Usable and Interpretable Human-Swarm Visualisations: A User Evaluation Study

Jediah R. Clark, Mohammad Divband Soorati & Sarvapali D. Ramchurn

Agents, Interaction and Complexity Research Group, University of Southampton, UK

ABSTRACT

In the field of swarm robotics, human factors research is deficient. Human-swarm interaction, the process in which humans work with collections of unmanned and autonomous aerial vehicles (UAVs) to conduct activities, require an operator to track and manage vast numbers of robotic agents. Due to increasing complexity in communication requirements, visualising a swarm is challenging and requires usable visualisation methods that do not detract from user-interpretability and swarm transparency. This study explores two initial swarm-displays for swarm geographical coverage and density – an individual point display, and a heatmap display. 100 users viewed a simulation of each display, provided subjective usability and acceptance ratings, and gave their display preferences related to eight contextual factors related to time constraints, swarm size, communication-constrained environments, displaying coverage and motion, time-criticality, error detection and transparency. It was found that heatmaps improved usability and acceptance, and were preferred for displaying coverage and motion, and when UAV numbers are high or when time is limited. Individual point displays, although being overall less usable, acceptable, and preferable, were still deemed as being a useful tool for detecting errors in swarm operation. It is therefore concluded that aggregated data displays are a promising display method for visualising swarm coverage and density. The study is limited due to participants not being able to interact with the displays, therefore further research is required to further test the effects reported in this study.

KEYWORDS

Human-swarm interaction, interface design, usability

Introduction

Autonomous robotic swarms - collections of unmanned autonomous vehicles (UVs) working together with a unifying purpose - promise to enhance remote operations where human activity is restricted (Schranz et al., 2020). Individual agents within these swarms have access to local information only, requiring a network of communication to build a global picture of the situation. Robotic swarms are currently or are expected to be used in a variety of domains such as agriculture, search & rescue, warehouse operation, military scenarios, environmental monitoring, and space exploration (Schranz et al., 2020). Further applications of swarm robotics could also aid in agricultural pest-control and medical applications such as targeted drug delivery (Dorigo et al., 2020).

There is high potential for a distributed multi-agent network to autonomously learn, adapt and conduct tasks dynamically in-line with an overarching objective, allowing robotic swarms to be amongst one of the most world changing technologies currently in development (Patil et al., 2013). Noted by Saffre et al. (2021) human factors enquiries into robotic swarms are currently in their infancy.

Although sharing many similarities to their single and multi-UV counterparts, human interaction requirements for robotic swarms differ due to several factors. For example, robotic swarms represent collections of lower complexity agents at a greater number and can reach numbers into the thousands (Roundtree et al., 2018; 2019). The move from individually operated multi-UV platforms to those of greater numbers of autonomous agents is challenging, as the ability for a human operator to manage each drone individually becomes less feasible as the number of UVs increase (Olsen & Wood, 2004). In-line with the 'fan out hypothesis', when being controlled individually, the ability to control a collection of UVs plateaus at between 4 and 12 agents – dependent on task parameters, how autonomous and how reliable the system is (Olsen & Wood, 2004). For swarms in greater numbers, aggregated dataflow, and large-scale supervision of swarm behaviour is required to maintain operational standards. Due to this, a human-operator is met with the challenge of trading off attention to individual drones in favour of overall swarm performance (Kolling et al., 2015). Under these circumstances, adding additional agents to the swarm does not contribute much additional cognitive complexity, as the human-operator is able to treat the swarm as a single entity (unless splitting the swarm up to be assigned to different tasks/locations).

Display methods, therefore, require an aggregation of data to communicate overall swarm state, or risk greater levels of complexity. As noted by many key automation researchers, poor feedback on autonomous agent behaviour and intentions can lead to increased errors (Norman, 1990) and a reduction in trust (Lee & See, 2004; McBride et al., 2014). Due to the move towards more aggregated, overall display methods of swarm behaviour and task processing, a trade-off should be made between trust, workload, and usability when swarms of collective individuals are being operated. Operators who wish to manage the tasks or analyse data being communicated by the swarm will need quick and efficient ways of building and/or applying pre-existing mental models about swarm state and current performance.

The challenge for human-swarm specialists is the relatively low availability of published works in human-swarm interaction research, and the limited capacity to test swarm robotics at high scales. This is partly due to the relatively early stage of artificial intelligence and machine learning development, as well as financial cost (Meshcheryakov et al., 2019; Outay et al., 2020). Further, there are few domains that can be learned from, as autonomous agent research has typically focused on singular physical entities within human-agent interaction (e.g., autonomous automotive vehicles, automation in aviation; Billings, 2018; Clark et al., 2019).

A major challenge for the development of swarm representations is how to display the swarm's coverage over geographical space (Kolling et al., 2013; Vasile et al., 2011). Due to multiple agents being present to confirm or deny the information model of the swarm, UV density is indicative of search efficiency and accuracy (Hamann & Reina, 2021). It is therefore critical that a human operator made aware of where the swarm is operating, as well as how many agents are operating in each location. This is to ensure that the swarm can be allocated across search areas more optimally and allow the system to make use of the human-operator's ability to process contextual factors related to the task and situational parameters (Hussein & Abbass, 2018). Location data should be provided in an accessible and usable format, so that operators are given a global picture that is both usable and does not require excessive attentional resources to process. It follows that aggregated data across agents may provide benefits at all levels of autonomy where a human-operator is present, whether the human operator is a supervisor (i.e., managing high-level plans and objectives) or an operator at a micro-level (i.e., issuing lower-level commands related to swarm movement and tasking; see. Hussein & Abbass, 2018).

To explore how human-swarm interaction may be enhanced or hindered by favouring aggregated density displays, this study examines the potential trade-offs for displaying individual Unmanned Aerial Vehicle (UAV) displays, representing each drone individually and aggregating UAV position and density via a heatmap based on density and position relative to their neighbours. It is hypothesised that heatmap displays would favour greater scalability of UAV numbers due to the fan-out hypothesis (Olsen & Wood, 2004), and therefore higher usability. However, this may come at the cost of reduced trust or transparency (due to less information; Norman, 1990; Lee & See, 2004) and result in lower user acceptance and decreased preference to use these displays. Additionally, contextual factors such as swarm size, time-pressure, communication-constrained environments, and the detection of errors may influence which display users prefer.

Method

Design

Two displays were developed to represent 50 UAVs in 2D space (see figure 1.). The left display in figure 1 shows individual UAVs represented by moving black dots. The right heatmap display shows a collective heatmap of UAV density (purple = lowest density, yellow = high density). Every agent update within each pixel-space was coded as dense (yellow), which decreased over-time as the simulation progressed (towards purple). The heatmap was presented at a medium reading threshold, meaning that a collection of closely neighbouring UAVs would be required to appear on the heatmap, defined by this threshold. Further developments of this display will allow users to manually select thresholds for density to adapt their displays to the task they are conducting. Displays representing swarm coverage were recorded in video format lasting 35 seconds. To mimic real-time agent updates in-line with connectivity and swarm dynamics, location data were provided only if the agent was able to send data to the operator. This was dependent on connectivity with the operator's base-station and with other agents in the swarm. This was fixed for both conditions.

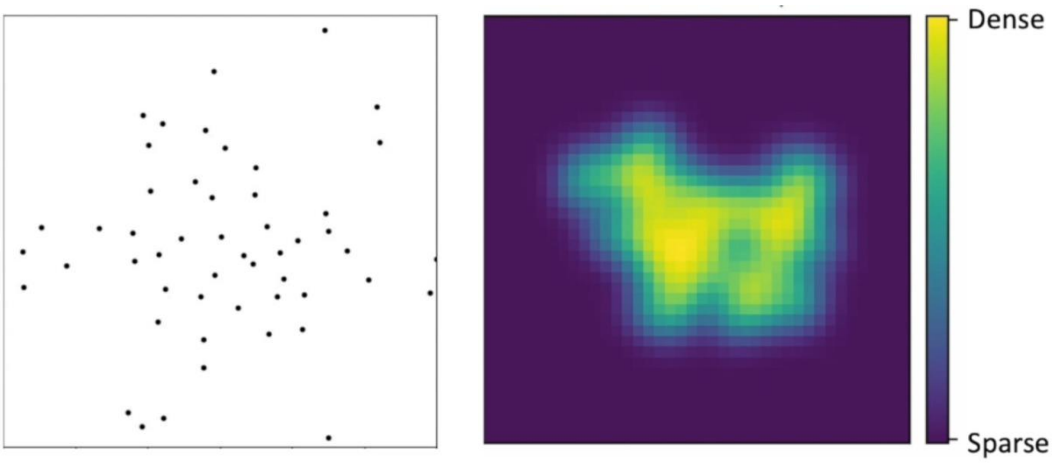

Figure 1: coverage display methods. Left = individual drone point display. Right = heat-map representing coverage. Images do not represent the same underlying truth due to random variance in simulation parameters

Participants

100 participants (62F, 37M, 1 Non-Binary) took part in an online visualization evaluation. Ethical approval was given by the University of Southampton ethics committee (ERGO number: 66360.A1).

Procedure

Participants accessed the study via Microsoft Forms where they were asked to fill out a consent form prior to being presented with an introductory video. Participants were given the following introduction: "You, as an operator, are going to observe the behaviour of a swarm of drones that are flying around. You don't have access to all drones (i.e., global communication is not available): operator and drones can only communicate with their neighbours, drones move around in the area, the operator receives updates from passing-by drones. When two agents move in a close proximity of each other, they exchange information about where other agents were last seen. Agents store the most recent locations of other agents. It takes longer for some agents to send/receive updates due to loss of connection with the rest of the swarm. We designed a set of questions to compare the usability of two visualisation methods. You are an operator, and your goal is to observe the swarm. You will see two videos and you will be asked to answer a set of questions for each display."

Participants were then introduced to each display method sequentially, and given the following descriptions:

- Display 1 – point display (individual drones): "This is how the operator sees the swarm in the first method. Drones are shown as points. With each update that the operator receives, the new location of agents appears or updates within the map. All agents are constantly moving around – drones that are not moving are 'lost drones' and you still see them where they were last observed."
- Display 2 – heat-map display: "The concentration of the swarm is shown in the heatmap. Instead of individual drones, the whole swarm is shown as a 'cloud'. With each update that the operator receives, the new concentration of the swarm appears or updates in the map."

Participants then watched a video of a swarm display showing individual UAVs, followed by a heatmap to represent swarm coverage. After viewing each video, participants were asked questions from the System Usability Scale (Brooke, 1996) and the System Acceptance Scale (van Der Laan et al., 1997) to measure how easily each could be interpreted. After watching both videos, participants were then asked to provide their preferred display given a variety of contextual factors:

- Which display is more practical to use with a larger swarm size?
- Which display is more practical to use in communication-constrained environments?
- Which display helps you to understand the swarm's motion & coverage of the area?
- In a time-critical situation, which display would you use?
- If time was not a concern, which display would you use?
- Which display helps in detecting errors in the behaviour of the swarm?
- In which display was the swarm's behaviour more transparent? (i.e., the behaviour was clearer & more predictable)
- Which method would you prefer to use for controlling the swarm?

Analysis

Preference responses to the contextual factors outlined above were analysed using eight chi-square goodness of fit analyses, corrected using the Bonferroni method (corrected $\alpha = .00625$). The system usability scale and system acceptance scale responses were analysed using two repeated measures ANOVAs.

Results

Responses to the eight questions prompting participants to provide their preferred display, given a contextual factor, were analysed using chi-square goodness of fit tests (Lancaster & Seneta, 2005). The alpha level was corrected using the Bonferroni method, according to the number of analyses conducted (Shaffer, 1995). The chi-square goodness of fit analyses showed that heatmaps were preferred when a larger swarm is being displayed, to display motion and coverage, and when time was a critical factor to task success (see Table 1). On the other hand, individual drone displays would be preferred for detecting errors in the swarm. There was no significant difference when communication would be constrained, when time was not a critical factor of task success, for being more transparent with the operator, and for controlling the swarm.

Table 1: Frequency of preferences and chi-square results for each contextual factor

Contextual Factor	Individual	Heatmap	$\chi 2$	p
Larger swarm size	21	79	33.64	.001*
Constrained communication	43	57	1.96	.162
Displaying motion and coverage	31	69	14.44	.001*
Time critical	28	72	19.36	.001*
Time non-critical	43	57	1.96	.162
Detecting Errors	74	26	23.04	.001*
Transparency	44	56	1.44	.23
For controlling the swarm	37	63	6.76	.009

Note. Bonferroni Corrected α = .00625 * = p < .00625

Two repeated measures ANOVAs (Girden, 1992) were conducted to identify whether coverage display method had an effect on System Usability Scores and System Acceptance Scores reported by participants. The analyses found a significant effect of display-type on usability scores ($F(1,99) = 22.53$, $p < .001$, $\eta_p 2 = .185$) and acceptance scores ($F(1,99) = 29.89$, $p < .001$, $\eta_p 2 = .232$) indicating that the heatmap display method had a greater level of usability and acceptance amongst participants when compared with displaying individual UAVs.

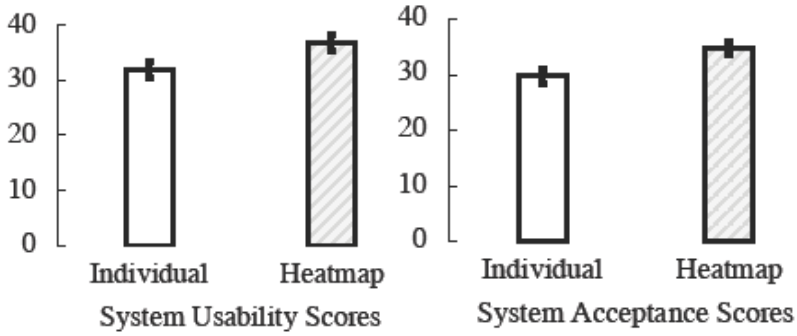

Figure 2: Bar graphs to show usability and acceptance scores for individual point-display and heatmap display. Error bars = 95% Confidence Intervals

Discussion

When visualising high numbers of individual robot agents, effective aggregated data displays may aid in collective human-swarm decision making (Kolling et al., 2015; Olsen & Wood, 2004). Due to a reduction of information, displaying swarm coverage through an aggregation like a heatmap (displaying swarm density) vs individual point-displays (where each drone is represented) may have trade-offs in-terms of their functionality and how they support operator usability, acceptance, and

in-turn, situation awareness and trust (Lee & See, 2004; McBride et al., 2014; Norman, 1990). This study identifies operational contexts that may have an effect in whether it's more appropriate to provide aggregated coverage displays, and whether users perceive this form of representing a swarm of 50 drones to be preferable than displaying individual agents.

The key-findings of this study were that heatmap displays were rated as being overall more usable (i.e., effective, efficient, and satisfying; Brooke, 1996) and acceptable (i.e., useful and satisfying; van Der Laan et al., 1997) to users than individual point displays, showing that users find heatmaps for swarm coverage to be easier to interpret and would likely use this type of display on a regular basis. Overall, heatmaps were rated as being preferable for displaying swarm coverage and motion. Contextual factors appear to contribute to these preferences. For example, heatmaps appear to be preferable when time constraints are high. Due to this it can be assumed that efficiency, one of the sub-categories of usability alongside effectiveness and satisfaction (Frøkjær et a., 2000), is likely to be a contributory factor towards participants' usability ratings. Participants often stated that heatmap displays as being easier to visualise, meaning that this form of display is likely to be higher in learnability and operability (Abran et al., 2003). Point-displays on the other hand were stated as being harder to visualise, likely due to the less efficient nature of displaying a non-connected network of individual entities.

Heatmaps were rated as being preferable when swarm sizes are larger, supporting previous notions that users require a novel type of interaction to deal with swarm complexity (Olsen & Wood, 2004). This finding shows that aggregated data displays are scalable, and more robust to increased UAV numbers compared to individual point displays. These effects may be explained by the number of drones visualised in this study. As swarm sizes increase, heatmaps may be more resilient to an increase in complexity as heatmaps provide a way of visualising the swarm as a single entity. These effects may dissipate, or not be present in studies with fewer UAV numbers.

Notably, there was no significant reduction in how 'transparent' the displays were, indicating that aggregating swarm geolocation data into heatmaps may not necessarily influence certain contributors towards the development of trust. This is likely due to the fact that users are able to visually represent geographical location to the same degree as individual point-displays, whilst also providing improvements to usability and acceptance, thus not necessarily reducing the amount of performance indicators (Lee & See, 2004).

The sole contextual factor that was rated as being preferable in favour of individual point displays was for detecting errors. This is likely due to the difficulty in displaying swarm dropouts via a heatmap – either displayed as static individual UAVs that have not updated their position, or through more explicit displays like colour changes. Due to this finding, it is recommended that diagnosis tools are provided alongside aggregated data displays for situations requiring further diagnosis. An operator would then be able to make use of this diagnosis tool to identify whether a certain region or certain UAVs are experiencing connectivity or battery issues.

It is worth noting that trust is not directly measured in this study, however, contributors towards situation awareness and trust (i.e., transparency, interpreting swarm coverage/motion, scalability and detecting errors) have been measured. Revisiting theoretical issues such as trust and situation awareness, this study shows that coordination is likely to improve during human-swarm interaction, and that these contributory factors addressed through the aggregation of geolocation displays via heatmaps. Further investigation is required regarding the relationship between the number of UAVs

and preference for heatmap displays, and the direct measurement of trust and situation awareness when these types of displays are utilised.

Conclusions

This user-study provides evidence that heatmap methods are more effective in addressing usability and acceptance in human-swarm interaction (see Figure 1). Swarm operators may make use of geolocation and density displays to improve their mental models on the locations, efficiency, and accuracy of swarm data processing. Further, in situations of larger swarm sizes, time-criticality and displaying motion and coverage, a heatmap has shown to have a higher preference amongst participants. Conversely, for detecting errors within the swarm, individual drone displays may be more appropriate. Notable limitations of this research include the omission of counterbalancing (due to the design of the data collection platform), which may introduce learning effects. However, due to the trials being relatively short, and participants were mostly naive to UV robotics, this effect is thought to be minor. Participants were also not able to interact with this early-stage interface, which may lead to additional variance in the results presented here, as task load, time-pressures, and overall task dynamicity will likely influence subjective report measures. Finally, further research is required for investigating how trust and situation awareness factors are directly influenced by these displays, and how further scalability (e.g., 100-1000 UAVs) may affect user preferences.

Acknowledgements

This work was conducted as part of the Trustworthy Autonomous Systems Hub, supported by the Engineering and Physical Sciences Research Council (EP/V00784X/1).

References

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, *11*(4), 325-338.

Billings, C. E. (2018). Aviation automation: The search for a human-centered approach. Boca Rotan, FL: CRC Press.

Brooke, J. SUS - A Quick and Dirty Usability Scale. *Usability Evaluation in Industry 1996*, *189*, 4-7.

Clark, J. R., Stanton, N. A., & Revell, K. M. (2019). Conditionally and highly automated vehicle handover: A study exploring vocal communication between two drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 699-715.

Dorigo, M., Theraulaz, G., & Trianni, V. (2020). Reflections on the future of swarm robotics. *Science Robotics*, *5*(49).

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 345-352. New York, NY: ACM.

Girden, E.R. *ANOVA: Repeated Measures*. (1992). Los Angeles, CA: Sage Publications.

Hamann, H.; Reina, A. Scalability in computing and robotics. (2021). *IEEE Transactions on Computers*.

Hussein, A., & Abbass, H. (2018, October). Mixed initiative systems for human-swarm interaction: Opportunities and challenges. In *2018 2nd Annual Systems Modelling Conference*, 1-8. New York, NY: IEEE.

Kolling, A., Sycara, K., Nunnally, S., & Lewis, M. (2013). Human swarm interaction: An experimental study of two types of interaction with foraging swarms. *Journal of Human-Robot Interaction*, *2*(2).

Lancaster, H.O.; Seneta, E. Chi-square distribution. In *Encyclopedia of biostatistics 2005*. Hoboken, New Jersey: John Wiley & Sons.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50-80.

McBride, S. E., Rogers, W. A., & Fisk, A. D. (2014). Understanding human management of automation errors. *Theoretical Issues in Ergonomics Science*, *15*(6), 545-577.

Meshcheryakov, R. V., Trefilov, P. M., Chekhov, A. V., Diane, S. A. K., Rusakov, K. D., Lesiv, E., Kolodochka, M.A., Shchukin, K. O., Novoselskiy, A. K. & Goncharova, E. (2019). An application of swarm of quadcopters for searching operations. *IFAC-PapersOnLine*, *52*(25), 14-18.

Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585-593.

Olsen Jr, D. R., & Wood, S. B. (2004). Fan-out: Measuring human control of multiple robots. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 231-238. New York, NY: ACM.

Outay, F., Mengash, H. A., & Adnan, M. (2020). Applications of unmanned aerial vehicle (UAV) in road safety, traffic and highway infrastructure management: Recent advances and challenges. *Transportation Research Part A: Policy and Practice*, *141*, 116-129.

Patil, M., Abukhalil, T., & Sobh, T. (2013). Hardware Architecture Review of Swarm Robotics System: Self-Reconfigurability, Self-Reassembly, and Self-Replication. *International Scholarly Research Notices*, *2013.*

Roundtree, K. A., Goodrich, M. A., & Adams, J. A. (2019). Transparency: Transitioning from human–machine systems to human-swarm systems. *Journal of Cognitive Engineering and Decision Making*, *13*(3), 171-195.

Roundtree, K. A., Manning, M. D., & Adams, J. A. (2018, September). Analysis of human-swarm visualizations. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 287-291. Los Angeles, CA: SAGE Publications.

Saffre, F., Hildmann, H., & Karvonen, H. (2021). The Design Challenges of Drone Swarm Control. In *International Conference on Human-Computer Interaction*, 408-426. Cham, Swizerland: Springer.

Schranz, M., Umlauft, M., Sende, M., & Elmenreich, W. (2020). Swarm robotic behaviors and current applications. *Frontiers in Robotics and AI*, *7*, 36.

Shaffer, J.P. Multiple hypothesis testing. *Annual review of psychology 1995*, *46*, 561–584.

Van Der Laan, J.D.; Heino, A.; De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*, 1–10.

Vasile, C., Pavel, A., & Buiu, C. (2011). Integrating human swarm interaction in a distributed robotic control system. In *2011 IEEE International Conference on Automation Science and Engineering*, 743-748. IEEE. New York, NY: IEEE.