Closing the feedback loop: the relationship between input and output modalities in human-robot interactions

Thesis submitted in partial fulfillment of the requirements for the M.Sc degree

By: Tamara Markovich
Closing the feedback loop: the relationship between input and output modalities in human-robot interactions

Thesis submitted in partial fulfillment of the requirements for the M.Sc degree

By: Tamara Markovich

Supervised by Professor Tal Oron-Gilad

April 2020
Abstract

Previous studies suggested that communication modalities used for human control and robot feedback influence human-robot interactions. However, they generally tended to focus on one part of the communication, ignoring the relationship between control and feedback modalities. We aim to understand whether the relationship between a user’s control modality and a robot’s feedback modality influences the quality of the interaction, in terms of subjective experience and task performance (Reaction times and accuracy). First, a preliminary experiment was conducted to identify intuitive gestures and vocal commands for controlling the robot, as well as intuitive robot movements and vocal statements for providing feedback to the user. This preliminary experiment led to the creation of a user-driven communication vocabulary for guiding the motions of a mobile robot. Based on these control and feedback vocabularies, we ran our main experiment for evaluating how the relationship between control modalities and feedback modalities impacts the human-robot interaction. In a laboratory Wizard-of-Oz experiment, participants were asked to guide a robot through a maze by using either hand gestures or vocal commands. The robot provided vocal or motion feedback to the users across the experimental conditions forming different combinations of control-feedback modalities. We found that the combinations of control-feedback modalities affected the quality of human-robot interaction (subjective experience and efficiency) in different ways. Participants showed less worry and were slower when they communicated with the robot by voice and received vocal feedback, compared to gestural control and receiving vocal feedback. In addition, they felt more distress and were faster when they communicated with the robot by gestures and received motion feedback compared to vocal control and motion feedback. We also found that providing feedback improves the quality of human-robot interaction: participants felt more distress and were less accurate when the robot didn’t provide feedback. In this paper, we detail the procedure and results of this experiment.

Keywords: Human-robot interaction • Feedback loop • Navigation task • Feedback by Motion Cues • Stimulus-response compatibility
Participation in Academic Conferences

Oral Presentation


Peer reviewed proceeding papers and Oral presentation

Acknowledgments

First, I would like to express my sincere gratitude to my supervisor: Prof. Tal Oron-Gilad for her continuous support, guidance, and counseling of my research with expertise knowledge and motivation. I would also like to thank Dr. Yisrael Parmet for his assistance with the statistical analysis during the research. In addition, I would like to thank Shanee Honig for her devoted guidance and her help with planning the experiment and with the writing. I would also like to thank Shai Givati for his technical support with the robot.

This research was supported in part by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative and by the Marcus Endowment Fund both at Ben-Gurion University of the Negev.
# Contents

Abstract ........................................................................................................................................... I
Participation in Academic Conferences ......................................................................................... II
Acknowledgments ............................................................................................................................. III
Contents .............................................................................................................................................. IV
List of Figures ..................................................................................................................................... VI
List of Tables ...................................................................................................................................... VI

1 Introduction .................................................................................................................................... 1

2 Related Work ................................................................................................................................. 2
   2.1 Feedback modalities .................................................................................................................. 2
   2.2 Human control modalities ......................................................................................................... 4
   2.3 The Interaction between Control and feedback modalities ..................................................... 6

3 Goals and Hypotheses ..................................................................................................................... 8

4 Preliminary experiment .................................................................................................................. 9
   4.1 Method ...................................................................................................................................... 10
      4.1.1 Overview .......................................................................................................................... 10
      4.1.2 Participants ...................................................................................................................... 11
      4.1.3 Procedure ....................................................................................................................... 11
      4.1.4 Analysis .......................................................................................................................... 12
   4.2 Results ..................................................................................................................................... 13
      4.2.1 First part .......................................................................................................................... 13
      4.2.2 Second part ..................................................................................................................... 17
   4.3 Final feedback and control design ............................................................................................ 20
   4.4 Summary ................................................................................................................................. 20

5 Method ........................................................................................................................................... 21
   5.1 Overview .................................................................................................................................. 21
   5.2 Participants ............................................................................................................................... 22
   5.3 Robot ......................................................................................................................................... 22
   5.4 Experimental Design ................................................................................................................ 22
   5.5 Procedure ................................................................................................................................ 25
   5.6 Measures .................................................................................................................................. 26
   5.7 Analysis .................................................................................................................................... 26
      5.7.1 Video analysis ................................................................................................................... 26
      5.7.2 Subjective Experience ..................................................................................................... 28
      5.7.3 Efficiency ........................................................................................................................ 29
5.7.4 Variables summary ................................................................. 29

6 Results................................................................................................. 30

6.1 Subjective Experience Analysis..................................................... 30
  6.1.1 Distress ............................................................................. 30
  6.1.2 Worry ........................................................................... 32
  6.1.3 Engagement.................................................................... 32
  6.1.4 SUS score......................................................................... 33

6.2 Efficiency....................................................................................... 34
  6.2.1 RT .................................................................................. 34
  6.2.2 Accuracy .......................................................................... 35

6.3 Preferences for communication modalities .................................. 36

6.4 Failures ....................................................................................... 38

7 Discussion.......................................................................................... 40

7.1 The Research Hypotheses.............................................................. 40

7.2 Run duration................................................................................ 42

7.3 Failures ....................................................................................... 43

7.4 Limitations .................................................................................. 43

8 Conclusion......................................................................................... 44

9 References......................................................................................... 46

10 Appendices....................................................................................... 49
List of Figures

Figure 1: the 4-d Multiple Resource Theory diagram (Wickens,2008) .......................................................... 6
Figure 2: Turtlebot3 burger ......................................................................................................................... 11
Figure 3: The distribution of participants selections of the command that they thought as most
suitable for each one of the gestures ........................................................................................................... 18
Figure 4: The distribution of participants selections of the messages that they thought as most suitable
for each one of the motion feedbacks ........................................................................................................... 19
Figure 5: Turtlebot3 burger after connecting the Bluetooth speaker to it ....................................................... 22
Figure 6: The maze. Left. A schematic representation of the maze. Each path starts and ends with the
same number. Right. A pictorial view of the maze markings on the floor .................................................... 23
Figure 7: A demonstration of the gestures the participants were asked to use in the gestural control
condition ......................................................................................................................................................... 24
Figure 8: Distress by feedback modality ........................................................................................................ 31
Figure 9: and distress by the control and feedback modality interaction ....................................................... 31
Figure 10: Worry by control and feedback modality ...................................................................................... 32
Figure 11: Engagement by feedback modality ............................................................................................... 33
Figure 12: SUS score by control modality ...................................................................................................... 34
Figure 13: RT by control modality ................................................................................................................ 34
Figure 14: RT by control and feedback modality ........................................................................................... 35
Figure 15: Accuracy by control modality ....................................................................................................... 36
Figure 16: Participants selections for the last trial ......................................................................................... 37
Figure 17: participants’ preferred combinations of control and feedback modalities .................................... 37
Figure 18: the number of runs distribution by number of misunderstandings ................................................ 38
Figure 19: Distress by number of misunderstandings .................................................................................. 39
Figure 20: Worry by number of misunderstandings ..................................................................................... 39
Figure 21: SUS score by number of misunderstandings ............................................................................... 40

List of Tables

Table 1: the possible commands and feedback messages ............................................................................... 12
Table 2: The set of gestures suggested by participants for each command (first online questionnaire)
that we used for the second questionnaire .................................................................................................. 14
Table 3: The set of motions suggested by participants for each command (first online questionnaire)
that we used for the second questionnaire .................................................................................................. 16
Table 4: The set of most common vocal commands suggested by participants for each command (first
online questionnaire) ..................................................................................................................................... 17
Table 5: The set of most common vocal feedback suggested by participants for each message (first
online questionnaire) ..................................................................................................................................... 17
Table 6: Gesture-user command matching percent ....................................................................................... 18
Table 7: Motion-robot message feedback matching percent .......................................................................... 20
Table 8: The final control design and feedback design ................................................................................... 20
Table 9: the subjects and behaviors used for tagging events during video analysis ....................................... 27
Table 10: Example of the time-event table .................................................................................................... 28
Table 11: Example of the Behavior Analysis table .......................................................................................... 28
Table 12: Tests of model effects for distress ................................................................................................ 31
Table 13: Tests of model effects for worry .................................................................................................... 32
Table 14: Tests of model effects for engagement ........................................................................................... 33
Table 15: Tests of model effects for SUS score ............................................................................................. 33
Table 16: Tests of model effects for lnRT ...................................................................................................... 34
Table 17: Tests of model effects for accuracy .............................................................. 36
Table 18: Tests of model effects for distress ............................................................ 39
Table 19: Tests of model effects for worry ............................................................... 39
Table 20: Tests of model effects for SUS score ......................................................... 40
1 Introduction

The “Feedback loop” is an important feature of interactive systems. It represents the nature of the interaction between a person and a dynamic system (a system like computer, car or robot that simply react to an input). The user provides input to the system in order to achieve a goal, gets output (feedback) from the system and interprets it. This interpretation affects the user’s next actions, beginning the cycle again (Dubberly et al., 2009). In human-human interaction, the feedback that one person receives from another when conversing can be viewed as a collaborative "feedback loop" that is necessary for successful conversations (Clark & Schaefer, 1989). According to Clark and Schaefer (1989), humans use language, gestures, and body language to inform their conversational partners that they understood their communication. To allow the speaker to adapt accordingly, the listener must provide evidence of understanding. Otherwise, the speaker will seek further evidence that the listener understood his communication using repair behavior. Similar interactions have been reported in human-computer interaction. Perez-Quinones and Sibert (1996) showed that human repair behavior in a human-computer dialogue, when a communication expectation is not met, is similar to the repair behavior found in human-human dialogues. Not providing any feedback about the computer’s current state produced some form of repair behavior.

Like in human-human and human-computer interaction, the feedback loop is also important in human-robot interactions. Robots interacting with humans should be able to react to commands given by the user as well as provide feedback back. Some studies emphasize the importance of giving feedback by robots to humans during the interaction (Mirmig et al., 2011; Mohammad & Nishida, 2007). The feedback can be provided in different modalities. Robots can provide information to the human by tactile devices, verbal feedback, and visual feedback like screens or lights and more. Similarly, humans can communicate with the robot by several modalities, including gestures, voice and using a touch screen.

In the literature, different communication modalities have been investigated for the robot’s feedback or the human’s control. For example, when investigating the robot’s feedback, Perrin et al. (2008) presented a qualitative and quantitative evaluation of different types of feedback used to communicate a robot’s navigational decisions to a human user. Redden et al. (2010) studied the influence of different modalities of the human’s commands (voice or manual) to the robot on task performance. While the studies above suggest that the type of communication modalities influence human-robot interactions, they provide an incomplete understanding of the interaction since they focused only on one part of the communication
(from the robot to the human or vice versa), ignoring the relationship between control and feedback modalities. According to Greenwald (1970), there are stimuli modalities that most compatibly mapped to certain response modalities. This statement is an extension of a principle called **Stimulus-Response (S-R) compatibility**. According to this principle, when the relation between displays and controls, or stimuli and responses, is direct and natural, it is described as being compatible. When S-R matchings are compatible, the responses are faster and more accurate than when they are incompatible (Proctor & Vu, 2006). Applying this idea to human-robot interaction, it is possible that there are compatibility effects between control modalities and feedback modalities that may affect the quality of the human-robot interaction by influencing the speed and accuracy of human response, as well as the subjective experience during the interaction.

This research aims to evaluate whether the relationship between a human's control modality and a robot's feedback modality during the interaction influences the quality of the interaction. More specifically, this study evaluates whether there are control modalities that are most compatibly mapped to certain feedback modalities in terms of speed, accuracy, and subjective evaluation. If so, we want to find this mapping. More than that, we want to find the control-feedback modalities combination that would produce the most efficient interaction, depending on the type of task.

## 2 Related Work

In the HRI literature, there are studies that explored the different modalities used to transfer information from human to robot (control) or from robot to human (feedback). In order to understand the interaction between these two directions of communication, it is important to first understand what is already known about each side of the communication.

### 2.1 Feedback modalities

The importance of feedback to the human-robot interaction has already been established. Mirnig et al. (2011) tested the influence of feedback on the successfulness of human-robot communication. In their experiment, the robot asked the participants for directions to a specific destination in a cardboard model town. When the robot provided feedback, it was more likely to be perceived as a social communication partner. In addition, participants that didn't receive feedback during the interaction stated that they would want to receive feedback from the robot. There was also a trend showing that the success rate was higher when feedback was provided, however, this result was not significant. In another study,
Mohammad and Nishida (2007) found a statistically significant improvement in task completion time when using feedback in human-robot interaction compared to a no-feedback condition. The type of sensory modalities chosen to provide feedback to the user can be visual, auditory, tactile, gestural feedback or a combination of different modalities. Several studies present a comparison between different feedback modalities.

Allwood and Cerrato (2003) investigated the role of gestural feedback (movements that are produced to show feedback including head movements, facial expressions, and other gestures) and verbal feedback expressions in human-human communications. Real dialogues between different customers and a travel agent were analyzed to find if gestural feedback expressions always co-occur with vocal feedback expressions or can occur on their own. They found that feedback is mostly expressed simultaneously by vocal and gestural means. A possible relationship between those two kinds of feedback is described: gestures can add redundancy, indicate a positive reinforcing attitude, weaken what has been said vocally or contradict it to denote sarcasm.

In human-robot interactions, Perrin et al. (2008) presented a comparison between six different ways to convey navigational information provided by a robot to a human. Subjects were asked to monitor the decisions of a simulated robot navigating in a virtual maze. Visual, auditory, and tactile feedback modalities were selected to suggest a direction of travel to a human user, who could then decide if he agrees or not with the robot’s proposition. The user had controlled the robot’s actions by a manual control (press button). They found that the visual pictograms (icons containing an arrow pointing left, right, or up) were the most liked feedback approach and the one that provided the quickest and best answers. A possible explanation for their finding is that the cues in this experiment were arrows, identical to the visual pictograms, which helped facilitate decision-making in the task.

Mohammad and Nishida (2007) investigated the effectiveness of using motion cues as a feedback mechanism and compared it to the use of auditory feedback. In their experiment, a human instructed a miniature robot using free hand gestures to help it follow a specific path to a goal in an environment that was projected on the ground. The results showed that there was no significant difference in the task completion accuracy and time or in the feeling of naturalness between these two modalities but there was a statistically significant improvement when using any of them compared with the no-feedback case. However, most subjects selected the motion modality as their preferred modality.
Another aspect of using feedback is the uncertainty of the feedback modality. The modality of feedback is uncertain if there are several possible modalities that can occur. Boulter (1977) used visual, auditory and tactile signals in a reaction task, where a non-discriminative response was required. He found that RT to a signal was longer if the modality of the signal was uncertain (when subjects did not know in which modality the signal would occur). Boulter theorized that a selective process is responsible for the uncertainty effect: when the modality of the signal was uncertain, the selective process was occupied with "identifying" the modality of an RT signal and thus increased RT.

2.2 Human control modalities

The modality channel that the user controls the system through is also an important aspect of human-computer or human-robot interaction. Common controls that were investigated in previous works are manual, verbal and gestural controls.

Verbal and gestural control are natural ways to communicate since we use speech and gestures to communicate in our everyday life. Because of that, it can be quicker and feel more intuitive than manual controls (Hu et al., 2003; Redden et al., 2010; Steeneken, 1996; Xu et al., 2007). Studies have investigated the importance and functional role of hand gestures as part of human-to-human and human-computer communication and tried to understand the relationship between speech and gestures. For example, Alibali and Heath (2009) investigated whether hand gestures are used for the facilitation of communication or for the speaker's needs only. They found that participants used gestures even when they could not see their listener, but the gestures that were used were less likely to depict semantic content related to speech than in case they could see their listener. Alibali and Heath concluded that speakers use gestures for both speaker-internal and communicative purposes. Mol et al. (2009) found that people have different gesture behavior depending on if their listener is human or artificial. When the participants believed they were talking to a computer, they produced fewer gestures and smaller movements than the participants who thought their listener is human. The authors mentioned two possible explanations for these findings. One explanation is that information in gestures is largely redundant with information in speech and people feel that the information given to a computer by speech is enough. Another explanation is that participants adapt less to an artificial listener that does not give feedback (like was found in Maes et al., 2007) and therefore were less informative in the gestural modality. A similar finding was shown in Beringer's (2001) research. He investigated the way users interact with an interface that can be controlled by speech and gestures. Participants were asked to plan a
trip using a system that could understand spoken language and gestures, without restricting
the gestures the subjects could use. Most of the subjects chose to interact with the system via
speech. Subjects found it unusual to gesture without a human dialog partner and not
necessary since the system understood them via voice input. Some of the subjects found it
difficult to use gestures without a clear demonstration of the recognizable ones. In the field
of HRI, Abich and Barber (2017) explored the effects of verbal and gesture methods of
communication on perceived workload, usability preferences, and expectations of robot
behavior while commanding a robot to perform a spatial-navigation task. The task consisted
of an autonomous robot navigating to specified locations and reporting reconnaissance and
surveillance information. The robot provided feedback via the visual display. The results
showed that speech elicited the lowest level of perceived workload but the usability
preference and expectations of robot behavior after interacting through each communication
condition was the same.

Several studies tested the effectiveness of vocal versus manual controls. In one experiment
by Baber (1992), soldiers were asked to enter reports into a battlefield database either using
pushbuttons or speech. There was no significant difference in the reporting time between
these two input modes. However, Steeneken (1996) showed that when operators have busy
eyes and hands, using speech control during tasks was effective. Redden et al. (2010)
compared voice control to manual control in the field of HRI. The task was driving a robot to
three different waypoints while writing sequential numbers, using speech control or manual
control. The results show that manual control was more effective than speech control for
continuous tasks such as turning, while in a discrete robotic task, voice commands reduced
operation time. Cassenti et al. (2009) also compared manual and vocal control in HRI, using
the “Wizard of Oz” paradigm. In their experiment, participants directed a robot to find a target
bomb in an indoor environment within a limited time, through manual control (mouse and
keyboard commands), restricted verbal control (verbal commands restricted to direction
information) or free verbal control (verbal commands with directions and object labels). The
manual condition was faster than the free verbal condition, while the manual and free verbal
conditions were not different from the restricted verbal condition. The authors proposed that
in free verbal condition there is a larger set of possible commands, so it is more difficult to
choose instructions and thus more time is required to formulate the instructions. Another
factor that possibly affected the results is the fact that participants didn’t believe they were
speaking to a robot.
Other studies compared between gestural controls and manual controls in HRI. Gesture interface is easy to use, can be used anywhere in the field of view of a camera, does not require special hardware, and allows a wide variety of commands (Hu et al., 2003). Xu et al. (2007) compared a gesture-based human-robot interface using a motion glove to a traditional joystick interface. They found that the gesture-based interface worked more efficiently and reduced the average task completion time compared to the joystick interface, although the subjective evaluation of its easiness of use was worse than in the joystick case. Radmard et al. (2015) compared the efficacy of a conventional keyboard and a gesture-based interface. In their experiment, the participants performed a collaborative task, using robotic telepresence system. The task required each collaborator to have continuous visual communication with each other, while one of them controlled the motions of the robot. One of the subjects was given a completed, labeled version of a 9-piece puzzle and instructed the other subject, faced with an empty puzzle board, to find and place each puzzle pieces on a board in a specific order. The findings indicated that the gesture-based interface provided smoother and more continuous control of the platform, but the keyboard provided better performance in terms of task completion time, ease of use, and workload.

2.3 The Interaction between Control and feedback modalities

According to Multiple Resource Theory (Wickens, 2002), sensory modalities play an important role in task performance. This theory suggests that there will be greater interference between two tasks to the extent that they share stages (perceptual/cognitive vs response), sensory modalities (auditory vs visual), codes (visual vs spatial), and channels of visual information (focal vs ambient). In Human-Robot communication, receiving feedback from the robot occurs at the perceptual stage, while commanding the robot belongs to the response stage. By applying Multiple Resource theory to human-robot communication, we can assume that the modalities used for both controlling and receiving feedback from the robot might affect the human performance and quality of the interaction.

Figure 1: the 4-d Multiple Resource Theory diagram (Wickens, 2008).
Greenwald (1970) suggested that there are stimuli modalities that most compatibly mapped to certain response modalities. This suggestion is based on the ideomotor theory and S-R compatibility effects. According to the ideomotor theory, actions are initiated by the anticipation of their sensory effects. For example, an image of a word's sound primes speech of the word and visual image of the written word primes writing. Greenwald applied this idea to S-R compatibility. According to Greenwald's theory, stimulus in a certain modality would activate the sensory images in this modality that in turn would initiate the response that produces this sensory effect. Under this hypothesis, there is ideomotor compatibility between stimulus and response that would produce a shorter RT. This hypothesis was tested in several experiments in which auditory and visual stimulus modalities were paired with spoken or manual responses. The results showed, as predicted, an interaction effect between stimulus and response modalities: compatible combinations (visual-manual, auditory-spoken) had shorter RT than incompatible combinations. Ideomotor interpretation of these results is that compatible combinations allow easier encoding because the proper response is selected without the need to translate the stimulus code into another modality.

Wickens et al. (1983) expanded the concept of stimulus-response compatibility to incorporate a mediating central processing (C) component (S-C-R): operators incorporate the stimulus information into a mental model of the system and then choose a response action. In their experiment, they compared task performance between combinations of auditory and visual stimulus modalities and speech and manual response modalities for spatial or verbal tasks. They showed that there is an association between a task's processing code (verbal or spatial) and stimulus-response modalities: for the verbal task, RT was shortest with the combination of auditory input and speech output. For the spatial task, RT was shortest with the combination of visual input and manual output.

Wang and Proctor (1996) also demonstrated these effects in an experiment. They used two types of visual stimuli: spatial-location stimuli (squares that appear either in a left or right position) and verbal stimuli (the words left or right), and two kinds of response modalities: manual responses (pressing either a left or right key) and vocal responses (speaking the word left or right into a microphone). In accordance with S-R compatibility effects, since the stimuli were visual, RTs were faster overall for manual responses than for vocal responses. In addition, supporting Wickens' findings, spatial stimuli with manual responses and verbal stimuli with vocal responses were more compatible than the two other combinations. In another experiment, two types of manual responses (keypresses and aimed movements) were examined in combination with the spatial-location and verbal stimuli. Aimed movements
refer to responses that were made by moving the index finger of the dominant hand on a touch screen, from a home box to one of two target boxes, located on the left or right side. For both the keypresses and aimed movements, compatibility was higher with spatial stimuli and lower with verbal stimuli than for vocal responses. When the aimed movements were compared with the keypresses, compatibility was higher for the verbal-movement and spatial-keypress conditions than for the spatial-movement and verbal-keypress conditions. These findings suggest that the size of the compatibility effect between manual response and spatial task depended on the type of manual response.

3 Goals and Hypotheses

In our current research, we aim to understand whether the mapping of human control modality to robot feedback modality during the interaction influences the quality of the interaction and if so, find the most compatible mappings. To accomplish this goal, participants were asked to perform a navigation task (helping a robot to get out of a maze by giving it navigation commands) while changing the modality of the commands and the robot’s feedback across the different conditions of the experiment. For controlling the robot, participants used hand gestures or vocal commands while the robot provided vocal or motion feedback (motion feedback means that the robot will be executing a predefined action in order to communicate its intention) or did not provide feedback at all.

Based on the literature we described in the previous section, we hypothesized the following hypotheses:

**H1: There are better subjective experience and task performance when the robot provides feedback.**

Previous literature emphasized the importance of providing feedback in human-robot interactions (Mirmig et al., 2011; Mohammad & Nishida, 2007). Based on that, we hypothesized that the efficiency in performing the task and the subjective experience would be better when the robot provides feedback than when the robot does not provide feedback.

**H2: There is an interaction effect between control modalities and feedback modalities.**

Greenwald (1970) showed that there are stimuli modalities that most compatibly mapped to certain response modalities. Our main hypothesis, based on Ideomotor theory and S-R compatibility effects, is that there is a compatibility effect between control modalities (response) and feedback modalities (stimulus). We suggested that different combinations of
control-feedback modalities would affect the quality of human-robot interaction in different ways, in terms of subjective experience and task performance.

**H3: Vocal control would be most compatibly mapped to vocal feedback and gestural control would be most compatibly mapped to motion feedback.**

Greenwald (1970) suggested that stimulus and response with the same sensory effects are compatible since it allows easier response selecting because there is no need to translate the stimulus code into another modality. Since the sensory effects of vocal response and vocal stimulus are the same, Greenwald found that they were compatible. Based on that, we expect to find that vocal control would be most compatibly mapped to feedback in the same modality. Similarly, since gestures and robot motions are both types of motions, we hypothesize that gestural control would be most compatibly mapped to motion feedback.

**H4: The efficacy of different combinations of control-feedback would be dependent on the task type.**

We mentioned experiments that demonstrate the efficacy of a combination of manual control and visual feedback for spatial tasks (Wickens et al., 1983; Wang & Proctor, 1996). We hypothesize that the efficacy of different combinations of control-feedback would be dependent on the type of task. Specifically, we expect that for a spatial task like maze navigation, the combination of gestural control (that can be considered as a type of manual response) and motion feedback (that can be considered as a type of visual feedback) would produce a more efficient interaction than the combination of vocal control and vocal feedback.

4 Preliminary experiment

In order to run the main experiment, we wanted first identify intuitive user commands and robot feedback methods that are worth considering. As Sarne-Fleischmann et al. (2017) noted, the success and acceptance of many robotic systems depends on the humans’ ability to utilize intuitive, easy-to-use and easy-to-remember commands. Similarly, the feedback provided by the robot also should be intuitive and easy-to-understand. However, there is very little research that determines which gestures or spoken commands people are likely to use or prefer to use when navigating a robot (Sarne-Fleischmann et al., 2017), or what feedback the robot should give in order to facilitate clear and intuitive communication (Ende et al., 2011, Kuno et al., 2007). Thus, we conducted a preliminary experiment to identify intuitive
control and robot feedback that would later be used to evaluate compatibility effects between control modalities and feedback modalities in human-robot interaction.

Nielsen et al. (2003) provided two approaches for obtaining intuitiveness measures for gestures:

- **Bottom-up**: presents commands and finds matching gestures.
- **Top-down**: presents gestures and finds which functions are matched.

We adopted these two approaches for intuitive gestures and vocal commands design and for intuitive motion and vocal feedback design.

### 4.1 Method

#### 4.1.1 Overview

The main goal of the preliminary experiment was to design intuitive commands and robot feedbacks for a robot navigation task. It was composed of two parts. The first part of the study was designed to create two vocabularies, one for gestural and vocal navigation commands and one for robot motion and vocal feedback. Fifty participants completed an online questionnaire where they were asked to propose:

1) Gestures and vocal commands that would make the robot perform a certain specified action
2) Ideas for how the robot can move and what it can say in order to communicate a certain specified message to the user

In the second part of the study, we validated the intuitiveness of the most common motion feedbacks and gestural commands that were suggested by participants in the first part, using a second online-questionnaire. In this part, we chose to focus on gestural commands and motion feedback design and not on vocal commands and feedback. Gestures can have a certain broadly understood interpretation (Gleeson et al., 2013). Similarly, a motion that is executed by a robot can be interpreted in various ways and not always be properly understood, especially if the human users do not have any previous knowledge of the movement patterns that will be used (Mohammad & Nishida, 2008). The same fifty participants that completed the first questionnaire participated also in the second one. In the second questionnaire, we presented participants with videos of the gestures and asked them to choose from a given list the command that they thought was most suitable for each gesture.
Then, we presented participants with videos of the motion feedbacks and asked them to choose from a given list the message that they thought the given motion feedback is conveying.

By analyzing the results of the two parts of the preliminary experiment, we have identified intuitive vocal and motion feedbacks, as well as gestures and vocal commands, that could be used to guide the motions of a mobile robot.

4.1.2 Participants

Fifty Industrial Engineering and Management students (32 female, 18 male), aged 20-28, from Ben-Gurion University of the Negev participated in this study. In compensation, they received extra course credit. Each subject participated in both parts of the experiment.

4.1.3 Procedure

Navigation task: In both parts of the experiment, we asked the participants to imagine that they are performing a navigation task with a TurtleBot3 burger robot (see Figure 2). They were given a general explanation about the robot’s motion abilities (moving forward, backward or rotating) and the modalities that they can communicate with the robot through (gestural or vocal control). In the navigation task we described, the user gives to robot four commands and then receives feedback from the robot. The robot executes these commands consecutively only after getting an order to start moving. When the robot finishes moving, the user gives it a new set of four commands and so on. The commands and feedback messages that were targeted using this task are displayed in Table 1.

Figure 2: Turtlebot3 burger
First part (the bottom-up approach): Using an online questionnaire that was administered via Google Form (see Appendix A), participants read the explanation about the robot and the navigation task. Then, they were presented, one-by-one, with a series of four actions that can be executed by the robot during the task (described as text and demonstrated using videos). For each action, participants were asked to propose a vocal command (in writing) and to record a video of themselves performing the gesture that they believe would communicate the robot to perform that action (control design). After uploading the video to the survey system, the participants were asked to explain how the gesture that they demonstrate in the video communicates the desired action. Then, the participants were presented with messages that the robot can communicate. They were asked to describe (in writing) vocal feedback and robot motions that could convey each message. For the motion feedback, they asked to explain why they chose this motion. One of the messages that the participants were presented with is a message about robot failure, but we didn’t use this type of message in our main experiment and therefore we won’t discuss it in this thesis.

Second part (the top-down approach): In the second online questionnaire (see Appendix B), we validated the intuitiveness of the most common gestural commands and motion feedbacks that were suggested by participants in the first questionnaire. After being given a reminder about the robot and the navigation task, participants were presented with a series of videos. In each video, they watched a person performs a certain gesture. Then, the participants were asked to choose the command that each gesture best represents from a closed list of commands and explain why this is the most appropriate command. Following that, the participants watched videos of the robot performing different motions, and for each video, they were asked to choose the message that they thought the robot’s motion communicated.

4.1.4 Analysis

First part: For each command, we calculated how many times each gesture was repeated by watching and analyzing the videos that were uploaded by the users. Based on that, we calculated the percent of subjects that proposed the same gesture as instructs the robot to
perform a certain action. We also calculated the percent of subjects that proposed the same vocal command. Similarly, we calculated the percent of subjects that proposed the same motion feedback and vocal feedback to communicate a certain desired message. Based on this data, we calculated an agreement score for each command and message in each modality, which represents the participants’ agreement among the gestures, vocal commands, motions and vocal feedbacks proposed by them (Wobbrock et al., 2005):

\[ A_r = \sum_i \left( \frac{P_i}{P_r} \right)^2 \]

\( A_r \) is the agreement score of command/message \( r \). \( P_i \) represents the number of the appearance of proposed gesture/motion \( i \) for command/message \( r \), and \( P_r \) is the number of total suggestions for command/message \( r \).

For example, for the command "turn right" there were 15 different suggestions for appropriate gestures. 5 participants suggested the same gesture \((P_1 = 5)\), 3 suggested another gesture \((P_2 = 3)\) and so on. \( P_r \) is the total number of suggestions that is the number of the participants \((P_r = 50)\). We calculated the agreement score with the following equation:

\[ A_r = \left( \frac{5}{50} \right)^2 + \left( \frac{3}{50} \right)^2 + \left( \frac{4}{50} \right)^2 + \left( \frac{14}{50} \right)^2 + \left( \frac{3}{50} \right)^2 + \left( \frac{9}{50} \right)^2 + \left( \frac{3}{50} \right)^2 + \left( \frac{2}{50} \right)^2 + \left( \frac{1}{50} \right)^2 + \left( \frac{1}{50} \right)^2 + \left( \frac{1}{50} \right)^2 + \left( \frac{1}{50} \right)^2 = 0.1428 \]

**Second part:** For each gesture, we calculated the percentage of participants that chose each command as represented by the displayed gesture. Since the most commonly selected command was the same for several gestures, we mapped to each command the gesture with the highest percentage. The same process was applied to participant selections of motion-feedback.

### 4.2 Results

#### 4.2.1 First part

The results of the first questionnaire for gestural control and motion feedback design are summarized in Table 2 and Table 3, respectively: for each command/message, the first two rows are the most common gestures/motions suggested by participants, while the third row is a gesture/motion that a-priori seemed to the experimenters as appropriate for expressing the desired command/message. The "suggestion percent" displays the percentage of participants that proposed the gesture/motion in each row. There were 12 different gestures but only 10 different motion feedbacks since there were some cases of overlap where the
same motion feedback was proposed for two different messages. The set of gestures/motions displayed in Table 2 and Table 3 was used as the base set for the second questionnaire.

**Gestural control design:** For the command "turn right", the most popular gesture was "an arm (from the elbow) descends from above to the right side of the body", proposed by 28% of the participants. For the command "turn left", the most proposed gesture was the same, but to the left side, proposed by 26% of the subjects. The agreement score for both commands is 0.14, which is quite low. The most common gesture for the "go forward" command was "an arm from the elbow is sent forward" with 60%. The agreement score is relatively high - 0.38.

For the "start moving" command the agreement score is 0.13 and the highest percentage of subjects (26%) suggested "hands clinging" as an appropriate gesture. The most common gestures for turning right, left and go forward were explained by the participants mainly by the fact that these are the gestures that they use when directing another person to certain destinations. The most common gesture for "start moving" was explained by being clear and simple, rather than by its connection to the command it represents. This is probably due to the fact that this type of command is less common than the other commands in human-human communication.

**Table 2: The set of gestures suggested by participants for each command (first online questionnaire) that we used for the second questionnaire**

<table>
<thead>
<tr>
<th>Command</th>
<th>Number</th>
<th>label</th>
<th>Gesture description</th>
<th>Start</th>
<th>End</th>
<th>Suggestion percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn right</td>
<td>1</td>
<td>Gesture 1</td>
<td>An arm (from the elbow) descends from above to the right side of the body</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Gesture 2</td>
<td>Pointing with the index finger to the right</td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Gesture 3</td>
<td>Pointing with the thumb to the right</td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td>10%</td>
</tr>
<tr>
<td>Turn left</td>
<td>4</td>
<td>Gesture 4</td>
<td>An arm (from the elbow) descends from above to the right side of the body</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td>26%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Gesture 5</td>
<td>Pointing with the index finger to the left</td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td>22%</td>
</tr>
</tbody>
</table>
Motion feedback design: For communicating understanding, the most commonly suggested motion feedback was "the robot moves one step forward and returns backward" (39%). The agreement score for this message is 0.2. According to the participants, this motion reminded them of a human nod that represents approval. In a similar way, for representing misunderstanding, most subjects (35%) suggested motions that reminded them of head shaking: "the robot turns to one side, then to the other side and returns". The agreement score for communicating misunderstanding is 0.18. For the messages "the command cannot be executed", the most suggested motions were the same: "the robot rotates once round", with 24%. The participants said that the rotation represents confusion which indicates that the robot cannot perform the requested action. Participants most suggested motion for "starting to move" was "the robot moves a little forward" (29%). Participants who chose this motion explained that it showed the robot's willingness to move. The agreement score for this message is 0.2.
Table 3: The set of motions suggested by participants for each command (first online questionnaire) that we used for the second questionnaire

<table>
<thead>
<tr>
<th>Message</th>
<th>Number</th>
<th>Label</th>
<th>Feedback description</th>
<th>Suggestion percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>The command was understood</td>
<td>1</td>
<td>Feedback 1</td>
<td>The robot moves one step forward and returns backward</td>
<td>39%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Feedback 2</td>
<td>The robot rotates one round</td>
<td>17%</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Feedback 9</td>
<td>The robot turns a little to right and returns</td>
<td>7%</td>
</tr>
<tr>
<td>The command was not</td>
<td>4</td>
<td>Feedback 4</td>
<td>The robot turns to one side, then to the other side and returns</td>
<td>35%</td>
</tr>
<tr>
<td>understood</td>
<td>5</td>
<td>Feedback 5</td>
<td>The robot moves one step backward and returns forward</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Feedback 10</td>
<td>The robot turns a little to left and returns</td>
<td>4%</td>
</tr>
<tr>
<td>The command cannot be</td>
<td>2</td>
<td>Feedback 2</td>
<td>The robot rotates once round</td>
<td>24%</td>
</tr>
<tr>
<td>executed</td>
<td>4</td>
<td>Feedback 4</td>
<td>The robot turns to one side, then to the other side and returns</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Feedback 7</td>
<td>The robot moves one step to the left and one step to the right</td>
<td>7%</td>
</tr>
<tr>
<td>Starting to move</td>
<td>3</td>
<td>Feedback 3</td>
<td>The robot moves a little forward</td>
<td>29%</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Feedback 2</td>
<td>The robot rotates one round</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Feedback 5</td>
<td>The robot moves one step backward and returns forward</td>
<td>12%</td>
</tr>
<tr>
<td>Robot failure</td>
<td>2</td>
<td>Feedback 2</td>
<td>The robot rotates one round</td>
<td>29%</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Feedback 6</td>
<td>The robot moves a little backward</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Feedback 8</td>
<td>The robot moves one step forward and returns backward twice</td>
<td>12%</td>
</tr>
</tbody>
</table>

The most common vocal commands and vocal feedbacks suggested by participants in the first questionnaire for each command/message are displayed in Table 4 and Table 5, respectively. The "suggestion percent" displays the percentage of participants that proposed the vocal commands/vocal feedbacks in each row.

Vocal control design: For the command "turn right", the most popular vocal command was "turn right", proposed by 64% of the participants. The agreement score for this command is 0.44. For the command "turn left", the most proposed vocal command was "turn left", proposed by 71% of the subjects. The agreement score for this command is 0.54. The most common vocal command for the "go forward" command was "go forward" with 38%. The agreement score is relatively low - 0.21. For the "start moving" command the agreement score is 0.17 and the highest percentage of subjects (40%) suggested "start" as an appropriate vocal command.

Vocal feedback design: For communicating understanding, the most commonly suggested vocal feedback was "understood" (37%). The agreement score for this message is 0.24. For
communicating misunderstanding, most subjects (40%) suggested vocal feedback: "did not understand". The agreement score for communicating misunderstanding is 0.2. For the messages "the command cannot be executed", the most suggested vocal feedback was "cannot execute", with 32% with agreement score of 0.25. Participants most suggested vocal feedback for "start moving" was "Starting to move " (29%). The agreement score for this message is 0.15.

Table 4: The set of most common vocal commands suggested by participants for each command (first online questionnaire)

<table>
<thead>
<tr>
<th>Command</th>
<th>Vocal commands</th>
<th>Suggestion percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn right</td>
<td>Turn right</td>
<td>64%</td>
</tr>
<tr>
<td>Turn left</td>
<td>Turn left</td>
<td>71%</td>
</tr>
<tr>
<td>Go forward</td>
<td>Go forward</td>
<td>38%</td>
</tr>
<tr>
<td>Start moving</td>
<td>Start</td>
<td>40%</td>
</tr>
</tbody>
</table>

Table 5: The set of most common vocal feedback suggested by participants for each message (first online questionnaire)

<table>
<thead>
<tr>
<th>Message</th>
<th>Vocal feedback</th>
<th>Suggestion percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>The command was understood</td>
<td>Understood</td>
<td>37%</td>
</tr>
<tr>
<td>The command was not understood</td>
<td>Did not understand</td>
<td>40%</td>
</tr>
<tr>
<td>The command cannot be executed</td>
<td>Cannot execute</td>
<td>32%</td>
</tr>
<tr>
<td>Start moving</td>
<td>Starting to move</td>
<td>29%</td>
</tr>
</tbody>
</table>

4.2.2 Second part

The results of the second questionnaire are displayed in Table 6 and
Table 7. The rows represent the gestures/motions that were presented in the second questionnaire. The gray cell in each row marks the most commonly selected command/message for each gesture/motion. Figure 3 and Figure 4 display the participants’ answers distribution for each gesture.

**Control design:** The gesture with the highest percent for "turn right" command was "pointing with the index finger to the right" (96%), while for "turn left" command it was "pointing with thumb to the left" (94%). Since we want the two gestures to be consistent in the final vocabulary, we decided to choose "pointing with thumb to the right" (94%) to represent the command "turn right" even though it did not have the highest percent. For the "go forward" command, the gesture with the highest percentage was "a tense arm rises from the bottom up" (96%). "hands clinging" was the gesture with highest percent for the "start moving" command with 92% agreement. The agreement scores among participants for all gestures were high, with an average agreement score of 0.79.
Table 6: Gesture-user command matching percent

<table>
<thead>
<tr>
<th>Gesture</th>
<th>turn right</th>
<th>turn left</th>
<th>go forward</th>
<th>start moving</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>62%</td>
<td>2%</td>
<td>2%</td>
<td>34%</td>
</tr>
<tr>
<td>2</td>
<td>96%</td>
<td>4%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>94%</td>
<td>4%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>4</td>
<td>2%</td>
<td>86%</td>
<td>2%</td>
<td>10%</td>
</tr>
<tr>
<td>5</td>
<td>6%</td>
<td>92%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>6</td>
<td>4%</td>
<td>94%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td>92%</td>
<td>8%</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td>68%</td>
<td>32%</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td>96%</td>
<td>4%</td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td>8%</td>
<td>92%</td>
</tr>
<tr>
<td>11</td>
<td>2%</td>
<td>0%</td>
<td>8%</td>
<td>90%</td>
</tr>
<tr>
<td>12</td>
<td>0%</td>
<td>0%</td>
<td>16%</td>
<td>84%</td>
</tr>
</tbody>
</table>

feedback design: the motion feedback whose meanings was most commonly agreed upon to convey understanding was "the robot moves one step forward and returns backward" (42%). "the robot turns to one side, then to the other side and returns" was chosen for misunderstanding (52%). For the message "the command cannot be executed", we saw something interesting: the motion feedback that had the highest percentage was "the robot turns a little to left and returns" (40%). This motion feedback was suggested by participants in the first questionnaire as feedback to notify about failure rather than the inability to perform. For "start moving", the matching feedback was "the robot moves a little forward" (36%). Participant agreement regarding the meaning of the presented motion feedbacks was much lower than for the gestures, with a 0.27 average agreement score. The difference in agreement scores between gestures and robot movements can attributed to the fact that we, as humans, use gestures to communicate with others in our daily lives, but motion feedback is not very common. Our lack of experience may have created greater variability in participants’ interpretation of motion feedback.
Table 7: Motion-robot message feedback matching percent

*We added in the questionnaire another option—“failure”, but we did not use this kind of feedback in the main experience.
4.3 Final feedback and control design

The control design and feedback design that were identified as most intuitive through this process are summarized in Table 8.

Table 8: The final commands and feedback design that were chosen to be used in the main experiment

<table>
<thead>
<tr>
<th>feedback</th>
<th>motion feedback</th>
<th>Vocal feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>The command was understood</td>
<td>The robot moves one step forward and returns backward</td>
<td>I understood</td>
</tr>
<tr>
<td>The command was not understood</td>
<td>The robot turns to one side, then to the other side and returns</td>
<td>I did not understand</td>
</tr>
<tr>
<td>The command cannot be executed</td>
<td>The robot turns a little to left and returns to the starting point</td>
<td>Cannot be executed</td>
</tr>
<tr>
<td>Start moving</td>
<td>The robot moves a little forward</td>
<td>Start moving</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>commands</th>
<th>Gesture</th>
<th>Vocal command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn right</td>
<td>Pointing with the thumb to the right</td>
<td>Turn right</td>
</tr>
<tr>
<td>Turn left</td>
<td>Pointing with the thumb to the left</td>
<td>Turn left</td>
</tr>
<tr>
<td>Go forward</td>
<td>A tense arm rises from the bottom up</td>
<td>Go forward</td>
</tr>
<tr>
<td>Start moving</td>
<td>Hands clinging</td>
<td>Start moving</td>
</tr>
</tbody>
</table>

4.4 Summary

The aim of this preliminary experiment was to design intuitive control and robot feedback for a ground mobile robot. These elements were intended to be used in our main experiment to evaluate compatibility effects between control modalities and feedback modalities in human-robot interaction. In the first part of the experiment, participants suggested gestures from their human-human communication experience. The robot motions that they proposed also shared resemblance to gestures used in human communication, like head nodding. This adaptation of gestures from human-human interaction to human-robot interaction suggests that participants may expect the robot to communicate to them like a social, human partner. These findings are consistent with previous research, suggested that an intuitive interaction with robots might be perceived similar to interacting with humans (Sarne-Fleischmann et al., 2017). In addition, Gleeson et al. (2013) found that observation of human-human interaction can be effective in determining what should be communicated in a given human-robot task and how communication gestures should be executed. In the second part of the experiment, when evaluating the intuitiveness of the common gestures/motions suggested in the first part, greater understanding of gestures was found compared to robot motion feedback understanding. A possible explanation for that difference can be the fact that humans use
gestures to communicate with others in their daily lives, whereas motion feedback is less common. However, previous research showed the effectiveness and naturalness of using motion cues as a feedback mechanism in collaborative robot navigation tasks (Mohammad & Nishida, 2008).

This preliminary experiment led to the creation of a user-driven communication vocabulary for guiding the motions of a mobile robot. Based on these control and feedback vocabularies, we ran our main experiment for evaluating how the relationship between control modalities and feedback modalities impacts the human-robot interaction.

5 Method

5.1 Overview
To evaluate how the relationship between control modalities and feedback modalities impacts the human-robot interaction, a WoZ experiment was conducted, using a within-subjects factorial design. The first factor was control modality, divided into two options: vocal control or gestural control. The second factor was feedback modality, divided into three alternatives: motion feedback, vocal feedback, and no feedback. Participants performed a navigation task, guiding a robot out of a maze by giving it navigation commands. The modality of the commands and the robot's feedback changed across the conditions of the experiment. Participants gave the robot four successive commands before they ordered it to start moving. The robot executed these commands consecutively once it received the order to start moving. When the robot had finished moving, the participant gave it a new set of four commands and so on. The maze had seven different entrance and exit points. Each participant performed the task seven times, each time using a different route and different combinations of control-feedback modalities.

5.2 Participants
Twenty-three students (12 female, 11 male), aged 21-27, from Ben-Gurion University of the Negev participated in this study. In compensation, they received extra course credit.

5.3 Robot
The experiment used a “Burger” variant of TurtleBot3, running ROS. The Burger variant has two wheels driven by servos and a ball bearing to keep its balance. Using the servos, the robot can go forward, backward, and turn around itself. The robot was teleoperated from a remote
A Bluetooth speaker was connected to the robot to enable it to provide vocal feedback (see Figure 5).

![Turtlebot3 burger after connecting the Bluetooth speaker to it](image)

**Figure 5: Turtlebot3 burger after connecting the Bluetooth speaker to it**

### 5.4 Experimental Design

*Maze Design:* The maze used in the experiment was marked on the floor using masking tape. There were seven different routes with different entrance and exit points (see Appendix F). We put corresponding numbers on each route’s entrance and exit point to clarify where each route begins and ends (see Figure 6). To construct the maze in such a way that all routes would not differ in their complexity, we used the choice-clue wayfinding model by Raubal and Egenhofer (1998). According to this model, the complexity of a wayfinding task is determined by counting the decision points where people have more than one option to continue the task. Thus, in the maze we built, all routes had the same number of decision points with more than one option (four decision points in each route). A run was defined as a navigation of the robot from a beginning point to an end point of a certain path (e.g., from 5 to 5).
Control and Feedback: The set of commands, described in Table 8, was used to guide the robot to turn right, turn left, go forward and start moving. The demonstration of possible gestural commands can be seen in Figure 7. An order to turn, made the robot turn ninety degrees to the right or left, respectively. An order to go forward advanced the robot for a constant distance of three tiles. The possible feedbacks options, also displayed in Table 8, were: providing the participant information about the robot’s understanding (i.e., if the command was understood or not), about its intention to start moving or about its ability to execute the command. The robot’s understanding or misunderstanding feedback appeared right after the participants gave the robot an order to turn right, turn left or go forward. The likelihood for command to be understood or not was set randomly (with 0.9 probability for the appearance of “understood” feedback and 0.1 for “not understood” feedback). If a certain command was not understood, the participant repeated it. In the no-feedback conditions, if a certain command was not understood, after receiving a command to start moving, the robot halted instead of executing the commands and the participant was asked to repeat the last sequence of four commands. The feedback about the robot’s intention to start moving appeared at the end of sequence of the four commands, after the participant ordered the robot to start moving. The feedback about the robot’s ability to execute the command appeared during the execution of the command and not immediately after the participant gave the problematic command. For example, if the participant asked the robot to perform an action that would cause the robot to clash into a maze wall, during the execution, the robot halted and informed the participant that the command could not be executed. In such a case, the participant gave the robot a new set of four commands.
Subjective measurement: for subjective experience measurement, at the end of each run (successful navigation of the robot to the end point of the current path) the participants filled out two online surveys that were administered via Google Form:

1. The System Usability Scale (SUS) (Brooke, 1996): A ten-item scale that gives a global view of subjective assessments of usability. SUS yields a single number (in range 0-100) representing a composite measure of the overall usability of the system. A system that receives a SUS score of 68 or above is considered usable (Bangor et al., 2009). We adjusted the scale to our goal in such a way that the questions referred to the interaction with the robot (see Appendix C). For example, instead of the item "I felt very confident using the system", we used "I felt very confident communicating with the robot". All items were measured on a five-point scale ranging from one (strongly disagree) to five (strongly agree). The SUS score was calculated for each run.

2. Stress State Questionnaire (DSSQ) (Matthews et al., 2002): Comprehensive assessment of subjective states in performance contexts, based on a factor model that differentiates 11 primary state factors, which cohere around three higher-order dimensions of task engagement, distress, and worry. We used twenty relevant items (see Appendix D). All items were measured on a five-point scale ranging from zero (strongly disagree) to four (strongly agree). We calculated the participant's engagement, distress and worry scores for each run. The possible worry score is in a range of 0-24, while distress and engagement scores are in a range of 0-28.

5.5 Procedure

The experiments were conducted in a Ben-Gurion University ABC lab. One experimenter was present during the study, responsible for instructing the participants on the maze task, the experimental procedure and running the study. First, participants were seated in front of a computer and read, via Google Form, about the maze navigation task. They were introduced
to the possible commands they could use and how to communicate them to the robot by gestures or voice commands. The experiment continued with a collection of participants' demographic data and consent forms. Before performing the task, the participants were trained on robot control. We asked them to navigate the robot from one point to another twice, one time using voice commands and the second time using gestures. During the training, the robot didn’t provide feedback. After training, participants performed the task for seven runs. In each run they used different route and different combination of control-feedback modalities, which were in a randomized order. Except for the last run, participants were instructed on which modality they should use and if they should expect the robot to provide feedback and if so the type of feedback to expect. In the last run, participants could freely choose the way they wanted to communicate with the robot as well as their preferred feedback modality (or without feedback).

Except for the no-feedback condition, the robot provided feedback in response to each command the participant gave. The feedback was provided under experimenter control. In the vocal feedback condition, the experimenter played pre-recorded vocal feedback recordings using R-studio. Since the motion feedback duration was longer than the recorded vocal feedback messages, we played three "beeps" using the "beepr" package before the vocal feedback. By this, the length in time of the two feedback modalities was similar and we could eliminate feedback duration as a possible confounding variable. In a probability of 0.9, understanding feedback was played and in a probability of 0.1 misunderstanding feedback was played. After playing four times of understanding feedback, "starting to move" feedback was played. In the motion feedback condition, the experimenter controlled the robot's movements using a keyboard via ROS platform and made the robot perform the appropriate motion feedback to the participants. To control the probability of understanding and misunderstanding feedback, the experimenter used pre-made sequences of "0" and "1" that contained four times "1" and a variable number of “0”. "1" represents understanding feedback and "0" represents misunderstand feedback. The sequences were constructed so that the probability of occurrence of "1" is 0.9 (and 0.1 for "0"). After the participants instructed the robot to start moving, the experimenter led the robot along the maze according to the participant's commands, using a keyboard and making it appear as the robot operates autonomously.

At the end of each run, the participants were asked to complete the SUS and DSSQ. By the end of the last run, they were given a post-experiment questionnaire that included questions about their preferences for communication modalities combinations (see Appendix E). All runs
were filmed by two video cameras for post-experiment analysis. Each experiment lasted between 50-60 minutes.

5.6 Measures

For each participant, we measured efficiency and subjective experience. To rate the efficiency, we measured the participant’s reaction time (RT): the average time from the moment the robot produced feedback till the following command the participant gave to the robot. We also measured accuracy by counting the number of errors participants made in performing the task (whether the subject led the robot on a longer path than the shortest possible path). We evaluated subjective experience using the custom SUS and DSSQ surveys that participants were asked to fill.

5.7 Analysis

The statistical analysis was done in SPSS. We used a generalized estimating equation model framework (GEE) including the fixed effects and one random effect which accounted for individual differences among participants.

5.7.1 Video analysis

During the experiment, each run was filmed by two video cameras- one was focused on the participant and the other one on the robot. Each participant performed the task seven times (23 participants*7 runs = 161 runs). Due to technical problems with the robot and the speaker, one participant didn't perform the last trial and another one performed the task only five times. In addition, the camera wasn’t recording for two runs. As a result, there were a total of 156 videos to analyze.

In order to create a video for each run in which the robot and the participant can be seen simultaneously, we used VideoPad-Video Editor software and combined the two recorded videos that we got from the cameras. To analyze the videos we made, we used Observer XT, professional software for the collection, analysis, and presentation of videos (Noldus et al., 2000). We watched each video and tagged the events that appeared during the observation, by marking the start time and end time and typing predefined key codes for subjects and behaviors. For example, if the participant gave the robot a command to turn right, we tagged this event’s start and end points. We also defined the subject as "Participant" and the behavior as "Turn right". If the event was motion failure, we defined the subject as "Failure" and the behavior as "Motion failure". The Possible subjects and behaviors that we defined can be seen
in Table 9. At each valid key press, the program calculated the elapsed time from the start of the observation and logged the time-stamped event record in the observational data file.

Table 9: the subjects and behaviors used for tagging events during video analysis

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant</td>
<td>Command to turn right</td>
</tr>
<tr>
<td></td>
<td>Command to turn left</td>
</tr>
<tr>
<td></td>
<td>Command to go forward</td>
</tr>
<tr>
<td></td>
<td>Command to start moving</td>
</tr>
<tr>
<td>Robot</td>
<td>Action: turn right</td>
</tr>
<tr>
<td></td>
<td>Action: turn left</td>
</tr>
<tr>
<td></td>
<td>Action: go forward</td>
</tr>
<tr>
<td></td>
<td>Action: start moving</td>
</tr>
<tr>
<td></td>
<td>Feedback: The command was understood</td>
</tr>
<tr>
<td></td>
<td>Feedback: The command was not understood</td>
</tr>
<tr>
<td></td>
<td>Feedback: The command cannot be executed</td>
</tr>
<tr>
<td></td>
<td>Feedback: Start moving</td>
</tr>
<tr>
<td>Reaction times (RT)</td>
<td>RT after understanding feedback</td>
</tr>
<tr>
<td>Failure</td>
<td>Vocal failure</td>
</tr>
<tr>
<td></td>
<td>Motion failure</td>
</tr>
</tbody>
</table>

For the analysis, we used two files that were provided by the system:

- The **time-event table** (Table10): this file includes a chronological listing of timestamped events.
- **Behavior Analysis table** (Table11): this file provides several calculations for each event in certain observation (run), including:
  - The duration of the observation
  - Mean duration of the event
  - The total number of event appearance
  - The total duration of all event appearance

For example, for the motion failure event in a certain observation, the file provides the duration of the observation, the mean duration of the motion failure in that run, the total number of motion failures and the total duration of all motion failures that appeared during the run.

Table 10: Example of the time-event table

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>00:00:00.684</td>
<td>00:00:00.000</td>
<td>842</td>
<td>0.58419</td>
<td>1.5783</td>
<td>Participant</td>
<td>GoForward start</td>
</tr>
<tr>
<td>284</td>
<td>00:00:03.500</td>
<td>00:00.000</td>
<td>590</td>
<td>3.4989</td>
<td>1.34207</td>
<td>Participant</td>
<td>GoForward start</td>
</tr>
<tr>
<td>627</td>
<td>00:00.04.042</td>
<td>00:00.00.04</td>
<td>842</td>
<td>4.84199</td>
<td>0.26</td>
<td>Participant</td>
<td>GoForward start</td>
</tr>
</tbody>
</table>
### Table 11: Example of the Behavior Analysis table

<table>
<thead>
<tr>
<th>Observations</th>
<th>Subjects</th>
<th>Behaviors</th>
<th>Mean duration</th>
<th>Total duration</th>
<th>Rate per minute (observation duration)</th>
<th>Total number</th>
<th>Duration</th>
<th>Start time</th>
<th>Stop time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Participant TurnRight</td>
<td>00:00:73</td>
<td>00:02:92</td>
<td>0.786687</td>
<td>4</td>
<td>05:06:08</td>
<td>14:32:31</td>
<td>14:37:36</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Participant TurnLeft</td>
<td>00:00:08</td>
<td>00:02:71</td>
<td>0.786687</td>
<td>4</td>
<td>05:06:08</td>
<td>14:32:31</td>
<td>14:37:36</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Participant GoForward</td>
<td>00:00:27</td>
<td>00:08:45</td>
<td>2.16329</td>
<td>11</td>
<td>05:06:08</td>
<td>14:32:31</td>
<td>14:37:36</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Participant Start</td>
<td>00:00:77</td>
<td>00:03:84</td>
<td>0.883308</td>
<td>5</td>
<td>05:06:08</td>
<td>14:32:31</td>
<td>14:37:36</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Robot MoveRight</td>
<td>00:03:22</td>
<td>00:09:66</td>
<td>0.580015</td>
<td>3</td>
<td>05:06:08</td>
<td>14:32:31</td>
<td>14:37:36</td>
<td></td>
</tr>
</tbody>
</table>

### 5.7.2 Subjective Experience

At the end of each run, the participants were asked to complete the SUS and DSSQ. One participant accidentally did not fill out the questionnaires.

**SUS score**

The SUS yields one factor- SUS score. The questionnaire includes 10 items, 5 are reverse-scored questions. The participant’s scores for each question ranged from 1 to 5. For calculating the SUS score for each run we converted the scores into new numbers: first we subtracted 1 from each score and then reversed the reverse-scored questions. The new scores added together and then multiplied by 2.5 to convert the original scores of 0-40 to 0-100.

**DSSQ factors**

The DSSQ consists of three factors: distress, worry, and task engagement. The questionnaire includes 20 items. We calculated for each participant the distress, engagement and worry scores by adding and subtracting the relevant item’s score:

- **Engagement calculation:** q3 + q7 + q9 + q17 - q1 - q12 - q20 + 12
- **Worry calculation:** q6 + q8 + q10 + q13 + q15 + q18
- **Distress calculation:** q4 + q11 + q19 - q2 - q5 - q14 - q16 + 16

We used GEE framework to calculate the effects of control modality, feedback modality and run duration (the time from the beginning of the navigating task until the robot reached the end exit of the maze) on distress, worry, engagement and SUS score. The second-order interaction control modality*feedback modality was included in the model. We chose to treat to run duration as independent variable and not as depended variable for several reasons. First, since there was different number of required commands in each of the seven possible paths, it affected the overall time that needed for robot navigation at each path. In addition, the duration of execution of different command and feedback modalities varies. Also, there were technical failures that affected the run duration and we could not control it. Therefore, we decided to measure the response time as a dependent variable and not the run duration.
Instead, since the run duration can affect the subjective experience, we treat this variable as independent variable.

5.7.3 Efficiency

RT

For RT analysis, we log-transformed RT, to lnRT, which has a normal distribution. We used GEE framework to calculate the effects of control modality and feedback modality (including the two-way interaction) on lnRT.

Accuracy

We used a binary variable, representing whether navigation errors were made during each run. When the participant navigated the robot in the shortest path, without deviations, the accuracy factor was marked as 1. For all other cases, i.e., when the participant did not navigate the robot in the shortest path it was marked as 0.

5.7.4 Variables summary

Using the data that obtained from the video analysis and the subjective questionnaires, we created two tables that we used for statistical and descriptive analysis:

- **Subjective preferences table**: each row represents a participant, including:
  - Last run communication modalities: The control and feedback modalities that the subject chose to use in the last run.
  - Preferred communication modalities: The control and feedback modalities that the subject chose as his favored via the post-experiment questionnaire.

- **Runs analysis table**: each row represents a run (7 runs for each subject), including:
  - Trial: trial number (1-7).
  - Path: path number (1-7).
  - Control modality: vocal or gestural control.
  - Feedback modality: vocal feedback, motion feedback or no feedback.
  - RT: the average time from the moment the robot produced feedback till the following command the participant gave to the robot.
  - Understanding RT: RT after understanding feedback.
  - Misunderstanding RT: RT after misunderstanding feedback.
  - Understanding number: total number of understanding feedbacks.
  - Misunderstanding number: total number of misunderstanding feedbacks.
Voice failures: total number of voice failures.
Motion failures: total number of motion failures.
Failures: total number of failures (voice and motion).
Min robot actions: the number of actions required on the shortest path.
Robot’s actions number: total actions number that the robot actually did according to subject orders.
Accuracy: Whether the participant navigated the robot in the shortest path.
Total Duration: run duration (from the first command till the robot reached the end point).
Engagement.
Distress.
Worry.
SUS score.

6 Results

6.1 Subjective Experience Analysis

6.1.1 Distress

Tests of model effects of distress are summarized in Table 12. Feedback modality had significantly contributed to distress (Chi^2(2) =12.393, p=0.002), as did the control modality and feedback modality interaction (Chi^2(2) =6.809, p=0.033). As can be seen in Figure 8, when there was no feedback, distress was higher (M=5.63, SD=0.645) than in the cases that included motion feedback (M=4.92, SD=0.636) or vocal feedback (M=3.80, SD=0.664). As can be seen in Figure 9, when the participants received motion feedback, they felt more distress if they communicated with the robot by gestures (M=5.56, SD=0.727) than by voice (M=4.28, SD=0.706). Run duration also significantly contributed to distress (Chi^2(1) =9.388, p=0.002). An increase in distress was correlated with longer runs (i.e., longer interaction duration).

Table 12: Tests of model effects for distress

<table>
<thead>
<tr>
<th>Source</th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>1.971</td>
<td>1</td>
<td>.160</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>12.393</td>
<td>2</td>
<td>.002</td>
</tr>
<tr>
<td>Control modality</td>
<td>.224</td>
<td>1</td>
<td>.636</td>
</tr>
<tr>
<td>Run Duration</td>
<td>9.388</td>
<td>1</td>
<td>.002</td>
</tr>
<tr>
<td>Control modality * Feedback modality</td>
<td>6.809</td>
<td>2</td>
<td>.033</td>
</tr>
</tbody>
</table>
6.1.2 Worry

Tests of model effects of worry are summarized in Table 13. The interaction between control modality and feedback modality had significantly contributed to worry ($\chi^2(2) = 10.866$, $p=0.004$). Figure 10 shows the interaction effect; when there was vocal feedback, there was
higher worry if the participants communicated with the robot by gestures (M=6.90, SD=0.877) rather than by voice (M=5.29, SD=0.887).

Table 13: Tests of model effects for worry

<table>
<thead>
<tr>
<th>Source</th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>15.899</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>1.222</td>
<td>2</td>
<td>.543</td>
</tr>
<tr>
<td>Control modality</td>
<td>.447</td>
<td>1</td>
<td>.504</td>
</tr>
<tr>
<td>Run Duration</td>
<td>.692</td>
<td>1</td>
<td>.405</td>
</tr>
<tr>
<td>Control modality * Feedback modality</td>
<td>10.866</td>
<td>2</td>
<td>.004</td>
</tr>
</tbody>
</table>

![Graph]

Figure 10: Worry by control and feedback modality

6.1.3 Engagement

Tests of model effects of engagement are summarized in Table 14. Feedback modality had significantly contributed to task engagement (Chi² (2) = 6.113, p=0.047). As can be seen in Figure 11, there was higher task engagement in motion feedback condition (M=20.68, SD=0.810) compared to no feedback (M=20.60, SD=0.823) or vocal feedback condition (M=19.74, SD=0.850).

Table 14: Tests of model effects for engagement

<table>
<thead>
<tr>
<th>Source</th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>304.779</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>6.113</td>
<td>2</td>
<td>.047</td>
</tr>
</tbody>
</table>
Control modality ... .313

6.1.4 SUS score

Tests of model effects of SUS score are summarized in Table 15.

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>393.751</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Control modality</td>
<td>7.821</td>
<td>1</td>
<td>.005</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>1.598</td>
<td>2</td>
<td>.450</td>
</tr>
<tr>
<td>Run Duration</td>
<td>3.225</td>
<td>1</td>
<td>.073</td>
</tr>
<tr>
<td>Control modality * Feedback modality</td>
<td>1.217</td>
<td>2</td>
<td>.544</td>
</tr>
</tbody>
</table>

Control modality had significantly contributed to SUS score ($\chi^2(1) = 7.821$, $p=0.005$). As Figure 12 shows, when the participants communicated with the robot by voice, SUS scores were higher ($M=78.50$, $SD=2.488$) compared to communication by gestures ($M=71.90$, $SD=2.441$).
6.2 Efficiency

6.2.1 RT

Tests of model effects for lnRT are summarized in Table 16.

Table 16: Tests of model effects for lnRT

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>36.728</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>2.146</td>
<td>1</td>
<td>.143</td>
</tr>
<tr>
<td>Control modality</td>
<td>9.952</td>
<td>1</td>
<td>.002</td>
</tr>
<tr>
<td>Control modality * Feedback modality</td>
<td>.402</td>
<td>1</td>
<td>.526</td>
</tr>
</tbody>
</table>

Control modality significantly contributed to lnRT (Chi² (1)=9.952, p=0.002). Participants were slower to respond in the vocal control condition, compared to the gestural control condition.
(Figure 13). The interaction between control modality and feedback modality had not significantly contributed to lnRT. However, when looking at pairwise comparisons of estimated marginal means of lnRT, we found some significant means differences. The means of the group that used vocal control and received motion feedback (VM) and the group that used gestural control and received motion feedback (GM) were significantly different (p=0.001). In addition, the mean of the group that used vocal control and received vocal feedback (VV) was significantly different (p=0.001) from the mean of the group that used gestural control and received vocal feedback (GV). The means of RT for these groups are demonstrated in Figure 14. The VM group were slower (M=0.82, SD=0.048) than GM group (M=0.65, SD=0.021) and VV group were slower (M=0.73, SD=0.025) than GV group (M=0.63, SD=0.043).

![Figure 14: RT by control and feedback modality](image)

6.2.2 Accuracy

Overall, accuracy rates were quite high (M=0.83, SD=0.374), meaning that participants erred only in about 17% of the runs. Tests of model effects for accuracy are summarized in Table 17. Feedback modality had significantly contributed to accuracy (Chi²(2) =6.428, p=0.04). As Figure 15 shows, participants erred more during the no feedback condition (M= 0.72, SD=0.061), compared to the vocal feedback condition (M=0.89, SD=0.04) and motion feedback condition (M=0.90, SD=0.039), that were not different from each other.
### Table 17: Tests of model effects for accuracy

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>979.287</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Control modality</td>
<td>.053</td>
<td>1</td>
<td>.818</td>
</tr>
<tr>
<td>Feedback modality</td>
<td>6.428</td>
<td>2</td>
<td>.040</td>
</tr>
<tr>
<td>Control modality * Feedback modality</td>
<td>2.371</td>
<td>2</td>
<td>.306</td>
</tr>
</tbody>
</table>

**Figure 15: Accuracy by control modality**

### 6.3 Preferences for communication modalities

In the last trial, participants could freely choose the way they wanted to communicate with the robot as well as the way they wanted the robot to communicate with them. The results are displayed in Figure 16. Six participants chose vocal control without feedback (VN), five participants chose vocal control and vocal feedback (VV), four chose gestural control and vocal feedback (GV), three chose gestural control without feedback (GN) and three chose vocal control with motion feedback (VM). No one of the participants chose gestural control with motion feedback (GM). Two participants didn't perform the last trial because of technical issues.
After completing the last trial, the participants filled out a post-experimental questionnaire, in which they asked to choose their preferred combination of control and feedback modalities. The results are displayed in Figure 17. Nine of the participants chose vocal control without feedback (9 out of 23), while the preferences of the rest of participants was divided relatively equal among the other possible control-modality combinations (four preferred vocal control and vocal feedback, three preferred vocal control and motion feedback, three preferred gestural control and vocal feedback, three preferred gestural control without feedback and one preferred gestural control with motion feedback).
6.4 Failures

There was no significant contribution of the number of failures to RT, accuracy or the subjective measures. However, we found that the number of misunderstandings had significantly contributed to distress (Chi² (5) =32.199, p=0.00), worry (Chi² (5) =11.119, p=0.49) and SUS score (Chi² (5) =43.329, p=0.00). Tests of model effects of distress, worry and SUS score are summarized in Table 18, Table 19 and Table 20, respectively. The subjective experiment was getting worse in accordance to the increase in the number of misunderstandings by the robot. We conducted the experiment in such way that the probability of appearance of misunderstanding feedback after each command that the participant gave to the robot was 0.1. In practice, the misunderstanding feedback appeared up to six times during one run. As demonstrated in Figure 18, looking at the runs that the robot provided feedback during them (134 runs), we see that in most runs (52) the robot didn’t understand the participant’s command only once. In 26 runs the robot didn’t provide misunderstanding feedback at all, in 36 runs the robot didn’t understand a command twice, in 15 runs the robot provided the misunderstanding feedback three times, in four runs the robot didn’t understand four commands and during one run the robot provided six misunderstanding feedbacks.
The participants felt more distress (Figure 19) and worry (Figure 20) as the robot provided more misunderstanding feedbacks during the run.

**Table 18: Tests of model effects for distress**

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>202.773</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Number of misunderstandings</td>
<td>32.199</td>
<td>5</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Figure 18: the number of runs distribution by number of misunderstandings**

**Figure 19: Distress by number of misunderstandings**

**Table 19: Tests of model effects for worry**

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>358.997</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Number of misunderstandings</td>
<td>11.119</td>
<td>5</td>
<td>.049</td>
</tr>
</tbody>
</table>
The SUS score was lower as the number of robot misunderstandings increased (Figure 21).

**Table 20: Tests of model effects for SUS score**

<table>
<thead>
<tr>
<th></th>
<th>Wald Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>24027.968</td>
<td>1</td>
<td>.000</td>
</tr>
<tr>
<td>Number of misunderstandings</td>
<td>43.329</td>
<td>5</td>
<td>.000</td>
</tr>
</tbody>
</table>

![Figure 20: Worry by number of misunderstandings](image)

![Figure 21: SUS score by number of misunderstandings](image)

**7 Discussion**

The aim of this experiment was to evaluate how the relationship between control modalities and feedback modalities impacts the human-robot interaction. Participants were asked to lead the robot to perform a maze navigation task while changing the modality of the commands and the robot's feedback across the different conditions of the experiment.
7.1 The Research Hypotheses

Our first hypothesis was that there are better subjective experience and task performance when the robot provides feedback (H1). The results support this hypothesis, showing that participants felt more distress and were less accurate when the robot didn’t provide feedback. These results are consistent with Mirnig et al. (2011) finding that showed a trend of higher success rate when feedback was provided, although this was not significant. At their experiment, participants that didn’t receive feedback during the interaction stated that they would want to receive feedback from the robot. However, in our experiment, when participants asked about their preference regarding control and feedback, most of the participants chose without feedback. A possible explanation for that result is that the feedback that participants received in our experiment did not include information that was required for task completion. Since the feedback did not contribute to completing the task, it wasn’t necessary for the participants that preferred to finish the task as quickly as possible.

Our second hypothesis was that there is an interaction effect between control modalities and feedback modalities (H2). Supporting this hypothesis, the results show that different combinations of control-feedback modalities affected the quality of human-robot interaction in different ways. We found a significant influence of control and feedback modalities interaction on the subjective experience (worry and distress). The interaction between control modality and feedback modality had not significantly contributed to InRT. However, when looking at pairwise comparisons of estimated marginal means of InRT, we found some significant means differences between different groups. That is, the control modalities and the feedback modalities affected how quickly participants responded to robot feedback.

Our third hypothesis was that vocal control would be most compatibly mapped to vocal feedback and gestural control will be most compatibly mapped to motion feedback (H3). Some findings support this hypothesis. The worry levels were low generally, but participants showed less worry when they communicated with the robot by voice and received vocal feedback, compared to gestural control and receiving vocal feedback. However, contrary to our expectations, participants felt more distress when they communicated with the robot by gestures and received motion feedback compared to vocal control and motion feedback. A possible explanation may be that the vocal commands were more intuitive and/or easier to remember. Memorizing the possible gestures may have been harder (Baun & Mackay, 2008), leading to higher cognitive load during this condition. Matthews et al. (2002, 2006) suggested that the distress response is linked to task workload and a sense of being overwhelmed by task demands is central to distress. If the task workload was greater in the gestural control
condition, it may explain the higher distress over the vocal control condition. However, as we expected, participants were faster when they received motion feedback and communicated with the robot by gestures compared to when they communicated by voice. When the participants communicated with the robot by voice and received vocal feedback, they were slower than in the case where they used gestures and received vocal feedback. This finding is also not consistent with our hypothesis since we expected to find that vocal control and vocal feedback would produce more efficient interaction. A possible explanation may lie in the type of task that we used in our experiment. Wickens et al. (1983) showed that there is an association between a task's type (verbal or spatial) and stimulus-response modalities: for the verbal task, RT was shortest with the combination of auditory input and speech output. For the spatial task, RT was shortest with the combination of visual input and manual output. Based on that, we hypothesized that gestural control (that can be considered a type of manual response) and motion feedback (that can be considered as a type of visual feedback) would produce a more efficient interaction (lower RT) while performing a spatial task like navigation. This can explain the results: when the participants communicated with the robot by voice and received vocal feedback, they were slower than in the case where they used gestures and received vocal feedback, since the use of gestures while performing a spatial task produced more efficient interaction, compared to vocal control.

Further to this, our fourth hypothesis was that the efficacy of different combinations of control-feedback would be dependent on the task type (H4). Partially confirming this hypothesis, we found that participants were faster when they controlled the robot by gestures rather than by voice and showed more task engagement when they received motion feedback in comparison to no feedback or vocal feedback. It seems that although the objective measurement (reaction time) supported the idea that gestural control is more effective when the task is spatial, the subjective experience shows the opposite. Even though both control modalities received high SUS scores (above 70) that indicate high usability (Bangor et al., 2009), the usability of the robotic system was perceived to be higher when the participants communicated with the robot by voice and not by gestures. Jakob (1993) described usability by five attributes: learnability, efficiency, memorability, low error rate or easy error recovery, and satisfaction. As mentioned by Bau and Mackay (2008), gesture-based interaction is ultimately more efficient for experts, while novices need extra support to learn. Since it possibly has been harder for our participants, that are novice in gesture-based interaction, to learn and remember the possible gestures compared to the vocal commands, this may be the factor that made participants rate gestural communication with lower usability than vocal
communication. In addition, perhaps they found it unusual to gesture without a human dialog partner and unnecessary since the system understood them via voice input, as was found in Beringer research (2001). Consistently, looking at post-experimental questionnaire results, we see that most of the participants chose vocal control (16 out of 23) with motion feedback (3), vocal feedback (4) or without feedback at all (9). It makes sense that participants preferred the control modality that they perceived as more usable.

7.2 Run duration
We found that the duration of the run (a run was defined as a navigation of the robot from a start point to an end point of a certain path) affected the subjective experience. An increase in distress was correlated with longer runs. The Distress was characterized by “an overload of processing capacity” (Matthews et al., 2002; Matthews et al., 1997) and tended to increase when participants experienced a loss of control over performance quality (Matthews et al., 1997). It is possible that as the duration of the run was longer, the participants experienced a loss of control that led to an increase in distress.

7.3 Failures
In our experiment there were two types of failures:

1. Technical failures:
   a. Motion failure: the robot didn’t move when it was asked to, stopped moving during the execution or didn’t stop moving after completion of the participant orders execution.
   b. Vocal failure: problems with providing vocal feedback.

2. Misunderstandings: the robot didn’t understand a command that the participant provided.

We did not find an influence of the number of technical failures on efficiency or on the subjective experience. However, we found that the number of misunderstandings affected the subjective experience, significantly contributed to distress, worry, and SUS score. The subjective experiment was getting worse in accordance to the increase in the number of misunderstandings by the robot. One of the attributes that describe usability by Jakob (1993) is low error rate. It makes sense that the more times the robot didn’t understand the user commands, the usability of the interaction with the robot was perceived to be lower. The distress increase can be explained by workload increase. When the robot misunderstood a certain command, the participants repeated the command and, at the same time, they had to
remember the commands they already gave. As we mentioned before, Matthews et al. (2002, 2006) suggested that the distress response is linked to task workload and a sense of being overwhelmed by task demands is central to distress. The workload increase that possibly caused by the robot misunderstandings and the need to memorize the given commands at the same time, may explain the distress increase. In addition, according to Matthews et al. (2002, 2006) worry correlates strongly with cognitive interference. It is possible that the need to repeat the last command when the robot didn’t understand caused cognitive interference and increased the participants' worry.

7.4 Limitations

Our experiment has some limitations that may have affected the results that were not conclusive regarding the question of how the mapping of human control modality to robot feedback modality influences the quality of the interaction. The most serious limitation of the present research may be the fact that the feedback that participants received did not include information that was required for task completion. On one hand, since the feedback was not necessary to task completion, it allowed us to examine a condition where there was no feedback at all. However, on the other hand, the fact that the feedback was unnecessary may affected the results. It makes sense that some of the participants, who perceived the feedback as unnecessary and preferred to complete the task as quickly as possible, preferred the condition where there was no feedback. The Lack of necessity of feedback may also have interfered with the influence of the different feedback modalities on the subjective experience and task performing. It is possible that using a task in which the feedback provided by the robot is critical to task success would produce more unambiguous results. However, it is important to note that although there were participants who preferred the conditions without feedback, most of the participants still preferred to have feedback. Also, we saw that participants felt more distress and were less accurate when the robot didn’t provide feedback.

Another limitation is related to navigation task design. The navigation task was conducted in such a way that the participant gave the robot four commands and then received feedback from the robot. The robot executed these commands consecutively only after getting a command to start moving. A more natural and simple way to conduct the task could be if the robot has executed each command immediately after receiving it. The problem with this case, which made us design the task as we had designed it in the first place, is that it makes the feedback unnecessary. It would be beneficial to think about a way to design more natural task, where the robot provides feedback that is necessary for task completion.
8 Conclusion

In conclusion, the aim of this experiment was to evaluate whether and how the relationship between control and feedback modalities impacts the human-robot interaction. We found that providing feedback improved the quality of the interaction. We also found that different combinations of control-feedback modalities affected the quality of the interaction in different ways. Participants showed less worry and were slower when they communicated with the robot by voice and received vocal feedback, compared to gestural control and receiving vocal feedback. In addition, participants felt more distress and were faster when they communicated with the robot by gestures and received motion feedback compared to vocal control and motion feedback. Some of our findings support the assumption that the type of task has influence when exploring the control and feedback modalities relationship impact. Participants were faster when they controlled the robot by gestures rather than by voice and showed more task engagement when they received motion feedback in comparison to no feedback or vocal feedback.

Our findings highlight the importance of investigating the relationship between control and feedback modalities for improving the quality of human-robot interaction, instead of focusing only on one part of the communication.

Future research should examine this question using a task in which the feedback provided by the robot is critical to task success. In addition, it would be beneficial to evaluate a verbal task to see if the type of task has influence when exploring the control and feedback modalities relationship impact on the human-robot interaction. Additionally, future research should investigate different communication modalities to generalize our findings to other modalities. Another interesting aspect to explore is the uncertainty of feedback modality. Previous research found that RT to a signal was longer if the modality of the signal was uncertain (Boulter, 1977). In our research, the participants were told before each run which type of feedback the robot would use so the modality of the feedback wasn’t uncertain. It would be interesting to see how the uncertainty of the feedback modality affects the quality of the interaction, depended on different combinations of control and feedback modalities. Also, it can be interesting to examine how different combinations of communication modalities affects the user's processing fluency, that is the ease with which information is processed (Jacoby & Brooks, 1984). In other words, it would be interesting to see whether the processing fluency is better in the compatible conditions.
9 References


Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3(2), 159-177.


10 Appendices

Appendix A: The online questionnaire used in the first part of the preliminary experiment ............... 50
Appendix B: The online questionnaire used in the second part of the preliminary experiment .......... 58
Appendix C: SUS questionnaire (original + adjusted) ........................................................................ 68
Appendix D: DSSQ questionnaire (original + translated) ................................................................. 69
Appendix E: Post-experiment questionnaire ....................................................................................... 70
Appendix A: The online questionnaire used in the first part of the preliminary experiment

Experiment short explanation:

Appendix A: The online questionnaire used in the first part of the preliminary experiment

Experiment short explanation:

The online questionnaire used in the first part of the preliminary experiment.

Consent form:

Consent form:

2. 

3. 

4. 

5. 

6. 

I confirm and agree to participate in the experiment.

I confirm and agree to participate in the experiment.

I confirm and agree to participate in the experiment.

I confirm and agree to participate in the experiment.

I confirm and agree to participate in the experiment.

I confirm and agree to participate in the experiment.
Demographics questionnaire:

9. שאלון רקע

10. מודל

Experiment detailed explanation:

مواد וSingleNode


10. מודל

Experiment detailed explanation:
מדגמה לאנפ במסד בימיה המוסל. המן הראות מבריע עלינו של הרחבון.
Control design questionnaire:

The four actions that can be executed by the robot during the task:

1. **Footstep:** The robot moves its body forward.

http://youtube.com/watch?v=siZkem2cjo
For each action, participants were asked the following questions:

27.艺术品的材质是什么？
28. 你对这个艺术品的看法如何？
29. 艺术品的形状和大小如何？

Scale:
1. Not at all small
2. Small
3. Medium
4. Large
5. Very large

5  4  3  2  1

Massa magna: $\Box$
Massa minima: $\Box$

For each action, participants were asked the following questions:

27.艺术品的材质是什么？
28. 你对这个艺术品的看法如何？
29. 艺术品的形状和大小如何？

Scale:
1. Not at all small
2. Small
3. Medium
4. Large
5. Very large

5  4  3  2  1

Massa magna: $\Box$
Massa minima: $\Box$
30. דרשה ממית מחשבהך לקרוא לוגריך ולהשאילו מהו מהירות הממשק.

5 4 3 2 1
מודר מוכים

31. המหมาย烟火 קרש ליבגון מבזגון פידע.

5 4 3 2 1
מודר מוכים

32. עליפס לגשם או הקודקוד הקולית המתחבש על התבקש כדר לזרוב לבועו והופעה.

33. שכרות:

1. הקודקוד הקולית שבחרת וראתי וראותשו תכונה.

5 4 3 2 1
מודר מוכים מודר

34. דרשנו ממית מחשבהך לקרוא לוגריך pomysו הקודקוד הקולית.

5 4 3 2 1
מודר מוכים
Feedback design questionnaire:

Explanation for feedback design questionnaire:

The messages that the robot communicate to a user:

1. The robot asks the user to do something.
2. The robot asks the user to confirm.
3. The robot confirms the user's action.
4. The robot modifies the user's action.

For each message, participants were asked the following questions:

For each message, participants were asked the following questions:
אוספי סיווגני ית מידה הסכנתה ע"ה האמירות הבאות להב''ג פירבך

הנאהוות: שבחרתם

1. הסכנתה חמורה מעביר התצרהован ההת possibilità הcelandית

2. rdr ראה המושבעה רכז לוחץ פירבך גהמהים

פיבך קולו

3. לעיל המושב הפגז התלהבות לפיכך לוחץ הפיבך הקולו

אוספי סיווגני ית מידה הסכנתה ע"ה האמירות הבאות להב''ג פירבך הקולו

שובחרתם: שבחרתם

1. הסכנתה הקולו שבחרתם נדיעה לא נוטשים

2. rdr ראה המושבעה רכז לוחץ ממאז הפיבך הקולו

58
Appendix B: The online questionnaire used in the second part of the preliminary experiment

Experiment short explanation:

Experiment detailed explanation:

Consent form:
כדי להימנע מהתקפה, דוֹרי מفحص בراتيج הלבני פלט שני יחידות ומרחアクセissant מתוכן עם הרובוטים ושימשו קצינים במחנה אימונים.

כדי להימנע מהתקפה, דוֹרי מفحص ב礓 הלוחрактиוות מתוכן עם הרובוטים ושימשו קצינים במחנה אימונים.

כדי להימנע מהתקפה, דוֹרי מفحص ברecha הלוחпрактиות מתוכן עם הרובוטים ושימשו קצינים במחנה אימונים.
Control design questionnaire (gestures):

Explanation for control design questionnaire:

The gestures that can be executed by the user during the task:

1.触摸

http://youtube.com/watch?v=9hitNNVx4cQ

2.触摸

http://youtube.com/watch?v=Z0cm1PjdlB
For each gesture, participants were asked the following questions:

65. Why is the coded gesture important for the task?
   [ ] Yes
   [ ] No
   [ ] Definitely yes
   [ ] Definitely no
   [ ] Somewhat yes
   [ ] Somewhat no

66. How does the coded gesture affect the task?
   [ ] Yes
   [ ] No
   [ ] Definitely yes
   [ ] Definitely no
   [ ] Somewhat yes
   [ ] Somewhat no

67. Does the coded gesture include information about the task?
   [ ] Yes
   [ ] No

68. How natural is the coded gesture?
   [ ] Very natural
   [ ] Somewhat natural
   [ ] Not very natural
   [ ] Completely unnatural

69. The coded gesture is:
   1. Establishing a connection with the experimenter.
   2. Establishing a connection with the computer.
   3. Establishing a connection with the experimenter.
   4. Establishing a connection with the computer.
   5. Establishing a connection with the computer.
   6. Establishing a connection with the computer.
   7. Establishing a connection with the computer.
   8. Establishing a connection with the computer.
   9. Establishing a connection with the computer.
   10. Establishing a connection with the computer.

70. The coded gesture:
   1. Establishing a connection with the experimenter.
   2. Establishing a connection with the computer.
   3. Establishing a connection with the experimenter.
   4. Establishing a connection with the computer.
   5. Establishing a connection with the computer.
   6. Establishing a connection with the computer.
   7. Establishing a connection with the computer.
   8. Establishing a connection with the computer.
   9. Establishing a connection with the computer.
   10. Establishing a connection with the computer.
Feedback design questionnaire:

Explanation for feedback design questionnaire:

The motion feedbacks that can be executed by the robot during the task:

2 Feedback

http://youtube.com/watch?v=Ho58um5Z0JQ
פידבק 1

http://youtube.com/watch?v=RYp3ieETZXw

פידבק 4

http://youtube.com/watch?v=Z00vMByq32s

פידבק 5

http://youtube.com/watch?v=NYOZ7zxf8S0

פידבק 3

http://youtube.com/watch?v=gc2nWY2MTVM

פידבק 10

http://youtube.com/watch?v=Wyw6FrR9SU
For each motion feedback, participants were asked the following questions:

131. How is the motion of the feedback message transmitted in the device? 
- Yes
- No

132. Is the feedback message clear?

133. The device transmits the feedback message in the form of a message or a sound?
- Yes
- No

134. Is there any confusion with the feedback?

An objective rating of the feedback message:
- No confusion - 1
- Minor confusion - 2
- Confusion - 3
- Significant confusion - 4
- Major confusion - 5

An subjective rating of the feedback message:
- No confusion - 1
- Minor confusion - 2
- Confusion - 3
- Significant confusion - 4
- Major confusion - 5

135. The feedback provided by the device is:
- Satisfactory - 1
- Partially satisfactory - 2
- Insufficient - 3
- Inadequate - 4
- Poor - 5

136. The device successfully prompts the user to perform the appropriate motion:
- Yes
- No

68
<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 I think that I would like to use this system frequently</td>
<td>=$^1$ אני רוצה להשתמש במערכת באופן распространенני</td>
</tr>
<tr>
<td>2 I found the system unnecessarily complex</td>
<td>=$^2$ אני מצאתי את מערכת זו שלא没必要ית高等</td>
</tr>
<tr>
<td>3 I thought the system was easy to use</td>
<td>=$^3$ אני הת htonsתי систем זו קל למשתמש</td>
</tr>
</tbody>
</table>
| 4 I think that I would need the support of a technical person to be able to use this system | =$^4$ אני חושב שאני צריךثبت תמיכה של מהנדס כדי להשתמש במערכת זו |}

### Hebrew

<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 I found the various functions in this system were well integrated</td>
<td>=$^5$ אני מצאתי שהункциוניות של מערכת זו יובאו בצורה יוצאת ל роли</td>
</tr>
<tr>
<td>6 I thought there was too much inconsistency in this system</td>
<td>=$^6$ אני חושב שהמערכת מכילה חוסר אחידות קיצוניים</td>
</tr>
<tr>
<td>7 I would imagine that most people would learn to use this system very quickly</td>
<td>=$^7$ אני מagina שהרוב人们 יוכלו ללמוד את אマル זה במהירות גבוהה</td>
</tr>
<tr>
<td>8 I found the system very cumbersome to use</td>
<td>=$^8$ אני מצאתי את מערכת זו מעכילה לעיך</td>
</tr>
<tr>
<td>9 I felt very confident using the system</td>
<td>=$^9$ אני_overlapבpatialy אמין במערכת זו</td>
</tr>
</tbody>
</table>
| 10 I needed to learn a lot of things before I could get going with this system  | =$^{10}$ אני Needed to learn a lot of things before I could get going with this system |}

### Hebrew

<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 I think that I would like to use this system frequently</td>
<td>=$^1$ אני רוצה להשתמש במערכת באופן распространенני</td>
</tr>
<tr>
<td>2 I found the system unnecessarily complex</td>
<td>=$^2$ אני מצאתי את מערכתי выс שוב הראשונה מה שנאיבית</td>
</tr>
<tr>
<td>3 I thought the system was easy to use</td>
<td>=$^3$ אני הת şekני מערכת זו קל למשתמש</td>
</tr>
</tbody>
</table>
| 4 I think that I would need the support of a technical person to be able to use this system | =$^4$ אני חושב שאני צריךثبت תמיכה של מהנדס כדי להשתמש במערכת זו |}

### Hebrew

<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 I found the various functions in this system were well integrated</td>
<td>=$^5$ אני מצאתי שהעשיוניות של מערכת זו יובאו בצורה יוצאת ל роли</td>
</tr>
<tr>
<td>6 I thought there was too much inconsistency in this system</td>
<td>=$^6$ אני חושב שהמערכת מכילה חוסר אחידות קיצוניים</td>
</tr>
<tr>
<td>7 I would imagine that most people would learn to use this system very quickly</td>
<td>=$^7$ אני מagina שהרוב人们 יוכלו ללמוד את אマル זה במהירות גבוהה</td>
</tr>
<tr>
<td>8 I found the system very cumbersome to use</td>
<td>=$^8$ אני מצאתי את מערכתי זה מעכילה לעיך</td>
</tr>
<tr>
<td>9 I felt very confident using the system</td>
<td>=$^9$ אניOverlapבpatialy אמין במערכת זו</td>
</tr>
</tbody>
</table>
| 10 I needed to learn a lot of things before I could get going with this system  | =$^{10}$ אני Needed to learn a lot of things before I could get going with this system |}

### Hebrew

<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 I think that I would like to use this system frequently</td>
<td>=$^1$ אני רוצה להשתמש במערכת באופן распространенני</td>
</tr>
<tr>
<td>2 I found the system unnecessarily complex</td>
<td>=$^2$ אני מצאתי את מערכתי זה复查高等高等</td>
</tr>
<tr>
<td>3 I thought the system was easy to use</td>
<td>=$^3$ אני הת şekני מערכתי זה קל למשתמש</td>
</tr>
</tbody>
</table>
| 4 I think that I would need the support of a technical person to be able to use this system | =$^4$ אני расположен ואת מערכתי זה מחמודד כדיstdexcept Чтобы            |}

### Hebrew

<table>
<thead>
<tr>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 I found the various functions in this system were well integrated</td>
<td>=$^5$ אני מצאתי שה עדיוניות של מערכתי זה יובאו بصورة יוצאת ל דליי</td>
</tr>
<tr>
<td>6 I thought there was too much inconsistency in this system</td>
<td>=$^6$ אני חושב שהמערכת מכילה חוסר אחידות קיצוניים</td>
</tr>
<tr>
<td>7 I would imagine that most people would learn to use this system very quickly</td>
<td>=$^7$ אני מagina שהרוב人们 יוכלו ללמוד את אマル זה במהירות גבוהה</td>
</tr>
<tr>
<td>8 I found the system very cumbersome to use</td>
<td>=$^8$ אני מצאתי את מערכתי זה מעכילה לעיך</td>
</tr>
<tr>
<td>9 I felt very confident using the system</td>
<td>=$^9$ אניOverlapבpatialy אמין במערכת זו</td>
</tr>
</tbody>
</table>
| 10 I needed to learn a lot of things before I could get going with this system  | =$^{10}$ אני Needed to learn a lot of things before I could get going with this system |}
<table>
<thead>
<tr>
<th>Q</th>
<th>Original (English)</th>
<th>adjusted (Hebrew)</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I found the task boring</td>
<td>נ�认ת המשימה היה מועלם</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>2</td>
<td>I felt relaxed</td>
<td>הרגשתי רגע</td>
<td>Distress</td>
</tr>
<tr>
<td>3</td>
<td>I was determined to succeed on the task</td>
<td>הייתנו نحو להצלחת המשימה</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>4</td>
<td>I felt tense</td>
<td>הרגשתי מتحمل</td>
<td>Distress</td>
</tr>
<tr>
<td>5</td>
<td>Generally, I felt in control of things</td>
<td>הביא כללו הרגישתי контрол</td>
<td>Distress</td>
</tr>
<tr>
<td>6</td>
<td>I reflected about myself</td>
<td>הרחורי לצעמי</td>
<td>Worry</td>
</tr>
<tr>
<td>7</td>
<td>My attention was directed towards the task</td>
<td>התמקדתי את תשומת לבו</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>8</td>
<td>I thought deeply about myself</td>
<td>התעמקתי בעצמי</td>
<td>Worry</td>
</tr>
<tr>
<td>9</td>
<td>I felt vigorous</td>
<td>הרגשתי אנרגטי</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>10</td>
<td>I thought about something that happened earlier today</td>
<td>עסקתי על מה שקרה מהплан</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>11</td>
<td>I found the task too difficult</td>
<td>מצאתי את המשימה קשה מדי</td>
<td>Distress</td>
</tr>
<tr>
<td>12</td>
<td>I found it hard to maintain my concentration for more than a short time</td>
<td>התמקתי לשומר על הרוכש במישנה</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>13</td>
<td>I thought about personal concerns and interests</td>
<td>עסקתי על חששות ואנייטרסיים</td>
<td>Worry</td>
</tr>
<tr>
<td>14</td>
<td>I felt confident about my performance</td>
<td>הרגשתי בטוח ביצואים של</td>
<td>Distress</td>
</tr>
<tr>
<td>15</td>
<td>I examined my motives</td>
<td>בדכתי את המטריך של</td>
<td>Worry</td>
</tr>
<tr>
<td>16</td>
<td>I felt that I could not deal with the situation effectively</td>
<td>הרגשתי שאיני יכול להנהדר עם כל חששכם בהמתקילים</td>
<td>Distress</td>
</tr>
<tr>
<td>17</td>
<td>I was motivated to do the task</td>
<td>הייתנו מלמוטציה לצלחת במשימה</td>
<td>Task Engagement</td>
</tr>
<tr>
<td>18</td>
<td>I had thoughts of Personal worries</td>
<td>עסקתי על דרישות חשושיבה ל.</td>
<td>Worry</td>
</tr>
<tr>
<td>19</td>
<td>I felt uncomfortable</td>
<td>הרגשתי שאלה בנה</td>
<td>Distress</td>
</tr>
<tr>
<td>20</td>
<td>I felt tired</td>
<td>הרגשתי עייף</td>
<td>Task Engagement</td>
</tr>
</tbody>
</table>
Appendix E: Post-experiment questionnaires

Can anyone think of any other questions or comments? (open-ended)

1. What do you think of the experiment?

2. What did you like/dislike about the experiment?

3. Would you be willing to participate in a similar experiment in the future?

4. Do you have any suggestions for future experiments?
Appendix F: the seven paths we used in the experiment
TABLE:

<table>
<thead>
<tr>
<th>מילה ממפה:</th>
<th>אינטראקציה רובוטים • לולאות משוב • מטלת יוזם • משוב יוזם • תחומים יידי-סביבתיים</th>
</tr>
</thead>
</table>

73
סultipartFile המסלול הפידבק: הקישר بين אופני תקשורת

המשויכ לביאטריקסית רובוט-אדם

הボード והкомית חלקי המвладשת קבלת תואר מוניסטר בנדסה

מאית: תמאה מרקוביץ'.

מנחת: פֶל טל אורון געלד

תאריך: 5/4/2020

חתימת המנהל: אי.יסר

הרשמה לתואר: 6/4/2020
ס出入无敌的ビジョン: との探し | 新の | 신의 | 신의

โดย | アド

หมายถึง | トマイ break

สวัสดี: พรอมสิ่ง