Vehicles. *Human Factors of Remotely Operated Vehicles: Vol. 7. Advances in Human Performance and Cognitive Engineering Research,* (pp. 359-371). Amsterdam: Elsevier.

Riley, J.M., & Strater, L.D. (2006). Effects of Robot Control Mode on Situation Awareness and Performance in a Navigation Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50(3)*, 540-544.

Riley, J.M., Strater, L.D., Chappell, S.L., Conners, E.S., Endsley, M.R. (2010). Situation Awareness in Human-Robot Interaction: Challenges and User Interface Requirements. In M.J. Barnes, & F.G Jentsch (Eds.), *Human-Robot Interactions in Future Military Operations* (pp. 171-191). Farnham: Ashgate Publishing.

Ruff, H.A., Narayanan, S., & Draper, M.H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple Simulated Unmanned Air Vehicles. *Presence: Teleoperators and Virtual Environments, 11(4)*, 335-351.

Salmon, P.M., Stanton, N.A., Walker, G.H., Baber, C., Jenkins, D.P., McMaster, R., et al. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science, 9(4)*, 297-323.

Squire, P., Trafton, G., & Parasuraman, R. (2006). *Human Control of Multiple Unmanned Vehicles: Effects if Interface Type on Execution and Task Switching Times*. Paper presented at the Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction, Salt Lake City, Utah.

Sterling, B.S., & Perala, C.H. (2007). *Workload, Stress, and Situation Awareness of Soldiers Who are Controlling Unmanned Vehicles in Future Urban Operations* (No. ARL-TR-4071). Aberdeen Proving Ground, MD: Army Research Laboratory, Human Research & Engineering Directorate.

Wheatley, S. (2004). *The Time is Right: Developing a UAV Policy for the Canadian Forces*. Paper presented at the CDAI-CDFAI 7th Annual Graduate Student Symposium, Royal Military College, Ontario.

Wickens, C.D., & Dixon, S.R. (2006). *Human Machine Interface Analysis of Unmanned Vehicle Systems* (No. HFD-06-1/MAAD-06-1). Savoy: University of Illinois at Urbana-Champaign, Institute of Aviation, Aviation Human Factors Division.

Wickens, C.D., Dixon, S.R., & Ambinder, M.S. (2006). Workload and Automation Reliability in Unmanned Air Vehicles. In N.J. Cooke, H.L. Pringle, H.K. Pedersen, & O. Connor (Eds.), *Advances in human performance and cognitive engineering research: Vol. 7. Human Factors of Remotely Operated Vehicles* (pp. 209-222). Amsterdam: Elsevier.

Annex A

Table 6 - Summary of Empirical Studies

Reference	n	Simulated or Real?	Military?	UV Type	Operator Capacity	Level of Automation	Mission Task.	Dependent Variable(s) (bold=result that were stat sig)	Metric(s)
Sterling & Perala, 2007	12	Simulated	Yes	Mixed	1-3	Partial static automation	MOUT reconnaissance of primary and secondary targets.	Type of military unit Type of UV	Workload Stress SA
Endsley & Kaber, 1999	30	Simulated	No	N/A	N/A	Level 1 – Level 10	General cognitive control task.	LOA	Workload SA Task Performance
Dixon & Wickens, 2003; Wickens et al., 2006		Simulated	Yes	UAV	1-2	Manual – Fully automatic	Track UAV waypoints while conducting target recognition.	**LOA** Interface Design - Audio	Task Performance
Wickens & Dixon, 2006		Simulated	Yes	UAV	1-4	Manual – Fully automatic	Track UAV waypoints while conducting target recognition.	Automation Reliability	Task Performance
Kaber et al., 2006	32	Simulated	Yes	Unmanned Underwater Vehicle	1	Manual-Fully Automatic Adaptive	Perform an underwater mine disposal task.	LOA (including adaptive) **Interface Design - Multimodal**	Workload SA Task Performance
Squire et al., 2006	12	Simulated	No	N/A	1-6	Waypoint-play -super play Adaptive	Use robots to capture a flag.	LOA OC	Task Performance
Parasuraman et al., 2009	16	Simulated	Yes	Mixed	2	Manual-automatic Adaptive	UAV target identification, while UGV rout planning and monitoring communications.	LOA	Workload SA Task Performance
Dixon et al, 2005		Simulated	Yes	UAV	1-2	Manual-Fully automatic	Track UAV waypoints while conducting target recognition.	Automation Reliability Model Development	Task Performance
Ruff et al., 2002	12	Simulated	Yes	UAV	1-4	Manual-MBC-MBE	Target recognition and identification while controlling	**LOA**	Workload

							UAVs in the different automation schemes.	**Operator Capacity** **Automation Reliability**	SA Trust Task Performance
Liu et al., 2009	60	Simulated	Yes	UAV	1-4	MBC-MBE	Monitor sensor data and determine accuracy of ATR	**LOA** **Operator Capacity**	Task Performance
Riley & Sarter, 2006	20	Real	No	UGV	1-2	Manual – semi-automatic – fully automatic	Navigate UGVs through a maze	**LOA** **Operator Capacity**	Workload SA Task Performance
Lif et al., 2007	12	Real	Yes	UGV	1-3	Adaptive automation (operators choice)	Navigate through a MOUT environment as quickly as possible	**Operator Capacity**	Task Performance Time spent in control mode Subjective interview
Adams, 2009	12	Real (in a simulated environment)	No	UGV	1,2,4	Manual-semi automatic	Navigate while avoiding obstacles in a room	**Operator Capacity**	Workload Task Performance
Levinthal & Wickens, 2005	42	Simulated	Yes	UAV	2,4	Manual-fully automatic	Management of UAVs and detection of targets	LOA Automation Reliability **Operator Capacity**	Task Performance
Chadwick, 2006	24	Simulated	No	UGV	1,2,4	Fully automatic	Monitor and recharge batteries, while monitoring navigation and performing target recognition and identification	**Operator Capacity** **Interface Design – Visual**	Task Performance
Baber et al., 2011	16	Simulated	No	Mixed	1,3,5	Manual	Control UV sensors and identify targets, while responding to system warnings	**Operator Capacity** **Interface Design - Multimodal**	Workload Task Performance
Crandall & Cummings, 2007	12	Simulated	No	UGV	2,4,6,8	Semi-automatic	Remove as many objects as possible from an unknown area.	**Operator Capacity**	Task Performance
Gunn et al., 2005	16	Simulated	Yes	UAV	1	Semi-automatic	Target recognition	**Interface Design – Visual, Haptic**	Workload Task Performance
Calhoun et al., 2007	12	Simulated	Yes	UAV	1	Semi-automatic	Manually manipulate sensor to locate a landmark, while the UAV navigated autonomously.	**Interface Design - Visual**	Workload SA Task

									Performance
Dehais et al., 2011	23	Real	No	UGV	1	Semi-automatic	Target localization and identification	**Interface Design - Visual**	Task Performance Physiological Measures
Chadwick, 2005	61	Real	Yes	Mixed	1-3	Manual	Urban search and rescue simulation	**Operator Capacity** **Interface Design - Visual**	Workload Task Performance
Chen et al., 2008	30	Simulated	Yes	Mixed	1,3	Manual – Semi-automatic	Parallel rout reconnaissance	**Operator Capacity** **Interface Design-Visual**	Workload Task Performance Usability
Chen, 2010	28	Simulated	Yes	Mixed	1,2	Semi-automatic	Find and navigate to both primary and secondary targets	**Interface Design - Visual**	Workload Task Performance
Donmez et al., 2009; Graham & Cummings, 2007	39, 44	Simulated	Yes	UAV	1,2	Semi-automatic	Monitor flight path and ensure the correct targets are engaged and the UAV(s) return to base within the allotted time	**Operator Capacity** **Interface Design - Audio**	Workload Task Performance
Maza et al., 2010	9	Simulated	No	UAV	N/A	N/A	Respond to general queries.	**Interface Design – Audio, Tactile & Multimodal**	Task Performance
Donmez et al., 2008	13	Simulated	Yes	UAV	Multiple	semi-automatic	Monitor flight path and ensure the correct targets are engaged and the UAV(s) return to base within the allotted time	**Interface Design - Haptic**	Workload Task Performance
De Crescenzio et al., 2009	12	Simulated	Yes	UAV	1	Manual-semi-automatic-automatic	Mission planning and supervision	Interface Design - Multimodal	Workload SA Questionnaire
Ferris & Sarter, 2008	15	Simulated	Yes	UGV	1	Semi-automatic	Responsible for command level task, as the vehicle commander for the first vehicle in a convoy.	**Interface Design - Multimodal**	Task Performance
Haas & Stochowiak, 2007	12	Simulated	Yes	UV	1	Semi-automatic	Target search	**Interface Design – audio, tactile & multimodal** **Mobile Environment**	Workload Task Performance

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) HumanSystems Incorporated 111 Farquhar Street Guelph, ON N1H 3N4	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC A REVIEW: GCEC JUNE 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Human-Robot Interaction Literature Review:

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Bray-Miners, Jordan; Ste-Croix, Chris; Morton, Andrew

5. DATE OF PUBLICATION (Month and year of publication of document.) March 2012	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 80	6b. NO. OF REFS (Total cited in document.) 53

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Contract Report

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Toronto
1133 Sheppard Avenue West
P.O. Box 2000
Toronto, Ontario M3M 3B9

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.) DRDC Toronto CR 2012-083

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

ABSTRACT: In the field of unmanned vehicle (UV) systems, researchers have been striving to inverse the human-robot ratio such that one operator can control multiple robots. This goal has not yet been accomplished for military applications, despite ongoing research. Research suggests that the human-robot interaction (HRI) that takes place while an operator is in control of one or more UVs needs to be improved before the ratio can be inversed. This literature review included 53 references to provide an overview of current HRI research dealing with the operation of UVs and to identify the key human factors (HF) issues when conducting research within this area.

The literature identified three key factors in HRI research related to operating UVs for military applications: operator capacity (that is, the number and type of UVs that a human operator controls or supervises), automation, and interface design. Within the literature HRI is most often measured through the three common metrics of situation awareness (SA), workload, and task performance. In general, research shows that increasing operator capacity increases workload and decreases SA, while the corresponding impact on performance has been shown to be inconsistent. Automation and multimodal interfaces have been shown to alleviate some of the increased workload and decreased SA as operator capacity is increased, however, there is a complex interaction between the three variables. The literature suggests that adaptive automation and adaptive interfaces are promising solutions to accommodate for this complex interaction, but further research and empirical studies are necessary before they can be implemented into military operations. Three additional characteristics of military applications also need to be investigated further: one operator in control of mixed UV platforms (i.e. UAVs and UGVs), operators controlling UVs in a mobile environment, and team coordination between multiple operators each in control of multiple UVs.

To help further research in this area, the new Human-Robot Interaction laboratory being built at DRDC—Toronto should consider investigating HF issues in the design of a multimodal adaptive interface for mixed UV military operations. In particular, due to gaps in the literature and the need for more detailed research in certain areas, studies should look at the interactions between operator capacity, adaptive automation, automation reliability, adaptive interfaces, mobile environments, and team coordination.

Resume: Dans le domaine des systèmes de véhicules sans pilote (UV), les chercheurs travaillent sans cesse à inverser le ratio humain-robot, de sorte qu'un seul opérateur puisse commander plusieurs robots. Cet objectif n'est pas encore atteint en ce qui concerne les systèmes militaires, et ce, en dépit de recherches continues. Selon les recherches, il faut améliorer l'interaction humain-robot (IHR) qui a lieu quand l'opérateur est aux commandes d'un ou de plusieurs UV avant de pouvoir inverser le ratio. Cette analyse documentaire porte sur 53 ouvrages et vise à donner un aperçu des recherches sur l'IHR concernant l'utilisation des UV qui sont actuellement en cours et à cerner les principaux enjeux liés aux facteurs humains (FH).

Dans les ouvrages étudiés, on a constaté que les recherches sur l'IHR portant sur le fonctionnement des UV militaires font ressortir trois principaux facteurs : la capacité de l'opérateur (c'est-à-dire, le nombre et le type de véhicules qu'un opérateur humain contrôle ou supervise), l'automatisation et la conception de l'interface. Dans les ouvrages analysés, l'IHR est la plupart du temps mesurée au moyen des trois paramètres communs que sont la connaissance de la situation (CS), la charge de travail et le rendement à l'exécution des tâches. En général, les recherches démontrent que l'amélioration de la capacité de l'opérateur fait augmenter la charge de travail et diminuer la CS. Cependant, l'incidence sur le rendement s'est avérée inégale. Il a été démontré que l'automatisation et les interfaces multimodales aident à atténuer quelque peu l'augmentation de la charge de travail et la diminution de la CS quand on accroît la capacité de l'opérateur. Toutefois, il existe une interaction complexe entre les trois variables. Les documents indiquent que l'automatisation adaptive et les interfaces adaptatives constituent des solutions prometteuses qui permettraient de faciliter cette interaction complexe, mais il faudra poursuivre les recherches et effectuer d'autres études empiriques avant de pouvoir les intégrer aux opérations militaires. On doit également poursuivre l'étude de trois autres caractéristiques des applications militaires : un seul opérateur aux commandes de diverses plateformes UV (p. ex. UAV et UGV), des opérateurs commandant des UV dans un environnement mobile et la coordination d'équipe en présence de multiples opérateurs commandant chacun plusieurs UV.

Pour contribuer aux recherches effectuées dans ce domaine, le nouveau laboratoire de l'interaction humain-robot, qui est en cours de construction à RDDC—Toronto devrait envisager d'étudier les questions liées aux FH dans la conception d'une interface adaptative multimodale pour les opérations militaires utilisant divers types d'UV. En particulier, comme il existe des lacunes dans la documentation et comme des recherches plus approfondies s'imposent dans certains domaines, les études devraient porter sur les interactions entre la capacité de l'opérateur, l'automatisation adaptative, la fiabilité de l'automatisation, les interfaces adaptatives, les environnements mobiles et la coordination d'équipe.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Human-Robot interactions; unmanned vehicles; human factors; workload