# Different Confirmation Methods of Baxter Robot for Deictic Gesture

Ilke Kas, Zhenyu Yang, and Jiayi Chen, *Case Western Reserve University, IEEE*

***Abstract*— This study investigates different confirmation methods employed by the Baxter robot for deictic gestures in Human-Robot Interaction (HRI). Motivated by the need to enhance user experience, we explore the impact of various confirmation modalities on the retrieval of a pointed object.**

## I. Introduction

In Human-Computer Interaction, certain scenarios necessitate confirming computer actions. For instance, using Homepod to instruct Siri to control a humidifier may lack feedback, leading to potential misinterpretations. To enhance user experience, we explore if incorporating diverse confirmation methods can mitigate variability and improve user interaction. This study investigates the impact of robot confirmation methods on user experience while the Baxter robot (the robot) retrieves a designated object.

We designed an experiment and analyzed the data to learn how different confirmation methods in HRI with gestures impact the user experience in terms of likability, animacy, etc. We proved that different kinds of confirmation methods play an important role in HRI, and we also found out which kind of confirmation method that participants have the best user experience with.

## II. Previous research

Most of the existing study in the literature focuses on improving the technicality of dynamic hand gesture recognition, hand pose estimation, eye tracking and object recognition in addition to the accuracy of completion of the task to enhance HRI [1][2]. One of the previous studies that make participants use deictic gestures in a Pick-and-Place task shows that any kind of confirmation gestures and methods improve the task completion accuracy and efficiency in addition to user experience [3]. However, this study only uses a visual confirmation method by showing the image of the object to the participants [3].

There are other studies which focus on the communication side of the HRI studies. One of these studies mentioned the impact of visual, auditory, tactile and other non-verbal cues to improve human-robot communication [4]. However, it does not have many experimental results regarding their difference in terms of the user experience.

## III. Proposed Study

The proposed study explores what robot confirmation methods are preferred by the users. The main goal of this exploration is to improve the user experience by analyzing the likability, anthropomorphism, animacy, perceived intelligence and safety of the different confirmation methods. For this purpose, we used four different methods:

- *Visual Confirmation Method*: The image of the pointed object is shown to the participant on the screen of the robot.
- *Verbal Confirmation Method*: The robot verbally asked participants the color of the object.
- *Body Movement Confirmation Method*: The robot used its arm to confirm the pointed object. It moves its arm towards the object pointed and it closes and opens its gripper several times to ask the pointed object.
- *No Confirmation Method (Baseline Method)*: The robot did not use any confirmation method.

In this work, we investigate the following exploratory experimental question: "How do the different confirmation methods of the robot (visual, verbal, body movement or no confirmation) affect the user experience during the retrieval of the pointed object?". We are expecting that there is a significant difference in user experience when the robot employs different confirmation methods (visual, verbal or body movement) during the retrieval of a pointed object using a deictic hand gesture (**Hypothesis 1**).

## IV. Methods

During the experiment, we deceived the participants by stating it can detect their deictic hand gesture to perceive the pointed object. However, during the experiment we controlled the robot by using the Robot Operating System (ROS) terminal ourselves by running previously decided commands [5]. We located the experimenter in front of the computer in a way that s/he can see the pointed object by the participant. In this way, s/he can run the corresponding command according to the pointed object.

### A. Visual Confirmation Method

For the visual confirmation method, we show the three different colored cube (red, blue and red) images on the screen of the robot by running the corresponding ROS command for this action [5].

### B. Verbal Confirmation Method

For the verbal confirmation method, we placed an unseen JBL speaker behind the robot. We connected one of the experimenters' phones to that JBL speaker to play three different audio for three different colored cubes according to the color of the pointed object.

### C. Body Movement Confirmation Method and Handing Over the Object

For the body movement confirmation method and handing over the object , we fixed the places of the cubes on the table firstly. After that, we saved 3 different arm routines by controlling the arm from our keyboard [5]. Every time a participant points to one of the cubes, we replayed the corresponding arm routine to that cube [5].

## V. Experiments

### A. Preparation

Before the experiment, we calculated the number of participants needed by using power analysis [6]. We used ANOVA test with effect size 0.5, alpha error probability 0.05, power 0.8, numerator df 3 and number of groups 4 [6]. This calculation gives us 48 participants [6]. We were able to find 30 participants for this experiment. Each takes 30 minutes to complete.

### B. Procedure

After participants signed informed consent and media released forms, they followed the following procedure 4 times for each different confirmation methods:

1. Participants point the object on the table
2. Participants confirm the object by using one of the four ways mentioned above
3. Participants say explicitly “Yes”/ “No” to confirm the object
4. Participants take the object given by the robot
5. Participants fill a survey given by experimenter

### C. Measurement

We use 18 questions from the Godspeed Questionnaire, which has 5 scale Likert assessing anthropomorphism, animacy, likability, perceived intelligence, and perceived safety [7]. Besides that, we asked one open-ended question: “What do you like/dislike about this interaction? ”.We used Qualtrics to perform this survey and get the results of it.

Besides that, we recorded the video of the participants during the interaction to perform face analysis for analyzing their emotions and to perform speech sentiment analysis.

#### *Face Analysis for Emotion Recognition*

Face Analysis is performed on the recorded video of the participants during the interaction. We used the Facial Expression Recognizer (FER) library which can recognize 6 different expressions [8]. However, we only used the dominant one of these three emotions during the whole interaction for the corresponding confirmation method: Neutral, Happiness and Sadness.

#### *Sentiment Analysis*

Sentiment Analysis performed on both of the speech data of the participants during the interaction and on the answers of the open-ended question in the survey. We used the Sentiment Analyzer of the Natural Language Toolkit (NLTK) [9]. To extract the speech data to text data, we used the AI tool for speech to text of Microsoft Clipchamp application [10]. Sentiment analysis gave us three different values of the collected data: Neutral, Positive or Negative.

## VI. Result

In light of the conducted experiments and analyses, the present study has yielded significant findings that contribute to our understanding of direct gesture confirmation methods in HRI, providing valuable insights into verbal confirmation, visual confirmation, no confirmation and body confirmation.

### A. Survey

In the survey, we use Analysis of variance (ANOVA) to analyze participants’ response to 18 different questions in the Likert scale. Through the analysis of the data collected from the survey, we found that there is a significant difference in user experience when the robot employs different confirmation methods (no confirmation, body, verbal, or visual) during the retrieval of a pointed object using a deictic hand gesture. We calculated the mean and one-way variance of the values obtained from 30 participants’ responses to 18 questions in each of the four confirmation methods. We also calculated that the test statistic values of the four different confirmation methods are all greater than the critical value (f crit = 1.642369), and the p-value of the four different methods are all far less than 0.001, which means all the results are significant.

The highest average value and the lowest variance value belong to the verbal confirmation method(average value = 4.16, variance = 0.7702, p-value = 4.31834E-). While the lowest average value and the highest variance value belong to the body confirmation method (average value =3.86 , variance = 0.9923). Participants favored verbal confirmation for the best user experience, while the body confirmation method resulted in the worst experience. Views on the body confirmation method were polarized, with strong dislike from some and lukewarm endorsement from others.

Participants prefer verbal confirmation due to its natural interaction, commonly experienced in daily life. Conversely, the body confirmation method’s extra movements lead to a less natural experience, contributing to participants’ lower acceptance and overall satisfaction.

### B. Face Analysis

Fig. 5-8, depict participants' emotions across different confirmation methods. Surprisingly, a notable proportion exhibits sad facial expressions. Upon reviewing the experiment, we attribute this to the camera angle – from up to down, resulting in low robustness of facial emotion recognition during our study.

### C. Speech Analysis

Fig. 9-11, shows the percentages of the sentiment analysis of the speeches of the participants during the experiment. From these figures, the most positive analysis belongs to the body confirmation robot while the most neutral analysis belongs to the no confirmation methods.

### D. Summarize

The proposed HRI interaction confirmation method utilizes a combination of verbal confirmation, visual confirmation, body’s confirmation, no confirmation and gesture recognition, providing users with a seamless and intuitive way to confirm actions.

Though the results of speech recognition analysis, facial emotion recognition analysis, p-value and ANOVA, verbal confirmation method is the popular method. Because this kind of confirmation method has a more direct, more intuitive way of interaction than other methods.

## VII. Conclusion And Future Works

In our experiment, the verbal confirmation method has the highest mean value in ANOVA, and the research of emotion recognition confirms this result. It shows that this kind of confirmation method gives participants more intuitive interaction, better interactivity and more seamless experience. No confirmation method also results with a good experience, but its low level of interaction would not improve the user experience of HRI.

In the future, as robots become integral across industries, HRI will extend beyond verbal communication. The evolving landscape of Virtual Reality (VR) technology will foster increased collaboration between robots and VR. Consequently, visual interaction and potentially unconfirmed interactions will gain prominence. Enhancing the seamlessness of alternative interaction methods is crucial for ensuring a superior user experience.

## VIII. FIGURES

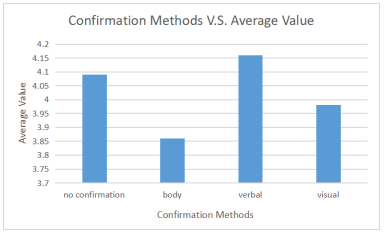


Fig. 1. The average value for four different confirmation methods

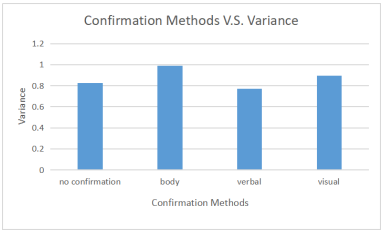


Fig. 2. The one-way variance for four different confirmation methods

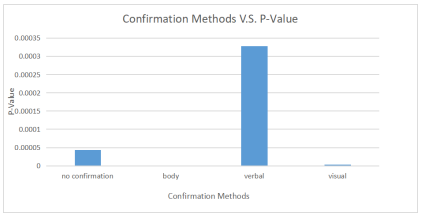


Fig. 3. The p-value for four different confirmation methods

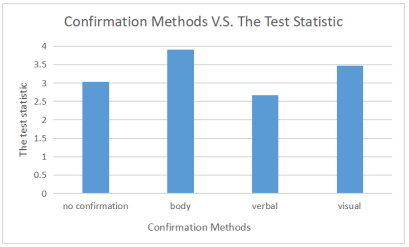


Fig. 4. The test statistic for four different confirmation methods

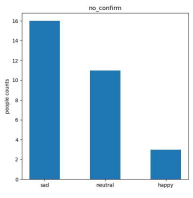


Fig. 5. no confirmation

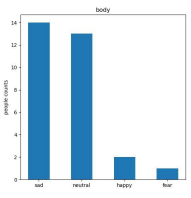


Fig. 6. body's confirmation

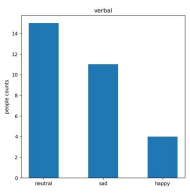


Fig. 7. verbal confirmation

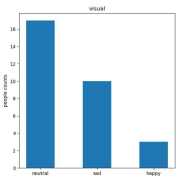


Fig. 8. visual confirmation

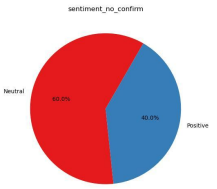


Fig. 9. no confirmation

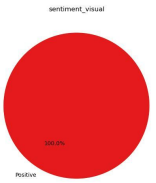


Fig. 10. visual confirmation

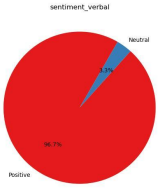


Fig. 11. verbal confirmation

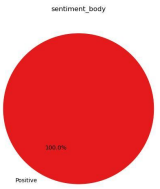


Fig. 12. body's confirmation

## REFERENCES


[1] Q. Gao, Y. Chen, Z. Ju and Y. Liang, "Dynamic Hand Gesture Recognition Based on 3D Hand Pose Estimation for Human–Robot Interaction," in IEEE Sensors Journal, vol. 22, no. 18, pp. 17421-17430, 15 Sept.15, 2022, doi: 10.1109/JSEN.2021.3059685. (https://ieeexplore.ieee.org/abstract/document/9427388)

[2] Li, W. et al. (2021). A Novel Gaze-Point-Driven HRI Framework for Single-Person. In: Gao, H., Wang, X. (eds) Collaborative Computing: Networking, Applications and Worksharing. CollaborateCom 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 406. Springer, Cham. https://doi.org/10.1007/978-3-030-92635-9_38 (https://link.springer.com/chapter/10.1007/978-3-030-92635-9_38#citeas)

[3] C. P. Quintero, R. Tatsambon, M. Gridseth and M. Jägersand, "Visual pointing gestures for bi-directional human robot interaction in a pick-and-place task," 2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Kobe, Japan, 2015, pp. 349-354, doi: 10.1109/ROMAN.2015.7333604. (https://ieeexplore.ieee.org/document/7333604)

[4] Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective. Urakami, Jacqueline, and Katie Seaborn. "Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective." ACM Transactions on Human-Robot Interaction 12.2 (2023): 1-21. https://dl.acm.org/doi/10.1145/3570169#d1e616

[5] Ros Robotics by example - university of houston–clear lake, https://sceweb.sce.uhcl.edu/harman/CENG5437_MobileRobots/Webitems2020/ROS_ROBOTICS_BY_EXAMPLE_SECOND_EDITION.pdf (accessed Dec. 12, 2023).

[6] "G*Power," Universität Düsseldorf: G*Power, https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower (accessed Dec. 11, 2023).

[7] C. Bartneck, "Godspeed questionnaire series: Translations and usage," International Handbook of Behavioral Health Assessment, pp. 1–35, 2023. doi:10.1007/978-3-030-89738-3_24-1

[8] P. Nagarajan, "Face emotion recognition (FER)," Medium, https://medium.com/mlearning-ai/face-emotion-recognition-fer-114ccb359604 (accessed Sep. 26, 2023).

[9] NLTK, https://www.nltk.org/ (accessed Dec. 11, 2023).

[10] "Microsoft," Microsoft Support, https://support.microsoft.com/en-au/topic/how-to-use-the-text-to-speech-feature-1aa3a474-dd42-40f0-803d-d9f78ada0387 (accessed Dec. 11, 2023).