

Evaluation of Interface Modality for Control of Multiple Unmanned Vehicles

Gloria L. Calhoun^{1(✉)}, Heath A. Ruff², Kyle J. Behymer^{2(✉)},
and Clayton D. Rothwell²

¹ Air Force Research Laboratory, 711 HPW/RHCI, Dayton, OH, USA

gloria.calhoun@us.af.mil

² Infoscitex, Dayton, OH, USA

{heath.ruff.ctr,kyle.behymer.1.ctr,

clayton.rothwell.ctr}@us.af.mil

Abstract. The U.S. Air Force envisions future applications in which a single human operator manages multiple heterogeneous unmanned vehicles (UVs). To support this vision, a range of play-based interfaces were designed by which an operator can team with autonomy (consisting of several intelligent agents/services) to manage twelve air, ground, and sea surface UVs performing security defense tasks for a simulated military base. To enable flexible delegation control, the interfaces were designed to enable the operator to use one or more of three control modalities in calling and editing plays that define UV actions. Specifically, each step defining a play could be completed: (1) manually, via mouse/click inputs, (2) by touching a touchscreen monitor, or (3) via speech commands. This paper reports results relevant to input modality from two experiments where operators were free to choose which modality to employ. Operators overwhelmingly used the mouse compared to the touchscreen or speech and were faster and more accurate with the mouse. Subjective data also favored the mouse modality with operators commenting that it was more intuitive to use with the play calling interfaces. Results are discussed and recommendations for further multimodal research are provided.

Keywords: Multimodal interfaces · Unmanned systems · Autonomous vehicles · Speech recognition · Touchscreen

1 Introduction

Multimodal interfaces have the potential to enhance human-autonomy interaction. For example, system inputs made with speech, gesture, and touch leverage natural human communication capabilities. Thus, multimodal interface concepts should be evaluated in conjunction with other technology advancements towards enabling single operator management of multiple heterogeneous unmanned vehicles (UVs). The Air Force

The rights of this work are transferred to the extent transferable according to title 17 U.S.C. 105.

Research Laboratory (AFRL) recently led a multi-service effort to integrate several autonomy advancements into a control station prototype referred to as “IMPACT” (Intelligent Multi-UxV Planner with Adaptive Collaborative/Control Technologies) [1]. The interfaces in the IMPACT system were designed to support a wide spectrum of human-autonomy control of multiple UVs (air, ground, and sea surface) as they perform dynamic security mission tasks defending a simulated military base [2]. At one extreme, the operator calls “plays” that define the actions of one or more UVs. With this play-based adaptable automation approach, the operator can quickly task UVs by specifying high level commands indicating the play type and location, and the autonomy determines all other parameters. For example, when an IMPACT operator calls a play to achieve air surveillance (play type) on a building (location), an intelligent agent recommends a UV to use (based on estimated time en route, fuel use, environmental conditions, etc.) and a cooperative control algorithm provides the quickest route to get to the building (taking into account no-fly zones, etc.). At the other end of the control spectrum, the operator can manually control UV movement with keyboard/mouse inputs or build plays from the ground up with minimal autonomy assistance. Between these two extremes of the control spectrum, the operator makes more inputs to the play, for instance specifying parameters and constraints that the autonomy may not be aware of (e.g., current visibility which can drive UV/sensor payload choice).

Besides providing the operator flexibility on the degree to which the autonomy assists with UV control, the interfaces in IMPACT were also designed to provide the operator flexibility in terms of which control modality could be employed to make inputs [1]. Specifically, plays could be called or edited: (1) via mouse/click inputs, (2) by touching a touchscreen monitor, or (3) via speech commands. These multimodal inputs support the overarching architecture that allows the operator to flexibly interact with autonomy at any time, employing any of the three control modalities. In other words, the interfaces were designed to support all three modalities for each step in utilizing the play-based interfaces. This approach was based on past research that has shown that most users prefer interfaces that are multimodal versus unimodal (e.g., 56–89% of users in an evaluation comparing spoken, written, and combined pen/speech input [3]). This preference reflects a number of advantages of having multiple control modalities available. First, the operator may prefer selecting which modality to employ, even alternating between input modalities to capitalize on the advantages of each modality [4]. Having multiple modalities available helps prevent the overuse of any individual mode. Also, some modalities may be more aligned with certain types of tasks or environmental situations [5]. For instance, making inputs with speech commands may not be ideal if the operator is involved in conversations or is in a noisy environment. Certain types of information are less amenable to vocal specification too (e.g., temporal relationships) [6]. Having multiple modalities available also allows the operator to leverage knowledge and past experience with respect to when and how to deploy a modality for the most efficient and accurate inputs [4].

Past research supports the use of multimodal interaction in UV applications. The utility of touch [7], speech-based input [8, 9], and “spatial dialog” (a combination of speech and touch input [10]) has been examined for single air UV control. In multi-UV control research by Levulis and colleagues [11], participants supervised a team of three air UVs and two manned helicopters traveling towards a landing zone to deploy ground

troops. Results showed that both touch and multimodal (touch and speech) input conditions were better than the speech-only condition in terms of task performance and subjective ratings of workload, situation awareness, and input usability. Their tasks focused on the input modality for monitoring and reporting status (e.g., classifying photographs, responding to instrument warnings, and addressing task queries), in contrast to the present research that emphasizes play-based UV control.

For play-based interfaces that establish respective human/autonomy roles in task completion, prior research has also shown the benefits of multimodal input. In a simulation demonstration of multiple input methods (keyboard/mouse, touch, and speech) for calling *single* air UV plays, pilots commented that the use of speech input was a natural method, but multimodal options should be available since certain tasks lend themselves to one mode versus another [12]. There were, however, concerns about the vocabulary training and memory requirements if the number of plausible plays and associated parameters is large. Additionally, speech commands should have a meaningful relationship (e.g., semantic) to their resulting actions [13].

Multiple air UVs were successfully controlled via plays called with speech recognition in a flight demonstration of a delegation control interface used in an urban mission scenario. Mean reaction time to mission events was significantly shortened with speech recognition, reflecting the ability for the operator to bypass cumbersome menu control steps [14]. This was viewed as especially advantageous during time critical mission phases. Similar advantages for play-based speech control were found in a simulation evaluation comparing multimodal inputs for control of three air UVs [15, 16]. Participants could either call plays with speech or by performing drag/drop actions to move symbology with the mouse or finger into “activity windows” used to construct plays for one or more UVs. Data on which control modality was used most frequently was not reported. However, participants commented that even though the ability to employ multiple input methods was useful, input would be even more flexible if the operator could switch between modalities while calling a single play. In other words, all modalities should be available in specifying a play such that the operator can flexibly switch between methods on a step-by-step basis [17].

The present paper will describe the multimodal play-based control approach used in two recent experiments employing the IMPACT simulation (Fig. 1).



Fig. 1. IMPACT simulation

Across the experiments, data were collected from fourteen participants familiar with unmanned vehicle operations and/or base defense missions. Operators received briefing and training, including multimodal control practice, for base defense mission related tasks involving simulated air, ground, and sea surface UVs. Each of the two experiments will be described separately, followed by a summary discussion. For each experiment, an overview of the play-based interfaces and methodology will be provided. This will be followed by report of the results that specifically pertain to which modality (mouse/click inputs, touch, or speech) was employed when interacting with the play-based interfaces, as well as other modality relevant objective and subjective data. However, first a brief overview of key elements of the play-based interfaces will be provided as well as methodology details common across the two experiments.

2 IMPACT Play-Based Interface Approach

Given that the previous research [12–17] supported relatively few plays (and primarily a single UV type), extensive development was required to implement play-based interfaces in IMPACT for heterogeneous UV control (see [2, 18]). The design and implementation process was incremental. Experiment 1 provided only a few play-based interfaces to control 6 UVs. In contrast, Experiment 2 featured refinements for the interfaces utilized in Experiment 1, as well as additional interfaces to better support control of 12 UVs. The differences in play-related interfaces between the experiments will be described in Sects. 3 and 4. As an introduction, this section provides an overview of common elements (see [2, 18] for more details).

2.1 IMPACT Simulation

In both experiments, operators sat at the IMPACT control station supported by an AFRL developed Fusion software framework for coordinating the communications from multiple systems and software components (for more details, see [19]). The control station was designed to support simulated single operator management of multiple heterogeneous UVs performing a base defense mission. (The mission details and tasking were informed by an earlier cognitive task analysis [20].) The operator's key task was to respond to mission events that required the reassignment of one or more UVs from normal patrol to instead either investigate a threat or perform other defensive measures (e.g., surveil the ammo dump every 30 min).

The control station for both experiments contained a keyboard, mouse, foot-pedal (for push-to-talk speech control), and a Plantronics GameCom Commander headset with boom microphone (for speech input and audio feedback). (Experiment 2 contained an additional foot-pedal for radio communication with a confederate sensor operator.) Both experiments employed four monitors (Fig. 1). The top center monitor presented a Tactical Situation Display (TSD) that included a geo-referenced map showing the locations of each UV and its associated on-going patrol route (white symbology, with gray UV symbols), route if under manual control (dark gray), or ongoing play. Symbology for each different play was presented in a unique color; if the play involved multiple

UVs, all the UVs and respective routes were presented in the same color. Lines depicting routes were coded to differentiate ongoing patrol and plays (solid lines) from plays being developed (dashed lines).

Both the left and right monitors were considered auxiliary displays (Fig. 1). The left monitor presented “help” information related to the mission and the right monitor presented imagery from each UV’s sensor (simulated via SubrScene Image Generator; www.subrscene.org/). While operators were presented payload information (as well as symbology showing each sensor’s field-of-view on the map), the operators’ responsibility in both experiments was to manage the movement and tasking of the UVs. Operators were briefed that a remotely located sensor operator (simulated in Experiment 1) was tasked to monitor and interpret the sensor imagery and that assessment information would be communicated from the sensor operator and/or commander via chat (and/or radio in Experiment 2).

The lower center sandbox (touch sensitive) monitor (Fig. 1) presented elements of the TSD, as well as several interfaces pertaining to play calling and management. This monitor provided a workspace for operators to interact with the UVs and autonomy support without obscuring the current state of the world (which was always visible on the TSD). Interactions involved either keyboard/mouse/touch inputs to manually control a UV’s movement (Experiment 2 only) or play-based interfaces to call and edit single- and multi-UV plays. In both experiments, three input modalities (mouse/click, touch, and speech) were implemented for operators to interact with the play-based interfaces. All three input modalities were identical in terms of the control actions initiated, as well as other control station feedback.

2.2 IMPACT Control: Pictorial Symbology and Speech Commands

Concise symbology (illustrated in Fig. 2) was used on the sandbox’s map and play interfaces to represent each UV and play type [18]. Each UV symbol was shape coded (e.g., air: plane form, ground: wheeled rectangle, surface: finned pentagon). Each play type was depicted by a circle with inner pictorial symbology representing the UV(s)’ task (e.g., plus sign to surveil either a point/location, line for road, and square for an area). The UV(s) associated with each play was represented by both shape and location coding on the play icon’s surrounding circle (e.g., air UV upper left versus ground UV lower left). Concise symbology was also designed to represent many play-related details

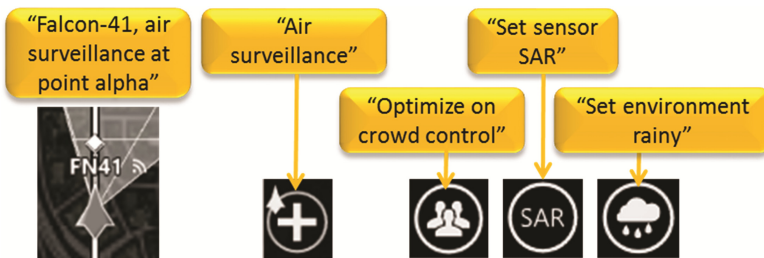


Fig. 2. Sample UV and play icons and associated speech commands

(Fig. 2) such as the target size, current environment, play priority, and what factors/constraints are pertinent to the play [18].

Each pictorial symbol utilized in the play-related interfaces presented UV/play related information and acted as a control element to initiate the play. Selecting the symbol with either a mouse click or single finger touch and release (“lift off” or “last contact” touching) changed the UV/play functioning in some manner, affording the operator the advantages of direct perception [21] and manipulation [22]. With these two manual input modes, the operator directly acted on the object of interest.

To implement the speech input mode, a companion speech command (either a word or phrase) was determined for each manual input. To illustrate, Fig. 2 provides the speech commands for a sample of icons used in the play interfaces for specifying play type and detail. During experiments, speech command reference information was displayed on the left auxiliary monitor. (In Experiment 1, the speech system had a vocabulary of 84 words and was capable of recognizing and parsing 2160 phrases. The system was expanded in Experiment 2 to 322 words with the capability of recognizing and parsing tens of millions of phrases.) In both experiments, a push-to-talk (PTT) approach was used to differentiate speech commands issued into a headset from other auditory communications. Operators signaled the Sphinx speech recognition system [23] to start processing the verbal input by either depressing a pedal on the floor or by clicking or touching a bar on the lower monitor. After the PTT switch was released, operators could confirm if the command was recognized from the auditory and visual feedback provided. (The visual feedback was presented on the PTT bar for 2 s before fading away and also displayed in a scrolling chat window exclusively dedicated to speech interaction).

2.3 IMPACT Familiarization for Experimental Participants

Experimental sessions began with operators completing a demographics questionnaire. Next, operators were given a simulation overview describing the project’s goals and introducing the concept of play calling. Operators were then seated at the IMPACT station and given a mission briefing that included:

- A description of the UVs they would be controlling, how each UV, its route, and its sensor footprint were represented on the map, and the tasks that each UV was responsible for performing in support of base defense operations.
- An overview of the base they would be defending including the base’s perimeter, sectors, critical facilities, patrol zones, and the named areas of interests in the area immediately surrounding the base.
- An explanation of their role as a multi-UV operator supporting base defense operations: in response to chat messages from a remotely located commander and sensor operator (played by confederates), they would be assigning high-level tasks to the UVs while the autonomous system components flew, drove, and operated the UVs.

The experimenter then provided the operator with a high-level description of the IMPACT interfaces. A detailed explanation of how plays could be called by using speech commands and/or clicking or touching icons designating play types and details was given, followed by providing time for operators to interact with the system until they

reported being familiar with the various interfaces and three input modalities. Additional methodological details are provided in the next two sections.

3 Experiment One

The first experiment was designed to evaluate the initial IMPACT interfaces in a 20-minute trial, as well as identify potential design improvements. Operators were provided 13 plays to task six UVs (three air, two ground, and one sea surface) in performing the base defense mission. Operators called plays using either speech commands (e.g., “air surveillance”) or making touch or mouse click inputs on a Play Creator interface (Fig. 3a). Once a play was initiated, operators selected a play location either with a speech command (e.g., “at the flight line”), from a dropdown menu of previously identified points, or by clicking on the map (see Fig. 3b). Further play details could be specified with speech commands and mouse/touch inputs by expanding the interface to show additional views of play-related options (Fig. 4). The operator could also view a list of active plays and the progress of plays related to several parameters. For more detailed descriptions of play calling and monitoring interfaces, as well as methodology and evaluation results, see [20].

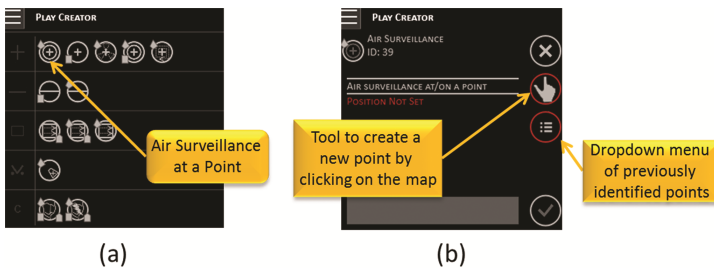


Fig. 3. Play Creator interface. (a) icons/buttons for each of Experiment 1’s thirteen plays. (b) methods for identifying play location [20]

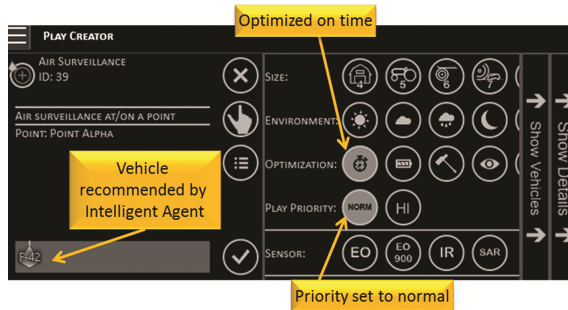


Fig. 4. Play Creator interface: Sample mechanisms for autonomy and operator to communicate UV and other constraints and details used in generating play plans [20]

3.1 Method

Participants. Seven volunteers from a U.S. Air Force Base participated. Three operators had prior experience flying UAVs (Predator, ScanEagle, Global Hawk, Shadow) as well as manned aircraft. Four operators were active Air Force security force personnel with experience conducting base defense operations in deployed environments (Afghanistan, Germany, Iraq, Kuwait, and Saudi Arabia). All operators were male and reported normal or corrected-to-normal vision and normal hearing.

Equipment. Six computers were used for IMPACT in Experiment 1 (a Dell T5610 & five Dell R7610 s running Windows 8.1). One computer ran IMPACT and the AMASE (AVTAS: Aerospace Vehicle Technology Assessment and Simulation - Multi-Agent Simulation Environment) vehicle simulation (used to simulate the UVs). One computer ran the test operator console and simulation for simulated entities in the sensor videos (Vigilant Spirit Simulation [25]), three computers ran two simulated (SubrScene) sensor videos, and one computer ran an XMPP (Extensible Messaging and Presence Protocol) Chat server for simulated communications. This IMPACT version used four 68.58 cm touchscreen monitors (Acer T272HUL; usable touch screen area: 59.69×33.66 cm; 2560×1440 resolution; tilted 45° from horizontal).

Procedure. After the general overview of the IMPACT simulation, mission-related tasks, and input modalities available for play calling, operators received a detailed briefing on the play-related interfaces available in Experiment 1. Next, training focused on providing operators with experience with each input modality. Operators received 12 chat messages asking them to call a play using a specific modality (e.g., “Using speech, call an air surveillance at Point Alpha”). Table 1 lists the exact sequence of twelve plays operators were asked to call during this portion of the training as well as the modality operators were instructed to use (4 plays for each modality).

Table 1. Play calling modality training

Method	Play
Speech	Air surveillance at point alpha
Speech	Normal full coverage patrol
Speech	Ground inspect at route brave
Speech	Air expanding square at point charlie
Touch	Air sector search at point alpha
Touch	Surface parallel search at area delta
Touch	Air parallel search at area bravo
Touch	Ground inspect at point charlie
Mouse	Air ground surveillance at point bravo
Mouse	Air inspect at route charlie
Mouse	Air surface parallel search at area delta
Mouse	Go highly mobile at point alpha

Operators were then trained on how to specify constraints, vehicles, and details when calling and/or editing a play. For all three input modalities, operators were asked via chat messages to call a specific play (e.g., “Using speech, call an air surveillance on Point Alpha, set sensor to EO, and optimize for low impact”), then make edits to the ongoing play (e.g., “Change the loiter type to a figure 8”). If an operator made a mistake, the experimenter provided feedback and the operator tried again until he had successfully completed the correct action. On average, training lasted one hour and was followed by a short break before the experimental scenario.

The goal of the 20-minute experimental scenario was to provide operators with the opportunity to exercise all of IMPACT’s capabilities within a realistic base defense scenario. Operators were informed that the scenario would begin with an air UV investigating a suspicious watercraft with all other UVs on a high alert patrol. Table 2 lists the exact sequence of mission events that occurred. Operators were instructed to respond to each chat message by calling one or more plays that best addressed the event. Operators were free to choose which of the three input modalities to employ in completing each step of the play calling process.

Table 2. Sequence of Mission Events

Event	Description
1	Operator receives a chat message from the sensor operator that the unidentified watercraft is a fishing boat
2	Operator receives a chat message from the commander to resume normal base defense operations/patrols.
3	Intelligent agent recognizes a serendipitous surveillance opportunity (an air UV is near a critical facility) and recommends a play (air surveillance at the critical facility) to the operator.
4	Operator receives a chat message from the commander to send an air UV to point charlie and to instruct other UVs to go highly mobile in response to a patrol reporting smoke at point charlie.
5	Operator receives a chat message from the sensor operator to send a ground UV to point charlie to get a better look.
6	Operator receives a chat message from the commander to provide the ground UV headed to point charlie with an air UV overwatch.

Once the experimental scenario was completed (approximately five plays called), operators completed paper questionnaires on the overall IMPACT system and its components. Then a semi-structured interview was conducted to capture additional feedback on IMPACT and its associated technologies including the three different input modalities. The entire procedure lasted approximately 3.25 h.

3.2 Results

Of the seven individuals who participated in the study, six completed the training and mission in the allotted time. Due to unanticipated time restrictions, the seventh operator was unable to complete the study and was eliminated from the data analysis. Due to the

small number of participants, UV operators and security force personnel data were not analyzed separately.

Subjective feedback on touch and speech was mixed. In general, operators seemed to like the idea of being able to execute plays via touch and speech. However, operators expressed concerns about the touchscreen’s lack of precision (an operator might touch an icon three times before the system registered it) and the speech system’s poor accuracy (an operator might utter the same command several times before it was recognized; the word error rate was 21.95% for in-grammar utterances).

Objective data were collected on the modality (mouse, touch, or speech) that operators used to call plays during the experimental mission (when operators choose which input modality to employ). Operators used the mouse much more frequently than touch or speech (see Fig. 5). Note that data for ‘speech’ is labeled “Speech/Mouse Confirm” because when operators used speech during the mission, they always used it in conjunction with the mouse: operators would initiate a play call with a speech command but execute the play by clicking the checkmark with the mouse instead of saying the speech command “Confirm” to execute the play. Actually, only two operators employed the speech modality during the experimental trial, and a different operator (only one) employed the touch input modality. Operators made a higher percentage of major errors (defined as failing to complete a play) when using touch than mouse or speech (see Fig. 5).

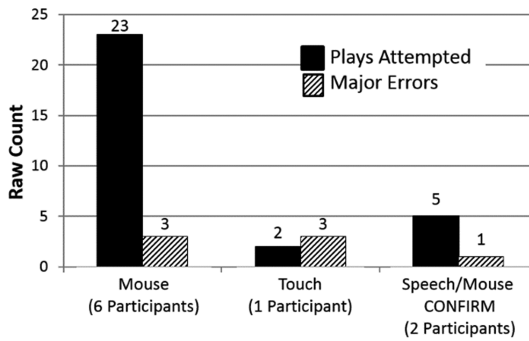


Fig. 5. Number of plays attempted and number of errors by input modality

Operators were also faster at completing plays using the mouse as compared to using touch or speech (Fig. 6a). However, this difference most likely reflected the reported problems operators had with touch input. In fact, when only plays correctly completed (i.e., no major errors) were examined, the mean difference between time to complete a play with the mouse and speech was just 2.5 s (see Fig. 6b). Note that operators never correctly completed a play using touch input.

Operators overwhelmingly used the mouse input modality compared to the touch or speech, and were faster and more accurate with the mouse as well. Several factors may have contributed to these results. Multiple operators had difficulties with the touchscreen registering inputs, but commented that if the touchscreen had worked better they would have been more likely to use it. For example, one operator stated, “Touchscreen could be extremely intuitive and quick if implemented correctly.”

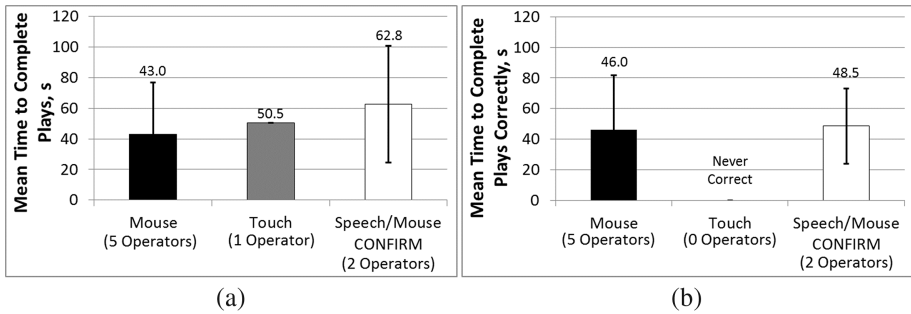


Fig. 6. Mean time to complete a play call by modality: (a) all plays, (b) plays called correctly. (Error Bars are the standard deviations.)

Several operators also spoke favorably of the speech input modality, especially the security force personnel, who mentioned that the speech commands were very similar to the dispatch calls they make during security force operations. However, this preference was not reflected in performance, as operators used the mouse more than speech to call plays. Several operators commented that they weren't completely familiar with the speech vocabulary, suggesting inadequate training. In the end, operators may have chosen to use the mouse modality due to its reliability; clicking a play icon with the mouse consistently resulted in the desired action, while touching a play icon or issuing a speech command often failed to register an input.

3.3 IMPACT Modifications

Based on the results of Experiment 1 (see Sect. 3.2), several modifications were made to improve IMPACT's touch and speech input modalities. For touch input, the main concern was that the sizes of selectable areas were too small (e.g., for play icons in Fig. 3a: 6.35 mm diameter circles with 1.59 mm separation). Although smaller targets (1.7 mm) were selectable in earlier research [26], MIL-STD-1472G [27] recommends a 15.2×15.2 mm area. To help aid play icon selection in Experiment 2, the diameter of the play icon's selectable area was increased slightly (7.94 mm diameter). Additionally, the touchscreen was replaced with a slightly larger one positioned at a lower tilt angle (see Sects. 3.1 and 4.1).

For the speech modality, the finite grammar was dramatically expanded to allow hundreds more ways to say things, resulting in a large increase in flexibility and naturalness. Commands were also added to support a more complex mission (i.e., more UVs, larger variety of play types, and ability to specify play details with speech). By modifying the speech pipeline, the operator in Experiment 2 could change symbology clutter level with speech and issue verbal queries to autonomy (e.g., "which vehicle can get to the flight line the fastest?" followed by aural and text responses). This process attempted to strike a balance between flexibility and accuracy, as changing from a closed- to an open-language vocabulary can increase recognition errors dramatically. Besides this expansion of the speech system language model, the acoustic model was changed, based on extensive testing.

4 Experiment Two

Besides the modifications described in Sect. 3.3, the simulation and play-based interfaces were expanded to provide operators 25 base defense related plays, in addition to two types of patrols. The assets were also increased to 12 UVs (four each of air, ground, and sea surface), with more variety in payload, including some with weapons. This involved making many changes to the symbology and interfaces used in Experiment 1, besides the modifications prompted by the operators' comments. Figure 7 illustrates the revisions made to the interfaces used in Experiment 1 to call a play and support operator-autonomy communications on play details. For more details on the interfaces for play calling and monitoring, see [24].



Fig. 7. Play Calling and Play Workbook used in Experiment 2 [24]

In addition to revising the interfaces used in Experiment 1, additional play based interfaces were added. Two of them provided other means to employ mouse and touch input modalities. One interface termed “radial menu” allowed operators to call plays directly from the map, instead of utilizing the play-calling interface illustrated in Fig. 7a. When the operators selected a location on the map (Fig. 8a) or a UV on the map (Fig. 8b), a radial menu appeared consisting of only the play options relevant to that location or UV (e.g., no ground based plays if a sea surface UV selected). The radial menu appeared with a right click of the mouse or if the operator touched the screen with a finger, continued the touch (i.e. lingered), and released upon feedback (white square presented) that the selection was registered.

The second additional play-calling interface was available in a task manager that maintained a list of tasks to be completed based on prompts in the chat window and previously defined (quick reaction checklist) steps for addressing certain mission events. As shown in the illustration (Fig. 9), the top row contains the mission prompt from chat (“Unidentified Watercraft...”) and each row below it shows established tasks that should be performed in response to such an event. Selecting each of these rows (a single touch and release, signaling left mouse click) called up a row below it that either displayed more detailed task text or a play button to select and call up the corresponding Play Workbook (Fig. 7b) to further support the play calling.

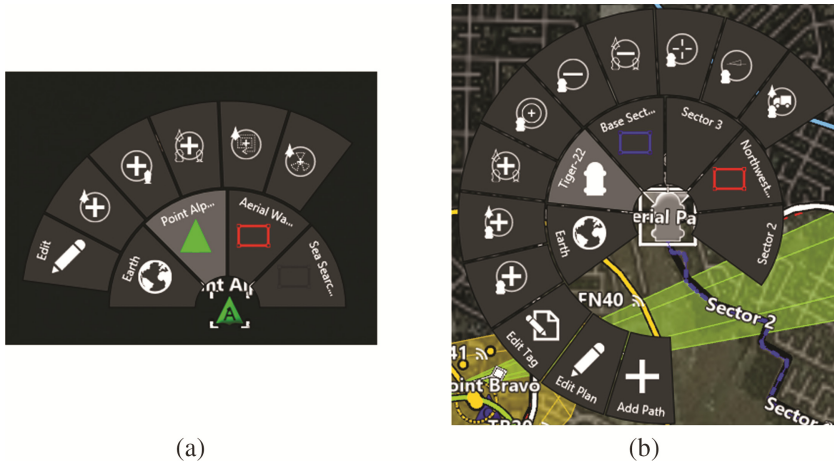


Fig. 8. Illustration of radial menu when: (a) UV selected or (b) location selected

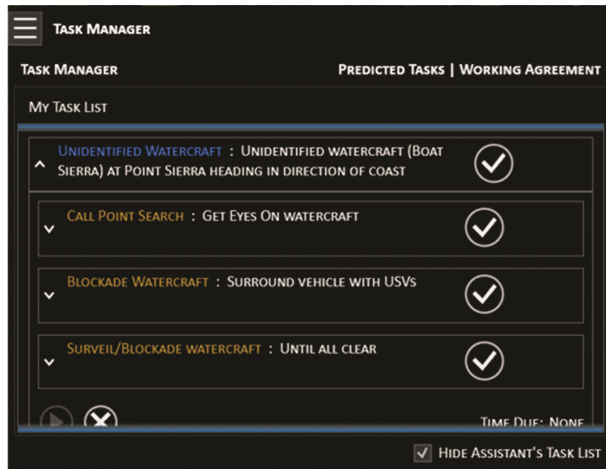


Fig. 9. Illustration of task manager that provided a mechanism for calling plays with mouse and touch input modalities

Experiment 2 also featured added interfaces that provided operators with further insight into the rationale of the plans generated by the autonomy for plays as well as the status of plays under development. These interfaces primarily presented information to the operator to monitor play calling and execution and are of less interest to this input modality-focused discussion. (For further details, see [2, 18, 24].)

In Experiment 2, the IMPACT prototype was compared to a baseline system (representing current state-of-the-art [25] that did not have play-based interfaces, including speech control. Four 60-minute experimental trials were conducted, two trials with each

system with the trials varying in mission complexity (number and timing of mission-related tasks). Trials were blocked by system and counterbalanced. Only data from the two trials conducted with the IMPACT system were relevant to exploring operators' modality choice when making inputs into the play-based interfaces.

4.1 Method

Participants. Eight volunteers with relevant military experience participated in this study, four active duty and four who had previously served. Six operators had prior experience piloting air UVs (Global Hawk, Predator, Reaper, Scan Eagle, Raven), one operator was a former Predator/Reaper SO, and one operator was an experienced security force and base defense expert. Seven operators were male (one female) and all operators reported normal or corrected-to-normal vision, normal color vision, and normal hearing. Operators' mean age was 43.6 years ($SD = 10.84$).

Equipment. The experimental configuration was expanded to four stations: the C2 Operator Station, the Sensor Operator Station, the Test Operator Console, and the Simulation Station. The Simulation Station used a Dell Precision T5400 and ran One Semi-automated Forces (OneSAF), a simulation tool that generated all friendly, neutral, unknown, and hostile forces during the experiment, with the exception of the UVs. The C2 Operator Station and Test Operator Console each used a Dell Precision T7910 while the Sensor Operator Station used a Dell Precision T5600. The C2 Operator Station, Sensor Operator Station, and Test Operator Console had identical monitor setups, with three Acer T272HUL LED touchscreens (2560×1440) and one (lower center) Sharp PN-K322B 4 K Ultra-HD LCD touchscreen (usable touchscreen area: 69.79×39.26 cm; 3840×2160 resolution; tilted 42° from horizontal). Three Dell Precision R7610 located in a different room provided the sensor feeds for the UVs (four feeds per machine).

Procedure. Operators received briefings and training, as described in Sect. 2.3, as well as training on the play-based interfaces added in Experiment 2. Practice trials were also conducted to provide operators familiarity on how mission events would be prompted (chat window and over the headset) and how to respond in trials featuring IMPACT play-based interfaces versus the baseline system (the latter not described here). In addition to how to respond to mission events, operators were familiarized with how to accomplish anti-terrorism measures assigned at the beginning of each trial (e.g., image four sides ("360") of a certain building at a set interval (accomplished by calling point inspect plays in a timely manner)). There were also other base defense tasks added in Experiment 2 (such as to provide a ground vehicle scout ahead coverage with an air UV), in addition to queries issued by the commander through chat (e.g., "How long would it take to get a Show of Force at Gate 3 in place?"). For most queries, it was more efficient for operators to issue speech commands to the autonomy, asking for relevant information to answer the query. Training to address all these tasks took an entire day to accomplish. On a second day, after refresher training, the four experimental trials were conducted and several debriefing questionnaires were administered.

4.2 Results

Operators' performance was better on multiple mission performance metrics with the IMPACT system as compared to the baseline system. For instance, operators were able to execute plays using significantly fewer mouse clicks with IMPACT compared to baseline. Operators also rated IMPACT higher on usability than the baseline. Detailed results will be published elsewhere [24]. Here, results will focus on operators' use of the three input modalities (mouse/click, touch, and speech).

In that Experiment 2's focus was on evaluating the IMPACT and baseline systems, operators were not asked to compare the modalities in questionnaires. Figure 10 illustrates this point by showing data from three Likert Scales addressing IMPACT features with respect to: how easy to use, how quick to learn, and potential value for future multi-UV operations. (Likert scales were five-point, ranging from 'Strongly Disagree' to 'Strongly Agree.') The administered questionnaire specifically had items addressing touch and speech input modalities. However, there was not a specific scale for mouse input. In Fig. 10, data for the mouse input modality are estimated from the scale for the Play Calling interface (Fig. 7a), as the experimenter observed that none of the operators used touch to interact with this interface. It cannot be determined, though, if the operators' ratings reflect the mouse modality and/or other features of the play calling interface (e.g., arrangement of play icons in the rows).

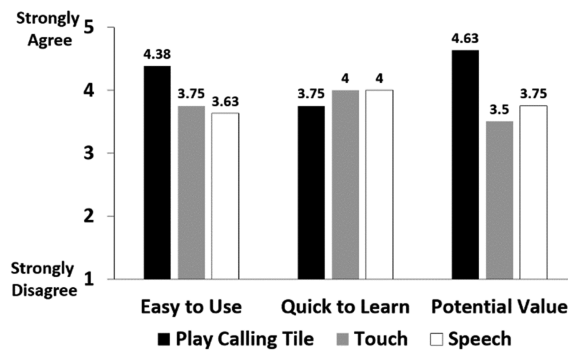


Fig. 10. Ratings related to each input modality

Four operators reported that there were difficulties making inputs with touch (e.g., citing lack of confidence and its sensitivity to "fat fingers"). One operator reported that after employing the mouse input, it didn't make sense to change to another modality. Comments pertaining to the speech input (from three operators) mentioned its unreliability and restrictive syntax. The word error rate was 23.38% for Experiment 2 (for in-grammar utterances). This error rate was only slightly better (~1.5%) from that achieved in Experiment 1, despite increases in both vocabulary and phrases. The word error rate across all utterances, both in-grammar and out-of-grammar, was 34.26%.

The degree to which operators did not employ touch and speech input is further illustrated in Fig. 11 that shows the number of plays called with each type of play interface, as well as the modality employed (given all interactions with the radial menu, play

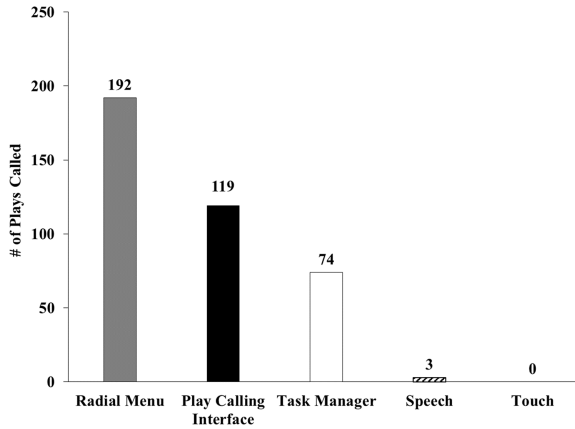


Fig. 11. Number of plays called with each play-based interface and input modality type

tile, and task manager were made with the mouse). Of the 388 play calls made during Experiment 2, only three used speech commands and none were made using touch input. These data also suggest that operators did not switch between modalities in completing steps in the play calling process. In fact, the tendency to use the mouse to confirm the speech input reported in Experiment 1 was not observed in Experiment 2. Rather, two operators who employed speech failed to make the confirmation response immediately after calling the play verbally (leaving the Play Workbook open for that play while calling other plays and then eventually closing the workbook for the verbally-called play), escalating the play calling completion time (mean = 14.23 min).

Results depicted in Fig. 12 show that every operator primarily used the mouse input modality. Also, seven of the eight operators employed the radial menu. However, the results also suggest the value of providing a variety of interfaces for calling plays as

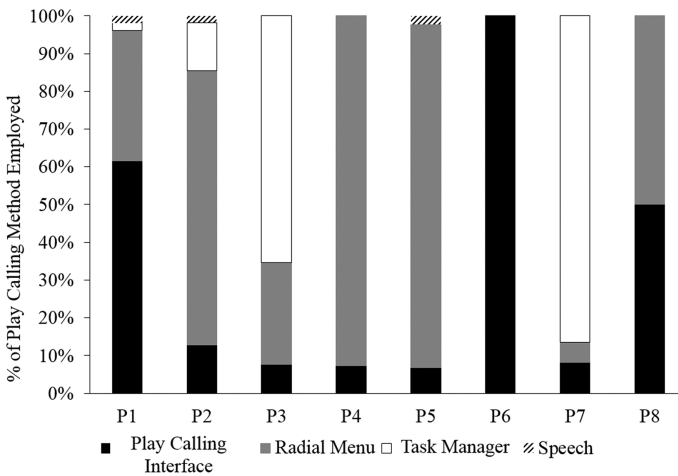


Fig. 12. Percentage of plays called with each play-based interface

three operators used the task manager to call plays (the primary mechanism for two of the operators) and three other operators primarily used the play-calling interface.

5 Discussion

Subjective data from the two experiments indicated that the mouse input modality was better than both touch and speech modalities for play-based management of multiple UVs. Operators preferred the mouse input modality for calling plays, commenting that the mouse was more intuitive for exercising the play calling interfaces. In both experiments, this preference was also evident in the frequency operators chose to employ the mouse for calling plays, rather than touch or speech. Use of mouse inputs proved to be an efficient input modality in the IMPACT simulation.

The other two modalities for play calling were problematic: operators cited the speech recognition rate and the touchscreen's lack of precision. While modifications were made in the implementation of these modalities after Experiment 1, these changes may have actually limited their utility. For instance, it is possible that more speech command training was needed to reap the benefits of expanding the speech model and vocabulary. (Note: a concern for vocabulary training requirements was raised in earlier play-related research [13].) It may also be that use of speech input is simply not ideal for play calling. As Cohen and Oviatt [6] explain, speech-based control is most ideal when the operator's hands and/or eyes are busy or when there is limited screen real estate to exercise control. With the tasking and mission employed in IMPACT, operators were able to devote attention to the play-based interfaces and employ mouse inputs efficiently.

With respect to the touch input modality, the changes to the monitor's size and tilt in Experiment 2 may have complicated touch entry by increasing the reach envelope even more, exceeding distances recommended in [26]. Touch input is more useful with smaller reach distances and when inputs are not frequent [28]. Given the missions were designed to require frequent play calling, it is logical that small manipulations of the mouse to position the cursor on various play related interfaces was more effective than reaching to touch interfaces or map locations on a large monitor. Over the course of hour-long missions (Experiment 2), issues of fatigue noted in other examinations of touch input (e.g., [28]) would have likely been observed. Operators may have also been hesitant to employ touch due to arm/hand movement occluding map/play symbology on the monitor.

These research results suggest that providing multiple input modalities for exercising IMPACT's play interfaces was not advantageous to the operators in these experiments. It could be viewed that these results suggest input modality should instead be optimized for specific tasks, rather than expending effort to enable multiple input modalities for every task type. However, this does not mean that only a single modality should be implemented for each task type. For instance, the present results indicate that the radial menu play calling interface on the map was most frequently employed, compared to the dedicated play calling interface, task manager, or utilizing speech control (Fig. 11). If direct interactions with the map (designating a location or UV) are ideal for calling a play, perhaps an integration of sketch and speech map inputs would be useful [29],

enabling the operator to draw on the map with a companion speech command to call a play (“loiter here for 10 min) or specify a play detail (“ingress here”). Continued use of the mouse for sketching is likely more convenient compared to switching modality to touch. However, a mechanism (e.g., button press or speech commands) would be needed to signal the start and end of the sketch input to differentiate it from other mouse inputs.

In addition to multiple modalities available for play calling, the operators were also free to choose modality for responding to query prompts. Despite the poor speech recognition rate, operators’ chose the speech modality to acquire most of the information needed to address these prompts. Additionally, several operators suggested that they found it useful to query the autonomy in this manner. It is recommended that this capability be expanded, along with improvements to the speech recognition system, to better support collaboration and joint problem solving between the human operator and autonomy.

Acknowledgments. This research supports the ASD/R&E Autonomy Research Pilot Initiative “Realizing Autonomy via Intelligent Adaptive Hybrid Control.”

References

1. Draper, M., Calhoun, G., Ruff, H., Frost, E., Evans, D., Hansen, M., Douglass, S., Spriggs, S., Bearden, G., Howard, M., Behymer, K., Patzek, M., Rowe, A.: Intelligent multi-unmanned vehicle planner with adaptive collaborative/control technologies. In: International Symposium on Aviation Psychology (2017)
2. Calhoun, G., Ruff, H.A., Behymer, K.J., Mersch, E.M.: Operator-autonomy teaming interfaces to support multi-unmanned vehicle missions. In: Savage-Knepshield, P., Chen, J. (eds.) *Advances in Human Factors in Robots and Unmanned Systems. Advances in Intelligent Systems and Computing*, vol. 499, pp. 113–126. Springer International Publishing, Switzerland (2017)
3. Oviatt, S.L., Olsen, E.: Integration themes in multimodal human-computer interaction. In: Shirai, K., Furui, S., Kakehi, K. (eds.) *Proceedings of the International Conference on Spoken Language Processing*, vol. 2, pp. 551–554. Acoustical Society of Japan (1994)
4. Oviatt, S.: Ten myths of multimodal interaction. *Commun. ACM* **42**(11), 74–81 (1999)
5. Turk, M.: Multimodal interaction: a review. *Pattern Recogn. Lett.* **36**, 189–195 (2014)
6. Cohen, P.R., Oviatt, S.L.: The role of voice input for human-machine communication. *Proc. Nat. Acad. Sci.* **92**, 9921–9927 (1995)
7. Durlach, P.J., Neumann, J.L., Bowens, L.D.: Evaluation of a touch screen-based operator control interface for training and remote operation of a simulated micro-uninhabited aerial vehicle. In: Cooke, N., Pringle, H., Pedersen, H., Connor, O. (eds.) *Human Factors of Remotely Operated Vehicles*, pp. 165–178. Elsevier, NY (2006)
8. Williamson, D.T., Draper, M.H., Calhoun, G.L., Barry, T.P.: Commercial speech recognition technology in the military domain: results of two recent research efforts. *Int. J. Speech Technol.* **8**(1), 9–16 (2005)
9. Calhoun, G.L., Draper, M.H.: Multi-sensory interfaces for remotely operated vehicles. In: Cooke, N., Pringle, H., Pedersen, H., Connor, O. (eds.) *Human Factors of Remotely Operated Vehicles*, pp. 149–163. Elsevier, NY (2006)

10. Chun, W.H., Spura, T., Alvidrez, F.C., Stiles, R.J.: Spatial dialog and unmanned aerial vehicles. In: Cooke, N., Pringle, H., Pedersen, H., Connor, O. (eds.) *Human Factors of Remotely Operated Vehicles*, pp. 193–208. Elsevier, NY (2006)
11. Levulis, S.J., Kim, S.Y., DeLucia, P.R.: Effects of touch, voice, and multimodal input on multiple-UAV monitoring during simulated manned-unmanned teaming in a military helicopter. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, p. 132 (2016)
12. Calhoun, G., Draper, M.H., Ruff, H., Barry, T., Miller, C.A., Hamell, J.: Future unmanned aerial systems control: feedback on a highly flexible operator-automation delegation interface concept. In: *AIAA Infotech @ Aerospace Conference*, AIAA-2012-2549, pp. 1–16 (2012)
13. Calhoun, G., Ruff, H., Miller, C., Murray, C., Hamell, J., Barry, T., Draper, M.: Flexible levels of execution-interface technologies (FLEX-IT) for future remotely piloted aircraft control applications. Technical report, AFRL-RH-WP-TR-2012-0077. Wright-Patterson Air Force Base, OH (2012)
14. Shively, J., Flaherty, S., Miller, C., Fern, L., Neiswander, G.: Delegation control in control of unmanned aerial systems (UAS). In: *Infotech @ Aerospace Conference*, AIAA-2012-2458 (2012)
15. Draper, M.H., Miller, C.A., Calhoun, G.L., Ruff, H., Hamell, J., Benton, J., Barry, T.: Multi-unmanned aerial vehicle systems control via flexible levels of interaction: an adaptable operator-automation interface concept demonstration. In: *AIAA Infotech @ Aerospace Conference*, AIAA-2013-4803 (2013)
16. Miller, C.A., Draper, M., Hamell, J.D., Calhoun, G., Barry, T., Ruff, H.: Enabling Dynamic Delegation Interactions with Multiple Unmanned Vehicles; Flexibility from Top to Bottom. In: Harris, D. (ed.) *EPCE 2013. LNCS (LNAI)*, vol. 8020, pp. 282–291. Springer, Heidelberg (2013). doi:[10.1007/978-3-642-39354-9_31](https://doi.org/10.1007/978-3-642-39354-9_31)
17. Calhoun, G., Draper, M., Miller, C., Ruff, H., Breeden, C., Hamell, J.: Adaptable automation interface for multi-unmanned aerial systems control: preliminary usability evaluation. *Proc. Hum. Factors Ergon. Soc.* **57**(1), 26–30 (2013)
18. Calhoun, G., Ruff, H., Behymer, K., Frost, M.: Human-autonomy teaming interface design considerations for multi-unmanned vehicle control. *Theor. Issues Ergon. Sci. Special Issue: Human-Autonomy Teaming* (2017, in press)
19. Rowe, A., Spriggs, S., Hooper, D.: Fusion: a framework for human interaction with flexible-adaptive automation across multiple unmanned systems. In: *International Symposium on Aviation Psychology* (2015)
20. Behymer, K., Rothwell, C., Ruff, H., Patzek, M., Calhoun, G., Draper, M., Douglass, S., Kingston, D., Lange, D.: Initial evaluation of the intelligent multi-UxV planner with adaptive collaborative/control technologies (IMPACT). Technical report, AFRL-RH-WP-TR-2017-0011. Wright-Patterson Air Force Base, OH (2017, in press)
21. Bennett, K.B., Flach, J.M.: *Display and Interface Design: Subtle Science*, Exact Art. CRC Press, Boca Raton (2011)
22. Shneiderman, B.: *Designing the User Interface: Strategies for Effective Human-Computer Interactions*, 2nd edn. Addison-Wesley, Reading (1992)
23. Huggins-Daines, D., Kumar, M., Chan, A., Black, A.W., Ravishankar, M., Rudnick, A.I.: Pocketsphinx: a free, real-time continuous speech recognition system for hand-held devices. In: *IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings (ICASSP)*, vol. 1, pp. I-185–I-188 (2006)

24. Behymer, K., Rothwell, C., Ruff, H., Patzek, M., Calhoun, G., Draper, M., Douglass, S., Kingston, D., Lange, D.: Second evaluation of the intelligent multi-UxV planner with adaptive collaborative/control technologies (IMPACT). Technical report, AFRL-RH-WP-TR-2017-TBD. Wright-Patterson Air Force Base, OH (2017, in press)
25. Feitshans, G.L., Davis, J.E.: Advanced net-centric combat unmanned air vehicle. Technical report, AFRL-RG-WP-TR-2011-0126, Wright-Patterson Air Force Base, OH (2011)
26. Sears, A., Shneiderman, B.: High precision touchscreens: design strategies and comparisons with a mouse. *Int. J. Man Mach. Stud.* **34**(4), 593–613 (1991)
27. United States Department of Defense: MIL-HDBK-759C: Handbook for Human Engineering Design Guidelines (1995)
28. Shin, G., Zhu, X.: Ergonomic issues associated with the use of touchscreen desktop PC. *Proc. Hum. Factors Ergon. Soc.* **55**(1), 949–953 (2011)
29. Taylor, G.B., Purman, P., Schermerhorn, G., Garcia-Sampedro, R., Hubal, K., Crabtree, A., Rowe, A., Spriggs, S.: Multi-modal interaction for UAS control. In: *SPIE defense + security*, p. 946802. International Society for Optics and Photonics (2015)

Engineering Psychology and Cognitive Ergonomics:
Cognition and Design
14th International Conference, EPCE 2017, Held as Part
of HCI International 2017, Vancouver, BC, Canada, July
9-14, 2017, Proceedings, Part II
Harris, D. (Ed.)
2017, XXII, 439 p. 147 illus., Softcover
ISBN: 978-3-319-58474-4