FLEXIBLE ROBOTIC CONTROL VIA CO-OPERATION BETWEEN AN OPERATOR AND AN AI BASED CONTROL SYSTEM

by

MANOLIS CHIOU

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Mathematics
College of Engineering and Physical Sciences
The University of Birmingham
20th September 2016

Abstract

Most robots used for demanding tasks in unpredictable environments, e.g. bomb disposal, are teleoperated (i.e. directly controlled by a human using a joystick). In contrast, autonomous robots are mostly limited to repetitive tasks or highly constrained environments. The reason for this is that the current state-of-the-art for robotics does not allow fully autonomous robots to deal with continuously changing and unstructured environments such as the hazardous real-world situations for which teleoperated robots are deployed. However, recent advances in the field make autonomous robots now able to deal reliably with a variety of tasks, for example navigate between two points on a map.

Demanding and safety critical tasks (e.g. search and rescue (SAR), hazardous environments inspection), which are currently teleoperated, could soon start to benefit from autonomous capabilities, such as algorithms for automatic robot navigation or algorithms for automatically constructing and updating a map of the environment. There is a big literature on Human-Robot-Interaction (HRI) field studies requesting for more autonomy to be implemented in such type of scenarios. Robots could usefully use Artificial Intelligence (AI) control algorithms to autonomously take control of certain functions when the human operator is suffering a high workload, high cognitive load, anxiety, or other distractions and stresses. In contrast, some circumstances may still necessitate direct human control of the robot. Thus, the problem can be addressed by the use of variable autonomy. Variable autonomy refers to the approach of incorporating several different levels of autonomous capabilities (Level(s) of Autonomy (LOA)) ranging from pure teleoperation (human has complete control of the robot) to full autonomy (robot has control of every capability), within a single robot. Another way to describe this might be âĂIJon-demand triggered autonomyâĂİ, for example whereby a situation may require the skills of the operator or the capabilities of the AI and the system should be able to switch from one state to the other.

This project addresses the problem of variable autonomy in tele-operated mobile robots. Previous research on variable autonomy has not yet fully answered fundamental questions such as: which is the optimum choice of an autonomy level (e.g. should the operator be in control of navigation or the robot?) at a given moment; which factors affect this choice and most importantly how will this be used to address the transition between autonomy levels. Many systems use only the operator's judgement, while some others try to aid such judgements using suggestions automatically made by the AI. A common feature in most previous variable autonomy studies, is that the robot has limited authority to take initiative for actions. This research aims in contributing by:

- Investigating which conditions affect the performance of either the operator or the robot.
- Identifying the tasks for which teleoperation is better (i.e. less errors, faster task completion etc. than autonomy and vice versa.
- Designing a mixed-initiative control algorithm for switching between the different autonomy levels in an optimal way.

Mixed-Initiative (MI) refers to the peer-to-peer relationship between the robot and the operator in terms of the authority to initiate actions and changes in the autonomy level. Research will be conducted and evaluated in a principled way by designing experiments with methods drawn from human factors, psychology, HRI and robotics

Acknowledgements

My acknowledgements here

Publications arising from this Thesis

- **Conference**

  1. Chiou, M., Hawes, N., Stolkin, R., Shapiro, K. L., Kerlin, J. R., & Clouter, A. (2015). *Towards the Principled Study of Variable Autonomy in Mobile Robots.* In IEEE International Conference on Systems, Man, and Cybernetics (SMC2015) (pp. 1053-1059)

- **Journal**

  1. Chiou, M., Hawes, N., Stolkin, R., Shapiro, K. L., Kerlin, J. R., & Clouter, A. (2015). Towards the Principled Study of Variable Autonomy in Mobile Robots. In Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on (pp. 1053-1059). http://doi.org/10.1109/SMC.2015.190

Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

1.1 Project summary

Robots in safety- and time-critical applications such as Search and Rescue (SAR) and hazardous environment inspection are predominantly teleoperated by one or more operators. The difficulty of such applications due to the unpredictable nature of the environment along with the required human abilities (e.g. critical decision making or communication with victims in SAR), dictates always following a human-in-the-loop paradigm.

Several field studies regarding the use of robots in these domains, have pointed out that these applications can benefit from robots that actively assist operators. Ideally what is required is a human-robot team system that dynamically benefits from the capabilities of both agents and at the same time counteract the weaknesses of each.

This project addresses the use of variable autonomy as a potential solution in blending the capabilities of humans and robots. A variable autonomy system is one in which control can be traded between the human operator and the robot by switching between different Levels of Autonomy (LOAs), such that agents can assist each other. More specifically this project addresses the issues of dynamically changing LOA on-the-fly (i.e. during task execution) using Mixed-Initiative (MI) control. MI refers to the peer-to-peer relationship between the robot and the operator in terms of the authority to initiate actions and changes in the autonomy level.

More specifically, in the scope of this research, we are particularly interested in mobile robots in SAR and inspection/exploration scenarios as they provide challenging testing fields for teleoperation and variable autonomy. This however, does not necessarily mean that other potential applications are excluded.

1.2 Autonomy versus teleoperation

The state of the art in autonomous robots has progressed greatly the recent years. Robotic systems are able to operate autonomously and robustly for long periods of time [66]. This is mostly limited in simple tasks, taking place in structured and relatively controlled environments such as offices, museums and warehouses.

Despite these advances, robots in safety and time critical applications such as Search and Rescue (SAR), nuclear decommissioning, and hazardous environment inspection were not able to demonstrate the same level of self-sufficiency. They are predominantly teleoperated by one or more operators. There are three main reasons imposing this limitation. Firstly, these tasks often involve environments that are highly unstructured and changing, e.g. a partially collapsed building after an earthquake. Secondly, the nature of these tasks requires specific human abilities such as critical decision making based on incomplete information (e.g. victim dead or alive; and risk assessment of certain actions); or communication with victims in SAR [29]. Thirdly, high consequence industries tend to be conservative and not to trust autonomy. For example in Fukushima Daiichi nuclear accident such lack of trust led the robots to be deployed in pure teleoperation despite having various autonomous capabilities [71]. Hence, following the always a human-in-the-loop paradigm [68].

According to Lichiardopol [61] “Teleoperation comprise a robot technology where a human operator (master) controls a remote robot (slave). The system is formed by two parts, the control module, called cockpit and the telemanipulator, the slave robot at the remote location.". In other words, teleoperation allows an operator to control a robot from a distance that can vary from very short distances (e.g. next room) to vast distances (e.g. other planets). The control is achieved by an Operator Control Unit (OCU) which usually is composed by a screen providing video feedback and a joystick for

1.2.1 Initial Hypotheses

There are two initial hypotheses that will guide the research and the experimental design initially. These two hypotheses are:

- The factors that affect operator's performance are cognitive workload and fatigue.
- The factors that affect robot's performance are: a) the uncertainty in information given (i.e. sensing) to an AI algorithm; b) the limited capabilities of an AI algorithm to process the given information; c) the complete absence of a specific algorithm and thus absence of a specific capability.

It is highly intuitive and also supported by the literature (as described in previous sections) that operator's performance are workload depended. One can argue plausibly that also SA affects performance. However, it is deficient compared to the workload hypothesis in two ways: a) a big percentage of workload is actually a product of mental effort to either acquire SA or maintain SA; b) contrary to workload, SA cannot be measured directly by the AI in real-time as it is either measured subjectively or by highly intrusive techniques such as SAGAT [**?**] that require operator to completely pause the task. Even with such methods, validation comes after the task is finished by comparing the real situation with the perceived SA.

Given the perfect algorithm and enough information about the environment, then the robot should be able to always outperform the operator as it will always find the optimum solution and execute it without errors. In reality however this is not the case. The AI algorithms used to provide the robot with its capabilities have limitations and are highly dependent upon the reliability of the sensors measurements. For example the performance of a SLAM algorithm will degrade in the case that the laser range finder sensor provides noisy measurements. Another case is an algorithm that even given quality input, has limited capabilities and does not produce good results (e.g. a path planning technique can compute a non-optimal path).Lastly, the robot may not be able to perform an action at all. For example not be able to perform a search task without the appropriate robot vision technique.controlling the movement of the robot.

Despite its advantages, teleoperation comes also with some fundamental disadvantages. As a result of remotely controlling a robot, the operator's Situation Awareness (SA) is impaired. Yanco et al. [105] define SA in Human-Robot-Interaction (HRI) as "the perception of the robots' location, surroundings, and status; the comprehension of their meaning; and the projection of how the robot will behave in the near future...". The operator can have difficulty in controlling the robot, because he does not have the same SA as if he was having direct visual contact with the scene. Sensor limits often give very narrow and low resolution fields of view which can increase the mental workload of the operator and overload his working memory as he needs to have a constant model of the environment. Also limitations in communication can cause information distortion and delays [53] (e.g. in commands) which can further increase workload. All these, in cases of time critical and safety critical applications can cause the operator to be disorientated and potentially make costly errors.

1.3 Improving teleoperation

1.3.1 Interfaces and telepresence

Research on improving teleoperation is mostly focused in two major areas, interfaces and telepresence. Interfaces have systematically been studied as a way to improve robot teleoperation and deal with its intrinsic disadvantages [24]. They can provide high level fused sensor information and a fairly detailed world model along with alert messages for critical events. All these can reduce mental effort by providing operator with improved SA. However, studies have shown that humans tend to rely too much on visual feedback and cues [102, 4], something that can overload their visual modality. Also the same factors can give them a false sense of confidence leading to careless errors. These factors make interface design a non trivial task. Research on the field of interfaces is not in the scope of this thesis. There is a rich literature on standards and guidelines for designing interfaces [103, 82, 73]. These recommended standards and guidelines have been adopted by the interface used in our system.

Telepresence is "the ideal of sensing sufficient information about the teleoperator and task environment, and communicating this to the human operator in a sufficiently natural way, that the operator feels physically present at the remote site." as defined by Sheridan [86]. Where teleoperator in our case is a mobile robot. Designing and applying telepresence systems is challenging and often requires specialized equipment. This is because human senses (e.g. tactile sense) are very complex to be sensed by the robot and to be transferred as an experience to the operator, with a high degree of fidelity [16, 17]. Furthermore, telepresence can help in reducing the mental effort that is required by an operator to acquire SA but cannot reduce errors caused by high workload. This is because, telepresence cannot in itself overcome the problems of an operator being overloaded by, e.g. several simultaneous cognitively complex tasks.

1.3.2 Variable autonomy and Mixed-Initiative

Synergistically with the use of advanced interfaces and telepresence as methods to improve teleoperation, equipping robots with autonomous capabilities can potentially tackle some of the intrinsic difficulties of the problem. Several field studies regarding the use of robots in the domains of interest, e.g. in 9/11 World trade center [23] or in DARPA robotics challenge [107], have pointed out that these applications can benefit from robots that actively assist operators. Towards this direction, research is mostly focused on automating certain individual capabilities such as navigation, perception, stairs climbing and tip-over prediction [62].

Ideally what is required is a human-robot team system that dynamically benefits from the capabilities of both agents and at the same time counteract the weaknesses of each. Variable autonomy can potentially offer a solution in blending the capabilities of humans and robots. A variable autonomy system is one in which control can be traded between the human operator and the robot by switching between different Levels of Autonomy (LOAs), such that agents can assist each other [26]. More specifically, this project is focusing on Mixed-Initiative (MI) control. Mixed-Initiative means that the robot and the operator both have the authority and capacity to initiate actions and changes in the LOA [25].

Such a system offers the potential to assist a human who may be struggling to cope with issues such as high workload; intermittent communications; operator multi-tasking; fatigue; and sleep deprivation. For example, a human operator might need to concentrate on a secondary task while temporarily devolving control to an AI which can autonomously manage robot navigation. This is something very common as robot operators have to convey Situation-Awareness (SA) information, e.g. to SAR task force team mates [70, 15].

The use of different LOAs in order to improve system performance is a challenging and open problem, raising a number of difficult and previously unexplored questions. For example: which LOA should be used under which conditions?; what is the best way to switch between different LOAs?; and how can we investigate the trade-offs offered by switching LOAs in a repeatable manner? This project explores such questions by conducting experiments within a rigorous multidisciplinary framework, drawing on methodologies from the fields of psychology and human factors, as well as engineering and computer science.

1.4 Contributions of this Thesis

Research which focuses on investigating dynamic LOA switching on mobile robots is fairly limited. Furthermore, the investigation of MI systems to address this dynamic switching is even more limited. This research has contributed by:

- Established a rigorous experimental framework in order to formally and systematically evaluate the benefits of combining the capabilities of both human and autonomous control in a dynamically mode-switching system. This framework draws on methodologies from the fields of psychology and human factors, as well as engineering and computer science.
- Designed a Human-Initiative variable autonomy system in which the human operator can dynamically switch between teleoperation (i.e. direct joystick control) and autonomy (i.e. robot navigates autonomously towards waypoints selected by the human).
- Provided statistically validated empirical evidence that HI outperformed teleoperated or autonomous systems in various circumstances.
- Designed a MI control system in which the both the operator (based on judgement) and the robot (based on empirically informed objective metrics) are able to switch between different LOA in an optimal way.
- Provided statistically validated empirical evidence that MI outperformed the HI system in various circumstances.
- Provided analysis on the interaction of the human operator with and using the HI and MI systems.

1.5 List of Publications

1.6 Thesis structure

Thesis structure

CHAPTER 2

Background

The research described in this thesis is highly multidisciplinary. Thus, this chapter is focused on providing the necessary background knowledge, review the relevant literature from multiple perspectives, identify the gaps and discuss some of the techniques and tools that have been used in this research.

2.1 Human-Robot interaction

In this section research from the perspective of Human-Robot interaction will be examined. Studies in the HRI field provide a valuable insight on addressing problems and designing robotic systems in which the human is expected to interact at some level with the robot, as in the case of telerobotics and variable autonomy control. Research done mostly on SAR robotics will be presented as it provides important HRI lessons and insights on mobile robots, learned in the field rather than in a lab. Moreover, it provides motivation and inspiration for the potential impact that increased autonomy robots can have upon disaster response and other relevant applications.

2.1.1 Human-Robot interaction definition and context

Human-Robot interaction is a newly emerged multidisciplinary field of study which according to Goodrich et al. [47] is "dedicated to understanding, designing, and evaluating robotic systems for use by or with humans". Teleoperation and variable autonomy control, belong to the category of remote HRI as the human and the robot are not in the same location but interact from a distance. The type of interaction depends on the LOA used. It can vary from direct motor commands in the case of pure teleoperation, to a higher and more abstract type of interaction in the case of more sophisticated autonomy. In the latter the human provides high level commands in the form of goals or the robot takes the initiative to execute a task while informing the operator.

In our work, of particular interest are some prominent factors that affect interactions between humans and robots [47]. These factors are briefly presented and put in context with respect to this thesis:

- **Level and behaviour of autonomy:** In our system multiple LOAs exist that are not fixed. They rather change dynamically (i.e. during task execution). Thus, we are interested in how each LOA affects interactions between the operator and the robot. Moreover, of particular importance are the interactions between the two agents in regard to the LOA switching capabilities of MI control.

- **Nature of information exchange:** In our research, the robot and the human are exchanging mostly three types of information. Firstly, they exchange spatial information relevant to SA (e.g. position on the map). Secondly, information regarding the state of the human (e.g. neglecting the robot) and the robot (e.g. robot unable to continue task). This in many cases is implicit and is done through observation or estimation of system's performance. The purpose of such information is to facilitate informed LOA switches. Thirdly, explicit information regarding the LOA transitions, the current LOA, and control; e.g. sound alerting the operator that the robot has taken control.

- **Structure of the team:** Our research is focused on the case of a human-robot team consisting of one mobile robot and one human operator. The use of MI control means that they both act as peers regarding the initiative to take control and perform actions.

In [81] a theory of HRI is proposed, and the roles of humans along with the necessary awareness/information that each role should have are defined. The roles are the following: supervisor; operator; mechanic; peer; and bystander. In the case of a MI robotic system, as the one in our work, the human takes simultaneously two roles: a) the role of supervisor in the planning and goal level; b) the role of the operator in the action (i.e. execution) level.
One of the earliest works on a complete HRI interaction taxonomy is presented in [106]. Among other categories, this taxonomy classifies HRI systems by the level of autonomy the robot has, the ratio of humans and robots on the team, the composition of the robot team (homogeneous/heterogeneous robots) and the time/space relationships of the team members (e.g. collocated robot and operator). Based on the above, an updated taxonomy is proposed by Yanco et. al [104] taking further into account classification based on the task type, the robot morphology, the interaction roles and the human-robot physical proximity.

2.1.2 Human-Robot interaction awareness

One very important notion in HRI is that of awareness, which in general is defined as “an understanding of the activities of others, which provides a context for your own activity" [31]. The work of Drury et al. [33] defines a framework for awareness in the HRI context. They argue that many proposed definitions of awareness are not precise enough for HRI. They propose a framework of definitions for different cases of HRI awareness. Relevant to our work, the one robot-one operator awareness is defined as: “Given one human and one robot working on a task together, HRI awareness is the understanding that the human has of the location, activities, status, and surroundings of the robot; and the knowledge that the robot has of the human commands necessary to direct its activities and the constraints under which it must operate". Also they make the hypothesis that performance of the human-robot teams is greatly affected by HRI awareness violations. The hypothesis was tested by observing and analyzing the results of the AAAI 2002 Robot Rescue Competition. They inferred that all of the critical errors during the competition were due to some form of awareness violation, especially on human-robot spatial awareness. In addition to that, we argue that a key role in performance plays also the human-robot awareness regarding robot's activities and goals. Transparency in robot actions can reduce mental effort and improve the interaction as operator does not have to guess about robot's intentions [45]. An example of problematic interaction because of miss-interpretation of robot's intentions, is the tendency operators have to override robot actions if they feel that they are not contributing towards the goal, even if they are [60]. These findings highlight the importance of a shared world model between the robot and the operator [36]. The research presented in this section, was taken into account in our system design to ensure that HRI awareness will not be a confounding factor in our experiments.

2.1.3 Human-Robot interaction field studies

The use of robots on the field, especially in applications like SAR and hazardous environment inspection, is becoming more frequent. This lead some of the researchers to publish their experiences and lessons learned regarding HRI in the real world or realistic robot deployment exercises. One of the most important papers of this type, is the paper of Murphy and Casper [23]. In their seminal work they analyze data taken from the 9/11 disaster at the World Trade Center, where rescue teams equipped with various types of robots were deployed. The post-hoc analysis of the data showed several key points for the future of research in field robots used in time and safe critical situations. One of the major findings was that cognitive fatigue due to lack of sleep and pressure can cause serious errors and greatly affect the abilities of human operators. They arguably suggest that priority must be given to the development of intelligent robots that can actively assist the operator. Other findings include the lack of proprioception information (i.e. information regarding the state of the robot). Also in many cases the lack of different types of sensors mounted on the robot greatly impaired SA. The operators had lack of information about the robot and the world state, something that added extra cognitive workload.

Murphy et al. [70] distil experience from numerous SAR field studies and suggest four important lessons. From our perspective two of them are relevant. Firstly, they infer that the major bottleneck in robot autonomy is SA and not navigation. The requirements proved to be far beyond the current robot's capabilities for autonomous information gathering and perception. Secondly, HRI must focus on how to exploit the robot as an active information source rather than robot control itself. This is further supported from clues in the literature that humans cannot effectively perform two tasks at the same time (e.g. navigation and search), something very common in teleoperation applications [22]. Further clues in [102] gathered by the feedback given from operators, point out that the navigation and the search task when done in parallel impair SA. We argue that by automating some of the control tasks like navigation or by providing increased autonomy in critical moments, the operator can concentrate more in SA and use the robot more efficiently to gather information or perform other tasks in parallel. Our claim is supported by the findings of Yanco et al. [105]. In this study, when operators use autonomous navigation, report that they are able to concentrate more on acquiring SA. The above reported human limitations, are the main reason why the current standard is to deploy two operators in the field [14], one having the role of driving the robot and the other the role of problem holder [68], which is acquiring SA and making inferences. More autonomy will eventually lead to one operator per robot or one operator controlling multiple robots.

In [69], a road-map is been proposed for improving SAR robotics by providing increased automatic capabilities in the form of a basic mixed-initiative system which would be able to assist the operator in navigation and in victim search task. Casper and Murphy [22] conducted HRI work-flow field tests as part of a realistic fire rescue training session. One of their findings was that there are several patterns in teleoparation that the user follows and that these patterns can be automated by the robots. For example in a building stair search scenario the pattern that the operator follows is: âĂIJclimb (5-10 steps), stop, rotate to look in the corners and up at the ceiling, repeatâĂİ. Lastly another important finding is that many collisions while

driving the robot, are due to the lack of different kinds of sensors, operator disorientation and communication delay. All these could be improved by the use of more intelligent robots and more informative interfaces in terms of spatial awareness.
Lastly, the findings reported in this section are further supported by the latest research. Yanco et al. [107] provide an in depth HRI analysis of the DARPA robotics challenge trials. They found that teams using more autonomy in they robots, performed generally better from the ones using less. They argue that using more autonomous capabilities can reduce the amount of control effort by the human operators, and thus improving performance by allowing more interactions in a higher level of abstraction.

2.1.4 Metrics for Human-Robot interaction and mobile robots

To evaluate the effectiveness of HRI, a variety of metrics is needed. In order to address this problem, Steinfeld et al. [89] proposed a framework of common metrics, mostly focused on task-oriented mobile robots. They identify and group the metrics into categories as navigation, perception, management, manipulation and social. These categories and metrics are selected because they are applicable to the full range of operation (i.e. the spectrum between pure teleoperation and autonomy). Olsen and Goodrich [74] also propose some metrics regarding HRI and autonomy performance. These metrics are: task effectiveness, neglect tolerance, robot attention demand, free time, fan out and interaction effort. One of the most important metrics from our perspective is neglect tolerance. It is a measure of “how the robot’s current task effectiveness declines over time when the robot is neglected by the user". Another proposed metric is robot attention demand (RAD), which is a measure of how much time the robot is demanding. Simply, it is "the fraction of total task time that a user must attend to a given robot". Low RAD will ensure that the operator will have the minimum workload and he is free to attend to more important tasks. Also they point out the importance of context acquisition time which is the time the operator needs to update his SA when switching from one task to another. The latter is also a major factor for increased workload and bottleneck in working memory.

The issue of how to select and evaluate metrics is been addressed by Donmetz et al. [30]. They propose guidelines for selecting metrics for human-automation studies and ways to evaluate them. They distinguish between different categories of factors that a researcher has to take into account while designing metrics. These categories are: experimental constraints, comprehensive understanding, construct validity, statistical efficiency, and measurement technique efficiency.

2.2 Interfaces

Interfaces constitute a vital part of any remotely controlled robotic system, regardless of the LOA. As seen briefly in chapter 1, interfaces have been widely researched as a way to improve teleoperation. Although the literature is vast, in this section the most important aspects that constitute a good interface from the robotic control perspective are going to be examined.

2.2.1 Interfaces as a way to improve teleoperation

In [24], Chen et al. give an overview on how interfaces can tackle some of the most profound issues in teleoperated robots. However, in many cases interfaces tend to alleviate the “symptoms" rather than solving the root of the problem. For example limitations in telecommunication can cause video frame rate to drop; and delays between commands. These factors increase mental effort and force the operator to adopt a “move and wait" strategy [38] that decreases overall task performance [28]. For counterbalancing these effects, a common approach is an interface with a predictive display. Predictive displays, make use of kinematic models to predict the effect of a given command to the movement of a the robot and display it to the operator. Matheson et al. [67] conducted experiments in order to evaluate interfaces that make use of predictive displays under command delay conditions. They report that predictive displays improve driving performance. Other studies by Ricks et al. [78] also suggest that predictive displays improve performance on a navigation task and reduce error. However, this type of approach can get as good as the model used to predict locomotion. Such type of models are imperfect since they cannot simulate fully the complexity of the real world and may fail under unexpected conditions. On the other hand, for this and for many other cases in which interfaces were studied as an improvement (e.g. communication failures), using increased autonomous capabilities give a more complete solution. This is due to the fact that instead of just improving teleoperation, they can take away almost completely the burden of control from the operator, and thus tackling the root of the problem.

2.2.1.1 Situation awareness and the use of maps

Interfaces are particularly useful when operating a remote robot as the main source of SA acquisition. This use, highlights the importance of a good interface, as operators tend to spend around 30% of their time not performing the task but gaining SA [105]. A wisely designed interface can help the operator to control the robot more easily, with less errors, and with lower workload due to the improved SA. According to Goodrich et al. [45] a good interface should apply some kind of attention management by highlighting and inform the human about important events. Also should remove the cognitive burden from the operator by translating the sensor data into some kind of meaningful world representation. The latter implies the use of a map constructed from fused sensor data, that provides the user with a world model. Human-Robot Interaction analysis that took place during the AAAI Robot Rescue Competition [103], suggests that the operator must be provided with more spatial information about the environment, preferably by the use of a map. Teams that used maps were more successful in navigation than the teams that used only the video streaming. For these reasons the use of both map and video feedback in mobile teleoperated robots is considered a standard. There are two main categories for interfaces that make use both of video and maps. One category is video-centric interfaces, in which video is the primary feature for acquiring SA and occupies most of the display space. The other category is map-centric interfaces with the map as the prominent visual feature. Drury et al. [32] report that video-centric interfaces are better for surroundings and activities awareness while map-centric are better for location and status awareness. Interfaces with 2-D maps tend to fall into the video-centric category while interfaces with 3-D maps tend to be in the map-centric group.

The interest in interface research is shifting towards improving the maps, mainly with the use of 3-D, augmented reality and video integration. In [102] two interfaces are compared in a search task. One interface provides a 3-D augmented reality map in which information from the camera, map and robot pose are fused. It also gives the operator the ability to insert objects and labels into this map. The other interface is mainly comprised by a 2-D map and two video feedback windows (front and rear cameras). The results show that the 3-D interface performed better in terms of unknown area explored. Another example of this, is the work of Bruemmer et al. [11] in which a 3-D virtual map outperforms a 2-D map with video feedback in terms of errors and workload. Nielsen et al. [73] conduct a variety of experiments between a standard interface with a 2-D map and an 3-D augmented reality interface. The 3-D interface proved to outperform the 2-D interface in almost every scenario, with faster task completion, less coalitions and lower workload. Also the 3D interface improved the ability of the operator to perform a search task while navigating. Lastly, they propose 3 principals for designing better interfaces: "1) present a common reference frame; 2) provide visual support for the correlation between action and response; and 3) allow an adjustable perspective". They argue that the improved performance of the proposed 3-D interface in contrast to the 2-D interface, originate from these principles.

2.2.1.2 Various features for a good interface

It is commonly reported [102, 105] that operators were confused on where they had been previously (e.g. path followed) and what they have seen (e.g. landmarks). This suggests that a useful feature for the interface map would be to remember the paths that the robot has previously followed. Also the operator should have the ability to insert icons of landmarks or important locations to the map. In this way the operator will not have to remember and think on the locations previously visited.

Another commonly reported issue [105, 103] that can be solved by interfaces is the off-center camera, when pan and tilt cameras are used. Operators tend to forget to recenter the camera after using it for a search task, something that causes obstacle collisions and disorientation. This can be solved if camera's orientation is shown on the interface, an approach followed by Baker et al. [4] as they provided pant and tilt indicators.

One particular challenge when designing interfaces, is that care must be taken for not to overload the visual modality of the operator. This is due to the fact that operators tend to rely and use video streaming and other visual cues extensively for obtaining SA [102, 4]. A good example of this, is that because of the focus of the operator on an egocentric view (as video feedback) other salient visual information (e.g. from the map) may not be attended, something that is called "cognitive tunnelling" [90]. Only vital information and with the correct timing must be presented [82]. Lastly it is very important for this information to be integrated into a common frame of reference [73, 82], as this was identified by Thomas et al. [90] as the main reason for cognitive tunnelling.

2.3 Variable autonomy

In this section the reader will be firstly introduced to the concept of having different autonomy levels and modes that vary throughout the whole spectrum, between teleoperation and full autonomy. Then, the literature is being grouped into the main variable autonomy control categories which are traded control, shared control and multiple LOA. The most recent tendency is for robotic systems with many autonomy levels that cannot easily fit into one category or the other.

2.3.1 The notion of different autonomy levels

The case of a human acting as a supervisor to any semi-autonomous system by directing and monitoring it, is defined as supervisory control [86]. When supervisory control is applied to a teleoperated robot, then it is described as a telerobot [87]. Arguably the elimination of the notion of supervisory control in robotics used in the real world, is a challenging endeavour. However what is required, as argued in section 2.1.3, is a human supervisor/operator with

less responsibilities as robots are able to utilize increased autonomy and take initiatives.
A very important concept for this project and for robotics and AI in general, is the concept of Level of Autonomy (LOA). It is the degree in which the robot, or any artificial agent, takes its own decisions and acts autonomously. It can vary from the level of pure teleoperation (human has complete control of the robot), to the other extreme which is full autonomy (robot has control of every capability), within a single robot. The earliest work on a hierarchical stratification for the LOAs was developed by Sheridan and Verplank [88]. This stratification is based on three elements: decision making in regard to selecting actions; performing actions; and information shared by the computer to the human:

1. Human does the whole job up to the point of turning it over to the computer to implement.
2. Computer helps by determining the options.
3. Computer helps determine options and suggests one, which human need not follow.
4. Computer selects action and human may or may not do it.
5. Computer selects action and implements it if human approves.
6. Computer selects action, informs human in plenty of time to stop it.
7. Computer does whole job and necessarily tells human what it did.
8. Computer does whole job and tells human what it did only if human explicitly asks.
9. Computer does whole job and tells human what it did and it, the computer, decides he should be told.
10. Computer does whole job if it decides it should be done, and if so tells human, if it decides he should be told.

Endsley et al. [35] build on the above stratification, to propose an updated taxonomy. This taxonomy was meant to have a wider applicability to a range of tasks and domains. It was formulated on the basis of four functions: a) monitoring, which includes acquiring SA regarding system status; b) generating/formulating options or task strategies for achieving goals; c) selecting/deciding on a particular option or strategy and d) implementing/carrying out the chosen option through control actions at an interface. The taxonomy of Endsley et al. [35] is the following:

1. **Manual:** The human performs all tasks including monitoring the state of the system, generating performance options, selecting the option to perform (decision making) and physically implementing it.

2. **Action support:** At this level, the system assists the operator with performance of the selected action, although some human control actions are required. A teleoperation system involving manipulator slaving based on human master input is a common example.

3. **Batch processing:** Although the human generates and selects the options to be performed, they then are turned over to the system to be carried out automatically. The automation is, therefore, primarily in terms of physical implementation of tasks. Many systems, which operate at this fairly low level of automation, exist, such as batch processing systems in manufacturing operations or cruise control on a car.

4. **Shared control:** Both the human and the computer generate possible decision options. The human still retains full control over the selection of which option to implement, however, carrying out the actions is shared between the human and the system.

5. **Decision support:** The computer generates a list of decision options, which the human can select from, or the operator may generate his or her own options. Once the human has selected an option, it is turned over to the computer to implement. This level is representative of many expert systems or decision support systems that provide option guidance, which the human operator may use or ignore in performing a task. This level is indicative of a decision support system that is capable of also carrying out tasks, while the previous level (shared control) is indicative of one that is not.

6. **Blended decision making:** At this level, the computer generates a list of decision options, which it selects from and carries out if the human consents. The human may approve of the computerâĂŹs selected option or select one from among those generated by the computer or the operator. The computer will then carry out the selected action. This level represents a high-level decision support system that is capable of selecting among alternatives as well as implementing the selected option.

7. **Rigid system:** This level is representative of a system that presents only a limited set of actions to the operator. The operator's role is to select from among this set. He or she cannot generate any other options. This system is, therefore, fairly rigid in allowing the operator little discretion over options. It will fully implement the selected actions, however.

8. **Automated decision making:** At this level, the system selects the best option to implement and carries out that action, based upon a list of alternatives it generates (augmented by alternatives suggested by the human operator). This system, therefore, automates decision making in addition to the generation of options (as with decision support systems).

9. **Supervisory control:** At this level, the system generates options, selects the option to implement and carries out that action. The human mainly monitors the system and intervenes if necessary. Intervention places the human in the role of making a different option selection (from those generated by the computer or one generated by the operator); thus, effectively shifting to the Decision Support LOA. This level is representative of a typical supervisory control system in which human monitoring and intervention, when needed, is expected in conjunction with a highly automated system.

10. **Full automation:** At this level, the system carries out all actions. The human is completely out of the control loop and cannot intervene. This level is representative of a fully automated system where human processing is not deemed necessary.

Choosing the appropriate LOA is not a trivial task. One has first to answer precisely the question of what should be automated in a system and to what extend? Parasuraman et al. [75] propose a model for types and levels of automation in order to aid choices related to the above question. They define 4 different classes of types/functions of automation. Here they are being presented in regard to the project's research scope:

- **Information acquisition:** This includes automation of low level control of sensors such as the movement of a camera pan tilt unit. Also, information acquisition automation can be on a higher level, such as data fusion from sensors.
- **Information analysis:** Algorithms for Simultaneous Localization and Mapping (SLAM) or localization fall under this category. Also, obstacle representations and other information presented with a high level of abstraction belong to this category.
- **Decision and action selection**: Robot's ability to take initiative and change the LOA or perform an action is one of the most crucial aspects of this project.
- **Action implementation:** The second most crucial factor in designing the autonomy for this project, was the capabilities that allows the robot to take actions in the world such as navigation.

Within each of these categories, automation can vary from a low to a high level. According to Parasuraman et al. [75] "any particular level of automation should be evaluated by examining its associated human performance consequences". The proposed evaluation criteria in [75] are: the human performance consequences; automation reliability; and the costs of decision/action consequences.

The taxonomies presented so far are useful in their own right, as they can be used for every system or domain that uses some form of automation. However, as Beer et al.[7] point out, they can only inform HRI up to a certain level. The reason is that they are not robotics specific. Beer et al. [7] combine and extend these taxonomies to produce one that is HRI specific. In their proposed taxonomy [7] (see FIG. 2.1) the following factors play a key role: task and environment; robot autonomy that can dynamically change and that lies in a continuum; and HRI specific variables that influence robot autonomy.

2.3.2 Shared control

One of the most widely applied LOA in robotics is shared control, "the merging of teleoperation and autonomous control in real time during task execution" [3]. In shared control, input from the operator is blended in real time with the robot's calculated movement in order to produce an improved output. Gains from such controllers come mostly in the form of safety and increased accuracy on performing the task.

Arguably, robotic manipulators and arms were able to make the most in terms of performance and safety, with the use of shared control. A study using large-scale industrial manipulators, [49] demonstrated that shared control can increase the performance of novice operators compared to pure teleoperation, and decrease the workload in both experienced and novice operators. Similar gains are seen in telesurgery [6]. Surgeons movements are combined with automatic control to produce a more smooth and safe movement (e.g. filtering out hand tremor) for the robotic manipulator. This improves overall safety, gives patients less pain and faster recovery.

An early application of shared control to mobile robots, has been in the form of safeguard teleoperation (also known as safe teleoperation). In safeguard teleoperation the operator is driving the robot, but the robot controller reacts in order to prevent commands that are unsafe. Krotkov et al. [59] implemented a safeguard controller to a lunar rover in order to account for time delays between commands. In their field experiments, safeguard teleoperation improved performance and safety in an exploration task. In [41] a safeguard controller is proposed for mobile robots deployed in unstructured environments. Urdiales et al. [91] implement a shared control robotic system by coupling the human joystick input with the robot's trajectories based on local efficiency factors. They report improved performance in a navigation task and a more homogeneous efficiency distribution between participants.

Robotic wheelchairs is another mobile robot application that benefits from shared control. It

LORA	Sense	Plan	Act	Description
Manual	H	H	H	The human performs all aspects of the task including sensing the environment, generating plans/options/goals, and implementing processes.
Tele-operation	H/R	H	H/R	The robot assists the human with action implementation. However, sensing and planning is allocated to the human. For example, a human may teleoperate a robot, but the human may choose to prompt the robot to assist with some aspects of a task (e.g., gripping objects).
Assisted Tele-operation	H/R	H	H/R	The human assists with all aspects of the task. However, the robot senses the environment and chooses to intervene with task. For example, if the user navigates the robot too close to an obstacle, the robot will automatically steer to avoid collision.
Batch Processing	H/R	H	R	Both the human and robot monitor and sense the environment. The human, however, determines the goals and plans of the task. The robot then implements the task.
Decision Support	H/R	H/R	R	Both the human and robot sense the environment and generate a task plan. However, the human chooses the task plan and commands the robot to implement actions.
Shared Control With Human Initiative	H/R	H/R	R	The robot autonomously senses the environment, develops plans and goals, and implements actions. However, the human monitors the robot's progress and may intervene and influence the robot with new goals and plans if the robot is having difficulty.
Shared Control With Robot Initiative	H/R	H/R	R	The robot performs all aspects of the task (sense, plan, act). If the robot encounters difficulty, it can prompt the human for assistance in setting new goals and plans.
Executive Control	R	H/R	R	The human may give an abstract high-level goal (e.g., navigate in environment to a specified location). The robot autonomously senses environment, sets the plan, and implements action.
Supervisory Control	H/R	R	R	The robot performs all aspects of task, but the human continuously monitors the robot, environment, and task. The human has override capability and may set a new goal and plan. In this case, the autonomy would shift to executive control, shared control, or decision support.
Full Autonomy	R	R	R	The robot performs all aspects of a task autonomously without human intervention with sensing, planning, or implementing action.

Figure 2.1: The HRI taxonomy proposed by Beer et al.[7]. Adapted and reprinted from [7].

allows for increased safety without compromising the development of the user's operating skills. Carlson and Demiris [19] propose a shared control method which combines safe trajectories from an AI generator with user intention prediction based on joystick commands. They report increased safety when the user is occupied with a distracting secondary task. They also report increased performance in the secondary task. In [20], a human factors evaluation of a shared control wheelchair is presented. Findings based on secondary task reaction time and eye tracking, suggest that shared control results in reduced cognitive workload. Carlson et al. [21] present an adaptive shared control system which modulates the level of assistance based on user's current behaviour. They report increased performance and increased user acceptance. Research from Parikh et al. [76] reveals that users also tend to prefer shared control compared to teleoperation and autonomous control, when operating a robotic wheelchair.

2.3.3 Traded control

Another common approach to variable autonomy is the switch between the two extremes in the LOA scale (i.e. pure teleoperation and autonomy). This is called traded control. Such an approach aims at using autonomy in tasks that the robot is capable of performing but can be tedious or difficult at circumstances for the operator. Teleoperation is used in the tasks that the robot is not able to perform. This approach constitutes the base of our research. However, as we will describe in a later section, our research tackles the how, why, and when of trading control dynamically. Here we will present a brief overview of traded control.

Kortenkamp et al. [57] present an architecture for traded control in which the robot queries the operator each time it finishes a simple action. Mano et al. [64] constructed a traded control system for SAR. They heuristically define situations that will make use of autonomy mode or teleoperation. These heuristics are based on the task and the operation environment (e.g. presence of obstacles, wireless communication quality). However, they do not address how the switch will take place and who has the responsibility for triggering the switch. A principled way of trading between teleoparation and autonomy is investigated in [83]. Robots in an assembly scenario query the operator about trading control after a model based cost-benefit analysis. Human and robot performance models are explicitly trained by repeating the assembly scenario several times. Lastly, a control architecture is created [56] to be used in an underwater robotic vehicle. This architecture allows operators to switch between teleoperation and semi-autonomous operation, in which a human supervisor is giving commands at the mission level.

2.3.4 Multiple levels of autonomy

Research on variable autonomy also considers robotic systems that implement several LOAs. In such cases distinction between levels is less clear, e.g. safe teleoperation is used as a different LOA from shared control. Commonly four LOAs are implemented: a) teleoperation in which the operator has full control; b) a safe mode in which the operator teleoperates the robot but the robot can take initiative to protect itself; c) shared control (as above); and d) an autonomous or semi-autonomous mode in which the operator gives high level commands to the robot (e.g. specific goals). The operators are able to switch between these modes mostly based on their own judgment. A typical example is the system in [10]. A mobile robot is presented which is capable of switching between different levels of autonomy at the operator's command. The system aids operator's judgment by providing on screen indications of blockage or motion resistance. More research on multiple LOAs will be presented from this project's perspective later on.

2.4 Dynamically switching Levels of Autonomy

The use of different LOAs leads to a series of open challenges. Which LOA should be used under which conditions?; Who should initiate switches in the LOA and based on which factors?; and how can we investigate the trade-offs offered by switching LOAs in a repeatable manner?

The majority of the robotics literature is focused on describing the engineering and/or computational details of new technologies, while comparatively few studies address the issues of rigorously evaluating how well a human can use such robots to carry out a real task. Additionally, the autonomous robotics literature has historically tended to be somewhat separated and distinct from the literature investigating the issues of teleoperation, with relatively little work specifically focusing on variable autonomy systems.

Research which focuses on investigating dynamic LOA switching on mobile robots is fairly limited. Furthermore, the investigation of Mixed-Initiative (MI) systems to address this dynamic switching is even more limited, as highlighted by Jiang and Arkin [55] and Chiou et al. [25]. A large part of the literature, e.g. [59, 11], is focused on comparing the relative performance of separate LOAs, and does not report on the value of being able to switch between LOAs. In contrast, our work specifically addresses the issues of dynamically changing LOA on-the-fly (i.e. during task execution) using either a MI or Human-Initiative (HI) paradigm.

In this section, research addressing the strategies for switching autonomy levels is discussed. More specifically, the discussion aims at pointing out the gaps in the literature regarding three key aspects: a) conducting rigorous experiments using a principled scientific framework; b) LOA switching initiated by the human operator (namely HI); and c) LOA switching initiated both by the AI or/and the operator (namely MI).

In contrast to the literature presented here, and to the best of our knowledge, our work is the first that exploits rigorous methodologies from psychology and human factors research to carry out a principled study of variable autonomy in mobile robots; the first mobile robot experiments that combine quantifiable and repeatable degradation factors for both human and robot; and the first work which formally and systematically evaluates the benefits of combining the capabilities of both human and autonomous control in the context of dynamically mode-switching systems.

2.4.1 Conducting variable autonomy experiments

Surprisingly, previous research on variable autonomy in mobile robots lacks a rigorous experimental framework that will allow for meaningful and repeatable scientific inference. Characteristically much of the published experimental work does not carefully control for possible confounding factors. These factors can vary from partially uncontrolled test environments (as in [65]), up to the absence of standardized training for human test-subjects as in [11, 39, 9]. It is particularly important to control for the training and experience of human test-subjects, as these factors are known to affect overall robot operating performance [13, 2]. Additional confounding factors include the robot having different speed limits in the different conditions tested [39], or different navigation strategies of human operators [25]. In contrast to our work, Nielsen et al. [72] report no significant primary task results due to large measurement variances, but they do present a method for systematically categorizing the different navigational strategies of human operators.

All of the papers discussed above make important contributions in their own right, and we do not intend to devalue such work in any way. However, across the related literature we note a deficiency of: a) rigorous statistical analysis; b) clarity on assumptions and hypotheses; c) precise and detailed descriptions of the experimental protocol followed; d) a formalized, coherent and repeatable experimental paradigm. In contrast, in disciplines such as psychology and human factors, the above criteria constitute standard practice.

An excellent example of related work, which does provide a rigorous protocol, statistical analysis and detailed description, is the work of Carlson et al. [21]. They validate an adaptive shared control system, while degrading task performance with the use of a secondary task. However, their work is focused on the use of a Brain-Computer Interface for robot control. Because this field is relatively young, and the problems are extremely difficult, [21] used a robot navigation task which was comparatively simplified, i.e. operators only control left-right movement of a robot using a keyboard.

2.4.2 Human-Initiative variable autonomy

A Human-Initiative (HI) variable autonomy system, is a system in which the human operator can dynamically switch between the different LOAs (e.g. teleoperation and autonomy) [26]. In such systems only the human operator has the authority to initiate LOA switches based on his/her judgment. This judgment is often aided by suggestions made by the AI. The robot adopts a passive role without any kind of authority to initiate a LOA switch. This can be problematic in cases where the operator's judgment is impaired, e.g. if they have incomplete situation awareness (SA). Also it can be possible that the operator simply does not realize that a change in LOA is possible or beneficial (e.g. when under high workload). Additionally, the task of considering whether to switch LOA can add extra workload to the operator.

In [46] a system is presented with different LOAs. However, the initial LOA choice cannot change on the fly. Building on [10], Baker and Yanco [5] presented a robotic system in which the robot aids the operator's judgment by suggesting potential changes in the LOA. However, the system was not validated experimentally. Marble et al. [65] conducted a SAR-inspired experiment in which, similar to our experiments, participants were instructed to switch LOA in order to improve navigation and search task performance. However, [65] was intended to be a usability study which explored the ways in which participants interacted with each different LOA. In contrast, our own work is additionally focused on evaluating and demonstrating the overall task-performance when LOA levels can be dynamically switched. As in our own work, [65] also incorporate secondary tasks into their experiments. However, in contrast to our work, the use of these secondary tasks was opportunistic in nature because participants were only instructed to perform them optionally. Hence, the secondary tasks in [65] do not degrade human performance on the primary task (steering the robot). Also, unlike our work, [65] did not incorporate any methods into their experiments for degrading the robot's autonomous performance in a controlled way. In [85], a robot was presented which could navigate autonomously to way-points specified by a human operator. This paper suggested that the performance of such robots might be improved, by enabling a human operator to teleoperatively intervene in situations such as navigating narrow corridors, where the authors anecdotally reported difficulties with autonomous navigation. However, performance of this system were not experimentally validated in [85].

Although it is out of this thesis scope, variable autonomy research in the field of multiple robots being controlled by a single operator, provides similar studies. However much of this research (e.g. [50, 44]) is focused on higher levels of abstraction than our work, e.g. planning or task allocation. Other experiments, e.g [79, 92], are focused on human factors issues such as gaining SA when controlling multiple robots, or how the operator interacts with as many robot as possible.

2.4.3 Mixed-Initiative

2.4.4 Definition and taxonomy of Mixed-Initiative control

In [55] Jiang and Arkin present an elaborate definition of MI control in the context of human-robot teams. They define MI as:

> "A collaboration strategy for human-robot teams where humans and robots opportunistically seize (relinquish) initiative from (to) each other as a mission is being executed, where initiative is an element of the mission that can range from low-level motion control of the robot to high-level specification of mission goals, and the initiative is mixed only when each member is authorized to intervene and seize control of it."

They also present the first taxonomy for MI robotic systems. Their taxonomy has three dimensions:

- **Span-of-mixed-initiative** which characterizes the control elements (initiatives) in which both agents are capable of initiating actions.
- **Initiative reasoning capacity** which characterizes the ability of an agent to reason about taking the initiative.
- **Initiative hand-off coordination** which characterizes the strategies used by the system when shifting initiative from one agent to the other.

Regarding the span-of-mixed-initiative, our system presented in section REF HERE is *mostly-joint* as both agents have initiative capacity over two of the control elements, navigation execution and LOA (ADD ALSO THAT THE ROBOT FIRST PLANS AND THEN EXECUTES). Regarding the initiative reasoning capacity, the system is characterized as *hybrid*. It has the ability to reason about initiating actions deliberately based on an online performance metric and reactively based on sensor input. Lastly the initiated actions are communicated explicitly to the agents via the control interface and sound notifications. Thus, in the initiative hand-off coordination dimension our system is characterized as *explicitly-coordinated*.

2.4.5 Mixed-Initiative control related systems

Our literature survey found several systems characterized by authors as MI. However given the comprehensive taxonomy in [55], we believe that many of them cannot be characterized as truly MI control systems. This is due to the fact that the human-robot team does not share any initiatives, e.g. in [40] only the operator is initiating actions based on system's suggestions. In this survey we will only present systems that have some form of true MI control.
Shared control is a widely researched LOA falling under the banner of MI control. Any mixed-initiative is restricted inside the shared control LOA and is robot navigation execution explicit. For example in safeguard teleoperation the robot initiative takes place reactively to prevent collisions.
Nielsen et al. [72] conduct experiments using multiple LOAs. However, the LOA is chosen during the initialization of the system and cannot change on the fly. Moreover, similar to shared control, the robot has only reactive initiative inside a specific LOA to prevent collisions. Lastly, initiative is not coordinated by any hand-off strategy. Multiple robotic configurations using multiple LOAs (teleoperation, safe mode, shared mode, autonomy) are tested in [11]. In shared mode the robot drives autonomously while accepting interventions from the operator. In safe mode the robot takes initiative only to prevent collisions. However, these LOAs can not change on the fly and robot's initiative is limited in safe mode. Few et al. [39] present a control mode in which the operator is giving directional commands to adjust robot's navigation by

using the joystick. The system offers limited initiative which is depended on the frequency of operator's interaction.

Research on MI systems that are able to switch LOA dynamically or having initiative capabilities not restricted on specific LOA, is fairly limited. Moreover, the MI systems proposed are either theoretical or not experimentally evaluated. Bruemmer et al. [12] present a theoretical multiple LOAs, MI system. This system is based on "the theory of robot behavior" (human understanding of the robot) and "the theory of human behavior" (robot understanding of the human). The latter proposes the use of readily available non intrusive workload cues from operator as an indication of poor performance. More specifically, it proposes the use of the frequency of human input and the number and kind of dangerous commands issued by the operator. This provides the robot with the capacity to initiate switches between the different LOAs. However, it can be argued that input frequency is not necessarily an indication of poor performance. Rather it reflects different operatorsâĂŹ driving styles. Adams et al. [1] propose a MI robot control architecture which relies on the detection of operator's emotional state. Initiative is mixed in all the levels of the system, i.e. in setting goals and constrains, planning and execution. Changes in control are initiated based on the operator's sensed state (e.g. boredom, stress, drowsiness, engagement). This requires user-specific models that can be challenging and impractical to create. Having a working system that mixes initiative in all the levels of abstraction, is an extremely challenging concept for the current state of the research. Lastly, in contrast to our work, [1] is not proposing any hand-off coordination strategies and the system is not experimentally evaluated.

In multiple robots studies, variable autonomy often lies in a higher level of abstraction compared to our work. Manikonda et al. [63] describe a multiple robots MI controller and testbed for human-robot teams in tactical operations. The agents in the system share information towards a common model of the world and other agents behavior. Based on these information they are able to initiate modifications in their goals and associated roles in the team. In [50] a MI approach is proposed in a multiple robots search task. Robots are equipped with the ability to initiate changes in their respective search areas (e.g. size of search area). These changes are reactively triggered by specific events, e.g. the human operator has identified an item of interest.

In summary, MI robotic systems found in literature offer limited initiative inside a predefined LOA. In the case of multiple robots, the MI lies in a higher level of abstraction, making assumptions about other layers, e.g. navigation. Moreover, contrary to our work, initiative actions from the robot are not based on task performance metrics. The robot controller is rather taking initiative by reacting to sensor input (e.g. obstacles).

To the best of our knowledge, our work is the first to use an online task performance metric and a rigorous experimental framework, to address the problem of switching LOA during task execution using MI control. Also the first to show in a principled way the benefits to human operator cognitive workload, of a robotic system that initiates dynamic LOA switches.

2.4.6 Human-Robot interaction with LOA switching robots

TO BE UPDATED AFTER NEW REAL-WORLD EXPERIMENT As seen in the previous sections, research which focuses on investigating dynamic LOA switching on mobile robots is (perhaps surprisingly) very limited [25]. Furthermore, very little previous literature has attempted to rigorously evaluate variable autonomy systems which are able to switch LOA on-the-fly [26]. Consequently, human interaction with a variable autonomy system remains predominantly unexplored in the prior literature. Studies which address similar applications to ours, e.g. SAR, have evaluated how operators interact with user interfaces [103, 4]. Other studies explored the human operator's interaction with the robot in order to exchange information [42], but did not explore issues of control. Other studies investigated the human operator's interaction with a robotic system, but were restricted to exploring a single LOA [11], and did not explore the issues of variable LOA. Interesting studies of robotic wheelchairs, which exploited autonomous navigation capabilities by using a shared control (mixed initiative) architecture, measured the interaction of the operator with the collaborative control system based on joystick activity [18]. In contrast to the above literature, this work specifically investigates issues of the interaction of a human operator with a variable autonomy (multiple LOA) system, specifically the use by human operators of LOA switching capabilities.

In [5] a system was presented which aids the operator's judgment by automatically suggesting potential changes in the LOA. However, unlike our work, no data were presented on the operator's interaction with this LOA switching controller, because the system was not validated experimentally. As seen in section 2.4.2, the collaborative control for the robot presented in [85] was not experimentally validated and thus no HRI were presented. Marble et al. [65] conducted a SAR-inspired experiment in which participants were instructed to switch LOA in order to improve navigation and search task performance. In contrast to our work, [65] did not investigate the human operator's interaction with the a robotic system in which LOA levels can be dynamically switched.

To the best of our knowledge and in contrast to the literature reported here, our work is the first on mobile robots that reports a principled analysis of the ways in which human operators interact with, and exploit the capabilities of, a robotic system in which LOA modes can be dynamically switched.

2.5 Measuring and inducing cognitive Workload

Cognitive workload is one of the major reasons of performance degradation in humans when conducting a task. As discussed in section 2.1.3, robot operators often suffer from errors and performance degradation due to workload. Thus, inducing and measuring workload it is of importance when evaluating variable autonomy robotic systems.

Various methods exist for measuring workload. These methods can be categorized in three main classes [37]: physiological measures; subjective measures; task performance measures. In this section each of them is briefly discussed.

2.5.1 Physiological measures

Physiological measures are based on the assumption than workload will cause a physical reaction to the body of the person experiencing it. Such measures include respiratory activity, heart rate, brain activity, eye blink, eye movement and pupil dilation. They are direct objective measures and non-intrusive in regard to the task. However, they significantly raise the complexity of experiments and data analysis. The data are hard to collect; often require specialized equipment or processing; and are prone to noise. For these reasons, physiological measures have not been used in this work.

An example of a physiological measure that offers promising potential is Electroencephalography (EEG). EEG measures the electrical activity of the brain with electrodes placed in standardized positions in the scalp. It has been successfully used as an on-line measure of workload and validated in many studies [99, 97, 96, 58, 77]. Compared to other physiological measurements it has three key advantages: a) offers much better temporal resolution on the order of milliseconds; b) there are evidence that EEG outperforms some of the other techniques for measuring workload on-line [98]; c) it is a measure of brain activity and as such can potentially provide much more information regarding the cognitive state of the operator than just workload. The latter makes EEG very appealing as an input to adaptive automation. For example EEG can be used on the operator to determine if a stimulus has being perceived [100], recognize emotions [80] or measure task engagement [8]. Disadvantages of EEG include the fact that the signal is prone to artifacts (e.g. created by perspiration or movement) and noise and thus has to be recorded with care. Also, has day to day variability [27] and is highly depended on the individual and hence not easy to generalize.

2.5.2 Subjective measures

Subjective measures come in the form of questioners and are the easiest and least intrusive methods to investigate workload. Participants, after the experiments, are asked to rate the workload they experienced in one or more scales. Thus, subjective measures cannot provide real-time assessment of workload. Also they are highly subjective as they capture the individual differences between operators.

Such measures have been used extensively through out the experiments in the form of NASA Task Load Index (NASA-TLX) [52]. It is the most popular and well validated technique

[51] of subjective workload measurement. It rates perceived workload in order to assess a technology or system. NASA-TLX is considered a standard in the relevant human factors or teleoperation studies. It is multidimensional as it attributes workload to the following factors: Mental; physical; temporal demands; frustration; effort; performance. It applies a weighting scheme in order to decrease the variability between different participants. However, in our experiments we used what is called "raw NASA-TLX", which is simply a NASA-TLX without the weighting scheme. Raw NASA-TLX was found to be equally accurate, although there is an ongoing debate about its use [51]. The reason the raw version was used, is that the weighting scheme is very tedious and time consuming. This leads to participants answering in random some of the pairwise comparisons of the weighting scheme, as observed during our pilots experiments.

Lastly, Yagoda [101] proposed a workload questioner to be used in conjunction with NASA-TLX in order to provide more insight on specific HRI factors that affect workload (e.g. team process). This technique is new and not validated to the best of our knowledge. Hence it has not been used.

2.5.3 Task performance measures

Task performance workload assessment is done by measuring the performance on a primary or a secondary task. The assumption behind measuring workload with this method is that humans have limited cognitive resources. Thus, performance on a secondary or on the primary task are an indication of workload and will degrade with difficulty.

Our experimental paradigm uses a secondary task in order to primarily induce and secondary to measure operator's workload. Thus, the secondary task must be relevant and mentally interfere, but not interact, with the primary task. According to the Multiple Resource Theory [95, 94] this happens when two or more tasks share the same resources. For example the same perceptual modality and/or the same processing stages. The primary task of operating a robot through an interface, is mainly visual and involves spatial reasoning. Hence the secondary tasks used in the experiments are visual and/or require some sort of spatial reasoning.

2.6 Statistics

TO DO IF NEEDED? Bethel, C. L., Murphy, R. R. (2010). Review of Human Studies Methods in HRI and Recommendations. International Journal of Social Robotics, 2(4), 347âĂŞ359. http://doi.org/10.1007/s12369-010-0064-9

Human-Robot Interaction: A Survey -> this paper has some good practices in experimental HRI studies that we actually follow. Should be mentioned either here or on the lit review about how other studies suck in the experimental framework.

CHAPTER 3

Towards the Principled Study of Variable Autonomy

Conducting experiments requires a rigorous scientific framework for yielding meaningful, statistically validated results, and for repeatability. Our literature survey, as reported in section 2.4.1, found out (perhaps surprisingly) the absence of such framework in variable autonomy experimental work. Thus, our first step was a pilot-study experiment towards defining such an experimental paradigm. This framework allowed for the rigorous validation of the systems reported in later chapters.

This chapter is based on our Systems, Man, and Cybernetics conference paper [25]. It describes a pilot experiment in which a variable autonomy robot completes a navigation task. It explores the comparative performances of the human-robot system at different autonomy levels under different sets of conditions. This is done from a MI system investigation perspective. Sensor noise was added to degrade robot performance, while a secondary task induced varying degrees of additional workload on the human operator. Carrying out these experiments and analysing the results, has highlighted the profound complexities of designing tasks, conditions, and performance metrics which are: principled; eliminate confounding factors; and yield scientifically rigorous insights into the intricacies of a collaborative system that combines both human and robot intelligences. A key contribution of this chapter is the description of lessons learned from attempting these experiments, and a variety of suggested guidelines for other researchers to consider when designing experiments in this context. Lastly, these lessons and guidelines, informed a framework for the scientifically principled development of a system for changing LOA on the fly. This framework was used at the later experiments as it enables for more robust and clean experimental designs.

3.1 Initial Hypotheses

TO BE HEAVILY UPDATED OR USED IN OTHER SECTION There are two initial hypotheses that will guide the research and the experimental design initially. These two hypotheses are:

- The factors that affect operator's performance are cognitive workload and fatigue.
- The factors that affect robot's performance are: a) the uncertainty in information given (i.e. sensing) to an AI algorithm; b) the limited capabilities of an AI algorithm to process the given information; c) the complete absence of a specific algorithm and thus absence of a specific capability.

It is highly intuitive and also supported by the literature (as described in previous sections) that operator's performance are workload depended. One can argue plausibly that also SA affects performance. However, it is deficient compared to the workload hypothesis in two ways: a) a big percentage of workload is actually a product of mental effort to either acquire SA or maintain SA; b) contrary to workload, SA cannot be measured directly by the AI in real-time as it is either measured subjectively or by highly intrusive techniques such as SAGAT [**?**] that require operator to completely pause the task. Even with such methods, validation comes after the task is finished by comparing the real situation with the perceived SA. Given the perfect algorithm and enough information about the environment, then the robot should be able to always outperform the operator as it will always find the optimum solution and execute it without errors. In reality however this is not the case. The AI algorithms used to provide the robot with its capabilities have limitations and are highly dependent upon the reliability of the sensors measurements. For example the performance of a SLAM algorithm will degrade in the case that the laser range finder sensor provides noisy measurements. Another case is an algorithm that even given quality input, has limited capabilities and does not produce good results (e.g. a path planning technique can compute a non-optimal path).Lastly, the robot may not be able to perform an action at all. For example not be able to perform a search task without the appropriate robot vision technique.

3.2 System description

Our experimental system used a Pioneer-3DX mobile robot equipped with a laser range finder sensor and a camera. Remote control was achieved by a wireless link to the Operator Control Unit (OCU) which comprised a laptop connected to a screen showing the control interface (see Figure 3.1) along with a joystick and a mouse. There is a rich literature on standards and guidelines for designing interfaces as seen in section 2.2. The interface of our system adopts these recommended standards and guidelines where possible. Note that the intention of this work was not to design a better interface, but to explore the performance variations of human versus autonomous control under various conditions.

Our system offers two LOAs: **Teleoperation:** the human operator drives the robot with the joystick, while gaining SA via a video feed from the robot's onboard camera. Additionally a 2D map generated by Simultaneous localization and Mapping (SLAM) using the laser, is displayed on the OCU. **Autonomy:** the operator gives high level navigation commands to the robot, which is responsible for executing them autonomously. The commands consist of the human operator clicking a desired destination on the 2D map. The robot autonomously navigates towards these destinations. The operator can switch between the two modes of operation at any time by simply using their chosen control method.

The system was developed in Robot Operating System (ROS). For autonomous navigation ROS's *navigation stack* was used. It is a robust state-of-the-art solution [66]. For SLAM

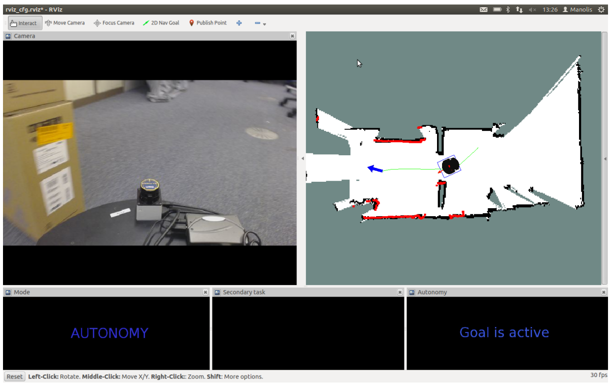


Figure 3.1: The control interface as presented to the operator. **Top**: video feed from the camera and the map of the environment. Inside the map, the position of the robot is visualized by the 3D model, the current goal by the blue arrow, the AI planned path by the green line, the obstacles' laser reflections by red and map walls with black. **Bottom**: the current autonomy mode, secondary task, and the status of the robot.

the *OpenSlam's GMapping* algorithm [48] through the ROS wrapper package called *slam gmapping* was used. Other in-house developed ROS software is responsible for teleoperation, LOA switching and providing information to the interface which is a tweaked version of ROS *rviz* package.

3.3 Pilot Experiment

The aim of the pilot experiment was to carry out a preliminary evaluation of how both human and machine intelligences perform on a navigation task, under ordinary and performance-degrading conditions. In particular, we sought to test how increased load on the robot (sensor noise) and increased load on the operator (more frequent secondary task demands) could affect performance during autonomous and teleoperation control.

3.3.1 Tasks and Robot Arena

The primary task was to drive the robot between two points in an obstacle course, as quickly and as accurately as possible (i.e. with the minimum number of collisions). This took place inside an arena of approximately 6.4 x 6 meters (see Figure 3.2). It is approximately equivalent

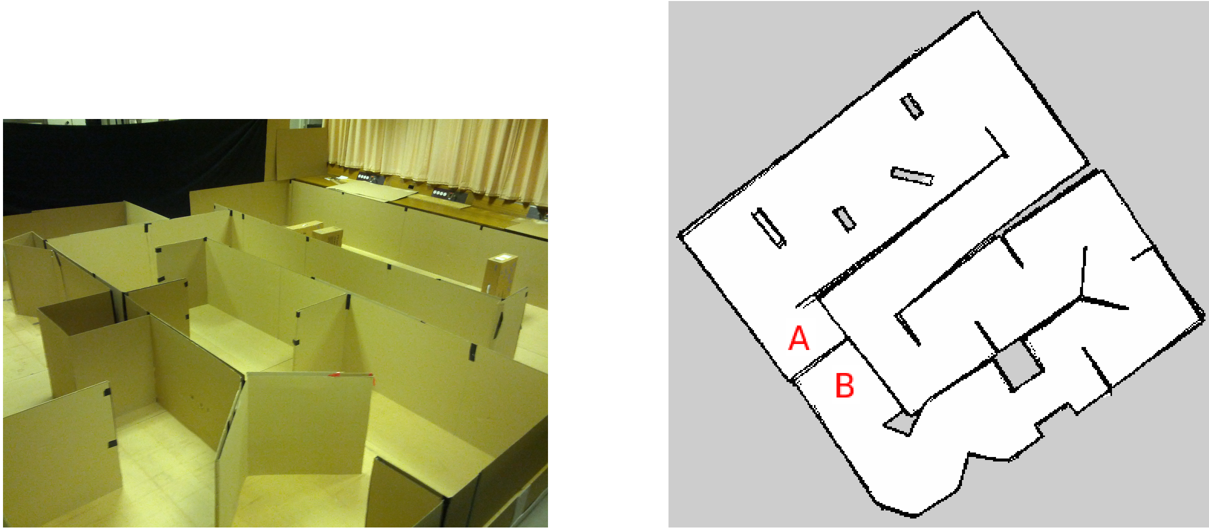


Figure 3.2: **Left**: the maze-like arena used in the experiment. **Right**: a SLAM map of the arena constructed using the robot's onboard laser, and displayed to the human operator on the OCU. Operators are asked to drive from point A to point B and then back again to point A.

to a yellow coded National Institute of Standards and Technology arena [54]. A secondary task was performed in parallel by the operator and was designed to induce additional workload. In the secondary task a target stimulus (a white dot on a black background) appeared at random intervals on the interface. The operator needed to respond to the presence of the target by pressing a joystick button. The difficulty of the task was controlled by the frequency with which the target appeared on the OCU screen. We considered the secondary task to be the performance degradation factor for the human operator, with two levels, low and high.

In terms of robot performance, the degradation factor (load) was the noise in sensor readings with two levels: low and high. High robot load was achieved by means of artificial Gaussian noise added periodically to the laser readings. In the low robot load, no artificial noise was added.

Four conditions were tested: **Teleoperation with high operator workload**: The operator had to manually drive the robot with a joystick while performing a high difficulty secondary task. Robot load was set to low. **Teleoperation with low operator workload**: The operator had to manually drive the robot with a joystick while performing a low difficulty secondary task. Robot load was set to low. **Autonomy with low load**: The operator gave high level commands for the robot to perform autonomously. He was only allowed to switch to teleoperation if it was absolutely necessary for the completion of the trial and only for a dictated period of time. For example if the robot was stuck in a corner, then the operator was required to switch to teleoperation to unstuck. Then, the operator was immediately required to switch back to autonomy. Once the robot was unstuck had to switch back to autonomy. Workload on the secondary task was set to low. **Autonomy with high load**: The operator gave high level commands for the robot to perform autonomously. He was only allowed to switch mode if it was necessary. Workload on the secondary task is set to low.

3.3.2 Participants and experimental design

The experiment had a between-group design with four groups corresponding to the experimental conditions. In total 28 participants took part, equally distributed among the groups when possible. This was based on their previous experience with driving, video games and operating robots.

All participants underwent a standardized training procedure. Before the start of the experimental session, participants were required to achieve a minimum performance standard. They had to complete a different obstacle course within a specific time limit, with no collisions, while not missing any secondary task responses. This ensured that all participants started the experiment having attained a minimum standard of proficiency.

Participants were instructed to perform both the primary task and the secondary task as quickly and accurately as possible to the best of their abilities. The map of the arena was not known to the robot or to the operators in advance. Before and during the experiments participants were not permitted to view the arena. Participants had to acquire all information about the arena from the video feed and a progressively acquired laser SLAM map, displayed on the OCU interface.

At the end of the session participants were asked fo complete an online NASA Task Load Index (NASA-TLX) [84] subjective workload/task difficulty questionnaire.

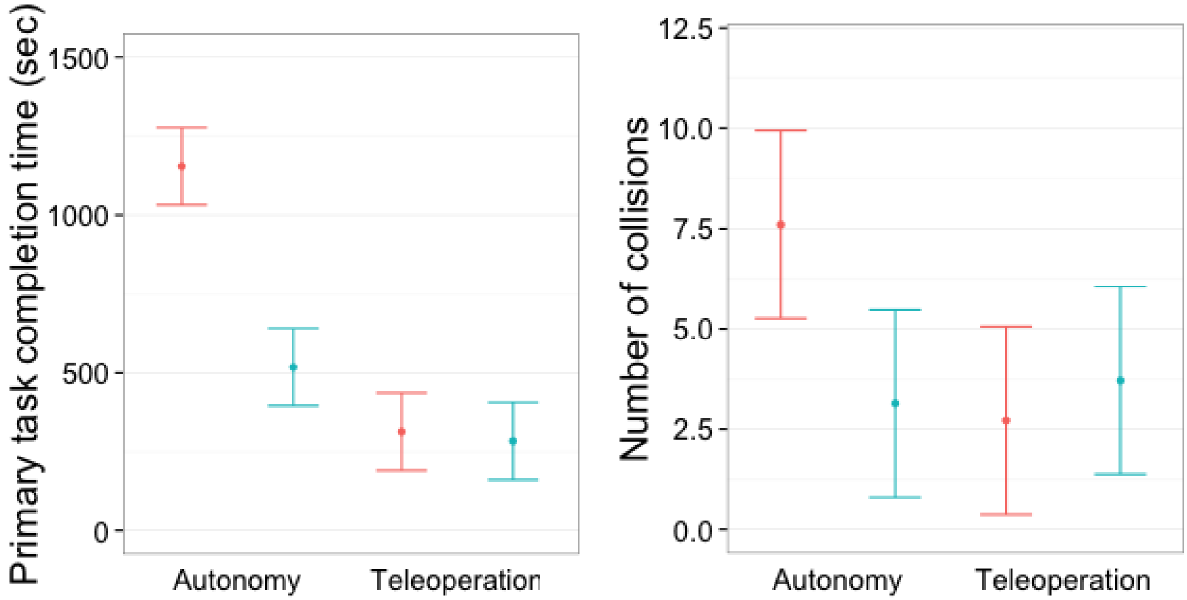
3.3.3 Results

Analysis was conducted on a series of metrics as elaborated in this section. A two-way ANOVA was conducted and Fisher's least significant difference (LSD) for the pairwise group comparisons. Data in some occasions violated ANOVA's assumptions for normality of distribution and homogeneity of variances. However, ANOVA has been proven to be robust in practice when such violations exist. The two factors are the control mode (autonomy and teleoperation) and load (low and high). We use the term "load" to describe the amount of degradation induced in the agent who is in control in each condition, i.e. load is the secondary task in teleoperation mode and added sensor noise in autonomy mode.

The effect of load on *primary task completion time* was significant, $F(1,22) = 15.525$, $p < .01$, as was the effect of mode, $F(1,22) = 40.377$, $p < .01$. The interaction of these two factors was also significant, $F(1,22) = 12.831$, $p < .01$ (see Figure 3.3a) and the statistical power was $power > .9$. Pairwise comparison revealed that high and low conditions in teleoperation did not have a significant difference. This means that secondary task difficulty does not seem to have an effect on speed performance. On the other hand autonomy-high with a $M = 1153.8$ performed much worse than autonomy-low ($M = 517.7$). The difference was significant at $p < .01$ suggesting that the added sensor noise had a big impact on autonomy performance. Autonomy-high was performed worse than teleoperation-high ($M = 313.5$) with significance $p < .01$. This means that the added noise degrades robot performance more than the high workload secondary task degrades the human operator's performance. A trend can be seen in which teleoperation-low ($M = 283.2$) performs better than autonomy-low ($M = 517.7$), but the difference is marginal $p = .052$.

Regarding the number of *collisions* (see Figure 3.3b), load, mode and their interaction did not have a significant effect ($power < .8$). However collisions seem to be trending towards autonomy-high having a higher number ($M = 7.6$) compared to the rest of the groups (teleop-high $M = 2.7$, teleop-low $M = 3.7$, autonomy-low $M = 3.1$). This is largely due to the fact that noise distorts the map and thus the robot's ability to autonomously navigate degrades.

Primary task score, compensates for the individual differences in speed-accuracy trade-off



(a) Primary task mean completion time per group. (b) Mean collisions among the different groups.

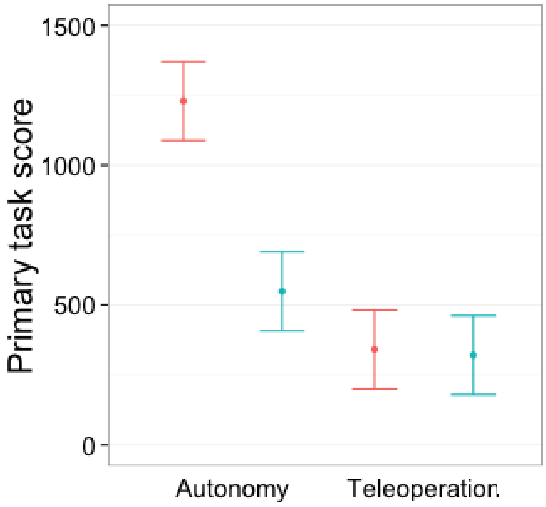
Figure 3.3: Red is high load, blue is low.

(i.e. time vs collisions). It was calculated by adding $10sec$ of penalty to the task completion time for every collision. With $power > .9$ the effects of load (*F(1,22) = 13.003,* $p < .05$), mode (*F(1,22) = 33.067,* $p < .01$) and interaction (*F(1,22) = 11.541,* $p < .05$) on primary task score were significant (see Figure 3.4a). The condition that had significant difference from the rest was autonomy-high ($M = 1229.8$, $p < .01$).

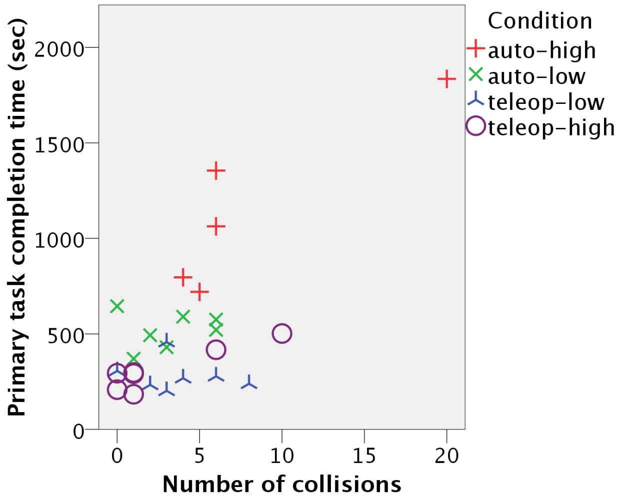
Individual participant performance on the two primary task metrics (see Figure 3.4b), shows no obvious groups. This is in terms of operator strategies favouring speed or accuracy over the other. It seems that there are some participants that perform generally worse and participants that perform generally better.

In terms of secondary task performance, i.e. *reaction time* (see Figure 3.5a), ANOVA ($power <$ .8) showed that the main effects for load and mode are not significant. The interaction however is significant with *F(1,22) = .4.638,* $p < .05$. Pairwise comparisons between the groups showed that the difference between teleoperation-high ($M = 0.588$) and teleoperation-low ($M = 0.717$) is marginally significant $p = .05$ with the rest of the comparisons not being significant.

The *rate of missed responses* (i.e number of missed responses over the number of targets presented on the display) in the secondary task was also measured. ANOVA with $power < .8$

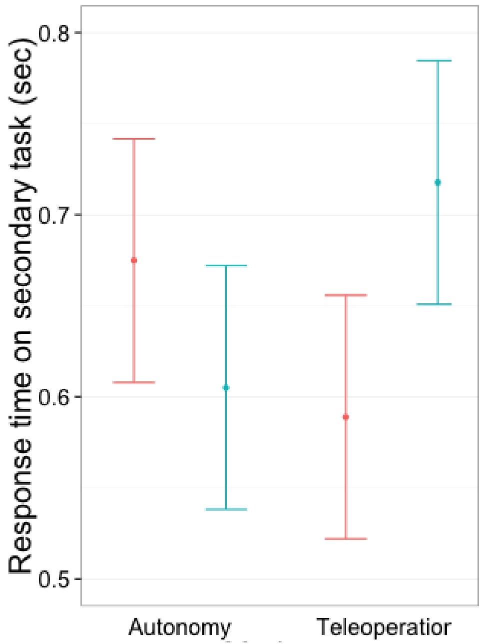


(a) Primary task score per group.

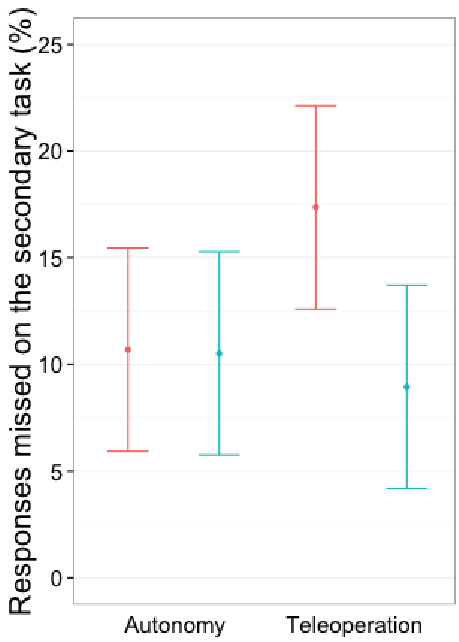


(b) Individual participant primary task performance.

Figure 3.4: In 3.4a red is high load, blue is low.



(a) Mean reaction time for each group.



(b) Percentage of missed responses for each group.

Figure 3.5: Secondary task performance. Red is high load, blue is low.

showed that differences were not significant. In Figure 3.5b the rate of missed responses is presented for reference. Playback of the recorded trials and informal discussion with the participants provided useful insight regarding missed responses. In most cases for teleoperation, they were occurred during a difficult navigation maneuver or when the operator was trying to acquire SA or both. During autonomy mode, greater rates of missed responses occurred during periods when a) AI performance dropped, e.g. the robot became stuck or lost localization; b) when a human operator was trying to infer what the robot was doing; c) when the operator was giving a command.

NASA-TLX scores suggest that participants found all of the different conditions equally difficult, as ANOVA ($power < .8$) showed no significant difference. This means that the degraded/high load trials were not difficult enough to be consciously perceived as such by the participants. It should be noted that the data from two participants in the conditions autonomy-high and autonomy-low were removed from this analysis as the trials were not completed.

3.3.4 Discussion

It has proven unexpectedly difficult to extract a clear analysis or definitive interpretation of the results. This is mainly due to the small number of participants and the high variance in performance between trials. There were a number of confounding factors that could have affected the latter and need careful consideration as discussed here and in section 3.4.

Firstly the degradation factors for the two agents did not caused an equivalent effect, as assumed during the experimental design. Results show that the sensor noise degrading effect on the robot, was higher than the effect of the high workload secondary task on the human. Thus, the secondary task needs to change or be adjusted to match the autonomy degradation factor in order for these to become meaningfully comparable. In general autonomy-high condition performs the worst in terms of primary task performance.

Counter-intuitively, participants were faster to respond to stimulus in the teleoperation high-workload condition than in the teleoperation-low condition. This can be due to the fact that participants may feel more engaged and alerted when they have to respond to a more frequent stimulus.

3.4 Guidelines for Future Experiments

This section presents observations and lessons learned for experimental design in the context of variable autonomy mobile robotics research. It is based on the results presented above.

1) Most of the arena should be simple for the robot to navigate autonomously. However, some parts should be impossible for the robot without human intervention. This will yield more significant results by avoiding variance in performance of the autonomy between trials.

2) The arena map should be fully known in advance, and presented to each participant in the same way with the same accuracy. This will minimize the variance in performance caused by different human operator strategies for exploring an unknown map.

3) Tasks have to be designed such that they genuinely necessitate the use of both teleoperation and autonomy in order to be better or successfully completed.

4) Extensive participants training makes experiments time consuming. On the other hand ensures participants trust the autonomous system and fully understand its capabilities and limitations. This can improve the human-robot interaction, and minimize variability between different participants. An example of a problematic interaction is the tendency that many human operators exhibit of overriding correct autonomous actions. This is because they mistakenly believe these actions are not contributing towards the goal [65].

4) A within-subjects design is more likely to control for individual differences. This is preferable over avoidance of possible training effects, given the variance of our current results. Moreover it is a more efficient use of resources.

5) SA affects performance but is very complex to measure. Measuring SA implicitly through task performance can be better suited than other methods in the current context.

6) Operator workload is difficult to measure in real time without using physiological techniques, such as electroencephalography [98]. Also, counter-intuitively, high workload does not necessarily cause task performance to degrade. Predictions on performance degradation during task execution, based on the measured workload level, can only be made if a correlation between them is found.

7) A MI system needs to *jointly* consider both (predicted) degraded task performance and the context and timing of the performance measurements. For example, if a human is attempting a very difficult manoeuvre or operation, this may produce degraded performance using a particular metric. However, giving control to the robot in this context could cause errors.

8) Degraded performance, jointly with context, might be simplified by designing experiments that use “idle time” as an additional performance metric. This can be defined as the time passed without any progress towards achieving a goal. In teleoperation, idle time includes times when an operator is neglecting the robot. In autonomy it includes time which passes without the AI actively achieving something, e.g. the robot sits stuck in a corner or stuck in front of an obstacle. This is similar to the “neglect time” metric proposed by Goodrich and Olsen [45]. However, it is simpler and more easily applied to multiple conditions and tasks.

3.5 Conclusion

Research in variable autonomy has predominantly focused on optimizing specific LOA, or comparing performance between them. A MI system which is optimally changing LOA on the fly, is an important domain that remains comparatively unexplored. Our initial experiment to address this provided some useful insights and interesting results.

The simple reaction time secondary task proved inappropriate, mainly because of its unclear impact on performance. On the other hand, the system’s performance degraded significantly with noise that distorts the robot’s model of the world. In order to objectively infer when a LOA switch is needed, the different experimental conditions should be able to degrade performance in a consistent and measurable way. A variety of different secondary tasks such as arithmetic tasks and SA tasks (e.g. the operator being asked to provide information to the experimenter during trials), can be investigated towards this direction. Also, tasks that can be cooperatively performed by both agents make potential candidates.

Designing and conducting principled experiments, which yield statistically meaningful insights in this context, is proving to be a challenging task. Mainly this is due to the intrinsic complexity of combining human factors with a robotic system. A multidisciplinary approach is required, taking into account robotics, human factors, psychology and HRI. There are a plethora of difficult confounding factors, that can distort results and make meaningful inference intractable. An important aim of this paper is to communicate to the research community the lessons learned, and guidelines for a principled experimental framework, that we have arrived at through carrying out our pilot-study experiment.

Constraining the physical conditions and the experimental design in appropriate ways might help in reducing complexity. However, such results might not be practically meaningful or useful if the environment is excessively sterile or artificially constrained. With the guidelines presented in this paper, we argue that a cleaner and more principled experimental design, rather than a restricted or sterilized one, can provide data that is meaningful to real-world robotic systems, but is also collected scientifically under measurable laboratory conditions.

CHAPTER 4

Experimental Analysis of a Variable Autonomy Framework for Controlling a Remotely Operating Mobile Robot

4.1 INTRO

This paper presents a principled experimental analysis of a variable autonomy control approach to mobile robot navigation. A Human-Initiative (HI) variable autonomy system is investigated, in which a human operator is able to switch the Level of Autonomy (LOA) between teleoperation (joystick control) and autonomous control (robot navigates autonomously towards waypoints selected by the human) on-the-fly. Our hypothesis is that the HI system will enable superior navigation performance compared to either teleoperation or autonomy alone, especially in scenarios where the performance of both the human and the robot may at times become degraded. We evaluate our hypothesis through carefully controlled and repeatable experiments using a significant number of human test-subjects.

The use of different LOAs in order to improve system performance is a challenging and open problem, raising a number of difficult questions. For example: which LOA should be used under which conditions?; what is the best way to switch between different LOAs?; and how can we investigate the trade-offs offered by switching LOAs in a repeatable manner? These questions need to be explored by conducting experiments within a rigorous multidisciplinary framework, drawing on methodologies from the fields of psychology and human factors, as well as engineering and computer science. Our previous work [25] highlighted the absence of such a framework in existing literature. Additionally it demonstrated the intrinsic complexity of conducting such experiments due to the high number of confounding factors and large variances in the results.

This paper develops from our previous work by designing and carrying out a principled experimental study to empirically evaluate the performance of a human-robot team when using a variable autonomy controller. More specifically it improves the experimental framework by: a) minimizing confounding factors, e.g. by using extensive participant training and a within-subject design; b) introducing a meaningful secondary task for human operators; and c) introducing a variable autonomy controller. We present formally-analysed, statistically-evaluated experimental evidence, to support the hypothesis that a variable autonomy system can indeed outperform teleoperated or autonomous systems in various circumstances.

In our experiments, we compare the performance of three different systems: 1) pure joystick teleoperation of a mobile robot; 2) a semi-autonomous control mode (which we refer to hereafter as the "autonomy" LOA) in which a human operator specifies navigation goals to which the robot navigates autonomously; 3) a Human-Initiative (HI) variable autonomy system, in which the human operator can dynamically switch between the teleoperation and autonomy modes using a button press. During experiments, human test subjects are tasked with navigating a differential drive vehicle around a maze-like test arena, with SA provided solely by a monitor-displayed control interface. At various points during the experiments, the robot's performance is degraded by artificially introducing controlled amounts of noise to sensor readings, and the human operator's performance is degraded by forcing them to perform a cognitively complex secondary task.

The experiments reported in this paper focus on the ability and authority of a human operator to switch LOA on the fly, based on their own judgement. We define this form of variable autonomy as Human-Initiative (HI), in contrast to Mixed-Initiative (MI) systems in which both the AI and the operator have the authority to initiate LOA changes. However, towards the end of this paper we additionally make suggestions for how the data, results and insights gathered during these experiments could be used to inform the design of a Mixed-Initiative (MI) system in future work.

4.2 APPARATUS AND ROBOTIC SOFTWARE

Our robot and environment were simulated in the Modular Open Robots Simulation Engine (MORSE) [34], which is a high fidelity simulator. The robot used was a Pioneer-3DX mobile robot equipped with a laser range finder sensor and a RGB camera. The robot is controlled by the Operator Control Unit (OCU), composed of a laptop, a joystick, a mouse and a screen showing the control interface (see FIG. 4.1).

In our previous work [25] we built a large maze-like test arena (see FIG. 4.2b and FIG. 4.3b), and carried out human-subject tests using a real Pioneer-3DX robot fitted with camera, laser scanner and WiFi communication to the remote Operator Control Unit. While demonstrating new methods on real robots is important, we observed that this can introduce difficult confounding factors, which can detract from the repeatability of experiments and the validity of collected data. For example, tests at different times of day or different weather, mean that daylight levels inside the lab change, affecting the video images observed by each test-subject. Different amounts of battery charge can cause top speed of the robot to vary slightly between different test-subjects. These and other factors led us to design the experiments reported in this paper using a high fidelity simulated robot and test-arena. As can be seen in FIG. 4.2 and FIG. 4.3, and comparing the real and simulated video feeds (FIG. 4.1 and FIG. 4.4), the simulation environment creates very similar situations and stimuli for the human operators as experienced when driving the real robot, but with a much higher degree of repeatability.

Our system offers two LOAs. **Teleoperation:** the human operator drives the robot with the

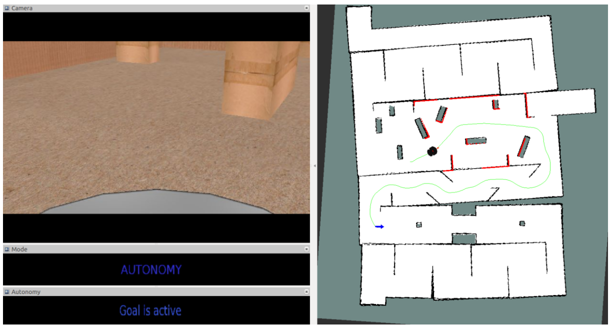
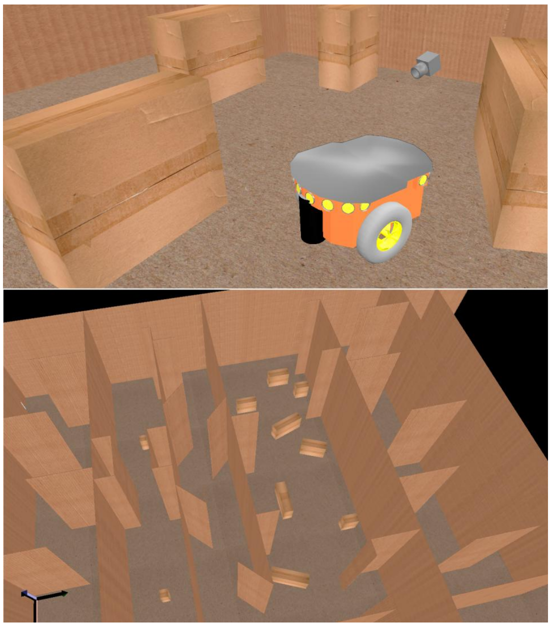


Figure 4.1: The control interface as presented to the operator. **Left**: video feed from the camera, the control mode in use and the status of the robot. **Right**: The map showing the position of the robot, the current goal (blue arrow), the AI planned path (green line), the obstacles' laser reflections (red) and the walls (black).

joystick, while gaining SA via a video feed from the robot's onboard RGB camera. Additionally a laser generated 2D map is displayed on the OCU. **Autonomy:** the operator clicks on a desired location on the 2D map, then the robot autonomously plans and executes a trajectory to that location, automatically avoiding obstacles. The system is a **Human-Initiative (HI)** system as the operator can switch between these LOAs at any time by pressing a joystick button. The software used was developed in Robot Operating System (ROS) and is described in more detail in [25].

4.3 EXPERIMENTAL DESIGN AND PROCEDURE

This experiment investigates to what extent circumstances in which the robot is under-performing, can be overcome or improved by switching control between the AI and the human operator. Such circumstances may include idle time, which is the time passed without any progress towards achieving a goal [25]. For example a robot being neglected by its operator when in teleoperation mode, or stuck due to a navigation failure in autonomy mode. Similar situations are quite common in real world robotics deployments [68]. For example, consider the case in which a robot operator must interrupt their control of the robot, to provide information to the SAR team leader or EOD team commander. Our hypothesis is that in such circumstances, trading control to another agent will improve the overall task performance of



(a) The simulated arena and the robot model used in the experiment.

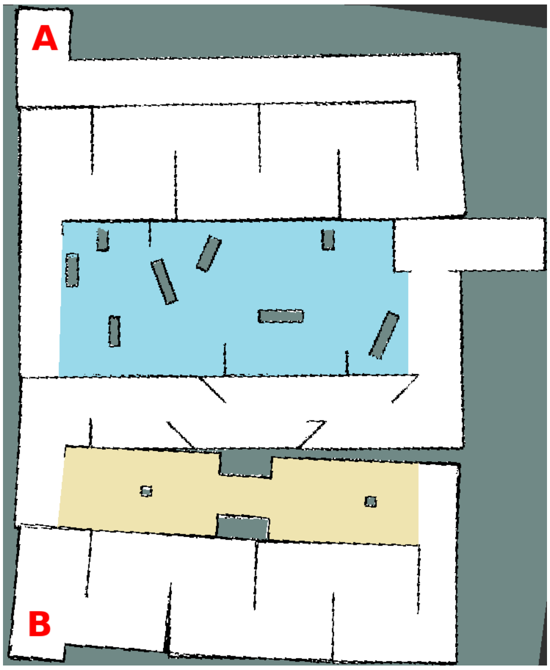
(b) The real arena and robot used in our previous experiment.

Figure 4.2: Note that the simulation recreates the real environment with a good degree of fidelity.

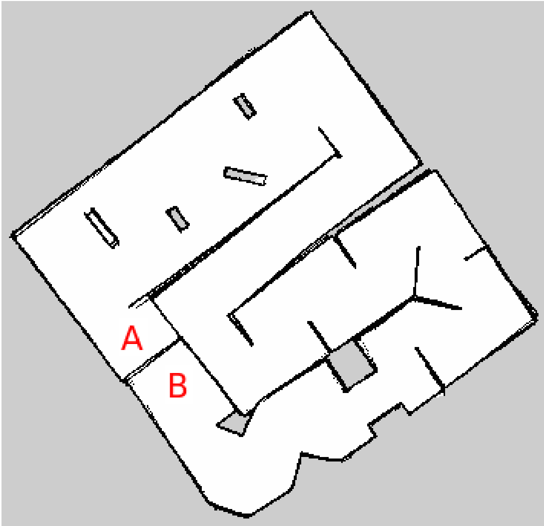
the system.

4.3.1 Experimental setup - operator control unit and robot test arena

In the work described in this paper, we used an identical OCU (see FIG. 4.4b) as that used in our previous experiments with a real robot [25]. A simulated maze was designed with dimensions of 11×13.5 meters (see FIG. 4.2a and FIG. 4.3a). It approximates a yellow coded National Institute of Standards and Technology arena [54]. As can be seen in FIG. 4.3b and FIG 4.4a, the data presented to the human operator via the OCU is almost identical to that experienced by human test subjects operating the real robot in a real arena in our prior work.

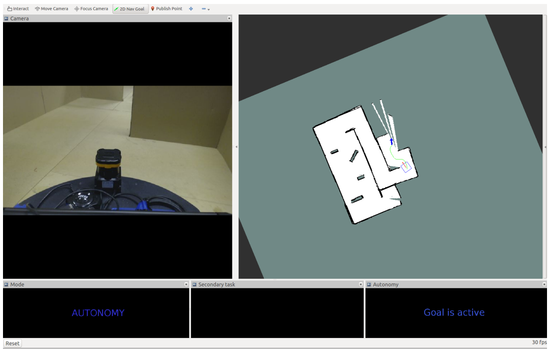


(a)

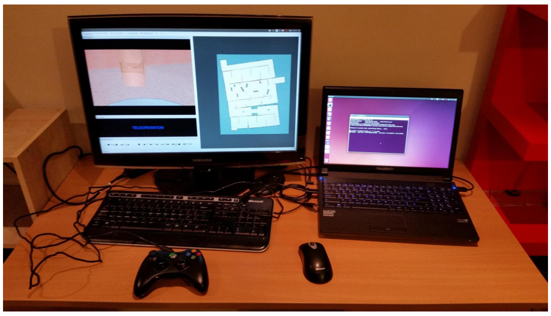


(b)

Figure 4.3: 4.3a: laser-derived SLAM map created in the simulation environment. Primary task was to drive from point A to B and back again to A. The yellow shaded region is where artificial sensor noise was introduced. The blue shaded region is where the secondary task was presented to the operator. 4.3b: laser-derived SLAM map generated by real robot in our previous experiment. Note the similarities between the real and simulated data.



(a)



(b)

Figure 4.4: 4.4a: the control interface as presented to the operator in our previous real world experiment. 4.4b: the Operator Control Unit (OCU), composed of a laptop, a joystick, a mouse and a screen showing the control interface. The same OCU was used in both experiments.

4.3.2 Primary and secondary tasks, and experimental test modalities

Each human test subject was given the primary task of navigating from point A in FIG. 4.3a (the beginning of the arena) to point B (the end of the arena) and back to point A. The path was restricted and one way, i.e. no alternative paths existed.

Two different kinds of performance degrading factors were introduced, one for each agent: artificially generated sensor noise was used to degrade the performance of autonomous navigation; and a cognitively intensive secondary task was used to degrade the performance of the human test subject. In each experimental trial, each of these performance degrading situations occurred twice, once on the way from point A to point B, and a second time on the way from point B back to point A. The two different kinds of degradations occurred separately from each other, as shown in FIG. 4.3a.

More specifically, autonomous navigation was degraded by adding Gaussian noise to the laser scanner range measurements, thereby degrading the robot's localization and obstacle avoidance abilities. For every experimental trial this additional noise was instantiated when the robot entered a pre-defined area of the arena, and was deactivated when the robot exited that area.

To degrade the performance of the human operator, their cognitive workload was increased via a secondary task of mentally rotating 3D objects. Whenever the robot entered a predefined area in the arena, the test subject was presented with a series of 10 cards, each showing images of two 3D objects (see FIG. 4.5). In half of the cards, the objects were identical but rotated by 150 degrees. In the other half the objects were mirror image objects with opposite chiralities. The test subject was required to verbally state whether or not the two objects were identical (i.e. yes or no). This set of 3D objects was previously validated for mental rotation tasks in [43].

For each human test subject, three different control modes were tested. In *teleoperation* mode, the operator was restricted to using only direct joystick control to steer the robot, and no use of the robot's autonomous navigation capabilities was allowed at any time. In *autonomy* mode, the operator was only allowed to guide the robot by clicking desired destinations on the

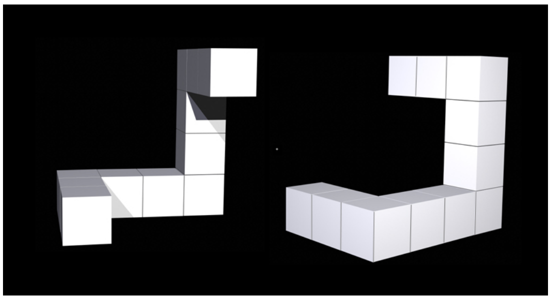


Figure 4.5: A typical example of a rotated 3D objects card.

2D map. The only exception was in the case of critical incidents such as the robot becoming stuck in a corner. Under such circumstances the experimenter would instruct the human operator to briefly revert to joystick control in order to free the robot so that the experiment could continue. In *Human-Initiative (HI)* mode, the operator was given freedom to switch LOA at any time (using a push-button on the joy-pad) according to their judgement, in order to maximize performance.

4.3.3 Participants and procedure

A total of 24 test subjects participated in a within-groups experimental design (i.e. every test subject performed all three trials), with usable data from 23 participants. A prior experience questionnaire revealed that the majority of the participants were experienced in driving, playing video games or operating mobile robots. Each test subject underwent extensive training before the experiment. This ensured that all participants had attained a common minimum skill level (which otherwise might lead to a confounding factor in later data analysis). Participants were not allowed to proceed with the experimental trials until they had first demonstrated that they could complete a training obstacle course three times, within a specific time limit, with no collisions and while presented with the two degrading factors (i.e. the secondary task and sensor noise). Each of the three training trials used a different control mode. Additionally, all participants were required to perform the secondary task separately (i.e. without driving the robot) in order to establish baseline performance.

During the actual experimental trials (testing the three different control modes), counterbalancing was used, i.e. the order of the three control modes was rotated (through six different possible permutations) for different participants. The purpose of this counterbalancing measure was to prevent both learning and fatigue effects from introducing confounding factors into the data from a within-groups experiment. Ideally, counterbalancing should have been done using 24 test-subjects (i.e. a multiple of 6). Unfortunately, due to technical reasons, only 23 out of our 24 human test-subjects yielded usable data, however our slightly imperfect counterbalancing over 23 subjects should still have eliminated most learning and fatigue effects from our statistical results. For the secondary task, different cards, but of equal difficulty [43], were used for each control mode, again to eliminate learning as a confounding factor in the test data.

Participants were instructed to perform the primary task (controlling the robot to reach a destination) as quickly and safely (i.e. minimising collisions) as possible. Additionally they were instructed that, when presented with the secondary task, they should do it as quickly and as accurately as possible. They were explicitly told that they should give priority to the secondary task over the primary task and should only perform the primary task if the workload allowed. Also they were told that there would be a score penalty for every wrong answer. This experimental procedure was informed by initial pilot study tests, with pilot participants, which showed that when people are instructed to "do both tasks in parallel to the best of your abilities", they either a) ignore the secondary task or b) choose random answers for the secondary task to alleviate themselves from the secondary workload, so that they can continue focusing on the primary task of robot driving. Lastly, participants were informed that the best performing individuals in each trial (using a weighted performance score based on both primary and secondary tasks) would be rewarded with a gift voucher. The purpose of this prize was to provide an incentive for participants to achieve the best score possible on both primary and secondary tasks.

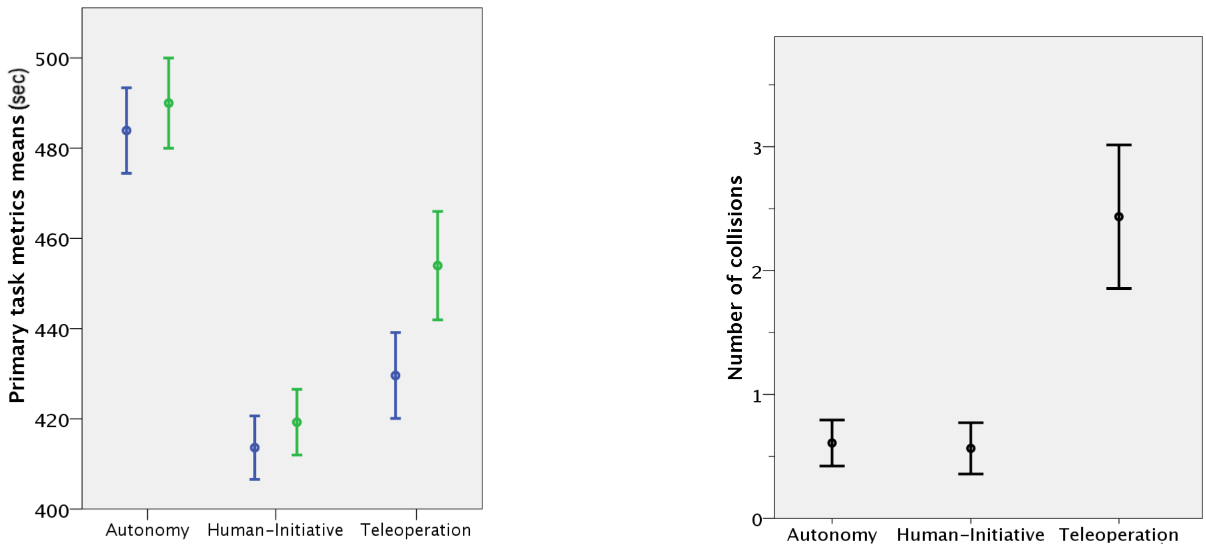
The human operators can only acquire situational awareness information via the Operator Control Unit (OCU) which displays real-time video feed from the robot's front-facing camera, and displays the estimated robot location (derived from laser scanner and SLAM algorithm) on the 2D SLAM map.

Our previous work [25] showed that a difficult confounding factor can be introduced by the fact that different test subjects may explore in different directions, thus revealing different information about the test arena at different times, as the robot's onboard laser SLAM progressively performs mapping. Additionally, real-time SLAM can produce maps of varying accuracy between trials. To overcome this confounding factor, all participants were given an identical and complete 2D map, generated offline prior to the trials by driving the robot around the entire arena and generating a complete SLAM map.

During each trial, a variety of data and metrics were collected: primary task completion time (time taken for the robot to travel from point A to point B and back again to point A (see FIG.4.1); total number of collisions; secondary task completion time; number of secondary task errors.

At the end of each experimental run, participants had to complete an online NASA Task Load Index (NASA-TLX) [84] questionnaire.

Statistical analysis was conducted on a number of metrics gathered during the experiments. A repeated measures one-way ANOVA was used, with a Greenhouse-Geisser correction in the cases that sphericity assumption was violated (i.e. that the variances of the differences between conditions/levels are not equal). The independent variable was the control mode with three levels. Fisher's least significant difference (LSD) test was used for pairwise comparisons given the a) clear hypothesis; b) predefined post-hoc comparisons; c) small number of comparisons. LSD is typically used after a significant ANOVA result to determine explicitly which conditions differ from each other through pairwise comparisons. Here we consider a result to be significant when it yields a p value less than 0.05, i.e. when there is less than a 5 percent chance that the observed result occurred merely by chance. We also report on the statistical power of the results. Power denotes the probability that a statistical significant difference will be found, if it actually exists. It is generally accepted that greater than 80 percent chance to find such differences constitutes a good power value. Lastly η^2 is reported as a measure of effect size.



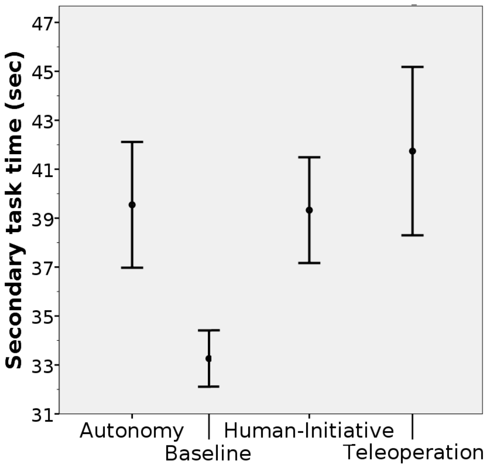
(a) (b)

Figure 4.6: Primary task results. 4.6a: average time to completion (blue) and score (green) combining time and collisions penalty. 4.6b: average number of collisions. In all graphs the error bars indicate the standard error.

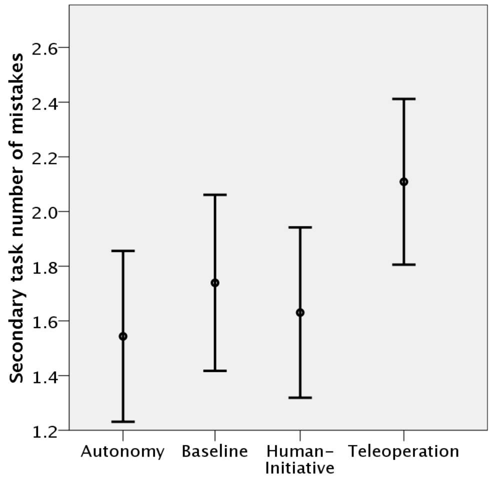
ANOVA for *primary task completion time* (see FIG. 4.6a) showed overall significantly different means with $F(1.275, 28.057) = 34.567$, $p < .01$, $power > .9$, $\eta^2 = .61$ between HI variable-autonomy ($M = 413.6$), autonomy ($M = 483.9$) and teleoperation ($M = 429.6$). Pairwise comparison reveals that pure autonomy performed significantly worse than the other two modes of operation with $p < .01$. Also HI variable autonomy performed significantly better than teleoperation ($p < .05$).

The effect of control mode on the number of *collisions* (see FIG. 4.6b) was significant, $F(1.296, 28.507) = 9.173$, $p < .05$, $\eta^2 = .29$ with a $power > .85$. Pure autonomy mode led to significantly ($p < .05$) fewer collisions ($M = .61$) than teleoperation ($M = 2.43$). HI variable autonomy mode ($M = .57$) also led to fewer collisions ($p < .01$) than teleoperation. HI and autonomy had no significant difference. Playback of the recorded trials revealed that in teleoperation most of the collisions occurred during the time of the secondary task. This was true for the participants that attempted to perform both tasks in parallel.

It is useful to be able to rank each trial according to an overall performance metric, which we refer to as the *primary task score*. This overall score is needed to be able to compare e.g. one human operator who achieves a very fast task completion time, but with many collisions,



(a)



(b)

Figure 4.7: Secondary task performance. 4.7a: average time to completion for one series of 3D objects. 4.7b: average number of errors for one series of 3D objects.

against another operator who achieves a slower time but with few collisions. We generate the primary task score by adding a time penalty, of 10 *sec* for every collision, onto the primary task completion time for each participant. This is inspired by the performance scores used in the RoboCup competitions [54]. FIG. 4.7a shows the mean primary task scores for each robot control mode. ANOVA analysis confirmed that control mode had a significant effect on the primary task score, *F(1.336, 29.403) = 19.342, p* $< .01$, $power > .95$, $\eta^2 = .47$. LSD test suggests that HI variable autonomy ($M = 419.2$) significantly ($p < .01$) outperforms both the pure autonomy mode ($M = 490$) and the pure teleoperation mode ($M = 453.9$). Note also that teleoperation appears to outperform autonomy ($p < .05$) in these experiments.

Secondary task completion time (see FIG. 4.7a) refers to the average time per trial, that the participants took to complete one series of the 3D object cards. ANOVA with $F(1.565, 34.420) = 7.821, p < .01, power > .85, \eta^2 = .26$, suggests that there is a significant difference between the mean secondary task completion times with and without also performing the primary task of controlling the robot. Participants performed significantly ($p < .05$) better in the baseline trial ($M = 33.2$) compared to their performance during robot operation. During robot operation, HI variable autonomy mode ($M = 39.3$), pure autonomy mode ($M = 39.5$) and teleoperation mode ($M = 41.7$) did not show statistical differences.

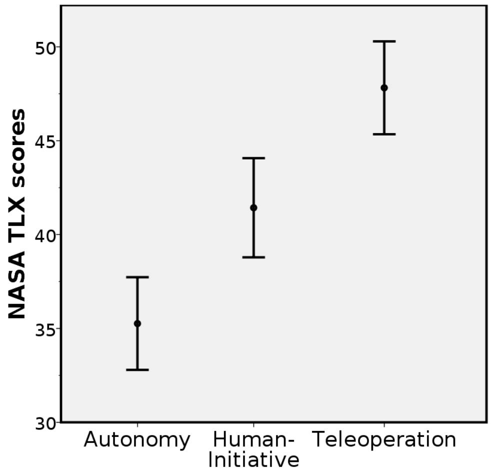


Figure 4.8: NASA-TLX score showing the overall trial difficulty as perceived by the operators.

No significant differences were observed between the different robot control modes with respect to numbers of secondary task errors (see FIG. 4.7b) according to ANOVA with $F(3,66) = 1.452, p > .05, power < .8, \eta^2 = .06$. Participants had $M = 1.7$ errors during baseline tests without operating the robot, $M = 1.6$ during HI variable autonomy mode, $M = 1.5$ in pure autonomy mode, and $M = 2.1$ in pure teleoperation mode.

Control mode had a significant effect on *NASA-TLX scores* (see FIG. 4.8) as suggested by ANOVA ($F(2,44) = 11.510, p < .01, power > .9, \eta^2 = .34$). Pairwise comparisons showed that autonomy ($M = 35.2$) was perceived by participants as having the lowest difficulty, as compared to HI variable autonomy mode ($M = 41.4$) with $p < 0.05$ and teleoperation mode ($M = 47.8$) with $p < 0.01$. HI variable autonomy is perceived as being less difficult than teleoperation ($p < 0.05$).

4.3.4 Discussion

In terms of overall primary task performance, HI variable autonomy control significantly outperformed both pure teleoperation and pure autonomy. This confirms our hypothesis that a variable autonomy system with the capability of on-the-fly LOA switching can improve overall performance of the human-robot team. In essence, it does so by being able to overcome situations in which a single LOA may struggle to cope. For example, external distractions to the operator such as the secondary task can be overcome by the operator switching from teleoperation to autonomy. In contrast, when autonomous control struggles to cope with noisy sensory information, the situation can be ameliorated by switching to teleoperation. From the Human-Robot Interaction (HRI) perspective, operators were able to successfully change LOA on-the-fly in order to maximize the system's performance. Since the LOA change was based on the operator's judgement, these experiments suggest that, given sufficient training, operators make efficient use of the variable autonomy capability. Additionally, note that autonomy generates significantly fewer collisions than teleoperation, however HI variable autonomy generates equally few collisions. This reinforces the conclusion that human operators can efficiently exploit autonomy by making smart decisions about switching between autonomy and teleoperation when most appropriate.

Regarding the secondary task, when performed in isolation from the primary task (during baseline testing), participants perform better. Since participants were instructed to focus on the secondary task whenever it was presented, this suggests that even having the primary task waiting on standby was enough to impair their performance on the secondary task. The absence of statistical differences across control modes in the secondary task time to completion and errors, suggests that a) the choice of control mode did not have any effect on secondary task performance; b) participants had the same level of engagement with the secondary task across trials.

NASA-TLX showed that autonomy is perceived as the easiest control mode, while HI is perceived as being easier than teleoperation. The fact that HI is perceived as more difficult than autonomy might perhaps reflect the cognitive overhead imposed on the operator by having to make judgements about switching LOA. This suggestion was further reinforced by observations made during trials and from informal conversations with participants. Most participants demonstrated a more laid-back attitude while using autonomy. However, participants stated that, while HI variable autonomy mode was "more stressful and demanding", it was also "more fun" due to a perception of increased engagement. For this reason, many participants expressed strong preference for HI variable autonomy over the other control modes. These observations are perhaps related to those of [93] which suggests that humans' "sense of agency" is improved when they interact more actively with a system.

4.4 THEORETICAL FRAMEWORK FOR DESIGNING A MI CONTROLLER

The results of these experiments yielded several insights for how to design a MI controller. The robot can be seen as a resource with two different agents having control rights: one agent is the human operator and the other is the robot's autonomous control system. At any given moment, the most capable agent should take control. Of particular importance is the ability of each agent to diagnose the need for a LOA change, and to take control (or hand over control) successfully. We assume that humans are able to diagnose when they need to intervene, given sufficient understanding of the system and the situation. On the other hand, it is not obvious how to enable the autonomous controller to detect when the human operator's performance is degraded, enabling the AI to robustly and automatically take control when it is needed. Automatic switching of control to an autonomous LOA would be important in situations where the human operator is too preoccupied with the primary cause of his or her performance degradation to voluntarily switch control to the robot.

In future work we propose to develop, test and analyse such an MI system. To make initial progress, it may be necessary to at first rely on naive assumptions, such as operators being willing to give control and the context and timing of a LOA change being appropriate [25]. We propose to carry out initial validation of our MI system using the same experimental design as reported in this paper, so that the MI can be compared against the HI system reported here. To be useful, the MI algorithm should provide the same level of performance or better, in terms of primary task completion, as compared to the simpler HI system.

Two different approaches are being investigated for the design of such MI algorithms. The first is focused on *task effectiveness*. The second is focused on using *machine learning techniques* on the HI data gathered during the experiments described in this paper.

In the first approach and more specifically in a navigation task, an online metric could express the effectiveness of *goal directed motion*. In the simplified case, this could be a function of speed towards achieving a desired goal position [74] or the number of collisions inside a thresholded time window. The general idea is that the metric should compare the current speed towards achieving a goal, with the optimal speed towards achieving the same goal.

Such proposals are limited, in that they rely on a variety of assumptions: the full map is known in advance, or the navigational goal lies inside an already known region; the robot's AI possesses a planner which is capable of reliably computing both the optimal path, and also the optimal velocities, from the current pose of the robot towards the goal; the agent to which the control will be traded, is capable of coping with the cause of performance degradation in the other agent.

An alternative approach is one of exploiting machine learning techniques in order to learn patterns of how human operators efficiently change LOA. The HI variable autonomy experiments reported in this paper, have enabled the collection of a variety of measurements that might be used as the training features of such a learning system. Such data includes: current mode of control at each time-step, positions and times of each change of LOA; time-stamped joystick logs; time-stamped series of velocity commands given to the robot; complete robot trajectories and information about periods of robot idle time.

4.5 Conclusion

This paper presented a principled and statistically validated empirical analysis of a variable autonomy robot control system. Previously, a comparatively small part of the robotics literature has addressed the issues of variable control. Previous studies have focused on the engineering and computer science behind building such systems; or on enhancing the human-robot interface; or investigated the ways in which humans interact with the system.

In contrast, this paper has made a variety of new contributions, including: showing how to carry out a principled performance evaluation of the combined human-robot system, with respect to completing the overall task; presenting clear empirical evidence to support the notion that variable autonomy systems may have advantages over purely autonomous or teleoperated systems for certain kinds of tasks; using rigorous methodologies, transferred from the fields of psychology and human factors research, to inform experimental design, eliminate confounding factors, and yield results that are statistically validated; demonstrates that human operators, when appropriately trained, make successful decisions about switching LOA, which efficiently exploit the contrasting strengths of both teleoperation and autonomous controllers. We must note here that our hypothesis and experimental paradigm are intended to be a starting point, from which more complex hypotheses and scenarios can be formulated. We believe this is the first study which has used truly scientifically repeatable experiments to support the continued development of variable autonomy mobile robots. Additionally, this paper has discussed the difficult issues involved in extending notions of variable autonomy from Human-Initiative (HI) to Mixed-Initiative (MI) robotic systems, and makes several suggestions for different approaches for building an autonomous MI switching algorithm. Developing such an MI system forms the subject of our ongoing research.

References

[1] Julie A. Adams, Pramila Rani, and Nilanjan Sarkar. Mixed initiative interaction and robotic systems. *AAAI Workshop on Supervisory Control of Learning and Adaptive Systems*, Tech. Rep.(WS-04-08):6–13, 2004. 32

[2] Miriam E. Armstrong, Keith S. Jones, and Elizabeth A. Schmidlin. Tele-Operating USAR Robots : Does Driving Performance Increase with Aperture Width or Practice ? In *59th Annual Meeting of the Human Factors and Ergonomics Society*, pages 1372–1376, 2015. 28

[3] Paul G. Backes. Dual-arm supervisory and shared control task description and execution. *Robotics and autonomous systems*, 12(1-2):29–54, 1994. 23

[4] Michael Baker, Robert Casey, Brenden Keyes, and Holly A. Yanco. Improved interfaces for human-robot interaction in urban search and rescue. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 3, pages 2960–2965, 2004. 5, 18, 33

[5] Michael Baker and Holly A. Yanco. Autonomy mode suggestions for improving human-robot interaction. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 3, pages 2948–2953, 2004. 29, 34

[6] G. H. Ballantyne. Robotic surgery, telerobotic surgery, telepresence, and telementoring. Review of early clinical results. *Surgical endoscopy*, 16(10):1389–402, 2002. 23

[7] Jenay M. Beer, Arthur D. Fisk, and Wendy A. Rogers. Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction. *Journal of Human-Robot Interaction*, 3(2):74–99, 2014. xi, 22, 24

[8] Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(5):B231–44, may 2007. 35

[9] David J. Bruemmer, Ronald L. Boring, Douglas A. Few, L. Julie, and Miles C. Waiton. âĂIJ I Call Shotgun !âĂİ: An Evaluation of Mixed-Initiative Control for Novice Users of a Search and Rescue Robot *. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 2847–2852, 2004. 28

[10] David J. Bruemmer, Donald D. Dudenhoeffer, and Julie L. Marble. Dynamic-Autonomy for Urban Search and Rescue. In *AAAI Mobile Robot Competition*, number WS-02-18, pages 33–37, 2002. 26, 29

[11] David J. Bruemmer, Douglas A. Few, Ronald L. Boring, Julie L. Marble, Miles C. Walton, and Curtis W. Nielsen. Shared Understanding for Collaborative Control. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 35(4):494–504, jul 2005. 17, 27, 28, 31, 33

[12] David J. Bruemmer, Julie L. Marble, Donald D. Dudenhoeffer, Matthew O. Anderson, and Mark D. Mckay. Mixed-initiative control for remote characterization of hazardous environments. In *36th Annual Hawaii International Conference on System Sciences*, pages 9 pp.–, 2003. 32

[13] David J. Bruemmer, Curtis W. Nielsen, and David I. Gertman. How training and experience affect the benefits of autonomy in a dirty-bomb experiment. In *The 3rd ACM/IEEE*

international conference on Human robot interaction (HRI), pages 161–168. ACM Press, 2008. 28

[14] Jennifer L. Burke and Robin R. Murphy. Human-Robot Interaction in USAR technical search: Two heads are better than one. In *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, pages 307–312, 2004. 13

[15] Jennifer L. Burke, Robin R. Murphy, Michael D. Coovert, and Dawn L. Riddle. *HumanâĂŞComputer Interaction*, (1-2). 6

[16] Darwin G. Caldwell, K. Reddy, O. Kocak, and A. Wardle. Sensory requirements and performance assessment of tele-presence controlled robots. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1375–1380, 1996. 5

[17] Darwin G. Caldwell, Andrew Wardle, and Michael Goodwin. Tele-presence: visual, audio and tactile feedback and control of a twin armed mobile robot. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 244–249, 1994. 5

[18] Tom Carlson and Yiannis Demiris. Collaborative Control in Human Wheelchair Interaction Reduces the Need for Dexterity in Precise Manoeuvres. In *Robotic Helpers: User Interaction, Interfaces and Companions in Assistive and Therapy Robotics, a Workshop at ACM/IEEE HRI 2008*, pages 59–66, 2008. 33

[19] Tom Carlson and Yiannis Demiris. Increasing robotic wheelchair safety with collaborative control: Evidence from secondary task experiments. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 5582–5587, 2010. 25

[20] Tom Carlson and Yiannis Demiris. Collaborative control for a robotic wheelchair: evaluation of performance, attention, and workload. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics*, 42(3):876–88, 2012. 25

[21] Tom Carlson, Robert Leeb, Ricardo Chavarriaga, and José Del R. Millán. Online modulation of the level of assistance in shared control systems. In *IEEE International*

Conference on Systems, Man and Cybernetics (SMC), number 2, pages 3339–3344, 2012. 25, 28

[22] Jennifer L. Casper and Robin R. Murphy. Workflow study on human-robot interaction in USAR. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1997 – 2003 vol.2, 2002. 13

[23] Jennifer L. Casper and Robin R. Murphy. HumanâĂŞRobot Interactions During the Robot-Assisted Urban Search and Rescue Response at the World Trade Center. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 33(3):367–385, 2003. 5, 12

[24] Jessie Y. C. Chen, Ellen C. Haas, and Michael J. Barnes. Human Performance Issues and User Interface Design for Teleoperated Robots. *Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 37(6):1231–1245, nov 2007. 5, 15

[25] Manolis Chiou, Nick Hawes, Rustam Stolkin, Kimron L. Shapiro, Jess R. Kerlin, and Andrew Clouter. Towards the Principled Study of Variable Autonomy in Mobile Robots. In *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*, pages 1053–1059, 2015. 6, 27, 28, 33, 39, 54, 55, 56, 57, 62, 67

[26] Manolis Chiou, Rustam Stolkin, Goda Bieksaite, Nick Hawes, Kimron L. Shapiro, and Timothy S Harrison. Experimental Analysis of a Variable Autonomy Framework for Controlling a Remotely Operating Mobile Robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016. 6, 29, 33

[27] James C. Christensen, Justin R. Estepp, Glenn F. Wilson, and Christopher A. Russell. The effects of day-to-day variability of physiological data on operator functional state classification. *NeuroImage*, 59(1):57–63, jan 2012. 35

[28] J. Corde Lane, C.R. Carignan, B.R. Sullivan, D.L. Akin, T. Hunt, and R. Cohen. Effects of time delay on telerobotic control of neutral buoyancy vehicles. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 3, pages 2874–2879, 2002. 15

[29] Lorin D. Dole, David M. Sirkin, Robin R. Murphy, and Clifford I. Nass. Robots Need Humans in the Loop to Improve the Hopefulness of Disaster Survivors. In *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN)*, pages 707–714, 2015. 2

[30] Birsen Donmez, Patricia E. Pina, and M. L. Cummings. Evaluation criteria for human-automation performance metrics. In *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, pages 77–82. ACM Press, 2008. 15

[31] Paul Dourish and Victoria Bellotti. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, number November, pages 107–114. ACM New York, 1992. 11

[32] Jill L. Drury, Brenden Keyes, and Holly A. Yanco. LASSOing HRI: analyzing situation awareness in map-centric and video-centric interfaces. In *2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 279–286, 2007. 16

[33] Jill L. Drury, Jean C. Scholtz, and Holly A. Yanco. Awareness in Human-Robot Interactions. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 912 – 918 vol.1, 2003. 11

[34] Gilberto Echeverria, Nicolas Lassabe, Arnaud Degroote, and Séverin Lemaignan. Modular open robots simulation engine: MORSE. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 46–51, 2011. 55

[35] Mica R. Endsley, D. B. Kaber, D.B. Endsley, M.R., and Kaber, Mica R. Endsley, and D. B. Kaber. Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3):462–492, 1999. 19, 20

[36] Xiaocong Fan and John Yen. Realistic cognitive load modeling for enhancing shared mental models in human-agent collaboration. In *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, pages 1–8. ACM Press, 2007. 12

[37] Eric Farmer and Adam Brownson. Review of workload measurement, analysis and interpretation methods. *European Organisation for the Safety of Air Navigation*, 33:1–33, 2003. 34

[38] William R. Ferrell. Remote manipulation with transmission delay. HFE-6(1):24–32, 1965. 15

[39] Douglas A. Few, David J. Bruemmer, and Miles Walton. Improved Human-Robot Teaming through Facilitated Initiative. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 171–176, sep 2006. 28, 31

[40] Alberto Finzi and Andrea Orlandini. Human-Robot Interaction through Mixed-Initiative Planning for Rescue and Search Rovers. In *Congress of the Italian Association for Artificial Intelligence*, pages 483–494, 2005. 31

[41] Terrence Fong, Charles Thorpe, and Charles Baur. A safeguarded teleoperation controller. In *IEEE International Conference on Advanced Robotics (ICAR)*, number No. LSRO2-CONF-2001-002, 2001. 23

[42] Terrence Fong, Charles Thorpe, and Charles Baur. Robot, asker of questions. *Robotics and Autonomous Systems*, 42(3-4):235–243, 2003. 33

[43] Giorgio Ganis and Rogier Kievit. A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes : Validation Data and Stimulus Set. *Journal of Open Psychology Data*, 3(1), 2015. 59, 61

[44] Michael A. Goodrich, Timothy W. McLain, Jeffrey D. Anderson, Jisang Sun, and Jacob W. Crandall. Managing autonomy in robot teams: observations from four experiments. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI)*, pages 25–32, 2007. 30

[45] Michael A. Goodrich and Dan R. Olsen. Seven principles of efficient human robot

interaction. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 4, pages 3942–3948, 2003. 12, 16, 50

[46] Michael A. Goodrich, Dan R. Olsen Jr., Jacob W. Crandall, and Thomas J. Palmer. Experiments in Adjustable Autonomy. In *IJCAI Workshop on Autonomy, Delegation, and Control: Interacting with Autonomous Agents*, pages 1624–1629, 2001. 29

[47] Michael A. Goodrich and Alan C. Schultz. Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007. 10

[48] Giorgio Grisetti, Cyrill Stachniss, and Wolfram Burgard. Improving grid-based slam with rao-blackwellized particle filters by adaptive proposals and selective resampling. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2432 – 2437, 2005. 42

[49] A. Hansson and M. Servin. Semi-autonomous shared control of large-scale manipulator arms. *Control Engineering Practice*, 18(9):1069–1076, 2010. 23

[50] Benjamin Hardin and Michael A. Goodrich. On Using Mixed-Initiative Control : A Perspective for Managing Large-Scale Robotic Teams. In *The 4th ACM/IEEE international conference on human robot interaction (HRI)*, pages 165–172, 2009. 30, 32

[51] Sandra G. Hart. Nasa-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006. 36

[52] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload*, 1(3):139–183, 1988. 35

[53] Peter F. Hokayem and Mark W. Spong. Bilateral teleoperation: An historical survey. *Automatica*, 42(12):2035–2057, dec 2006. 4

[54] A. Jacoff, E. Messina, B.A. Weiss, S. Tadokoro, and Y. Nakagawa. Test arenas and performance metrics for urban search and rescue robots. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 3396–3403, 2003. 43, 57, 64

[55] Shu Jiang and Ronald C. Arkin. Mixed-Initiative Human-Robot Interaction : Definition , Taxonomy , and Survey. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 954–961, 2015. 27, 30, 31

[56] T.W. Kim and J. Yuh. Development of a real-time control architecture for a semi-autonomous underwater vehicle for intervention missions. *Control Engineering Practice*, 12(12):1521–1530, 2004. 26

[57] David Kortenkamp, Debra Keirn-Schreckenghost, and R. Peter Bonasso. Adjustable control autonomy for manned space flight. In *IEEE Aerospace Conference proceedings*, volume 7, pages 629–640, 2000. 25

[58] Christian A. Kothe and Scott Makeig. Estimation of task workload from EEG data: new and current tools and perspectives. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*, volume 2011, pages 6547–51, jan 2011. 35

[59] Eric Krotkov, Reid Simmons, Fabio Cozman, and Sven Koenig. Safeguarded teleoperation for lunar rovers: From human factors to field trials. In *IEEE Planetary Rover Technology and Systems Workshop*, 1996. 23, 27

[60] G. J. M. Kruijff, Janicek Miroslav, Benoit Keshavdas, Shanker Larochelle, Hendrik Zender, N. J. J. M. Smets, T. Mioch, M. A. Neerincx, J. van Diggelen, F. Colas, M. Liu, F. Pomerleau, R. Siegwart, V. Hlavac, T. Svoboda, T. PetrÄścek, M. Reinstein, K. Zimmermann, F. Pirri, M. Gianni, P. Papadakis, A. Sinha, P. Balmer, N. Tomatis, R. Worst, T. Linder, H. Surmann, V. Tretyakov, S. Corrao, S. Pratzler-Wanczura, and M. Sulk. Experience in system design for human-robot teaming in urban search & rescue. In *Proceedings of 8th International Conference on Field and Service Robotics*, pages 1–14, 2012. 12

[61] S Lichiardopol. A Survey on Teleoperation. *Technische Universiteit Eindhoven, Department Mechanical Engineering, Dynamics and Control Group*, Tech. Rep.(DCT 2007.155), 2007. 2

[62] Yugang Liu and Goldie Nejat. Robotic Urban Search and Rescue: A Survey from the Control Perspective. *Journal of Intelligent and Robotic Systems*, 72(2):147–165, mar 2013. 6

[63] Vikram Manikonda, Priya Ranjan, Zachary Kulis, and Calhoun Drive. A Mixed Initiative Controller and Testbed for Human Robot Teams in Tactical Operations. In *AAAI Fall Symposium*, pages 92–99, 2007. 32

[64] Hayato Mano, Kazuyuki Kon, Noritaka Sato, Masataka Ito, Hisashi Mizumoto, Kiyohiro Goto, Renajit Chatterjee, and Fumitoshi Matsuno. Treaded control system for rescue robots in indoor environment. In *IEEE International Conference on Robotics and Biomimetics*, pages 1836–1843, 2009. 25

[65] Julie L. Marble, David J. Bruemmer, Douglas A. Few, and Donald D. Dudenhoeffer. Evaluation of supervisory vs. peer-peer interaction with human-robot teams. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, volume 00, pages 1–9, 2004. 28, 29, 34, 49

[66] Eitan Marder-Eppstein, Eric Berger, Tully Foote, Brian Gerkey, and Kurt Konolige. The office marathon: Robust navigation in an indoor office environment. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 300–307, 2010. 2, 41

[67] Adrian Matheson, Birsen Donmez, Faizan Rehmatullah, Piotr Jasiobedzki, Ho-kong Ng, Vivek Panwar, and Mufan Li. The Effects of Predictive Displays on Performance in Driving Tasks with Multi-Second Latency: Aiding Tele-Operation of Lunar Rovers. In *Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 21–25, sep 2013. 15

[68] Robin R. Murphy. Human-Robot Interaction in Rescue Robotics. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 34(2):138–153, may 2004. 2, 13, 56

[69] Robin R. Murphy, John G. Blitch, and Jennifer L. Casper. AAAI/RoboCup-2001 Urban Search and Rescue Events: reality and competition. *AI Magazine*, 23(1):37–42, 2002. 13

[70] Robin R. Murphy and Jennifer L. Burke. Up from the Rubble: Lessons Learned about HRI from Search and Rescue. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, 49(3):437–441, sep 2005. 6, 13

[71] Keiji Nagatani, Seiga Kiribayashi, Yoshito Okada, Kazuki Otake, Kazuya Yoshida, Takeshi Nishimura, Tomoaki Yoshida, Eiji Koyanagi, Satoshi Tadokoro, Mineo Fukushima, and Shinji Kawatsuma. Emergency Response to the Nuclear Accident at the Fukushima Daiichi Nuclear Power Plants using Mobile Rescue Robots. *Journal of Field Robotics*, 30(1):44–63, 2013. 2

[72] Curtis W. Nielsen, Douglas A. Few, and Devin S. Athey. Using mixed-initiative human-robot interaction to bound performance in a search task. In *IEEE International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pages 195–200, 2008. 28, 31

[73] Curtis W. Nielsen, Michael A. Goodrich, and Robert W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *IEEE Transactions on Robotics*, 23(5):927–941, 2007. 5, 17, 18

[74] Dan R. Olsen and Michael A. Goodrich. Metrics for evaluating Human-Robot Interactions. In *PERMIS*, 2003. 14, 68

[75] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans*, 30(3):286–297, 2000. 22

[76] S.P. Parikh, Jr. V. Grassi, V. Kumar, and Jr. Jun Okamoto. Usability Study of a Control Framework for an Intelligent Wheelchair. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 4745–4750, 2005. 25

[77] Lawrence J. Prinzel, Frederick G. Freeman, Mark W. Scerbo, Peter J. Mikulka, and Alan T. Pope. Effects of a Psychophysiological System for Adaptive Automation on Performance, Workload, and the Event-Related Potential P300 Component. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4):601–613, 2003. 35

[78] Bob Ricks, Curtis W. Nielsen, and M Goodrich. Ecological displays for robot interaction: a new perspective. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 3, pages 2855–2860, 2004. 15

[79] Jennifer M. Riley and Laura D. Strater. Effects of Robot Control Mode on Situation Awareness and Performance in a Navigation Task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 50, pages 540–544, 2006. 30

[80] Kristina Schaaff and Tanja Schultz. Towards an EEG-based emotion recognizer for humanoid robots. In *18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 792–796, sep 2009. 35

[81] Jean Scholtz. Theory and evaluation of human robot interactions. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*. IEEE, 2003. 11

[82] Jean Scholtz, Jeff Young, Jill L. Drury, and Holly A. Yanco. Evaluation of human-robot interaction awareness in search and rescue. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 2327–2332 Vol.3, 2004. 5, 18

[83] Brennan Sellner, Reid Simmons, and Sanjiv Singh. User modelling for principled sliding autonomy in human-robot teams. In *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III*, pages 197–208. Springer Netherlands, 2005. 25

[84] David Sharek. A Useable, Online NASA-TLX Tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55(1):1375–1379, 2011. 44, 62

[85] Jian Shen, Javier Ibanez-Guzman, Teck Chew Ng, and Boon Seng Chew. A collaborative-shared control system with safe obstacle avoidance capability. In *IEEE Conference on Robotics, Automation and Mechatronics*, volume 1, pages 119–123, 2004. 29, 30, 34

[86] Thomas B. Sheridan. Telerobotics. *Automatica*, 25(4):487–507, 1989. 5, 18

[87] Thomas B. Sheridan. Teleoperation, telerobotics and telepresence: A progress report. *Control Engineering Practice*, 3(2):205–214, 1995. 18

[88] Thomas B. Sheridan and William L. Verplank. Human and computer control of undersea teleoperators. *MIT Man-Machine Systems Laboratory*, 1978. 19

[89] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael A. Goodrich. Common metrics for human-robot interaction. In *1st ACM SIGCHI/SIGART conference on Human-Robot Interaction*, HRI '06, pages 33–40, 2006. 14

[90] Lisa C. Thomas and Christopher D. Wickens. Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(4):336–340, 2001. 18

[91] C. Urdiales, A. Poncela, I. Sanchez-Tato, F. Galluppi, M. Olivetti, and F. Sandoval. Efficiency based reactive shared control for collaborative human/robot navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3586–3591, 2007. 23

[92] Aberto Valero-Gomez, Paloma de la Puente, and Miguel Hernando. Impact of Two Adjustable-Autonomy Models on the Scalability of Single-Human/Multiple-Robot Teams for Exploration Missions. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(6):703–716, 2011. 30

[93] Wen Wen, Atsushi Yamashita, and Hajime Asama. *Plos One*. 67

[94] Christopher D. Wickens. Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, 3(2):159–177, 2002. 36

[95] Christopher D. Wickens. Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, jun 2008. 36

[96] Glenn F. Wilson, Jared D. Lambert, and Chris A. Russell. Performance Enhancement with Real-Time Physiologically Controlled Adaptive Aiding. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(13):61–64, jul 2000. 35

[97] Glenn F. Wilson and Chris A. Russell. Operator Functional State Classification Using Multiple Psychophysiological Features in an Air Traffic Control Task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(3):381–389, 2003. 35

[98] Glenn F. Wilson and Christopher A. Russell. Real-Time Assessment of Mental Workload Using Psychophysiological Measures and Artificial Neural Networks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4):635–643, 2003. 35, 50

[99] Glenn F. Wilson and Christopher a. Russell. Performance Enhancement in an Uninhabited Air Vehicle Task Using Psychophysiologically Determined Adaptive Aiding. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(6):1005–1018, dec 2007. 35

[100] Brad Wyble, Patrick Craston, and Howard Bowman. Electrophysiological Feedback in Adaptive Human Computer Interfaces. *Technical Report, Computing Lab, University of Kent at Canterbury*, (8-06), 2006. 35

[101] Rosemarie E. Yagoda. Development of the Human Robot Interaction Workload Measurement Tool (HRI-WM). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(4):304–308, sep 2010. 36

[102] HA Yanco, Michael Baker, R Casey, and Brenden Keyes. Analysis of human-robot

interaction for urban search and rescue. In *IEEE International Workshop on Safety, Security and Rescue Robotics.*, pages 22–24, 2006. 5, 13, 17, 18

[103] Holly Yanco, Jill Drury, and Jean Scholtz. Beyond Usability Evaluation: Analysis of Human-Robot Interaction at a Major Robotics Competition. *Human-Computer Interaction*, 19(1):117–149, jun 2004. 5, 16, 17, 33

[104] Holly A. Yanco and Jill Drury. Classifying Human-Robot Interaction: An Updated Taxonomy. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 3, pages 2841–2846, 2004. 11

[105] Holly A. Yanco and Jill Drury. "Where am I?" Acquiring situation awareness using a remote robot platform. In *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, volume 3, pages 2835–2840, 2004. 4, 13, 16, 17

[106] Holly A. Yanco and Jill L. Drury. A Taxonomy for Human-Robot Interaction. In *AAAI Fall Symposium on Human-Robot Interaction*, pages 111–119, 2002. 11

[107] Holly A. Yanco, Adam Norton, Willard Ober, David Shane, Anna Skinner, and Jack Vice. Analysis of Human-robot Interaction at the DARPA Robotics Challenge Trials. *Journal of Field Robotics*, 32(3):420–444, 2015. 5, 14

.1 Abbreviations

- Search and Rescue (SAR)
- American Association for Artificial Intelligence (AAAI)
- Artificial Intelligence (AI)
- analysis of variance (ANOVA)
- Electroencephalography (EEG)
- Human-Robot-Interaction (HRI)
- Level(s) of Autonomy (LOA)
- Mixed-Initiative (MI)
- Human-Initiative (HI)
- NASA Task Load Index (NASA-TLX)
- Operator Control Unit (OCU)
- reaction time (RT)
- Robot Operating System (ROS)
- Situation Awareness (SA)
- Simultaneous Localization and Mapping (SLAM)