The Impact of Different Levels of Autonomy and Training on Operators' Drone Control Strategies

JIN ZHOU, Duke University, USA
HAIBEI ZHU, Duke University, USA
MINWOO KIM, Duke University, USA
M. L. CUMMINGS, Duke University, USA

Unmanned Aerial Vehicles (UAVs), also known as drones, have extensive applications in civilian rescue and military surveillance realms. A common drone control scheme among such applications is human supervisory control, in which human operators remotely navigate drones and direct them to conduct high-level tasks. However, different levels of autonomy in the control system and different operator training processes may affect operators' performance in task success rate and efficiency. An experiment was designed and conducted to investigate such potential impacts. The results showed us that a dedicated supervisory drone control interface tended towards increased operator successful task completion as compared to an enhanced teleoperation control interface, although this difference was not statistically significant. In addition, using Hidden Markov Models, operator behavior models were developed to further study the impact of operators' drone control strategies as a function of differing levels of autonomy. These models revealed that people with both supervisory and enhanced teleoperation control training were not able to determine the right control action at the right time to the same degree that people with just training in the supervisory control mode. Future work is needed to determine how trust plays a role in such settings.

CCS Concepts: • **Human-centered computing** *HCI theory, concepts and models*; *Empirical studies in HCI*;

Additional Key Words and Phrases: Levels of Autonomy, Skill-Based Training, Drone, Unmanned Aerial Vehicle, Hidden Markov Model, Supervisory Control

ACM Reference Format:
Jin Zhou, Haibei Zhu, Minwoo Kim, and M. L. Cummings. 2018. The Impact of Different Levels of Autonomy and Training on Operators' Drone Control Strategies. *ACM Trans. Hum.-Robot Interact.* 1, 1 (August 2018), 15 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The commercial market of unmanned aerial vehicles, known as UAVs or drones, has significantly increased because of their various sizes, feasibility of remote control and their ability to conduct high-level tasks [13]. According to the Federal Aviation Administration (FAA), the population of civilian drone systems is expected to reach 7 million in 2020 [21]. Considering human operators play an important role in such drone control systems, investigating the interaction between human

Authors' addresses: Jin Zhou, Duke University, 304 Research Rd, Durham, NC, 27708, USA, jin.zhou@duke.edu; Haibei Zhu, Duke University, 304 Research Rd, Durham, NC, 27708, USA, haibei.zhu@duke.edu; Minwoo Kim, Duke University, 304 Research Rd, Durham, NC, 27708, USA, minwoo.kim@duke.edu; M. L. Cummings, Duke University, 304 Research Rd, Durham, NC, 27708, USA, mary.cummings@duke.edu.

2573-9522/2018/8-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

operators and drone control systems benefits drone operators' performance and the future design of such human-in-the-loop systems.

With technical developments in both software and hardware, more highly automated systems are available to Unmanned Aerial System (UAS) designers. However, more automation does not necessarily bring better performance. It is important for system designers to understand how to balance the degree of autonomy with the degree of desired control, keeping in mind human control and cognitive limitations. The design space for such systems includes understanding the role of mental workload, automation reliability, and the system's expected costs and benefits from possible choices of decision/action implementations [14, 17].

One important application area that needs more principled design considerations is the use of drones for first-person inspection tasks including inspection of roofs [24], power lines [25], and other major infrastructure like oil pipelines [8]. Such tasks require an operator to navigate to the site of interest, then determine where the object of interest is at this site, and then move from global, high-level control of the drone to more low-level control of the camera on the drone. As a result, any ground control station that supports these tasks must provide such functionality. Some commercial-off-the-shelf systems do this through enhanced teleoperation, which is the lowest form of supervisory control [22]. For these systems, the automated flight control system provides some altitude, speed and altitude stability, but operators still have the ability to rate command heading changes as well as altitude and speed changes through throttle and joystick manipulations [16].

Other more advanced drone systems can fly through waypoint commands, a higher level of supervisory control. In this higher autonomy system, a three-dimensional point in space is commanded by the operator and the vehicle automatically determines what flight control changes are needed to achieve this point in space. Waypoint-directed supervisory control reduces operator workload but also does not allow for the same flexibility and agility in maneuvering an aircraft [7]. Because of the need to have a vehicle achieve a very specific position in inspection tasks, the enhanced teleoperation mode is generally preferred for the inspection task, although some manufacturers include both the low and high levels of autonomy (enhanced teleoperation vs. waypoint control).

Given that drone operators can have access to a variety of autonomy modes, as well as different levels of training in learning to operate such systems, an experiment was conducted in a drone inspection scenario. Additionally, operators' strategies in interacting with the remote drone control systems were modeled utilizing Hidden Markov Models. The goal of these experiments was to determine whether the mode of autonomous control was resilient to degree of training, and what operator error modes would be encountered in both enhanced teleoperation and waypoint control.

2 EXPERIMENT

Two different drone control interfaces were designed for experimental devices based on two control modes: enhanced teleoperation control (ETC, a lower level of automation) and the waypoint supervisory control (WSC, a higher level of automation) modes. Figure 1 shows the interface for the ETC mode, which includes five main components: a power button, a status bar, joysticks, a map window and a camera view. Specifically, operators are able to obtain the environmental information surrounding a drone through a live camera stream. Operators gain global situation awareness by checking the real-time drone position on the mini-map. The ETC interface is a virtual depiction of typical drone ground control stations that resemble gaming consoles. In the ETC interface, operators use two joystick simulators to navigate drones. In the left joystick, the middle bar changes altitude, and two buttons marked with arrows cause the drone's rotation. The right joystick is responsible for horizontal movement.

Figure 2 shows the interfaces for the WSC mode, which contains two interface, a Supervisory ontrol interface (Figure 2a) and an Inspection Model interface (Figure 2b). The Supervisory Control

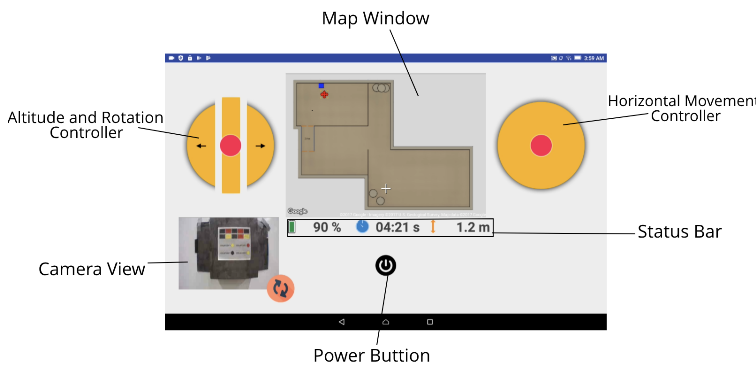


Fig. 1. ETC Interface.

interface is the main navigation mode and has four main components: a status bar, a map, a control panel, and a mode switch button. With such an interface, operators can set waypoints to build overall flight plans through a touchscreen interface and then execute the customized flight plan represented by the waypoints. Operators switch to the Inspection Mode when they want to see the world immediately surrounding the drone. This mode is similar but not the same as ETC as it only allows for nudge control [16], which only allows operators to make very small positional changes in altitude and heading. In ETC, operators can directly influence the rate of change which provides for greater control but also an increased risk of losing control. In WSC, operators check their global position by reverting to the Supervisory Control display as opposed to consulting the mini-map as in ETC.

2.1 Experiment Subjects and Procedure

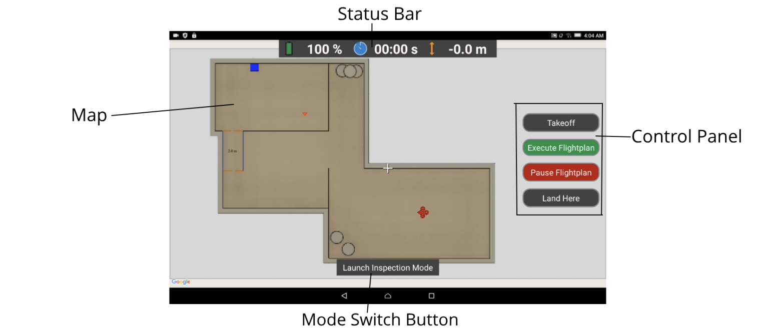
Participants were randomly assigned to 3 training groups. Participants in Group 1 were trained using ETC control, and participants in Group 3 were trained using WSC, while participants in Group 2 were trained using both ETC and WSC, as shown in Table 1. These training programs will be discussed more in the following section.

It is important to note that while participants in Group 2 were trained on both interfaces, they only used the WSC interface in the final test. Previous research has shown that actual Air Force UAV operators perform better when they have both versions of ETC and WSC training as opposed to just WSC training [23]. Thus, our expectation was that Group 2 would perform at least as well as the WSC-only participants.

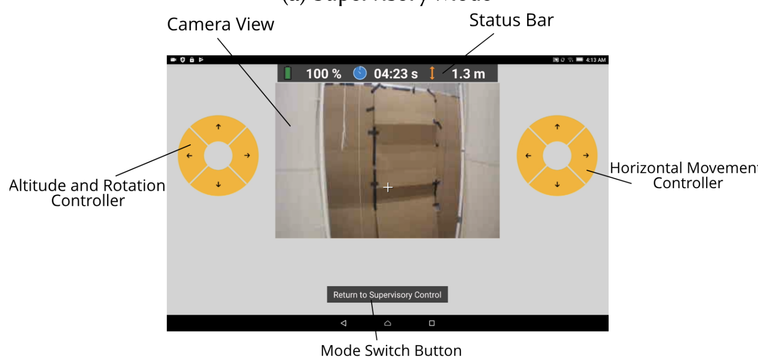
Table 1. Experimental treatments

	Group 1	Group 2	Group 3
Training Program	ETC	ETC + WSC	WSC
Estimated Training Time (mins)	100	120	40
Testing Interface	ETC	WSC	WSC

In each test session, a participant controlled one drone to conduct a navigational task. Before beginning the test mission, participants were first briefed about the assumed task scenario of the test mission, in which they needed to control a UAV into a building which was contaminated by



(a) Supervisory Mode



(b) Inspection Mode

Fig. 2. WSC Interfaces.

radioactive material from reactors due to an earthquake. They needed to reach a control panel to read key information on the status of the reactor, then safely fly back to the takeoff location for recovery of the vehicle. All participants' operations via the interfaces were recorded for data analysis. After test sessions, participants were debriefed about their performance by the experimenter and were paid based on the estimated average run time, which ranged from $25 to $50.

In the experiment, A tablet computer (Lenovo Tab 2 A10-70, 1920 × 1080 pixels) was used as hand-held device for both interfaces. The type of the UAV used in the experiment was a Parrot AR 2.0, and open-source software Paparazzi was used to customize the drone's settings. To track the drone's location, 26 Vicon Vero cameras and 4 Vicon Vantage cameras (for better coverage than Vero cameras to open areas) were installed in the experiment environment as replacements for GPS.

2.2 Training Programs and Test Sessions

Each training program consisted of six basic training modules with a checkride module. The first basic training module briefly explained basic concepts and knowledge about drones. Module 2 introduced the drone control interface, which was different according to which group a participant belonged. Module 3 explained how to take off and land a drone. Module 4 taught participants how to navigate a drone. Module 5 explained how to control the camera installed on the drone. Module 6 provided general advice and tips on possible emergency handling in situations that could occur

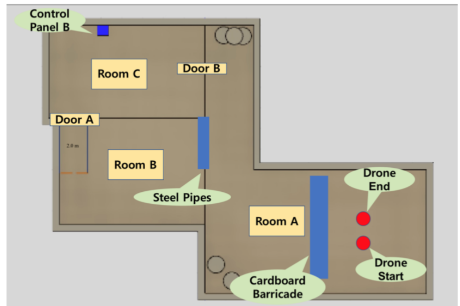


Fig. 3. Map for Environment Settings.

during operation. The final checkride module was designed to train participants in preparation for avoiding various obstacles that they might encounter during the test sessions.

Each of the 6 modules contained a self-paced learning part, where participants were given tutorial slides to study. All slides can be found at http://hal.pratt.duke.edu/training-modules. In addition, modules 3-5 included hands-on practice, where participants were given opportunities to practice flying a drone, based on the instructions from the tutorial slides. Different experimental groups were provided different training programs as illustrated in Table 1, and are detailed further in [10].

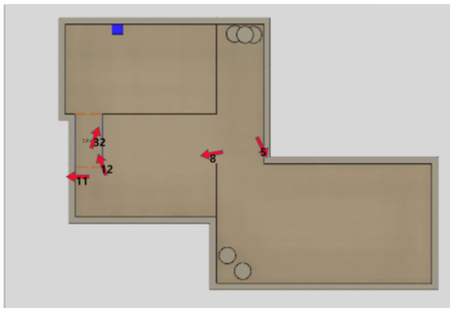
As mentioned previously, the final objective task in the test scenario was to reach and read a control panel in a partially-collapsed room in a building damaged by an earthquake. The map of the test environment is shown in Figure 3. The starting point of the drones in test sessions was in room A. The trajectory of the drone passed through room B and then through a narrow ventilation shaft. This shaft opened into room C, which had a control panel on a wall to be read by the operator via the drone camera. After this primary task, operators were to navigate the drone back to the starting point.

As is typical in actual recovery drone operations, the map in Figure 3 reflected the world before the earthquake, and operators were warned that the actual world may not match the world after the earthquake. To this end, we added a barricade and steel pipes as static obstacles as seen in Figure 3, however, these obstacles were not shown on any drone control interface. In addition, one dynamic change occurred during the experiment. After each participant read the control panel in room C, an explosion sound occurred, and the shaft that formed the entryway was sealed shut by a falling fake wall. At the same time, another "wall" on the other side of the room fell, allowing the participant to escape through a newly-formed gap labeled door B in Figure 3. This scenario reflects unexpected events that require operators to improvise and adapt under uncertain environment conditions. All experimental material and pictures of the training and testing rooms are available at http://hal.pratt.duke.edu/drone-piloting- research.

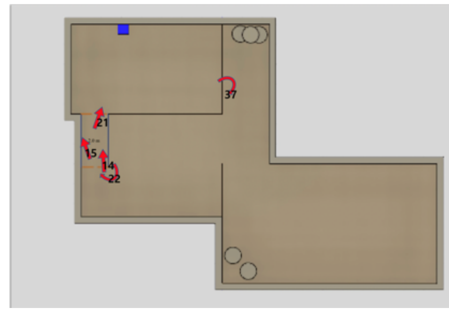
3 EXPERIMENT RESULTS

In the experiment, 4 out of 38 subjects were excluded from analysis for either poor quality of data or failure in data retrieval. All 34 valid subjects' age range was from 21 to 41, with an average of 25.53 and a standard deviation of 3.70. Out of 34 participants, 27 were males and 7 were females. In addition, due to some external factors like battery depletion, only 26 subjects completed the test. For all statistical analysis, a significance level of $\alpha = 0.05$ was used.

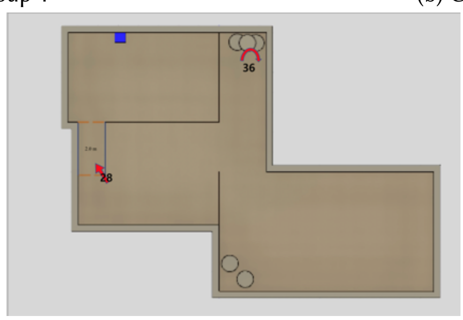
Table 2 shows the number of pass and fail attempts for each group. Participants in Group 3 presented the highest success rate of $6/8 = 0.75$, followed by participants in Group 1 ($5/10 = 0.5$) and Group 2 ($3/8 = 0.38$). However, likely due to the small sample size, the generalized Fisher exact



(a) Group 1



(b) Group 2



(c) Group 3

Fig. 4. Crash Points.

(Fisher-Freeman-Halton) test did not indicate statistical significance between the pass/fail ratio between groups ($p = 0.302$). It is worth noting that those with only supervisory control were more than twice as successful that those with both manual and supervisory control training.

Table 2. Pass/Fail

	Group 1	Group 2	Group 3
Pass	5	3	6
Fail	5	5	2
Total (Complete)	10	8	8
Success Rate	50%	38%	75%

In terms of where people crashed drones, for the 12 crashes, 8 crashes occurred near the shaft as shown in Figure 4, which indicates drones' movements at the moment of crashing with red arrows. By examining the screen recording videos from those participants, all 8 participants had a hard time aligning their drones to the 40 inches-wide entrance of the shaft because of the delay in video stream and wobble caused by reflected winds from the rotors.

Table 3 illustrates how much time participants spent in Inspection Mode through the three major phases of the experiment. Inspection Mode allowed participants a maximum view of the first-person camera, which allowed them to best assess the actual environment. It also then allowed participants to access the Nudge Control function which would allow people to make small but safe

Table 3. Average Percentage of time spent on Inspection Mode (% (sample size))

	Group2	Group3
Start to Shaft	56.21 (9)	71.50 (9)
Shaft Passage	70.50 (6)	91.75 (8)
Collapse To Finish	76.13 (4)	88.01 (6)

position adjustments to the vehicle. Time in Inspection Mode means that people were definitely not engaging the higher-level Supervisory Control mode that controlled the vehicles through waypoints.

When looking at Figure 4 and Table 3, it is clear that Group 2 did not accurately assess the risk of maneuvering through the ventilation shaft, choosing to spend much less time in Inspection Mode than those in Group 3. Not surprisingly, operators in Groups 2 and 3 tended to spend more time in Inspection Mode after the explosion, which represents an important state change with an increase in uncertainty. While it was expected that operators would spend more time in Inspection Mode after the explosion, one question this result raised is why those people with no training in ETC and little training in Nudge Control preferred overwhelmingly to use Nudge Control for basic navigation? This will be addressed in a subsequent section.

For those people who successfully finished the course, Table 4 shows the summary statistics of completion times. Similar to the success rate result, while participants in Group 3 had the lowest course completion times, the one-way ANOVA test revealed no significant difference ($p = 0.744$).

Table 4. Mission Completion Time

	Group 1	Group2	Group3
Average	7:49	7:21	6:58
Minimum	5:22	5:23	5:27
Median	8:13	6:42	7:10
Maximum	10:05	10:00	8:06
Standard Deviation	2:10	2:22	1:05
Sample Size	5	3	6

The differences in performances between the groups were expected since WSC has already been shown in similar environments to lead to improved task times and fewer crashes [7]. However, we expected that the WSC group would do significantly better than the ETC group who only lagged the WSC group slightly in number of crashes and completion times. Moreover, because of the previous Air Force UAV operator research that showed mixed training to be advantageous [23], it was surprising that Group 2 people did not perform better.

Given that the statistical results did not provide enough insight into the underlying nature of why additional training didn't lead to better performance, additional information was needed about how people went about their tasks of controlling the drones. To this end, human operator models were needed for further investigation of operators' behavior patterns and strategies in such drone control scenarios.

4 OPERATOR BEHAVIOR MODELLING THROUGH HIDDEN MARKOV MODELS

Human operator behavior models are important tools for behavior analysis in unmanned aerial system (UAS) settings. Through these models, we can determine if observed behaviors match expected behaviors, investigate operators' strategies to identify points of inefficiency or error, study both endogenous and exogenous factors that impact operator behavior patterns, and study how automation can improve operators' performance and success rate in task performance [3, 4]. Models for operators undergoing different training can help us learn the effect of training factors, including how different technologies influence people's ability to master skills.

In general, existing computational modeling techniques can be divided into 3 categories: symbolic models, architecture-based models and statistical models [2]. The first two models tend to be deductive, and their use of a priori definition of rules or cognitive processes make them unsuitable for behavior analysis because of the complexity of the decisions and uncertainty in the environment [3]. By contrast, statistical models tend to be inductive and data-driven. In UAS settings, data is generated by human interaction with the ground control station, which then forms behavioral models.

One popular statistical learning tool, a Hidden Markov Model (HMM) is a two-layer stochastic model used for model generative sequences characterized by a set of observable sequences [20]. HMMs have been used extensively in the past in speech recognition [18], stock market prediction [6], and modeling driver behavior [15]. HMMs are based on a set of unobserved underlying states amongst which transitions can occur and each state is associated with a set of possible observations [9]. Human operator behaviors share similar properties. Hidden states represent higher-level operator information processing states, while transitions between these underlying states show operators' strategies in navigating the drone. Because of this ability to connect observed behavior data to an underlying latent process, generally interpreted as the cognitive patterns [11], the HMM approach was selected as the modeling framework for this effort.

4.1 HMM Structure

The Hidden Markov model usually has two layers, including a directly observable layer, known as observations or emissions, and a hidden layer, which contains certain number of hidden states. Each hidden state in the hidden layer represents a special combination of observations with different proportions. The HMM can be formally defined as a tuple [19]:

$$H = \{S, V, A, B\}$$

In this notation, $S = \{S_1, S_2, ..., S_N\}$ represents N different hidden states, $V = \{V_1, V_2, ..., V_M\}$ represents M different observations. Connection probabilities are also important elements in the notation, here $A = \{a_{ij} \geq 0\}$ is a $N \times N$ transition probability matrix, where $a_{ij} = P\{S_j^{t+1}|S_i^t\}$, $i, j = 1, 2, ..., N$, and $B = \{b_{ik} \geq 0\}$ is a $N \times M$ emission probability matrix, where $b_{ik} = P\{V_k|S_i\}$, $i = 1, 2, ..., N$, $k = 1, 2, ..., M$.

For developing the operator HMMs, training data were collected from the experiment sessions introduced in the previous section. An example of how data was matched from an interface to an observable stated, Table 5 depicts 15 ETC operations. Table 6 depicts 15 WSC operations, observations 1-7 are from the supervisory mode and 8-15 are from the inspection mode.

4.2 Model Training and Selection

An unsupervised Hidden Markov Model training method of multi-sequence Baum-Welch algorithm was used in model training [1]. However, before finalizing our final HMM models, we need to select the number of hidden states, which is represented by K. Given 15 observations for both ETC and WSC, the potential range for K is $2 - 15$. For HMM model training, the emission probabilities were

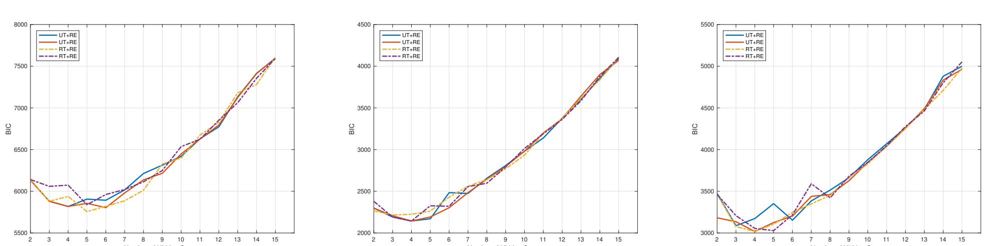
Table 5. Observations (Emissions) of HMM from ETC

Index	1	2	3	4	5
Observation	Takeoff	Land Here	View Switch	Rise	Descent
Index	6	7	8	9	10
Observation	Move Forward	Move Backward	Move Left	Move Right	Move Forward Left
Index	11	12	13	14	15
Observation	Move Forward Right	Move Backward Left	Move Backward Right	Rotate Left	Rotate Right

Table 6. Observations (Emissions) of HMM from WSC

Index	1	2	3	4	5
Observation	Takeoff	Land Here	Execute Flight plan	Add Waypoint	Adjust Waypoint
Index	6	7	8	9	10
Observation	Launch Inspection	Return Supervisory	Rise	Descent	Move Forward
Index	11	12	13	14	15
Observation	Move Backward	Move Left	Move Right	Rotate Left	Rotate Right

initialized with random assignments, and transition probabilities were initialized with uniform and random assignments.



(a) Group 1 (b) Group 2 (c) Group 3

Fig. 5. BIC curves of HMM modes for all 3 groups. UT means initialize transition probabilities with uniform distribution, RT means initialize transition probabilities with random distribution, RE means initialize emission probabilities with random distribution.

The Bayesian information criterion (BIC) was used for picking the most appropriate number of hidden states. Models with the lower BIC values are preferred [18]. Figure 5 shows BIC curves for all 3 groups. Based on such BIC curves, $K = 5$ is the most appropriate hidden states number for the

Group 1 model, which balances between the model complexity and the model likelihood. Similarly, for both Group 2 and Group 3 models, $K = 4$.

4.3 Operator Behavior Models

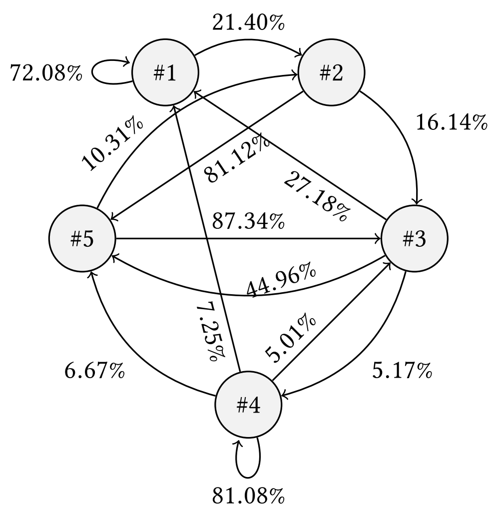
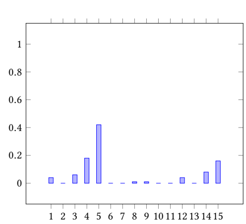
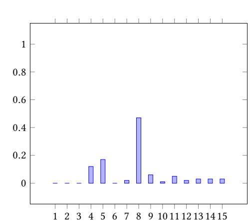


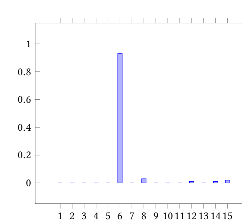
Fig. 6. Human operator behavior HMM model for Group 1. Five hidden states are interpreted as #1 Vertical Movement, #2 Vertical Movement + Move Left, #3 Move Forward, #4 General Lateral Movement, and #5 Oblique Movement.



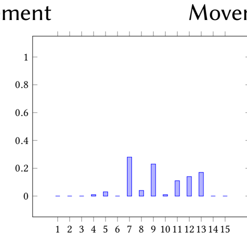
(a) Hidden #1: Vertical Movement



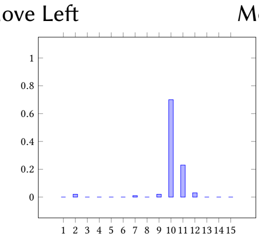
(b) Hidden #2: Vertical Movement + Move Left



(c) Hidden #3: Move Forward



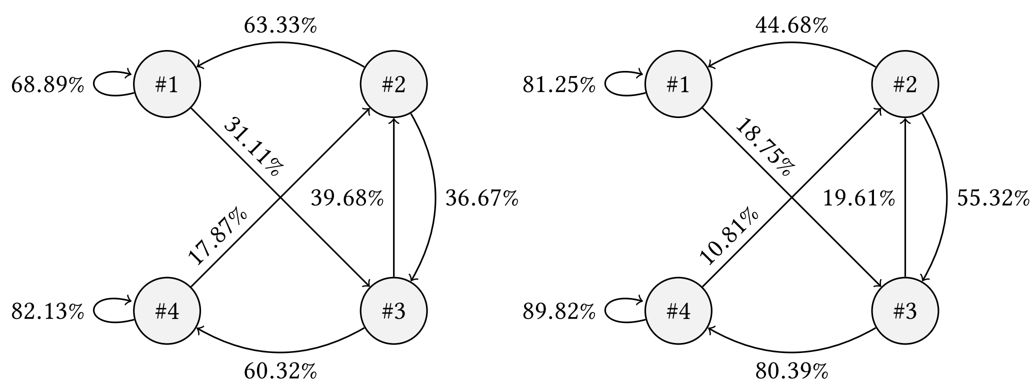
(d) Hidden #4: General Lateral Movement



(e) Hidden #5: Oblique Movement

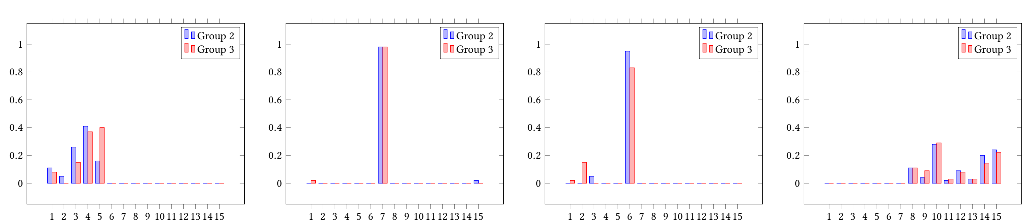
Fig. 7. Emission probabilities of the human operator behavior HMM model for Group 1.

Based on the model training and selection described previously, Figure 6 shows the selected operator behavior model structure for operators in Group 1. Emission probabilities for each hidden state are shown in Figure 7. One limitation of such an approach is that hidden states must be interpreted in light of corresponding emission probabilities, shown in Figure 7. Understanding that this process is a subjective one, we interpreted these five hidden states as #1 Vertical Movement, #2 Vertical Movement + Move Left, #3 Move Forward, #4 General Lateral Movement, and #5 Oblique Movement. Thus it appeared that generally operators had five different behavioral clusters while controlling drones in ETC mode that represented the goal states of the operators.



(a) Group 2 (b) Group 3

Fig. 8. Human operator behavior HMM models for Group 2 (left) and 3 (right). Four hidden states are interpreted as #1 Manage Flight Plan, #2 Return Supervisory Mode, #3 Launch Inspection Mode, and #4 Manual Control.



(a) Hidden #1: Manage Flight Plan (b) Hidden #2: Return Supervisory Mode (c) Hidden #3: Launch Inspection Mode (d) Hidden #4: Manual Control

Fig. 9. Emission probabilities of the human operator behavior HMM models for Group 2 and 3.

Figure 8 shows the selected operator behavior model structures for both Groups 2 and 3. Figure 8a is the operator behavior model for operators in Group 2, who were trained on both ETC and WSC, and Figure 8b is the behavior model for participants in Group 3, who were trained on only WSC. The interpretation for each hidden state was determined by their emission probabilities, which are shown in Figure 9. These two models share similar corresponding emission probabilities, which means that both models contain similar cognitive operator behavior states, which is not surprising given that people were using the same interface. These four hidden states are interpreted as #1 Manage Flight Plan, #2 Return Supervisory Mode, #3 Launch Inspection Mode, and #4 Manual

Control. These states represent higher functionalities that those in Figure 6, which is expected given the higher level of autonomy (and reduced control) in the WSC interface.

5 DISCUSSION

Based on the models described in the previous section, we can investigate operators' cognitive strategies and behavioral patterns exhibited in the test sessions. When examining HMM models of operator behavior, it is important to analyze two major components. The first is the emission probabilities from each hidden state. These probabilities give insight into the cluster of individual behaviors that form an overall cognitive grouping. Transition probabilities form the next critical assessment points, as these represent the likelihood of moving from one state to the other. The transition probabilities indicate cognitive flow and overall operator strategies while the emission probabilities represent the mechanics of lower level actions.

In looking at the basic differences between the models in Figures 6 and 8, one significant difference is that, not surprisingly, the cognitive models for the two different modes of control, ETC vs. WSC, have fundamentally different structures. The ETC model in Figure 6 shows five states (Group 1), while those participants with just WSC in the final test (Groups 2 and 3) exhibited a 4-state model (Figure 8). ETC is a higher workload mode of control, thus it is expected that this lower form of supervisory control would generate more states. The WSC design was intended to reduce workload, which can be seen in the HMMs. It was also designed to more explicitly support higher levels of cognitive reasoning like navigation and fine-grained nudge control, so it is not surprising that the cognitive states match the core functionalities designed into the interface.

5.1 The Impact of Training on Operator Strategies

Given that the performance results from Groups 2 and 3 were different from what was anticipated, it is worth examining these two models much more closely. Figure 8 and Figure 9 reveal that although the fundamental hidden states are the same between Groups 2 and 3, which is expected since they used the exact same interface, their different training programs produced different transition and emission probabilities.

Figure 8 demonstrates that people in Group 3 spent less time transitioning to the Manage Flight Plan state as compared to those in Group 2. Group 3 people with just WSC training monitored the vehicle more in the Supervisory Mode interface, and what is likely a critical difference, when they went into Inspection Mode, Group 3 people more often went into Nudge Control and made minor corrections to the vehicle's position, as compared to Group 2. In contrast, when participants with both WSC/ETC training in Group 2 went into Inspection Mode, they were 20% less likely to make any corrections.

Group 2 tended to switch to Inspection Mode just to check the surroundings and instead of using Nudge Control to change the position of the drone, they elected to enter a new waypoint through the Flight Plan Mode, which took more time and was less precise for short term changes. Indeed, it appeared Group 2 participants were overly focused on pinpointing the exact spots they wanted to set waypoints, which may indicate distrust concerning the stability and accuracy of the waypoint function than Group 3. As a result, these unnecessary worries not only slowed down their pace, but also likely led to more crashes. The preference of controlling the vehicle through waypoints instead of Nudge Control likely was a major contributor to the more than double crash rate for Group 2, as they did not make important corrections that kept the vehicle out of trouble. The question is then why would significantly more training, especially in ETC, negatively influence operators in this way?

Recall that Group 2 had the most training of any group, with 80 more minutes of training than the WSC Group 3, which included 30 minutes more of actual hands-on flight training. Group 2 participants were taught how to operate the drone first in ETC for various functions. Then once successful in this mode, Group 2 participants moved to WSC, which was an easier control mode. While we thought that the extra ETC training would give Group 2 an advantage using Nudge Control since ETC and Nudge Control are similar, this training had the opposite effect and made people in Group 2 use Nudge Control less than those in Group 3. Table 3, which shows how much time Groups 2 and 3 spent in Inspection mode illustrates that even though Group 2 had significantly more training time in how to position the vehicle at a fine-grained level through ETC, they avoided the closely-related Nudge Control mode more than those with just the WSC training that focused on higher level navigation.

The Group 2 problem of transitioning into a mode but not really taking any action was also seen in both transitions to and actions in the Flight Plan Mode. As seen in Figure 8, Group 2 participants had a clear preference for using the Flight Plan Mode (63.33% as opposed to 44.68% in Group 3). In this mode, operators could set and adjust waypoints, as well as change the altitude of the drone from a god's eye view, thus reducing workload. Curiously, while Group 2 operators preferred to go into this mode 19% more than Group 3, the emission probabilities reveal a lack of action for Group 2. These probabilities tell us that Group 2 participants were slightly more likely to add new waypoints than Group 3, but that Group 3 participants were more than 20% likely to adjust waypoints. This analysis suggests that while Group 2 preferred to use the Flight Plan Mode more than Group 3, they actually took fewer actions in terms of plan modification which ultimately made them more inefficient (slower) and also more likely to crash.

5.2 The Role of Trust

The HMM analysis reveals that Group 2 was less likely to use the Nudge Mode, a critical mode for moving the drone through restricted environments, but more likely to use the Flight Plan Mode, albeit in a sub-optimal manner. These actions suggest that Group 2 may have had too much trust in the flight control system, as they did not feel the need to reassess and update flight plans as much as Group 3, and also did not use nudge control as much. Another viewpoint could be that the extra time in ETC training may have made people reluctant to use any semblance of manual control since it was very high in cognitive and physical effort.

In effect, having ETC training could have made the Group 2 participants distrust nudge control even though ETC and nudge control are not the same. Thus, the dual training program could have caused mode confusion, the inability of an operator to recognize the actual state of automated control [5], This dual training approach also could have caused negative transfer of training [12], in that because ETC and Nudge Control were similar, but not the same, participants were not able to understand the nuanced differences and gravitated to the easier process of adding a waypoint.

5.3 Limitations of HMMs

The HMM models described in the previous section were developed based on sequential operation data which did not include any temporal factors. The transition probabilities in such models indicated the sequential transitions among cognitive states, which did not explicitly consider the duration in such states. Thus, a major limitation of such HMM models was that state duration information was lost. However, the development of Hidden semi-Markov models based on such data will be considered as part of future work since Hidden semi-Markov models are able to address this limitation by considering temporal factor [3]. However, significantly more data is needed for such applications.

One other limitation of this approach is the degree of subjectivity that accompanies such unsupervised learning approaches. Subjectivity is introduced in defining which states constitute the observed states from user clicks on an interface, which is not always straightforward. Then, once the "hidden" states are numerically defined, understanding just what such clusters mean is not always clear or even unanimous between researchers. Thus, others viewing these results might have a different interpretation so more work is needed to determine how to best interpret such models.

6 CONCLUSION

Drones, especially those that can execute first-person inspection tasks, exhibit enormous potential across many applications. In order to investigate drone operators' performance in navigation and inspection tasks in supervisory drone control systems with increasing autonomy, an experiment was conducted. This experiment focused on the potential impact from the different operator training processes as well as two different drone control interfaces with different levels of autonomy of enhanced teleoperation and waypoint supervisory control.

The experiment results showed that operators who completed tasks with a supervisory control interface at a higher level of autonomy tended towards better performance than those who used a lower level interface, but only when they did not have training in lower level, enhanced teleoperation control. Operator behavior models developed through Hidden Markov Models allowed us to understand that people with both supervisory and enhanced teleoperation control training were not able to exert the right control at the right time to the same degree that people with just training in supervisory control. Future research will examine the role of trust in such situations as it was not clear from the results whether operators with more training trusted the control system too much or whether they did not trust the system enough.

ACKNOWLEDGMENTS

This research was sponsored by the Office of Naval Research under the Science of Autonomy program. Dr. Alex Stimpson, Ben Welton, Rohini Sharma, Akash Jain, Sam Liu, Justin Havas, and Mihir Dutta were all instrumental parts of a very complicated testing process.

REFERENCES

[1] Leonard E. Baum and Ted Petrie. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *The Annals of Mathematical Statistics* 37, 6 (Dec. 1966), 1554–1563. https://doi.org/10.1214/aoms/1177699147

[2] Yves Boussemart. 2011. *Predictive models of procedural human supervisory control behavior.* Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA.

[3] Yves Boussemart and M. L. Cummings. 2011. Predictive models of human supervisory control behavioral patterns using hidden semi-Markov models. *Engineering Applications of Artificial Intelligence* 24, 7 (2011), 1252–1262. https://doi.org/10.1016/j.engappai.2011.04.008

[4] Yves Boussemart, M. L. Cummings, Jonathan Las Fargeas, and Nicholas Roy. 2011. Supervised vs. Unsupervised Learning for Operator State Modeling in Unmanned Vehicle Settings. *Journal of Aerospace Computing, Information, and Communication* 8, 3 (2011), 71–85. https://doi.org/10.2514/1.46767

[5] Jan Bredereke and Axel Lankenau. 2002. A Rigorous View of Mode Confusion. In *Computer Safety, Reliability and Security. (SAFECOMP 2002)*, Stuart Anderson, Massimo Felici, and Sandro Bologna (Eds.), Vol. 2434. Springer, Berlin, Heidelberg, 19–31. https://doi.org/10.1007/3-540-45732-1_4

[6] Ahani E. Bridget and O. Abass. 2011. A sequential monte carlo approach for online stock market prediction using hidden markov models. *Journal of Modern Applied Statistical Methods* 10, 2 (2011), 669–675. https://doi.org/10.22237/jmasm/1320121440

[7] M. L. Cummings, K. Jackson, P. Quimby, and D. Pitman. 2012. Field Testing of a Quad Rotor Smartphone Control System. *International Journal of Micro Air Vehicles* 4, 3 (2012), 165–177. http://hdl.handle.net/1721.1/86945

[8] Cristina Gómez and David R. Green. 2017. Small unmanned airborne systems to support oil and gas pipeline monitoring and mapping. *Arabian Journal of Geoscience* 10, 9 (May 2017), 202. https://doi.org/10.1007/s12517-017-2989-x

[9] Aditya Gupta and Bhuwan Dhingra. 2012. Stock market prediction using Hidden Markov Model. *IEEE 2012 Students Conference on Engineering and Systems* (2012), 1–4. https://doi.org/10.1109/SCES.2012.6199099

[10] Minwoo Kim. 2018. *The impact of skill-based training across different levels of autonomy for drone inspection tasks.* Master's thesis. Duke University, Durham, NC.

[11] Vianey Leos-Barajas and Théo Michelot. 2018. An Introduction to Animal Movement Modeling with Hidden Markov Models using Stan for Bayesian Inference. *ArXiv e-prints* (June 2018). arXiv:1806.10639

[12] Dahai Liu, Elizabeth L. Blickensderfer, Nikolas D. Macchiarella, and Dennis A. Vincenzi. 2008. Transfer of Training. In *Human Factors in Simulation and Training*, Peter A. Hancock, Dennis A. Vincenzi, John A. Wise, and Mustapha Mouloua (Eds.). CRC Press, Boca Raton, FL, 49–60.

[13] Gonzalo Pajares. 2015. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Engineering & Remote Sensing* 81, 4 (2015), 281–330. https://doi.org/10.14358/PERS.81.4.281

[14] Raja Parasuraman and Thomas B. Sheridan. 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 (May 2000), 286–297. http://dx.doi.org/10.1109/3468.844354

[15] Alex Pentland and Andrew Liu. 1999. Modeling and prediction of human behavior. *Neural Computation* 11, 1 (Jan. 1999), 229–242. https://doi.org/10.1162/089976699300016890

[16] David Pitman and M. L. Cummings. 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 (2012), 18. http://dx.doi.org/10.1155/2012/768180

[17] Julie C. Prinet, Andrew Terhune, and Nadine B. Sarter. 2016. Supporting Dynamic Re-Planning In Multiple Uav Control: A Comparison of 3 Levels of Automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (Dec. 2016), 423–427. https://doi.org/10.1177/1071181312561095

[18] L. R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 2 (Feb. 1989), 257–286. https://doi.org/10.1109/5.18626

[19] L. R. Rabiner and B. Juang. 1986. An introduction to hidden Markov models. *IEEE ASSP Magazine* 3, 1 (1986), 4–16. https://doi.org/10.1109/MASSP.1986.1165342

[20] Narayanan Ramanathan. 2006. Applications of Hidden Markov Models. http://www.cs.umd.edu/~djacobs/CMSC828/ApplicationsHMMs.pdf

[21] R. Schaufele, L. Ding, Nick Miller, H. Barlett, Michael Lukacs, and D. Bhadra. 2017. *FAA Aerospace Forecast: Fiscal Years 2017–2037*. Technical Report. Washington DC.

[22] Thomas B. Sheridan. 1992. *Telerobotics, Automation, and Human Supervisory Control*. MIT press, Cambridge MA.

[23] Greg Zacharias and Mark Maybury. 2011. *Operating Next-Generation Remotely Piloted Aircraft for Irregular Warfare*. Technical Report. United States Air Force Scientific Advisory Board, Washington DC.

[24] Junjie Zhang, Jaewook Jung, Gunho Sohn, and Michael Cohen. 2015. Thermal Infrared Inspection of Roof Insulation Using Unmanned Aerial Vehicles. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-1/W4 (2015), 381–386. https://doi.org/doi:10.5194/isprsarchives-XL-1-W4-381-2015

[25] Yong Zhang, Xiuxiao Yuan, Yi Fang, and Shiyu Chen. 2017. UAV Low Altitude Photogrammetry for Power Line Inspection. *ISPRS International Journal of Geo-Information* 6, 1 (2017), 14. https://doi.org/10.3390/ijgi6010014