Adapting Mixed Reality Robot Communication to Mental Workload

Nhan Tran
Colorado School of Mines, Department of Computer Science
nttran@mines.edu

ABSTRACT

This paper describes early work in the intersection of Mixed Reality for Human-Robot Interaction and Brain-Computer Interface fields. Our research seeks to answer these two questions: (1) How do different types of mental workload impact the effectiveness of different robot communication modalities? (2) How can a robot select the effective communication modality given information regarding its human teammate's level and type of mental workload?

KEYWORDS

human-robot interaction, extended reality, mixed reality, augmented reality, brain-computer interface

ACM Reference Format:
Nhan Tran. 2020. Adapting Mixed Reality Robot Communication to Mental Workload. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI '20 Companion), March 23–26, 2020, Cambridge, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3371382.3377438

1 INTRODUCTION AND BACKGROUND

Mixed reality technologies that integrate virtual objects into the physical world have sparked research interest in the Human-Robot Interaction (HRI) community [13] because they enable better exchange of information between people and robots, in order to improve shared mental models, calibrated trust, and situation awareness [10]. Consider a scenario presented in [14] that involves a human teammate and an armless robot like a wheelchair or drone. While mounting arms on these robots can be mechanically infeasible or cost-intensive, mixed reality visualizations of robot arms can simply and cheaply enable these robots to gesture as they have a physical arm. Previous work established a taxonomy of such forms of *mixed reality deictic gesture*, including physical gestures, augmented reality (AR) annotations, and combinations thereof [12, 14]:

- Egocentric gestures: Physical gestures performed by the speaker.
- Allocentric gestures: AR gestures annotating the speaker's target referent from the addressee's perspective (e.g., an AR circle or arrow drawn around or pointing to an object).
- Perspective-free gestures: Gestures that change how all observers perceive the world, that are not tied to the perspective of any one agent (e.g., projecting a light on an object).
- Ego-sensitive allocentric gestures: AR gestures indicating the speaker's referent within the addressee's perspective but performed as if generated from the speaker's perspective (e.g., a robot pointing with a simulated AR arm).
- Ego-sensitive perspective free gestures: Gestures that change how all observers perceive the world, but that are performed as if generated from the speaker's perspective (e.g., projecting an arrow from the robot to its referent).

Hirshfield et al. [3] suggest several contextual factors that may influence the scenarios in which mixed reality deictic gestures can become helpful to human teammates: teammates' cognitive load may dictate whether they are capable of accepting new information; and their auditory and visual perceptual load may dictate the most effective modality to accompany or replace natural language. These neural correlates of cognitive and perceptual states can be collected using a neurophysiological sensor such as functional Near-Infrared Spectroscopy (fNIRS). fNIRS, a lightweight and non-invasive device, is gaining popularity in the Human-Computer Interaction community [9], as it offers several advantages over other brain-computer interface (BCI) technologies such as greater spatial resolution, higher signal-to-noise ratio, and better practicality for use in normal working conditions [8]. While there has been some work combining AR and neurophysiological technologies like the electroencephalography (EEG) [2], to our knowledge, there has been no previous attempt to combine these technologies in the context of human-robot communication. Our work seeks to integrate mixed reality and fNIRS technologies to inform real-time adaptation of robots' deictic language and mixed reality gestures, in order to answer the following key research questions:

(1) How do different types of mental workload impact the effectiveness of different robot communication modalities?
(2) How can a robot select the effective communication modality given information regarding its human teammate's level and type of mental workload?

2 CURRENT WORK

2.1 System Architecture

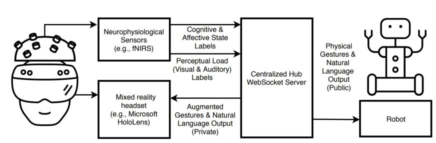

Figure 1: System architecture.

We have developed a robust communication interface, as shown in Figure 1, that enables duplex data transmission among the mixed

HRI '20 Companion, March 23–26, 2020, Cambridge, United Kingdom

ACM ISBN 978-1-4503-7057-8/20/03.
https://doi.org/10.1145/3371382.3377438

reality headset Microsoft HoloLens, the Pepper robot from SoftBank Robotics, and the fNIRs system (in progress). Setup involves starting the WebSocket server on a centralized computer and connecting with the WebSocket client on the HoloLens side, and with Pepper and fNIRs ends. After all clients connect to the same WebSocket server, they are capable of publishing and subscribing to real time messages to each other via bidirectional connection.

The fNIRS component, developed by our collaborator at a nearby university, handles raw data from the sensor and outputs a multilabel vector consisting of four labels (workload, negative affect, auditory perceptual load, and visual perceptual load) from a multilabel long short-term memory (LSTM) classifier every second. These labels are sent to and processed by the centralized server, which then communicates the appropriate decision to both the robot and the mixed reality headset. In the next few weeks, we plan on integrating all components with the Distributed Integrated Affect Reflection and Cognition (DIARC) architecture to leverage its affect processing and deep natural language features [7].

2.2 Experiment

To investigate our two overarching research questions, we are preparing to run an experiment to explore how human teammates perceive mixed reality deictic gestures, and how such gestures interact with the teammates' perceptual and cognitive load (as measured with fNIRS). In particular, we are interested in these effects when allocentric mixed reality deictic gesture is compared to or paired with complex natural language expressions.

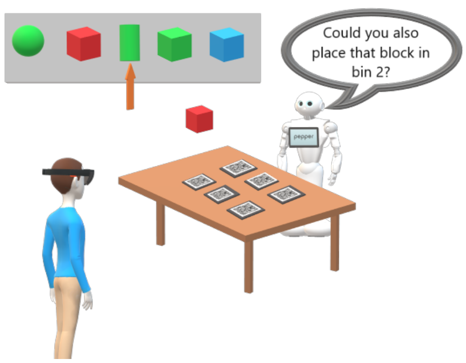

Figure 2: Our experimental setup.

Our experimental design uses a dual-task paradigm oriented around a tabletop pick-and-place task. Participants view this task through the Microsoft HoloLens, allowing them to see virtual bins overlaid over the markers on the table, as well as a panel of blocks above the table that changes every few seconds (Figure 3). As shown in Figure 2, the Pepper robot is positioned behind the table, ready to interact with the participant.

The task consists of a set of 12 within-subject rounds, counterbalanced using a Latin square design. Over the course of these rounds, we systematically vary the communication modalities the robot uses to refer to target blocks, as well as participants' expected cognitive and perceptual load, which we manipulate using techniques such as those proposed by Beck and Lavie [1], who do so by varying the discriminability between target and distractor stimuli (cp. [4]).

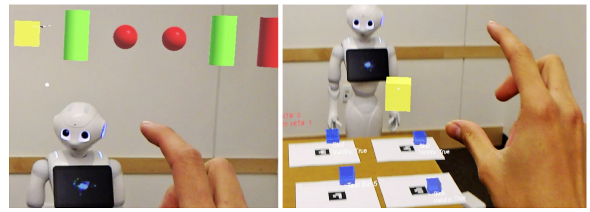

Figure 3: Experiment in progress.

To formulate our hypotheses regarding the effect of communication strategies on mental workload and vice versa, we refer to prior theoretical work on human information processing, including Multiple Resource Theory [11], the Perceptual Load model [5], and the Dual-Target Search model [6], which allow us to predict which type of gesture or visual aid would be most appropriate for a teammate under various levels of cognitive and perceptual load.

The data from this experiment will help us craft high level design guidelines regarding use of specific communication modalities in contexts with specific expected perceptual and cognitive load profiles. However, a single communication modality is unlikely to be sufficient in any given context, as these load profiles may dynamically change within or between tasks. Instead, we argue that an adaptive system is needed, which can be sensitive to both the current context and the predicted effect of potential choices of communication modality. Therefore, we plan on developing such an adaptive model for communication modality selection using probabilistic modeling techniques such as Markov Decision Processes (MDP).

Our reward function for this MDP is expected to be informed by (1) the level of workload the participant is under, (2) the importance of the participant's current task in relation to the new task, and (3) the discriminability that a gesture would have –and that a new target would have if described through language–with respect to both the existing visual search target and the distractors.

3 FUTURE WORK

Following these two upcoming studies, we will investigate other types of visual gestures: Ego-sensitive allocentric gesture and Ego-sensitive perspective free gesture. Future work will consider (1) how these visual gestures will be perceived when accompanied by natural language under different levels of cognitive and perceptual load; (2) how robots can generate these different gestures so that they are easily discriminable from both background visual stimuli, and other task targets; and (3) creating an efficient architecture that handles high speed data transfer between the neurophysiological and mixed reality subsystems.

ACKNOWLEDGEMENT

This research was funded in part by NSF grant IIS-1909864.

REFERENCES

[1] Diane M Beck and Nilli Lavie. 2005. Look here but ignore what you see: effects of distractors at fixation. *Journal of Experimental Psychology: Human Perception and Performance* 31, 3 (2005), 592.

[2] Tathagata Chakraborti, Sarath Sreedharan, Anagha Kulkarni, and Subbarao Kambhampati. 2017. Alternative modes of interaction in proximal human-in-the-loop operation of robots. *arXiv preprint arXiv:1703.08930* (2017).

[3] Leanne Hirshfield, Tom Williams, Natalie Sommer, Trevor Grant, and Senem Velipasalar Gursoy. 2018. Workload-driven modulation of mixed-reality robot-human communication. In *Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data*. ACM, 3.

[4] Nilli Lavie. 1995. Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance* 21, 3 (1995), 451.

[5] Nilli Lavie. 2006. The role of perceptual load in visual awareness. *Brain research* 1080, 1 (2006), 91–100.

[6] Alex Muhl-Richardson, Katherine Cornes, Hayward J Godwin, Matthew Garner, Julie A Hadwin, Simon P Liversedge, and Nick Donnelly. 2018. Searching for two categories of target in dynamic visual displays impairs monitoring ability. *Applied cognitive psychology* 32, 4 (2018), 440–449.

[7] Matthias Scheutz, Thomas Williams, Evan Krause, Bradley Oosterveld, Vasanth Sarathy, and Tyler Frasca. 2019. An overview of the distributed integrated cognition affect and reflection diarc architecture. In *Cognitive Architectures*. Springer, 165–193.

[8] Abdul Serwadda, Vir V Phoha, Sujit Poudel, Leanne M Hirshfield, Danushka Bandara, Sarah E Bratt, and Mark R Costa. 2015. fNIRS: A new modality for brain activity-based biometric authentication. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 1–7.

[9] Erin Treacy Solovey, Audrey Girouard, Krysta Chauncey, Leanne M Hirshfield, Angelo Sassaroli, Feng Zheng, Sergio Fantini, and Robert JK Jacob. 2009. Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 157–166.

[10] Daniel Szafir. 2019. Mediating Human-Robot Interactions with Virtual, Augmented, and Mixed Reality. In *International Conference on Human-Computer Interaction*. Springer, 124–149.

[11] Christopher D Wickens. 2002. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* 3, 2 (2002), 159–177.

[12] Tom Williams. 2018. A framework for robot-generated mixed-reality deixis. In *Proceedings of the 1st International Workshop on Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)*.

[13] Tom Williams, Daniel Szafir, Tathagata Chakraborti, and Heni Ben Amor. 2018. Virtual, augmented, and mixed reality for human-robot interaction. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 403–404.

[14] Tom Williams, Nhan Tran, Josh Rands, and Neil T Dantam. 2018. Augmented, mixed, and virtual reality enabling of robot deixis. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 257–275.