Framework for Adaptive Human-Robot Interaction Initiation for Domestic Environments

Hewa Pelendage Chapa Sirithunge

(168060C)

Degree of Doctor of Philosophy

Department of Electrical Engineering

University of Moratuwa
Sri Lanka

December 2020

Framework for Adaptive Human-Robot Interaction Initiation for Domestic Environments

Hewa Pelendage Chapa Sirithunge

(168060C)

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

Department of Electrical Engineering

University of Moratuwa
Sri Lanka

December 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate 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 conducted research for the PhD thesis under my supervision.

Signature of the Supervisor(s): Date:

Prof. A.G. Buddhika P. Jayasekara

DEDICATION

To my beloved family and all the people who strive hard to make the world better than today

Abstract

Intelligent robot companions contribute significantly in improving living standards in the modern society. Therefore human-like decision making skills and perception are sought after during the design of such robots. On the one hand, such features enable a robot to be easily handled by a non-expert human user. On the other hand, the robot will have the capability of dealing with humans without causing any disturbance by the its behavior. Mimicking human emotional intelligence is one of the best and reasonable ways of laying the foundation for such an emotional intelligence in robots. As robots are widely deployed in social environments today, perception of the situation or the intentions of a user prior to an interaction is required in order to be proactive. Proactive robots are required to understand what is communicated by the human body language prior to approaching a human. Social constraints in an interaction could be demolished by this assessment in this regard.

This thesis addresses the problem of perceiving nonverbals in human behavior and fusing human-environment semantic representations with a robot's cognition before interacting with humans. The novelty lies in laying the background to relate nonverbal human behavior and the features of the environment to generate context-aware interactive responses during robot-initiated interaction. That informs the robot about its environment. Toward this end, we introduce novel methods of perceiving human nonverbals and spatial factors in the environment which make up a context in which we integrated that knowledge to determine appropriate responses from a robot to assist its user. Visual information acquired by a vision sensor was analyzed, and the level of emotional engagement demanded by the user's nonverbals was evaluated, before a robot initiates an interaction. After such an analysis, a robot's conversational and proxemic behavior was adjusted to maintain an empathetic relationship between the user and the robot. Our algorithms efficiently sustained the empathy between user and robot so that the interaction resembles human-human interaction to a larger extent. To assist the main problem, we formulated novel methods to recognize human nonverbals such as postures, gestures, hand poses, psychophysiological behavior of humans and human activities, and decode the emotional hints displayed to the outside world. In support of this work, we conducted a series of human studies to explore the patterns in human behavior which could be perceived by a proactive robot using its cognitive capabilities.

We introduce separate systems which can decode the sentiments of humans using observable cues based on accepted social norms. We detail the meanings of human nonverbals by observing human behavior over time and evaluating the context for any patterns in behavior. Ambiguities in human behavior and random, meaningless behaviors could be omitted through this approach. This approach further omits the negative effect of human responses that can be faked, such as facial expressions and words. Finally we introduce an adaptive approach towards robot-initiated human-robot interaction by letting a robot observe a context and generate responses while changing its responses continuously as human behavior changes. We first developed algorithms based on a limited number of observable human cues and decoded their sentiments based on modern psycho-physiological interpretations of human behavior. Next, we expanded such approaches towards multiple observable human cues. Finally we integrated observations from the human and the environment which create the context during HRI (Human-Robot Interaction). Hence we integrated all the recognition approaches to perceive a complete scenario which comprises the user, robot and the environment.

Upon unimodal systems to identify these features independently, we propose a multi modal approach to evaluate above features together to understand a scenario. Through this approach, we took an effort to make proactive behavior of a social robot more instinctive, ethical and socially acceptable or simply, humanlike. We evaluate this approach by means of physical experiments in simulated social and domestic environments and demonstrate its performance in appropriate occasions as determined by a robot according to the formulated criteria of perceiving a context.

Keywords- **Nonverbal behavior, Interaction initiation, Proactive robots, Social Human-robot Interaction**

ACKNOWLEDGMENT

It is with great pleasure that I acknowledge the support and contribution of all the mentors and colleagues who assisted me in successfully completing my PhD thesis.

First I would like to express my sincere gratitude to my advisor Dr. Buddhika Jayasekara. It was a privilege to work under his guidance and I thank him for his support, encouragement and patience. Without his tireless efforts, this thesis would never have been accomplished. Also I would like to extend my gratitude to my progress review panel, Prof. Ruwan Gopura and Dr. Chandima Pathirana for their insightful comments and helpful suggestions. I specially acknowledge the efforts put by the thesis examination panel.

Furthermore I would like to thank the following people who helped me undertake this research.

- Arjuna Srimal, Sajila Wickramaratne and the members of Intelligent Service Robotics Group (ISRG), especially Gayashan Porawagamage, Viraj Muthugala, Sahan Kodikara, and Ravindu Bandara, Department of Electrical Engineering, University of Moratuwa.
- My parents and Nikolas Dahn for their kind support and patience through out the journey.
- Dear teachers and my dear friends who set me off on the road to this PhD a long time ago should accept my sincere gratitude.
- The staff of the Robotics and Automation Laboratory in the Department of Electrical Engineering

-People who were generous with their time and effort in participating in experiments related to research.

- The staff and my friends in the department of Electrical Engineering for their immense support.
- My family, my friends, numerous to mention, and everybody who were with me in making this endeavor a success.

This work was partially supported by the University of Moratuwa Senate Research Grants SRC/CAP/17/03 and SRC/LT/2018/20. Furthermore, some of the publications originated from this thesis were partially supported by travel grants awarded by the National Science Foundation of Sri Lanka, IEEE Robotics and Automation Society and European Association for Artificial Intelligence.

TABLE OF CONTENTS

Declaration **i**

Dedication **ii**

Abstract **iii**

Acknowledgments **v**

Table of Contents **xiii**

List of Figures **xx**

List of Tables **1**

1 Introduction **1**

1.1 Development of assistive robots 1

1.2 Building Attentive Robots 5

1.2.1 Study of human behavior 5

1.2.2 "Meet, Perceive and Act" 7

1.2.3 Multimodal perception 7

1.3 Problem Statement 8
1.4 Thesis Contributions 8
1.4.1 Identifying observable human behavior 8
1.4.2 Recognizing components embodied in a context 9
1.4.3 Decoding nonverbals to interpret user situation 10
1.4.4 Adapting to a situation 11
1.5 Thesis Overview 11

2 Literature Review **13**

2.1 Nonverbal Interaction during HRI 13
2.1.1 Features to observe 14
2.1.2 Earlier Approaches 15
2.1.3 Modern Developments 17
2.1.4 Systems with Adaptive Perception of Situation based on Human Cues 19
2.2 Outlook 26

3 System Design **28**

3.1 Extracting information 28
3.2 Robot Platform 28
3.3 Overview of the system 29

4 Study of Human Behavior **33**

4.1 Study 1: Human Interest towards Robot Initiated Human-Robot Interaction . . . 33

4.1.1 Monitoring Human Behavior . . . 33

4.1.2 Results of the experiment . . . 35

4.1.3 Observations and Discussion . . . 38

4.2 Study 2: Conversational Preferences in Social Human-Robot Interaction . . . 41

4.2.1 Setting . . . 42

4.2.2 Results and discussion . . . 46

4.2.3 Observations and Discussion . . . 49

4.2.4 Conclusions and Implications . . . 52

4.3 Study 3: Human Gaze Behavior . . . 54

4.3.1 Understanding Human Behavior: Evaluate Attention . . . 54

4.3.2 Time scales of human actions . . . 54

4.3.3 Monitoring Gaze Behavior . . . 55

4.3.4 Results of the experiments and Discussion . . . 57

4.3.5 Finding Answers to previous questions . . . 64

5 Recognition of nonverbals **66**

5.1 Psychophysiological Aspects of Human Behavior . . . 66

5.1.1 Simple Human Activities 66
5.1.2 Criteria used to analyze activity space 67
5.1.3 Analyzing Extracted Information 69
5.1.4 Decision Evaluation 70
5.2 Hand pose 75
5.2.1 Recognizing hand pose 75
5.2.2 Experiment and results 79
5.3 Pose recognition and robot-initiated interaction 81
5.3.1 Interpretation of Fuzzy Information 85
5.4 Interpreting movements of the body 90
5.4.1 Measuring Behavioral Responses 90
5.4.2 Observable Nonverbal Human Behavior 91
5.4.3 Decision Making Criteria 93
5.4.4 Monitoring Human Behavior 95
5.4.5 Experiment 1 98
5.4.6 Experiment 2 98
5.4.7 User Responses Towards Robot 99
5.5 Recognition of Hand and Body Pose 103
5.5.1 Changes in Behavior 106
5.5.2 Mechanism used to differentiate postures 108

5.5.3 Behavior Evaluation Model 113

5.5.4 Posture based Interaction Decision Making 114

5.5.5 Valence of a situation 115

5.5.6 Decision-making in dynamic user behavior 117

5.5.7 Experiment and Results 118

5.5.8 Results and Discussion 122

5.6 Recognition of complex human activities 127

5.6.1 Robot, user and the context 127

5.6.2 Features of an activity 131

5.6.3 Recognition of activities 132

5.6.4 Learning new activities 135

5.6.5 Experiment and Results 137

5.6.6 Setting 137

5.6.7 Results 139

5.6.8 Conclusion 144

6 Decoding nonverbals and Modelling User Behavior 145

6.1 Identification of Gaze 145

6.1.1 Determination of Gaze direction and Gaze time 146

6.1.2 Return of Gaze 146

6.2 Identification of Friendly Gestures 147

6.3 Identification of Changes in Pose 148

6.3.1 Evaluation of attention 150

6.3.2 Interpretation of Fuzzy Parameters 152

6.3.3 Decision Making 152

6.3.4 Experiment and Results 153

6.3.5 Conclusions 156

7 Determining Robot's Proactive Behaviors: Proxemics 157

7.1 Probabilistic analysis 158

7.2 Decision making criteria 159

7.3 Results and Discussion 162

7.3.1 Observations and discussion 163

7.3.2 Conclusions 167

8 Modeling a context: Autoregressive Model 168

8.1 Level of Interest 169

8.2 AR model to calculate LOI 170

8.3 Decision making criteria 172

8.4 Results and Discussion 173

8.5 Observations and Discussion 176

8.6 Conclusions 180

9 A Social Robot with Experience and Learning 181

9.0.1 Semantic Mapping of the features 183

9.0.2 Robot behavior 184

9.0.3 Selection of participants and creating training data 184

9.1 Conclusions 187

10 Conclusions 189

10.1 Implications for Theory 193

10.2 Implications for Design 195

10.3 Considerations for future 196

References 201

LIST OF FIGURES

1.1 An example of a cooking robot 2

1.2 An airport cleaning robot 2

1.3 Example scenarios where the required level of situation-awareness differs. 4

1.4 An example domestic environment is shown. 10

1.5 Different types of settings and components associated with HRI . 11

3.1 Hardware setup used in experiments. 29

3.2 Overall system 31

4.1 Poses used as behavioral changes 36

4.2 An example scenario during the experiment 36

4.3 A comparison of the types of behavioral changes observed during a relaxing task 38

4.4 A comparison of the types of behavioral changes observed while user was engaged 38

4.5 A typical conversation between the robot and its user 44

4.6 An example scenario during the experiment 47

4.7 A situation in which the user was having a drink 47

4.8 A stacked graph drawn for the comparison of conversational preferences with the two conditions; when the user is alone and when surrounded by few people 51

4.9 conversational preferences during the selected tasks while the domestic area and people in the surrounding were kept constant . 53

4.10 Phases in a typical encounter prior to an interaction 55

4.11 Two scenarios during the experiment 57

4.12 Existence or nonexistence of gaze during HHI is plotted against each user 58

4.13 The percentage (%) of encounters with any kind of gaze plotted againt each occasion 59

4.14 The average duration of existed gaze was plotted against the delay 59

4.15 The average delay of gaze was plotted against the occasion 60

4.16 The delays of encounters where an averted gaze was present . . . 60

4.17 The delays of encounters where a continued gaze was present . . . 61

4.18 The delays of encounters where an averted gaze was present . . . 62

4.19 Responses observed from users during HHI are plotted against each individual 63

5.1 Arrangement of activity zones near a person 67

5.2 Dimensions and joints used to separate zones 68

5.3 Approach of QASA 70

5.4 Various occasions tested with the 2 systems:ASA and OASA . . . 72

5.5 Boxplot from the results obtained for ASA and QASA 74

5.6 Criteria used to determine length, width and height of the mesh . 76

5.7 The 3D mesh created out of the 'activity zones' established around a human 77

5.8 Utilization of zones during a certain gesture 77

5.9 The set of gestures used to recognize using the proposed approach 78

5.10 Depth images corresponding to the gestures from a-k 78

5.11 Types of Standing poses used in the experiment 82

5.12 Types of Sitting poses used in the experiment 82

5.13 Routes followed by robot during various situations. 85

5.14 Input and output membership functions used to find the interaction demanding 86

5.15 An occasion with 'high' interaction demanding 88

5.16 An occasion with 'low' interaction demanding 89

5.17 Skeletals extracted from a standing and a seated person 92

5.18 Nonverbal behavioral responses observed during the experiment . 93

5.19 The decision grid 95

5.20 Two occasions from the experiment 99

5.21 Two occasions in which adaptive approach behavior was implemented 100

5.22 A comparison between the average feedback scores received for direct approach and adaptive behavior 100

5.23 Two occasions from the experiment 102

5.24 Overview of the system 105

5.25 A set of body postures encountered daily 107

5.26 A set of arm postures encountered daily 107

5.27 Body angles derived from the limbic vectors of a person with stretched hands 109

5.28 Limbic vectors of a seated person 109

5.29 Body angles derived from limb vectors of a seated person as in Fig. 5.28 112

5.30 A semantic map illustrating the components that make valence; hand and body poses 115

5.31 Marginal values chosen for the valence to determine appropriate approach behavior of the robot 117

5.32 A comparison of the limbic arrangement for regular sitting posture with different chairs 121

5.33 Two occasions during experiment 3 125

5.34 This box and whisker plot shows the mean values of feedback scores received from users for the robot's proactive behavior with IDMM and without IDMM 125

5.35 A semantic map which shows the co-relation between various aspects of a human activity 128

5.36 Objects and aspects related to each semantic feature 129

5.37 The relationship between various steps involved in an activity . . 129

5.38 The decision flow of the approach during learning and recognizing activities. 130

5.39 Upper and lower limits of speeds related to each speed range . . 131

5.40 A typical conversation between the robot and its user while learning an activity 135

5.41 A map of the domestic area used for the study 137

5.42 Different occasions encountered during the experiment 140

5.43 Recognition accuracies of the 4 occasions considered in the experiments 143

5.44 Recognition accuracies when two occasions considered 143

6.1 Change of behavior encountered before and after approaching a human 148

6.2 Spatial vectors related to the body joints used to differentiate two poses 149

6.3 Input membership functions of the gaze parameters 150

6.4 Input membership functions of the gesture parameters 151

6.5 Output membership functions of each parameter 151

6.6 Output Membership Functions of second fuzzy evaluation 151

6.7 Rule Base to evaluate Gaze Level 152

6.8 Rule base for Attention level evaluation 152

6.9 Two occasions from a typical domestic environment 155

7.1 Approach of QASA 158

7.2 The top view of the HRI scenario 160

7.3 Distances marked considering maximum occupancy regions of the selected set of tasks 164

7.4 Two occasions encountered during PHN 165

7.5 This graph shows the mean values of feedback scores received from users for the robot's behavior in the three considered occasions . . 166

8.1 A semantic map which shows the co-relation between various aspects in human behavior 169

8.2 Features used as the inputs to the AR model are shown. 171

8.3 Marginal values to choose the *nature of conversation* and *proxemics* of the robot. 173

8.4 Marginal values to choose the *orientation* of the robot. 173

8.5 Values obtained for LOI_1, LOI_2 and LOI_3 during ordinary situations encountered 175

8.6 Occasions involved during SEA 179

8.7 The mean values of feedback scores received from users for the acceptance of robot's behavior in three occasions 179

9.1 The action-state model 182

9.2 Decision flow of the robot platform 182

9.3 The interrelations of semantic features during the Observation Plan and the Monitoring Plan. 183

9.4 Observational parameters considered in the system for robot decisioning. 184

9.5 Q-value test for 4 occasions while training 186

9.6 Q value tests for 3 occasions during training are shown. (The user feedback scores plotted against each user 187

LIST OF TABLES

4.1 Results of the experiment 36

4.2 T-Test for the comparison among each situation: while relaxing and while engaged when the person initiating the interaction was the robot and another human. 37

4.3 Some of the tasks selected for the study 43

4.4 A comparison of conversational preferences by the type of interaction when the user was alone and when with few people around 46

4.5 ANOVA test for the comparison of percentage frequencies of each interaction type in each area of the social environment 48

4.6 t- Test for the comparison among each type of interaction when the user was alone and with few people around 49

4.7 ANOVA test for the comparison of the frequencies of each conversational preference during each task 52

4.8 T-test for the comparison of gaze parameters among HHI and HRI-week 3 64

5.1 Experiment results 73

5.2 Results of the experiment obtained for two occasions: while users were posing for the sensor and recognition accuracy of hand pose while the users were seated. 80

5.3 Results of the experiment 87

5.4 An analysis of the confusion matrix generated from the results of experiment 2 102

5.5 Observations made through time 112

5.6 Valence assigned for each hand and body posture depending on the general interpretation of emotional state behind each posture 116

5.7 Results of experiment 01 119

5.8 Results of experiment 03 123

5.9 Observations made through time 141

6.1 Experiment Results 154

7.1 The set of zones and their corresponding probabilities of occupancy obtained for each task 161

7.2 Results of ten scenarios during the experiment and average feedback score received for each task 165

7.3 T-test for the comparison of user feedback scores in initial and final attempts 166

8.1 Calculated values based on observations 177

8.2 Results of the experiment 178

Chapter 1

INTRODUCTION

1.1 Development of assistive robots

Today, robots are deployed in our environments to make most of the tasks easier. Such robots are expected to have a general sense of the outcomes of their behavior in human-robot collaborative environments. Many robots entering social environments are expert in only one or a few given specific tasks. Two such smart robots used for cooking and cleaning are shown in Fig. 1.1 and Fig. 1.2. Cleaning robots [1], rescue robots [2], shopping assistants [3] and healthcare robots [4] are some examples for task-specialized robots which require lesser overall emotional intelligence. However with the deployment of robots in social environments, robotic systems have become a demonstrator of social and emotional interaction among humans and robots [5, 6]. Therefore in the decades to come, the hospitality and emotional intelligence of artificial agents are expected to increase with their wide variety of social applications. Most significantly, the robots have to match between a particular situation and their emotional behavior [7].

Some situations urge a robot to take decisions regarding the pattern of interaction expected. Such situations request a conscious observation. Such robotic systems are already used in teaching, childcare and other social domains. These systems still require development to adjust according to unexpected situations where human intentions change. Simulating and comprehending

Figure 1.1: An example of a cooking robot

diverse social behaviors of humans is a challenging, and yet sought after feature to be engraved into a robot's personality [8]. In contrast, there are humanoid robots which can replicate typical human characteristics in the form of physical appearance, movements, words, facial expressions, etc. An example system in which robot's responses evolved by means of an adaptive behavior is presented in [9].

Figure 1.2: An airport cleaning robot

Human ascribes upon robots must be considered before planning this present robot's intelligence. For instance, creators in [10] researched how robots can enhance the personal satisfaction of senior citizens by being a partner to defeat loneliness. Here, individuals' mentality that robots are performing overwhelming undertakings but still lacking in intelligent conduct, could be changed with such an approach. For example [11] are models for fall detection of elderly by robots. Thus, robots are acknowledged by numerous communities as a partner having the required capacities installed within. In such situations, following a decorum; essentially a "**robotiquette**" acknowledged by people is expected [12, 13]. Somehow numerous reactive techniques for human-robot domains have been set up so far, but yet not many proactive strategies were settled [14].

On the other hand, we live in a complex environment filled with various distinct phenomena. In the mean time, we revolve around other people we daily meet and incidents that happen in the surrounding. We react to people and phenomena around ourselves in various means, based on the nature of a particular situation. Our reactions are visible to outside through behavior. Out of such reactions, words, expressions and physical movements are powerful drivers of human behavior [15]. Perception of such behaviors by such robots is beneficial in this regard.

Preference-based collaboration with robots is a demanding viewpoint in human-robot shared situations [16]. In such circumstances, adopting supportive behaviors which rely on the insight, instead of playing out a requested task are required. Fitting such smart conduct is as critical as challenging. This is because of the multifaceted and complex nature of human conduct and perceiving such practices, outcoming the difficulties in innovation and environment. Accordingly human-level prediction of situations still needs enhancement [17].

A set of observable cues extracted from humans and their environment can be used as demonstrators for a perception model to simulate a human-robot scenario. Hence the level of interactivity within a situation can be determined. Nonverbal features of perceiving a situation are the most prominent and effective for an evaluation of the situation prior to an interaction. Understanding nonverbal behavior or the body language elevates decision making capabilities related to the interaction initiation by a robot, as the system outputs a measure of the emotional state in the human-robot encounter. Two such encounters are shown in Fig. 1.3. In the first instance in Fig. 1.3(a), the user gives instructions to follow. Hence the perception of the object placement, user's gestures and voice instructions will be adequate to perform the task. In contrast, Fig. 1.3(b) shows an occasion in which the user was unaware of the robot's presence. In both the situations, the state of the user is determined by the intentions of him/her as well as factors that exist in the surrounding environment. As Fig. 1.3 (b), objects and people make the environment, while intentions of the user are responsible for the factors within him/herself which might also affect the emotional state of that situation. In order to deliver a service in a polite and socially accepted manner, the robot can adjust it's approach behavior, dialogs and other actions regarding interaction to match the situation. User's engagement in the task and interest for an interaction have to be predicted by the robot based on the factors

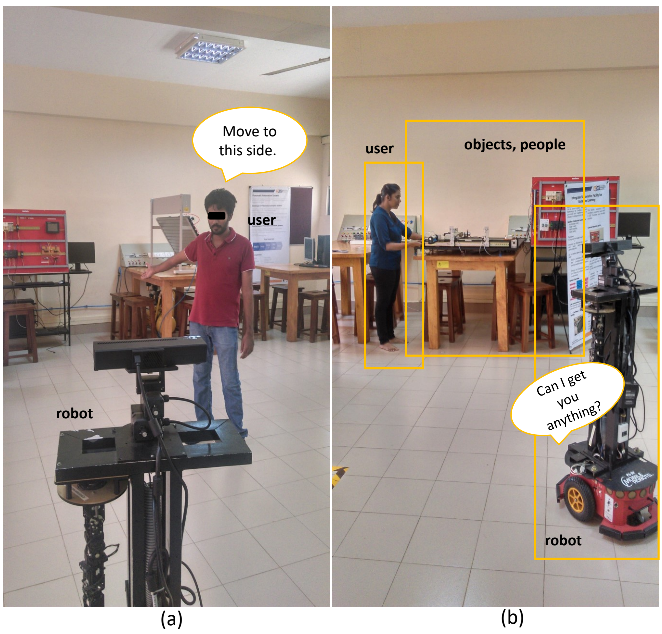


Figure 1.3: Example scenarios where the required level of situation-awareness differs.

visible in the scenario. Hence perception of emotional, social, psychological and other aspects of the situation become prominent during this type of scenario. The algorithms outlined within this thesis allow a robot to observe such a context, decode human nonverbals and generate context-aware responses to initiate an interaction to support its user at its maximum.

1.2 Building Attentive Robots

Theory of Mind (ToM) has been developed in order put various attributes in mind together, to reason out a certain behavior and it has often been used to evaluate abnormal or awkward behaviors which have been deviated from the accepted level of behavior [18, 19]. The Theory of Mind has been used in several architectures to perceive human intention associated with a specific situation [20], but these approaches still estimate user's preferences based on a limited number of cues from its environment. Henceforth adequate evaluation of a situation is critical for emotional intelligence.

1.2.1 Study of human behavior

A human's intention alters involuntarily upon the factors that prevailed already in the surrounding. This perception will be based on various parameters including the individual's beliefs. Hence, the user's reaction upon an interaction which has been initiated by the robot will take different forms depending on the scenario. Responses that are most likely to be displayed from a human during a human-robot interaction scenario, can be used to assess human behavior during the study [21]. Such observable responses are listed below.

- Gaze - Maintaining or returning to original gaze e.g: looking at the robot and/or looking away

- Gestures - Using mainly hand gestures e.g: waving hand, calling in
- Postures - Changes in existing posture e.g: sitting to standing posture
- Utterances - Verbal responses e.g: “Hello”, “May I get you something?”
- Movements - Random or intentional movements associated with activity
- Expressions - Facial expressions e.g: smile, disgust, frown

These responses devote to perceive and evaluate attitudes, attention, expectations, subjective norms and perceived behavioral control mainly as explained by the theory of planned behavior.

There are many cognitive and psychophysical theories to explain human behaviors as well as the behavior of a robot. Even so, we still lack a model to perceive and predict appropriate behaviors for both the human and the robot during a human-robot encounter. Humans make an instant evaluation of the surrounding and its people before approaching anyone. As robots are widely deployed in social environments, a similar perception of the situation around its human user prior to an interaction is required. Social constraints during an interaction could be demolished by this assessment. Through this research, we tried to utilize such theories and build a criteria for multimodal cognition of a social robot to interact with humans.

As the basic idea of this thesis is to evaluate the requirement of providing the robot with the ability of understanding situations or ‘cognition’, we deployed these theories to investigate how to lay a justifiable background for this. Hence the cognition in a situation will relate the connection between human’s state and his/her behavior. Such an approach facilitates a dynamic interplay of flexibility and adaptation in robot.

1.2.2 "Meet, Perceive and Act"

Many of the today's robotic systems are intended to act according to a certain goal. But there are occasions at which 'not acting' is the most appropriate behavior to suit the condition in that environment. Hence perception of the situation is important in determining such behavioral requirements.

1.2.3 Multimodal perception

In an normal environment people use to interact with others and the environment using more than one modality. They speak, use gestures and look at things to interact with environment and other humans. By listening to different voice tones, looking at face gazes and arm movements people understand communication cues. Interaction among humans and animals tend to show a multimodal behavior due to their biological build. Studies regarding superior colliculus have revealed that different senses are segregated initially at the neural level. When those senses reach the brain, sensory signals are converged to the same target area in the superior colliculus. It also receives signals from the cerebral cortex and in turn modulates resultant behavior. Also, studies show that about 75% of neurons leaving the superior colliculus are multisensory. These particulars encourage the multimodal interaction in HRI especially in social environments. Aside of the concepts of unimodal and multimodal interfaces, what really matters is the implementation of those concepts. It requires the assistance of the field of sensory data fusion. Furthermore Statistical data analysis goes hand in hand with data fusion. Therefore we tested our algorithms upon robot's perception with a thorough analysis of statistical data gathered.

After an insightful evaluation of its context, including its user, a robot will be capable of generating context-aware responses towards a human, rather than acting upon a predefined set of behaviors.

1.3 Problem Statement

This thesis addresses the problem of engraving adequate insight and social intelligence into proactive social robots to initiate interactions with its human user. To achieve our objectives in this regard, we tackle the following three problems. First, we address the problem of recognizing features that make up a context and have an impact upon the human-robot interaction scenario. Within this scope, we considered both human features such as nonverbals and environment features such as space and objects. Second, we introduced simple models which can analyze a certain human-robot encounter to assess the favorability of the situation for an interaction. Initially, such models were limited to a small number of features and, the complexity of models was enhanced and the number of observed features were increased as the third solution to our problems. Finally, these context-awareness models were improved to adapt dynamic user behavior and tackle individual differences in human behavior. We investigate how a robot can perceive the state of its user and the environment to generate appropriate responses in order to support the user. Furthermore we develop cognitive architectures for a social robot to enhance empathetic behavior of a robot during HRI, based on our findings. Adaptive models were based on these cognitive architectures to perceive human intent and generate proactive responses to match the scenario.

1.4 Thesis Contributions

1.4.1 Identifying observable human behavior

A situation between a human and a robot consists of the *robot* itself, the *user* (human) and the *environment* around the robot and human. When the robot intends to perceive such a situation, it first has to identify interactive factors within itself, the environment and the user. Factors within the robot itself include the dialog patterns the robot generates, maintaining an interactive distance in between, and displaying appropriate behavior, etc. Factors within the user will be numerous, but emotions, social norms, beliefs, personality traits, activities the user involved in and other psychophysiological factors contribute majorly to decide the level of interaction readiness within a human. Objects and other humans in the surrounding, obstacles, etc. make the list of factors in the environment which have to be perceived by the robot.

We conducted several user studies to first monitor how humans respond towards such different scenarios within their environment. Human behavior in daily encounters during human-human interaction (HHI) were evaluated and compared with their responses towards social robots placed for their assistance during human-robot interaction. Most of the human studies were conducted by means of wizard-of-oz (WoZ) experiments. A typical setup during an experiment is shown in Fig. 1.4. In this scenario, the user is involved in a desk activity in the bed room. In order to understand the whole situation, the robot has to be knowledgeable on three aspects; itself, the user and the surrounding environment. Factors related to these three aspects are marked as 1, 2 and 3 respectively. Implications derived from these human studies were used in the development of cognitive models for social robots in the next stages of research.

1.4.2 Recognizing components embodied in a context

A context is made of its people and the objects. Such attributes and the semantic relationships between them have to be correctly perceived and mapped for their meanings or general interpretations to understand a particular scenario. This requirement is the same during human-robot interaction. Therefore recognizing the compnents which make up a context is the next stage of this research. We focused on recognizing human features such as posture, movements, upperbody gestures and several other nonverbals to a great extent in this research,

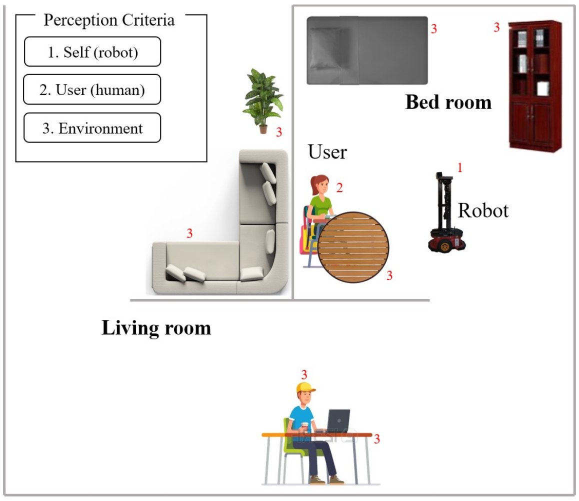


Figure 1.4: An example domestic environment is shown.

but we used existing methods to recognize the environment as well. The reason for this is that, it is important to perceive both these aspects: user and the environment in order to have a complete understanding of a scenario.

1.4.3 Decoding nonverbals to interpret user situation

There are numerous components associated with a scenario which nonverbally signals the requirements or sentiments within that scenario. Such components are illustrated in Fig. 1.5. Semantic relationships between each component have to be understood to model the situation and generate appropriate responses from the robot. Such an understanding is required to ensure maximum user satisfaction by not violating user expectations during robot-initiated interaction with the user and in the meantime, to serve any requests of the user. Hence the robot can get 'experience' by interacting with people and evaluating their responses during different occasions. This 'experience' will help cognitive models to generate more context-aware actions from the robot. Therefore we developed our Cognitive models based on these components of a context.

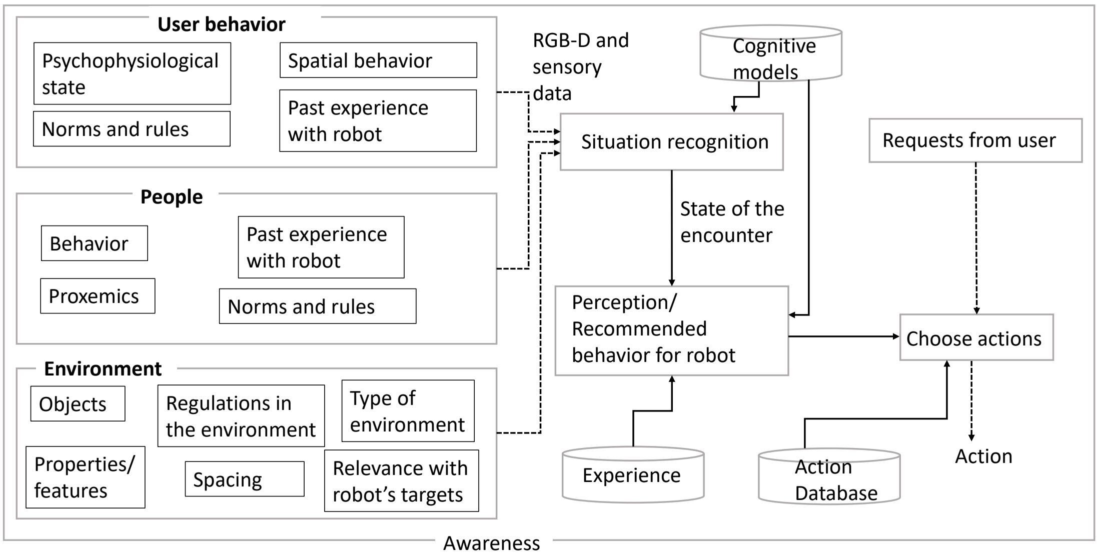


Figure 1.5: Different types of settings and components associated with human-robot encounters and the process of decision making are shown.

1.4.4 Adapting to a situation

Complexity of human behavior outstands the complexities in environments during proactive HRI. Therefore perceiving individual differences in users' personalities and the nature of performing activities is important in achieving intelligent proactive behavior of a robot. Besides, there are dynamic changes in the environment such as the placement of objects, time of the day etc. Hence we improved our cognitive models into adaptive approaches which are able to perceive such dynamic changes in a context in the end.

1.5 Thesis Overview

Overview of this thesis can be summarized as follows.

- Chapter 2- Literature Review. Existing systems for interaction with humans and possible improvements are discussed. Furthermore, the proposed system and the research gap is presented.

- Chapter 3- Presents the design of the system. Overview of the proposed system and hardware used in the implementation of proposed system are presented.
- Chapter 4- Human studies conducted in support of the research are explained. These studies include experiments, wizard-of-oz studies and surveys.
- Chapter 5- Novel approaches introduced during this research for the recognition of nonverbals within humans are explained. Among these are gesture, pose, movements, activity space and human activity recognition.
- Chapter 6- A fuzzy-logic based approach was introduced to decode multiple nonverbals and evaluate human attention towards the presence of a robot.
- Chapter 7- The previous approach was extended to an autoregressive model to perceive human behavior to evaluate a situation prior to an interaction initiated by the robot.
- Chapter 8- This chapter introduces an approach to determine proxemics and orientation of the robot during an interaction scenario.
- Chapter 9- A learning and adaptive approach is developed as an extension of the previous systems. This system is based on reinforcement learning and is an integration of previous recognition techniques as well.
- Chapter 10- Concludes the thesis.

Chapter 2

LITERATURE REVIEW

2.1 Nonverbal Interaction during HRI

Capacity of making inviting discussions at right occasions is indispensable in accomplishing a proactive behavior of a robot. Robots are required to have insight to decide when to cooperate and when not to. When starting communication, this natural or human-like conduct upgrades joint effort among robots and the non-master. So as to improve the union between human clients and robots, robots which can replicate human practices so as to assume the role of a close contact, for example, 'companion' and 'parental figure' are being produced [22–24]. However, the focus of these methodologies is restricted to the investigation and implementation of strategies that ought to be encapsulated into the robots. This helps robots to keep up a fluid interaction with their users. However, how does a robot identify key features to observe during a human-robot encounter?

To find answers to this question, researchers have been trying to establish common criteria to model human behavior. A large number of the current frameworks utilize facial expressions and body postures and additionally voice so as to decide the interaction demanding or the emotional state of a situation [25]. In spite of the fact that voice can be utilized as a factor which demands interaction from an outsider, this is just conceivable after the user speaks. That is, after the user initiates the interaction. Meantime, a few methodologies have been introduced to assess the expressions and body postures during an interaction so as to judge user's likeliness towards an interaction with the robot [26, 27]. However, these realities can be assessed simply after the commencement of the interaction. Perceived information is used to decide upon the continuation of interaction flow thereafter. In any case, postures are useful in distinguishing the interaction demanding of the human subject before beginning a conversation.

Which factors to be selected to analyze a situation is an area which requires attention. As a situation consists of the robot and its surrounding, factors from all these three aspects will need to be analyzed. Even so, development of a criteria which can bring all these factors to a common pool is still confusing and challenging. Difficulty of quantifying emotional factors for the analysis can be stated as the reason which retards the progress in respective field of research.

2.1.1 Features to observe

Environmental features

Orientation and locomotion are profound collaborators in 'interest' during an interaction among humans. This fact remains the same during human-robot interaction as well. Locomotion and attribution of body-based movement as an interpretation of own intentions are committing factors during HRI [28]. In [29] provides an example of proactive obstacle avoidance in dynamic environments is given. Such systems are examples for intelligent agents with situation-awareness based on spatial behavior of subjects. Spatial arrangement of the two conversant is an important fact to determine the interactivity of a situation. Furthermore, the number of outsiders and placement of objects affect both directly and indirectly towards the emotional state of a situation. The relationships between people and behavior of the people in the surrounding further affect the responses of a human to a certain situation. Simply, all the subjects in the environment which are part of a person's cognitive mapping have an impact on his/her behavior [30]. Perception of natural environment such as symbol anchoring [31] and tracking dynamic obstacles [32] accounts for a proper understanding of the environment. However we exclude literature about robot's perception upon the *environment* and *robot* itself in this review.

Features within a robot

Certain features and practices encapsulated in robots have an effect on individuals' willingness to participate in at least a brief interaction with the robot. Work explained in [33] has displayed a lot of social standards for robot conduct (a 'robotiquette') which is convenient and agreeable to people. As indicated by that, the conceptual space of HRI expects a robot friend in a home to 'do the correct things' and meet their expectations comfortably. Moreover, constant execution of the robot which pursues human social traditions and standards are bound to be acknowledged for a long span by its users [34]. In [35], the robot is exemplified with the ability to express emotions with a humanoid face and demonstrate attention by turning towards the subject. Creators have theorized that these highlights were non-negligible prerequisites for a viable social communication between a human and a robot. Anyway this framework has ignored enhancing the capacity of the robot to predict or realize user situation before using the previously mentioned two highlights while having an interaction. This approach gives an example of socially intelligent behavior of the robot as well. Even so it focuses on the behavior of the robot rather than that of the human. As per the model proposed in [36], **perception** and **evaluation** are constantly considered as critical attributes in human-robot interaction.

2.1.2 Earlier Approaches

At the point when these robots are required to perform as domestic companions, starting a conversation at right events is very demanding among users. A great part of the users lean toward interaction by means of voice or friendly conversations [37]. Such insightful behavior upgrades the attachment and connection between the user and the robot [22–24, 38, 39]. In this way, gaining that emotional intelligence is a critical viewpoint concerning social situations.

There have been various psycho-physiological approaches to gain emotional intelligence through perception of behavior. Cognitive methods to understand the interactivity during a conversation have already been developed. These address the problem of identifying the intent of the conversational partner based on verbal cues and facial expressions. For example, in [40], the engagement of a human is assessed by the head-nod during a conversation.

In [41], a robot which was an insightful weight reduction mentor has been presented. This is an excellent example for behavior perception exemplified in the robot itself and subsequently has been deployed to reduce obesity among participants. In this case, the robot has been accepted for a long term interaction as well. Anyway this framework wasn't completely skilled to perceive rather general user situations other than monitoring physical health. By incorporating instinctive behavioral aspects in emotional situations into a robotic architecture, a higher emotional intelligence as well as a greater user acceptance can be ensured [42].

There are numerous automated frameworks to shape a discussion between a human and a robot, yet the ability of these frameworks are restricted simply after the initiation of an interaction, not before the interaction [43]. Mimicing human behavior is exceedingly applauded in cooperative discussions [44]. This methodology is prominent for keeping up interaction which has already been started, not prior to an interaction. Satake et. al in [45] enhances this scenario by taking various user cues into scrutiny. Forecasting of walking direction was valuable in estimating whether the user showed any interest for a conversation. 'User unaware' failure was abrogated in this method by moving toward the user before starting a conversation. In 'user unaware' failure, the user leaves the situation without realizing that the robot approaches to interact with him/her.

Furthermore, in [45], a model to define an approach behavior for a robot to initiate a conversation with dynamic users has been proposed. The system was first intended to use in shopping malls to prevent 'user unaware' failure when a robot approaches a walking user. This mechanism was effective to predict the intention of dynamic users. Such approaches might not perform in smaller spaces such as domestic environments, as users adopt much less speeds and short trajectories in general.

Method proposed in [46] computed the 'degree of interest' of a user towards an interaction with a robot by evaluating the subject's attention and the distance from the robot to its subject by using a fuzzy logic based system. This system could perceive ambiguous situations with a higher accuracy with the help of linguistic variables that account for the fuzziness. A major concern while having a conversation is the distancing between the two conversant. Accordingly, the robot must decide upon the appropriate distancing for their communication before initiating a conversation. Personal zones suitable for such circumstances were presented in [47] and [48]. Hence a robot may choose which zone to enter while having an interaction, depending on the displayed interest of the user.

2.1.3 Modern Developments

A significant number of the past work have investigated the elements which influence human-robot interaction [33]. However exploring methods of initiation and maintaining smooth flow of a robot-initiated interaction, have been impeded due to practical and theoretical constraints. One under-investigated factor in this manner is perceiving nonverbal human behavior before communication. Once the robot is able to evaluate nonverbal human behavior at a certain scenario, it could evaluate the likelihood of initiating an interaction based on the encounter. In this manner, the impression of user situation is fundamental in such an event. Conversely, the absence of predictability and the transparency in numerous advanced mechanical frameworks have adversely influenced human's trust and dependence on robots [49]. A verbal approach for interaction is shown in [50]. During this approach, a robot which reacts dynamically to a visitor is deployed. Conversational opening is used as a critical influencer in maintaining visitor engagement. It further analyzed user's upper-body postures, facial expressions and head movements as demonstrators of engagement. Likeliness of the robot being acknowledged as a conversational partner relies upon the surrounding and also the emotional state of the user. Moreover, it is imperative to investigate human inclinations towards a communication with a robot since human behavior before another human could take a different form in comparison to that before a robot.

In [51], nonverbal user engagement is approximated using initiator and responder gaze times, face orientation and feedback times for the two speakers as cues. Bodily postures of humans that are also an interesting parameter just as the gaze time were excluded in this method.

Mimicking real-world HRI behaviors and simulating scenarios based on observable information are key features during human-robot collaboration (HRC) [52]. A predictive and adaptive system based on learning by demonstration is explained in [53]. This system is an example for robots with supportive behavior for the user but it does not identify many salient features which portray human intention. A situation-conscious model is proposed in [54] to improve the design and interactive capabilities of an industrial robot. These findings show that the impact of social and spatial environment have to be considered in order to design a context-aware robot.

In [55], an android system with the capability of monitoring noverbal behavior of humans. However this cannot generate physically appealing behaviors such as a robot. There has been systems which were capable to match the appearance and demeanor as well [56]. Even though such systems cover psychological and physical aspects in interaction, there should be emotional aspects as well in order to behave appropriately in social environments.

A context-sensitive approach to anticipate the human behavior while the human is followed by a robot has been proposed in [57]. Although this anticipated human motion while walking, prediction of human behavior during stationary situations becomes highly dependent of the emotional state and the task of that human. Hence anticipation of stationary situations is a complex process. There are CPU-intensive approaches to recognize human activities such as [58]. However less CPU-intensive techniques to monitor human behavior are admired when real-time decisions have to be taken. Requirement of lesser pre-processing in such techniques become advantageous in implementing them in real-time.

Perceiving emotional cues that shows affect is important in avoiding misbehaviors of robots and improving acceptance in human community [59]. Findings suggest that perception of nonverbal behavior positively impacts HRC and hence the understandability of the robot is increased as well [60]. A situated interaction method which uses behavioral cues such as proximity, velocity, sound and posture information is presented in [61]. A virtual companion adjusted himself according to the situation and outputs an engagement score at each occasion. Even so only the gaze behavior of the virtual companion was evolved based on behavioral cues of the user.

2.1.4 Systems with Adaptive Perception of Situation based on Human Cues

An intuitive, multimodal approach for interaction was introduced in a museum guide robot in [62]. This robot used audio and video information to shift attention among humans and to monitor its eye and hand gestures while speaking. This robot further interacts with multiple users at a time and uses a series of facial expressions based on users' interest. This method further determined focus of attention to be paid on each person based on the time when a person has last spoken, distance between the person and the robot, and its position relative to robot's front. Therefore this method can be stated as a first step towards perceiving a situation for adaptive social behavior.

Mead et. al proposed a method to evaluate the perception of distance in [63]. During this approach, robot used gestures and verbal responses of the user to determine the mutual distance. In [64] a probabilistic approach has been adopted to analyze the engagement in verbal cues and gestures used by a human during interaction. However both these approaches are possible only after initiating an interaction by the user. Hence the task performance of a human-robot team can considerably be improved by mutually understanding nonverbal cues responsible for the situation [65]. Therefore perception of nonverbal cues related to human behavior and relating various cues together to determine the state of the situation are the attention-seeking requirements during HRI.

An affective robot was deployed to interact naturally with customers in a shopping mall and to provide shopping information [66]. It further tried to build a rapport between customer and the robot by remembering customers who repeatedly visit. This was an effort to identify which kind of robots were required by people to in their shopping routine. This is an example for a special scenario where a market-related situation perception was used. However actual human behavior will be more complex than in a shopping environment. When robots are deployed in our ecosystems, they have to be self sufficient physically [67] as well as instincts in order to play the role of an active companion.

A framework to develop mixed-initiative approach for specifying the relationship between dialog structure and task structure using generic interaction patterns is proposed in [68]. Action-oriented dialogs were generated by a robot depending on the current task of its user. Objects from the environment and the task of a person were linked to find relationship in between, and hence conversation between the robot and the human was adapted accordingly during an object-grasping task. The work proposed in [69] is an example for a robot creating a map and maintaining a knowledge database through interaction with people in natural language.

In [70], a social robot was used to handle emotionally charged health care situations. According to the findings of this research, people's perception of robot was affected by how robots cope or how they think robots can cope with their emotions. In the end, improved intelligence of robots, like in these mechanisms, will increase people's acceptance of robots for longer interaction.

During the study of psychological benchmarks in HRI, in many forms of human-robot interaction there is almost nothing gained functionally by using a humanoid [71]. Furthermore, intelligence and capabilities become prominent in a robot upon its appearance. There are also contexts where the humanlike form might work against optimal human-robot interaction. For example, an older person might not want to be seen by a robot with a humanlike appearance when being helped to the washroom. In addition, people may dislike a robot that looks exactly human but lacks a humanlike behavioral repertoire [72].

An effort to model interactivity of an encounter using four types of connection events which incorporate gesture and speech: directed gaze, mutual facial gaze, conversational adjacency pairs and back channels was taken in [73]. This uses initiator and responder gaze times to determine gaze and gaze is used with the other three parameters to determine user engagement during interaction. This is an example for using both verbal and nonverbal cues to perceive an interaction scenario. Adaptive speech control mechanism as a response for conditions in the environment has been proposed in [74]. However these are an example for evaluating a single or fewer number of cues from a situation.

User-awareness has been taken into consideration in [75] for safety and acceptance reasons. Spatial relationships, objects and dynamic gait behavior of a human were considered as features of the setting. Another mechanism for an approach behavior was proposed in [76]. But in this approach the tradeoff between user's and robot's gaze were used as a cue to evaluate the situation. Here, situation-awareness was used to determine approach behavior prior to an interaction.

Museum guide robots observing visitors to find an appropriate situation to guide has been proposed in [77]. This perception was then used by the robot to locate himself in a spatial-oriental arrangement with a human. The purpose of this behavior was to allow humans perceive their participation. A conversation was possible with a visitor only when both parties; robot and visitor establish a common belief to share a conversation. In this method, the robot believed that a visitor is interested in an exhibit if their face and body are oriented towards the exhibit for a certain duration. Similarly, their requirement to know further about the exhibit was determined by if they maintain face and body orientation towards the robot for a certain duration. But in social environments, human encounters will be more complex than those in a museum. Hence much more cues will be associated with user behavior.

A mobile robotic system which adaptively attend its user based on walking/sitting behavior of its user has been deployed in [78]. This system uses Finite State Machines (FSM) to model transitions of the user's state. The robot tracks a person's position and orientation to determine his/her state. The decision of the system based on these states was the path followed to approach the particular human. Here, three states: 'initial', 'walking' and 'sitting' were used by the FSM to model the transition of the person's state. Accordingly, dynamic or static behaviors in humans have a great impact upon the interactivity of a situation. It considerably affect the approach behavior of the robot as well. Yuan et. al [79] evaluate the applicability of deep learning approaches to adapt to and predict comfortable proxemic behavior during interaction. The model estimated the discomfort when approaching its user. Distance between participant and robot, angle of approach, gender, age, previous experience with robots, preferred writing hand, and pet ownership was chosen as the inputs to the algorithm. Even though the research was about a comparison of different deep learning algorithms, the approach taken for the proxemic behavior was itself novel and situation-based. The model was capable of predicting proximity distances that were suitable in the context of HRI.

In [80], authors have proposed a system which used Regulatory Focus Theory, user's psychological state and game performance information to determine user's stress while playing a game. Hence the robot adapted its behavior accordingly. Different kinds of body-based gestures and speech speeds of the robot were regulated based on the observations on the scenario. Furthermore this is an example for using body-based behaviors as a demonstrator of internal state of humans. [81] is and example for human's sensitivity towards empethetic and emotional features in a robot's speech. This fact can be associated with the overall performance of a robot as well in addition to speech alone.

In [28] the question 'How do the spatial distances and orientations of a user in relation to a robot vary throughout a cooperative task performed in a home-like environment?' has been addressed. Hence a spatial conduct for the robot was developed during this approach. This spatial management behavior was capable of active monitoring and dynamically reacting to each others' movement and position changes. This used the definitions of proxemics introduced by Hall [82] for interpersonal distancing and Kendon's F-formation arrangement for orientation between two persons [83]. The robot was intended to learn and find objects that were missing from its original location. For this purpose, the robot had to follow users who were willing to show the objects. Spatial formation of the human and the robot was analyzed during 'following the user', 'showing an object' and 'validating'. As different formations could be observed during each occasion, it could be deduced that there should be a perception upon each occasion during an interaction.

In [84], a robot that distributed flyers to pedestrians has been developed. This focuses on achieving the target by identifying the behavior of pedestrians and determining appropriate approaching mechanisms. This intended to approach a human being non-obstructive to the receiver. Approach direction was computed after a detailed observation on actual scenarios involving human-human interactions. Hence an entire distribution plan was deployed within a single situation.

Most of the findings were based on a limited number of tasks selected from the environment and the user itself. In a real-life scenario, this number will be much higher than the number of factors considered in the present systems. Therefore a maximum number of parameters must be observed from the user and his/her environment before the decision-making process of a robot. Therefore these systems could not replicate all parts of the HHI into the HRI scenario.

These methods were based on the assumption that people prefer the same rules of interaction with the robot as they do when interacting with humans. There can be certain cultures and social groups in which there are alterations regarding this fact. Hence such communities would react to robots in a different manner. In addition, behavior adaptation is as important as behavior monitoring in such a scenario. Several other factors which might influence interaction such as the gender, previous experience and familiarity with the robot were not considered within the context of many methods analyzed during this review. Furthermore there should be a common platform which can analyze psychophysiological, social, cultural, and other aspects of a situation. Finding relationships between these aspects and the emotional state of a situation is still found to be challenging. Unveiling the intentions behind the web of various human behaviors is another requirement in modeling cognitive models for social robots.

User-centered Design

User-centered design confronts designers to mold an interface around the capabilities and requirements of the operators. Rather than displaying information that is gathered around sensors and technologies which produce it, a user-centered design unites such various information in ways that set the goals, tasks, and requirements of the users. As a result of the user-centered design, we can extraordinarily diminish blunders and enhance productivity without requiring significantly new technical capabilities. Alongside user-focused structure, also comes enhanced user acceptance and satisfaction as a side advantage, by removing much of the frustration upon limitations of present technologies. User centered designs provide measures to support humans and then humans will work better with robots. The requirement of user's adaptation according to the limitations in technology is abolished during this approach.

In order to avoid people calling a robot explicitly, proactive robotic systems which can offer their help voluntarily are required. Therefore the user will not have to formulate his/her behavior to suit the capabilities of the robot, as the robot can read the intentions of its user. Simply, the proactivity of the robot reduces the user's effort. A robot with cognitive skills engraved into its personality traits is more reliable [85]. Aspects where a general sense is required from the robot are shown in Fig. 1.5. Each aspect; user, robot and environment is subdivided into smaller elements which affect the situation. For example, factors within the user are categorized into psychological attributes, physiological attributes, social norms and rules etc. Factors in the environment are divided into people and objects. Both spatial and emotional attributes of each aspect are considered.

Psychological state is the reason behind human behavior. This psychological state is displayed to the outside through both verbal and nonverbal behavior. Body-based behavior is the result of cognitive processes developed in human brain. Behavior can be analyzed as an interplay of mental states and actions. Simply, thoughts and emotions provoke actions. In addition, cognitive elements such as facial expressions, verbal phrases, etc. fall under 'behavior' which includes both verbal and nonverbal aspects. Furthermore, brain activities such as internal states of mind, cognition and emotions are responsible for one's actions. Proper interaction between brain activity and actions, not only makes him perceive the world around him, but also enables the others around to perceive him.

Behavioral responses of humans can either be voluntary or sometimes involuntary. Furthermore, many involuntary behaviors are nonverbal. There is a number of psychological theories behind both voluntary and involuntary human behavior. Among these various theories, the theory of planned behavior and the theory of reasoned action map human actions and their thoughts in a rather reasonable and a justifiable basis [86]. Hence a method to perceive and combine these factors associated with a situation to predict the emotional state of a certain situation is required.

Introducing an established relationship between these various factors which are difficult to quantify, has been challenging at all the times. This is the reason to develop systems which evaluate only a limited number of factors so far. Often these systems either evaluate human factors or environmental factors alone.

2.2 Outlook

Following challenges have been disclosed during the study of present proactive robotic systems.

- Robust human detection and tracking systems to analyze gait and nonverbal behavior are not in par with advanced conceptual design of such systems.

- Establishing a relationship between intent and observable human cues has been challenging and lacks conceptual basis.

- Monitoring the environment in parallel with human behavior has been difficult and relating environment factors and human factors upon a certain behavior has been challenging.

- Deceiving and ironical human behaviors which do not match the shown intent with the real intent cannot be differentiated by existing robots.

- Bringing all observable cues to a combinable common platform in order to make interaction decisions, still needs further development. In other words, unwrapping the social and emotional attributes in physically observable cues still lacks conceptual basis.

Hence emotionally intelligent agents need development optimizing these challenges on the way.

Summary: Most earlier approaches which intended to evaluate human attention or the interest towards an interaction focused on the face i.e. head nod or gaze. Verbal cues and facial expressions came next. In addition to movements, approach behaviors played an important role in finding human intention in dynamic environments such as museums and shopping malls. Some systems utilized the body orientation as a cue when recognizing gaze was difficult. Systems which used multiple cues as input, utilized fuzzy logic to evaluate vagueness in portraying intentions through behavior. Other approaches used predefined set of interpretations for most observable cues and appropriate robot's behavior. Previous studies further explored that higher emotional intelligence in robots increased user acceptance. Existing frameworks to interpret human behavior consider only a limited number of observable human cues. Therefore only a few aspects of a situation were effectively perceived by these mechanisms. On the other hand, establishing a common criteria to assess a situation, considering all the aspects: human, robot and environment, is still confusing and lacks conceptual basis. The key features of humans that could reveal their internal state of mind have been identified. In addition, different attributes: emotional and spatial, during a human-robot encounter have also been identified. This will create a user-centered perception of a robot, which will demolish the requirement of the human user to adopt a restricted behavior to accommodate the robot's limited perceptive skills. Therefore it has been identified that intelligent and adaptive decision-making skills in dynamic environments make robots more appealing and effectively performing in human-robot collaborative environments.

Chapter 3

SYSTEM DESIGN

3.1 Extracting information

Visual information from user's environment are captured from Microsoft Kinect sensor. There are 3 coordinate spaces used in the Kinect camera as color, depth and skeleton spaces. Color space captures pixel values of the captured image as RGB information, depth space gives pixel values which can be mapped to a distance and skeleton space uses above 2 spaces to give coordinates of a joint as (x,y,z) coordinates. The skeleton data consists of a set of joints. For gesture recognition, in-built features of the sensor such as palm recognition and joint tracking were used. The sensor creates the depth map of the body by analyzing a speckle pattern of IR laser light. Then body parts are inferred using a randomized decision forest to map depth images to body parts. The mean shift algorithm is used to compute modes of probability distributions and finalize the joints to make the skeletal.

3.2 Robot Platform

Moratuwa Intelligent Robot (MIRob) platform shown in Fig. 3.1 is used for the implementation of the proposed system. It consists of a Pioneer 3DX mobile robot platform. The robot has two sonar sensor arrays one in the front and one in the back. The base can reach maximum speed of 1.2 m/s and carries a payload of

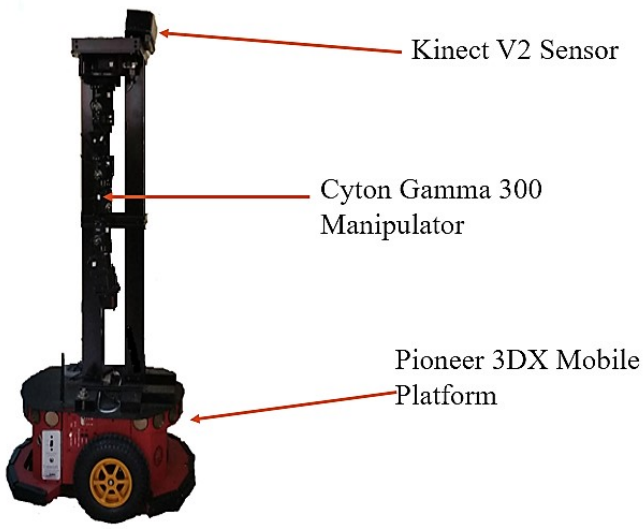


Figure 3.1: Hardware setup used in experiments.

up to 17 kg. An aluminum structure is constructed on top of the base to increase the height of the robot to a match the height of a human. A Cyton Gamma 300 manipulator is installed on robot for object handling. The manipulator has 7 DOFs and a 1 DOF gripper. It can handle a maximal payload of 300 g. On the top, a Kinect sensor is placed to capture vision information from the user's environment. Google Voice API was used to generate voice responses to initiate an interaction.

Proposed concepts have been implemented on the MIRob platform. Navigation maps required for locomotion were created with Mapper3 Basic software. Skeletal representation of human body was extracted as 3D co-ordinates of feature points in Kinect sensor. The experiments were conducted in an artificially created domestic environment with the participation of users in a broad age gap.

3.3 Overview of the system

The functional overview of the proposed system is shown in Fig. 1. Shown in blue are the information extracted from the users and the environment such as object attributes. Green arrows denote the robot's responses towards the user after an inspection of the context. The system evaluates unclarifiable characteristics such as postures considered by humans often before initiating an interaction with another. This enhances the decision making abilities related to interaction initiation. The system outputs a method of identifying emotional situation of a human, using his body parameters and spatial information. The system has separate functional units to recognize pose, gestures, movements, human activities and other human behaviors.

All the work explained in this thesis, were generated based on the information collected by the robot. MiRob collects visual information upon its user and the environment using the RGB-D sensor. Human body-related information is extracted by the module called 'Body Parameter Extractor (BPE)' and space-related information such as measurements related to objects, dimensions and distances were collected by the 'Spatial Parameter extractor'. Extracted information for the entire period of observation is stored in the 'Data Recording System (DRS)'. Data recorded in the DRS are used for the functioning of the next modules or calculate new parameters which will be important for decision making. Hence the DRS contains both calculated parameters and raw information. Both these types of information are analyzed and used by the 'Body Parameter Analyzer (BPA)' to evaluate user-related behaviors and by the 'Spatial Parameter Analyzer (SPA)' to evaluate space and distance-related settings within the context. Parameters analyzed by the Body Parameter Analyzer are used by all these modules: 'Posture Identifier (PI)', 'Gesture Identification Module (GIM)', 'Activity Recognition Module (ARM)', 'Human Behavior Evaluator (HBE)' and 'Context Evaluator (CE)'. These modules are responsible to recognize different human behaviors and nonverbals, and decode the sentiments wrapped inside such behaviors.

For instance, Posture Identifier is responsible to recognize a predefined set of postures (both body and hand postures) based on the information derived

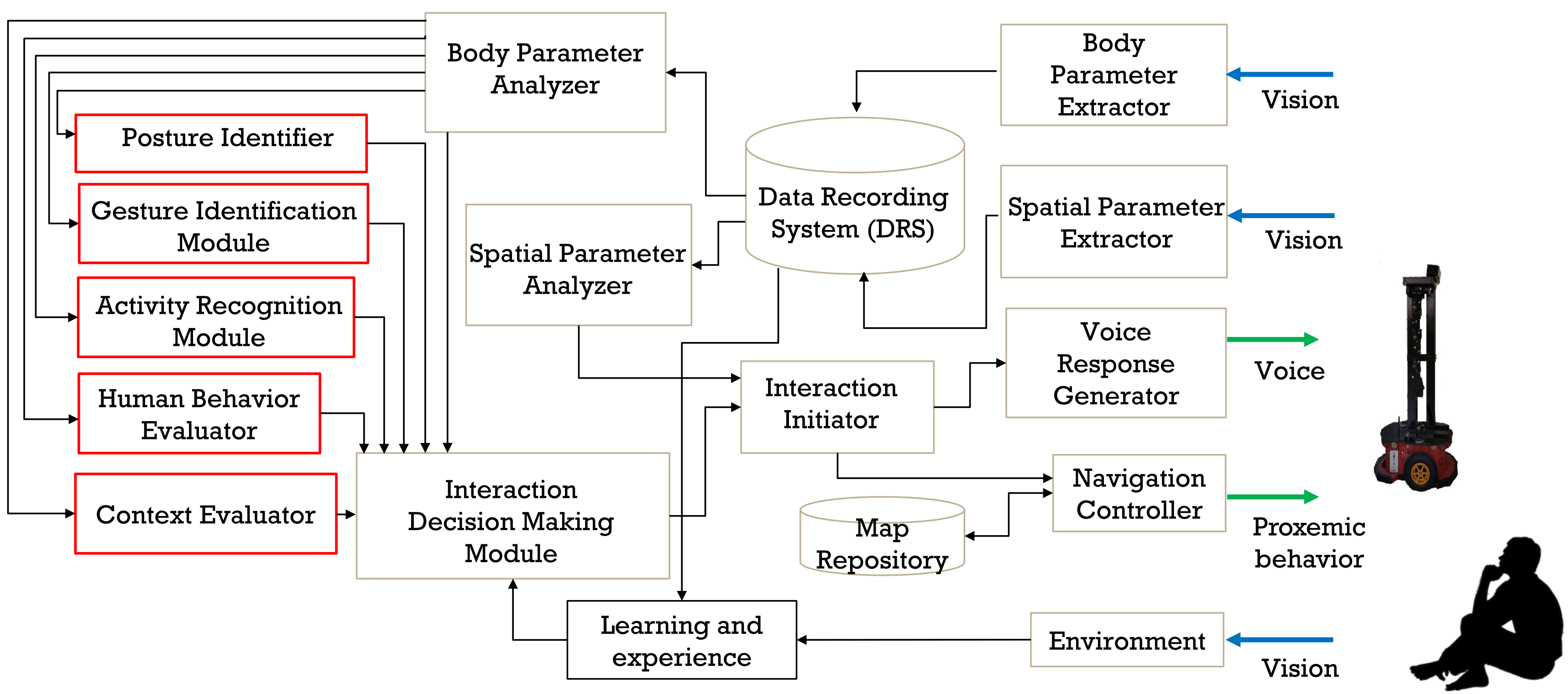


Figure 3.2: Overall system

from the RGB and depth data. Similarly, Gesture Identification Module recognize a predefined set of hand gestures such as friendly and diectic gestures which are often utilized in an interaction. The Activity Recognition Module recognizes daily encountered human activities which play an important role in determining a person's current state and decide the level of engagement of that person. Unclarifiable human behaviors such as activity-related movements and meaningless or random movements are monitored by the Human Behavior Evaluator. Context Evaluator examines other parameters in the context such as the activity space of the subject (user) and verbal responses from the user, if the user speaks first. Features recognized by these modules are fed into the 'Interaction Decision Making Module (IDMM)'. All the cognitive models which map observations and their semantics to a meaningful interpretation of the context are included in the IDMM. Information stored in the DRS for the present scenario are sent to the 'Learning and Experience' module to create a memory for the robot in order to use the current experience and adapt its responses to the context accordingly, in the future. Appropriate responses of the robot, as determined by the IDMM are converted into actions by the 'Interaction Initiator (II)'.

Two major responses of a robot were considered within the scope of this thesis. They are **verbal responses** and **proxemics**. 'Verbal Responses' include the utterances made by the robot and 'proxemics' includes mutual distancing and orientation between the robot and its user. Utterances are generated by the 'Voice Response Generator' based on predefined keywords and phrases. 'Navigation Controller' takes care of the low level implementation of proxemic decisions. Maps required for navigation are stored in the 'Map Repository' and are accessed by the robot to move within the context and generate appropriate proxemic behavior.

Chapter 4

STUDY OF HUMAN BEHAVIOR

4.1 Study 1: Human Interest towards Robot Initiated Human-Robot Interaction

We conducted a wizard-of-oz study to explore both verbal and nonverbal human responses towards the presence of a service robot. We further investigated the responses of humans towards an interaction initiated by the robot itself. Behavioral changes were used as a mediator to perceive user situation and the level of interest in engaging in the interaction initiated by the robot. Responses recorded in the study can be used to improve a robot's perception of human behavior in order to be a successful companion to the human (the user), without violating user expectations.

This human study is intended to explore human tendencies toward an interaction in different situations. Hence, it will provide means of engraving social intelligence into robot behavior before utilizing them in social environments.

4.1.1 Monitoring Human Behavior

The experiment was conducted in a simulated social environment within the laboratory. Participants of this experiment were students and outsiders in the age range 23 to 60 (Mean-31.59, SD-10.53). There were 26 participants and all the participants were in good mental and physical health condition to make decisions. The effects of gender and the age of the user were not considered within the scope of this study. Before the experiment, participants were given instructions regarding the process and the behavior of the robots in the laboratory upon arrival. Users were allowed to engage in a selected task or just relax as part of the experiment. They were made aware of the presence of the robot but they were unaware that the interaction with the robot was a part of the experiment. The experiment was conducted using MiRob shown in Fig. 3.1.

This experiment was conducted to monitor the behavioral responses of users towards robot-initiated interactions. Since the responses depend on the priority given by the user to his/her current activity, we have selected two tasks: reading an instruction sheet and sitting relaxed. It is assumed that reading is a task with high priority while sitting relaxed is of low priority. Participants were asked to perform one task at once. Before the experiment, the participant was aware of the robot but he/she was not aware that the robot is going to interact with him/her. Robot was remotely navigated towards a particular participant while he/she is engaged in the task. The robot initiates a conversation after approaching a particular user. User responses were recorded from the moment robot starts approaching the user. Each user was allowed to perform both the tasks mentioned before. Path planning and navigation tasks of the robot were autonomous while response generation and tracking a user were teleoperated by the experimenter. Hence the teleoperator instructs the robot where to approach and what to say. A single user participated in an experiment and the interaction process was repeated for each participant separately. Furthermore a single participant was asked to perform both the tasks at separate times.

At first, the user was asked to relax and his/her responses towards an interaction initiated by another human were recorded. Again, responses towards the human when the user was engaged in the given task were recorded. This human-human interaction scenario was examined with human-robot scenario in order to be compared. In both the occasions, the external party initiated the conversation. Same participants were selected to study the human-human scenario as well. In human-human occasion, one of the experimenter approached the participant (user) in order to initiate the interaction process.

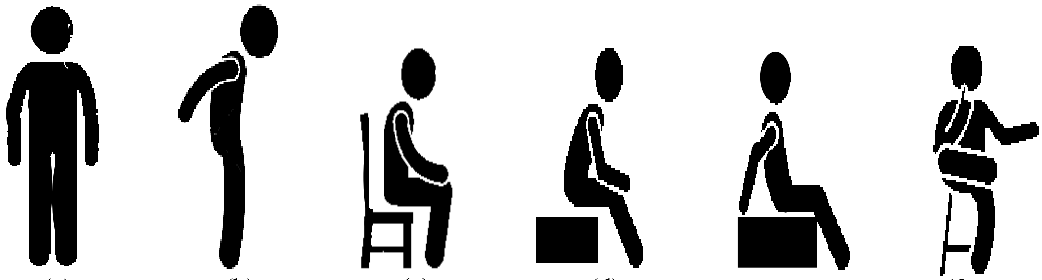
As MIRob was monitored by a human operator, its responses were generated in correspondence with the responses from the user. During the experiment, responses from the robot include approaching and verbal cues only. User responses were recorded in the form of a video using a camera fixed in the experiment area and later analyzed for behavioral changes manually. If the user responses demonstrate no interest towards an interaction, such as looking at the robot when he/she notices the robot, but again looks away and engage in his/her current task, robot was instructed to leave the situation without interaction. In such a situation, it is assumed that the user does not prefer to interact.

Independent variables used in the study are the two tasks of the user. The response towards the interaction initiated by the robot is the dependent variable during analysis stage. Furthermore, it is assumed for comparison purposes that there is no significant difference between the groups that are used in the experiment.

The existence of mutual gaze was identified by the 'yaw' of participants head, as measured by the Kinect SDK. All the other behavioral changes were identified manually by the operator. For the ease of analysis and to avoid missing any behavioral changes, scenarios were recorded using the external camera mounted on a wall. Gestures considered during the study were 'waving hand', 'open palm' and 'calling towards the user'. Poses considered during the study are given in Fig. 4.1 and two occasions during the experiment are shown in Fig. 4.2.

4.1.2 Results of the experiment

Responses of a human were categorized into five major groups; *Gaze*, *gestures*, *pose*, *verbal responses* and *facial expressions*. Here, the *Gaze* is



(a) (b) (c) (d) (e) (f)

Figure 4.1: (a) - (f) shows the poses that were identified as behavioral changes during the study. (a) and (c) indicate standard standing and sitting postures respectively. Other poses are deviations from standard postures.

(a) (b)

Figure 4.2: An example scenario during the experiment. (a) The user was sitting relaxed (b) User looks at the robot while the robot was approaching her.

the maintenance of mutual gaze at least for a shorter duration. It is used as a parameter to measure the user's attention towards the robot. *Gestures* include meaningful movements in hands. Smile and frowning are the *facial expression* considered considered during the experiment. Out of the 26 participants, the percentage of participants who used each type of behavior is shown in Table 4.1. Percentage number of participants with a certain response, for both the interaction scenarios: human-robot interaction and human-human interaction

Table 4.1: Results of the experiment

	Robot				
	Gaze	**Gestures**	**Pose**	**Verbal responses**	**Emotions**
Relaxed	100	46	12	62	92
Engaged	92	0	15	88	58
	Human				
Relaxed	85	35	19	96	81
Engaged	85	0	27	100	58

Here the number of participants for each occasion is given as a percentage.

under both user situations; when relaxed and when engaged. A single scenario during the experiment is demonstrated in Fig. 4.2.

Fig. 4.3 and Fig. 4.4 graphically present these results. Fig. 4.3 displays the percentage number of participants who adopted each type of behavior while the interaction was initiated by the robot itself. This was compared with the percentage number of participants who followed the same type of behavior while the interaction was initiated by another human, with the same user. Variation of responses based on the interactive partner (the robot) when the user was in a relaxing situation can be seen here. Human behavioral responses toward a human and the robot were compared. Same behavior was observed when the user was busy with a certain task as illustrated in Fig. 4.4. Here, the task was reading some additional information regarding the experiment. This task was selected so that users could be kept engaged with the exact real mindset that they are in the middle of an important task which they are reluctant to stop.

Table 4.2 shows the results of a t-test performed to test the deviation between the user responses for the robot and another human during both the occasions: while relaxed and while engaged. The percentage number of participants displayed each type of behavior was taken as the data for the t-test.

Table 4.2: T-Test for the comparison among each situation: while relaxing and while engaged when the person initiating the interaction was the robot and another human.

vspace1em

Task	T- Scores	Robot	Human
While relaxing	mean	62.31	60.95
	variance	1289.94	1324.49
	dof	4	
	P	0.45	
	t	2.13	
While engaged in a task	mean	50.77	53.85
	variance	1755.92	1678.99
	dof	4	
	P	0.23	
	t	2.13	

4.1.3 Observations and Discussion

From the results displayed in Table 4.1 and Fig. 4.3, behavioral changes observed during HRI and HHI are compared while the user was in a relaxing situation. During this scenario, *gaze* was the type of behavior adopted by almost all the users. In human-robot scenario, all the users (100%) maintained gaze, but in human-human scenario, this dropped to 85% and the highest adopted type of behavior was *verbal responses*.

Maintenance of *gaze* and *expressions* have the least deviation when both the

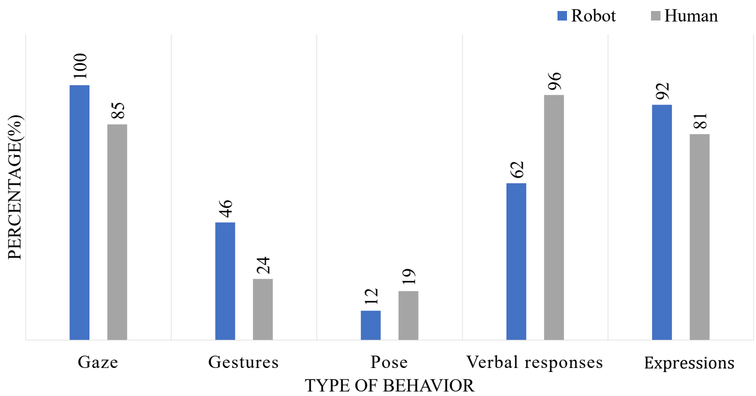


Figure 4.3: A comparison of the types of behavioral changes observed before and just after initiating an interaction by a human and a robot. In this case, the user was in a relaxing situation without engaging in any task.

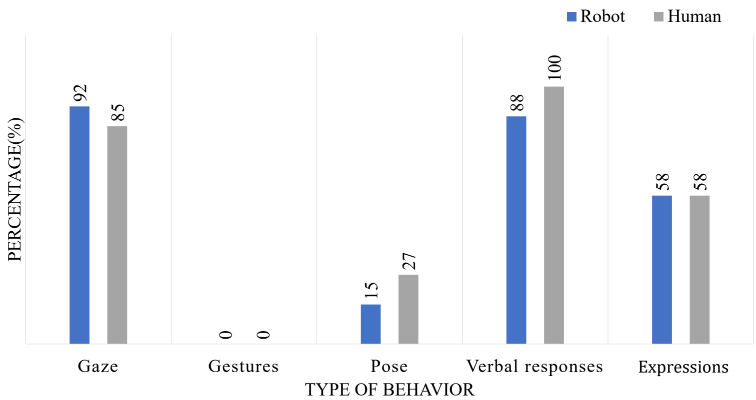


Figure 4.4: A comparison of the types of behavioral changes observed before and just after initiating an interaction by a human and a robot. In this case, the user was engaged in a given task.

situations were considered; human-human and human-robot scenarios. Simplicity and the ease of adoption of these behaviors are likely to be the reason for this. In general, users have tend to adopt behaviors which involve lesser movements. $Gaze$, $verbal\ responses$ and $facial\ expressions$ received the highest percentage during both the scenarios due to this reason. As $gaze$ and $facial\ expressions$ are most likely to be involuntary, the deviation between human-human and human-robot scenarios considering $gaze$ and $expressions$ was less.

When compared human-human situation with human-robot situation, out of the number of expression that humans display, only 'smile' could be observed during the whole process of the experiment. This could be due to the fact that we considered only the interaction initiation process but not the continuation of interaction. When $gestures$ and $expressions$ are considered, 22% have used more gestures and 11% more expressions upon robots during interaction. Hence users have followed a more displaying friendliness towards robots than for humans.

From the results shown in Table 4.1 and Fig. 4.4, the trend has slightly been deviated from the previous case. Although users engaged in a task during this situation, they have tried to respond towards the interaction process. In both human-human and human-robot scenarios, the highest percentage of users adopted $gaze$ and $verbal\ responses$. Adoption of $gestures$ and $expressions$ has dramatically been dropped when the user situation changes from 'relaxed' to 'engaged'. No $gestures$ were involved in support of the interaction when the user was engaged.

From Fig. 4.3 and Fig. 4.4, it can clearly be seen that the least adopted type of behavior was $pose$. Hence it can be deduced that humans are more comfortable with simple behavioral changes such as $gaze$ and $verbal\ responses$. As seen from the charts, $gestures$ and $pose$ are least preferred behavioral changes. Possible reasons for this are the complexity of such behaviors and the laziness within humans to break their inertia.

As seen from the study, on the one hand when the users were relaxed, robots received more attention and friendlier responses during interaction initiation. On the other hand, it was also robots which received higher negligence when users were engaged or busy.

Most important and unexpected patterns in user behavior were demonstrated from the t-test shown in Table 4.2. In both user situations; when relaxed and when engaged, $p>0.05$. In this case $p= 0.45$ and $p= 0.23$. Hence the significance of the effect in the both the user situations becomes of interest. In both these situations, null hypothesis cannot be rejected. Hence it can be considered that there is no significant difference between the user response toward a robot and a human despite the user situation. It can be said that, as a whole humans tend to respond towards social robots just in the manner they respond other humans.

Although most prominently displayed human behaviors were used for the study, only a few of them were dominant in the context of interaction initiation according to these observations. From Fig. 4.3 and 4.4, *gaze*, *verbal responses* and *expressions* were the frequently adopted by many users. *Verbal responses* were limited to smaller clauses or sentences such as 'hello', 'hi' or other types of greetings and openings. The sentences became even smaller when the user was engaged or busy.

According to the results of the t-test, there was no significant difference in the responses towards the robot and the human in both situations. Therefore, it can be seen that humans preferred robot-initiated interaction whenever they were ready to interact. In this case, the user readiness (to which extent the user is ready to interact) was defined by the task of that particular user; where a relaxed user was ready to interact and an engaged user was not. In a typical domestic scenario, there can be other factors which affect user readiness. Furthermore, there was a tendency in users to treat robots in the same way as they treat other humans. Also, users were not reluctant to neglect the robot when busy. Therefore an open minded behavior was observed from users in the presence of robot. Another interesting fact revealed during the study was that users preferred gestures and expressions upon robot rather than for the human. This was more similar to human responses towards a child. Results of this study suggest that the trends in human behavior regarding interaction initiation are different from the trends that are observed during an interaction process. Therefore there is a requirement of emotional intelligence in social robots to perceive this difference.

4.2 Study 2: Conversational Preferences in Social Human-Robot Interaction

Likeliness of the robot being accepted as a conversational partner depends on the environment as well as the current task of the user. We investigate how the nature of interaction initiated by the robot affects its acceptance by a human. Subsequently, a set of affect based types of conversation were selected and implemented on a service robot in a simulated social environment in which few users were present or only a single user was present. This study is intended to find human tendencies towards interaction in different situations and hence will provide means of engraving social skills into robot behavior before utilizing in human environments. We conducted a wizard-of-oz study to explore the nature of human conversation with the presence of a service robot when several factors in the environment vary. Factors which are more likely to have an influence upon the conversational preference of humans are selected. The type of interaction preferred by the user was used as a mediator to perceive user situation and the level of interest towards the interaction initiated by the robot. Responses observed during experiments were used to upgrade existing robots' perception of human behavior. Furthermore we intend to lay a justifiable basis to bring several such observable factors that can be used by a robot to evaluate an encounter. We considered observable cues from humans as well as their surrounding, which were likely to have an impact towards responses generated by a human during a certain scenario.

4.2.1 Setting

The experiment was conducted in a simulated social environment in the laboratory. Participants were students, non academic staff members of the university and some outsiders in the age range 19-58 (Mean-28.45, SD-9.02) who volunteered the study. There were 37 participants and they were in good health condition without any physical defects which will alter their reactions during the study. More than half of the participants did not have a technical background in education, majors or research related to engineering. The gender of the user was not included within the scope of this study. Upon arrival, the users were given instructions regarding the tasks they should complete but they were not aware of the fact that they are intended to talk to the robot but they are instructed to respond towards the robot if the robot initiates an interaction. They were not knowledgeable about the exact intention of the experiment because that will cause a bias response from users towards the robot. Hence the participants were instructed to perform a given activity in the way they are used to perform that before. The experiment was conducted with MIRob.

The selected user was allowed to engage in a certain task and the robot was allowed to approach the user to initiate a conversation with him/her. The user was advised to complete a certain task and if the robot talks to him/her, to talk back. The set of tasks to be performed by the users was predefined. There were separate lists of tasks to be performed in living room, bed room and kitchen. The participant or the user was knowledgeable on the tasks that are to be performed in each living area. The tasks were selected so that at least three tasks were performed in each area. These tasks are the most common to that particular social or domestic environment and few tasks selected for the study are listed in Table 4.3. While the user was engaged in a task, the robot was remotely guided towards him/her and was allowed to initiate an interaction in the ways given

Table 4.3: Some of the tasks selected for the study

Living Area	**Task**
Living room	Resting while sitting
	Reading while sitting
	Having a snack
	Watching television
	Engaged in a desk activity
	Engaged in a conversation
Bed room	Tidying up
	Resting
	Engaged in a desk activity
Kitchen	Cleaning
	Preparing a meal
	Having breakfast

below. The map of the environment was predefined in the simulation. Therefore the robot navigated to the target positions and its orientation which were defined by the operator. In this scenario, the target position of the robot was a point within the interactive area near the user. For the ease of future referencing, these types of conversational preferences are abbreviated as follows.

NI - No interaction

GRT - Greeting

SER - Asking to deliver a service

TLK - Small talk

CON - Long conversation

How a conversation is categorized into these types are shown in Fig. 4.5. As the conversation extends, the type of conversation shifts from NI to a CON. The robot will not talk to its user during NI. In GR, the robot only greets the person and goes away. The greeting will just be a single sentence saying ‘good morning’, ‘hey’, ‘hello’, etc. In SER, the robot asks to deliver something for the user, as an assistance to his/her current task. This will be approximately

Conversation begins

R: Good Morning!
H: A very good morning!

Greeting

R: You want me to get something for you?
H: No, Thank you.

Asking to deliver a service

R: Okay. Are you tired?
H: No, I'm just relaxing.
R: Okay.

Small talk

R: Is there something you want me to talk about?
H: Hmm, There's nothing specific, but I'd like a conversation.

R: May I play a song?
H: I would rather hear a poem.

R: Good choice. A poem about nature?
H: That'll be great. Go on.

.
.
.
.

R: Have a nice day!
H: You too!

Long Conversation

Conversation ends

Figure 4.5: The nature and the length of conversations determine the 'type of conversation' existed at a certain occasion.

four sentences maximum in the entire conversation. In the TLK, the robot will say a few additional sentences other than greeting and sometimes will ask if the user wants something. Such a conversation consisted of about 5-7 sentences. All the conversations longer than that were considered as CON. Such conversations covers a broader scope of topics as well as these existed for a longer duration. Therefore the duration of the conversation depended on the type of conversation robot had with a user. In all the occasions, the robot stopped continuing the conversation depending on the curiosity of the user to engage with the robot or when the conversation seems to disturb the user. During the experiment, robot had a CON with its user and in the end a survey was conducted to know the actual preference of the user. Users were shown how the type of conversations are categorized and were asked to select their preference at that particular occasion despite the conversation he/she already had with the robot.

The robot initiated a conversation despite the task of the user and user responses towards that interaction were recorded. Voice responses were monitored remotely by an operator without the knowledge of the user. Furthermore, a single participant was asked to perform all the tasks listed in the experiment separately in different occasions. Each task was performed twice; when the user was alone and when few others were present. The second occasion replicates a typical domestic or a social environment in which family members or few other known persons were present around. In the experiment, participants in a single setting were acquaintances. For instance, if there were few people around the user at the time of the conversation, all these people were acquaintances but were not related to each other.

MIRob was remotely controlled by a human operator. Robot responses were generated with respect to the user response in each occasion. It is expected to assess the effect of considered factors in the surrounding upon human conversational preferences, in both qualitative and quantitative manners. An important fact to be considered in this case was that when few other people were present around, they were not involved in the interaction process except when the user was having a conversation with them. Voice responses were generated after the robot approached the user. Path planning and navigation of the robot were autonomous while tracking of the user and generation of voice responses were teleoperated by the human operator. Therefore the operator instructed the robot where to approach and what to speak. None of the persons in the environment participated in the conversation with the robot, except for the intended participant. The responses of the robot during the experiment include only maintaining a socially interactive distance between the robot and the user, and voice. If any of the users does not respond the robot, the robot was instructed to leave without causing any distraction. In such a situation, the robot assumed that the user does not prefer to interact.

Independent variables used in the experiment were the task of the user, domestic area and the presence of others in the surrounding. The conversational preference was the dependent variable during analysis. The assumption made during the study was that there is no significant difference between the groups used for comparison purposes.

4.2.2 Results and discussion

After the experiment, the conversational preferences of users were analyzed using statistical methods. The first question of interest was whether there is a difference in user responses depending on the number of people in the surrounding. Table 4.4 shows a comparison of the percentage frequency of each type of interaction for the two occasions; when the user was alone and when few people were around. This study was intended for all the tasks listed in Table 4.3. As seen from the results, there is an increase in demanding a service or limiting the conversation just for a greeting when few other people were present in the surrounding. As seen from this information, the demand for interaction types NI,

Table 4.4: A comparison of conversational preferences by the type of interaction when the user was alone and when with few people around

	Alone	**With people around**
NI, GRT, SER	64%	79%
TLK, CON	36%	21%

GRT, SER have been increased by 15% when the number of people around the user has increased from zero to a few. An example scenario from the experiment is given in Fig. 4.6 when the user was alone. A scenario when there were people around is shown in Fig. 4.7. In this situation the user preferred a long conversation when she was alone and a service when there was a second person in the kitchen. Such behavioral changes were recorded during the experiment.

Table 4.4 shows an ANOVA test performed on the same data for the comparison of percentage frequencies of each type of conversational preference in each area of the social environment. The test was performed to analyze how the tendency towards each type of conversational preference changes when the domestic area changes. Percentage usage of the types of interaction are calculated and compared. First test was implemented for the case when only the user was alone in the considered environment and the second test for the case when few

(a) (b) (c)

Figure 4.6: An example scenario during the experiment is shown. (a) The user was in the kitchen, having a drink (b) The robot approached user and initiated a conversation (c) Interaction continued.

Figure 4.7: A situation in which the user was having a drink and in the surrounding there was another human without an interaction with the user. Robot approached the user and initiated a conversation.

others were present in the surrounding, in addition to the user and the robot. From the test, it was intended to find the differences in conversational preferences when condition of the surrounding with regard to the peer (whether the user was alone or there were few others around) was kept constant. Furthermore it was expected to find whether there is a change in user behavior upon where the user is, despite whether he/she is alone or with few people around.

Table 4.5 shows the results of a t-test performed to test the deviation between the preference of each type of interaction when alone and when surrounded by a few. Changes in demand for each type of interaction in the said two occasions were analyzed without an involvement of other types of conversational preferences. Frequency of the type of interaction in each domestic area was taken as data for the t-test.

Shown in Table 4.5 are two ANOVA tests performed on the same set of data to test the deviation between the conversational preferences during the list of selected tasks while the user was alone and with one/few people around. The frequency of using each conversational preference during these tasks was

Table 4.5: ANOVA test for the comparison of percentage frequencies of each interaction type in each area of the social environment

	Alone	
	Mean	**Variance**
Living room	20	107.39
Bedroom	20	127.99
Kitchen	20	286.66

ANOVA Test

	SS	**DOF**	**F**	**p-value**
Between groups	0	2	0	1
Within group	2088.22	12		
Total	2088.22	14		

	With people around	
	Mean	**Variance**
Living room	20	116.73
Bedroom	20	309.39
Kitchen	20	263.53

ANOVA Test

	SS	**DOF**	**F**	**p-value**
Between groups	0.1333	2	0.00029	0.9997
Within group	2756.8	12		
Total	2756.93	14		

calculated and analyzed for the deviations between each group. Here the groups were the conversational preferences from NI to CON and the frequencies were listed according to the tasks listed in Table 4.3. In both the situations in Table 4.5: when alone and when surrounded by few people, the F-critical value was 2.539.

4.2.3 Observations and Discussion

From the results displayed in Table 4.5, it can be seen that the tendency of the user towards a friendly interaction reduced when there were people around. Even though these people were not directly involved with him/her, their presence influenced the reactions of the user towards robot. A perceived behavioral control could be observed within the user due to such changes in the surrounding. From these results, behavioral changes observed when the user was alone and when few people were around were analyzed separately. From the first ANOVA test, a p-value of 1 ($\geq$ 0.05) and an F-value of NI could be observed for comparing conversational preferences within each social area; living room, bed room and the kitchen. Therefore, the fact that 'there is a significant difference in the

Table 4.6: t- Test for the comparison among each type of interaction when the user was alone and with few people around

Type of interaction	T-scores	Alone	With people
NI	mean	20.67	22.33
	variance	210.33	16.33
	dof	2	
	P	0.852	
	t	4.302	
GRT	mean	8	12.67
	variance	7	16.33
	P	0.034	
	t	4.302	
SER	mean	35.33	44.33
	variance	82.33	54.33
	P	0.046	
	t	4.302	
TLK	mean	21.67	11.67
	variance	100.33	9.33
	P	0.131	
	t	4.302	
CON	mean	14	8.67
	variance	9	8.33
	P	0.246	
	t	4.302	

conversational preferences with the social area when the user was performing a task alone' cannot be accepted. In the same way, from the second ANOVA test in Table 4.5; when few people were around, p-value of 0.9997 ($\sim$1) and an F-value of 0.00029 ($\sim$0). Hence the fact that 'there is a significant difference in conversational preferences when the user was surrounded by a few people in the surrounding' also cannot be accepted. From the two tests, we could observe that there is no significant effect of the type of living area upon conversational preference of a particular user but his task.

Most important and unexpected patterns in user behavior were demonstrated from the t-test shown in Table 4.6. For all the conversational preferences except GRT and SER, p>0.05. Hence the significance of the effect in the cases 1 and 2 becomes of interest. A probable reason for this is that, in almost all the occasions NI was preferred, the user gave prominence to the task despite how many people were around. In such situations, the user gave prominence for relaxation by means of conversation, rather than the task. An example was when the user was in a phone call or a desk activity. In such a situation, user will not prefer to be interacted. As a whole, conversational preferences at the two ends; NI and 'having a friendly conversation' (TLK and CON) had no influence from the living area but middle interaction types (GRT and SER) had. For GRT and SER, where p=0.034 and p=0.046, the null hypothesis could not be accepted. Hence it can be concluded that there exists a significant difference in the conversational preference for GRT and SER, when the living area changes.

Significant rises and drops in conversational preferences were observed with the change in the number of people around. This is demonstrated in the Fig. 4.8. People became more introvert with the presence of other humans. The expectancy of service increased when there were few people around. Therefore, a significant increase for SER was observed when the user situation changed from 'alone' to 'with few people around'. As TLK and CON are rather friendlier types of interaction, these were preferred by the users mostly when they were alone. The

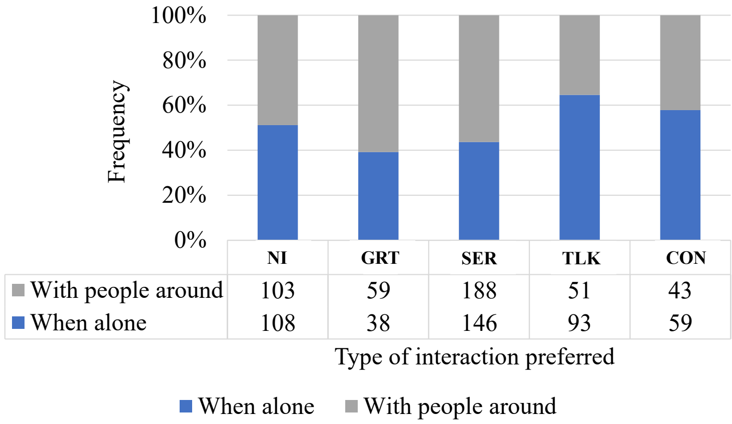


Figure 4.8: A stacked graph drawn for the comparison of conversational preferences with the two conditions; when the user is alone and when surrounded by few people. The type of interaction is plotted against the frequency of each type of interaction preferred in above two the occasions.

highest percentage difference for these two occasions was observed in TLK. A possible reason for this is that TLK is the most flexible type of interaction which a user can have without getting disturbed to his/her task.

According to the two ANOVA tests in Table 4.7, in both the cases; when the user was alone and was with one/few people around, F-values (4.442, 21.979) were larger than F-critical (2.539). Hence in both these cases the null hypothesis can be rejected. Hence the assumption that 'there is no significant difference between each type of conversational preference during the selected set of tasks' was declined. Therefore it can be deduced that the preferences for NI to CON were significantly different when the given tasks were considered. Furthermore in both the occasions, p-values (0.035, 7.053E-11) were smaller than the alpha variable (0.05). This also suggests that the individual variables were statistically significant. During 'with people' situation, the F-value (21.979) was significantly larger than the F-critical (2.539). Hence the joint effect of all the variables together is larger than that when the user was 'alone'.

Another fact observed during the study was that the existence of a significant difference in conversational preferences based on the task. This is examined in

Table 4.7: ANOVA test for the comparison of the frequencies of each conversational preference during each task

	Alone	
Groups	**Mean**	**Variance**
NI	24.32	516.63
GRT	8.56	27.67
SER	32.88	317.20
TLK	20.95	256.49
CON	13.29	102.87

ANOVA Test

	SS	**DOF**	**F**	**p-value**
Between groups	4338.20	4	4.442	0.0035
Within groups	13429.51	55		
Total	17767.71	59		

	With people around	
Groups	**Mean**	**Variance**
NI	23.20	162.64
GRT	13.29	65.69
SER	42.34	193.02
TLK	11.49	25.40
CON	9.68	53.73

ANOVA Test

	SS	**DOF**	**F**	**p-value**
Between groups	8800.10	4	21.979	7.053E-11
Within groups	5505.23	55		
Total	14305.33	59		

the chart in Fig. 4.9. In Fig. 4.9, the frequency of the users who used each type of interaction is plotted against each type of interaction while the user was engaged in the selected task. In all the occasions, the domestic area was the living room and the user was alone in the environment. As seen from the chart, there were significant differences in user's conversational preferences when their task changed. For example, few users have chosen NI while resting but many users have chosen NI while making a phone call. As a whole, there was a considerable variation in conversational preferences in the six tasks considered here. People were comfortable with only certain types of interaction when most tasks were considered. This fact was confirmed by the results shown in Fig. 4.9.

4.2.4 Conclusions and Implications

In this study, the conversational preference was used as a major contributer to perceive human interest and attention towards the robot while some factors

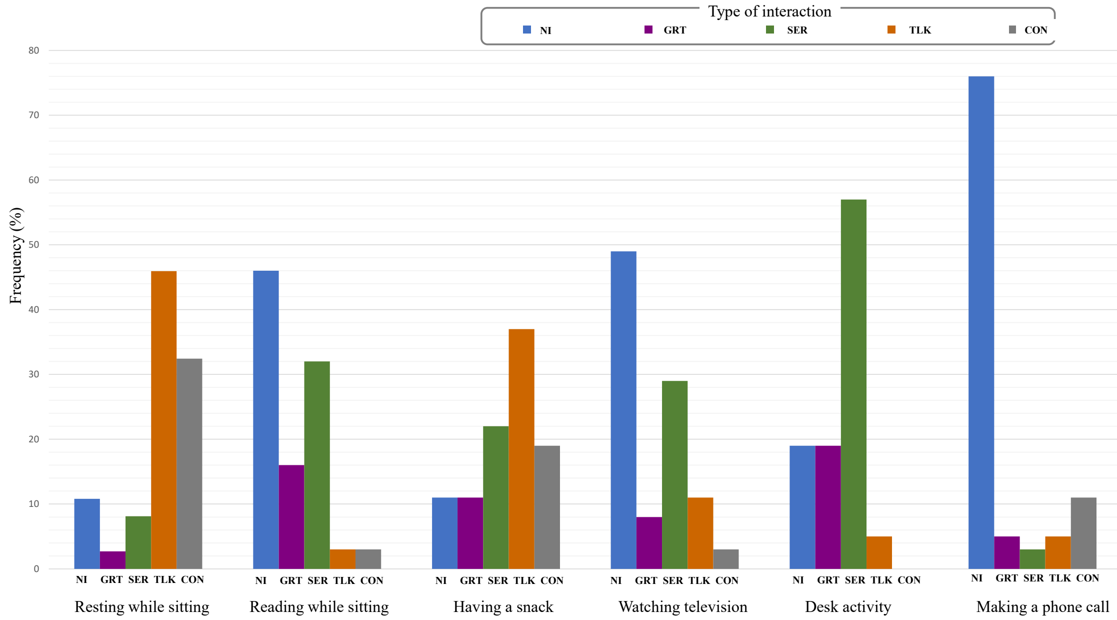


Figure 4.9: This graph depicts how the users picked up conversational preferences during the selected tasks while the domestic area and people in the surrounding were kept constant. Here, the domestic area was the living room and the user was alone in the area.

in the environment or factors within the user change. As we have studied the behavior of humans among acquaintances, their responses will be friendlier in the presence of family or relatives (e.g: a domestic environment) and less friendly in the presence of strangers (e.g: a public space). Therefore this study can be used to find tendencies of humans in general and to derive those in other occasions.

The experiment was intended to reveal the relationship between surrounding factors and conversational preferences of humans. Interesting facts regarding conversational preferences based on the changes related to the user and the surrounding were revealed during the analysis of data. The findings of the study are expected to be used to rebuild modern interaction mechanisms among humans and robots, so that the two conversant are motivated towards a sustaining conversation. Results show that there are considerable effects of factors in the surrounding and the user, upon the conversational preference of a user at that particular time. Moreover, despite age differences, these factors have become prominent in deciding conversational preferences during a particular moment.

Furthermore these findings can be made useful in developing adaptive robotics systems which are expected to be used in social environments.

4.3 Study 3: Human Gaze Behavior

From the existing literature, we could observe that many robotic systems developed so far have considered 'gaze' as a parameter to evaluate a certain situation between a human and a robot. Therefore we tried to find the role of 'gaze' during an interaction to find how important it is to evaluate this feature and what other facts can be explored by analyzing gaze.

During this study, we tried to find answers to the following questions.

1) How humans use gaze-related features as actions during human-human interactions?

2) How similar features were associated with human-robot interactions?

3) How similar are HHI and HRI in terms of gaze behavior?

For that we conducted a series of experiments with dyads of humans, and robots and humans to find patterns in their gaze behavior and hence find answers to the above questions.

4.3.1 Understanding Human Behavior: Evaluate Attention

4.3.2 Time scales of human actions

In this study, we have selected 'gaze' behavior which is also a cognitive act. Gaze can sometimes be an automatic behavior and sometimes, a controlled behavior [87] where the reaction time may take up to about 1.3 s. Alongside this, three parameters involved in gaze: response time, duration, averted/not were analyzed during this study. These occur at different time scales when initiating an interaction. Fig. 4.10, the *delay* represents the time taken by the responder to react, once a 'change in situation' was detected. When considering gaze, the 'change in situation' was the presence of an outsider who seeks attention. *Duration* of gaze relates to the entire time the responder looks at the approacher. *Averted gaze* is when the responder returns his gaze back at his previous task. These features associated with gaze can say a lot to an outsider about the inner intents and priorities of the responder. Hence these three features of gaze during human-human interaction (HHI) have been investigated thoroughly in our experiments.

4.3.3 Monitoring Gaze Behavior

The experiment was conducted in study areas and laboratories inside the university with the participation of 47 graduate students, postgraduate students, and the members of non-academic staff in an age range of 25-51 years (mean of 38.6 and SD of 11.31). Nearly half of the participants have not had a

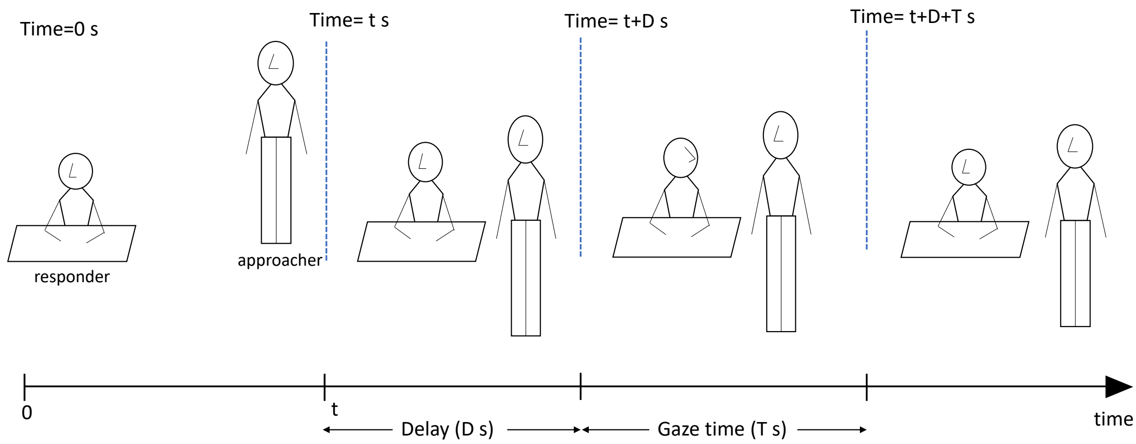


Figure 4.10: Phases in a typical encounter prior to an interaction are shown. At time= 0, the approacher intends to approach the target person: responder. At t=t s, he approaches the responder. The responder notices his presence, but takes another D s to give attention. The responder looks at him for a duration of T s and averts his gaze to continue his work.

strong technical background. The experiment was conducted in 2 stages: HHI (Human-human interaction) and HRI (Human-robot interaction).

Initially participants were asked to be seated and read guidelines of the experiments in the form of a document opened on a laptop on the table. While each participant was reading, a person previously known to them, but not a relative, approached him/her. The participant's reaction to the approaching person were recorded. These observations were analyzed under stage 1: HHI. The same procedure was repeated in stage 2. But this time, the approacher was a robot. Stage 2 was conducted once a week over a period of 3 weeks to give the participants some time to adapt to social robots. We terminated the set of experiments once we noticed several trends in HRI which were similar to HHI.

Stage 2: HRI was conducted over 3 weeks as 3 occasions. These occasions are HRI-week 1, HRI-week 2 and HRI-week 3 in which the experiment was conducted in the first, second and third week respectively. The reason to conduct only stage 2 along 3 weeks is that human-robot interaction process is not normalized as human-human interaction even at present. Therefore we gave participants time to adapt to the human-robot collaborative environments. Therefore we repeated the same experiment thrice within 3 consecutive weeks. The only difference in stage 2 was that the approacher was the robot, wherein the approacher was another human in stage 1.

Therefore, in the end there were 4 occasions considered as follows. HHI- Both the approacher and the participant were humans. HRI (HRI-week 1, HRI-week 2, HRI-week 2)- The interaction took place between human participant and robot. In all 4 occasions, the interaction was dyadic. We kept people in the surrounding who were not involved in the interaction scenario away as much possible as, human responses might have an influence from outsiders. The interaction scenario was stopped before a conversation between the two parties begins. In all the occasions, the parameters associated with the gaze were recorded using a video camera set up in front of the setting. At a later stage, this video was analyzed to extract required features during the interaction. Two example scenarios from HHI and HRI from the experiment are shown in Fig. 4.11.

4.3.4 Results of the experiments and Discussion

We analyzed whether gaze existed or not in a scenario and if existed, we recorded parameters or properties of existed gaze behavior. In the end, we looked for the existence of other responses towards an interaction between two participants in addition to gaze. An in-depth analysis of our observations is given in the following paragraphs.

Several labels have been used for the ease of referring parameters related to gaze. These are mentioned below.

Continued gaze - The user looked at the approacher and continued looking

(a) (b)

Figure 4.11: Two scenarios during the experiment are shown. (a) HHI scenario is shown. Both the responder and approacher are humans. (b) An HRI scenario is shown. The approacher is a robot while the responder is human. In both the occasions, the responses generated by the responder after noticing the presence of the approacher were recorded.

until the interaction ends

Averted gaze - The user looked at the approacher and averted gaze after a while.

Delay - The time taken by the participant to look at the approacher after the person approached him/her.

Duration - The duration of existed gaze before the participant averted gaze.

0 and 1 in graphs denote non-existence and existence of a considered parameter respectively.

Fig. 4.12 shows the presence of gaze parameters during HHI. Even when at work, more than half of the participants (55%) maintained gaze with the approacher. That is many participants signaled that they give attention to the approacher. Even so more than half of them (54%) averted gaze and gave their attention back to work. However it is interesting that most individuals tried to be interactive and not to be disturbed at the same time.

The existence of continued/averted gaze as a percentage in each of the 4

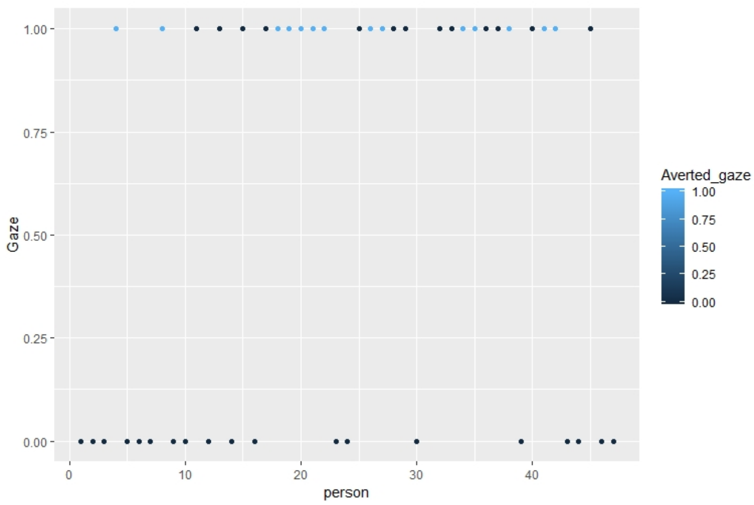
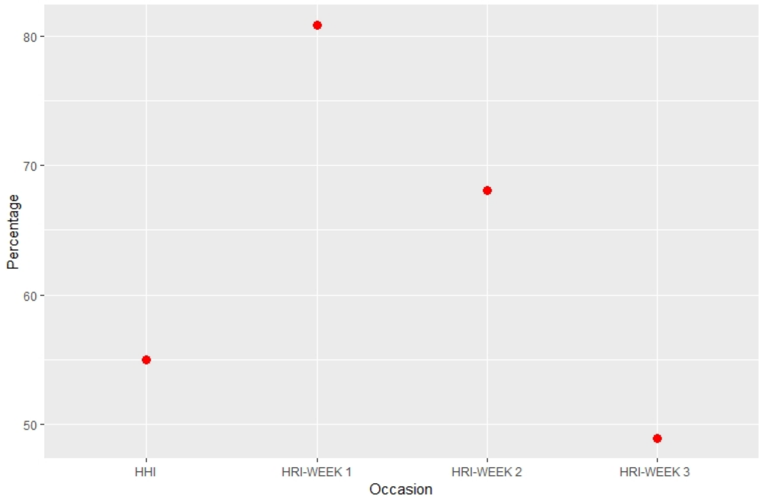


Figure 4.12: Existence or nonexistence of gaze during HHI is plotted against each user. If the user averted his/her gaze, such gazes are marked in light blue while continued gazes are marked in dark blue.

occasions is shown in Fig. 4.13.

Considering scenarios with averted gaze, the duration of gaze is plotted against the average delay of the gaze for the four occasions in Fig. 4.14. From this, it can be seen that, when humans are engaged robots received much higher negligence from humans. Similar to the properties of averted gaze in Fig. 4.14, Fig. 4.15 shows the properties of continued gaze.



[!b]

Figure 4.13: The percentage (%) of encounters with any kind of gaze (continued or averted), out of the total encounters considered, is plotted against the occasion: HHI, HRI-week 1, HRI-week 2 and HRI-week 3.

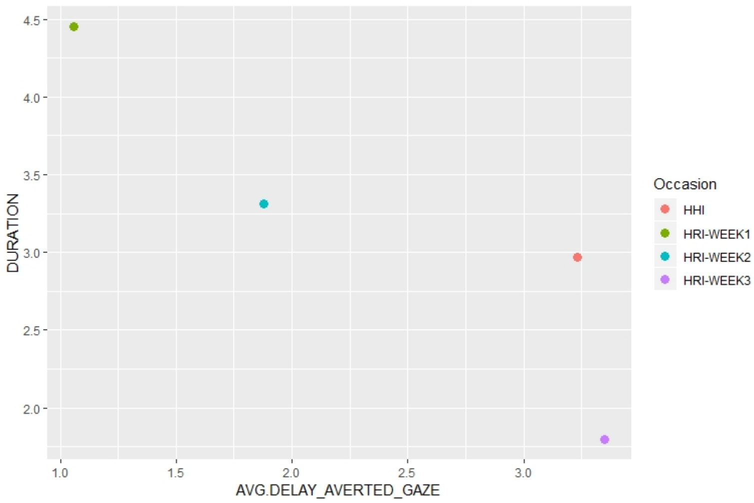


Figure 4.14: Of the total encounters where averted gaze could be observed, the average duration of existed gaze was plotted against the delay. All the units are in seconds (s).

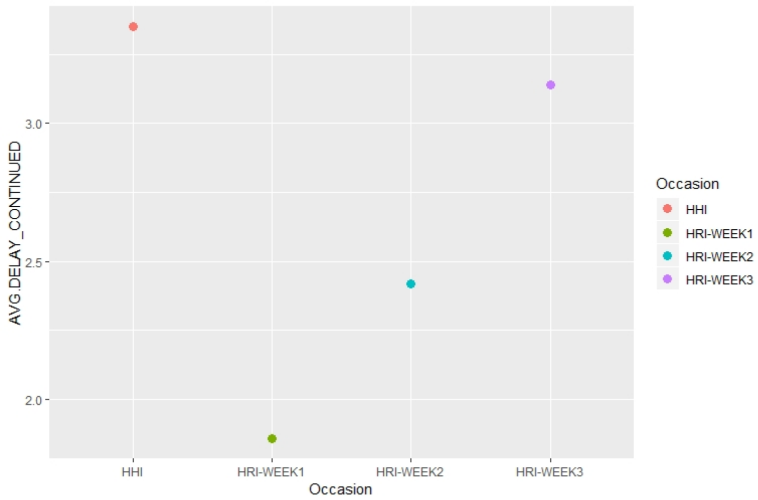


Figure 4.15: Of the total encounters where continued gaze could be observed, the average delay of gaze was plotted against the occasion: HHI, HRI-week 1, HRI-week 2 and HRI-week 3. The average delay was measured in seconds (s).

In continued gaze, HHI and HRI-week 3 recorded only a slight numerical difference.

The delays of averted gaze in the four occasions are plotted against the occasion in Fig. 4.16, with the markers scaled according to the duration of each gaze. It

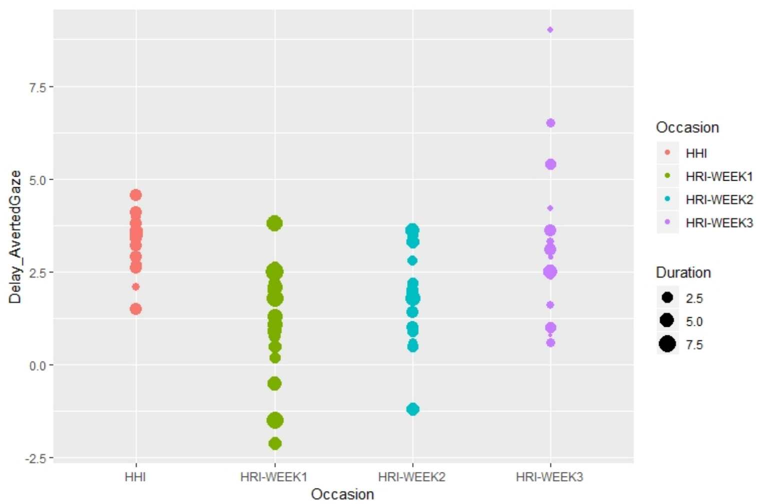


Figure 4.16: The delays of encounters where an averted gaze was present. The box plots are drawn to identify the behaviors of parameters and their evolution with time and experience. The markers are drawn to the scale of the duration of gaze.

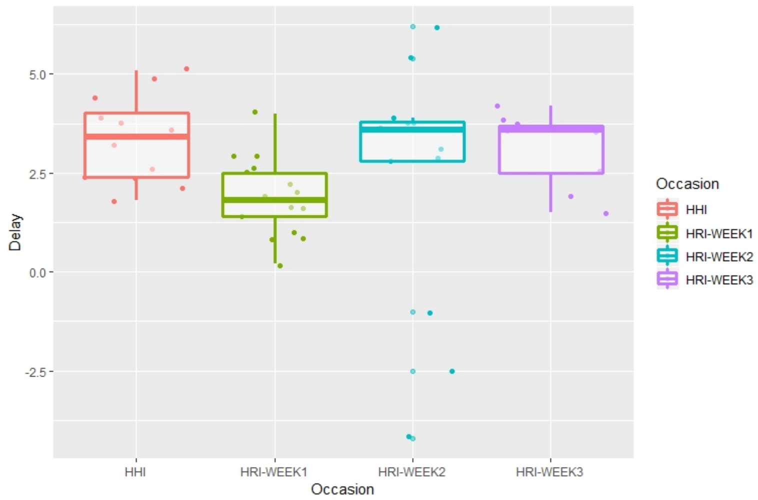


Figure 4.17: The delays of encounters where a continued gaze was present. The box plots are drawn to identify the behaviors of parameters and their evolution with time and experience.

can be observed that most delays accumulated between 0 to 3.75 s. The durations of the gaze lied within this range also recorded an average value in general. If the two outliers are excluded, the trends in HHI and HRI-week 3 were similar to each other.

Fig. 4.17 illustrates boxplots showing the gaze delays in encounters where a continued gaze was present for the four occasions. When the average delay of the four occasions were considered, HRI-week 1 recorded the least average delay in continued gaze. This value gradually increased by week 3 and, HHI and HRI-week 3 resulted in close average delays in continued gaze. The trend in HHI and HRI in terms of gaze delay, at the end of the 3 weeks. A possible reason for the scatter of data in week 2 is that people are getting familiar with social robots in their environments after week 2. The scatter of data was again normalized in week 3. This can be the reason to observe negative delays in gaze as well. The unfamiliarity of humans with robots resulted in giving much more attention to robot's behavior when they approach.

Fig. 4.18 plots several boxplots to identify the behavior of gaze delay of encounters where a continuous gaze was present.

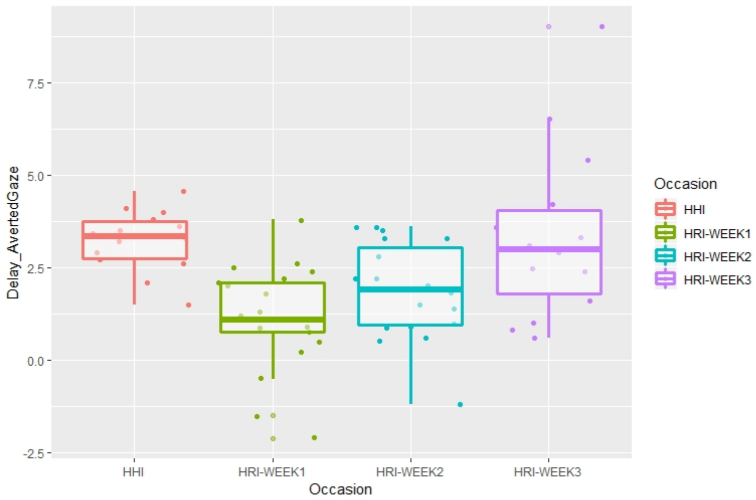


Figure 4.18: The delays of encounters where an averted gaze was present. The box plots are drawn to identify the behaviors of parameters and their evolution with time and experience. The markers are drawn to the scale of the duration of gaze.

Similar to the previous graphs, duration of averted gaze also recorded similar trends in HHI and HRI-week 3 in terms of the average duration and the pattern of data scatter. HRI-week 2 recorded data considerably scattered as in previous graphs. As seen from the box-plots, the trend in delays of averted gaze received numerically slightly different values in HHI and HRI-week 3. These trends were further analyzed statistically in Table 4.8 to derive into conclusions upon similarities between HHI and HRI.

In addition to gaze-related behaviors, several other responses towards the interaction were observed during the study. These were voice responses such as greetings (e.g: "Hey", "Hello, how are you?"), smile, changes in body pose and changes in hand pose. Fig. 4.19 illustrates these behaviors adopted by the participants to react to the approacher, in addition to gaze.

It can be seen that humans use several other responses to support an interaction even when engaged. Proactive robotic systems have a challenge in perceiving such behaviors as well as analyzing gaze behavior.

Most important and unexpected patterns were demonstrated from the results

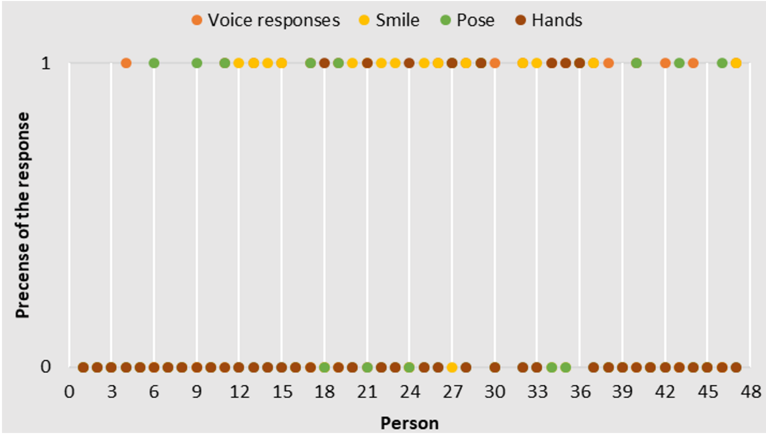


Figure 4.19: Responses observed from users during HHI, in addition to gaze are plotted against each individual.

of the T-test shown in Table 4.8. For both continued gaze delay and averted gaze delay p>0.05. Hence the significance of the effect in the cases: continued gaze delay and averted gaze delay, becomes of interest. But this significance is slight in continued gaze delay. A probable reason for this is if users are interested they paid attention for both robots and humans. But if they are not interested in an interaction, they will tend to neglect the robot for long. The null hypothesis: There is no difference between the two groups in terms of the gaze parameter, was rejected in averted gaze duration as p=0.0079 (p<0.05). Therefore it can be assumed that there is a significant difference between human responses towards humans and robots in terms of averted gaze duration. This difference can also be seen in Fig. 4.16. Even though the mean values of averted gaze duration in HHI and HRI- week 3 are close, the actual durations in HRI were scattered over a wide range of values. Familiarity and past experience with robots and personal opinions on robots can be stated as a probable reason for this observation. Humans still consider a robot as a 'machine' rather than a 'companion'. Such perceptions can also support this observed behavior.

Table 4.8: T-test for the comparison of gaze parameters among HHI and HRI-week 3

Gaze parameter	T-scores	HHI	HRI-week 3
Continued gaze delay	mean	0.855	0.602
	variance	2.473	1.717
	dof	46	
	P	0.0535	
	t	1.678	
Averted gaze delay	mean	0.963	0.998
	variance	2.425	3.949
	dof	46	
	P	0.434	
	t	1.678	
Averted gaze duration	mean	0.885	0.536
	variance	2.072	1.106
	dof	46	
	P	0.0079	
	t	1.678	

4.3.5 Finding Answers to previous questions

As mentioned in the beginning of section 4.3, the objectives of this set of experiments were to find answers to following questions and the answers that were derived from the study are stated.

1) How humans use gaze-related features as actions during human-human interactions?

During the study we observed the existence/nonexistence of gaze and if gaze existed, the delay and the duration of that gaze. These parameters associated with gaze were numerically measured as a part of the study. In addition, four other responses were observed; voice, smile, pose changes and hand movements. From the results, it could be observed that gaze was the most frequently used behavior during HHI, upon all these observed behaviors. Hence exploring gaze-related behaviors received much higher attention while analyzing results.

2) How similar features were associated with human-robot interactions?

Once humans were comfortable with social robots, similar responses they use towards other humans could be observed towards robots as well. Comparison of results in Fig. 4.19 testify to this. According to these two studies, when users were busy, robots received lesser attention than humans, as a result lesser responses as well. Somehow the behaviors observed in front of robots were not different from those in front of humans, expect in number.

3) How similar are HHI and HRI in terms of gaze behavior?

Out of the three parameters considered: continued gaze delay, averted gaze delay and averted gaze duration, we could observe similarities between HHI and HRI in continued gaze delay and averted gaze delay. Averted gaze duration during HHI and HRI showed a significant difference. Therefore we can say that there are dissimilarities between HHI and HRI in terms of gaze. It is important to identify these changes as an insight for the future cognitive models of robot assistants. It is also important in evaluating the human acceptance of social robots as well.

Results confirmed the fact that 'gaze' play an important role in HHI as well as HRI with some observable variations. These findings will help future proactive robots to take situation-cautious decisions in human-robot collaborative environments. Trends in human-human interaction that could be adopted in human-robot interaction are suggested as the outcomes of this study.

Chapter 5

RECOGNITION OF NONVERBALS

5.1 Psychophysiological Aspects of Human Behavior

5.1.1 Simple Human Activities

Utilization of the space around a human vary with the movements involved in the activity. Most human activities involve hands. Movements involved with hands also vary depending on the handedness of that particular person performing the activity. Therefore, in this section we introduce a model to monitor spatial behavior of hands which make a convenient feature for observation and perceive simple human activities.

Space around the user, except his back, is divided into 50 regions for the ease of identification of the task, the human is engaged in. These are shown in Fig. 5.1. One or few regions will be associated for a certain activity. Generally, only an identified number of regions is covered during a certain task. Activities that involve gestures which are not directly interpreted by a robot can be made certain using this fact. Information is analyzed to examine which regions are mostly visited by joints such as elbow and palm, which are mostly responsible for gestures. Finally, the Interaction Decision Making Module outputs the task identified by the system after the period of observation.

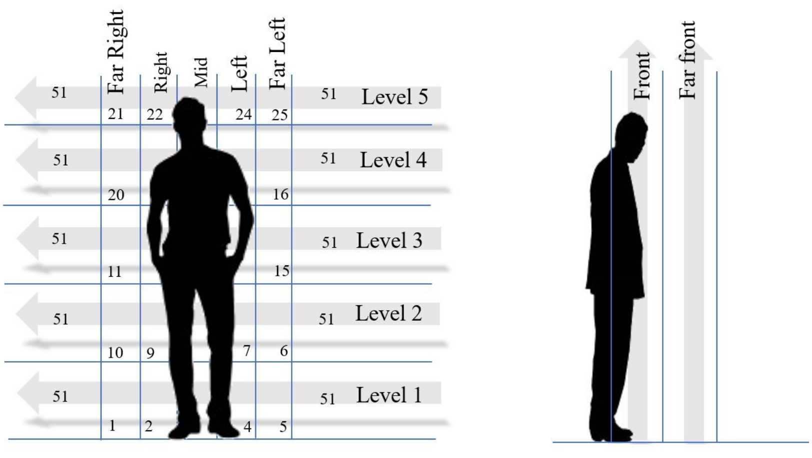
Two approaches are implemented for comparative assessment of activity space. Modules called Activity Space Analyzer(ASA) and Quatitative Activity Space

Analyzer(QASA) are used for this. In ASA, only the touched portions of activity space are analyzed while in QASA, the percentage of times each portion is visited or the frequency of visits, is also analyzed simultaneously.

5.1.2 Criteria used to analyze activity space

Front and side personal space of the human is divided into sub spaces called 'zones' as in Fig. 5.1. For accurate prediction of behavior, 50 zones are identified. The larger the number of zones, the higher the capability to track even smaller movements. Even so, training of data is difficult when the number of zones increase. Therefore 50 regions are selected to match these 2 requirements. Coordinate locations of elbow and palm are tracked and their respective zone is recorded. This is due to the fact that sholder, elbow and palm are the most responsible joints for upper body gestures. Among them, sholuder coordinates do not change considerably during activity. Therefore, elbow and palm joints are selected for this system. This way, all the zones that each joint visits through out the period of observation,t are recorded. Zones are numbered as below.

($level1$, $far\ right$, $front$) - Zone 1 ($level1$, $right$, $front$) - Zone 2 ($level2$,



(a) (b)

Figure 5.1: Arrangement of activity zones near a person (a) and (b) shows the division of side and front space which altogether makes 50 zones. Any region outside the marked boundaries is categorized as the 51^{st} zone.

right, *far front*) - Zone 34 (*level*5, *far left*, *far front*) - Zone 50 *all other areas* - Zone 51

At the end of t seconds, the percentage of visits to each zone by elbow joint and the palm joint are calculated separately. Set of zones does not depend on the whole body movement of the person, as zones are generated relative to specific body joints as in Fig. 5.2. The optimum number of zones required for accurate activity recognition is determined empirically.

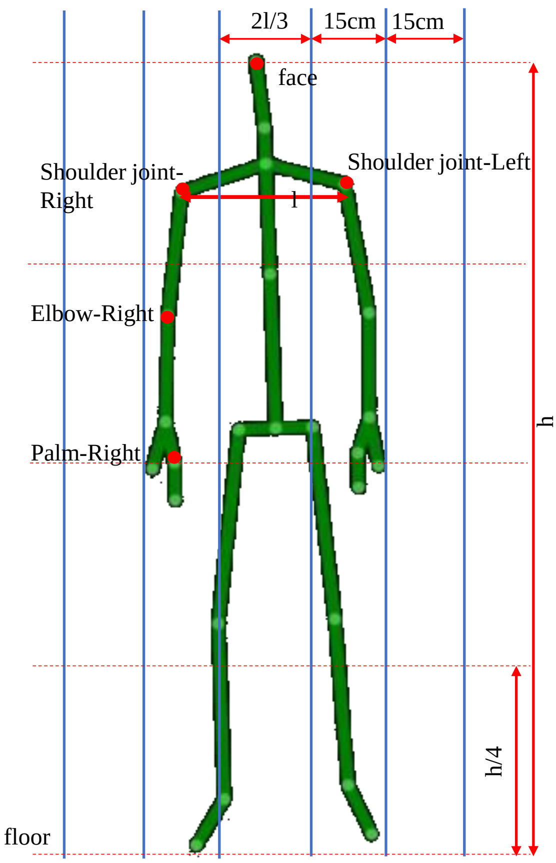


Figure 5.2: Dimensions and joints used to separate zones are shown in the skeletal diagram. The distance between right and left shoulder joints is denoted as l and the height from the floor to face joint is denoted as h. Values used for length and width are marked.

5.1.3 Analyzing Extracted Information

Out of the 2 approaches, ASA uses zones visited by each joint as inputs. QASA takes the zones as well as the number of visits to each recorded zone as a percentage. An example scenario is given below.

Consider an occasion in which the user was walking. Zones recorded during the walk were as follows.

Right elbow $=\{20, 51\}$

Right palm $= \{11,36,51\}$

Activity Space Analyzer (ASA)

This system takes the input sets shown above. While walking, hands move freely forwards and backwards. Therefore the area covered by the palm is greater than that of elbow joint. Therefore elbow visits zone 20, while palm visits zones 11 and 36. Backward movement of joints is the reason for having zone 51 as an input in both scenarios.

Quantitative Activity Space Analyzer (QASA)

Inputs to the system are as in Fig. 5.3. Not only the zones but also their percentage out of the total number of visits to all the zones are taken as inputs to this system. This allows the user to distinguish tasks which use same activity space but have a different pattern in behavior.

Zone	1	2	3	……….	20	……….	48	49	50	51
Percentage	0%	0%	0%		48%		0%	0%	0%	52%

(a)

Zone	1	2	……….	11	……….	36	……….	49	50	51
Percentage	0%	0%		56%		23%				21%

(b)

Figure 5.3: Approach of QASA. It records the zones each joint visits and finally calculates the percentage visit to each zone. (a) and (b) gives the zones and respective percentage of visits of each zone for elbow joint and palm respectively.

5.1.4 Decision Evaluation

A feed forward neural network was used to relate zones visited during a particular task. In ASA, inputs to the neural network were the set of zones numbered from 1-51, visited during the period of observation. Zones related to elbow and palm were given as separate inputs. QASA takes the percentage of times each zone was visited(visiting frequency), in addition to the inputs of ASA. The output from this neural network was the closest task out of the specified tasks within the system. Input variables of 17 persons during each task performed over 6 minutes are taken to train the system.

The robot was placed in the domestic environment and the experiment was carried out under 2 cases. In both the cases same activity set was performed by the 11 participants and same activity was performed 10 times by a single participant in order to identify the behavior of the system accurately. A set of 9 standing tasks were selected for the experiment. These tasks are listed below.

- Task 1: Walking, hands aside
- Task 2: Standing, arms crossed
- Task 3: Standing, scratching head
- Task 4: Standing, heavy object in hand
- Task 5: Standing, desk activity

- Task 6: Standing, brushing teeth
- Task 7: Standing, holding phone in the ear
- Task 8: Standing, waving hand towards robot
- Task 9: Standing, touching face

Among these, were tasks recognized as difficult to be distinguished by other activity recognition techniques. Scratching head, brushing teeth and holding a phone in the ear can be stated as examples. Tasks were numbered to avoid confusion with other tasks. During practical implementation, 20 s was used for t. Two frames were captures per second.

Case 1

ASA is implemented. Each Participant was allowed to perform tasks numbered from 1-9. System observed the task for 20 s after the person is tracked. Spatial arrangement of right elbow and palm were recorded and analyzed by the system as described in previous sections. Output of the system was compared with the actual activity. Eleven participants were asked to perform the same task in their own way. These eleven participants were asked to perform all the tasks ten times, with a time gap in each trial, to validate the system. Ten trials are selected because the regions covered during the same task by same user in two different times will be different.

Case 2

QASA is implemented and the same procedure as case 1 is followed. The number of inputs to the neural network was increased as the percentage distribution of 6 most visited zones are taken in addition to the zones itself. All the other factors are same as case 1.

Several occasions analyzed by the 2 approaches are shown in Fig. 5.4. Information used by the Data Analyzer during each scenario in Fig. 5.4 are given in Table 5.1. The zones recorded during the corresponding activity during the 2 cases are given under 'Zones in ASA' and 'Zones in ASA' columns. Number of times the activity was correctly identified was given under 'Recognition Accuracy'. To find the relationship between the tasks and activity space during both cases, mean values of the results are compared in Fig. 5.5.

In Fig. 5.5, the mean values for tasks 1-8, during case 2 is greater than that of case 1. Therefore QASA seems more appropriate for these 8 tasks. Recognition of task 9 is more accurate in case 1 but the overall accuracy of task 9 even in case 2 is not high. 'Touching face' is often done by humans involuntary, but habitually. Therefore, various humans perform task 9 in various ways. This made the task identification complex.

All the below facts were supported by the results presented in the box plot in Fig. 5.5. Results of the study confirms the fact that the activity space is a demanding feature for human activity recognition. Many tasks that were confused by complex techniques could be recognized with a satisfactory accuracy, by the second approach(QASA) proposed.

(a) (b) (c)

Figure 5.4: Various occasions tested with the 2 systems (a) An occasion in which the person is reading a book. This is regarded as a desk activity (Task 5) which keeps the forearms just below the chest. Marked in red, are elbow and palm joints considered by the system. Corresponding skeletal diagram used to locate these joints and to create the mesh is shown on left. (b) The user is waving hand towards the robot (Task 8) (c) User standing with his hands crossed (Task 2).

Table 5.1: Experiment results

Task	Zones in ASA		Zones in QASA	
	DA Information	Recognition Accuracy (out of 11)	DA Information	Recognition Accuracy (out of 11)
1	elbow: 14 palm: 7, 14, 39	6	elbow: 14(100%) palm: 7(93%), 14(3%), 39(4%)	10
2	elbow: 14, 39 palm: 13	8	elbow: 14(88%), 39 (12%) palm: 13(100%)	10
5	elbow: 14 palm: 13, 38	6	elbow: 14(100%) palm: 13(17%), 38(83%)	10
6	elbow: 16, 17 palm: 18	8	elbow: 16(17%) , 17(83%) palm: 18(100%)	11
7	elbow: 14 palm: 17, 18, 43	7	elbow: 14(100%) palm: 17(56%), 18(41%), 43(3%)	9
8	elbow:14, 15, 17 palm: 17, 16, 42, 41, 18	9	elbow:14(73%), 15(6%), 17(21%) palm: 17(7%), 16(39%), 42(49%), 41(1%), 18(4%)	10

Certain tasks which do not use activity zones repeatedly, could be identified by SA. Examples are tasks 2, 3, 4 and 9. These tasks use fewer zones as well as fewer number of visits to each zone. Therefore recognition accuracies of ASA and QASA are not highly deviated. This rule does not apply for tasks with repetitive movements such as 1, 6 and 8. During these, specific but fewer number of zones were visited constantly. Therefore, these tasks were difficult to be differentiated just by zones, but by the percentage of visits per each zone. By comparing 2 approaches as in Fig. 5.5, it can be observed that the mean recognition accuracy of most of the activities by QASA is larger than that of ASA. Therefore, QASA is a suitable approach for many activities except for the exclusions discussed above.

Summary: This model presents a comparison of the results of two approaches that use activity space. The first approach called 'Activity Space Analyzer' is used to study the relationship between the usage of activity space and a certain task. Second approach which is called 'Quantitative Activity Space Analyzer', is to find the best approach to use activity space for task recognition. An experiment

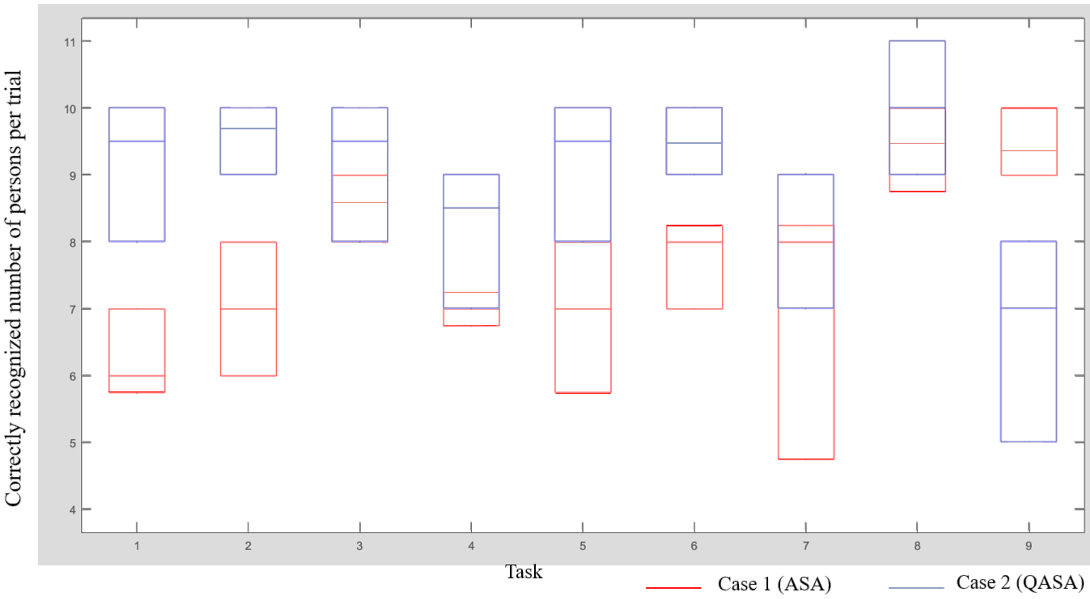


Figure 5.5: Boxplot from the results obtained for case 1 and 2. Box represent inter quartiles. The tasks are numbered from 1-9. y- axis denotes the number of times tasks of 11 participants were recognized accurately (out of 10 performance trials made by each person for a single task).

was conducted to assess the usefulness of each approach towards recognition of certain tasks and restrictions in each approach is discussed. Furthermore, results confirm that the suitability of each approach depends on the type of task but Quantitative Activity Space Analyzer showed positive results for many activities involved. The key factors found during two approaches were highlighted in the discussion. The results of these two approaches can be used to improve existing activity recognition techniques with an additional feature called ‘activity zones’.

5.2 Hand pose

Gestures made by hands are a way of utilizing the perceived environment. In another perspective, hands are a rich source of body language and affect an on-going task and inner (psychological) state as well. Gestures used when relaxed and engaged are different. Meantime, gestures used in human activities are different in each activity as well. The intention of this research is to develop a system which can identify the psychological or emotional state as well as possible physical activities in order for a social robot to gain several aspects of situation-awareness.

5.2.1 Recognizing hand pose

The approach used to recognize gestures was evolved from the approach proposed in section 5.1. The same concept has been adopted in this section to differentiate and recognize gestures. The space around a person while engaged in a certain task or just relaxing, is divided into smaller regions to identify each portion of space separately. Hence this approach is simply how a person fills the space around him/her, by means of a gesture.

The space occupied by gestures is different from each other when different gestures are considered. The criteria used to identify the activity space and set boundaries in length, width and height of the considered space are shown in Fig. 5.6. The front and side space of a person is divided into two as 'front' and 'far front'. Each of these regions is 20 cm wide. These lengths were chosen to be adequate in differentiating the selected gestures through trial and error. The higher the number of zones, the greater the capability to track even gestures that are alike. However training of data will be cumbersome when the number of zones increase. The mesh created by selecting the optimum number of zones in this way is shown in Fig. 5.7 and activity zones are numbered from 1 to 70. Space outside these smaller regions is numbered as 71. Different gestures occupy different zones and Fig. 5.8 demonstrates this fact. Optimum number of zones required for accurate recognition of gestures is determined through trial and error.

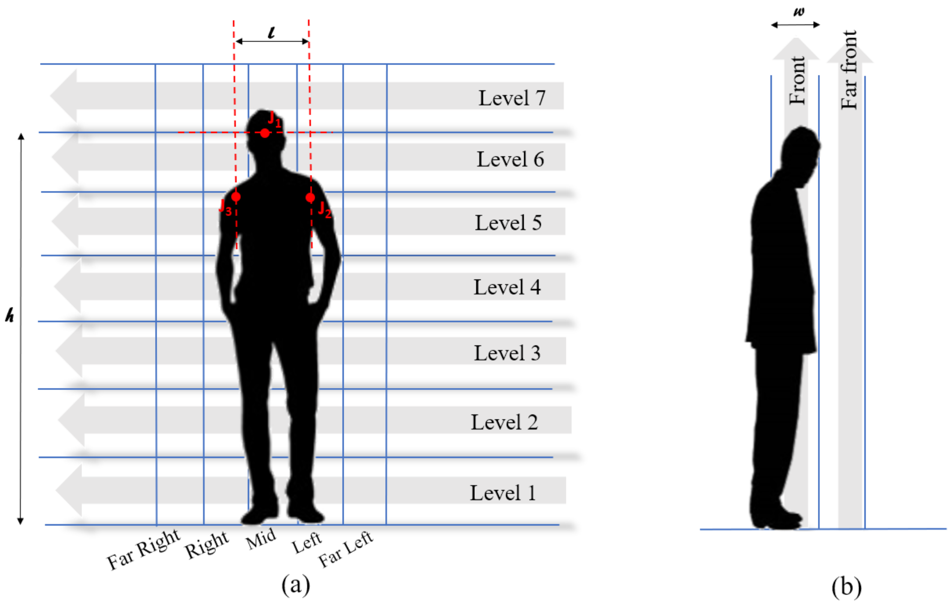


Figure 5.6: Criteria used to determine length, width and height of the mesh are shown. (a) The height from the ground level to the head joint (J_1) is divided into six levels as shown. The vertical space is divided into five regions as 'Far right', 'right', 'mid', 'left' and 'far left' and the length of a single region is taken as 2/3 of the distance between two shoulder joints of the person (J_2 and J_3). (b) The plan view of a person is shown.

Figure 5.7: The 3D mesh created out of the ‘activity zones’ established around a human subject is shown. The zones are numbered from 1 to 71 for the ease of differentiation of portions of 3D space.

Space consumed by the hand

Human hand has two limbs: fore arm and the upper limb. Limb from the shoulder to elbow makes the upper limb and the elbow to hand tip makes the forearm. Hence we consider the spatial formation made by these two limbs by means of shoulder, elbow and wrist locations. The zone occupied by each joint

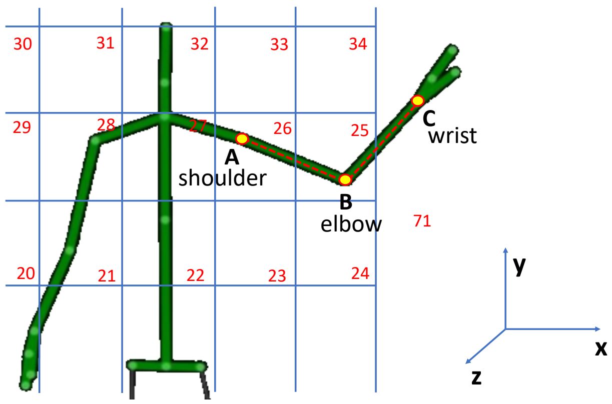


Figure 5.8: Utilization of zones during a certain gesture is shown. The shoulder, elbow and wrist occupy the zones 26, 25, 71 respectively.

was identified in order to recognize the gesture of the hand. Fig. 5.8 shows how these zones were filled by these body joints while making a gesture. The set of gestures considered during this research was shown in Fig. 5.9. The reason for choosing a spatial approach to recognize these gestures is demonstrated in Fig. 5.10. Basically, two shoulder joints hold the grid. When the person walks, the grid will follow him/her immediately.

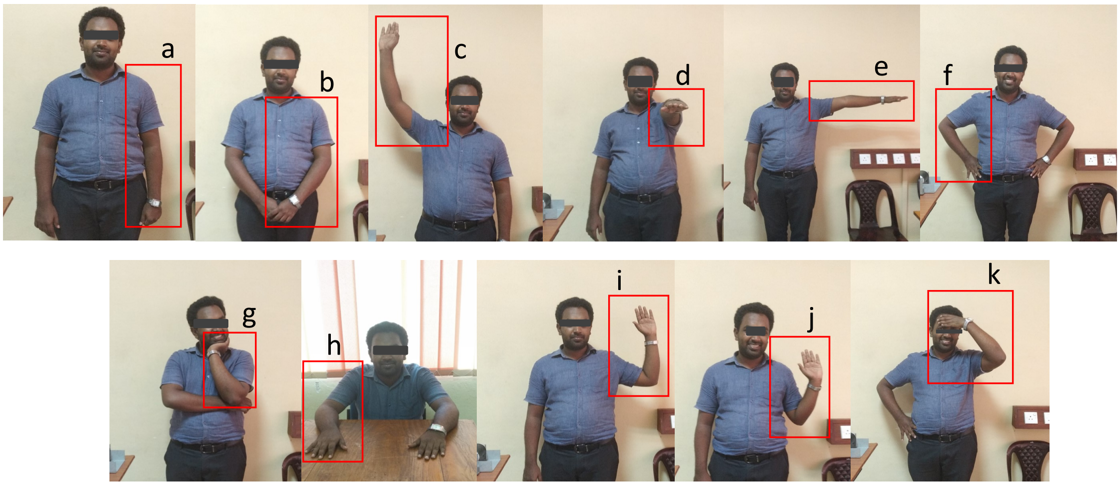


Figure 5.9: The set of gestures used to recognize using the proposed approach are marked with a rectangle colored in red. These gestures are named from a to k for the ease of reference.

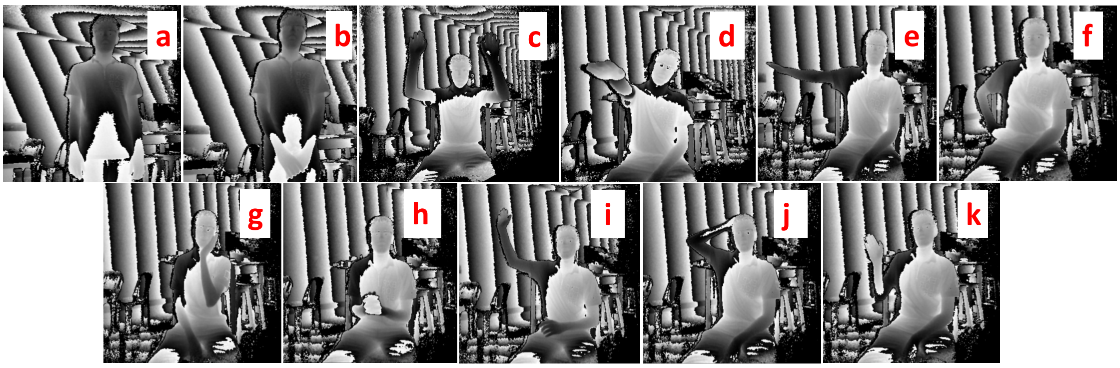


Figure 5.10: Depth images corresponding to the gestures from a-k are shown. Relatively deep points are shown in black while relatively shallow points are shown in white. As seen, each gesture has equal or unequal depths in shoulder, elbow and wrist.

5.2.2 Experiment and results

Experiments are carried out with the participation of 14 individuals on an age gap of 25-51 years (mean of 35.6 and SD of 10.7). A deep neural network (DNN) was trained to map the set of zones occupied by the shoulder, elbow and the wrist, to the corresponding gesture. We used a DNN of 6 layers and 24 neurons in the first layer. Training data set consisted of gesture information of 34 human subjects. The model resulted in a training accuracy of 97.05%. Considering the gesture shown in Fig. 5.8, the inputs and the output can be stated as follows.

Inputs: 26, 25, 71

Output: j

Here inputs denote the set of zones occupied by the shoulder, elbow and wrist respectively. First we asked each participant to represent each gesture in front of the camera. Hence the participants were not covered by any obstacles in the surrounding. Then we asked the same participant to repeat the gesture while seated. The participant was allowed to sit on a chair with a table in front in order to evaluate the system performance with a few obstructions. The intention of this task was to find the impact of obstacles and self-occlusion upon the behavior of the recognition system. The recognition accuracy was calculated as a percentage of correctly recognized gestures out of all performed gestures. In each of the two occasions, participant was asked to perform the gesture twice. The reason for this is that in different occasions, we alter the same gesture, sometimes significantly.

Results of the experiment are shown in Table 5.2. The accuracy of the proposed algorithm is given in two occasions and the correctly recognized number of gestures out of 28 trials is given in the brackets. Gestures with most of the body parts adequately visible to the camera showed a higher recognition accuracy. Highest accuracy was obtained for ‘a’ while posing for the sensor. The recognition accuracy of the same gesture while seated showed a relatively lower value as most

Table 5.2: Results of the experiment obtained for two occasions: while users were posing for the sensor and recognition accuracy of hand pose while the users were seated.

Gesture	Accuracy (%) Posing for display	Accuracy (%) During an activity
a	100.0 (28)	82.1 (27)
b	89.2 (25)	85.7 (24)
c	89.2 (25)	85.7 (24)
d	96.4 (27)	85.7 (24)
e	85.7 (24)	82.1 (23)
f	82.1 (23)	85.7 (24)
g	96.4 (27)	89.2 (25)
h	92.8 (26)	85.7 (24)
i	92.8 (26)	82.1 (23)
j	85.7 (24)	78.5 (22)
k	53.5 (15)	39.3 (11)

parts of the hand were covered by the person him/herself when performing this gesture.

As a whole, the recognition accuracies of gestures were lower when seated. A probable reason for this behavior of the system is the self-occlusion resulted by the person's posture and obstruction by nearby objects such as tables, and chairs. Except for 'k', the approach seemed to give promising results in gesture recognition. As most of these gestures are displayed by humans while standing, where the system showed a satisfactory recognition accuracy. As this approach can be used real-time, it has the possibility of being utilized for behavior monitoring in many applications. In seated arrangements, the sensor was set at an angle from the direction the participant was facing. The reason for this is that the most of the participant was visible to the sensor in such an arrangement. Otherwise tables or other objects will obstruct the vision of the sensor. Except for 'k' which was confused with 'g', the approach showed promising results for the recognition of gestures a-j.

Summary: In this section, we identified a method to recognize the body language of hands by means of the spatial orientation of hands. The 'activity space' around a human was divided into smaller units called 'zones' and a 3D mesh was generated around the particular subject with the combination of a number of zones. Location of major joints of human hand: shoulder, elbow and wrist, were recorded by means of these zones. The set of zones occupied by each joint vary in each gesture and this feature was used to distinguish each gesture. An artificial neural network was used to map the location of joints into the corresponding gesture. The experiments were conducted in simulated laboratory environment to evaluate the system. In the end of the experiment we found that the concept of 'activity space' is an interesting aspect to explore human body language by means of gestures. Hence this method can be used to recognize most of the frequently used gestures in social settings. Implementation of this approach showed promising results in identifying selected 11 human gestures with a decent accuracy.

5.3 Pose recognition and robot-initiated interaction

When relaxing postures are considered in domestic environments, often humans tend to attain unusual postures such as leaning on chairs and sitting on floor. Although relaxation of a person is directly related to his/her posture, there's no exact method of defining that relationship. Therefore, in this section, we present how commonly used poses of standing and sitting postures were recognized and interpreted to relate them with the relaxation level of a user. Out of the selected, some poses which can exactly be clarified as 'sitting' or 'standing', are used as references. Other poses which are deviated from these reference poses are divided into categories, considering how these poses are used by humans in their casual behavior. For example, in working or writing a person may sit on a chair with his spine bent forward. The same person will sit with his spine erect, during an ordinary situation. Such often encountered postures are considered to be evaluated by the proposed system.

There are 9 postures considered in the present system, as in Fig. 5.11 and Fig. 5.12. At the end of t seconds, spatial parameters are analyzed prior to body parameters. This stage evaluates the distances to the user within the observed

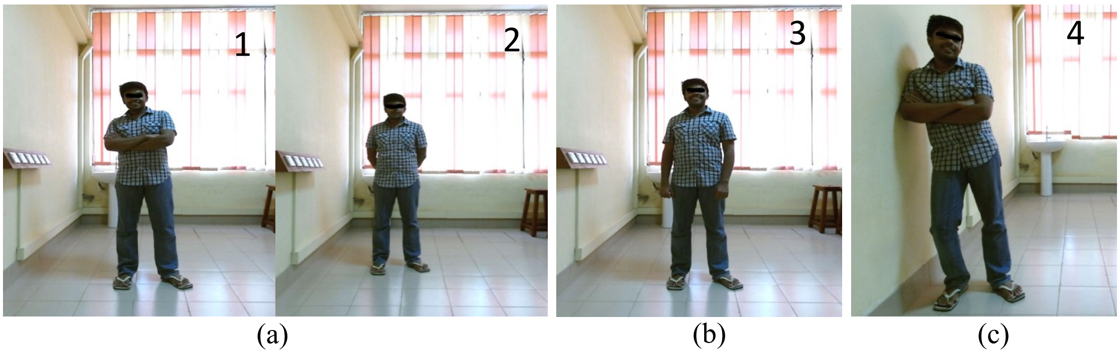


Figure 5.11: Types of Standing poses used in the experiment (a) Standing pose 1,2 with 'Low' Interaction Readiness (b) Pose 3 with 'Medium' Interaction Readiness (c) Pose 4 with 'High' Interaction Readiness.

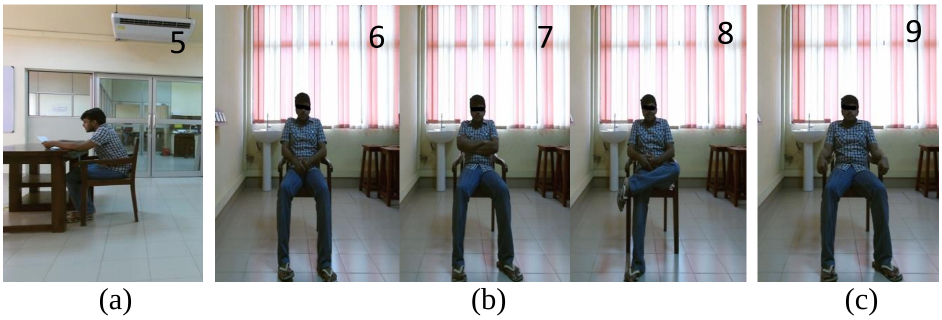


Figure 5.12: Types of Sitting poses used in the experiment (a) Pose 5 with 'Low' Interaction Readiness (b) 'Medium' Interaction Readiness in poses 6, 7, 8 (c) 'High' Interaction Readiness in pose 9.

duration of time and determines the walking direction of the user. If the user is not walking away, further information from DRS is processed. Angular information is extracted once every second and the rate of change of these angles are also calculated and stored in DRS. Extracted and calculated information in every second is stored in this way during t seconds. At the end of t seconds, these values stored in the DRS are compared with the corresponding values throughout previous (t-1) seconds. Final decisions are made after the analysis of information within these t seconds. At the t^{th} second, Pose Identifier (PI) analyses user's pose using the extracted joint angle data within that second. Most probable 9 postures to be found in a domestic environment including 4 standing and 5 sitting postures are defined in PI.

A variable called ‘Interaction Readiness’ is used along with the rates of change of joint angles as inputs to the fuzzy system from PI. A measurement of natural tendency for interaction in each pose is defined as the ‘Interaction Readiness’ of that particular pose. Interaction readiness is considered to be high when the user is highly interested to interact with an outsider.

In relaxing standing postures, torso will sag slightly to a side and legs will slightly be bent. Therefore pose 4 in Fig. 5.11 is used as the most relaxing standing position. In convention, when the user covers his chest with arms as in pose 1, it’s a sign of interaction rejection. Therefore, pose 1 is considered to be a least suitable occasion for interaction. The torso of a resting pose will be well balanced, but will not curl up forward as in fear. Therefore, user leaning back, as in pose 9 in Fig. 5.12 is used as the most relaxing sitting pose, while pose 5 has high tension and is therefore considered as a less relaxing pose.

Considering the above reasoning, pose 3 in Fig. 5.11 is the defined standard standing pose of a user in this system, while 1 and 2 have ‘Low’ interaction readiness and pose 4 has a ‘High’ interaction readiness wherein humans normally tend to lean during long term interaction. In the same way, in sitting poses, 6,7,8 in Fig. 5.12 are considered as normal sitting poses while 5 has ‘Low’ interaction readiness and 9 has ‘High’ interaction readiness. The standard poses 3,6,7,8 have ‘Medium’ interaction readiness. In other words, these 4 poses are used as reference poses. The system further determines the user’s walking direction through comparison of stored distance values in DRS. Joint angular speeds and whether the user is still or walking towards or walking away were considered before measuring user readiness. These parameters were analyzed in the ‘Information Processing’ stage. Interaction readiness and highest angular speed throughout t seconds recorded were input to a fuzzy inference system to decide the interaction demanding level of the particular user. This phase met within Interaction Decision Making Module (IDMM). IDMM made the decision regarding interaction and outputs the time to wait and the social distance that should be maintained before interaction. In relaxed body language, fluid movements are made. These movements do not prevail over a time. This is the reason to observe user's body parameters stored in DRS for a t duration.

After evaluation, if interaction demanding of the user is high, robot waited for a predefined time and then moved towards the user until the distance between two conversant becomes equal to the output from Interaction Initiator. With this step, conversation initiation factors are achieved and hence it's deduced as a favorable condition for interaction. This is performed in the final phase.

During each body information extraction stage, angular movement was calculated and stored in DRS. During information processing, information stored in DRS were used. IDPI decided whether the user was in a standing or a sitting position using such information regarding joint angles. After this evaluation only, the data from DRS was analyzed for pose selection. A feed forward neural network is used to relate body angles to a particular posture. Inputs to the neural network are the set of body angles α, β, θ and γ where the variables are shoulder joint angle, elbow joint, hip joint and knee joint respectively. The output from this neural network was the closest pose out of the specified poses. If the angular speeds were constantly changing, the maximum recorded speed was obtained as the second input of the fuzzy system.

Here, in this system the sleeping relaxing poses were excluded as humans often do not prefer long term communication in these poses. Often when seated, people involuntarily shake their legs or flung their legs out to music. Therefore $\frac{d\gamma}{dt}$ was not considered during the analysis of joint angular speeds, to measure interaction level. Here, $\frac{d\gamma}{dt}$ is the rate of change of knee joint angle. This avoids misinterpretation of the occasion as a 'busy user'. IDMM had two inputs as the interaction readiness and the angular speed. For example, consider the following inputs; Input 1 = HIGH and Input 2 = $max\{\frac{d\alpha}{dt}, \frac{d\beta}{dt}, \frac{d\theta}{dt}\}$. Input 1 indicates the interaction readiness of the user after analyzing the pose information and Input 2 gives the maximum recorded angular speeds out of the 4 joints considered. Given above is an occasion like pose 4 of Fig. 5.11. IDPI output for this scenario will be 'Medium' while all the angular movements are close to zero. Hence Input 2 is recognized as 'Very Slow'. The mutual distancing between robot and the user are determined as in Fig. 5.13. S_1 and S_2 denotes distances traveled by a robot during 'high' and 'low' interaction demanding respectively.

5.3.1 Interpretation of Fuzzy Information

In this approach, some poses are given more weight as 'relaxing' positions while some are considered less relaxing when a domestic environment is considered. After pose identification from DI, interaction readiness is decided from that pose. This is the final phase of the PI. IDMM is implemented by Mamdani type fuzzy inference system to evaluate the level of interaction demanding of the user. The fuzzy logic controller in FIDMM takes interaction readiness and highest angular movement as input. Singleton membership functions are used to represent poses identified from the neural network.

Output is the extent to which the user demands interaction. This is illustrated as 'Interaction Demanding' and is scaled from 'Very Low' to 'Very High'. Centre

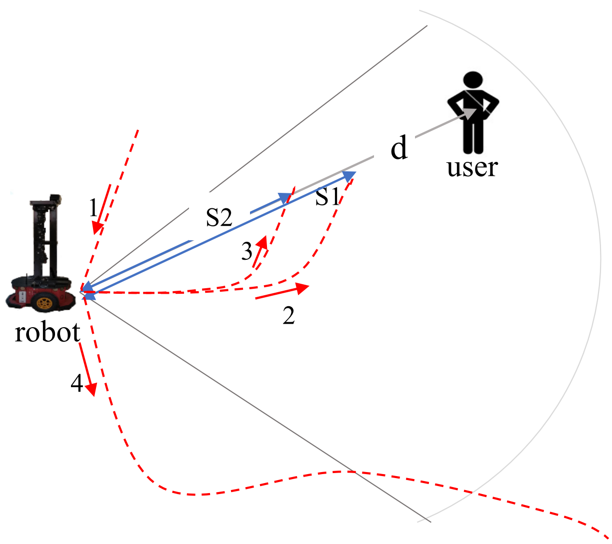


Figure 5.13: Routes followed by robot during various situations.

of area method is used for defuzzification of this output. Membership functions of corresponding inputs and outputs are shown in Fig. 5.14.

Angles related to a similar posture of two users will be slightly different from each other. Therefore, defining a boundary to define a posture will be inaccurate. To interpret vagueness in this situation, several fuzzy labels are defined for postures. Standing position 1 and 2 were considered to be highly unlikely to interact and were categorized under fuzzy label 'LOW' interaction readiness during standing while 4 is considered to be highly interactive. Therefore, that posture is categorized under 'HIGH' interaction readiness. If the angular movement was greater than 95°, the system identified the user as 'busy'. In this case 95° was the angular speed limit for interaction and was denoted by $\frac{d\phi}{dt}_{max}$. Therefore, only angular movements below the above limit are considered in applying the rule base. Otherwise, the user is approached by applying the Interaction Initiator. The rule base was defined using the natural tendency in considered postures. Rule base for this fuzzy inference system is shown in Table I. VL, L, M, H and VH denoted fuzzy labels 'Very Low', 'Low', 'Medium', 'High' and 'Very High' respectively.

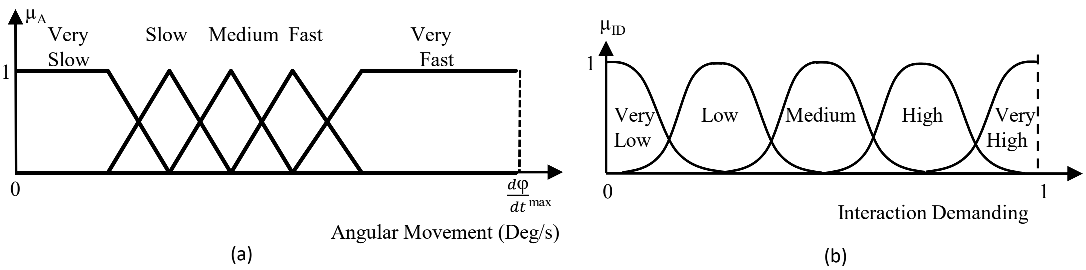


Figure 5.14: (a) represents the input membership function corresponding to angular movement and (b) represents the output membership function which corresponds the level of interaction demanding.

Table 5.3: Results of the experiment

Posture	Closest posture	Extracted Information			Walking direction of user	Interaction decision	Existing Distance	Interactive distance (m)	User Attitude
		Interaction Readiness	Angular movement (deg/s)	Distance (m)					
Sitting, palms on lap	6	Medium	0	2.1	none	yes	reduce	0.8	satisfied
Leaned on chair	9	High	79	1.61	none	yes	reduce	0.8	satisfied
Sitting, arms crossed	7	Medium	0	1.89	none	yes	reduce	1	satisfied
Sitting, one leg on another	7	Medium	0	1.71	none	yes	reduce	0.8	satisfied
Standing, arms crossed	2	Low	0	1.72	none	wait	reduce	1.5	satisfied
Sitting, arms on chair	9	High	1	2.1	none	yes	reduce	0.8	satisfied
Leaning to a wall	4	High	0	1.81	none	yes	reduce	0.8	satisfied
Working in a table	5	Low	77	2.6	none	wait	reduce	1.5	satisfied
Standing	1	Low	8	2.51	none	wait	reduce	1.5	satisfied
Leaning to a wall	4	High	6	1.8	none	yes	reduce	0.8	satisfied
Sitting, hands on chair	9	High	0	2.3	none	yes	reduce	0.8	satisfied
Sitting, hands on lap	6	Medium	0	1.67	none	yes	reduce	0.8	satisfied
Sitting, one leg on another	8	Medium	-1	2.7	none	yes	reduce	0.8	satisfied
Working on a table	5	Low	0	1.4	none	wait	reduce	1.5	satisfied
Sitting, one leg on another	8	Medium	0	1.56	none	yes	reduce	0.8	satisfied
Working on a table	5	Low	58	2.1	none	no	-	-	satisfied
Running	-	-	151	-	none	no	-	-	satisfied
Exercising	-	-	155	-	none	no	-	-	satisfied
Standing, speaking on mobile	-	-	11	2.39	negative	-	-	-	satisfied
Standing	-	-	9	2.2	negative	-	-	-	satisfied
Sitting, working on mobile	6	Medium	5	1.98	none	yes	reduce	1	dissatisfied
Sitting, arms crossed	8	Medium	0	1.45	none	yes	reduce	0.8	dissatisfied
Standing, hands on back	1	Low	0	1.48	none	wait	reduce	1.5	dissatisfied

Results and discussion

The experiment has been carried out with 8 persons aged from 25 to 58 years (Mean of 36.34 and SD of 11.21). The neural network was trained for 9 postures of 27 persons to identify each pose. As sitting and standing positions do not deviate much on age, the results did not show a considerable difference for the same posture of an adult and a youth. Experiment results for several users are given in Table 5.3. Two occasions during the experiment are explained below.

A user was selected and asked to sit in a preferred relaxing pose as in Fig. 5.15. Angular speed was zero and therefore angular movement was 'Very Slow' and the output from layer 4 was 'High'. If the pose changed from one to another within the considered time interval, the rates of change of angles in considered joints must become high. The system decided that the user is busy. This type of occasion was met when the user was exercising or running. Fig. 5.16 shows an occasion where the user was walking away from robot. Hence interaction demanding was decided to be zero. In that case the robot did not reach the user but continued wandering around. Table 5.3 presents the results of the experiment conducted with 8 persons selected out of a selected sample and user satisfaction regarding interaction initiation process followed by this system. One difficulty faced during this experiment was that differentiation of the two standing positions, 'hands on

(a) (b)

Figure 5.15: (a) MIRob approaches a user with 'high' interaction demanding state. The skeletal tracked during that position is shown on right in (b).

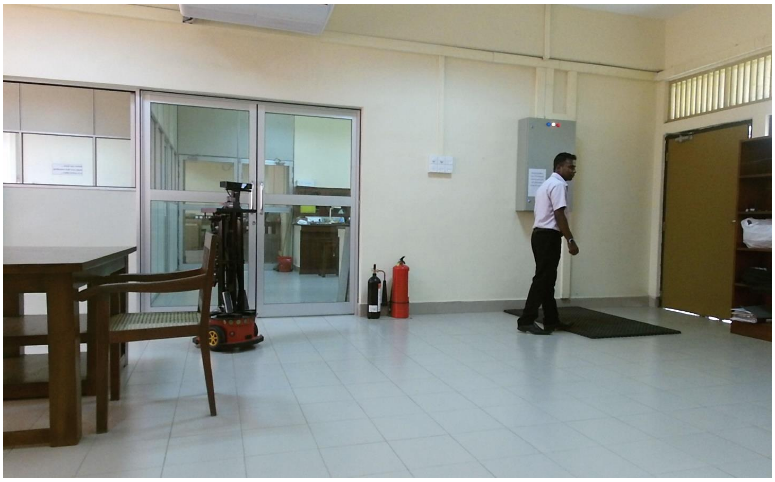


Figure 5.16: The user walks away and MIRob does not approach the user after analyzing information for 5 seconds.

back' and 'hands crossed' was cumbersome as a result of equal range of angles obtained in both the occasions. Therefore, it was difficult to differentiate poses 1 and 2 in Fig. 5.12 using only angles. We could observe dissatisfied users when the robot decided to interact with them even though the user was engaged in a task. Except for such exclusions, general behavior was identified and interaction process took place with higher user satisfaction.

Summary: A fuzzy based attention evaluation system is presented. The system takes angular movements as well as human pose as inputs to assess the level of attention given by a human towards a robot while working. Experiments and results are presented to evaluate the behavior of the proposed model. But there are other factors except angular movements and pose, which may affect human attention. This can be stated as a limitation in the presented system.

Before going for direct conversation, the robot will be able to assess the level of busyness and the suitability of having a conversation depending on the scenario. The major improvement of this system is the involvement of a nonverbal mechanism to perceive favorable human behavior for interaction. Over the existing systems, this system has the capability to define user readiness for interaction through complete analysis of posture information without being framed to face and emotional details of the human. Furthermore, the introduced system does not rely on verbal instructions from user for interaction initiation.

5.4 Interpreting movements of the body

We conducted a study to explore the tendencies in users approving his/her robotic companion under different conditions. Under the circumstances, the activities humans engage in may vary and activities deploy various bodily movements. During this study, users were allowed to engage in a set of activities and movements of the users were tracked for a certain duration. Properties in these movements such as speed, frequency and fanning were used to determine the engagement of a particular user. As the engagement in an activity reduces the interaction demanded by the user from outsiders, degree of affect in a highly engaged activity will be low and vice versa. This fact was evaluated during the study using various activities that humans usually engage in during daily chorus. Depending on the degree of affect, factors considered before the interaction were evaluated. The type of conversation, proxemics between the two conversant and the approaching path were the factors considered here. Finally this approach was compared with an ordinary interaction scenario by initiating a direct interaction with the human without considering behavioral aspects for the same situations. These two approaches were compared by obtaining user feedback during each scenario.

Hence the objectives of this work focus on identifying the effect of factors which affect the perception of nonverbal human behavior and human responses towards a situation-aware robot with the above mentioned capability.

5.4.1 Measuring Behavioral Responses

A qualitative study will base on immeasurable human features such as facial expressions, staring, etc. As a robot cannot perceive such features without analyzing body geometry, facial contours, etc., we focused on quantitatively studying the motion of human body for a period of time, before going for decisions regarding his/her behavior. Hence we translated behavioral observations into discrete figures and statistical outputs in order to perceive their meaning.

Evaluation of verbal responses will only be possible after starting a conversation. In this scenario, the robot requires to evaluate human behavior before an interaction. Therefore only nonverbal cues were selected to evaluate situation for the appropriateness of an interaction.

5.4.2 Observable Nonverbal Human Behavior

Three major features included in bodily movements were used as cues to perceive the situation during this study. These features are mentioned below.

1. Speeds of selected body joints
2. Positioning of body joints
3. Maximum occupied areas around the individual

Out of the observable human cues, the above cues were used to analyze a situation. Body-based movements were used as the factors which drive robot's decisions regarding interaction. The most fit variables that can define motions of a human were selected for the study. The robot perceives its environment through qualitative information derived from quantitative information.

The above features were measured quantitatively in order to determine such emotional parameters associated with the movement. Here, the emotional parameters include the priority given to the task, usage of activity space and user's engagement. All the above factors are determined mostly by the task of an individual. Even so, humans adopt both rational and irrational behavior in the same environment. Here, rational behavior is the influence by emotions or thoughts while irrational behavior is not. Therefore there can be illogical movements associated with irrational human behavior. Random movements in hands and legs are an example for such movements. But these movements prevail over a very short duration. Misinterpretation of such behavior is omitted by analyzing the usage of *maximum occupied areas* around the individual over the period of observation.

Body joints considered for monitoring human behavior in this way, are shown in Fig. 5.17 (a). These joints include head, spine-base and elbow, wrist and ankle of right and left sides of the body. Joints which are mostly utilized during a task were tracked during this study. Variables associated with each motion are shown in Fig. 5.17 (b). These variables are the distance from a reference and joint speeds. Here, the vertical going through the spine-base was chosen as the

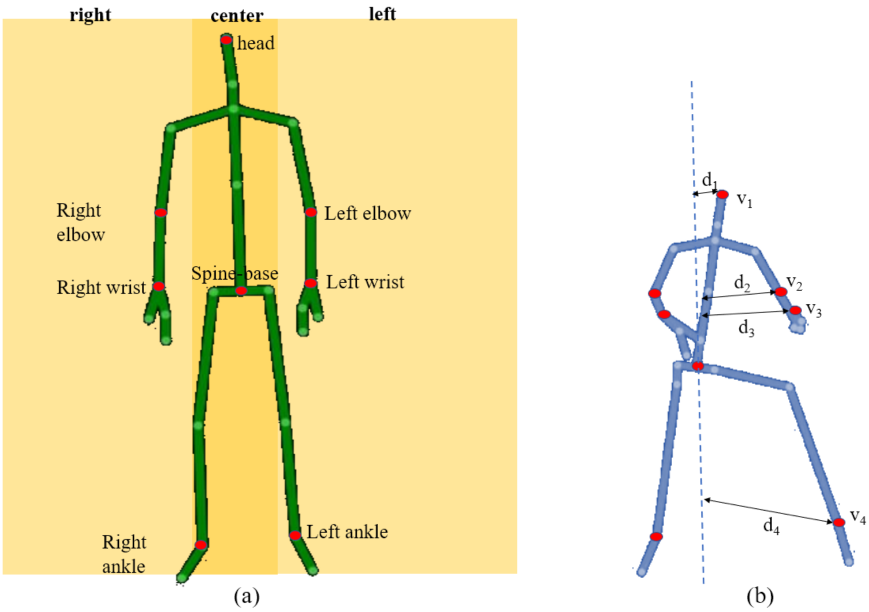


Figure 5.17: Skeletals extracted from a standing and a seated person are shown. The joints used to track movements are marked and the right, center and left regions are shown in (a). (b) shows the distances measured from the line going through the spine-base of the person. Distances to head joint, left elbow, left wrist and left ankle are marked as d_1, d_2, d_3 and d_4. The corresponding joint speeds are marked as v_1, v_2, v_3 and v_4. Similarly, the distances and speeds of joints on the right side are also considered for analysis.

reference to measure distance, as spine-base is least subjected to movements due to the inertia of human body.

These variables observed through out the period of observation, are plotted against time in Fig. 5.18. This shows the *speed*, *fanning* and *maximum occupied areas* of the joint with the highest speed. The reason for this is that this joint is the most actively used during the task. Hence robot considers adjusting the interaction scenario using these properties associated with the user. As the robot determines proxemic behavior in addition to conversation, analysis of spatial behavior of the user will play an important role in being situation-cautious.

5.4.3 Decision Making Criteria

Two decisions; proxemic behavior and conversational preference to suit the occasion were taken after an analysis of the observed cues. The proxemic behavior included the approach direction or the path to be followed and the mutual

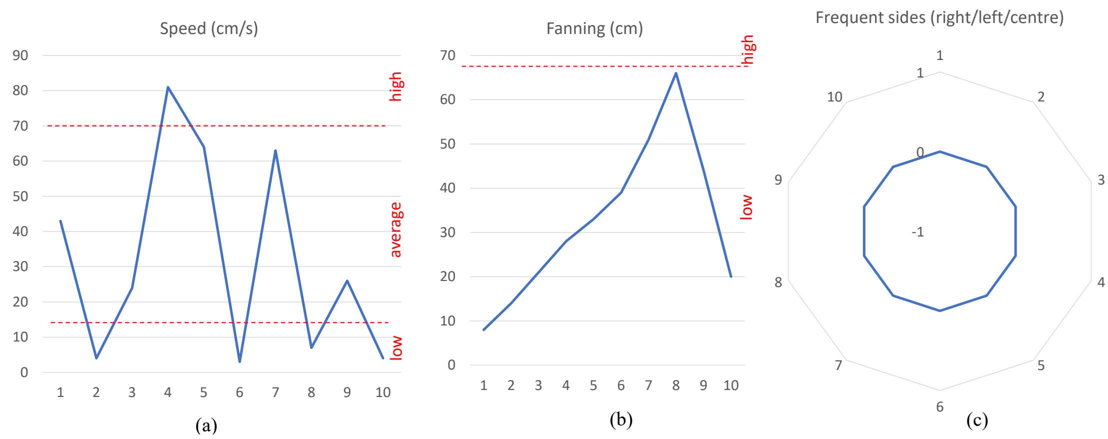


Figure 5.18: Nonverbal behavioral responses observed during the experiment (a) joint speed (here, the speed of the right wrist is shown) (b) the distance from the vertical drawn through the spine-base joint to the considered joint is marked against the time of observation (c) a radial graph showing the highest occupied area. Areas were denoted as right, center and the left of the user. Right, center and left were denoted as 1, 0 and -1 respectively. The responses observed by the robot over 10 s are shown.

distance between the human and the robot. Conversational preference included the type of conversation to have with the user. Conversational preferences are categorized according to the length of the conversation.

Determining the proxemics

The approach direction was chosen so that the maximum occupied region out of right, center and left, was least obstructed by the robot. For example, if the maximum occupied region was right, robot approached the user on his/her left and vise versa. If the center was occupied maximumly, the robot approached the user on the right. If both right and left were equally occupied during the task, the robot approached from front. The accepted interpersonal distance for a mutual interaction is within 1-1.5 m. Therefore we chose an interpersonal distance of 1.2 m for a still user. Otherwise the interpersonal distance was calculated as in (5.1). The units are in meters.

$$Interpersonal\ distance = 1.2 + maximum\ fanning\ to\ the\ front\ (m) \tag{5.1}$$

Determining the conversational preference

Types of conversations robot chose to have with each user were as in the previous experiments. A typical example for such conversations are given in Fig. 4.5 Here we omitted long conversation during this study.

The conversational preference was based on the speed of the joints and fanning. This was due to the fact that maximum occupied area will have no impact upon the conversation that the user intend to have with the robot. Joint speeds and fanning were chosen as the demonstrators of engagement in a specific task. Values of these variables were categorized as ‘high’, ‘average’ and ‘low’ for the ease of analysis. The exact figures for these boundaries were chosen by trial and error, repeating the experiment for a number of times. Conversational preferences were decided using the criteria given in Fig. 5.19.

5.4.4 Monitoring Human Behavior

The experiment was conducted in a simulated social environment within the laboratory. Participants of this experiment were students and some outsiders in the age range 23 to 28 (Mean-25.64, SD-1.68). There were 17 participants and all the participants were in good mental and physical health condition to make decisions. Participants were given instructions regarding the process and the behavior of the robots in the laboratory upon arrival. Users were allowed to engage in a selected set of tasks as part of the experiment. The experiment was conducted using MIRob.

The experiment was conducted to monitor the behavioral responses of users towards robot-initiated interactions. The robot initiated the interaction after considering the movement-based behavior of its user. Movements of head joint, spine joint, right and left wrist, elbow and foot joints were monitored for a period of time. Decisions of the robot included appropriate proxemic and conversational behavior.

Since the responses depend on the priority given by the user to his/her current

		Joint speed		
		LOW	AVERAGE	HIGH
Fanning	LOW	GRT 1	STLK 3	SER 2
	HIGH	NI 0	STLK 3	NI 0

Figure 5.19: The decision grid: Conversational preference chosen at each occasion are shown.

activity, we have selected a set of tasks. These set of tasks are as follows. Tasks encountered in a typical domestic and lab environments were selected for the study. These tasks are as follows.

- Engaged in a desk activity
- Resting on a chair
- Engaged in an exercise
- Cleaning floor
- Standing relaxed
- Having breakfast
- Engaged in lab work
- Listening to a song
- Making a phone call

Robot was remotely navigated towards a particular participant while he/she was engaged in the task. Before approaching, the robot observed the behavior of the individual for a duration of 10 s. The robot initiated a conversation after approaching that particular user. Each user was allowed to perform atleast three tasks listed above. Path planning and navigation tasks of the robot were autonomous while response generation and tracking a user were teleoperated by the experimenter. Hence the teleoperator instructed the robot where to approach and what to say. A single user participated in the experiment at a time and the interaction process was repeated for each participant separately.

As MIRob was monitored by a human operator, its responses were generated with respect to the responses from the user. During the experiment, the responses of the robot towards its user include proxemic and verbal cues only.

If the situation was not favorable for interaction in the form of displayed nonverbal behavior of the user, robot was instructed to leave the situation without interaction. In such a situation, it is assumed that the user does not prefer to interact.

Independent variables used in the study were the movements made by the user. The response of humans towards the interaction initiated by the robot was the dependent variable during analysis stage.

Joint coordinates of a particular user were measured by the Kinect SDK. All the behavioral changes were identified according to the explained criteria, through the algorithm. The map of the environment was predefined in the simulation. Therefore the robot navigated to the target positions and its orientation were defined by the operator.

As each participant was asked to perform three out of the nine tasks mentioned above, 51 different scenarios were encountered during the experiment. MIRob was allowed to observe each individual once the activity was started. After the evaluation of user behavior according to the model, the approach behavior of the robot was rated by the participant. User rated the robot's behavior based on the convenience or the discomfort he/she felt as the robot approached. Here, the period of observation, 10 s, was determined so that an adequate amount of data were obtained for the analysis and so that the user does not feel uncomfortable by a longer observation. Visual information was extracted at a rate of 2 per sec. Proxemic decisions were taken so that highly engaged areas were least obstructed.

The study included two experiments. One to evaluate user responses for a direct approach behavior with a fixed conversational and proxemic behavior. The second as a situation-based approach behavior which is referred o as adaptive approach behavior.

5.4.5 Experiment 1

The robot directly approached the user with a predefined proxemic behavior and initiated a predefined conversation (direct approach method). This direct approach method was used in experiment 1. Hence the approach behavior of the robot was not adaptive in this experiment. The robot always reached the user from his/her front and the conversation was TLK. A mutual distance of 1.2 m was kept between robot and user during the interaction process. After the interaction with the robot, the user was asked to rate the behavior of the robot in the form of a feedback score (out of 10). This approach scenario is shown in Fig. 5.20 (a). In Fig. 5.20 (a) and (b), the robot approached the user from front, with a mutual distance of 1.2 m in between. Fig. 5.20 (c) and (d) shows the approach behavior of the robot while the user was dining. As the right hand did most of the work during the activity, the maximum occupied area was user's right. Hence robot approached from left. As the user keeps his hands in the front region, hands are the farthest to the body. Therefore the robot counted that distance before keeping the mutual distance. Hence the mutual distance was greater than that in Fig. 5.20 (b).

5.4.6 Experiment 2

This experiment was conducted in two occasions as follows. Occasion 1: The robot approached the user after an analysis of his/her behavior. Hence the approach direction, mutual distance and the conversational preference were adapted according to the conclusions derived after the observation. The user was asked to rate the behavior of the robot as in experiment 1. Meantime, the participant's expected behavior from the robot was also recorded.

Occasion 2: After participating at occasion 1, user was allowed to take part in experiment 1 again and give a feedback score.

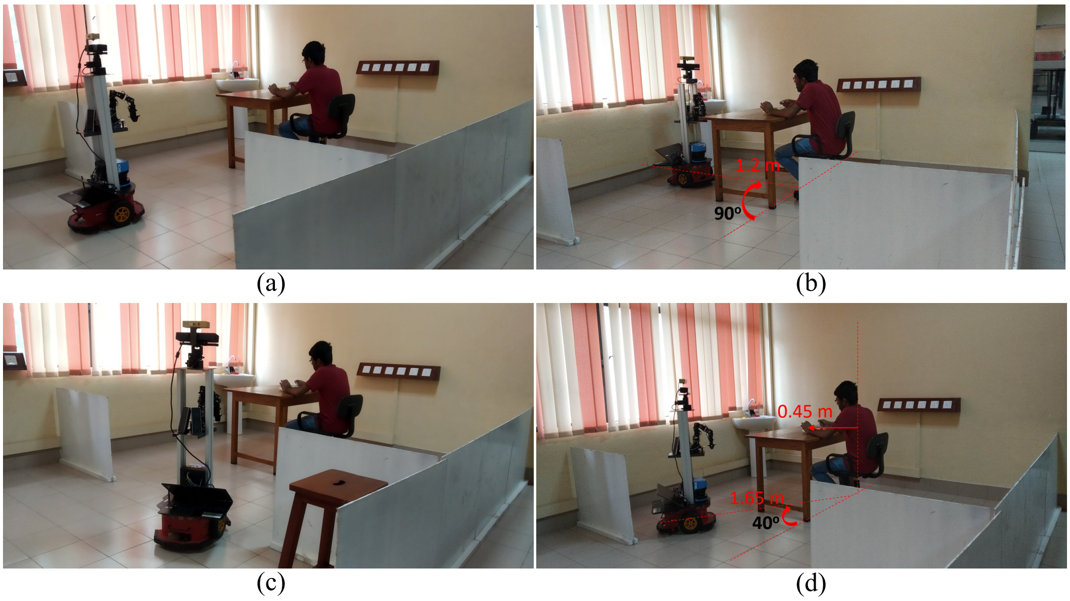


Figure 5.20: Two occasions from the experiment (a) and (b) show the direct approach behavior. (c) and (d) show the adaptive approach behavior after an analysis of the movements made by the user.

Adaptive behavior of the robot during two tasks could be observed in Fig. 5.21. Fig. 5.21 (b) shows the adaptive approach from user's right, as the front region was obstructed by his hands. In Fig. 5.21 (c) and (d), the robot approached from front, as right and left were equally occupied by the user. This decisioning system is as described in 5.4.3.

The three feedback scores received for experiment 1 and occasion 1 and 2 in experiment 2, were compared to find human trends towards each type of interaction.

5.4.7 User Responses Towards Robot

A comparison of feedback scores for the experiment 1 and occasion 1 in experiment 2 is shown in Fig. 5.22. When the feedback scores are considered, occasion 1 in experiment 2 has received a higher feedback score during most of the tasks. The users were impressed with the fact that robot respected their

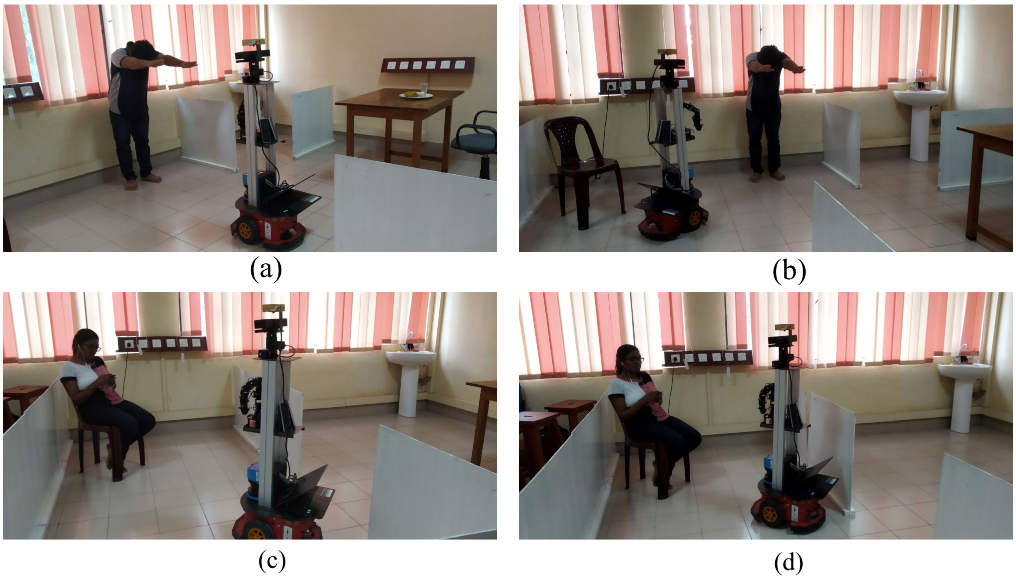


Figure 5.21: Two occasions in which adaptive approach behavior was implemented. (a) The robot observes the user for a duration. The user was doing an exercise. (b) After the observation, robot approached the user from right and initiated the conversation. (c) The robot observed a user listening to a song. (d) robot approached the user from front and initiated interaction.

activity space while approaching that user. Another reason for higher feedback score was the selection of appropriate verbal interaction. In tasks such as desk

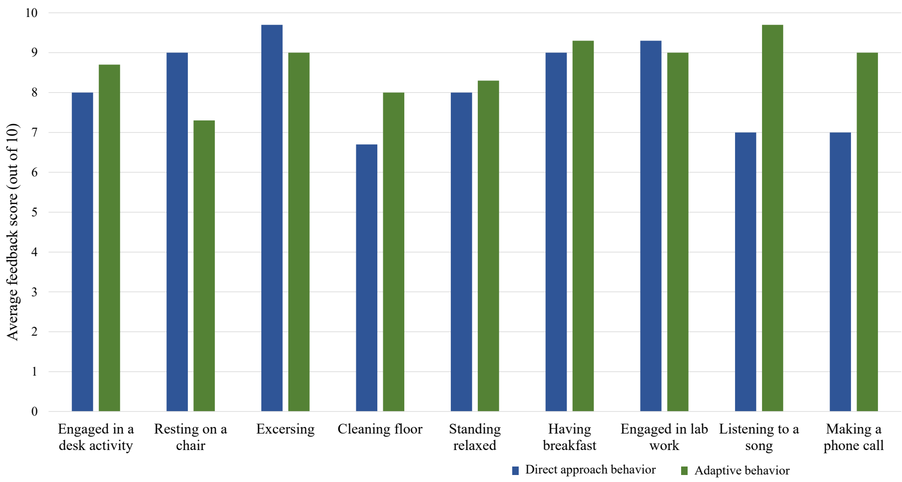


Figure 5.22: A comparison between the average feedback scores received for the behavior of the robot during only the direct approach only vs adaptive behavior after comparing both the situations; direct approach and adaptive behavior.

activity and resting on a chair, where less bodily motions could be observed, users preferred shorter conversations. This preference remained same for tasks such as exercising and cleaning the floor, where speedy and wide motions could be observed. Motions were at the two ends (very fast or very slow) when the users were highly engaged in their task. Motions of average speed and width could be observed for usual tasks where user engagement is not very high. In such scenarios, the robot used conversations such as a SER or TLK. In the tasks: 'resting on a chair', 'exercising' and 'engaged in lab work', second experiment received a lower feedback score and the difference was considerable in 'resting on a chair'. The reason behind this trend was that users preferred to talk to the robot while 'resting'. But the conversational preferences during experiment 1 and occasion 2 in experiment 2, were TLK and GRT respectively. Hence it can be observed that there is a requirement of comprehending the complete task, in addition to behavior.

A comparison of feedback scores for the occasion 1 and 2 in experiment 2 is shown in Fig. 5.23. After users were allowed to encounter with the proxemic-aware approach behavior of the robot in experiment 2, the feedback scores for experiment has been decreased considerably. It could be observed that the users were impressed with the situation-awareness of the robot. This trend had an anomaly in 'resting on a chair' as in the previous comparison. Again, the reason for this is that users preferred a longer conversation but movement-based analysis suggested an interaction limited to just a few words.

In experiment 2, a confusion matrix was created in order to evaluate user satisfaction regarding the evaluation of the robot. This is shown in Table 5.4. The experiment was designed to meet assumptions mentioned below. In the confusion matrix, Cohen's kappa value was calculated with linear weighting. Used weights were equal in the confusion matrix. Since kappa values for the systems were over 0.6, it can be proved that the systems was satisfactorily capable of making good judgments that substantially agree with those of the user. Hence the adaptive

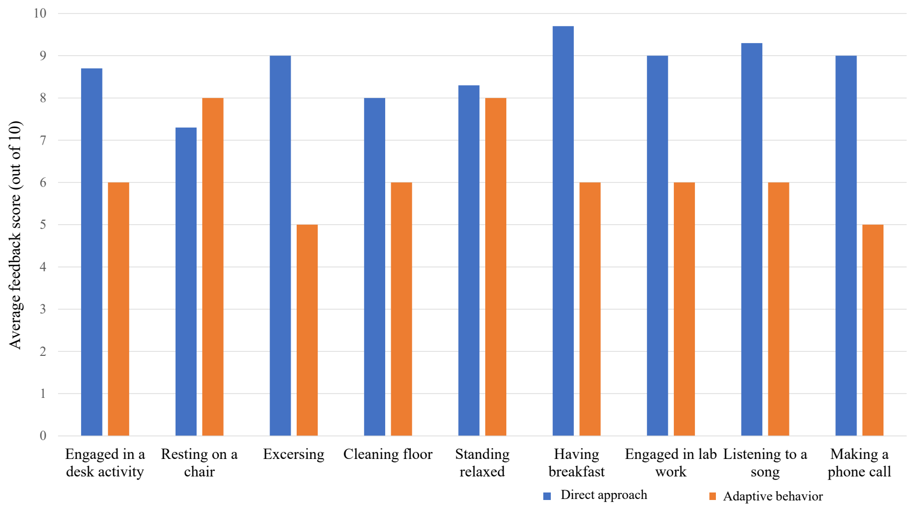


Figure 5.23: Two occasions from the experiment (a) and (b) show the direct approach behavior. (c) and (d) show the approach behavior after an analysis of the movements made by the user.

Table 5.4: An analysis of the confusion matrix generated from the results of experiment 2

Observed Cappa	Standard Error	.95 Confidence Interval	
0.722		Lower Limit	Upper Limit
Method 1	0.0841	0.5572	0.8868
Method 2	0.082	0.5613	0.8827
0.9382	maximum possible unweighted kappa, given the observed marginal frequencies		
0.7696	observed as proportion of maximum possible		

approach scenario can well replace direct approach scenarios utilized by most of the robotic platforms.

Summary: This study evaluated observable human cues such as body-based movements as a demonstrator of degree of engagement in a certain situation. For comparison purposes, responses from humans during an interaction initiated without any pre-concerns of the situation were analyzed in support of the major experiment. The experiment was intended to explore the requirement of situation-awareness of a robot deployed in a social environment. Interesting facts regarding human intentions on proxemics and conversational preferences were revealed during the study.

Results confirm the fact that the capability of the robot to perceive human behavior makes an impact to sustain interaction between robot and its user. Moreover, there are still other aspects in humans which may affect the degree of engagement in a certain situation, such as the internal state of mind, traditions, beliefs and norms followed by that human. However as these aspects are non-observable, there is still a huge space for novel methods to evaluate similar psychophysiological behavior in humans.

This proved to be helpful in determining appropriate approach behavior for the robot when a human was encountered. A robot evaluated human behavior in all physical, social and emotional aspects according to this experiment, the robot could behave with situation-awareness. Hence user dissatisfaction occurred due to inappropriate behaviors of the robot such as disturbing the user without invitation or invading the personal space have been eliminated.

5.5 Recognition of Hand and Body Pose

We present a Body Geometry-based approach for arm and body posture recognition, which was then used to determine the user situation which can be used for a number of postures and for the realtime use of a robot.

Recognized arm and body postures were used as nonverbals which communicates internal state of mind of a user. Observing a user for a certain duration gave an insight to the robot to determine how 'stressed or relaxed' the person is. Finally, based on this knowledge of user behavior, appropriate behaviors such as approach behavior and verbal responses that should be adopted by the robot were determined. Hence we tried to launch an insightful robot-initiated human-robot interaction during daily encounters in social environments.

Arm and body posture recognition techniques which can recognize an adequate number of postures with a higher accuracy are scarce. Furthermore, above systems evaluate a number of parameters for evaluating the possibility of an interaction but integration of posture and gaze factors are not adequately utilized in monitoring user behavior. Therefore, this section presents a model to interpret user situation before interaction by analyzing postural behavior by means of arm and body-based geometry derived through postures encountered during each scenario.

The proposed system adopts a novel Geometry-based approach for differentiating human arm and body postures. Basic body postures considered here were sitting, standing and bending, which were most frequently adopted by humans in domestic environments. In addition to that, frequent relaxing postures are considered as well. The same approach was applied to recognize often encountered arm postures as well. In addition to that, postural behavior of the user was monitored by the system to identify user's interest towards interaction by evaluating his/her body language. This approach enabled the robot make intuitive decisions rather than adopting a behavior just based on a predefined set of actions. This evaluation determined the appropriate behaviors for the assistive robot. We considered the nature of conversation and approach distance as the 'approach behaviors' of the robot.

Therefore this experiment had two contributions.

1) A novel method to recognize a number of arm and body postures.

2) Evaluation of the effect of considering arm and body postures to generate human-aware responses by a robot prior to an interaction. In this context, the robot's responses were related to its approach behavior which included mutual distancing and the type of conversation.

The functional overview of the overall system is shown in Fig. 5.24. This system evaluates factors that can be used to describe the nonverbal interaction demanding of a user, such as postural behavior. This 'postural behavior' includes

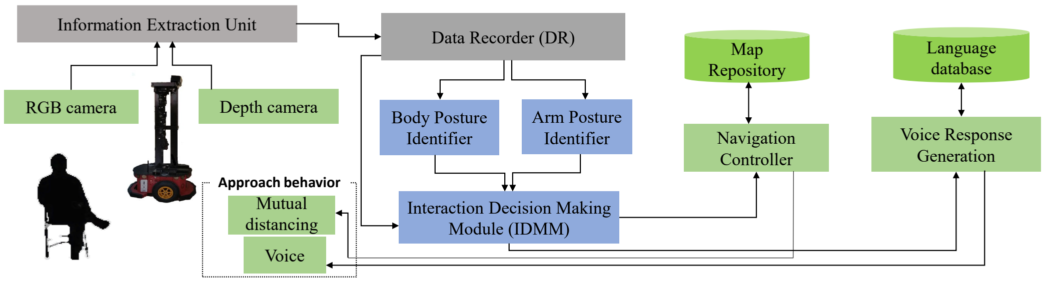


Figure 5.24: Overview of the system

the changes in body posture and arm posture of a person over time. It is intended to enhance the capability of understanding user situation for deciding whether or not an interaction is appropriate at that particular moment.

This system took spatial orientation of limbs as input to determine a particular posture. Limbs critical in forming a certain posture were used as vectors that make the posture arrangement. Joint information required to form such vectors was extracted from an individual by the Information Extraction Unit through RGB-D camera. Extracted information and calculated parameters were stored in the Data Recorder (DR) for further analysis at a later stage. After observing the user for a predefined period of time, Posture Identifiers (Body Posture Identifier and Arm Posture Identifier) were accessed to determine the current posture of its user and postural changes made by that user during the period of observation, T. Arm and body postures determined by Arm and Body Posture Identifiers were fed into the Interaction Decision Making Module (IDMM) to determine whether the postural behavior of the human is favorable for interaction. In addition to the outputs from Posture Identifiers, data recorded in DR are used to take decisions regarding interaction initiation, by IDMM.

Decisions regarding initiation of interaction with the user are then transfered to Navigation Controller and Voice Response Generation Modules to take basic steps toward interaction such as moving toward the user and greeting. Navigation Controller was used to achieve a socially interactive distance between the two conversant. In this case, conversant will be human and the robot. Maps required for navigation within the specified environment were held in Map Repository. Voice responses were generated from the Voice Response Generation Module as the final stage of initiating the interaction. Sentences for the conversation between the user and robot are generated here.

5.5.1 Changes in Behavior

To decode user behavior, the robot will observe him/her at a distance in our approach. Hence, neither the observation nor the interaction will be intrusive. For example, a person may distance herself from someone by leaning away. All such behaviors are controlled by the brain according to [88] and indicates whether a person is interested in or getting ready to avoid an encounter. Another similar example is that a person may rub his face or cheeks when in need of getting rid of a situation. Therefore hands and the body are important to be considered in reading body language. In this research we considered a number of arm and body postures to be distinguished by a robot before initiating an interaction with a human. A set of often encountered body and arm postures are shown in Fig.5.25 and 5.26 that we tried to recognize using the proposed concept. Hence the robot might be able to interact with a human without territorial violations and disturbing behaviors.

Set of body postures shown in Fig.5.25 identified by the proposed approach are abbreviate as follows.
STA- Standing
STDUP-Standing, with one leg up
LEA-Leaning forward
PIK- Picking up something
CRO-Crouch
SEA- Seated
SEAF- Seated, with spine bent forward
SEAB- Seated, with spine bent backward

Figure 5.25: A set of body postures encountered daily are shown. Each posture is called by an abbreviated form of their actual name and their actual names are explained in section 5.8.1.

MED- Meditation

SEAUP- Seated, with one leg up

BEN- Seated, with legs up and bent

Similarly, the set of arm postures shown in Fig.5.26 identified by the proposed approach are abbreviated as follows.

DWN- Hand down

FRO- Hand in front

UP- Hand up

Figure 5.26: A set of arm postures encountered in daily chorus are shown. Each posture is called by an abbreviated form of their actual name and their actual names are explained in section 5.8.1.

FOR- Hand forward

SID- Hand to side

WAI- Hand on waist

CHK- Hand on cheek

DES- Hand as in a desk activity

WAV- Waving hand

HEA- Hand on head

5.5.2 Mechanism used to differentiate postures

To detect a posture uniquely from a set of postures, orientation of vectors drawn along selected limbs in human body was considered. In other words, each limb that a posture, becomes a vector. A vector is formed by combining two adjacent joints in the skeleton. Once a person is tracked, his/her skeletal arrangement is extracted to analyze the positioning of selected set of vectors in 3D space. This is demonstrated in Fig. 5.27 and Fig. 5.28. Fig. 5.27 shows two vectors- fore arm vector and upper arm vector, which are used to determine arm posture. To determine the spatial orientation of the body, 3 major limbs; spine, femur bone and knee are considered. A vector is drawn across the limb, joining the end joints of that limb, according to the extracted skeleton. The angles this vector makes with the XZ plane (horizontal) and Y are also measured. Spine, femur and tibia vectors shown in Fig. 5.28 are the vectors used to determine the body posture. Front view of a skeletal while sitting is shown in Fig. 5.28.

Arm Postures

As already explained, three limbic vectors used to derive arm posture are marked in Fig. 5.27. Following equations are used to calculate each vector considered. A, B, and C denote the coordinate positions of shoulder, elbow and

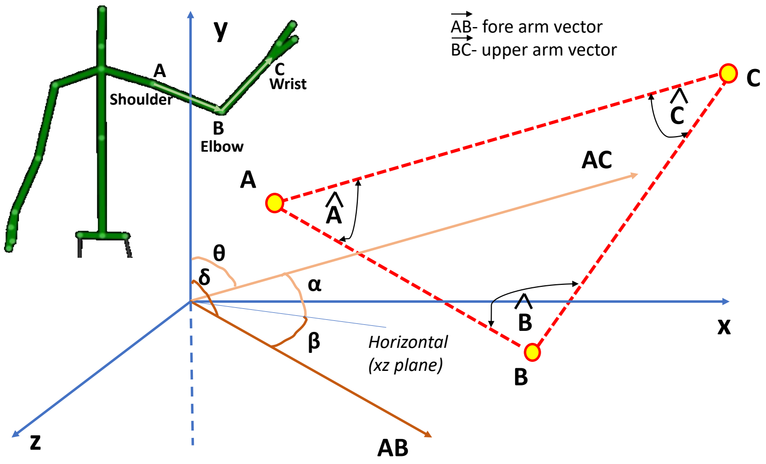


Figure 5.27: Body angles derived from the limbic vectors of a person with stretched hands are shown. Coordinate system as seen from the camera is shown as X-Y-Z. The angles formed by the 3 limbic vectors with the XZ-plane are marked as α and β while the angles made with the XY-plane are marked as θ and δ. These angles are used to recognize the arm posture. A, B and C denote shoulder, elbow and wrist joints.

wrist joints respectively.

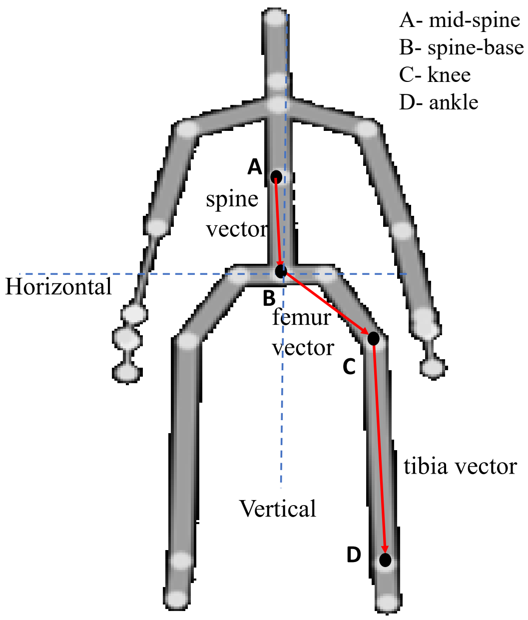


Figure 5.28: Limbic vectors of a seated person are shown here. The limbic vectors are derived by joining A (spine-mid), B (spine-base), C (knee) and D (ankle) joints.

$$\vec{AB} = (x_{elbow} - x_{shoulder})\vec{i} + (y_{elbow} - y_{shoulder})\vec{j} + (z_{elbow} - z_{shoulder})\vec{k} \tag{5.2}$$

$$\vec{BC} = (x_{wrist} - x_{elbow})\vec{i} + (y_{wrist} - y_{elbow})\vec{j} + (z_{wrist} - z_{elbow})\vec{k} \tag{5.3}$$

Where AB and BC denote the vectors drawn along the upper arm and the fore arm. x, y and z denote the coordinates in i, j and k directions respectively in usual vector notation. Labels *shoulder*, *elbow* and *wrist* denote the shoulder, elbow and wrist joints. The set of inputs required to identify the particular arm posture is represented by H as follows.

$$H_i = \{\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha, \beta \mid i = 1, 2, ..., n\} \tag{5.4}$$

Here, i denotes the i^{th} posture considered out of n number of postures. $\hat{A}$, $\hat{B}$, $\hat{C}$, δ, θ, α and β are measured as marked in Fig. 5.27. δ and θ are the angles measured from the vertical (XY-plane) drawn through the spine-base joint of the skeletal. α and β are angles measured in 3-dimensional space from the horizontal (XZ-plane) drawn through the spine-base joint of the skeletal.

For example, the set of angles made by limbic vectors corresponding to a standing human posture can be represented as follows.

$H_i = \{18°, 148°, 14°, 2°, 17°, -70°, -87°\}$ if the i^{th} arm posture is considered as an ordinary standing posture: STA with hands down: DWN as shown in Fig. 5.26.

Body Postures

Angles related to knee and femur bones are calculated for the right hand side of the body. Orientation of these three vectors is different from posture to posture. Even though other limbic vectors that change with respect to posture are present, these three vectors are critical in differentiating two postures that are slightly deviated from each other. Furthermore, majority of the human body responsible for a posture is included in the above three vectors. Vectors along the limbs in arms are not considered as hands are often moved by humans even without a specific need to do so. Therefore, arms are not considered for body posture recognition during this study. The set of angles denoted by P_i in (7) includes the angles used to calculate the orientation of limb vectors shown in Fig. 5.28.

Limbic vectors required to calculate these angles are shown in Fig. 5.29 and these vectors are calculated as follows.

$$\vec{AB} = (x_{spine\_base} - x_{mid\_spine})\vec{i} + (y_{spine\_base} - y_{mid\_spine})\vec{j} + (z_{spine\_base} - z_{mid\_spine})\vec{k} \tag{5.5}$$

$$\vec{BC} = (x_{knee} - x_{spine\_base})\vec{i} + (y_{knee} - y_{spine\_base})\vec{j} + (z_{knee} - z_{spine\_base})\vec{k} \tag{5.6}$$

$$\vec{CD} = (x_{ankle} - x_{knee})\vec{i} + (y_{ankle} - y_{knee})\vec{j} + (z_{ankle} - z_{knee})\vec{k} \tag{5.7}$$

Where AB, BC and CD denote the vectors drawn along the spine, femur and tibia limbs. x, y and z denote the coordinates in i, j and k directions respectively in usual vector notation. Labels *mid_spine*, *spine_base*, *knee* and *ankle* denote the spine-mid, spine-base, knee and ankle joints marked in Fig. 5.29.

$$P_i = \{\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2, \beta_2 \mid j = 1, 2, ..., m\} \tag{5.8}$$

Here, $\hat{A}$, $\hat{B}$, $\hat{C}$, α_1, β_1, $\hat{D}$, $\hat{E}$, $\hat{F}$, α_2, β_2 are as marked in Fig. 5.29. j denotes the j^{th} posture considered out of m number of postures. A set of actual values obtained for a specific posture are shown below.

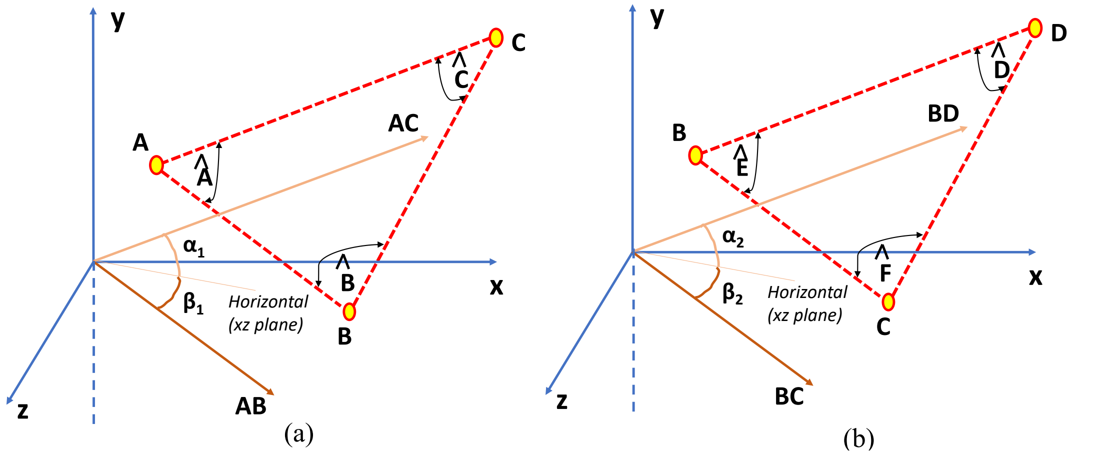


Figure 5.29: Body angles derived from limb vectors of a seated person as in Fig. 5.28 are shown here. Coordinate system as seen from the camera is shown on bottom left. The angles formed by the 3 limb vectors with the xz plane are marked as $\alpha_1, \beta_1, \alpha_2, \beta_2$. (a) Orientation of the upperbody vectors is shown by the triangle ABC (b) Orientation of the lowerbody vectors shown by the triangle BCD.

$$P_i = \{6°, 162°, 12°, -82°, -87°, 5°, 167°, 8°, -77°, -73°\}$$

if i^{th} body posture is considered as 'standing' or STA in Fig. 5.25.

Identifying postures in the above manner is a function of the Body Posture Identifier and Arm Posture Identifier. A deep neural network was trained to relate the above body angles in P_i and H_i, to a particular posture. Inputs to the neural network are $\hat{A}$, $\hat{B}$, $\hat{C}$, α_1, β_1, $\hat{D}$, $\hat{E}$, $\hat{F}$, α_2 and β_2 in Body Posture Identifier. The output from this neural network is the closest posture out of the specified m postures. This output is considered for all sets of angles recorded each second throughout T s. This is similar in Arm Posture Identifier as well. Here, the inputs to the neural network are $\hat{A}$, $\hat{B}$, $\hat{C}$, δ, θ, α and β in Arm Posture Identifier and the output from this neural network is the closest posture out of the specified arm postures.

Table 5.5: Observations made through time

time (s)	t= 1	t= 2	t= 3	t= 4	t= 5	.....	t= T
Body Posture	CHK	CHK	STA	STA	STA	.....	STA
Hand Posture	DES	DES	DES	DES	DES	.....	DES

5.5.3 Behavior Evaluation Model

Posture change over a timeline

During certain activities, a human changes his/her posture at least once. Therefore, posture changes must be analyzed for an adequate duration. The system observed its user for T s from the moment a user's skeletal was identified. DR records raw information such as coordinate positions of specific joints and calculated parameters such as joint vectors throughout this period. At the end of T s, set of information recorded in DR is analyzed for posture identification and decision evaluation. An example of this evaluation for the period of observation equal to i s is shown in Table 5.5.

In Table 5.5, arm and body postures perceived by the model in each second throughout the period of observation are shown. This postural behavior is taken into consideration prior to the interaction. Hence the decision making is affected by this 'postural behavior' of the user. The 'initial posture' was recorded at t= 1 s and the 'final posture' was recorded at t= T s.

Availability of the user

The intention of the assistive robot is to uplift the mental condition of its user, through friendly interaction at user's leisurely or lonely hours. Therefore, the robot's interaction approach must consider appropriate situations to achieve this outcome. IDMM assesses the availability of the user by analyzing the change of his/her behavior by means of posture over time. The idea behind this evaluation is as follows.

According to behavioral sciences, there's a trend in humans to change their behavior when somebody is around. Therefore, the user's response toward the robot changes, when he/she sees the robot nearby. A common example for this type of situation is that humans slant their spine (SEAB in Fig. 5.25) when seated for a long duration, but when some one comes near, they adjust their postures to standard sitting posture (SEA in Fig. 5.25) with erect spine. Humans notice such changes in behavior in someone when they intend to interact with that person. Instead of the human, it is the assistive robot who assesses such incidents in domestic environment in this context.

5.5.4 Posture based Interaction Decision Making

A number of postures are found in humans while engaged in various activities in domestic environments. Out of these, often found 11 (m=11) postures shown in Fig. 5.25 were recognized using our approach. These include 'standing' and 'seated' postures and postures in-between. Postures such as LEA and PIK can also be found while performing a certain task. These are referred to as 'intermediate' postures. Bending down to pick a fallen object is an example of a task with such postures. Awkward postures that are unique to humans, which are especially found in children, are not considered here. Postures SEAUP and STAUP have two different angles for right and left femur and knee vectors respectively. In such instances, training data sample has postures with both right and left sides included. The reason is that only the main postures are responsible to assess the availability of a human for interaction. Limbic vectors vary during each posture considered here. Similarly there are 10 arm postures (n=10) which were selected to be recognized using the proposed approach.

As a prior step before interaction, it's important for the robot to evaluate the user's posture to gain knowledge on current situation of the user. On the one hand, posture is a measure of the emotional condition of a person. On the other hand, postures are related to a user's current activity according to [89]. Therefore, the robot is expected to evaluate posture as a measure of interactivity of a particular person. Although there aren't strict rules on the relationship between posture, emotional state and physical activities, there are conventional scenarios where user situation can be approximated by analyzing posture information.

Interaction decisions are made after analyzing posture changes according to behavior evaluation model. Information required by the IDMM are given in the set denoted by H_i and P_i in (5.4) and (5.8).

5.5.5 Valence of a situation

Once the postures observed in a single scenario are known, a variable called 'valence' is calculated to assess the 'internal state' of that person quantitatively. To calculate valence, we categorized postures into 3 groups namely, 'relaxing poses', 'neutral poses' and 'stressed poses'. This is shown in Fig. 5.30. Postures adopted during relaxing situations were categorized as 'relaxing poses' and are assigned a positive valence. Similarly, postures adopted during stressed situations were categorized as 'stressed poses' and are assigned a negative valence. Standard postures were assigned the value zero for valence. This categorization was done after analyzing some simulated domestic encounters according to the generally accepted social norms. Valence for each posture was determined after this analysis and the values assigned for the valence are given in Table 5.6. Assigning valence

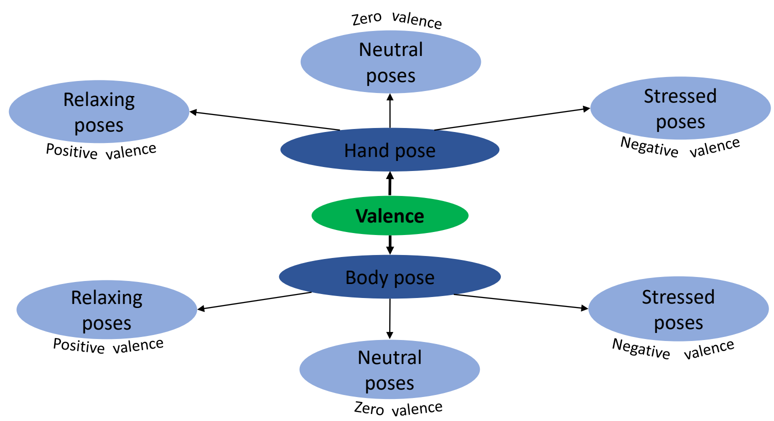


Figure 5.30: A semantic map illustrating the components that make valence; hand and body poses. Each pose is categorized into three groups; relaxing, neutral and stressed. This categorization is based on the emotions associated with each pose in general human perception.

and categorizing postures according to their emotional state were empirical and based on how humans react in social encounters in general. Hence we simulated such encounters and arranged arm and body postures in an order of their positive emotional state. Values for the valence were assigned after this step.

In a single scenario, a person adopts both body posture and hand posture. In such a case, the valence of the scenario was calculated as follows.

$$Valence = k_1 * Valence_{Body\ Posture} + k_2 * Valence_{Hand\ Posture} \tag{5.9}$$

$Valence_{Body\ Posture}$ and $Valence_{Hand\ Posture}$ indicate the valence corresponding to body posture and hand posture as given in Table 5.6. k_1 and k_2 indicate the weights given to body posture and arm posture respectively in determining the *valence* of a situation.

Interaction decisions made by the robot based on the calculated valence are shown in Fig. 5.31. The values to separate margins between each type of conversation and mutual distancing was empirical and were decided by repeating a few trial experiments before implementing the IDMM. Corresponding type of conversation and the mutual distancing appropriate for a particular instance were determined according to the value of the valence as shown in Fig. 5.31. The margins for each approach behavior are set by simulating a number of occasions in domestic encounters and human responses if a human approaches another human instead of the robot. This simulation helped in determining the values which were used for the valence in each arm and body posture shown in Table 5.6. Postures were listed in order of their relaxing or stressed nature as per the

Table 5.6: Valence assigned for each hand and body posture depending on the general interpretation of emotional state behind each posture

Arm Posture	HEA	FOR	SID	DES	DWN	FRO	UP	WAI	CHK	WAV	
Valence	-8	-6	-4	-2	0	2	4	6	8	10	
Body Posture	MED	LEA	PIK	SEAF	SEAB	STA	SEA	CRO	STDUP	SEAUP	BEN
Valence	-10	-8	-6	-4	-2	0	0	2	4	6	8

human perception and then a numerical value for the valence is assigned.

5.5.6 Decision-making in dynamic user behavior

In a dynamic user behavior, several changes in body and arm postures could be observed. In such occasions, the initial and final postures observed during the period of observation, T were considered in decision-making. The conversation types used in this work are the same as in Fig. 4.5.

For instance, if the valence was increased during a certain occasion and the approach behavior corresponding to the final valence is 'Greeting' and 1.5 m, this is incremented to 'offering a service' and 1.8 m. A combination of a mutual distance and the corresponding conversation is referred to as a 'set'. Similarly, if the difference between the initial and final valences is a decrement, the approach behavior is demoted by 1 set. For instance, if the final valence is 'greeting' and

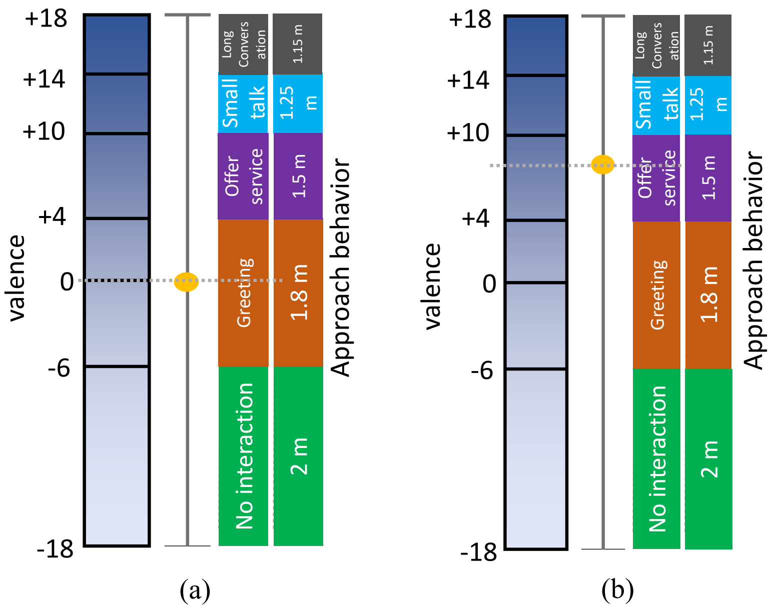


Figure 5.31: Marginal values for valence to determine appropriate approach behavior of the robot. (a) Initial valence is marked as 0 and this corresponds to 'Greeting' at a distance of 1.8 m as the appropriate approach behavior (b) Final valence increased to +8 which resulted in an approach behavior of 'asking for a service' and keeping a mutual distance of 1.5 m.

1.5 m, this is demoted to 'no interaction' and 2 m. If the difference between the two valences: Initial valence and the final valence, is an increment, the approach behavior is promoted by 1 set. This decision making criteria is given in the algorithm below.

Algorithm 1 Decide appropriate approach behavior

Require: Initial valence, final valence
Ensure: Mutual distance, type of conversation (Approach behavior)
  Arm and body posture transformations known
  if final valence = initial valence **then**
    Select the corresponding set of approach behavior
  else
    if final valence <initial valence **then**
      Select the corresponding set of approach behavior for the final valence
      Decrement corresponding approach behavior by 1 set
    else
      if initial valence <final valence **then**
        Select the corresponding set of approach behavior for the final valence
        Increment corresponding approach behavior by 1 set
      end if
    end if
  end if

5.5.7 Experiment and Results

A set of experiments were conducted in support of the proposed concepts of posture recognition and robot's decisioning.

The robot was placed at a predefined location in the map and was allowed to wander within the specified map. The routes that were covered by the robot are determined by a remote experimenter. Once a body was tracked, the robot stops its motion and starts observing. During this observation, DR stores information and Body Posture Identifier and Arm Posture Identifier are initiated. After the recognition of postures, IDMM is initialized to take interaction decisions. Once the robot decides an appropriate type of conversation and the mutual distance, this is achieved by the remote experimenter as in wizard-of-oz experiments. Hence the experimenter instructs the Voice Response Generation Module to utter the statements that he types. Navigation Controller takes care of maintaining the mutual distance decided by the IDMM. We implemented the proposed concept to the right side of the body and the corresponding gesture was performed by the right hand.

The set of experiments are explained below.

Determination of T (Experiment 01)

An experiment was conducted to determine a reasonable value for T. Therefore, a set of domestic activities were selected and a person was allowed to engage in a particular activity for a period of time. During that period, posture changes and the time gap between two consecutive different postures were recorded. For the experiment, 11 users aged from 26-59 years (SD-12.83 and mean-31.63) participated. Details of the activities selected are given in Table 5.7. Average values for the variables observed during each activity are given. Here, a single user was observed for 30 minutes unless the user walked away in the middle of observation. To determine T, we observed for only body postures because the hands are dynamic and change faster than the body in general.

Recognition of Arm and Body Postures (Experiment 02)

A deep neural network (DNN) was trained to map the set of angles made by the limbic vectors to the corresponding **arm posture**. We used a DNN of 3 layers and 50 neurons in the first and second layers. Training data set consisted

Table 5.7: Results of experiment 01

Activity	Postures observed	Average time between 2 consecutive body postures(s)	Minimum and maximum times between 2 consecutive body postures (s)
1 Standing, waiting	STA, STDUP	3.8	2, 8
2 Sitting, relaxing	SEA, SEAF, SEAB, SEAUP, BEN	5.1	3, 9
3 Desk activity(reading/studying)	SEA, SEAF	6.4	3, 12
4 Cooking	STA	6.1	4, 16
5 Making a phone call, sitting	SEA, SEAF, SEAB, SEAUP, BEN	8.4	5, 76
6 Making a phone call, standing	STA	-	-
7 Standing, engaged in lab work	STA, LEA	5.3	3, 16

of posture information of 21 human subjects. The inputs and the output of this network can be stated as follows.

Inputs: $\hat{A}, \hat{B}, \hat{C}, \delta, \theta, \alpha, \beta$

Output: Corresponding hand posture out of {DWN, FRO, UP, FOR, SID, WAI, CHK, DES, WAV, HEA}

The module: Arm Posture Identifier performed the above functions.

Another DNN was used to map the set of limbic vectors to the corresponding **body posture**. This DNN also consisted of 3 layers and 50 neurons in the first and second layers. Data related to body posture was also taken from the same subjects as for the hand posture.

Inputs: $\hat{A}, \hat{B}, \hat{C}, \alpha_1, \beta_1, \hat{D}, \hat{E}, \hat{F}, \alpha_2, \beta_2$

Output: Corresponding body posture out of {STA, STDUP, LEA, PIK, CRO, SEA, SEAF, SEAB, MED, SEAUP, BEN}

Different seating arrangements make a body posture slightly different from each other. Furthermore, people tend to keep their legs at different angles when seated. Fig. 5.32 shows three examples for this. Such instances were also considered while collecting data for the training data set.

The module called Body Posture Identifier performs the above functions.

Implementation of IDMM (Experiment 03)

Before implementing IDMM, the robot was remotely navigated towards the participant adopting the approach behavior as 'offering a service' and 1.5 m. User was asked to give a feedback score for the robot's behavior by considering user satisfaction. This is used as the ground truth for this experiment. Then the entire scenario was performed again and the valence of the set of postures was

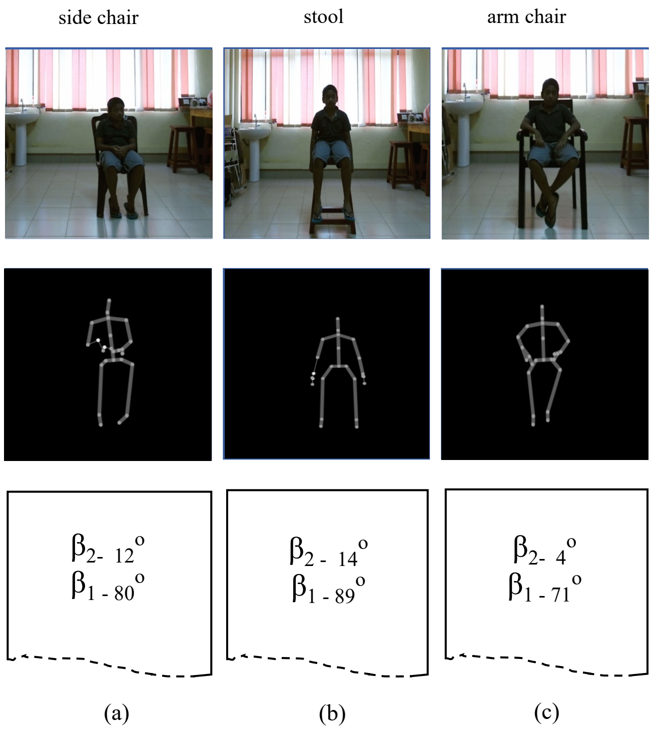


Figure 5.32: A comparison of the limbic arrangement for regular sitting posture (SEA) when the seating arrangement changes. This occurs due to the difference in dimensions of each chair type. (a), (b) and (c) represents the user posture, the corresponding skeletal image and 2 angles β_1 and β_2 for side chair, stool and arm chair respectively.

calculated for each scenario according to (8). Then the robot was commanded by a remote experimenter to adopt the approach behavior determined by the IDMM. IDMM decided upon an approach behavior by evaluating the *valence*. For the experiments, we assigned 1 for k_1 and k_2 in (8), assuming that the arm and body postures contribute equally for the current user situation. Then the participant was asked to give a feedback score for the approach behaviors of the robot during the two occasions. Users were asked the reason to give their corresponding feedback score and important remarks during reasoning are highlighted in the discussion.

5.5.8 Results and Discussion

Determination of T (Experiment 01)

According to results presented in Table 5.5, several postures could be observed during a single activity. Results obtained for a single user during the listed activities are shown in the table. All the postures that were observed from each user are listed under 'postures observed'. Number of postures encountered during an activity increases with the duration of the activity. Therefore, when the person was allowed to engage in an activity for a long duration, up to 6 postures could be observed. In contrast, during short term tasks such as standing/waiting or making a phone call, only few postures were observed. Number of posture changes were observed in 'making a phone call while sitting'. This is because the person finds more relaxing postures when the duration of the call is longer. Out of the results obtained, it was at the 8^{th} second on average, the person started to change the posture. That is the average time between two postures was 8 s. Therefore, 8 s was used as the practical value for the period of observation, T. This value was used for the period of observation in rest of the experiments.

Recognition of Arm and Body Postures (Experiment 02)

The model developed to recognize hand posture resulted in a training accuracy of 97.05% and testing accuracy of 96%. The model to recognize body posture resulted in a training accuracy of 98.5% and a testing accuracy of 97.5%. These two models were integrated to achieve the outcomes of the IDMM.

User responses prior to interaction (Experiment 03)

Table 5.8 shows the results of the experiment 03: implementation of the IDMM. 'Initial Postures' column indicates the set of body and hand postures observed

Table 5.8: Results of experiment 03

Initial Postures (Body pose | Hand pose)	Transformation (Body pose | Hand pose)	Description	Valence	Robot responses	Feedback score (IDMM)	Feedback score
1 (STA | DWN)	-	Person walking slowly.	0	Greeting, 1.8 m	9	5
2 (STDUP | WAI)	-	Person thinking, relaxed.	+8	Offering a service, 1.5 m	9.5	9.5
3 (LEA | DWN)	-	Cleaning the floor.	-8	No interaction, 2 m	5	4
4 (PICK | WAV)	-	Picking something up, then stalled to wave at robot.	+4	Offering a service, 1.5 m	10	10
5 (CRO| DWN)	-	Cleaning the floor.	+2	Greeting, 1.8 m	7.5	4.5
6 (SEA | DES)	-	Seated, relaxed.	-2	Greeting, 1.8 m	9	5.5
7 (SEAF | DES)	-	Seated, relaxed.	-6	Greeting, 1.8 m	9.5	6
8 (SEAF | WAV)	-	Seated, waves at robot.	+6	Offering a service, 1.5 m	10	10
9 (MED | FRO)	-	Meditating.	-16	No interaction, 2 m	9	6.5
10 (SEAUP | CHK)	-	Seated relaxed with one leg up. Hands on cheek.	+14	Long conversation, 1.15 m	10	5
11 (BEN | FOR)	-	Seated, hand forward.	+2	Greeting, 1.8 m	7	8
12 (STA | DWN)	(LEA | FRO)	Was seated, then stands and relaxes.	0 -> -6	No interaction, 2 m	10	6.5
13 (SEA | DWN)	(STA | FRO)	Was seated, then stands.	0 -> +2	Greeting, 1.8 m	10	6
14 (SEA | DWN)	(SEAF | DES)	Seated, then bends forward.	0 -> -6	No interaction, 2 m	7.5	7.5
15 (SEA | DWN)	(SEAB | DES)	Seated, then bends backward and relaxes.	0 -> -4	No interaction, 2 m	9.5	7
16 (STA | DWN)	(STA | WAV)	Standing, then waves at robot.	0 -> +10	Long conversation, 1.15 m	10	6
17 (LEA | DWN)	(STA | UP)	Exercising.	-8 > +4	Small talk, 1.25 m	4	9
18 (STA | DWN)	(STA | HEA)	Standing, then scratches head.	0 -> -8	No interaction, 2 m	7	6.5
19 (STA | UP)	(STA | FOR)	Exercising.	+4 -> -6	No interaction, 2 m	10	4
20 (STA | DWN)	(STA | WAI)	Standing, then engaged in deep thought.	0 -> +6	Small talk, 1.25 m	8	5.5

by the robot at the beginning of time T. 'Transformation' column gives the set of body posture and hand posture at the end of T. The column 'Description' gives a detailed image of the user situation. The valence of each scenario is given in the 'Valence' column. The type of conversation and the mutual distance are given as a set in the 'Robot Responses' column. 'Feedback score (IDMM)' shows the feedback (out of 10) received by the user in each scenario when the IDMM was implemented. 'Feedback score' gives the score given by the same user for the same occasion when the ground truth was implemented. When the user was stationary, no transformation was found in the initial set of postures and therefore the column is kept empty. When the users had posture changes over the period of observation, the finally observed set of postures was given in under the 'Transformation'.

In the scenario 1, the initial postures were (STA | DWN) and there were no transformations in the postures for the entire duration. During the scenario, the person was walking slowly. Therefore the person was standing (STA- valence 0), holding hands down (DWN- valence 0). This recorded a valence of 0 when both hand and body postures were considered, which resulted in a greeting and a mutual distance of 1.8 m according to the criteria in Fig. 5.31. As the user was walking, he preferred a greeting rather than a conversation. But once the robot spoke, he stopped to greet back. This occasion is shown in Fig. 5.33 (a). The approach behavior received a high feedback score as 9. The feedback received for the third scenario, was low compared to the previous occasions. Here, the person was cleaning the floor, where he was engaged, as perceived by the robot. Therefore the robot's decision was not to have any interaction any interaction, even though the user preferred to have a friendly conversation in such a task.

It could be observed from Fig. 5.34 that the user rated the system with IDMM with a higher feedback score in general and the mean value of this was 8.56. That of the system without IDMM was 5.52. Both the systems received a maximum feedback score of 10 while the minimum scores received by the two systems were

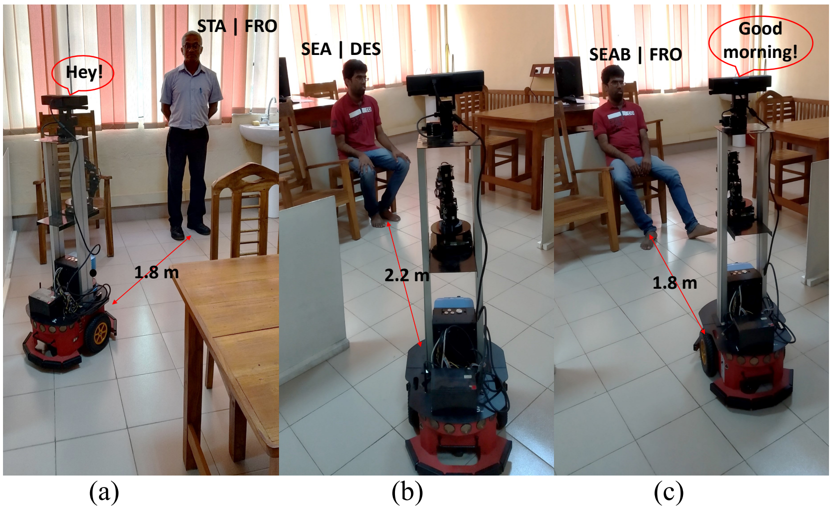


Figure 5.33: Two occasions during experiment 3 is shown. (a) Occasion 1: Responses of the user when the approach behavior: 'greeting at 1.8 m' was adopted by the robot (b) Occasion 2: Robot observes the user at a distance (c) Final position of the robot and the behavior of the user. Recognized postures and mutual distances are marked.

4 and 3.5. From the results, it could be observed that the participants preferred a much higher intelligence from the robot if possible.

Facts observed during these occasions can be summarized as follows. Interactivity of a human strongly depends on activity that the particular human is engaged in at the moment. As a clue of the activity, posture changes are

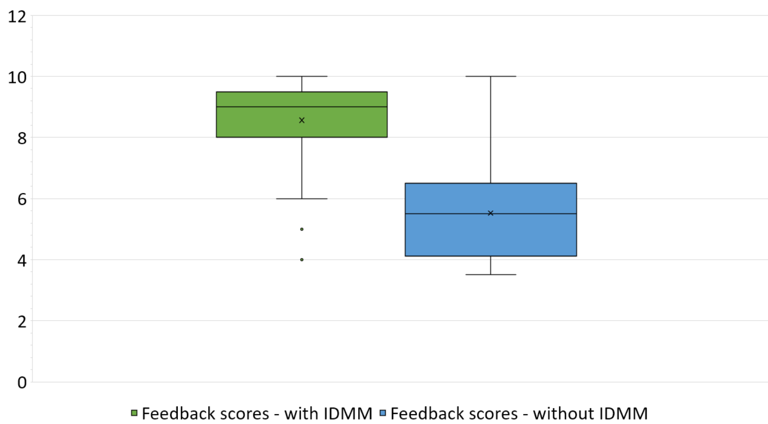


Figure 5.34: This box and whisker plot shows the mean values of feedback scores received from users for the acceptance of the robot's proactive behavior in the two occasions: with IDMM and without IDMM. Error bars represent the minimum and maximum scores received.

considered during this work. Using posture changes and gaze as behavioral measurements of human readiness toward interaction, showed positive results towards interaction decision making by a robot. In some occasions, where the feedback scores were average, the users preferred to speak more even though they were engaged. For instance, in the tasks such as cleaning or just doing nothing. In such tasks, recognition of the task, becomes important. This is an implication derived from this study. Sometimes, they did not prefer to be distracted.

User feedback was low in some occasions because it is not only the pose which contribute to determine a human's state. According to the behavior of the model it can be observed that postures such as sleeping, running can also be identified using this approach. We excluded such postures to ensure the comfort of in our experiments.

Summary: A method has been presented to identify often met human postures using a simplified method rather than using hardware intensive complex mechanisms. The proposed method was based on orientation of limbs and proved to be an efficient method to recognize a selected set of postures with a considerable accuracy and these postures cover a majority of domestic postures that are encountered during daily chorus.

After recognizing arm and body postures, the robot went through a logical argument about user's behavior before deciding upon an action. One major improvement of the system was the implementation of a novel vector-based approach for posture recognition using spatial orientation of limbs and the other was the utilization of a non-verbal mechanism for evaluation of user situation. Over the existing systems, presented system has the ability to perceive user situation using uncountable features such as posture and postural changes which are deviated from often used characteristics to measure the attention such as facial expressions and emotional responses.

As a whole, a robot could be able to generate adaptive and proactive responses based on a user's postural behavior. Furthermore the proposed system proved to be a convenient mechanism to identify a defined set of postures and to perceive nonverbal interaction demanding of a human, at a distance. This fact was confirmed by the experiments and the results presented.

5.6 Recognition of complex human activities

We present an approach to integrate several features that contribute to an activity in order to increase the probability of recognizing the correct activity. These features include behaviors adopted by a person while performing the activity, objects in the surrounding and the location of the activity. The approach is based on a semantic map drawn by relating how these features make an activity.

5.6.1 Robot, user and the context

Therefore recognizing an activity requires a logical analysis on how activity-related factors such as movements and configurations, related objects, location and attributes of objects are linked to each other. Therefore not only the performer, but also his/her surrounding takes part in an activity. Even so, there can be individual differences in utilizing such features during an activity due to various reasons. Therefore allowing some space for such differences is crucial in activity recognition approaches.

Fig. 5.35 illustrates the semantic features considered in this approach to recognize an activity. As in Fig. 5.35, five features have been semantically mapped in forming an activity. Features observed in the performer are called 'user parameters' and those observed within the environment are referred to as 'environment parameters'. These are pose, movements, related minor and major objects and the utilization of space. 'Pose' includes the body and hand poses of the individual performing an activity. Here the 'body pose' corresponds to

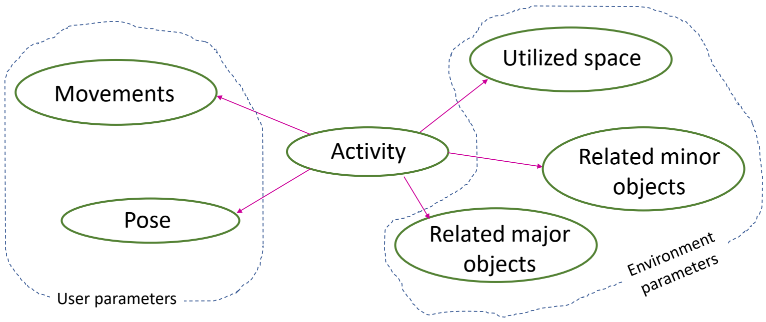


Figure 5.35: A semantic map which shows the co-relation between various aspects of a human activity. Features that can be observed during an activity are divided into ‘user parameters’ and ‘environment parameters’.

the pose of the performer and the ‘hand pose’ corresponds to the pose of his/her hands. ‘Movements’ represent the speed of the body and hands. We used the speed of the spine joint of the performer as the body speed. The reason for this is that the spine joint lies in the center when the entire body is considered and the spine makes up most of the torso which is involved in most activities. Pace of the hands was also considered as almost all the activities involve hands, in comparison to other body parts. We considered the domestic area under ‘utilization of space’ and four domestic areas; living room, kitchen and the bed room were considered here. Larger objects such as tables, chairs, beds, stools (furniture) etc were considered as ‘major objects’ and much smaller objects such as kitchenware, cooking aid, computer, stationary and books, etc were considered as ‘minor objects’. It can be interpreted that major objects accommodate minor objects. For the ease of understanding, examples for each semantic feature is given in Fig. 5.36. Here, the n number of body poses and m hand poses given in Fig. 5.25 and Fig. 5.26 were identified as a previous work of this research which is explained in section 5.8.

There are some common complexities attached to activities. Sometimes people engage in more than one activity at once. For instance, a person may cook dinner while listening to music. In such situations, recognizing both the activities is important to get an overview of the context. In addition, it is equally important to know which activity has to be given much attention. Furthermore some activities

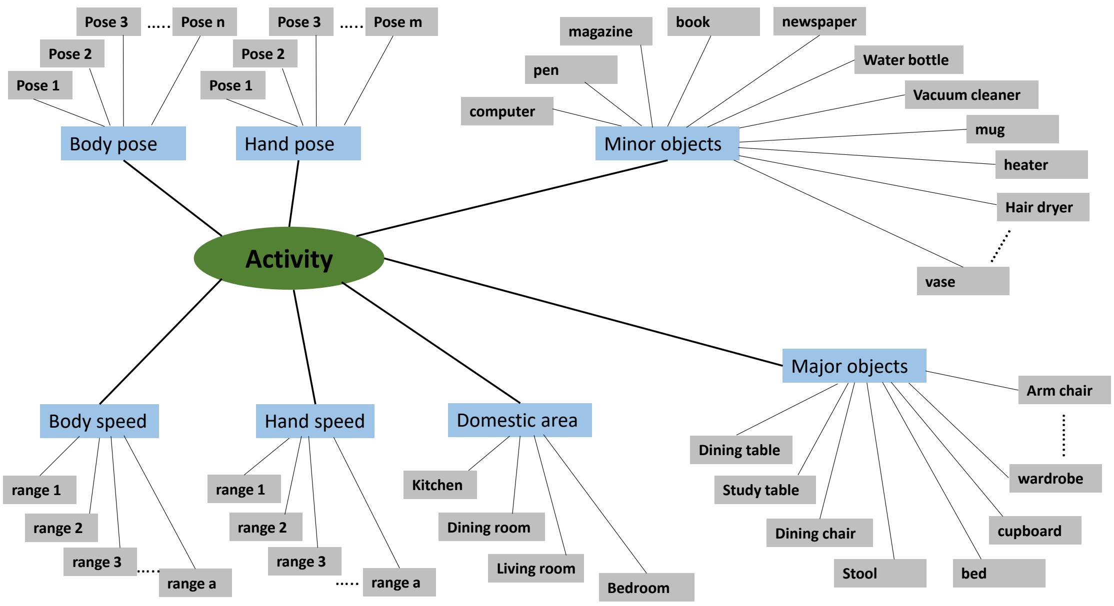


Figure 5.36: Objects and aspects related to each semantic feature are shown. Hand and body poses are numbered and, objects and domestic area are labeled for the ease of referencing. Body and hand speeds are categorized into a number of groups which include a range of speeds in one category.

will have different features over time. In such activities, the ability to forecast the activity's future is required in making activity-related decisions. Such a situation has to be closely monitored and analyzed before recognizing and forecasting a task. This analysis is sought after in distinguishing closely related tasks. For instance, doing a yoga exercise and seating relaxed may take the same form unless observed for a while. How these aspects are related to each other is shown in Fig. 5.37 as a flow diagram.

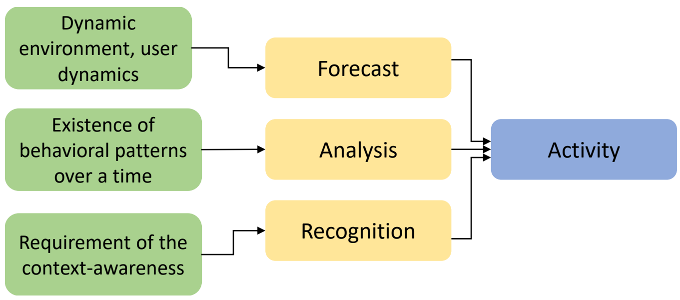


Figure 5.37: The relationship between various steps involved in an activity is illustrated here.

Therefore in our semantic approach for activity recognition, a robot makes several guesses upon an activity in case there are several possibilities for an activity. While the task having the maximum probability of being performed is selected, the algorithm outputs other possible tasks as well.

There can be more than one component in semantic features during an activity. For instance, while reading poses such as seated with spine leaned backwards (SEAB), seated straight (SEA) or seated with spine leaned forward (SEAF) can be observed as the 'body pose'. Similarly, various objects can be around as well. The person use different seats such as the arm-chair or a dining chair. Therefore we have kept space to update such information regarding an activity. Fig. 5.38 illustrates how the proposed approach works.

According to Fig. 5.38, there are three aspects of the features associated with an activity. These are forecast, analysis and recognition. For accurate prediction of an activity, the behavior of dynamic components such as the environment and

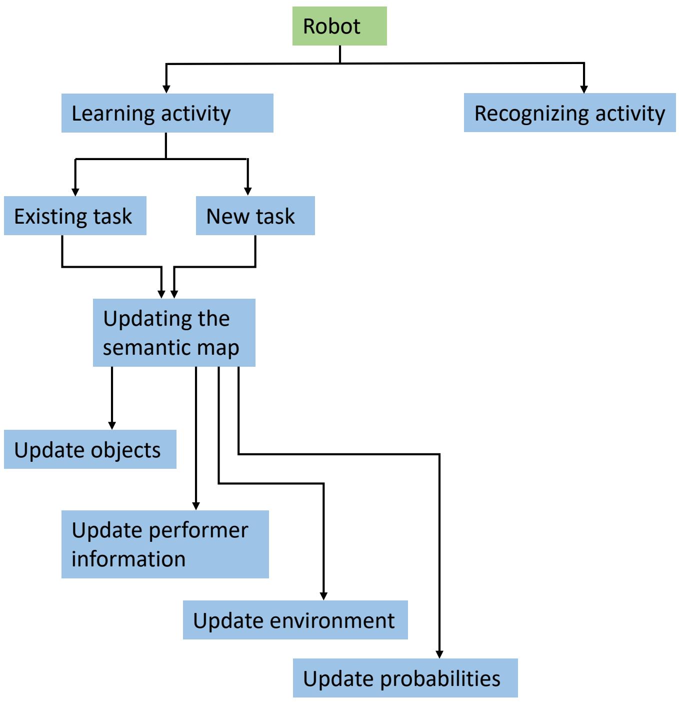


Figure 5.38: The decision flow of the approach during learning and recognizing activities.

individuals have to be forecast properly, patterns of behavior have to be analyzed, and the context have o be recognized adequately. Here, the 'context' corresponds to everything around a person under observation such as the objects being utilized, others in the surrounding, and the nature of the activity's environment.

5.6.2 Features of an activity

If the numerical value of speed is considered, there can be hundreds of different values for various activities. Besides, the speed of doing work might have differences depending on the person itself. To reduce the effect of such slight differences, we considered a range of speed for an activity instead of considering value by value. These ranges are shown in Fig. 5.39. Maximum value for the speed was determined empirically by analyzing the speeds recorded for different tasks performed by different individuals.

Assume that the set of observations made at a time during a certain activity for body pose, hand pose, range corresponds to body speed, range corresponds to hand speed, major object, and minor object are *body pose* i, *hand pose* i, *body speed_range* i, *hand speed_range* i, *domestic area*, *major object* i and *minor object* i respectively. These observations given to the system are denoted by the set S in (5.10).

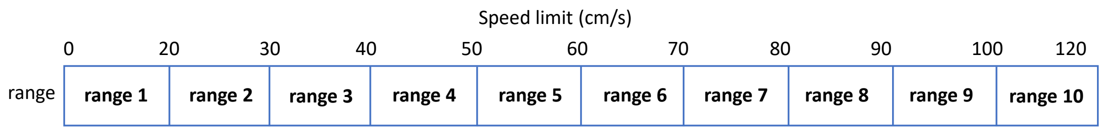


Figure 5.39: As the movements are unique to individuals, instead of considering the speed, a speed range is considered. On average we divided the range from still position to the maximum speed of 120 cm/s into 10 ranges. The maximum speed limit was found empirically by allowing people to perform several activities before the camera and recording their movement speeds.

$$S = \{body\ pose, hand\ pose, range\ corresponds\ to\ body\ speed, range\ corresponds\ to\ hand\ speed, domestic\ area, major\ object, minor\ object\ 1, minor\ object\ 2, minor\ object\ 3\} \tag{5.10}$$

5.6.3 Recognition of activities

Once these inputs are recorded, the number of appearances of the particular input in its corresponding semantic feature is calculated. Then the probability of appearance of that input is calculated as given in the following equations. The number of appearances of the input in the already recorded set of body poses is denoted by $f_{body\ pose\ i}$ and the total number of body poses recorded throughout the lifetime is denoted by $\sum_{i=1}^{n} f_{body\ pose\ i}$.

$$P_{body\ pose\ i} = \frac{f_{body\ pose\ i}}{\sum_{i=1}^{n} f_{body\ pose\ i}} \tag{5.11}$$

The same meaning of notations were used to calculate the probabilities of other semantic features as well. The probability of having the i^{th} body pose is denoted by $P_{body\ pose\ i}$ and there are n number of different body poses considered throughout the method. All the information stored when learning the tasks are searched through the 'activity database' to calculate this probability.

$$P_{hand\ pose\ i} = \frac{f_{hand\ pose\ i}}{\sum_{i=1}^{m} f_{hand\ pose\ i}} \tag{5.12}$$

Similarly, the probability of having the i^{th} hand pose is denoted by $P_{hand\ pose\ i}$ and there are m number of different body poses considered through out the method.

$$P_{minor\ object\ i} = \frac{f_{minor\ object\ i}}{\sum_{i=1}^{a} f_{minor\ object\ i}} \tag{5.13}$$

The probability of having the i^{th} minor object is denoted by $P_{minor\ object\ i}$ and there are a number of different minor objects considered through out the method.

$$P_{major\ object\ i} = \frac{f_{major\ object\ i}}{\sum_{i=1}^{b} f_{major\ object\ i}} \tag{5.14}$$

The probability of having the i^{th} major object is denoted by $P_{major\ object\ i}$ and there are b number of different major objects considered through out the method.

$$P_{body\ speed\_range\ i} = \frac{f_{body\ speed\_range\ i}}{\sum_{i=1}^{c} f_{body\ speed\_range\ i}} \tag{5.15}$$

The probability of having the i^{th} speed range is denoted by $P_{body\ speed\_range\ i}$ and there are c number of different values recorded as speed range through out the method.

$$P_{hand\ speed\_range\ i} = \frac{f_{hand\ speed\_range\ i}}{\sum_{i=1}^{c} f_{hand\ speed\_range\ i}} \tag{5.16}$$

The probability of having the i^{th} speed range for hands is denoted by $P_{hand\ speed\_range\ i}$ and there are c number of different values recorded as speed range through out the method as for the hand speed range.

According to our approach, there could be up to three minor objects depending on the context. We considered the closest three objects during this study for the ease of processing information and to avoid unrelated objects.

Assuming the semantic features are independent from each other, P(A), the probability of an activity to be the selected was calculated as follows.

$$
\begin{aligned}
P(A) = & P_{body\ pose\ i} * P_{hand\ pose\ i} * P_{minor\ object\ 1\ i} * P_{minor\ object\ 2\ i} * P_{minor\ object\ 3\ i} * \\
& P_{major\ object\ i} * P_{domestic\ area} * P_{body\ speed\_range\ i} * P_{hand\ speed\_range\ i}
\end{aligned}
\tag{5.17}
$$

When there were two activities recorded with a considerable probability, both the tasks were selected to be considered as taken place. But we chose a minimum probability for an activity to take place for real and this probability was chosen by repeating experiments over and over again. We denote this probability as $P(A)_{min}$ for future referencing. Once P(A) calculated for a set of inputs, S is greater than $P(A)_{min}$, the corresponding activity (or activities) is selected as the most probable activity that the performer was doing. If there were only one object was detected in the vicinity of the performer, $P_{minor\ object\ 1\ i}$ is calculated and $P_{minor\ object\ 2\ i}$ and $P_{minor\ object\ 3\ i}$ are omitted. In such an instance, P(A) is calculated as in (9).

$$
\begin{aligned}
P(A) = & P_{body\ pose\ i} * P_{hand\ pose\ i} * P_{minor\ object\ 1\ i} * P_{major\ object\ i} * P_{domestic\ area} * \\
& P_{body\ speed\_range\ i} * P_{hand\ speed\_range\ i}
\end{aligned}
\tag{5.18}
$$

$$
\begin{aligned}
P(A) = & P_{body\ pose\ i} * P_{hand\ pose\ i} * P_{major\ object\ i} * P_{domestic\ area} * P_{body\ speed\_range\ i} * \\
& P_{hand\ speed\_range\ i}
\end{aligned}
\tag{5.19}
$$

If no nearby minor objects were found, P(A) is calculated as in (10). Similarly, if any object was not recognized, the corresponding probability was omitted when calculating P(A).

5.6.4 Learning new activities

While learning, the robot is allowed to ask questions from its user to learn new activities. The robot initiates a conversation with its user to gather information regarding an activity which cannot be recognized using the existing information. A typical conversation between the robot and the user when learning an activity is shown in Fig. 5.40. Keywords used to identify semantic features are marked in bold letters. Once all the required inputs or the corresponding set S is identified through the conversation, a database is updated with the acquired information and the corresponding activity. This database is referred to as 'activity database'.

We used this method to identify the following 21 activities.

- Desk activity
- Sweeping
- Cleaning floor
- Standing relaxed
- Standing stressed (hands on hip)

Robot : Hello there!
Human: Hello!
Robot : Mind if I ask you a few questions?
Human: Go ahead.
Robot : What are you **doing**?
Human: I'm **reading** a **newspaper**.
Robot : I see. What other **objects** do you use while reading?
Human: A **pen** and a **cup of tea** may be.
Robot : Okay, thank you. Another question. **Where** do you prefer when reading?
Human: Here, in the **living room** and sometimes in **bed room**.
Robot : **How** do you like to read?
Human: Here in the **study table** or on the **sofa**.
Robot : Well, thank you. You helped me understand your actions in the future.
Human: You're welcome!
Robot : Have a nice day!

Figure 5.40: A typical conversation between the robot and its user when the robot learns an activity. Robot acquires relevant information by searching for the 'keywords' marked in brown.

- Cooking
- Sleeping
- Combing hair
- Seated relaxed or loll (hands hanging loosely)
- Seated stressed (hands on head)
- Making a telephone call
- Meditation
- Walking
- Ironing clothes
- Folding clothes
- Picking something up from above
- Picking something up from a surface
- Picking something up from floor
- Stretching, seated (loll or hands hanging loosely)
- Stretching, standing

We selected activities which are daily encountered in a domestic environment. It is possible to teach new activities and related features once the environment changes. For instance 'engaged in lab work' is an activity that can be observed in a laboratory environment and various objects such as electronic equipment, lap tops, soldering iron, and circuit boards etc. could be observed in the surrounding. Either the user may update the activity database voluntarily or the robot will update it itself through a friendly conversation. A typical such conversation is given in Fig. 5.40. For the ease of handling information, minor objects were labeled with its original name. For instance the laptop is labeled as ‘laptop’ in the activity database. Major objects were labeled using letters in the English alphabet as shown in Fig. 5.41. The environment consisted of four domestic areas: living room, kitchen, dining hall and bed room.

5.6.5 Experiment and Results

5.6.6 Setting

The ways of performing an activity change from community to community or the external influences such as family. Therefore two experiments were conducted to measure the performance of the system under two circumstances. One circumstance is when a robot encounters many individuals as in a social environment. This was called experiment 1. The other is when a robot accompanies members in a domestic environment, i.e. four members in a family. This was referred to as experiment 2. We used Google Voice API to initiate the conversation between the robot and the human. Voice is converted to text in order to update the activity database. We used MIROB, a social robot of height

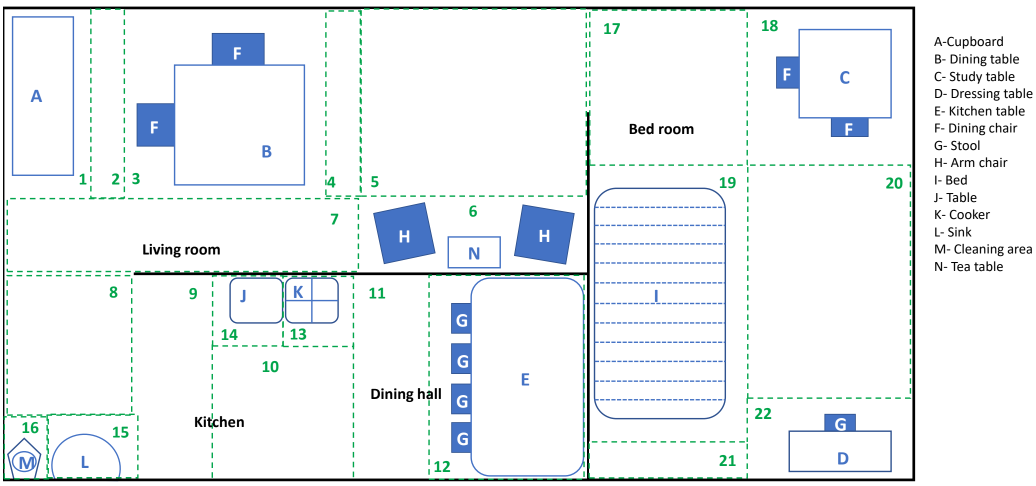


Figure 5.41: A map of the domestic area used for the study. Letters from A-M denote objects and numbers from 1-24 mark blocks of free space including and without including objects. Note: Objects are not to scale.

1.25 m developed at Robotics and Control Laboratory, University of Moratuwa. A web cam was used to detect objects with YOLO.

Experiment 1

We allowed selected 31 individuals of age 27 to 51 (mean age-31.5, SD-11.6) to perform the list of activities given in section 5.6.4. The activities were performed in a simulated domestic environment and its arrangement is illustrated in Fig. 5.41. The information in the activity database is used to recognize the activities performed by each individual. The same setting was used to update the activity database. Once we have updated the database with around 400 example scenarios (20 different occasions of each activity), we tested the recognition accuracy of the proposed approach and this stage was referred to as stage 1. As there were only a limited number of examples stored in the activity database by the end of stage 1, we allowed the robot to learn other different occasions of activities by updating the activity database manually and allowing the robot to learn by itself through conversation. Once around 750 examples were collected within the activity database, we tested the accuracy of the system with the same set of participants plus 10 more (mean age-35.5, SD-9.3). This was referred to as stage 2.

Previous two stages were performed based on the objects found by using their predefined positions as in the Map. Even so, in a social environment or a domestic area, object positions might change when they are being used. Therefore predicting that an object must be at a certain place becomes impractical in such situations. Hence we used a realtime object recognition system to identify nearby objects during an activity. We used YOLO V3 to recognize objects around a person and repeated stage 1 and stage 2 to test the accuracy of the system. We developed these two stage by expanding already collected data in the activity database, in order to cover a maximum number of scenarios. Without changing the positions of objects used for activity recognition 'with predefined objects', we added several other objects during activity recognition 'with realtime object recognition'. As a result, the existing activity database could be expanded during the activity recognition 'with realtime object recognition'.

Experiment 2

We selected 4 participants (age 26-female, 28-male, 50-male, 52-female) out of the group of participants and updated the activity database from the beginning. Then we evaluated the accuracy of the system at stage 2 with realtime object recognition. We repeated the same procedure as in experiment 1, to update the activity database and finally evaluate the system performance. As the number of participants was limited to four, we included 450 different examples in the activity database (at least 20 examples for each task) to determine the recognition accuracy.

5.6.7 Results

By repeating several trial experiments, $P(A)_{min}$ was decided to be 0.008 to confirm an activity.

Several occasions during the experiment are shown in Fig. 5.42. As an example, occasion (b) in Fig. 5.42 shows a user engaged in a desk activity. The set of observations made by the system, S corresponding to this occasion at stage 1 'with predefined objects' can be stated as in (9).

$$S = \{SEAB, DES, range\ 1, range\ 1, living\ room, C(study\ table), cup, bottle, book\} \tag{5.20}$$

The probabilities of each element in the set in activity database were {0.45, 0.8, 0.6, 0.5, 1.0, 0.5, 0.5, 0.6} respectively. This resulted in a P(A) of 0.0162 for the activity to be 'desk activity'. When all the elements of the set were considered,

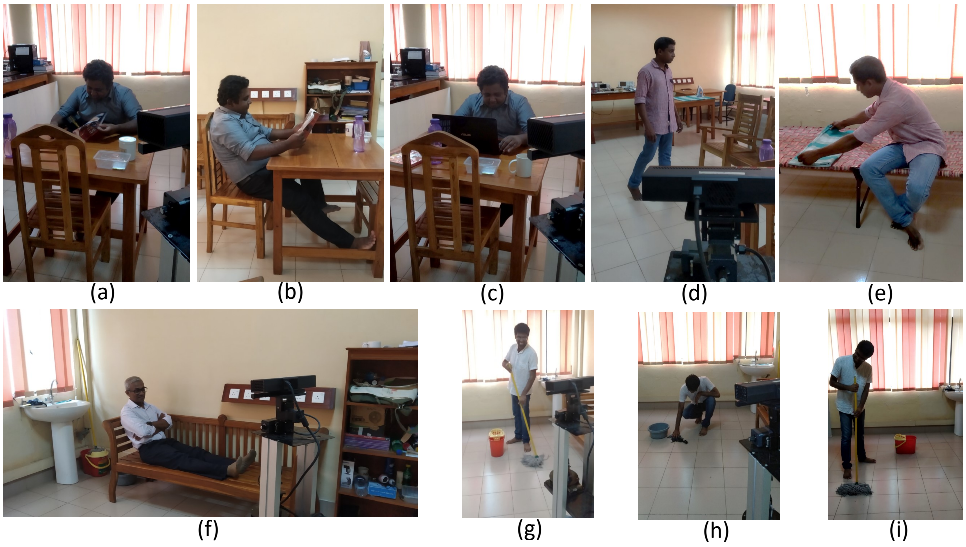


Figure 5.42: Different occasions encountered during the experiment. (a) User engaged in a desk activity (b) User engaged in a desk activity in a different posture (c) User engaged in a desk activity with different objects (A computer is added) (d) User was walking (e) User was folding clothes (f) User was seated, relaxed (g) The user was cleaning the floor (h) The user was cleaning the floor in a different posture. Bounding boxes show the objects recognized during realtime object recognition.

this also recorded a probability of 0.000814 for the activity to be 'seated relaxed', which is quite small in comparison with the highest probability recorded (0.0162). Hence the activity was recognized as 'desk activity'.

The set of observations made by the system, S corresponding to the same occasion at stage 1 'with realtime object recognition' can be stated as in (10).

$$S = \{SEAB, DES, range\ 1, range\ 1, living\ room, C(study\ table), book, bottle, \_\} \tag{5.21}$$

Here, the cup was not adequately visible for the system. Hence not recognized. But the set fulfilled the minimum requirement of at least having 4 correct elements to recognize the task. Hence the probabilities of each element in the set in activity database were $\{0.5,\ 0.9,\ 0.6,\ 0.45,\ 0.8,\ 0.2,\ 0.6,\ \_\}$ respectively. This resulted in a P(A) of 0.0116 for the activity to be 'desk activity'. When all the elements of the set were considered, this also recorded a probability of 0.00021 for the activity to be 'seated relaxed', which is even smaller than in stage 1 'with predefined objects', in comparison with the highest probability recorded (0.0116). Hence the activity was recognized as 'desk activity'. Probabilities of some elements have been reduced, in contrast to what is expected. The reason for this is that the number of appearances of new objects/ speeds have been added to the database with learning. For instances such as Fig. 5.42 (c), new objects have been introduced to the context. To recognize such contexts better, realtime object recognition is important. That is one major reason why we compared the two approaches: with predefined objects and with realtime object recognition, for performance.

Recognition accuracies of the system with predefined objects and realtime object recognition (Experiment 1) are shown in the Table 5.9. These results are plotted in a graph for the ease of visibility in Fig. 5.43. Results of the experiment 2 are plotted in Fig. 5.44.

From the results of experiment 1, it can be seen that realtime object recognition increased the recognition accuracy of activities as well. This is important especially when there are no objects involved or multiple objects are involved during an activity.

Table 5.9: Observations made through time

Activity	Recognition accuracy (%)			
	With predefined objects		With realtime object recognition	
	Stage 1	*Stage 2*	*Stage 1*	*Stage 2*
Desk activity	35	72	76	84
Sweeping	42	59	61	79
Cleaning floor	14	34	49	67
Standing relaxed	55	79	89	94
Standing stressed (hands on hip)	52	76	86	91
Cooking	18	64	66	74
Sleeping	34	76	81	86
Getting ready	15	41	63	65
Seated relaxed or loll (hands hanging loosely)	38	72	75	87
Seated stressed (hands on head)	31	54	76	83
Making a telephone call	42	86	84	89
Meditation	41	67	76	83
Walking	29	73	86	94
Ironing clothes	21	59	83	89
Folding clothes	19	47	73	82
Picking something up from above	27	78	81	88
Picking something up from a surface	31	42	68	51
Picking something up from floor	26	61	61	69
Stretching, seated	39	70	77	93
Stretching, standing	45	79	89	96

In stage 1 with 'with predefined objects', most activities were recognized with lower accuracy. The reason for this is that the database lacks adequate observations regarding each task. As a result, the recognition accuracies were between 14 and 45%. Amongst these activities 'standing stretched' received the highest recognition accuracy whereas this activity does not involve a larger variation in associated objects or postures. In contrast, 'cooking' involves a larger variation in utilized objects as well as a greater number of individual differences in postures and joint speed. Therefore recorded a lower recognition accuracy in the beginning.

As standing postures are clearly visible, the recognition was easier. Also, recognition of activities was less cumbersome, in activities involving stretched hand postures. The reason for this is that such hand postures reduce self-occlusion, which is a main obstacle to recognize a posture precisely.

In general, activities which involve standing body postures and stretched hand postures were recognized with a better accuracy than the activities with seated body postures and covered hand postures. Hence 'visibility' plays important role during this recognition process. Incorrect recognition of hand postures, body postures and objects and hardware limitations are found to be the other reasons for the difficulty in recognizing some activities.

It can be seen from Fig. 5.43, both the occasions: with predefined objects and with realtime object recognition, stage 2 recorded a higher accuracy than stage 1. The obvious reason for this is the accuracy of the system increased as it stored more and more different scenarios during each activity. In other words, the accuracy increased as the experience was greater.

People's speed of doing work differ from each other. As there are three-four people live in a social environment, teaching such deviations to robot became easier to customize robot's behavior to that particular domestic context. Hence the robot receives the capability of adapting its capabilities to match the context.

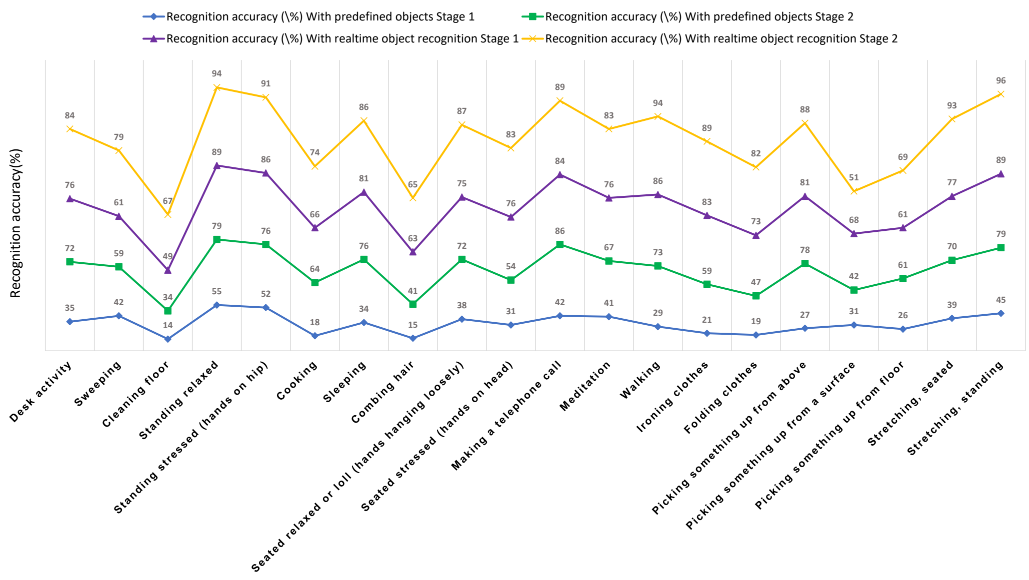


Figure 5.43: Recognition accuracies of the 4 occasions: Stage 1 with predefined objects, Stage 2 with predefined objects, Stage 1 with realtime object recognition and Stage 2 with realtime object recognition are plotted against the corresponding activity.

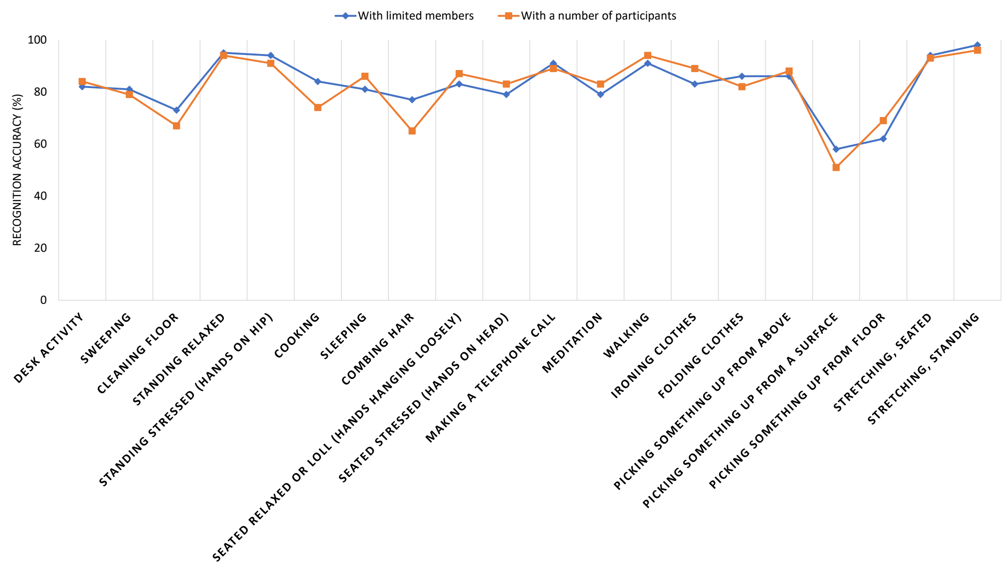


Figure 5.44: Recognition accuracies of the two occasions: with limited number of participants and with a set of participants (stage 2 with realtime object recognition) are plotted against the corresponding activity.

The system showed a satisfactory performance in recognizing most activities when the context was adequately visible. In other occasions, one major hindrance for the system performance was self occlusion and hardware limitations of vision cameras to recognize people in reduced lighting conditions.

From Fig. 5.44, it can be seen that some activities that were difficult to be recognized in experiment 1, showed an improvement in recognition in experiment 2. Hence this personalized approach was successful for activities such as 'cleaning floor (73%)', 'cooking (84%)' and 'combing hair (77%)' which were recognized before with accuracies of 67%, 74% and 77% respectively. The recognition accuracies of the rest of the activities were not very deviated from that of the experiment 1. Hence it can be seen that some activities have personality traits which have to be distinguished by an automated agent. Still in activities such as 'combing hair', recognition of the object (e.g: the comb) was difficult due to occlusion. Hence were not recognized with a higher accuracy as the activities which do not involve objects.

5.6.8 Conclusion

Most recent activity recognition techniques rely on spatio-temporal behavior and trajectories of a performer during an activity. Although such techniques recognized some tasks with a reasonable accuracy, maintaining that accuracy for complex tasks becomes difficult. Furthermore new tasks are being added to our lives as new technologies invades our homes. Therefore we need a more structured and intelligent representation of activities rather than looking for patterns. We introduced a semantic space which includes features from the performer: pose and movements, features from the surrounding: the location where the activity is performed, and features from objects: related major and minor objects. We consider these features could capture meaningful details of an activity based on the context not being restricted to the performer itself. Furthermore we kept space to introduce new activity as they come along and update related semantic features. This enhanced the capability of the system and the understanding of the surrounding in the process of recognizing an activity.

Chapter 6

DECODING NONVERBALS AND MODELLING USER BEHAVIOR

A model is developed to perceive user attention considering a number of important parameters that humans deploy to show attention towards someone nearby. These include major components of nonverbal behavior such as gaze, changes in pose and friendly gestures. In addition, above major categories are sub divided into minor properties as well. For example, gaze is analyzed in 3 sub categories as gaze direction, gaze time and whether the gaze is returned to original position by the user.

6.1 Identification of Gaze

The system recognizes gaze, friendly gestures and change in pose after the user has been notified of robot's presence. Out of these, gaze is analyzed under 3 sub categories as gaze time, gaze direction and gaze return. Gaze direction is zero if the user is looking directly at the camera mounted on robot. Gaze return is the extent to which the user turned his head back to original direction after looking at the robot. In the same way, friendly gestures are sub divided into happiness, gesture speed and gesture time wherein gesture speed and time are not involved with smile. How friendly gestures are evaluated is given in section 5.5. Change in pose is evaluated separately without sub categories. These 3 are analyzed by a fuzzy system to evaluate the level of attention given by the user in each scenario.

6.1.1 Determination of Gaze direction and Gaze time

Gaze is determined by the deviation of user's yaw angle of the head. This is taken from the point of view of camera mounted on robot. Yaw of the head of the person's skeletal tracked by the robot is denoted as θ_1.

$$|\theta_1| \leq \theta_{gaze\_lim} \tag{6.1}$$

where θ_{gaze_lim} is the angular limit for the gaze direction to be towards robot. If the condition in (1) is fulfilled, the system identifies the user to be looking at the robot.If the user satisfies the above requirement for t_{gaze} seconds, the gaze time will be t_{gaze}. The total duration system extracts information is t where $t_{gaze} \leq t$.

6.1.2 Return of Gaze

When the user is not interested in interacting with the robot, he will look away though he first looked at the robot.

$$|\theta_{1\_final} - \theta_{gaze\_lim}| \leq \Delta\theta_{lim} \tag{6.2}$$

Here, θ_{1_final} is the yaw at t^{th} second and $\Delta\theta_{lim}$ is the angular limit after which the gaze is decided to be returned. Above scenario is analyzed using the criteria given by (2). Gaze is not returned while the above criteria is fulfilled. In situations where the user looks at the robot and then looks away, many robots interpret such gaze as a favorable factor for interaction. But in real life,the condition is the opposite. Analyzing the above initial and final angular values, avoids a situation in which the robot tries to interact without the consent of its user. $|\theta_{1\_final} - \theta_{gaze\_lim}|$ decides the extent to which the user has turned his head back. All the 3 parameters, gaze direction, gaze time and return of gaze are analyzed by a fuzzy system to interpret gaze level. Gaze level is the extent to which the person's gaze is favorable from the point of view of an observer.

6.2 Identification of Friendly Gestures

Smile and waving hand are used as the parameters which defines friendly gestures. Built in open/closed palm property in Kinect is used as one condition to look for waving hand.

$$y_{wrist} - y_{handtip} \geq y_{lim} \tag{6.3}$$

where y_{wrist} and $y_{handtip}$ are y coordinates of wrist joint and hand tip respectively as seen by the camera. y_{lim} is the angular limit after which the gesture will not be considered as waving hand. The requirement in (3) must be fulfilled together with open/closed palm property for 'waving hand' gesture. Smile is detected using positions of the facial features in color and infra-red space, especially mouth left corner and mouth right corner.

Gesture_speed is evaluated by movement of open palm across the X-plane within a second of time. The speeds through out t is considered and average value is used as the input for fuzzy system. For example, *gesture_speed* at the 1^{st} second out of t duration is shown in (4) below.

$$Gesture\_speed = \{x_{wrist}\}_{t=1} - \{x_{wrist}\}_{t=0} \ cm/s \tag{6.4}$$

Gesture_time is the duration at which hand waving is detected. *Happinenss* is measured using smile and waving hand. *Happiness* is considered 'high', if both smile and waving hand are present. *Happiness* is 'average' if only one out of smile and waving hand is present. *Happiness* is 'low' if none of the two are present. All the three parameters, *gesture_speed*, *gesture_time* and *happiness* are required to evaluate gesture level of the user. Here, gesture level is the extent to which the person's friendly gestures are favorable for an interaction with an outsider.

6.3 Identification of Changes in Pose

When a human notice another human expecting interaction with him, involuntarily he changes pose. This is analyzed as an input for the fuzzy system in ALE. This fuzzy system takes, changes in pose into account with gaze level and gesture level talked in the previous section. This scenario is illustrated in Fig. 6.1. Skeletal corresponding to the particular pose is shown on right of the image. Vectors shown in Fig. 6.2 are used to determine how the initial skeletal is changed from the final. IEU keeps track of the following vectors for further

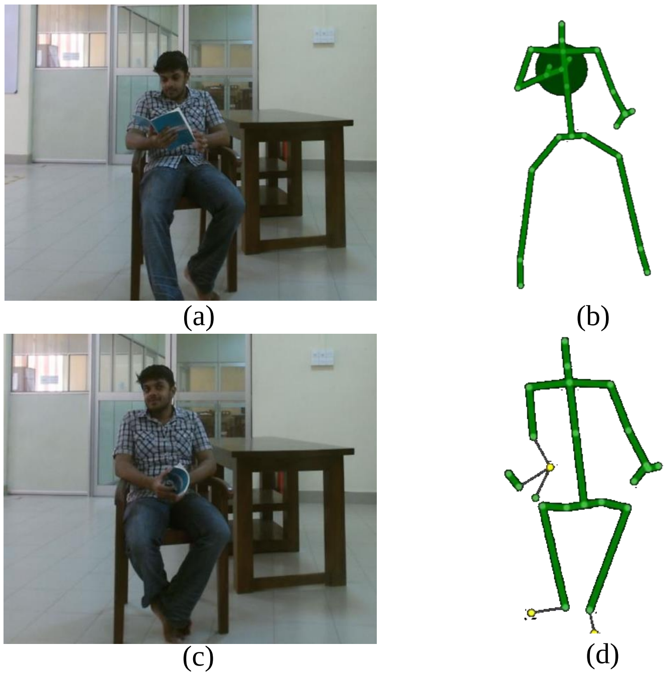


Figure 6.1: Change of behavior encountered before and after approaching a human. In (a) Initially the user is engaged (b) skeletal which corresponds to that pose (c) change of pose when somebody reaches (d) Skeletal of the final pose.

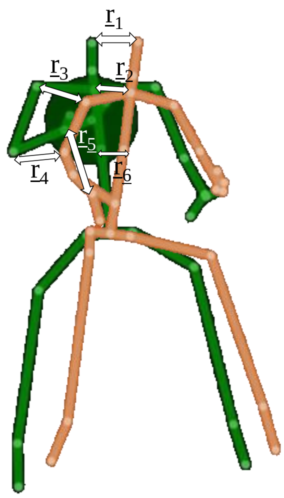


Figure 6.2: Spatial vectors related to the body joints used to differentiate two poses of the same user. The initial pose and the final pose are shown in green and orange respectively.

analysis in GRU. Vectors shown in the figure are expanded in (5)-(10) below.

$$\underline{r}_1 = \Delta x_{head}\underline{i} + \Delta y_{head}\underline{j} + \Delta z_{head}\underline{k} \tag{6.5}$$

$$\underline{r}_2 = \Delta x_{neck}\underline{i} + \Delta y_{neck}\underline{j} + \Delta z_{neck}\underline{k} \tag{6.6}$$

$$\underline{r}_3 = \Delta x_{shoulder}\underline{i} + \Delta y_{shoulder}\underline{j} + \Delta z_{shoulder}\underline{k} \tag{6.7}$$

$$\underline{r}_4 = \Delta x_{elbow}\underline{i} + \Delta y_{elbow}\underline{j} + \Delta z_{elbow}\underline{k} \tag{6.8}$$

$$\underline{r}_5 = \Delta x_{wrist}\underline{i} + \Delta y_{wrist}\underline{j} + \Delta z_{wrist}\underline{k} \tag{6.9}$$

$$\underline{r}_6 = \Delta x_{spine}\underline{i} + \Delta y_{spine}\underline{j} + \Delta z_{spine}\underline{k} \tag{6.10}$$

For further clarification, consider vector $\underline{r}_1$. If i, j, k components of the vector satisfy below conditions, a considerable change has been observed in user's pose. Here,Δx,Δy and Δz are differences between final and initial x, y and z coordinates of the particular vector.

$|\Delta x_{head}| \geq \Delta x_{head\_lim}$, $|\Delta y_{head}| \geq \Delta y_{head\_lim}$,

$|\Delta z_{head}| \geq \Delta z_{head\_lim}$

The other 5 vectors are also analysed in the same way as above. Vectors are related to joints in the upper body which are most probably involved in pose changes. Number of vectors which satisfy the above condition is selected as the input for the fuzzy system. Angular and magnitude limits for condition checking in this research are derived through experimentation.

6.3.1 Evaluation of attention

Impact of gaze, friendly gestures and change in pose towards the attention of the user is evaluated from a fuzzy system. Before evaluation of these 3 parameters, gaze and friendly gesture levels are evaluated by another fuzzy system. Inputs to evaluate *gaze level* are illustrated in Fig. 6.3. Triangular membership functions are used for the ease of handling information, as the system is involved with a number of parameters. In the same way, Fig. 6.4 shows the input parameters and membership functions used to determine the *friendly gesture level*. *Gaze level* and *friendly gesture level* are used with *change in pose* as inputs for the second fuzzy system. Evaluation of this system is shown in Fig. 6.5.This system evaluates the overall behavior related to attention. Output membership functions of the system are illustrated in Fig. 6.6.

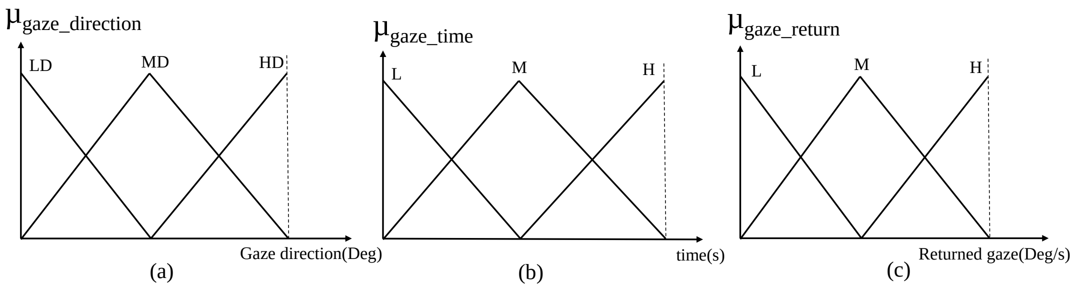


Figure 6.3: Input membership functions of the gaze parameters (a) Gaze direction (b) Gaze time (c) Level of returned gaze Fuzzy labels: L, M, H, LD, MD, HD denote Low, Medium, High, Less deviated, Medium deviated and Highly deviated respectively.

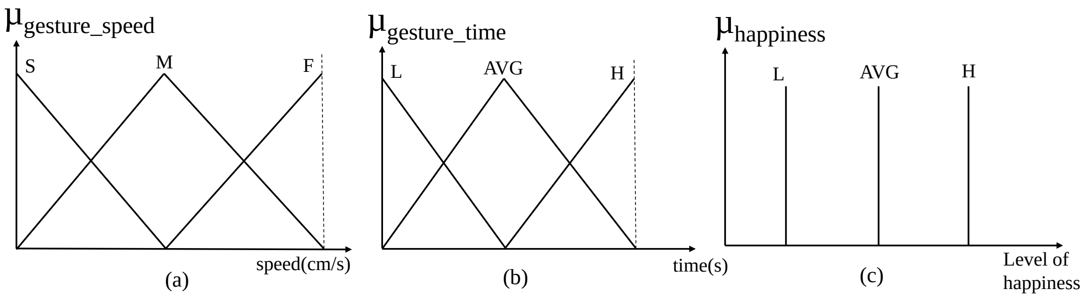


Figure 6.4: Input membership functions of the gesture parameters (a) Gesture speed (b) Gesture time (c) Level of happiness Fuzzy labels: S, M and F for Slow, Medium and Fast respectively for Gesture Speed. L, AVG and H denotes Less, Average and High for Gesture time. L, AVG and H denote Low, Average and High respectively for Level of Happiness.

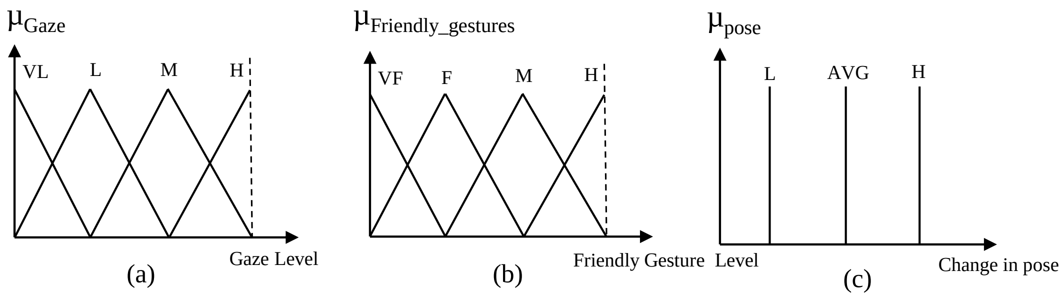


Figure 6.5: (a) Output from Gaze parameters (b) Output from Friendly gestures (c) Change in Pose. These are used as the input parameters of second fuzzy evaluation Fuzzy labels: VL, L, M, H denote Very low, Low, Medium and Very high for Gaze Level. VF, F, M, H for Very few, Few, Medium and High in friendly gestures. L, AVG, H denote Less, Average and High for Change in Pose.

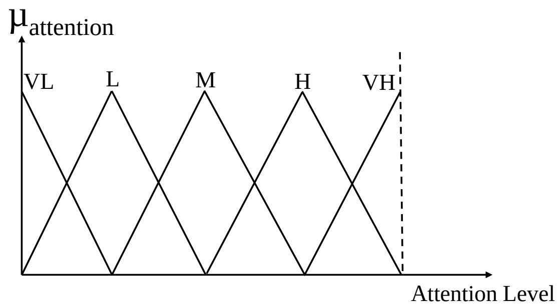


Figure 6.6: Output Membership Functions of second fuzzy evaluation Fuzzy labels: VL, L, M, H, VH, VVH denote Very low, Low, Medium, Very high and Very Very High.

6.3.2 Interpretation of Fuzzy Parameters

Gaze level shown in Fig. 6.5(a) is evaluated from the input parameters in Fig. 6.3. The rule base for this is shown in Fig. 6.7. Some of the rules are omitted in actual implementation. For example, When the *gaze_direction* is 'low', *Gaze_return* automatically becomes 'medium' or 'high'. Therefore it does not become 'low' at the same time. When the *Gaze_direction* is 'high', *Gaze_return* does not become 'high' at the same time. Rule base for *Friendly Gesture Level* evaluation is also performed in the same way. Rule base to evaluate *attention level* is given in Fig. 6.8. Boundaries for membership functions were determined through implementation.

6.3.3 Decision Making

Final decisions are taken by considering *gaze level*, *friendly gesture level* and *change in pose* to evaluate *attention level*. Input parameters of this fuzzy system are shown in Fig. 6.5 and the output membership functions are illustrated in Fig. 6.6. The type of interaction with the user may vary according to the

	Gaze Direction								
	LD			MD			HD		
Gaze return	L	M	H	L	M	H	L	M	H
Gaze time									
L	-	L	VL	M	L	L	M	M	-
AVG	-	L	L	M	L	L	H	M	-
H	-	M	H	M	M	L	H	H	-

Figure 6.7: Rule Base to evaluate Gaze Level

	Gaze Level															
	VL				L				M				H			
Friendly gesture level	VF	F	M	H	VF	F	M	H	VF	F	M	H	VF	F	M	H
Change in pose																
L	VL	VL	L	M	L	L	M	H	L	L	M	H	M	M	H	VH
AVG	VL	VL	L	M	M	M	M	H	M	M	M	VH	M	M	H	VVH
H	VL	L	M	H	M	M	H	H	M	M	M	VH	M	H	VH	VVH

Figure 6.8: Rule base for Attention level evaluation

attention level determined by the robot. Factors such as whether to reach the user or not, duration of conversation etc. can be determined from the measured *attention level*.

6.3.4 Experiment and Results

The experiment is carried out with 11 persons in a broad age gap from 25 to 58 years (Mean of 35.4 and SD of 11.4).The robot was placed at several locations of a domestic environment and it was allowed to wander within the specified map. Once a body is tracked, the robot stops its motion and records extracted information for consecutive 8 seconds which is denoted by t in previous sections. Two occasions during the experiment are explained.

Results obtained from the proposed system

Results obtained by this method for several daily encountered tasks are given in Table 6.1. Here, few tasks are considered due to the fact that the parameters considered in the system does not vary much depending on the task. Even so, evaluation of gestures, return of gaze can nullify the error occurred by not recognizing the tasks individually. Out of the 11 participants, the variation of their decisions for the same task if another human was engaged in each task is shown in the column named 'Humans with same decision'. In this scenario, the situation of the user depends on the task. Extracted information and calculated parameters are given. Approach decision to reach the user, making the mutual distance 1.5 m , is taken only when the interaction decision is 'yes'. Robot decides to interact only during 'Low', 'High' or 'Very high' attention levels are encountered. The system decision is compared with the decision of same group of participants. In this case, a participant is allowed to observe the human user behavior and his decision on initiating an interaction is recorded. In the same way, all the participants are allowed to take their own decisions regarding the

Table 6.1: Experiment Results

Situation	Yaw (°)	Gaze return (°)	Gaze time (s)	Gestures	Speed (°/s)	Gesture time (s)	Changing vectors	Attention level	Approach Decision	Humans with same decision (out of 11 participants)	Accuracy
1 Sitting on a chair, task free	25	-	-	-	-	-	1	VL	-	11	100%
2 Sitting on a chair, task free	15	-	-	smile	-	-	2	L	1.5m	9	81.80%
3 Sitting, reading a book	0	45	2.1	-	-	-	0	L	-	10	90.90%
4 Working on a table	0	5	3.05	smile, waves hand	14	1.6	3	H	1.5m	10	90.90%
5 Working on a table	5	40	1.1	-	-	-	4	H	1.5m	4	36.40%
6 Sitting, in a call	0	40	2.3	smile, waves hand	26	1.8	2	M	1.5m	0	0.00%
7 Walking around	0	-	3.1	smile	-	-	6	H	1.5m	9	81.80%
8 Walking around	0	5	1.9	waves hand	24	2.1	6	VH		10	90.90%
9 Walking around	35	40	1.7	smile	-	-	6	L	-	11	100%
10 Walking away	45	-	-	-	-	-	6	VL	-	10	90.90%

suitability of an interaction during each task. During 'Sitting and making a call', human users did not try to interact, but the user's behavior made the system decide it as a favorable occasion for interaction initiation due to friendly gestures. This type of confusions are present due to the incapability of the system to identify tasks separately. During task 5, 'working on a table', due to the considerable change in pose, 4 users decided it as a favorable occasion for interaction if there's a necessity to interact. Otherwise the decision of the robot contradicts normal human behavior. An example scenario from the experiment is shown in Fig. 6.9.

Comparison of results of a human study and the system

A human study was conducted with the same 11 participants, to identify human tendency in the same scenario, if a human was in place of the robot. In other words, the human-robot interaction scenario was compared with a human-human scenario. This is shown under 'Humans with same decision' column and percentage accuracy under 'Accuracy' column in Table 6.1. Despite the stated issues, the system performed satisfactorily during interaction scenarios.

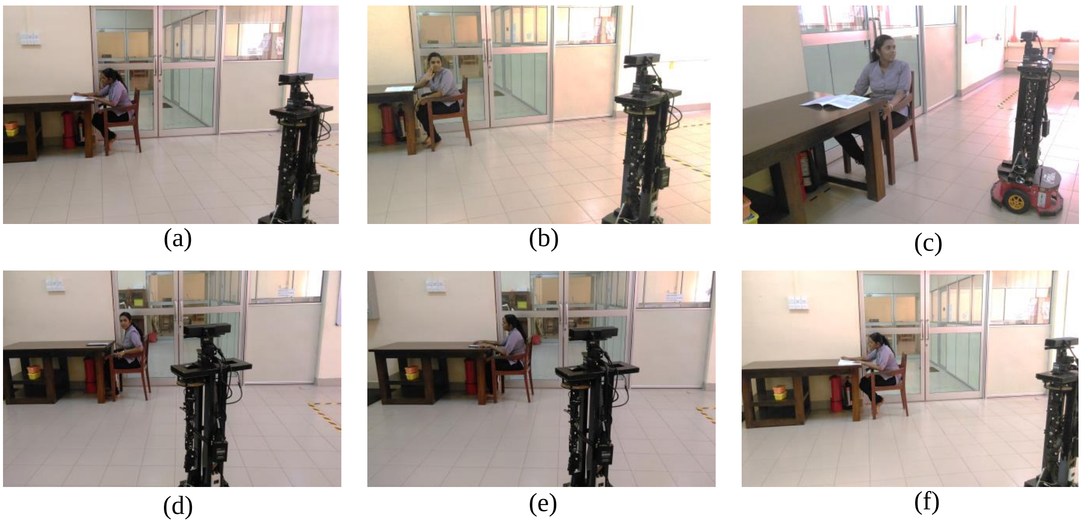


Figure 6.9: Occasion 01 (a) User at the time of detection (b) User smiles and waves hand when she notices robot's presence (c) Robot approaches user Occasion 02 (d) User sees the robot (e) User looks away (f) Robot does not approach the user.

6.3.5 Conclusions

A method has been introduced to interpret a human's attention towards the presence of a robot wherein the robot approaches the user with the intention of initiating an interaction. The intention of the robot is starting a conversation with the user by analyzing posture and body movements related to joints. Before starting a conversation, the robot will be able to assess the attention given by the user towards robot. The major improvement of this system is making use of a nonverbal mechanism to identify favorable human behavior for interaction. Furthermore, the proposed system does not rely on verbal instructions from the user for interaction initiation. Results show that the system is capable of understanding user situation without verbal communication with the user, as in many commonly used service robots. Furthermore the system was able to perceive user attention towards robot satisfactorily through just observation.

Chapter 7

DETERMINING ROBOT'S PROACTIVE BEHAVIORS: PROXEMICS

This work is about how proxemic behavior of a social robot is determined through a probabilistic evaluation of activity space of a person. Through this work, an effort was taken to minimize the problem of formulating a relationship between task related human behavior and appropriate proxemics for his/her robot companion. This work includes an analysis of the usage of local areas within the activity space to determine probable areas to approach a human without causing any disturbance to that person. This work demonstrates a design space for proxemic-based developments in social robotics. We used the concept of 'activity zones' discussed in section 5.1.1 to implement this approach.

We present a geometry-based evaluation of wrist movements of a human in order to predict the movements of that human to determine an appropriate approach behavior for a robot which intends to initiate an interaction with that human. Unlike most of the present methods, a robot observes its user for a specific duration before making decisions regarding its approach behavior through this method. The approach direction, orientation and mutual distance to be kept are included in the 'approach behavior' in this method.

7.1 Probabilistic analysis

After the frequency of visits to each zone is calculated at the end of t s, the probability of each zone to be visited during the same activity under same conditions - **probability of occupancy**, is calculated. (1) shows the equation used for this evaluation.

$$P_{zone\_i} = \frac{f_{zone\_i}}{\sum_{j=1}^{51} f_{zone\_j}} \tag{7.1}$$

where i=1,2,...,51. Here, P_{zone_i} is the probability of occupancy of zone i at the end of t s. f_{zone_i} and f_{zone_j} denote the frequency of visits to zone i and j respectively during t. The values obtained for P_1 to P_{51} are compared before making interaction decisions. Often, only a few zones are visited during a certain task. Therefore the probability of the rest of the zones will be zero. Fig. 7.1 shows the probabilities of occupancy for each zone visited while the user was 'standing'. In this scenario, zones 11, 36 and 51 have been occupied for 56%, 23% and 21% of the time respectively. Hence, $P_{zone\_11} = 0.52$, $P_{zone\_31} = 0.23$ and $P_{zone\_51} = 0.21$ according to (1). This probability provides the tendency of the user to initialize the same zone again during the task.

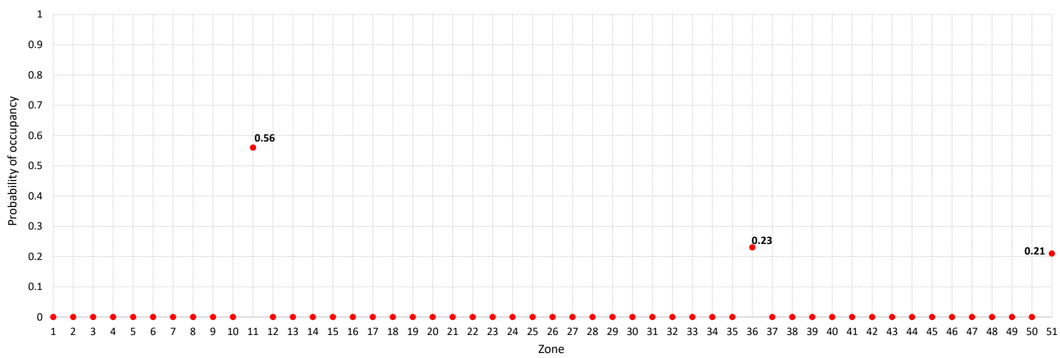


Figure 7.1: Approach of QASA. It records the zones each joint occupies and finally calculates the percentage visit to each zone. The observation was made upon an individual who was standing, relaxed.

7.2 Decision making criteria

Movements of the individual are recorded in the form of probabilities calculated using (1) for the ease of figurative or quantitative comparison.

The distance between robot and user-d was calculated with respect to the numerical value of the highest probability of occupancy after observation. The equations used to calculate this distance is shown in (2) and (3). It is assumed that a minimum gap of 1.0 m has to be kept in order to follow the social norms. This marginal value for interpersonal distance- 1.0 m was chosen as in [45]. If i is the zone with the maximum probability of occupancy and P_{zone_i} is the corresponding probability of occupancy,

if the highest occupancy is in $front$,

$$d = 1.0 + P_{zone\_i} * 1.0 \quad m \tag{7.2}$$

if the highest occupancy is in $far\ front$,

$$d = 0.2 + 1.0 + P_{zone\_i} * 1.0 \quad m \tag{7.3}$$

These equations were formulated so that an adequate mutual distance is kept to least distract the highly occupied zones.

The orientation of robot after approaching the user is calculated in a similar manner. The orientation of the robot at two occasions are shown in Fig. 7.2. In ordinary scenarios, maintaining an orientation close to 90° is rare. Therefore we chose a range from $0 - 80°$ for θ.

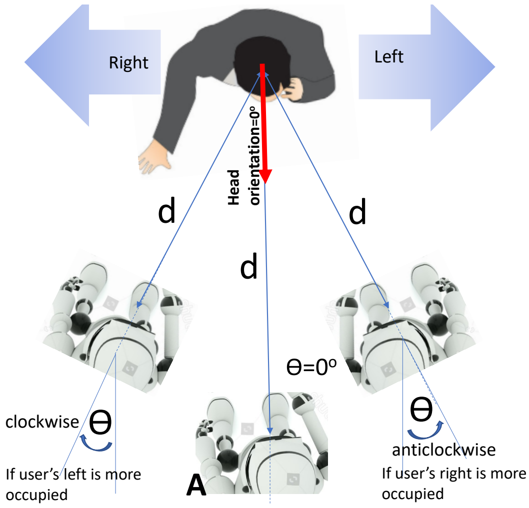


Figure 7.2: The top view of the HRI scenario is shown. The orientation of the robot is marked as θ. θ is measured clockwise (if reaching user from his right) or anticlockwise (if reaching user from his/her left) from the vertical.

$$\theta = P_{i\_max} * 80^{\circ} \tag{7.4}$$

Here, P_{i_max} is the zone recorded with the highest occupancy during observation. This allows the robot to maintain an orientation proportional to the highest probability of occupancy. As the robot positions himself on the side (right/left) of least occupancy, the robot will face towards the zone with highest occupancy from the side with least occupancy. This is similar to the human behavior when dealing with somebody highly engaged. During such occasions, the outsider will be cautious about the movements of the opponent while continuing interaction. The higher the occupancy of a region, the greater the repulsion of a robot from that region. Furthermore, the higher the occupancy of a region, the lesser the robot tends to choose the opposite region. θ is measured clockwise or anticlockwise from the horizontal drawn at 0^o head orientation as marked in Fig. 7.2. From (4), it is expected to scale the angle of deviation with respect to the occupancy. Hence the robot will deviate much from highly occupied areas to least disrupt the user. Using (2), (3) and (4), the robot identifies the approach direction and mutual distancing by considering the user behavior. Hence the robot tries to follow an etiquette-based behavior by respecting its user's personal space.

The criteria used for making major decisions regrading the approach direction, orientation and mutual distance to keep between the user and robot are given in algorithm 1 below.

When choosing the direction to approach user according to the algorithm, for a maximum occupancy in $far\ right$ allows the robot to approach user from $left$. Similarly, for maximum occupancy in $right$, mid, $left$ and $far\ left$ allows the

Algorithm 1 Decision making criteria

Data: zone corresponding to P_{i_max}, P_{i_max}, θ
Result: Interaction decisions: approach direction, orientation, interpersonal distance

```
initialization; while A user is present do
    if right side is occupied then
        Reach the user from left
    else
        if left side is occupied then
            Reach the user from right
        else
            if 'Front' region is occupied then
                Keep an interpersonal distance as in (2)
                Keep an orientation of θ as calculated from (1)
            else
                if 'Far front' region is occupied then
                    Keep an interactive distance as in (3)
                    Keep an orientation of θ as calculated from (1)
                else
                end
            end
        end
    end
end
```

Table 7.1: The set of zones and their corresponding probabilities of occupancy obtained for each task

Task	(I, P_{zone _i})									
DES	(15, 0.76)	(14,0.24)								
CLN	(11, 0.01)	(12,0.03)	(13,0.39)	(14,0.19)	(15, 0.1)	(16,0.05)	(20,0.06)	(36,0.04)	(37,0.08)	(45,0.05)
EXR	(11,0.06)	(13,0.05)	(14,0.01)	(15,0.06)	(16,0.32)	(17,0.11)	(18,0.05)	(19,0.11)	(20,23)	
WTV	(11,0.07)	(32,0.1)	(35,0.02)	(36,0.67)	(42,0.09)	(45,0.05)				
SEA	(11,0.04)	(32,0.12)	(36,0.76)	(37,0.03)	(42,0.05)					
LAB	(7,0.1)	(8,0.11)	(13,0.3)	(14,0.34)	(15,0.1)	(38,0.05)				
PHN	(17,0.12)	(18,0.18)	(19,50)	(22,0.03)	(23,17)					
LAP	(14,0.20)	(15,0.80)								
CK	(7,0.14)	(8,0.15)	(14,0.30)	(15,0.06)	(38,0.08)	(39,0.04)				
LEA	(11,0.50)	(20,0.50)								

robot to approach user from $far\ left$, $far\ right$, *zone* 51 $from\ right$ and $right$ respectively. The gap between the $front$ and $far\ front$ regions (0.20 m) must be kept if the highest number of movements were recorded in $far\ front$ region. This way, the robot gets the opportunity to avoid highly engaged areas within the user's personal space which makes an interaction adaptive and situation-cautious.

7.3 Results and Discussion

Experiments were conducted with the participation of 21 individuals of 25-50 years (mean of 26.3 and SD of 8.96).

Tasks encountered in a typical domestic and lab environments were selected for the study. The list of selected tasks is shown below. The task names are shortened for the ease of future reference.

- DES- Engaged in desk activity
- CLN- Cleaning floor
- EXR- Exercising
- WTV- Watching television
- SEA- Seated, relaxing
- LAB- Engaged in lab work
- PHN- Making a phone call
- LAP- Working on laptop
- CK- Cooking
- LEA- Leaning to a wall

Each participant was asked to perform the ten tasks mentioned above. Hence 210 different scenarios were selected during the experiment. First, MIRob was allowed to approach each user in direction A in Fig. 7.2 where orientation is $0°$ and d= 1.2 m. Then MIRob was allowed to observe each individual once the activity was started. After the evaluation of user behavior according to the model, robot navigated towards the user keeping an orientation, a mutual distance and an approach direction as determined by the proposed model. This approach behavior of the robot was rated by the participant. By comparing this approach behavior with that in direction A in Fig. 7.2. Hence approaching the user from front as in A, is used as the ground truth during this experiment. User rated the robot's behavior based on the convenience or the discomfort he/she felt as the robot approached. Here, the period of observation or t was taken as 10 s. This value was determined so that an adequate amount of data were obtained for the analysis. Visual information was extracted at a rate of 5 sets of data per sec. This data included the positions of user's body joints. The head orientation was obtained at the first instance only in order to avoid the algorithm becoming computer intensive. Proxemic decisions were taken so that highly engaged areas were least obstructed. To evaluate the performance of the system, we conducted the same experiment two more times with a gap of 7 and 14 days. The feedback scores received at each stage were analyzed. Robot was allowed to wander around in a given map and it stopped and started observation once a human is tracked.

7.3.1 Observations and discussion

Zones recorded for a single occasion and the calculated probabilities for each zone during each task are shown in Table 7.1. Only the zones with a nonzero frequency of visits are given. The zones with the highest probability of occupancy was chosen and the distance from the user was calculated according the decision making criteria. Fig. 7.3 shows the distances calculated for the 10 tasks in Table 7.1, considering the highest probability of occupancy. Except for the zones given

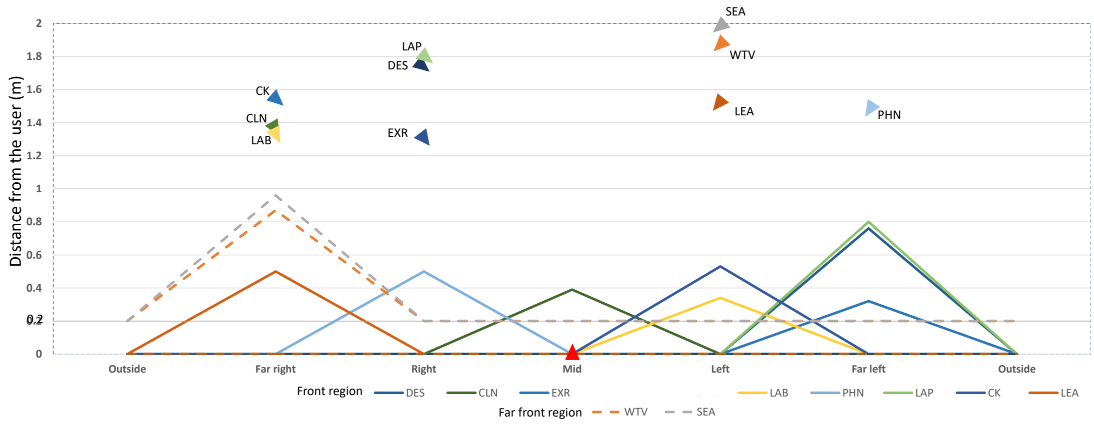


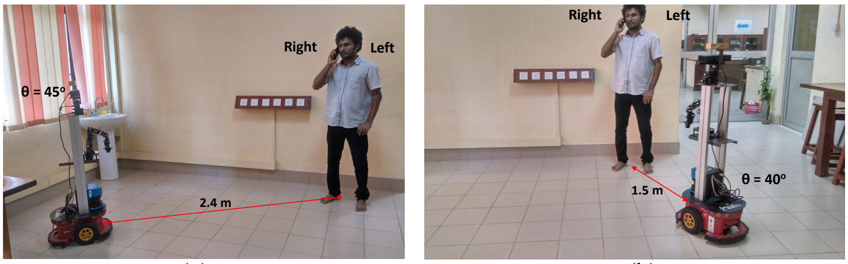
Figure 7.3: Distances marked considering maximum occupancy regions of the selected set of tasks are marked here. The red triangle denotes the position of the user and the other colored triangles denote the position and orientation of the robot during each task. The size of the triangle is not scaled with the size of the user or robot but the orientation is marked by measuring θ. Colored contours represent the restricted area for the robot to invade while approaching its user.

in Table 7.1, all the other zones received a zero occupancy probability since these zones were never occupied during the activity. After selecting P_{i_max} for each task, the approach direction, orientation and the distance between user and robot were calculated. Results obtained for each occasion are given in Table 7.2. The mean feedback score received from the users at their first encounter with the robot is given here. The position and orientation of the robot with respect to its user after implementing the model are marked in Fig. 7.3. Shown in dotted and dashed lines are restricted area including the region with the maximum occupancy. Fig. 7.4 shows the implementation of the model during PHN as given in Table 7.2.

According to Table 7.1 and Table 7.2, while the user was performing a desk activity: DES, zone 15, which is on the $Far\ left$ region, recorded the maximum probability of occupancy. Therefore after implementing Algorithm 1, the approach direction was obtained as $Right$. Similarly, the orientation was $0.76 * 80°$. Hence the figures for d and θ were 1.76 m and 61°. Robot behavior during this situation received an average feedback score of 8.17 out of 10. During all the other instances except WTV and SEA, the behavior of the robot received feedback scores above 8.

Table 7.2: Results of ten scenarios during the experiment and average feedback score received for each task

Task	P_{i_max}, **Region**	**Approach direction**	**d (m)**	**Orientation** (ˆ{∘})	**Average feedback score (out of 10)**
DES	0.76, Far left	Right	1.76	61	8.17
CLN	0.39, Mid	Far right	1.39	31	9.31
EXR	0.32, Far left	Right	1.32	26	9.33
WTV	0.67, Far right	Left	1.87	54	5.83
SEA	0.76, Far right	Left	1.96	61	6.07
LAB	0.34, Left	Far right	1.34	27	8.31
PHN	0.50, Right	Far left	1.5	40	9.50
LAP	0.80, Far left	Right	1.8	64	9.05
CK	0.53, Left	Far right	1.53	42	9.30
LEA	0.50, Far right	Left	1.5	40	8.33



(a) (b)

Figure 7.4: Two occasions encountered during PHN in Table 7.2. (a) Positions of the robot and the user initially as the robot observes user (b) Positions of the robot and user after approaching the user. The orientation and the interpersonal distances during these two occasions are marked.

The feedback scores received when the experiment was repeated after 7 and 14 days of the first experiment were given in a box and whisker plot in Fig. 7.5. We intended to analyze how user acceptance regarding the behavior of the robot evolve with experience. The mean values have shown an increase with time as users get the experience of this user-aware behavior of the robot. This could clearly be seen in the tasks: DES, WTV, SEA, LAB, and LAP. CLN, EXR, PHN and CK recorded higher feedback scores from the beginning and therefore there was only a slight difference between the scores received for these tasks afterwards.

Table 7.3 represents the results of a t-test performed to analyze the differences between the feedback scores received by the robot for its approach behavior in initial and final stages. Initially, it is assumed by the null hypothesis that there is no significant difference between the compared groups. From these results it can be seen that $p < 0.05$ in DES, WTV, SEA, LAB, LAP, and LEA. Hence it can

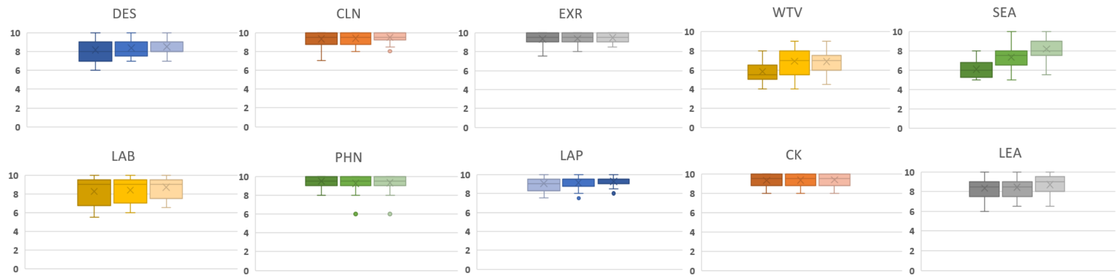


Figure 7.5: This graph shows the mean values of feedback scores received from users for the acceptance of robot's behavior in the three occasions: initial stage, after 7 days of initial stage and after 14 days of initial stage. The error bars represent the minimum and maximum scores received.

be deduced that the null hypothesis cannot be accepted. It shows that there is a significant improvement in the feedback scores in these set of tasks. In contrast, for the tasks; $p<0.05$ in CLN, EXR, PHN, and CK null hypothesis is accepted. Hence it can be seen that there was no significant improvement in the feedback

Table 7.3: T-test for the comparison of user feedback scores in initial and final attempts

Task	t-scores	Initial stage	final stage
DES	Mean	8.17	8.5
	Variance	1.55	0.85
	P	0.011	
CLN	Mean	9.31	9.404
	Variance	0.84	0.51
	P	0.081	
EXR	Mean	9.33	9.48
	Variance	0.51	0.26
	P	0.081	
WTV	Mean	5.83	6.86
	Variance	0.98	1.6
	P	0.0084	
SEA	Mean	6.07	8.17
	Variance	0.98	1.68
	P	0.0000016	
LAB	Mean	8.31	8.5
	Variance	2.26	1.8
	P	0.036	
PHN	Mean	9.5	9.31
	Variance	0.325	0.886
	P	0.179	
LAP	Mean	9.048	9.26
	Variance	0.597	0.29
	P	0.012	
CK	Mean	9.31	9.38
	Variance	0.562	0.572
	P	0.093	
LEA	Mean	8.33	8.71
	Variance	1.28	0.964
	P	0.049	

scores. But it can already be seen that the feedback scores received for CLN, EXR, PHN and CK are much higher. In all the cases considered here, dof and $t_{critical}$ was 20 and 1.724.

In general, the minimum feedback scores have been increasing with time, with an exception in PHN and CK. Maximum feedback scores either stayed constant or increased with time. From these trends it can be concluded that there was a significant improvement in determining appropriate approach behavior of a robot and the performance of the system could make a positive effect upon the users' acceptance of robots as well.

7.3.2 Conclusions

Various tasks utilize activity space differently. This fact was deployed in this research to determine an appropriate approach behavior for a robot to suit a situation. The 'approach behavior' included approach direction, orientation and mutual distance between the person and the robot. This chapter is about probabilistic evaluation of human activity space. The probability of a certain region being utilized by a person in the future, is assessed by a robot before deciding to approach a person from that side. The region with the highest probability of being occupied by a person is identified and avoided when approaching that person. This probabilistic analysis of human activity space was implemented on a social robot in order to determine appropriate proxemic behavior to approach humans. Feedbacks upon robot's behavior were taken to evaluate the functionality of the proposed mechanism. Results obtained during various situations in a social environment were used to validate the system. An etiquette for a robot based on proxemics is developed in this study with the feature called 'activity zones'. The empirical results of this study could be used to reveal the human response towards a robot with proxemic-based etiquettes.

Chapter 8

MODELING A CONTEXT: AUTOREGRESSIVE MODEL

In this work, a method to identify the interest of a human towards an interaction, based on observable nonverbal cues has been proposed. These cues were applied to an autoregressive (AR) model to obtain a quantitative measure for the displayed interest of the human towards an interaction. This measure was referred to as the 'Level of interest (LOI)'. This evaluation will be deployed by the robot itself to initiate an interaction with its user, if the situation is favorable for an interaction. The interaction was followed by decisioning upon which behaviors and etiquettes robot has to follow, in order to make the scenario more comfortable and humanlike. The system evaluates unclarifiable characteristics such as movements adopted by humans often before initiating an interaction with an outsider. This enhances the decision making abilities related to interaction initiation as the system outputs a measure of the emotional state in a human encounter.

Even so there are other factors in a human which can display their inner state to the outside world. Therefore, we present a mechanism based on autoregressive(AR) models considering a number of nonverbal human cues related to different aspects. We used multiple AR models to represent different psycho-physiological aspects of human behavior uniquely. Hence the behavior that is likely to be expected from the robot at that particular situation is determined by these AR models. This model determines robot's 'behavior' by means of three variables: orientation with the user, proximity and the nature of conversation to have with the user.

8.1 Level of Interest

The relationship between observable human cues and planned psychophysical behavior of a robot according to the observations is mapped in Fig. 8.1. 'Level of interest (LOI)' defines to which extent a person is interested in an interaction with the robot. Human interest was classified under three aspects: mutual gaze, gestures and movements. According to the features associated with each aspect, LOI changes. Therefore three different LOIs have been introduced to distinctly identify what is communicated by each aspect. Depending on the magnitude of three LOIs psycho-physiological behavior of the robot is tuned. For instance, if the magnitudes of LOIs show a less user interest towards an interaction, the robot

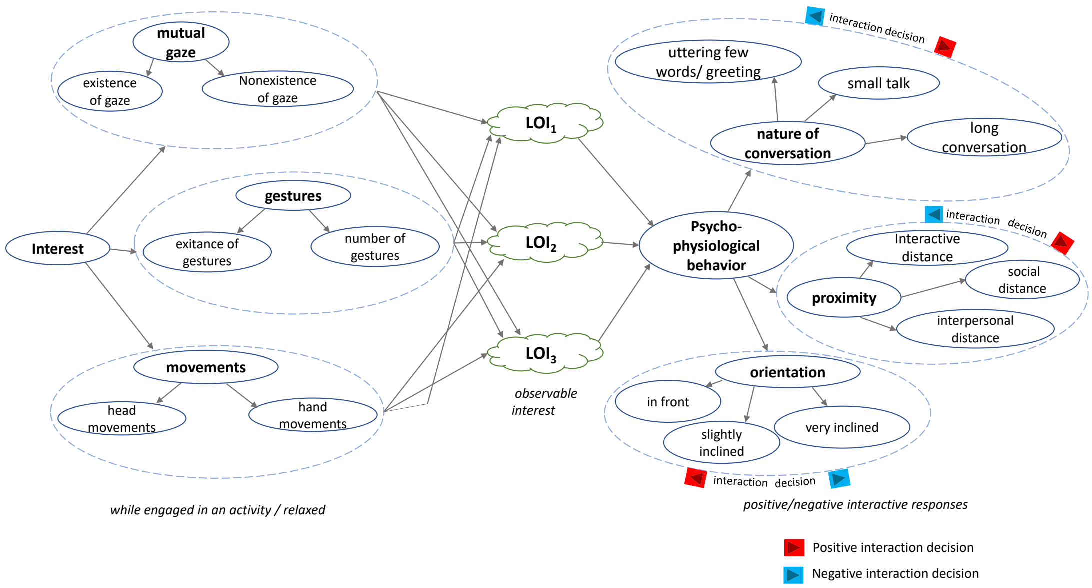


Figure 8.1: A semantic map which shows the co-relation between various aspects in human behavior considered during this study and responses generated from the robot in correspondence with its observations. The directions of increment and decrement of interactivity are marked in red and blue respectively. Interactivity increases as the situation turns favorable for interaction.

may decide to utter a few words, without going for a long conversation. Similarly, the robot may adjust positivity or negativity in interaction decision, based on the user behavior. This will facilitate an encounter with least discomfort due to the violation of expectations of both parties. Furthermore 'mutual gaze' will have a significant effect upon the proximity between two conversant while 'gestures' having the least effect. Therefore we used multiple levels of interest which can contribute to determine robot's behavior in different amounts.

In our approach, most distinctly perceivable and often used cues were chosen as the input features to assess the situation. Observable features; *mutual gaze*, *gestures* and *movements* were used as a representation of interest for an interaction. These three parameters are sub categorized as follows.

- Mutual gaze: existence of gaze, non-existence of gaze, averted gaze
- Gestures: nonexistence of gestures, fewer gestures, a number of gestures
- Movements: head movements, hand movements

The system further analyzes *gaze* for the existence, non-existence (looking away) or if averted. Similarly, *gestures* were evaluated according to the number of gestures used, and movements as speeds on head and hands. These parameters are measured and used as the inputs of an autoregressive model to determine the level of interest of an encounter.

8.2 AR model to calculate LOI

As several features can be used in combination to make different interaction decisions, LOI is sub divided into three, as LOI_1, LOI_2 and LOI_3. These are chosen based on similar features that are used as observable human cues which display interest. As usually done in AR models, we assumed that these similar variables are linearly related. Gaze (α) was taken as a parameter to determine LOI_1. Friendly gestures (θ) and walking speed (β) are considered in the calculation of LOI_2, while initial head (γ) and hand (ϕ) movements determine LOI_3. Fig. 8.2 illustrates the minimum and maximum ends of each input of the AR model. The equations (1), (2) and (3) are used to calculate LOI_1, LOI_2 and LOI_3.

$$LOI_1 = a_1.\alpha \tag{8.1}$$

$$LOI_2 = b_1.\beta + b_2.\theta + b_3 \tag{8.2}$$

$$LOI_3 = c_1.\gamma + c_2.\phi + c_3 \tag{8.3}$$

Here, a_1, a_2, a_3, b_1, b_2, b_3, c_1 and c_2 are constants, values of which were determined empirically before the experiment. Predefined example scenarios were modeled using (1)-(3), assigning 1 for all the above constants. Depending on the interactivity of the situation, these values were changed until socially acceptable interaction decisions were made by the model. Magnitudes of the three LOIs are evaluated before making decisions regarding interaction with the user. Psycho-physiological behavior required to initiate the interaction is determined in this way. Three features which determine robot's behavior are found. These are the *orientation* between robot and user, *proximity* and the

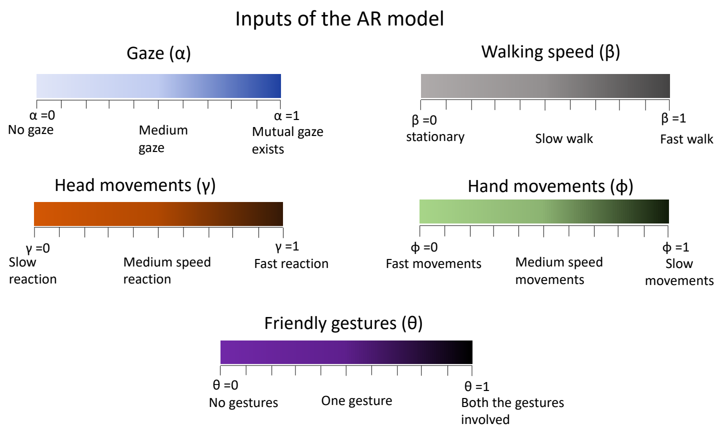


Figure 8.2: Features used as the inputs to the AR model are shown.

nature of conversation. Here, the *nature of conversation* was decided by the duration of the conversation. The distance that has to be maintained between the robot and the user was determined by the *proximity*. *Orientation* determines to which extent the robot should face the user. These parameters are determined as per their sub categories as follows.

- Nature of conversation: greeting (few words), small talk (few sentences), long conversation (large number of sentences)
- Proximity: social distance, interactive distance, interpersonal distance
- Orientation: directly towards user (in front), slightly inclined towards user, very inclined

Therefore the *nature of conversation* was decided as a greeting or a small talk or else a long conversation depending on the values obtained for LOI_1, LOI_2 and LOI_3. Similarly, *proximity* was determined as an accepted social distance or an interactive distance or else as an interpersonal distance and *orientation* as an angle directly towards the user or inclined from the user.

8.3 Decision making criteria

Based on the features used to calculate LOI_1, LOI_2 and LOI_3, interaction decisions were taken after considering some of the values obtained for LOI_1, LOI_2 and LOI_3. All three are considered when determining the *proximity*. The *nature of conversation* was determined by LOI_1 and LOI_2. LOI_1 and LOI_3 determined the *orientation* of the robot. LOI_2 is omitted here since the gestures exist only for a small duration so that the proxemic behavior of the robot will not interrupt user behavior. The marginal values required to determine these output parameters are shown in Fig. 8.3 and Fig. 8.4. These marginal values were selected by comparing the values of LOI_1, LOI_2 and LOI_3 during non-interactive

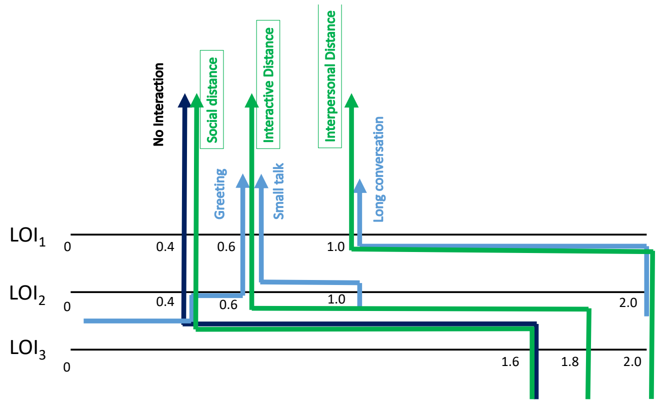


Figure 8.3: Marginal values to choose the *nature of conversation* and *proxemics* of the robot.

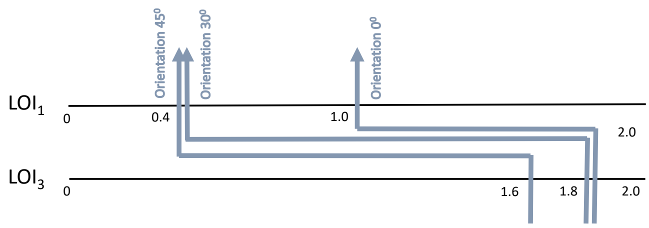


Figure 8.4: Marginal values to choose the *orientation* of the robot.

situations in which humans were alone, with interactive situations when the robot was present. In any case where LOI_1, LOI_2 and LOI_3 associated with the scenario do not fall under any of the shown classifications, the prominence is given to LOI_1. Reason for this is that gaze is a prominent feature during interaction than any other feature considered. The interaction decision was a 'no' if any 'averted gaze' or 'walking away' condition encountered. The ability to trace both stationary and dynamic human behaviors can be stated as an advantage of this method.

8.4 Results and Discussion

It must be noticed that words and sentences preferred during each conversation, were preprogrammed in the algorithm. In-built gaze recognition and open-palm property in Kinect sensor were used to identify gaze angle and 'open-palm' and 'waving hand' gestures respectively. Side-to-side movement of wrist and elbow joint positions and the open-palm property associated with the gesture were considered to recognize the 'waving hand' gesture. Gaze is extracted from the 'face yaw' property associated with the Kinect sensor. All the other movements were extracted from the skeletal representation of human body from the Kinect sensor.

Behavior of the proposed model was evaluated during a number of scenarios. For that, the model was implemented on MIRob, and the responses generated from the system during each scenario were evaluated using a feedback score given by the participant. Feedback score indicates a value out of 10, depending on the user's satisfaction upon robot's behavior. The experiment was conducted with the participation of one person at a time. He/she was allowed to perform a given task (the performer). Then MIRob was allowed to monitor the behavior of that person using the proposed model and finally, the interaction responses were generated. Hence the robot decided the *orientation* and *proximity* to reach the observed user and the *nature of conversation* which was appropriate at the occasion as deduced by the model, after the evaluation. This behavior of the robot was reviewed by the performer according to the appropriateness of behavior in the particular situation. Tasks selected for this experiment were the same as in Fig. 8.5. The task names are shortened for the ease of future reference as follows.

- DES- Engaged in desk activity
- CLN- Cleaning floor
- EXR- Exercising
- WTV- Watching television
- SEA- Sitting relaxed
- LAB- Engaged in lab work

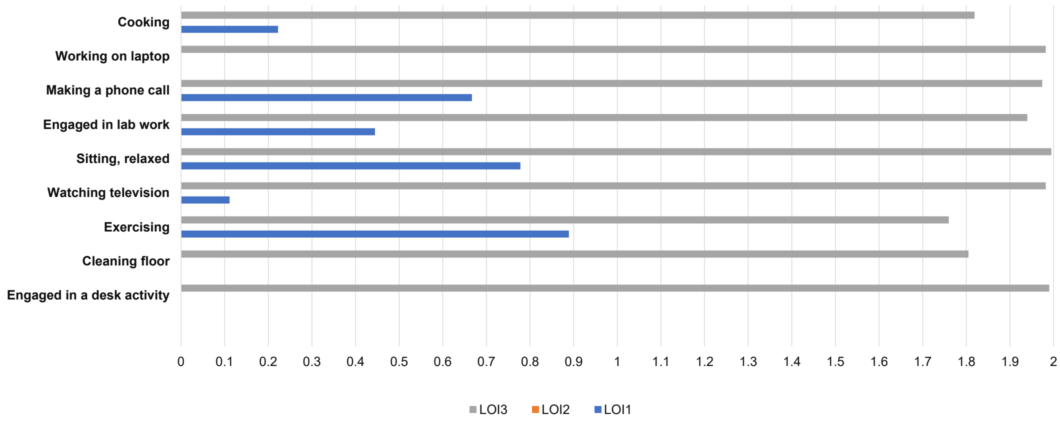


Figure 8.5: Values obtained for LOI_1, LOI_2 and LOI_3 during ordinary situations encountered in a typical domestic and lab environment are shown. LOI_2 was zero as there were no friendly gestures observed when people are engaged in work and they were not walking (gestures= none, speed of movements=0).

- PHN- Making a phone call
- LAP- Working on laptop
- CK- Cooking

Tasks encountered in a typical domestic and lab environments were selected for the study. Out of the 21 participants, each had to perform the 9 tasks in the above list. Therefore 189 different encounters were reviewed and the performer was asked to give a feedback score for the robot's behavior by comparing his/her expectations about the encounter.

During implementation, for the two stages of gaze; existing gaze and averted gaze the value of a_1 were 1 and -1 respectively. In the same way, b_1 was -1, 0 and 1 for the three occasions; walking away, stationary and walking towards the robot. b_2, c_1 and c_2 were chosen to be 1 and b_3 and c_3 were chosen to be 0 for all the occasions. These values were chosen by considering the human trends in behavior when similar conditions appear.

The movement of spine joint was used to track the walking direction of the user. The interpersonal, interactive and social distances used in this study were

1.2 m, 1.5 m and 1.8 m respectively. Angular values for the *orientation* were 0° (in front), 30° (slightly inclined) and 45° (very inclined). These values were chosen based on acceptable behaviors during human-human interaction. The period of observation, t was chosen to be 10 s and this value was chosen so that an adequate amount of data is collected for the evaluation process to begin. Vision information was extracted 5 times a second. The maximum frame rate for RGB-D processing of visual information was 30 fps. To evaluate the performance of the system, we conducted the same experiment two more times with a gap of 7 and 14 days. The feedback scores received at each stage were analyzed. An average feedback score was calculated in the end for each category of task.

8.5 Observations and Discussion

Results obtained for several tasks performed by 14 participants are given in Table 8.1. Values calculated for LOI_1, LOI_2 and LOI_3 by the model are also given. Decisions corresponding to the observations and calculations made by the model during selected scenarios are shown in Table 8.2. A feedback upon the behavior of the robot was taken from the participant during each interaction. Facts such as the convenience in the situation initiated by the robot, following social etiquettes and violations of user expectations were considered by the users to give feedback. Finally, the average feedback score received upon robot's behavior for the task category is given.

Consider the first occasion in Table 8.1 where the user was engaged in a desk activity (DES). Here, the *gaze* was not averted. Hence α was 1 for this occasion. This gives an LOI_1 of 0.78. As there were no *gestures* associated with the user behavior, LOI_2 was zero. As the user was stationary, and only very slow hand movements (of 6 cm/s) were recorded, LOI_3 was 1.96. Neither these values belonged to 'no interaction' region in Fig. 8.3 and Fig. 8.4 nor 'walking away' or 'averted gaze' was recorded in this situation. Therefore the interaction decision

Table 8.1: Calculated values based on observations

Task	Gaze		Gestures		Walking	Movements($\circ$/s)			LOI_1	LOI_2	LOI_3	Interaction
	Angle ($\circ$)	Averted/ not	Gestures involved	no. of gestures	direction	Hand	Spine	Head				Decision (y/n)
DES	10	1	none	0	0	6	0	1.5	0.78	0	1.96	y
DES*	5	-1	Op	1	0	0	0	2	0.89	0.5	1.99	n
CLN	10	1	none	0	1	16	9	8	0.44	0	1.88	y
CLN*	25	1	none	0	0	24	12	16	1.00	1	1.80	y
EXR	0	1	none	0	0	28	1	39	0.67	0	1.67	y
WTV	20	1	none	0	0	2	0	3	0.78	0	1.98	y
WTV*	15	-1	Wav	1	0	12	0	9	0.44	0.5	1.90	n
SEA	5	1	Wav, Cal	2	0	5	1	3	0.89	1	1.96	y
SEA*	10	1	none	0	0	0	0	1	0.22	0	2.00	n
LAB	25	1	none	0	0	9	3	3	0.00	0	1.94	n
PHN	15	1	none	0	0	9	1.5	6	1.00	0	1.93	y
PHN*	5	1	Op	1	-1	12.5	16	9	1.00	0.5	1.89	n
LAP	0	1	none	0	0	14	0	1	1.00	0	1.93	y
CK	45	1	none	2	0	34	5.5	0.2	1.00	1	1.83	y

* denotes the same activity performed by a different participant.

was 'yes'. By comparing LOI_1, LOI_2 and LOI_3 with the marginal values found from the study, interaction decisions were taken. These decisions corresponding to the tasks in Table 8.1, are shown in Table 8.2. Average feedback score for the first implementation of the experiment is given in Table 8.2. Accordingly, an interpersonal distance for *proxemics*, a small talk as the *conversation* and an *orientation* of 0° towards the user were obtained as the modes of interaction for the situation. This received a feedback score of 7. This was not more, as the user was less interested in relaxing while engaged in work. DES received an average feedback score of 7.89 as there were 21 different performances for each task. However, as a whole, the user was satisfied by the behavior of his robot companion. For the purpose of explanation, scenarios with feedback scores closer to the average feedback score of the task were chosen. Fig. 8.6 shows an occasion related to the HRI scenario encountered during the experiment.

According to these results, despite few exceptions, evaluations of the model received higher user satisfaction.

Fig. 8.7 illustrates how the feedback scores evolved when the same set of experiments was repeated after a period of 7 and 14 days respectively after the first implementation of the system. We took an effort to reveal the tendencies in user acceptance upon such a system with situation-awareness. In general,

Table 8.2: Results of the experiment

Task	Modes of interaction	Feedback score (out of 10)	Average feedback score
DES	smal, intp, 0°	7	7.89
DES*	-	10	
CLN	greet,intp, 0°	9	8.74
CLN*	smal,intr, 30°	8	
EXR	smal, intp, 0°	4	4.89
WTV	smal, intp, 0°	5	5.02
WTV*	-	10	
SEA	smal,intr, 0°	9	9.02
SEA*	-	8	
LAB	-	10	9.19
PHN	smal, intp, 0°	5	4.48
PHN*	-	10	
LAP	smal, intp,	9	9.30
CK	smal, intp, 0°	10	8.93

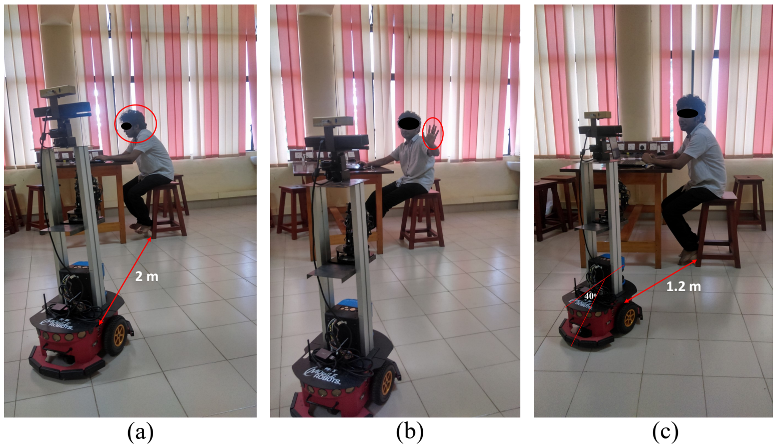


Figure 8.6: Occasions involved during SEA (a) the user was unaware of the robot initially (b) When he notices robot, he maintains mutual gaze, and waves hand towards the robot (c) Robot approaches the user after analyzing the situation. Here the *orientation* was 45°.

user feedback for the tasks: DES and SEA showed a noticeable increase in 1.18 and 0.55 points respectively. A similar trend, but less in magnitude, could be observed in the average feedback scores received for CLN, CK, LAB and LAP. The scores improved by 0.16, 0.17, 0.17 and 0.26 for CLN, CK, LAB and LAP respectively. Although the scores of these tasks improved insignificantly, the scores have already been higher, unlike in DES and SEA. The more the experience with the situation-cautious behavior of the robot, the more users were delighted. An interesting trend in user behavior was revealed for the tasks: EXR, WTV

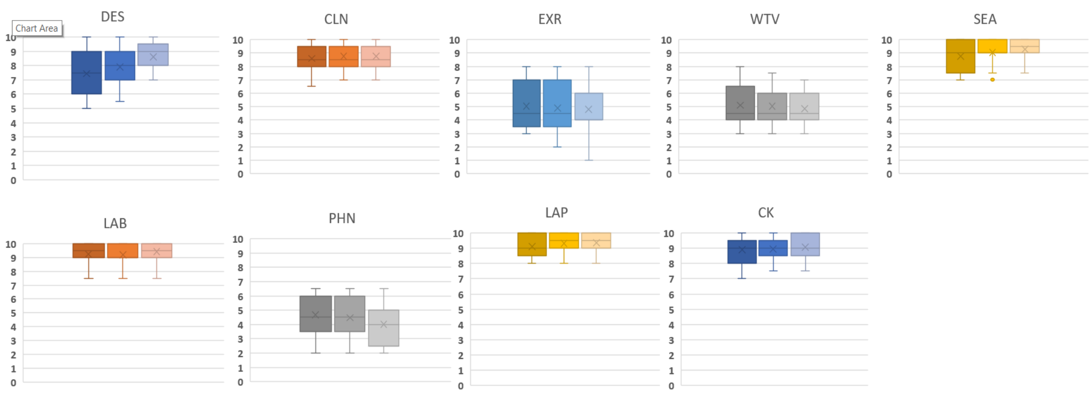


Figure 8.7: This graph shows the mean values of feedback scores received from users for the acceptance of robot's behavior in the three occasions: initial stage, after 7 days of initial stage and after 14 days of initial stage. The error bars represent the minimum and maximum scores received.

and PHN. The three tasks: EXR, WTV and PHN recorded drops of 0.26, 0.27 and 0.67 points respectively in the three occasions considered. A possible reason for this is that user's didn't expect to be distracted by the robot when they were highly engaged. This expectation increased with their knowledge of the capabilities of the robot, especially the implemented situation-cautious behavior.

8.6 Conclusions

This chapter presents an autoregressive model based on nonverbal human cues which demonstrate interest for an interaction with outsiders. Cues which are confusing to be measured in a quantitative manner were selected for the study. These features were used by a robot to determine whether to interact with the person or not and if interacting, how to behave with the human. This model was capable of determining appropriate nature of conversation and proxemics for the robot. This proved to be helpful in determining appropriate behavior for the robot when a human was encountered. As the system evaluates human behavior in all physical, social and emotional aspects, the robot could behave with situation-awareness. Hence user dissatisfaction due to inappropriate or disturbing behaviors such as disrupting the user without invitation or invading the personal space have been eliminated through this approach. This fact is validated using the results of the experiment and suggestions for the improvement of future robotic models for initiating interaction with humans are stated. Results further confirmed that autoregressive modeling of human behaviors to be producing satisfactory evaluation of a situation in proactive HRI.

Chapter 9

A SOCIAL ROBOT WITH EXPERIENCE AND LEARNING

In the previous chapters, techniques for the recognition of the following nonverbal human cues were developed and our models used to perceive an encounter were based on these recognized cues.

- Psychological aspects of human behavior
- hand pose
- body pose
- friendly gestures
- movements
- time of the day
- activity space
- activities

We used fuzzy logic-based models, supervised machine learning and autoregressive models to map observable nonverbals in humans during our first attempts to understanding human behavior prior to robot-initiated interaction. Robots can be intuitive about its decisions regarding interaction by adapting to the situation. Meantime, a situation comprises of the user and the factors which prevail in the surrounding. These factors in the surrounding also influence the internal state of mind of a human and are partially displayed to the outside world though his/her behavior.

In our model, the model parameters are as follows.

- Environment: human-robot encounter or the setting
- Agent: Robot
- States: Specific to user
- Rewards: User behavior

These are shown in the action-state model in Fig. 9.1. The decision flow of the system after implementing this model is shown in Fig. 9.2. The robot will first observe its user and the context to make an 'observation plan'. During this

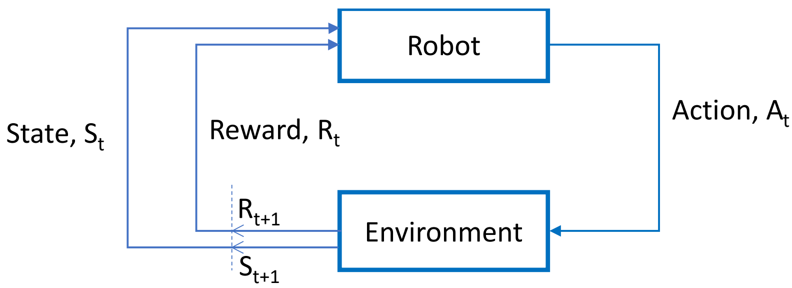


Figure 9.1: The action-state model

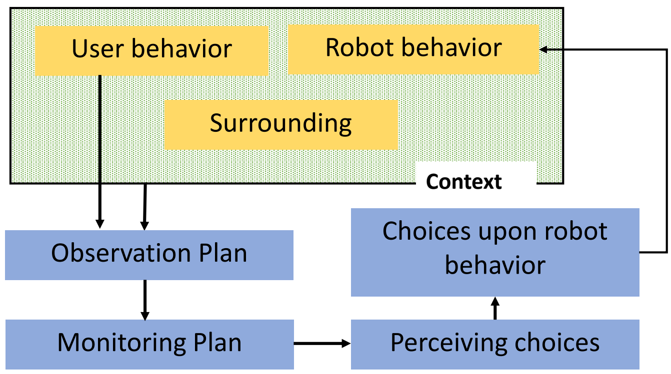


Figure 9.2: Decision flow of the robot platform

stage it will identify the observations that have to be recorded for further use. 'Monitoring plan' includes the observations which have to be monitored for long. For instance, the immediate distance between the robot and the user, before the robot approaches him/her. Then the robot analyzes its observations according to the trained model, to take decisions regarding its own behavior. These actions are monitored at the 'Perceiving choices' and 'Choices upon robot behavior' stages. Once the user's interaction demanding or the requirement is perceived, the robot initiates an interaction. Hence changes its behavior.

9.0.1 Semantic Mapping of the features

Fig. 9.3 shows how various human cues are being used in our cognitive model. Stage 1 or 'Environment Observation' falls under the Observation Plan in Fig. 9.2 and stage 2 or 'Learning' falls under 'Monitoring Plan'. The robot first observes its surrounding and some behaviors of its user under stage 1. These behaviors are whether any user exists or not, if exists whether the user walks away, and whether the user moves in a wider range of distance. The robot will not try to interact with a user who walks away or with movements that cover a wider range (probably engaged or busy). After stage 1, the robot explores stage 2, by trying to perceive the situation. For that it will look for the user's gaze behavior (gaze angle and whether the gaze is averted or not as determined in chapter 8),

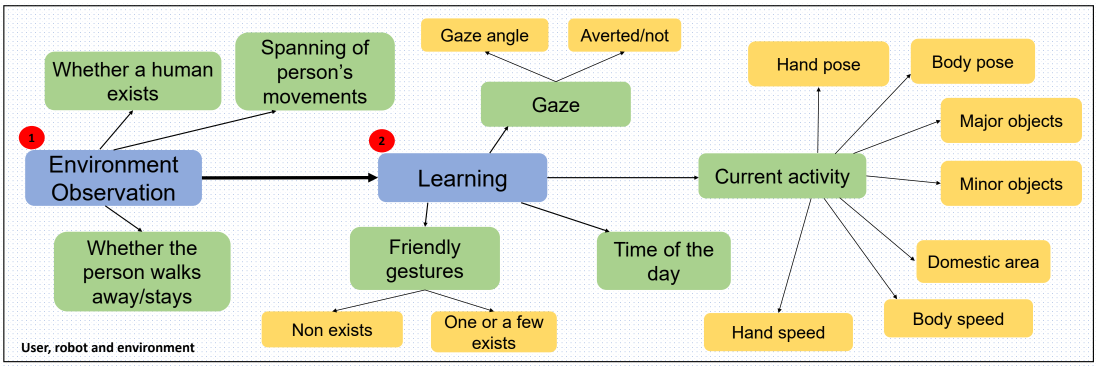


Figure 9.3: The interrelations of semantic features during the Observation Plan and the Monitoring Plan.

current activity (as determined in chapter 5.6), friendly gestures (as determined in chapter 5.5) and the time of the day. We considered 'time of the day' as an input feature to identify personality traits in humans to recognize when a person wants to relax the most. We divided a day in to 5 time categories from 1 to 5 as: Early morning (until 7 am), Morning (7am - 12pm), Afternoon (12pm - 4pm), Evening (4pm - 6pm) and Night (6pm - 12am).

These features included in stage 2 were used as the states of a particular situation.

9.0.2 Robot behavior

Once the state-action model is executed, the model outputs three actions of the robot appropriate for the situation. These are the orientation, mutual distance and the length of the conversation. The learning algorithm outputs whether to increase or decrease each of these variables. Fig. 9.4 illustrates how these variables change according to the robot's decisions.

9.0.3 Selection of participants and creating training data

We selected eight users who are different in age, and have different traits of personality. Hence we selected people with friendly/extrovert personality as well as introvert personality. In addition we selected a child, 2 youths and 4 ordinary

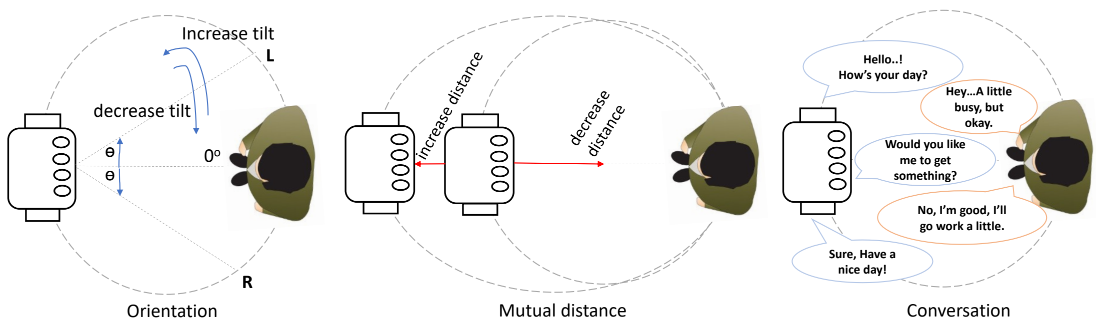


Figure 9.4: Observational parameters considered in the system for robot decisioning.

adults of age range 12 to 60 (mean-38.4, SD- 9.6). The participants were selected as: user 1- with friendly/extrovert behavior, user 2- with a less friendly/introvert behavior, user 3- an ordinary adult, user 4- an ordinary adult, user 5- an ordinary adult, user 6- youth, user 7- youth and user 8- child.

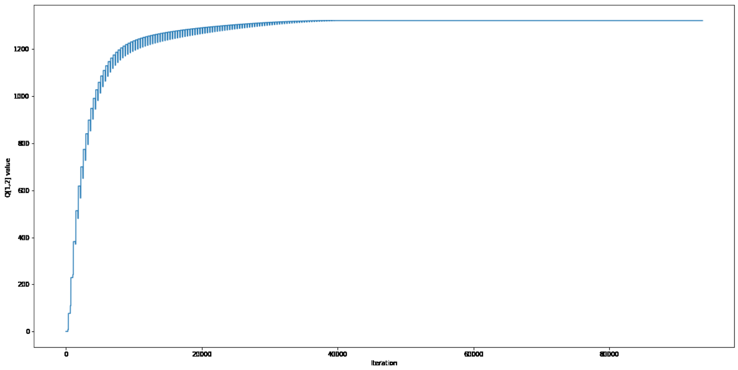
We asked selected one participant at a time and asked to simulate certain behaviors in his/her own way and recorded the cues in Fig. 9.3 under 'Learning'. Then we inquired what types of robot responses were preferred at each occasion. This is how data was collected for training. Similarly, data was collected for each participant separately.

Fig. 9.5 shows Q-value tests for Q[1,2], Q[33,0], Q[33,2] and Q[45,3] respectively. We used a learning rate which starts with 0.4 and deceases by 0.00001 in each iteration, a discount factor of 0.9 abd 10 maximum steps of 10 per episode.

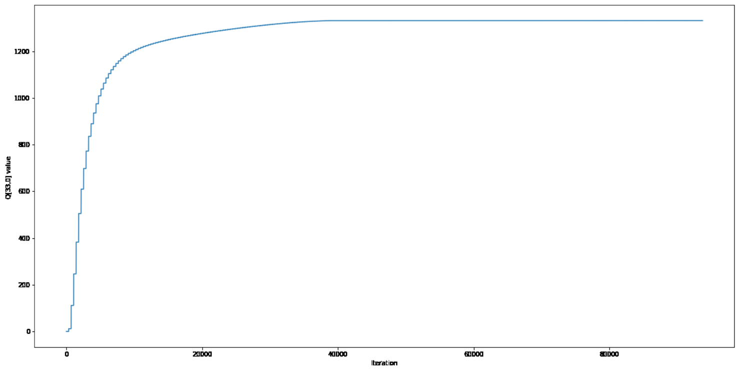
We collected data for consecutive 7 days. First, we trained the model with a limited set of data after 1 day of the data collection. In the end, we showed the robot responses for each user and asked them to rate (out of 10) the robot's responses according to his/her satisfaction upon robot's behavior. Similarly we repeated the same process after 4 days and 7 days of data collection. We plotted the user feedback score for these three occasions: after 1 day, 4 days and 7 days. This is shown in Fig. 9.6.

According to the results, the adaptive behavior improved as the robot gains more and more experience. Thus increases user satisfaction. Participants except the child, was least satisfied with less experienced robot behavior (day 1), while they were mostly satisfied with the experienced, adaptive behavior at a later stage (day 7). A probable reason for this is that children expect the least capabilities from a robot, hence have the least expectations upon not getting disturbed.

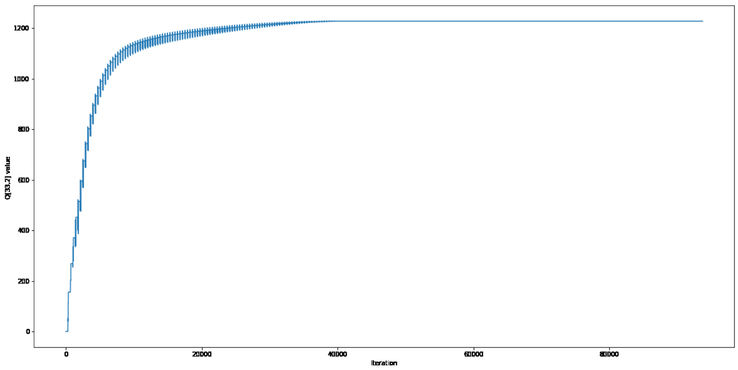
During this stage, postures were identified by the 'Posture Identifier' explained in 3. Similarly, gestures and activity were recognized by the modules:



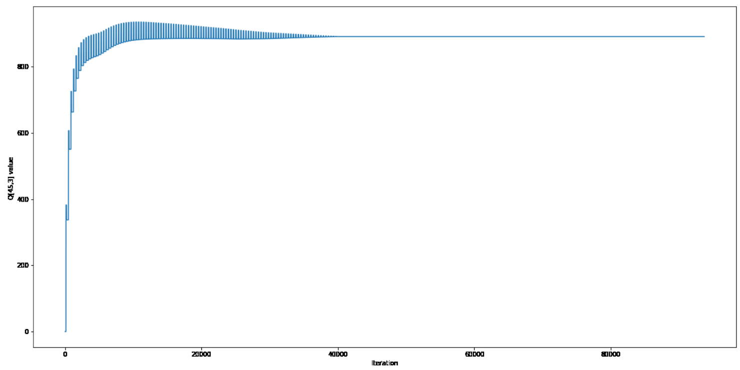
(a)



(b)



(c)



(d)

Figure 9.5: Q value tests for 4 occasions during training are shown. (a), (b), (c) and (d) show Q[1,2], Q[33,0], Q[33,2] and Q[45,3] respectively against the iteration.

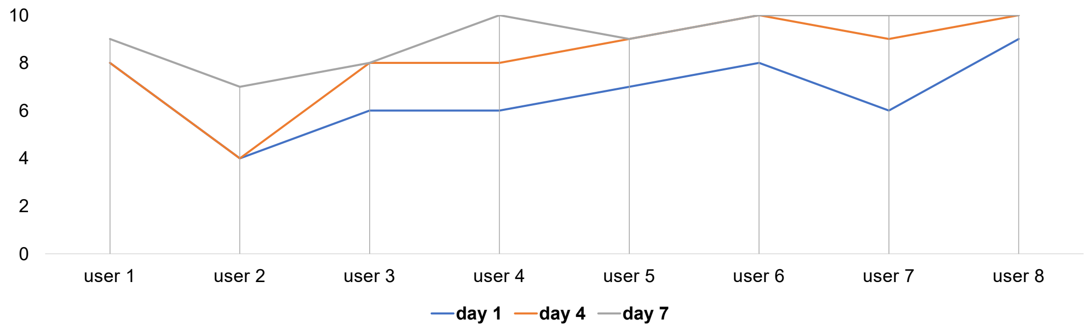


Figure 9.6: The user feedback scores over three days during the period of collecting training data are plotted against each user. The feedback scores are given out of ten.)

'Gesture Identification Module' and 'Activity Recognition Module' respectively. Movements were monitored by the 'Human Behavior Evaluator'. Objects in the surrounding and the context were recorded by the 'Context Evaluator'. Decisions regarding proxemics and conversations were taken by 'Interaction Decision Making Module' after considering the outputs of the previous modules. Memory on previous occurances was maintained by 'Learning and Experience' unit. All the information regarding interaction and the context was stored in 'Data Recording System' through out the time. 'Interaction Initiator' was responsible for converting the decisions of the IDMM into real-world actions.

9.1 Conclusions

The previous chapters put forward various methods to recognize human behaviors and built cognitive models to decode human behavior by observing different features. In this chapter, all the previous recognition techniques were incorporated to be used in a single cognitive model to perceive traits in different encounters. In addition to the features identified in these models, two new features of a situation: the person and the time of the day, which are more particular to an individual's personality were introduced to the cognitive model in this chapter. We trained a deep Q-network with these features. We collected the training data by simulating different scenarios with each user and taking a feedback from them about the nature of interaction they would like to have. We improved the model by increasing the number of human cues under observation. The Q-network proved to be reliable in training a social robot to generate user-specific responses towards an interaction. The responses of the robot included the orientation, mutual distancing, and the length of a conversation. The test results confirmed the fact that the users preferred the personalized behavior of the robot upon a predefined generalized behavior.

Chapter 10

CONCLUSIONS

There are many cognitive and psychophysical theories to explain human behaviors as well as the behavior of a robot. Even so, we still lack a model to perceive and predict appropriate behaviors for both the human and the robot during a human-robot encounter. Humans make an instant evaluation of the surrounding and its people before approaching anyone. As robots are widely deployed in social environments, a similar perception of the situation around its human user prior to an interaction is required. Social constraints during an interaction could be demolished by this assessment. Through this work, we took an effort to report functional units which come into play during such an encounter. This further identified the cues that were utilized by such intelligent agents to simulate and evaluate outcomes of its environment. As a result, we identified the requirement of a unified platform of robot's cognition during human-robot encounters and based on that, we developed multimodal methods of perceiving an encounter or to gain situation-awareness in order to generate instinctive responses towards humans. Understanding nonverbal signals shown by people associated with a situation is the most sought-after means of situation-awareness. Verbal signs are fairly immediate in conveying messages to the outside. However this is conceivable only after the initiation of a conversation by a person. This way, nonverbal means of perceiving a situation are demanding and less aggravating in a specific social situation. Without going for a direct interaction, situation-based adaptive behavior gives a robot more human-like personality based on society's accepted social manners. Consequently the ability to interpret emotional state of a scenario and evaluate that scenario for the appropriateness of an interaction makes HRI more adaptive and less disturbing for humans around. We derived several methods to recognize nonverbal human behaviors such as movements, postures, gestures, proxemics and the usage of activity space. The reason for using such methods without using the already existing systems is that these methods are capable of recognizing nonverbal cues with a reasonable accuracy and can be used for the realtime evaluation of a scenario. Realtime evaluation of a scenario is crucial in adapting to an encounter and generating appropriate responses through out the interaction scenario. These methods were based on Geometry of the human body and that Geometrical assessment was reliable than most of the existing methods to monitor human behavior. This was another reason to use such approaches during our research.

We identified that the following challenges have been faced by present proactive robotic systems.

-Robust human behavior detection and tracking systems dedicated to analyze nonverbal behavior are not in par with the advancement of the conceptual design of similar systems.

-Establishing a relationship between the intent and the observable or audible human cues has been challenging and lacks conceptual basis.

-Ironical human behaviors such as ‘smile to show disgust or sarcasm’ which do not match the shown intent with the real intent cannot be differentiated by the existing models.

-Bringing all observable or audible cues to a common platform which can analyze cues of different disciplines, still needs further development in order to make satisfactory interaction decisions.

Hence emotionally intelligent agents and their cognitive models need development optimizing these challenges on the way. Through this work, we tried to address the above challenges to develop the situation-awareness models of robots to a socially acceptable level. Through human studies we further established the requirement of observing nonverbals associated with a scenario to interpret that scenario and identified which cues we should look for. As the next stage, we developed models to combine these cues, decode them and perceive a situation through logical reasoning.

In our models, we began with only two cues: pose and movements, and expanded the number of cues to more than ten including a person's hand pose, body pose, gestures, movements, current activity, usage of activity space, and mutual gaze. Fuzzy logic-based assessments shown an improvement in the user acceptance for robots since such approaches could assess the vagueness in human nonverbals and handle a number of variables at once. As the complexities grew in an encounter, fuzzy-based assessment upon the scenario was inadequate. We used auto-regressive models in such situations to evaluate multiple factors associated with the scenario. AR models proved to be more convenient that fuzzy-based approaches in analyzing events that are different from each other and belong to various aspects: robot, user and the environment. Defining a rule base to assess these multiple aspects was confusing in our previous approaches and AR models demolished this difficulty to a great extent. We analyzed recognition accuracies and user feedback scores statistically to evaluate our methods.

In the end we extended our model into an adaptive approach so that it could adjust the robot's behaviors according to the dynamics in the factors within a context. Such factors included user responses and changes in observable cues within the context. Adapting to such changes in context, immensely enhansed the capability of the system to replicate dynamic human behaviors during an interaction. This improvement could be proven by the user feedbacks received upon the robot's proactive behaviors in this regard.

Our findings and human studies revealed that, helping each other or altruistic behavior, taking optimum decision, friendly or cooperative behavior, are the most important, yet challenging due to the lack of the skill of “situation-awareness”. Making right assumptions and predictions upon a complex scenario still needs improvement. Hence this “artificial awareness” contribute to a great deal of robot’s acceptance by humans in a majority of their environments. Human tendencies during experiments showed that humans were delighted with the context-cautious behavior of robots [24]. Therefore such skills can create a lively personality within the robot. Hence the perception of robot as a ‘machine’ will be transformed into an idea of a ‘creature’ with the use of such intelligence upon small but emotional encounters. The capability of such robots to model the inexplainable emotional relationship between human psychology and behavior is still questionable when compared with human cognitive capabilities. Limitations of existing models to simulate and integrate multiple cues for decision-making have been identified and our models to perceive a situation by means of nonverbal behavior were improved based on these findings.

The following key points could be highlighted from the research.

- Understanding nonverbal signals shown by people associated with a situation is the most sought-after means of situation-awareness.

- Verbal signs are fairly immediate in conveying messages to the outside. However this is conceivable only after the initiation of a conversation by a person. This way, nonverbal means of perceiving a situation are demanding and less aggravating in a specific social situation.

-Without going for a direct interaction, situation-based adaptive behavior gives a robot more human-like personality based on society’s accepted social manners. Consequently the ability to interpret emotional state of a scenario and evaluate that scenario for the appropriateness of an interaction makes HRI more adaptive and less disturbing for humans around.

- Modern systems evaluate cues from the two aspects of a situation: user and the environment to gain situation-awareness. Ability of a robot to perceive

nonverbal cues plays an important role in gaining awareness of that situation. Even so, most of the modern approaches use verbal cues and properties of voice to understand a situation during interaction. But this will not possible before an interaction. Hence, the evaluation of a number of nonverbal observable cues is required for situation-aware proactive interaction.
- Limitations of existing models to simulate and integrate multiple cues have been identified and possible future improvements that have to be implemented upon cognitive models have been suggested.
-In summary, capability of existing robotic systems to make proactive decisions based on the cues belonging to multiple disciplines is far below that of a human. Therefore we discussed the requirement of developing cognitive intelligent systems using a multimodal approach or a 'Unified theory of cognition' in this regard.
-As an effort towards this, we combined a number of cues which could be observed by a robot at a distance to get an idea of a situation prior to an interaction. Although it yielded successful outcomes, yet, there are other cues related to many aspects of a human. Developing a cognitive model to perceive such contrasting cues using a model has a lot of space to grow in the future.

We would suggest the following implications from our findings to improve future research in the field.

10.1 Implications for Theory

Even though our models evaluate a number of human psychophysiological cues of a human, still there are so many factors which have been under-evaluated by proactive models. Therefore this does not replicate all parts of HHI into HRI scenario. Among all, the current activity of a person plays an important role in determining his/her priorities at the moment. Therefore understanding the task of the user plays an important role in gaining situation-awareness for a social robot.

Our work was mostly based on the assumption that people prefer similar proxemic rules when interacting with robots as they do while having interactions with other humans. Several factors which influence personal space such as the gender, previous experience and familiarity with the robot were not considered within the context of these experiments. Furthermore, human cues are sometimes deceiving as well. Humans use ironical responses in certain situations and perception of such behaviors is still difficult for a robotic platform. For example, differentiating a smile which was meant for humiliation and a smile showing affection are yet difficult to be differentiated by a robot.

Human tendencies during the experiments showed that humans were delighted with the context-cautious behavior of robots. Therefore such skills can create a lively personality within the robot. Hence the perception of robot as a 'machine' will be transformed into an idea of a 'creature' with the use of such intelligence upon small but emotional encounters.

In the experiments we simulated a prototype of a social environment. We believe that there will be other factors which could be influential in human preferences upon an interaction. Hence such factors within the environment, robot and the user can also be considered in development of cognitive robots. In addition, patterns in speech, physical appearance and personality traits of the robot might influence the acceptance of the robot. Therefore evaluation of such factors is important as well. The conceptual design of a robot's cognition must consider these factors before implementing them in general applications.

The findings of our studies were based on the assumption that humans reaction for HRI is similar to HHI. However there are communities that have a different opinion of robots and such incidents may create complications in perceiving such encounters. It is important to overcome these limitations to define a rather completed physical, behavioral and emotional states of a human-robot encounter.

Activities which require object's attributes such as the material, color, nature of the surface etc. cannot be recognized with our approach. Even though such activities are rare in a domestic environment, some special environments such as laboratories, work sites and children's homes can be benefited by expanding the semantic features in activity recognition up to object attributes as well.

From the results we could see that there is an interesting relationship between objects and activities. Therefore in addition to monitoring individual behavior, human-object interaction must clearly be perceived in order to recognize an activity.

Although we did not consider the voice as a semantic feature to interpret a situation, voice makes recognizing some activities easier. For instance, when making a phone call, listening to the person's voice is more promising than recognizing the hand pose and the presence of the phone. Therefore considering 'voice' as another semantic feature can be stated as a design guideline for future proactive robots.

Lastly, there are numerous gestures and major and intermediate postures associated with hands and the body. In our approaches, we recognized only a selected number of such features. Therefore, to have an idea of a situation entirely, recognition of the remaining postures will be advantageous. This will be another implication derived from this research.

10.2 Implications for Design

Based on our findings, we can state that this evaluation offers better means of determining an appropriate approach behavior for a robot to initiate interaction with a user based on nonverbal cues displayed. As users prefer not getting disrupted by the behavior of robots, the first design guideline proposed by the study is improving the sense of respecting the personal space of an individual based on both physical and emotional aspects. It is vital to design social robots that do not violate user expectations. This will gently allow the user to establish the idea of 'accompanying a social companion' upon 'using a machine' in his work. The feature of 'maintaining appropriate approach behavior' can be used as an 'etiquette' for social robots deployed in human environments. Therefore this can be introduced as the second design guideline for robots.

The third guideline proposed by the study is to consider human cues as much as possible. It is important to get a clear representation of the situation. In addition, considering more factors from the surrounding as well as the user can be a plus for a better situation awareness. Therefore it is required to equip the robot with hardware having adequate capabilities. Reaction towards the presence of the robot may differ from user to user. Therefore perception of different psychophysiological human behavior will help make a situation-aware robot companion. This shall be considered as the forth guideline derived from the study.

We considered approach behavior and initiating a conversation as a proactive robot's responses upon an encounter. Often, movements of humans are generated as a result of their tasks. Recognition of the task is an important aspect to determine appropriate approach behavior in this regard. It is important to predict the dynamic user behavior during a task in order to determine the suitability of a conversation at a particular situation. This will be the fifth design guideline derived from the study.

10.3 Considerations for future

Most common areas in which situation-awareness is critical and will take life or death decisions can be listed as warfare, self-driving vehicles, rescue, health care, disaster management, and surveillance in atmosphere and space. In addition, the most common applications of situation-aware robots can be found in domestic and social environments, shopping malls, museums, elderly and child care. Development of task specialist robots offers only a narrow AI. In order to broaden the applications of AI, we need situation-aware robots which can theoretically come up with solutions in even unfamiliar situations. A general theory of behavior adoption is required to cope in such scenarios. This is a suggestion to convert machine autonomy into a less stressful but useful with the ability of perceptive intelligence. Human judgment upon robots and robot's judgment upon humans can be taken to a successful level of cooperation through such an approach. When dealing with uncertainties associated with input sensors a Multi-Modal Perception (MMP) model can be used. But real time processing of multiple human inputs is one of the most difficult challenges out there in this research area. Range of sensor inputs for human interaction is much larger than for most other robotic domains in current research. Computer vision methods used to process human-oriented data such as facial expression and gestures recognition will need to be developed in order to handle a wide array of possible inputs and scenarios. Outcoming such difficulties is critical in developing faster, realtime cognitive robots.

Limitations of existing models to simulate and integrate multiple cues have to identified further and existing cognitive models have to be extended. In summary, capability of existing robotic systems to make proactive decisions based on the cues belonging to multiple disciplines is still far below that of a human. Therefore we discuss the requirement of developing cognitive intelligent systems using a multimodal approach or a 'Unified theory of cognition' in this regard.

LIST OF PUBLICATIONS ORIGINATED FROM THE THESIS

- **Chapa Sirithunge**, A. G. Buddhika P. Jayasekara, and D. P. Chandima. Proactive Robots with the Perception of Nonverbal Human Behavior: A Review. *IEEE ACCESS Multidisciplinary Journal.* 7, pp.77308-77327, IEEE, 2019.
- **Chapa Sirithunge**, H. M. Ravindu T. Bandara, A. G. Buddhika P. Jayasekara, D.P. Chandima, "A Probabilistic Evaluation of Human Activity Space for Proactive Approach Behavior of a Social Robot". In: Paladyn Journal of Behavioral Robotics 12 (2020), pp. 102–114.
- **Chapa Sirithunge**, A. G. Buddhika P. Jayasekara, D. P. Chandima, "An Evaluation of Conversational Preferences during Human-Robot Interaction in Social Environments". In: Accepted for publications in Hindawi Journal of Applied Bionics and Biomechanics (2021).
- **Chapa Sirithunge**, H. M. Ravindu T. Bandara, A. G. Buddhika P. Jayasekara, D. P. Chandima and H. M. Harsha S. Abeykoon. A Study on Robot-Initiated Interaction: Towards Virtual Social Behavior. *Social Robotics: Technological, Societal and Ethical Aspects of Human-Robot Interaction* , pp. 37-60, Springer, 2019.
- **Chapa Sirithunge**, H. M. Ravindu T. Bandara, A. G. Buddhika P. Jayasekara, D. P. Chandima, "Situation Awareness for Proactive Robots in HRI". In: IEEE/RSJ International Conference on Intelligent Robots and Systems 2019 (IROS). IEEE. 2019, pp. 7813–7820.
- **Chapa Sirithunge**, K. S. Priyanayana, H. M. Ravindu T. Bandara, Nikolas Dahn, A. G. Buddhika P. Jayasekara, D. P. Chandima, "Tell me more! A Robot's Struggle to Achieve Artificial Awareness". In: The

29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). IEEE. 2020, pp. 60–66.

- **Chapa Sirithunge**, M. A. Viraj J. Muthugala, A. G. Buddhika P. Jayasekara, D. P. Chandima, “A Wizard of Oz Study of Human Interest towards Robot Initiated Human-Robot Interaction”. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). IEEE. 2018, pp. 515–521.

- **Chapa Sirithunge**, Viraj Muthugala, A. G. Buddhika P. Jayasekara, D. P. Chandima, “Interpretation of Interaction Demanding of a User based on Nonverbal Behavior in a Domestic Environment”. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE. 2017, pp. 1–8.

- **Chapa Sirithunge**, K. S. Priyanayana, H. M. Ravindu T. Bandara, A. G. Buddhika P. Jayasekara, D. P. Chandima, “A Space-Filler Mechanism to Recognize Hand Pose”. In: IEEE International Conference on Industrial and Information Systems (ICIIS 2019). IEEE. 2019.

- **Chapa Sirithunge**, A. G. B. P. Jayasekara and D. P. Chandima, ”Human attention estimation by a domestic service robot using upper body skeletal information,” 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), 2017, pp. 1-6.

- **Chapa Sirithunge**, A. G. Buddhika, P. Jayasekara and D. P. Chandima, ”Effect of activity space on detection of human activities by domestic service robots,” In: TENCON 2017 - 2017 IEEE Region 10 Conference, 2017, pp. 344-349.

- **Chapa Sirithunge**, P. H. D. Arjuna S. Srimal, A. G. Buddhika P. Jayasekara, “Identification of Friendly and Deictic Gestures in a Sequence Using Upper Body Skeletal Information”. In: TENCON 2017-2017 IEEE Region 10 Conference. IEEE. 2017, pp. 3051–3056.

- **Chapa Sirithunge**, Kalhan S. Boralessa, W. Hasitha Eranga, A. G. Buddhika P. Jayasekara, D. P. Chandima, K. T. Manjula U. Hemapala, “Role of Gaze in Human-Robot Collaborative Environments”. Manuscript submitted for review (2020).

- **Chapa Sirithunge**, Nikolas Dahn, Gayashan Porawagamage, A. G. Buddhika P. Jayasekara, D. P. Chandima, “Interpretation of Body Language for Situation-awareness in Proactive HRI”. Manuscript submitted for review (2020).
- **Chapa Sirithunge**, A. G. Buddhika P. Jayasekara, D. P. Chandima, “Context-aware Activity Recognition through Semantic Mapping”. Manuscript submitted for review (2021).

REFERENCES

[1] J. L. Jones, N. E. Mack, D. M. Nugent, and P. E. Sandin, "Autonomous floor-cleaning robot," Apr. 18 2017. US Patent 9,622,635.

[2] L. Pineda, T. Takahashi, H.-T. Jung, S. Zilberstein, and R. Grupen, "Continual planning for search and rescue robots," in *IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pp. 243–248, IEEE, 2015.

[3] N. Doering, S. Poeschl, H.-M. Gross, A. Bley, C. Martin, and H.-J. Boehme, "User-centered design and evaluation of a mobile shopping robot," *International Journal of Social Robotics*, vol. 7, no. 2, pp. 203–225, 2015.

[4] T. Tanioka, K. Osaka, R. Locsin, Y. Yasuhara, and H. Ito, "Recommended design and direction of development for humanoid nursing robots perspective from nursing researchers," *Intelligent Control and Automation*, vol. 8, no. 02, p. 96, 2017.

[5] J. Albo-Canals, A. B. Martelo, E. Relkin, D. Hannon, M. Heerink, M. Heinemann, K. Leidl, and M. U. Bers, "A pilot study of the kibo robot in children with severe asd," *International Journal of Social Robotics*, vol. 10, no. 3, pp. 371–383, 2018.

[6] S. Andrist, X. Z. Tan, M. Gleicher, and B. Mutlu, "Conversational gaze aversion for humanlike robots," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 25–32, ACM, 2014.

[7] J. Goetz, S. Kiesler, and A. Powers, "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *Proceedings of the 12th IEEE international workshop on robot and human interactive communication*, pp. 55–60, IEEE Press Piscataway, NJ, 2003.

[8] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: a survey," *International Journal of Social Robotics*, vol. 5, no. 2, pp. 291–308, 2013.

[9] J. Bongard, "Using robots to investigate the evolution of adaptive behavior," *Current opinion in behavioral sciences*, vol. 6, pp. 168–173, 2015.

[10] H. A. Karim, A. M. Lokman, and F. Redzuan, "Older adults perspective and emotional respond on robot interaction," in *4th International Conference on User Science and Engineering (i-USEr)*, pp. 95–99, IEEE, 2016.

[11] D. Stavrou, D. G. Eliades, C. G. Panayiotou, and M. M. Polycarpou, "Fault detection for service mobile robots using model-based method," *Autonomous Robots*, vol. 40, no. 2, pp. 383–394, 2016.

[12] P. Liu, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "A model for generating socially-appropriate deictic behaviors towards people," *International Journal of Social Robotics*, vol. 9, no. 1, pp. 33–49, 2017.

[13] I. Pedersen, S. Reid, and K. Aspevig, "Developing social robots for aging populations: A literature review of recent academic sources," *Sociology Compass*, vol. 12, no. 6, p. e12585, 2018.

[14] M. G. Rashed, D. Das, Y. Kobayashi, and Y. Kuno, "A study on proactive methods for initiating interaction with human by social robots," *Asian Journal for Convergence in Technology*, 2018.

[15] J. Navarro and M. Karlins, *What Every Body Is Saying*. William Morrow Paperbacks, 2008.

[16] C. A. Smarr, T. L. Mitzner, J. M. Beer, A. Prakash, T. L. Chen, C. C. Kemp, and W. A. Rogers, "Domestic robots for older adults: attitudes, preferences, and potential," *International journal of social robotics*, vol. 6, no. 2, pp. 229–247, 2014.

[17] J. D. Mayer, P. Salovey, and D. R. Caruso, "Target articles:" emotional intelligence: theory, findings, and implications"," *Psychological inquiry*, vol. 15, no. 3, pp. 197–215, 2004.

[18] B. Scassellati, "Theory of mind for a humanoid robot," *Autonomous Robots*, vol. 12, no. 1, pp. 13–24.

[19] S. K. Card, *The psychology of human-computer interaction.* CRC Press, 2017.

[20] O. C. Görür, B. S. Rosman, G. Hoffman, and S. Albayrak, "Toward integrating theory of mind into adaptive decision-making of social robots to understand human intention," 2017.

[21] C. Breazeal, "Social interactions in hri: the robot view," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 2, pp. 181–186, 2004.

[22] S. Strohkorb and B. Scassellati, "Promoting Collaboration with Social Robots," in *The Eleventh ACM/IEEE International Conference on Human Robot Interation*, pp. 639–640, IEEE Press, 2016.

[23] G. Hoffman and C. Breazeal, "Robotic partners' bodies and minds: An embodied approach to fluid human-robot collaboration," *Cognitive Robotics*, 2006.

[24] M. N. Nicolescu and M. J. Mataric, "Learning and interacting in human-robot domains," *IEEE Transactions on Systems, man, and Cybernetics-part A: Systems and Humans*, vol. 31, no. 5, pp. 419–430, 2001.

[25] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, pp. 1334–1345, Nov. 2007.

[26] M. P. Michalowski, S. Sabanovic, and R. Simmons, "A spatial model of engagement for a social robot," in *9th IEEE International Workshop on Advanced Motion Control, 2006.*, pp. 762–767, IEEE, 2006.

[27] C. H. Chen and P. S.-P. Wang, eds., *Handbook of pattern recognition and computer vision.* River Edge, NJ: World Scientific, 3rd ed ed., 2005.

[28] H. Hüttenrauch, K. S. Eklundh, A. Green, and E. A. Topp, "Investigating spatial relationships in human-robot interaction," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 5052–5059, IEEE, 2006.

[29] Göller, Michael and Steinhardt, Florian and Kerscher, Thilo and Zöllner, J Marius and Dillmann, Rüdiger, "Proactive avoidance of moving obstacles

for a service robot utilizing a behavior-based control," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5984–5989, IEEE, 2010.

[30] R. M. Downs and D. Stea, *Image and environment: Cognitive mapping and spatial behavior*. Transaction Publishers, 1974.

[31] M. W. and, "Proactive multimodal perception for feature based anchoring of complex objects," in *2007 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, pp. 1069–1074, Dec 2007.

[32] A. Sabnis, G. Arunkumar, V. Dwaracherla, and L. Vachhani, "Probabilistic approach for visual homing of a mobile robot in the presence of dynamic obstacles," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 9, pp. 5523–5533, 2016.

[33] K. Dautenhahn, "Socially intelligent robots: dimensions of human–robot interaction," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704, 2007.

[34] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[35] A. Bruce, I. Nourbakhsh, and R. Simmons, "The role of expressiveness and attention in human-robot interaction," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, vol. 4, pp. 4138–4142, IEEE, 2002.

[36] J. Scholtz, "Theory and evaluation of human robot interactions," in *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pp. 10–pp, IEEE, 2003.

[37] C. Ray, F. Mondada, and R. Siegwart, "What do people expect from robots?," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 3816–3821, IEEE, 2008.

[38] M. Heerink, B. Vanderborght, J. Broekens, and J. Albó-Canals, "New Friends: Social Robots in Therapy and Education," *International Journal of Social Robotics*, vol. 8, pp. 443–444, Aug. 2016.

[39] B. Robins, K. Dautenhahn, R. T. Boekhorst, and A. Billard, "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?," *Universal Access in the Information Society*, vol. 4, pp. 105–120, Dec. 2005.

[40] C. L. Sidner, C. Lee, L.-P. Morency, and C. Forlines, "The effect of head-nod recognition in human-robot conversation," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 290–296, ACM, 2006.

[41] C. D. Kidd and C. Breazeal, "Robots at home: Understanding long-term human-robot interaction," in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 3230–3235, IEEE, 2008.

[42] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An ethological and emotional basis for human–robot interaction," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, pp. 191–201, 2003.

[43] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: how robots might shape participant roles using gaze cues," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 61–68, ACM, 2009.

[44] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, "Where to look: a study of human-robot engagement," in *Proceedings of the 9th international conference on Intelligent user interfaces*, pp. 78–84, ACM, 2004.

[45] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to approach humans?: strategies for social robots to initiate interaction," in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 109–116, ACM, 2009.

[46] E. Aguirre, M. Garcia-Silvente, A. González, R. Paúl, and R. Munyoz, "A Fuzzy System for Detection of Interaction Demanding and Nodding Assent Based on Stereo Vision," *Journal of Physical Agents*, vol. 1, no. 1, pp. 15–26, 2007.

[47] A. Fern'ndez-Baena, A. Susin, and X. Lligadas, "Biomechanical Validation of Upper-Body and Lower-Body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments," pp. 656–661, IEEE, Sept. 2012.

[48] M. L. Walters, D. S. Syrdal, K. L. Koay, K. Dautenhahn, and R. Te Boekhorst, "Human approach distances to a mechanical-looking robot with different robot voice styles," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 707–712, IEEE, 2008.

[49] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011.

[50] K. Pitsch, H. Kuzuoka, Y. Suzuki, L. Sussenbach, P. Luff, and C. Heath, ""the first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in hri," in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pp. 985–991, IEEE, 2009.

[51] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE multimedia*, vol. 19, no. 2, pp. 4–10, 2012.

[52] E. Grigore, A. Roncone, O. Mangin, and B. Scassellati, "Preference-based assistance prediction for human–robot collaboration tasks," in *Intelligent Robots and Systems (IROS)*, 2018.

[53] E. C. Grigore, O. Mangin, A. Roncone, and B. Scassellati, "Predicting supportive behaviors for human-robot collaboration," in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 2186–2188, International Foundation for Autonomous Agents and Multiagent Systems, 2018.

[54] T. Stipancic, B. Jerbic, and P. Curkovic, "A context-aware approach in realization of socially intelligent industrial robots," *Robotics and computer-integrated manufacturing*, vol. 37, pp. 79–89, 2016.

[55] A. M. von der Pütten, N. C. Krämer, C. Becker-Asano, K. Ogawa, S. Nishio, and H. Ishiguro, "At the cafe—exploration and analysis of people's nonverbal behavior toward an android," *Geminoid Studies: Science and Technologies for Humanlike Teleoperated Androids*, pp. 375–397, 2018.

[56] S. Andrist, B. Mutlu, and A. Tapus, "Look like me: matching robot personality via gaze to increase motivation," in *33rd annual ACM conference on human factors in computing systems*, pp. 3603–3612, ACM, 2015.

[57] J. S. Hu, J. J. Wang, and D. M. Ho, "Design of sensing system and anticipative behavior for human following of mobile robots," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 4, pp. 1916–1927, 2014.

[58] C. Liang, L. Qi, and L. Guan, "Motion energy guided multi-scale heterogeneous features for 3d action recognition," in *Visual Communications and Image Processing (VCIP), 2017 IEEE*, pp. 1–4, IEEE, 2017.

[59] M. Moradi, M. Moradi, and F. Bayat, "On robot acceptance and adoption a case study," in *2018 8th Conference of AI & Robotics and 10th RoboCup Iranopen International Symposium (IRANOPEN)*, IEEE, 2018.

[60] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.

[61] B. van den Brink, Z. Yumak, *et al.*, "Social gaze model for an interactive virtual character," in *International Conference on Intelligent Virtual Agents*, pp. 451–454, Springer, 2017.

[62] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, "Towards a humanoid museum guide robot that interacts with multiple persons," in *Humanoid Robots, 2005 5th IEEE-RAS International Conference on*, pp. 418–423, IEEE, 2005.

[63] R. Mead and M. J. Matarić, "Proxemics and performance: Subjective human evaluations of autonomous sociable robot distance and social signal understanding," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 5984–5991, IEEE, 2015.

[64] R. Mead and M. J. Matari, "Autonomous human–robot proxemics: socially aware navigation based on interaction potential," *Autonomous Robots*, vol. 41, no. 5, pp. 1189–1201, 2017.

[65] M. Strait, C. Canning, and M. Scheutz, "Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance," in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 479–486, ACM, 2014.

[66] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "An affective guide robot in a shopping mall," in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pp. 173–180, IEEE, 2009.

[67] A. Birk, "Robot learning and self-sufficiency: What the energy-level can tell us about a robot's performance," in *European Workshop on Learning Robots*, pp. 109–125, Springer, 1997.

[68] J. Peltason and B. Wrede, "Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns," in *Proceedings of the 11th annual meeting of the special interest group on discourse and dialogue*, pp. 229–232, Association for Computational Linguistics, 2010.

[69] A. Bauer, K. Klasing, G. Lidoris, Q. Mühlbauer, F. Rohrmüller, S. Sosnowski, T. Xu, K. Kühnlenz, D. Wollherr, and M. Buss, "The autonomous city explorer: Towards natural human-robot interaction in urban environments," *International Journal of Social Robotics*, vol. 1, no. 2, pp. 127–140, 2009.

[70] M. L. Spekman, E. A. Konijn, and J. F. Hoorn, "Perceptions of healthcare robots as a function of emotion-based coping: The importance of coping appraisals and coping strategies," *Computers in Human Behavior*, vol. 85, pp. 308–318, 2018.

[71] P. H. Kahn, H. Ishiguro, B. Friedman, and T. Kanda, "What is a human?-toward psychological benchmarks in the field of human-robot interaction," in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 364–371, IEEE, 2006.

[72] M. L. Walters, D. S. Syrdal, K. Dautenhahn, R. Te Boekhorst, and K. L. Koay, "Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion," *Autonomous Robots*, vol. 24, no. 2, pp. 159–178, 2008.

[73] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, pp. 375–382, IEEE, 2010.

[74] E. Martinson and D. Brock, "Improving human-robot interaction through adaptation to the auditory scene," in *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, pp. 113–120, IEEE, 2007.

[75] O. A. I. Ramírez, H. Khambhaita, R. Chatila, M. Chetouani, and R. Alami, "Robots learning how and where to approach people," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 347–353, IEEE, 2016.

[76] K. Fischer, L. C. Jensen, S.-D. Suvei, and L. Bodenhagen, "Between legibility and contact: The role of gaze in robot approach," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 646–651, IEEE, 2016.

[77] M. A. Yousuf, Y. Kobayashi, Y. Kuno, A. Yamazaki, and K. Yamazaki, "How to move towards visitors: A model for museum guide robots to initiate conversation," in *RO-MAN, 2013 IEEE*, pp. 587–592, IEEE, 2013.

[78] S. Oishi, Y. Kohari, and J. Miura, "Toward a robotic attendant adaptively behaving according to human state," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 1038–1043, IEEE, 2016.

[79] Y. Gao, S. Wallkötter, M. Obaid, and G. Castellano, "Investigating deep learning approaches for human-robot proxemics," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 1093–1098, IEEE, 2018.

[80] A. Cruz-Maya and A. Tapus, "Adapting robot behavior using regulatory focus theory, user physiological state and task-performance information," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 644–651, IEEE, 2018.

[81] J. James, C. I. Watson, and B. MacDonald, "Artificial empathy in social robots: An analysis of emotions in speech," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 632–637, IEEE, 2018.

[82] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold Jr, M. Durbin, M. S. Edmonson, J. Fischer, D. Hymes, S. T. Kimball, *et al.*,

"Proxemics [and comments and replies]," *Current anthropology*, vol. 9, no. 2/3, pp. 83–108, 1968.

[83] A. Kendon, *Conducting interaction: Patterns of behavior in focused encounters*, vol. 7. CUP Archive, 1990.

[84] C. Shi, S. Satake, T. Kanda, and H. Ishiguro, "A robot that distributes flyers to pedestrians in a shopping mall," *International Journal of Social Robotics*, vol. 10, no. 4, pp. 421–437, 2018.

[85] S. Petisca, J. Dias, P. Alves-Oliveira, and A. Paiva, "Emotional sharing behavior for a social robot in a competitive setting," in *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on*, pp. 730–735, IEEE, 2016.

[86] T. J. Madden, P. S. Ellen, and I. Ajzen, "A comparison of the theory of planned behavior and the theory of reasoned action," *Personality and social psychology Bulletin*, vol. 18, no. 1, pp. 3–9, 1992.

[87] W. Schneider, S. T. Dumais, and R. M. Shiffrin, "Automatic/control processing and attention.," tech. rep., Illinois Univ Champaign Human Attention Research Lab, 1982.

[88] A. J. Gruber and R. J. McDonald, "Context, emotion, and the strategic pursuit of goals: interactions among multiple brain systems controlling motivated behavior," *Frontiers in behavioral neuroscience*, vol. 6, p. 50, 2012.

[89] T. Fu and A. Macleod, "Intellichair: An approach for activity detection and prediction via posture analysis," in *Intelligent Environments (IE), 2014 International Conference on*, pp. 211–213, IEEE, 2014.