

Comparative Study of Voice-Activated vs Controller-Based Manipulation in Pick-and-Place Robotics

Neel Girish Bahadarpurkar
MS Robotics Engineering
Worcester Polytechnic Institute
Worcester, Massachusetts
nbahadarpurkar@wpi.edu

Harshal Suresh Bhat
dept. name of organization (of Aff.)
Worcester Polytechnic Institute
City, Country
email address or ORCID

Brendan Byrne
BS Robotics Engineering & Computer Science
Worcester Polytechnic Institute
Marshfield, Massachusetts
bsbyrne@wpi.edu

Jack Lynch
MS Computer Science
Worcester Polytechnic Institute
Shrewsbury, Massachusetts
jmlynch@wpi.edu

Paul Raynes
MS Robotics Engineering
Worcester Polytechnic Institute
Worcester, Massachusetts
psraynes@wpi.edu

Abstract—This study compares the performance of voice instructions against mechanical controls in a simulated work setting. Participants in a Unity-based simulation are charged with commanding autonomous robots to fetch certain things using mechanical controllers or vocal instructions. The NASA-TLX approach is used in the experiment to assess job completion time and obtain user input on workload and preferences. The study intends to determine if voice control may improve productivity and usability in real-world applications, hence leading to a better understanding of human-robot interaction in industrial settings.

I. INTRODUCTION

With autonomous robots becoming more prevalent in work-place settings, determining the most efficient means of commanding them becomes an important task. As human workers may have their own tasks to perform in addition to commanding the machines, finding an efficient control method that can reduce the overall workload of the human operator becomes important. Currently, mechanical controls are the most widespread, as they are both easy and cheap to implement. Another control method that has become more popular has been voice communication. While voice communication seems to be more efficient on paper, there can of course be issues in certain scenarios. Our project aims to determine if voice commands are preferable to normal basic commands in a work environment by creating a simulation with both mechanical and voice commands to determine which one people prefer.

II. RELATED WORK

Past research work has investigated the efficacy of various robot control techniques, focusing on the comparison between manual control and voice commands. Bremner and Leonards (2015) examined the integration of speech and gestures for

robotic communication. In contrast to manual control techniques, which need more direct and accurate input, their study revealed that mixing voice commands with gestures improves both robot likeability and efficiency, making interactions more intuitive for users.

Comparing autonomous control to voice control, Vatsal and Hoffman (2018) investigated a wearable robotic forearm intended for cooperative tasks. Their research showed that although voice instructions had their uses, people preferred autonomous control when doing repeated and difficult activities since it was faster and more efficient. This implies that although voice instructions have their advantages, there are situations in which they might not be as effective as alternative approaches.

A framework for classifying voice-controlled and manual human-robot conversation systems was created by Berzuk and Young in 2022. Their study demonstrated the benefits of multimodal communication (speech, gestures, etc.), demonstrating that in some situations, these systems perform better than only voice- or manual-based control because they allow for more flexible and adaptive interactions.

Finally, in 2017, research from the CARIS lab showed that voice-controlled robots are now more widely available in industrial settings thanks to natural language commands, which greatly increase the usability of voice-controlled robots for non-expert users. This method contrasts with manual control methods, which demand more concentration and training.

These studies highlight that although voice command systems provide a more natural interface, depending on the job complexity and user expectations, there are situations in which manual control or autonomous systems may provide greater performance or preference. This context guides our study's exploration of whether voice commands can function

more effectively in a workplace setting than conventional mechanical controls.

III. PROPOSED METHOD

To determine the effectiveness of voice controls in a supervisory control scheme, our method will compare voice controls to a typical mechanical control scheme. We will create a 3D task environment which will simulate a potential real-life usage for supervisory control in the Unity engine. This simulation will contain four rooms, one which the participant's character in the simulation will work from, as well as three rooms which house three different autonomous agents, which all have an opening leading to the user's room. The user will be unable to enter the robots' rooms. Within each room will be nine objects, each with a unique shape and color combination that can be collected by the robots and brought to the user through the opening. Each of these shapes will be labelled 1 through 3 depending on which agent's room it came from. There will also be nine unlabeled objects within the user's room. Outside of each room will be a display which the user can use to control the robots mechanically.

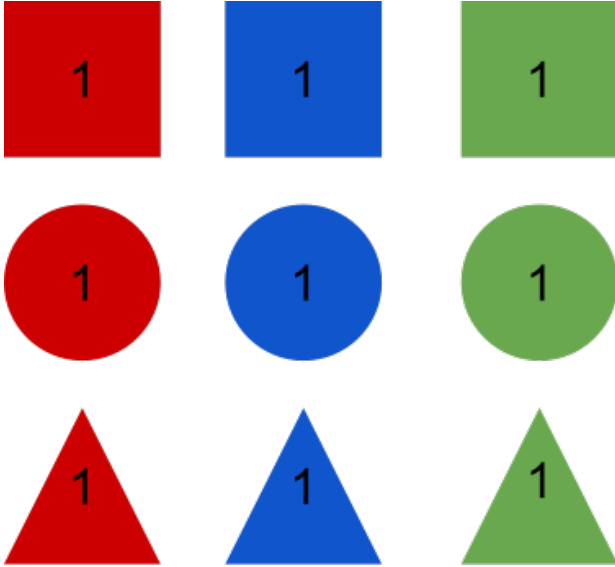


Fig. 1. Example of what the shapes will look like

IV. EXPERIMENT

During the experiment, participants will be shown a sequence of shapes that they need to construct using an object from their room as well as one from the three adjacent rooms. As the participant cannot enter the adjacent rooms, they must command the agents to provide the correct objects for them. Once a set number of sequences are completed by the user, the time it took to accomplish the task will be recorded for evaluation.

There will be two versions of the simulation. The first version will use the displays found outside the rooms of the robots which the user will utilize to obtain the needed objects for the sequence. This will simulate typical mechanical controls for



Fig. 2. Example of a shape sequence

the automata. In the second version of the simulation, instead of the displays, voice commands will be used. Participants will be requested to speak into their microphones to request objects from one of the three robots. An example command could be "Robot 1, bring me the red square".

Evaluation of both methods will be conducted by comparing the times needed to complete the sequences in both the mechanical and voice activated trials. In addition to the time, feedback from the participants in the study will also be collected to analyze both workload levels as well as the overall user experience. A subjective usability test such as the NASA-TLX or a Likert scale will be used to collect this feedback.

V. PRELIMINARY RESULTS

Our first deliverable for this project is a basic unity simulation which will be built upon to create our final simulation. Currently, the simulation is incredibly simple and only serves to be a foundation from which the rest of the simulation can be built upon. This simulation features a rectangular block which acts as the robot and two platforms, a red one on the right and a blue one on the left. With the A and D keys, the block moves between the two platforms. This is to give an idea of how the robot will move between various objects and points. As the experiment will involve robots going between various points to pick up and deliver objects.

As of right now, the movement is a bit buggy, as the block attempts to clip into the platform. This is a result of the MoveTowards function, which moves an object to the position of another object. However, this attempts to move the block to the center of mass of the platform, which causes a 'clipping' effect as the block attempts to move inside the platform. While this is merely a stylistic thing, it does have the potential to become an issue down the line if the robots are unable to make it to their destination due to this clipping. The chances of this becoming an issue is slim, but it is still worth mentioning as a possible issue.

VI. SCHEDULE

TABLE I
TABLE TYPE STYLES

Task	Start Date	Due Date
Make Unity Simulation Tests	<i>Table column subhead</i>	<i>Subhead</i>
copy	More table copy ^a	

ACKNOWLEDGMENT

The preferred spelling of the word "acknowledgment" in America is without an "e" after the "g". Avoid the stilted

expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

REFERENCES

Please number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955.
- [2] J. Clerk Maxwell, *A Treatise on Electricity and Magnetism*, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
- [3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] K. Elissa, “Title of paper if known,” unpublished.
- [5] R. Nicole, “Title of paper with only first word capitalized,” *J. Name Stand. Abbrev.*, in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” *IEEE Transl. J. Magn. Japan*, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p. 301, 1982].
- [7] M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.

IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove the template text from your paper may result in your paper not being published.