2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Towards a Situated, Multimodal Interface for Multiple UAV Control

Geraint Jones, Nadia Berthouze, Roman Bielski, Simon Julier

Abstract— **Multiple autonomous Unmanned Aerial Vehicles (UAVs) can be used to complement human teams. This paper presents the results of an exploratory study to investigate gesture/speech interfaces for interaction with robots in a situated manner and the development of three iterations of a prototype command set. A command set was compiled from observing users interacting with a simulated interface in a virtual reality environment. We discovered that users find this type of interface intuitive and their commands tend to naturally group into both 'High-Level' and 'Low-Level' instructions. However, as the robots moved further away, the loss of depth perception and direct feedback was inimical to the interaction. In a second experiment we found that using simple heads up display elements could mitigate these issues.**

I. INTRODUCTION

Through their ability to quickly cover large tracts of ground from an aerial perspective, Unmanned Aerial Vehicles (UAVs) can aid the activities of human teams in a number of situations. We are particularly interested in the use of UAVs to complement Wilderness Search and Rescue (WiSaR) operations [1]. The benefits of UAVs in this domain are well-established: they can rapidly acquire aerial imagery even in adverse environments and multiple UAVs can collect data from multiple vantage points simultaneously.

An important role of UAVs is their ability to act as an extension of the human team's natural senses [2]. Such an extension is useful when assisting a human team in an operation known as "Hasty Search" [1], where UAVs complement rescuers by providing a view "over the hill" [1] and opportunistically look for clues. To optimize the Hasty Search, the interface between the human team and the robots must support fine grained interactions. A major obstacle to this combination of human and robot faculties in current interfaces is a lack of situation awareness on behalf of the operator [3]. Operators are often unaware of where the robot is in the environment, or are unaware of the exact nature of the robots immediate surroundings. The operators understanding of the 'situatedness' of rescue robots is a major problem with SaR (Search and Rescue) [4]. Additionally, piloting robots places heavy demands on users' cognitive and perceptual resources [5], thus currently deployed systems require many operators to control a single vehicle [6] — typically at least a pilot and a sensor operator [1]. The delegation of low level control to autonomous flocking and tasking behaviours allows one user to control multiple UAVs at once [7]. Speech and gestures provide a natural, intuitive means to communicate with autonomous agents of this type in a situated manner, since this emulates to some degree human-human interactions [8].

The aim of this study was to investigate an interaction paradigm for situated UAV control, by utilising naturalistic gesture/speech commands which leverage the physical characteristics of the environment to support input/output in an intuitive manner: users can see and point to agents, terrain and points within the space, and have an inherent understanding of where these system elements are in real space. Specifically, the aim was to gather key requirements for a multi-modal system that can be subsequently implemented. A series of tests were run in a large field-of-view virtual environment using a Wizard-of-Oz experimental protocol, as simulations offer greater control, ease of replicability and increased flexibility to alter aspects of the interface and the agents' behaviours [9]. The study led to 3 iterations in the design of the interface. The first, described in Section II, was a freeform exploratory study (Experiment 1). The purpose of this interface was to investigate a) if the general concept was sound and b) to extract commonalities from users' naturalistic behaviours to develop a command language. Using the feedback generated we developed two further interfaces, both described in Section III. The second iteration (Experiment 2, condition 1) tested this generated language. The third iteration (Experiment 2, condition 2) further refined the interface to address issues elicited from the earlier interfaces through the use of graphical display elements.

II. EXPERIMENT 1 - EXPLORING THE COMMAND SPACE

Experiment 1 was exploratory in nature. The intent was to observe users naturalistically interacting with the swarm. Additionally, requirements and potential usability issues were elicited through post trial interviews and video analysis. From this experiment, a set of commands was compiled, along with some implications for the design of the interface.

A. Method

1) Participants: 10 participants undertook this study, (n=6 females, n=4 males).

2) Materials/apparatus: The experiment was presented to participants in a CAVElike Virtual Reality Environment (VRE). This kind of stereoscopic volume is qualitatively different from using a monitor and keyboard, as the wide

---

The work in this paper was partially supported under the EPSRC-funded project "SUAAVE: Sensing Unmanned Autonomous Aerial Vehicles (EP/F064179/1)

G. Jones, R. Bielski and S. Julier are with the Department of Computer Science, University College London, Gower St, London, WC1E 6BT, UK; N. Berthouze is with the UCL Interaction Centre, Department of Computer Science, University College London, Gower St, London, WC1E 6BT, UK., `geraintrjones@googlemail.com`, `n.berthouze@ucl.ac.uk`, `roman.battersea@gmail.com`, `S.Julier@cs.ucl.ac.uk`

978-1-4244-5040-4/10/$26.00 ©2010 IEEE

TABLE I
VIDEO CODING CATEGORIES, COUNTS AND INTERRATER RELIABILITY SCORES

Command Modality / Type	Coding Pattern	Mean Commands Rater 1	Mean Commands Rater 2	Interrater Reliability
Voice: High-Level	Any command that assumes a level of navigational autonomy eg, 'Go to the Tree', 'Search those Rocks' etc.	6.30 (7.90)	4.27 (5.76)	0.82
Voice: Low-Level	Any command that directly steers a UAV.eg:'Left','Right'	8.84 (15.1)	5.86 (6.44)	0.93
Gesture: Pointing	Gestures characterised by raising an arm and indicating a direction with it, normally one hand used	4.38 (4.74)	5.86 (6.44)	0.53
Gesture: Herding	Gestures characterised by moving arms through space as if pushing UAVs in the desired direction, often bimanual, palms often flattened perpendicular to the direction of limb motion	12.27 (24.70)	17.62 (31.41)	0.93

field of view and freedom of movement allow a much more immersive experience.

Participants were presented with a simulated world consisting of a flat grassy plain and partially cloudy sky, populated with landmarks (see fig. 1) and representations of UAVs, the behaviour of which was based on the boids algorithm [10]. Users' commands were interpreted by the experimenters, and enacted on the system through a separate experimenter interface. Audio/Visual data was captured from two sources to provide rear and side views of participants.

3) Procedure: The participants took part in 5 trials in the VRE, 1 practice and 4 experimental. Each trial had either 3 or 9 landmarks (each participant experienced two of each, order determined pseudorandomly), and a swarm of 10 UAVs. One of the landmarks was the target, the others distractors. Participants were instructed to guide the UAVs to search the environment for a missing person, who could be found at a landmark. The urgency of the task was emphasised, as SaR scenarios are highly time critical [4].

Participants were informed if a landmark was a distractor or a target when the UAVs were flying above it by a short message in the centre of their visual field. The decision was taken not to include a more complex data presentation format, so as to focus the investigation on control rather than sensor fusion and data presentation.

In worlds with 3 landmarks, all 3 were of the same type. Worlds with 9 landmarks contained 3 of each type. The positions of the landmarks, and which of those should be designated the target were pseudorandomly generated. The landmark type for the trials with 3 landmarks was also pseudorandom, as was the landmark target type in the trials with 9 landmarks.

Fig. 1. Types of landmarks (trees, rocks, wood pile).

B. Results and Analysis: Video Data

Participants' video data was examined to ascertain potentially interesting areas and a coding scheme (see table 1) was developed for investigating these. The coding was performed independently by two raters and concordance ratings were generated using intraclass correlation.

a) Gesture and Speech: Command gestures were classified into two main types — *pointing gestures* (where participants indicated an area in the world with an outstretched arm), and *herding gestures*, where participants moved a limb through space as if pushing or herding the UAVs. Vocal commands were classified as *Low-Level* or *High-Level*. Low-Level commands steer the UAVs with direct physical orders (such as "left" or "right"), while commands assuming a certain level of navigational autonomy on behalf of the robots were considered to be High-Level (eg, "Go to the tree").

The ratings were very strongly in agreement for Low-Level vocal commands, High-Level vocal commands and herding gestures, with single measures intraclass correlations (ICC) of .934, .816 and .934 respectively, all significant at a p threshold of less than 0.01. The agreement for pointing gestures was less strong, but still very significant, with a single measures ICC of .526 ($p < 0.01$). The herding gestures were by their nature more gross than the pointing gestures, whereas sometimes participants used fine pointing motions, articulating the wrist rather than directing the forearm. This made discrimination less robust, possibly leading to the lower agreement between raters.

b) Cross Modal Coupling: The analysis shows that vocal and gestural commands were generally delivered in concert — each command in a modality was associated with a command in the other modality. Total number of gestures is very strongly correlated (using Cohen's bounds, [11]) to total number of vocalisations, $r(35) = .930p < 0.01$; the vast majority of speech commands were accompanied by a gesture, although there is not a one to one correspondence between gesture and speech commands. On average, 3 gestures were executed for every 1 vocalisation, mainly repetitions. It was also noted that the High-Level vocal commands were associated with pointing type gestures, whereas the Low-Level vocal commands were associated with herding type gestures. As such it appeared that pointing type gestures coupled with High-Level vocal commands represent a High-Level instruction, and herding commands coupled with Low-

Level vocal commands a Low-Level instruction.

To examine the relationship between gesture and vocal command types, correlations between each command type were performed. The following ratio measures of command type over command modality were considered: Low-Level vocal commands over total vocal commands (v_low/v_total), High-Level vocal commands over total vocal commands (v_high/v_total), herding gestures over total gestures(g_herd/g_total) and pointing gestures over total gestures (g_point/g_total). In two trials, users made no vocalisations — as such no ratio measures could be calculated for voice in these two trials.

The results using the ratio measures were as expected, supporting the hypotheses formed during the subjective video viewing. v_low/v_total is correlated positively with g_herd/g_total ($r(33) = .447, < 0.01$) and negatively with g_point/g_total ($r(33) = -.420, < 0.05$), and vice versa for v_high/v_total, indicating that herding gestures are associated with lower level vocal commands, and pointing gestures are associated with higher level commands.

c) Interaction Effort: It appeared that the participants using a greater proportion of higher level commands seemed to perform the task with less interaction effort than those using more lower level commands. There is a positive correlation between total number of vocal commands with the ratio of Low-Level vocal commands, and a negative correlation for High-Level commands ($r(33) = .460$ and $r(33) = -.460$ respectively, both $p < 0.01$). This is also the case for gestures, if pointing gestures ($r(35) = .613$ correlation with total gestures) are considered High-Level and herding ($r(33) = -.619$) low level as suggested by the findings above. This indicates that higher level commands involve less interaction effort.

C. Results and Analysis: Interview Data

Semi-structured interviews were conducted following the final trial in order to probe users' experiences of the interaction. 41 unique issues of interest were found, with 14 occurring for more than one participant. As a rough metric, we considered that the greater the number of times an issue was seen, the more pertinent it is in terms of implications for the design. The most frequently occurring issues are discussed below.

d) Depth Perception (5 occurrences): Many users had difficulty with depth perception and occlusion — as the distance to the robots increased, participants found it increasingly difficult to judge the location and speed of the robots. A typical comment was "Gauging distance was difficult. [once you] get to a certain point." It is likely that some of the feedback issues (discussed below) were related to this problem. Not only did loss or degradation of visual contact negatively impact the users' only real feedback channel, but also the loss of motion perspective at range meant the judgements of the robots speed became inaccurate, making them appear stationary (and therefore nonresponsive) from certain viewpoints. Although the VRE-induced perceptual distortions and simplified environment are likely to exacerbate these issues [12], they are inherent to the human vision processing system. Therefore, the same issues will apply in a real physical environment. Ultimately, there comes a point when the usable interaction range is constrained by the limits of the visual system. Indeed, performance will fall off over distance rather than having a sharp delineation of operating range, which is potentially more frustrating. It follows that any system wanting to make use of a situated interface of this nature where the interactive space extends beyond the users immediate personal environment needs to account for this depth perception problem.

e) Limited vocabulary set (4 occurrences), strict syntax (4 occurrences): Participants used short, simple constructs to control the robots. The interview data suggests this is due to the expectation that the system will not be able to understand natural language, but will know some simple commands. Comments included: "I would expect a limited vocabulary set" and "...just assumed they were robots and they weren't that bright!" A number of participants mentioned that they "talked to them like controlling a dog." This level of speech seemed to be quite natural for participants, possibly because the limited autonomy of the robots was quite similar to that of an animal; it has been argued that dogs provide a good model of robotic agents that have some autonomy but lack the cognitive and linguistic abilities of humans [13].

A limited speech set and simple syntax would reduce the complexity of the system, making it less resource hungry and more robust. Since the data suggests users will accept, maybe even prefer, this kind of system, it seems reasonable to attempt to implement a simple, reduced instruction set with a lightweight syntax.

f) Egocentric Control Scheme (6 occurrences): Many of the participants made reference to the robots being aware of their (the users') current position, and how important this was to their interaction.

The facing of the UAVs was not considered by any users — all participants seemed to assume that the robots not only knew the users position, but were also able to translate their frame of reference to that of the users. It is interesting that users overwhelmingly assumed a low level of verbal communication ability from the UAVs, but took this locational and directional ability for granted. Again, it is possible that an animal schema is being used - animals are good at finding their way, but are not normally known for their conversational skills. Alternatively, this could be an aspect of technology exposure — positioning data and location based services are widely used, and often considered 'better' than humans. Speech recognition on the other hand has historically been perceived as poor. As such, the users may have a "Machines are good at knowing where they are, but bad at talking" model.

g) Feedback (5 occurrences): Participants were unsure of whether the robots had received commands, and whether they had understood them. One participant said that it was "Difficult to know how my gestures affected the robots."

Since the only channel of feedback was the motion of the robots in response to commands, it was not clear to participants that the UAVs were responding accurately to their commands at long distances. The loss of depth perception led participants to believe that UAVs had reached a distant target and stopped, even though the UAVs were still en route. The response to commands to return were not immediately evident in such situations ; perceived size change in the UAVs at that distance was not great enough for the participants' perceptual systems to infer motion.

Additionally, considering the earlier observations about the assumed abilities (or lack thereof) of voice recognition, it is likely that the assumption of imperfect and limited command discrimination for an unknown command set contributed to the desire for explicit feedback about the system state.

h) Splitting Swarm (3 occurrences): Splitting the swarm was an issue raised, not only at interview, but also during both training and task sessions. This is encouraging, as it indicates that multiple robot control may be intuitively understood. However, the data also indicates that participants imagined themselves actively splitting swarms into identifiable groups, and then controlling the groups essentially independently. This is not consistent with the idea of higher level tasking systems such as suggested by [14]. These systems would assign tasks according to optimising or at least satisficing algorithms. As such, users would not have to explicitly split swarms and control subswarms, but would provide a series of high-level orders, the details of which would be handled by the tasking system. Although this may be more operationally efficient, it is possible there may be a detrimental effect on users.

III. EXPERIMENT 2: TESTING THE COMMAND SET

Experiment 2 was performed to test the command set based on data from the previous experiment, and to examine the effects of modifications to the interface which were aimed at addressing the issues uncovered in experiment 1. The modifications are discussed below.

i) Command Set: Taking into account the lower interaction effort of High-Level commands in experiment 1, the command set in experiment 2 gave primacy to a "go to the..." command based on a family of similar commands observed, which consisted of saying "go to the landmark" and pointing to the desired landmark (with an extended forearm, to aid discrimination). This command was demonstrated to users first. Then, the user was shown the Low-Level commands. The Low-Level commands observed in the video were largely steering-type commands, and four were codified — 'left', 'right','forward' and 'come back to me'. The first three consist of pushing the palms of both hands towards the desired direction as if attempting to push or herd the robots. The final command begins with the arms held against the sides of the body, bent at the elbow so the forearm is perpendicular to the body and with the hands palms up. The user then raises their palms towards them in a beckoning gesture.

j) False Shadows: To address the problem of depth perception, false shadows were implemented, following the work of [15] As illustrated in Figure 2, these were graphical regions projected onto the ground.

Fig. 2. False shadows

A. Method

1) Participants: 9 new participants undertook this study, (n=1 females, n=8 males).

2) Materials/apparatus: The study was run using the same materials and apparatus as previously. However, the set of gesture commands described above was provided to the participants. The experimenters would only respond to this command set when controlling the system's behaviour.

3) Procedure: The procedure of the study was identical to experiment 1, with the addition of instructions on the command set, which were delivered immediately preceding the practice trial for each participant. Participants 1–5 did not receive the false shadows, while 6–9 did.

B. Results and Analysis: Video Data

The videos were coded in the same manner as for experiment 1, permitting direct comparison between the measures used. All comparisons are Mann-Whitney U tests unless otherwise noted, as the data violates assumptions of normality, showing a strong positive skew.

Participants in experiment 2 used significantly less commands than participants in experiment 1, in both vocal $(U(73) = 482, Z = -2.04, p < 0.05)$ and gestural modalities $(U(73) = 304, Z = -4.00, p < 0.01)$, see Fig. 3.

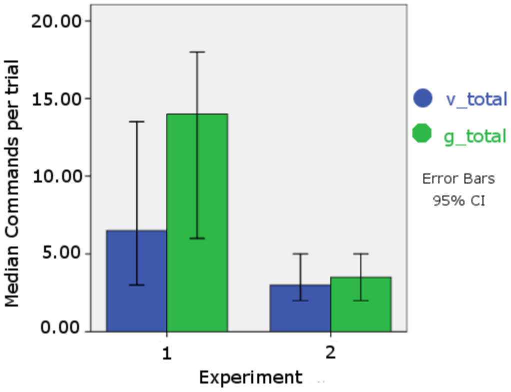

Fig. 3. Median commands per trial in experiment 1 vs experiment 2

Comparisons were also performed between the ratio measures (v_low_t, v_high_t, g_herd_t and g_point_t) for each experiment to test whether participants used a higher proportion of High-Level vocal commands and pointing commands in experiment 2. It was found that participants used a significantly higher proportion of High-Level commands in experiment 2 than in experiment 1 in both modalities, $U(71) = 241$, $Z = -4.90, p < 0.01$ for High-Level vocal commands and $U(73) = 221$, $Z = -5.16, p < 0.01$ for pointing type gesture commands. The Low-Level commands showed the inverse ($U(71) = 241$, $Z = -4.90$, $p < 0.01$ for Low-Level voice commands and $U(73) = 246$, $Z = -4.89$, $p < -0.01$ for herding type gestures).

As can be seen from fig. 3, the gestures per vocal command in this experiment was much closer to 1, indicating better command coupling.

It is interesting to note that despite the requirements for strict syntax, many users utilised a number of variations on the instructed commands, especially for 'go to the...', which actually assumed quite a high level of semantic understanding on behalf of the voice recogniser in some cases. This supports assertations [16] that users expect quite sophisticated recognition in multimodal gesture/speech interfaces.

C. Results and Analysis: Interview Data

a) To See with Robot Eyes: The feature request overwhelmingly elicited from participants was to be able to see the sensor data from the robots — participants wanted to "see with the robots' eyes", and other participants wanted to see more than just "there/not there" . This is in contrast to experiment 1, where this desire was only elicited from one participant. It is likely that the feedback and depth perception issues were more important to participants, and so overshadowed this need. It could be argued that this issue is in part an artefact of the experimental setup, since the binary found/not found feedback was largely a placeholder to control the trials, rather than a genuine attempt at portraying that aspect of the interface. Nevertheless, this issue does suggest that users should be provided with at least some level of data from the robots, rather than simply having a remote sensor operator parse the data and inform the users. It is possible that reduced or processed data may be more useful to the user than raw feeds from the UAVs. Whilst some of the participants explicitly wanted access to the video feeds from all the robots, it has been argued that it would be difficult for users to actually monitor outputs of this nature [1]. The level and modality of the data presentation requires further study.

b) Low-Level Gestures and Commands: Participants preferred the 'Go to the...' command, and in many cases used it almost exclusively. This was echoed in the interview data with many participants stating that they failed to see the need for the other commands. There were one or two exceptions however, who stated that the inclusion of the Low-Level commands were important for them. Indeed, it appeared that users dropped into using Low-Level gestures when they perceived the robots were not following their instructions properly — almost as if using them for diagnostic purposes. Users often tried out these gestures in the practice trials also, but then suddenly stopped using them in test trials. It is possible that the lower level gestures were perceived as being less efficient during the practice, leading to lack of use in the real trials. However, the suddenness of the dropoff argues against this kind of learning effect.

It is possible that the Low-Level commands give more immediate feedback as to whether the command was understood or not. The depth perception issues coupled with the swarm dynamics based pathfinding meant that with 'go to the' commands there was less certainty as to whether commands had been understood. This is consistent with 'debugging' use in the second experiment. Since in this experiment the users had a defined set of commands, they were more certain in using the command even if the feedback was not immediate.

c) Depth Perception: Depth perception problems did not seem to be reduced by the new gesture set, as the issue was elicited for 3 of the five participants in Experiment 2 in the no shadows condition, a similar proportion to in experiment 1 (5 of 10). However, the issue was not commented upon by any of the 4 participants in the with shadows condition. Indeed, one participant remarked that without the shadows "I have no idea where they are flying in the sky..." and another said that "The red hexagons were really useful. Without that... wouldn't know if they were flying off into!".

d) Feedback: A similar frequency of feedback being raised as an issue was seen in the no shadows condition of experiment 2 as in experiment 1 (two of five vs five of ten, or 40% vs 50%), with no occurrences in the shadows condition of experiment 2. This would appear to indicate that the false shadows were efficacious at providing feedback.

e) Splitting: This occurred as an issue for twice as many participants across experiment 2 as experiment 1 - three in the no shadows condition and three in the shadows condition. It is possible that as less cognitive resources were being consumed by controlling the robots, the participants were able to give more thought to the overall structure of the task. As with the 'See with robot eyes' issue, it is possible that the splitting desire was overshadowed in experiment 1 by the more pressing concerns of feedback, loose command set etc.

f) Intuitive: A number of participants made positive comments across both conditions of experiment 2, to the effect that the interface was intuitive or easy to use. Whilst this would appear to validate the command set and general paradigm as being intuitive to use, it should be considered that there are other factors that may be effecting this outcome. It is possible that the command set and false shadows were seen as design features, which could have led to the increase in positive comments due to socially conditioned effects. The freeform commands in the first experiment may not have been subject to this effect.

IV. DESIGN IMPLICATIONS

A. Commands

Commands should be well defined but should allow some flexibility.

It can be seen from the difference in performance between experiment 1 and 2 that a well defined command set is necessary. However, the system must be robust enough to handle a certain level of syntactic variation, where the semantic nature of the command is constant but the exact wording is slightly different.

High-Level commands require less interaction effort and so should be given primacy. Lower level commands to deal with edge cases can be supported without detracting from the High-Level commands.

B. Feedback

Feedback is very important to users. As [17] state, autonomous agents need to be able to 'talk' as well as 'listen'. Users need to be informed the system has received and understood a command. This feedback can be provided by placing interface elements into the world. Feedback would be particularly important in real world situations since the variances and novelty of real world terrain may leave the robots in any number of what [18] describes as "subtle failure modes.'

C. Depth Perception

Depth perception is an issue with this kind of interface. This can be addressed with the use of false shadow information. Shadows with straight edges and angles seem to be more efficaous than smooth, elliptical shadows, however this requires further testing.

D. Data Abstraction

Participants wanted more data available in the interface. Future designs should provide data at a lower level of abstraction.

V. CONCLUSION AND FUTURE WORK

The experiments show that a situated, multimodal interface is potentially viable in a SaR domain. However, the robots in this study moved as a single unit. The interface needs to be tested with a more advanced tasking algorithm that presents the user with agents moving independently of each other. It would also be useful to test the interface against other types of interface, such as mobile devices or laptops, to ascertain whether the situated nature of the interface provides any operational benefits.

ACKNOWLEDGMENTS

The work in this paper was partially supported under the EPSRC-funded project "SUAAVE: Sensing Unmanned Autonomous Aerial Vehicles" (EP/F064179/1).

REFERENCES

[1] J. L. Cooper and M. A. Goodrich, "Towards Combining UAV and Sensor Operator Roles in UAV-Enabled Visual Search," in *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction*. Amsterdam, NL: ACM New York, NY, USA, March 2008, pp. 351–358.

[2] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA, USA: MIT Press, 1st October 1992.

[3] J. L. Drury, J. Richer, N. Rackliffe, and M. A. Goodrich, "Comparing Situation Awareness for Two Unmanned Aerial Vehicle Human Interface Approaches," MITRE Corporation, Bedford, MA, USA, Tech. Rep. A652654, 2006. [Online]. Available: http://handle.dtic.mil/100.2/ADA456256

[4] J. Burke, R. Murphy, M. Coovert, and D. Riddle, "Moonlight in Miami: Field Study of Human-Robot Interaction in the Context of an Urban Search and Rescue Disaster Response Training Exercise," *Human-Computer Interaction*, vol. 19, no. 1–2, pp. 85–116, June 2004.

[5] J. Casper and R. R. Murphy, "Human-Robot Interactions During the Robot-Assisted Urban Search and Rescue Response at the World Trade Center," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 33, no. 3, pp. 367–385, June 2003.

[6] J. L. Franke, V. Zaychik, T. M. Spura, and E. E. Alves, "Inverting the Operator/Vehicle Ratio: Approaches to Next Generation UAV Command and Control," in *Proceedings of AUVSI Unmanned Systems North America*, Melbourne, Australia, March 2005.

[7] N. Chambers, J. Allen, L. Galescu, and H. Jung, "A Dialogue-Based Approach to Multi-Robot Team Control," in *Proceedings from the 2005 International Workshop on Multi-Robot Systems*, Washington DC, USA, 14–16 March 2005, pp. 257–262.

[8] H.-J. Böhme, T. Wilhelm, J. Key, C. Schauer, C. Schröter, H.-M. Groß, and T. Hempel, "An Approach to Multi-Modal Human-Machine Interaction for Intelligent Service Robots," *Robotics and Autonomous Systems*, vol. 44, no. 1, pp. 83–96, July 2003.

[9] D. R. Olsen Jr and S. B. Wood, "Fan-Out: Measuring Human Control of Multiple Robots," in *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, Vienna, Austria, 2004, pp. 231–238.

[10] C. W. Reynolds, "Flocks, Herds and Schools: A Distributed Behavioral Model," *ACM SIGGRAPH Computer Graphics*, vol. 21, no. 4, pp. 25–34, July 1987.

[11] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum Associates, 1988.

[12] D. Drascic and P. Milgram, "Perceptual Issues in Augmented Reality," in *Proceedings of SPIE*, ser. Stereoscopic Displays and Virtual Reality Systems III, vol. 2653, San Jose, CA, USA, 30 January – 2nd February 1996, pp. 123–134.

[13] K. Dautenhahn, "Robots We Like to Live With?! — A Developmental Perspective on a Personalized, Life-Long Robot Companion," in *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2004)*, Okayama Japan, 20–22 September 2004, pp. 17–22.

[14] J. L. Baxter, E. K. Burke, J. M. Garibaldi, and M. Norman, "Multi-Robot Search and Rescue: A Potential Field Based Approach," *Autonomous Robots and Agents*, vol. 76, p. 9, 2007.

[15] J. Wither and T. Höllerer, "Pictorial Depth Cues for Outdoor Augmented Reality," in *Proceedings of the Ninth IEEE International Symposium on Wearable Computers (ISWC05)*. Osaka, Japan: IEEE Computer Society, 18–21 October 2005, pp. 92–99.

[16] D. Perzanowski, A. C. Schultz, W. Adams, E. Marsh, and M. Bugajska, "Building a Multimodal Human-robot Interface," *IEEE Intelligent Systems*, vol. 16, no. 1, pp. 16–21, Jan-Feb 2001.

[17] H. Jones and S. Rock, "Dialogue-based Human-Robot Interaction for Space Construction Teams," in *Proceedings of the IEEE Aerospace Conference .*, vol. 7, Big Sky, MT, USA, 9–16 March 2002, pp. 7–3645–7–3653.

[18] R. R. Murphy, "Human-Robot Interaction in Rescue Robotics," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 34, no. 2, pp. 138–153, May 2004.