Augmenting First-Person Swarm Teleoperation with Multisensory Feedback

Anonymous Author(s)

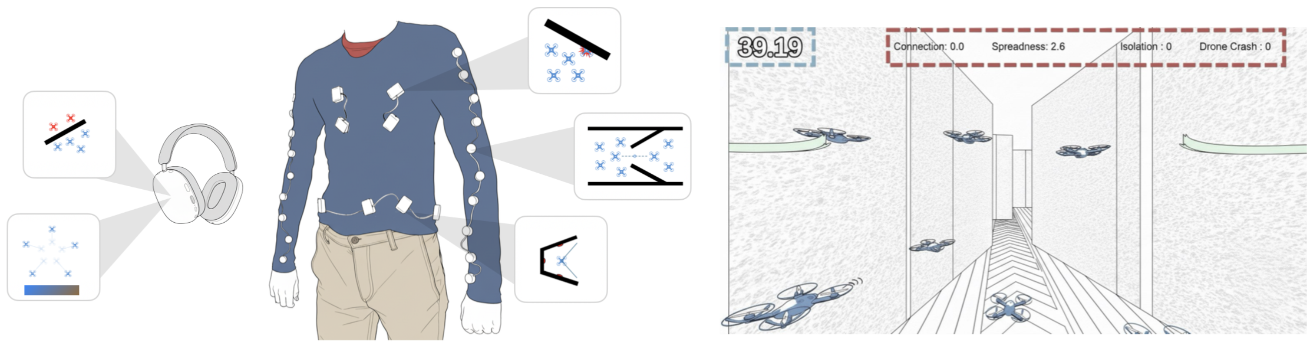


Figure 1: Our proposed multisensory feedback system. Left: A conceptual illustration of our mapping strategy, where critical swarm-state information—such as cohesion, obstacle proximity, and agent isolation—is translated into intuitive haptic patterns on a wearable jacket and spatial audio cues. Right: The teleoperator's first-person view (FPV) from a drone within the swarm, showing the visual interface used for navigation.

Abstract

This paper evaluates how multisensory feedback can improve human teleoperation of aerial swarms. Effective oversight in swarm missions requires user interfaces that convey the swarm state and environment information. While Top-Down View (TDV) naturally provides this, it is impractical in real-world contexts. First-Person View (FPV), by contrast, aligns with the distributed nature of swarms but limits awareness of the swarm state. To address this limitation, we designed a multisensory interface that augments an FPV interface with haptic and audio feedback. Specifically, we propose a novel mapping that compresses the complex swarm state into intuitive cues, using key metrics conveyed through a vibrotactile jacket and spatial audio. Through a user study, we show that while TDV perspectives achieve the best performance, multisensory feedback enabled significant performance increases for FPV perspectives. These results suggest that multisensory feedback can effectively mitigate the limitations of FPV perspectives, improving performance and enabling more reliable human-swarm teleoperation.

CCS Concepts

• **Human-centered computing → User studies**; • **Computer systems organization → External interfaces for robotics**.

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:

Anonymous Author(s). 2018. Augmenting First-Person Swarm Teleoperation with Multisensory Feedback. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Aerial swarms have potential to revolutionize applications such as search and rescue, environmental monitoring, and infrastructure inspection [1]. This has been spurred on by advances in autonomous operations [2]. However, human oversight remains essential for sensitive operations, especially those concerning human lives [3]. This require an effective interface that conveys knowledge of the swarm's internal state and the surrounding environment [4]. A third-person or Top-Down View (TDV) naturally fulfills this and is known to provide high levels of environmental awareness [5]. However, creating this view with a swarm member would result in a single point of failure, negating robustness, one of the key advantages of swarms [6], and real-time reconstructions do not yet provide the same quality of information in real-time [7].

In contrast, First-Person View (FPV) perspectives can leverage the distributed and robust nature of aerial swarms [8]. FPV perspectives from single drones have also been shown to produce higher levels of embodiment that can improve performance [9, 10]. However, FPV perspectives can also degrade Situational Awareness (SA) of the swarm state leaving the operator unaware of the other drones outside the immediate camera feed [11].

In this work, we seek to address this informational gap by investigating how FPV perspectives can be enhanced with auditory and haptic feedback to better convey the swarm state. The key challenge lies in the mismatch between the singular perspective of humans, and the distributed morphology of a swarm, which can disperse through-out it's environment. Additionally, there are no established methods for non-visually conveying the state of many agents to a single teleoperator [12]. To overcome this, we compress the swarm state through a set of metrics derived from the Olfati-Saber algorithm[13].

These metrics provide a holistic representation of the swarm state and are conferred to the teleoperator through a novel mapping to audio and haptic cues. Specifically, we map connectivity (the swarm's cohesion and risk of splitting), inter-agent distance (the distance between neighbours in the swarm), drone isolation (when an agent disconnects from the swarm), predicted paths (expected trajectory of each agent based on user-inputs), and obstacle forces (repulsive forces applied to each agent). By communicating this representation through haptic and audio feedback, we aim to preserve the robustness and practicality of FPV perspectives while restoring awareness of the swarm state, thereby reducing cognitive load and increasing performance in teleoperation tasks.

This work makes several key contributions. First we introduce a novel mapping strategy to translate the swarm state into intuitive haptic and audio cues. We then present a controlled user study to assess the relative benefit of these cues for both FPV and TDV perspectives when piloting a simulated swarm through an obstacle course. A user-centered evaluation then assesses the perceived effectiveness of these metrics. The findings of these studies demonstrate that TDV perspectives achieve higher levels of performance than FPV perspective for swarm teleoperaiton tasks. However, multi-sensory feedback was found to have a significant increase in pilot performance, reducing the time to complete complex tasks while also reducing the number of crashes, drone isolation, and cognitive load. These findings suggest that multi-sensory feedback offers an effective method to augment FPV perspectives, improving task performance and enabling more reliable swarm teleoperation.

2 Prior Work

2.1 Methods for Conveying Emergent Behavior in Swarms

Swarm systems are composed of multiple simple agents that interact locally to produce emergent, coordinated group dynamics. A wide range of swarm algorithms have been developed, often drawing from natural systems such as bird flocks, insect swarms, and schools of fish [14–16]. However, humans struggle to recognize emergent swarm behavior such as fragmentation [17] or flocking[18]. In their review of Human Interaction with Swarms, Kolling et al. [4] highlight swarm visualization and understanding of dynamics as a key area of future work.

Several approaches exist for defining which agents in a swarm are considered neighbors and will interact with each other. One of the most common approaches is the *metric* method, which considers neighbors as those that are within a defined metric distance from each other [14]. Another approach defines neighbors based on their *visibility*, where agents only interact with others that are within line-of-sight, not obscured by obstacles or other agents [15, 16], Fig. 2A. However, it is important to note that due to this, in the presence of obstacles swarms can fragment, with some agents becoming isolated or forming into separate sub-groups, Fig. 2B.

Prior work has sought to improve operator awareness by augmenting the visual and auditory channels. Some visual approaches have used augmented reality to overlay data [19] or provided multiple viewpoints to give operators better spatial context of the swarm [8]. The auditory channel has also been explored, with research showing that audio cues can effectively convey swarm information [20]; however, this work focused on ground-based swarms, which have simpler dynamics than their aerial counterparts. In this work, we adopted the Olfati-Saber flocking algorithm [13] as the foundation for swarm behavior in our simulation, leveraging its interpretable potential-field structure to extract meaningful metrics such as *deviation energy* for connectivity. Our work goes beyond prior implementations by mapping this algorithmic information into intuitive haptic feedback.

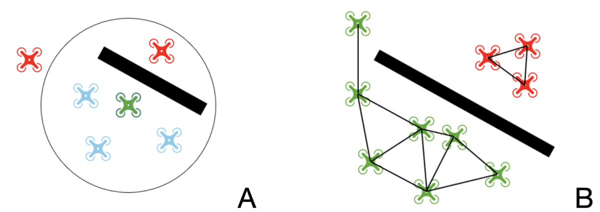


Figure 2: Swarm connectivity and neighbor definitions. (a) A visualization of swarm disconnection, where agents have split into two isolated groups (green and red) due to an obstacle. (b) A depiction of how neighbors are determined for a selected drone (green). Neighbors (blue) must be within the interaction range and not be occluded by obstacles. Non-neighbors are shown in red.

2.2 Haptics for Swarm Teleoperation

When a human teleoperates a swarm, maintaining situation awareness is essential for making informed decisions based on the swarm's state [21]. However, depending on the level of autonomy and mission complexity, visual displays can lead to cognitive overload [22]. Prior research shows that haptic feedback enhances situational awareness, particularly under high mental workload [12]. For instance, in aviation, haptics have been used to offload cognitive demands from overburdened visual and auditory channels to improve pilot awareness [23]. Similarly, the wearable Tactile Situation Awareness System (TSAS) used torso vibrations to help helicopter pilots hover in low-visibility and demonstrated improved response time and reduced cognitive load [24, 25]. Similar benefits of haptics have been found in drone teleoperation, where haptic force feedback improves control and reduces crashes [26]. Other work has focused on wearable interfaces, such as the FlyJacket, to provide more natural and intuitive drone control through haptics [27].

Despite these advances, gaps remain—particularly in scenarios where a single operator supervises a group or swarm of robots. Most existing research focuses on one-to-one control or tightly coupled human–automation systems [28]. In contrast, "many-to-one" scenarios, where a human oversees multiple autonomous agents, are less explored [29]. Similarly, Haas et al. (2009) developed a multimodal interface for a soldier-swarm convoy mission that relied on simple, low-level cues [30]. In another approach, McDonald et al. (2017) explored using a non-wearable, desktop haptic device to let a single operator control a multi-robot team as a deformable shape [31]. A key limitation of this method, however, is that it only communicates the swarm's overall shape. These examples highlight how existing haptic systems for swarms tend to provide feedback on task-related metrics (like basic alerts or physical shape) rather than conveying high-level information about swarm state and emergent swarm behavior. Yet, for effective human–swarm teaming, shared understanding and intuitive communication are essential [32].

Building on these insights, our study extends the application of haptic feedback into the domain of swarm teleoperation, specifically addressing the underexplored "many-to-one" challenge. Our work uniquely focuses on swarm-level feedback rather than single-agent feedback. To address this challenge, we developed a haptic jacket that provides spatially distributed feedback across the operator's torso. This wearable system translates critical, swarm-wide behaviors, such as disconnection, agent isolation, and overall connectivity, into distinct tactile patterns. This approach enables operators to perceive emergent swarm dynamics through haptics, which is particularly beneficial when their visual channel is saturated, as is common in FPV teleoperation.

3 Methods

3.1 User Study

We conducted a between-subjects study with 40 participants (28.2 ± 8.8 years, 26 male, 14 female). Participants controlled a simulated aerial swarm through an obstacle course in Unity using an *Xbox controller*. They were randomly assigned to one of two conditions: (1) *Haptic & Audio Feedback*, receiving spatial audio cues (50–65 dB) and vibrations from a haptic jacket, or (2) *Visual Only*, where they wore the devices but received no feedback beyond visual information on the screen. All participants were required to wear a Medium or Large t-shirt for device compatibility; those with hearing disorders were excluded. The study was approved by MIT IRB (2501001530).

Two viewpoints were tested: *Top-Down View (TDV)* and *First-Person View (FPV)*, with presentation order counterbalanced. Each participant completed two tasks—(1) basic navigation and (2) navigation with collectible objects—under both viewpoints, repeated twice, for a total of 8 trials. For each crash, a 5-second penalty was added. Completion required that all drones stayed connected and all stars were collected. After each condition, participants completed questionnaires including NASA-TLX [33], demographic and experience questions, and feedback on the feedback modalities.

In total, 290 flights (out of 320) were retained after excluding trials with outlier completion times or crash counts ($>2\sigma$ per task-viewpoint combination). During each trial, we recorded swarm state (positions, velocities, neighbors, forces), camera orientation, and user control inputs.

3.2 Swarm Model and Status Metrics

The aerial swarm consisted of 20 simulated drones, each with a maximum velocity of 2.1 m/s, maximum acceleration of 1.7 m/s^2, and a field of view (FOV) of 84°. Swarm coordination was implemented using the Olfati-Saber algorithm [13], which regulates inter-agent distances and avoids obstacles through potential fields. Neighboring drones were defined as those within 1.3× *the desired separation radius* and not occluded by obstacles. While the algorithm promotes a unified swarm, occlusions could fragment the network into sub-groups. As per the Olfati-Saber algorithm, obstacles are modeled as virtual agents exerting repulsive forces, this facilitates navigation through cluttered environments, but can cause swarm fragmentation or lead to agents being trapped in local minima around non-convex obstacles.

3.2.1 Operator Perspectives. Two control perspectives were implemented to evaluate the effectivness of multisensory feedback with different visual perspectives. In the *FPV* condition, participants embodied a single drone that acted as the swarm leader. Unlike follower drones, this leader was not subject to inter-agent forces, which facilitated maneuvering but increased collision risk. Its influence on followers was weighted five times higher than that of a regular agent. Haptic feedback reflected obstacle forces acting on the leader drone, filtered according to their directional relevance: only forces within 45° of the acceleration vector were conveyed, expanding to 72.5° when obstacle proximity increased. In the *TDV* condition, participants controlled the swarm as a cohesive unit by setting a reference velocity applied to all members of the currently selected network. Users could switch networks via controller inputs. Obstacle forces were aggregated across drones and clustered using DBSCAN [34]. A haptic warning was generated when a directional cluster represented at least one-quarter of the swarm.

3.2.2 Status Metrics. To maintain situational awareness, the interface displayed real-time swarm indicators (Fig. 3):

- **Spreadness.** Average inter-agent distance, adjustable by the operator (bounded between 1–5 m).
- **Connection.** Likelihood of swarm fragmentation, derived from the normalized deviation energy $\tilde{E}(q)$ [13]. After Min–Max normalization ($E_{\min} = 0.3$, $E_{\max} = 0.7$), the metric reached 1 when the swarm split into separate networks.
- **Trajectory Prediction.** Each drone's future trajectory was simulated 1 s ahead (25 steps, $\Delta t = 0.04$ s), updated at ∼13 Hz. Predicted collisions were highlighted in red to signal imminent risk.
- **Crashes and Isolation.** Counters displayed the number of drones lost to collisions or disconnected from the controlled network.
- **Obstacle Forces.** Force vectors computed from obstacle avoidance were mapped to haptic cues on the torso (see Table 1).

3.3 Haptic and Audio Feedback

The haptic jacket developed for this project incorporates 32 vibrotactile actuators, comprising Linear Resonant Actuators (LRA) and Voice Coil Actuators (VCA) using the *VibraForge* toolkit[35]. Ten LRAs were positioned along the length of each arm, eight VCAs were arranged in a ring around the waist, and four VCAs were placed on the torso, Fig. 4. To compliment this haptic feedback, spatial audio was delivered through headphones. The torso and waist actuators were assigned to events affecting the collective swarm state, such as collisions or nearby obstacles, while the actuators on

Figure 3: Visual interfaces. (A) Top-Down View (TDV) of the swarm in the Pillar zone. Obstacles are in black; red arrows mark the path. (B) First-Person View (FPV) from an embodied drone with propellers visible in green. Metrics and elapsed time are displayed in the overlays.

the arms were used to convey disconnection from the main swarm body. This configuration was selected to provide the operator with distinct body regions and senses to easily identify between localized cues.

This mapping strategy was informed by prior research on spatially distributed vibrotactile cues, which has shown that such designs improve situational awareness[36]. To further enhance perception, vibration intensities were adjusted from *VibraForge*'s default electrical linearity to a perceptually linear profile, following the Weber-Fechner law [37]. The system was integrated with *Unity*, achieving an end-to-end latency of 20 ms.

When the inter-agent distance between some neighbors in the swarm increased, suggesting fragmentation was imminent, users experienced vibrations propagating down their arms to mimic the sensation of something separating from the swarm. If separation did occur and a agent was isolated from the swarm, this was indicated by an audio cue and a change of color in the isolated agent. If a crash was imminent, determined by the trajectory prediction, the controller would start to vibrate. If the user did not correct and a crash occurred then the user would experience a strong vibration on their torso. If the user modified the inter-agent distance of the swarm and thus its overall size, this was accompanied by an changing frequency audio cue from brown to blue noise. The actuators around the waist were reserved for conveying environment information through obstacle repulsive forces. A complete description of these mappings is provided in Table 1.

Figure 4: Locations of the haptic actuators used in the experiment 10 LRA on each arm which represent swarm connection (red boxes), 8 VCA around the waist which represent the obstacle forces, (blue boxes) 4 VCA on the torso (green boxes) which represent drone crashes.

3.4 Obstacle Course

The obstacle course was designed to assess the operator's ability to control swarm trajectory, regulate spread, and maintain cohesion while navigating through obstacles in a *3.5 m* wide corridor (Fig. 5). It began with an *Entry Gate* featuring two narrow gaps, one inverted, which created local minima in the obstacle avoidance potential field and required reducing swarm spread to pass. A subsequent *Zigzag* section tested maneuverability and rapid adjustments, followed by the *Pillars*, 24 evenly spaced obstacles in an 8×3 array that slowed the swarm and demanded heightened situational awareness. The *Diamond* section encouraged the operator to deliberately split and re-merge the swarm, and the course concluded with an *End Gate* that required precise control of spread for successful passage. This structured sequence provided a comprehensive assessment of swarm control capabilities.

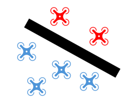
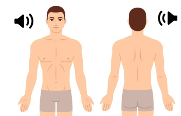
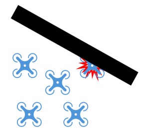
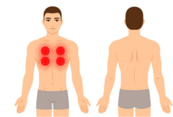
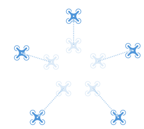
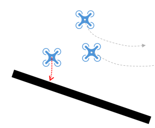
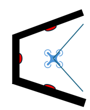
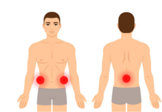
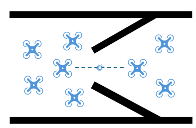
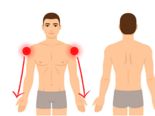
4 Results

4.1 Haptics vs Non-Haptics

The most significant effect of Haptics was observed in the FPV configuration during the collectible task. The task was better performed across all the metrics: reduction in crashes, reduction in time completions time and reduced time having drone isolation time.

Quantitatively, the number of crashes in the FPV collectible task dropped from 2.58 ± 1.73 without haptics to 1.32 ± 1.58 with haptics ($p < 0.01$). Isolation time also improved, decreasing from 10.08 ±

Table 1: Summary of the Mapping of the Different Features with Extra Information

Feature	Extra Information	Drone configuration	Haptic-Audio
Isolated Drone(s)	Spatial audio emits sound from the drone's location; drone's color changes for visual identification.		
Drone Crash	Vibration on the torso if any crashes happen. Small firework animation at the crash position.		
Change of Spreadness	Audio transitions from brown to blue noise based on spreadness value during user adjustments.		Brown-Blue Noise
Trajectory Prediction	Trajectory lines turn red and trigger controller vibration. Intensity increases as predicted crash gets closer.		
Obstacles Force(s)	FPV: Feel nearby obstacles around the embodied drone. TDV: Feel swarm-level force cues. Threshold set to 25% of swarm density.		
Connection	Represents the likelihood of swarm splitting or isolated drones emerging; Vibration going from shoulder (0.1) to hand (1)		

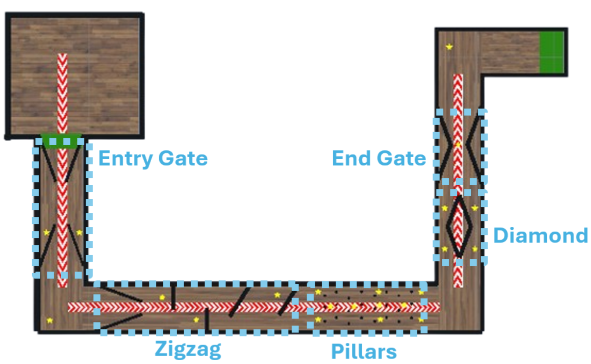


Figure 5: Obstacle course with the different sections labelled. Red arrows indicated the direction to complete the obstacle course. Stars shown in yellow were only available in the Collectible task.

11.79 seconds to 4.67 ± 7.79 seconds ($p < 0.05$). The completion time was likewise reduced from 177.49 ± 50.53 seconds without haptics to 152.55 ± 51.22 seconds with haptics ($p < 0.05$).

These improvements suggest that haptic feedback contributed to more stable and cautious navigation in the more demanding FPV condition. In contrast, TDV showed minimal gains from haptics, possibly due to the enhanced spatial awareness TDV already provides.

4.2 NASA-TLX Workload

In terms of perceived workload, haptic feedback significantly reduced mental workload in most conditions. For the *collectibles task* in FPV, participants reported a workload of 10.83 ± 3.76 without haptics and 8.04 ± 3.87 with haptics ($p < 0.01$). Under TDV, workload was further reduced, with 8.13 ± 3.51 in the non-haptic condition and just 5.69 ± 3.12 with haptics ($p < 0.01$) as shown in Fig. 7

Similarly, in the obstacle task, FPV users experienced 8.60 ± 2.71 workload without haptics and 6.86 ± 3.79 with haptics ($p < 0.05$). In TDV, the difference was not statistically significant ($p > 0.05$), with workload values of 6.79 ± 2.99 (NH) and 5.54 ± 3.50 (H).

Furthermore, when comparing task types, the Collectible task elicited higher ratings in mental demand, frustration, and perceived performance (inverted success scale) than the Obstacle task. These results indicate that participants perceived the Collectible task as more cognitively demanding and complex (Fig. 7).

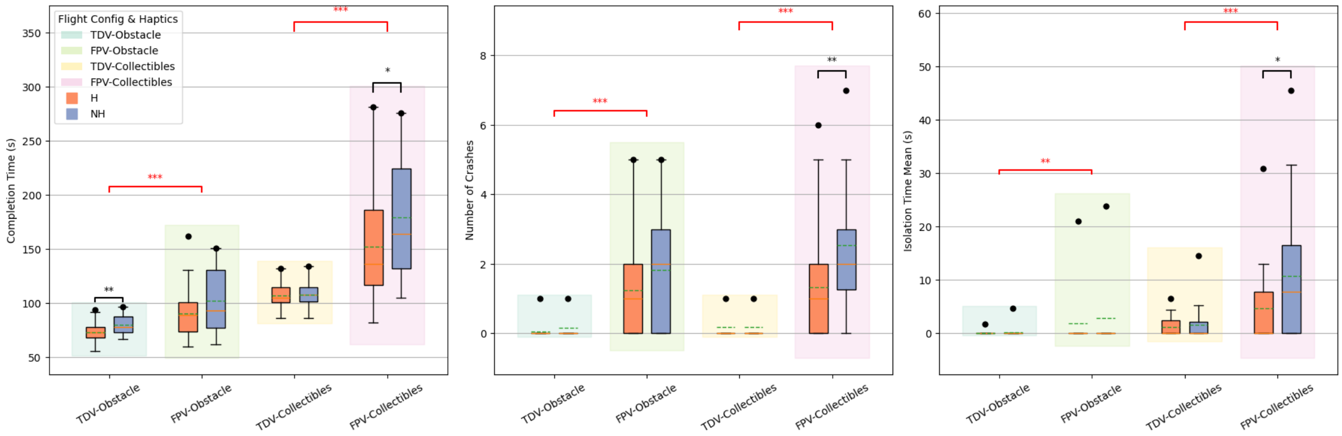


Figure 6: TDV significantly reduced completion time and crashes in both tasks ($p < 0.001$), and reduced isolation time in the Collectible task ($p < 0.001$) and Obstacle task ($p < 0.05$). Haptic feedback improved completion time in the TDV-Obstacle task ($p < 0.01$) and FPV-Collectible task ($p < 0.05$), reduced crashes in the FPV-Collectible task ($p < 0.01$), and decreased isolation time ($p < 0.05$). The black dot represents the max value for each configuration.

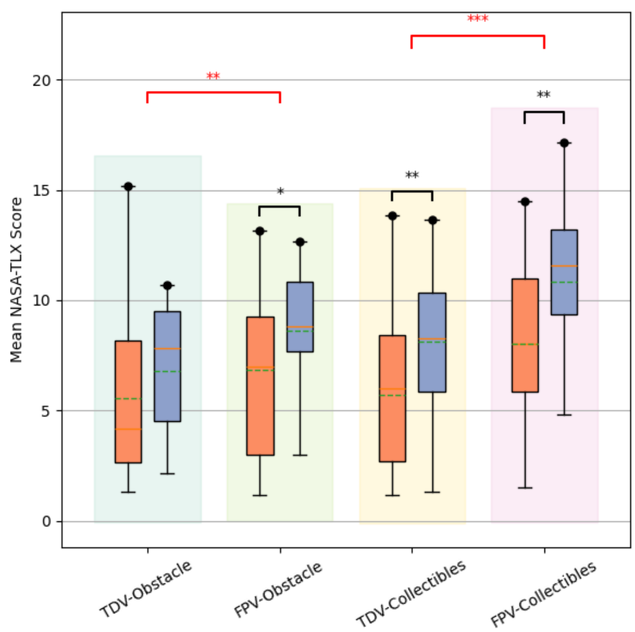


Figure 7: TDV significantly reduced workload according to NASA-TLX for Obstacle ($p < 0.01$) and collectibles ($p < 0.001$). Haptic reduces workload for FPV-Obstacle ($p < 0.05$), TDV-Collectibles and FPV-Collectibles ($p < 0.01$). NASA-TLX score of 0 is good and 100 bad.The black dot represents the max value for each configuration.

4.3 TDV vs. FPV

Table 2 summarizes the performance differences between the TDV and FPV across both tasks. Overall, TDV demonstrated superior performance: participants completed tasks significantly faster, experienced fewer crashes, and reduced periods of drone isolation. These trends are consistent across both the obstacle and collectible tasks (see also Fig. 6).

Table 2: Comparison of FPV and TDV performance (mean ± SD).

Measure / Task	FPV	TDV	Signif.
Completion Time (s)			
Obstacle	96.40 ± 26.75	76.41 ± 9.41	$p < 0.001$
Collectible	165.17 ± 52.61	107.38 ± 11.31	$p < 0.001$
Crashes / Flight			
Obstacle	1.53 ± 1.57	0.10 ± 0.30	$p < 0.001$
Collectible	1.90 ± 1.77	0.17 ± 0.38	$p < 0.001$
Isolation Time (s)			
Obstacle	2.32 ± 5.77	0.09 ± 0.58	$p < 0.01$
Collectible	7.54 ± 10.34	1.34 ± 2.55	$p < 0.001$
NASA-TLX			
Obstacle	7.75 ± 3.37	6.12 ± 3.31	$p < 0.01$
Collectible	9.37 ± 4.04	6.89 ± 3.51	$p < 0.001$

5 Discussion

This work has demonstrated that Haptic feedback further enhanced performance in FPV—particularly in the more demanding Collectible task—by reducing crashes, isolation time, and mental workload, while offering moderate benefits in TDV where situational awareness was already high.

This study explored how control perspectives (FPV and TDV) and sensory feedback modalities (visual, haptic, and audio) influence performance and perceived workload in swarm teleoperation. The results indicate a clear advantage of the TDV interface, which enabled participants to complete tasks faster, with fewer crashes and reduced drone isolation. However, haptic feedback demonstrated a notable impact, particularly in the more cognitively demanding FPV condition.

5.1 Haptic Feedback Enhances Performance Under Cognitive Load

While TDV consistently outperformed FPV, the haptic jacket proved most effective in the FPV collectible task, which posed the greatest mental demand. Participants with haptic feedback experienced significantly fewer crashes, shorter isolation periods, and improved completion times. These findings suggest that haptics can offload cognitive burden by conveying critical swarm state information—such as proximity to obstacles or swarm disconnection—through peripheral sensory channels. This allows users to redirect their visual attention to higher-level objectives, such as navigation and object collection. NASA-TLX ratings further support this, showing lower mental demand and frustration in haptic conditions.

In contrast, the impact of haptics in TDV was limited. Since TDV already provides a global overview of swarm dynamics, the additional haptic cues offered minimal performance gains. This implies that haptics are most useful when visual resources are saturated or when operating in first-person conditions where global context is lacking.

5.2 Trust and Control Burden in FPV

Despite the immersive control offered by FPV, participants struggled to maintain swarm cohesion and avoid collisions. This was largely due to the embodied drone being exempt from swarm forces, placing the full responsibility of guidance and obstacle avoidance on the operator. As a result, participants often engaged in micromanagement, spending significant cognitive effort "babysitting" the swarm rather than focusing on mission objectives.

This suggests the need for increased autonomy in swarm behavior during FPV teleoperation. Future approaches could incorporate mechanisms such as pheromone trails or dynamic leader-following, [38] enabling the swarm to self-organize around the leader without constant user correction. Such mechanisms would allow the operator to focus on exploration or inspection tasks, trusting the swarm to maintain cohesion and avoid obstacles independently.

5.3 Study Limitations and Future Work

Several limitations and avenues for future development were identified. While the integration of haptic feedback improved both performance and workload, participants noted that its spatial resolution and intensity discrimination were limited. These shortcomings occasionally made directional cues less clear, suggesting that future designs should consider more refined actuator placement and improved intensity scaling to enhance precision.

Another limitation concerned the relevance of the chosen viewpoints. The tasks in this study favored the top-down perspective because no occlusions or confined spaces were present. In real-world scenarios such as tunnel inspection or indoor navigation, however, the first-person perspective may provide crucial advantages. Beyond viewpoint selection, practical deployment introduces additional constraints. Real-time swarm control depends on low-latency, high-bandwidth communication, yet wireless links are often subject to interference and delay. Even small disruptions can destabilize decentralized algorithms like Olfati-Saber, leading to fragmentation. Localization presents another challenge: GPS is unreliable in cluttered or indoor environments, motion capture is impractical at scale, and onboard relative sensing methods are vulnerable to drift. These factors complicate the transfer of laboratory findings to operational conditions [39, 40].

A further issue arose in the top-down condition, where participants occasionally received haptic signals generated by the swarm's autonomous obstacle-avoidance responses. These cues were a normal consequence of the system's behavior but were often interpreted as urgent warnings in contexts where user intervention was not possible or necessary. The resulting ambiguity sometimes led to frustration, and in a few cases caused participants to disregard haptic feedback altogether.

Finally, the study did not fully explore the potential of alternative haptic modalities. Additional swarm-level information—such as Olfati-Saber forces, density variations, or contraction cues—could prove valuable for enhancing situational awareness. Early attempts to directly map these signals into the haptic channel revealed the risk of overwhelming operators with excessive information, particularly those without prior training. More selective encoding strategies, such as emphasizing critical contraction events at narrow passages, may offer a path forward while avoiding cognitive overload.

6 Conclusion

The findings from this study demonstrate that multisensory feedback produces significant changes in operator performance and workload during FPV teleoperation of aerial swarms. While TDV consistently achieved higher absolute performance, FPV scenarios showed significant benefits from haptic and audio feedback, significantly reducing the task completion time, number of crashes, and measured workload. These effects were most pronounced in more complex tasks, where FPV perspectives suffered from limited environmental awareness. Together, these results demonstrate that multisensory feedback can be an effective method for offsetting the limited information available in FPV teleoperation of complex systems, approaching the performance of impractical TDV interfaces. This highlights the potential of multisensory feedback to enable effective and safe teleoperation of aerial swarms in real-world environments.

6.0.1 LLM Disclosure. LLMs were used to convert photos into lineart for the teaser figure.

References

[1] Marco Dorigo, Guy Theraulaz, and Vito Trianni. Swarm Robotics: Past, Present, and Future [Point of View]. *Proceedings of the IEEE*, 109(7):1152–1165, July 2021.

[2] Xin Zhou, Xiangyong Wen, Zhepei Wang, Yuman Gao, Haojia Li, Qianhao Wang, Tiankai Yang, Haojian Lu, Yanjun Cao, Chao Xu, and Fei Gao. Swarm of micro flying robots in the wild. *Science Robotics*, 7(66):eabm5954, May 2022. Publisher: American Association for the Advancement of Science.

[3] MD Moniruzzaman, Alexander Rassau, Douglas Chai, and Syed Mohammed Shamsul Islam. Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey. *Robotics and Autonomous Systems*, 150:103973, April 2022.

[4] Andreas Kolling, Phillip Walker, Nilanjan Chakraborty, Katia Sycara, and Michael Lewis. Human Interaction With Robot Swarms: A Survey. *IEEE Transactions on Human-Machine Systems*, 46(1):9–26, February 2016.

[5] Valeria I. Petkova, Mehrnoush Khoshnevis, and H. Henrik Ehrsson. The Perspective Matters! Multisensory Integration in Ego-Centric Reference Frames Determines Full-Body Ownership. *Frontiers in Psychology*, 2, 2011.
[6] Ryotaro Temma, Kazuki Takashima, Kazuyuki Fujita, Koh Sueda, and Yoshifumi Kitamura. Third-Person Piloting: Increasing Situational Awareness using a Spatially Coupled Second Drone. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pages 507–519, New Orleans LA USA, October 2019. ACM.
[7] Zhihao Jia, Bing Wang, and Changhao Chen. Drone-NeRF: Efficient NeRF based 3D scene reconstruction for large-scale drone survey. *Image and Vision Computing*, 143:104920, March 2024.
[8] Benjamin Jarvis, Charbel Toumieh, and Dario Floreano. First-Person View Interfaces for Teleoperation of Aerial Swarms. *IEEE Robotics and Automation Letters*, pages 1–8, 2025.
[9] Matteo Macchini, Manana Lortkipanidze, Fabrizio Schiano, and Dario Floreano. The impact of virtual reality and viewpoints in body motion based drone teleoperation. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pages 511–518, 2021.
[10] Maximilian Jablonowski. Beyond drone vision: the embodied telepresence of first-person-view drone flight. *The Senses and Society*, 15(3):344–358, 2020.
[11] Riley Jennifer M. Situation awareness in human-robot interaction: Challenges and user interface requirements.
[12] M. S. Prewett, L. R. Elliott, A. G. Walvoord, and M. D. Coovert. A Meta-Analysis of Vibrotactile and Visual Information Displays for Improving Task Performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(1):123–132, January 2012. 71 citations (Crossref) [2024-11-14].
[13] R. Olfati-Saber. Flocking for multi-agent dynamic systems: algorithms and theory. *IEEE Transactions on Automatic Control*, 51(3):401–420, March 2006. Conference Name: IEEE Transactions on Automatic Control.
[14] Craig W Reynolds. Flocks, Herds, and Schools: A Distributed Behavioral Model.
[15] Sara Brin Rosenthal, Colin R. Twomey, Andrew T. Hartnett, Hai Shan Wu, and Iain D. Couzin. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112(15):4690–4695, April 2015. Publisher: Proceedings of the National Academy of Sciences.
[16] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *Proceedings of the National Academy of Sciences*, 105(4):1232–1237, January 2008. Publisher: Proceedings of the National Academy of Sciences.
[17] Aymeric Hénard, Etienne Peillard, Jérémy Rivière, Sébastien Kubicki, and Gilles Coppin. Human perception of swarm fragmentation. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 250–258, Boulder CO USA, March 2024. ACM.
[18] Phillip Walker, Michael Lewis, and Katia Sycara. Characterizing human perception of emergent swarm behaviors. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 002436–002441, October 2016.
[19] Aymeric Hénard, Étienne Peillard, Jérémy Rivière, Sébastien Kubicki, and Gilles Coppin. Towards Augmented Reality Support for Swarm Monitoring: Evaluating Visual Cues to Prevent Fragmentation. In *24th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2025)*, Daejon, South Korea, October 2025.
[20] James McLurkin. Speaking Swarmish: Human-Robot Interface Design for Large Swarms of Autonomous Mobile Robots.
[21] Aya Hussein and Hussein Abbass. Mixed Initiative Systems for Human-Swarm Interaction: Opportunities and Challenges. In *2018 2nd Annual Systems Modelling Conference (SMC)*, pages 1–8, October 2018.
[22] Christopher D. Wickens. Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, June 2008. 1261 citations (Crossref) [2024-11-14].
[23] Sam Chin, Noam Eisen, Brittany Bishop, Andrew Liu, Katya Arquilla, and Joe Paradiso. Evaluating Head-Mounted Visual-Haptic Displays for Recovery from Unusual Flight Attitudes under Normal and Visually-Degraded Conditions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 68(1):1473–1476, September 2024.
[24] Angus H. Rupert, Ben D. Lawson, and Jared E. Basso. Tactile Situation Awareness System: Recent Developments for Aviation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):722–726, September 2016.
[25] J. Christopher Brill, Ben D. Lawson, and Angus Rupert. Tactile Situation Awareness System (TSAS) as a Compensatory Aid for Sensory Loss. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1):1028–1032, September 2014. 1 citations (Crossref) [2024-02-23].
[26] Bingjian Huang, Zhecheng Wang, Qilong Cheng, Siyi Ren, Hanfeng Cai, Antonio Alvarez Valdivia, Karthik Mahadevan, and Daniel Wigdor. AeroHaptix: A Wearable Vibrotactile Feedback System for Enhancing Collision Avoidance in UAV Teleoperation, 2024. Version Number: 2.
[27] Carine Rognon, Stefano Mintchev, Fabio Dell'Agnola, Alexandre Cherpillod, David Atienza, and Dario Floreano. FlyJacket: An Upper Body Soft Exoskeleton for Immersive Drone Control. *IEEE Robotics and Automation Letters*, 3(3):2362–2369, July 2018. Conference Name: IEEE Robotics and Automation Letters.
[28] Abhinav Dahiya, Alexander M. Aroyo, Kerstin Dautenhahn, and Stephen L. Smith. A survey of multi-agent human–robot interaction systems. *Robotics and Autonomous Systems*, 161:104335, 2023.
[29] Steven Nunnally, Phillip Walker, Mike Lewis, Nilanjan Chakraborty, and Katia Sycara. Using Haptic Feedback in Human Robotic Swarms Interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1):1047–1051, September 2013.
[30] Ellen Haas, Susan Hill, Christopher Stachowiak, and Mary Anne Fields. Designing and Evaluating a Multimodal Interface for Soldier-Swarm Interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 53(4):259–263, October 2009. Publisher: SAGE Publications Inc.
[31] Samuel J. McDonald, Mark B. Colton, C. Kristopher Alder, and Michael A. Goodrich. Haptic Shape-Based Management of Robot Teams in Cordon and Patrol. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pages 380–388, Vienna Austria, March 2017. ACM.
[32] Aya Hussein, Leo Ghignone, Tung Nguyen, Nima Salimi, Hung Nguyen, Min Wang, and Hussein A. Abbass. Towards Bi-Directional Communication in Human-Swarm Teaming: A Survey, March 2018. arXiv:1803.03093 [cs].
[33] Sandra G. Hart and Lowell E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Peter A. Hancock and Najmedin Meshkati, editors, *Advances in Psychology*, volume 52 of *Human Mental Workload*, pages 139–183. North-Holland, January 1988.
[34] Martin Ester, Hans-Peter Kriegel, and Xiaowei Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.
[35] Bingjian Huang, Siyi Ren, Yuewen Luo, Qilong Cheng, Hanfeng Cai, Yeqi Sang, Mauricio Sousa, Paul H. Dietz, and Daniel Wigdor. VibraForge: A Scalable Prototyping Toolkit For Creating Spatialized Vibrotactile Feedback Systems, February 2025. arXiv:2409.17420 [cs].
[36] Hong Z. Tan, Seungmoon Choi, Frances W. Y. Lau, and Freddy Abnousi. Methodology for Maximizing Information Transmission of Haptic Devices: A Survey. *Proceedings of the IEEE*, 108(6):945–965, June 2020. 28 citations (Crossref) [2024-02-23].
[37] Miyuki Morioka and Michael J. Griffin. Magnitude-dependence of equivalent comfort contours for fore-and-aft, lateral and vertical hand-transmitted vibration. *Journal of Sound and Vibration*, 295(3):633–648, August 2006.
[38] Caroline Harriott, Adriane Seiffert, Sean Hayes, and Julie Adams. Biologically-Inspired Human-Swarm Interaction Metrics. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58:1471–1475, October 2014.
[39] Luka Kramarić, Niko Jelušić, Tomislav Radišić, and Mario Muštra. A Comprehensive Survey on Short-Distance Localization of UAVs. *Drones*, 9(3):188, March 2025. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.
[40] Yunes Alqudsi and Murat Makaraci. UAV swarms: research, challenges, and future directions. *Journal of Engineering and Applied Science*, 72(1):12, January 2025.

Received 30 September 2025