Authors' accepted manuscript. Paper presented at the 31st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (Naples, Italy, Aug. 29 - Sept. 02, 2022).
The final published version will be available at https://ieeexplore.ieee.org/xpl/conhome/1000636/all-proceedings

Towards evaluating the impact of swarm robotic control strategy on operators' cognitive load

Anita Paas[1,2], Emily B. J. Coffey[1], Giovanni Beltrame[3], David St-Onge[2]

Abstract— **The use of multi-robot systems is increasing in disaster response, industry, transport, and logistics. Humans will remain indispensable to control and manage these fleets of robots, particularly in safety-critical applications. However, a human operator's cognitive capacities can be challenged and exceeded as the sizes of autonomous fleets grow, and more sophisticated AI techniques can lead to opaque robot control programs. In a user study (n = 40), we explore how autonomous swarm intelligence algorithms and novel tangible interaction modalities relate to subjective and physiological indices of operator cognitive load (i.e., NASA Task Load Index, heart rate variability). Our findings suggest that there are differences in workload across conditions; however, subjective and cardiac measures appear to be sensitive to different aspects of cognitive state. The results hint at the potential of both tangible interfaces and automation to engage operators and reduce cognitive load, yet show the need for further validation of workload measures for use in studying and optimizing human-swarm interactions.**

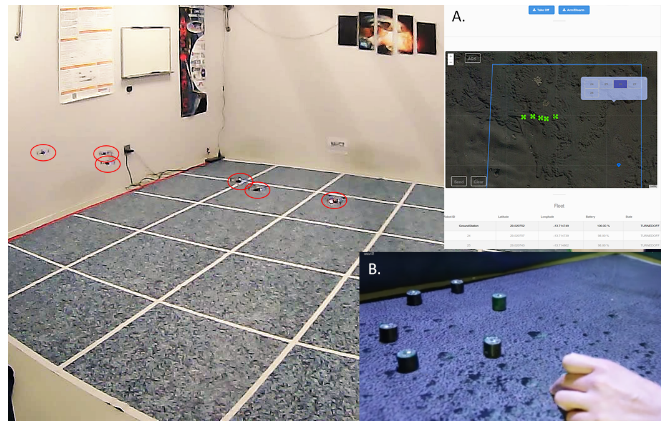

Fig. 1: Six micro-UAVs (Crazyflies - red circles) in the flying arena during a session. View from the participants' position. A. A sample view of the tablet interface. B. a sample view of the tangible interface using the Zooids robots.

I. INTRODUCTION

Unmanned robotic teams are growing in popularity for many real world scenarios, such as search and rescue, medicine, and space exploration. Despite the recent technological advancements, there are still numerous problems that need to be addressed for robust multi-robot deployments. The general problem of systematically designing reliable, robust, and resilient large-scale multi-robot systems remains open, and is one of the most important challenges of the decade [1]. On the path to fully autonomous swarms of robots, humans will act as monitors and controllers for the robot fleet, as technology gaps prevent us from performing all of the required functions autonomously. The control of a networked pool of autonomous robots by an individual or by a team is challenging because of the complexity of the coordination task [2] and of the information exchanged. The difficulty of ensuring operational performance is compounded when incoming information is scattered, delayed, or unreliable. These factors lead to increased pressure on human operators' cognitive resources and their ability to maintain situational awareness, detect problems, and make successful decisions.

There is a long history of research in the neurobiological correlates of task performance-related human cognition (and its limitations) using techniques like functional magnetic resonance imaging and magnetoencephalography. These techniques are used to measure brain activity associated directly with cognitive processes and are valuable for neuroscientists to study the neurobiological bases of cognitive load, yet are impractical for most user applications as they are bulky, expensive, and limit participant movement. However, portable equipment is becoming available that can measure physiological proxies of human workload and stress states. One of the most promising biosignals that fulfill the criteria of being non-invasive, wearable, reliable, and inexpensive is heart-rate variability [3], [4]. Heart-rate variability (HRV) can be used to objectively evaluate and compare operator states caused by task conditions.

Our long term goal is to create control modalities that maximize the ability of an operator to achieve mission objectives using a large fleet of robots. Proxies of cognitive load can be monitored during missions and used to improve their design. In multi-robot systems, we identify two ways in which the operator's cognitive load can be reduced: by giving more intelligence and autonomy to the robotic system, and by designing a command interface that is intuitive and natural to the operator. In this paper, as a stepping stone towards our final goal, we investigate the impact of autonomous swarm behaviors and (possibly more intuitive) tangible interfaces on operator cognitive load in multi-robot missions.

We experimentally compare two types of interfaces (graphical and tangible) and two robot control modalities

*We acknowledge the support of the Canadian Space Agency (CSA) (19FAPOLA32)

[1]Anita Paas and Emily B. J. Coffey are with the Department of Psychology, Concordia University, Montréal, Canada a_paas@live.concordia.ca emily.coffey@concordia.ca

[2]Anita Paas and David St-Onge are with the Department of Mechanical Engineering at École de Technologie Supérieure, Montréal, Canada david.st-onge@etsmtl.ca

[3]Giovanni Beltrame is with the Department of Computer and Software Engineering at Polytechnique Montréal, Canada giovanni.beltrame@polymtl.ca

(individual waypoints and group self-deployment) in a user study (n = 40) in which operators controlled a fleet of six micro-quadcopters during an exploration mission. We hypothesize that the tangible interface demands more involvement, but reduces cognitive load on the user (H1), and that greater autonomy of the robotic fleet reduces operator cognitive load (H2). We use subjective ratings and heart rate variability to monitor the effects of these interface variations.

We believe this work underscores the potential for complementary information provided by self-reported and physiologically-based measures of workload in common use in human-robot interface research, particularly to evaluate and guide the development of new operator interfaces. In the following sections, we first consider relevant previous work II, then give an overview of the robotic system deployed in section III. The user study is covered in section IV, and the presentation of the results in section V.

II. Related work

Human-Swarm Interaction (HSI) is a special case of Human-Robot Interaction (HRI), due to the large numbers of units involved and the presence of emergent group behaviours [5]. Controlling multiple robots can be cognitively challenging, and its complexity is influenced by (a) the required level of attention of the operator [6], [7], (b) the degradation of the robots' performance when left unattended [8], and (c) the need to manage the coordination between robots to perform a task [9]. While robot teams are the subject of many interaction studies, HSI has particular considerations due the fundamental clash between a centralized control element and the distributed nature of swarms [5]: a human operator issues commands to a swarm whose behaviours are governed by local interactions and self-organization.

In the emerging field of HSI, user studies are often employed to investigate workload and performance [10]. Generally, these studies focus on specific interface media and many are conducted in simulation only, suffering from a reality-gap [11]. Such studies have demonstrated that swarm behaviours are more challenging to visualize than deterministic and predictable control strategies [12]. It is worth noting that the collective movement of a swarm can convey non-trivial, swarm-behavioural information to an operator [12], and an effective communication strategy is key [13]. Although humans are generally good at recognizing patterns of collective motion [14], it is important to design swarm motion dynamics [15] that are compatible with human cognitive skills of interpretation and to which the operator can properly react.

In particular, since human attention can fluctuate and the capacity of human working memory is limited, the number of robots a single operator can control is also limited [6], [7]. In fact, in tasks where operators have to recognize a common type of swarm behaviour (e.g. flocking), the operators report using a holistic approach to the perception of collective motion inherent to emergent swarm behaviours. Walker [14] observed operators applying strategies such as "unfocusing their eyes" and "watching for a global pattern to emerge". Those strategies are a response to the increased demands of swarm interaction [16], [17]. Following this line of research, Podevijn et al. [18] successfully showed that the number of robots does not influence the cognitive load required from a user if the control is performed on the swarm as a whole.

Most current research on interfaces deals with tablets and phones, but some studies consider tangible interfaces as well. Research shows that tangible interfaces can benefit cognitive performance, aiding performance through support for memory and by decreasing cognitive load [19], [20]. Furthermore, there is evidence that tangible interfaces promote user engagement and immersion into the scenario or problem [21]. Finally, for some problems, participants are more likely to find a solution to a problem when using a tangible interface compared to a tablet [22].

III. Swarm robotic system

Our research interest targets exploration missions with unmanned aerial vehicles (UAVs). While we conducted several experiments in the field [23], prior design and validation in an indoor arena, in which experimental parameters can be better controlled, is preferable. This work is based on indoor micro-UAVS (Crazyflies[1]). Figure 2 shows the experimental overview and illustrates some of the features of the robotic system, including communication between robots, safety considerations, and human-robot communication. All technical aspects of the robotic system are thoroughly explained in the work of St-Onge, et al. [24].

A. Deployment ecosystem

In several remote operation scenarios, such as planetary exploration [23], sudden changes in the environment may arise too quickly for the operator to send each robot a timely command, due to communication delays. To limit the exploration task to a manageable area, and for the more practical reason of working within the limitations of the arena, we used a virtual fence to set limits before take-off (Fig. 2*C*). Additional layers of security are required to cope with communication and human errors. For example, when setting a goal, the destination's distance is verified to be reachable in less than 5 seconds (e.g., less than 5 meters, if the UAV maximum velocity is 1m/s). This verification is done on input, before sending the command, and again verified by the receiving robot. As the robots generate their own trajectories, they can use the known locations of their neighbours to avoid each other, using a decentralized collision avoidance algorithm [25]. To minimize risk of collision, we also flew UAVs at different altitudes.

In our user study, the goal was to locate simulated hidden ground features within the flying area using two different deployment methods. In the self-deployed group control mode, participants assigned a marker (hotspot) around which the UAVs distributed themselves and searched for the features autonomously. In the individual waypoint deployment method, participants assigned destination positions to each UAV and the UAVs searched for features along their path.

[1] `https://www.bitcraze.io/products/crazyflie-2-1/`

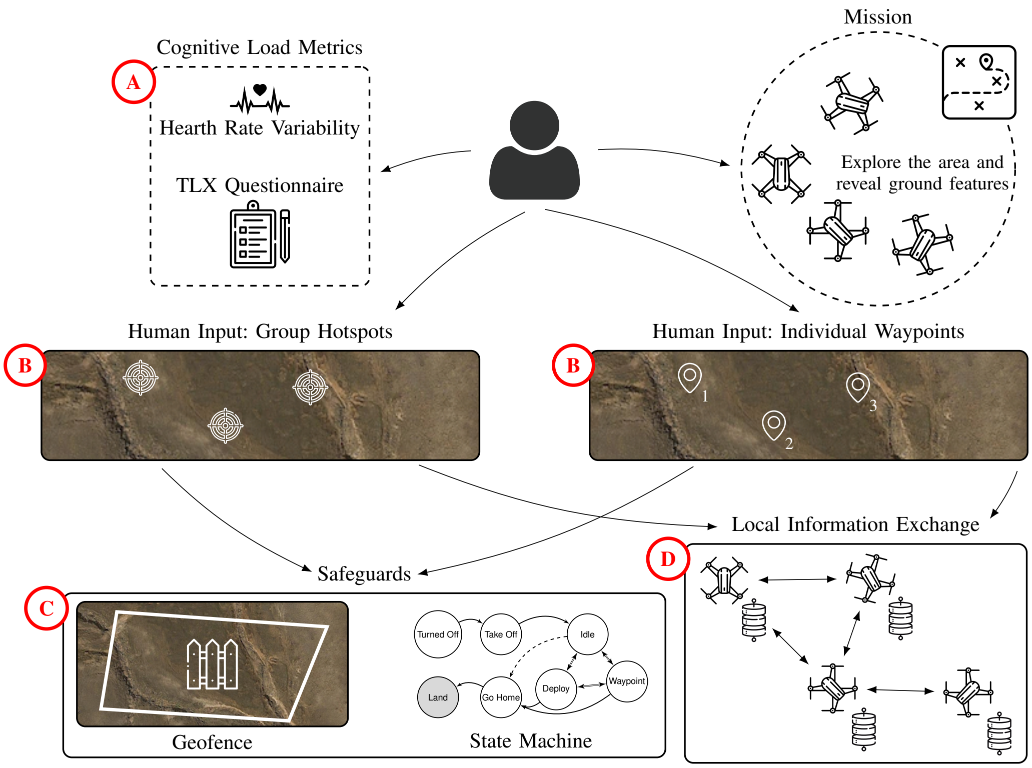

Fig. 2: Experiment overview: Operator cognitive load is measured by self-report and physiology (A). The operator controls the swarm in each of two modes - Group Hotspots and Individual Waypoints (B). During the missions, a geofence acts as a safety feature and a state machine provides coordination between robots (C). Consensus is achieved via a distributed database (D).

B. Tangible interface

Taking a step back, we looked at the origin of the geographic information systems and mission planners; the physical map. Collaborative logistics plans have been successfully completed using a simple map for centuries. With the help of miniature robots' design, a portable localization system, and image processing, we designed a smart, tangible, map-based command center for exploration missions [26].

During the experiment, we used table-top robots, called *Zooids* [27]. The Zooids are a group of small cylindrical robots, 2.6 cm in diameter and 5 cm in height, localized using structured light emitted by a ceiling projector. We selected the Zooids for the minimal setup time, their low cost of manufacturing, their open-source controller code, and the simplicity of their manipulation.

The command centre consisted of a tent under which Zooid robots imitated the movements of the UAV fleet, by moving over a map of the deployment area. The Zooids were equipped with short-range radio communication devices and an RGB LED on the top. Each robot knew its position and, after initialization, was assigned a unique micro-quadcopter counterpart (i.e., a Crazyflie) on the field. Communication with the fleet passed through a central communication node for long-range transmission. The LED color showed the aircraft battery level in real time (green, yellow, or red) with different blinking patterns for different behavioural states. For individual waypoint control (Fig. 2*B* right), command was sent to a flying robot by holding its wheeled counterpart to the desired position for two seconds. For group self-deployment (Fig. 2*B* left), the tangible interface has specially-marked robots with which a user can input focal points. Using these focal points, a region of interest is created by the union of these focal points and the robots distribute evenly around this region.

IV. Method

A. Study design

To establish a controlled environment, we conducted our study indoors using a fleet of six micro-quadcopters with global localization. The code for the robot behaviour is available online[2]. Details of the algorithm for the deployment behaviour and the software infrastructure for the general control of the swarm are presented in [24]. Details of the algorithm for the individual versus group control strategies are presented in [23]. Participants were asked to stand in a designated area at the border of our flying arena. For immersion reasons, no glass or net was added between them and the flying arena. Each participant completed the four conditions in random order, controlling the fleet from: 1. a tablet using group control (self-deployment), 2. a tablet using individual waypoints, 3. a tangible interface using group control (self-deployment), and 4. a tangible interface using individual waypoints. When using the tablet interface, participants stood facing the flying arena and controlling the swarm required little movement. When using the tangible

[2] `https://github.com/MISTLab/ROSBuzz`

interface, participants controlled the swarm at a table beside the flying arena and looked up to observe the UAVs, moving the Zooids with their hands.

The objective of each of the four missions was to find as many hidden ground features (simulated) as possible in the flying area, up to a maximum of five features, within five minutes. In the tablet/group condition, participants manipulated a user interface on a tablet to assign one or two markers around which the UAVs distributed themselves autonomously. When in the tablet/individual condition, participants again used a tablet but assigned a position to each UAV individually. Commands were entered into the interface and transmitted to the fleet when the participant pressed a "send" button. In the tangible interface setting, UAVs were controlled using Zooids that replicated the movements of the robots in the flying arena. When using the tangible/group control mode, one or two Zooids (used to input a region of interest) were picked up and placed at the desired location around which the UAVs distributed themselves evenly. In the tangible/individual condition, each UAV was controlled by moving and holding its respective table-top robot to the desired location until the command had been received (indicated by LED lights changing colors). While being held, the Zooids cannot illustrate the drone motion, but as soon as the operator releases it, the Zooid resumes its behaviour by going back to the respective drone's current location on the map. Once released, the table-top robot mimicked the movement of the UAV.

The experiment was performed by 40 adult participants. All participants (F: 13 M: 27 Age: 18-55 years) were able to complete the four missions. The user study and recruitment process had the approbation of the ethical committee of all three universities involved in the project. After each of the four conditions, participants were asked to rate their subjective workload level with the NASA Task Load Index (NASA-TLX) [28] for self-assessment of their interaction. As an objective, physiological measure of cognitive workload, heart rate variability was analyzed.

B. NASA Task Load Index

Subjective ratings of workload were recorded using a slightly modified version of the NASA-TLX with the following dimensions and questions: Effort (How hard did you have to work?), Frustration (How insecure, discouraged, irritated, stressed, and annoyed were you?), Hurried (How hurried or rushed was the pace of the task?), Mental Demand (How mentally demanding was the task?), Physical Demand (How physically demanding was the task?), and Successful (How successful were you in accomplishing what you were asked to do?). Participants rated each dimension for each task on a 7-point Likert scale. Recent studies have used the NASA-TLX not only as a combined metric, but as a multidimensional tool, considering dimensions separately [29], [30]. Based on these studies, we investigated each factor separately and combined.

C. Heart rate variability

Heart rate variability (HRV) was measured using a Biopac MP35 with a 3-lead ECG electrode set (SS2L, Biopac Systems Inc.). Measurements were taken according to the Biopac Student lab Pro manual [31] at 1000 Hz, with a 35 Hz low pass filter applied during recording. 3M Red Dot electrodes with a built-in abrader (3M, ID: 7000128699) were placed on the recently cleaned right wrist and both ankles. To analyze the ECG data, we used the Python package *hrvanalysis* made for the Aura Healthcare project [32]. Out of the 40 participants, 8 experienced technical issues with the BioPac recording and were removed from the analysis. A first step of outlier removal, ectopic beats removal, and linear interpolation was conducted to prepare the remaining datasets. We used the SDNN measure (i.e., the standard deviation of the interval between normal heartbeats) to investigate cognitive load, as research has shown that SDNN decreases with increased cognitive load [33], [34] and that there is a positive correlation between SDNN and cognitive performance [35].

V. RESULTS

In the user study, we compared a classic tablet-based mission planner with a novel tangible interface and validated the comparison with two sets of control algorithms in a controlled environment. For all analyses, assumptions were met for statistical procedures, and a Bonferroni correction was used as a conservative method of type I error control. The results of the mission goals (i.e., number of features found and number of actions performed; see Fig. 3) show that participants were able to achieve the mission goals with a high rate of success in each condition. The analysis of variance (ANOVA) on the average number of features found showed significant main effects of interface and control algorithm and a significant interaction between interface and control algorithm (see Table I). Follow-up t-tests for showed that participants found more features in the tablet/group, tablet/individual, and Zooid/group conditions compared to the Zooid/individual condition (p-values <0.004 for all comparisons). Analyses on the average number of actions (i.e., control inputs) performed showed significant main effects of interface and control algorithm, and a significant interaction (see Table II). Follow-up t-tests showed a significant difference between all conditions (p <0.001 for all comparisons), with the most actions performed in the Zooid/group condition. For above and all subsequent analyses, a Bonferroni correction was used due to multiple t-tests in the analyses. The corrected p-value that would indicate statistical significance is 0.0083.

TABLE I: Number of Features Found

	F-value	p-value	df
Interface	5.91	0.02	(1, 35)
Control algorithm	27.50	<0.001	(1, 35)
Interaction	6.90	0.013	(1, 35)

Figure 4 illustrates the subjective ratings for each dimension of the NASA-TLX for each combination of interface

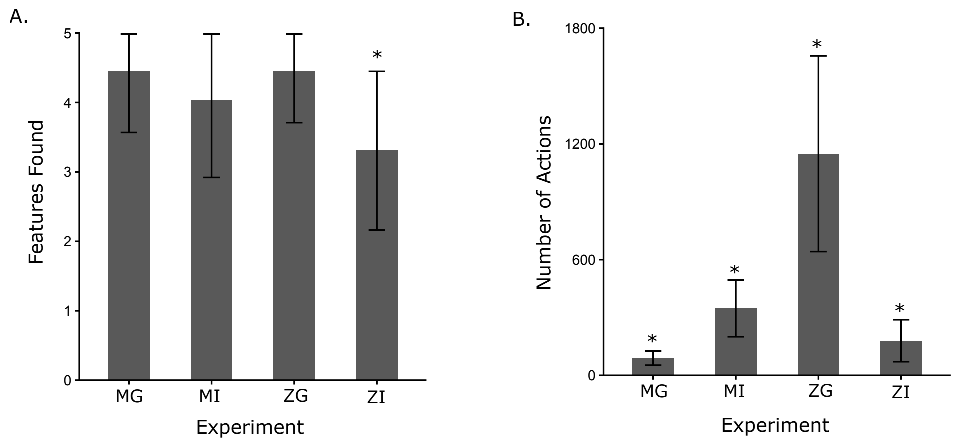

Fig. 3: Measures of mission success across interfaces and control modes. A. Average number of features found (maximum = 5), and B. Average number of actions performed for each experimental condition, as follows: tablet-based Mission Planner (M), tangible Zooids interface (Z), self-deployment group control mode (G), and individual waypoint control mode (I). Error bars represent standard deviation.* represents a significant difference between that condition and all other conditions. For all significant differences, p <0.008.

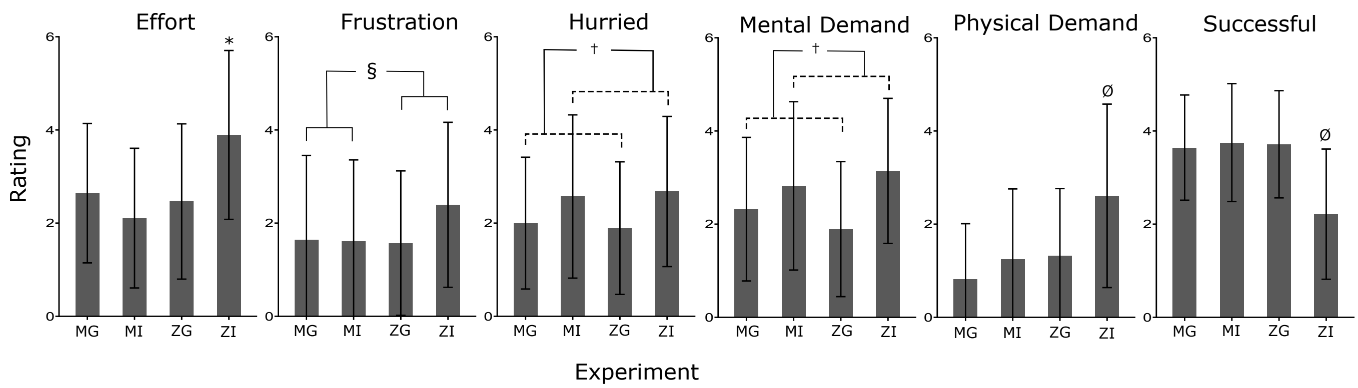

Fig. 4: Subjective measures of operator workload. Average and standard deviation of the Likert-scale results (from 0 to 6) for the NASA-TLX dimensions each participant completed after each of their missions. The scores are grouped by condition: tablet-based Mission Planner (M), tangible Zooids interface (Z), self-deployment group control mode (G), and individual waypoint control mode (I). Error bars represent standard deviation. * represents a significant difference from MI and ZG, but not MG; § represents a significant difference between M and Z conditions, regardless of deployment; † represents a significant difference between G and I control modes, regardless of interface; Ø represents a significant difference from all others. For all significant differences, p <0.05

TABLE II: Number of Actions Performed

	F-value	p-value	df
Interface	94.74	<0.001	(1, 35)
Control algorithm	67.70	<0.001	(1, 35)
Interaction	179.71	<0.001	(1, 35)

and control algorithm. As seen in Figure 4, the individual waypoint control was rated more mentally demanding and more hurried than the self-deployed group control mode. Participants rated the Zooid interface more frustrating than the tablet interface. The Zooid/individual condition was rated as requiring most effort and least successful of the conditions. This is complementary to the performance results, which showed that fewer features were found in the Zooid/individual condition compared to the rest. The combined NASA-TLX scores were analyzed in a 2 (interface) x 2 (control algorithm) ANOVA. This analysis showed a significant main effect of interface ($F(1, 27) = 5.14$, $p = 0.032$) and a significant main effect of control algorithm ($F(1, 27) = 11.28$, $p = 0.002$). In the combined metric, participants rated the Zooid interface higher than the tablet interface, and rated the individual waypoint control higher than the self-deployed group control.

While the performance results and subjective ratings tend to indicate that a higher level of cognitive load is required to operate the tangible interface, the HRV measurement does not corroborate this conclusion. Figure 5 shows the average SDNN for each condition. The 2 (interface [tablet/tangible]) x 2 (control algorithm [individual/group]) ANOVA of the SDNN data showed significant main effects of interface and control algorithm, as well as a significant interaction (see Table 3). Follow-up t-tests confirmed the significant main effects and showed significant differences between tablet and Zooid interfaces at both levels of control algorithm (p <0.001 for all significant comparisons) but no other significant differences. We also explored correlations between the NASA-TLX ratings (both the combined NASA-TLX and each dimension individually) and SDNN. There were no significant correlations between the combined NASA-TLX ratings and SDNN for any of the experimental conditions

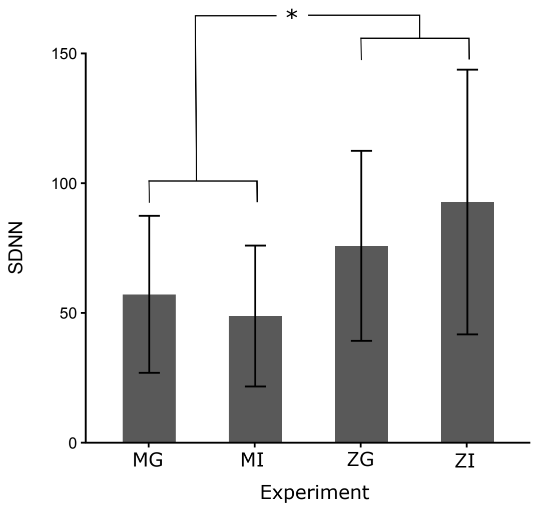

Fig. 5: Physiologically-based measures of operator workload. Heart rate variability presented as the standard deviation of the average normal-to-normal intervals for each condition: tablet-based Mission Planner (M), tangible Zooids interface (Z), self-deployment group control mode (G), and individual waypoint control mode (I). Based on previous work, higher SDNN is associated with lower cognitive load [33], [34], [35]. Error bars represent standard deviation. * represents a significant difference between M and Z conditions, regardless of deployment.

TABLE III: SDNN

	F-value	p-value	*df*
Interface	33.21	<0.001	(1, 26)
Control algorithm	2.18	0.151	(1, 26)
Interaction	7.05	0.013	(1, 26)

(all p-values >0.06). There was a significant negative correlation between the Successful dimension (i.e., the rating of successful performance on the task) and SDNN for the Zooid/individual condition only ($r(26) = 0.42$, $p = 0.024$). Our primary measure was SDNN due to its use in similar works [33], [34], [35]. We also explored other relevant HRV metrics known to relate to cognitive workload, including pNN50 (which showed the same pattern of results as SDNN), RMSSD, LF/HF, LF (all of which showed a significant main effect of interface only), and HF (which showed no significant results).

VI. Discussion and Conclusions

The robotic swarm presented in this paper allowed us to compare, in a controlled setting, a classic tablet-based mission planner with a novel tangible interface and validate the comparison with two sets of control algorithms – individual and group. Participants controlled the UAVs with a tablet and tangible Zooids to find hidden simulated features. We expected that the tangible Zooid interface would demand more involvement but reduce cognitive load on the user (H1). We found that the Zooid interface did result in more involvement, as shown by a higher number of actions in the Zooid/group condition. Also in-line with our hypotheses, the results from the SDNN measure suggest that the Zooid interface may be less cognitively demanding, as SDNN was higher when participants were using the Zooid interface compared to the tablet interface. Previous research has shown that SDNN decreases when cognitive load increases [33], [34]. We should also consider that the SDNN results for the Zooids may be confounded with physicality, as the Zooid interface required participants to bend over the table and reach for and move the Zooids, whereas with the tablet interface, participants stood while controlling the tablet. In contrast, the combined subjective ratings from the NASA-TLX and the ratings in the frustration dimension showed that participants rated the Zooid interface higher (i.e., more demanding across the six dimensions, and more frustrating) than the tablet interface. In other contexts, it has been noted that while performance, subjective, and objective measures are all valuable measures, they do not always track together, and may differ in their sensitivity and variability across research domains and task scenarios [36], [37], [38]. These results underscore the need for further work to assess which subjective and objective measures are most apt to measure operators' cognitive load during swarm robotic control.

We predicted that greater autonomy of the robotic fleet would reduce operator cognitive load (H2). Robotics engineers generally agree that more embedded intelligence allows for a decrease of dependency on the operator, thus most likely reducing the operator task load. As shown in Fig. 4, the control mode in which UAVs organised themselves based on a single general user input (i.e., self-deployment group control mode) was perceived by participants as being less physically and mentally demanding, required less time to complete the task, and overall less cognitively demanding than the individual waypoint control mode, as demonstrated by the combined ratings on the NASA-TLX. As shown in Fig. 3, increased automation by the robots did not consistently reduce the amount of work (i.e., the number of actions) across interfaces, nor did it consistently improve performance on the mission goals (i.e., the number of features found). This result may be due to the mission, task environment and even the specifics of how automation is implemented.

Our work represents a first human user study with a swarm of UAVs, simultaneously looking at subjective and objective measures of cognitive workload, across conditions that differ in their level of automation and user interface. The results hint at the potential of both tangible interfaces and automation to engage operators and reduce cognitive load, yet to be fully realized, all aspects of a particular application must be considered. In the present case, specific design choices in the interface implementation and the nature of the control algorithms, as well as the degree of familiarity people have with tangible vs. tablet-based controls, would all affect the user's cognition and experience. To derive a more general solution to the problem of assessing cognitive load in HSI, we require a solid methodology that will allow us to test and guide each design choice during development, based on performance, subjective and objective measures. Important next steps will be to validate additional physiological measures (e.g., pupil dilation, skin conductivity) and their combinations as indicators of workload in the context of swarm control applications, and evaluate their relationship to users' subjective experience and to objective measures of mission success. With more accurate wearable sensors and a flexible user-centric control methodology, we expect to increase the safety and effectiveness of human-swarm teams across many applications.

ACKNOWLEDGMENT

We would like to thank Alexandre Piot and Camille Bouhour for protocol testing, and Alexandre Piot and Vivek Shankar Varadharajan for assistance with data collection.

REFERENCES

[1] G.-Z. Yang, J. Bellingham, P. E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, *et al.*, "The grand challenges of science robotics," *Science Robotics*, vol. 3, no. 14, p. eaar7650, 2018.

[2] L. Chérif, V. Wood, A. Marois, K. Labonté, and F. Vachon, "Multitasking in the military: Cognitive consequences and potential solutions," *Applied cognitive psychology*, vol. 32, no. 4, pp. 429–439, 2018.

[3] N. Mogilever, L. Zuccarelli, F. Burles, G. Iaria, G. Strapazzon, L. Bessone, and E. B. Coffey, "Expedition cognition: a review and prospective of subterranean neuroscience with spaceflight applications," *Frontiers in human neuroscience*, vol. 12, p. 407, 2018.

[4] B. Cowley, M. Filetti, K. Lukander, J. Torniainen, A. Henelius, L. Ahonen, O. Barral, I. Kosunen, T. Valtonen, M. Huotilainen, *et al.*, "The psychophysiology primer: a guide to methods and a broad review with a focus on human–computer interaction," *Foundations and Trends® in Human–Computer Interaction*, vol. 9, no. 3-4, pp. 151–308, 2016.

[5] A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis, "Human Interaction with Robot Swarms: A Survey," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 9–26, 2016.

[6] M. Lewis, "Human Interaction With Multiple Remote Robots," *Reviews of Human Factors and Ergonomics*, vol. 9, no. 1, pp. 131–174, 2013.

[7] M. Lewis, J. Wang, and P. Scerri, "Teamwork Coordination for Realistically Complex Multi Robot Systems," in *NATO Symposium on Human Factors of Uninhabited Military Vehicles as Force Multipliers*, 2006, pp. 1–12.

[8] D. R. Olsen Jr and S. Bart Wood, "Fan-out : Measuring Human Control of Multiple Robots," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienna, 2004, pp. 231–238.

[9] J. Wang and M. Lewis, "Assessing Coordination Overhead in Control of Robot Teams," *Systems, Man and Cybernetics*, pp. 2645–2649, 2007.

[10] C. E. Harriott, A. E. Seiffert, S. T. Hayes, and J. A. Adams, "Biologically-inspired human-swarm interaction metrics," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 58, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2014, pp. 1471–1475.

[11] G. Podevijn, R. O'Grady, C. Fantini-Hauwel, and M. Dorigo, "Investigating the effect of the reality gap on the human psychophysiological state in the context of human-swarm interaction," *PeerJ Computer Science*, vol. 2, no. 82, 2016.

[12] F. Levillain, D. St-Onge, E. Zibetti, and G. Beltrame, "More Than the Sum of its Parts: Assessing the Coherence and Expressivity of a Robotic Swarm," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2018, pp. 583–588.

[13] G. Dietz, J. L. E, P. Washington, L. H. Kim, and S. Follmer, "Human perception of swarm robot motion," in *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, 2017, pp. 2520–2527.

[14] P. Walker and M. Lewis, "Characterizing Human Perception of Emergent Swarm Behaviors," in *Intl. Conference on Systems, Man, and Cybernetics (SMC)*. Budapest, Hungary: IEEE, 2016, pp. 2436–2441.

[15] S. Kerman, D. Brown, and M. A. Goodrich, "Supporting human interaction with robust robot swarms," *Proceedings - 5th Intl. Symposium on Resilient Control Systems, ISRCS 2012*, pp. 197–202, 2012.

[16] S. K. Card, *The Psychology of Human-Computer Interaction*, 2017.

[17] T. B. Sheridan, *Humans and automation : system design and research issues*. Santa Monica, CA, USA: Human Factors and Ergonomics Society, 2002.

[18] G. Podevijn, R. O'Grady, N. Mathews, A. Gilles, C. Fantini-Hauwel, and M. Dorigo, "Investigating the effect of increasing robot group sizes on the human psychophysiological state in the context of human-swarm interaction," *Swarm Intelligence*, vol. 10, no. 3, pp. 1–18, 2016.

[19] O. Shaer and E. Hornecker, *Tangible user interfaces: past, present, and future directions*. Now Publishers Inc, 2010.

[20] D. Kirsh, "Thinking with external representations," *AI & society*, vol. 25, no. 4, pp. 441–454, 2010.

[21] M. J. Kim and M. L. Maher, "The impact of tangible user interfaces on designers' spatial cognition," *Human–Computer Interaction*, vol. 23, no. 2, pp. 101–137, 2008.

[22] F. Vallée-Tourangeau, S. V. Steffensen, G. Vallée-Tourangeau, and M. Sirota, "Insight with hands and things," *Acta psychologica*, vol. 170, pp. 195–205, 2016.

[23] D. St-Onge, M. Kaufmann, J. Panerati, B. Ramtoula, Y. Cao, E. B. J. Coffey, and G. Beltrame, "Planetary exploration with robot teams: Implementing higher autonomy with swarm intelligence," *IEEE Robotics Automation Magazine*, vol. 27, no. 2, pp. 159–168, 2020.

[24] D. St-Onge, V. S. Varadharajan, I. Švogor, and G. Beltrame, "From Design to Deployment: Decentralized Coordination of Heterogeneous Robotic Teams," *Frontiers in Robotics and AI*, vol. 7, no. May, pp. 1–16, 2020.

[25] M. Shahriari, I. Svogor, D. St-Onge, and G. Beltrame, "Lightweight collision avoidance for resource-constrained robots," in *IEEE/RSJ Conference on Intelligent Robots (IROS)*, 2018.

[26] D. St-Onge, V. S. Varadharajan, and G. Beltrame, "Tangible robotic fleet control," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 2019, pp. 2387–2389.

[27] M. L. Goc, L. H. Kim, A. Parsaei, J.-d. Fekete, P. Dragicevic, and S. Follmer, "Zooids : Building Blocks for Swarm User Interfaces," in *UIST*, Tokyo, 2016.

[28] D. A. Coelho, J. N. O. Filipe, M. Simões-marques, and I. L. Nunes, "The Expanded Cognitive Task Load Index (NASA-TLX) applied to Team Decision-Making in Emergency Preparedness Simulation," *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference.*, vol. 4959, no. 2015, 2014.

[29] I. Albuquerque, A. Tiwari, M. Parent, R. Cassani, J.-F. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "Wauc: A multi-modal database for mental workload assessment under physical activity," *Frontiers in Neuroscience*, vol. 14, p. 1037, 2020.

[30] K. Kühnlenz and B. Kühnlenz, "Motor interference of incongruent motions increases workload in close hri," *Advanced Robotics*, vol. 34, no. 6, pp. 400–406, 2020.

[31] B. system Inc., *Biopac Student Lab PRO Manual*, Biopac system Inc.

[32] Association Aura, "Aura healthcare project website," Last visited 10/2019. [Online]. Available: https://www.aura.healthcare/

[33] V. Villani, B. Capelli, C. Secchi, C. Fantuzzi, and L. Sabattini, "Humans interacting with multi-robot systems: a natural affect-based approach," *Autonomous Robots*, vol. 44, no. 3, pp. 601–616, 2020.

[34] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis," *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 2015.

[35] G. Forte, F. Favieri, and M. Casagrande, "Heart rate variability and cognitive function: a systematic review," *Frontiers in neuroscience*, vol. 13, p. 710, 2019.

[36] L. M. Mazur, P. R. Mosaly, L. M. Hoyle, E. L. Jones, and L. B. Marks, "Subjective and objective quantification of physician's workload and performance during radiation therapy planning tasks," *Practical radiation oncology*, vol. 3, no. 4, pp. e171–e177, 2013.

[37] T. M. Gable, A. L. Kun, B. N. Walker, and R. J. Winton, "Comparing heart rate and pupil size as objective measures of workload in the driving context: initial look," in *Adjunct proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications*, 2015, pp. 20–25.

[38] D. Tao, H. Tan, H. Wang, X. Zhang, X. Qu, and T. Zhang, "A systematic review of physiological measures of mental workload," *International journal of environmental research and public health*, vol. 16, no. 15, p. 2716, 2019.