Conference or Workshop Item:

Millard, Alan Gregory orcid.org/0000-0002-4424-5953 and Williams, James (Accepted: 2017) Conversational human-swarm interaction using IBM Cloud. In: UK-RAS Conference on Robotics and Autonomous Systems, 12 Dec 2017, UWE Bristol Exhibition and Conference Centre (ECC). (In Press)

Conversational human-swarm interaction using IBM Cloud

Alan G. Millard and James R. Williams
alan.millard@york.ac.uk, james.williams@uk.ibm.com

1 Introduction

Swarm robotics is an approach to the coordination of large numbers of robots that has become an increasingly popular field of research in recent years [1], not least because properly engineered robot swarms are scalable, flexible, and robust, making them an attractive alternative to single-robot systems in many application domains [2]. Since its inception, the field of swarm robotics has grown beyond its roots in purely decentralised control inspired by social insect behaviour [3], now often utilising hybrid centralised/decentralised control architectures that incorporate human operators who guide swarm actions during tasks such as firefighting [4], or the localisation of radiation sources [5].

This kind of human-swarm interaction has attracted significant interest from the research community, spawning an entire sub-field of its own that investigates how human operators, supervisors, and team-mates can interact with robot swarms and receive feedback from them [6]. To date, human-swarm control methods such as the use of graphical user interfaces [5][7] and spatial gestures [8][9][10] have received much attention, but there has been little investigation into the potential of controlling swarm robotic systems with an operator's voice. The few studies that have explored this idea [11][12] are restricted to the use of specific predefined phrases that the human operator is required to learn, resulting in interactions that are unnatural in comparison to the way a human would normally express themselves in speech. In this paper, we present a novel architecture for conversational human-swarm interaction that addresses these issues, allowing swarm robotic systems to be engineered in such a way that a human operator can guide a swarm using spoken dialogue in a more natural manner.

2 An architecture for conversational human-swarm interaction

IBM Cloud [13] is a cloud platform-as-a-service (free to use for educators and students via the *Academic Initiative for Cloud offer* [14]) that provides a catalogue of services which developers can use to build applications. As part of this catalogue, IBM offer a number of machine-learning-based services such as language translation, image recognition, and sentiment analysis. We employ the *Watson Conversation* service [15], which allows developers to define a dialogue tree for a virtual agent/chatbot that can engage with users on a particular topic. The developer defines a set of *intents* (goals the user may want to achieve), a set of *entities* (classes of objects relevant to the user's intents), and a set of *rules* that define how the chatbot should respond based on the entities and intents recognised in the user's input. In addition, the developer provides a set of example phrases for each intent, as well as synonyms for each entity, which are used to train the natural language classifier. The Watson Conversation service can also be configured to ask the user for further information if their intent is unclear, thus making the system robust to the flexibility of natural dialogue.

To conversationally interact with a swarm of robots, we can define the human-swarm interaction as a dialogue using the Watson Conversation service. A human operator's goals can be encoded as intents (e.g. "select robots with at least 50% battery", "add robots to swarm three", "swarm three perform flocking"), and domain objects can be defined as entities (e.g. "robot", "battery", "swarm three"). The dialogue tree provided by the developer links these intents and entities to operator feedback via natural language responses (like a traditional chatbot), as well as domain-specific responses that can be used to guide the robotic system.

2.1 Core architecture

The core architecture for the proposed conversational interaction is shown in Figure 1 (left). The *Human-Swarm Interaction Controller* (HSIC) is the interface to the human operator commanding the swarm, which streams voice recordings to the *Watson Speech to Text* service [16] running in the

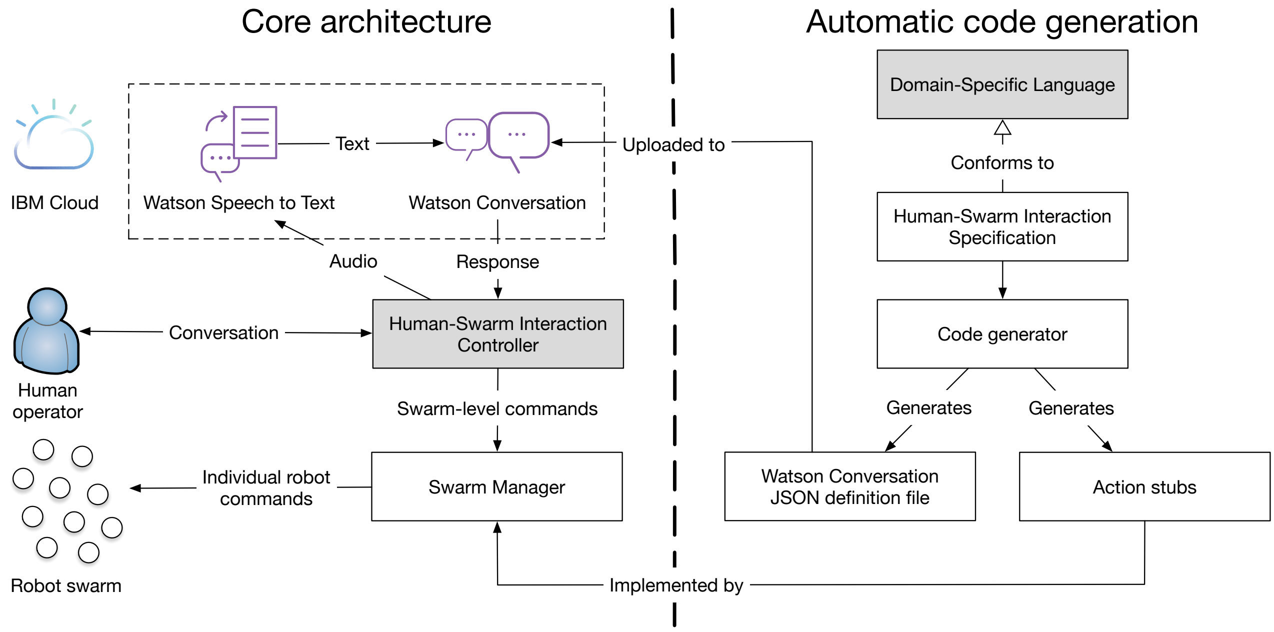

Figure 1: System architecture. *Left:* Core architecture, *Right:* Automatic code generation.

IBM Cloud. This service analyses the audio and converts it into text, then passes it to the Watson Conversation service running the domain-specific dialogue tree, which interprets the human operator's intent and responds accordingly. The HSIC then either requests further information from the operator (if required by the dialogue tree), or dispatches complete commands to the *Swarm Manager* — a target-specific piece of software written by the developer (e.g. a ROS master [17]) that interprets these swarm-level commands and decomposes them into instructions that are sent to the individual robots, triggering decentralised autonomous swarm behaviours.

2.2 Automatic code generation

Developing code for the Watson Conversation definition and the Swarm Manager independently may lead to inconsistencies between the two that are only discovered at run-time. To address this, we propose the use of a Domain-Specific Language (DSL) and a code generator that allows swarm engineers to develop both of these artefacts from a single specification (see Figure 1, right), thus ensuring consistency whilst reducing development effort. The code generator produces a JSON file that can be uploaded to the Watson Conversation service running in the IBM Cloud, and corresponding Swarm Manager code stubs for each of the swarm actions defined in the dialogue. The developer then completes the implementation of these stubs to define how the swarm-level commands are decomposed into messages that are sent to the individual robots. When the Swarm Manager receives a command from the HSIC, it invokes the corresponding auto-generated action stub, and the swarm is instructed accordingly. To minimise development effort, a library of common, composable actions can be used in the DSL. These actions might include the ability to select specific robots, assign a name to a selection of robots for later reference, and common swarm behaviours such as flocking and aggregation.

3 Conclusion

In this paper, we have presented an extensible architecture for defining a conversational human-swarm interaction model, which is sufficiently generic to be applied in many target domains. Using a DSL, developers can easily define how a human operator's intents are mapped to swarm-level commands, enabling rapid development of conversational communication. The flexibility of the resulting human-swarm interaction facilitates the use of natural spoken dialogue, which is particularly advantageous in high-pressure situations where human operators may struggle to recall specific predefined commands, and can instead interact with a robot swarm quickly and conversationally to achieve their goals.

References

[1] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, “Swarm robotics: a review from the swarm engineering perspective,” *Swarm Intelligence*, vol. 7, no. 1, pp. 1–41, 2013.

[2] E. Şahin, “Swarm robotics: From sources of inspiration to domains of application,” in *International workshop on swarm robotics*. Springer, 2004, pp. 10–20.

[3] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm intelligence: from natural to artificial systems*. Oxford University Press, 1999.

[4] A. M. Naghsh, J. Gancet, A. Tanoto, and C. Roast, “Analysis and design of human-robot swarm interaction in firefighting,” in *17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2008, pp. 255–260.

[5] S. Bashyal and G. K. Venayagamoorthy, “Human swarm interaction for radiation source search and localization,” in *Swarm Intelligence Symposium (SIS)*. IEEE, 2008, pp. 1–8.

[6] A. Kolling, P. Walker, N. Chakraborty, K. Sycara, and M. Lewis, “Human interaction with robot swarms: A survey,” *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 1, pp. 9–26, 2016.

[7] J. McLurkin, J. Smith, J. Frankel, D. Sotkowitz, D. Blau, and B. Schmidt, “Speaking swarmish: Human-robot interface design for large swarms of autonomous mobile robots.” in *AAAI Spring Symposium 2006: To Boldly Go Where No Human-Robot Team Has Gone Before*, 2006, pp. 72–75.

[8] J. Nagi, A. Giusti, L. M. Gambardella, and G. A. Di Caro, “Human-swarm interaction using spatial gestures,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 3834–3841.

[9] G. Podevijn, R. O’Grady, Y. S. Nashed, and M. Dorigo, “Gesturing at subswarms: Towards direct human control of robot swarms,” in *Conference Towards Autonomous Robotic Systems*. Springer, 2013, pp. 390–403.

[10] S. Nagavalli, M. Chandarana, M. Lewis, and K. Sycara, “Multi-operator gesture control of robotic swarms using wearable devices,” in *Tenth International Conference on Advances in Computer-Human Interactions (ACHI)*, March 2017.

[11] S. Pourmehr, V. M. Monajjemi, R. Vaughan, and G. Mori, ““You two! Take off!”: Creating, modifying and commanding groups of robots using face engagement and indirect speech in voice commands,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 137–142.

[12] G. Kapellmann-Zafra, J. Chen, and R. Groß, “Using Google Glass in Human–Robot Swarm Interaction,” in *Conference Towards Autonomous Robotic Systems*. Springer, 2016, pp. 196–201.

[13] IBM Cloud. [Online]. Available: https://www.ibm.com/cloud/

[14] Academic Initiative for Cloud offer. [Online]. Available: https://developer.ibm.com/academic/resources/cloud-offer/

[15] IBM Watson Conversation service. [Online]. Available: https://www.ibm.com/watson/services/conversation/

[16] IBM Watson Speech to Text service. [Online]. Available: https://www.ibm.com/watson/services/speech-to-text/

[17] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “ROS: an open-source Robot Operating System,” in *ICRA Workshop on Open Source Software*, vol. 3, no. 3.2, 2009, p. 5.