Towards the Effectiveness of Ambiguous Spatial Descriptions in Human Robot Interaction

Christopher D. Wallbridge
Cardiff University and University of Plymouth
Cardiff, UK
wallbridgec@cardiff.ac.uk

Alex Smith
Bristol Robotics Laboratory
Bristol, UK
alex.smith@brl.ac.uk

Manuel Giuliani
Bristol Robotics Laboratory
Bristol, UK
manuel.giuliani@brl.ac.uk

Chris Melhuish
Bristol Robotics Laboratory
Bristol, UK
chris.melhuish@brl.ac.uk

Tony Belpaeme
IDlab - imec, Ghent University and University of Plymouth
Ghent, Belgium
tony.belpaeme@ugent.be

Séverin Lemaignan
Bristol Robotics Laboratory
Bristol, UK
severin.lemaignan@brl.ac.uk

1 INTRODUCTION

The nuclear decommissioning site Sellafield in the UK is estimated to cost £162 billion and require 100 years to clean up [9]. The UK nuclear industry is looking to reduce the time and cost by investing in technology such as advanced robot tele-operation, which can reduce the amount of human exposure to high radiation environments. Such challenging and dangerous environments require high situational awareness. One method we can use to help enhance this awareness is by having robots using spatial referring expressions.

Spatial referring expressions are used by people to describe an object by its location. For example an assistive robot may use an expression like "You left your keys *under* the folder *on* the desk" [10]. They are one of the main ways that people describe objects, and are often used even when the object could be identified in another way, such as by using colour. Even if we have another way of identifying the object –such as a circle around the object in an augmented reality system– people often prefer to still have a description [13].

The usual assumption in robotics is that a unique description for the object is best [2]. We refer to these descriptions as "non-ambiguous" as they create a description that identified a single object or location without the possibility of confusion. These systems cite Gricean Maxims [3], especially taking a very rigid approach to the maxim of manner. In a highly complex environment this can result in a combination explosion as more and more objects need to be checked, and their relations to other objects. From there a lot of work has been conducted to try to reduce this explosion [4, 6, 8].

Both generating and evaluating referring expressions are often based on upon a single direction of communication, such as in Williams and Scheutz [14]. This often neglects the communication that a description is for [5]. When we look at communications between two participants we can see this process is actually highly dynamic [12]. Social linguistics suggests that this natural way of communication is allows for the *least collaborative effort* [1] and that full descriptions to reach complete alignment are only necessary when there is difficulty [7]. Wallbridge et al. [11] looks into using a dynamic method of generating spatial referring expressions and see some efficiency gains over using a non-ambiguous description method that seem to disappear over time. In the game like task presented there was also a preference for the non-ambiguous descriptions.

2 METHODOLOGY

We proposed expanding upon the work in Wallbridge et al. [11] by looking at using a dynamic description method –under specified, ambiguous descriptions with follow up repair– to the task of cleaning up nuclear waste. By using a physical robot in a physical task we increase the complexity of description required. It also creates a longer task, so that we can see if non-ambiguous descriptions may become better than dynamic methods over time. What follows is a small preview of the methodology and results that were obtained.

We designed a scenario in which users tele-operate a robot to sort barrels of nuclear waste for disposal. Users were given an interface to control the robot, and multiple camera views in which to view a room where a number of barrels were placed. The robot was able to provide descriptions on which barrels needed to be removed to the participant piloting the robot. Participants would see two rounds with non-ambiguous and dynamic descriptions.

We built a classifier for the dynamic system by looking at human-human interaction. While one participant would pilot the robot as normal another would provide a description, based on the robots understanding of the environment. Based on the work in Wallbridge et al. [12] we identified four categories of tasked based communication: negate, contextual, localised non-ambiguous and positive. We applied these categories to the state of the interaction –based on distance to target, change in distance to target, magnitude of motion from the previous position, yaw required, change in yaw and required change in yaw– to train the classifier.

We measured the time participants took to complete each round of the scenario, as well as the distance travelled. We also gave participants a questionnaire to measure their preferences and perception of the robot.

3 RESULTS AND DISCUSSION

We saw a significant difference between the dynamic and non-ambiguous conditions on time taken (Paired t-test: t = -8.301, df = 29, p < 0.001, mean of dynamic = 390.33s (sd = 92.41) mean of non-ambiguous = 629.53s (sd = 164.35)). We also saw that of our 31 participants, 27 preferred the dynamic condition, only 3 preferred the non-ambiguous condition, and 1 showed no preference.

Our results suggest that by using ambiguity and follow up repair we can increase usability and efficiency over trying to generate a non-ambiguous description when tele-operating a robot. This is in addition to the already identified benefit of reducing combination explosion.

4 ACKNOWLEDGEMENTS

This work was supported by UK Engineering and Physical Sciences Research Council (EPSRC No. EP/R02572X/1) for the National Centre for Nuclear Robotics (NCNR).

REFERENCES

[1] Herbert H Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22, 1 (1986), 1–39.

[2] Robert Dale and Ehud Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science* 19, 2 (1995), 233–263.

[3] H Paul Grice, Peter Cole, Jerry Morgan, et al. 1975. Logic and conversation. *1975* (1975), 41–58.

[4] John D Kelleher and Geert-Jan M Kruijff. 2006. Incremental generation of spatial referring expressions in situated dialog. In *Proceedings of the 21st COLING and the 44th annual meeting of the ACL*. Association for Computational Linguistics, 1041–1048.

[5] Emiel Krahmer and Kees Van Deemter. 2012. Computational generation of referring expressions: A survey. *Computational Linguistics* 38, 1 (2012), 173–218.

[6] Séverin Lemaignan, Raquel Ros, E Akin Sisbot, Rachid Alami, and Michael Beetz. 2012. Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics* 4, 2 (2012), 181–199.

[7] Martin J Pickering and Simon Garrod. 2006. Alignment as the basis for successful communication. *Research on Language and Computation* 4, 2-3 (2006), 203–228.

[8] Raquel Ros, Séverin Lemaignan, E Akin Sisbot, Rachid Alami, Jasmin Steinwender, Katharina Hamann, and Felix Warneken. 2010. Which one? grounding the referent based on efficient human-robot interaction. In *19th International Symposium in Robot and Human Interactive Communication*. IEEE, 570–575.

[9] James Temperton. 2017. Inside Sellafield: how the UK's most dangerous nuclear site is cleaning up its act. https://www.wired.co.uk/article/inside-sellafield-nuclear-waste-decommissioning

[10] Jette Viethen and Robert Dale. 2008. The use of spatial relations in referring expression generation. In *Proceedings of the Fifth International Natural Language Generation Conference*. Association for Computational Linguistics, 59–67.

[11] Christopher David Wallbridge, Séverin Lemaignan, Emmanuel Senft, and Tony Belpaeme. 2019. Generating Spatial Referring Expressions in a Social Robot: Dynamic vs Non-Ambiguous. *Frontiers in Robotics and AI* 6 (2019), 67.

[12] Christopher David Wallbridge, Séverin Lemaignan, Emmanuel Senft, Charlotte Edmunds, and Tony Belpaeme. 2018. Spatial Referring Expressions in Child-Robot Interaction: Let's Be Ambiguous!. In *4th Workshop on Robots for Learning (R4L) - Inclusive Learning @HRI 2018*.

[13] Tom Williams, Matthew Bussing, Sebastian Cabrol, Elizabeth Boyle, and Nhan Tran. 2019. Mixed Reality Deictic Gesture for Multi-Modal Robot Communication. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 191–201.

[14] Tom Williams and Matthias Scheutz. 2017. Referring expression generation under uncertainty: Algorithm and evaluation framework. In *Proceedings of the 10th International Conference on Natural Language Generation*. 75–84.