Mobile Museum Guide Robots Able to Create Spatial Formations with Multiple Visitors

(複数鑑賞者の身体配置を創りだす移動ミュージアムガイドロボット)

2013 年 9 月

埼玉大学大学院理工学研究科（博士後期課程）

理工学専攻（主指導教員　久野義徳）

Mohammad Abu Yousuf

Dedicated to my loving family, for their endless support

Acknowledgements

For the work which is presented here I am particularly grateful to my supervisor, Professor Yoshinori Kuno, who has provided enthusiastic advice, support and guidance at every stage of my studies. I am grateful to him for his constant encouragement, and his stimulating ideas.

I am grateful to Dr. Yoshinori Kobayashi who has been always ready to help, discuss, and enlighten me with regard to all aspects of statistical analysis. I would also like to thank all my Computer Vision Lab members at Saitama University.

I owe much gratitude to all my co-authors for their hard work, thoughtful, and valuable contributions including Professor Kiichi Yamazaki and Professor Akiko Yamazaki.

Many thanks are also due to my family for their continuous support and encouragement, including my wife Siffat Ayat who supported me every ways she could, my daughter Farhin Yousuf, and last but not least to my parents.

I also offer many thanks to several students, and friends of Saitama University who made this work possible through their crucial assistance, and ideas.

This research work was partly supported by JSPS KAKENHI (grant number 21300316, 23252001), Japan Science and Technology Agency (JST), Core Research for Evolutional Science and Technology (CREST), whose support is greatly appreciated.

Abstract

The development of robot is entering a new stage where the focus is placed on interaction with people in their daily environments. The concept of service robot is rapidly emerging. The service robot will act as a peer providing mental, communicational, and physical support. Such interactive tasks are of importance for allowing robots to take part in human society. Many robots have already been applied to various fields like hospital, school, day care center, museum and so on in daily environment. In museum context, guide robot needs to interact with the visitors in a natural way. Although much research has already been conducted in the area of non-verbal communication between a guide robot and human, such as facial expression, eye-gaze, and gesture commands, whether create and control spatial formation with the multiple visitors is also a fundamental function for the museum guide robots that remain unexplored. Drawing upon psychological and sociological studies on the spatial relationship between human, it is considered that museum guide robots should have also the capability to create and control spatial formation in various situations.

The research questions that we seek to address in this area are as follows: What are the constraints to create spatial formation with the visitors? How does guide robot create spatial formation with the visitors before start its explanation of any exhibit? Can the robot attract visitor's attention to creates spatial formation properly? Can the guide robot indentify interested

bystanders and invite them into an ongoing explanation session and thus reconfigure spatial formation? Can the robot reconfigure spatial formation during explaining multiple exhibits collectively? How does guide robot initiate conversation with the multiple visitors?

This dissertation seeks to find answers to these questions by incrementally exploring the constraints to create spatial formation and developed an integrated model to configure spatial formation with the visitors in various situations. We began by observing and videotaping scenes of actual museum galleries where human guides explained exhibits to multiple visitors. Based on these analyses of the video, we developed a mobile robot system able to create and control spatial formation while guiding multiple visitors inside the gallery from one exhibit to another. We evaluate the guide robot system in a series of study that focuses on different situations where guide robot creates spatial formation with the visitors. The first study focused on designing a model to create spatial formation by analyzing the constraints of spatial formation before start the explanation by the guide robot. The effectiveness of the guide robot system was confirmed through the experiments.

In a museum context, when a guide explains any exhibit to a small numbers of visitors, many other visitors who are not participating in the current explanation may stand around the explanation area. Among them who demonstrate interest in the explanation are considered to be interested bystanders. A museum guide robot needs to identify interested bystanders and invite them into ongoing explanation session. Thus to deal with the bystanders, we extended our model and evaluate its performance through a series of experiments in the next. These experiments focused on designing three steps process of identifying and inviting interested bystanders into ongoing explanation session. Change of robot's body orientation plays an important role to reconfigure the spatial formation again. This dissertation also seeks to find the answer how robot changes the position and orientation of the visitors by rotating its own body from one exhibit to another while explaining multiple exhibits collectively. Experimental results suggest that repositioning and reorientation of the visitor's body are reasonable indications of the visitor's intention of spatial reconfiguration. Finally, we have presented a model to initiate conversation with the visitors. Museum guide robot should observe visitors to find those who may want to be guided and initiate conversation with them. We developed a model that describes the constraints and expected behaviors in the phase of initiating conversation. We conducted an evaluation experiment that demonstrates that our model significantly improves the robot's performance in initiating conversation.

This research contributes to the design of a mobile museum guide robotic system that is capable to create and control spatial formation with the visitors in different situations.

Keywords

Spatial formation, F-formation, O-space, pause and restart, human-robot interaction (HRI), museum guide robot, attention control, particle filter, head turn, face direction, verbal actions, interested bystanders, body orientation, body position, transactional segment, joint transactional segment, initiation of conversation and evaluation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Recent works in robotics have enabled us to start developing humanoid robots that interact with people and support their regular activities. Current studies have explored that humanoid robots are suitable for communicating with humans. Human-robot interaction (HRI) is an interdisciplinary research field aimed at improving the interaction between human beings and robots and to develop robots that are capable of functioning effectively in real-world domains, working and collaborating with humans in their daily activities. HRI lies in the intersection of robotics, artificial intelligence social sciences and psychology [1] (Figure 1.1). For robots to be accepted into the real world, they must be capable to behave in such a way that humans do with other humans. Bartneck & Forlizzi [2] propose the following definition of a social robot: "*A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact*". This definition does not say what would actually be humans' normal expectations with regard to socially interactive robots. Fong et al. [3] provide some indications by describing socially interactive robot characteristics: A socially interactive robot may express and/or perceive emotions, communicate with high-level dialogue, learn and/or recognize models of other agents, establish and maintain social relationships, use natural cues (gaze, gestures, etc.), exhibit distinctive personality and character, and learn or develop social competencies.

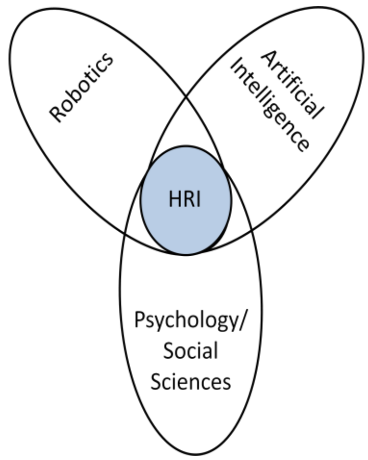


Figure 1.1: HRI is at the intersection of Robotics, Artificial Intelligence and Psychology/Social Sciences.

Service robots are envisioned to coexist with humans and to fulfill various kinds of tasks. In the last few years there has been a substantial progress in the field of service robots. A variety of mobile robots that are designed to operate in environments populated by humans has already been developed. These robots, for example, have been deployed in hospitals, office buildings, department stores, and museums. Although a number of significant challenges remained unsolved related to the social capabilities of service robots, the service robot that can create and control spatial formation with the human is also an important research issue in the realm of natural HRI.

HRI could benefit our society in multiple ways. Many robots have already been applied to various fields in daily environments. Assistive and health-care robotics can improve the quality of life of the elderly or physically impaired people, as our aging population is growing and there is a limited human health-care workforce available. For applications such as service robots, the use of robots in homes, offices, museums, schools, or stores can increase the efficiency of people's work, providing new services and improving the quality of life. In line with this prospect, we have developed a mobile museum guide robot that we believe to be a promising application. There are several studies on a guide robot for a museum, shopping mall that presents an object (e.g., an exhibit or merchandise) to visitors [4, 5, 6, 7]. In closed environment, such as a home, an elementary school or an office, robots interact with a limited group of people [8, 9, 10].

Social interaction between humans takes place in the spatial environment on a daily basis. We occupy space for ourselves and respect the dynamics of spaces that are occupied by others. We know both anecdotally [11] and from architectural theories such as Space Syntax [12] that the organization of space can generate and structure the activities of those who inhabit it. This is not to suggest that space determines behavior, but rather that there is an interaction between spatial structures and the kinds of social activities enacted within them.

1.1 Objectives

When people enter into interaction, they tend to place themselves in a spatial-orientational arrangement such that each is facing inward around the space to which each has immediate access. When this kind of particular spatial formation occurs, they can feel that they are participating in the conversation; once they perceive their participation, they will try to maintain this spatial formation. In this dissertation, our main concern is to develop a mobile museum guide robot that is capable to create and control such kind of spatial formation with the visitors. Based on the spatial relationship between people [13, 14, 15], it is considered that museum guide robots should also form an appropriate spatial relationship with the visitors in various situations [16, 17, 18, 19]. Specifically the objective can be broken down into the following categories:

Our proposed mobile museum guide robot is capable:

- To create and control spatial formation with the multiple visitors before start its explanation of any exhibit.
- To reconfigure spatial formation while moving from one exhibit to another.
- To reconfigure spatial formation when interested bystanders incorporate into an ongoing explanation.
- To reconfigure spatial formation while explaining multiple exhibits collectively to multiple visitors.
- To create spatial formation during initiation of interaction with the visitors by judging their behaviors.

1.2 Motivation

The development of robots acting as museum tour-guides is a motivating challenge, so that a considerable number of mature robotic systems have been developed during the last decade. Drawing upon psychological and sociological studies on the spatial relationship between human, it is considered that museum guide robots should have also the capability to create and control spatial formation in various situation. In a museum context, in human-human interaction, museum guide and the visitors group themselves into lines or circles and take place in the spatial environment and thus create spatial formation. Based on how humans create spatial relationships among themselves, it has been argued that such museum guide robots should also form an appropriate spatial relationship with people in various situations. The ability to naturally create and control spatial formation is very important for a mobile museum guide robot because in this way the visitors can acquire equal, direct, and exclusive access to the target object, just as with a human guide.

There has been extensive research in the field of psychology pertaining to human positioning and the attraction of attention during human interaction. The research results in sociology: spatial formation by Kendon and 'pause and restart' by Goodwin have suggested us how a guide robot should behave when starting explanation. Kendon states that, "A spatial formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access" [13]. This kind of spatial formation is also known as *F-formation*. Goodwin discussed some systematic ways in which speakers capture the attention of a recipient [20]. When a speaker begins to utter a sentence, and if s/he finds recipients are not gazing towards him or her, the speaker can use the techniques of "restart" and/or "pause" in the delivery of the utterance.

A museum guide robot should have visitors come around the exhibit that it will explain and attract their attention when it starts explaining the exhibit. If the visitors are away from the exhibit, it cannot start explanation. Even though they are around the exhibit, if they are not looking at the exhibit or the robot, they may not be ready to listen to the robot's talk. So, guide robot needs to be confirmed that visitors are around the exhibit and also needs to attract the attention of the visitors before start its explanation. There may be some other cases where the guide robot escorts visitors to several exhibits in a museum. This means that the spatial formation or F-formation arrangement, once configured in front of an exhibit, should disappear while the group is moving from one exhibit to another, and then should be reconfigured in front of the next exhibit. Moreover, according to

Kendon, an existing spatial formation may gain or lose participants, undergoing dynamic reconfiguration as it does so [13]. If the guide robot wants to change the position and body orientation of the visitors to give them better access to a target object and thus reconfigure spatial formation, it is recommended that the robot should rotates its own body towards the target object. Initiation of conversation is also an important concern for social service robots like museum guide robots. A conversation can start only when both guide robot and visitors have established a common belief that they will share the conversation. In human-human interaction, to initiate conversation in a typical situation, one would stop at a certain distance orienting toward the target person, speak a greeting word, and find that they are engaging in a conversation. People do this unconsciously in daily life. On the other hand, in human-robot interaction, it is difficult for a robot to initiate conversation in such a way that humans do frequently in daily basis. Museum guide robot needs to create spatial formation to initiate conversation.

So, museum guide robot needs to know every detail of human behavior to establish spatial formation with the visitors in different situations. Thus, the question becomes: how does a museum guide robot create and control spatial formation with the multiple visitors in different situation during showing tour performance to them? This dissertation addresses this question through developing a spatial formation model for mobile museum guide robots.

1.3 Research Contribution

This research resulted in a complete spatial formation model for a museum guide robot. Research contributions include the following:

- An experimental paradigm for studying how mobile museum guide robots create and control spatial formation with the multiple visitors by observing their position, body orientation, and face direction before start its explanation of any exhibit (Chapter 3).
- An experimental paradigm for studying how guide robots attract the attention of the visitors at the beginning of explanation (Chapter 3).
- An experimental paradigm for studying how guide robots reconfigure spatial formation with the visitors when robots along with the visitors move from one exhibit to another (Chapter 3).
- An experimental paradigm for studying how robots identify interested bystanders around itself and invite them to join into ongoing explanation session and thus reconfigure spatial formation with existing and newcomers visitors (Chapter 4).
- An experimental paradigm for studying how guide robots change the position and body orientation of the visitors by rotating its own body from one exhibit to another while explaining multiple exhibits collectively (Chapter 5).
- An experimental paradigm for studying how robots identify interested visitors about the exhibit and initiate conversation with them (Chapter 6).

1.4 Structure of Thesis

Chapter 2- Background. The next part concerns the review of relevant literature which first considers robots and human-robot interaction, then comparable and relevant findings about guide robots and spatial formation from the fields of Human-Robot Interaction (HRI) research.

Chapter 3- A Empirical framework to Create Spatial formation by guide robot Before Beginning the explanation. In chapter 3, the constraints to create spatial formation are described and then the proposed modeling of interaction to create spatial formation at the beginning of explanation is presented. Following that, robotic system overview for the museum guide robot is presented. Eventually conducted evaluation experiments and experimental results are reported.

Chapter 4- Reconfiguration of Spatial Formation While Interested Bystanders Join into Ongoing Explanation. In chapter 4, proposed robotic system for museum guide robots is extended to indentify interested bystanders and invite them to join into ongoing explanation session. Model of incorporating interested bystanders is presented in this chapter. Experiments with the guide robot and experimental results are show at the end of this chapter.

Chapter 5- Reconfigure Spatial Formation During Explanation of Multiple exhibits Collectively. The effects of rotating robot's own body from one exhibit to another and thus reconfiguration of spatial formation again are illustrated in chapter 5. The constraints of robot's body orientation are also reported in this chapter. Experimental design to evaluate the robotic system and results are described at the end of this chapter.

Chapter 6- Spatial Formation Model to Initiate Interaction. Detail human-human behaviors at the moment to initiate interaction or conversation are analyzed in this chapter to develop a model to initiate interaction. The proposed model of expected behavior of the guide robots to initiate interaction with the visitors is explained in this chapter. Detail experimental procedure and experimental results are shown at the end of this chapter.

Chapter 7- Conclusion. Conclude the thesis with a summary of the concepts and frameworks introduced in the thesis followed by the potential future work and application.

Chapter 2

Background

2.1 Robots and Human-Robot Interaction

A robot can be loosely defined as a “re-programmable multi-functional manipulator" [21] although there are several conflicting, detailed definitions. Re-programmability distinguishes robots from other automatic machines. From the inception of the term by a playwright, people have conceived robots as human-like machines which can think and move just like human beings. However, the first commercial application of robots did not look much like humans. From the 1960s, when they were first introduced by General Motors to an automobile assembly plant [21], industrial robots were developed and widely installed in manufacturing facilities all over the world. In most cases, these industrial robots were placed in a confined area of the factory floor and performed pre-defined, albeit re-programmable, material handling and manipulation tasks. Although these robots have re-programmability, they are controlled in a similar manner to other industrial equipment, such as computerized numerical control machines or unmanned guided vehicles. The rapid progress of computer and communication technologies brought intelligent robots which have advanced computing power compared with traditional industrial robots. Because intelligent robots can accomplish complex and varied tasks, the role of human control over the robot also dramatically increased. With the advent of intelligent robots, human-robot interaction (HRI) has gained importance as a research topic. This interaction has been deemed necessary to design and build effective robot systems which include human users.

2.2 HRI: Situation and Context

The term Situation denotes the physical environment where the interaction takes place, including the physical shape and size of the immediate area, the location, size and orientation of obstacles, and the location, posture, orientation and position of humans (e.g. standing, seated, behind desk, against wall etc.) within the immediate area. The purpose and intention of a Human-Robot Interaction (HRI) are separately defined as the Context of an HRI or more simply Context. In general, whenever the context of an interaction between a human and a robot is referred, this should be taken as referring to the robot context of an HRI. The robot context normally refers to the current context, which is normally the task which the robot is currently carrying out (e.g. fetching, carrying, navigating, verbal communication, physical interaction etc.). In this thesis, the term scenario is used to encompass both the context and situation of a HRI. Figure 2.1 illustrates how the robot's view of a particular context will always be limited by what is accessible to its sensors and perceptual abilities. Some aspects of individual human technological limitations, or specifically by design choices, limitations and human preferences. Therefore only a limited perception of the overall (human centered) context will be apparent to the robot at any time. Humans may either be performing a task, or any number of non-task based or other activities. Many aspects of these are not accessible or even perceived by the robot. For a domestic or service robot there is no real distinction between an activity and a task, as (currently) service robots only undertake activities in order to accomplish tasks. Therefore each robot activity can be considered to be a (sub-) task.

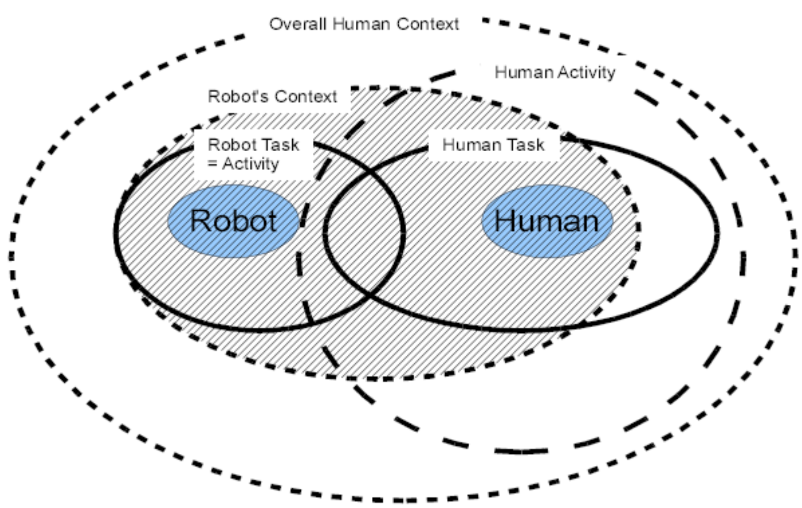


Figure 2.1: The robot's perception of human contexts, tasks and activities (shaded) will always be limited

2.3 Taxonomies of HRI

Human-robot interaction studies are closely related with human-computer interaction, as most modern robotic systems employ hardware and software components used in other common computing systems. However, the peculiarity of human-robot interaction, in comparison to other areas of human-computer interaction, is that the robot interacts with the world and is in physical contact with the human operator. Human-computer interaction primarily deals with user interface technologies, such as keyboard and mouse input devices and visual/auditory interfaces, which human operators use. However, this physical interaction is confined to input and output activities used to support the completion of computer-based tasks. In comparison with virtual world-based personal computers, robots exist in the real world. This actual physical existence creates potential problems with safety, as well as physical constraints created by human bodies. The sensory and motor abilities of humans, as well as their limitations, pose a greater challenge for human-robot interaction than in other human-computer interaction areas. These challenges become even more apparent when robots are developed for users with sensory and/or motor impairments. Various types of robots exist, making human interaction with these robots a variable problem. Thus, it would be useful to produce some form of taxonomy to compare existing and future applications of human-robot interaction within a unified frame of reference. Yanco and Drury [22] compiled various attempts of creating taxonomies of human-robot interaction, based on the studies performed to date. Among the various views of human-robot interaction, the present work employs the dimensions shown in Table 2.1 to characterize a specific application of human-robot interaction. The mobile museum guide robot to be used in this thesis research can be defined as a semi autonomous mobile museum guide service robot with the aforementioned dimensions.

2.4 Types of Robot

In HRI, three types of robot are used: Mechanoid, Humanoid and Android robot. The definitions of Mechanoid and Humanoid robots used here are based on the definitions for animated agents adopted by [23] and for Android robots from [24]:

Mechanoid - a robot which is relatively machine-like in appearance. In both live and video based HRI trials described in this thesis, a robot described as mechanoid will have no overtly human-like features.

Humanoid - a robot which does not have a realistic human-like appearance and is readily perceived as a robot by human interactants. However, it will possess some human-like features, which are usually stylised, simplified or cartoon-like versions of the human equivalents, including some or all of the following: a head, facial features, eyes, ears, eyebrows, arms, hands, legs. It may have wheels for locomotion or use legs for walking.

Table 2.1: Dimensions of robot taxonomies

Dimensions	**Components**	**Examples**
Application areas	Industrial robots Service robots	Assembly and transport robots Home/ office service robots, rehabilitation/healthcare robots, museum guide robot
Autonomy	Fully autonomous Semi autonomous Non-autonomous	Humanoid robots with ideal artificial intelligence. Most currently available robots. Tele-operated robots
Mobility	Mobile robots Stationary robots	Robots on vehicles (e.g., Mars Exploration Rover), walking robots. Fixed manipulators (e.g., MANUS arm, stationary assembly robots)

Android - a robot which exhibits appearance (and behavior) which is as close to a real human appearance as technically possible. The eventual aim is to create a robot which is perceived as fully human by humans, though the best that can be achieved currently is for a few seconds under carefully staged circumstances. The main scientific purpose of current android robots is usually to investigate interaction and social aspects of human cognition, rather than for use as domestic or servant robots [25].

In this dissertation, a humanoid robot Robovie 3 is used as a mobile museum guide robot.

2.5 Related Robotic Works

The overall goal of this research is to robotic system for a mobile museum guide that is capable to create and control spatial formation properly with the multiple visitors in different situation. As such, this thesis draws on work from many fields, including different guide robots, human-robot interaction, and spatial formation. In this chapter the most relevant research is presented in the field of HRI.

2.5.1 Guide Robot Overview

The idea of enhancing the museum experience by the use of robots has been pursued by several research groups. There is growing interest among robotics researchers in facilitating the experience of multiparty visitors in museum. There are mainly three fields of study concerning museum guide robots. The first one consists of studies focusing on the robot's autonomous movement. The second and third one consists of studies focusing on verbal and non-verbal communication with the guide robot. The literature survey below encompasses a review of related work from all three fields, with a focus on the social contexts.

2.5.1.1 Autonomous Movement of Robots

There have been several museum guide robot projects based on robot's autonomous movement [26, 27, 28]. Faber et al. gave emphasis on mechanical and electrical details to develop their mobile humanoid tour guide robot Robotinho [29]. Wheeled robots have been deployed as museum tour guides or on large fairs [30, 31, 32, 33]. They mainly focused on the autonomy of the (non-humanoid) robots and did not emphasize the interaction part so much; we want to build a robot that behaves human-like during the interaction. Recently, Shiomi et al. studied if tour guiding robots should drive forward or backward, facing the visitors, during navigation to keep them interested in the tour [34]. ROBITA turns its head towards the person to whom it addresses [35]. This robot also moves its head towards the person when the person starts talking to the robot. Yet, while very informative, such development is not been based detailed observations of human-human interaction.

The tracking of persons using laser-range sensors and cameras has been investigated, e.g. by Cui et al. [36], Schulz [37], and Spinello et al. [38]. Some teleoperated robots are being developed as a prototype of autonomous robots, known as the WOZ (Wizard of OZ) approach [39, 40]. Topp *et al.* also address the dynamic, joint movement of a robot and its user [41]. Yoda and Shiota take the need for safety in passing a human in a hallway as motivation to develop control strategies for the robot to adhere [42]. Three types of encounters were anticipated as test cases for their control algorithm, i.e. a standing, a walking, and a running person. Though all these researches concentrate on the autonomous capability of the robot, very few analyzed the interaction part between human and robot.

2.5.1.2 Verbal Communication with the Visitors

Recently, museum guides have been using questions that promote conversation and interaction, and likely draw visitors' attention and interaction. Focusing on this trend, we have found that guides engage in various strategies being posing questions visitors. More specifically, when asking a question to multiple visitors, the guide distributes his or her gaze before posing a question. There is an interest on the relationship between gaze and questions in multiparty interaction within the field of conversation analysis. For instance, Sacks et al. pointed out that one technique for selecting the next speaker is to pose a question to the intended next speaker [43]. This may include gazing towards him or her. Lerner points out that there is some limitation in assigning gaze to selecting a next speaker. He observes that the function of gaze as selecting a next speaker works when the recipient is aware of the gaze [44].

McNeill described two types of cues used for sharing attention: implicit and explicit cue [45]. Explicit cues are those also known as "deictic" references. Examples of these references are gaze and pointing behavior accompanied by verbal references. Robots have also used deictic gestures to draw others' attention to information in the environment [46]. Spexard et al. presented integration of a localization and mapping system with a spoken dialog system for a joint attention during a tutoring situation. In this research, their robot recognizes "where" and "what" in a room based on partner's explicit dialog input and robot's current position [47].

Lang et al. apply an attention system in which only the person that is currently speaking is the person of interest [48]. Okuno et al. also follow the strategy to focus the attention on the person who is speaking [49]. They apply two different modes. In the first mode, the robot always turns to a new speaker, and in the second mode, the robot keeps its attention exclusively on one conversational partner. The system developed by Matsusaka et al. is able to determine the one who is being addressed to in the conversation [50]. The model developed by Thorisson focuses on turn taking in one-to-one conversations [51]. Nakadai *et al.* developed a robot that tracks a speaking person [52]. While all these researches are impressive, so far very few studies have revealed how a guide robot should create spatial formation with multiple visitors in different situations.

2.5.1.3 Non-verbal Communication with the Visitors

It is assumed that social robots may engage in “natural” interaction with humans, i.e., interaction in the same way as humans do with other humans. The use of human-like body properties for robots has been studied for providing greater naturalness in the interaction. Often, studies have focused on the interaction after robots meet people. Matthies et al. [53] developed a robotic system for museum that incorporates body language, gesture, and facial expression. Shiomi et al. studied the group attention control (GAC) system that enables a communication robot to simultaneously interact with many people [54].

Much research has already been conducted in the area of non-verbal communication between a robot and a human, such as facial expression, eye-gaze, and gesture commands [55, 56, 57, 58, 59]. However, only little research has been done in the area of developing a robotic system that is able to interact with multiple persons appropriately. Bischoff and Graefe [60] presented a robotic system with a humanoid torso that is able to interact with people using its arms. Kuzuoka et al. and Scassellati et al. studied on the use of pointing gestures [61, 62]. Three kinds of nonverbal communicational, body movement behaviors to be observed in face-to-face encounters were differentiated by Scheflen and Scheflen [63]. As for the gaze and head orientations, although they are considered to relate to subordinate involvement, it is known that they play significant roles in communication [64]. Therefore, various studies have investigated the effect of gaze and head orientation on social interaction between a human and a robot [65, 66]. Breazeal showed that the robot's eye movement is the most relevant cue in determining if a person could successfully direct its attention to a certain object [67]. Mutlu et al. showed that head orientation (gaze) is an important cue for a subject to determine his/her role among addressee, bystander, and overhearer [68]. Sidner et al. [6, 69, 70] developed a penguin robot and examined how users responded to the robot under two conditions, both within the context of the robot explaining an exhibit: 1) the robot continuously gazes towards the visitor/user, 2) the robot moves its head and arms occasionally during the explanation. Under the second condition, user attention increased, as users responded to the robot's head movement and gaze direction by changing their own gaze and head direction. Breazeal focused on emotion, and the results suggest the importance of nonverbal interaction between humans and robots [71]. Several (non humanoid) museum tour-guide robots that make use of facial expressions to show emotions have already been developed. Schulte et al. used four basic moods for a museum tour-guide robot to show the robot's emotional state during traveling [72]. Mayor et al. used a face with two eyes, eyelids and eyebrows (but no mouth) to express the robot's mood using seven basic expressions [73]. Bennewitz et al. have proposed a robot that chooses a person among multiple visitors by considering their positions and frequency of speech obtained through facial image processing, and then turns its gaze toward him or her [4].

Previous research in human-robot interaction has looked at aspects of joint-attention [74]. Joint-attention between people is established by directing gaze toward an object, which is understood by others who respond to it by shifting their attention toward the same object. These implicit cues have been the focus of research in joint attention in robotics (e.g. [75, 76, 77]). Breazeal et al. suggested that implicit cues are particularly important in communicating internal states and can significantly improve human-robot interaction [65].

In another study, the research led by Wang [78] considered user evaluation of a humanoid robot head movement. Their experiment consisted of four variations of head movement: motionless head, smooth tracking head, tracking head without smooth movements, and avoidance behavior. The users evaluated the latter two as "enjoyable." In another study, Also, Shiomi et al. [79] have conducted a longitudinal study on human-robot interaction at Science Museum. These studies, however, did not report on the spatial relationship between guide robot and the visitors. Kuno et al. showed that a robot can move its head communicatively during explanation [80, 81]. Yamazaki et al. examined how human participants non-verbally respond to a robot when the robot's head turns and gaze are coordinated to its talk at transition-relevant instances (TRPs) [82]. Moreover, Yamazaki et al. discussed the effectiveness of the guide robot-initiated questions. When robot posed questions toward the visitor during its talk, it was able to draw the visitors' attention and elicit a positive response [83]. In all their research, however, they considered only situations that occur after the robot starts its explanation. They assumed that the visitors were already in the proper position to enjoy the explanation. However, guide robots need to bring about such situations in order to really work as guides effectively. This research, therefore, is concerned with the issue of creating spatial formation properly by the guide robot in different situation before start its explanation.

Researchers also studied the phenomenon of *engagement*. Engagement is a situation where people listen carefully to an interlocutor's conversation. A model has been developed for robots' gaze behavior [6] as well as people's gaze behavior for recognizing the engagement state [84, 85]. The major difference between the initiation of interaction and engagement is that the latter addresses a phenomenon occurring after people and robots have established a common belief that they share the conversation. In contrast, the phenomenon of initiation of interaction, which this research addresses, concerns the situation before or just at the moment that they establish the common belief that they are sharing a conversation. In this dissertation, we proposed a model that precisely describes the constraints and expected behaviors for the phase of initiation of interaction.

2.5.2 Position-Based Interaction Overview

In order to support visitor's multiparty experience, museum guide robots should be able to create spatial formation and attract visitor's attention at the beginning of explanation as an expert human guide does. We have limited conceptual tools for thinking about how the physical aspects of a setting influence interactions between people. One promising framework is Kendon's F-formation system of spatial organization [13]. Kendon suggested that when people shift attention, they move their gaze. If they continue to attend to their new focus of attention, they re-orient their body orientation and form a space where people's attention focus together. As Kendon's study suggested, position and body orientation are used for sharing continuous and prolonged attention, while gaze and pointing are used to communicate instant attention.

Detailed analyses are beginning to emerge of how people interact around particular shareable technologies in public spaces, such as interactive walls [86 ,87] and multi touch tabletops [88, 89]. F-formations are already known to researchers in human-robot interaction (HRI), having been discussed in relation to distributed technologies such as virtual environments [90] and video conferencing [91]. However, for the purposes of this research, studies that have used spatial formations in the analysis of co-located interactions by people around technology are more pertinent. In their analysis of how children build physical programs with the AlgoBlock system, Suzuki and Kato [92] described how periods of collaborative working were negotiated by children through standing up and bringing their transactional segments into alignment, whereas when they wished to watch the consequences of running a program, they would turn away from each other and sit down facing the screen. Hornecker [93] has been influenced by Kendon and by Suzuki and Kato in developing her concept of embodied constraints, which suggests that people can be encouraged to collaborate or not through material, hardware and software constraints and affordances. However, the details of what these constraints and affordances might be in particular situations remains to be worked out. Morrison and colleagues [94] have carried out perhaps the most empirically-grounded study of the impact of a technological intervention on the structure of F-formations in their comparison of hospital ward rounds carried out before and after the introduction of electronic patient records.

There has not been significant work concerning controlling the position of people by robots during interaction with them. E.T. Hall [14, 15] suggests that the area around an individual can be divided into zones depending on the nature of their current interaction. Argyle and Dean found that eye-contact is a "*component of intimacy*" and interrelates with physical proximity between two people, e.g., "*reducing eye-contact makes greater proximity possible and the greater proximity reduces eye-contact*" [95]. These studies have influenced various HRI studies. Brooks and Arkin controlled the distance between a humanoid robot and people depending on the emotional factors of the robot [96]. Walters investigated the effect of the appearance, behavior, task content, and situation of a robot on the distance between people and the robot [97].

A number of studies looked at trajectories of robot movement that keep a comfortable distance from people. Scheflen [98] proposes micro-territories called "spots," "cubits,""k-spaces," "locations," and "modules" to characterize units of space that generally determine common distances in face-to-face interaction, dimensions of furniture, seating configurations, and room layouts. The distances involved here are highly dependent on the dimensions of the human body, suggesting that the size of a socially interactive robot should be taken into account when anticipating the spatial factors that will affect its interactions. The robot Kismet [99] implicitly used an interactor's spatial configuration in a set of reactive behaviors. These behaviors included seeking, avoiding, calling, and greeting people based on the distance, speed, and sound of interactors. Sisbot et al. developed a path-planning algorithm that takes into account people's positions and orientations in order not to disturb them [100]. Walters et al. studied about the distance that it keeps while talking to people [101] affected their levels of comfort. Tasaki et al. developed a robot that chooses a target person based on people's distance [102]. Gockley et al. found that following the direction of a person (instead of the person's trajectory) is a good way for a robot to follow a person [103] and create feelings of "being together". Kahn et al. suggested the importance of aligning one's physical movements with others, which people often exhibit when walking together [104].

Birdwhistell [105] believed that behavior of posture or bodily movements in relation to social and communicational processes can be understood and interpreted as an external visible and observable code which maintains and regulates relationships between humans. Goffman [106] proposed that elements of interactions can be studied to gain an in situ natural understanding of events that happen in encounters when people continuously exchange signals of behavior. This would aid the understanding of how "people routinely achieve order in their interactions with one another".

Michalowski et al. revealed the relation between the robot's environment and the state of the person's participation in the interaction [107]. Hüttenrauch et al. found that people follow the F-formation in their interactions with robots [16]. Although their observations revealed that humans establish an O-space towards a robot, they did not study how a robot should establish its spatial relationships. Kuzuoka et al. studied the effect of body orientation and gaze in controlling the F-formation [108]. However, they only considered the change in orientation of a robot's lower body part to reconfigure an F-formation, and their robot can interact with only one person at a time. There are other key factors left unexamined that can nevertheless play important roles in the establishment of an appropriate spatial formation. Yamaoka et al. focused on positions and body orientations when implementing their information presenting robot, but their robot interacted with only one person [17]. In contrast, our proposed guide robot is capable to interact with multiple visitors. Nakauchi and Simmons [18] present another approach by first collecting empirically data on how people stand in line. They use these data then to model a set of behaviors for a robot that needs to get into, wait and advance in a queue for being serviced along with other people there. Butler and Agah [109] varied a robot's movement behaviors and evaluated in a user study how robot speed and robot distance were perceived by users. Using Hall's interpersonal distances as defining the interaction, Pacchierotti *et al* [110] devised an algorithm that allows robots to pass people in hallways. Koay et al. investigated participant preferences for a robot's approach distance with respect to its approach direction and appearance. Their results show that participant preferences change over time as they habituate to the robot [111]. However, no interactive task between the robot and the user was administered.

2.6 Social Spaces

E.T. Hall [14, 15] developed a conceptual framework known as "proxemics" that is concerned with human perception and use of space. Hall observed that human social spatial distances varies by the degree of familiarity between interacting humans and the number of interactors. Where interactions took place between individual humans, who were familiar with each other and in a private situation, the distances taken relative to each other by the interactors tended to be smaller, and where the encounter was in a more public situation, with larger groups or less familiar individuals, the distances taken between interactors tended to be larger. He proposed a basic classification of distances between individuals (Table 2.2). Figure 2.2 shows four spaces during human-human interaction.

Burgoon & Jones [112] in their review of proxemic research, categories the factors which affect human proximity as external or Environmental factors of an encounter (location, crowded, boundaries, territory, etc.), internal or interactant factors (status, age, friendship, gender, etc.) and the Nature of the interaction. This last category includes the purpose, intention and state (e.g. ignoring, talking, passing, angry, happy etc.) of those interacting, and also includes manipulations of the proxemic distance to accomplish subtle aims (dominate, reward, punish, ingratiate, etc.). Gillespie & Leffler [113] have reviewed many of the published studies into human-human proxemics and have concluded that most of the observed variation in social distances between communicating humans can be accounted for by the relative status of the interactants. In general, the higher the relative status of one of the people in an interaction, the more distance other relatively low status individuals will keep from them. One the other hand, relatively high status individuals will not respect the social spaces of other lower status individuals to the same degree.

Table 2.2: Human-human space zones

Space	Range	Situation
Intimate space	0-50cm	unmistakable involvement with another body (lover or close friend)
Personal space	50cm-120cm	comfortable separation, interaction with friends
Social space	120cm-350cm	reduced involvement, interaction with non-friends
Public space	>350cm	outside circle of meaningful involvement, public speaking

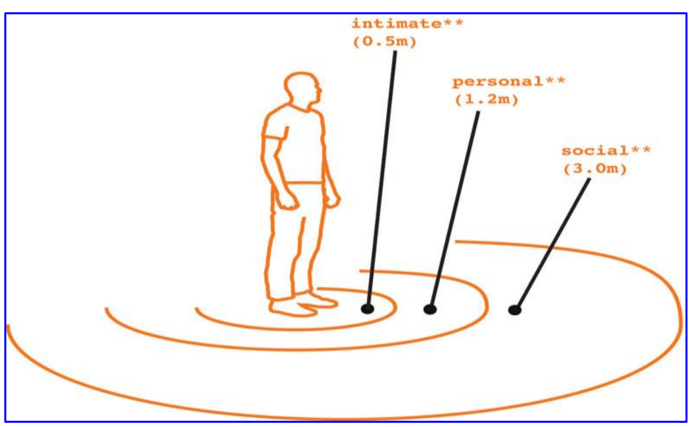


Figure 2.2: Proxemic space classification

Increasing distance naturally results in degraded thermal, olfactory, visual, and aural sensations between interactors. Voice volume increases with distance between individuals, while intimacy of conversational content decreases to a public nature. It might be postulated that the most co-present HRI exchanges and reciprocal adaptations between a human and a robot will happen in the *social* and the *personal spaces*. The *public space* is of interest as this seems like an appropriate distance to perhaps try to signal that an exchange can or is about to happen. The *social* and the *personal spaces* seem appropriate in theory to facilitate both the communication and the exchange of goods (for example the manipulation with a robotic arm). The *intimate space* seems to be better suited for exchanges with, e.g. mental commit robots like the seal-robot Paro [114], where touch is an intended interaction modality, resulting in the system giving off heat that can be felt. Specific distances between interactors actually vary by culture, gender, status, age, familiarity, relationship, pose, etc. [115]. Proxemics is an important aspect of human-robot social interaction as well, primarily due to the physical embodiment of robots, and also because there was evidence from virtual environments and existing HRI research indicating that humans respected personal space with regard to robots. Walters et al. [116, 40] have found that these spaces are generally applicable to human-robot interaction. Among them, the setting of personal space (*PS*) and social space (*SP*) could be the main issue of creating spatial formation.

2.7 What is Spatial Formation?

People often group themselves into clusters, lines, or circles, or into various other kinds of patterns. These patterns may be highly fluid or they may be relatively sustained. When such a pattern is sustained it will be referred as *formation*. *Spatial formation* is one kind of formation. According to kendon

[13], “A *spatial formation* arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access” (Figure 2.3). Such a pattern can be seen in the circle of the free-standing conversational group. Here the participants stand so that they all face inwards to a small space which they cooperate together to sustain and which is not easily accessible to others who may be in the vicinity. This kind of spatial formation is also known as *F-formation* (or facing formation [117]).

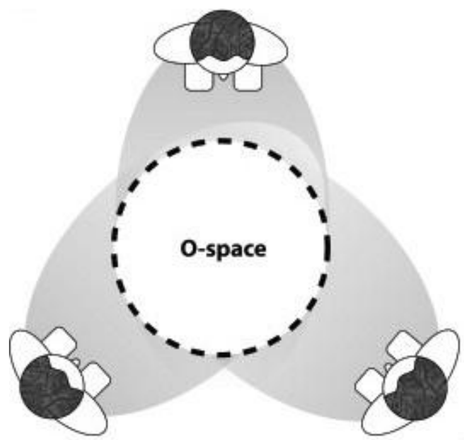


Figure 2.3: Spatial formation

People generally organize themselves in F-formations so they can have a platform for their collaborative activity. In doing so there are some practical geometrical constraints. Goffman said, “When people enter into a interaction, they tend to place themselves in a spatial-orientational arrangement such that each is facing inward around a space to which each has immediate access” [118]. By establishing such a little “knot” or huddle, the participants have at their exclusive disposal a domain or arena where their communicative transactions can be conducted. Such spatial-orientational arrangements arise because they create the conditions within which participants can effectively exchange the glances, gestures, and words out of which conversations are constructed.

Activity is always located. A person doing something always does it somewhere and his/her doing always entails a relationship to the space which has in it the objects or people with which the doing is concerned. We may imagine, thus, a space extending in front a person which is the space s/he is currently using in whatever his/her current activity may be. This space will be referred to as the individual's transactional segment (Figure 2.4). It is the space into which he looks and speaks. The size of this space can vary depending upon the activity in which they are engaged. The location and orientation of the transactional segment is limited by how the individual places his body, how he orients it and his limbs. When two or more persons come to do something together, they are liable to arrange themselves in such a way that their individual transactional segments overlap to create joint transactional segment (Figure 2.4 (b)). In Figure 2.4(b) two persons create a joint transactional segment (strip region) by overlapping their own transactional segment. This joint transactional space, which is the space between the interactants over which they agree to maintain joint jurisdiction and control, will be called an *o-space* (Figure 2.4(b)). Whenever such an o-space is created we have a spatial formation. Kendon claims that the orientation of the lower portion of the body is dominant in forming an o-space, compared to the effect of the upper body segments such as the head or upper part of the body below the neck (hereafter, called the upper body). On the basis of this claim, Schegloff proposed the concept of "body torque" which means "*different or diverging orientations of the body segments above and below two major points of articulation--the waist and the neck*" [119]. Schegloff claimed that the orientation of the lower part of the body relates to "dominant involvement" of the person, and the orientation of the shoulders and face relates to "subordinate involvement".

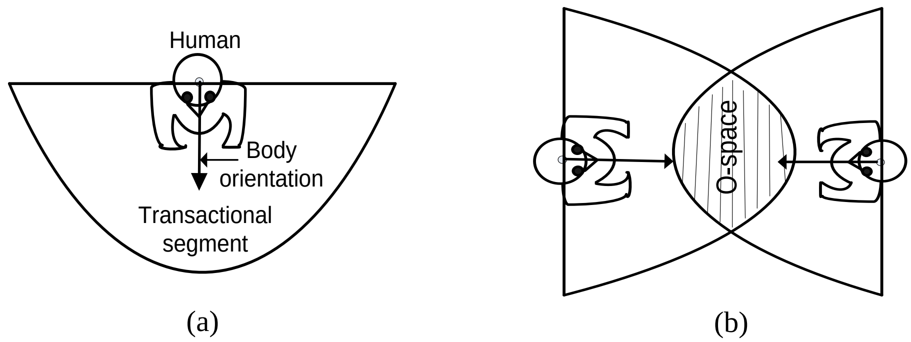


Figure 2.4: (a) Transactional segment and (b) Joint transactional segment

2.7.1 Domains of Spatial formation

Kendon suggested in his observations of a spatial formation systems that people generated three concentric rings or spatial domains [120, 121] (Figure 2.5). The innermost space, the o-space, is created by an overlap of the participants' transactional segments. It is the space into which people project their voices and gazes. In other words, it is a small scale interactional area where overt and explicit actions are located so that everybody within the circle of participants can have an easy and direct access to the ongoing exchange. The second spatial domain, the so called p-space [120, 122], which can be seen in every spatial formation, is a narrow zone where the bodies and personal belongings of people interlocked in spatial formation are placed. The third zone of space intimately associated with the occurrence of spatial formation is thought to be less clearly defined than the o- or p-spaces. Called the r-space, it stretches behind the backs of the spatial formation participants and envelopes the whole encounter in such a way that people inside it are bound into a clear-cut participation unit that is distinct from the rest of the environment. The concept for an r-space comes from observations of the behavior of outsiders. As we saw, when an outsider is going to become a member of an existing system, he approaches the system but stops a little distance away, before being invited in by the current participants.

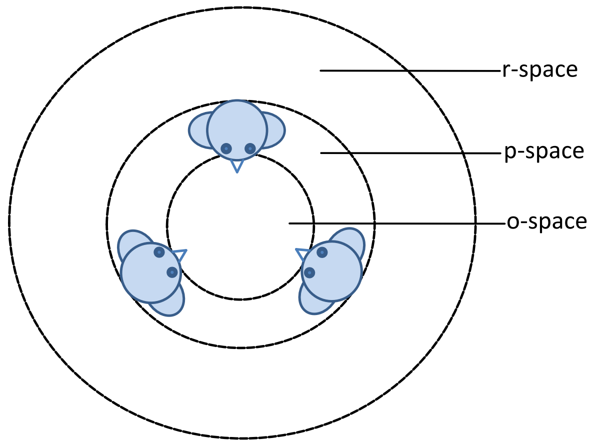
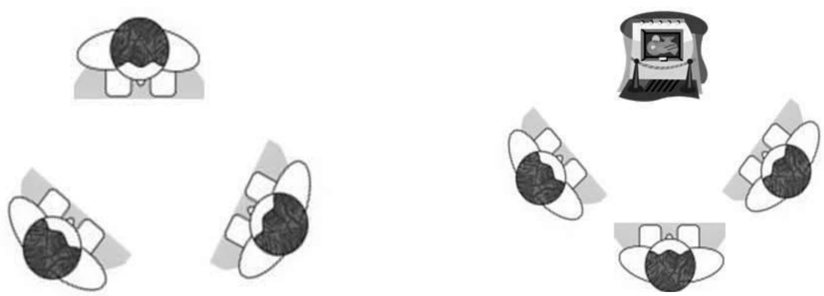


Figure 2.5: Domains of spatial formation

2.7.2 Category of Spatial formation

McNeill categorized the spatial formation into two types: social and instrumental [123, 124] (Figure 2.6). A social spatial formation consists of only people and is identical to Kendon's original definition. On the other hand, an instrumental spatial formation includes a physical object as an element, upon which participants normally gaze. This paper is concerned with the instrumental spatial formation since this is relevant to a museum.



(a) (b)

Figure 2.6: Types of spatial formation. (a) Social and (b) Instrumental

2.7.3 Shapes of Spatial Formation

Kendon showed in addition that some joint activities and spatial interactions are supported by certain spatial formation system *arrangements*, and thus often are encountered in prototypical situations. In a circular arrangement, three participants normally face to a common space and applicable for free-standing groups of three or more (Figure 2.7(a)). In the *Vis-à-vis* arrangement (Figure 2.7(b)) two participants normally face one another directly; a *L-Shape arrangement* in which two contributors are positioned so that the frontal surfaces of their bodies fall on the two arms of an 'L' (Figure 2.7(c)) usually indicates a joint system in which something is shared in the o-space, e.g. an object of interest. As a last arrangement Kendon mentions the *Side-by-side* configuration (Figure 2.7(d)) where the two participants are standing closely together and face the same way. This arrangement is said to occur often in situations where both interactors are facing an outer *edge*, e.g. given externally by the environment in the form of a table, a wall, a kitchen sink, or similar. For HRI it is important to notice that all spatial formation arrangements support a *triadic* relationship between the two interactors and an object of shared interest.

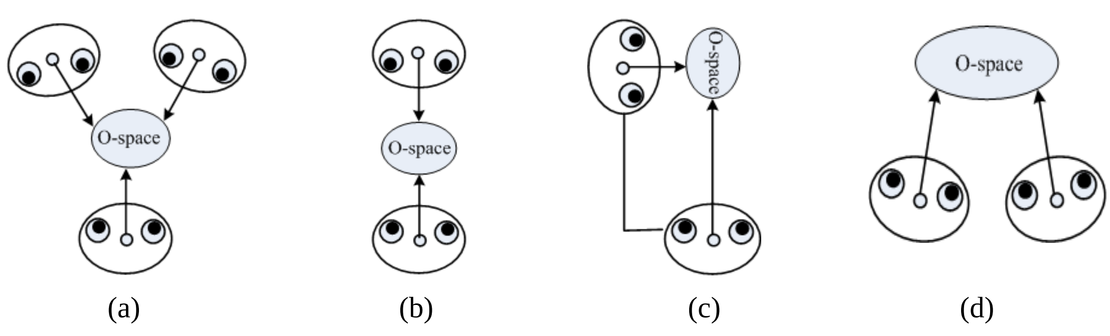


Figure 2.7: Spatial formation arrangement. (a) Circular arrangement, (b) Vis-à-vis arrangement, (c) L-shape arrangement, and (d) Side-by-side arrangement

2.8 Guide-Visitors Interaction in Museum

We worked on developing a museum guide robot through the following procedures. We analyzed some video footage recorded at the National Japanese American Museum in Los Angeles in order to find out how expert human guides explain the exhibit/s towards multiple visitors. We analyzed these data from the perspective of creation of spatial formation. We are especially interested in how museum guide create spatial formation with the visitors. The videos record human guides performing part of their normal work explaining exhibits to a small group of visitors. From our analysis it emerges that there are ways that guides systematically present explanations to multiple visitors. We then develop a museum guide robot that can create spatial formation with the visitors in different situation. Transcript (1) and Figure 2.8 convey a typical fragment of such human guide behavior at the museum.

In transcript (1), MG moves to another exhibit (Figure 2.8(a)). After arriving near the exhibit, MG clears his throat in the second line. "Clearing throat" is one way of indicating that he is waiting for the visitors to come to the exhibit. In line 3, MG employs a pause of 5.0 seconds. By this time the visitors are following MG (Figure 2.8(b) and Figure 2.8(c)). At the 5th line, MG deploys "restart" and a "pause" of 0.2 seconds while asking an involvement question to the visitors ("Have you all heard of picture brides?"). At line 7, some visitors move their heads vertically in response to MG's question. From lines 9-12, V1, V2, and V3 offer verbal responses to MG's question. In lines 13 and 14, MG asks the visitors to come closer (Figure 2.8(d)), an indication that they should form a proper F-formation (Figure 2.8(e)).

Transcript (1) (Picture bride)
Data: Collected at National Japanese American Museum
MG: Guide, V1,V2, V3: Visitors

1 MG: Okay, let's come over here ((Guide moves to
another exhibit))
2 MG: ((clears his throat))
3 (5.0) ((people follow MG))
4 MG: Okay, so you: ah: heard of uh picture bri::des?
5 <Have you all heard picture bri:des? (0.2)
6 See the picture up the::re, those are picture
bri::des.
7 (2.0) ((some people move their heads vertically))
8 MG: You all heard of- you never heard of picture
bri::des,
9 V1: No:::,=
10 V2: =(Not me.)
11 (0.8)
12 V3: [Uh-huh.
13 MG: [Or if you wanna come closer this way so that
14 other (.) people could leave uh: on the other side.s

Symbols Used in the Transcript	
(( ))	Vocalizations which are difficult to convey in text
(5.0) (2.0) (0.8) (0.2)	Pauses are timed in seconds and inserted within parentheses
,	Slight rising tone
:	Stretched sound
<	Speeding up the pace of delivery
?	Final rising tone which may(or may not) indicate a question
-	Short untimed pause within an utterance
=	Overlap
(did)	Guess at unclear word
[	Simultaneous utterances
.	Stopping fall in tone, with some sense of completion

(a) Guide moves to another exhibit

(b) Visitors follow guide

(c) A visitor moves to her appropriate position

(d) Guide asks visitors to come closer

(e) F-formation is formed

Figure 2.8: Human guide and visitors' interaction at actual museum

2.9 Chapter Summary

In this chapter, we presented several aspects of research that we performed prior to developing the spatial formation model for a mobile museum guide robot (which is described in detail in the next chapter). From the background studies and analysis of various recorded video footage of guide-visitor interaction in a museum, we can derive the following interaction patterns for designing a mobile museum guide robot:

1) Museum guide Robots should have the capability to establish a proper spatial formation with the visitors before start its explanation.
2) Robots should attract visitor's attention at the beginning of explanation.

3) Guide robots should identify interested bystanders around itself and invite them into ongoing explanation and thus reconfigure spatial formation again.
4) Guide robots should reconfigure spatial formation while explaining multiple exhibits collectively.
5) Guide robots should create spatial formation during initiation of interaction with the visitor's by judging their behaviors.

The next four chapters will describe a series of empirical studied that are contextualized in and motivated by our objective to design a mobile museum guide robot that is capable to create and control spatial formation with the visitors. The next chapter will describe the detail constraints to create spatial formation and a model will be presented to establish spatial formation at the beginning of explanation by museum guide robot.

Chapter 3

A Empirical Framework to Create Spatial Formation by Guide Robot

Humans are trained in social norms, taught to the young members of a social group [125]. For robot interaction behavior it remains an open question which social norms will be established over time. Equally undeveloped are social norms and rules that robots should know and act upon in posture and positioning. Such behaviors in communicative and interactional encounters that are interpreted as orderly are said to be socially appropriate [13], i.e. they are characterized by being perceived as ordered affairs that go mostly unnoticed and are handled without consciously reflecting about them. From studies in contexts such as museums, a rich descriptive picture is emerging of how people find out how to use technologies by watching those nearby and create engagement and participation through per formative interaction [126], but also how the current generation of museum interactive has tended to priorities constrained interactions and individual use [127]. Brignull and Rogers [128] noted how physical aspects of the environment could influence the likelihood of people engaging with a large display at an event in a public space. They suggest for example, placing a display in a location with a constant flow of people. They also discuss how other people can create social affordances within a space: the so called 'honey pot' effect. Hornecker has described how interactive museum exhibits can index into the surrounding context [129]. While this research is impressive, so far very few studies have revealed how a robot should behave to initiate interaction with multiple visitors before starting its explanation.

3.1 Guide Robot System

The purpose of our research is to develop a spatial formation model for mobile museum guide robots that can create and control spatial formation with the visitors in different situations. Thus, for serving our purpose, we have developed a robotic system that can establish spatial formation with the visitors. In the following sections, we discuss the architecture of our robotic systems and its behaviors.

3.1.1 System Overview

Yonezawa et al. [130] have developed a vision system that can detect multiple visitors' gaze directions simultaneously and they presented a robot with a system that deploys appropriate speech and actions depending on the gaze direction of people looking at an advertisement. Although these studies pertain to robots interacting with multiple visitors through speech and action, they are not based on analysis of actual human guides and visitors in interaction. Based on our findings from the interaction of human guide and visitors, we developed a museum guide robot system utilizing a humanoid robot Robovie-R Ver.3 (Vstone), which is a research platform for human-robot communication (Figure 3.1).

(a)

Parameter	Specification
Size	108cm x 50cm x 52cm
Weight	35kg
DOF	17 (Eyes:2, Neck:3, Arms:4x2, Base:2 wheels)
Servo Motors	VS-SV1150x7, VS-SV3310Jx4, MICRO STD/Fx4
Motor	Maxon Brushless Motorx2
Sub CPU board	VSRC003HV (ARM7 60MHz)
I/O	Touch sensorx11, USB Camerax2, Mono microphonex2, speakerx1.
Battery	12V 28Ah

(b)

Figure 3.1: (a) Robovie 3 and (b) Properties of Robovie 3

3.1.1.1 Hardware Configuration

The robot can move via wheels installed on the bottom, and can move its head and arms by controlling its joints. Its head, which incorporates eye cameras and an ear microphone, moves along three axes (Yaw, Roll and Pitch) like a human head. Our system utilizes two general-purpose PCs, connected by a wired network in the current implementation. In our vision system we incorporate one omni-directional camera and two laser range sensors (URG-04LX by Hokuyo Electric Machinery) (Figure 3.2). Omni-directional camera is attached to a pole installed on the back of the robot, and can detect and track visitor’s faces and their face direction.

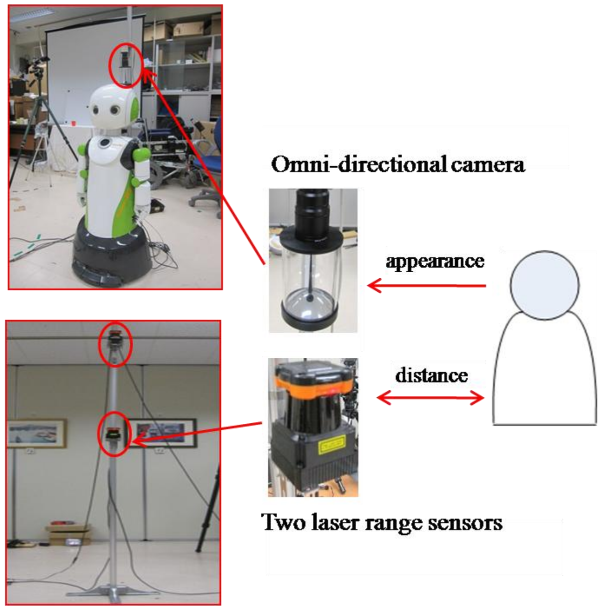


Figure 3.2: Vision system

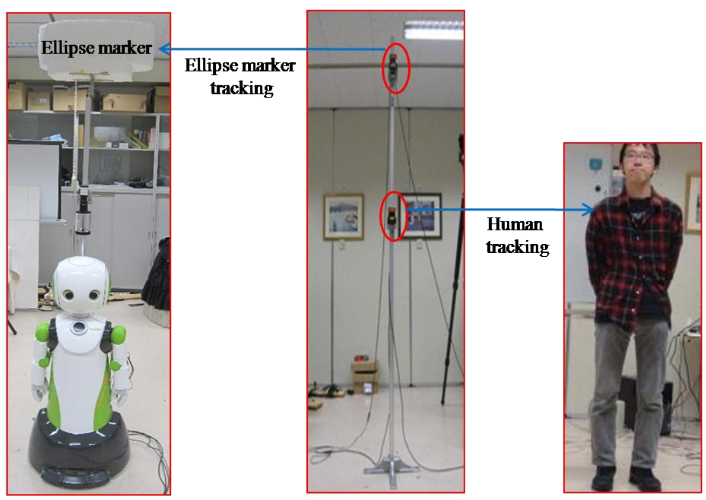


Figure 3.3: Tracking of ellipse marker and human

The two laser range sensors are attached to another long pole which is kept at a fixed position just in front of the experimental area. We also attached an ellipsoid marker to the robot's body, and put one of the two laser range sensors at the marker's height on the long pole. This laser range sensor detects the ellipsoidal marker to obtain the position and orientation of the robot, while the other laser range sensor is used to track the position and body orientation of the visitors (Figure 3.3). The overall system overview is shown in Figure 3.4.

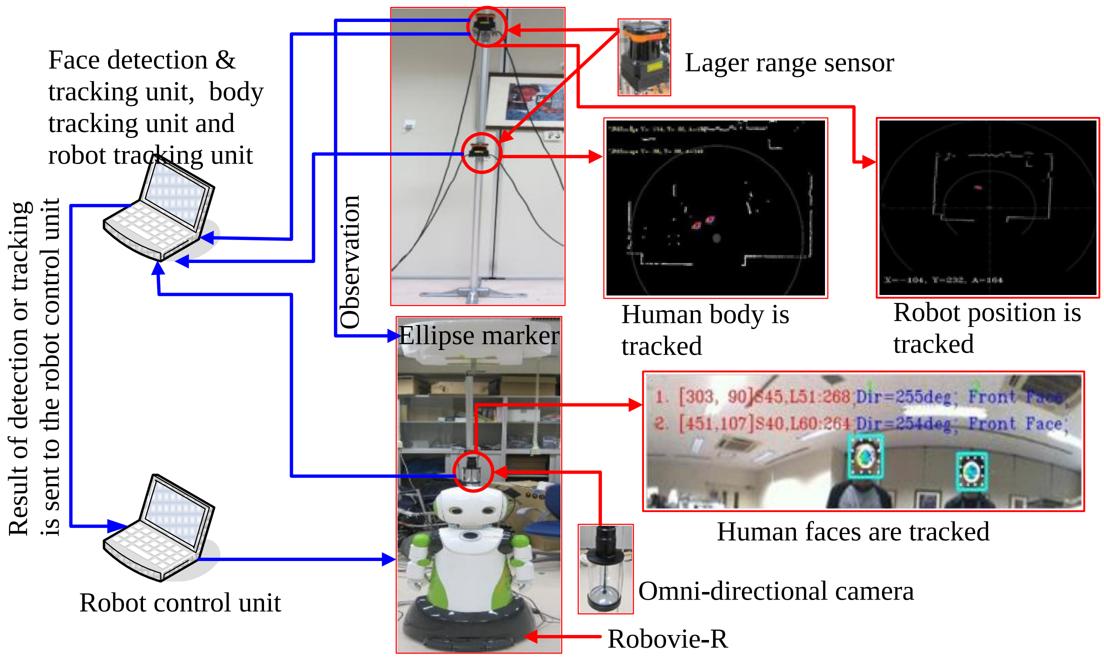


Figure 3.4: System Overview

3.1.1.2 Software Configuration

Our system consists of four software units: the face detection and tracking unit, the human body tracking unit, the robot position tracking unit, and the robot control unit. One of the two PCs runs the robot control program, while omni-directional camera and laser range sensors are connected to the other PC which runs the face detection and tracking unit, human body tracking unit, and robot position tracking unit. The results of the detection and tracking are sent to the robot control unit. During its explanation of exhibits, the robot performs predetermined bodily non-verbal actions, such as facing towards the visitors, gesturing with its hands, and pointing to the exhibits.

The timing of these nonverbal behaviors is programmed based on analyzing the videotaped interactions between human guides and visitors at the museum.

3.1.2 Tracking Framework

While several methods for tracking people using laser range sensor have been proposed, many of these use laser range sensors distributed in the environment, and track only positions of people [131, 36, 27]. In contrast, our system can track the position and body orientation of visitors and robot. In particular, over the last few years, the particle filter framework is reported to be effective for tracking people [132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143]. We employ a particle filter frame work [134] to track visitors' positions and the orientation of bodies based on the position data captured from the laser range sensor. We also use particle filter framework incorporating Ada-Boost-based cascaded classifiers to detect and track visitor's head based on panorama images captured from omni-directional camera. In the following section, we will describe the details of our tracking method based on the particle filter and Ada-Boost-based cascaded classifiers. We use the background subtraction technique to detect tracking targets. First, we capture the background initial image when no one in the room. Then, we find the difference between the background initial image and the current laser image to detect tracking targets.

3.1.2.1 Particle Filtering

In this section we briefly give an overview of the particle filter framework. The particle filter is a Bayesian sequential importance sampling technique, which recursively approximates the posterior distribution using a finite set of weighted samples. A set of samples can be used to approximate non-Gaussian distribution and they are propagated by a state transition model for each recursion. It thus allows us to realize robust tracking against observation noise and abrupt changes of target's motion.

Suppose that a state of a target at time t is denoted by the vector $\mathbf{x}_t$, and that the observation of camera image at time t is denoted by the vector $\mathbf{z}_t$. Then all the observations up to time t is $\mathbf{Z}_t = \{\mathbf{z}_1, \ldots, \mathbf{z}_t\}$. Assuming the Markov process enables us to describe a prior probability $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ at time t by

$$P(x_t \mid Z_{t-1}) = \int P(x_t \mid x_{t-1}) P(x_{t-1} \mid Z_{t-1}) dx_{t-1} \qquad (1)$$

where $P(\mathbf{x}_{t-1} \mid \mathbf{Z}_{t-1})$ is a posterior probability at time $t-1$, and $P(\mathbf{x}_t \mid \mathbf{x}_{t-1})$ is a state transition probability from $t-1$ to t. Assuming that $P(\mathbf{z}_t \mid \mathbf{Z}_{t-1})$ remains constant, a posterior probability $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ at time t is described by

$$P(x_t \mid Z_t) \propto P(z_t \mid x_t) P(x_t \mid Z_{t-1}) \qquad (2)$$

where $P(\mathbf{z}_t \mid \mathbf{x}_t)$ is a likelihood and $P(\mathbf{x}_t \mid \mathbf{Z}_{t-1})$ is a prior probability at time t. Tracking is then achieved by calculating the expectation of posterior probability $P(\mathbf{x}_t \mid \mathbf{Z}_t)$ at each time.

In the particle filter framework, the probability distribution is approximated by a set of samples $\{s_t^{(1)}, \ldots, s_t^{(N)}\}$. Each sample $s_t^{(n)}$ representing a hypothesis has the weight $\pi_t^{(n)}$ representing a corresponding discrete sampling probability. The hypothesis evaluation, which is also called as the sample evaluation, is to compute the weight $\pi_t^{(n)}$ by considering the observation likelihood corresponding to the sample $s_t^{(N)}$. A set of samples is then updated by the following procedures at each time.

1. Sampling:

Select samples $\{s_{t-1}^{(1)}, \ldots, s_{t-1}^{(N)}\}$ in proportion to weight $\{\pi_{t-1}^{(1)}, \ldots, \pi_{t-1}^{(N)}\}$ corresponding to sample $\{s_{t-1}^{(1)}, \ldots, s_{t-1}^{(N)}\}$.

2. Propagation:

Propagate samples $\{s_{t-1}^{(1)}, \ldots, s_{t-1}^{(N)}\}$ with state transition probability $P(\mathbf{x}_t \mid \mathbf{x}_{t-1} = s_{t-1}^{(n)}) t{-}1)$ and generate new samples $\{s_t^{(1)}, \ldots, s_t^{(N)}\}$ at time t.

3. Weight computation:

Compute weights $\pi_t^{(n)} \approx P(\mathbf{z}_t \mid \mathbf{x}_t = s_t^{(n)})$ corresponding to sample $s_t^{(n)}$ by evaluating a likelihood through camera images($n = 1,2, \ldots, N$).

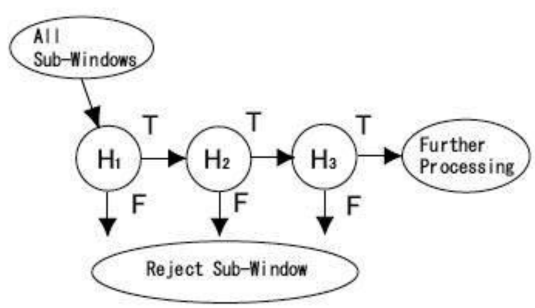
3.1.2.2 AdaBoost-based cascade classifier

Numerous methods for detecting faces in general images have been proposed. Among them, the AdaBoost-based face detector using Haar-like features has become popular because of its accuracy and robustness against observation with low resolution or varying illumination conditions. The AdaBoost-based classifier consists of linearly connected weak classifiers. Viola and Jones arranged the classifiers in a cascade structure and proposed an efficient computation technique for Haar-like features [144]. Though the training of AdaBoost-based cascaded classifiers (hereafter referred to in this chapter as the 'cascaded classifier') requires huge amount of time, the cascaded classifier rapidly detects a face because most of non-face target regions are rejected in an early stage of the cascade. This cascade is effective in the evaluation phase even in the particle filter framework.

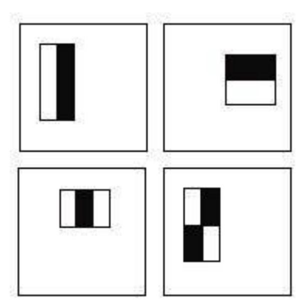
In Figure 3.5(a), H_i represents a strong classifier. Each strong classifier classifies an input image into a positive or a negative. Only positive images are used as the input of the next strong classifier. At each stage, a strong classifier is trained to detect almost all face images while rejecting a certain fraction of non-face images. For instance, the classifier at each stage is trained to eliminate 50% of the non-face images while falsely eliminating is only 0.1% of the face images. After passing 40 stages, we can then expect a false alarm rate about $0.5^{40} \approx 9.1\times10^{-13}$ and a hit rate about $0.999^{40} \approx 0.96$. Thus the face detector detects almost all the face images and rejects almost all the non-face images. A strong classifier $H_i(x)$ at each stage of the cascade consists of many weak classifiers $h_t(x)$ (Figure 3.5(b)). This can be described as follows:

$$H_i(x) = sgn\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3}$$

where T is the number of weak classifiers and $\alpha_t = log\frac{1-\varepsilon_t}{\varepsilon_t}$. We note that ε_t is an error rate specified in the training phase. Each weak classifier $h_t(x)$ evaluates a target image region by using Haar-like features. The weak classifier performs that the sum of the intensity of pixels located within the black rectangles is subtracted from the sum of the intensity of pixels located within the white rectangles. The AdaBoost algorithm selects efficient features to classify the target image region among a huge variety of features.



(a) Cascade of classifiers

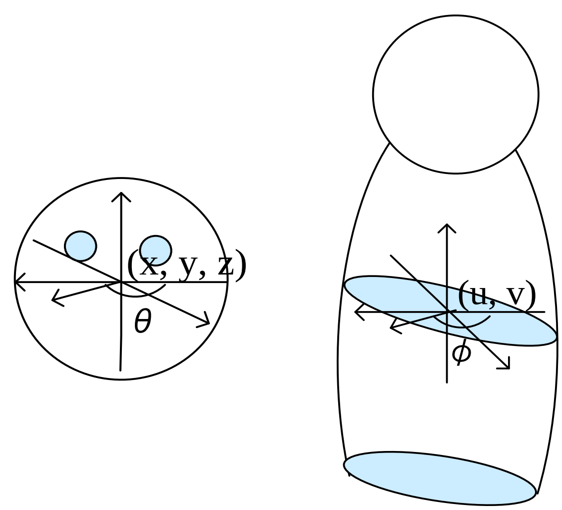


(b) Examples of features

Figure 3.5: Cascade classifier

3.1.3 Modeling Human as Tracking Target

The visitor is modeled as shown in Figure 3.6. We assume that the laser range sensor is placed horizontally at the visitor's shoulder level so that the contour of the visitor's shoulder can be observed as part of an ellipse. We then use the ellipse as a model to track the position and the direction of the visitor's body. The visitor's head position and direction are also tracked using the omni-directional camera. We use an ellipsoid as a model to track the visitor's head. We assume a coordinate system represented by the X- and Y- axes aligned on the ground plane, with the Z-axis representing the vertical direction from the ground plane. We assume that visitors walk without tilting their heads, and that therefore the orientation of a human head can be identified by θ representing the rotation around the Z-axis. Thus, the final model of the tracking target is represented by the center coordinates of the ellipse $[u,v]$, rotation of the ellipse Φ, center coordinates of the ellipsoid $[x,y,z]$ and rotation of the ellipsoid θ.



(a) Human head model (b) Human body model

Figure 3.6: Human model as tracking target

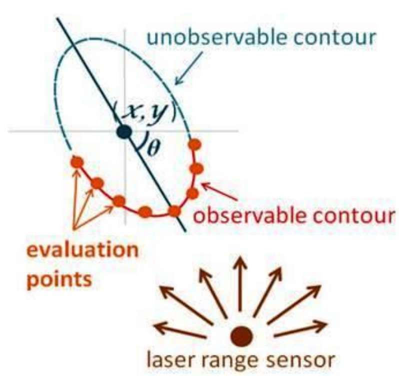
3.1.4 Likelihood Evaluation

I will now discuss the evaluation of the samples based on the observations of the laser range sensors and the omni-directional camera.

3.1.4.1 Evaluation Based on Laser Image

Position data captured with the laser range sensor is mapped onto the 2D image plane (what we call a "laser image") and used for the likelihood evaluation of a body. The likelihood evaluations of the samples are based on the contour similarity between the model and a visitor's upper body that is partially observed with the laser range sensor. The contour observation model is shown in Figure 3.7(a). The likelihood of each sample is evaluated as shown in Figure 3.7(b). The likelihood is evaluated from the maximum distance between evaluation points and the nearest distance data using:

$$w_{t,lser}^{i} = exp\left(\frac{-d_{max}^{2}}{\sigma_d}\right) \qquad (4)$$



(a) Contour observation model

(b) Evaluation based on maximum distance

Figure 3.7: Likelihood evaluation by laser image

where $w_{t,laser}^{(i)}$ is the likelihood score based on the laser image, and $d_{\max}$ is the maximum distance between evaluation points and the nearest distance data. At each time instance, once the distance image is generated from the laser image, each distance d_n is easily obtained. The σ_d is the variance derived from d_n.

3.1.4.2 Evaluation Based on Omni-directional Camera Image

We use the variables $x_t^{(i)}, y_t^{(i)}, z_t^{(i)}$ and $\theta_t^{(i)}$ in a sample $s_t^{(i)}$ for evaluating the likelihood using AdaBoost-based classifiers [144]. Our human tracking system can recognize *u,v* and *Φ* (position and orientation of the target person in the world coordinate system). Our system can also recognize the position and orientation of the robot by using the ellipsoidal marker in the world coordinate system. Thus the system can calculate the position and orientation of the target person with respect to the robot because the omni-directional camera is fixed on the pole behind the robot. In other words, we can calculate the position and orientation of the target person with respect to the omni-directional camera because its position and orientation (external parameters of the camera) can be determined when the robot position and orientation are localized. (Because the internal parameters are calibrated in advance, we can use the same internal parameters through the experiment.) Therefore, we can project the position of the target person's face onto the camera images by assuming the parameter z (which means target person's height).

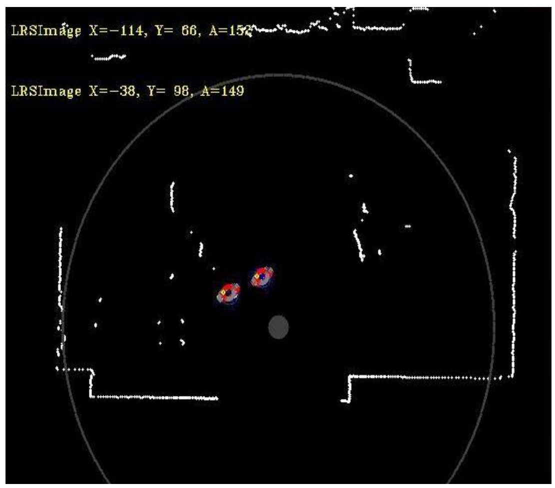
We evaluate samples by applying the AdaBoost-based classifiers over the projected image region and the number of stages passed in the cascade is employed as the likelihood of a human head. While an outline evaluation procedure is given below, a detailed account of the evaluation method is provided in [145]. Note that we use seven classifiers and that each classifier is trained respectively to detect a human head from a particular direction, such as front, 90^0 left, 90^0 right, and so on.

1) Project the sample's variables $x_t^{(i)}, y_t^{(i)}, z_t^{(i)}$ and $\theta_t^{(i)}$ at time t onto the omni-directional camera image and obtain a projected position and size of the head. By considering the location and the direction of a head, we can also calculate the direction of a human head relative to the omni-directional camera $\theta_{t,cam}^{(i)}$.
2) Extract the square image region corresponding to the head and extend it into a normal image based on its position and size.
3) Resize the extracted image region to obtain an image region $g_t^{(i)}$ (e.g. 24x24 sizes) as the input of the cascaded classifier.
4) Select a classifier by considering the direction of a human head relative to the omni-directional camera $\theta_{t,cam}^{(i)}$. For example, if we use three classifiers such as front, 90^0 left, 90^0 right, then the front is selected in the case of $-45^0 \leq \theta_{t,cam}^{(i)} \leq 45^0$, the 90^0 left is selected in the case of $45^0 \leq \theta_{t,cam}^{(i)} \leq 135^0$, and so on.
5) Apply the selected classifier to the image region $g_t^{(i)}$ and obtain the likelihood score of a human head $w_{t,cam}^{(i)}$.

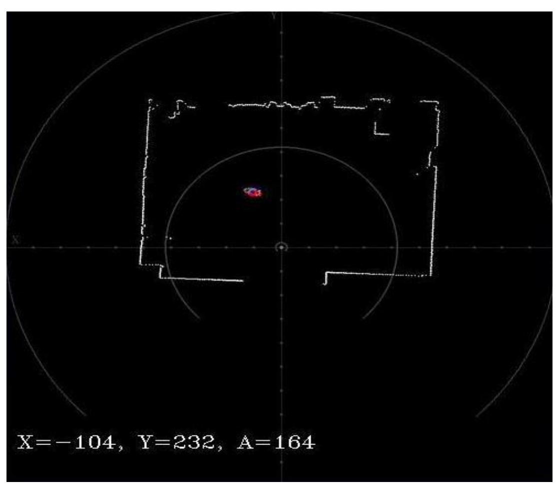
Basically, we track multiple target persons by using the particle filter independently. When a person is occluded by another person, the system terminates the tracking and when the person appears again, the system starts tracking again. We use the total likelihood value to decide whether the target person is occluded or not.

An example of tracking is shown in Figure 3.8. The result of human head tracking using the omni-directional camera is shown in Figure 3.8(a). The system also tracks the human body and the ellipsoidal marker using the particle filter. The lower and upper range sensors in Figure 3.4 are used to track human bodies and the ellipsoidal marker on the robot, respectively. We use the same tracking technique for tracking both humans and marker.

(a) Human head tracking based on particle filter



(b) Human body tracking using particle filter



(c) Ellipsoidal marker tracking using particle filter

Figure 3.8: Example of tracking

The red ellipses in Figure 3.8(b) and Figure 3.8(c) show the human bodies and the ellipsoidal marker tracked by the system, respectively.

3. 2 Proposed Modeling of Interaction

3.2.1 Model of Spatial Formation

The spatial formation arises whenever two or more people sustain a spatial relationship in which the space between them is one to which they have equal, direct, and exclusive access. The constraints to establish a proper spatial formation are as follows:

Constraint of Proximity:

- Distance between the robot and the visitors: Ranges from 90 cm to 120 cm (Figure 3.9(a)).
- Distance between the robot and the exhibits: About 110 cm, fixed in all cases for all exhibits (Figure 3.9(a)).

Constraint of Visitor's Face Direction:

- Visitor's face direction: Should be towards the robot or the exhibit (Figure 3.9(b)).

Constraint of Body Orientation:

- Visitor's body orientation: Should be in the direction between the robot and the exhibit (Figure 3.9(c)).
- Robot's body orientation: The robot turns its body 30^0 towards the exhibit to explain the exhibit (Figure 3.9(d)).

Constraint of Robot's Field of View (FOV):

- Robot's field of view (FOV): All angles between the robot's body orientation and each vector from the robot to each visitor should not be over 75 degrees (Figure 3.9(d)). The angle between the robot's body orientation and the vector from the robot to exhibit is fixed as 25 degrees.

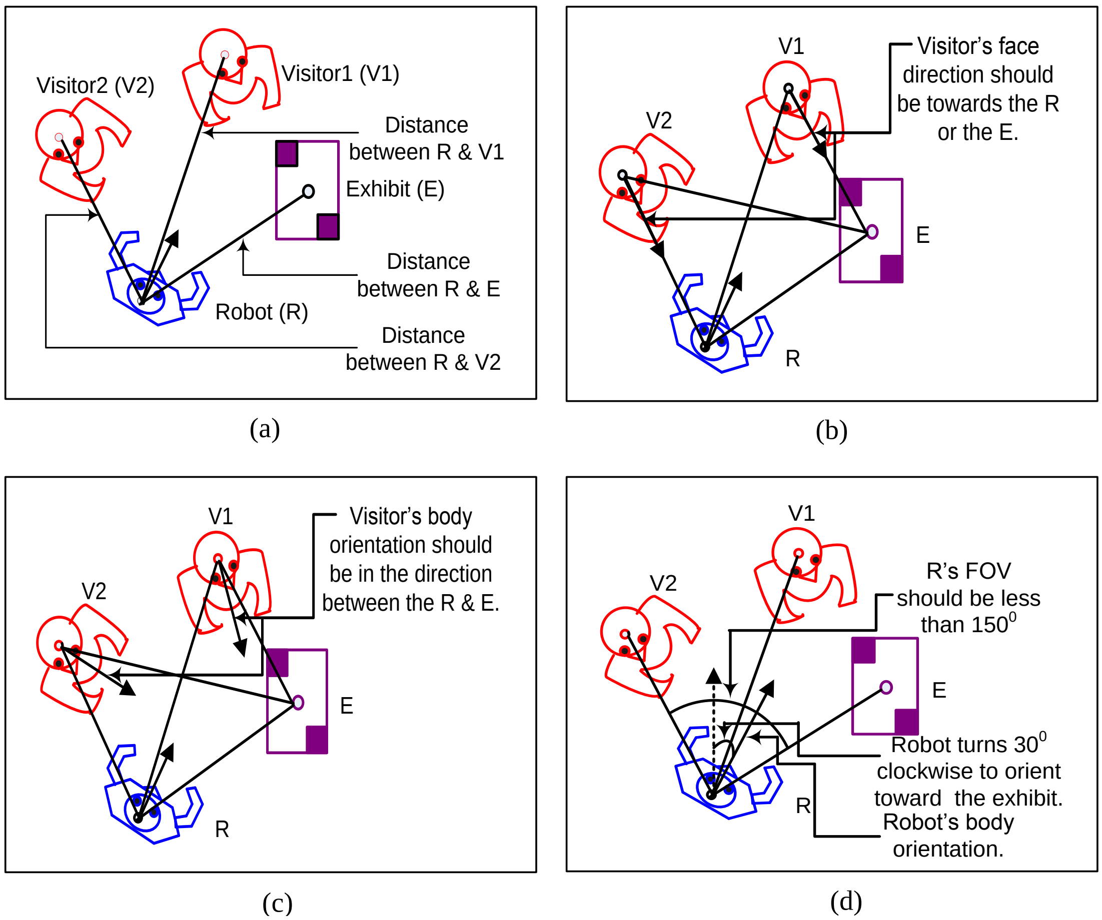


Figure 3.9: Spatial arrangement to create spatial formation

3.2.2 Model to Achieve Mutual Gaze

Goodwin discussed some systematic ways in which speakers capture the attention of a recipient [20]. When a speaker begins to utter a sentence, and if s/he finds recipients are not gazing towards him or her, the speaker can use the techniques of "restart" and/or "pause" in the delivery of the utterance. "Pause and restart" strategy plays an important role to establish spatial formation properly. The robot observes the face direction and body orientation of the visitors before beginning its explanation of any exhibit. If, at the beginning of the speaker's (robot) turn, the face direction is detected as not directed toward the robot or exhibit, or body orientation is detected as not in the direction between the robot and the exhibit, the robot employs the "pause and restart" strategy. The format of "restart" in a sentence is as follows:

Sentence of the speaker: *[Fragment] + [New Beginning]*

*X*______________

For a "pause," the speaker (robot) starts the sentence, waits until the gaze of its recipient is secured, and then continues the sentence. The format is as follows:

[Beginning] + [pause] + [Continuation]

*.............X*________

In our system, we implemented a combined version of "pause and restart," the format of which is as follows:

[Beginning] + [pause] + [new Beginning]

*.............X*_______

In this format, the speaker (robot) starts the sentence (beginning), waits (pause) until the gaze of its recipient is secured, and then restarts (new beginning) the sentence again. Before starting its explanation of an exhibit, the robot observes the direction of the visitors' faces. If it finds that upon the start of its turn the visitors' orientation is not directed towards the exhibit and they are not gazing at the robot either (we experimentally allow 15 degrees tolerances about the orientation decision), the robot employs a "pause and restart" according to the format just outlined. The solid line below the sentence structure indicates that the recipient is gazing toward the speaker, no line indicates that the recipient is gazing elsewhere, and the 'X' marks the point at which the recipient's gaze reaches the speaker. The dotted line represents the time required for the recipient to move his/her gaze from some other position to the speaker. Figure 3.10 illustrates the procedure of pause and restart.

The solid line below the sentence structure indicates that the recipient is gazing toward the speaker, no line indicates that the recipient is gazing elsewhere, and the 'X' marks the point at which the recipient's gaze reaches the speaker. The dotted line represents the time required for the recipient to move his/her gaze from some other position to the speaker.

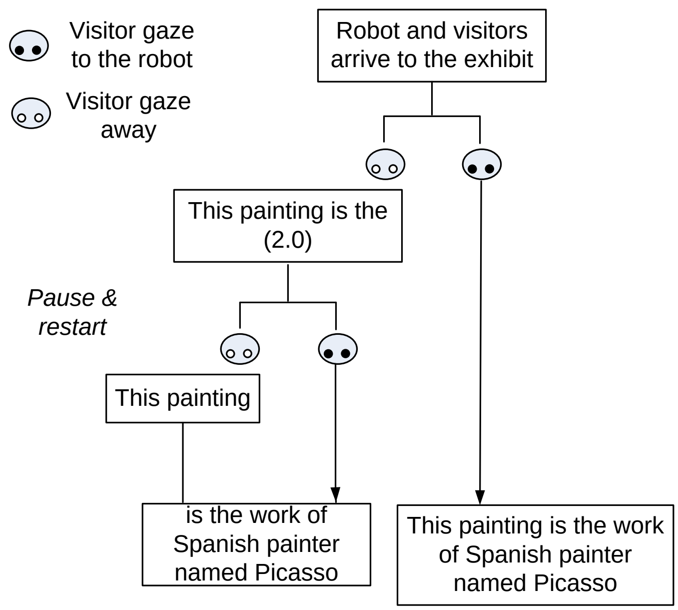


Figure 3.10: Procedure to implement "pause and restart"

Figure 3.11 shows the graphical representation of the proposed model to attract visitor's attention. In Figure 3.11(a), guide robot noticed that V1's body orientation and V2's face direction was not in the direction between the robot and the exhibit at the beginning of its explanation. The robot then turned its head towards them and employed "pause and restart" to attract visitor's attention (Figure 3.11(b). This caused a change in the visitor's body orientation and face direction; V1 and V2 turned their body and head respectively towards the robot and mutual gaze was established (Figure 3.11(c)).

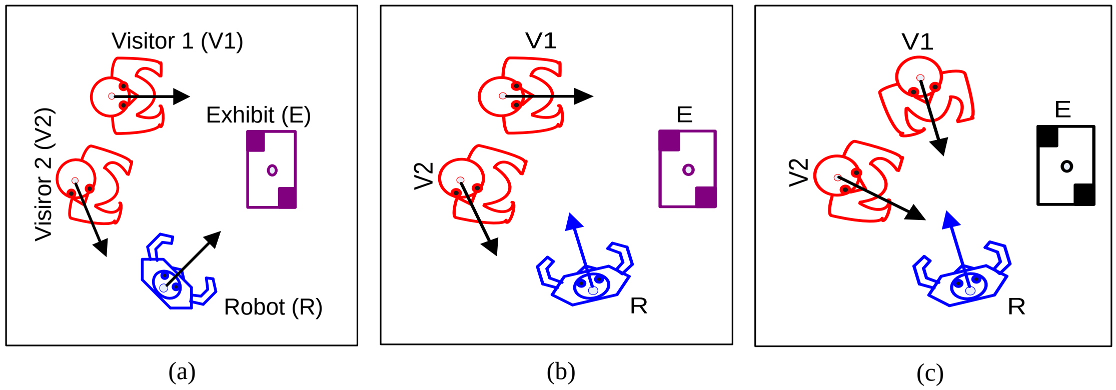


Figure 3.11: Modeling to attract visitor's attention

In order to implement "pause and restart," we consider the first sentence of each script explaining the exhibits, as in the following:

Script 1: First sentence for explaining the first exhibit, "Te Nave Nave Fenua"

Robot: This is a- (2.0) This is a famous work of Gauguin.

Script 2: First sentence for explaining the second exhibit, "Still Life with Skull"

Robot: This painting-(2.0) This painting is the work of the Spanish painter named "Picasso"

Script 3: First sentence for explaining the third exhibit, "The Vegetable Float"

Robot: This float is- (2.0) This float is made of vegetables.

Script 4: First sentence for explaining the fourth exhibit, "Landscape Painting".

Robot: This famous work- (2.0) This famous work is a typical landscape of South France.

In our proposed model, the guide robot deploys “pause and restart” depending on the situation. In all scripts, a restart with a preceding pause of 2 seconds is used. We have determined this length from observations of human guides as described in section 2.8 in chapter 2.

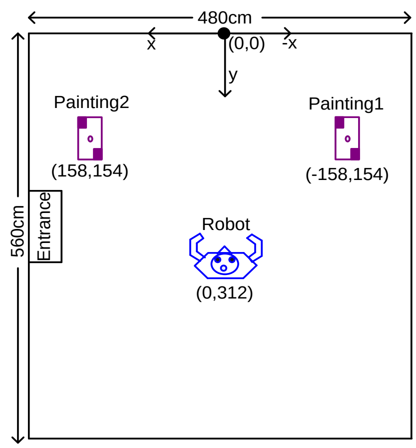
After arriving at the predefined position for explanation of the exhibits, the robot should examine whether or not a proper spatial formation has been established by measuring the positioning data provided by the vision system discussed in section 3.1. If not, the guide robot takes the initiative to create spatial formation using the proposed model described in section 3.2. After creating a proper spatial formation, the guide robot starts its explanation.

3.3 Evaluation Experiments

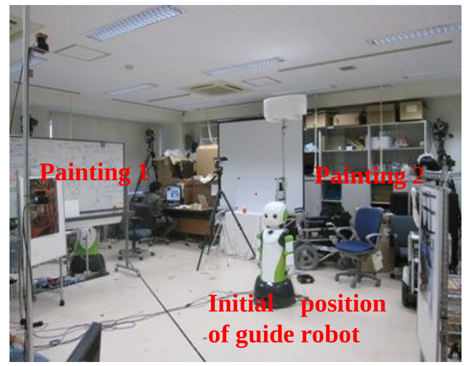
To test the robot’s effectiveness, two experiments were performed in a laboratory. The experiment was performed to test the robot’s effectiveness at creating spatial formation with the visitors before start its explanation. In the experiment, guide robot followed the proposed model described in section 3.2.

3.3.1 Experimental 1: Create Spatial Formation Considering the Constraints of Proximity, Face Direction and FOV

In the first experiment, guide robot creates spatial formation considering constraints of proximity, visitor’s face direction and constraint of robot’s FOV (described in section 3.2.1) with the visitors before start its explanation. In this experiment we did not consider the constraints of body orientations.



(a)



(b)

Figure 3.12: Experimental settings: (a) Schematic diagram of experimental area. (b) Guide robot with two paintings.

3.3.1.1 Experimental Environment

The experiment was conducted in our laboratory, under the assumption that it was a small art museum. The experimental area was restricted to 480cm x 560cm area. In the experiment, two paintings were placed in the experimental area and the robot with the proposed system waited in the middle of the experimental area as shown in Figure 3.12. The long pole with laser range sensor was placed in front of the experimental area.

3.3.1.2 Experimental Procedure

A total of 16 people (10 males, 6 females) participated in the first experiment. We formed 8 groups with 2 members in each group. Among the 8 groups, 4 groups participated in the experiment with the mobile guide robot system outlined above, and the remaining 4 groups did so with a robot which was not equipped with the capacity to form an F-formation. Participants were not informed of which robot was which. Before beginning the experiment, we provided participants with the scenario and instructions, as well as a brief idea about the F-formation and the "pause and restart" strategy.

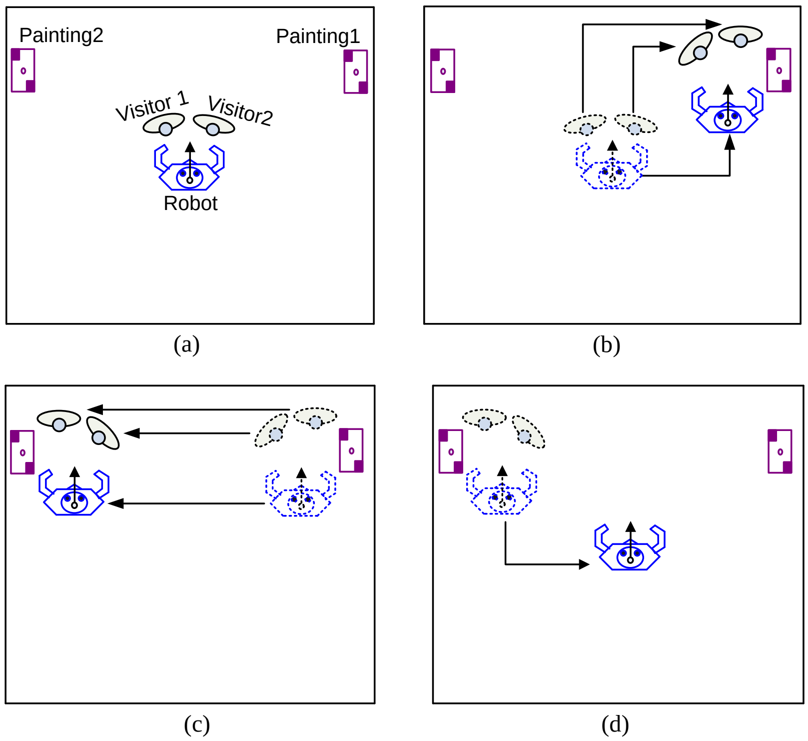


Figure 3.13: Schematic diagram for experiment 1

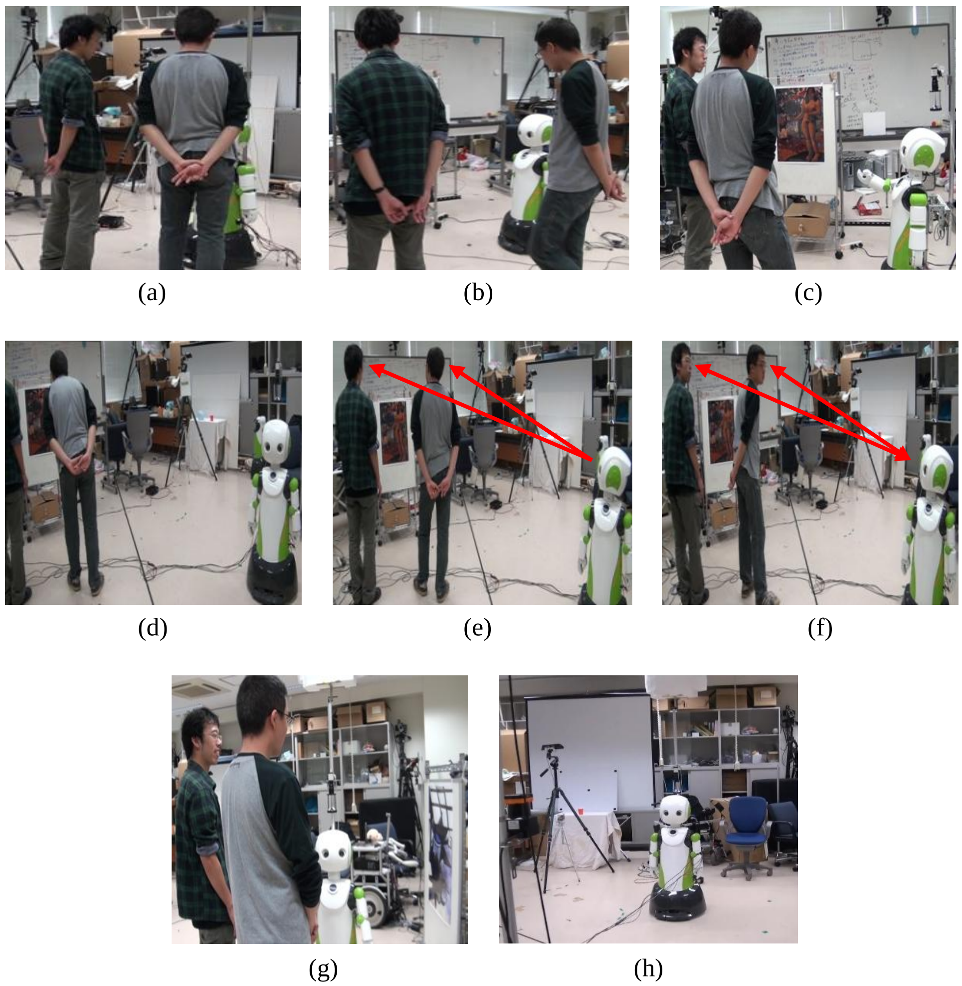
Initially, the robot with the proposed system waits in the middle of the experimental area and turns its head from side to side in order to display availability to visitors. A schematic diagram of the main tasks performed by the robot is given in Figure 3.13. The order of these main tasks is as follows:

1) When the robot finds visitors coming into its immediate vicinity, it says, "May I explain these paintings to you?" If the visitors' gaze turns toward the robot's direction for three seconds, the robot system considers the visitors to be highly interested in the exhibits (Figure 3.13(a)). This length is also empirically determined as in the pause length described in section 3.2.2.
2) The robot then guides the visitors to the first exhibit (Figure 3.13(b)).

3) The robot verifies the distance between itself and the visitors in terms of the body positions (described in section 3.2.1). It verifies distance using position data provided by the laser sensors.
4) If the visitors are not within the range, the robot says to the visitors, "Please come closer" or "Please move back a little" depending on the situation.
5) Next, the robot verifies the direction of the visitors' faces using data provided by the omni-directional camera. If the direction is towards the robot or the exhibit, the robot begins to explain the first exhibit. If not, the robot starts its explanation with a "pause and restart" (described in section 3.2.2). After completing its explanation, the robot moves to the next exhibit and at the same time invites the visitors to follow along (Figure 3.13(c)).
6) The robot repeats tasks (3) to (5) for all the exhibits.

The robot returns to its initial position and waits for more visitors to arrive (Figure 3.13(d)).

Figure 3.14 shows the details of the experiment using the proposed method. In Figure 3.14(a), the guide robot finds two visitors who are interested in the paintings and says, "May I explain these paintings to you?"Having confirmed their interest, the robot moves towards the first painting while saying, "Please follow me," and the visitors do so (Figure 3.14(b)). In Figure 3.14(c), the robot and the visitors arrive at the first painting. After verifying the proximity and face direction constraints of the visitors, the robot begins its explanation. After explaining the first painting, the robot moves to the second exhibit, but the visitors are still looking at the first painting (Figure 3.14(d)). The robot thus turns its head towards the visitors (the red arrow in Figure 3.14(e)), indicating that it is waiting for the visitors to approach the second painting. Mutual gaze is established (the red arrow in Figure 3.14(f)), where upon the visitors move towards the second painting (Figure 3.14(g)). After arriving at the second painting, the robot again verifies constraints of proximity and constraints of face direction of the visitors. The robot then continues explaining the second. Finally, it returns to its initial position upon completing its explanation (Figure 3.14(h)).



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

Figure 3.14: Experimental scene of first experiment.

3.3.1.3 Experimental Conditions

In the experiment, the robot based on our proposed model was compared with a robot that did not employ the proposed model.

a) *Proposed Robot*: The robot behaves based on the model outlined in this paper.

b) *Conventional Robot*: The robot begins its explanation after finding the faces of visitors. It does not care whether or not a proper spatial of F-formation is formed, nor does it utilize the 'pause and restart' strategy. It explains the exhibits with the same preprogrammed nonverbal behaviors as the proposed robot.

As mentioned above, the proposed robot explained the al two exhibits to 4 groups, and the conventional robot did so for another 4 groups. After the experiments, we asked participants to fill out a questionnaire. They were asked to subjectively rate the robot's effectiveness on a seven-point Likert scale, with the range: 1-very ineffective, 2-ineffective, 3-somewhat ineffective, 4-undecided, 5-somewhat effective, 6-effective, 7-very effective. The questionnaire items were as follows:

1) Evaluating the effectiveness of F-formation system.
2) Evaluating the effectiveness of "pause and restart."
3) Overall evaluation of the robot.

3.3.2 Experimental 2: Create Spatial Formation Considering the Constraints of Proximity, Body Orientation, Face Direction, and FOV

In the first experiment, we did not consider the constraint of visitor's body orientation to create spatial formation. According to Kendon [13], body orientation plays important role for the creation of spatial formation. Moreover, we provided the participants with a brief idea about the F-formation and the "pause and restart" strategy before the experiment. This might give them some preconceived ideas. Also, we did not record the sensor data and could not objectively analyze the behaviors of the robot and the participants. Thus, we performed the second experiment.

The second experiment was conducted in the same laboratory as first experiment. In the experiment, four paintings (P1, P2, P3, and P4) were placed in four corner of the area and the robot with the proposed system waited in the middle of the experimental area as shown in Figure 3.15. The long pole with laser range sensor was placed in front of the experimental area.

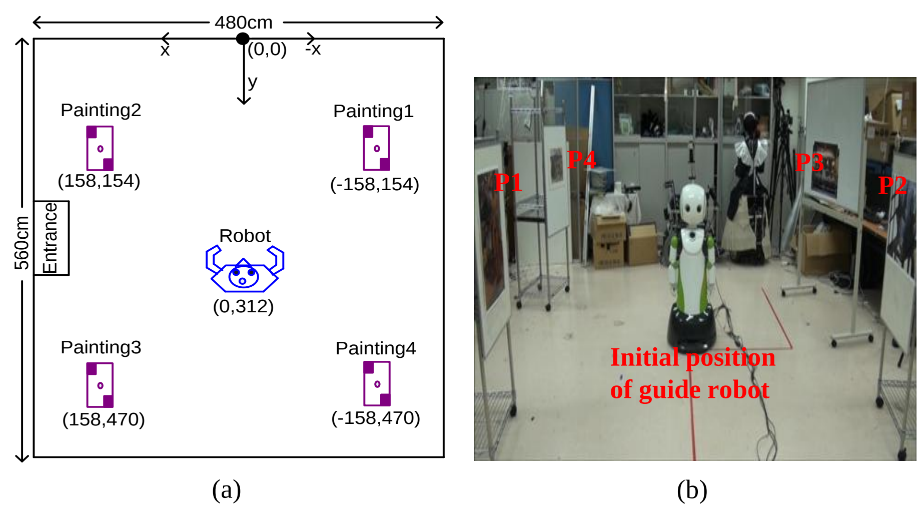


Figure 3.15: (a) Overview of the experimental area. (b) Mobile guide robot and four paintings.

3.3.2.1 Experimental Procedure

A total of 16 graduate students (average age is 29) from Saitama University participated in the experiment. We formed 8 groups with 2 members in each group. The experiment followed a "within subject" design. The robot explained four paintings to all groups. Among the 8 groups, Group A (groups no. 1, 3, 5, and 7) participated in sessions where the robot explained the first two paintings as the proposed mobile guide robot system outlined above, and the remaining two paintings as a robot not equipped with the capacity to form a spatial formation. On the other hand, Group B (groups no. 2, 4, 6, and 8) participated in sessions where the robot did this in reverse. Participants were not informed of which robot was which.

3.3.2.2 Guide Robot Behaviors

Initially, the robot with the proposed system waits in the middle of the experimental area and turns its head from side to side in order to display availability to visitors. A schematic diagram of the main tasks performed by the robot is given in Figure 3.16. The order of these main tasks is as follows:

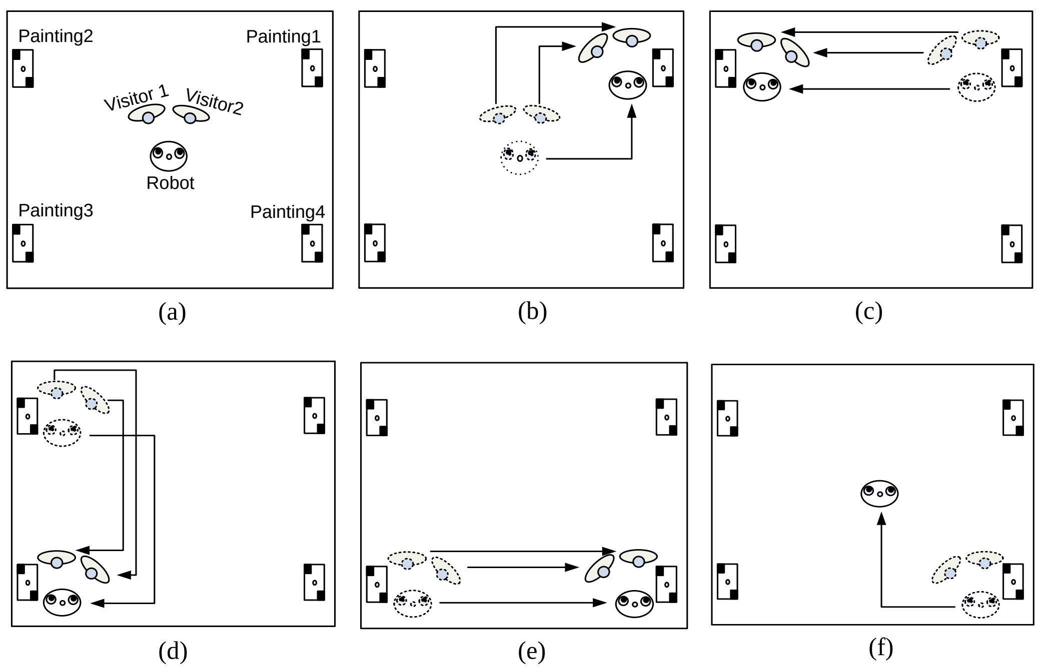


Figure 3.16: Schematic diagram of main tasks for experiment 2

1) When the robot finds visitors coming into its immediate vicinity, it says, "May I explain these paintings to you?" If the visitors' gaze

turns toward the robot's direction for three seconds, the robot system considers the visitors to be highly interested in the exhibits (Figure 3.16(a)).

2) The robot then guides the visitors to the first exhibit (Figure 3.16 (b)).
3) After arriving at the predefined position near the first exhibit, the robot follows the following steps to establish a proper spatial formation, in terms of body and face information (described in section 3.2.1).
   i) First, the robot verifies the distance between itself and the visitors using data provided by the laser range sensors.
   ii) If the visitors are not within range, the robot turns its head towards the visitors and says to them, "Please come closer" or "Please move back a little" depending on the situation.
   iii) Next, the robot turns 30^0 clockwise to orient towards the first exhibit.
   iv) Then, the robot verifies the body orientation of the visitors using data provided by the laser range sensors.
   v) If the visitors' body orientation is not in the direction between the robot and the exhibit, the robot turns its head towards the visitors and starts its explanation using "pause and restart" (described in section 3.2.2).
   vi) Next, the robot verifies the direction of the visitors' faces using data provided by the omni-directional camera.
   vii) If the face direction is towards the robot or the exhibit, the robot begins to explain the first exhibit. If not, the robot starts its explanation with a "pause and restart" (described in section 3.2.2).

4) After completing its explanation, the robot moves to the next exhibit and at the same time invites the visitors to follow along (Figure 3.16(c), Figure 3.16(d), Figure 3.16(e)).

5) The robot repeats task (3) to explain the next exhibit.

Finally after explaining all four exhibits, the robot returns to its initial position and waits for more visitors to arrive (Figure 3.16(f)). Figure 3.17 shows the behavioral protocol of the guide robot.

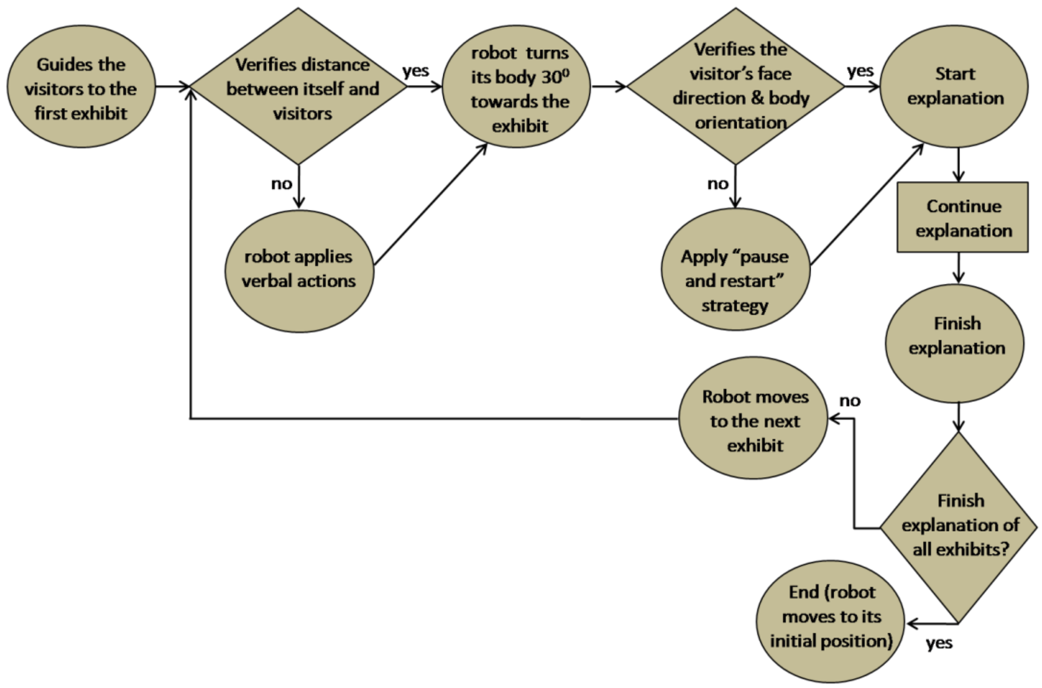


Figure 3.17: Behavioral protocol of guide robot.

We videotaped all sessions. In addition, we recorded all laser range finder and camera data so that we could obtain the exact motions of the robot and the participants for later analysis. Figure 3.18 shows the details of the experiment using the proposed method. In Figure 3.18(a), the guide robot finds two visitors who are interested in the paintings and says, "May I explain these paintings to you?"Having confirmed their interest, the robot moves towards the first painting while saying, "Please follow me," and the visitors do so (Figure 3.18(b)). In Figure 3.18(c), the robot and the visitors arrive at the first painting. After verifying all the constraints to create a proper spatial formation, the robot begins its explanation. The red circle in Figure 3.18(c) indicates that a proper spatial formation has been formed. After explaining the first painting, the robot moves to the second exhibit, and performed the same activities as it did while explaining of first painting Figure 3.18(d). The guide robot along with the visitors then moves to the third and fourth paintings respectively and explained those to the visitors (Figure 3.18(e) and Figure 3.18(f)). Finally, it returns to its initial position upon completing its explanation.

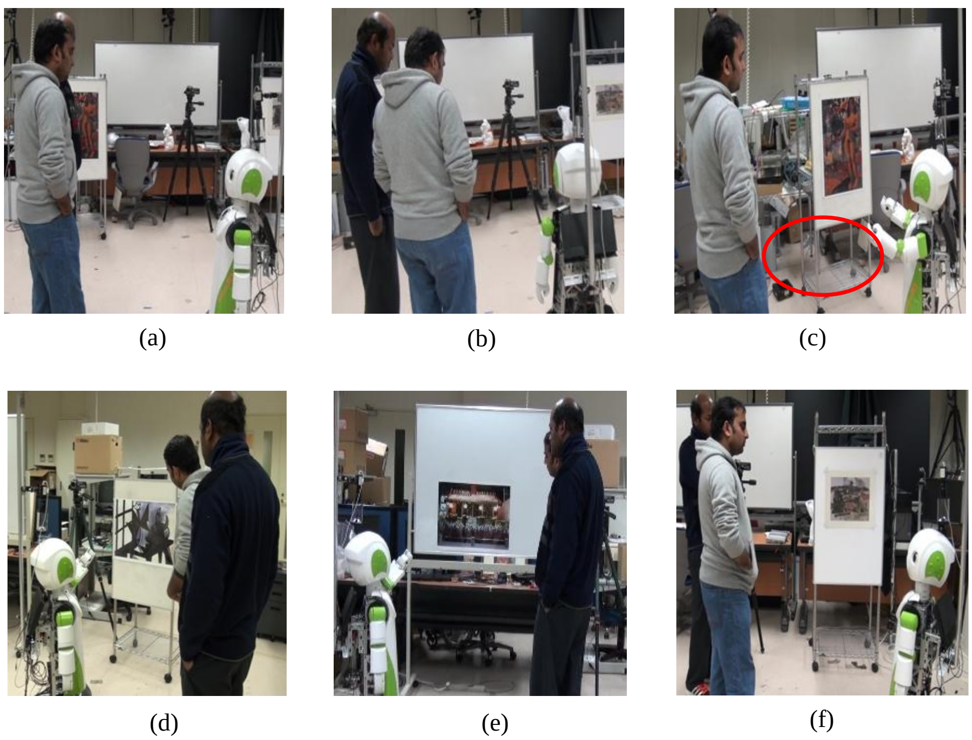


Figure 3.18: Example scenes from the second experiment.

3.3.2.3 Experimental Condition

In the experiment, the robot based on our proposed model was compared with a robot that did not employ the proposed model.

a) *Proposed Robot:* The robot behaves based on the model outlined in this paper.

b) *Conventional Robot:* The robot begins its explanation after finding the faces of visitors. It does not care whether or not a proper spatial formation is formed, nor does it utilize the "pause and restart" strategy. It explains the exhibits with the same preprogrammed nonverbal behaviors as the proposed robot.

After the experiments, we asked participants to fill out a questionnaire. They were asked to subjectively rate the robot's effectiveness on a seven-point Likert scale, with the range: 1-very ineffective, 2-ineffective, 3-somewhat ineffective, 4-undecided, 5-somewhat effective, 6-effective, 7-very effective. The questionnaire items were as follows:

1) Did you think that the robot attended to you adequately during its explanation?
2) Did you think that the robot was able to attract your attention to listen to its explanation?
3) Overall evaluation of the robot.

3.3.2.4 Hypothesis and Prediction

According to the discussion, we expect that it is very important for the mobile museum guide robot to be aware of the development of spatial formation which is estimated from spatial arrangement. At the same time, it is also important for the mobile guide robot to be conscious to employ "pause and restart" depending on the situation in order to achieve mutual gaze. Thus, our hypothesis states that, if we become successful in implementing our concept, the proposed method would produce one of the most suitable interactions. Based on this, we predicted that the proposed model does better than the other method that we have mentioned in this chapter.

3.4 Experimental Results

After the two experiments, we examined the results from the following three viewpoints:

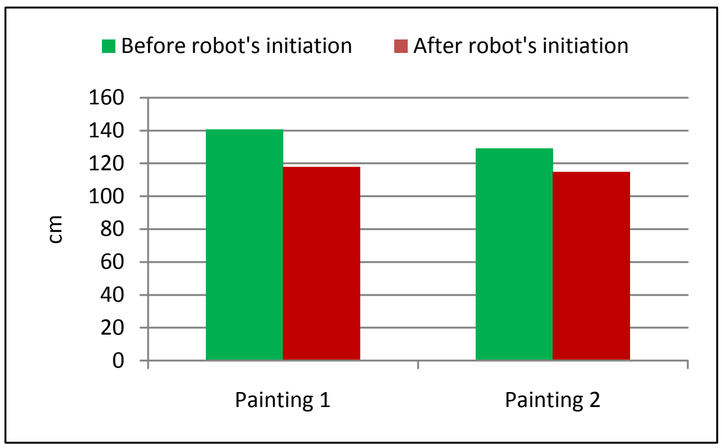
1) Autonomous capability of the robot: Can the robot correctly judge the situation and behave properly according to our proposed model? (Sections 3.4.1, and 3.4.2)
2) Effectiveness of the robot's actions: Can the robot's actions make the participants form a proper spatial formation and attract their attention at the beginning of explanation? (Sections 3.4.1, and 3.4.2)
3) Subjective evaluation: Do the participants prefer the proposed robot? (Section 3.4.3)

3.4.1 Control of Visitor's Standing Position

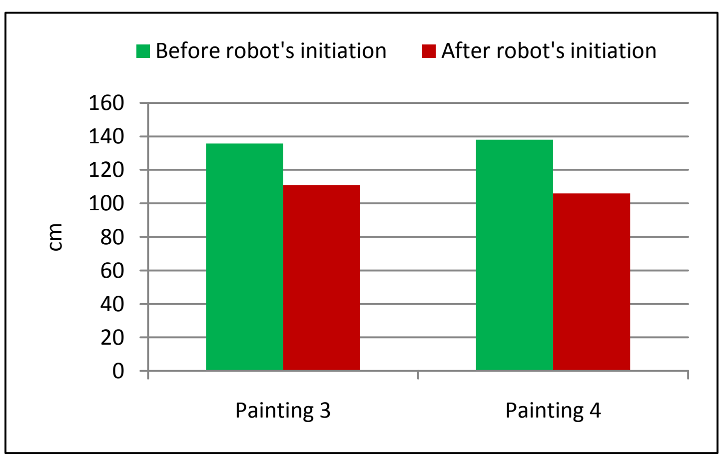
We recorded all sensor data from all sessions in the second experiment for analysis. As covered in section 3.3.2, the robot explained the first two paintings as the proposed robot and the remaining two as a conventional robot to Group A (groups no. 1, 3, 5, and 7) and in the reverse order to Group B (groups no. 2, 4, 6, and 8). In handling Group A, the robot took the initiative to control visitors' standing position (i.e., the robot said to the visitors, "Please come closer" or "Please move back a little" depending on the situation) in 3 cases out of 4 before starting its explanation of painting 1, and in 2 cases out of 4 before starting its explanation of painting 2. In dealing with Group B, the robot did so in 2 cases out of 4 before starting its explanation of painting 3, and in all 4 cases before starting its explanation of painting 4.We examined whether or not the robot's decisions were correct by analyzing the laser range finder data recorded during the experiment. We found that participants were in the proper area in 3 cases (Group A) and 2 cases (Group B), where the robot did not need to control their standing position. In cases where the robot needed to control visitors' standing position (5 cases for Group A, 6 cases for Group B), the robot's decisions were uniformly correct. Thus, the robot's decisions pertaining to standing position assessment ranked a success rate of 100%.

We then examined whether or not the participants moved to form a proper spatial formation after the robot's action. In the 5 cases for Group A, out of 10 participants 6 were out of proper spatial formation range, and after the robot's action 5 of these 6 moved inside the range. In the 6 cases for Group B, out of 12 participants 9 were out of proper range, and after the robot's action 6 of these 9 moved inside the range. The total success rate was therefore 73% (the robot corrected 11 out of 15). Figure 3.19 shows the average distance between the guide robot and the visitors before and after taking initiative by guide robot while explaining paintings as a proposed robot. For painting 1 and 2 (Figure 3.19(a)), the results show that the average distances between the robot and the visitors were respectively 140.6 cm and 129 cm which are out of range (the proper range to form a spatial formation is 90 to 120 cm). After the robot's initiative, the average distances between them are 117.5 cm and 114.8 cm respectively. For painting 3 and 4 (Figure 3.19(b)), after robot's initiative, the average distances are 111 cm and 106 cm respectively which were 135.8 and 138 cm respectively before robot's initiative. All participants of both groups in all 16 cases (robot explained first two paintings to Group A and rest two paintings to Group B as proposed robot, so 4x2x2=16 cases) were within robot's FOV.

On the other hand, when the robot explained the paintings as a conventional robot would, we found from the sensor data that 7 out of 16 participants in Group A (each group experienced an explanation of 2 paintings, so 8x2=16 participants) were out of range while the robot explained paintings 3 and 4, and 10 out of 16 participants in Group B were out of the range while the robot explained paintings 1 and 2. Of the total of 17 participants who were out of range (7 in Group A, 10 in Group B), only 6 moved inside the range to form a proper spatial formation when the robot began its explanation. The "success rate" for the conventional robot was thus 35% (6 out of 17, although here we only use the term "success rate" for convenience since the robot did not try to do correct the visitors). Table 3.1 and Table 3.2 summarize the results of guide robot's capability to control visitor's standing position.



(a)Painting 1 and painting 2



(a)Painting 3 and painting 4

Figure 3.19: Average distance between robot and visitors.

Table 3.1: Success rate of robot's autonomous capability to control visitor's standing position.

TR	NC/TNC	Robot took initiative to control visitor's standing position	Success rate of robot's decision (%)
Proposed	11/16	11	100%

Table 3.2: Success rate of effectiveness of robot's action to control visitor's standing position.

TR	NC/TNC	POR/NP	PMIR/POR	TSR
Proposed	11/16	15/22	11/15	73%
Conventional	13/16	17/26	6/17	35%

TR: Type of Robot
NC: Number of Cases
TNC: Total Number of Cases
POR: Participants Out of Range before Robot's Action
NP: Number of Participants
PMIR: Participants Moved Inside the Range after Robot's action
TSR: Total Success Rate

Figure 3.20 shows a typical example of changing a visitor's standing position. In Figure 3.20(a), the robot noticed that the distance between both visitors and the robot was more than 120 cm. It then turned its head towards them (blue arrow) and said "Please come closer" (Figure 3.20(b)). This caused visitors to come closer to the robot (Figure 3.20(c)). Red circle in Figure 3.20(c) indicates that, proper spatial formation is created. In the current system, the robot tries only once to establish a proper spatial formation. Since the robot can recognize the current formation, it can repeat the action until a proper spatial formation is established.

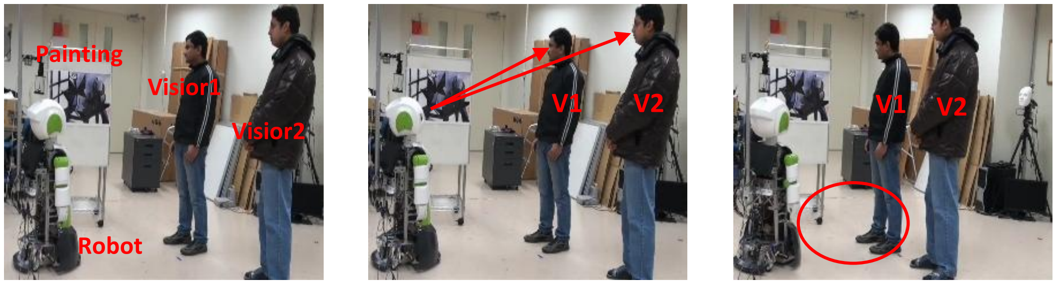


Figure 3.20: Example of participants changing his standing position after the robot's verbal action.

3.4.2 Control of Visitor's Body Orientation

From the recorded sensor data of the second experiment, we counted the number of visitors who successfully changed their body orientation to a direction towards (meaning, orientated to the space between) the robot and the painting after the robot employed a "pause and restart" at the beginning of its explanation. As mentioned earlier, the robot performed as the proposed model in explaining the first two paintings to Group A (groups no. 1, 3, 5, and 7) and the last two paintings to Group B (groups no. 2, 4, 6, and 8). In 2 out of 8 cases in Group A (each group experienced an explanation of two paintings, so 2x4=8) and 3 out of 8 cases in Group B, the robot noticed that some visitors' body orientations were not in the direction between it and the painting. The robot employed "pause and restart" in all 5 cases, meaning its success rate at deploying the strategy was 100%.

We then examined whether or not the participants successfully changed their body orientation in the direction between the robot and the painting after the robot's actions. In 2 cases among those of Group A, the robot noticed 2 out of the 4 participants' orientations were not in the direction between the robot and the painting, while in 3 cases among those of Group B it noticed 3 out of the 6 participants were incorrectly orientated. After the robot's actions, both of the incorrectly orientated participants in Group A and 2 of the 3 in Group B changed their body orientations appropriately. The total success rate was therefore 80% (the robot corrected 4 out of 5).

On the other hand, when the robot explained the paintings as a conventional robot would, we found from the recorded sensor data that out of 16 participants in Group A (each group experienced an explanation of two paintings, so 8x2=16 participants), 3 were not orientated towards the robot and the painting, while out of the 16 participants in Group B 1 was not correctly orientated. Out of a total of 4 incorrectly orientated participants, 2 changed their body orientations towards the robot and the painting just after the robot began its explanation. The conventional robot thus scored a "success rate" of 50% (2 out of 4, although here we only use the term "success rate" for convenience since the robot did not try to do correct the visitors). Table 3.3 and Table 3.4 summarize the results of guide robot's capability to control the visitor's body orientation.

Table 3.3: Success rate of robot's autonomous capability to control visitor's body orientation

TR	NC/TNC	Robot employed "pause and restart" to control visitor's body orientation	Success rate of robot's decision (%)
Proposed	5/16	5	100%

Table 3.4: Success rate of effectiveness of the robot's action to control visitor's body orientation

TR	NC/TNC	POI/NP	POC/POI	TSR
Proposed	5/16	5/10	4/5	80%
Conventional	3/16	4/6	2/4	50%

TR: Type of Robot
NC: Number of Cases
TNC: Total Number of Cases
POI: Participant's Oriented Incorrectly before robot's action
NP: Number of Participants
POC: Participant's Oriented Correctly after robot's action
TSR: Total Success Rate

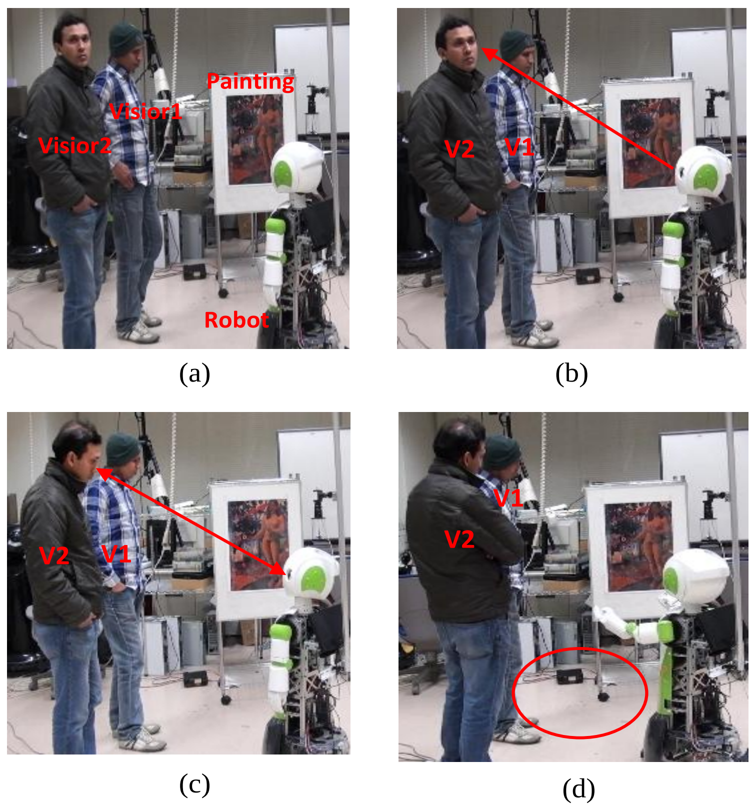


Figure 3.21: Example of participant changing his body orientation after the robot's use of "pause and restart".

Figure 3.21 shows a typical example of a visitor changing his body orientation in the direction between the robot and the painting after the robot's use of "pause and restart." In Figure 3.21(a), the robot noticed that the visitor's (V2) body orientation was not in the direction between the robot and the painting. It then turned its head towards V2 and employed "pause and restart" to attract the visitor's attention (Figure 3.21(b)). This caused a change in the visitor's body orientation; V2 turned his body towards the robot and mutual gaze was established (red double arrow in Figure 3.21(c)). V2 fully turned his body towards the robot (Figure 3.21(d)) and the robot continued its explanation. Red circle in Figure 3.21(d) indicates that a proper spatial formation is formed.

3.4.3 Control of Visitor's Face Direction

Here we examined those visitors whose body orientations towards the robot and the painting but whose faces were directed elsewhere during second experiment. We counted the number of such visitors from the recorded sensor data. When performing as the proposed model, the robot noticed 2 cases out of 8 in Group A (again, each group experienced a 2-painting explanation, so 4x2=8 cases) and 4 cases out of 8 in Group B where some visitors' face directions were not towards the robot or the painting. The robot employed "pause and restart" in all 6 cases, so its deployment success rate was100%.

We then examined whether or not the participants successfully shifted their attention towards the robot or the painting after the robot's actions. In the 2 cases in Group A, the robot noticed 3 out of 4 participants' face directions were not towards itself or the painting, while in the 4 cases in Group B it noticed 4 out of the 8 participants' faces were directed elsewhere. After the robot's actions, 2 of the 3 participants with faces directed elsewhere in Group A and 3 out of the 4 in Group B changed their face directions towards the robot or the painting. The total success rate was therefore 71% (the robot corrected 5 of the 7 participants).

Table 3.5: Success rate of robot's autonomous capability to control visitor's face direction

TR	NC/TNC	Robot employed "pause and restart" to control visitor's face direction	Success rate of robot's decision (%)
Proposed	6/16	6	100%

Table 3.6: Success rate of effectiveness of the robot's action to control visitor's face direction

TR	NC/TNC	PFDI/NP	PFDC/PFDI	TSR
Proposed	6/16	7/12	5/7	71%
Conventional	4/16	5/8	1/5	20%

TR: Type of Robot

NC: Number of Cases

TNC: Total Number of Cases

PFDI: Participant's Face Directed Incorrectly before Robot's Action

NP: Number of Participants

PFDC: Participant's Face Directed Correctly after Robot's action

On the other hand, when the robot explained the paintings as a conventional robot would, we found from the recorded sensor data that 2 out of 16 participants in Group A (again, each group experienced an explanation of 2 paintings, so 8x2=16 participants) had faces not directed towards the robot or the painting, while 3 out of 16 participants in Group B had faces directed elsewhere. Out of the total of 5 whose face directions were not towards the robot or the paintings, only 1 shifted their face appropriately just after the robot started its explanation, for a "success rate" of 20% (1 out of 5, although here we only use the term "success rate" for convenience since the robot did not try to do correct the visitors). Table 3.5 and Table 3.6 the results of guide robot's capability to control the visitor's face direction.

Although the number of cases is small and we cannot yet draw any definite conclusions, the results suggest that the "pause and restart" strategy can have a considerable effect. Since the robot is able to recognize the visitors' body orientations and face directions, it is possible to modify the robot so that it may repeat the same action or take alternative actions as necessary to attract the attention of visitors.

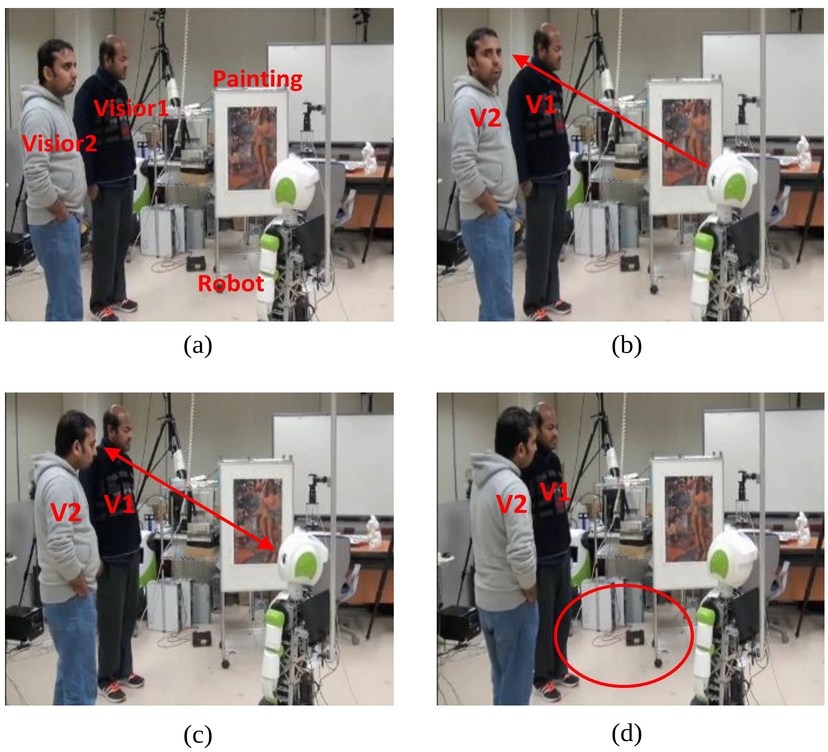


Figure 3.22: Example of participant changing his face direction after the robot's use of "pause and restart".

Figure 3.22 shows a typical example of a visitor changing his face direction towards the guide robot after the robot's use of "pause and restart." In Figure 3.22(a), the robot noticed that the visitor's (V2) face direction was not in the direction between the robot and the painting. It then turned its head towards V2 and employed “pause and restart” to attract the visitor’s attention (Figure 3.22(b)). This caused a change in the visitor’s face direction; V2 turned his face towards the robot and mutual gaze was established (red double arrow in Figure 3.22(c)). The guide robot continued its explanation (Figure 3.22(d)). Red circle in Figure 3.22(d) indicates that a proper spatial formation is formed.

3.4.4 Subjective Evaluation

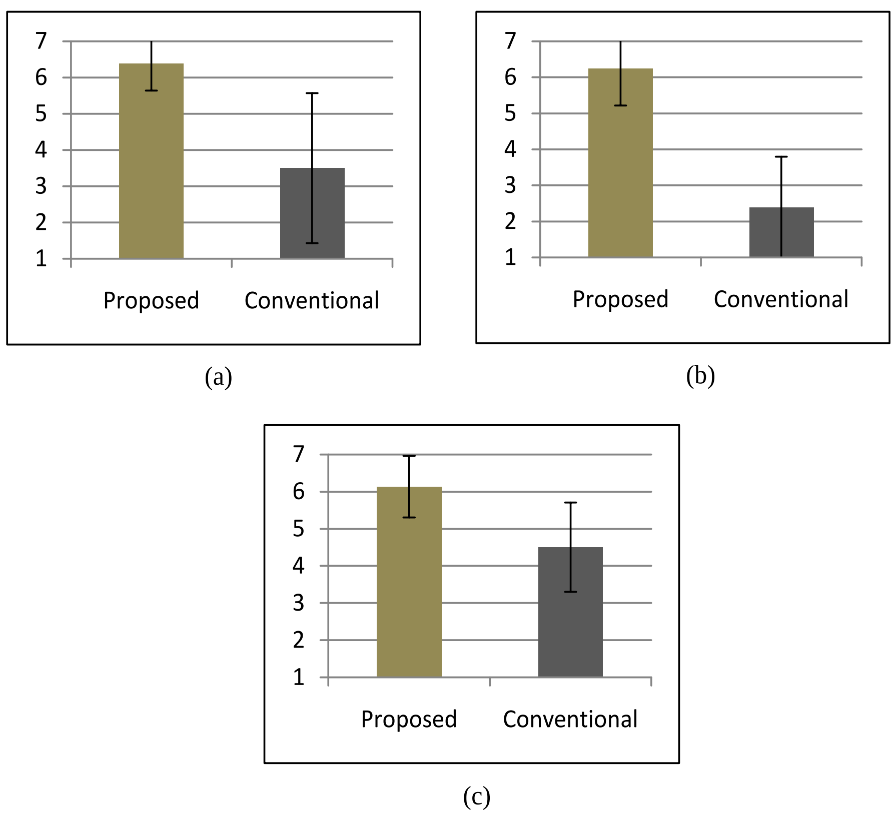


Figure 3.23: Result of subjective evaluation of first experiment.

The subjective evaluation of the first experiment is as follows. The data were analyzed by paired t-test. Figure 3.23(a) shows the results of the effectiveness of the F-formation. The t-test shows a statistically significant difference between the two conditions *(t(14) = 3.69, p=0.0024)*. Figure 3.23(b) shows the results of the effectiveness of the “pause and restart” strategy. The t-test shows a statistically significant difference between the two conditions *(t(14) = 6.27, p=0.0001)*. Figure 3.23(c) shows the result of the overall evaluation about the robot. The t-test shows a statistically significant difference between the two conditions *(t(14) = 3.15, p=0.0070)*.

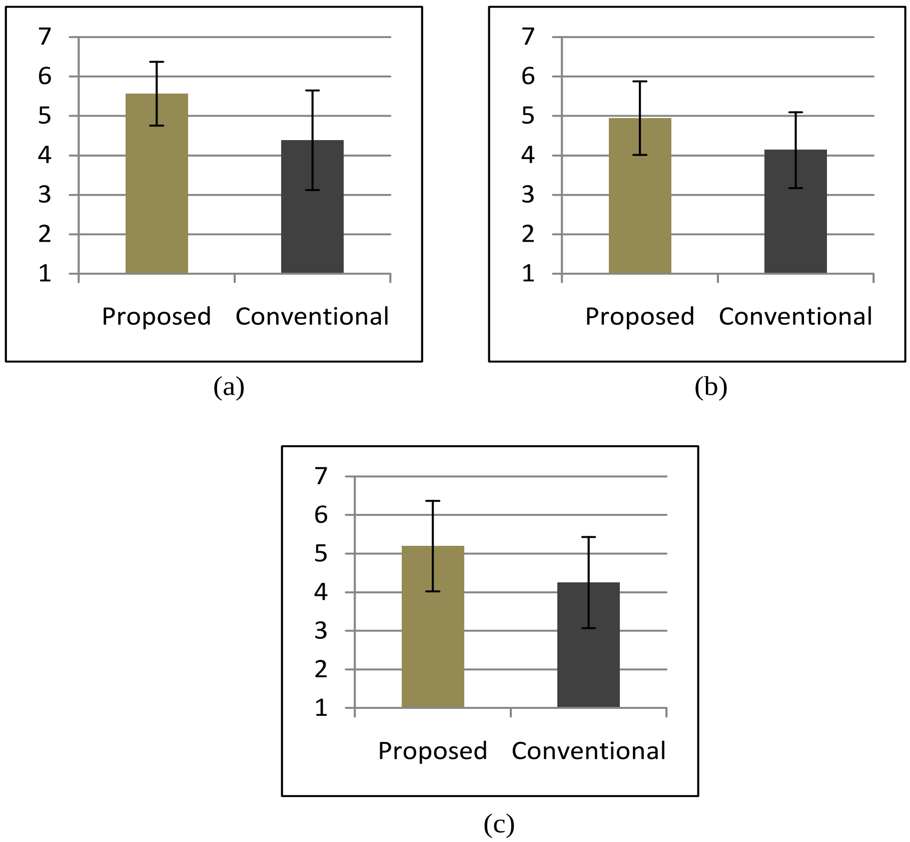


Figure 3.24: Results of subjective evaluation of second experiment.

We also conducted a subjective evaluation of the second experiment among the participants. For the second experiment, the data were analyzed by one-way analysis of variance (ANOVA) with the experimental conditions as the within subject factor (or repeated measures). In the experiment, two types of robots (the proposed robot and the conventional robot) were investigated. “Type of Robot” is the factor or independent variable in this experiment. Since two types of robots were compared, the factor “Type of Robot” has two levels. We adopted a significance level (p-value) of 0.05 or less to determine statistically significant differences among the conditions. The results of the second experiment are as follows.

For the question “Do you think that the robot attended to you adequately during its explanation?” (Figure 3.24(a)), repeated measures of ANOVA revealed a significant difference between the two conditions ($F(1,15)=15.08$, $p = 0.0014$). The same was true of the question “Did you think that the robot was able to attract your attention to listen to its explanation?” (Figure 3.24 (b)), where repeated measures of ANOVA also showed a significant difference ($F(1,15)=8.59$, $p=0.0103$). Finally, the participants’ overall evaluation (Figure 3.24(c)), here too repeated measures of ANOVA showed a significant difference between the two conditions ($F(1,15)=6.03$, $p=0.0266$).

These results suggest the participants’ preference for our proposed robot, and moreover demonstrate the significance and utility of guide robots being able to establish a proper spatial formation and to employ the “pause and restart” strategy.

3.5 Chapter Summary

A museum guide robot needs to have people group around the exhibit that it will explain; it must also attract their attention when it starts explaining the exhibit. If people are away from the exhibit, it cannot begin its explanation. Likewise, even if they are near the exhibit, if they are not looking at the exhibit or the robot they may not be ready to listen to the robot’s talk. Clearly, sociological studies such as Kendon’s on the spatial formation and Goodwin’s on the “pause and restart” strategy can offer helpful suggestions as to how a guide robot should behave when starting its explanation. Through observations at a museum we confirmed that human guides effectively form a proper spatial formation and use “pause and restart” to attract visitors’ attention when necessary and thus create a proper spatial formation at the beginning of explanation. We then proceeded to develop a robot based on these findings.

The experimental results shown above indicate that museum guide robot that performed according to our proposed model to create spatial formation at the beginning of explanation, can effectively form a proper spatial formation, and moreover that the robot’s use of “pause and restart” can effectively attract visitors’ attention.

3.5.1 Limitations

In this chapter, we have considered the case where the robot can create spatial formation before start its explanation. There are many other situations where the guide robot needs to reconfigure spatial formation with the visitors. According to Kendon, an existing spatial formation may gain or lose participants, undergoing dynamic reconfiguration as it does so [13]. The next chapter focuses how the guide robot can identify and invite any bystanders into an ongoing explanation.

Chapter 4

Reconfiguration of Spatial Formation While Interested Bystanders Join into Ongoing Explanation

A spatial formation may gain participants. Goodwin and Goodwin [146] examine participation frameworks that include not only speaker and hearer but also a potential hearer such as a bystander. By participation framework, Goodwin and Goodwin make clear how participants change their orientation and engagement. Indeed, a complete turnover of participants may occur in some instance without any discontinuity in the spatial formation system itself. Reconfiguration of a spatial formation depends on how both the current members and outsiders cooperate in maintaining the integrity of the system's boundary. A new participant or outsider only becomes a member of an existing spatial formation system through cooperation action between him/herself and the members of the existing spatial formation system. A member of, or a participant in, a spatial formation is anyone whose transactional segment (described in chapter 2) falls over the o-space without intersecting the body of any other individual and who takes part in the adjustments of spacing and orientation by which the space is sustained.

4.1 Who is Bystander?

A bystander is anyone who observes a situation but may or may not know what to do. In other words, bystanders are those whose presence the participants acknowledge and who observe the conversation without being participants in it. In a museum context, when a guide explains any exhibit to a small numbers of visitors, visitors normally stand within the range of personal space and many other visitors who are not participating in the current explanation may stand around the robot within the social or public space. Among these, those who demonstrate interest in the explanation are considered to be interested bystanders. Figure 4.1 shows the different forms of conversational participants. The body orientation and face direction play an important role in establishing and maintaining conversational participant roles. In Figure 4.1 bystander observes the conversation between the speaker and listener without being participants in it by turning his/her body and face towards them. Here, speaker and listen all together formed a

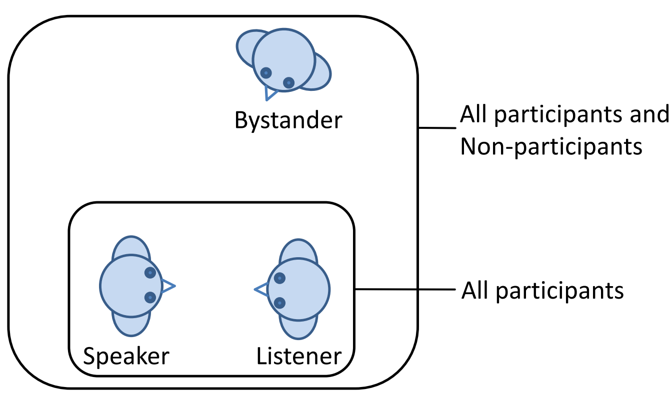


Figure 4.1: Different forms of conversational participants.

spatial formation. To be a member of current spatial formation, bystander needs to wait until the invitation by any current member.

4.2 Model of Incorporating Interested Bystanders

During explanation of any exhibit to a small group of visitors, guide robots need to be careful about the others visitors who are around itself. If the guide robot identifies any visitors as interested bystanders, it may stop its explanation for a while and invite the interested bystanders to join in the ongoing explanation session. After the bystanders join, the guide robot reconfigures the spatial formation and starts its explanation again. A spatial formation may be said to have begun as soon as the spatial-orientational cooperation which sustains an o-space can be observed. This is usually easy to determine, for when a new spatial formation is established the participants move towards one another in a decided fashion and they also move to a location that is new for all of them. In order to identify interested bystanders for interaction, the guide robot should have the capability to carry out the following three steps:

4.2.1 The Robot should Identify the Visitors Around itself

As noted by Pacchierotti et al., interactions take place within distances up to 350 cm [147]. The area outside that zone mainly hosts one-way communication. In our approach, the robot is able to identify visitors, if they are found:

a) Within social space (120 cm to 350 cm) and
b) Within robot's FOV (maximum 150^0).

4.2.2 The Robot should Assess Visitors' Intentions

After locating visitors within social distance and its own FOV, the robot needs to identify those visitors who are interested in its ongoing explanation. The following two features are helpful in identifying an interested visitor:

a) Visitor's body orientation: gives an indication of where the visitor might be looking.

b) Visitor's face direction: crucial for the robot to estimate whether or not a visitor is interested.

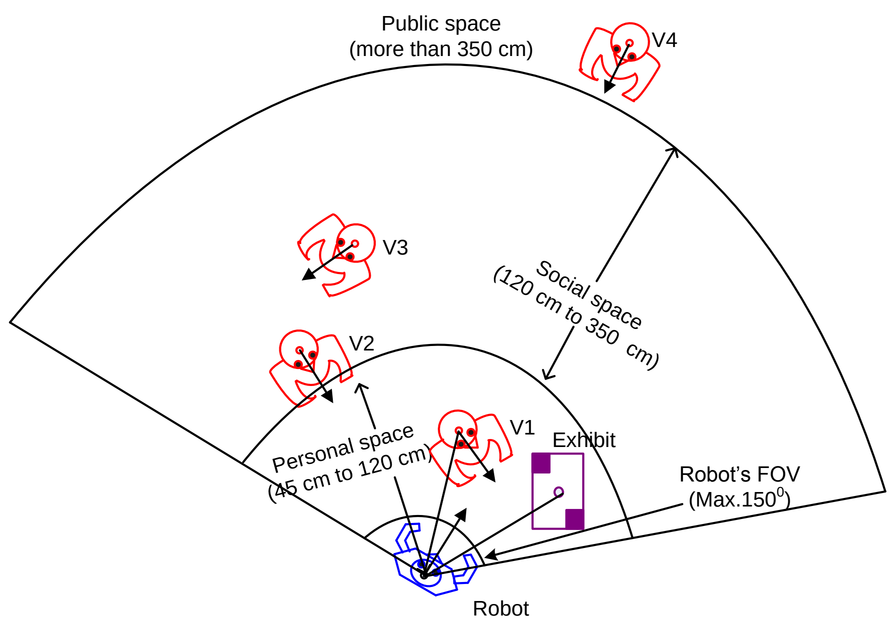


Figure 4.2: Selection of interested bystanders

Therefore, if the robot finds any visitor displaying interest in the explanation, determined by their standing for three seconds or more within social distance and the robot's own FOV while changing their body orientation and face direction towards the robot or the painting, it identifies that visitor as an interested bystander. Figure 4.2 shows which visitors should be selected as an interested bystander. In Figure 4.2, the robot explains an exhibit to V1. During the explanation, the robot observes V2 and V3 within both social space and its FOV, and V4 in public space. Among the three visitors, only V2 is considered as an interested bystander because s/he is found within social distance and FOV and his/her body orientation and face direction are towards the robot.

4.2.3 Approaching Interested Bystanders Appropriately

Once the robot has made its assessment, it needs to approach the interested bystander. The following actions are performed by the robot after identifying an interested bystander:

i) The robot stops its current explanation for a moment and turns its head towards the interested bystander. This serves as the robot's acknowledgement of the bystander's presence.

ii) The robot then turns its body towards the interested bystander. After shifting its body orientation, the robot needs to give the impression that it is establishing eye contact. With no information available on the target visitor's height, the robot assumes the person to be 160 cm tall, and tilts its head accordingly. We assume that this action by the robot can make him/her feel eye contact.

iii) The robot next invites the interested bystander to join in the ongoing explanation by saying "Hi, if you are interested in this exhibit please join us" and thus initiates a verbal interaction with the interested bystander.

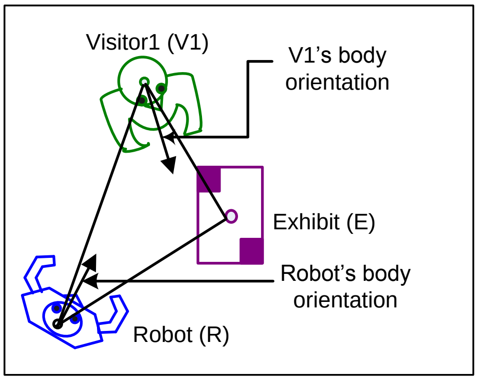
iv) Finally, the robot turns its body back away from the interested bystander to resume explanation.

Figure 4.3 shows a schematic diagram of how the robot invites interested bystanders to join into an ongoing explanation. The guide robot (R) was explaining an exhibit to visitor 1 (V1) in Figure 4.3 (a). During explanation, R noticed another visitor (V2) was observing its explanation by turning his/her body and face towards it (Figure 4.3 (b)). By observing his/her body orientation and face direction, R treated him/her as an interested bystander. R then turned its own body towards V2 and invited him/her to join into ongoing explanation session (Figure 4.3 (c).). Finally V2 moved towards the existing spatial formation and by overlapping his/her own transactional segment, s/he became the member of new spatial formation.

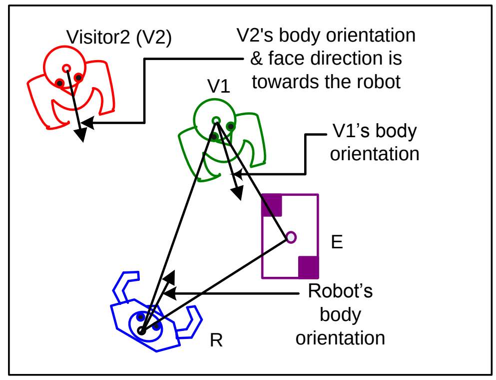
The main issue that is addressed in this section is the guide robot's detection of visitors around itself, its assessment of their intentions towards the robot, and its ability to act towards them in a natural way.

4.3 Evaluation Experiment

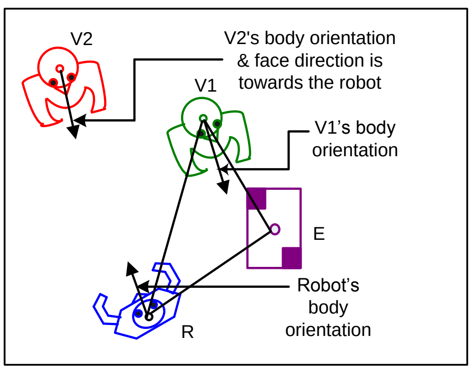
We performed an experiment in a laboratory to test the robot's effectiveness. We used the guide robotic system which is described in chapter three. The experiment was performed in order to test the robot's ability at indentifying and inviting interested bystander into ongoing explanation session. In the experiment, guide robot followed the proposed model described in section 4.2. Figure 4.4 shows the schematic diagram of the experimental setting. Six paintings (P1-P6) were placed in the experimental area, under the assumption that it was a small art museum. Among all paintings, P2 to P6 were hung on the wall at the same height. Only P1 was placed in the experimental area without hanging on the wall. Proposed guide robot explained only P1 to the visitors.



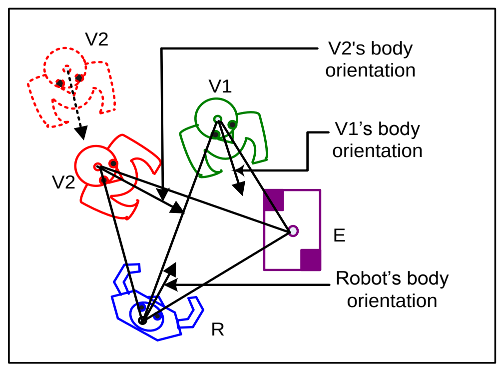
(a) R explains exhibit to V1



(b) R finds V2 as interested bystander



(c) R turns its body towards V2 and invites him to join



(d) R reconfigures spatial formation with V1 and V2

Figure 4.3: Incorporation of an interested bystander into existing spatial formation.

The experimental setup was designed as follows:

1) A total of 9 graduate students (average age is 28) at Saitama University participated in the second experiment. We formed 3 groups with 3 members in each group. Each group participated in the experiment just once.
2) At the start of the session, one participant from a group was told to enter the experimental area.
3) The remaining two participants were told to enter the experimental area after the robot had already commenced its explanation to the first participant.

In the experiment, the robot followed the proposed model outlined in section 3.2 in chapter 3 to configure a spatial formation. We videotaped all four sessions and recorded all laser range finder and camera data for later analysis.

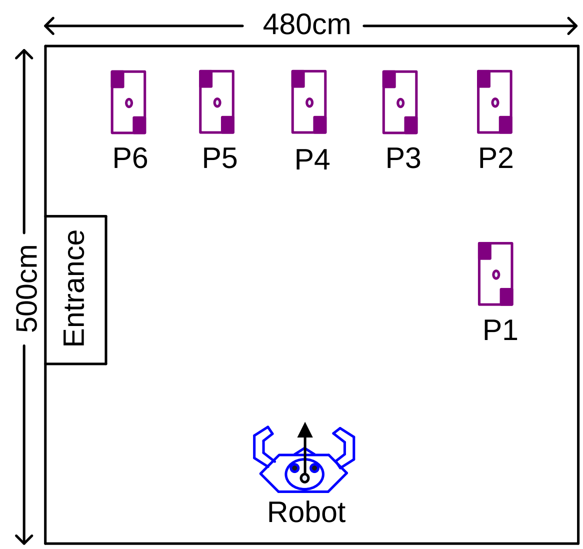


Figure 4.4: Schematic representation of the experimental setting.

4.4 Experimental Results

As mentioned in section 4.3, we formed 3 groups with 3 members in each group and each group participated in the experiment just once. At the start of each session, one participant (V1) was told to enter the experimental area. The remaining two participants (V2 and V3) were told to enter the experimental area after the robot had begun its explanation to the first participant. The robot took the initiative to invite the interested bystanders to join its explanation in all 3 cases (each group participated in the experiment once).

We examined whether or not these decisions were correct from the recorded sensor data. We found that 1 out of the 2 bystander participants (V2 and V3) in each case was in social space and within the robot's FOV, while his/her body orientation and face direction were towards the robot or the painting. A total of 3 out of 6 bystander participants in the experiment were in social space and within the robot's FOV with body orientation and face direction towards the robot or the painting. So, the success rate of the robot taking the initiative to invite interested bystanders ranked at 100%. After the robot's invitation, all 3 participants (interested bystander) from 3 groups moved towards the guide robot and placed their body orientation in between the robot and exhibit and thus existing spatial formation is reconfigured with two participants.

In the first and second cases, one of the bystander participants (V2 or V3) was in social space and within the robot's FOV but did not have a body orientation or face direction towards the robot or the painting; consequently the robot did not consider him to be interested bystanders and did not take any initiative to invite them into the explanation that it had begun. In the third case, one of the bystander participants (V2 or V3) was in social space and had a body orientation and face direction towards the robot, but the robot did not consider him an interested bystander nor correspondingly take any initiative to invite them because they were not within the robot's own FOV.

We then examined whether or not the participants moved into the current spatial formation after the robot's action. After the robot's action, all 3 of the participants identified by the robot as interested bystanders successfully moved to the current spatial formation. There was thus a success rate of 100%. In all three cases the robot proceeded to verify all the parameters mentioned in section 3.2 in chapter 3 to reconfigure a spatial formation with one more participant. In this implementation, the robot was programmed so that it could only identify and invite one bystander in a session. Of course, effective guide robots should have the capability to identify and invite more than one bystander at a time, and this is a topic we are pursuing.

Figure 4.5 shows example scenes from the experiment, revealing the details of the identification and invitation of an interested bystander by the guide robot. In Figure 4.5(a), after verifying all the parameters mentioned in section 3.2 in chapter 3 to configure an spatial formation, the robot begins its explanation of the painting "Te Nave Nave Fenua" to a visitor (V1). During the explanation session, the robot notices two more visitors (V2 and V3) enter in the experimental area (Figure 4.5(b)). Both of them are in social space and within the robot's FOV. Among these two, V2 shows interest in the explanation by turning his body towards the robot, facing it and waiting until the robot looks at him. V2 moves a little bit towards the robot (Figure 4.5(c)). In Figure 4.5(c), V2's transactional segment (a space extending in front of a person which is the space he/she is currently using) does not overlap with the current spatial formation between the robot, V1 and painting. When the robot finds that the body orientation and face direction of V2 are directed towards it, the robot stops its current explanation for a moment and turns its head towards the bystander, V2. This serves as the robot's acknowledgement of the interested bystander's presence. The robot then changes its body orientation towards V2 (Figure 4.5(d)), which V2 reciprocates with a spatial-orientational move of his own, and which allows V2 access to the current spatial formation. The yellow double arrow in Figure 4.5(d) indicates that mutual gaze is established. This action by the robot establishes a non-verbal interaction with V2. The robot then invites V2 to join in the ongoing explanation, and thus creates a verbal interaction with V2 (Figure 4.5(e)). After the robot's invitation, V2 moves to join the current spatial formation (Figure 4.5(f)). The robot then turns back to its previous position (Figure 4.5(g), here V2 contributes his transactional segment to current spatial formation).

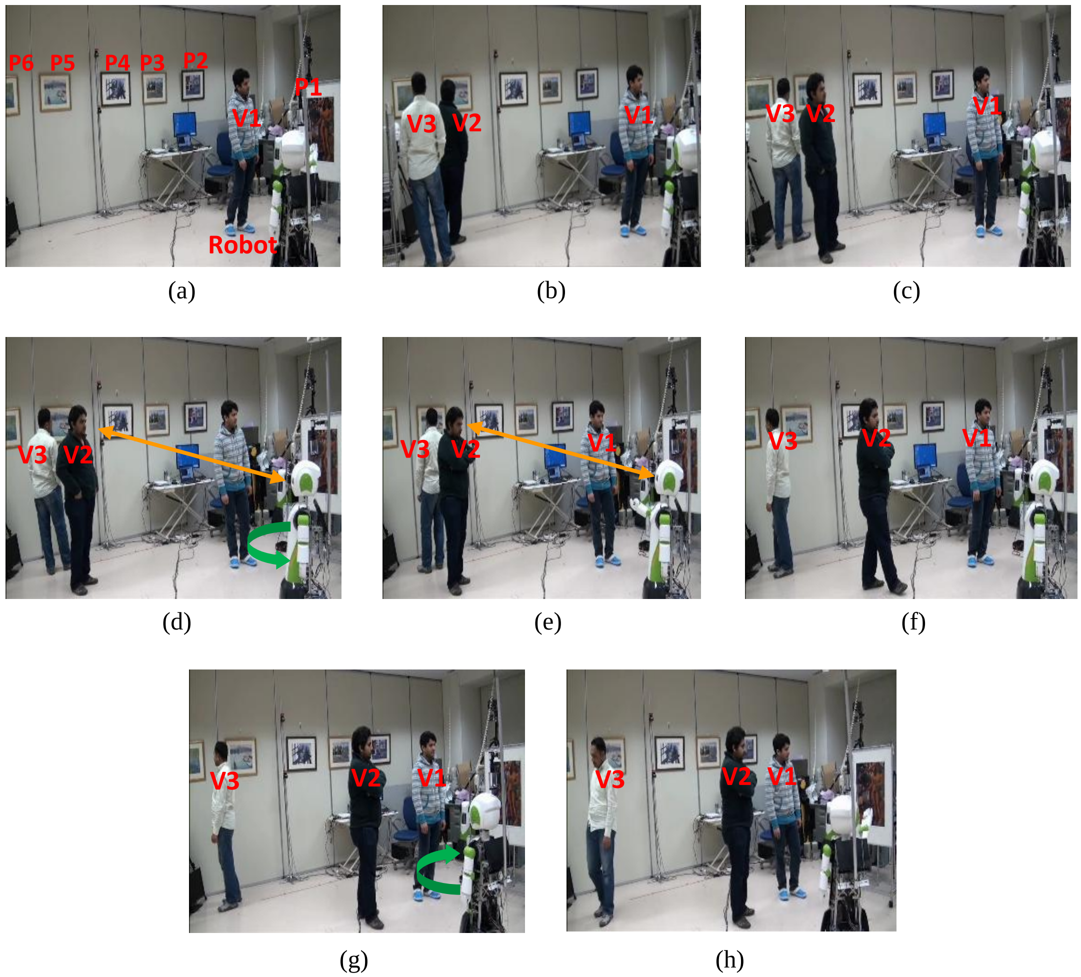


Figure 4.5: Identification and invitation of interested bystander.

The robot again verifies all the parameters mentioned in section 3.2 in chapter 3 to reconfigure the spatial formation, which now has two participants, and then continues its explanation (Figure 4.5(h)). In this scene, the robot did not consider V3 an interested bystander because his body orientation and face direction were not towards itself or the painting. So the guide robot neither turned its body towards V3 nor invited him to join into ongoing explanation session.

4.5 Chapter Summary

This chapter demonstrates how the spatial formation system of behavioral relationship that sustains an o-space may be continued despite change in participants. It also exemplifies how the arrangement of a spatial formation changes as it adapts to these variations in participants. The model of entry an interested bystander described are seen where the interested bystander approaches an existing spatial formation system on his/her own initiative. Where an interested bystander is called over by a current member (guide robot), the members of the current spatial formation system adjust their arrangement appropriately as the interested bystander approaches first.

We have conducted an experiment in our laboratory to evaluate our proposed framework to incorporate interested bystander into existing spatial formation. Our evaluation results showed that our model significantly improves the robot's performance in incorporating interested bystanders into ongoing spatial formation and thus reconfigure spatial formation again.

4.5.1 Limitations

Although our guide robot has achieved good performance in terms of creating spatial formation with the multiple visitors and identifying and inviting interested bystanders into existing spatial formation, there is a limitation of our guide robot. This chapter's and previous chapter's study deal with the case where the guide robot explains single exhibit to multiple visitors at a time. So, we should consider the case where the guide robot explains multiple exhibits collectively to multiple visitors. We also need to investigate how the guide robot reconfigures spatial formation while it changes its own body orientation from one exhibit to another during explanation of multiple exhibits collectively.

Chapter 5

Reconfiguration of Spatial Formation during Explanation of Multiple Exhibits to Multiple Visitors

To create and sustain a dynamic physical group during interaction with people is one of the essential features of museum guide robots that are physically co-located with people. To create a spatial formation there are some practical geometrical constraints. For example, people must stand facing one another in order to have full visual auditory access to each other's verbal and non-verbal conversational moves. In the case of a mobile museum guide robot, for visitors to observe both the robot's explanation and the target object without difficulties, the robot and visitors both should have equal, direct, and exclusive access to the space where the target object exists. We also have to consider the cases where a guide robot changes its body orientation from one exhibit to another during explaining of multiple exhibits collectively. This means that the spatial formation arrangement, once configured in front of an exhibit, should disappear when the guide robot rotates its own body towards another exhibit, and should reconfigure in front of the next exhibit. Therefore, the robot and the visitors must dynamically reconfigure the spatial formation arrangement each time they face at a target exhibit.

Therefore, in this study, we investigated how a guide robot can actively configure the appropriate spatial formation in the course of the explanation. We are especially interested about the affect of robot's body orientation. Our research questions are as follows:

1) Can the robot establish a proper spatial formation at the beginning of explanation?
2) Can the robot reconfigure the spatial formation again during the ongoing explanation session?

5.1 Modeling of Interaction with Visitors

In our previous study (described in chapter 3), we developed a position model for a mobile museum guide robot that establishes spatial formation before start its explanation. However, our previous model cannot deal with the case in which guide robot explains multiple exhibit collectively; more specifically guide robot and the visitors pay attention to multiple targets at a time. Thus, we updated our position model based on our observation. Our previous model consisted of six constraints for establishing spatial formation: (1) proximity to visitors, (2) proximity to exhibit, (3) visitor's field of view, (4) visitor's body orientation, (5) robot's body orientation, and (6) robot's field of view. To deal with the case where the robot needs to pay attention to multiple exhibits, we updated some of these constraints.

5.1.1 Model of Spatial Position

Our previous study (described in chapter 3) showed the importance of the constraint of proximity to the visitors and exhibit. This proposed modeling of spatial position also supports this constraint. Thus, we defined the constraint of proximity as follows.

5.1.1.1 Constraint of Proximity

- Distance between the robot and the visitors: Ranges from 90 cm to 120 cm (Figure 5.1(a)).
- Distance between the robot and the exhibits: About 110 cm, fixed in all cases for all exhibits (Figure 5.1(a)).

5.1.1.2 Constraint of Visitor's Face Direction

Our previous study (described in chapter 3) showed that the constraint of a visitor's Face direction is important for spatial formation. This proposed modeling of spatial position also supports this constraint. Thus, we defined the constraint of visitor's face direction as follows.

- Visitor's face direction: Should be towards the robot or the exhibit (Figure 5.1(b)).

5.1.1.3 Constraint of Body Orientation

- *Visitor's body orientation*: Each angle between each pair of vectors from the robot to each exhibit is calculated. The visitor's body orientation should be in the direction between the two vectors that make the larger angle among all combinations (Figure 5.1(c) and Figure 5.1(d)).

- *Robot's body orientation*: When the robot pays attention to one exhibit, its body orientation follows the vector from the robot to the exhibit. On the other hand, the robot's body orientation follows the median vector between the two vectors from the robot to each exhibit when it pays attention to two exhibits (Figure 5.1(e)).

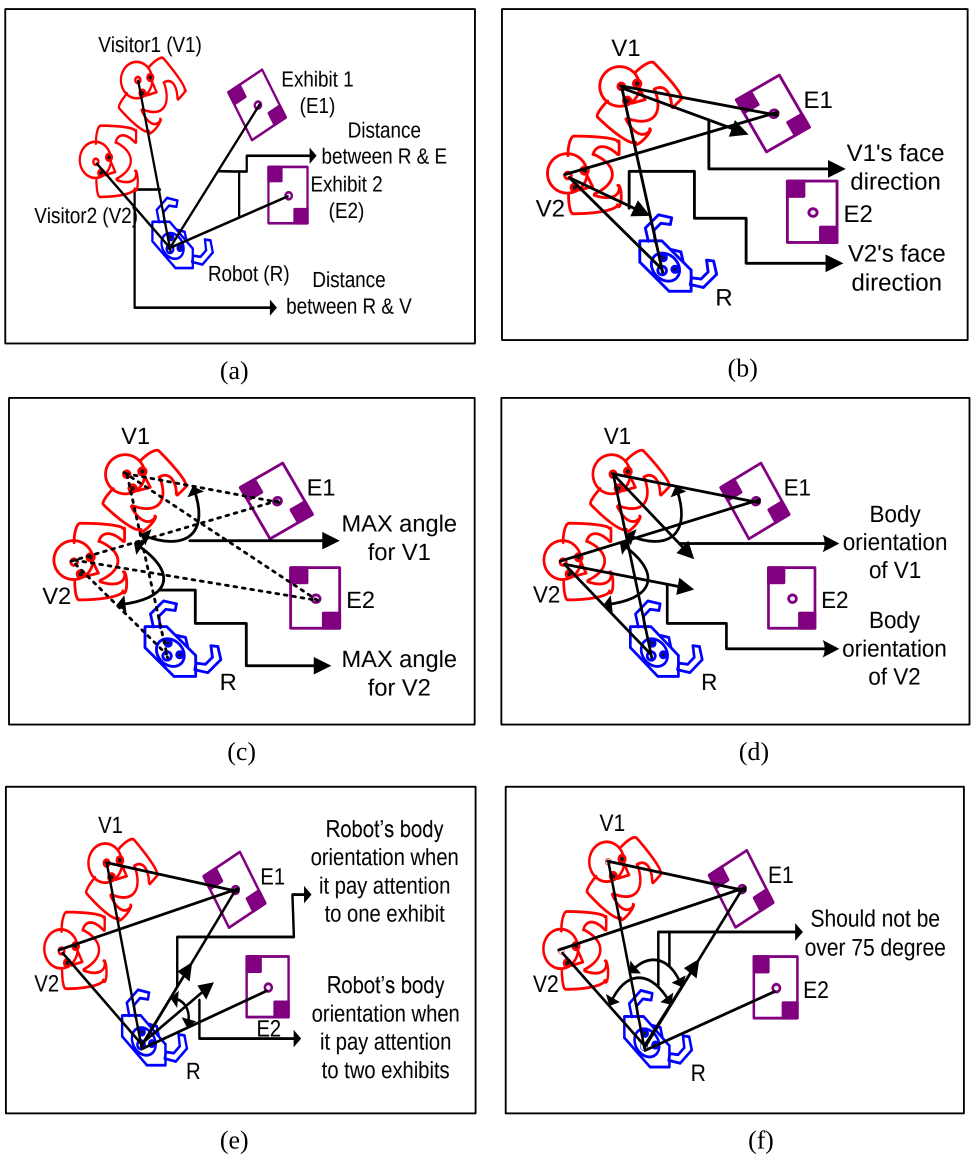


Figure 5.1: Spatial arrangement to explain multiple exhibits.

5.1.1.4 Constraint of Robot's Field of View

- All angles between the robot's body orientation and each vector from robot to each visitor and exhibit should not be over 75 degrees (Figure 5.1(f)).

5.1.2 Model to Attract Visitor's Attention

Our previous study (described in chapter 3) showed that, guide robot needs to attract visitor's attention at the beginning of explanation. Current study also supports this constraint. We enabled our mobile guide robot system to detect and monitor the visitor's face direction and to interpret this as an indicator of their attention. For our system, we have used the "pause and restart" procedure as a means to achieve mutual gaze. The robot employs the "pause and restart" if it finds that visitors are not gazing or visitor's body orientation are not towards it. The format of "pause and restart" which is implemented in our system is as follows:

[Beginning] + [pause] + [new Beginning]

In this format, the speaker (robot) starts the sentence (beginning), waits (pause) until the gaze of its recipient (visitor) is secured, and then restarts (new beginning) the sentence again.

5.2 Tasks Performed by Guide Robot

Initially, the robot with the proposed system waits in the middle of the experimental area. The order of these main tasks is as follows:

1) When the robot finds visitors coming into its immediate vicinity, it says, “May I explain the paintings to you?” If the visitors’ gaze turns toward the robot’s direction for 3 seconds, the robot system considers the visitors to be highly interested. The robot then guides them to the exhibits1 & 2.
2) After arriving at the predefined position near the first two exhibits, the robot follows the succeeding steps to establish a proper F-formation.
   a) First, the robot verifies the distance between itself and the visitors using data of laser range sensors. If the visitors are not within range, the robot says, “Please come closer” or “Please move back a little” depending on the situation.
   b) Next, the robot turns towards the first of the two exhibits which it is going to explain.
   c) Then, the robot verifies body orientation of the visitors. If the visitors’ body orientation is not in the direction between the robot and the exhibit, the robot starts its explanation using “pause and restart” (described in section 2.2).
   d) Next, the robot verifies the face directions of the visitors using data provided by the omni-directional camera. If the face direction is towards the robot or the exhibit, the robot begins to explain the exhibit 1. If not, the robot starts with a “pause and restart”.
   e) After completing explanation of exhibit 1, the robot turns towards exhibit 2 and explains it.
   f) After completing its explanation of exhibit 2, the robot moves back a little bit and turns towards the median vector between the two vectors from the robot to each of the two exhibits. The robot then gives a comparison explanation between two exhibits.

3) After completing its explanation of first two exhibits, the robot moves to explain next two exhibits and at the same time invites the visitors.
4) Robot repeats task 2 to explain the next exhibits.

5.3 Experiment with Multiple Exhibits

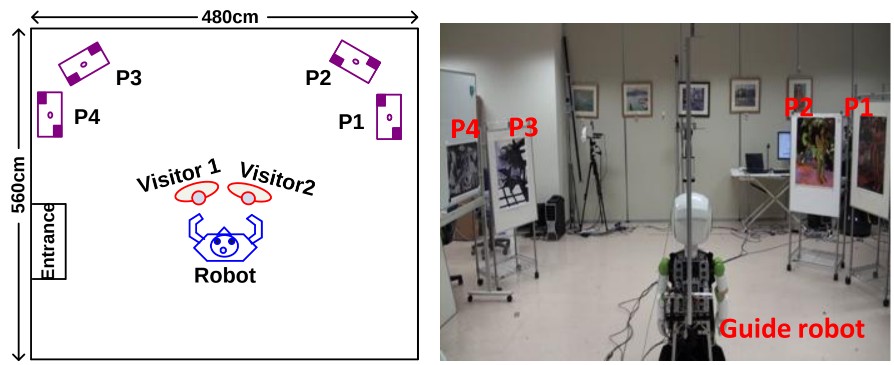
We conducted an experiment to verify that our proposed model based on explanation of multiple exhibits is effective to reconfigure spatial formation with the visitors. The guide robotic system which is described in chapter three is used for experimental purpose.

5.3.1 Experimental Design

We conducted an experiment to verify the effectiveness of our proposed robot. 4 paintings were placed in the area of size 480cmx500cm (Figure 5.2). In the experiment, the robot based on our proposed model was compared with a robot that did not employ the proposed model.

a) Proposed Robot: Robot behaves based on the model outlined in this paper.
b) Conventional Robot: Robot begins its explanation after finding the faces of visitors. It does not rotate its body from one exhibit to another during explanation, thus does not care about the configuration of spatial formation.

A total of 12 graduate students (6 groups with 2 members in each group) participated in the experiment. The experiment followed within-subject design, and the robot explained four exhibits to all groups. The robot explained the first two paintings as the proposed robot and the remaining two as a conventional robot to Group A (groups no. 1, 3, and 5) and in the reverse order to Group B (groups no. 2, 4, and 6). Participants were not informed of which robot was which. We videotaped all sessions. In addition, we recorded all laser range sensors and camera data so that we could obtain the exact motions of the robot and the participants.



(a) (b)

Figure 5.2: (a) Schematic diagram of experimental area. (b) Experimental setting with guide robot.

5.3.2 Results

After the experiments, we examined the results from the following three viewpoints:

1) Can the robot correctly judge the situation and behave properly according to our proposed model? (Sections 5.3.2.1, and 5.3.2.2)
2) Can the robot's actions make the participants form a proper spatial formation, attract their attention and again reconfigure the spatial formation? (Sections 5.3.2..1, and 5.3.2.2)
3) Do the participants prefer the proposed robot (subjective evaluation)? (Section 5.3.2.3)

5.3.2.1 Spatial Formation Transformation

Distance between the robot and the visitors:

We found from the sensor data that some visitors were not in proper position in 5 cases out of 6 (2 visitors in each group, so total 12/2=6cases) cases. The robot took initiative to control visitor's standing position in all 5 cases. Thus the success rate of the robot's decisions was 100%. In 5 cases, 7 visitors out of 10 were out of range, and after the robot's action, 6 of these 7 moved inside the range. The total success rate was therefore 86% (the robot corrected 6 out of 7).

On the other hand, when the robot explained the paintings as a conventional robot, 6 visitors in both groups were out of range. Of the total of 6 visitors, only 2 moved inside the range when the robot began its explanation. The "success rate" for the conventional robot was thus 33% (2 out of 6, although here we only use the term "success rate" for convenience since the robot did not try to do correct the visitors).

Body orientation of the visitors:

From the recorded sensor data, we counted the number of visitors who successfully changed their body orientation to a direction between the robot and the painting after the robot employed a "pause and restart" at the beginning of its explanation. In 3 out of total 6 cases, the robot noticed that some visitors were not properly oriented. The robot employed "pause and restart" in all 3 cases, meaning its success rate at deploying the strategy was 100%. In 3 cases, 4 visitors out of 6 were incorrectly oriented. After the robot's action 3 of them changed their body orientations appropriately. The total success rate was therefore 75% (the robot corrected 3 out of 4).

On the other hand, when the robot explained the exhibits as a conventional robot, 5 visitors in both groups were incorrectly oriented. Out of 5, 2 changed their body orientations towards the robot and the painting just after the robot began its explanation. The "success rate" for the conventional robot was thus 40% (2 out of 5).

Face direction of the visitors:

In 4 out of 6 cases, the robot noticed that some visitors face direction were not towards the robot or the exhibit. The robot employed "pause and restart" in all 4 cases, meaning its success rate at deploying the strategy was 100%. In 4 cases, 6 out of 8 visitors' faces were directed elsewhere. After the robot's action, 4 of them changed their face direction towards the robot or the exhibit. The success rate was 66% (robot corrected 4 out of 6).

On the other hand, when the robot explained the exhibits as a conventional robot, 4 visitors' face direction in both groups were not properly directed. Out of the 4, 2 shifted their faces properly just after the robot started its explanation for a success rate of 50% (2 out of 4).

5.3.2.2 Reconfiguration of Spatial Formation

The robot turns towards the exhibit 2 after explaining exhibit 1 to initiate the reconfiguration of an F-formation. After explaining exhibit 2, it moves back a little bit and turns towards the median vector of the two exhibits.

We found from the sensor data that, 7 visitors out of the total 12 repositioned themselves when the robot turns its body towards the exhibit 2 after explaining exhibit 1, 9 out of 12 changed their position when the robot moves back a little bit and turns towards the median vector of two exhibits.

Therefore, the total success rate of the visitors' repositioning was 67% (7+9=16 out of 12+12=24).

In addition, 9 visitors out of 12 reoriented themselves towards the exhibit 2 when the robot turns its body towards the exhibit 2 after explaining exhibit 1 and 10 visitors out of 12 shifted their body orientation in response to the robot when it moves back a little bit and turns its body towards the median vector of two exhibits. Therefore, the total success rate of the visitors' body reorientation was 79% (9+10=19 out of the 24).

We coded the following items for each participant based on our video data. In addition, we checked when each item begins and ends.

- Attentional targets: Which painting did the robot talk about?
- Gaze: Which painting did a participant look at?
- Turn: Which painting did a participant finally turn toward?

Figure 5.3 shows a sample from our coding. In the initial state, the guide robot talked about painting 1 (P1) by turning its body towards the P1 and at the same time visitor 2 (V2) also turned his body orientation to P1. At first, the guide robot began to turn towards P2. As soon as the guide robot fully turned its body toward the P2, V2 also began to turn his body orientation from P1 to P2. Finally, both the guide robot and V2 completed their turning towards the P2 and guide robot began to talk about P2.

These results imply that if the robot wants to move visitors to give them better access to a target object, it is recommended that the robot rotates its body. Moreover, repositioning and reorientation of body are reasonable indications of the visitor's intention of spatial reconfiguration. Once a spatial formation is configured in front of P1, disappeared when the robot rotated its body toward P2. So, to reconfigure spatial formation in front of P2, the visitors need to turn their body toward the P2. These results demonstrate that shifting robot's body orientation is an effective attempt to change the body orientation of the visitors.

5.3.2.3 Subjective Evaluation

After all sessions, we asked the participants to fill out a questionnaire for each condition (described in section 4). Table 5.1 shows the questionnaire items.

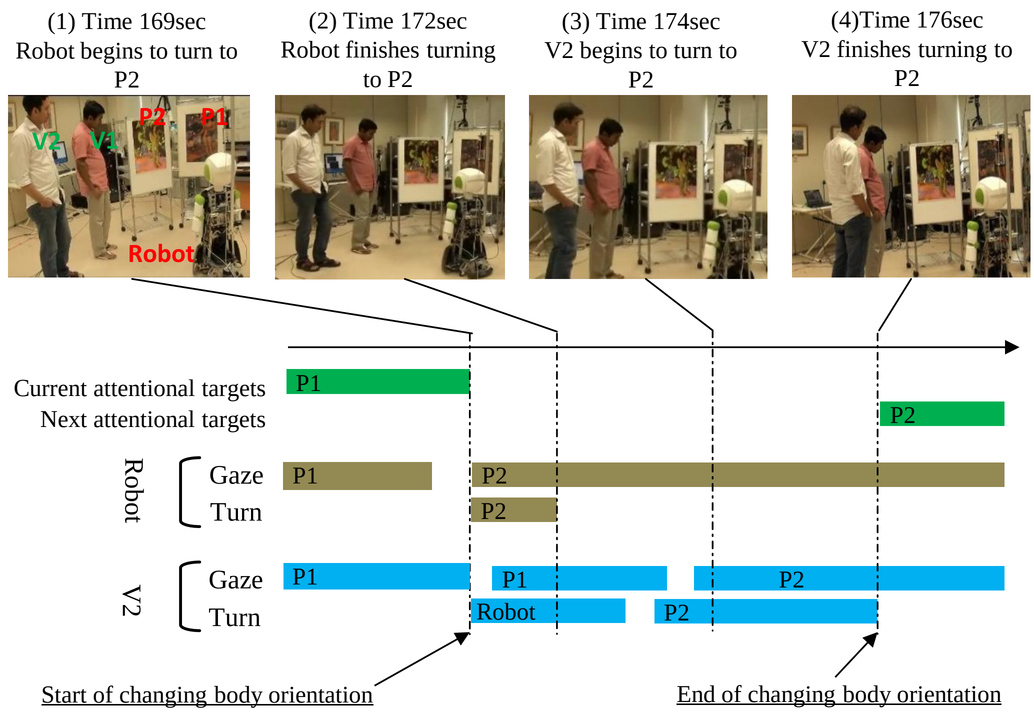


Figure 5.3: Example coding of behavior data (participants' body orientation changes from P1 to P2).

The measurement was a simple rating on a Likert scale of 1 to 7 with the range: 1-very ineffective, 2-ineffective, 3-somewhat ineffective, 4-undecided, 5-somewhat effective, 6-effective, 7-very effective (Figure 5.4). For the question "Did you feel that the robot was attentive to you during explanation?," repeated measures of ANOVA revealed a significant difference between the two conditions ($F(1, 11) = 6.8$, $p = 0.024$). For the question "Did you feel that the robot's action make you more involvement in the explanation?," ANOVA also showed a significant difference ($F(1, 11)$

=8.5, p=0.014). Finally, for the "Did you like the robot?", here too repeated measures of ANOVA showed a significant difference between the two conditions ($F(1, 11) = 10.38$, $p = 0.008$). These results suggest the participants' preference for our proposed robot and demonstrate the significance of guide robot being able to configure a proper spatial formation.

Table 5.1: Questionnaire items

No.	Questions
Q.1	Did you feel that the robot was attentive to you during explanation?
Q.2	Did you feel that the robot's action make you more involvement in the explanation?
Q.3	Did you like the robot?

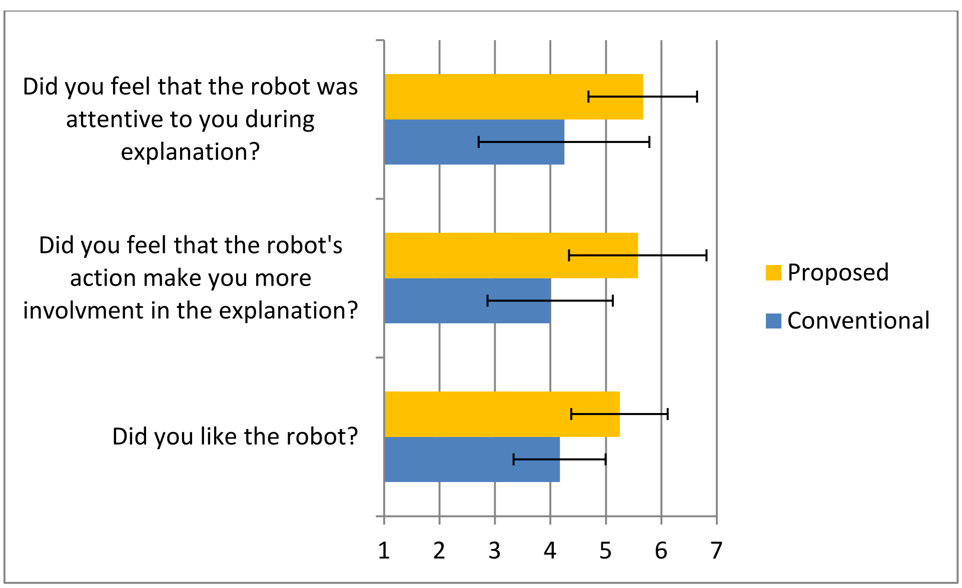


Figure 5.4: Results of subjective evaluation.

5.4 Chapter Summary

In our previous study, we developed a spatial model for a guide robot to create a proper spatial formation to explain a single exhibit to the visitors. This model was extended to multiple exhibits in this study. We also investigated that the robot can change the position and body orientation of the visitors by rotating its own body during ongoing explanation and thus reconfigure the spatial formation arrangement. This ability is important for a museum guide robot because it encourages visitors to naturally assume the appropriate position for observing the target exhibit while the robot is explaining.

5.4.1 Limitations

So far, we assumed that the people and the guide robot have already met and started conversation. We assumed that the guide robot and visitors have established a common belief that they share the conversation, in the real world this is not generally the case. The next chapter addresses the situation before or just at the moment when they establish the common belief that they are sharing a conversation.

Chapter 6

Spatial Formation Model to Initiate Conversation

Recent work in social robotics has enabled us to start developing humanoid robots that interact with people and support their regular activities. Current studies have explored that humanoid robots are suitable for communicating with humans. Social interaction between humans takes place in the spatial environment on a daily basis. We occupy space for ourselves and respect the dynamics of spaces that are occupied by others. In human-human interaction, to initiate conversation in a typical situation, one would stop at a certain distance orienting toward the target person, speak a greeting word, and find that they are engaging in a conversation. People do this unconsciously in daily life. On the other hand, in human-robot interaction, it is difficult for a robot to initiate conversation in such a way that humans do frequently in daily basis. For a social robot, initiating conversation is one of its fundamental capabilities. The robot needs to know every detail of human behavior to initiate conversation with humans.

In human communication studies, there are not many studies about how humans initiate conversation beyond the basic facts that they stop at a certain distance [14], start the interaction with a greeting [118], share a recognition of each other's participation state [148], and arrange themselves in a spatial formation [13]. Recent studies have started to reveal more detailed interaction, e.g., side participants stand close to the participants and often become the next participant [149], but this new knowledge has so far been limited. Some robots were developed with the capability to encourage people to initiate interaction [150, 151, 152]. Those robots wait for people to approach them first, which is one strategy for robots to initiate interaction. We propose a museum guide robot that can interact in a natural way at the moment of initiation of conversation as shown in Figure 6.1. We specifically focused on spatial formation and gaze, which have been discussed as important factors in the literature on human communication [153].

6.1 Modified System Architecture

Although there have been many studies focusing on tracking people in public space [27, 154, 155], only a few have attempted to localize a robot and track people simultaneously [28, 156]We have implemented our proposed model on a humanoid robot so that it can appropriately initiate conversation with visitors. We use a humanoid robot Robovie-R Ver.3 in this experiment. The robot can move via wheels installed on the bottom, and can move its head and arms by controlling its joints. Its head, which incorporates eye cameras and an ear microphone, moves along three axes (Yaw, Roll and Pitch) like a human head. In order to estimate the location and orientation of the visitors in the vicinity, we employ a system using what we call a sensor pole on which two laser sensors are attached [154]. The system is capable of tracking the location and orientation of the robot and the visitors independently at the same time.

An overview of the sensor pole is shown in Figure 6.2. In this study, we propose a sensor pole consisting of two lager range sensors attached to a pole. Bothe the robot and visitors can be tracked easily just by placing this sensor pole in the corners of the rooms where the robot will work. Two Logicool USB cameras are installed on the chest of the robot body to detect and track the visitors' faces and compute their directions. Assuming the human height 160 cm, the USB cameras are tilted up accordingly so that faces can be tracked accurately. In addition, science our system needs to perform in real-time situation, we developed our system on the GPGPU platform. Our system consists of three software units: the human and robot position and orientation tracking unit, the face detection and tracking unit, and the robot control unit. During its explanation of exhibit, the robot performs predetermined bodily non-verbal actions, such as facing towards the visitors, gesturing with its hands, and pointing at the exhibit.

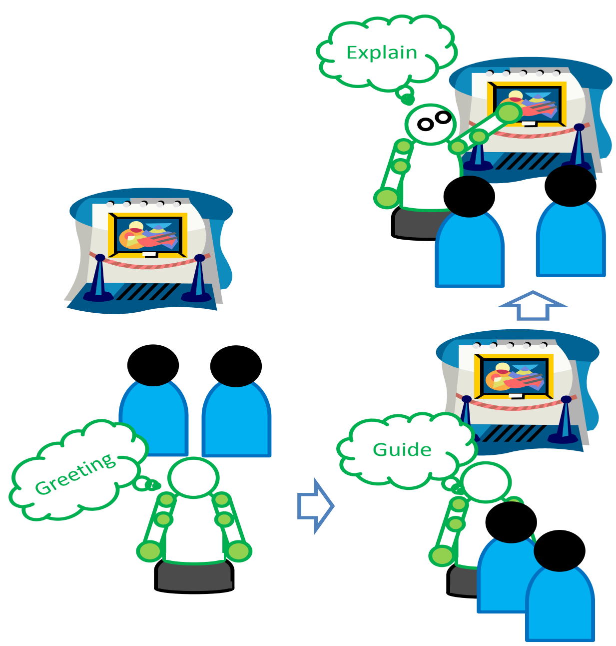


Figure 6.1: Conceptual image of a museum guide.

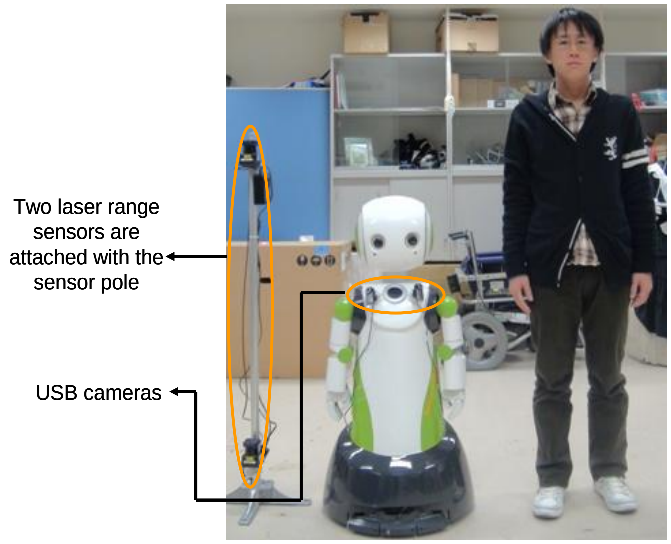


Figure 6.2: The sensor pole consists of two laser range sensors.

6.1.1 Tracking Position and Orientation of Visitors

A laser range sensor can measure the distance to the object on a horizontal plane and then map the distance data onto an image plane, as shown in Figure 6.3(left). This laser range sensor is installed at the shoulder level of visitors. The outline shape of a visitor's shoulders can then be observed as shown in Figure 6.3(right).

This outline portion of the visitor's shoulders can be considered as part of an ellipse. Thus we use an ellipse for the model of the tracking target

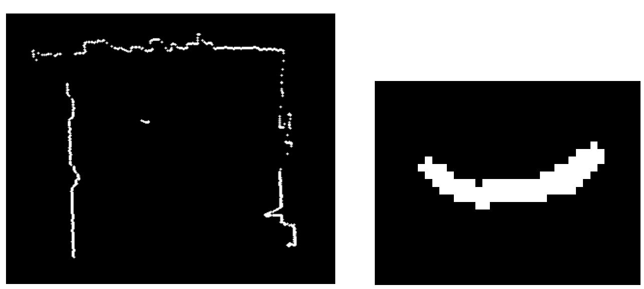


Figure 6.3: Distance-mapped image generated by the laser range sensor.

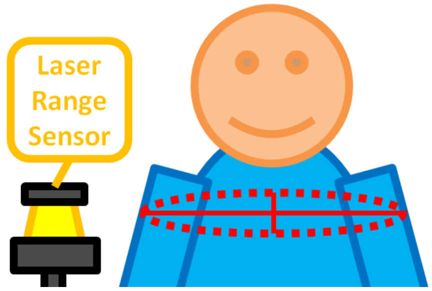


Figure 6.4: The shoulder outline can be modeled as an ellipse

(Figure 6.4). The system tracks the locations and orientations of visitors by using a particle filter framework.

We assume a coordinate system with its X and Y-axes aligned on the ground plane. The model of the tracking target is then represented with the center coordinates of the ellipse[x, y] and a rotation of the ellipse θ. These parameters are estimated in each frame by the particle filter. Using regular particle filters, we represent posterior distribution by a set of weighted samples that are propagated by a motion model and evaluated by an observation model. Here, we employ a simple random walk model for state propagation, and we evaluate samples based on the o observations of the

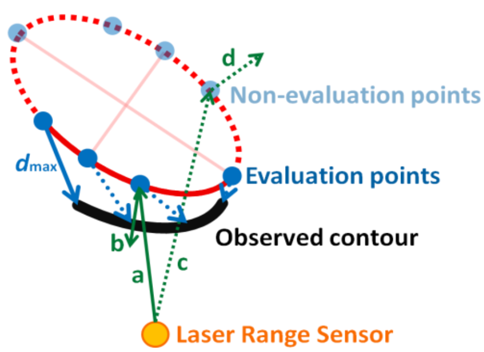


Figure 6.5: Evaluation model formed by fitting an ellipse to the shoulder outline obtained by the laser range sensor.

laser range sensor. The likelihoods of the samples are evaluated from assessing the contour similarity between the model and the body outline partially observed by the laser range sensor. The contour observation model is shown in Figure 6.5.

Specifically, i-th sample is generated at time t for each contour obtained from the image mapped from the distance data. The normal vectors of each point (the blue points in Figure 6.5), such as 'b' and 'd', are assumed as shown in Figure 6.5. The vectors from the position of the laser range sensor to the points are assumed to be 'a' and 'c'. The system then calculates the inner product of the vectors for each point. If the result of the inner product is negative (a.b<0), the point is able to be observed by the laser range sensor. These observable points are dealt with as evaluation points (the deep blue points in Figure 6.5). Conversely, a positive inner product (c.d>0) indicates that the point is not able to be observed by the laser range sensor. Next, the distance between each evaluation point and the observed contour is calculated (d_n). If the maximum distance of all d_n is labeled as d_{max}, the likelihood of i-th sample at time t is calculated using equation 4 (described in chapter 3):

$$w_{t,laser}^{i} = exp\left(\frac{-d_{max}^{2}}{\sigma_d}\right) \quad (4)$$

The σ_d is decided on the basis of the distribution of d_n. By calculating the expectation value across the samples, the system can estimate a visitor's position and orientation.

In our system, parallel computation of each evaluation point in each sample can be performed by using CUDA (Compute Unified Device Architecture) [157]. Consequently, the system performance will not be adversely affected by the number of visitors being tracked. Figure 6.6 shows the processing time of each frame as the number of tracking targets increases. As can be seen, the processing time is not significantly increased as the number of visitors increase.

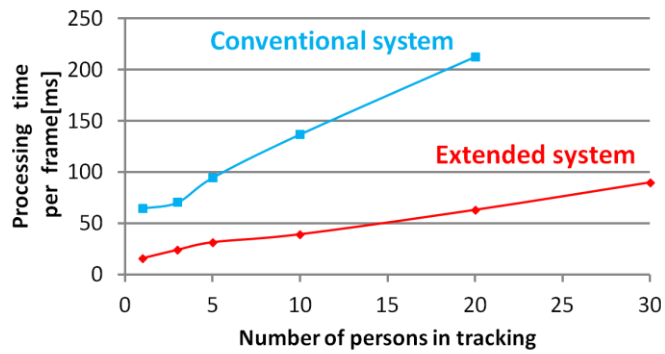


Figure 6.6: The processing time per frame compared to the number of persons being tracked. The blue and red lines indicate the time needed for the CPU and GPU respectively.

6.1.2 Tracking Location and Orientation of Robot

To be effective, our system must be capable of tracking the location and orientation of the robot and the visitors independently at the same time. However, it is difficult to track the position and orientation of the robot by using the methods outlined above. Therefore, we added another laser range sensor to the Sensor Pole to observe the characteristic outline shape of the robot. We found that the most varied shapes are observed when the laser range sensor is placed at the visitors' ankle level (about 15cm off the floor). Though the upper body of our robot is roughly similar in shape to that of a human, the lower body is quite different. In particular, the base section of the robot is shaped like a large circle which forms a vastly different shape from that of a visitor's ankles as shown in Figure 6.7 and Figure 6.8. By observing these outlines the system is able to readily distinguish the robot from visitors.

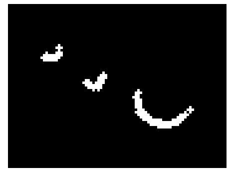


Figure 6.7: Human legs and the robot's base are distinctly observed.

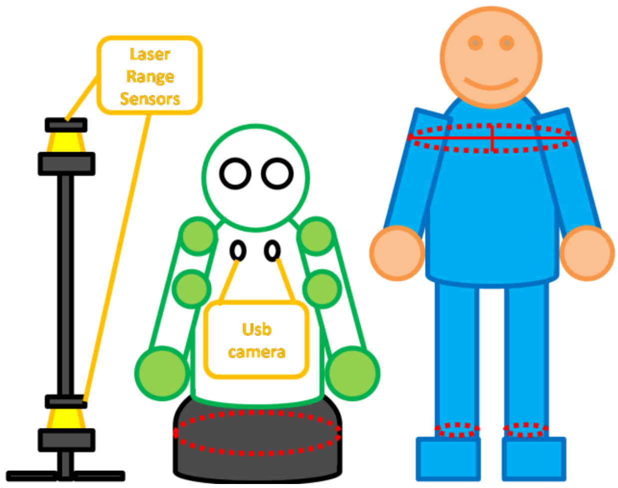


Figure 6.8: Sensor pole employs laser range sensor to track the position of the robot by using shape difference cues.

Therefore, we set the second laser range sensor at ankle height to estimate the location of the robot. Consequently, the two laser range finders on our "Sensor Pole" are set at human shoulder height and ankle height. Although a simple mechanism, in that it consists simply of two laser range sensors affixed to a pole, this "Sensor Pole" device can estimate the location and orientation of visitors and the robot at the same time without requiring concern for sensor fusion.

As mentioned above, the contour shape of the lower part of our robot is very different from that of the visitors. It is more similar to a perfect circle rather than an ellipse. Therefore, the contour of the guide robot is modeled as a circle for estimating its location. However, unlike the case with an ellipse, when we track a robot by using a circle outline model, it becomes difficult to estimate the orientation of the robot.

Therefore, we employ an electronic compass installed in the guide robot to obtain the robot's orientation (Figure 6.9). Specifically, the system obtains the orientation from the electronic compass independently while tracking the position of the robot. Once the initial orientation is memorized, the system computes the difference between the current direction and the memorized direction to recognize the rotation angle. Consequently, the combination of the Sensor Pole and the digital compass enables us to estimate the location and orientation of both visitors and the robot at the same time.

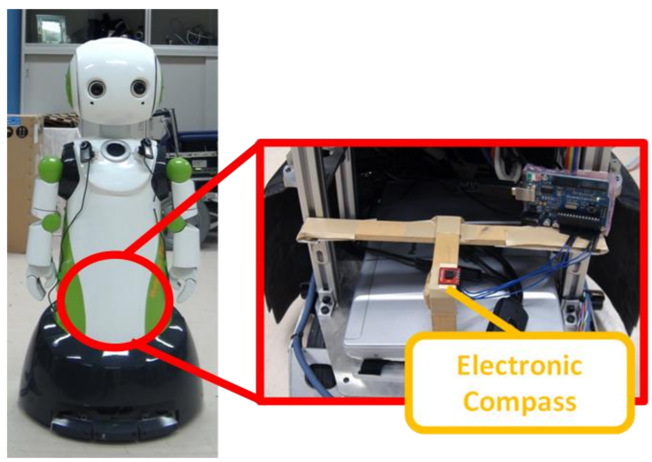


Figure 6.9: The electronic compass is installed in the guide robot.

6.1.3 Tracking Face Direction of Visitors

The visitor's head position and direction are tracked using the two Logicool USB cameras. We use an ellipsoid as a model to track the visitor's head. The visitor's head is modeled as shown in Figure 6.10(a). We assume a coordinate system represented by the X- and Y- axes aligned on the ground plane, with the Z-axis representing the vertical direction from the ground plane. We assume that visitors walk without tilting their heads, and that therefore the orientation of a human head can be identified by θ representing the rotation around the Z-axis. Thus, the final head model of the tracking target is represented by center coordinates of the ellipsoid $[x,y,z]$ and rotation of the ellipsoid θ.

We use the variables $x_t^{(i)}, y_t^{(i)}, z_t^{(i)}$ and $\theta_t^{(i)}$ in a sample $s_t^{(i)}$ for evaluating the likelihood using AdaBoost-based classifiers [144]. We evaluate samples by applying the AdaBoost-based classifiers over the projected image region and the number of stages passed in the cascade is employed as the likelihood of a human head. While an outline evaluation procedure is given below, a detailed account of the evaluation method is provided in [145]. Note that we use seven classifiers and that each classifier is trained respectively to detect a human head from a particular direction, such as front, 90^0 left, 90^0 right, and so on.

Figure 6.10: (a) Human head model (b) Human face tracking based on particle filter.

1) Project the sample's variables $x_t^{(i)}, y_t^{(i)}, z_t^{(i)}$ and $\theta_t^{(i)}$ at time t onto the USB camera image and obtain a projected position and size of the head. By considering the location and the direction of a head, we can also calculate the direction of a human head relative to the USB camera $\theta_{t,cam}^{(i)}$.

2) Extract the square image region corresponding to the head and extend it into a normal image based on its position and size.
3) Resize the extracted image region to obtain an image region $g_t^{(i)}$ (e.g. 24x24 sizes) as the input of the cascaded classifier.
4) Select a classifier by considering the direction of a human head relative to the USB camera $\theta_{t,cam}^{(i)}$. For example, if we use three classifiers such as front, 90^0 left, 90^0 right, then the front is selected in the case of $-45^0 \leq \theta_{t,cam}^{(i)} \leq 45^0$, the 90^0 left is selected in the case of $45^0 \leq \theta_{t,cam}^{(i)} \leq 135^0$, and so on.
5) Apply the selected classifier to the image region $g_t^{(i)}$ and obtain the likelihood score of a human head $w_{t,cam}^{(i)}$.

The result of human face tracking using the USB camera is shown in Figure 6.10(b). Basically, we track multiple target persons by using the particle filter independently. When a person is occluded by another person, the system terminates the tracking and when the person appears again, the system starts tracking again. We use the total likelihood value to decide whether the target person is occluded or not.

6.2 Modeling of Initiation of Interaction

We propose a model for a humanoid robot to appropriately initiate conversation.

6.2.1 Identifying Visitors' and Robots' Transactional Segments

Kendon suggested that humans make transactional segments that they can use to look at things, talk with others about them, and touch them [158]. In this paper, we define a human's transactional segment as the space in front of the person where there are no obstacles (objects and humans). When the angle between the forward direction of the human's body and the vector from the human's body center to an object/person is less than 90 degrees, and the distance between the human and the object/person is less than 200cm, the object/ person is identified as the human's target of attention (Figure 6.11). To initiate interaction, the guide robot needs to confirm that visitors' and its transactional segments overlap to create a joint transactional space. Once the joint transactional space is created, the robot and the visitors may agree to maintain this space and the robot can initiate conversation.

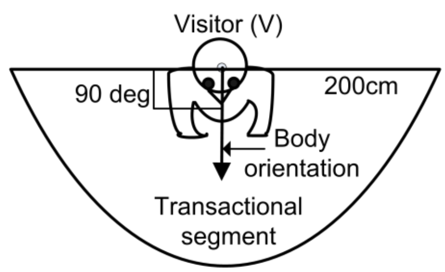


Figure 6.11: Transactional segment.

6.2.2 Control of Spatial Position and Initiation of Conversation

A conversation can start only when both guide robot and visitors have established a common belief that they will share the conversation. The museum guide robot should choose the visitors who may be interested in the exhibit and want to listen to its explanation. The robot considers that visitors are interested in the exhibit if their body and face orientations are towards the exhibit for a while (five seconds in our experiment). It considers that visitors are interested in the robot and would like to listen to its explanation if their body and face orientations are towards the robot for a while (five seconds in our experiment). If the robot finds such visitors, the robot proceeds to the next step to start conversation. For simplicity, we illustrate how the robot acts when there are two visitors in the following.

6.2.2.1 All Visitors are looking Towards the Robot

When all (two) visitors seem to pay attention to the guide robot (Figure 6.12(a)), the robot carries out the following steps to initiate conversation:

1) The robot turns its face and body towards the visitors. The angle between the two vectors from the robot to each visitor is calculated. The robot's body orientation follows the median vector between the two vectors (Figure 6.12(b)).
2) The robot then moves towards the visitors along the calculated median vector and stops at a position 100cm to 130cm from the visitors (Figure 6.12(b)). The transactional segment of the robot (TSR) and those of visitors 1 and 2 (TSV1 and TSV2) overlap to create a joint transactional space (Figure 6.12(c)). When the joint transactional space is created, the participants (the robot and the visitors) agree to maintain and control this space and start conversation. The robot then greets the visitors with an expression like, "Welcome," while turning its head towards them and says, "May I explain the exhibit to you?"
3) Next the robot moves to the appropriate position from where it can explain the exhibit comfortably (Figure 6.12(d)) and at the same time guides the visitors to the exhibit if the distance between the visitors and the exhibit is more than 120cm. The robot then starts the explanation of the exhibit. During its explanation of exhibit, the robot performs predetermined bodily non-verbal actions such as facing

towards the visitors, gesturing with its hands and pointing to the exhibits.

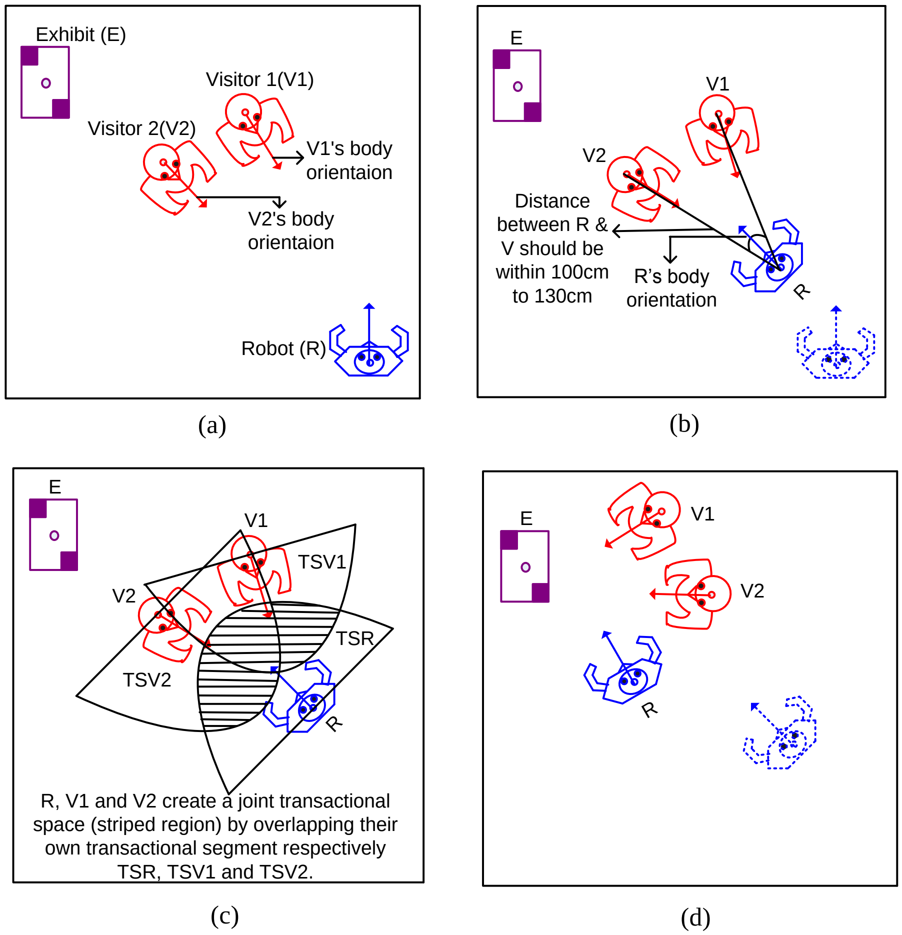


Figure 6.12: Two visitors looking toward the guide robot.

6.2.2.2 Some of the Visitors are Looking Towards the Robot

When a visitor is looking towards the robot and the other is looking towards the exhibit (Figure 6.13(a)), the robot goes after the subsequent steps to initiate conversation:

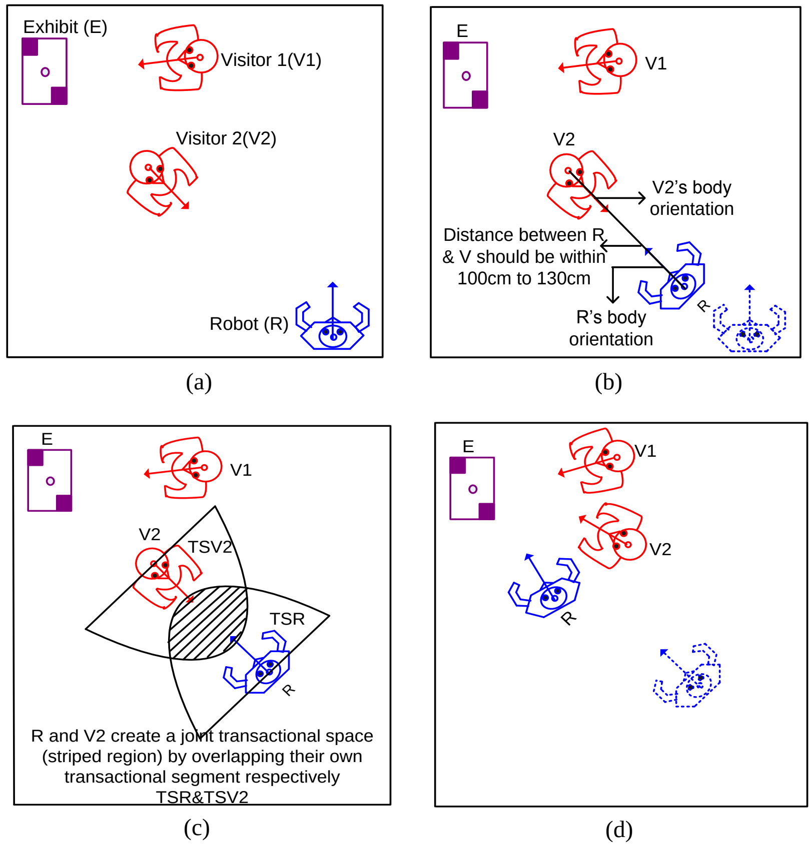


Figure 6.13: One visitor looking toward the robot and another toward the exhibit.

1) The robot first pays attention to the visitor looking at the robot by turning its face and body towards him/her. The robot's body orientation follows the vector from the robot to the visitor. The robot then moves towards the visitor and stops at a position 100cm to 130cm away from him/her (Figure 6.13(b)).
2) They (the robot and the visitor) establish a joint transactional space by overlapping their transactional segments (Figure 6.13(c)). By creating the joint transactional space, they reach a belief that they can share conversation. The robot then greets the visitor with the expression, "Welcome," and says, "May I explain the exhibit to you?"

3) Next the robot moves to the appropriate position from where it can explain the exhibit comfortably (Figure 6.13(d)) and at the same time guides the visitor to the exhibit if the distance between the visitor and the exhibit is more than 120cm. If the robot finds the other visitor (who is looking toward the exhibit) keep facing towards the exhibit, the robot invites him/her also to join the guide. The robot then starts the explanation of the exhibit. During its explanation of exhibit, the robot performs predetermined bodily non-verbal actions.

6.2.2.3 All Visitors are Looking Towards the exhibit

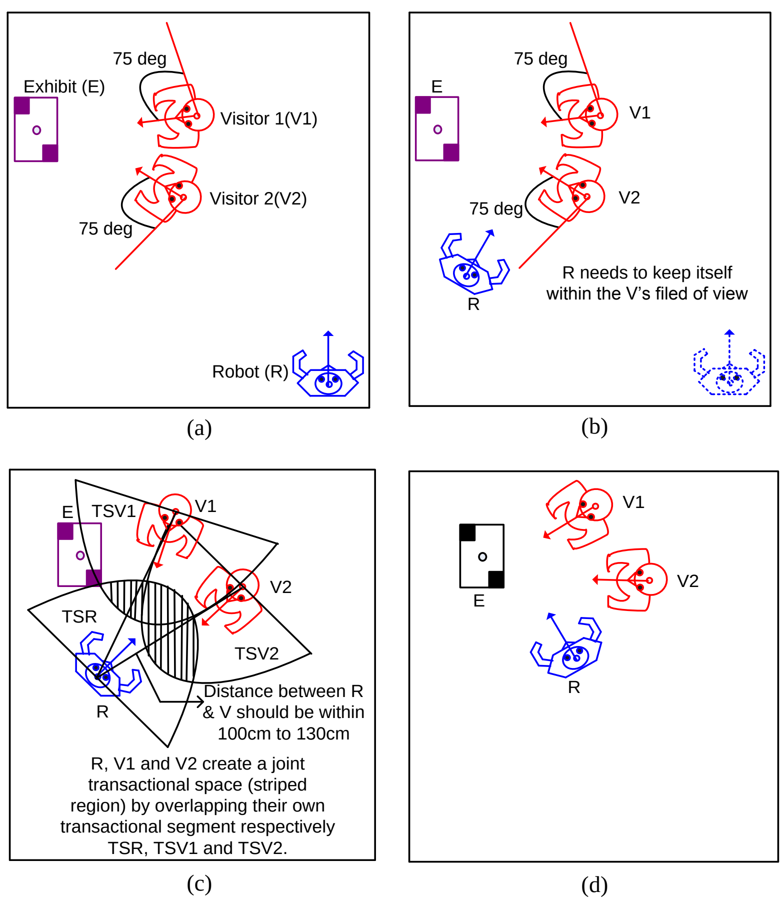


Figure 6.14: Two visitors looking towards the exhibit.

The guide robot takes the following steps when it finds two visitors are looking towards the exhibit (Figure 6.14(a)):

1) When the guide robot finds visitors keep facing towards the exhibits for five seconds, the robot considers the visitors to be highly interested in the exhibit. In this case, the robot first approaches the visitors. The robot moves toward the visitors and stops near them. If all or any visitors notice the robot's arrival and turn and keep their faces towards the robot, this turns out to be the first or the second case, respectively. Thus, the robot carries out the actions for each case. Otherwise, the robot needs to stop near the visitors within their field of view (Figure **6.14**(b)). In this work, visitor's field of view is set to a 150 degree fan-shaped area in front of the body of the visitor (Figure 6.14(b)). In this case, the robot should also be at a position 100cm to 130cm away from the visitors (Figure 6.14(c)).
2) When both robot and visitors meet gaze each other, they create a joint transactional segment (Figure 6.14(c)). The situation can be interpreted that they have reached an agreement to participate in the conversation. The robot then greets the visitors with the expression, "Welcome," and says, "May I explain the exhibit to you?"
3) Next the robot moves to the appropriate position from where it can explain the exhibit comfortably (Figure 6.14(d)). The robot then starts the explanation of the exhibit. During its explanation of exhibit, the robot performs predetermined bodily non-verbal actions.

6.3 Experiment with Guide Robot

We conducted an experiment to verify that our proposed model is useful for a mobile museum guide robot.

6.3.1 Experimental Area

The experiment was conducted in our laboratory, under the assumption that it was a small art museum. The experimental area was restricted to 350cm x 350cm area. One painting was placed in one corner of the area and the robot with the proposed system waited in another corner (Figure 6.15). The sensor pole was placed to cover the experimental area.

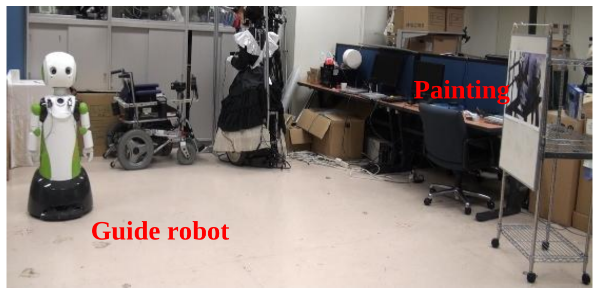


Figure 6.15: Experimental situation.

6.3.2 Experimental Condition

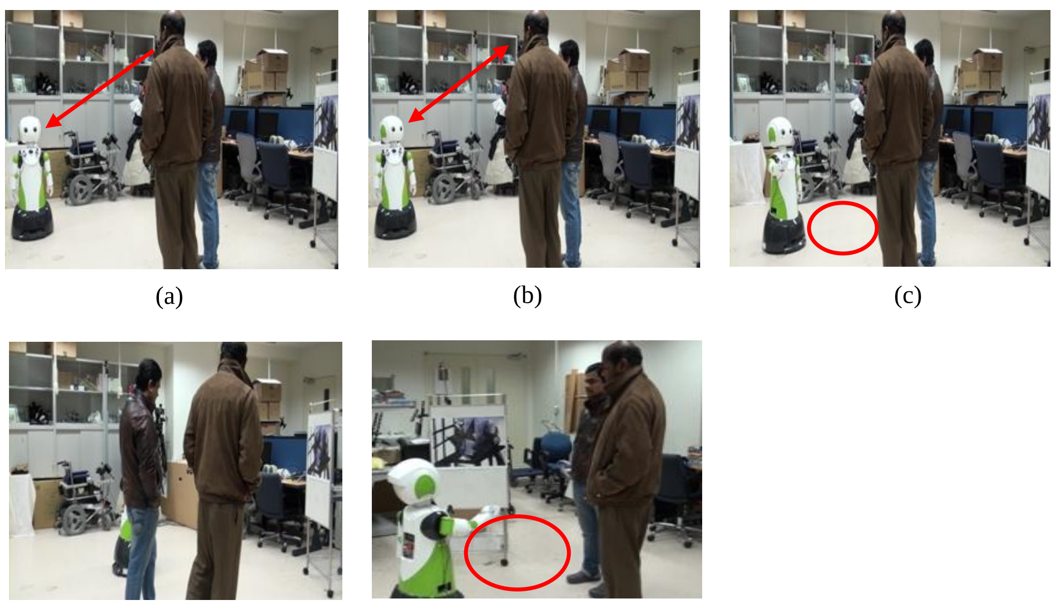
The robot based on our proposed model was compared with a robot that did not employ the proposed model.

a) *Proposed Robot:* Robot behaves based on the model outlined in this paper. During its explanation of painting, the robot performs predetermined bodily non-verbal action. The timing of these nonverbal behaviors is programmed based on analyzing the videotaped interactions between human guides and visitors at the museum
b) *Conventional Robot:* When the robot finds any visitor, it immediately moves to a location that is appropriate for explaining the painting and

starts explanation. It explains the exhibit with the same preprogrammed non-verbal behaviors as the proposed robot.

6.3.3 Procedure

A total of 20 graduate students (average age is 30) participated in the experiment. We formed 10 groups with 2 members in each group. All students participated in the experiment group wise. They were asked to play the role of visitors who visit the art museum in various ways so that they could fully judge the appropriate behavior of each robot. The experiment used a within-subject design and the order of experimental conditions was counterbalanced. Participants were not informed of which robot was which.



(a) (b) (c)

(d) (e)

Figure 6.16: Example scenes from the experiment.

Figure 6.16 shows experimental scenes using the proposed model. Two visitors are showing their interest by turning their bodies and faces towards the guide robot (Figure 6.16(a)). In Figure 6.16(b), the guide robot turns its head towards the visitors. This serves as the robot's acknowledgement of the interested visitors' presence. The red double arrow in Figure 6.16(b) indicates that mutual gaze is established. The guide robot then moves towards the visitors and greet them (Figure 6.16(c)). Red circle in Figure 6.16(c), indicates that a spatial formation if created. In Figure 6.16(d), the robot guides the visitors towards the exhibit. Finally the robot starts its explanation (Figure 6.16(e)). In Figure 6.16(e), guide robots reconfigure spatial formation in front of the painting before start its explanation.

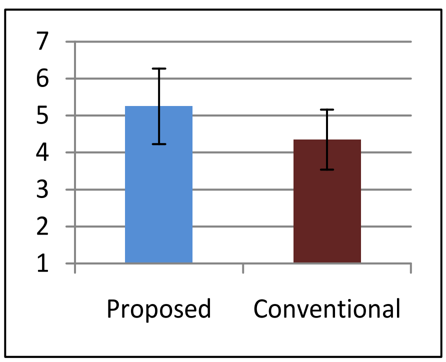
6.4 Experimental Results

After the experiment, we asked participants to subjectively rate the robot's effectiveness on a seven-point Likert scale, with the range: 1-very ineffective, 2-ineffective, 3-somewhat ineffective, 4-undecided, 5-somewhat effective, 6-effective, 7-very effective. The questionnaire items are shown in Table 6.1.

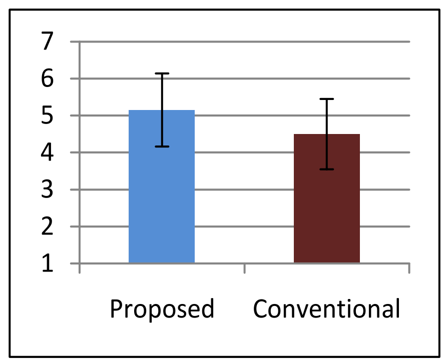
Table 6.1: Questionnaire items

No.	Questions
Q.1	Do you think that the robot was able to greet you properly?
Q.2	Do you think that the robot attended to you adequately?
Q.3	Overall evaluation about the robot.

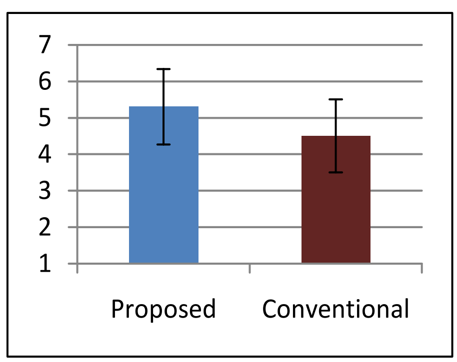
In the experiment, two types of robots (the proposed and the conventional robot) were investigated. We conducted a subjective evaluation of the experiment among the participants. The data were analyzed by one-way analysis of variance (ANOVA), with the experimental conditions as the within subject factor (repeated measures). The results of the experiment are as follows. For the question "Do you think that the robot was able to greet you properly?" (Figure 6.17(a)), the repeated measures of ANOVA reveal a significance difference between the two conditions *(F(1,19=12.93, p=0.0019)*. The same was true for the question. The same was true for the question "Do you think that the robot attended to you adequately?" (Figure 6.17(b)), where the repeated measures of ANOVA also show a significant difference (*F(1,19)=6.04, p=0.0236*). Finally, for the participants' overall evaluation (Figure 6.17(c)), here too the repeated measures of ANOVA show a significant difference between the two conditions (*F(1,19)=7.79, p=0.0116*).



(a)



(b)



(c)

Figure 6.17: Result of subjective evaluation.

We compared the proposed model with another baseline model (conventional robot). These results suggest the participants' preference for our proposed robot, and moreover demonstrate the significance and utility of guide robots being able to initiate interaction appropriately with visitors.

6.5 Chapter Summary

In real world, in human-human interaction, to initiate conversation, one would stop at a certain position toward the target person, speak a greeting word, and find that they are engaging in a conversation. In HRI, we believe that the capability of a robot to naturally initiate conversation is one of the major capabilities to be implemented in future museum guide robots. Although in many other research projects it is assumed that people and robots have already met and started interaction, in the real world this is not generally the case. Or, perhaps at an early deployment phase robots might not need to initiate interaction by themselves, as people would be interested in novel robots and walk over to the area in front of the robot. In such a case, robots don't need to know the constraints of spatial configuration in the initiation of interaction. In contrast, when robots actually start to work in the real world, it needs to know the constraints of spatial configuration in the initiation of interaction. First meeting scenario occurs frequently in the case of museum tour guide. Therefore the capability to initiate conversation is vital for museum guide robots. In this study, we proposed a mobile museum guide robot system that can judge the behavior of visitors and initiate interaction with them accordingly. There are four contents in our proposed model: identifying interested visitors, greetings, guide, and explanation. This study paves the way for research on initiation of interaction by guide robots although our proposed model has yet been tested in a specific scenario in laboratory experiments.

6.5.1 Limitations

An important limitation of our current systems is that it deals with the two participants at a time. Therefore, we do not know, what is effect of robot's behaviors on the people in a multiple people scenarios, though our current system can track the position and orientation of maximum ten people. Moreover, we evaluate the system in a highly controlled environment; therefore, we do not know how it will perform in the real environment. In the real environment, different types of visitors visit the museum, where some of them may not be interested about the exhibit but only curious about the robot. In that case, robot may treat those visitors as interested visitors about the exhibits.

Chapter 7

Conclusion

As robots enter the everyday physical world of people, it is important that they abide by society's unspoken social rules such as respecting people's personal spaces. In this thesis, we explore issues related to human personal space around robots, beginning with a review of the existing literature in human-robot interaction regarding the dimensions of people, robots, and contexts that influence human-robot interaction (HRI).

Service robot plays very important role in human world. Establishment of spatial formation is a vital component for service robots. We analyzed video footages recorded at the National Japanese American Museum in Los Angeles in order to find out how expert human guides explain the exhibit/s towards multiple visitors. We analyzed these data from the perspective of creation of spatial formation. We are especially interested in how museum guide create spatial formation with the visitors.

The primary focus of this dissertation is to develop a mobile museum guide robot that is capable to create and control spatial formations with the visitors in different situations. Towards these larger goals, this work has made a set of methodological, theoretical, and practical contributions. The methodological contributions include an interdisciplinary, integrated process for designing, building and evaluating social behavior for social robots. These contributions are listed in Section 7.1. The theoretical contributions advance our understanding of human communicative mechanisms from a computational point of view and of people's responses to theoretically based manipulations in these mechanisms when they are enacted by social robots. Section 7.2 summarizes these contributions. The practical contributions include the computational models of social behavior created for the empirical studies, which are described in Section 7.3. To designing this spatial formation model, we faced a number of technical and methodological challenges that remain significant bottlenecks in advancing the effective design for museum guide robots. Section 7.4 discusses these central challenges and provides a vision for how future work might address them. The last section in this chapter provides the closing remarks.

7.1 Methodological Contributions

This dissertation presents an incremental process for studying and designing human communicative mechanisms and a demonstration of an interdisciplinary research approach that combines techniques and methods from several research domains such as sociology, psychology, and human-robot interaction. This process is helps to extract the design, and behavior variables from human social behaviors. In addition, it also helps to create computational representations based on extracted variables. This work created a number of experimental paradigms in which these computational behavioral models were manipulated and evaluate through subjective and quantitative measures that revealed the social and cognitive outcomes. Table 7.1 lists these contributions.

Table 7.1: Methodological contributions

Context	Contributions
All studies	A theoretically and empirically grounded, interdisciplinary process for designing, implementing, and evaluating spatial formation model for mobile museum guide robots.
Chapter 3	An experimental framework for studying how museum guide robot creates and control spatial formation in terms of distance constraints, face direction constraints, body orientation constraints, and FOV constraints with the visitors before start its explanation. An experimental framework for studying how guide robot escorts visitors to several exhibits and then reconfigure spatial formation in front of another exhibit. An experimental framework for studying how museum guide robot attracts the attention of the visitors using "pause and restart" strategy at the beginning of its explanation.
Chapter 4	An experimental framework for studying how museum guide robot identifies interested bystander around itself during its ongoing explanation, invite him/her into the ongoing explanation session, and thus reconfigure spatial formation again.
Chapter 5	An experimental framework for studying how museum guide robot changes the position and body orientation of the visitors by rotating its own body from one exhibit to another and thus reconfigure spatial formation during explanation of multiple exhibits collectively at a time.
Chapter 6	An experimental framework for studying how museum guide robot initiates conversation with the visitors by judging their behaviors.

7.2 Theoretical Contributions

The theoretical contributions of this work consist of a sets of new knowledge extracted from a psychology, sociology, cognitive science and human-robot interaction fields that helps a deeper understanding of human attention control mechanisms as applied to robots and their social/cognitive outcomes. Table 7.2 provides a detailed list of these contributions.

Table 7.2: Theoretical contributions

Context	Contributions
All studies	Evidence that robot's verbal action and employment of "pause and restart" lead to significant social and cognitive outcomes, particularly better feeling of attraction toward the robot, stronger feeling of contributing to create spatial formation with the guide robot.
Chapter 3	Evidence that successful establishment of spatial formation requires: visitors should stand within robot's FOV, with in 120 cm from the guide robot, and by turning their faces and body orientations towards the robot. Evidence that robot's verbal action and employment of "pause and restart" affects the visitors to stand in proper position and to turn their faces and body orientations towards the guide robot and thus configure spatial formation.
Chapter 4	Evidence that museum guide robot can identify interested bystander if they are found within social space and within robot's FOV. Evidence that robot's body turning towards the interested bystander and robot's verbal action affect him to be included into existing

	spatial formation. Evidence that turning back of robot's body orientation towards the exhibit again allows the interested bystander to coordinate a spatial-orientaional move of his/her own, and which allow him/her access to the existing spatial formation and thus reconfigure spatial formation.
Chapter 5	Evidence that museum guide robot can change the position and body orientation of the visitors by rotating its own body from one exhibit to another during explanation of multiple exhibits collectively and thus reconfigure spatial formation.
Chapter 6	Evidence that guide robot can identify the body orientation, face direction, and the standing position of the visitors in different situations and move towards them , greet them, and guide them to the exhibit accordingly. Evidence that robot's movement action and greeting behavior make the visitors to feel them comfortable to initiate conversation.

7.3 Technical Contributions

The technical contributions of this dissertation include a set of design, and behavioral variables for spatial formation mechanism of robots, and computational models of incremental spatial formation process that are created for empirical studies of this dissertation.
Table 7.3 illustrates a detailed list of these contributions.

Table 7.3: Technical contributions

Context	Contributions
All studies	Conceptually designed and implemented a spatial formation model for a mobile museum guide robot to perform the human robot interaction studies. A set of design, and behavioral variables for designing a spatial formation model for museum guide robots.
Chapter 3	A computational model of creating spatial formation before starting of robot's explanation is programmed in C, C++, and Opencv library functions.
Chapter 4	A computational model of identifying and inviting interested bystander into existing spatial formation system is programmed in C, C++, and Opencv library functions.
Chapter 5	A computational model of changing the position and body orientation of the visitors by rotating robot's own body from one exhibit to another during explanation of multiple exhibits collectively is programmed in C, C++, and Opencv library functions.
Chapter 6	A integrated computational model of initiating interaction with multiple visitors, guide them to the exhibit, and explain the exhibit is programmed in C, C++, and Opencv library functions.

7.4 Future Works

In this dissertation, we attempted to design a spatial formation model to museum guide robots-with understanding of human social behavior. To develop the spatial formation model, we used knowledge and methods from a variety of research areas and made a number of design decisions on what knowledge and methods to use and how to use these resources. There are still several issues that have not been addressed in the current model. Some of these issues are discuss in the following.

- **Generalizability:** We tested our proposed model for a specific scenario where two people were assigned to interact with the guide robot. Therefore, its generalizability is limited. It is possible that context affects the preference of a guide robot's behavior. The strategies are required to interact with only the interested visitors among the multiple visitors. More studies are needed to explore the detail behavior of the guide robot during interaction with the interested visitors among a group of visitors in a busy museum.
- **Limited interaction:** The guide robot actions we designed for the dissertation are limited and may be different from the behavior of an actual museum guide. Robots may need to show different types to behavior depending on the various complex situations. We need to explore more complex situation to create spatial formation and design appropriate behaviors for those situation.
- **Lose of participants from existing spatial formation:** In this dissertation, we considered the case, where guide robot can identify and invite interested bystanders around itself and thus reconfigure spatial formation again. However, existing spatial formation may also lose the participants also. We need to explore the situation where people move away from the existing spatial formation and design appropriate robot's behavior to reconfigure spatial formation after change of number of participants.
- **Size of the exhibits:** In this research, we used medium size of exhibits (paintings) for the experiments. However, museum guide needs to

interact with the visitors while explaining large or small size of exhibits. In that case the standing position of the guide robot and the visitors may be different. We need to explore the case where guide robot explains large or small size of exhibits to the visitors and design spatial positional model to create spatial formation.

- **Controlled laboratory setting:** The research approach presented here used controlled laboratory experiments to understand the social and cognitive outcomes of the designed spatial formation model. Whether these outcomes could be obtained in less controlled environments is unknown. To prove the current approach will be effective beyond controlled laboratory settings, future work needs to also situate designed behaviors in real-world scenarios and contexts.

7.5 Published Papers from the Study

The following publications are a direct consequence of the research carried out during the elaboration of the thesis, and give an idea of the progression that has been achieved.

Journals

- Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Keiichi Yamazaki, and Akiko Yamazaki, “A Mobile Guide Robot Capable of Establishing Appropriate Spatial Formations”, IEEJ Transactions on Electronics, Information and Systems, Vol. 133, No. 1, pp. 28-39, January 2013.
- Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Akiko Yamazaki, and Keiichi Yamazaki, “Development of a Mobile Museum Guide Robot That Can Configure Spatial formation with Visitors”,

Intelligent Computing Technology, 8th International Conference, ICIC 2012, Vol. 7389, pp. 423-432, LNCS, Springer, Heidelberg, 2012.

International Conferences

- Mohammad Abu Yousuf, Yoshinori Kobayashi, Akiko Yamazaki, Yoshinori Kuno, "Implementation of F-Formation and "Pause and Restart" for a Mobile Museum Guide Robot", Proc. Of International Workshop on Multimodality in Multispace Interaction (MiMI), pp. 81-92, Japan, 2011.
- Mohammad Abu Yousuf, Yoshinori Kobayashi, Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, "A Mobile Guide Robot Capable of Formulating Spatial Formation", Proc. of 18th Korean-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp. 274-280, Japan, 2012.
- Mohammad A. Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Keiichi Yamazaki, Akiko Yamazaki, "Establishment of Spatial Formation by a Mobile Guide Robot", Proc. of 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp.281-282, Boston, USA, 2012.
- Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Keiichi Yamazaki, and Akiko Yamazaki, "Model of Guide Robot Behavior to Explain Multiple Exhibits to Multiple Visitors", Proc. of International Session of 30th Annual Conference of the Robotics Society of Japan (RSJ2012), pp. 83-86, Sapporo, Japan, 2012.
- Mohammad Abu Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Akiko Yamazaki, and Keiichi Yamazaki, "How to Move Towards Visitors: A Model for Museum Guide Robots to Initiate Conversation", 22nd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2013), pp. 587-592, Gyeongju, Korea, 2013.

7.6 Closing Remarks

I believe that the ability to naturally control spatial formations is important for a museum guide robot because in this way the visitors can acquire equal, direct, and exclusive access to the target object, just as with a human guide. In this dissertation, I proposed a model for a museum guide robot to create an appropriate spatial formation, known as an F-formation, in order to initiate interaction with multiple visitors in different situations. Based on our analysis of videos collected at real museums, I have found that a human guide always creates spatial formation with the visitors. Based on this, I developed a museum guide robot system that is able to move from one exhibit to another in a museum gallery, to appropriately establish spatial formation with the visitors in different situations.

In this process, I employed methods and knowledge from a variety of disciplines (such as psychology, sociology, cognitive science, and robotics) and made a number of design decisions that were grounded in theory and empirical data. While further work remains in order to improve the validity of these decisions and the generalizability of the results, this dissertation provides a major step towards designing an important social capability for mobile museum guide robot robots using a theoretically and empirically grounded methodology. I believe that our proposed museum guide robot is able to offer an attractive guide performance.

Appendix A

Data Collection Technique

This appendix includes the methods or techniques used in this dissertation to gather the data from the human-human and human-robot interactions studies. To collect the data, we have analyzed some video footage recorded at the actual museums in order to find out how expert human guides explain the exhibit/s towards multiple visitors as well as questionnaire methods. The questionnaire data is collected in terms of Likert scale.

A. Video Analyzing Based Method Observation and analyzing of video footages of human-human interaction is a useful data gathering technique at any stage during the system development. Early in design, observation helps designers understand the users' context, tasks, and goals. Observation conducted later in the development, e.g., in evaluation, may be used to investigate how well the developing prototype supports these tasks and goals. Observation and analysis may also take place in the field, or in a controlled environment. In the former case, individuals as they go about their day-to-day tasks in the natural settings. In the latter case, individuals are observed performing specified tasks within a controlled environment. In this dissertation, we have used video footages collected from actual museum to gather observational data. We analyzed these data from the perspective of creation of spatial formation. We are especially interested in how museum guide create spatial formation with the visitors.

B. Questionnaire Based Method Questionnaires are well-establish technique for collecting demographic data and users' opinions. Effort and skill are needed to ensure that questions are clearly worded and the data collected can be analyzed efficiently. Questionnaires can be used on their own or in conjunction with other methods to clarify or deepen understanding. It is important that questions are specific; when possible, closed questions should be asked and a range of answers offered, including a 'no opinion' or 'none of these' option.

C. Likert Scale A likert scale is a psychometric scale commonly involved in research that employs questionnaires. Likert scales are used for measuring opinions, attitudes, and beliefs, and consequently they are widely used for evaluating user satisfaction with systems. The purpose of this rating scale is to elicit a range of responses to a question that can be compared across respondents. This is good for getting people to make judgments about things, e.g., how easy, how effective, and such like. The success of Likert scales relies on identifying a set of statements representing a range of possible opinions and this scale is more commonly used because identifying suitable statements that respondents will understand easily. When designing Likert scale, issues that need to be addresses include: how many points are needed on the scale, how should they be presented, and in what form? Many questionnaires are used seven or five-point scales and there are also three point scales. Longer range is better, when asking respondents to make subtle judgments. In this dissertation, we have used 7-point Likert scale to collect the participants' impression on robot's behaviors using questionnaire method where '1' stands for definite no/very ineffective, and '7' stands for definitely yes/very effective (see in Figure A.0.1 as an example).

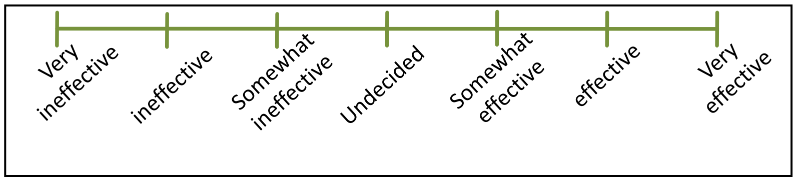


Figure A.0.1: Likert scale

References

[1] K. Dautenhahn. Methodology & themes of human-robot interaction: A growing research field. *Int. J. of Advanced Robotic Systems*, vol. 4(1), pp.103–108, 2007. 1

[2] 1.C. Bartneck and J. Forlizzi. A design centered framework for social human-robot interaction. In *IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004)*, Okayama, Japan. pp. 591- 594, 2004. 1

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A Survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(4-3):143–166, 2003. 1

[4] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke. Towards a humanoid museum guide robot that interacts with multiple persons. In *Proc. of IEEE-RAS International Conference on Humanoid Robots (Humanoids2005)*, pp. 418–423, 2005. 3, 19

[5] 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 *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2009)*, pp. 109–116, 2009. 3

[6] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh. Where to look: A study of human-robot engagement. In *Proc. of International Conference on Intelligent User Interfaces (IUI2004)*, pp. 78–84, 2004. 3, 19, 20

[7] H. Kuzuoka, K. Pitsch, Y. Suzuki, I. Kawaguchi, K. Yamazaki, A. Yamazaki, Y. Kuno, P. Luff, and C. Heath. Effect of restarts and pauses on achieving a state of mutual orientation between a human and a robot. In *Proc. of ACM International Conference on Computer Supported Cooperative Work (CSCW2008)*, pp. 201–204, 2008. 3

[8] T. Fukuda, Y. Nakauchi, K. Noguchi, and T. Matsubara. Sequential human behavior recognition system for cooking support robot. *JSME J. of Robotics and Mechatronics*, Vol.17, No.6, pp. 717–724, 2005. 3

[9] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro. Interactive robots as social partners and peer tutors for children: A field trial. *J. of Human Computer Interaction,* Vol. 19, No. 1-2, pp. 61-84, 2004. 3

[10] H. Asoh, S. Hayamizu, I. Hara, Y. Motomura, S. Akaho, and T. Matsui. Socially embedded learning of the office- conversant mobile robot Jijo-2. *International Joint Conference on Artificial Intelligence (IJCAI),*Vol. 2, pp. 880-885, 1997. 3

[11] T. Erickson. From interface to inter place: The spatial environment as a medium for interaction. *Proc. of Conference on Spatial Information Theory (ECSIT'93)*, pp.391-405, 1993. 3

[12] B. Hillier. *Space is the machine.* Cambridge University Press, Cambridge, 1996. 3

[13] A. Kendon. *Conducting interaction–patterns of behavior in focused encounters*. Cambridge University Press, Cambridge, 1990. 3, 5, 6, 21, 25, 28, 37, 110

[14] E. T. Hall. *The Hidden dimension: Man's use of space in public and private*. The Bodley Head Ltd, London, UK, 1966. 3, 22, 25, 110

[15] E. T. Hall. Proxemics. *Current Anthopology*, vol. 9, pp. 83-108, 1968. 3, 22

[16] H. Hüttenrauch, K. S. Eklundh, A. Green, and E. A. Topp. Investigating spatial relationships in human-robot interactions. In

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006), pp. 5052-5059, 2006. 4, 24

[17] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. How Close? A model of proximity control for information-presenting robots. In *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2008)*, pp.137-144, 2008. 4, 24

[18] Y. Nakauchi and R. Simmons. A social robot that stands in line. In *Proc. of* In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2000)*, pp. 357–364, 2000. 4, 24

[19] F. Yamaoka, T. Kanda, H. Ishiguro and N. Hagita. Developing a model of robot behavior to identify and appropriately respond to implicit attention-shifting. In *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2009)*, pp. 133–140, 2009. 4

[20] C. Goodwin. Restarts, pauses, and the achievement of a state of mutual gaze at turn beginning. *Sociological Inquiry*, vol. 50, issue 3-4, pp. 272-302, 1980. 5

[21] A. Ghosal. *Robotics: Fundamental concepts and analysis*. Oxford University Press, 2006. 10

[22] H. Yanco, and J. Drury. A taxonomy for human-robot interaction. In *Proc. of AAAI Fall Symposium on Human-Robot Interaction*, pp. 02-03, 2002. 13

[23] L. Gong, and C. N. Nass. When a talking-face computer agent is half-human and half- humanoid: human identity and consistency preference. *Human Communication Research*, vol. 33(2). pp. 163-193, 2007. 13

[24] K. MacDorman, and H. Ishiguro. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, vol. 7(3). pp. 297-337, 2006. 13

[25] H. Ishiguro. Android Science - Towards a new cross-interdisciplinary framework. In *Proc. of CogSci Workshop: Toward Social Mechanisms of Android Science*, pp. 1-6, 2005. 15

[26] F. Dellaert, D. Fox, W. Burgard, and S. Thrun. Monte carlo localization for mobile robots. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA1999)*, 1999. 16

[27] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita. Laser tracking of human body motion using adaptive shape modeling. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2007)*, pp.602-608, 2007. 16, 42, 110

[28] D.F. Glas, T. Kanda, H. Ishiguro and N. Hagita, "Simultaneous people tracking and localization for social robots using external laser range finders," In Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2009)*, pp.846-853, 2009. 16, 110

[29] F. Faber, M. Bennewitz, C. Eppner, A. Gorog, C. Gonsior, D. Joho, M. Schreiber, and S. Behnke. The humanoid museum tour guide robotinho. In Proc. of IEEE International Symposium on Robot and Human interactive Communication (RO-MAN2009), pp. 891-896, 2009. 16

[30] I. Nourbakhsh, C. Kunz, and T. Willeke. The mobot museum robot installations: A five year experiment. In Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2003),* pp. 3636-3641, 2003. 16

[31] R. Siegwart, K. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux,X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, et al. Roboxat Expo. 02: A large-scale installation of personal robots. *J. of Robotics and Autonomous Systems*, 42(3-4):203–222, 2003. 16

[32] S. Thrun, M. Beetz, M. Bennewitz, W. Burgard, A. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, et al. Probabilistic algorithms and the interactive museum tour-guide robot Minerva. *Int. J. of Robotics Research*, 19(11):972-999, 2000. 16

[33] W. Burgard, A. Cremers, D. Fox, D. H¨ahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Int. J. of Artificial Intelligence* 114(1-2):3-55, 1999. 16

[34] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita. A larger audience, please!: encouraging people to listen to a guide robot. *In Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2010)*, pp. 31-38, 2010. 16

[35] Y. Matsusaka, S. Kubota, T. Tojo, K. Furukawa, and T. Kobayashi. Multi-person conversation robot using multi-modal interface, In *Proc. SCI/ISA*, pp. 450- 455, 1999. 16

[36] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. Tracking multiple people using laser and vision. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS*2005), pp. 1301-1306, 2005. 16, 42

[37] D. Schulz. A probabilistic exemplar approach to combine laser and vision for person tracking. In *Proc. of International Conference on Robotics: Science and Systems (RSS2006)*, 2006. 16

[38] L. Spinello, R. Triebel, and R. Siegwart. Multimodal detection and tracking of pedestrians in urban environments with explicit ground plane extraction. In Proc. of *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2008),* pp. 1823-1829, 2008. 16

[39] A. Green, H. Hüttenrauch, and K. S. Eklundh. Applying the wizard-of-Oz framework to cooperative service discovery and configuration.

In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2004)*, pp. 575- 580, 2004. 16

[40] M. L. Walters, K. Dautenhahn, R. Boekhorst, K. L. Koay, C. Kaouri, S. Woods, C. Nehaniv, D. Lee, and I. Werry. The Influence of Participants' Personality Traits on Personal Spatial Zones in a Human-Robot Interaction Experiment. In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2005)*, pp. 347- 352, 2005. 16, 27

[41] E. A. Topp and H. I. Christensen. Tracking for following and passing persons. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005),* 70-76, 2005. 16

[42] M. Yoda and Y. Shiota. The mobile robot which passes a man. In *Proc. of the IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN1997)*, pp.112-117, 1997. 16

[43] H. Sacks, E.A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, vol.50, no.4, pp.696-735, 1974. 17

[44] G.H. Lerner. Selecting next speaker: The context-sensitive operation of a context-free organization. *J. of Language in Society*, vol.32, no.2, pp.177-201, 2003. 17

[45] D. McNeill, *Psycholinguistics: a new approach,* 1987. 17

[46] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita. Humanlike conversation with gestures and verbal cues based on a three-layer attention-drawing model. *Connection science*, 18(4), pp. 379-402, 2006. 17

[47] T. P. Spexard, S. Li, B. Wrede, M. Hanheide, E. A. Topp, and H. Hutenrauch. Interaction Awareness for Joint Environment Exploration. In *Proc. of International Symposium on Robot and*

Human Interactive Communication (RO-MAN2007), pp. 546--551, 2007. 17

[48] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G.A. Fink, and G. Sagerer. Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot. In *Proc. of the International Conference on Multimodal Interfaces (ICMI2003)*, pp. 28-35, 2003. 17

[49] H. Okuno, K. Nakadai, and H. Kitano. Social interaction of humanoid robot based on audio-visual tracking. In *Proc. of the International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE2002)*, pp. 725-735, 2002. 17

[50] Y. Matsusaka, S. Fujie, and T. Kobayashi. Modeling of conversational strategy for the robot participating in the group conversation. In *Proc. of the European Conference on Speech Communication and Technology*, pp. 2173-2176, 2001. 18

[51] K. R. Th´orisson. Natural turn-taking needs no manual: Computational theory and model, from perception to action. In *Multimodality in Language and Speech Systems*, Vol. 19 pp. 173-207, 2002. 18

[52] K. Nakadai, K. Hidai, H. Mizoguchi, H. G. Okuno, and H. Kitano. Real-time auditory and visual multiple-object tracking for robots. In *Proc. International Joint Conference on Artificial Intelligence*, pp. 1425–1432, 2001. 18

[53] M. Nieuwenhuisen, J. Gaspers, O. Tischler, and S. Behnke. Intuitive multimodal interaction and predictable behavior for the museum tour guide robot robotinho. In *Proc. of IEEE-RAS International Conference on Humanoid Robots (Humanoides2010)*, pp. 653-658, 2010. 18

[54] M. Shiomi, T. Kanda, S. Koizumi, H. Ishiguro, and N. Hagita. Group attention control for communication robots with wizard of OZ approach. *In Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2007)*, pp.121-128, 2007. 18

[55] C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati. Active vision systems for sociable robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 31(5):443.453, 2001. 18

[56] S. Li, M. Kleinehagenbrock, J. Fritsch, B. Wrede, and G. Sagerer. BIRON, let me show you something: Evaluating the interaction with a robot companion. In *Proc. of the IEEE International Conference on Systems, Man, and Cybernetics (SMC2004)*, pp. 2827-2834, 2004. 18

[57] O. Rogalla, M. Ehrenmann, R. Z¨ollner, R. Becher, and R. Dillmann. Using gesture and speech control for commanding a robot assistant. In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2002)*, pp. 454-459, 2002. 18

[58] R. Stiefelhagen, C. F¨ugen, P. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel. Natural human-robot interaction using speech, head pose and gestures. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2004)*, vol. 3, pp. 2422-2427, 2004. 18

[59] T. Tojo, Y. Matsusaka, T. Ishii, and T. Kobayashi. A conversational robot utilizing facial and body expressions. In *Proc. of IEEE International Conference on Systems, Man, and Cybernetics (SMC2000)*, pp. 858-863, 2000. 18

[60] R. Bischoff and V. Graefe. Dependable multimodal communication and interaction with robotic assistants. In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2002)*, pp. 300-305, 2002. 18

[61] H. Kuzuoka, S. Oyama, K. Yamazaki, K. Suzuki and M. Mitsuishi. Gesture Man: A mobile robot that embodies a remote instructor's actions. In *Proc. of ACM International Conference on Computer-supported cooperative work (CSCW2000)*, pp. 155-162, 2000. 18

[62] B. M. Scassellati, *Foundations for a theory of mind for a humanoid robot*, ed: Massachusetts Institute of Technology, 2001. 18

[63] A. E. Scheflen and A. Scheflen. *Body language and the social order – communication as behavioral control*. Englewood Cliffs, N.J.: Prentice-Hall, Inc, 1972. 18

[64] C. Goodwin. *Conversational organization: Interaction between speakers and hearers*. Academic Press, 1981. 19

[65] C. Breazeal, C. D. Kidd, A. L. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2005)*, pp. 383-388, 2005. 19, 20

[66] B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: modeling and evaluation of human-like gaze behavior. In *Proc. of IEEE-RAS International Conference on Humanoid Robots (Humanoids2006)*, pp. 518-523, 2006. 19

[67] 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, 2004, pp. 181–186, 2004. 19

[68] 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 *Proc. of ACM/IEEE International Conference on Human-Robot Interaction* (HRI2009), pp. 61–68, 2009. 19

[69] C. L. Sidner, C. Lee, C. D. Kidd, and C. Rich. Explorations in engagement for humans and robots. *Artificial Intelligence,* vol.*166*, issues 1-2, 140-164, 2005. 19

[70] C. L. Sidner, C. Lee, L. P. Morency, and C. Forlines. The effect of head-nod recognition in human-robot conversation, In *Proc. of ACM SIGCHI/SIGART International Conference on Human-Robot Interaction (HRI2006)*, pp. 290-296, 2006. 19

[71] C. Breazeal. Emotion and sociable humanoid robots. *Int. J. of Human-Computer Studies*, vol. 59, pp.119-155, 2003. 19

[72] J. Schulte, C. Rosenberg, and S. Thrun. Spontaneous short-term interaction with mobile robots in public places. In *Proc. of the IEEE International Conference on Robotics & Automation (ICRA1999)*, vol. 1, pp. 658-663, 1999. 19

[73] L. Mayor, B. Jensen, A. Lorotte, and R. Siegwart. Improving the expressiveness of mobile robots. In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN2002)*, pp. 325-330, 2002. 19

[74] C. Moore, P. J. Dunham, and P. Dunham (Ed.). *Joint Attention: Its origins and role in development*. Lawrence Erlbaum Associates, 1995. 19

[75] H. Kozima and H. Yano. A robot that learns to communicate with human caregivers. *In Proc. of International Workshop on Epigenetic Robotics*, 2001. 20

[76] Y. Nagai, K. Hosoda, A. Morita, and M. Asada. A constructive model for the development of joint attention. *Connection Science*, vol. 15, no. 4, pp. 211– 229, 2003. 20

[77] B. Scassellati. Theory of mind for a humanoid robot. In *Proc. of IEEE/RSJ International Conference on Humanoid Robots*, pp. CD–ROM, 2000. 20

[78] E. Wang, C. Lignos, A. Vatsal, and B. Scassellati. Effects of head movement on perceptions of humanoid robot behavior, In *Proc. of ACM SIGCHI/SIGART International Conference on Human-Robot Interaction (HRI2006)*, pp. 180-185, 2006. 20

[79] Shiomi, M., Kanda, T., Ishiguro, H., and Hagita, N. Interactive humanoid robots for a science museum. In *Proc. of ACM SIGCHI/SIGART International Conference on Human-Robot Interaction (HRI2006)*, pp. 305-312, 2006. 20

[80] Y. Kuno, H. Sokiguchi, T. Tsubota, S. Moriyama, K. Yamazaki, and A. Yamazaki. Museum guide robot with communicative head motion. In *Proc. of IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2006)*, pp. 33-38, 2006. 20

[81] Y. Kuno, K. Sadazuka, M. Kawashima, K. Yamazaki, A. Yamazaki, and H. Kuzuoka. Museum guide robot based on sociological interaction analysis. In *Proc. of Computer/Human Interaction (CHI2007)*, pp.1191-1994, 2007. 20

[82] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proc. of ACM Computer/Human Interaction (CHI2008)*, pp. 131-140, 2008. 20

[83] A. Yamazaki, K. Yamazaki, T. Ohyama, Y. Kobayashi, and Y. Kuno. A techno-sociological solution for designing a museum guide robot: regarding choosing an appropriate visitor. In *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2012)*, pp.309-316, 2012. 20

[84] Y. I. Nakano and R. Ishii. Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In *Proc. of International Conference on Intelligent User Interfaces*, pp. 139-148, 2010. 21

[85] C. Rich, B. Ponsler, A. Holroyd and C. L. Sidner. Recognizing engagement in human-robot interaction. In *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2010)*, pp. 375-382, 2010. 21

[86] G. Jacucci, A. Morrison, G. T. Richard, J. Kleimola, P. Peltonen, L. Parisi, and T. Laitinen. Worlds of information: designing for engagement at a public multi-touch display. In *Proc. of Computer/Human Interaction (CHI2010)*, pp. 2267-2276, 2010. 21

[87] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko. It's Mine, Don't Touch!: interactions at a large multi-touch display in a city centre. In *Proc. of ACM Computer/Human Interaction (CHI2008),* pp.1285-1294, 2008. 21

[88] X. Cao, S. E. Lindley, J. Helmes, and A. Sellen. Telling the whole story: anticipation, inspiration and reputation in a field deployment of Tell Table. In *Proc. of ACM Conference on Computer Supported Cooperative Work (CSCW2010)*, pp. 251-260, 2010. 21

[89] E. Hornecker. "I don't understand it either, but it is cool" Visitor interactions with a multi-touch table in a museum. In *Proc. of IEEE international workshop on tabletop '08*, 113-120, 2008. 21

[90] D. Lloyd, S. Benford, and C. Greenhalgh. Formations: explicit group support in collaborative virtual environments. In *Proc. of Symposium on Virtual reality Software and Technology (VRST1999),* pp.162-163, 1999. 22

[91] N. Yamashita, K. Hirata, S. Aoyagi, H. Kuzuoka, and Y. Harada. Impact of seating positions on group video communication. In *Proc. of ACM Conference on Computer Supported Cooperative Work (CSCW2008)*, pp.177-186, 2008. 22

[92] H. Suzuki, and H. Kato. Interaction-level support for collaborative learning: AlgoBlock—an open programming language. In *Proc. of International Conference on Computer Support for Collaborative Learning (CSCL1995)*, pp.349-355, 1995. 22

[93] E. Hornecker. A design theme for tangible interaction: embodied facilitation. In *Proc. of European Conference on Computer-Supported Cooperative Work (ECSCW2005)*, pp.23-43, 2005. 22

[94] C. Morrison, M. Jones, A. Blackwell, A. Vulysteke. Electronic patient record use during ward rounds: a Qualitative Study of Interaction Between Medical Staff. *Critical Care,* 12 (6), 2008. 22

[95] M. Argyle and J. Dean. Eye-contact, distance and affiliation. *Sociometry*, vol.28, no.3, pp. 289–304, 1965. 22

[96] A. G. Brooks and R. C. Arkin. Behavioral overlays for non-verbal communication expression on a humanoid robot. *Autonomous Robots*, vol.22, no.1, pp. 55–74, 2007. 22

[97] M. L. Walters. The design space for robot appearance and behaviour for social robot companions. In *Ph.D. Thesis*, University of Hertfordshire, 2008. 23

[98] A. E. Scheflen. *Micro-Territories in Human Interaction*. Mouton, pp. 159–173, 1975. 23

[99] C. Breazeal, *Designing Sociable Robots*. MIT Press, 2002. 23

[100] E. A. Sisbot, R. Alami, T. Simeon, K. Dautenhahn, M. Walters, S. Woods, K. L. Koay, and C. Nehaniv. Navigation in the Presence of Humans. In *Proc. of IEEE-RAS International Conference on Humanoid Robots (Humanoids2006),* pp. 181-188, 2005. 23

[101] 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 *Proc. of IEEE International*

Symposium on Robot and Human Interactive Communication (RO-MAN 2008), pp. 707-712, 2008. 23

[102] T. Tasaki, S. Matsumoto, H. Ohba, M. Toda, K. Komatani, T. Ogata, and H. G. Okuno. Dynamic communication of humanoid robot with multiple people based on interaction distance. In *Proc. of IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2004),* pp. 71-76, 2004. 23

[103] R. Gockley, J. Forlizzi, and R. Simmons. Natural person following behavior for social robots. *In Proc. of ACM/IEEE International Conference on Human-robot interaction (HRI2007)*, pp. 17-24, 2007. 23

[104] P. H. Kahn, Jr, N. G. Freier, T. Kanda, H. Ishiguro, J. H. Ruckert, R. L. Severson, and S. K. Kane. Design patterns for sociality in human robot interaction, *In Proc. of ACM/IEEE International Conference on Human-robot interaction (HRI2008),* pp. 97-104, 2008. 23

[105] R. L. Birdwhistell. *Introduction to kinesics: an annotation system for analysis of motion and gesture*. Louisville, Ky.: University of Louisville Press, 1952. 23

[106] E. Goffman. *Interaction Ritual - Essays on face-to-face behavior*. New York: Anchor Books, 1967. 23

[107] M. P. Michalowski, S. Sabanovic, and R. Simmons. A Spatial model of engagement for a social robot. In *Proc. of IEEE International Workshop on Advanced Motion Control*, pp. 762-767, 2006. 24

[108] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki. Reconfiguring spatial formation arrangement by robot body orientation. In *Proc. of ACM/IEEE International Conference on Human-Robot Interaction (HRI2010)*, pp. 285-292, 2010. 24

[109] J. T. Butler, and A. Agah. Psychological effects of behavior patterns of a mobile personal robot. *Autonomous Robots*, vol. 10, pp. 185-202, 2001. 24

[110] E. Pacchierotti, H. I. Christensen, P. Jensfelt. Human-robot embodied interaction in hallway settings: A Pilot Study. In *Proc. of IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN2005)*, pp. 164-171, 2005. 24

[111] K. L. Koay, D. S. Syrdal, M. L. Walters, K. Dautenhahn. Living with robots: investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *Proc. of IEEE International Conference on Robot and Human Interactive Communication (RO-MAN2007)*, pp. 564-569, 2007. 25

[112] J. K. Burgoon, S. B. Jones. Toward a theory of personal space expectations and their violations. *Human Communication Research*, vol. 2(2). pp. 131-146, 1976. 25

[113] D. L. Gillespie, A. Leffler. Theories of Non-Verbal Behavior: A critical review of proxemics research. *Sociological Theory*, vol.1, pp. 120-154, 1983. 25

[114] T. Shibata, K. Wada, K. Tanie. Subjective evaluation of a seal robot in the National Museum of Science and Technology in Stockholm. In *Proc. of IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2003)*, pp. 397-402, 2003. 27

[115] M. L. Patterson, "Spatial factors in social interactions," *Human Relations*, vol. 21, pp. 351-361, 1968. 27

[116] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. te Boekhorst, C. Nehaniv, I. Werry, and D. Lee. Close encounters: Spatial distances between people and a robot of mechanistic appearance. In *Proc. of IEEE-RAS International Conference on Humanoid Robots*, pp. 450–455, 2005. 27

[117] R. P. McDermott and D. R. Roth. The Social organization of behavior: interactional approaches. *Annual Review of Anthropology*, vol.7 , no.3, pp.321-45, 1978. 28

[118] E. Goffman. *Behavior in public places: Notes on the social organization of gatherings*. Free Press, 1963. 28, 110

[119] E. A. Schegloff. Body torque. *Social Research*, vol.65, no.3, pp. 535–596, 1998. 29

[120] A. Kendon. The F-formation System: The Spatial organization of social encounters. *Ma-Environment Systems*, vol. 6, pp. 291-296, 1976. 30

[121] A. Kendon. Spatial Organization in Social Encounters: The F-formation System. *In studies the Behavior of Social Interaction*, Peter de Ridder Press, 1977. 30

[122] A. E. Scheflen and N. Ashcraft. *Human Territories- How we behave in space-time*. Englewood Cliffs, N.J.: Prentice-Hall, 1976. 30

[123] D. McNeill, ed. Gesture, gaze, and ground. *Lecture notes in computer science*, *Springer-Verlag*, pp. 1–14, 2005. 31

[124] M. Bono. Units of analysis for understanding conversational structure— f-formation —. *J. of Japanese Society for Artificial Intelligence*, vol.23, no.4, pp. 545–551, 2008. 31

[125] M. Argyle. *Social interaction*. Paperback, London: Tavistock Publications Ltd, 1973. 37

[126] R. Meisner, D. V. Lehn, C. Heath, A. Burch, B. Gammon, and M. Reisman. Exhibiting performance: co-participation in science centers and museums. *Int. J. of Science Education*, vol.29 (12), pp. 1531-1555, 2007. 37

[127] C. Heath and D. V. Lehn. Construing interactivity: enhancing engagement with new technologies in science centers and museums. *Social Studies of Science,* vol. 38, pp. 63-96, 2008. 37

[128] H. Brignull, and Y. Rogers. Enticing people to interact with large public displays in public spaces. In *Proc. of INTERACT*, pp. 17-24, 2003. 37

[129] E. Hornecker. Interactions around a contextually embedded system. In *Proc. of ACM international Conference on tangible and Embedded Interaction (TEI2010)*, pp.169-176, 2010. 38

[130] T. Yonezawa, H. Yamazoe, A. Utsumi, and S. Abe. Gaze Roboard: Gaze-communicative guide system in daily life on stuffed-toy robot with interactive display board. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2008)*, pp.1204-1209, 2008. 38

[131] J. Cui, H. Zha, H. Zhao and R. Shibasaki, "Fusion of detection and matching based approaches for laser based multiple people tracking, In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2006)*, vol.1, pp.642-649, 2006. 42

[132] S. Birchfield. Elliptical head tracking using intensity gradients and color histograms. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, p. 232– 237, 1998. 42

[133] T. Chateau, V. G. Belille, F. Chausse, and J. T. Lapreste. Real time tracking with classifiers. In *Proc. of International Workshop on Dynamical Vision in conjunction with ECCV*, pp. 218-231, 2006. 42

[134] M. Isard, and A. Blake. Condensation – conditional density propagation for visual tracking. *Int. J. of Computer Vision*, vol.29, no.1, pp.5-28, 1998. 42

[135] G. Loy, L. Fletcher, N. Apostoloff, and A. Zelinsky. An adaptive fusion architecture for target tracking. In *Proc. of International Conference on Automatic Face and Gesture Recognition*, pages 261–266, 2002. 42

[136] K. Nickel, T. Gehrig, R. Stiefelhagen, and J. McDonough. A joint particle filter for audio-visual speaker tracking. In *Proc. of International Conference on Multimodal Interfaces*, pages 61–68, 2005. 42

[137] K. Nummiaro, E. Koller-Meier, and L. V. Gool. An adaptive color-based particle filter. *Image and Vision Computing*, vol. 21(1), pp.99–110, 2003. 42

[138] K. Okuma, A. Taleghani, N. De Freitas, J. J. Little, and D. G. Lowe. A boosted particle filter: Multi target detection and tracking. In *Proc. of European Conference on Computer Vision*, pages 28–39, 2004. 42

[139] P. Prez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. In *Proc. of the IEEE*, vol. 92(3), pp.495–513, 2004. 42

[140] J. Sherrah and S. Gong. Fusion of perceptual cues for robust tracking of head pose and position. *Pattern Recognition*, vol. 34(8), pp.1565–1572, 2001. 42

[141] A. Sugimoto, K. Yachi, and T. Matsuyama. Tracking human heads based on interaction between hypotheses with certainty. In *Proc. of Scandinavian Conference on Image Analysis*, pages 617–624, 2003. 42

[142] Y. D. Wang, J. K. Wu, and A. A. Kassim. Particle filter for visual tracking using multiple cameras. In *Proc. of IAPR Conference on Machine Vision Applications*, pages 298– 301, 2005. 42

[143] C. Yang, R. Duraiswami, and L. Davis. Fast multiple object tracking via a hierarchical particle filter. In *Proc. of International Conference on Computer Vision and Pattern Recognition*, pages. 212–219, 2005. 42

[144] P. Viola, and M. Jones. *Rapid Object Detection using a Boosted Cascade of Simple Features.* In *Proc. of IEEE Computer Society*

Conference on Computer Vision and Pattern Recognition (CVPR2001), vol. 1, pp.511-518 (2001). 44, 48, 118

[145] Y. Kobayashi, D. Sugimura, Y. Sato, K. Hirasawa, N. Suzuki, H. Kage, and A. Sugimoto. 3D head tracking using the particle filter with cascaded classifiers. In *Proc. of British Machine Vision Conference*, pp. 37-46, 2006. 49, 118

[146] C. Goodwin and M. H. Goodwin. Participation. In *A Companion to Linguistic Anthropology*, edited by D. A. Maldan, MA, Blackwell, pp. 222- 244, 2004. 82

[147] E. Pacchierotti, H. I. Christensen, and P. Jansfelt. Evaluation of passing distance for social robots. In *International Symposium on Robot and Human Interactive Communication (RO-MAN2006)*, pp. 315-320, 2006. 84

[148] H. H. Clark. *Using Language*, Cambridge University Press, 1996. 110

[149] Y. Katagiri, M. Bono, and N. Suzuki. Conversational inverse information for context-based retrieval of personal experiences. *Lecture Notes in Computer Science*, vol. 4012, pp. 365-376, 2006. 110

[150] K. Dautenhahn, M. walters, S. Woods, K. L. Koay, C. L. Nehaniv, A. Sisbot, R. Alami, and T. Simeon. How may I serve you?: a robot companion approaching a seated person in a helping context. In *Proc. of the ACM SIGCHI/SIGART International Conference on Human-robot interaction (HRI2006)*, pp. 172–179, 2006. 110

[151] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, and N. Hagita. Humanoid robots as a passive-social medium: a field experiment at a train station. In *Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI2007)*, pp. 137-144, 2007. 110

[152] N. Bergström, T. Kanda, T. Miyashita, H. Ishiguro, and N. Hagita. Modeling of natural human-robot encounters. In *Proc. International Conference Intelligent Robots and Systems (IROS2008),* pp.2623-2629*,* 2008. 110

[153] T. M. Ciolek, A. Kendon. Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry*, vol. 50, pp. 237–271, 1980. 110

[154] Y. Kobayashi and Y. Kuno. People tracking using integrated sensors for human robot interaction. In *Proc. International Conference on Industrial Technology (ICIT2010)*, pp. 1597-1602, 2010. 110

[155] K. Nakamura, H. Zhao, R. Shibasaki, K. Sakamoto, and N. Suzukawa. Visualizing passenger flow in railway station using laser scanners. In *Proc. of International Conference on Computers in Urban Management and Urban Planning*, Ref 220, 2006. 110

[156] M. Montemerlo, W. Whittaker, and S. Thrun. Conditional particle filters for simultaneous mobile robot localization and people-tracking. In *Proc. of International Conference on Robotics and Automation (ICRA2002)*, pp. 695-701, 2002. 110

[157] Lozano and K. Otsuka. Real-time visual tracker by stream processing. J. Signal Processing Systems, vo. 57(2), pp. 285-295, 2009. 114

[158] A. Kendon. Features of the structural analysis of human communicational behavior. In *Walburga von Raffler Engel, ed., Aspects of Nonverbal Communication.* pp. 29-43, 1980. 119