DEVELOPMENT OF AN INTELLIGENT VISION ENHANCED MULTIMODAL HUMAN-ROBOT INTERACTION FOR SERVICE ROBOTS IN DOMESTIC ENVIRONMENT

Pilippu Hewa Don Arjuna Shalitha Srimal

(168052E)

Degree of Master of Science

Department of Electrical Engineering

University of Moratuwa
Sri Lanka

August 2017

Development of an Intelligent Vision Enhanced Multimodal Human-Robot Interaction for Service Robots in Domestic Environment

Pilippu Hewa Don Arjuna Shalitha Srimal

(168052E)

Thesis submitted in partial fulfillment of the requirements for the degree Masters in Electrical Engineering

Department of Electrical Engineering

University of Moratuwa

Sri Lanka

August 2017

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgment any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The above candidate has carried out research for the M.Sc. thesis under my supervision.

Signature of the Supervisor: Date:
(Dr. A.G.B.P. Jayasekara)

Abstract – In the recent past, domestic service robots have come under close scrutiny among researchers. When collaborating with humans, robots should be able to clearly understand the instructions conveyed by the human users. Voice interfaces are frequently used as a mean of interaction interface between users and robots, as it requires minimum amount of work overhead from the users. However, the information conveyed through the voice instructions are often ambiguous and cumbersome due to the inclusion of imprecise information. The voice instructions are often accompanied with gestures especially when referring objects, locations, directions etc. in the environment. However, the information conveyed solely through gestures is also imprecise. Therefore, it is more effective to consider a multimodal interface rather than a unimodal interface in order to understand the user instructions. Moreover, the information conveyed through the gestures can be used to improve the understanding of the user instructions related to object placements.

This research proposes a method to enhance the interpretation of user instructions related to the object placements by interpreting the information conveyed through voice and gestures. The main objective of this system is to enhance the correlation between the user expectation and the placement of the object by interpreting uncertain information included in user commands. Furthermore, several human studies have been carried out in order to understand the factors that may influence and their level of influence on the object placement. The proposed system is capable of adapting and understanding, according to the spatial arrangement of the workspace of the robot as well as the position and the orientation of the human user. Fuzzy logic system is proposed in order to evaluate the information conveyed through these two modalities while considering the arrangement, size and shape of the workspace. Experiments have been carried out in order to evaluate the performance of the proposed system. The experimental results validate the performance gain of the proposed multimodal system over the unimodal systems.

***Keywords*- Human Robot Interaction, Natural Language Understanding, Deictic Gestures, Fuzzy Spatial Terms**

ACKNOWLEDGMENTS

It is with great pleasure that I acknowledge the assistance and contribution of all the people who helped me to successfully finish my Master's thesis.

First, I extend my sincere gratitude to my research supervisor Dr. Buddhika Jayasekara who provided me with his continuous support and assistance throughout the course of this thesis. Without his advices and encouragement, this thesis would never have been accomplished. I also would like to thank my progress review committee members, Dr. Chandima Pathirana and Dr. Ruwan Gopura for delivering their guidance and valuable comments for successfully continuing my thesis work throughout the past year.

I specially acknowledge the efforts put into reviewing this thesis by Prof. Nalin Wickramarachchi and Prof. Chandimal Jayawardena and thankful for the comments and suggestions.

I would not have been able to complete this thesis successfully, if it was not for the overwhelming assistance given by the individuals in my research group. I sincerely thank and appreciate Viraj Muthugala, Chapa Sirithunge and Sajila Wickramarathne for their support and suggestions in pursuing my Masters thesis.

Also, my appreciation goes to the staff and all of my friends in the Department of Electrical Engineering for their invaluable support, specially in taking part in the conducted experiments and surveys.

Finally, I would like to extend my deepest gratitude to my family. The blessings of my mother, father and sister undoubtedly helped me in making this endeavor a success.

This work was supported by University of Moratuwa Senate Research Grant Number SRC/CAP/16/03.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

INTRODUCTION

Usage of robots has had a overwhelming growth over the past few years [1]. The bloom in industrial usage as well as in the domestic service based tasks have exerted upon the requirement of robots that can be used with less hassle. The requirement for domestic service robots have increased specially due to the rapid growth in the number of elderly people in the community and decrease in the number of people who are eligible to take care of them [2]. Fig.1.1 shows the number of elderly population which is estimated till 2050 [3], while Fig. 1.2 shows the sales of service robots that is also estimated till 2050 [1]. The increase in both graphs provide evidence to the close correlation of the requirement of service robots versus the increase in the elderly population.

The area of domestic service robots has become a novel interest of the research community [4–9]. In this context, service robots designed specially for domestic usage can serve with a higher integrity. Pertaining to the fact that the domestic users are not overly cohabitated with technology, it is crucial that the interaction process between the user and the robot to be effortless. Most common method

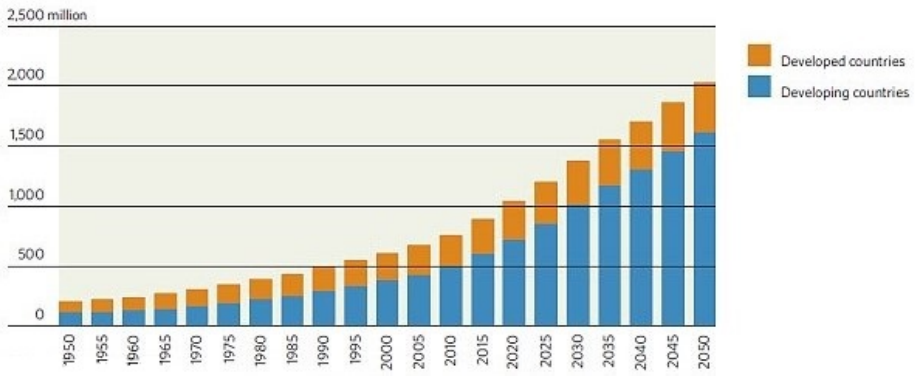


Figure 1.1: Number of elderly population in the world estimated till 2050.

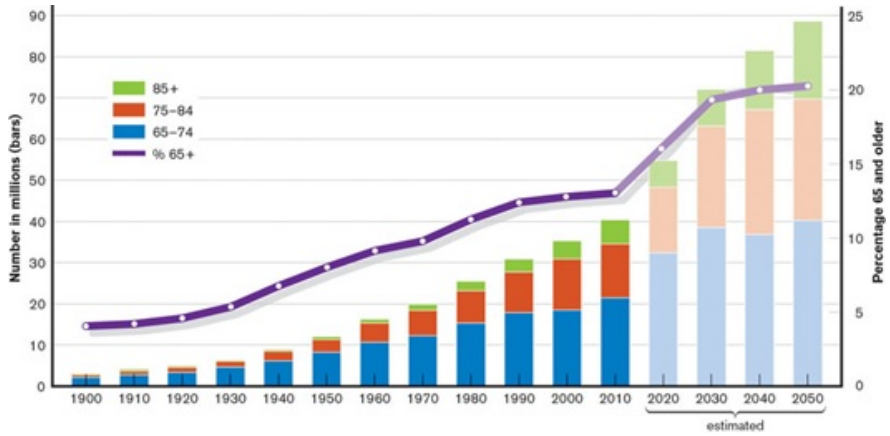


Figure 1.2: The sales of service robots up till 2010 and estimated values till 2050.

of interaction between humans is through voice commands. But when using such commands humans tend to use imprecise information which has vague meanings. Furthermore, when it comes to spatial information humans use more gesture based information. Even when using gesture based information there can be errors that can occur when extracting such information. Which leads to multimodal interaction methods containing both voice and hand gesture based commands which has engendered more human like interaction capabilities in service robots allowing them to clearly understand the conveyed user ideas.

This research proposes a method to interpret and understand the voice based user commands when placing objects on a surface using multimodal interaction. The system is improved to handle different shapes of surfaces and interpret spatial terminology accordingly. It takes robot's reachability to areas on the table as well as objects that are already on the table into consideration. Furthermore, the system has the capability to analyze the user's orientation and decide the reference frame for the object placement. Experiments have been conducted and the results have been analyzed to identify the behavioral capabilities and the performance of the system. The main objective of this system is to enhance the correlation between the user expectation and the placement of the object by interpreting uncertain information included in user commands. These concepts have further engendered the idea of enhancing the quality of the interaction between humans and the service robots.

1.1 Problem Statement

The service robotic systems that are deployed in the domestic environment needs to understand instructions that are conveyed by the human users clearly. Humans incorporate uncertain information in instructions which makes it essential for such systems to interpret these qualitative information. When explaining spatial information humans tend to use hand gestures more often. That helps to reduce the number of cumbersome voice commands used in order to explain a certain task. The robotic systems needs to understand these hand gestures which enhances the requirement of a multimodal system which can interpret the information conveyed through voice commands as well as hand gestures. Furthermore, when interpreting uncertain information in voice commands there are environmental and user aspects that has to be considered by such domestic robotic systems.

1.2 Thesis Overview

The overview of this thesis can be summerised as following,

- Chapter 2 - Literature review.
- Chapter 3 - Presents six user studies that has been performed in order to analyze how a human user would behave in a object manipulation scenario. Furthermore these studies helps to understand the factors that might influence the meaning of spatial terms in vocal user commands. Each study is presented in four sub sections 'Arrangement', 'Procedure and Matrix', 'Analyzing of Data' and 'Results and Discussion'.
- Chapter 4 - Presents the design of the system.
- Chapter 5 - Explain how the system understands the uncertain information in user commands.
- Chapter 6 - Introduces how the factors that were identified in the user studies can be applied in building a system that is more human-like in

behavioral aspects.

- Chapter 7 - Presents the experiments that has been carried out in order to test the performance of the system and the obtained results are discussed.
- Chapter 8 - Concludes the thesis.

Chapter 2

LITERATURE REVIEW

2.1 Human-like Robotic Assistants

Researchers have started brainstorming and scouting for robots with more human-like interaction abilities, which can assist the users in their day to day activities like cooking and other household aiding [4]. For example, In Fig. 2.1 shows Zora robot that helps as receptionist at hospitals [10]. Another example is the ASIMO (Shown in Fig 2.2) which is a humanoid robot created by Honda which can aid in performing several daily tasks [11]. Furthermore, while providing physical support, they must also be able to give cognitive assistance to the users [1, 5].

Even though the autonomous systems have advanced over the years they still lack the capability to carry out tasks entirely by themselves. In the car industry, robots have excelled to a level that majority of the human work force is replaced with autonomous robotic arms that are capable of working extended hours with higher integrity. In contrast to the industrial robots, the domestic service robots are par below the expected performance. Robots that are capable of cooking for example are not quite up to the task. Furthermore, robots who can perform tasks like surgery are a long way from reaching the accepted performance levels. Hence some collaboration with humans is required. These robots are more like assistants and rather need to follow the instructions given to them by the human master [12]. It is imperative that the robots should be designed in a way that the operation of a service robot in a domestic environment does not require the users to be overly cohabitate with technology [6].

Figure 2.1: Zora robots become receptionist and helps patients in Belgium hospitals

Figure 2.2: ASIMO by Honda

Figure 2.3: Motoman by Yaskawa Electric which has 'human-like' flexibility of movement demonstrates cooking [13].

2.2 Human Robot Interaction Methods

The most common method of interaction between humans and robots is through voice commands. Voice interaction mainly has two levels. They can either use short strict command set where the user has to memorise or else use natural language which is less cumbersome from the users' end [12,14]. Humans tend to use more qualitative information than the quantitative information when interacting with each others. These qualitative instructions mostly include uncertain terms like "near", "far", "close to" [15]. Therefore the robot should have the capability to respond after properly evaluating the uncertain information.

There has been several research work performed on the effect of using hand gestures and non verbal cues on human users [16–19]. As stated earlier the most common method of interaction between human and robot is through voice commands. But when explaining spatial information humans tend to include hand gestures along with voice commands more often [5]. So a considerable portion of natural human–human interaction happens through hand gestures. This will significantly reduce the number of words required to explain a task. For

an example when a robot is being asked to fetch an item from identical number of items in the vicinity, direct pointing is much easier rather than giving verbal instructions. This type of behavior should be embodied in the system so that it can be used with less hassle. If each and every command has to be given by voice it will become a tedious task. Furthermore, elderly people gradually get deprived of their ability to correlate among vocal instructions and spatial locations. Thus the deployment of verbal commands along with gesture based interaction could bolster the motivation of the elderly to easily get accustomed to these systems [5, 20, 21]. Despite the available resources much research has to be done in order to build more human like robots in domestic environment [4].

When using hand gestures, one common approach is to use fixed set of hand gestures which convey a predefined set of commands [22]. Which are used in place for the verbal commands. These are good where there are difficulty in giving commands using verbal commands like underwater or while driving [19, 23]. But in order to give a lot of commands to the system the user has to remember a lot of hand gestures which is far from the natural way of using hand gestures. Furthermore, using hand gestures alone will also not be ideal due to several other factors as well. One major concern is the extraction of the hand gesture itself. The pointing of hand gestures varies according to the person and the corresponding target. Creating a generic method will exert a lot of erroneous extractions. These may be either human or machine errors. But when hand gestures are used alongside voice commands, errors can be minimized by fusing the two inputs. So using a multimodal approach can provide better results than a unimodal approach [24, 25].

2.3 Understanding Uncertain Information

There have been many attempts to understand and interpret uncertain information based on spatial and environmental factors. Fuzzy logic based approach has been proven to give a successful solution to the complex task of understanding such information [15, 26–28]. But the main drawback in those attempts is that they only employ unimodal interaction capabilities. Furthermore, understanding of the spatial information alter depending on the external factors like

space arrangement. A successful attempt to understand the uncertain information was performed in [29]. In this system the robot uses the spatial data of the environment as well as the experience model to enhance and understand the user commands. But still the lack of gesture based interaction is visible when going towards a more human like approach.

Using gestures to interact has been incorporated by several researchers. These systems use various types of hand gesture extraction methods. Most common methods are to attach sensors to the hand or else to use external sensors to extract gestures. Even though the usage of body attached sensors are acceptable for field work, it is not an ideal method for day to day activities [12]. Pointed gesture identification has also been developed allowing the interaction between the human and robot more human like [5, 6, 30, 31]. These systems use RGB - D camera to extract the information of the hand pointing. However; in most cases the skeletal tracking system available in OPEN NI using depth sensors is used to obtain the information on the body joints with higher accuracy. In household environment, objects are often kept on planar surfaces like a table. The object manipulation on a table has been implemented on the scale of domestic service robots in [32]. Here the 3D depth data is used to identify the tabletop and segmentation is used to understand the objects placed on the table. This method has shown to be more effective. But this system lacks the ability to understand and respond to uncertain information which is highly critical when collaborating with humans.

2.3.1 Uncertainty in Spatial Terms

In domestic environments it is necessary for the robot to be able to understand the spatial terms which are associated with object manipulation. Most common requirement would be to pick or place items on a surface like a table [32, 33]. Specially when referring to an area on a table there are certain associated set of verbal terms. These terms usually doesn't have strict boundaries [34]. Terms like "Left", "Right" and "Center" can be pointed out as examples. In [34] work has been done in order to identify areas on a surface using spatial terms like "Middle", "Front" and "Left". The system uses strict boundaries to separate the spatial areas. Even though in [34] the system identifies the objects which are there in

a certain area on a surface, it lacks the capability to pin point a location when placing an item. In order to achieve higher accuracy each fixed area have to be separated in to smaller areas and so on. It will be tedious task if the user has to recursively segment the areas into smaller portions.

Understanding constraints in space is important when it comes to domestic service robots [35, 36]. Mainly because tasks that are associated with object placement on a surface is connected with understanding of space distribution. For an example, the location of the placement of an object will not be the same for an empty table as well as a table that has some objects already on it. When it comes to placement of the object on a table that has objects , the spatial terms that is used has to be altered accordingly. This type of operation is natural to a human being, but robots are still to achieve this level of human like behavior. Furthermore, when navigating around a table in order to place the object, the obstacles that are around the table has to be taken in to consideration. As humans will always move around a table avoiding these obstacle but will also change the placement of the objects depending on the reachability to certain areas on the table. As an example, if a certain area on the table is not reachable due to the obstacles around it, they may keep the object on a place that is closes to the requested location yet reachable. These are among the novel approaches to be considered in this research.

2.4 Summary of Literature Review

Fig. 2.4 shows the summary of the literature review. The current systems that are capable of understanding spatial terms can only select objects rather than place them on a surface. The scope of this research is to develop a system that can place objects after understanding the frequently used spatial terms in the natural language. In order for the system to be more human-like in interaction point of view they should have the ability to understand hand gestures as well as vocal commands. This research plan to incorporate multimodal interaction capabilities in th system. Rather using wearable technology this system will include external imaging sensors.

Previous systems lack the ability to adapt to various spatial attributes that may influence the understanding of the spatial terms. Here the system uses both spatial attributes and user aspects when determining the location of the object placement. These factors will be incorporated in to the system in a hierarchical manner.

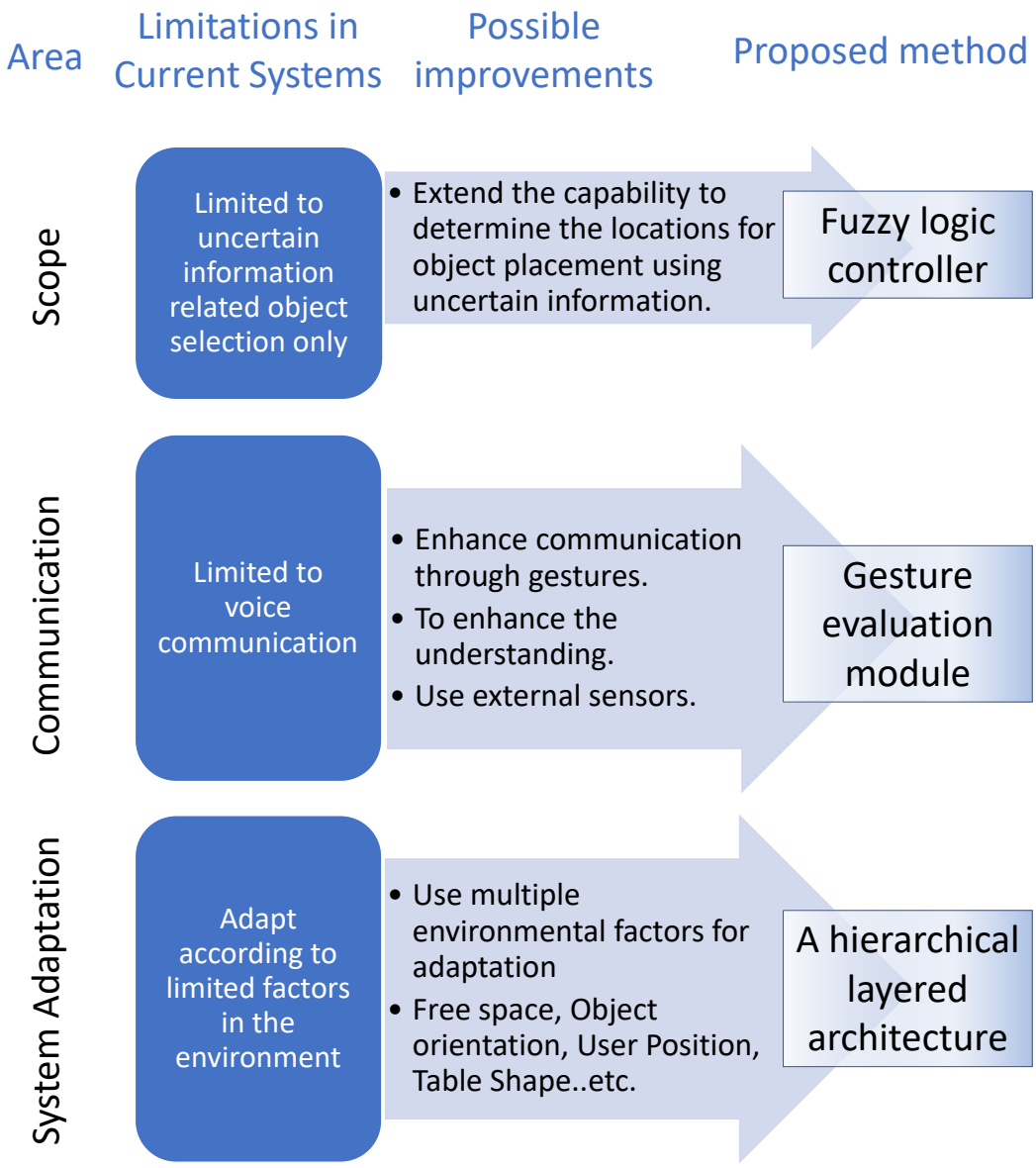


Figure 2.4: Summary of the literature review.

Chapter 3

HUMAN STUDIES ON UNDERSTANDING BEHAVIORAL CONCEPTS

When designing a system that has the same behavioral aspects of a human being, it is important to analyze the behavioral aspects of real human subjects. This research contains several human studies that help in understanding different key factors that might be important to be incorporated into the system to obtain a much more human like behavioral qualities.

Set of example scenarios are shown by Fig. 3.1. In figure (a), the user asks the participant to place the object on the right side of the table. So the participant needs to understand the area on the table that is referred by the term "right" and proceed with the placement. Human participant is able to understand these spatial terms as well as their boundaries. In figure (b) the location and the orientation of the user has been altered. So the participant has to understand how to alter the spatial areas and their boundaries accordingly. In figure (c) and (d) the table shape is different from two previous cases. So the participant has to redefine the spatial areas by considering the table shape. When figures (c) and (d) are compared the difference in the table size also has to be considered. In figure (e) there is an object that is on the table and in figure (f) the reachability of the participant is being limited by two obstacles that are near the table. The participant needs to consider these limitations and alter the meaning of the spatial terms accordingly. The aim of the user studies is to understand how a human user will react to above constraints and alterations when placing the object.

Figure 3.1: Example scenarios for human studies.

Figure 3.2: (a) shows a participant with the table and (b) shows the distribution of objects on the table from the point of view of the participant.

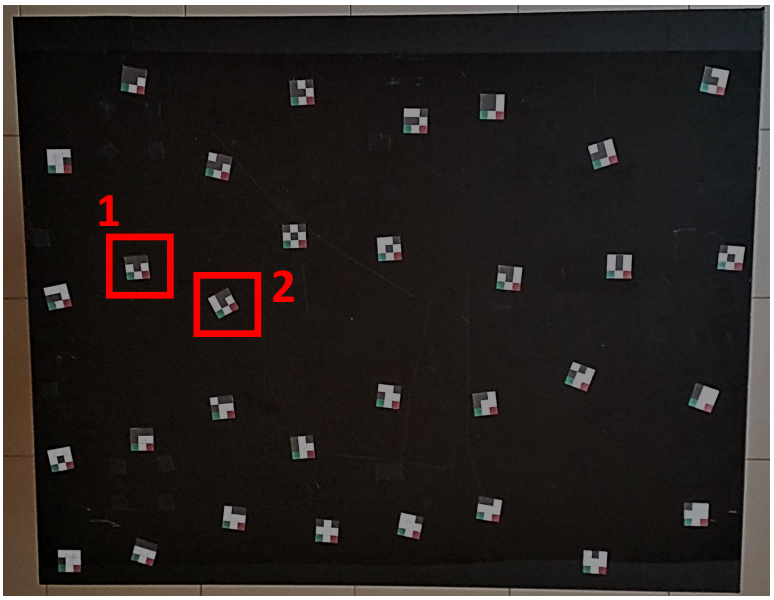


Figure 3.3: Effect of the user rating.

3.1 Understanding of Spatial Terminology

Humans use a certain set of words when it comes to explaining spatial locations. Specially when it comes to object manipulation on a surface. The main idea of this study is to understand the different set of words that are used in object manipulation scenarios in a domestic environment. Further to understand the boundaries of the identified set of words.

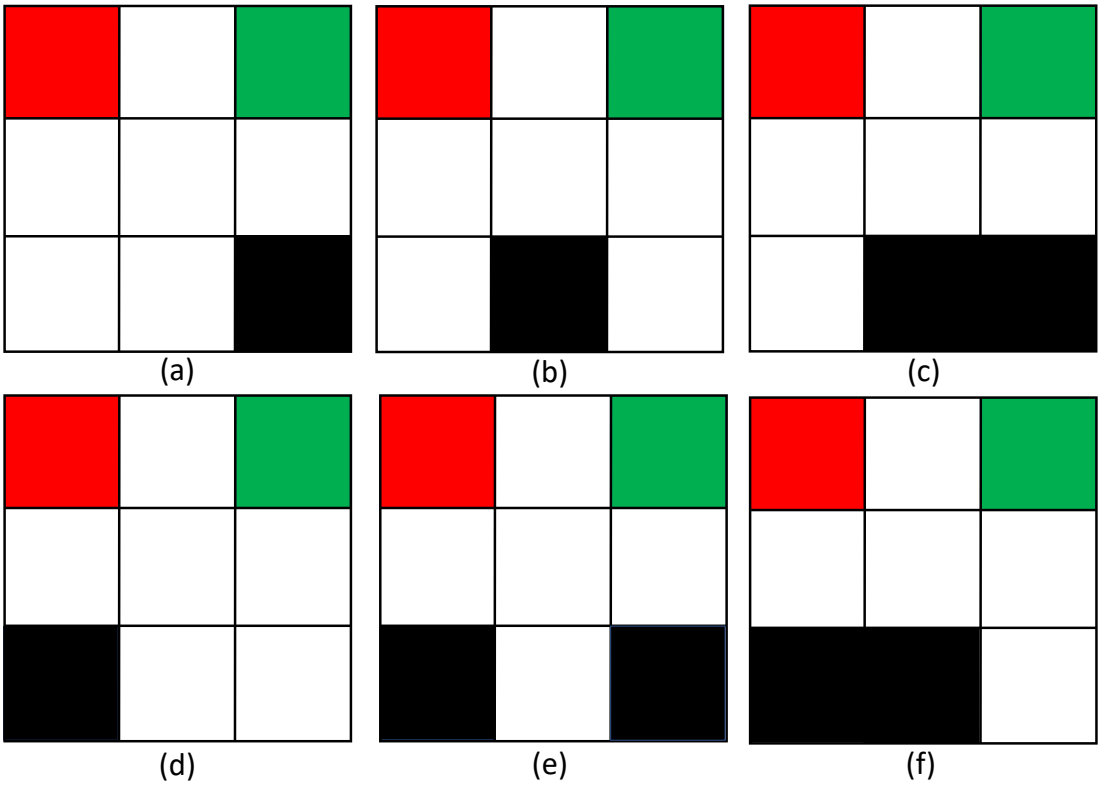


Figure 3.4: The color coded cards that were used for tracking the placement of locations.

3.1.1 Arrangement

The set up for this study is shown by Fig.3.2. A rectangular table of 910mm in length and 710mm in width was used. In order to mark the locations on the table set of coded cards were used as shown by Fig.3.2 (b). An overhead camera was used to identify the locations of the 32 coded cards. These cards were binary coded so that each card can be identified individually. Example set of cards are shown in Fig. 3.4. Here (a) to (f) represents the binary coded cards 1 to 6. Two red and green color squares were used to track the card's orientation and location. The black squares represents binary value 1 while white square represents binary value 0. Using this color coded method the location of each card can be tracked with respect to the table.

Here the table was kept at the center of an empty room in order to eliminate other effects that could have altered the meaning of the spatial terms. These effects are considered in the coming section of the user studies. Twenty six human subjects participated in this study with a mean age of 24.55 years and a standard deviation of 10.07 years.

3.1.2 Procedure and Metrics

The users were asked to explain the locations of each object as if they would explain it to another human user. In order to make sure that the participants get a clear view of the table, they were asked to stand while participating in the study as shown by Fig.3.2 (a). The positions of the cards were randomly distributed over the surface of the table and they were different from participant to participant as shown by Fig.3.3. The positions were recorded manually rather than using a speech to text software. Because using a software can give errors when converting to text. In order to achieve a higher accuracy, it is important to record the instructions clearly. Furthermore, the users were asked to give a rating to each explanation depending on the confident that they have regarding the categorization of the objects. This rating varied from 1 to 5, where 1 is the minimum rating and 5 denoted the highest rating. An example for this rating is shown by Fig.3.3. Here both objects are categorized as "Left" but the object that is denoted by 1 has a rating of 5 while the object that is denoted by 2 has a

Table 3.1: Usage of Area Terms

Area Term	Frequency
Middle	88
Left	102
Right	106
Front	87
Back	85
Corner	152
Edge	168

lower rating of 2. This is due to the fact that participant is more confident that the object denoted by 1 falls in to the spatial category "Left", while the object denoted by 2 is in between the key areas "Left" and "Middle".

3.1.3 Analyzing of Data

The location of each object was analyzed along with the spatial term that was used to explain it. The identified locations are summarized as in Table 3.1. There were 44 location that the users failed to classify clearly. Furthermore, there were several synonyms that were used by the users. They were also classified with the relevant spatial terms in the 3.1. For an example the spatial term "Center" is a synonym to the spatial term "Middle" and is categorized under the same term "Middle". Locations of objects of each classified area term is shown by Fig.3.5. The X and Y axis are shown in Fig.3.5(a). Furthermore, the diameter of the marker in the figures, depend on the confident factor. If the confident factor is high the marker diameter is also large. This particular method helps in understanding the distribution of the spatial terms better.

3.1.4 Results and Discussion

The obtained results from the study shows that there are certain areas that were classified using spatial terms. These areas do not have strict boundaries and their concentration diminishes when going from one area to another. The averages results are shown in Fig. 3.6. Fig. 3.6 (a) shows the summary of key area terms. Here 'C' denotes the area term "Corner" while highlighted in green is the area term "Edge". Fig. 3.6 (b) shows the 3D distribution of area term "Right".

Table 3.2: Area Terms Used in Study 2

Area Term	Subsection
Middle	-
Left	-
Right	-
Front	-
Back	-
Corners	Front Left
	Front Right
	Back Left
	Back Right
Edge	Left
	Right
	Front
	Back

As mentioned earlier there are no strict boundaries for these area terms and the results obtained from the study has been averaged in order to obtain a smooth curve. Fig. 3.6 (c) and (d) shows the 3D distribution of area terms "Middle" and "Corner" respectively.

3.2 Effect of the User Location and Orientation

3.2.1 Arrangement

This study was carried out by changing the location and the orientation of the user. The arrangement used in the previous study was used here with a single table. For this study 29 personnel participated with a mean age of 29.95 and standard deviation of 13.22 years.

3.2.2 Procedure and Metrics

Participants usually use two reference frames when deciding the location to place the object . These two frames can either be the participant's own reference frame or it can be the reference frame with respect to the person who is issuing the command(user). The main idea of this section of the study was to identify

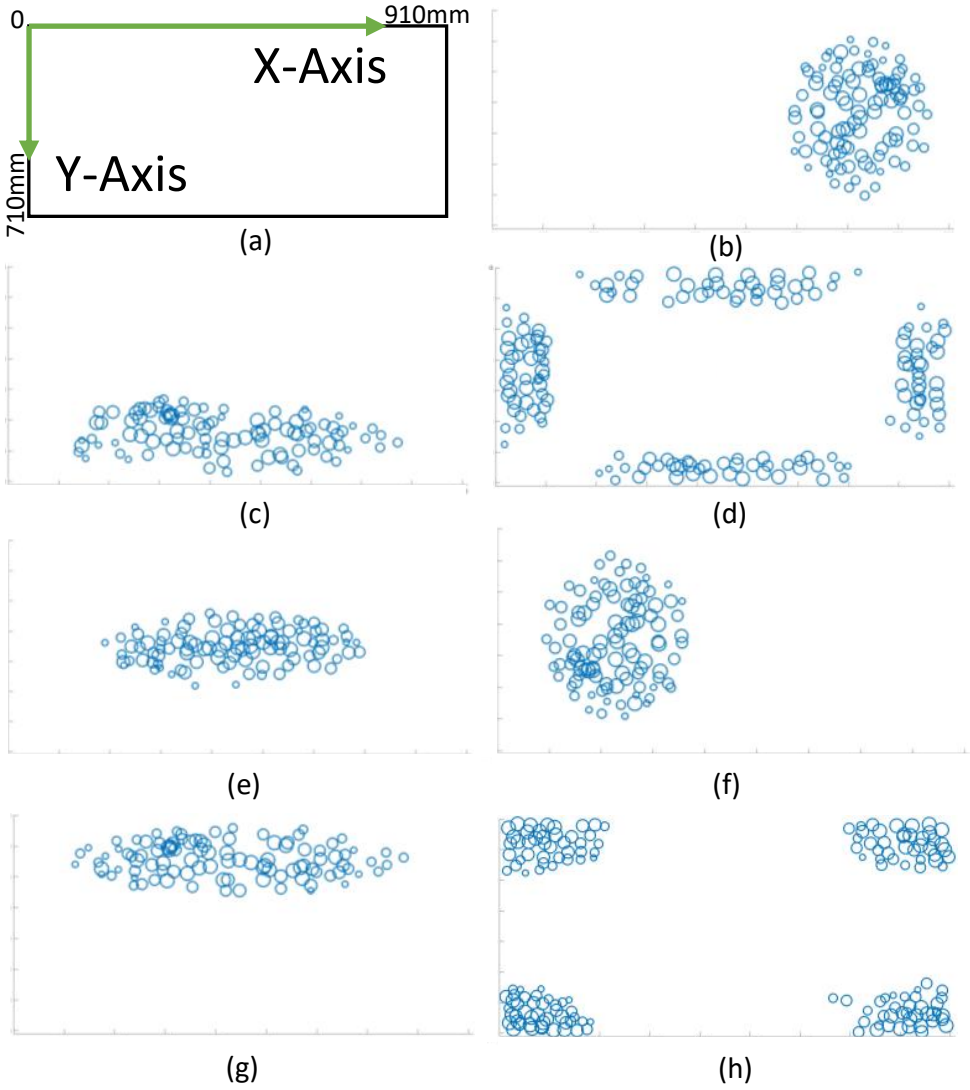


Figure 3.5: Shows the obtained results of the study 1. Here (a) represents the X and Y axes of the table. Distribution of the spatial terms "Right", "Front", "Edge", "Middle", "Left", "Back" and "Corner" are represented by (b), (c), (d), (e), (f), (g) and (h).

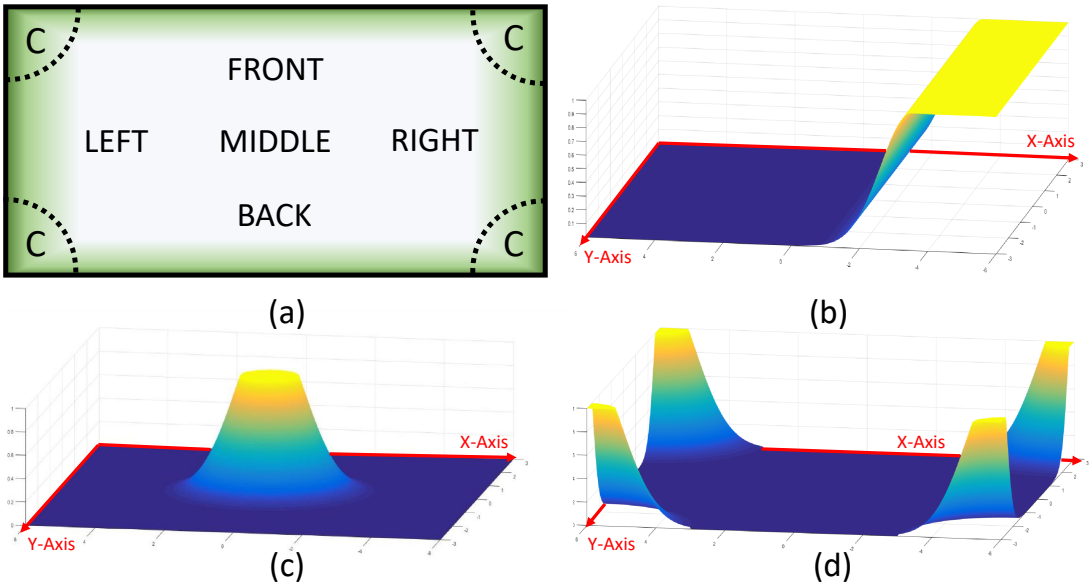
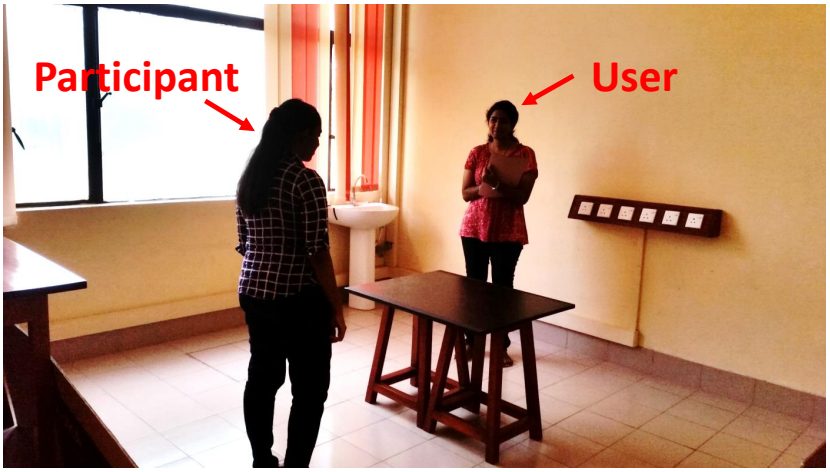


Figure 3.6: (a) shows the key area classifications of a table. (b),(c) and (d) shows "right", "middle" and "corner" area density distributions respectively.

the reference frame used by the participants for the table depending on whether the user is looking at the table or not(referred to as the attention of the user). This particular scenario can be further explained using the sub figures of Fig.3.7. Figure (a) and figure (b) depict cases where the participant has categorized the user's orientation as paying attention to the table. Figure (c) depicted a scenario where the participant found the user's orientation as not paying attention to the table. Here the user gave commands to the participants to place the object while looking at the table and looking away from the table. Each participant received 10 commands where the user was in 10 different locations and orientations with respect to the table. The participant was not asked to categorize whether the user was paying attention or not while conducting the study. After completing the 10 placements for one participant, the user repeated the locations and orientations that was performed earlier and the participant's perception regarding the user attention was recorded. This particular step was imperative in order to ensure that the perception of the participant is not affected by the true intention of the study. The given commands to place the items on table locations included spatial key terms from Table 3.2. The location of the object placement was tracked using the overhead camera while the given commands were manually recorded. The orientation and the location of the user was recorded with respect to the table. The participant's response about the attention of the user was also recorded for



(a)

(b)

(c)

Figure 3.7: The effect of the user's orientation.

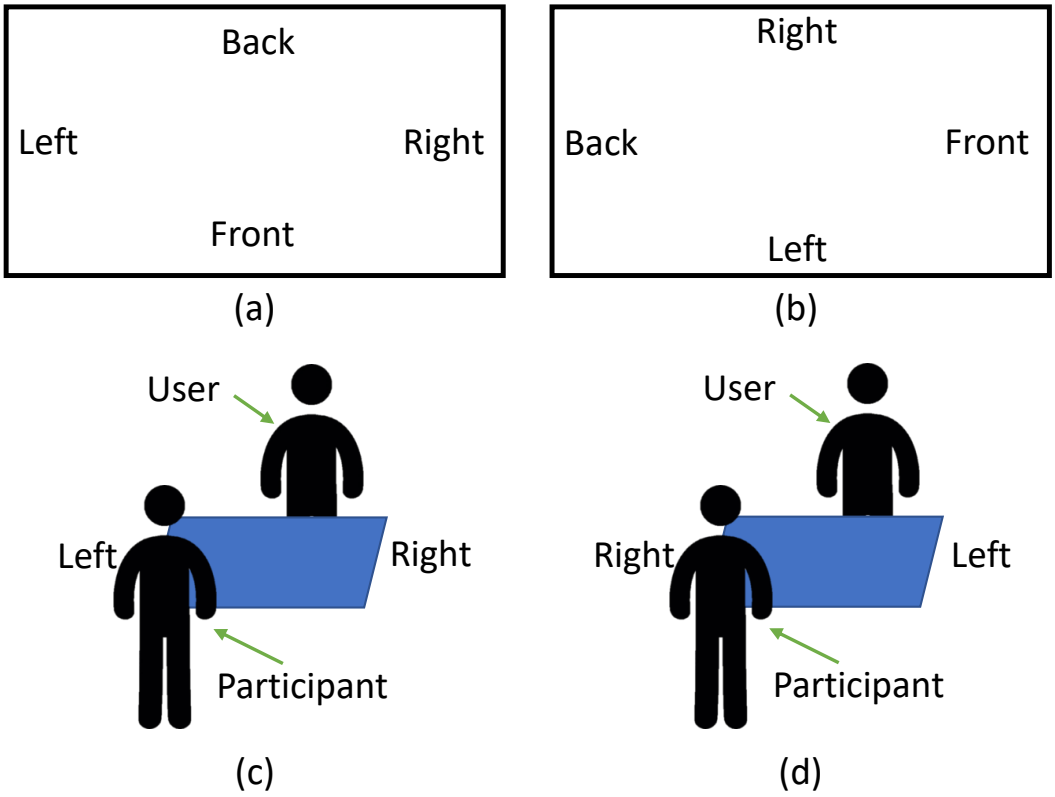


Figure 3.8: Effect of the reference frame.

each entry.

3.2.3 Analyzing of Data

This study produced 290 individual placements. The results from this study were analysed by categorising them with respect to the orientation of the user. Out of the 290 results, in 141 results the participants thought that the user was paying attention while 143 were categorised as the user was not paying attention. There were 6 occasions where the participants could not specify whether the users' were paying attention or not. The Fig. 3.9 shows the summary of the results obtained from this study.

3.2.4 Results and Discussion

Fig. 3.8 (c) and (d) shows the two reference frames that the participants used to refer when deciding the spatial areas on the table. In figure (c) the "Left" and

"Right" of the table is marked with respect to the participant's reference frame. While, in figure (d) it is marked with respect to the user. The participants decided whether the user is paying attention to the table or not, by observing the head orientation of the user. If the user was looking at the table, they marked the case as user was paying attention. Even if the user was not looking directly at the table, but looking at the direction of the table, majority of the participants marked the user as paying attention to the table. Furthermore, in most of the cases the participants judged the reference frame of the user by considering the body orientation of the user but not necessarily depending on the direction where the head was facing.

Although there was a clear effect of the user's attention on the participants' placement of objects, using the results obtained through study 3, a decision cannot be made as to when the participant uses the reference frame of the user. There were 141 occasions where the participants thought that the user was paying attention to the table. Out of them, only 39% of the times they used the user's reference frame. This result was influenced majorly by the fact that majority of the users did not categorize the table areas like shown in Fig. 3.8 (b). Only two users on two occasions decided that the table area classification can be done in that manner shown in Fig. 3.8(b). So even though the user was at a position close to the table and was categorised as paying attention, the participants used their own reference frame to place the item on the table as shown in Fig. 3.8(a). But it was clear that the participants used their own reference frame when the user is not paying attention to the table.

3.3 Effect of the Surface Area

3.3.1 Arrangement

For this study three tables were used instead of one. The three tables were identical in shape and were 0.75, 1 and 1.25 times in size in proportion to the table that was used for the first study. The arrangement of the table is same as for the study explained in section 3.1. Here also 32 binary color coded cards were used as objects. The purpose of this study was to identify the effect of the table size

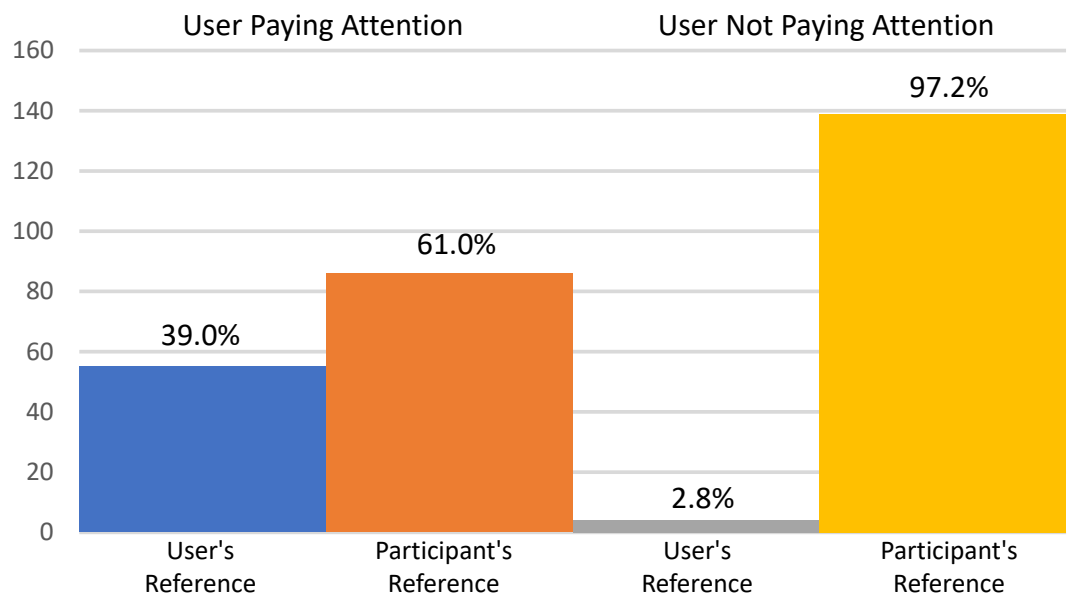


Figure 3.9: The results of the user's attention.

on the object placement. When selecting the table sizes it was decided to keep the object size comparatively smaller to the table size because the participants have a safety concern about the object so they did not place it too close to the edge of the table. If the table size is too small compared to the object size, this effect may interfere with the results.

3.3.2 Procedure and Metrics

Here also the recording of the vocal explanations was done manually and an overhead camera was used to record the locations of the objects. Participants were asked to place the items on 13 key locations as summarized in Table 3.2. For this study 25 personnel participated and they were asked to place the items on the three tables mentioned in section 3.3.1. In order to stop the effect of fatigueness and tiredness, the participants were called in randomly and performed the test on one table at a time. The locations of the objects were saved using the overhead camera alongside the location that the participants were asked to place the item.

3.3.3 Analyzing of Data

The results were analysed after categorising them into the each area term and size of the table. When recording details about the participants, their handedness was also recorded. The results were categorised by handedness and area terms. One way ANOVA test was performed on both X and Y axes coordinates of the object placement locations. The set of selected results are given in Fig. 3.10 as box plots. Here the results were scaled from 0 to 100 in order to be able to compared among different table sizes.

3.3.4 Results and Discussion

One of the main factors that influenced the placement of the object was the handedness of the participant. This was clearly seen when the object was asked to be placed on the far corners and edges of the table, where users had to reach out more, to place the object on the table. Fig. 3.10 (c) and (d) shows the box plot diagrams for X coordinate variation for the spatial term "Back Edge" of the table.

While figure (c) shows the placement for left handed participants, figure (d) shows the placements for the right handed participants. The enlarged diagram of the figure (c) is shown in Fig. 3.11. 'N', 'L' and 'S' represents Normal, Large and Small tables that were used. The large table was 1.25 times the normal table while the small table was 0.75 times the normal table as mentioned in Section 3.3.1. Maximum and the minimum values are shown by top and bottom whiskers in the plot shown in Fig. 3.11. The red color lines represents the median of the data. The upper flat line represents the 75th percentile of the data while the lower flat line represents the 25th percentile of the data. The upper notch of the angled lines represent the upper confidence limit of 97% confidence for median while the lower notches of the angled lines represents the lower confidence limit of 97% confidence for median. For example for the normal sized table maximum value is around 54.5 while the minimum value is approximately 44.5. The median is close to 48. So it can be said with a 97% confidence that the median is approximately between 45 and 50.5. All the values are represented as % distance so that they can be compared with each other.

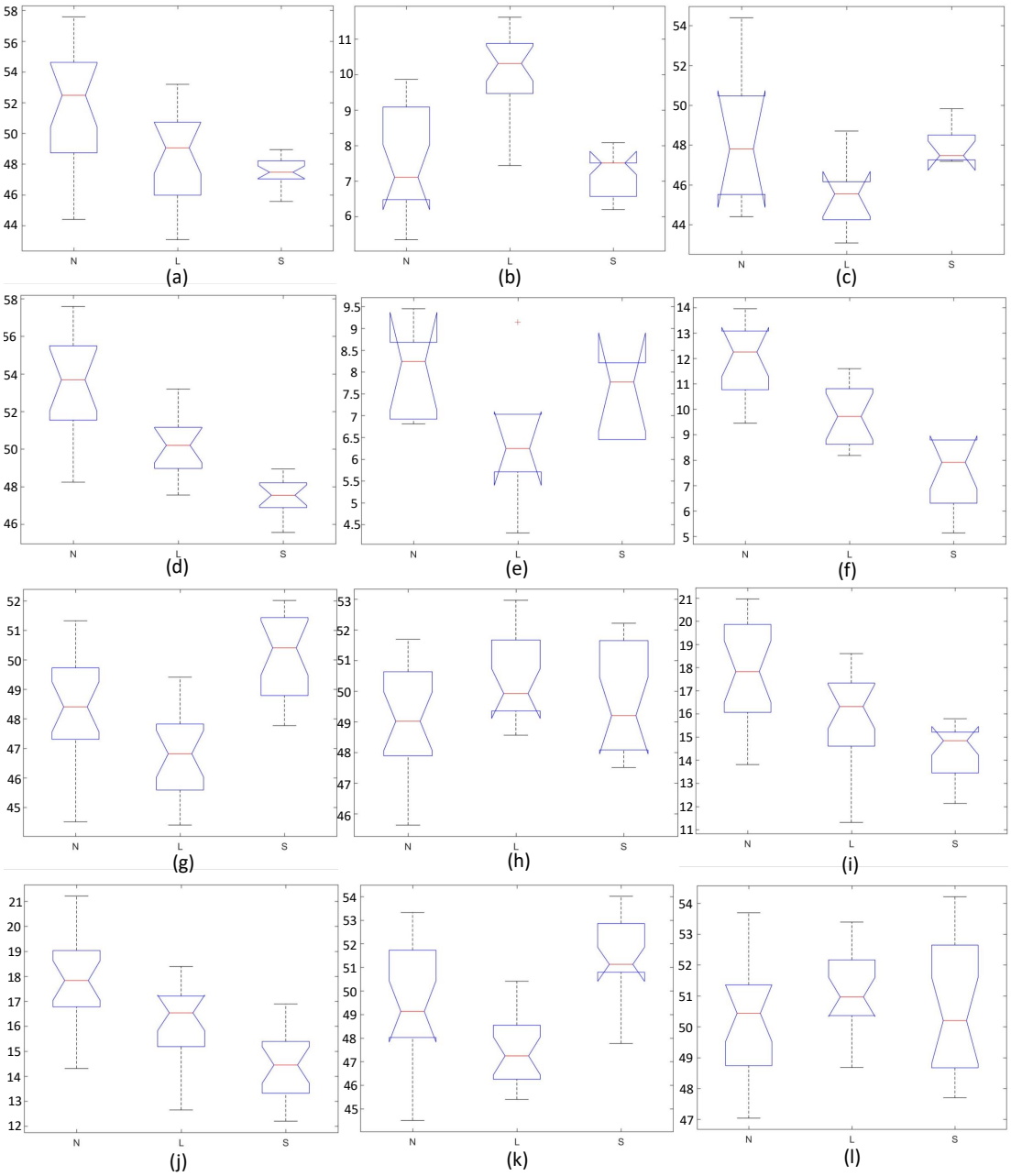


Figure 3.10: Box plot diagrams for selected results of the study 2 and 4 are shown here. From (a) to (h) the box plots are related to results from study 2 while (i) to (l) are related to results from study 4. All the results were scaled from 0 to 100 for comparison. Here "N" is the table size of 910mm*710mm while "L" is 1.25 times and "S" is 0.75 times the size of table"N". (a) and (b) shows the results for X and Y axis coordinates of the spatial term "Back Edge". (c) and (d) shows the left and right hand X coordinate results of the spatial term "Back Edge". (e) and (f) are the results for left and right hand X coordinates of area term"Back Left Corner". (g) and (h) shows the X and Y coordinate results of spatial term "Middle". (i) and (j) represents the results of X and Y coordinates of the spatial term "Back Left Corner". (k) and (i) are related to the X and Y coordinates of the spatial term "Middle".

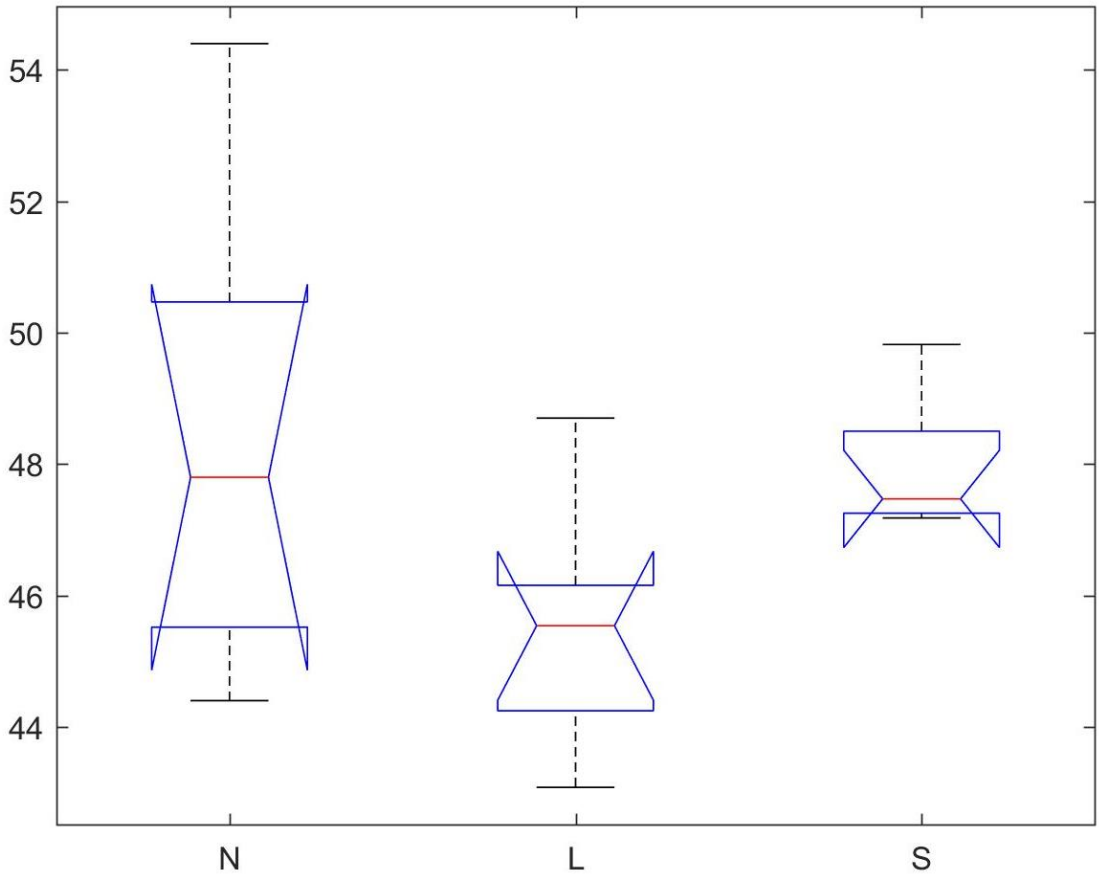


Figure 3.11: The box plot diagram of X coordinate for spatial term “Back Edge” for left handed participants

When analyzing two diagrams in Fig. 3.10 (c) and (d), significant different in object placement can be seen among the right handed and left handed participants. Where the left handed participants have placed the item to the left side of the table whereas the placements of the right handed participant have shifted to right side of the table. In Fig. 3.10 diagrams (g) and (h) depicts the X axis and Y axis distributions for spatial term middle respectively. The high variance present in the X axis compared to Y axis is due to the effect of handedness.

3.4 Effect of the Table Shape

3.4.1 Arrangement

This section of the study was conducted using 4 different shapes of tables. They are namely "Circular", "Oval", "Square" and "Rectangular". This study was performed as an extension to the study explained in section 3.1. The tables that were used for the experiment are shown in Fig. 3.12. They were arranged in the same manner as in the study explained in section 3.1. Here also 32 color coded objects were placed on the table and the users were asked to categorise them into spatial areas.

3.4.2 Procedure and Metrics

Twelve subjects participated in the study and they were asked to categorise the objects on each table. In order to avoid the fatigueness of the users, they were given one table at a time.

3.4.3 Analyzing of Data

The locations of the objects were analysed along with the spatial terms and the summary of the spatial terms are shown by Table 3.3. The visual distribution of area terms are shown by Fig. 3.13 where the results has been averaged to get a better representation.

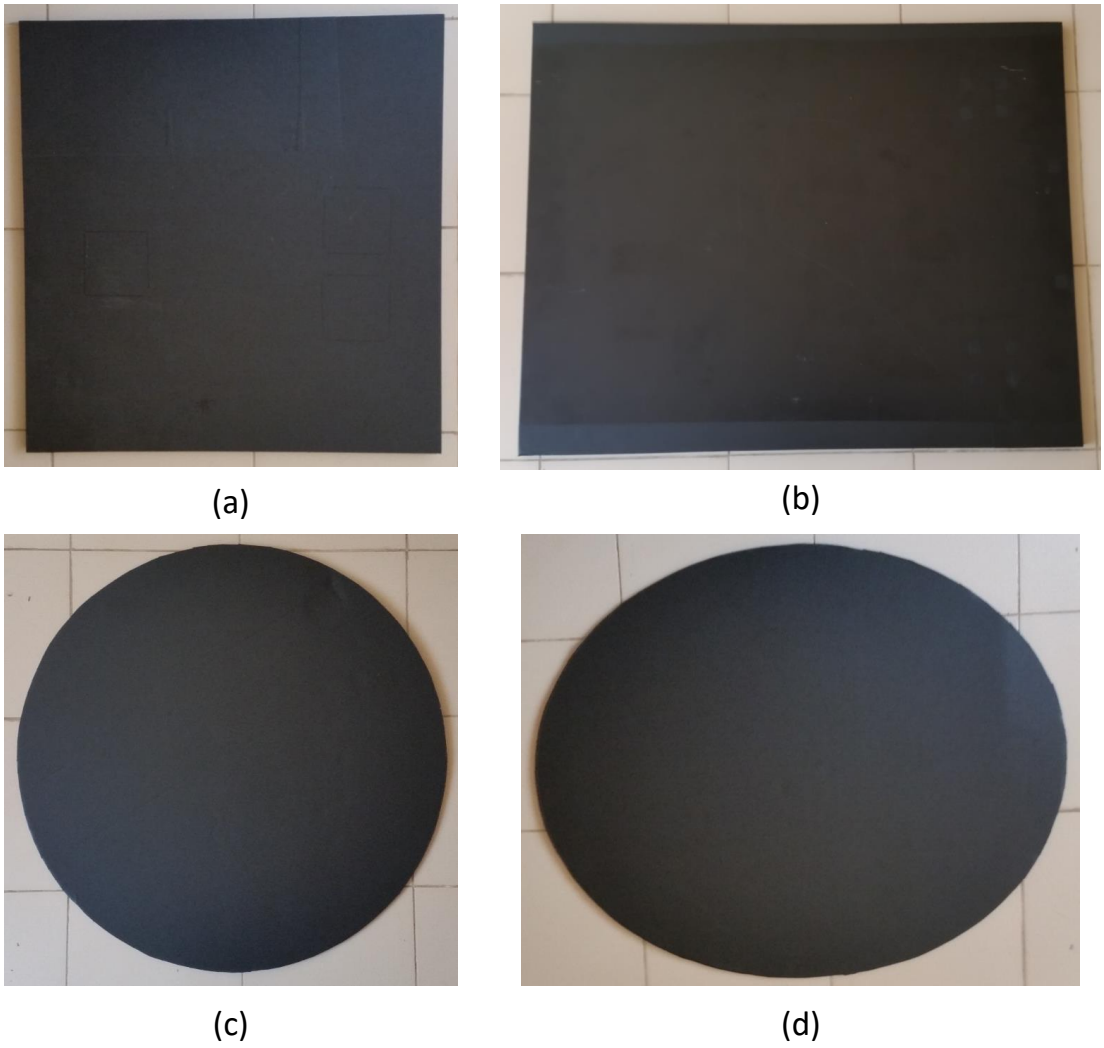


Figure 3.12: (a), (b), (c) and (d) shows the square, rectangular, circular and oval tables that were used for the study.

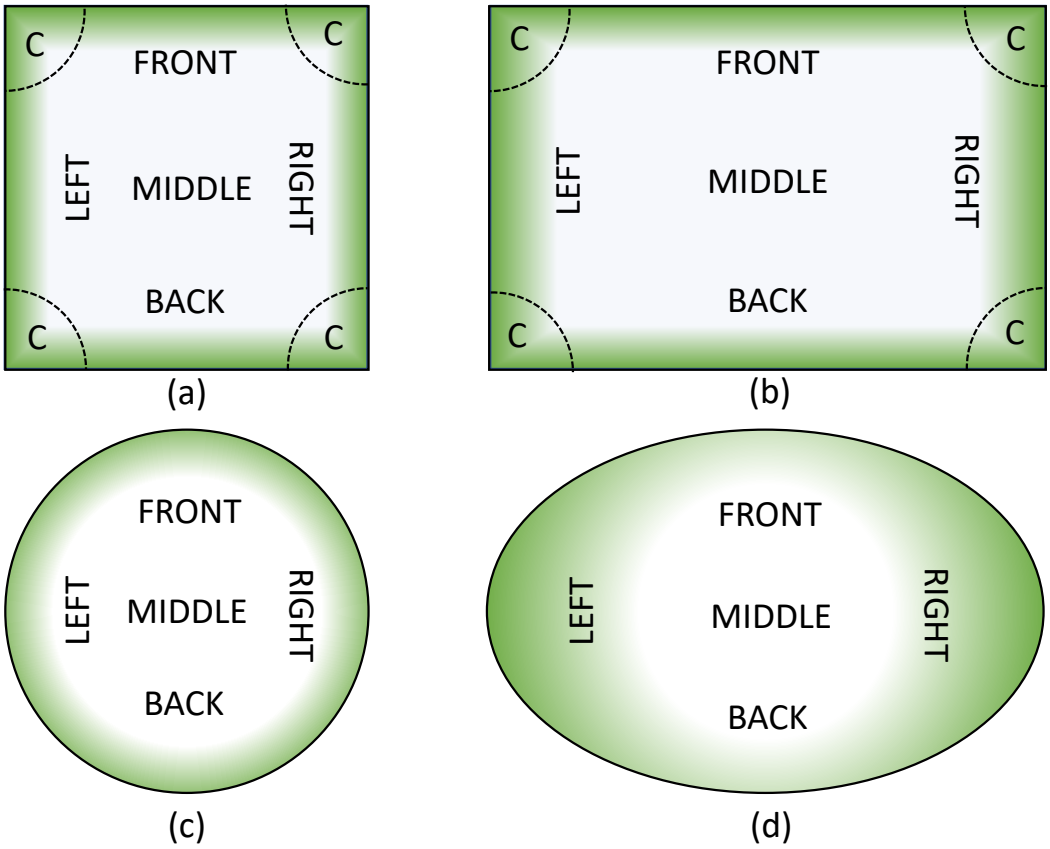


Figure 3.13: Averaged results for effects of the table shape. Figures (a), (b), (c) and (d) shows the identified spatial area terms for square, rectangular, circular and oval table shapes respectively. Highlighted in green color is the area term "Edge".

Table 3.3: Results for Effects of Table Shape

Table Shape					
Area Term	Subsection	Square	Rectangular	Circular	Oval
Middle	-	✓	✓	✓	✓
Left	-	✓	✓	✓	✓
Right	-	✓	✓	✓	✓
Front	-	✓	✓	✓	✓
Back	-	✓	✓	✓	✓
Corners	Front Left	✓	✓	-	-
	Front Right	✓	✓	-	-
	Back Left	✓	✓	-	-
	Back Right	✓	✓	-	-
Edge	Left	✓	✓	✓	✓
	Right	✓	✓	✓	✓
	Front	✓	✓	✓	✓
	Back	✓	✓	✓	✓

Figure 3.14: The table setup that was used to the study the effects of the objects on the table.

3.4.4 Results and Discussion

Here also the usage of spatial terms were similar to the study explained in section 3.1. But the key factor was that the tables Oval and Circular did not show case any area called “Corners”.

3.5 Effect of the Objects on the Table

When placing an object on the table, the effect of objects that are already on the table has be considered. The aim of this study is to examine the effect on the placement location of the object that is exerted by the objects that are already on the table.

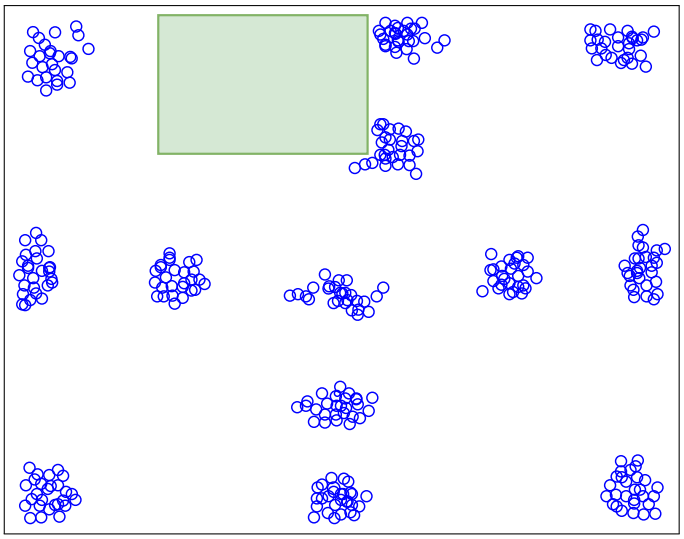


Figure 3.15: Results for table setting (a). Here the corresponding areas are shown as in Fig. 3.13 (b).

3.5.1 Arrangement

For this study the table that was used in study explained in section 3.1 was used. The table was kept in an empty room to avoid the external effects that might influence the object placement. Twenty six personnel were given the task of placing the objects on the table. The participants has a mean age of 27.65 years with a standard deviation of 8.94 years.

3.5.2 Procedure and Metrics

The participants were asked to place an object on 13 key locations on the table that were identified in the first study. These key locations are given in Table 3.2. Object (a book) was kept on one of the three locations shown by Fig. 3.14 (a), (b) and (c). The placements were recorded using a top mounted camera and the results were recorded alongside the occupied area on the table and the spatial terms.

3.5.3 Analyzing of Data

The results were analysed after mapping the placement locations of all users for each object (book) location on the table. The 3 table maps are presented in

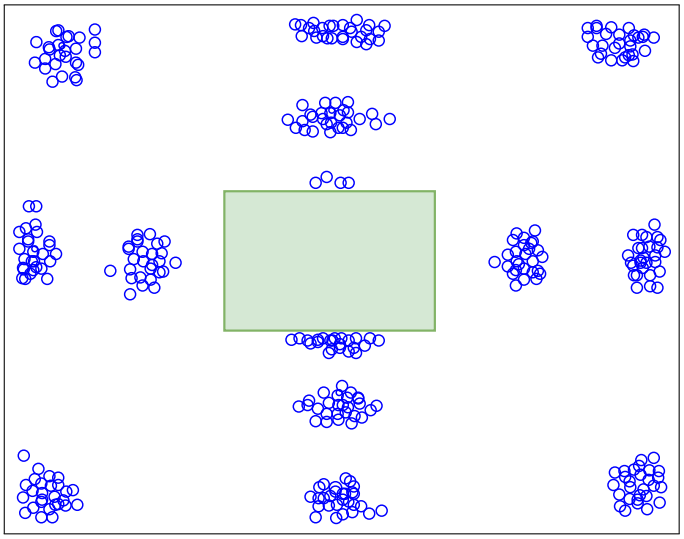


Figure 3.16: Results for table setting (b). Here the corresponding areas are shown as in Fig. 3.13 (b).

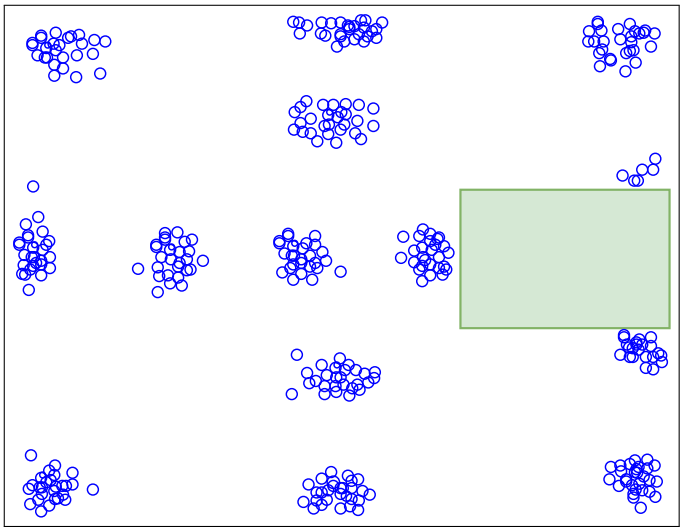


Figure 3.17: Results for table setting (c). Here the corresponding areas are shown as in Fig. 3.13 (b).

Fig. 3.15, Fig. 3.16 and Fig. 3.17. They corresponds to table settings that are shown in Fig. 3.14 (a), (b) and (c) respectively.

3.5.4 Results and Discussion

The results from this study are analysed after comparing with the results from section 3.1. Here it is clear that the users kept the objects avoiding the book that was on the table. One noticeable effect was the spatial terms that were most affected were the ones which are closer to the object that was on the table. This effect diminished when going from the nearest spatial term to terms that are further away. For example in Fig 3.15, spatial terms “Back” and ”Back Edge” are the most affected. But terms like “Front” the effect of the object can hardly be seen. Even in Fig. 3.16, spatial terms like “Corners” and “Edges” minimally affected. The effect of the handedness is also visible in the placements, where in spatial terms “Back Edge”, “Back Left” and “Back Right” the placements are more scattered when compared to the terms that are on the front of the table.

3.6 Effects of the Restrictions for Reachability

3.6.1 Arrangement

For this study, the participants were asked to place the object on different places on the table. But here the surrounding area of the table was obstructed by obstacles. The idea of this section is to understand how the participants alter their placement of the object depending on the restricted reachability to the required table area. Mainly to determine whether the participants will keep the item on the same location as in section 3.1 or whether they would change the location depending on the restrictions. The same three tables that were used for the study explained in section 3.3.1 , were used here. Object placement scenario with restricted reachability is depicted in Fig 3.18.

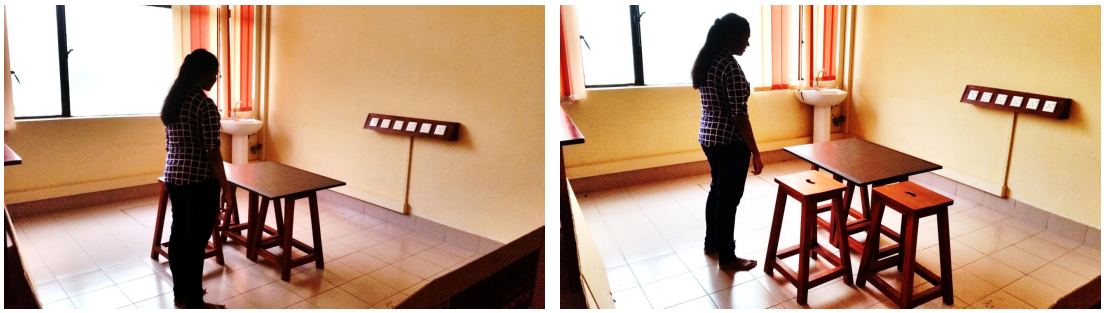


Figure 3.18: Effect of the restricted reachability.

3.6.2 Procedure and Metrics

In Fig 3.18 the participant is asked to place the objects on the table where the front left corner of the table was obstructed by two obstacles. Fig 3.18 (b) shows the scenario where the front right corner of the table was blocked using the same two obstacles. To analyze the above mentioned scenarios, 23 participants were asked to place the objects on 13 spatial areas shown in Table II. These locations of the object placement were saved and the results were analyzed.

3.6.3 Analyzing of Data

This section of the study produced 897 object placement locations. The results were analyzed after categorizing the placement coordinates into the spatial area and the table size. One way ANOVA test was performed for both X and Y axis of the placements in each category. The box plots that show key factors of this study is shown in Fig. 3.10

3.6.4 Results and Discussion

Box plot diagrams (i) to (l) in Fig.3.10 shows the effect of the restricted reachability from the obstacles that are shown by Fig.3.18 (a). Since the used obstacles did not exceed the height of the table, the placements on the front left corner didn't have much effect. But the effect was clearly visible in the locations that are on the middle and back side of the table and to the left. Fig.3.10(i) and Fig.3.10(j) shows the X axis and the Y axis distributions for the back left corner

of the table respectively. It is clearly visible that all the points has been moved to right side of the table as well as to the front side of the table when they are compared with the (e) and (f) box plots that are of the same location without the reachability restrictions. This effect is higher in the larger table than the smaller table. But the box plots shown by (k) and (l) show that there is not that much of difference when compared with the (g) and (h) which are of the spatial area "middle". So it can be concluded that the effect diminishes when it reaches the side of the table that is not obstructed.

The results of this study confirms to the fact that the reachability of the user to the table has a considerable effect on the object placement task. And the effect changes with the size of the table. This becomes evident when comparing the three tables in the box plot diagrams in the Fig.3.10. In the smaller table a narrow distribution among the placements can be seen. This is due to the fact that in the smaller table participants had the opportunity to reach out to places without much hassle. This also resulted in ideal object placements having a high correlation with the user expectation.

3.7 Summary of Human Studies

This chapter provided six human studies that were helpful in understanding the behavioral aspects of human users when it comes to object manipulation on a table. The results of these studies are used in designing a system that has the capability to behave more like a human user. This type of performance makes it easy for users to predict the behavior of the robotic system, hence making it easy for them to get accustomed to using this system.

First study, which is understanding the spatial terminology, helps to design a system that accepts the frequently uttered spatial terms when placing objects. An extension to this study has been carried out by incorporating different table shapes. The effects that exert upon the object placement because of the location of the person who is issuing the commands has also been studied, which will help to determine the reference frame that shall be used when understanding the user commands.

Furthermore, the effects of the objects that are on the table together with the effect of restricted reachability has been studied. The restricted reachability study shows how to place the object when certain areas on the table are not reachable, while objects on the table study shows how to avoid occupied areas on the table as well as to how to alter the placements when they are near to the occupied areas. Finally the effect of the surface area has been studied to understand how different sizes of tables will affect the object placements. The effects that were identified using these studies has been incorporated in the designed system and is explained in detail in Chapter 6.

The result of each study can be summerised as in Table 3.4.

Table 3.4: Summary of Human Studies

Effect	Results of the Human Study
3.1 Understanding of Spatial Terminology	Frequently used spatial terms and their area boundaries were identified.
3.2 User Location and Orientation	Effect of the user's location and orientation on the placement of the object were identified. Two reference frames - Robot's reference frame and the user's reference frame. The reference frame depends on the attention of the user. The distribution of spatial terms depend on the users location as well as the orientation.
3.3 Surface Area	The handedness of the user mainly influence the object placements. The larger table depicts wider distribution of object locations. The centroid of each area term seems to be distributed linearly depending on the length and the width of the table.
3.4 Table Shape	Frequently used spatial terms and their area boundaries with respect to 4 table shapes rectangular, square, oval and circular were identified.
3.5 Objects on the Table	Users avoided the objects that are on the table when placing the items. The effect of the objects are more on the areas that are closer to them. The effects diminishes over the distance.
3.6 Restrictions for Reachability	The unreachable areas were avoided by the users. Has a similar effect as 3.5 study. The effect diminishes when moving away from the unreachable areas.

Chapter 4

SYSTEM DESIGN

4.1 System Overview

The system contains three major modules. They are Uncertain Information Understating Module (UIUM) , Interaction Manager (IM) and Action Manager (AM). The output from depth sensor of the Kinect is used to extract the position and the orientation of the user, which is performed using the Visual Information Extraction Module (VIEM). The skeleton analysis of the user's body is performed using the Kinect SDK in order to extract the pointed hand gestures. The RGB camera that is available with Kinect is used to extract the information regarding the objects that are on the table by processing the images. An array of microphones exacts the voice commands that are issued by the user. These commands are then converted to text and processed using the Keyword Understanding Module (KUM). The KUM used a built-in data base that contains words which helps to understand the keywords that are in the user's commands. This is explained in detail in section 4.2. IM manages the interaction between the human user and the robot. The two data sets from the VIEM and the KUM is fed to the IM. IM uses these data feeds to evoke the relevant submodule in

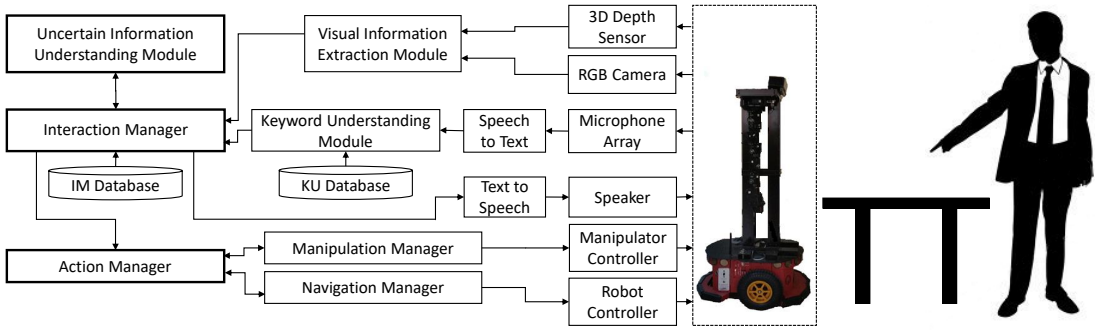


Figure 4.1: System overview.

UIUM to understand the uncertain information in user's commands. This module is explained in detail in Chapter 5. AM manages the high-level control of the robot. It uses Manipulation Manager (NM) to handle the placement of the object on the table, while Navigation Manager is used to navigate the robot around the table. NM or Navigation Manager Database contains the map of the room. Furthermore, the location and the orientation of the table is marked in the NM maps. These improves the navigation capacities of the robot which results in a much more precise maneuverer around the table. The low-level control of the robot platform and the robot manipulator is handled using the Robot Controller and the Manipulator Controller respectively.

4.2 Understanding Vocal Commands

There are two main parts in each voice command that were identified. They are referred to as keywords. One was action keywords and the other was the spatial keywords. Action keywords are the keywords that instructs the user what action to perform. As an example, words like "move", "place" and "keep" falls into this category. The other type is the spatial keywords. They define the spatial location where the object should be placed. For an example, words like "middle of", "left", "right" can be given. The keyword database (KU Database) also contains synonyms for these keywords that are stored along with these keywords. For an example term "center" can be pointed out as a synonym to the term "middle". Using this kind of approach allows the users not to be limited by a certain set of commands and remember a strict set of vocal commands.

4.3 Understanding Pointing Hand Gesture Location

The system uses skeletal information of the body joints, obtained from a 3D depth sensor to calculate the position where the user is pointing. In most of the previous systems that has been implemented, the position information has been obtained by extending the vector that goes through the elbow joint and the wrist or palm joint [6]. But humans usually point an object by aligning their palm

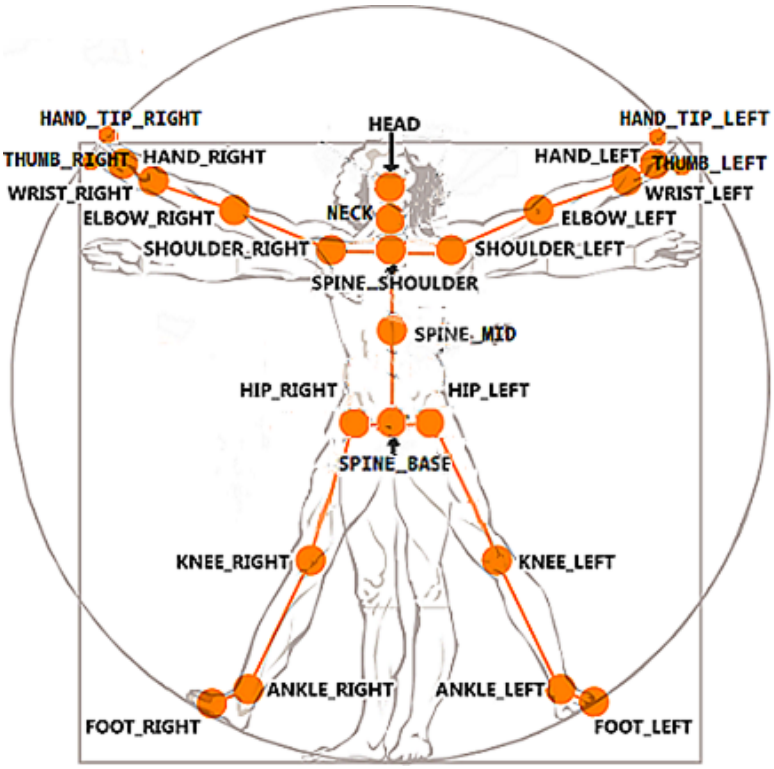


Figure 4.2: The tracked joints of the body by the Kinect.

along with the eye. So the most effective joints to be used are the head joint and the palm joint [30]. So the vector that goes through these joints is extended to get the intersection point with the table surface. This point is taken as the location where the user is pointing. When the system receives an action keyword, it calculates this vector. System considers both right hand and the left hand since the user can be right handed or left handed.

4.3.1 Kinect

A Kinect sensor was used to obtain the 3D skeletal tracking of the body. This sensor provides 3D depth information that can then be processed to estimate the skeletal joints of the human body. Researchers have found that Kinect sensor provides robust 3D data when compared with many available sensors in the market [37, 38]. The joints that are tracked by the Kinect are shown by Fig. 4.2 [39]. As mentioned earlier the vector that goes through the "HEAD" and "HAND_TIP" is considered.

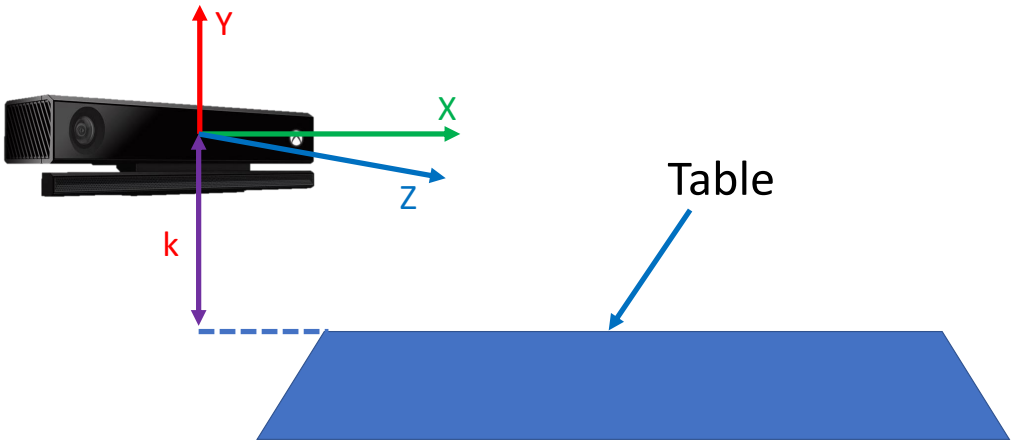


Figure 4.3: The X, Y and Z axes with respect to the Kinect. The placement of the table with respect to the Kinect.

4.3.2 Tracking of User

For the parts where the users orientation and location is tracked, the system uses the "SHOULDER_LEFT" and "SHOULDER_RIGHT" joints of the skeletal. The Kinect is mounted on a pan-tilt unit that allows the system to stay focused on the user.

4.3.3 Obtaining Pointed Location

The Kinect produces the location of each skeletal joint with respect to the Kinect's position. The X, Y and Z axes are shown in Fig. 4.3. The height of the table is a fixed value and is denoted by "k". The X and Z location values of the pointed hand gesture location is given by 4.1 and 4.2. The value "t" in those two equations are given by 4.3.

$$X = X_0 + t(X_1 - X_0) \tag{4.1}$$

$$Z = Z_0 + t(Z_1 - Z_0) \tag{4.2}$$

$$t = (k - Y_0)/(Y_1 - Y_0) \tag{4.3}$$

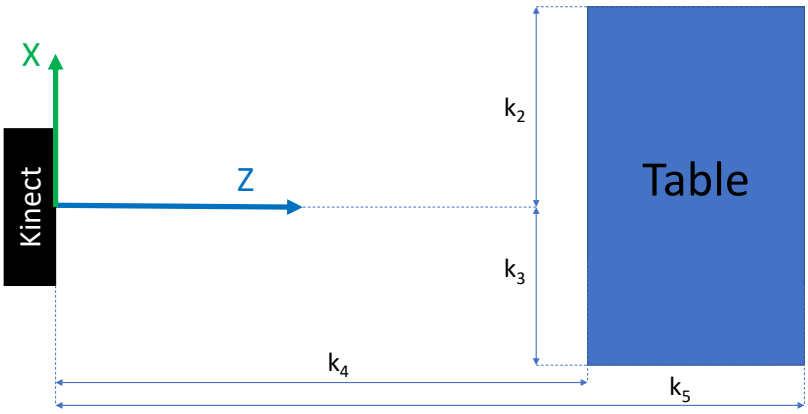


Figure 4.4: Top view of the Kinect with respect to the table.

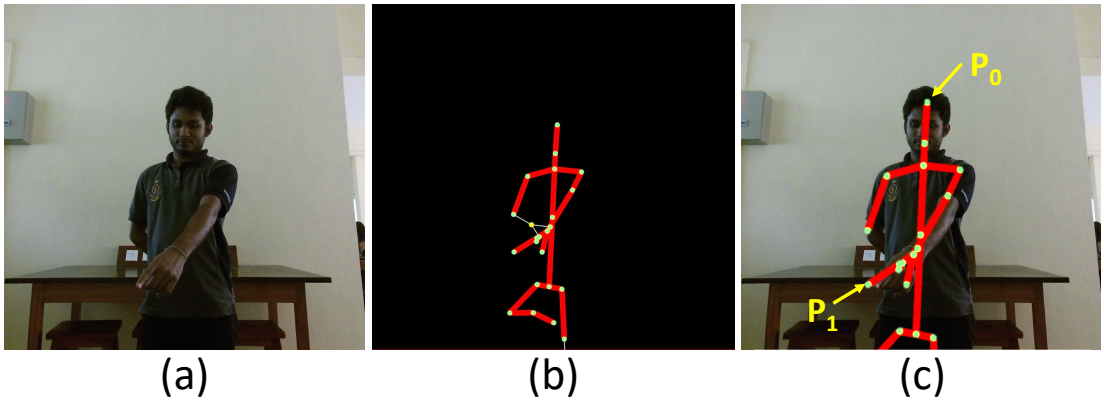


Figure 4.5: (a) shows the pointed hand gesture of the user as captured by the RGB camera of the Kinect. (b) shows the joint skeleton of the users body that is output by the Kinect SDK. (c) shows the image that is obtained after overlaying the two images.

An example is shown by Fig.4.5. Here figure (a) shows the RGB image of the user and figure (b) shows the tracked skeleton of the user. The vector that is marked using points $P_0(X_0, Y_0, Z_0)$ and $P_1(X_1, Y_1, Z_1)$ is extended to get the intersecting location on the table using the equations 4.1, 4.2 and 4.3. In order to eliminate the extraction of random hand gestures, the system only extracts a hand gesture where the vector that is extended is pointed towards the table. Which is called a valid hand gesture. Hence the X and Z values should be such that $k_3 < X < k_2$ and $k_4 < Z < k_5$. The k_2, k_3, k_4 and k_5 are marked on Fig. 4.4. Furthermore, the extraction of the hand gesture is only performed after receiving a user command that contains an action keyword.

Chapter 5

UNDERSTANDING UNCERTAIN INFORMATION IN USER COMMANDS

The users can instruct the system using three different types of commands. They are summerised in Fig.5.1. The first type is the commands that use only vocal instructions. Here the user can explain the spatial location using spatial terms. The second type is using hand gestures. Here the user can point out the location where the object has to be placed. The last type is the combined commands where the user can use both the voice commands and the hand gestures to give multi-modal inputs to the system.

5.0.1 Uncertain Information Understanding Module

This module interprets uncertain information that are used in the user commands. There are three sub modules that are within the UIUM for each type of user command. Each submodule contains two fuzzy inference systems for X and Y axes. The "Base case" refers to a instance where the placement values are calculated without any restrictions. For each submodule, the placement will be calculated for the base case and then be modified according to the restrictions that are available in each scenario. Here the X-Axis is explained in detail and

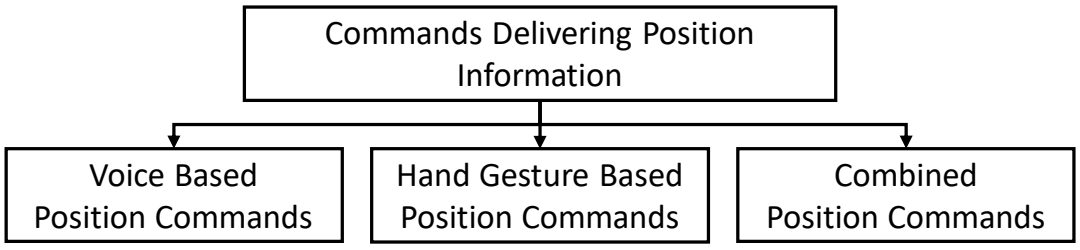


Figure 5.1: Command types that are used by the user.

Table 5.1: Rule Base For Fuzzy Modules 1 And 2

Input Memberships	Voice or Gesture				
	LE	L	M	R	RE
Output Memberships	VL	L	M	R	VR

the Y-Axis performs in the same manner.

5.0.2 Module 1 - For Voice Based User Commands

This module solely deals with voice commands. In order to be categorized as such the user commands has to have both action keywords and spatial keywords without a valid hand gesture. This module contains a single input single output fuzzy inference system to interpret the table locations with reference to the user commands. Fig. 5.2(a) shows the input membership functions and Fig. 5.2(b) shows the output membership functions for the fuzzy inference system. The rule base is given in the Table 5.1. The input fuzzy membership curves are singleton as voice commands directly request a specific area on the table. For an example, a command like "Keep the object on the left of the table" can be given. Here "Keep" is the action keyword and "left" is the spatial keyword. Since "left" denotes an exact area on the table, the input for the fuzzy inference system will be 0.25 for X axis and 0.5 for Y axis.

5.0.3 Module 2 - For Hand-gesture Based User Commands

This module is for the commands that are given using only the hand gestures. For example, "Place the object there" and pointing out the location can be given. This example scenario is shown by Fig.5.3 (a). In Fig.5.3 (b) the robot moves forward and places the object on the requested location. Here the user command contains an action keyword but no spatial keywords. These types of commands are categorized as hand gesture based user commands. When the system receives an user command with action keywords the system extracts the hand gesture. If there is a valid hand gesture but no spatial keywords; the system registers the command as module 2. A valid hand gesture is selected by considering the place the hand is pointing. If the hand is pointed towards the table then it is taken as a valid hand gesture command. If not, the system will ignore the hand gesture.

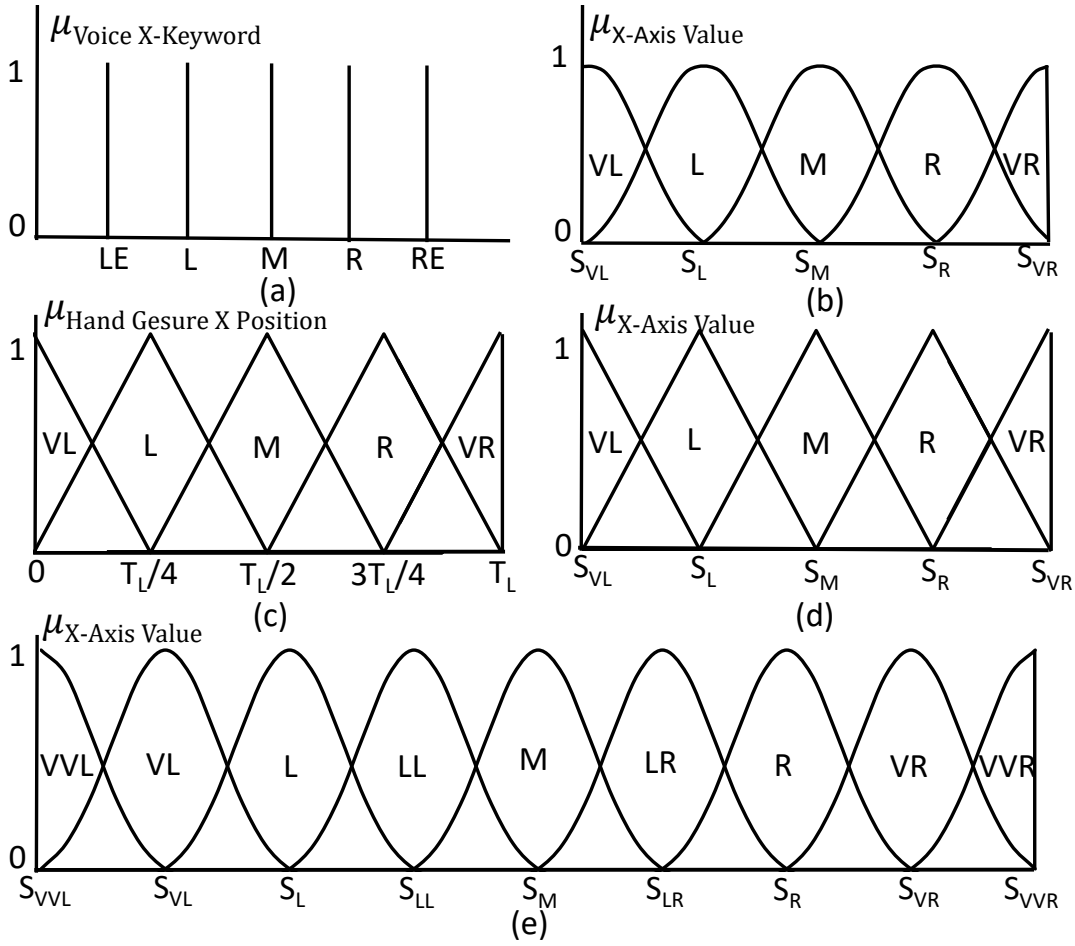


Figure 5.2: (a) and (b) represents the input and output membership functions for the module 1 respectively. (c) and (d) represents the input and output functions of module 2 respectively. (e) represents the output membership functions of module 3. Fuzzy labels are defined as LE:Left+Edge, L:Left, M:Medium, R:Right, RE:Right+Edge, VVL:Very Very Left, VL:Very Left, LL:Light Left...etc.

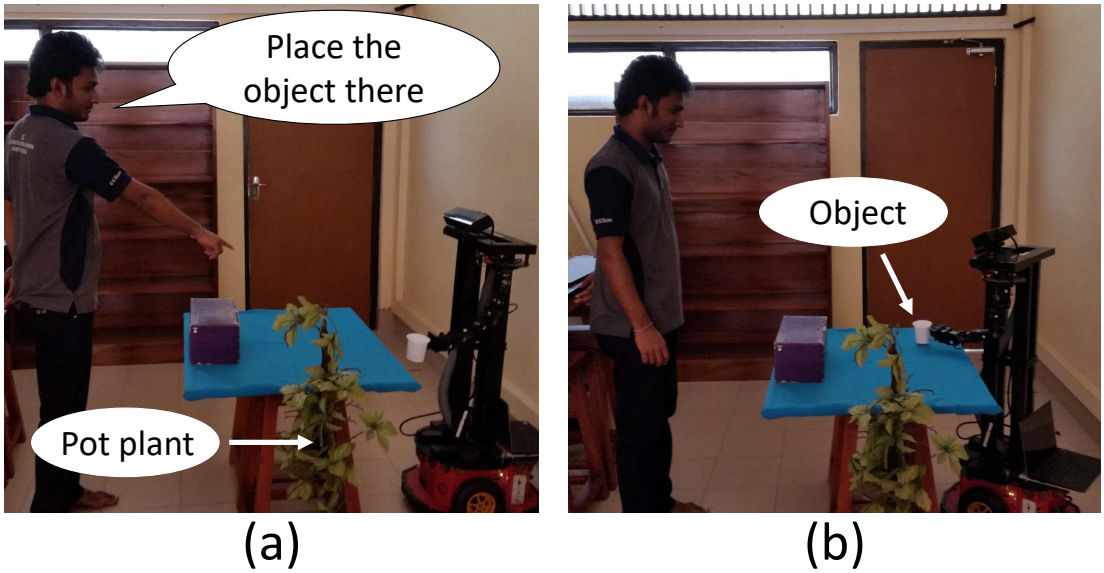


Figure 5.3: An example scenario for hand gesture based user command.

Having this type of behavior allows the system not to extract erroneous hand gestures or irrelevant hand gestures, or else register a random hand movement as a pointed gesture. If the user command contains action keywords without any spatial keywords or valid hand gestures the system will ignore the command and ask the user to give the command again. The input membership curves are shown by Fig. 5.2(c) and the output membership curves are shown by Fig. 5.2(d). Here triangular curves have been used since the measurements are performed as linear distance in millimeters. Furthermore, there is no strict requirement for a fuzzy based system. The extracted hand gesture point is a exact location on the table and the placement can be performed on the extracted location directly. In order to apply the modifications that will be explained in the coming sections, it is essential to have a fuzzy based system. The rule base for the system is given in Table 5.1.

5.0.4 Module 3 - Combined User Commands

This module handles user commands with both vocal and hand gesture positional information which are referred to as combined positional commands. Here if the user command contains both action keyword and spatial keyword alongside a valid hand gesture the user command is taken as a combined user command and is handled using module 3. This module contains a two input one output fuzzy

Table 5.2: The Rule Base For Fuzzy Module 3

Input Membership		Voice Commands				
		Left Edge	Left	Middle	Right	Right Edge
Hand Gesture	Very Left	VVL	VL	L	Clarify	Clarify
	Left	VL	L	LL	LR	Clarify
	Middle	L	LL	M	LR	R
	Right	Clarify	LL	LR	R	VR
	Very Right	Clarify	Clarify	R	VR	VVR

inference system. The input membership functions are shown by the Fig. 5.2 (a) and (c) while the output membership functions are shown by the Fig. 5.2 (e). The input membership functions are same as the ones used in Module 1 and 2. The rule base for this module is given in the Table 5.2. As an example, a user command like "Keep the object on the right side of the table" can be given alongside a hand gesture which point towards the table. The system will extract "Keep" as the action keyword and the "right" as the spatial keyword. The hand gesture will be used to calculate the location on the table where the user is pointing.

Chapter 6

SPATIAL CONCERNS

6.1 Concerns for Space Properties

6.1.1 Table Size

The system needs to have the ability to adjust to different sizes of tables since the study that was explained in Section 3.3.1 shows evidence that the system should have the capability adapt to such variations. In order to incorporate this behavior in the system, the fuzzy membership curves are adjusted according to the size of the table. T_W and T_L which are the width and the length of the table respectively, are taken into consideration when deciding the place for the object placement.

6.1.2 Table Shape

The most common table shapes found in domestic environment can be categorized into four basic shapes as shown in Fig. 7.7, namely 'Square', 'Rectangular', 'Circular' and 'Oval'. This modification process takes place before the UIUM decides the location for the object placement. For square and rectangular table shapes this modification is irrelevant since T_L and T_W is uniform along the whole span of the table. But for the other two table shapes the values of T_L and T_W has to be recalculated depending on the size and shape of the table, since there is no exact shape for oval tables. Their dimensions can alter from design to design where for a specific maximum length of the table, the maximum width can change.

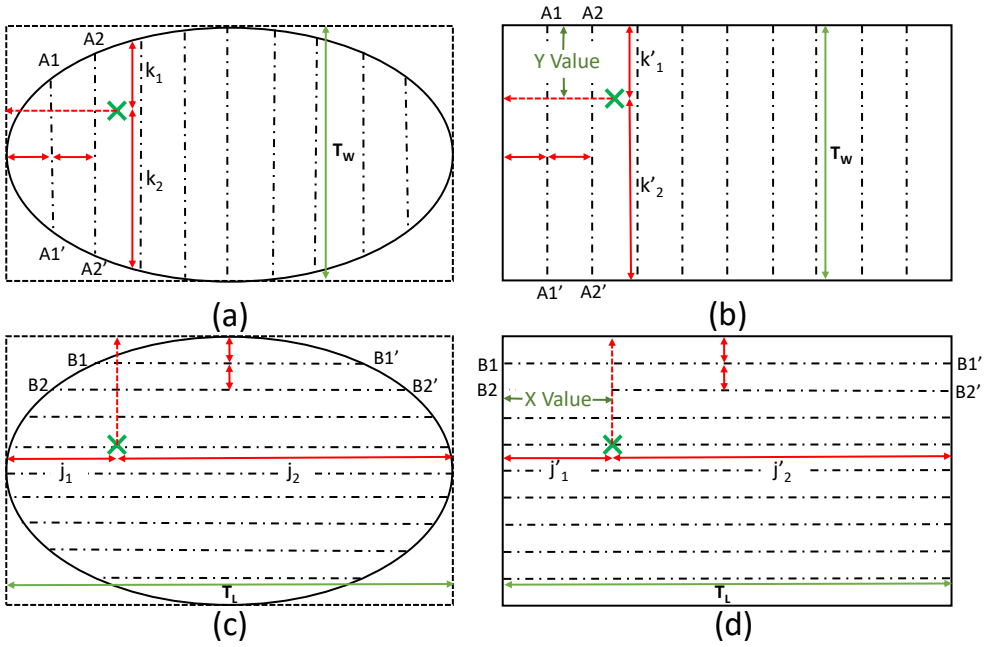


Figure 6.1: Recalculating of pointed hand gesture location for oval and circular shaped tables.

In order to build a system that can handle different shapes of oval tables following method is used. This is performed by expanding the table shape in both X and Y axis into a rectangle that can encapsulate the table shape. It contains two submodules for X and Y axes. The X axis submodule expands the table shape in X direction and Y axis submodule expands the table shape in Y direction.

An example Y-Axis transformation is shown by Fig. 6.1(a) and (b), where the oval shaped table is expanded to represent a rectangle. "X" mark on the table as shown in figure (a) depicts the hand gesture pointed location on the table. The transformation is performed so that dotted lines represented by $A1 - A1'$ and $A2 - A2'$ are stretched linearly to obtain $A1 - A1'$ and $A2 - A2'$ in figure (b). The Hand gesture point "X" is also relocated so that $k_1/k_1' = k_2/k_2'$. Figure (c) and (d) shows transformation for the X-Axis value for a hand gesture.

The advantage in using this type of approach is that this system can adapt to variety of table shapes, without impractically depending on a mathematical equation to model the parameters of each and every shape of table. After the location placement of the object has been determined using the UIUM, the location is recalculated using the backwards transformation in order to find the location

on the actual table.

6.2 Effect of Dynamic Space Constraints

6.2.1 Objects on the Table

This step determines the distribution of unoccupied area on the table. When placing objects in a requested location, both the value and distribution of free space is considered equally important. The occupation area of obstacles in each classified area is calculated using image processing as shown in Fig. 6.3. This calculation is performed using the point of view of the robot. Since humans perceive the world through the point of view of their eyes, for a human–like system a camera mounted right on top of the table will not justify this phenomenon. In Fig. 6.3, area occupied by obstacles are marked by red and the total table surface is marked by green.

From the study explained in Section 3.5, the requirement is to place the object on the table taking into consideration the distribution of the space. In the study it was found that the effect of the obstacles on the table affects placement closer to them more than the placements that are away from them. For example if there is an object on the middle of the table, areas close to the middle will be affected more than the areas that are closer to the edges of the table. This type of behavior has to be incorporated in the system to follow more human-like behavior.

The system follows a process where the fuzzy curves are moved according to the space that is obstructed by objects, in order to move the placement location of the object. In Fig. 6.2 (a) output membership functions are for the empty table while in the (b) output membership curves the "Left – L" and "Middle – M" membership functions has been shifted to the right side because of the object (book) that is on the table.

The modification factors for obstacles are calculated using (6.1). In order to get the distribution of the obstacle free space, a weighting factor $K_i^{ObjectFree}$ is calculated for each output membership function. The output membership

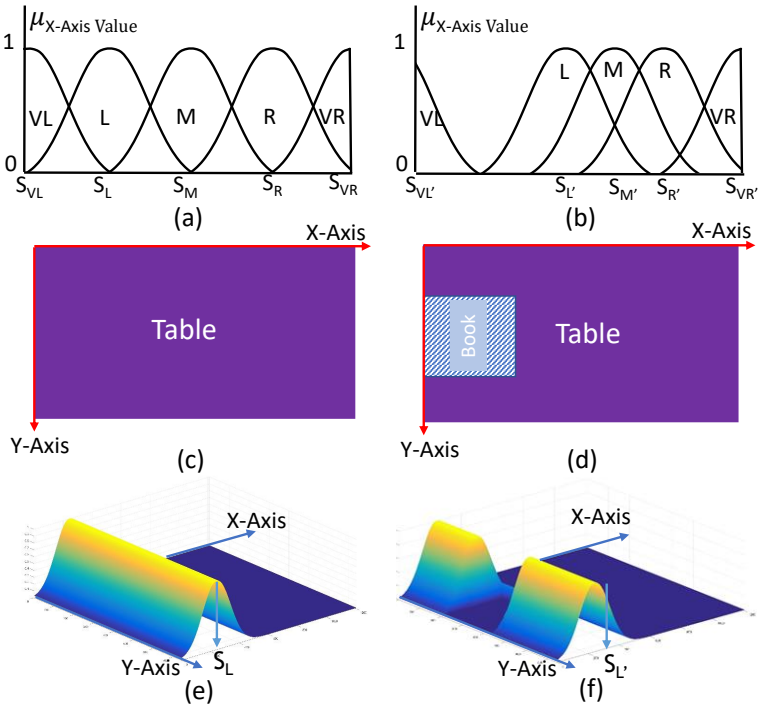


Figure 6.2: Shift in the fuzzy output curved due to the objects on the table.

functions values S_i is modified accordingly. For Module 1 and 2, $i = \{$VL, L, M, R, VR$\}$ and for module 3, $i = \{$VVL, VL, L, LL, M, LR, R, VR, VVR$\}$.

$$K_i^{\text{Object Free}} = C_i^{\text{With Objects}} - C_i^{\text{Without Objects}} \tag{6.1}$$

When calculating this value the 3D distribution of the output membership functions are considered. For an example the 3D distribution of output membership function "Left" in module 1 is shown in Fig. 6.2(e). When calculating the modification factor, this distribution is taken as a solid volume and the (X,Y) coordinates of the center of gravity of this volume is calculated, which is denoted by $C_i^{WithoutObjects}$. When there is an object on the table (the book in this case) the value of this distribution at those occupied areas are taken as zero. The resulting distribution is shown in Fig. 6.2(f). The value $C_i^{WithObjects}$ is calculated for this distribution in the same manner by obtaining the centroid. So each curve is shifted using 6.1. The shift of each curve reflects how much area the object (book) has affected each curve. So Curve that represents "Middle" will have a lower shift when compared to the curve representing "Left" and curves representing "Right" and "Very Right" will have zero or negligible shift. This type of

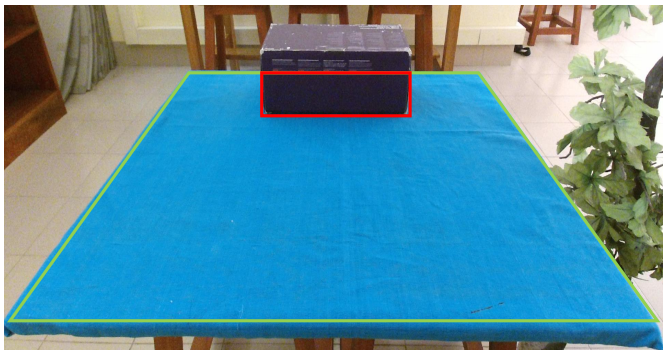


Figure 6.3: An example for extracting occupied area by objects on the table.

system behavior not only helps to avoid placing the object on other objects on the table but also encapsulate the behavior of a human being where the placement of the object has been shifted a certain amount from the original value, in near locations to the object on the table. For an example, if the robot was asked to place the object on the middle of the table, the placement location will be moved slight to the right side of the table. This type of system behavior tally with the results obtained from the study in Section 3.5.

6.2.2 Reachability of the Robot

The reachability of the robot is the area that can be manipulated by it. Here the arm reach and the position of the robot is considered. The extent to which robot can get closer to the table varies according to the placement, orientation and height of the table. The results from the study 3.6 shows the effect that restrictions in reachability has on a human participant when placing the object. Human beings has a complex anatomy that allows them to bend and reach different locations on the table. But when it comes to a robot the reachable location on a certain height is fixed and there are areas that the robot cannot reach. Since the height of the table is fixed, the reachability for a certain location depends on the surface area of the table and the obstacles around it. The Navigation Manager lets the robot reach different places on the table by reaching out from different directions around the table, but some areas may not be reachable anyhow.

For this modification also the 3D distribution of the output membership functions are considered. The calculations are performed in the same manner as for the obstacle modification factor. In each output curve 3D distribution is consid-

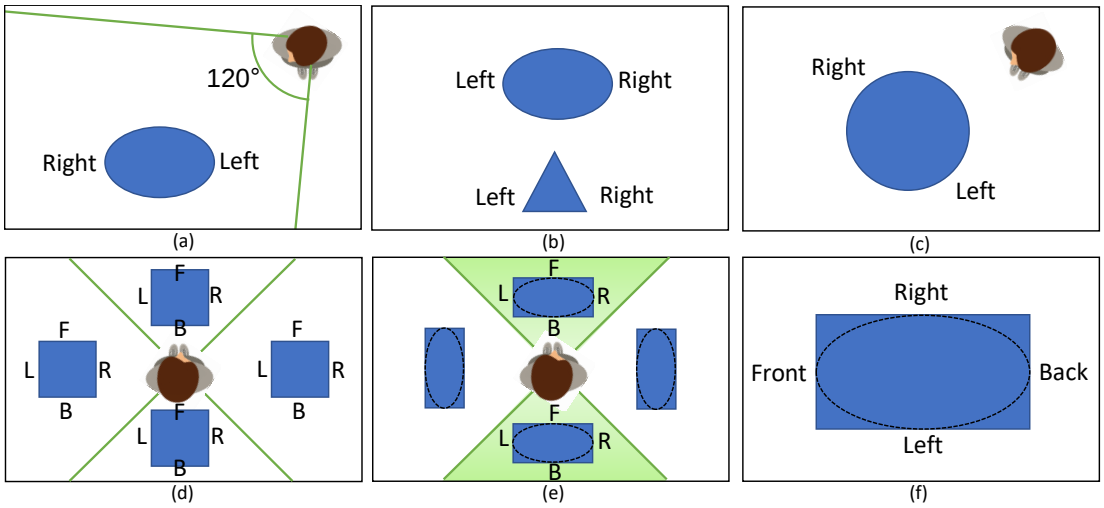


Figure 6.4: Effect of the User's orientation is shown here. Where area terms are represented as "L"-Left, "R"-Right, "F"-Front and "B"-Back.

ered and the value of the unreachable areas are taken as zero. So the placement in those areas will be avoided. Furthermore, there will be shift in each output curve near the unreachable areas. This behavior matches with the results obtained from Section 3.6. Where the study shows that even-though some locations are reachable, human participants were reluctant to reach out and place objects on or near unreachable areas. Modification factor reachability $K_i^{Reachability}$ is calculated using 6.2.

$$K_i^{Reachability} = C_i^{RR} - C_i^{NR} \tag{6.2}$$

Where C_i^{RR} is the center of gravity on the XY plane of the 3D distribution of the output membership functions with restricted reachability. C_i^{NR} is the center of gravity on the XY plane of the 3D distribution of the output membership functions without restricted reachability.

6.3 Special Attention

6.3.1 Concerns for User Location and Orientation

The Kinect sensor was mounted on a pan-tilt unit using which the sensor was directed at the user to track the user inside the room as mentioned in Section 4.3. The collected data includes the position, body orientation and the head orientation of the user which were the main factors that were identified during the user study to have an effect on the reference frame. This section mainly focuses on determining the reference frame which is used by the user. The head orientation of the user was used to estimate whether the user was paying attention to the table or not. As shown in Fig. 6.4(a), if the table is in the 120 degree angle then it was considered as the user was paying attention to the table. The reference frame of the user was determined by body orientation of the user. The left side of the table was taken as same as the left hand side of the user. For circular shaped table this was applied without any bounds since there were no strict sides as shown by Fig. 6.4(c). For square shaped tables, spatial terminology allocation is divided into four quadrants as shown by Fig. 6.4(d). The suitable quadrant is selected depending on the orientation of the user with respect to the table.

For oval and rectangular shaped tables the orientation of the users body with respect to the table was divided into four quadrants as shown by Fig. 6.4 (e). But here only two quadrants that are highlighted, use the reference frame with respect to the user. The other two cases use the robot's reference frame. This was due to the fact that from the user study it was found out that the users did not categorise the spatial keywords for these two tables as shown by Fig. 6.4 (f). For the cases where the user was not paying attention to the table, the reference frame of the robot was used to place the object. So as shown by Fig. 6.4 (b), the robots reference frame was used to place the object. But this also followed the same quadrants laws as stated earlier.

6.3.2 Calculating the Placement Position

In order to find the membership values for S_i the density distribution of the classified areas is used.

Equation 6.3 shows the modification for reachability for membership function values S_i. Here the effect of the two constraints for placing the objects, free space and reachability is averaged in order to get the final value. This is because the effect of both the scenarios affect equally to the placement of the object.

$$S_i = Si_{Base} + (K_i^{ObjectFree} + K_i^{Reachability})/2 \tag{6.3}$$

Where Si_{Base} value is calculated using the most preferred areas classifications from the user study. This value is calculated for each output membership function respectively.

6.3.3 Safety Distance

Final step is to determine the safe positioning of the object. Each object has a safety clearance that is required in order to manipulate the object as well as to ensure that the object will not come in contact with other objects. Another concern is about the safety of the object near the edges and corners of the table. So from the closest edge the item needs to have a minimum distance of $D_E + D_S$. Where D_E is Edge Clearance and D_S is Safety Clearance. The item is placed in the position where the centroid of the object is as close as possible to the calculated position. In Fig. 6.5 this value is denoted by D_C. The highlighted green area shows the possible area for the object to be placed after considering D_S and D_E. The item will be placed where D_C is minimum.

6.4 Example Scenario

This section explains an example scenario that uses the total system capabilities. In this scenario the user is using both pointing hand gestures and voice

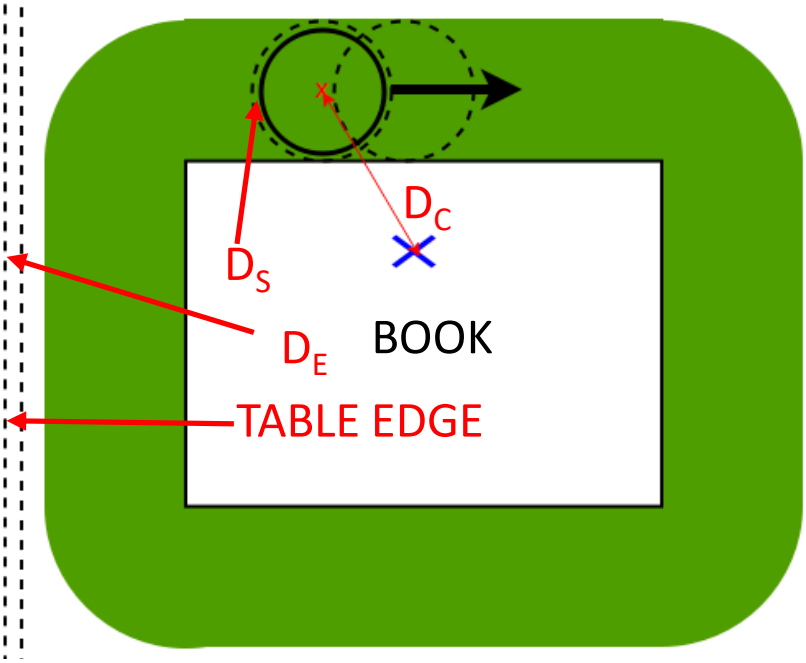


Figure 6.5: Safety concerns when placing the object.

commands as shown by Fig 6.6. The user command is "Keep the object in the middle of the table". So the action keyword is "Keep" and the spatial keyword is "Middle". Since there is a valid hand gesture which points towards the middle of the table, submodule 3 of UIUM is selected.

The pointed hand gesture location is calculated using the skeletal information of the user. Since the user is right handed, the right hand is considered for calculations. The handedness is identified using the hand that is pointing towards the table. The tracked skeletal is shown by Fig. 6.8. The 'Eye - Hand Tip' vector is calculated and considering the location of the table, the pointed hand gesture location is (-95mm,-102mm). Fig. 6.7 shows the robot's point of view of the user.

The system considers the user as paying attention to the table. So the spatial area term classifications will be taken as shown in Fig 6.4 (b). In this particular scenario the attention of the user is not considered because a valid hand gesture is available.

The Kinect is then tilted to analyze the objects that are on the table. Fig. 6.8 shows the image that is taken of the table. With this image the system performs image processing to find the occupied areas on the table.

Figure 6.6: Example scenario using both voice and hand gesture based commands.

Figure 6.7: Robots point of view of the user.

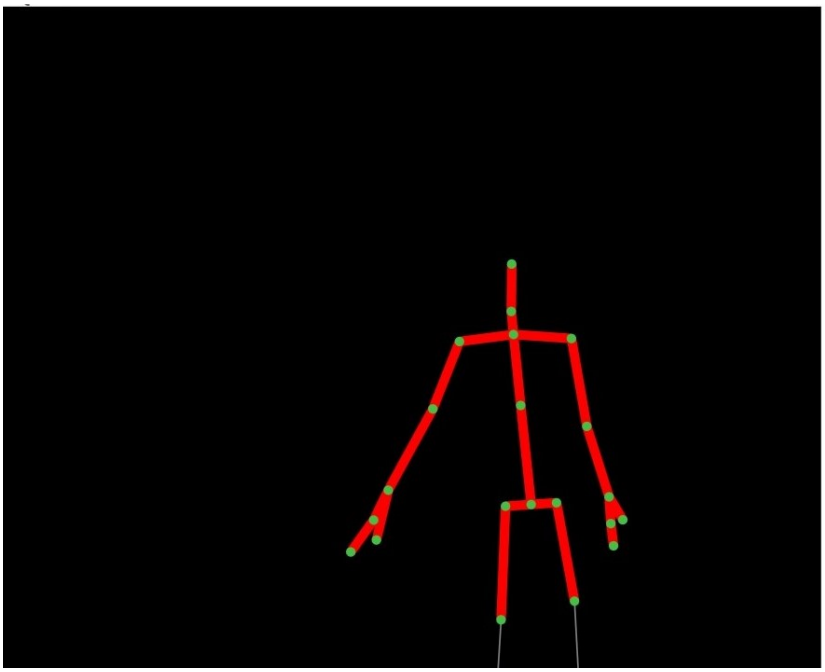


Figure 6.8: Tracked skeleton of the user's body.

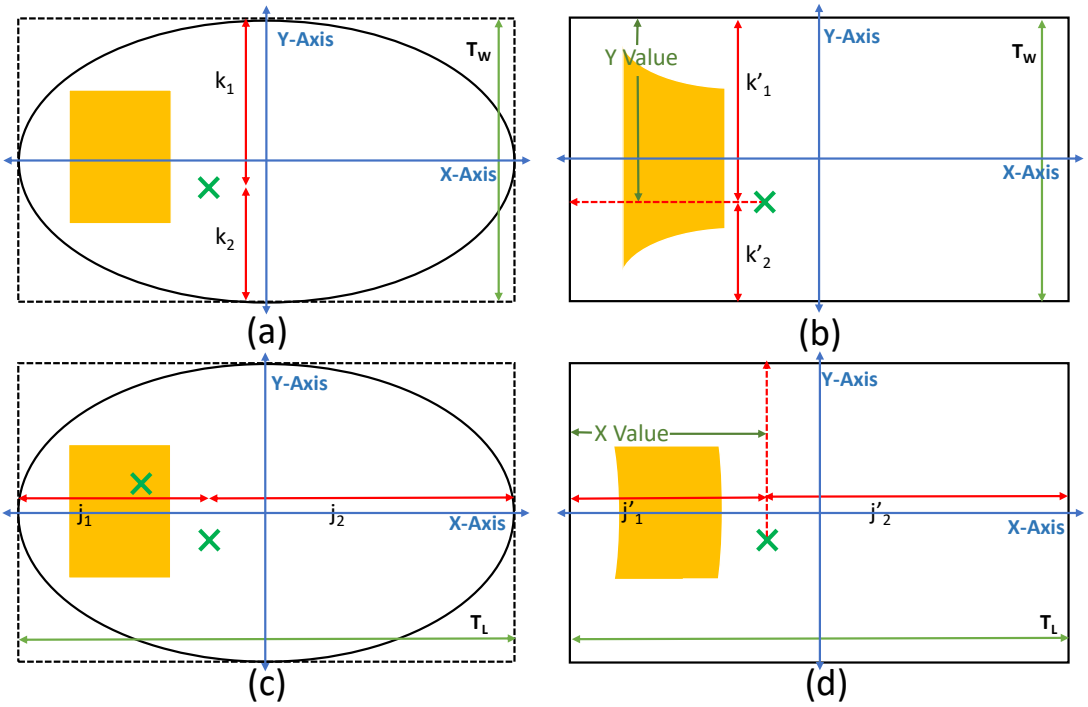


Figure 6.9: Y and X axis transformations from oval shaped table to a rectangular shaped table.

Next the system searches for obstacles that may be around the table which will restrict the reachability of the robot to certain area on the table. But since there are no obstacles available, the total surface of the table is taken as reachable.

In order to use the module 3 of UIUM the table has to be converted to rectangular shape. The conversion is performed separately for the two axes. The X axis conversion is shown in Fig. 6.9 (c) and (d). The green "X" shows the hand gesture location which is also recalculated. In Fig. 6.9 (d), the hand gesture point is recalculated so that $j_1/j_2 = j_1'/j_2'$. The areas that are occupied and unreachable are also converted using the same algorithm. In Fig. 6.9 (c) the occupied area (by the book) is shown by yellow color and in Fig. 6.9 (d) the same area is shown by yellow but recalculated.

The table in Fig. 6.9 (b) is fed to the UIUM module 3 Y axis while Fig. 6.9 (d) is fed to the X axis. The output membership functions are modified as mentioned in Section 6.2.

The inputs to the UIUM includes "Middle" for X axis spatial term and $-103mm$ as X axis hand gesture value. For Y axis the spatial input is "Middle" and hand gesture input is $-112mm$. The hand gesture location is recalculated according to the table conversion shown in Fig.6.9.

After calculating the placement location for rectangular table, the location is recalculated to the oval shape table using the reverse process of the method shown in Fig. 5.3. The final placement location is calculated as $(-73mm, -82mm)$. The system then examines whether this placement will violate any safety concerns. Since there are no obstructions the system will proceed with the placement as shown by Fig. 6.12. After the placement is complete the robot will move back to the starting position as in Fig. 6.13.

Figure 6.10: The robots point of view of the table.

Figure 6.11: The robot moves forward to place object.

Figure 6.12: The robot completes the placement of the object.

Figure 6.13: The robot is returning to the starting position

Chapter 7

RESULTS AND DISCUSSION

7.1 Hardware Implementation

The system has been implemented on the MIROB platform in [40]. The MIROB consists of a 'Pioneer 3DX' mobile platform alongside 'Cyton Gamma 300' manipulator. The MIROB platform with a Kinect sensor connected on top of a pan tilt unit is shows in Fig. 7.1.

7.2 Experimental Setups

The experiments were conducted using two setups. In the first setup the main focus was on the system's ability to alter the placements according to objects that are on the table and the consideration for restricted reachability of the user so a single table was used that is rectangular in shape. In the second setup the results were obtained mainly focusing on the system's ability to consider different table shapes and the user's orientation and position so four tables were used as explained in section 7.2.2. Multimodal capabilities of the system were analysed in both setups using voice, hand gesture and combined user commands.

7.2.1 Experimental Setup 1

The experiments were carried out using different arrangements of the table, under both limited free space and reachability conditions. In order to limit the free space available, objects were kept on the table as obstacles. Limitations for reachability were achieved by introducing obstacles in the map near the table,

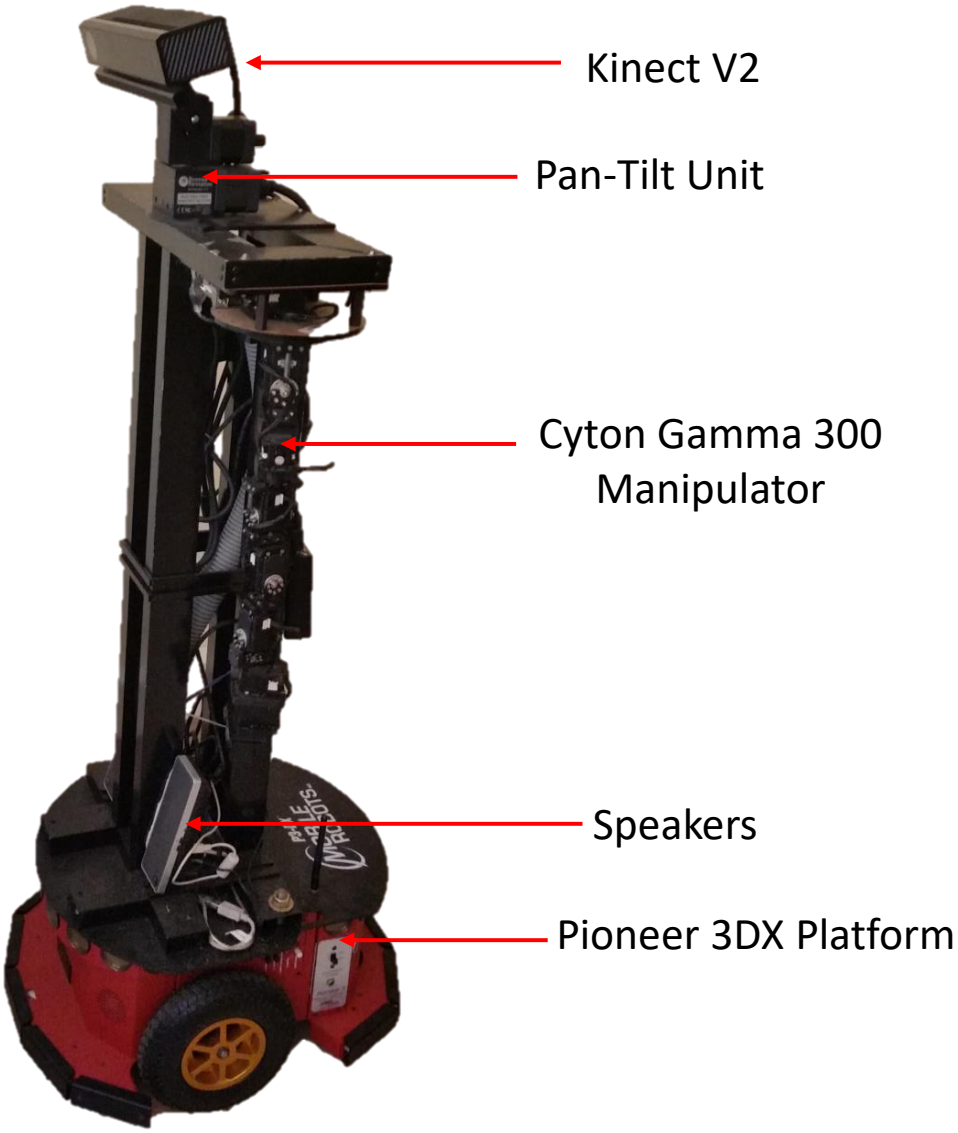


Figure 7.1: MIROB - The hardware that was used to implement the system.

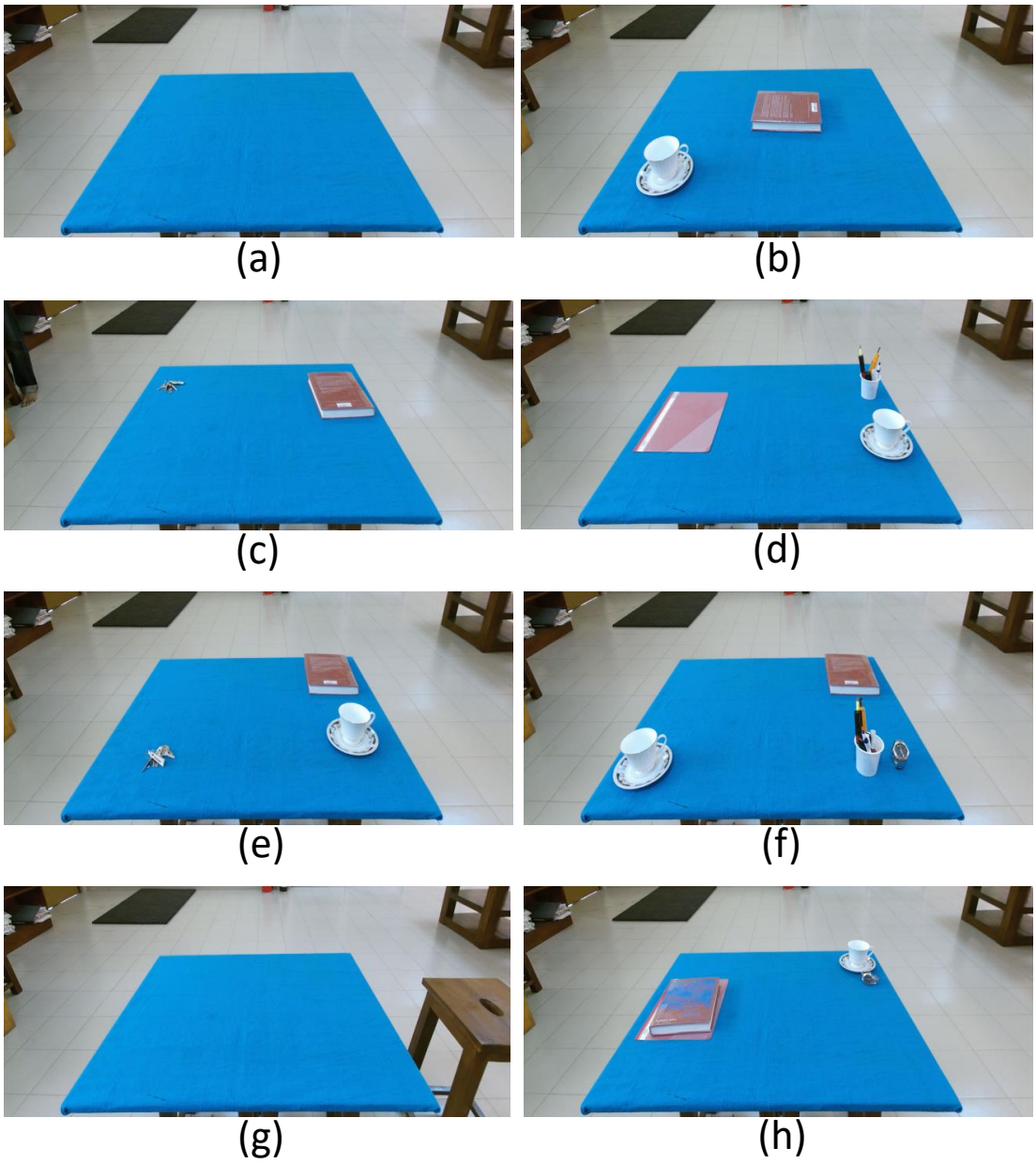


Figure 7.2: Figures (a) to (h) shows the different table settings that were used for the selected set of commands given in the Table 7.1. Setting 1 corresponds to figure (a) and so on.

where the robot's reach to the table was limited. Fig. 7.2 shows the arrangements that are relevant for the commands given in Table.7.1. Fig. 7.2 shows the robot's point of view of the table where figure (a) correspond to the arrangement 1 in the Table 7.1 and so on.

Experiments were conducted with 12 participants whose mean value of age was 39.11 years with a standard deviation of 15 years. Each participant were told to

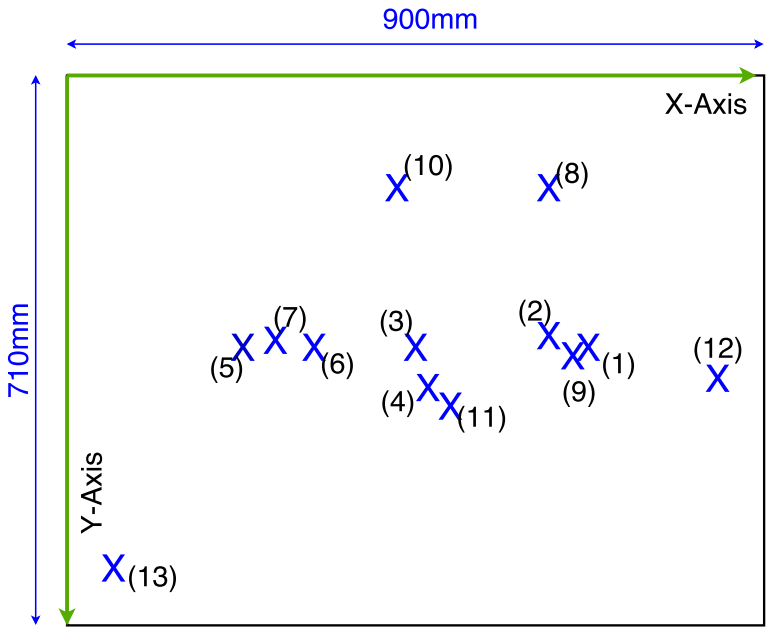


Figure 7.3: Placement of objects on the table for setup 1.

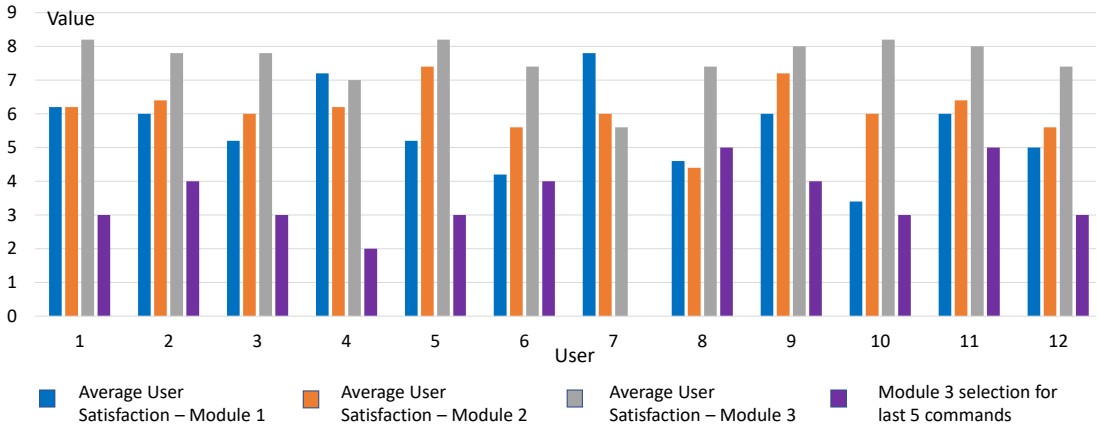


Figure 7.4: User satisfaction for setup 1.

give 20 commands. First 5 commands using module 1, next 5 commands using module 2 and another 5 commands using module 3. For the last 5 commands, users were given the choice to select the preferred module depending on their experience on the previous commands. Table 7.1 presents some of the commands which showed the key features of the system, where it shows the input keywords for both X-Axis and Y-Axis fuzzy inference systems and the membership values for the output. Hand gesture inputs are denoted by "H.G.". All the distance values are given with reference to the X and Y axis marked in Fig. 7.3. The table used for the experiments is 900 mm in length and 710 mm in width. The relevant submodule of the UIUM is selected automatically by the system according to the type of position information available in the user instructions. The selected UIUM submodule number is given under the column "UIUM". The positions where the objects were placed by the robot is marked by "X" in Fig. 7.3. Feedback of the user is given under the tab "User Rating". This denotes the satisfaction of the user regarding the positions of the object placement from an scale 1 to 10 where 1 being least satisfied. The final placement of the object is denoted by the tab "System Output" where the object is placed after considering for safety and edge clearance. In Fig. 7.4, the average user rating for each module is given. Also the number of times where module 3 was selected during the last 5 commands are given in the forth column in each table

7.2.2 Experimental Setup 2

In this section the experiments were carried out in an artificially created room as shown by Fig. 7.5. The room is 3.8m in length and 3.2m in width. Fig. 7.6 shows the map of the room. The table placements are shown by dotted lines. Each user was instructed to issue 20 commands using the 3 submodules as explained in Chapter 5. The users were asked to issue 10 commands for the first module; 5 commands considering the user attention and the remaining 5 commands using the reference frame of the robot . Module 2 and module 3 were subjected for 5 commands each. After each placement, the users were asked to rate the performance of the robot. The rating was from 1 to 10, where 1 is the minimum satisfaction and 10 is the maximum satisfaction. The experiment was conducted with 22 participates, with a average age of 27.5 years and a standard deviation of 7.2 years. The tables used for the experiments were of four basic

Figure 7.5: Shows the room setup that was used for experimental setup 2.

shapes as shown by Fig. 7.7. Even though the system has the ability to adopt to different sizes and shapes of tables, a set of fixed shaped tables were used that has a fixed height. The limiting factor here was the availability of different shapes of tables at disposal. In order to place the object with higher accuracy the system needs to know the exact location of the table with respect to the room walls. So the location of the table was mapped to the robots navigation database. The maps for the robot were created using Mapper 3 software.

In Table 7.2, shape of the relevant table is given by "T.S.", corresponding to the table shapes that are given in Fig. 7.7. The extracted X and Y axes spatial keywords are given in the same table as well. "H.G." refers to the extracted location of the hand gesture, where the values are given in mm with respect to the axes marked in Fig. 7.7. "UIUM" shows the relevant UIUM submodule that was selected for the given command. "VIEM" refers to the extracted user's position and orientation with respect to the axes given in the Fig. 7.6. Here the location is given by meters and the direction of the user is given by degrees. The direction is given with respect to the arrow marked on the Fig. 7.6 where the angles are measured in the clockwise direction. The final output of the system after consideration for the safety concerns of the object is given as "System Output" and are measured in mm with respect to the axes marked in Fig. 7.6.

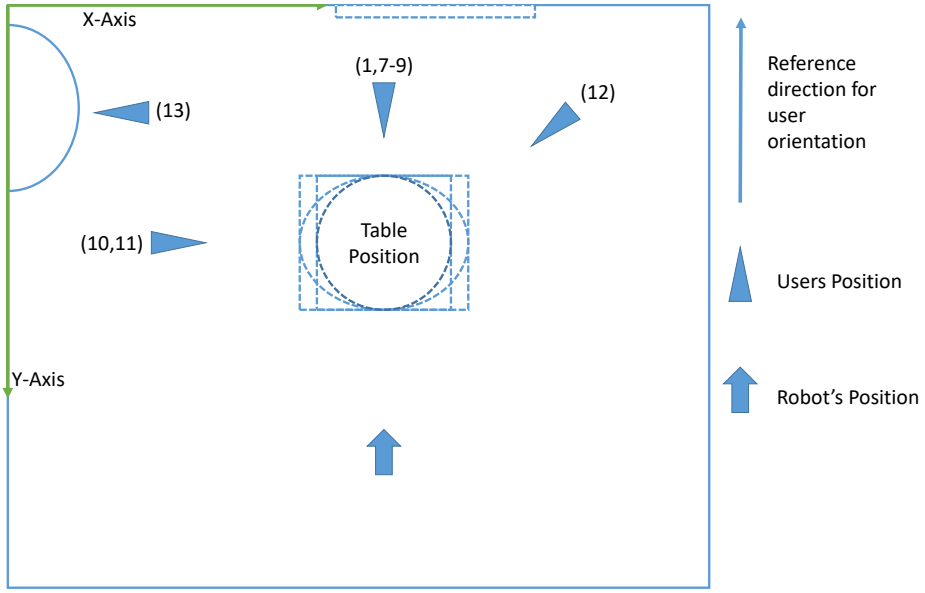


Figure 7.6: Shows the map of the room that was used for experiments in setup 2. The reference direction for the user's orientation is marked here. Orientation and the position of the user is marked using a triangle for the relevant commands in the Table III.

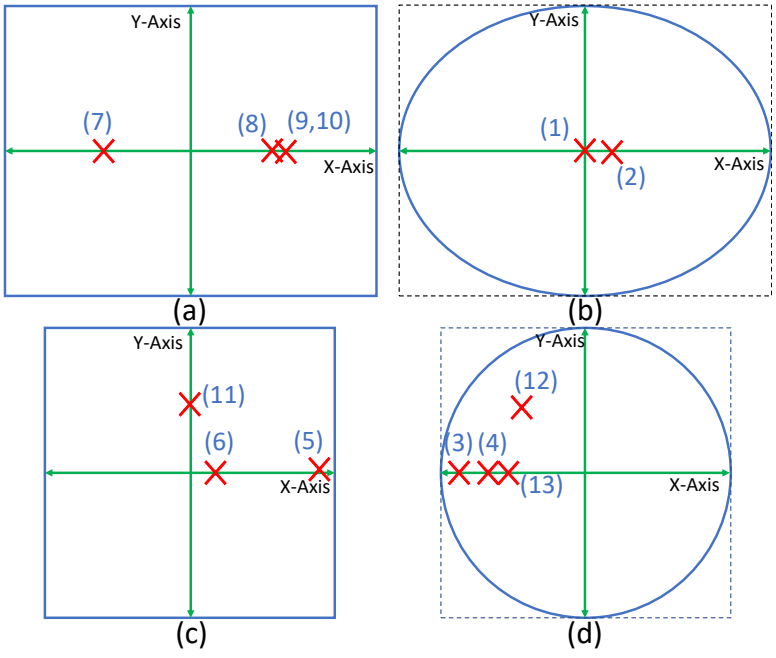


Figure 7.7: Different table shapes that were used during the experimental setup 2 are shown here. Rectangular, Oval, Square and Circular shaped tables are represented by (a),(b),(c) and (d) respectively. The object placements are represented by 'X' on each table.

7.3 System Evaluation

This section analyzes the obtained results in the above case studies where the system performance was tested.

7.3.1 Basic Multimodal System

In module 1 and 3 the system showcases the ability to provide interpreted numerical values required for the object placement task after evaluating the uncertain terms. For example in experimental setup 1, command 1 uses module 1 and keyword "Right" to obtain position $(673, 355)$ on the table.

Some locations on table cannot be reached just using the voice commands. This scenario is visible when commands 5 and 6 in experimental setup 1, are compared. Further commands 1 and 2 in experimental setup 2 are compared, same outcome is visible. Specially the locations which do not tally with the classified areas (e.g. area between two spatial terms) are only reachable by incorporating hand gestures. This further is evident when the Where in command 2 in experimental setup 2, by using hand gestures, the user was able to have a much more flexible placement position that will otherwise not be possible.

Another main concern when using hand gestures is that there can be human and system errors when extracting the hand gesture position information. This can endanger the safety of the object which is to be placed. But when a fuzzy inference system is used this type of errors can be minimized. Command 8 in experimental setup 1 presents an example for this type of case where the hand gesture pointed coordinates are outside of the table. Specially when both voice and hand gesture information are present, even in scenarios where the extracted hand gesture is erroneous can produce a satisfactory outcome following solely the voice commands. In command number 3 in experimental setup 2, the user has given a lower rating due to the fact that the system extracted the hand gesture a little deviated from the expected position. But in the command number 4 in experimental setup 2, the usage of the voice command alongside the hand gesture gave a more satisfying result. When using multimodal approach, the users can obtain object placements with a higher confident. This is due to the fact the

system takes into consideration the voice command as well and not solely depend on the hand gestures.

In most cases, the sole usage of voice commands was not sufficient to achieve the desired location for positioning. But in cases where the item is to be placed in a specific location like the center of the table, using voice commands gave more satisfying results. This is visible when the user tried to place the item on the center of the table in commands 3 and 11 in experimental setup 1. Here usage of voice commands gives a more satisfied outcome.

Commands 5 and 7 in experimental setup 1, show cases that the placement done with multimodal system input gives more satisfaction due to the precision placement. Further when considering the satisfaction of the users given in Fig. 7.4, it is clear that users obtained a more satisfied results when using Module 3 and prefer it to the other two modules. It can be noted that 10 out of 12 users have a higher user rating for module 3 over the other two modules. Furthermore, a clear majority of users selected module 3 for their last 5 commands. This factor is evident in experimental setup 2 as well, the average user rating for the module 1 when the user orientation is taken into consideration was 7.6 while without the user orientation consideration was 4.2 which was considerably lower. From the three submodules of the UIUM, module 3 obtained the highest average user rating of 8.4.

7.3.2 Effect of Dynamic Space Constraints

Restricted Reachability

With reference to the obtained results it can be shown that when subjected to the reachability restrictions, system is capable of changing the position of item placement accordingly. This can be verified by comparing commands 1 and 2 in experimental setup 1. In both cases system is given the same voice command but the final position of placement has been shifted 51mm along the negative X direction and 15mm along negative Y direction after modifying for safety of the object as the item cannot be placed at the location of the given system output due to limitations in reachability.

Objects on the Table

Interpreted quantitative values for object placement vary accordingly when the free space on the table gets limited. This can be verified using command 3 and 4 in experimental setup 1, which used module 1. The only change between the two scenarios is the limitation in free space due to the object that is present in table setting 2. But the resulting object position has been shifted along the negative X axis and positive Y axis.

7.3.3 Special Attention

Concerns for User Location

The effect of the user's orientation is considered when the system is supplied with only voice based positioning information. When the system receives hand gestures based positioning information, the system decides the reference frame of the object placement based on the hand gesture. The consideration for user location and orientation is presented in experimental set-up 2. In the command 7, the user asks the robot to place the item on the left side of the table while effect of the user's orientation is not taken into consideration. This results in a low user rating for the placement because the user refers the left side of the table differently from the robot's left side. For a similar scenario in command 9 the user's orientation is taken into consideration. This results in a much more satisfying outcome. Further in command 8 where the system receives a hand gesture alongside the voice command the reference frame is decided by the system using the hand gesture. Here the user gives a higher rating for the placement. So it can be highlighted that using multi-modal interaction, the ambiguity in the reference frame has been diminished.

In command 9, the user is facing the table with 90 degree orientation. If the user's orientation were to be considered when placing the object, the area terms has to be categorised as shown by Fig. 6.4(f). So the system ignores the user's orientation and place the object using the robots reference frame. But in a similar scenario when a square shaped table is used, the user's orientation is taken into consideration. In command 12 a circular table is used while the user is paying

attention to the table. So here the placement is performed using the reference frame of the user. Command 13 presents an example to a scenario where the user is not paying attention to the table. So the object placement is performed using the robot's reference frame.

Safety Distance

Commands 12 and 13 in experimental setup 1, are example cases for concerns about the safety of the objects. Even though the positions obtained by the system are inside the table boundaries the objects were kept taking into consideration the safety factors.

7.3.4 System Limitations

An unsatisfied user rating was detected on command 9 in experimental setup 1, which may be due to the height of the objects. As the robot uses its own point of view, the space behind the obstacle would seem limited than it really is. This could lead to misinterpretations in object placement. Specially in similar scenarios, users felt that objects has to be placed a little bit away from the obstacles that are taller. To overcome this limitation 3D depth analysis of the table surface can be incorporated. One other limitation was that the robot had to start from the same start position each time because the coordinates of the Kinect sensor's position had to be known in order to obtain hand gesture pointed location.

Table 7.1: Experiment Results for Setup 1

X-Axis

Command No.	UIUM	Table Setting	Commnd Keywords	H.G. (mm)	Membership Values									System Output(mm)	User Rating
					VVL	VL	L	LL	M	LR	R	VR	VVR		
1	1	1	Right	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	673	7
2	1	7	Right	-	-	0.000	0.250	-	0.496	-	0.692	1.031	-	622	8
3	1	1	Middle	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	450	9
4	1	2	Middle	-	-	-0.003	0.255	-	0.495	-	0.755	1.000	-	446	8
5	1	1	Left	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	227	5
6	2	1	-	315	-	0.000	0.250	-	0.500	-	0.750	1.000	-	319	8
7	3	1	Left	307	0.000	0.125	0.250	0.375	0.500	0.625	0.750	0.875	1.000	269	9
8	2	3	-	930	-	0.000	0.252	-	0.500	-	0.730	1.014	-	622	8
9	3	4	Left Edge	895	-0.003	0.126	0.252	0.375	0.500	0.624	0.744	0.881	1.003	653	9
10	3	5	Middle	458	-0.004	0.127	0.253	0.375	0.500	0.625	0.745	0.877	1.003	426	7
11	2	6	-	451	-	-0.015	0.273	-	0.504	-	0.737	1.006	-	455	4
12	3	8	Right Edge	881	-0.007	0.124	0.264	0.376	0.500	0.625	0.746	0.879	1.001	801	8
13	2	2	-	5	-	-0.004	0.257	-	0.493	-	0.757	1.000	-	60	8

Y-Axis

Command No.	UIUM	Table Setting	Commnd Keywords	H.G. (mm)	Membership Values									System Output(mm)	User Rating
					VVL	VL	L	LL	M	LR	R	VR	VVR		
1	1	1	Middle	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	355	7
2	1	7	Middle	-	-	0.000	0.250	-	1.500	-	0.750	1.000	-	340	8
3	1	1	Middle	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	355	9
4	1	2	Middle	-	-	0.000	0.249	-	0.501	-	0.746	1.003	-	407	8
5	1	1	Middle	-	-	0.000	0.250	-	0.500	-	0.750	1.000	-	355	5
6	2	1	-	341	-	0.000	0.250	-	0.500	-	0.750	1.000	-	355	8
7	3	1	Middle	343	0.000	0.125	0.250	0.375	0.500	0.625	0.750	0.875	1.000	346	9
8	2	3	-	50	-	0.000	0.252	-	0.501	-	0.730	1.014	-	149	8
9	3	4	Middle	340	-0.001	0.122	0.250	0.378	0.500	0.624	0.746	0.876	1.003	365	9
10	3	5	Front	378	0.000	0.124	0.250	0.372	0.500	0.628	0.747	0.878	1.003	149	7
11	2	6	-	381	-	-0.015	0.273	-	0.502	-	0.737	1.006	-	391	4
12	3	8	Front Edge	17	-0.003	0.120	0.253	0.377	0.506	0.626	0.750	0.875	1.001	355	8
13	2	2	-	573	-	-0.004	0.257	-	0.499	-	0.757	1.000	-	640	8

Table 7.2: Experiment Results for Setup 2

Cmd. No	T.S.	User Command	X-Axis Keywords	Y-Axis Keywords	H.G. (mm)	UIUM	VIEM (m,m,deg.)	System Output (mm)	User Rating
1	b	Keep the object on the center of the table	Middle	Middle	-	1	(0.5, 1.9, 180)	(5, -2)	10
2	b	Keep the object on the center	Middle	Middle	(81, 22)	3	-	(43,8)	9
3	d	-	-	-	(-349, 17)	2	-	(-305, 9)	6
4	d	Place the object on the left of the table	Right	Middle	(-345, -3)	3	-	(-282, 0)	9
5	c	-	-	-	(362, 19)	2	-	(306, 8)	8
6	c	Keep the object on the left of the table	Left	Middle	(-35, 8)	3	-	(62, 2)	8
7	a	Place it on the left	Left	Middle	-	1	(0.4, 1.9, 180)	(-218, 0)	3
8	a	Place it on the left of the table	Left	Middle	(149, -12)	1	(0.6, 1.7, 180)	(186, 0)	9
9	a	Keep the object on the left	Left	Middle	-	3	(0.4, 1.8, 180)	(215, 0)	8
10	a	Place the object on the left side of the table	Left	Middle	-	1	(1.1, 1, 90)	(214, 3)	7
11	c	Keep it on the left of the table	Left	Middle	-	1	(1, 1.3, 90)	(0, 167)	8
12	d	Can you place it on the right of the table	Right	Middle	-	1	(2.8, 0.7, 135)	(-117, 115)	9
13	d	Keep the object on the left side of the table	Left	Middle	-	1	(0.6, 0.6, 270)	(-165, 0)	9

Chapter 8

CONCLUSIONS

A multi-modal interactive system has been proposed that can place objects effectively on the desired positions of a table using a fuzzy based approach. A modular Uncertain Information Understanding Module (UIUM) is introduced which is able to handle the placement of objects on a surface using voice and/or gesture based positioning information. The system can interpret uncertain user commands using hand gesture positioning information and spatial factors like distribution of free space available and the reachability of the robot to certain areas on the table. Safety concerns for object placement has also been incorporated making the system more reliable. The deployed voice command interface uses no strict grammar rule. This system improves the human-like object placement capability of the robot improving the human robot interaction.

The summary of main contributions are listed as follows,

- Multimodal approach for object placement
  - This research introduces a novel approach for object placement using multimodal inputs of both voice based commands and hand gestures. The previous system has used multimodal approach to draw attention to items or to select items from a group of items.
  - A modular UIUM has been introduced that can understand the spatial terminology in voice commands using hand gestures.
  - The system has the capability to act for both unimodal and multimodal commands depending on the availability of information.
- Six human studies have been carried out in the real world scenario
  - This research contains six human studies to understand how a human

user would behave in a object placement situation and to understand what factors may affect the placement of the objects. All the studies have been carried out in real world scenarios. The results has been analysed and discussed.

- Understanding of uncertain spatial terms in voice commands when manipulating objects
  - The proposed system uses fuzzy based approach to understand the uncertain spatial terms. The system uses two fuzzy system for X and Y axes to determine the location of placement.
- Spatial factors as well as user aspects has been considered when understanding the user commands and performing the placement
  - Objects that are on the table and the reachability of the user are taken into consideration when placing the objects.
  - The system has the capability to adapt to different shapes and sizes of tables.
  - The user's location and orientation is taken into account when deciding the reference frame for understanding the user commands.

The main conclusions of the research is as follows,

- Multimodal approach with both hand gestures and voice commands produce a better user satisfaction. In Fig.8.1 shows the average user satisfaction for each module out of 10.
- Ability of the system to adjust for the restrictions in space is important for robotic system with human like behavioral capabilities. Two of such important limitations include objects that on the table and obstacle that are around the table which may limit the navigation of the robot when reaching out to certain areas on the table.
- It is important for the robotic system to be able to perform the placement taking into account different shapes of table. From the conducted studies 71.4% of user satisfaction was detected for the multi-modal system without

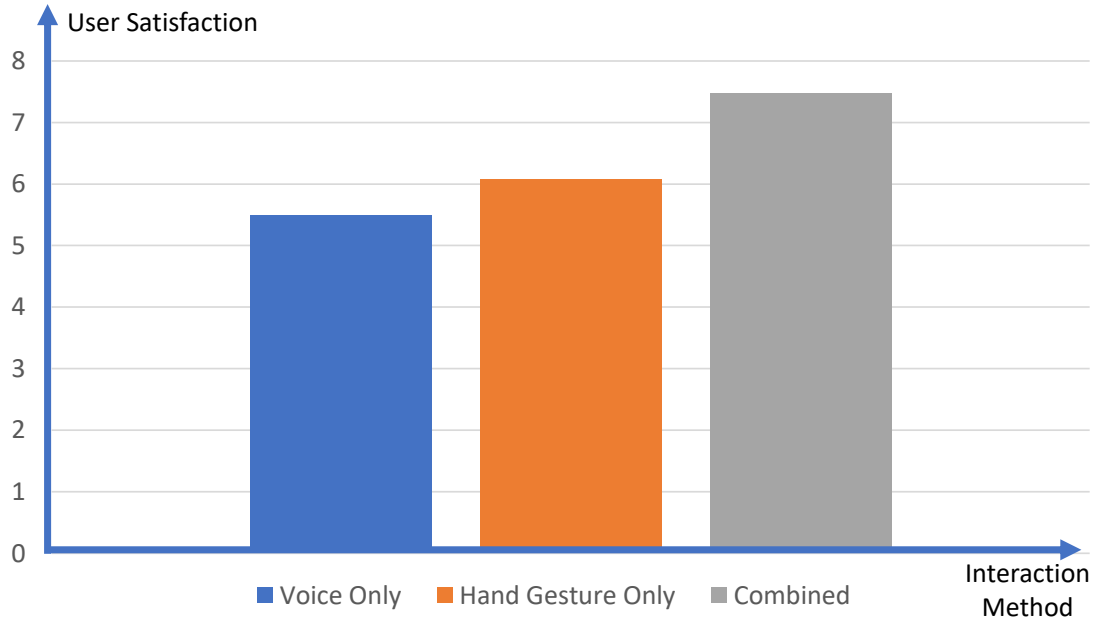


Figure 8.1: User satisfaction in unimodal and multimodal systems.

consideration for obstacles on the table and restrictions in reachability. The user satisfaction was improved to %84.61 when these restrictions were taken into consideration.

- Having the ability to consider the robot's position and orientation with respect to the table is important when deciding the reference frame of the user commands because the reference frame changes with the orientation and the position of the robot itself.

The effect from the height of the table was not considered during this research, even though the effect of the height of both the user and the table can have a significant effect on the final placement of the object. Furthermore, the effect of the handedness of the user was not implemented in the system even though the effect was analysed during the studies. These effects can be considered in the future developments of the system. The objects that were used in the studies as wells as in the experiments were not significantly larger in height. But the visibility of the area behind an object can determine the space that is available for manipulating the objects. In order to improve the capabilities of the system these effects acan be improved in the future implementations.

LIST OF PUBLICATIONS

---

[1] M. A. Viraj J. Muthugala, P. H. D. Arjuna S. Srimal, and A. G. Buddhika P. Jayasekara. Enhancing interpretation of ambiguous voice instructions based on environment and users intention for improved human friendly robot navigation. *Applied Sciences (Accepted)*, 2017.

[2] H. P. Chapa Sirithunge, P. H. D. Arjuna S. Srimal, and A. G. Buddhika P. Jayasekara. Identification of friendly and deictic gestures in a sequence using upper body skeletal information. In *TENCON IEEE Region 10 Conference (Submitted)*. IEEE, 2017.

[3] P. H. D. Arjuna S. Srimal and A. G. Buddhika P. Jayasekara. A multi-modal approach for enhancing object placement. In *National Conf. on Technology Management*. IEEE, 2017.

[4] P. H. D. Arjuna S. Srimal, M. A. Viraj J. Muthugala, and A. G. Buddhika P. Jayasekara. Deictic gesture enhanced fuzzy spatial relation grounding in natural language. In *Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on (Presented)*. IEEE, 2017.

[5] P. H. D. Arjuna S. Srimal, M. A. Viraj J. Muthugala, and A. G. Buddhika P. Jayasekara. Identifying spatial terminology and boundaries for human robot interaction: A human study. In *Moratuwa Engineering Research Conference (MERCon) (Presented)*. IEEE, 2017.

REFERENCES

[1] Ineternational Federation of Robotics, "Excecutive summery service robots 2016," *World Robotics - Service Robots*, 2016.

[2] W. H. Organization, *World report on ageing and health.* World Health Organization, 2015.

[3] S. Webb, "Our aging world: Un predicts there will be more pensioners than children by 2050," http://www.dailymail.co.uk/news/article-2211191/Our-ageing-world-UN-predicts-pensioners-children-2050.html, [Online; accessed 05-Aug-2017].

[4] J. Forlizzi and C. DiSalvo, "Service robots in the domestic environment: a study of the roomba vacuum in the home," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction.* ACM, 2006, pp. 258–265.

[5] H. Osawa, J. Orszulak, K. M. Godfrey, M. Imai, and J. F. Coughlin, "Improving voice interaction for older people using an attachable gesture robot," in *19th Int. Symp. in Robot and Human Interactive Communication.* IEEE, 2010, pp. 179–184.

[6] M. Eldon, D. Whitney, and S. Tellex, "Interpreting multimodal referring expressions in real time," in *Int. Conf. on Robotics and Automation*, 2016.

[7] T. Borangiu, "Advances in robot design and intelligent control," *Switzerland: Springer Int. Publishing*, 2014.

[8] P. Tsarouchi, S. Makris, and G. Chryssolouris, "Human-robot interaction review and challenges on task planning and programming," *Int. Journal of Computer Integrated Manufacturing*, vol. 29, no. 8, pp. 916–931, 2016.

[9] L. Fortunati, "Moving robots from industrial sectors to domestic spheres: A foreword," in *Toward Robotic Socially Believable Behaving Systems-Volume II*. Springer, 2016, pp. 1–3.

[10] H. Melkas, L. Hennala, S. Pekkarinen, and V. Kyrki, "Human impact assessment of robot implementation in finnish elderly care," in *International Conference on Serviceology, Tokyo, Japan*, 2016, pp. 6–8.

[11] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent asimo: System overview and integration," in *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, vol. 3. IEEE, 2002, pp. 2478–2483.

[12] J. Liu, Y. Luo, and Z. Ju, "An interactive astronaut-robot system with gesture control," *Computational intelligence and neuroscience*, vol. 2016, 2016.

[13] C. BATES, "How do you like your pancakes? robot cooks breakfast at exhibit," http://www.dailymail.co.uk/sciencetech/article-1089904/How-like-pancakes-Robot-cooks-breakfast-exhibit.html, [Online; accessed 05-Aug-2017].

[14] G. Gemignani, M. Veloso, and D. Nardi, "Language-based sensing descriptors for robot object grounding," in *Robot Soccer World Cup*. Springer, 2015, pp. 3–15.

[15] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "Enhancing human-robot interaction by interpreting uncertain information in navigational commands based on experience and environment," in *2016 IEEE Int. Conf. Robotics and Automation (ICRA)*, 2016, pp. 2915–2921.

[16] I. de Kok, J. Hough, D. Schlangen, and S. Kopp, "Deictic gestures in coaching interactions," in *Proceedings of the Workshop on Multimodal Analyses enabling Artificial Agents in Human-Machine Interaction*. ACM, 2016, pp. 10–14.

[17] P. Yan, B. He, L. Zhang, and J. Zhang, "Task execution based-on human-robot dialogue and deictic gestures," in *Robotics and Biomimetics (ROBIO), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1918–1923.

[18] S. Sathayanarayana, R. Kumar Satzoda, A. Carini, M. Lee, L. Salamanca, J. Reilly, D. Forster, M. Bartlett, and G. Littlewort, "Towards automated understanding of student-tutor interactions using visual deictic gestures," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 474–481.

[19] H. Cochet and J. Vauclair, "Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences," *Laterality: Asymmetries of Body, Brain and Cognition*, vol. 19, no. 3, pp. 278–301, 2014.

[20] P. Yuksel and P. J. Brooks, "Encouraging usage of an endangered ancestral language: A supportive role for caregivers? deictic gestures," *First Language*, 2017.

[21] K. Ouwehand, "Effects of observing and producing deictic gestures on memory and learning in different age groups," *ico*, 2016.

[22] J. Sattar and G. Dudek, "Underwater human-robot interaction via biological motion identification." in *Robotics: Science and Systems*, 2009.

[23] H. J. Fariman, H. J. Alyamani, M. Kavakli, and L. Hamey, "Designing a user-defined gesture vocabulary for an in-vehicle climate control system," in *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. ACM, 2016, pp. 391–395.

[24] C. Matuszek, L. Bo, L. Zettlemoyer, and D. Fox, "Learning from unscripted deictic gesture and language for human-robot interactions." in *AAAI*, 2014, pp. 2556–2563.

[25] F. Cruz, G. I. Parisi, J. Twiefel, and S. Wermter, "Multi-modal interdration of dynamic audiovisual patterns for an interactive reinforcment lerning scenario," in *RO-MAN, 2016 IEEE*. IEEE, 2016.

[26] A. B. P. Jayasekara, K. Watanabe, K. Kiguchi, and K. Izumi, "Interpretation of fuzzy voice commands for robots based on vocal cues guided by user's willingness," in *2010 IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2010, pp. 778–783.

[27] S. Schiffer, A. Ferrein, and G. Lakemeyer, "Fuzzy representations and control for domestic service robots in golog," in *Intelligent Robotics and Applications*. Springer, 2011, pp. 241–250.

[28] A. B. P. Jayasekara, K. Watanabe, and K. Izumi, "Understanding user commands by evaluating fuzzy linguistic information based on visual attention," *Artificial Life and Robotics*, vol. 14, no. 1, pp. 48–52, 2009.

[29] M. A. V. J. Muthugala and A. G. B. P. Jayasekara, "Interpreting fuzzy linguistic information in user commands by analyzing movement restrictions in the surrounding environment," in *2015 Moratuwa Engineering Research Conference (MERCon)*, April 2015, pp. 124–129.

[30] K. Nickel and R. Stiefelhagen, "Pointing gesture recognition based on 3d-tracking of face, hands and head orientation," in *Proceedings of the 5th int. conf. on Multimodal interfaces*. ACM, 2003, pp. 140–146.

[31] L. Perlmutter, E. Kernfeld, and M. Cakmak, "Situated language understanding with human-like and visualization-based transparency," in *Robotics: Science and Systems*, 2016.

[32] J. Stückler, D. Droeschel, K. Gräve, D. Holz, J. Kläß, M. Schreiber, R. Steffens, and S. Behnke, "Towards robust mobility, flexible object manipulation, and intuitive multimodal interaction for domestic service robots," in *Robot Soccer World Cup*. Springer, 2011, pp. 51–62.

[33] S. Lemaignan, R. Ros, E. A. Sisbot, R. Alami, and M. Beetz, "Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction," *Int. Journal of Social Robotics*, vol. 4, no. 2, pp. 181–199, 2012.

[34] J. Tan, Z. Ju, and H. Liu, "Grounding spatial relations in natural language by fuzzy representation for human-robot interaction," in *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE Int. Conf. on*. IEEE, 2014, pp. 1743–1750.

[35] K. Charalampous, I. Kostavelis, and A. Gasteratos, "Recent trends in social aware robot navigation: A survey," *Robotics and Autonomous Systems*, 2017.

[36] C. McGinn, A. Sena, and K. Kelly, "Controlling robots in the home: Factors that affect the performance of novice robot operators," *Applied Ergonomics*, vol. 65, pp. 23–32, 2017.

[37] T. Nowack, S. Lutherdt, S. Jehring, Y. Xiong, S. Wenzel, and P. Kurtz, "Detecting deictic gestures for control of mobile robots," in *Advances in Human Factors in Robots and Unmanned Systems*. Springer, 2017, pp. 87–96.

[38] P. Kondaxakis, K. Gulzar, and V. Kyrki, "Temporal arm tracking and probabilistic pointed object selection for robot to robot interaction using deictic gestures," in *Humanoid Robots (Humanoids), 2016 IEEE-RAS 16th International Conference on*. IEEE, 2016, pp. 186–193.

[39] Microsoft, "Microsoft kinect documentation," https://msdn.microsoft.com/en-us/library/microsoft.kinect.aspx, [Online; accessed 05-Aug-2017].

[40] M. V. J. Muthugala and A. B. P. Jayasekara, "Mirob: An intelligent service robot that learns from interactive discussions while handling uncertain information in user instructions," in *Moratuwa Engineering Research Conference (MERCon), 2016*. IEEE, 2016, pp. 397–402.