Designing planning and control interfaces to support user collaboration with flying robots

Article *in* The International Journal of Robotics Research · March 2017

DOI: 10.1177/0278364916688256

Daniel Szafir
University of Colorado Boulder

Bilge Mutlu
University of Wisconsin–Madison

Article

The International Journal of Robotics Research
1–29

DOI: 10.1177/0278364916688256
journals.sagepub.com/home/ijr
SAGE

Designing planning and control interfaces to support user collaboration with flying robots

Daniel Szafir[1], Bilge Mutlu[2] and Terrence Fong[3]

Abstract

Robots are becoming increasingly prevalent and are already providing assistance in a variety of activities, ranging from space exploration to domestic housework. Recent advances in the design of sensors, motors, and microelectromechanical systems have enabled the development of a new class of small aerial robots. These free-flying robots hold great promise in assisting humans by acting as mobile sensor platforms to collect data in areas that are difficult to access or infeasible to instrument. In this work, we explored the design of interfaces that support users in working with free-flying robots to accomplish tasks including inventory logistics and management, environmental data collection, and visual inspection. Extending prior work in control interfaces for ground robots, we conducted a formative study in order to identify key design requirements for free-flyer interfaces. We designed several realistic tasks for use in evaluating human–robot interaction within the context of indoor free-flyer operation. We implemented three prototype interfaces that each provide varying degrees of support in enabling remote users to work with a flying robot to plan, communicate goals, accomplish tasks, and respond to changes in a dynamic environment. An experimental evaluation of each interface found that the interface designed to support collaborative planning and replanning using an interactive timeline and three-dimensional spatial waypoints significantly improved users' efficiency in accomplishing tasks, their ability to intervene in response to spontaneous changes in task demands, and their ratings of the robot as a teammate compared to interfaces that support low-level teleoperation or waypoint-based supervisory control. Our results demonstrate the utility of a data-driven design process and show the need for free-flyer interfaces to consider planning phases in addition to task execution. In addition, we demonstrate the importance of providing interface support for interrupting robot operations as unplanned events arise.

Keywords

Aerial robotics, cognitive human-robot interaction, cognitive robotics, cognitive robotics, control architectures and programming, field and service robotics, programming environment, simulation, interfaces and virtual reality, simulation, interfaces and virtual reality, simulation, interfaces and virtual reality, telerobotics, virtual reality and interfaces

1. Introduction

Robots can fulfill many functions, such as providing domestic assistance (Fong et al., 2003), mediating human communication (Rae et al., 2015), and working in automated assembly lines (Mandfield, 1989). This work focuses on the potential of a new class of *free-flying robots* to serve as collaborative assistants, enabling users to accomplish tasks that require high degrees of mobility. Free-flying robots hold great potential due to their unique abilities to freely traverse and survey environments. Unmanned aerial vehicles (UAVs) or 'drones', such as the General Atomics MQ-1 Predator and MQ-9 Reaper,[1] may provide aid in modern military operations (Sharkey, 2009) but have a number of limitations, including expense, size, and a reliance on fixed-wing flight, which present challenges to maneuverability and data collection. However, a new class of smaller, agile robotic systems designed to operate within human environments is rapidly emerging from a growing body of research on micro air vehicles (MAVs) (Krajník et al., 2011; Pfeil et al., 2013; Sharma et al., 2013), robotic airships (Bueno et al., 2002; Elfes et al., 1998), and space exploration robots (Fong et al., 2013). These free-flying robots are already providing assistance in tasks such as journalism (Corcoran, 2014), inspecting disaster sites (Reavis and Hem, 2011), and tracking sporting events from a unique vantage point (Lavigne, 2014). Furthermore, ongoing research

[1]University of Colorado Boulder, CO, USA
[2]University of Wisconsin–Madison, WI, USA
[3]NASA Ames Research Center, Moffett Field, CA, USA

Corresponding author:
Daniel Szafir, University of Colorado Boulder, 1111 Engineering Drive, Boulder, CO 80309, USA.
Email: daniel.szafir@colorado.edu

and development aims to expand the use of these robots into many additional domains in the near future, including construction (Irizarry et al., 2012), utilities (Wang et al., 2010), search and rescue (Goodrich et al., 2008), and space exploration (Bualat et al., 2015; Fong et al., 2012, 2013).

All current or near-future scenarios of use for free-flying robots require some degree of user control over the robot. For example, users will need to communicate with robots to convey their goals, assign tasks, and examine collected data. A substantial body of research has explored various forms of interaction for user control of mobile ground robots, examining issues related to robotic teleoperation (Ferrell and Sheridan, 1967; Sheridan and Verplank, 1978; Sheridan, 1986), proposing new paradigms such as supervisory or collaborative control (Fong et al., 2001b), and investigating theoretical aspects of task analysis, teamwork, interdependence, joint activity, and coactive design (Johnson et al., 2011). Prior work has also examined interface design for large fixed-wing UAVs, such as those used in military operations. However, relatively little is known regarding how to extend such work to the design of interfaces for free-flyers.

Current free-flyer interfaces, which typically afford manual teleoperation or waypoint planning via two-dimensional maps, suffer from several limitations. Teleoperation places a large burden on the operator and does not support concurrent work (Chen et al., 2007). Waypoint paradigms allow limited planning a priori, but they typically do not support dynamic replanning as task or environmental conditions change. Additionally, two-dimensional maps do not support the type of dynamic mobility in three-dimensional space that make free-flyers unique. Finally, both paradigms focus on supporting low-level aspects of robot navigation and obstacle avoidance, rather than supporting a human–robot team in accomplishing tasks as a whole, which requires combining aspects of navigation with planning and higher-level task execution strategies.

The goal of this work is to inform the design of free-flyer interfaces that can support users in accomplishing realistic mobile sensing and data collection tasks. To achieve this goal, we followed a user-centered, data-driven design approach that consisted of two main phases. First, we studied human–human teams to gain a better understanding of the types of support users need in collaborating with a distal assistant that can move throughout three-dimensional space while providing a video feed, pictures, and other sensor data. Second, we used the insights gained in the first phase to design three free-flyer interfaces that we evaluated in a laboratory experiment to understand the various trade-offs between designs that afford different forms of user support.

In this paper, we first review related work on developing robot control and management interfaces and discuss the current state of the art in free-flyer interface design. Next, we present a formative exploration aimed at better understanding user needs for support in working with a collaborator to accomplish the types of tasks with which free-flyers are envisioned to provide assistance. We then present the design and evaluation of three interfaces, two of which are inspired by current free-flyer interface designs and a third that is directly informed by results from our formative exploration. We conclude with a discussion regarding the implications of this work for the design of free-flyer interfaces, including trade-offs found across design paradigms and open questions that can guide future exploration in this space.

1.1. *Summary of contributions*

The primary goal of this work is to explore how various design elements might support user interactions with free-flying robots. In following this objective, we make the following contributions:

- present key design requirements and insights obtained through a formative study that address several limitations of existing methods;
- describe the design and implementation of three end-to-end free-flyer control interfaces, each of which extends existing state-of-the-art paradigms;
- detail an experimental evaluation of how well these interfaces support human–robot teaming across a range of realistic tasks;
- introduce a set of novel, concrete design guidelines for developers to enable more efficient and responsive interaction with free-flyers.

These contributions significantly advance our understanding of the interplay between interface design and operational performance for human interaction with free-flying robots.

2. Related work

Research in developing interfaces to support human–robot collaboration has a long history that intersects several fields including robotics, human–computer interaction, and cognitive psychology. In this section, we review related work on developing robot control and supervisory interfaces and then discuss current approaches toward developing interfaces for free-flying robots.

2.1. *Control, collaboration, and autonomy*

Many traditional approaches toward developing robotic systems have focused on replacing human work in certain applications. Such methods aim to introduce automation to reduce the need for humans to engage in undesirable, dull, repetitive, or dangerous work and make such work faster, safer, and cheaper. Early approaches often relied on the concept of *function allocation*, an attempt to characterize relative strengths and weaknesses of humans and machines to determine appropriate divisions of labor (Bradshaw et al., 2004). The notion of function allocation can be traced to early work in human factors from the 1950s (Fitts, 1951).

Table 1. Levels of automation popularized by Sheridan and Verplank (1978), adapted from Johnson et al. (2011) and Parasuraman et al. (2000).

Level	Description
High 10.	The computer decides everything, acts autonomously, ignoring the human.
9.	The computer informs the human only if it decides.
8.	The computer informs the human only if asked.
7.	The computer executes automatically, then necessarily informs the human.
6.	The computer allows the human a restricted time to veto before automatic execution.
5.	The computer executes a suggestion if the human approves.
4.	The computer suggests an alternative.
3.	The computer narrows the selection down to a few.
2.	The computer offers a complete set of decision/action alternatives.
Low 1.	The computer offers no assistance: human must take all decisions and actions.

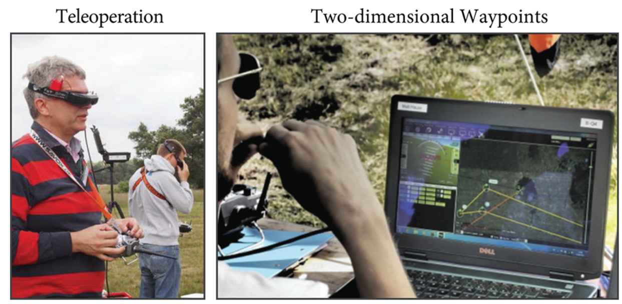

Fig. 1. Two main paradigms currently dominate free-flyer interfaces: direct teleoperation (left) and waypoint navigation using two-dimensional maps (right). Both images show users operating free-flyers at the 2014 International Micro Air Vehicles Conference and Flight Competition.

Contemporary work evaluating automation (Bright, 1955) and examining the impact of automated systems on the human labor force (Bright, 1958) proposed the notion of *levels of automation*. This notion first appeared in the form of a rubric with 17 levels describing aspects of how a mechanized system might respond to direct input from the user and changes in the environment, even anticipating and responding to user needs. This idea was later refined and popularized in pioneering work on the development of human control for manipulators in space exploration and underwater robotics throughout the 1960s, 1970s, and 1980s (Ferrell and Sheridan, 1967; Sheridan, 1986; Sheridan and Verplank, 1978). This body of work promoted the notion of 10 levels of automation ranging from *Level 1*, in which users are responsible for all work, to *Level 10*, where a fully autonomous system can perform the entire task without any input or intervention from the user (Table 1). As part of understanding how users might interact with systems of varying levels of automation, this work also advanced the concept of *supervisory control*, a paradigm in which humans allocate tasks to semi-autonomous systems, monitor system operation, and intervene when necessary.

The concept of supervisory control has inspired a great deal of research into new paradigms for working with robots by pushing back against the traditional notion of direct control in which robots must be manually 'teleoperated' by users. Several researchers have noted that the original concept of supervisory control was limited to defining a static distribution of work between a user and a system and thus have proposed notions of *adjustable* (Dorais and Kortenkamp, 2001), *sliding* (Dias et al., 2008), and *adaptive* (Parasuraman et al., 2000) autonomy as more dynamic approaches in which the distribution of work can be adjusted on the fly. *Mixed-initiative interaction* takes a similar approach in dynamic allocation of task elements with a focus on how agents (humans or machines) might negotiate to opportunistically determine appropriate roles to best accomplish the task (Allen et al., 1999). Common examples of mixed-initiative interactions in the human–robot interaction (HRI) literature include a robot requesting intervention from a human upon failing a task (Murphy et al., 2000) or determining what actions to pursue based on human emotional state (Adams et al., 2004). Further extensions promoting the model of 'human as collaborator', rather than 'human as controller', can be found in the paradigm of *collaborative control* (Fong et al., 2001b), which encourages human–robot dialog as a mechanism for adapting human–robot teamwork.

An alternative approach is represented by *shared control*, in which a robot can support direct teleoperation by attempting to predict user intentions while augmenting received inputs in a process that is invisible to the user. This approach has shown promise for users operating dextrous manipulators (Dragan and Srinivasa, 2013), surgical robots (Nudehi et al., 2003; Okamura, 2004), wheelchairs (Pires and Nunes, 2002), and brain–computer interfaces for robotic control (Mulling et al., 2015), although it faces the challenge of correctly deducing user intent from actual input. Shared control holds much in common with user-directed guarded motion (or 'safeguarded' motion), one of the three operational modes outlined by Murphy (2014) for disaster robotics, which also include direct teleoperation and waypoint delegation.

2.2. State of the art in free-flyer interfaces

Two main forms of interaction dominate the current trends in free-flyer interfaces (Figure 1). The first trend is represented by direct teleoperation using joysticks at a control station with a video display. While certain platforms attempt to increase operator situational awareness by providing first-person viewing of the robot's video feed, these systems – as with any teleoperation paradigm – are limited in several ways: they place a high burden on the operator; they do not support the user in achieving any other work concurrently; and they require users to have a great deal of skill in operating a robot that has potentially unfamiliar degrees of freedom. As an example of the skill required to teleoperate free-flyers, the Federal Aviation Administration in the United States initially considered regulations that would require commercial MAV operators to obtain a pilot's license (Harwell, 2014).

The second trend typically provides a two-dimensional overhead map view and allows users to plot waypoints outlining a free-flyer flight path. Popular interfaces include the Paparazzi Ground Control Interface (Brisset and Hattenberger, 2008) and the open-source Mission Planner waypoint delegation interface[2] that integrates with ArduPilot systems.[3] While these systems provide limited support for planning free-flyer operations, they typically do not support replanning or adjustments during flight and are inherently limited by providing a two-dimensional representation of the environment that does not match the three-dimensional capabilities of free-flyers. Additionally, these representations typically express flight plans at a large scale relative to the size of the robot, rather than a scale more appropriate for proximal operations within human environments.

Other interfaces combine a live video display with a two-dimensional map, allowing users to take manual control if problems arise during waypoint navigation, similar to large fixed-wing UAV control stations (Lorite et al., 2013). Combining video with map information is used not only for terrestrial MAVs, but also for the operation of the Smart-SPHERES free-flyers on board the International Space Station (ISS) (see Figure 2, adapted from Fong et al., 2013). However, simply presenting a map and a robot video feed may lead to error-prone context switching, as the user must attempt to make cognitive inferences regarding what is being shown in the video and the map representation. Furthermore, while users may be able to alternate between manual control and automatic waypoint navigation, utilizing these resources often requires transitioning between binary modes, which has been characterized by prior work as 'chaotic and a high risk activity' (Johnson, 2014).

Several approaches have sought to address the limitations of these systems, examining how to better overlay pictures and video onto maps (Drury et al., 2006; Goodrich et al., 2008), developing control systems using mobile devices or multi-modal interfaces (Quigley et al., 2004), and designing immersive interfaces in which the user wears a virtual head-mounted display and the free-flyer becomes the user's 'floating head' (Higuchi and Rekimoto, 2013). Additional work has advanced the notion of 'perceived first-order control', which bears similarities to shared control or user-directed guarded motion approaches, allowing users to 'nudge' a free-flyer using gestures on a mobile touchscreen to practice fine control while ensuring safe operation (Pitman and Cummings, 2012). Prior work has also combined a three-dimensional environmental map with a robot video feed while presenting a visualization of the robot's flight path that can be edited by the user (Johnson et al., 2012). This approach bears the greatest resemblance to the designs that we explore in the current work, particularly in the recognition of the importance of a three-dimensional representation of the work environment rather than a two-dimensional map.

All of the approaches discussed above are still limited in focusing solely on aspects of free-flyer navigation, which represent only one facet of successful human–robot collaboration (Bauer et al., 2008). While issues such as localization and hazard detection or avoidance are critical for the deployment of free-flyers, new research methods using advanced vision algorithms and embedded sensors are increasingly able to address low-level flight navigation problems (Achtelik et al., 2009; Bachrach et al., 2011; Scaramuzza et al., 2014; Winkvist et al., 2013). However, advancement in sensors or automation will not be able to address higher-level human–robot interdependencies regarding task and activity, such as how to make decisions regarding task procedure and priority or how to respond to unexpected events. Prior research in joint action has demonstrated that mechanisms for forming and sharing task representations are critical to collaborative work (Sebanz et al., 2006). Although research on human interaction with ground robots has begun to examine shared mental models of tasks (Huang et al., 2015; Mutlu et al., 2014), interfaces for free-flying robots designed to date are not task- or activity-oriented.

In this work, we argue that free-flyer interfaces should focus on supporting task and teamwork interdependencies to support users in accomplishing high-level environmental survey and data collection tasks, rather than solely the task of navigation. In this view, human operators and supervisors are essential teammates that can help to plan robot tasks, monitor execution, and intervene when necessary or desired, while systems support the user by automating low-level aspects of free-flyer navigation.

3. Research approach

Our investigation of the design space for free-flyer interfaces that support effective human–robot teaming builds on the *coactive design* methodology (Bradshaw et al., 2013; Johnson, 2014). The coactive design process, which focuses on supporting human–robot interdependence through *observability* ('making pertinent aspects of one's status, as well as one's knowledge of the team, task, and environment observable to others'), *predictability* ('one's actions should be predictable enough that others can reasonably rely on them when considering their own actions'), and *directability* ('one's ability to direct the behavior of others and complementarily be directed by others') (Johnson et al., 2011, 2014) has been successful in developing effective shared control systems.

In developing coactive design, Johnson (Bradshaw et al., 2013; Johnson, 2014) argues that prior approaches inspired by the levels-of-autonomy concept are fundamentally limited for several reasons. For example, improving autonomy often comes at the cost of increasing system opacity. Moreover, modifying the role of automation in joint human–machine systems changes the nature of activities and the interaction in complex ways (i.e., the 'substitution myth',

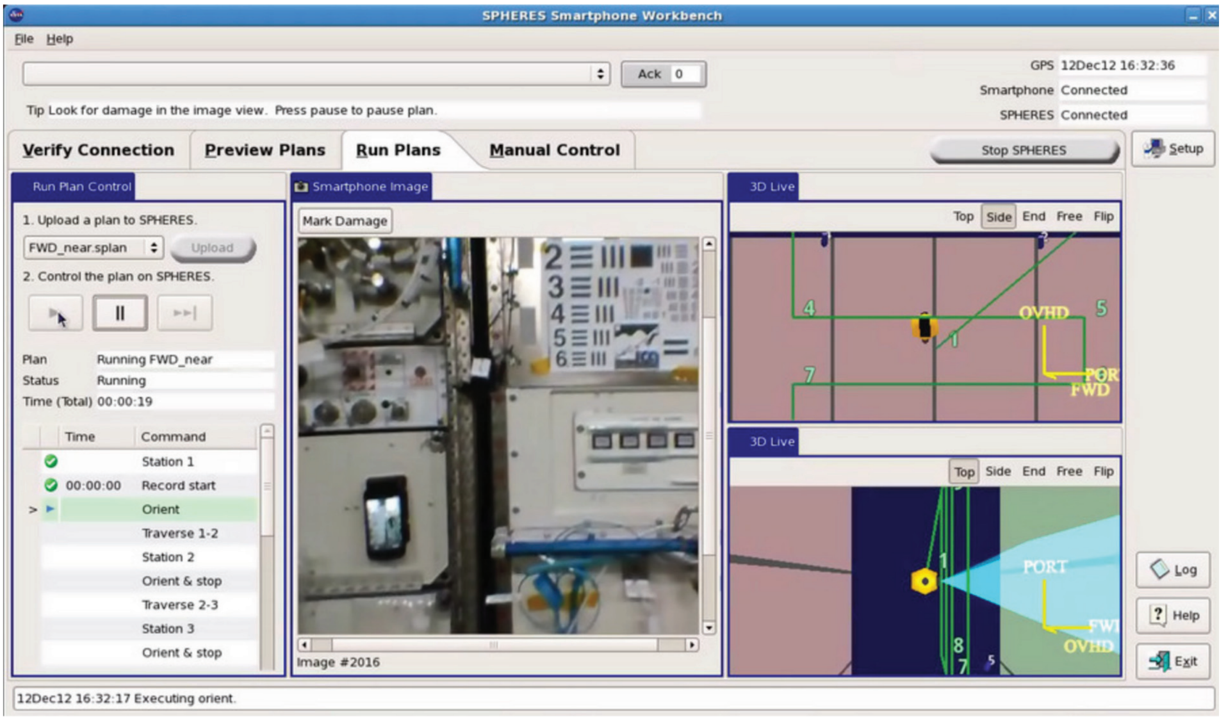

Fig. 2. The Smart-SPHERES Workbench enables ground controllers to operate free-flying robots on-board the International Space Station. However, this system does not support dynamic replanning and only supports a binary transition between waypoint navigation and manual control. Figure adopted from Fong et al. (2013).

as described by Christoffersen and Woods, 2002). Overall, the notion of a single dimension of autonomy with multiple levels may not provide sufficient guidance for design. Instead, Johnson (2014) proposes a more nuanced definition of autonomy as consisting of both *self-sufficiency* ('the degree to which an entity can take care of itself'), and *self-directedness* ('the degree of freedom from outside constraints'), while advancing a focus on teamwork and interdependence.

This coactive design approach combats the notion that robotic systems should replace humans to engage in dull or repetitive work or that increasing autonomy is always an appropriate goal. It instead recognizes that 'all useful robotic endeavors… are really human endeavors', highlighting human involvement in the design, planning, and use of autonomy, and argues that 'striving for full autonomy is ignoring the contextual understanding and creativity people bring to a problem' (Johnson, 2014).

Our work was directly informed by the principles of coactive design, particularly by its focus on teamwork and collaboration with users as an intrinsic part of successful task execution. However, we note that the current coactive design approach suffers from three key limitations, which we discuss below.

The first limitation is the reliance on an 'interdependence analysis' table (Johnson et al., 2014). The interdependence analysis table is the fundamental tool of the coactive design methodology, which a designer uses when performing task analysis to understand potential interdependence relationships and determine observability, predictability, and directability requirements. While such a tool, which draws on traditional task analysis techniques, may be useful for identifying viable team roles and interface requirements during the design process, it relies on subjective evaluations by a designer rather than using a more data-driven approach. For example, part of the interdependence analysis table involves assessment of the 'capacity to perform' and 'capacity to support', which relies solely on the designer's best judgment and may be difficult to accurately gauge when developing novel systems the capabilities of which are still evolving.

Second, while the coactive design methodology has led to very successful interfaces to support human–robot teamwork and collaboration, for example placing the Institute for Human & Machine Cognition team using a coactive design methodology first in the 2014 DARPA Virtual Robotics Challenge (Johnson et al., 2014) and second in the 2015 DARPA Robotics Challenge finals (Szondy, 2015), it has so far been limited to only designing interfaces for real-time task execution. Prior work has shown that human–robot teams often need support not only when coordinating actions in real time but also during planning phases that can occur at some point prior to execution or in certain phases during execution (Shah and Breazeal, 2010; Shah et al., 2011). Planning phases can be especially important to enhance team performance by helping to build shared mental models regarding the goals, expectations, abilities, and roles of each teammate while clarifying how communication will occur (Stout et al., 1999).

The third limitation involves the lack of appropriate representations of design elements dictated by the coactive design process for the end user. Communicating to users why a system is behaving in a certain manner based on the interdependence relationships predicted in the design process may further advance the coactive design goals of observability and predictability.

This study sought to address each of these limitations. We addressed the first limitation by collecting information in a formative study as part of a data-driven approach toward identifying the type of support a user might require (Section 4). The second limitation is explored in the design of interfaces that support planning, plan adjustments, and replanning before and during task execution (Section 5). The third limitation is addressed in the development of a visual approach to communicating interdependence relationships to the user via an interactive timeline, which we implement in one of our prototype systems (Section 5.3).

4. Formative study

To ground our work in a better understanding of remote collaboration and inform the design of control interfaces for free-flyers, we first conducted a formative study of human–human teams in the form of a design exploration with no a priori hypotheses. In the exploration, we gathered data on how users sought to accomplish a variety of tasks with a distal experimental confederate in a setup that mimics archetypical tasks in which free-flyers are envisioned to provide support.

4.1. Tasks and Setup

In the formative exploration, participants worked with a remote confederate (i.e., a trained experimenter) to accomplish three archetypical free-flyer tasks in an unfamiliar indoor environment. Our methodology bears similarities to prior work in the realm of mobile ground robotics that sought to identify collaborative user behaviors within the context of disaster search and rescue (Eberhard et al., 2010). The confederate carried a tablet running a video-chat application that allowed participants to view the remote environment and communicate with the confederate. The participant had to work with and instruct the confederate to accomplish three tasks: (1) *exploring* the unfamiliar environment in order to create a detailed written or drawn description of the remote location, (2) *inspecting* a number of boxes in the environment and recording the bar codes of boxes that appeared damaged, and (3) *monitoring* the confederate while recording any ‘mistakes’ made as the confederate assembled a puzzle. The confederate, imitating general free-flyer constraints, only followed very low-level instructions that a robot could feasibly execute (e.g., simple translation and rotation commands) and could not interact with the environment in any physical manner aside from moving the puzzle pieces during the monitoring task.

This formative exploration followed the model of an open-ended elicitation study, an increasingly popular method for HRI research (Cauchard et al., 2015). The study tasks were designed to raise questions related to interdependence, such as those noted by Johnson (2014): ‘What is the robot doing?’; ‘What is it going to do next?’; ‘How can I get it to do what I need?’ Such questions ‘highlight underlying issues of observability, predictability, and directability’ and elicit participant responses to obtain data regarding how users might intuitively attempt to answer such questions (Johnson, 2014). The first task was inspired by applications such as space telerobotics (Bualat et al., 2013) and wilderness search-and-rescue (Goodrich et al., 2008) where users direct robots to explore and collect data from unfamiliar or uncertain environments. The second task was designed as an analog to tasks involving robotic support for logistics management (Bualat et al., 2015; Fong et al., 2013) where robots may help locate tools or identify items or areas needing repair. Both the first and second tasks were designed to require a great deal of movement throughout the environment. For example, the boxes requiring inspection were located at a variety of position and heights, forcing the confederate to change ‘altitudes’. Additionally, both the first and second tasks were designed to be open-ended in an attempt to generate a large pool of participant behaviors; the first task allowed users to collect an arbitrary amount of data as part of their response detailing the environment, while in the second task, it was up to the participant to decide what ‘damaged’ meant. The third task attempted to simulate a monitoring task in which a free-flying robot might be utilized as a mobile camera, for example, recording critical operations or complicated assembly procedures, and required participant vigilance and situational awareness to watch for any ‘mistakes’ made as the confederate assembled two puzzles using tangram pieces. For each participant, the confederate constructed one puzzle ‘correctly’ and one puzzle ‘incorrectly’ based on the ‘solution’ provided to participants.

4.2. Participants

We recruited a total of nine participants (five male, four female) from the University of Wisconsin–Madison campus for this study, which was approved by the University of Wisconsin–Madison Education and Social/Behavioral Science Institutional Review Board (IRB). However, the digital video file for one of the male participants was corrupted and thus excluded from the analysis. These participants represent novice users, rather than experts such as pilots or flight controllers, as true experts in free-flyer operation are still quite rare due to the novelty of the technology. However, the actions of these novice users can still be extremely valuable in providing insights regarding how to

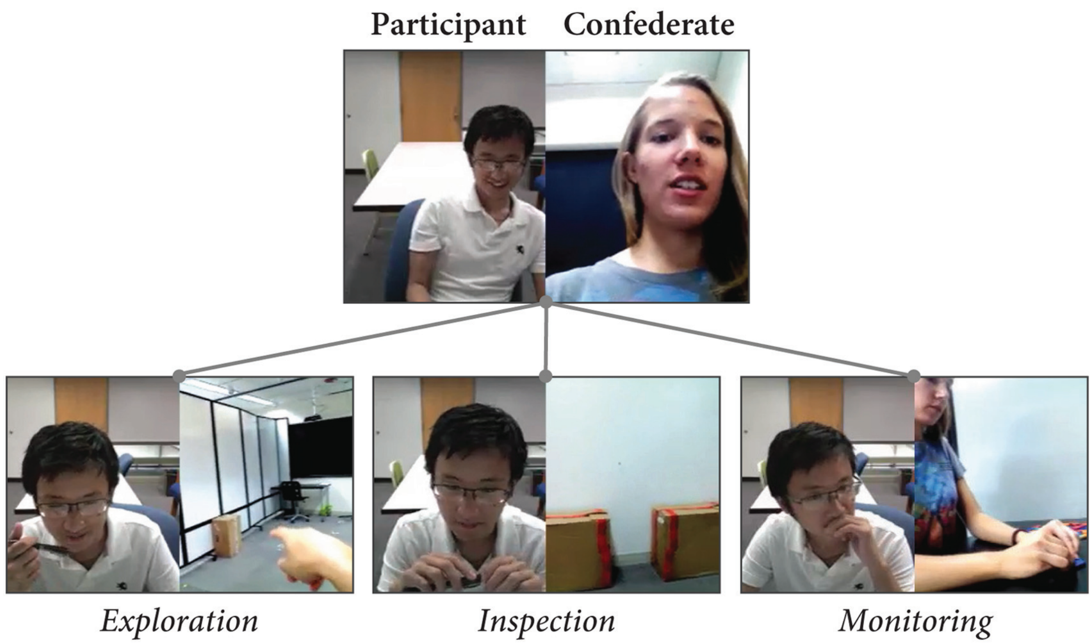

Fig. 3. In the formative exploration, participants used a video-chat system to direct a confederate, who modeled free-flyer behaviors, to undertake exploration, inspection, and monitoring tasks. Videos were then analyzed to determine high-level participant behaviors in working with the confederate and to generate guidelines for interface design.

design more intuitive systems that might decrease training times and better support overall interaction. The study lasted approximately 30 minutes, for which participants were compensated US$5.00.

4.3. *Analysis and findings*

Video recordings of each interaction were analyzed to identify high-level patterns of user behaviors (Figure 3). After a period of iteration examining and refining observed behaviors, eight dominant patterns emerged for classifying participant activities. Two coders annotated video data from each interaction with these eight classifications, which could consist of a single utterance (often seen in sequential commands) or a series of back-and-forth utterances between the participant and confederate (often seen in refinement behaviors). Data was divided evenly between the coders, with an overlap of 12.5% of the data coded by both. Inter-rater reliability analysis revealed substantial agreement between the raters (Cohen's $\kappa = .72$) (Landis and Koch, 1977).

The eight observed behaviors, which fall under three broad categories of 'directives', 'repair requests', and 'miscellaneous', are described in detail below. These behaviors bear similarities to language acts such as references and installments described in prior work in conversational grounding (Clark and Brennan, 1991) but focus on the domain of free-flyer interaction.

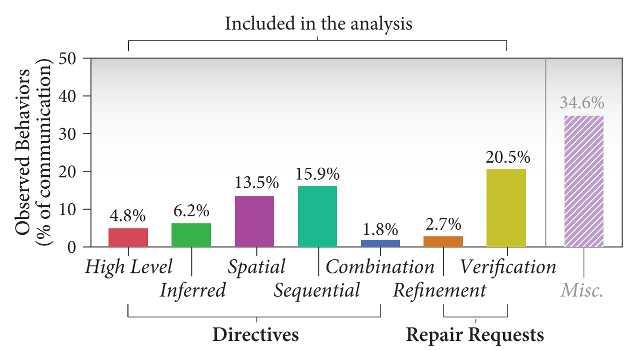

Fig. 4. The percentages of spatial, sequential, combination, inferred, and high-level commands, refinement and verification processes, and miscellaneous chatter observed in the videos out of total communication. Miscellaneous chatter was not included in our analysis.

4.3.1. Directives. Approximately 40% of all operator–performer interactions involved directives by the operator of different forms and levels of specificity to be followed by the performer. Specifically, we identified five different types of directives: (1) *high-level commands* represented requests that were too abstract for the confederate (or a free-flying robot) to interpret and carry out (e.g., 'I'd like you to explore the room'). (2) *Inferred commands* represented requests that the confederate interpreted based on potentially ambiguous but low-level input from the participant (e.g., participants asking the confederate to 'zoom in' on a detail, which was inferred to mean that participants wanted the confederate to move closer to the object, as there was no zoom feature on the camera). (3) *Spatial commands* represented the participant making a discrete request that referenced places and objects in the distal environment (e.g., 'Can you rotate the view so that I can see the top of that tall box on your left?'). (4) *Sequential commands* represented the participant making a plan of discrete actions with the confederate (e.g., 'I'd like you to walk to the center of the room, then raise the camera to eye level, and then slowly turn in place for a full 360-degree circle'). (5) *Combination commands* represented behaviors that appeared to aggregate both spatial and sequential cognitive processes (e.g., 'Let's move on… I want to look at the next box over in the right corner near the door. I'd like you to approach it and then lower the camera to the height of the box'.)

4.3.2. Repair requests. Many directives were followed by 'repair' actions that aimed to refine directives that may have been under-specified or confirm directives that may have been miscommunicated or misunderstood. More specifically, *refinement* represented the negotiation and arbitration process between participants giving a high-level command and eventually determining the correct spatial or sequential command(s) that the confederate could actually carry out to accomplish participant goals (e.g., in response to the previous high-level command, the confederate would respond 'OK, what specifically would you like me to do?'). *Verification* requests represented the times when the confederate and participant communicated to ensure that spatial commands or series of sequential commands were correctly achieving participant goals (e.g., participant: 'to the right a little bit more…'; confederate: 'like this?'; participant: 'OK, perfect'). Often, this process resulted in new commands if breakdowns occurred.

4.3.3. Miscellaneous. In addition to directives and repair requests, we observed several different forms of extraneous social chatter between participants and the confederate, consisting of interactions such as introductions and goodbyes at the beginning and end of each session or off-task conversation.

4.4. Design implications

Figure 4 shows the percentage of each type of action reported by the video coders out of total participant–confederate communication. While miscellaneous discussion was reported most frequently (average count: $M = 58.75$, $SD = 25.95$), it was largely excluded from the analysis. Because we are primarily concerned with developing general-purpose interfaces that can support users working with free-flying robots to accomplish a broad set of tasks, we analyzed participant behaviors across all three tasks (exploration, inspection, and monitoring), rather than identifying specific language discrepancies between each task. Two major design implications emerged from examining participant behaviors, as discussed below.

4.4.1. Need to communicate system capabilities. The 'gulf of execution' is a concept from the domain of cognitive engineering that describes gaps that exist between user goals and the affordances traditional software systems provide in achieving such goals (Norman and Draper, 1986). This 'gulf' is one part of the 'human action cycle', a model that has been successful in describing human interaction with computer systems, but has not been widely adapted when describing human interaction with robotic systems. Our formative study revealed a 'gulf of execution' where participants had difficulty communicating their intentions to a confederate operating under a set of constraints and capabilities similar to free-flying robots. High-level commands (average count: $M = 8.13$, $SD = 4.55$) and refinement actions (average count: $M = 4.50$, $SD = 2.88$) made up 11.40% of total on-task communication, which may indicate teamwork inefficiencies where time is wasted as participants struggled to generate actionable commands that a free-flyer would be able to feasibly understand. Overall, a large portion of such interaction occurred during an initial adjustment phase, with 81.5% of high-level commands occurring in the first half of participant interactions.

Our findings indicate that participants required a period of adaptation as they built shared understanding with the confederate. The large number of verification requests (average count: $M = 34.75$, $SD = 17.24$), representing 20.50% of total communication, reveal that this process was not fully successful as participants were never able to fully develop a mental model that enabled them to confidently reason about system capabilities and generate acceptable commands to accomplish their desired tasks. Overall, this data indicates a clear need for communicating robot capabilities, limitations, and actionable inputs from the beginning of the interaction in an unambiguous fashion. While existing systems may provide limited information regarding system capabilities (e.g., a manual may describe that moving a joystick will increase thrust), such information is largely tied to *maneuvering* the robot. Alternatively, our data revealed that such support is often needed in a more robust manner linked to robot *activities* (e.g., higher-level tasks such as taking pictures or collecting environmental data) to better help users map their intentions to robot actions. The human action cycle suggests that communicating this information might mitigate the issues observed in the formative study, decreasing training and adoption times while increasing user confidence when working with the system. Likewise, the principles of coactive design suggest that providing explicit support regarding possible robot actions might increase system directability.

4.4.2. Need to support spatial and temporal reasoning. Participants appeared to utilize both spatial and sequential commands frequently to accomplish the study tasks (spatial: $M = 22.88$, $SD = 10.47$, sequential: $M = 27.00$, $SD = 8.55$). Every single participant used both types of commands, occasionally even using combined commands ($M = 3.00$, $SD = 2.78$) with elements of both spatial and temporal cognitive processes, pointing to a need to support three-dimensional spatial reasoning about the work environment while concurrently supporting temporal planning and replanning. This design recommendation is also aligned with the coactive design principle that providing a three-dimensional view of the robot's environment may enhance system observability, while supporting sequential planning and replanning may aid in system predictability. While prior work, such as the interdependence analysis table used by Johnson et al. (2012), has highlighted the importance of providing a three-dimensional representation of the environment, the need to simultaneously support both spatial and temporal reasoning is a unique implication from our data-driven approach.

5. System and interface design

Our overall research aim is to develop a greater understanding of interface design requirements and enhance the support that general-purpose control, tasking, and collaboration systems provide for user interactions with free-flying robots. Design decisions often require trade-offs in terms of the interactive affordances the system provides, the amount and type of information the system presents, and the degree of coupling between the system and its user. To better understand such trade-offs, we undertook an iterative design process to develop three alternative free-flyer interfaces that implement different design elements. This process started by developing interfaces similar to existing systems (e.g., mimicking direct teleoperation systems and existing waypoint planners). These base systems were then extended in various ways to better align with the design implications uncovered in the formative study (e.g., changing a two-dimensional waypoint system into a three-dimensional map). The final versions of the three interfaces serve as reference designs that sample from the overall interface design space; each extends a prior free-flyer robot interface paradigm (teleoperation or waypoint planning) by providing different forms or levels of support regarding the ways in which it communicates system capabilities and supports (or lacks support for) spatial and temporal reasoning and planning (Table 2). Figure 5 shows screenshots from the interfaces.

The first system combines principles of teleoperation with approaches from shared control and user-directed guarded motion, resulting in a *safeguarded teleoperation* (Fong et al., 2001a) design that maximizes directability while protecting users who may lack an understanding of system capabilities, potentially at the cost of observability and predictability. The second interface extends traditional two-dimensional *waypoint delegation* interfaces by providing a three-dimensional spatial representation of the work environment. Finally, the third interface supports waypoint planning and replanning in a three-dimensional representation of the environment while also providing an editable, visual timeline representing different interdependencies between the user and the robot. This *collaborative interface* most fully supported the types of teamwork observed in the formative study and both design implications. By developing and testing three interfaces that provide reference points within the overall design space of free-flying-robot interfaces, we can gain a greater understanding of the relative importance of the design implications identified in the formative study and discover how various design elements contribute to the usability and effectiveness of general-purpose interfaces. We detail each of the three designs below.

5.1. *Safeguarded teleoperation interface*

The safeguarded teleoperation interface (Figure 5, top) provides a live video display from the robot and allows users to direct a free-flyer using a controller. This interface shares much in common with traditional free-flyer teleoperation interfaces but includes additional support for indoor operation. Traditional teleoperation interfaces provide joysticks that must be manually controlled to direct free-flyer thrust, pitch, roll, and yaw. Continuously performing manual control is extremely challenging; while experts with such controls can demonstrate complex aerobatic maneuvers (as seen in the rise of 'drone racing'; Murphy, 2016), it is often difficult for novice users to even maintain a stable hover. Newer free-flying platforms (e.g., the DJI Phantom 4[4]) are beginning to automate aspects of this control, enabling autonomous navigation to specified locations; however, they typically rely on GPS information that may be unavailable in indoor environments. In our experience, we have found that even with modern systems, users still have great difficulty navigating and avoiding collisions when operating within indoor environments. The safeguarded teleoperation interface we developed automates low-level aspects of controlling thrust, pitch, roll, and yaw using a proportional-integral-derivative (PID) controller (Ang et al., 2005). A PID controller takes the form

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau)\, d\tau + K_d \frac{d}{dt} e(t)$$

where K_p, K_i, and K_d represent constant parameters for tuning the proportional, integral, and derivative gains, e represents the error given an actual and desired value (in this case, the robot's position), and t is the time.

Using this controller, users can specify a desired location for the robot, causing the robot to fly to that location and hover until a new location is entered. This control scheme attempts to support the first design implication (mapping controller inputs to robot actions) while achieving a balance in managing the potential for the type of repair request behaviors observed in the formative study. It frees the user from having to constantly correct minute errors in thrust, pitch, roll, or yaw in a manual fashion, instead allowing users to control the desired altitude, location, and orientation of the robot with D-pads on an Xbox 360 Controller.[5] This input is fed to the PID controller on the back end,

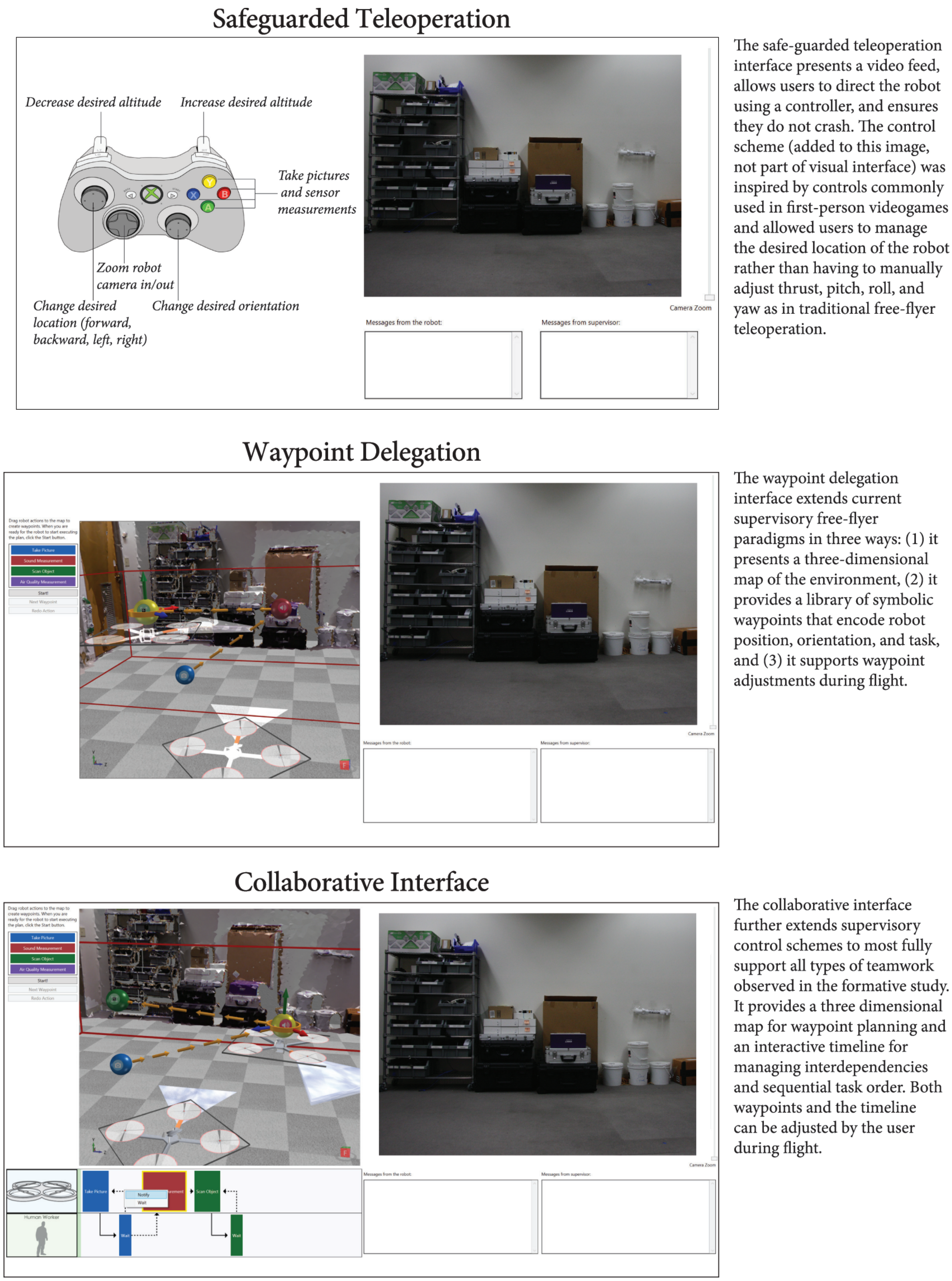

Fig. 5. We designed, implemented, and evaluated three interfaces to explore user teamwork with free-flyers: a safeguarded teleoperation interface (top), a waypoint delegation interface (center), and an interface designed to most fully support the collaborative behaviors from the formative study (bottom).

Table 2. Support provided by each interface prototype based on the design implications from the formative study.

Design implications	Interface design		
	Safeguarded teleoperation	*Waypoint delegation*	*Collaborative*
1. Communicate system capabilities	Robot abilities paired to teleoperation controls	Explicit waypoint-action command library	Explicit waypoint-action command library
2.a. Support spatial reasoning	—	Interactive three-dimensional map of robot environment	Interactive three-dimensional map of robot environment
2.b. Support temporal reasoning	—	—	On-the-fly editable timeline

which causes the robot to fly to the desired location. This scheme, which also provides users with a first-person video feed from the front of the robot, is very similar in feel to controlling first-person video games. Furthermore, it ensures safe handling of the robot for users who lack a prior understanding of robot capabilities and limitations, as built-in safeguards ensure that desired input locations do not place the robot too close to walls, ceilings, or objects in the environment. These safeguards take the form of simple thresholds; each dimension has maximum and minimum values that clamp user input if they try to alter the desired location beyond those bounds.

5.2. *Waypoint delegation interface*

The waypoint delegation interface (Figure 5, middle) was inspired by participants' use of spatial commands in the formative study as well as interfaces such as the Paparazzi Ground Control Interface and Mission Planner described in Section 2.2. Similar to these interfaces, it allows the user to plan a route for a free-flyer by placing waypoints prior to flight. However, this interface supports placing waypoints in a three-dimensional map of the work environment, rather than the two-dimensional overhead maps common to traditional free-flyer interfaces. This interface supports adding, modifying, and deleting waypoints not only during a planning phase prior to flight, but at any time during task execution. Additionally, waypoints are explicitly linked to robot actions, such as taking a photograph or sensor reading, rather than simply denoting locations for the robot to visit.

In this interface, users can view and navigate through a three-dimensional map of the work environment created by scanning a room using an Xbox Kinect[6] and assembling this data to form a point-cloud using the Point Cloud Library,[7] which includes open-source tools for two-dimensional/three-dimensional image and point-cloud processing. For the purposes of developing this interface prototype, the environment map was constructed a priori. However, in practice, free-flyers may be able to build environment maps on the fly using simultaneous localization and mapping techniques (Winkvist et al., 2013). Users can create waypoints by selecting a robot action, such as taking a picture or sensor measurement, from a list and dragging it into the environment in which they would like the robot to perform the task. Ray-casting is used to determine where

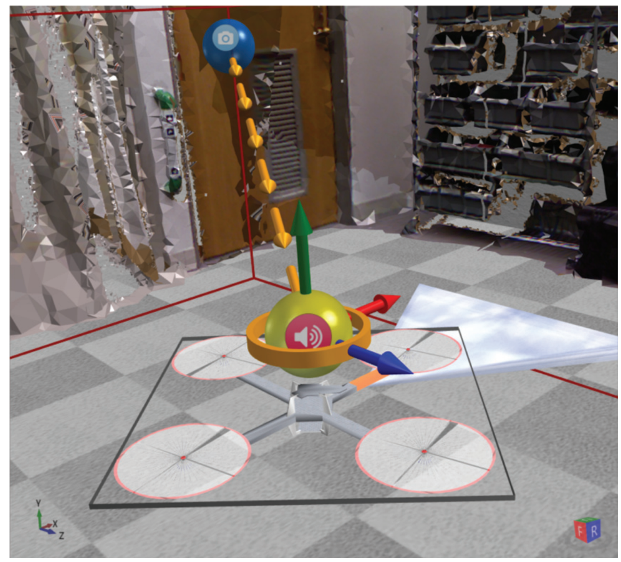

Fig. 6. Waypoints are represented by spheres that signify where a robot will perform an action. Above, a 'Sound Measurement' waypoint is selected, showing a visualization of how the robot will be positioned and oriented when taking the sound measurement, along with draggable handles to adjust desired robot position or orientation.

to place the initial waypoint in three-dimensional space from the two-dimensional cursor coordinates where the user drops the action.

Waypoints are represented by spheres that symbolize robot actions using color-coding and a visual icon that serves as an action glyph. Once initially placed, the user can modify the waypoint by dragging manipulators that move the waypoint in the x, y, or z direction. Additionally, the user can drag a circular manipulator to change the desired orientation of the robot at that waypoint. To help contextualize the robot's position and orientation while completing the action in the environment, a three-dimensional visualization of the robot, including a representation of the viewing area of the robot's camera, is presented whenever the user selects or manipulates a waypoint (Figure 6). A path of arrows links each waypoint to show the overall flight path the robot will follow.

Users may need to adjust the task plan or intervene to modify robot position and orientation at any time. To provide real-time control over robot behavior, waypoints can be added, modified, or deleted at any time prior to or during robot flight. For example, the user may plan a waypoint to take a picture of an object in the environment, but during execution, when the robot flies toward the waypoint, the details of the object in question may not be discernible from the video feed. In this situation, the user could manipulate the position or orientation of the waypoint that the robot was currently hovering at to cause real-time changes to the robot's position and orientation. Similarly, if during execution planned waypoints become extraneous, they can be deleted, while new waypoints can be added if task or environmental demands change. As in existing systems, new waypoints are always added at the end of the queue.

Once the user is ready to execute a flight path of planned waypoints, the robot will automatically fly to the first waypoint and begin the corresponding task (e.g., collecting data or taking a picture). When the task is completed, the robot will continue to hover at the waypoint and will present the collected data to the user via a text message in a notification dialog or, in the case of an image, via a pop-up window. The user must press a button to advance the robot to the next waypoint, representing an implicit constant dependency the robot has on the human to determine when the particular task at each waypoint has been accomplished to the user's satisfaction. The robot relies on a PID controller, similar to that described in the safeguarded teleoperation interface, to autonomously navigate from its current position to the desired position indicated by the current waypoint. This controller similarly ensures that the robot flies in a safe manner and does not come too close to obstacles, walls, or the ceiling, relying on the same thresholds as in the safeguarded teleoperation system. The three-dimensional representation of the environment visually demarcates these thresholds by outlining the volume where it is safe for the robot to fly with a wireframe overlaid on the three-dimensional environment map (see red wireframes in Figures 5 and 6). The system does not allow the user to position waypoints beyond the wireframe boundaries, instead clamping them to the maximum or minimum value.

Overall, this design aims to follow the design recommendations of communicating system capabilities (in a different fashion than in the safeguarded teleoperation interface) and supporting spatial reasoning. By providing users with a discrete list of actions that the robot can accomplish, we introduce a design constraint to help guide participant understanding of how they might work with the robot and what robot capabilities are available from the beginning of the interaction. By providing an interactive, three-dimensional representation of the environment, we better support spatial thinking regarding how to work with a teammate that can freely move in three-dimensional space.

5.3. Collaborative interface

The waypoint delegation interface only partially addresses the second design recommendation from the formative exploration; while it supports spatial reasoning, it was not designed to explicitly support the type of temporal cognitive processes observed in participant sequential commands. As a result, we designed a third interface that extends the waypoint delegation interface with an interactive timeline that supports dynamic task reordering and visually depicts interdependence relationships between the user and the robot.

This collaborative interface (Figure 5, bottom) provides a three-dimensional map of the environment in which users can place waypoints by dragging a selection from a list of robot actions, just as in the waypoint delegation interface. Similarly, users can add, modify, or delete waypoints (whose positions are clamped to avoid wall collisions as in the waypoint delegation and safeguarded teleoperation systems) during planning or while the robot is flying, with the robot autonomously navigating to each waypoint using a PID controller. However, unlike the waypoint-delegation interface, the collaborative interface includes an interactive task timeline at the bottom of the screen. This timeline explicitly shows robot actions and human interdependencies and supports dynamic re-sequencing of task actions. Every time the user adds a new waypoint, a corresponding action block is added to the timeline. The user can select and drag these blocks to re-order the timeline, changing the planned flight path in terms of waypoint orderings. These changes propagate to the arrows depicting the flight path in the three-dimensional virtual environment. Providing sequential task re-ordering may allow users to more dynamically respond to changes in task or environment; while the waypoint delegation interface always places new waypoints at the end of the queue, the collaborative interface allows users to rearrange the order of waypoints at any time.

In this design, the timeline explicitly encodes interdependence relationships, which can include waiting on the user to indicate that the robot should move to the next waypoint (the only dependency supported by the waypoint delegation interface), automatically advancing to the next waypoint when the robot has finished the planned task while notifying the user using visual and audible cues, or simply moving on to the next waypoint automatically. Certain robot actions are linked to inherent interdependencies by default. For example, given the difficulty for a robot to understand whether or not a picture captured what the user intended, the picture action automatically generates an interdependence relationship in which the robot waits on the user to approve the photograph before advancing to the next waypoint. Other actions, such as taking a sound measurement, contain no inherent interdependencies and would allow the robot to automatically advance by default. However, the user can change these default dependencies as desired by right-clicking on an action block in the timeline. This user input opens a context menu where the user can specify that, following this action, the robot should wait on user approval, notify the user while advancing, or simply move on to the next waypoint as soon as the robot is finished. In the timeline, actions (which for the robot correspond to waypoints) are visually depicted as blocks, with arrows showcasing interdependence relationships. For example, the timeline at the bottom of the collaborative interface in Figure 5 shows that the robot will first take a picture and then wait for approval from the user, with lines going from the ‘Take Picture’ block on the robot timeline to the ‘Wait’ block on the user timeline. The ‘Wait’ block has two arrows, one that leads back to the original ‘Take Picture’ block (in the event that the user did not approve the picture and requires a different one to be taken), and another one that advances to the ‘Sound Measurement’ block, which is the next robot action in this plan.

Overall, this collaborative interface aims to extend the benefits provided by the waypoint delegation paradigm by additionally supporting participant abilities to sequentially order tasks, fully operationalizing the design implications from the formative study. By supporting temporal planning in addition to spatial reasoning and communicating system capabilities, this design aims to enhance observability by explicitly encoding interdependence relationships in a way that can be customized by the user.

5.4. Implementation

During execution, each interface controlled the robot by sending desired three-dimensional coordinates via WiFi to a unified backend server. In the safeguarded teleoperation interface, the initial set-point (i.e., desired robot location) matched the starting location of the robot, and the user could alter the desired location using the controller as described previously. For the waypoint delegation and collaborative interfaces, the desired location was indicated by the current waypoint. The backend server was responsible for translating these coordinates into flight commands (e.g., thrust, roll, pitch, and yaw) that the robot could understand and sending these commands over radio via a serial connection with a USB adapter. To accomplish this goal, the server monitored the position of the robot using a single camera mounted in the ceiling of the environment. This camera was equipped with a wide-angle lens to be able to view the entirety of the experimental room and an infrared filter to isolate the infrared LEDs (light-emitting diodes) that were installed on the robot. Distortion from the wide-angle lens was corrected using a field-of-view fish-eye model (Devernay and Faugeras, 2001). The infrared LEDs were arranged on the robot in the form of an isosceles triangle such that the server could directly deduce the robot’s position and orientation while the area of the triangle corresponded to the robot’s altitude.

Four PID controllers independently managed robot thrust, roll, pitch, and yaw using the current robot position identified from the ceiling-mounted camera and the desired position indicated by the positions sent from the interface. This process enabled the robot to autonomously navigate to desired locations within the environment, while simultaneously protecting the robot by placing bounds on the maximum and minimum allowed goal locations. These bounds were selected to prevent the robot from flying too close to walls or the ceiling. In the waypoint delegation and collaborative interfaces, these bounds were visualized with a red wireframe in the three-dimensional environment map, as described above. The safeguarded teleoperation interface did not provide visual feedback regarding where these bounds were located, although users would receive instruction that such bounds were in place in the envisioned use of the interface.

The AForge library,[8] a C# framework for computer vision and image processing algorithms, was used in scanning the augmented reality tags for the inventory logistics task. The Helix Toolkit library,[9] which extended the three-dimensional functionality of the Windows Presentation Foundation, was used in presenting the three-dimensional maps of the environment in the waypoint delegation and collaborative interfaces. Finally, the EmguCV library,[10] a cross-platform .Net wrapper for the popular OpenCV image processing library, was used in filtering the video from the ceiling-mounted camera and identifying the location of the robot. All other code for extending these frameworks, developing the interfaces, managing the backend server, and directing network communication was custom-written in C#.

6. Evaluation

The formative exploration regarding the types of support that users might require when working with a distal assistant moving in three-dimensional space led to several design implications for free-flyer interfaces. We actualized these findings by implementing three complete free-flyer interfaces to gain further understanding of how differing levels of support and interdependence might help users better task, manage, and collaborate with free-flying robots. We then designed and conducted an in-person laboratory experiment in which participants utilized these interfaces to interact with a flying robot, which operated in a remote environment, to accomplish a series of realistic tasks. In this experiment, we examined how well the various interface design elements supported participant abilities to effectively work with the robot to accomplish their tasks as well as how the interfaces affected participant perception of the robot as a teammate.

6.1. Hypotheses

We developed four hypotheses regarding how specific interface design elements might shape operator work practices and experiences when interacting with a free-flying robot, based on prior research findings and the results of our formative study.

6.1.1. Hypothesis 1. Support for spatial reasoning and planning leads to more efficient work. Prior work has emphasized the importance of spatial reasoning in forming mental models (Tversky, 1991). Similarly, our formative study found that users often directed the confederate using spatial commands. As a result, the spatial reasoning support provided by an interface with robot capabilities must be aligned in all stages of interaction, not just the execution stage. We predicted that providing three-dimensional support for spatial reasoning when forming plans for aerial robots would lead to improved plans that require fewer adjustments during execution, thus leading to more efficient work overall.

6.1.2. Hypothesis 2. Support for temporal reasoning and planning improves responses to unplanned events. Our formative study showed that participants issued temporal commands at a similar rate as spatial commands, showing that both types of reasoning were necessary for completing their tasks. In real-world operations, it is extremely unlikely that users will be able to reliably form perfect plans that account for all possible contingencies during operations, especially in complex domains such as space exploration. We predicted that introducing interactive design elements that support temporal reasoning processes by enabling dynamic task reordering during planning and execution stages would allow users to respond more quickly to unplanned changes in task or environmental conditions.

6.1.3. Hypothesis 3. Support for pre-flight planning and flight automation reduces user cognitive load. Prior work has shown that increased automation levels (i.e., interfaces that support a management or shared-control approach) can reduce user mental workload over manual teleoperation paradigms (Crandall et al., 2002; Ruff et al., 2002). Planning prior to robot operation can enable a greater degree of automation; instead of the user manually directing the robot, the robot can execute the user's plan under supervision. Thus, we predicted that providing support for pre-flight planning, which enables a greater degree of automation during execution, would decrease user cognitive load by requiring less constant user intervention and attention.

6.1.4. Hypothesis 4. Support for greater autonomy increases perception of the robot as a team member. Prior work has shown that users will attribute social characteristics to robots and that such perceptions of agency can affect perceived robot usability and safety (Fong et al., 2003; Forlizzi, 2007; Sauppé and Mutlu, 2015). We predicted that design elements that support greater levels of robot autonomy and increase user–robot interdependence relationships would lead to perception of the robot as more of a teammate and less of a tool.

6.2. Experimental design

We conducted a 3 $\times$ 1 between-participants laboratory experiment to evaluate how well each of the three interface designs supported users in working with a free-flying robot. The independent variable represented what type of interface participants used: (1) *safeguarded teleoperation*, (2) *waypoint delegation*, and (3) *collaborative interface*. Dependent variables included objective measures of task completion efficiency and responses to unplanned events as well as subjective scores regarding participant cognitive load and ratings of the robot's perceived collaborative role.

6.3. Experimental tasks

Six experimental tasks were developed for the user to carry out, representing a variety of realistic activities that place different types of interdependency requirements on the human–robot team. These six tasks fall into three categories: three *planned tasks*, two *unplanned tasks*, and one *secondary task*.

6.3.1. Planned tasks. Three tasks were fully described to participants in the experimental instructions. These tasks required visiting locations throughout the robot's environment at a variety of heights to take measurements or make observations. As participants completed these tasks, they were instructed to fill in responses on a digital worksheet that provided an abstract, two-dimensional overhead view of where the tasks needed to be performed (Figure 8). Participants were able to plan how they would like to accomplish these tasks, which could be carried out in any order the participants chose, and were instructed that completing these tasks (and the unplanned tasks described below) were their main priority. These tasks included the following.

- *Sound Survey*: This task required participants to measure sound levels at four vents located in the remote environment. Participants were instructed to locate the vents and take a sound measurement as close to the vents as possible to get an accurate reading. They then needed to compare the measured values against a provided baseline measurement to determine whether or not the vent required repair. This step forced a dependency such that the user had to wait on the robot for the sound measurements, while the robot would be free to continue new tasks after taking the measurement.
- *Pipe Inspections*: This task required participants to work with the robot to visually inspect four pipes located in the robot's environment. Participants were instructed to examine a serial number on each pipe, use the number they observed to look up the pipe's installation date in a provided list, and then save a picture of any pipe that had been installed prior to a certain date. This task introduced a dependency in which the robot had to wait on the user before advancing, as whether or

not a picture of the pipe would be required would not be known in advance.

- *Inventory Logistics*: This task required participants to locate a ‘misplaced’ box in the warehouse environment. Participants needed to work with the robot to scan a number of boxes that were marked with augmented reality (AR) tags indicating the contents of the box and a box identification number. Once the missing box was found, participants were instructed to record the box identification number. This task was intended to place a series of interdependency relationships on the human–robot team and require extensive collaboration in order to navigate among the boxes, bring the robot to a point at which it could scan the AR tags, and make decisions regarding whether or not the box had been located.

Fig. 7. The experimental environment was set up to mimic that of a warehouse in which a free-flying robot might soon be deployed. A modified AscTec Hummingbird (highlighted) was used as the experimental robot platform.

6.3.2. Unplanned tasks. Participants were also instructed that either the robot or a ‘warehouse supervisor’ might ask them to accomplish additional tasks during the experiment. They were instructed to give any such additional tasks equal priority to the three planned tasks described above. However, they were not instructed regarding what such tasks might entail, only that details would be provided if the robot or supervisor requested their assistance. These requests came in the form of text messages. Prior to the experiment, participants were shown two dialogs on their screen: one over which the robot would communicate any data measurements or requests for aid and another one over which the supervisor would communicate. These dialogs were read-only, and the participant could not enter text to send messages to the robot or the supervisor. An audible alert sounded whenever the participant received a message from the supervisor or a help request from the robot. All participants received one request from the supervisor and one request from the robot, which were triggered at set times while the robot was in flight. These requests were to perform the following tasks.

- *Air-Quality Measurement*: This task represented a request by the supervisor to check the air quality in the remote environment. Participants were sent a text message indicating that sensors in the room had detected a potential ammonia leak, that the supervisor would like confirmation of the leak from the robot, and that the data from the next air-quality check would automatically be sent to the supervisor. This task was inspired by a recent incident on the ISS in which a false alarm indicated a potential ammonia leak, forcing the American modules to be evacuated (Dunn, 2014). In such events, whether on the ISS or in a terrestrial industrial environment in which toxic chemicals may be present, free-flyers may assist by assessing the environment while maintaining human safety. Users could not plan for this task, which requested that they spontaneously take an air-quality measurement anywhere in the experimental room.
- *Robot Re-alignment*: This task represented instances in which robots may recognize that internal errors or faults have occurred and require assistance from users. In this task, the robot sent a text message to the user requesting help in re-aligning its internal map of the environment. Users were sent a picture of the current video feed from the robot and had to click on five points, similar to tasks in which users must click on corners to align three-dimensional scans or help calibrate computer vision systems. While both this task and the air-quality measurement task interrupted users, this task did not require the user to re-direct the robot in any manner.

6.3.3. Secondary task. In addition to the main planned and unplanned experimental tasks described above, users were given a secondary task with instructions to complete it whenever they felt they had spare time. This secondary task was designed to measure user free time and thus indirectly gauge the attentional demands and cognitive load the work placed on the user (Brunken et al., 2003; van Gerven et al., 2006). Several tasks, such as performing mental arithmetic or reading email, have been suggested in prior literature in HRI and human factors (Crandall et al., 2002; Lee et al., 2001; Olsen and Goodrich, 2003). We adapted such tasks to the context of operating a free-flyer in a warehouse environment in an attempt to enhance the realism of the experiment.

The secondary task in this study presented users with a pair of line graphs that ostensibly showed internal robot temperatures across prior robot operations. Users were instructed that as battery temperatures increase, robot flight performance decreases. Therefore, users were asked to act as quality control and click on the graph portraying the higher *average* battery temperature. Once a graph was selected, a new pair would appear. Users were able to complete as many of these secondary tasks as they chose, but their instructions emphasized that they should only work on such tasks when otherwise unoccupied. As users had

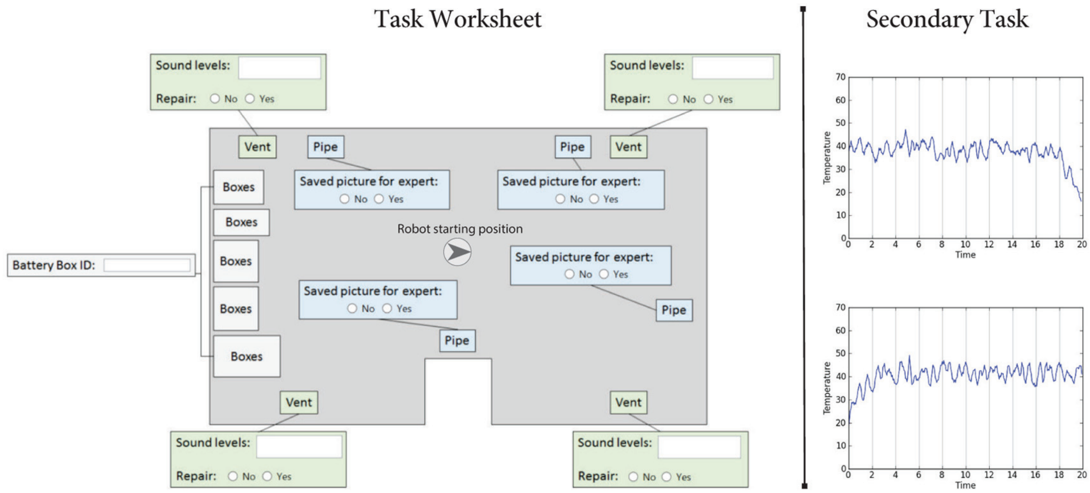

Fig. 8. A digital worksheet, representing a two-dimensional floor plan, guided participants through the planned tasks and served to capture their answers. These tasks were distributed throughout the robot's environment. A secondary task was designed to gauge how much concurrent work each interface design afforded. Participants were instructed to select the graph showing the higher average battery temperature (the bottom graph in this stimuli). Once selected, new stimuli appeared.

complete freedom in choosing which tasks to work on, this secondary task was given a lower priority than the main experimental tasks to ensure that participants did not simply use the entire experimental time completing secondary tasks. Each comparison task was designed to take users a maximum of 20 s, which enabled them to quickly perform the task during any available free time and smoothly switch back to the main tasks as necessary. The stimuli were generated by dividing a continuous Perlin noise signal (Perlin, 1985) into two equal segments. The values within each segment were then adjusted using the process described by Albers et al. (2014) such that the average Y-value of the segments differed by five. The location of the 'correct' graph was randomized for each stimulus. An example stimulus is shown in Figure 8.

6.4. Experimental environment, platform, and apparatus

This experiment sought to replicate realistic demands that will be placed on human–robot teams as free-flyers integrate into environments such as warehouses or the ISS. The environment in which the robot operated was designed to mimic that of a small warehouse. The environment measured 300 square feet and contained various clutter including pipes, boxes, and shelves along all outer walls (Figure 7). The interior of the room was open for the robot to fly, and the robot lacked abilities to detect and avoid obstacles (other than the safe-guarding algorithms described above that prevented wall and ceiling collisions).

We used a modified Ascending Technologies (AscTec) Hummingbird quadrotor UAV[11] as the experimental platform (highlighted in Figure 7). As the Hummingbird does not have a native video camera, an unlocked Samsung Galaxy S4 Mini[12] commercial cellphone was attached to the robot, with a periscope lens to enable forward-facing vision. The Android IP Webcam application[13] was utilized to stream video from the phone to the user interfaces. The Hummingbird was also modified to include three infrared LEDs for tracking, as described in Section 5.4.

Several variables in interface presentation were controlled across study conditions. The size of the presented video feed was constant across each interface, however, the location of this feed on the user's screen changed due to the increased demands on screen space in the waypoint delegation and collaborative interfaces. Similarly, the size of the three-dimensional environment map was constant across the waypoint delegation and collaborative interfaces, although its location varied between the two interfaces, as the collaborative interface required space for the interactive timeline. The sizes of the robot and supervisor dialog were consistent across all three interfaces. All interfaces supported zooming-in the robot's camera using a slider next to the video feed; users of the teleoperation interface could additionally use the controller's D-pad for this purpose (see Figure 5, top).

The main robot interface was presented on a monitor directly in front of participants. A secondary monitor to their right contained the digital worksheet for participants to record their progress on the planned tasks (e.g., marking down the sound measurements, identifying which pipes needed to be replaced, and recording the identification number of the missing box) as well as complete any secondary tasks when they were otherwise unoccupied (Figure 9). Participants using the waypoint delegation and collaborative interfaces utilized a mouse and keyboard for all interactions with the main interface, to fill in the worksheet, and to complete secondary tasks. Participants using the safe-guarded teleoperation interface utilized an Xbox 360 Controller for the main interface while needing to switch to a

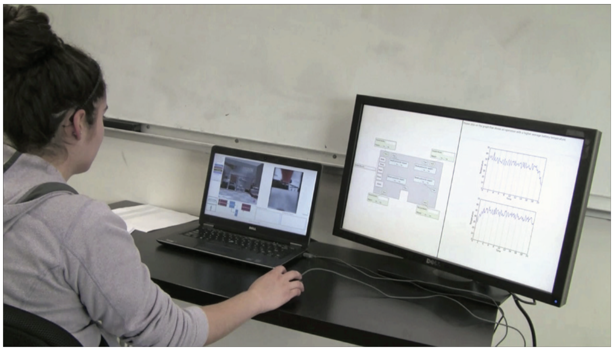

Fig. 9. In the experiment, participants used an interface on a main monitor in front of them to work with the robot. A secondary monitor to their right presented a task worksheet and secondary tasks.

mouse and keyboard to fill in the worksheet and complete any secondary tasks.

The cellphone attached to the robot enabled the user to view a live video feed that supported zooming and allowed the user to save pictures. All other sensor readings were simulated based on robot position. For example, when participants requested a sound measurement, the server determined which vent was closest to the robot and reported data accordingly such that two vents would always report sound levels indicating a repair was needed while the other two would always report normal readings. The same two vents required repair across all participants. Similarly, the air-quality measurement would always report data indicating that the ammonia leak was a false alarm. Each sound measurement took 10 s to complete, while air-quality measurements took 20 s. When taking measurements, the interfaces presented the user with a progress bar that indicated remaining measurement time and allowed the user to cancel the measurement at any time prior to completion. Participants were prevented from taking multiple measurements simultaneously. The sensor measurement delay was instituted to increase the realism of the task and present potential opportunities for free time during which participants might complete secondary tasks or make plan adjustments during execution.

6.5. Experimental procedure

The study took 30 to 45 minutes and consisted of six main phases: (1) introduction, (2) instruction, (3) planning, (4) execution, (5) evaluation, and (6) conclusion.

At the beginning of the procedure, the experimenter obtained informed consent and brought the participant into a controlled room separated from the robot environment. Participants were seated at a table, informed that they would be working with a robot in a distant warehouse, and given an instruction packet detailing the experimental tasks. While the participant read through the instructions, the experimenter left the room and set up the experiment, ensuring that the robot had a fresh battery, was positioned in an identical starting location for each participant, and that all network connections had been successfully initialized.

In phase 2, the experimenter re-entered the participant room and verified that the participant understood the experimental tasks. The experimenter then administered a tutorial with a dummy version of the interface that participants would use. This interface was not connected to the robot, but enabled users to become familiar with interface functionality. Participants were required to demonstrate knowledge regarding how to navigate the robot and how to instruct the robot to perform each possible measurement. Once the experimenter had verified that the participant understood the interface and answered any final questions, they started up the real interface that was connected to the robot and left the room so that the participant could begin the planning phase.

In phase 3, participants were given as much time as they desired to plan how they would accomplish the tasks. For both the waypoint delegation and the collaborative interfaces, users were able to set waypoints for the robot. The collaborative interface additionally supported the user in re-arranging the order of robot actions on a timeline and defining interdependence relationships, as described above. The safeguarded teleoperation interface did not support either spatial or temporal planning, however, users were still given time, if they desired, to think about how they would accomplish their tasks. Whenever the user felt ready, they could press on a button to begin the execution phase of the experiment.

In phase 4, the robot took off from a set location (marked with blue tape in Figure 7) and could be guided by the user to navigate around the environment and perform a variety of data collection and sensing tasks. In the safeguarded teleoperation condition, users directed the robot using the game controller, while in the waypoint delegation and collaborative interfaces, the robot flew to user-designated waypoints, as described in Section 5. Participants had 10 minutes of flight time to accomplish as many of the experimental tasks as possible in whatever order they chose while filling in the task worksheet and completing any secondary tasks. The supervisor's air-quality request was triggered three minutes into the execution phase, while the robot re-alignment request occurred after six minutes. A timer was always visible to indicate to participants how much flight time they had remaining. Upon reaching the 10 minute limit, the robot automatically landed, the interface closed, and the experimenter re-entered the participant's room. The 10-minute flight time limit was instituted to place demands on participant resources in completing the tasks and to align with platform battery limitations.

In phase 5, the experimenter administered assessments to gauge perceived workload and asked participants evaluate their experience and perception of the robot. In phase 6, the experimenter collected demographic information, debriefed the participant, and paid them US$5.00 for their time.

6.6. Participants

A total of 38 participants from the University of Wisconsin–Madison campus were recruited for this study, which was approved by the University of Wisconsin–Madison Education and Social/Behavioral Science IRB. Two participants were excluded from analysis due to not following the instructions of the experiment, resulting in a total of 36 participants (12 participants, six females and six males, per condition). Participant ages ranged from 19 to 50 with an average age of 24.42 ($SD = 6.17$). On a seven-point scale, participants reported a moderate prior familiarity with robots overall ($M = 3.53$, $SD = 1.83$) but a low familiarity with small flying robots ($M = 2.64$, $SD = 1.64$). As a result, these participants mainly represent novice users; however, one participant noted that they had a great deal of prior experience operating model airplanes and had obtained a private pilot's license. Another participant noted that they had prior experience working with machine automation. Furthermore, participants overall reported a moderate amount of prior familiarity with video games ($M = 4.22$, $SD = 1.87$), which may be relevant to the operation of the safeguarded teleoperation interface.

6.7. Measures and analysis

Objective (performance/task) and subjective (user-perception) measurements captured the outcomes of the user interaction with the three interfaces. Several measures were derived from prior work investigating metrics for effective HRI (Crandall et al., 2002; Olsen and Goodrich, 2003; Steinfeld et al., 2006).

Six main objective measures captured aspects of the effects of interface design elements on the human–robot team. As users in any condition could choose not to create a plan and simply start flying the robot, *planning time* was used to determine whether participants actually utilized the additional planning support provided by the waypoint delegation and collaborative interfaces. Communication *bandwidth* was measured in terms of how much information, such as commands to move the robot, save a picture, or take a sensor reading, was sent by the user to the robot in order to accomplish task goals. In all conditions, the maximum rate of communication was 20 Hz; however, information was only communicated to the robot as inputs changed (e.g., holding forward in the safeguarded teleoperation interface would only count as one command, not a stream of several commands). The safeguarded teleoperation interface applied default dead zones from the SlimDX framework[14] to ensure that small changes in joystick movements did not result in additional commands. *Efficiency* was calculated as overall task completion divided by bandwidth, computed by taking a weighted average of all tasks completed successfully, adding correctly completed secondary tasks, and dividing by the number of commands issued by the user. *Intervention response effectiveness* and *intervention response time* addressed how well users were able to react to unplanned tasks. *Secondary task* completion rates were examined to assess cognitive workload.

Subjective measurements included the NASA task load index (NASA-TLX) (Hart and Staveland, 1988) to measure perceived cognitive workload and responses to questionnaire items that were grouped to construct a scale for measuring participant perception of the robot's role (teammate vs tool). This scale consisted of three Likert-type items that asked participants to rate their agreement (from 1 for 'strongly disagree' to 7 for 'strongly agree') regarding how much they felt that they assisted the robot, supervised the robot, and managed the robot (Cronbach's $\alpha = .748$). Higher scores on this scale were interpreted as the participant viewing the robot more as a teammate or assistant (i.e., a semi-autonomous agent requiring some degree of user oversight), while lower scores indicated that the participant viewed the robot more as a tool (i.e., a device to be directly operated, rather than supervised, to accomplish tasks). Participants also provided open-ended responses regarding their impressions of the robot, the interface they used, and their overall experience.

We analyzed the collected data captured as continuous variables using a one-way independent measures analysis of variance with the main experimental manipulation as a fixed effect. To test our hypotheses, we conducted planned, pairwise comparisons across experimental conditions using Tukey's honest significant difference (HSD) test to control for Type I errors. Data from count measures were analyzed using Fisher's exact test.

6.8. Results

Figure 10 provides a visual summary of the main results of this experiment, which are described below in relation to the respective hypothesis. Performance measures across experimental tasks are reported in Table 3.

6.8.1. Hypothesis 1. The first hypothesis predicted that participants would be able to accomplish work more efficiently (i.e., successfully completing more tasks with fewer commands over the same time period) using interfaces that supported spatial reasoning and planning prior to execution. While the waypoint delegation and collaborative interface offer a much greater degree of support for planning, participants were free to start the execution phase of the experiment whenever they desired in all conditions. As one of the major design implications from the formative study was for interfaces to support both planning and execution, we first verified that participants understood that such support was offered and checked whether they made use of it by analyzing the time spent planning in each condition. We found a significant effect of interface design on planning time, $F(2,33) = 24.28$, $p < .001$, with a pairwise

Table 3. Completion rates, success/failure, and completion times for each experimental task, reported as *M* (*SD*) or number completed/total.

Experimental task	Performance measure	Interface design: *Safeguarded teleoperation*	*Waypoint delegation*	*Collaborative*
Sound Surveys	Completed (*of 4*)	2.92 (1.51)	3.00 (1.13)	3.25 (0.87)
	Correct (*of 4*)	2.67 (1.56)	2.75 (1.42)	3.00 (0.85)
	Time on task (s)	135.99 (94.09)	159.28 (101.26)	177.96 (91.24)
Pipe Inspections	Completed (*of 4*)	2.25 (1.60)	1.83 (1.47)	2.92 (1.16)
	Correct (*of 4*)	1.92 (1.56)	1.50 (1.57)	2.58 (1.08)
	Time on task (s)	139.22 (125.48)	135.41 (115.05)	193.55 (118.67)
Inventory Logistics	Number of Participants Completed	8/12	9/12	4/12
	Number of Participants Correct	7/12	8/12	4/12
	Time on task (s)	65.84 (59.99)	74.34 (79.62)	18.38 (30.34)
Air-Quality	Number of Participants Completed	12/12	6/12	11/12
	Intervention Time (s)	78.68 (86.50)	256.96 (123.46)	136.63 (100.86)
Robot Realign	Number of Participants Completed	12/12	12/12	12/12
	Intervention Time (s)	17.63 (13.03)	14.44 (10.92)	10.16 (30.34)
Secondary Task	Completed	3.92 (7.45)	5.25 (6.84)	6.50 (7.37)
	Correct	3.75 (7.14)	4.25 (6.31)	5.33 (6.15)
Overall Performance	Task Completion Score	74.72 (32.76)	54.83 (42.09)	79.81 (20.87)
	Efficiency	0.03 (0.02)	0.15 (0.14)	0.14 (0.09)

comparison revealing that users of the waypoint delegation and collaborative interfaces spent significantly more time planning than those using the safeguarded teleoperation interface at the $p < .05$ level. Relative to participants using the safeguarded teleoperation interface, participants spent on average an additional 11.98 minutes ($SD = 7.13$ minutes) planning with the waypoint delegation interface and 12.97 minutes ($SD = 5.14$ minutes) planning with the collaborative interface.

To test the hypothesis, we first analyzed *bandwidth*: how many commands each participant sent to the robot. We found a significant effect of interface design on bandwidth, $F(2, 33) = 34.00, p < .001$. As an initial test of our hypothesis, a pairwise comparison across conditions found that participants in the waypoint delegation ($M = 644.83$, $SD = 613.35$) and collaborative ($M = 749.08$, $SD = 313.00$) interface conditions sent significantly fewer commands to the robot than those in the safeguarded teleoperation condition ($M = 2804.50$, $SD = 1047.20$) at the $p < .05$ level. Cohen's effect size value suggests a high practical significance for both the waypoint ($d = 2.52$) and collaborative ($d = 2.66$) interfaces in reducing bandwidth.

We next examined task completion, success/failure rates, and completion times to understand task performance. These results are reported in Table 3. Time on task for tasks other than the interventions was particularly difficult to measure because participants were free to work on tasks in whatever order they chose. As a result, reported time values for sound surveys, pipe inspections, and inventory logistics are estimates calculated based on the time between participant answers in the task worksheet. For instance, if a participant filled in a sound survey answer followed by a pipe inspection answer, we assume that, after completing the sound survey, the participant began working on the pipe inspection, estimating the pipe inspection time as the time between reporting the two answers in the worksheet. As these assumptions may be incorrect (e.g., if a participant switched to a different task halfway through another task), we report the task time estimates but do not include them as a means of testing our hypotheses.

To account for the variance caused by participant choice in which tasks to focus on and by the varying number of subtasks within each task, as well as to better understand the ability of each interface to support general-purpose use across all tasks, we created a composite performance metric of *efficiency*. To calculate efficiency, we first created an overall task completion score (ranging from 0 to 100) by computing a weighted average of the percentage of main tasks completed correctly (weights assigned based on the number of subtasks within each task) and adding correctly completed secondary tasks. For example, a participant who correctly completed 75% of the planned and unplanned tasks and correctly completed three secondary tasks would receive an overall task completion score of 78. This scoring was inspired by educational assessment methods, where a student's cumulative grade point average might weigh grades from several courses, whose final grade calculations themselves may average several quizzes, tests, and assignments along with possible extra-credit points (see Sadler and Tai, 2007, and Stricker et al., 1994, for discussions and examples of such weighting systems in education). On average, we found that participants using the collaborative interface ($M = 79.81$, $SD = 20.87$) scored highest in our overall task performance score, followed closely by users of the teleoperation interface ($M = 74.72$, $SD = 32.76$), with users of the waypoint delegation interface ($M = 54.83$, $SD = 42.09$) scoring relatively lower.

Hypothesis 1 is primarily concerned with efficiency, which we calculated by taking this overall performance score and dividing it by bandwidth, creating a normalized ratio of overall task completion to commands sent. We found a significant effect of interface on efficiency, $F(2, 33) = 5.11$, $p = .012$. Pairwise comparisons between teleoperation and the waypoint delegation interface as well as between teleoperation and the collaborative interface provided support for Hypothesis 1, showing that spatial reasoning and planning significantly improved efficiency, the ratio of tasks completed successfully to commands sent, at the $p < .05$ level. Specifically, we found large effect sizes when comparing efficiencies for both the collaborative interface compared to the teleoperation interface ($d = 1.69$) and the waypoint delegation interface compared to the teleoperation system ($d = 1.2$). These results show that the teleoperation interface had a higher average task completion score than the waypoint interface, as shown in Table 3, but at the expense of sending many more commands. On the other hand, the collaborative interface scored highly both in the task completion score and in the efficiency metric.

6.8.2. Hypothesis 2. The second hypothesis predicted that design elements that supported temporal reasoning and planning would enable participants to better respond to spontaneous, unplanned events that require new teamwork. To test Hypothesis 2, we primarily analyzed data between the collaborative interface (which supported temporal reasoning via an interactive timeline) and the waypoint interface (which lacked this design element), although data for the teleoperation interface is included for completeness. Prior to hypothesis testing, we conducted a manipulation check to determine that participants were indeed responding to spontaneous requests. We found that all participants responded to the robot's request for re-alignment (which did not require any teamwork, interaction with the robot, or replanning) and did so relatively quickly (no significant differences were found across conditions at the $p < .05$ level; see Table 3), indicating that participants understood when interruptions occurred asking them to perform new tasks.

To determine whether interface design elements affected participant ability to respond to unplanned events that required actual intervention on their part, we analyzed participants' *intervention response effectiveness* and *intervention response time* regarding participant abilities and time taken to successfully respond to the supervisor's spontaneous request. For intervention response effectiveness, we used Fisher's exact test to determine whether the percentage of participants that successfully intervened and worked with the robot to perform an air-quality check in response to the supervisor's request differed significantly across interface conditions and found a significant main effect, $p = .008$. Fisher's exact test was used in place of a Pearson chi-square as the data was very unequally distributed across conditions (all teleoperation participants completed the air-quality check; all but one participant succeeded using the collaborative interface; only half of the participants succeeded in the waypoint condition).

We also found a significant effect of interface design on intervention response time, $F(2, 33) = 16.90$, $p < .001$. A pairwise comparison showed that participants took significantly longer to respond in the waypoint condition than in the collaborative interface (Cohen's $d = 1.08$). These results provide support for Hypothesis 2; design elements for temporal reasoning and planning improved intervention effectiveness and response time to spontaneous events. Interestingly, the results also showed that the teleoperation interface led to faster intervention response times than the waypoint interface (Cohen's $d = 1.67$), although no difference was found between the collaborative and teleoperation interfaces at the $p < .05$ level.

6.8.3. Hypothesis 3. The third hypothesis predicted that pre-flight planning and greater levels of flight automation would reduce participant cognitive load, thus we expected to see lower levels of cognitive load for the waypoint and collaborative interfaces compared with teleoperation. We assessed perceived cognitive load using the NASA-TLX. However, the analysis found only a marginal effect of interface used on perceived workload, $F(2, 33) = 2.72$, $p = .081$, and did not find any significant differences in pairwise comparisons at the $p < .05$ level.

Performance on the secondary task can act as an additional, indirect measure of workload. Secondary task completion is reported in Table 3. While participants often completed secondary tasks (on average, each participant completed 4.44 secondary tasks correctly), our analysis did not show a significant effect of the interface used on the number of secondary tasks completed, $F(2, 33) = 0.38$, $p = .684$, or the number of secondary tasks *correctly* completed, $F(2, 33) = 0.18$, $p = .833$. Thus, we did not find evidence to support Hypothesis 3.

6.8.4. Hypothesis 4. The fourth hypothesis predicted that interfaces that supported a greater level of automation would increase user perception of robot agency, thus we

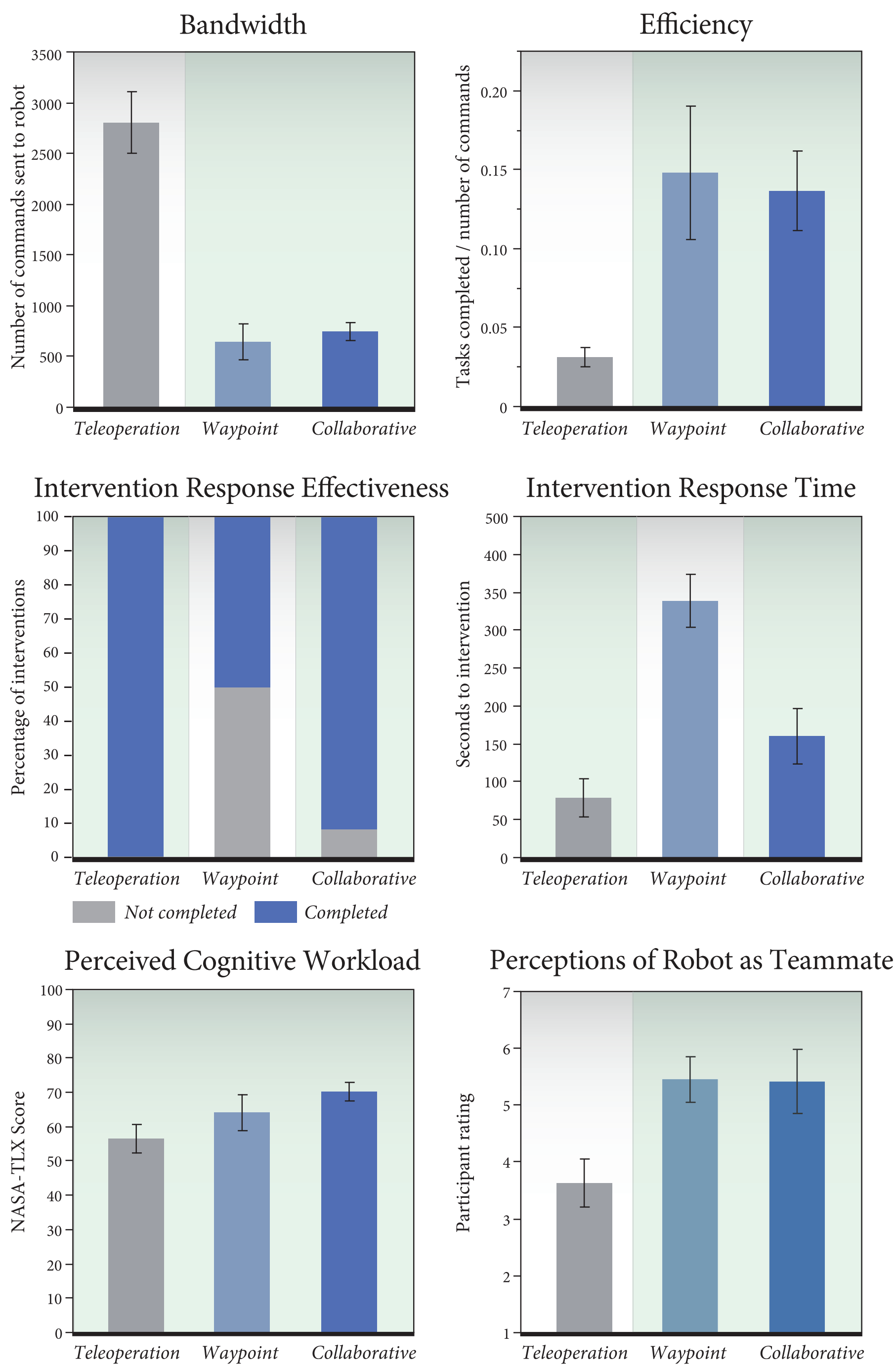

Fig. 10. Results from the objective and subjective measures. Green-shaded regions represent comparisons that were not statistically different at the $p < .05$ level based on Tukey's HSD test results. Overall, results supported Hypothesis 1 (bandwidth and efficiency), Hypothesis 2 (responsiveness), and Hypothesis 4 (perception of the robot), but not Hypothesis 3 (cognitive workload).

anticipated that participants in the waypoint and collaborative interface conditions would perceive the robot to be more of a partner or assistant and less of a tool compared with participants using the teleoperation interface. We found a significant effect of interface design on participant ratings of their perception of the robot, $F(2, 33) = 5.04$, $p = .012$. To test Hypothesis 4, we conducted a pairwise comparison across all conditions, which revealed that participant ratings of the robot as a teammate were significantly higher in the waypoint delegation ($M = 5.46$, $SD = 1.40$, Cohen's $d = 1.29$) and collaborative interfaces ($M = 5.42$, $SD = 1.94$, Cohen's $d = 1.05$) compared with the teleoperation interface ($M = 3.63$, $SD = 1.44$) at the $p < .05$ level, providing support for Hypothesis 4.

7. Discussion

The experimental results reaffirm our earlier argument that interface design decisions can have a significant impact on user collaboration with free-flying robots. We found that the presence, or absence, of specific design elements significantly affected user experience and overall work performance.

Our analysis provided support for Hypothesis 1; design elements that provide spatial reasoning and planning abilities can improve performance for certain metrics. In our evaluation, the interfaces with planning, a three-dimensional environment map, and three-dimensional waypoints (i.e., the waypoint delegation and collaborative interfaces) outperformed the teleoperation system in terms of bandwidth and efficiency, both important considerations for deployments where communication may be limited or intermittent, such as within the domain of space exploration. However, on average, participants scored higher in our overall task completion metric with the teleoperation interface than with the waypoint interface. At the same time, participants using the collaborative interface scored highest in the task completion metric, possibly indicating that spatial reasoning and planning is important, but is most beneficial when combined with support for temporal reasoning. This conclusion matches our second design implication from our formative study regarding the importance of *both* spatial and temporal support.

Our findings from the testing of Hypothesis 2 further reinforces the notion that simultaneously supporting spatial and temporal reasoning is important, showing that the interactive timeline, a design element for supporting spatio-temporal reasoning, was critical in enabling participants to quickly intervene and respond to spontaneous changes in task or environmental conditions. When examining intervention responses to unplanned events, we found no difference in responses between teleoperation and the collaborative interface, while participants were significantly slower and less effective at responding using the waypoint delegation system. Thus, when considering real operations in complex, dynamic environments, we see that support for spatial reasoning may only be beneficial when support for temporal reasoning is also provided.

The necessity of ensuring proper spatio-temporal support for free-flyer interfaces is a crucial design implication. Many current systems focus entirely on execution (as in the case of teleoperation), which may lead to inefficiencies, or only support (limited) spatial planning (as in the case of current two-dimensional waypoint planners) – a brittle design that risks failing in unplanned circumstances. In particular, many participants that used the waypoint delegation interface noted the challenge of using an interface that did not provide support for temporal reasoning, as expressed in the following excerpts from open-ended participant responses (each marked with participant ID number and the interface used).

> *P889 [Waypoint]:* 'The waypoint system made it difficult to follow supervisor instructions, because any new commands were added to the end of the list of waypoints.'

> *P222 [Waypoint]:* '…I also could not figure out how to interrupt the established waypoint in order to take the requested air-quality measure without erasing all of the previously established markers.'

> *P924 [Waypoint]:* 'The lack of the ability to insert a new task into the middle of the linked list of tasks or to change the type of task currently running resulted in a situation that required one to throw away the entire plan thereafter to deal with a sudden change…having to throw out my plan to deal with the air-quality emergency was beyond frustrating.'

These statements, along with our overall results, strongly advocate for designing interfaces that support plan adjustments and dynamic replanning during task execution.

While not directly related to objective task performance, we also investigated how interface design elements might alter user perception of the robot. We found evidence supporting Hypothesis 4, indicating that an interface designed for user collaboration with a free-flying robot better supported perception of teamwork, compared with a teleoperation interface that led to users perceiving the robot as more of a tool. User perception of free-flyers may be an interesting avenue for future research to explore, as prior work has indicated that such mental model formation affects human–robot coordination and joint activity (Bradshaw et al., 2009).

Hypothesis 3 remains unconfirmed; the results did not provide sufficient evidence for any difference in cognitive workload across conditions, measured by either the NASA-TLX or the raw secondary task performance. Three potential confounding factors may provide an explanation regarding the lack of support for Hypothesis 3. First, combining the strict time limit for the execution phase of the experiment with the requirement that participants quickly learn to use a novel interface and work with an unfamiliar robot to accomplish many tasks may have placed a great burden on cognitive load for all participants, resulting in a *ceiling effect* in measures of cognitive load and overshadowing any potential effects of interface design.

A second possible explanation stems from the between-participants design of the study, which does not allow us to control for individual differences in task completion rates. While we observe differences in secondary task completion rates as predicted by Hypothesis 3 (i.e., most secondary tasks completed by users of the collaborative interface and least by users of the safeguarded teleoperation interface), the data has too much variance to draw a statistical conclusion. Collecting data from a larger population, extending the total task time, or simplifying main tasks such that participants had more free time to complete secondary tasks may have granted us sufficient statistical power to understand whether interface design had an effect on cognitive load. Alternatively, we could have introduced rewards to motivate additional secondary task completion, potentially yielding a larger effect size. However, such an incentive may have reduced main task completion, which was the central focus of our study.

A within-participants design would have enabled participants to act as their own control regarding the variance in secondary task completion. However, a within-participants design would not be feasible for this study due to the difficulty of mitigating transfer effects such as learning and fatigue. In a within-participants design, it is unclear how to give users freedom in accomplishing tasks in their desired order without introducing learning effects (e.g., learning about the location of objects within the unfamiliar environment, learning about robot capabilities and flight speeds, and getting practice completing each task). Similarly, as the experiment apparently introduced high levels of cognitive load, it is likely that, in a within-participants design, fatigue could have diminished user performance on later trials, introducing additional noise in our data.

As a final potential confounding factor, the effects of interface design on perceived workload may have been influenced by participants' prior familiarity with video games, especially as the control scheme for the safeguarded teleoperation interface resembled interaction with first-person games. This explanation is supported by frequent comparisons that were made to video games in open-ended responses across all conditions.

> *P922 [Collaborative]:* 'I tend to use inverted controls when I (rarely!) play 1st person video games.'

> *P769 [Collaborative]:* 'I kept expecting the controls to work slightly differently, more like in Minecraft[15] I guess.'

> *P448 [Teleoperation]:* 'It honestly felt like a slightly challenging video game (well… I was using a controller after all).'

> *P522 [Teleoperation]:* 'Hook this bad boy up to a ps3 controller, and you'll get millions of proficient operators for this machine.'

As participants may have perceived similarities to their past experiences with video games, any unexpected alteration (e.g., not supporting inverted controls, using a particular game controller, and so on) may have increased the variance of reported cognitive workload. While the goal of this study was to inform the design of interfaces that better support teamwork and task completion overall, future studies might specifically investigate how free-flyer interfaces can be designed to reduce cognitive workload, possibly drawing on evaluation methods for predicting interface-induced workload across design elements (Zhang et al., 2016). Alternatively, studies might further explore trade-offs in designing interfaces to mimic video game interfaces, as matching prior user experiences and mental models by leveraging domains such as video games may diminish training times and enhance system intuitiveness.

7.1. Limitations and future work

In this work, we strove to design a variety of realistic experimental tasks. While we believe these tasks improve the ecological validity of our findings, their use introduces a challenge as we currently know little about how to measure and compare work across such diverse tasks and subtasks. For instance, looking at average performance across tasks, we see that participants in the collaborative interface completed (and completed correctly) more sound surveys and pipe inspections than participants using the teleoperation or waypoint-delegation interfaces. However, they also spent longer on these tasks and performed relatively worse on the inventory logistics task. Without understanding the inherent difficulties in these tasks, it is difficult to know how to best interpret aggregate performance. In this work, we used standardized measures (e.g., NASA TLX) where appropriate, but because we felt that it was important to analyze work completed and commands sent across all experimental tasks, we developed our own task completion score and efficiency metric. As we look to design interfaces for robots whose abilities enable human assistance across an ever-wider range of tasks, we believe that more work will be needed to develop appropriate ways of measuring aggregate usability and making comparisons of performance on specific tasks.

Overall, the results from our measures support the notion that the design elements for spatio-temporal reasoning implemented in the collaborative interface represented a positive trade-off in terms of efficiency and responsiveness. However, we note that, in our experiment, we sacrificed some aspects of experimental control (e.g., allowing participants to work on tasks in the order they chose) to better mimic realistic free-flyer deployments and further increase ecological validity. Certain aspects of our design (e.g., we controlled for the size of user interface elements, such as robot video feed, across conditions, although element position varied slightly in each interface) may have introduced experimental confounds or increased observed variance. Balancing experimental control with generalization to real-world settings is a delicate task, and future work is needed to more fully explore the design space for free-flyer interfaces.

Social aspects of teamwork represent one avenue that such future studies might examine more fully. While Hypothesis 4 examined the effects of interface design on perception of the robot as a teammate, the main focus of this study was on efficiency, as defined by total task completion versus commands issued, and responding to spontaneous changes in task/environment conditions. However, in the formative exploration, participants exhibited a great deal of social conversation, coded as 'miscellaneous' chatter, throughout the interaction. Similarly, participant responses in the evaluation study hint at potential for social interaction and enjoyment, even in the teleoperation condition.

> *P448 [Teleoperation]:* 'I essentially had a "robot buddy" who I could steer around and give directions to…'

> *P579 [Waypoint]:* 'If [the robot] has a name, it would be very adorable.'

> *P769 [Collaborative]:* 'I like the robot…I [sic] was fun!'

As prior studies have demonstrated that even ground robots designed for functional use and work (as opposed to robots designed for social companionship) will be viewed as social actors in domestic and manufacturing environments (Forlizzi, 2007; Sauppé and Mutlu, 2015), future studies might further explore the role and effects of social interaction in human–robot teaming involving free-flyers.

This study was also limited in scope to only exploring distal control and collaboration with a single free-flyer. While recent work has begun to examine proximal interactions for main users, such as using gestures (Cauchard et al., 2015; Ng and Sharlin, 2011) and examining the communication of intent (Szafir et al., 2014, 2015), the development of robust interactive systems that enable users to effectively direct and collaborate with co-located free-flying robots remains an open problem, as users will have direct visual access to the robot and be present in the work environment while interacting via the control interface. Similarly, additional research is needed to explore the scalability of free-flyer interface designs for working with multiple robots, which may find inspiration in similar studies of fan-out for mobile ground robots (e.g., Olsen Jr and Wood, 2004).

Each of our interface prototypes could be further enhanced to increase usability for real-world deployment. Participants provided several suggestions in their open-ended responses, including developing better systems to log collected data and providing a small two-dimensional 'mini map' that tracked robot location (another feature often found in video games). Further improvements might combine the collaborative design demonstrated here with the 'nudge' system developed by Pitman and Cummings (2012) for more precise control, provide a finer degree of control over planned paths (possibly in the manner proposed by Johnson et al., 2012), or integrate the video feed directly into the three-dimensional environment. Additional development is also needed to integrate the systems presented here, which used a three-dimensional map of the environment developed a priori, with other work examining how robots might develop such maps on the fly. Future studies might explore how to support planning and execution as a free-flyer builds a map in real time, or may investigate new workflows, such as first having a free-flyer autonomously scout an area to build a map, then letting the user construct a task plan (possibly while the robot recharges), and finally supporting collaborative teamwork to execute the plan.

7.2. Lessons learned

Throughout this work, we learned a great deal about potential strengths and weaknesses of our approach as well as where we believe gaps exist in our current understanding of HRI.

Our approach sought to address three limitations that we identified with coactive design: an over-reliance on designer intuition (e.g., when developing the interdependence analysis table), a focus solely on supporting real-time task execution, and a lack of transparency in communicating interdependence relationships to end users. To address the first limitation, we conducted the formative study, which allowed us to take a data-driven approach toward identifying the types of support a user might require. The second limitation was explored through prototyping the waypoint delegation and collaborative interface systems, both of which supported planning in addition to real-time execution. The development of the collaborative interface addressed the third limitation by visually communicating interdependence relationships (e.g., after each task the robot could autonomously advance, wait for user approval, or advance while notifying the user) via an interactive, editable timeline. While these extensions allowed us to collect new data and provided several valuable insights that may not have arisen otherwise, we recognize that there are both strengths and weaknesses to our approach and that more work is needed to integrate it with existing approaches and tools, such as the traditional coactive design methodology, heuristic evaluations, or cognitive walkthroughs.

Our formative study allowed us to collect data with real users, but we found it difficult to set up a simulated interaction that was truly analogous to real operations. For instance, our formative study required users to inspect targets at different heights, but the range of heights had to be constrained such that all targets were within reach of the confederate, an artificial limitation compared with an actual flying robot. Additionally, the interaction inherently involved two humans conversing, a high-bandwidth form of communication that may not align with lower-bandwidth HRIs. Future work could examine ways of conducting similar studies in which participants more closely control confederate actions or confederate capabilities more closely align with free-flyers. For instance, participants could be presented with an existing free-flyer interface that translates their commands to text messages or haptic signals for the confederate, who could carry an existing free-flyer using a pole or mount it to a gantry system to increase the simulated flight range. Despite these limitations, we believe our data-driven approach uncovered design implications that have not been reported by prior studies, such as the need to support both spatial and temporal reasoning.

We believe that our exploration of planning phases and visualizing interdependence relationships shows that providing such support will be central to improving the effectiveness of free-flyer interfaces. However, further work could explore both aspects in greater detail. For instance, we learned that planning could improve user efficiency, but we only explored planning support for waypoint-based interfaces. It may be fruitful to explore how planning systems could be designed for teleoperation paradigms, which are typically only designed in the context of real-time task execution.

Our system relied completely on users in developing task plans, which may place an undue burden on the user or lead to sub-optimal or infeasible plans. Interfaces for real-world use in complex environments, such as the ISS, may need to link user-centered planning interfaces with plan and schedule optimization systems, enabling a paradigm in which feasible plans are automatically generated that can be approved or adjusted by the user. Alternative visualization systems might also be explored, for instance, depicting interdependence relationships as graph structures rather than (or in addition to) the timeline used in the collaborative interface. Future designs might also support other stages such as supporting post-execution activities (e.g., data analysis) that may require understanding information from both the planning and execution stages and provide guidance for developing future plans.

Overall, we argue that our data-driven approach should complement, rather than compete with, alternative methods and that designers seeking to build future interfaces might best be supported by combining methodologies. For example, early in the design process, developers might use an interdependence analysis table from coactive design and conduct a formative exploration, as described in our approach, to validate designer assumptions using real data.

In addition to learning about our approach, we also found gaps in current knowledge in the HRI literature. In particular, we found that prior work was often of little help in constructing our experimental tasks and measures. We found few examples of studies that tested HRIs over a variety of realistic tasks and none that explored tasks within the context of flying robots operating within human environments. The lack of established tasks and measures introduced a challenge for measuring task performance; for example, in our study, how does completing two sound-survey measurements compare with completing an air-quality measurement in terms of work? As interface design decisions will inevitably involve trade-offs for performance on specific tasks, we felt that it was necessary to get a sense of overall productivity and thus developed our efficiency metric based on total work completed divided by commands issued. This process was inspired by grading systems for courses (Kulkarni et al., 2015), which may create aggregate scores by assigning weights across a variety of different tasks that may have unknown difficulties. Past HRI work has also used composite metrics, such as those utilized by Dragan et al. (2013) and Szafir et al. (2015), but as robots continue to progress in general-purpose utility, more work is needed to determine appropriate metrics in gauging performance over a variety of tasks with relative difficulties that may not be well understood and that may place varying burdens on human–robot teams (e.g., see Ezer et al., 2013).

7.3. Design implications

Overall, our results point to three specific design implications for improving robot interfaces. First, interface affordances should match both robot capabilities and desired levels of user interaction to avoid gulfs of execution. For instance, free-flying robots have the ability to navigate environments in a three-dimensional fashion, yet many current interfaces only support specifying waypoints in two dimensions. Similarly, we found that users are mostly interested in goal-directed actions within an environment (e.g., taking pictures), rather than maneuvering the robot within the environment. However, most current interfaces focus mainly on supporting robot movement instead of providing support at the level of activities and tasks. We recommend an aggregate solution, such as our symbolic waypoints that encode a desired robot position, orientation, and action. Second, interfaces should support both planning and execution stages, ideally within a single cohesive tool. Existing interfaces typically prioritize execution, while ignoring or only providing limited support for task planning. Third, interfaces must support dynamic replanning during execution. It is inevitable that unplanned changes will occur in task constraints and environmental conditions; interfaces must provide flexible support for users to interrupt execution and respond to such changes. This support can take many forms, such as allowing users to change the position of planned waypoints during execution or enabling an existing plan to be restructured on the fly.

8. Conclusion

In this work, we explored the design of interfaces to support users in tasking and directing free-flying robots in human environments. In a formative exploration, we studied human–human teams to generate design requirements for supporting users working with a flying robotic assistant to accomplish archetypal free-flyer sensing and data-collection tasks. Our findings, along with prior research in interfaces for mobile ground robots, informed the design of three prototype interfaces that were implemented as part of an end-to-end system for free-flyer control. We evaluated these interfaces in an in-person laboratory experiment that introduced a series of novel, realistic tasks for studying human interactions with a distal free-flyer operating within an indoor environment. We found that a collaborative interface that supported spatial waypoints in three dimensions as well as an interactive timeline for task planning and replanning significantly improved users' efficiency in accomplishing tasks, their ability to intervene in response to spontaneous changes in task demands, and their rating of the robot as a teammate. These findings demonstrate the potential of a data-driven approach for designing for teamwork and interdependence, and offer practical insights into addressing limitations in the design of current flying-robot interfaces.

Acknowledgements

We thank Brandi Hefty and Ross Luo for their assistance in this research.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported, in part, by the National Science Foundation (NSF) (grant number CNS-1135949) and the National Aeronautics and This research was supported by a NASA Space Technology Research Fellowship (award NNX12AN14H).

Notes

1. General Atomics: http://www.ga-asi.com/aircraft-platforms
2. Mission Planner: http://planner.ardupilot.com/
3. ArduPilot: http://ardupilot.org http://ardupilot.org
4. DJI Phantom 4: https://www.dji.com/phantom-4
5. Xbox 360 Controller: http://www.xbox.com/en-US/xbox-360/accessories/controllers
6. Xbox Kinect: http://www.xbox.com/en-US/xbox-one/accessories/kinect
7. Point Cloud Library: http://pointclouds.org/http://pointclouds.org/
8. AForge library: http://www.aforgenet.com/
9. Helix Toolkit library: https://github.com/helix-toolkit
10. EmguCV library: http://www.emgu.com/wiki/index.php/Main_Page
11. AscTec Hummingbird: http://www.asctec.de/en/uav-uas-drones-rpas-roav/asctec-hummingbird/
12. Samsung Galaxy S4 Mini: http://www.samsung.com/uk/consumer/mobile-devices/smartphones/galaxy-s/GT-I9195ZKABTU
13. IP Webcam: https://play.google.com/store/apps/details?id=com.pas.webcam
14. SlimDX: https://slimdx.org/
15. Minecraft: https://minecraft.net/

References

Achtelik M, Bachrach A, He R, et al. (2009) Autonomous navigation and exploration of a quadrotor helicopter in GPS-denied indoor environments. In: *Proceedings of ICRA'09*.

Adams JA, Rani P and Sarkar N (2004) Mixed-initiative interaction and robotic systems. In: *Workshop on supervisory control of learning and adaptive systems, nineteenth national conference on artificial intelligence (AAAI-04)*, San Jose, CA.

Albers D, Correll M and Gleicher M (2014) Task-driven evaluation of aggregation in time series visualization. In: *Proceedings of the 2014 ACM annual conference on human factors in computing systems*. ACM, pp.551–560.

Allen JE, Guinn C and Horvitz E (1999) Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications* 14(5): 14–23.

Ang KH, Chong G and Li Y (2005) PID control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology* 13(4): 559–576.

Bachrach A, Prentice S, He R, et al. (2011) RANGE – robust autonomous navigation in GPS-denied environments. *Journal of Field Robotics* 28(5): 644–666.

Bauer A, Wollherr D and Buss M (2008) Human–robot collaboration: A survey. *International Journal of Humanoid Robotics* 5(01): 47–66.

Bradshaw JM, Feltovich P, Johnson M, et al. (2009) From tools to teammates: Joint activity in human-agent-robot teams. In: *International conference on human centered design*, pp.935–944.

Bradshaw JM, Feltovich PJ, Jung H, et al. (2004) Dimensions of adjustable autonomy and mixed-initiative interaction. In: *Agents and computational autonomy*, pp.17–39.

Bradshaw JM, Hoffman RR, Woods DD, et al. (2013) The seven deadly myths of "autonomous systems". *IEEE Intelligent Systems* 28(3): 54–61.

Bright JR (1955) How to evaluate automation. *Harvard Business Review* 33(4): 101–111.

Bright JR (1958) Does automation raise skill requirements. *Harvard Business Review* 36(4): 85–98.

Brisset P and Hattenberger G (2008) Multi-UAV control with the paparazzi system. In: *HUMOUS 2008, conference on humans operating unmanned systems*.

Brunken R, Plass JL and Leutner D (2003) Direct measurement of cognitive load in multimedia learning. *Educational Psychologist* 38(1): 53–61.

Bualat M, Barlow J, Fong T, et al. (2015) Astrobee: Developing a free-flying robot for the International Space Station. In: *AIAA SPACE 2015 conference and exposition, AIAA SPACE Forum*. AIAA, 4643pp.

Bualat M, Fong T, Allan M, et al. (2013) Surface telerobotics: Development and testing of a crew controlled planetary rover system. In: *AIAA space*.

Bueno SS, Azinheira JR, Ramos J Jr, et al. (2002) Project AURORA: Towards an autonomous robotic airship. In: *Proceedings of IROS'02*, pp.43–54.

Cauchard JR, Jane L, Zhai KY, et al. (2015) Drone & me: An exploration into natural human-drone interaction. In: *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, pp.361–365.

Chen JY, Haas EC and Barnes MJ (2007) Human performance issues and user interface design for teleoperated robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 37(6): 1231–1245.

Christoffersen K and Woods DD (2002) How to make automated systems team players. *Advances in Human Performance and Cognitive Engineering Research* 2: 1–12.

Clark HH and Brennan SE (1991) Grounding in communication. *Perspectives on Socially Shared Cognition* 13: 127–149.

Corcoran M (2014) Drone journalism takes off. Available at: http://www.abc.net.au/news/2012-02-21/drone-journalism-takes-off/3840616 (accessed 11 August 2014).
Crandall JW, Goodrich M, et al. (2002) Characterizing efficiency of human robot interaction: A case study of shared-control teleoperation. In: *IEEE/RSJ international conference on intelligent robots and systems*, pp.1290–1295.
Devernay F and Faugeras O (2001) Straight lines have to be straight: Automatic calibration and removal of distortion from scenes of structured environments. *Machine Vision and Applications* 13(1): 14–24.
Dias MB, Kannan B, Browning B, et al. (2008) Sliding autonomy for peer-to-peer human-robot teams. In: *10th intelligent conference on intelligent autonomous systems (IAS 2008)*, Germany, pp.332–341.
Dorais G and Kortenkamp D (2001) Designing human-centered autonomous agents. In: Kowalczyk R, et al. (eds) *Advances in Artificial Intelligence* (*Lecture Notes in Computer Science*, vol. 2112). New York, NY: Springer, pp.321–324.
Dragan AD and Srinivasa SS (2013) A policy-blending formalism for shared control. *The International Journal of Robotics Research* 32(7): 790–805.
Dragan AD, Lee KC and Srinivasa SS (2013) Legibility and predictability of robot motion. In: *Proceedings of HRI'13*, pp.301–308.
Drury JL, Richer J, Rackliffe N, et al. (2006) Comparing situation awareness for two unmanned aerial vehicle human interface approaches. Technical report, DTIC document, Defense Technical Information Center, Fort Belvoir, VA.
Dunn M (2014) Astronauts flee US side of space station but no sign of leak. Available at: http://wtop.com/world/2015/01/ammonia-leak-at-space-station-poses-no-danger-to-crew/ (accessed 15 May 2015).
Eberhard KM, Nicholson H, Kübler S, et al. (2010) The Indiana "Cooperative Remote Search Task" (CReST) Corpus. In: *Proceedings of the seventh international conference on language resources and evaluation (LREC'10)*.
Elfes A, Siqueira Bueno S, Bergerman M, et al. (1998) A semi-autonomous robotic airship for environmental monitoring missions. In: *Proceedings of ICRA'98*, pp.3449–3455.
Ezer N, Sandor A and Vos G (2013) Human-robot interaction (HRI): Metrics and methods – A literature review. NASA Technical Report (in press), National Aeronautics and Space Administration, Washington, DC.
Ferrell WR and Sheridan TB (1967) Supervisory control of remote manipulation. *IEEE Spectrum* 4(10): 81–88.
Fitts PM (ed) (1951) *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.
Fong T, Berka R, Bualat M, et al. (2012) The human exploration telerobotics project. In: *Proceedings of GLEX'12*.
Fong T, Micire M, Morse T, et al. (2013) Smart SPHERES: A telerobotic free-flyer for intravehicular activities in space. In: *Proceedings of AIAA space'13*.
Fong T, Nourbakhsh I and Dautenhahn K (2003) A survey of socially interactive robots. *Robotics and Autonomous Systems* 42(3): 143–166.
Fong T, Thorpe C and Baur C (2001a) A safeguarded teleoperation controller. In: *IEEE international conference on advanced robotics (ICAR)*.
Fong T, Thorpe C and Baur C (2001b) *Collaborative control: A robot-centric model for vehicle teleoperation*. The Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.
Forlizzi J (2007) How robotic products become social products: An ethnographic study of cleaning in the home. In: *Proceedings of HRI'07*. ACM, pp.129–136.
Goodrich MA, Morse BS, Gerhardt D, et al. (2008) Supporting wilderness search and rescue using a camera-equipped mini UAV. *Journal of Field Robotics* 25(1–2): 89–110.
Hart SG and Staveland LE (1988) Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology* 52: 139–183.
Harwell D (2014) This government rule could cripple commercial drone flight. Available at: https://www.washingtonpost.com/blogs/the-switch/wp/2014/11/25/this-government-rule-could-cripple-commercial-drone-flight/(accessed 19 December 2014).
Higuchi K and Rekimoto J (2013) Flying head: A head motion synchronization mechanism for unmanned aerial vehicle control. In: *Proceedings of CHI'13 extended abstracts*. New York, NY: ACM.
Huang CM, Cakmak M and Mutlu B (2015) Adaptive coordination strategies for human-robot handovers. In: *Proceedings of robotics: Science and systems*.
Irizarry J, Gheisari M and Walker BN (2012) Usability assessment of drone technology as safety inspection tools. *Journal of Information Technology in Construction* 17: 194–212.
Johnson M (2014) *Coactive design: Designing support for interdependence in human-robot teamwork*. PhD Thesis, Delft University of Technology, Delft, The Netherlands.
Johnson M, Bradshaw JM, Feltovich PJ, et al. (2011) The fundamental principle of coactive design: Interdependence must shape autonomy. In: De Vos M, et al. (eds) *Coordination, Organizations, Institutions, and Norms in Agent Systems VI* (*Lecture Notes in Computer Science*, vol. 6541). New York, NY: Springer, pp.172–191.
Johnson M, Bradshaw JM, Feltovich PJ, et al. (2014) Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction* 3(1).
Johnson M, Carff J and Pratt J (2012) Coactive design for human-MAV team navigation. In: *International micro aerial vehicle conference and competition*.
Krajník T, Vonásek V, Fišer D, et al. (2011) AR-drone as a platform for robotic research and education. In: *Research and education in robotics – EUROBOT 2011*, pp.172–186.
Kulkarni C, Wei KP, Le H, et al. (2015) *Peer and self assessment in massive online classes*. New York, NY: Springer International Publishing, pp.131–168.
Landis JR and Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33: 159–174.
Lavigne P (2014) Eyes in the sports sky. Available at: http://espn. go.com/espn/otl/story/_/id/10974559/drones-use-ucla-mlb-sta rting-show-more-sports-fields-coaches-embrace-technology (accessed 11 August 2014).
Lee JD, Caven B, Haake S, et al. (2001) Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43(4): 631–640.

Lorite S, Muñoz A, Tornero J, et al. (2013) Supervisory control interface design for unmanned aerial vehicles through GEDIS-UAV. In: *Human-computer interaction: Human-centred design approaches, methods, tools, and environments*, pp.231–240.

Mandfield E (1989) The diffusion of industrial robots in Japan and the United States. *Research Policy* 18(4): 183–192.

Mulling K, Venkatraman A, Valois JS, et al. (2015) Autonomy infused teleoperation with application to BCI manipulation. In: *Proceedings of robotics: Science and systems*, Rome, Italy.

Murphy M (2016) There's now a drone racing league that feels like pod racing from Star Wars. Available at: http://qz.com/602230/theres-now-a-drone-racing-league-that-feels-like-pod-racing-from-star-wars/ (accessed 1 September 2016).

Murphy RR (2014) *Disaster Robotics*. Cambridge, MA: MIT Press.

Murphy RR, Casper J, Micire M, et al. (2000) Mixed-initiative control of multiple heterogeneous robots for urban search and rescue. Technical report, University of South Florida, Tampa, FL.

Mutlu B, Andrist S and Sauppé A (2014) Enabling human-robot dialogue. In: Markowitz J (ed.) *Robots that Talk and Listen*. Boston, MA: De Gruyter.

Ng WS and Sharlin E (2011) Collocated interaction with flying robots. In: *Proceedings of RO-MAN'11*, pp.143–149.

Norman DN (1986) Cognitive engineering. In: Norman DN and Draper SW (eds) *User Centered System Design: New Perspectives on Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 31–61.

Nudehi SS, Mukherjee R and Ghodoussi M (2003) A haptic interface design for minimally invasive telesurgical training and collaboration in the presence of time delay. In: *42nd IEEE conference on decision and control*, pp.4563–4568.

Okamura AM (2004) Methods for haptic feedback in teleoperated robot-assisted surgery. *Industrial Robot: An International Journal* 31(6): 499–508.

Olsen DR and Goodrich MA (2003) Metrics for evaluating human-robot interactions. In: *Proceedings of PERMIS*, pp.4–12.

Olsen DR Jr and Wood SB (2004) Fan-out: Measuring human control of multiple robots. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp.231–238. ACM.

Parasuraman R, Sheridan TB and Wickens CD (2000) 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.

Perlin K (1985) An image synthesizer. *ACM Siggraph Computer Graphics* 19(3): 287–296.

Pfeil K, Koh SL and LaViola J (2013) Exploring 3D gesture metaphors for interaction with unmanned aerial vehicles. In: *Proceedings of IUI'13*, pp.257–266.

Pires G and Nunes U (2002) A wheelchair steered through voice commands and assisted by a reactive fuzzy-logic controller. *Journal of Intelligent and Robotic Systems* 34(3): 301–314.

Pitman D and Cummings ML (2012) Collaborative exploration with a micro aerial vehicle: A novel interaction method for controlling a MAV with a hand-held device. *Advances in Human-Computer Interaction* 2012: 15 pp.

Quigley M, Goodrich M, Beard RW, et al. (2004) Semi-autonomous human-UAV interfaces for fixed-wing mini-UAVs. In: *2004 IEEE/RSJ international conference on intelligent robots and systems*, pp.2457–2462.

Rae I, Venolia G, Tang JC, et al. (2015) A framework for understanding and designing telepresence. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pp.1552–1566. ACM.

Reavis B and Hem B (2011) Honeywell T-Hawk aids Fukushima Daiichi disaster recovery: Unmanned micro air vehicle provides video feed to remote monitors. *Honeywell Aerospace Media Center*. Honeywell International Inc., 19 April 2011. Available at: http://honeywell.com/News/Pages/Honeywell-T-Hawk-AidsFukushima-Daiichi-Disaster-Recovery.aspx (accessed 12 October 2012).

Ruff HA, Narayanan S and Draper MH (2002) Human interaction with levels of automation and decision-aid fidelity in the supervisory control of multiple simulated unmanned air vehicles. *Presence: Teleoperators and Virtual Environments* 11(4): 335–351.

Sadler PM and Tai RH (2007) Weighting for recognition: Accounting for advanced placement and honors courses when calculating high school grade point average. *NASSP Bulletin* 91(1): 5–32.

Sauppé A and Mutlu B (2015) The social impact of a robot co-worker in industrial settings. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp.3613–3622. ACM.

Scaramuzza D, Achtelik MC, Doitsidis L, et al. (2014) Vision-controlled micro flying robots: From system design to autonomous navigation and mapping in GPS-denied environments. *IEEE Robotics & Automation Magazine* 21(3): 26–40.

Sebanz N, Bekkering H and Knoblich G (2006) Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences* 10(2): 70–76.

Shah J and Breazeal C (2010) An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 52(2): 234–245.

Shah J, Wiken J, Williams B, et al. (2011) Improved human-robot team performance using Chaski, a human-inspired plan execution system. In: *Proceedings of the 6th international conference on human-robot interaction*, pp.29–36. ACM.

Sharkey N (2009) Death strikes from the sky: The calculus of proportionality. *IEEE Technology and Society Magazine* 28(1): 16–19.

Sharma M, Hildebrandt D, Newman G, et al. (2013) Communicating affect via flight path exploring use of the laban effort system for designing affective locomotion paths. In: *Proceedings of HRI'13*, pp.293–300.

Sheridan TB (1986) Human supervisory control of robot systems. In: *1986 IEEE international conference on robotics and automation*, pp.808–812. IEEE.

Sheridan TB and Verplank WL (1978) Human and computer control of undersea teleoperators. Technical report, MIT Man-Machine Systems Laboratory, Cambridge, MA.

Steinfeld A, Fong T, Kaber D, et al. (2006) Common metrics for human-robot interaction. In: *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction*, pp.33–40. ACM.

Stout RJ, Cannon-Bowers JA, Salas E, et al. (1999) Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 41(1): 61–71.
Stricker LJ, Rock DA, Burton NW, et al. (1994) Adjusting college grade point average criteria for variations in grading standards: A comparison of methods. *Journal of Applied Psychology* 79(2): 178–183.
Szafir D, Mutlu B and Fong T (2014) Communication of intent in assistive free flyers. In: *Proceedings of HRI'14*. ACM, pp.358–365.
Szafir D, Mutlu B and Fong T (2015) Communicating directionality in flying robots. In: *Proceedings of HRI'15*, pp.358–365. ACM.
Szondy D (2015) South Korea's team KAIST wins 2015 DARPA robotics challenge. http://newatlas.com/darpa-drc-finals-2015-results-kaist-win/37914/(accessed 11 November 2016).
Tversky B (1991) Spatial mental models. *Psychology of Learning and Motivation* 27: 109–145.
Van Gerven PW, Paas F, van Merriënboer JJ, et al. (2006) Modality and variability as factors in training the elderly. *Applied Cognitive Psychology* 20(3): 311–320.
Wang B, Chen X, Wang Q, et al. (2010) Power line inspection with a flying robot. In: *Proceedings of CARPI'10*. IEEE, pp.1–6.
Winkvist S, Rushforth E and Young K (2013) Towards an autonomous indoor aerial inspection vehicle. *Industrial Robot: An International Journal* 40(3): 196–207.
Zhang W, Feltner D, Shirley J, et al. (2016) Unmanned aerial vehicle control interface design and cognitive workload: A constrained review and research framework. In: *Proceedings of the IEEE international conference on systems, man, and cybernetics (SMC'16)*.