Modeling of Face-to-Face Behavior in Scenario-based Interaction among Humans and Robots

場面別のインタラクションにおける人およびロボットとの対面行動モデルに関する研究

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Human Informatics

The University of Tsukuba

University of Tsukuba
筑波大学

Ph.D Program in
EMPOWERMENT INFORMATICS

Yadong Pan

March 2017

Acknowledgement

First of all, I would like to thank my supervisor Prof. Kenji Suzuki for his great guidance and support throughout my Ph.D studies. He is also a wonderful teacher in sharing extraordinary human qualities throughout daily life. It's a real honour for me to be a member in his laboratory, the Artificial Intelligence Laboratory in University of Tsukuba.

I would like to thank my lab-colleagues for their kind participation in the experiments, as well as their cooperation in development of useful technologies for my research. I have gained so much knowledge and power from all those excellent people, which supports me in overcoming numerous obstacles I have been facing through my research.

I would like to thank Empowerment Informatics Program in University of Tsukuba, one of MEXT programs for Leading Graduate Schools, which educates me to really improve my interdisciplinary ability, presentation ability, and frontline ability. I am so much grateful to the faculty members in this program for their great effort in arranging interdisciplinary courses, research internship, and guidance on job-hunting.

Nevertheless, I am also grateful to all groups in cooperation with my researches including Hankyu-Hanshin Daiichi Hotel Group, as well as Aichi Prefectural Colony and Department of Psychology in Keio University. I have been inspired a lot by sharing works with them.

Last but not the least, I would like to thank my dear parents, and my beloved wife Xia Yun for the greatest love. All the support they have provided me over the years is the best gift anyone has ever given me. "Having you, life is always my treasure."

Abstract

This dissertation summarizes a study on face-to-face interaction. The core issue is to create model of facing and face-to-face behavior (mutual action of facing) according to scenarios. The model of facing is studied based on two aspects of information: (i) geometric information of position of head as well as head-orientation, (ii) human perception during having social interaction. Experiments for modeling face-to-face are conducted under non-instructed and instructed human-human interaction, with subjects' face and gaze detected by using wearable devices. For each subject, a probability density chart is generated which shows how he/she distributes head-orientation during looking at others' face. The feature of all subjects' behavior under each specific scenario is fitted with Gaussian distribution. Based on the Gaussians, several ranges of head orientation could be calculated, which contribute a majority to visual focus of attention. Such head-gaze mapping provides theoretical support to the development of several automatic systems that measure face-to-face interaction based on head-orientation. Those systems can be used to create applications to human-robot and human-human interaction. In case of human-robot interaction, I focus on robot-assisted therapy for children with autism spectrum disorder, and social robots that provide information service in a hotel's public space. Similar therapeutic activity is conducted for describing face-to-face in human-human interaction. Throughout this study, it is suggested that face-to-face is performed in different ways according to scenarios, and face-to-face - as a fundamental social behavior - can be used to describe social interaction.

List of Figures

List of Tables

Contents

Chapter 1
Introduction

1.1 Motivations

Starting with an interest in study of face-to-face[1] - a phenomenon that covers various disciplines - we shall be concerned at the beginning with the outline of such research scope. One of the biggest challenges is to identify the behavioral aspects that can be measured in the sense of underlying motivation. An early attempt by Eliot Chapple indicated individuals perform with certain patterns and characteristics in their periods of activities so called face-to-face interactions (Chapple and Coon, 1942). Through a further research, Poyatos summarized the disciplinary aspects of face-to-face interaction in a triple structure language-paralanguage-kinesics, which attempts to suggest the full complexity of the face-to-face research scope (Poyatos, 1983). Those two works - in terms of research motivation - stand surely as the leading attempts for this dissertation.

Regarding definition of face-to-face behavior/interaction[2], theoretically there exist descriptions rather than definitions in the varied disciplines. In sociology, face-to-face interaction is one of the basic elements of the social system, forming a significant part of individual socialization and experience gaining throughout one's life time (Kendon et al, 1975). Many theorists infer that in human society, face-to-face could be considered as the top standard that establishes social interaction (Nardi and Whittaker, 2002). On one hand, face-to-face behavior engages more human senses than mediated communication (Schement and Ruben, 1993), which indicates the reason why sociology scholars often take into account face-to-face as an averseness of mediated communication. On the other hand,

face-to-face is one of the most effective forms in verbal communication, through which people could persuade or motivate others (Mills et al, 2006). This can be explained also through Poyatos' triple structure (1983).

Face-to-face behavior in an element of body-language. Body-language - also known as kinesics - functions as reflecting human's inner aspects through physical movements. For example, facial expressions and body postures have been considered as the common approaches which are indicative of interpreting emotions (Kret et al. 2011; Mondloch, Nelson and Horner, 2013). Oculesics including eye movement, gaze and eye-related nonverbal communication is a high level for delivering feelings (Sullivan, 2009). Proxemics categorizes social relationships according to measurable spatial distances (Hall, 1966; Lambert, 2004). Whereas, very few contributions have been focusing on the relation between physical features of face-to-face behavior and its social functions.

Based on the literatures, in this dissertation, face-to-face is described as a behavior/interaction that is both geometric and perceptual. To further explain such description, I pick up the one-way behavior called facing from the mutual face-to-face behavior, and use Figure 1 to show the meaning. If two persons are facing each other, it is called face-to-face. To conduct researches, first, it is necessary to see how such behavior is performed by individuals in terms of measurable and geometric body movements, as an extension to Eliot Chapple's early attempt where face-to-face was a general concept in social activities (Chapple and Coon, 1942). Second, face-to-face should be studies based on perception in social interaction in order to reach the discussions on human's inner aspects.

1: Face in this dissertation is defined as front of the human head or humanoid robot head, which could be perceived as a cue for interaction.
2: The two expressions are unified in this dissertation, because the behavior talked about is mutual and with perceptual aspects, compared with individuals' physical body-movement.

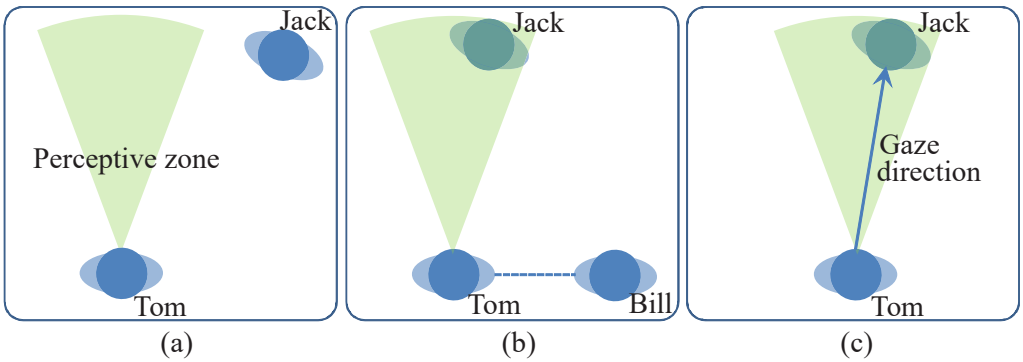

Figure 1. Description of the behavior "facing"
(a): Jack is located at a position which is away from the area where Tom could perceive based on his current head orientation. (b): Although Jack entered Tom's perceptive area, Tom does not really have a interaction with Jack, but with another person Bill. (e.g. Tom talks with Bill via telephone). (c): Tom interacts with Jack by making eye-contact. In such case, we say Tom is facing Jack.
In all cases, Tom' the perceptive area is a cone-shaped space around his head-orientation and with a scenario-based semi-vertex angle. The angle is discussed in the later sections in this dissertation.

1.2 Objectives

As mentioned in the previous section, the study of face-to-face covers various disciplines. Even we are going to consider it as a physical and perceptual behavior rather than a general social phenomenon, the expression of such behavior still varies according to different kinds of interaction. Think about the following two circumstances: (i) You make a behavior initiatively; (ii) You are asked to make the behavior. Under such two circumstances, your behavior might not be with the same physical expressions. Similarly, in case of (i) you see a person makes a behavior, and (ii) you see a robot makes the behavior, your response might be different. To summarize those kinds of different circumstances or situations, the concept called scenario is included in this research.

Scenarios are the descriptions that distinguish different subjects, their background information, the environmental and contextual conditions where they act, sequences of their actions, their targets and goals (Go and Carroll, 2004a). Considering scenarios is an effective way in the designing process throughout various disciplines such as human-computer interaction, engineering of requirements, object-oriented design, and strategic planning (Go and Carroll. 2004b; Jarke, Bui and Carroll, 1998). In social psychology researches, it is indicated that although individuals' uncertain and distributing perceptions might be logically summarized into stability, sometimes people adjust their perception to face group instability (Schon, 1967; Nisbett and Wilson. 1997). To fulfill such gap in task analysis, scenario has been used as a powerful tool.

Regarding face-to-face interaction, I summarized some scenarios in Figure 2. We see face-to-face happen in human-human, human-robot, and robot-robot interaction; We see face-to-face happen in different places such as home and public spaces; It could also happen within different spatial zones (e.g. personal zone which represents the distance from 0.45m to 1.2m, and social zone which is from 1.2m to 3.6m) as Hall defined in his work (1966). In this dissertation, the face-to-face study covers most of the scenarios shown in Figure 2. Although it is quite difficult to conduct comparative study among different scenarios due to different environment or context of the interaction, the study could reach some important objectives:

- Objective 1: Modeling face-to-face based on human-human interaction

To model face-to-face, we first need to model the one-way facing. The model of facing in this research is a probability density map that shows how people distribute head orientation during facing someone. The model is investigated under non-instructed and instructed human-human interactions, where human might have different awareness and therefore different body movement when acting with the same behavior.

Experiments for modeling of face-to-face are conducted under human-human interaction. The feature of face-to-face in human-human interaction could be used as important reference for human-robot interaction, because (i) Analyzing the patterns on how human behave natural behavior in front of robots could generate knowledge to understand robots as social agent. (ii) Human's response would be activated when robots could perform and perceive human behavior as same as human (Eyssel et al, 2012).

- Objective 2: Applying the model in the development of face-to-face measuring systems

Camera based and infra-red based systems could be developed by referring to the proposed model. Each system should have functions that adjusting the criteria for the judgment of face-to-face according to different scenarios. What's more, comparative study between system's and human's judgment on the same targets should be conducted. This would help to fulfill the gap between geometric correct and perceptual correct.

- Objective 3: Measuring face-to-face under different scenarios

A direct target of measuring face-to-face is to provide data support to the research on behavior science. Especially for those people who have difficulty in having social interaction, measuring their face-to-face behavior might be helpful for perceiving their state-of-health (Ono et al, 2012).

From another perspective, since it is difficult to generate a constant general model of face-to-face across scenarios, it is with highly importance to obtain feedback to a model when it is applied under different scenarios, which could be used to further understand and enhance the model.

Figure 2. Different scenarios of face-to-face

1.3 Contributions

1.3.1 Contribution to the research filed of visual focus of attention

The model of face-to-face in this dissertation considered human face as the visual target, which contributes to a specific aspect of the researches on visual focus of attention (VFOA). Recent works on VFOA suggested some factors that may affect the relation between head-orientation and VFOA, such as density of visual targets (Ba and Odobez, 2009), previous status of head pose and gaze (Sheikhi and Odobez, 2012), and sound information (Stiefelhagen, Yang, and Waibel, 2001). However, in those works, VFOA was not really based on subjects' awareness in interaction with others, and the study on scenarios was missing. In this dissertation, face-to-face is with more specific VFOA which is human face. This could connect the behavior analysis to the awareness of having social interaction. What's more, throughout experiments with human subjects, I take into account a contextual factor - instructed/non-instructed scenario - which could affect the relation between head-orientation and looking-at-face.

In the modeling of face-to-face, I separated the non-instructed and instructed scenario to conduct experiments. It is important to see the features of behavior under each specific scenario, because those two scenarios are considered to be used for different applications. For example, non-instructed scenario could be planned in public spaces during people's social activities (e.g. ball games), whereas instructed scenario might be useful as a way of training social orienting capability for children with developed disorders (e.g. children with autism spectrum disorder). The behavior analysis results in this research could provide theoretical and data supports to the design of those applications.

The experiment results under human-human interaction in this research is similar to a previous human-computer interaction research conducted by Fang et al (2013), in which the author investigated how human subjects distribute their head orientation during watching ultra high definition television. Such comparison could help to extend the understanding of VFOA into a wider scope.

1.3.2 Contribution on system engineering

To evaluate the particular use of the proposed face-to-face model, it is applied in the design of several systems including camera-based and infra-red (IR) based systems. Those systems automatically analyze face-to-face behavior with much higher speed than doing the analysis by human. What's more, the system can generate probability density of head orientation in any certain range. The data of face-to-face event on the time line and the probability density chart of head orientation are supposed to be used by psychologists and therapists to save workload on processing videos.

Each system also has its certain benefit for practical use. For example, the camera-based systems using robot-vision, head-mounted camera (first-person view camera), or 360-degree camera could release heavy instrumentation from users' body. The IR based system is robust against changes on lighting condition and speed of motion. However, under different scenarios, the systems should apply different criteria of face-to-face. This will be discussed in the later sections.

Potential subjects for using the face-to-face measuring systems include children with autism spectrum disorder (ASD) as well as neurotypical people. For children with ASD, features (e.g. time, cue) of their face-to-face behavior in specific therapeutic activities could help therapists to understand the children's affective response in social interactions. It would be also very interesting to investigate how their face-to-face behavior relates to other social behavior (e.g. smile, hand touch) by using the systems in this dissertation together with other behavior computing technologies.

Besides the automatic systems introduced in this dissertation, it is also possible to develop anthropomorphic robots by referring to the proposed model of face-to-face. This could help to enable robots with capability that having intuitive multimodal communication with human beings.

1.3.3 Contribution on body-language

As mentioned in section 1.1, this work is an early attempt that describes face-to-face in both geometric and perceptual way. The geometric way of understanding face-to-face extends the theory in Chapple's work (1942), and the perceptual way - which considers the status that having social interaction as basis - sets up a bridge between physical behavior action and analysis of social interaction, as well as between face-to-face and other behavior. It could therefore enhance the understanding of face-to-face in terms of body language.

1.3.4 Additional contribution: socially assistive robots in public space

The modeling of face-to-face in this dissertation is mainly targeting on interaction in laboratory environment or home space. As a supplementary section, I analyzed the multimodal interaction consists of face-to-face behavior and verbal information in a hotel public space using humanoid robot as agent. The robot provides information service to the hotel guests like what the hotel staff do.

Previous researches already tested the use of robots in station (Hayashi et al. 2007), shopping mall (Kanda et al. 2009), education center (Tanaka, Cicourel, and Movellan. 2007), museum (Thrun et al. 1999) and elderly care center (Sabelli et al. 2011). Those researches have a common feature that people may have different motivations during encountering with robots, and the motivations might not be focusing on information of the public space. I considered that it is important to control such conditions and to find a context that people do require some information related to the public space, so that the human-robot interaction there could be more related to the robots' practical effectiveness. In my research, the hotel public space - where people pass through from entrance of the hotel to the front desk to get information - could provide good opportunity for the study.

From the perspective of face-to-face, in the hotel public space, the robot faces each guest when greeting him/her in order to start direct interaction. Throughout comparative study between direct and indirect interaction (where the robot did not face the guests), we could understand how human's attention is changed by robot's behavior.

Chapter 2
Related Works

2.1 Face-to-Face Behavior

In human-robot interaction, several previous researches have preliminarily investigated whether face-to-face behavior could affect human's attention towards robots. As a positive evidence, Van Breemen et al (2003) reported a work which focused on the development of a robot that interact with human with emotional and body-language feedback. Especially the robot turned to the person who spoke to it by rotating its head and body with smile, no matter what the person said. The experiment results indicated that such "turn-to-speaker" behavior was highly appreciated among various people, and might have effect that favoring the natural interaction between the robot and human. On the contrary, there are also researchers suggest that face-to-face might not be a significant way to trigger human-robot interaction. In Bruce et al's work (2002), the authors presented an experiment in human-robot social interaction with motivation that quantifying the tendency of people in starting initial interaction with a robot. The robot used in the experiment had a humanoid face, and was capable to turn its head to track the person and greet him/her. Throughout analysis on people's interests towards the robot's behavior, the authors suggested that those features did not increase people's tendency on having interaction with robots. In summary of the two introduced studies, regarding face-to-face in human-robot interaction, different results have inferred that the scenario in such experiments should be controlled more in order to gain further knowledge.

In human-human interaction, the description of face-to-face is confused. Most typically, researchers defined face-to-face as an alternative expression for eye-contact, or as a general social phenomenon against mediated interaction. The importance of face-to-face was shown throughout various researching scopes such as infants (Blehar, Lieberman, and Ainsworth. 1977), autism (Dawson et al. 1998), and teamwork strategy (Rocco, 1998). As for understanding face-to-face in a spatial (geometric) way, Kendon is a first challenger through his work (Kendon, 1976) in which he defined F-formation that inspires face-to-face interaction, although Kendon considered individual's interactive orientation as referring to multimodal factors including head orientation, shoulder orientation and feet layouts. Similarly as Kendon's understanding, recently some technologies based on individuals' spatial relation have been developed to measure face-to-face behavior. Cattuto (2010) provided an active radio frequency identification device that assessed mutual proximity in a distributed fashion by exchanging low power radio packets. The device was attached to the upper torso of the subjects. By using the device, Cattuto analyzed the dynamics of face-to-face interaction networks. The results inspired the understanding towards some phenomena driven by face-to-face interactions, such as the spreading of information. Similarly, Otsuka et al (2009) proposed a system called Business Microscope that sensed the activities of people in a community and provided visual feedback to users. Name-tag shaped sensor nodes were used to measure face-to-face interaction. The data collected by the sensor network terminal was processed by the server and displayed in a topographic map that showed the frequencies of activities in the community. By using the system, the authors attempted to improve business performance. What's more, Watanabe et al (2006) presented a head-mounted device called Visual Resonator. The wearer of such interface can hear the voice or auditory information only from the spatial area which he/she is facing, and can send his/her voice only towards the area which he/she is facing. Watanabe's work could be used to better fulfill information exchange through face-to-face. Those systems are technically useful, while the face-to-face behavior they addressed on did not have a clear model of face-to-face that refer to either geometric or perceptual information. In this dissertation, the modeling of face-to-face will fulfill such gap.

2.2 Head-Orientation and VFOA

Regarding the behavior of facing, ideally, a person A's head orientation should directly point to a person B's face when we say A is facing B, as shown in Figure 3(a). However, such ideal condition does not really happen a lot, especially when A is not close to B. It would be quite usual that A's head orientation points to the outer area of B's face like Figure 3(b). In such condition, A could still use gaze-movement to calibrate his attention onto B's face, and feel himself facing B. The factor that determines if A could use a natural gaze-movement to pay attention to B is the angle between his head orientation and the link between A and B's faces. It is considered that the angle should have a range for supporting natural interaction, which means, within the certain range, head orientation could contribute to a majority of visual focus of attention (Stiefelhagen and Zhu. 2002). In other word, using the head pose as a surrogate for gaze is possible (Robertson and Reid. 2006).

In the meeting context, which is a typical context for analyzing VFOA, researchers have been studying the head-gaze mapping based on a geometric model shown in Figure 4. Stiefelhagen et al (2002) used a Hidden Markov Model to describe the sequence of head pose. The limitation of this work was that the head pose which was categorized using k-mean clustering into several discrete values. Stiefelhagen's approach was based on the seat arrangement and scenario of the experiment, where four participants sit around a table and engaged in a conversation. The number of visual targets for each participant was limited according to discrete spatial areas. More recently, with a similar experiment setting, Otsuka et al (2005) proposed a dynamic Bayesian network that recognized the VFOA of people based on the status of head orientation and utterance during the conversation. However, in Otsuka's research, the participants' behavior was captured using an overhead mounted device, and the status of utterance was coded by human coder by looking into the experiment video, which might influence the participants' natural behavior and the precision of behavior analysis.

Several factors may affect the performance of the model. In Ba et al's work (2009), it was indicated that good VFOA recognition could only be achieved if the visual targets of a person were well separated in the head pose angular space. The result was therefore depending on the person’s position in the meeting room. Sheikhi and Odobez (2012)

proposed a more contextual model, which suggested that previous head and gaze status could affect the relation between head pose and VFOA. Assuming that when a person is focusing on a target, when his/her gaze shift happens by incorporating the previous gaze target, the previous head pose estimated through recursion over a short time period would evolve and introduce an evolution of what the head pose mean of the target itself should be. Stiefelhagen, Yang, and Waibel (2001) have demonstrated how focus of attention can be predicted based on knowledge of who is currently speaking, and how this audio-based prediction can be improved by taking the history of utterances into account. In an experiment, it was observed that participants' focus of attention had been predicted correctly in 63% of the frames by using audio information only. By using both head pose and sound, focus of attention could be detected in 76% of the frames in recorded meetings.

In terms of social attention, head orientation is not only dominant when people behave but also an important factor in perceiving others' social behavior. In some early attempts, it was suggested that human perceive another person's attention based on both the person's head orientation and gaze direction, and the actual perception is processed as in middle of those two directions (Cline, 1967; Anstis et al. 1969; Maruyama and Endo, 1983). By looking into the functional activity of human brain, a research by Perrett et al (1992) has indicated that individual cells in the superior temporal sulcus region of the brain could respond to synergy of eye, head and body position. Langton (2000) also suggested that information derived from the head orientation is not completely inhibited when gaze direction is visible. These works provide evidence for head-orientation based methods to model face-to-face interaction, as what we perceive from others may reflect or even affect on how we behave towards others.

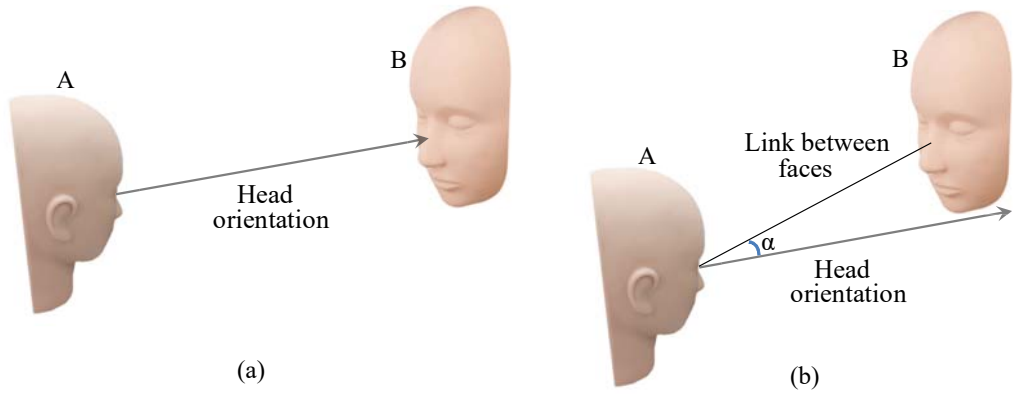

Figure 3. Different patterns of the behavior "facing"
(a): direct facing. (b): the way usually people act in

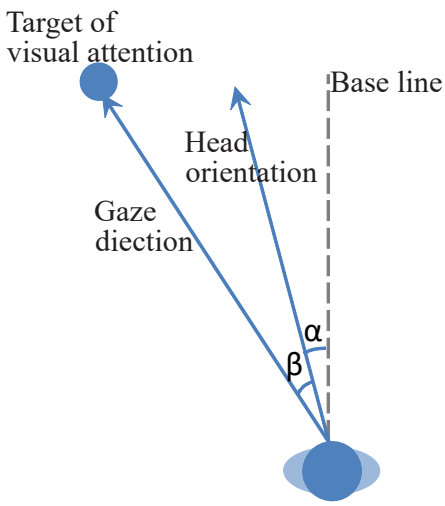

Figure 4. A geometric model of gaze (for the pan angle)
Base line can be body orientation or any user-defined orientation according to the scenario.

2.3 Facing and other Social Behavior

As mentioned in section 1.1, it is necessary to investigate how face-to-face relates to human's inner aspects. One possible way is to study how face-to-face corresponds to other social behavior that could be described with knowledge of emotion. Researchers working on smile have firstly challenged such topic. From the emotion's point of view, researchers suggested that smiles of enjoyment happens while one' head was turning toward another person (Kraut, and Johnston. 1979). In contrast, smiles of embarrassment can be observed as one's gaze and head orientation were moved away from another person (Keltner, 1995). From the physical gesture's point of view, it was also indicated that spontaneous smile tended to occur when one is facing another person (Cohn, et al. 2004).

Regarding the relation between face-to-face and other social behavior, the study is not yet spread conducted, although there has been technologies that can measure social behavior with emotional aspects such as hand shake (Iida and Suzuki. 2011), and hug (Nuñez, Uchida, and Suzuki. 2013). This could be a challenge topic for the next step.

Chapter 3
Modeling of Facing and Face-to-Face

3.1 A Simple Model with Avatars

I conducted an experiment to investigate people's eye-gaze while facing someone. As shown in Figure 5(a), I pasted several pictures of a regular face (online citation a) on the wall (size of each picture was 8.5cm×10.5cm). The arrangement of the pictures was set up following three directions: vertical up, vertical down and horizontal right. There was also a blue mark at the center position. During the experiment, the participant's head was fixed at a position where he/she could straightly face to and keep a 1.2-meter distance from the blue mark. It was ensured that the participants could clearly see the pictures. Each participant was asked to finish the following process:

1) Keep facing to the blue mark without turning head.
2) Use only gaze movement to look at each picture.
3) Answer a question - Do you think it is natural if you pay attention to a person with this pose. (Yes/No) - while looking at each picture.

Since the position of each picture was pre-defined with different gaze-angle (calculated based on 1.2-meter distance between participant's head and the blue mark), by doing this experiment, we could get a threshold of gaze-angle for supporting each participant's behavior that facing the image. 20 participants (12 men and 8 women) participated in this experiment, aged between 22 and 58 (M=31.8, SD=12.4). Each person was asked to look at the pictures using only gaze movement, and to answer the question. During one person's

experiment, other participants were not allowed to enter the experiment space, and were not instructed to know the content of experiment until their own terns. There were totally 3 sessions for each participant. In each session, the participant chose one certain direction (vertical up, vertical down, or horizontal right), and started looking at the pictures on the chosen direction one-by-one from using smaller gaze-angle to bigger one. After the experiment, I recorded his maximum gaze-angle to which he/she answered YES in each session. An overview of all means and standard deviations is shown in Figure 5(b).

The results of the previous experiment revealed that on all the three directions, participants' average acceptance of maximum gaze-angle was within 12 and 15 degrees. People's perception on the other direction - horizontal left - could be considered as equivalent to horizontal right because of the symmetrical structure of the face. According to the experiment results, in this study, I accepted 12 degree as a basic threshold, and created a perceptive zone (Figure 6) to identify the behavior of facing someone. Referring to this perceptive zone, we defined the following behavior:

- One-way facing (person A facing person B): B's face is within A's perceptive zone.
- K-degree face-to-face interaction between person A and B: Each person's face is within the perceptive zone of the other person.

In these definitions, the position of the face is defined as the middle of the two eyes. This position is also the vertex of the perceptive zone. In this dissertation, when we use the term "face someone", it means to face someone's face according to the perceptive zone. The concept of k-degree face-to-face is mainly used within a range from 45cm to 120cm, which is called personal zone according to the proxemics defined by Hall (1966) and Lambert (2004). The threshold of k-degree face-to-face is not fixed. 12-degree face-to-face is considered a basic criterion. However, we need to adjust the threshold according to different scenarios.

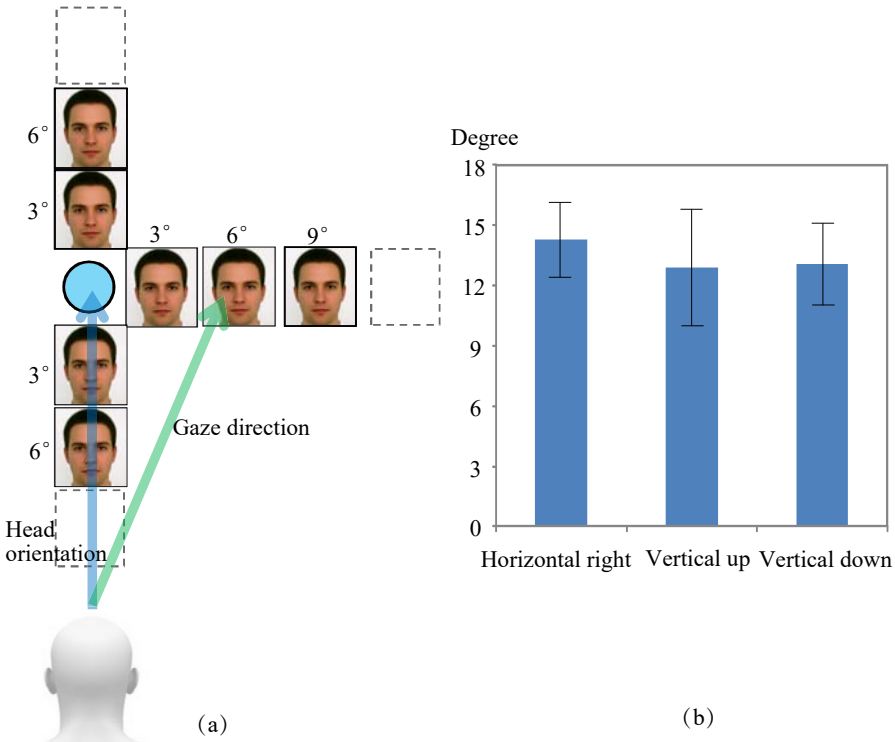

Figure 5. An experiment using avatar faces on the wall.
(a): Experiment design. (b): Average values of natural gaze direction.

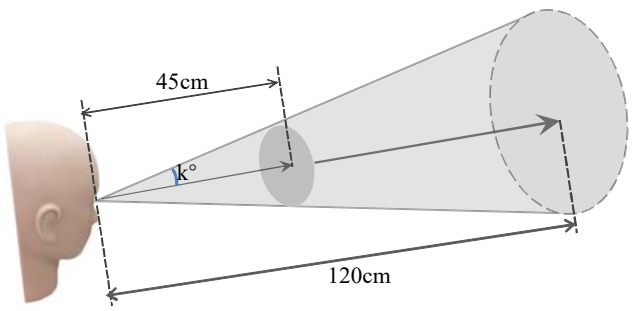

Figure 6. A cone-shaped perceptive zone around one's head orientation

3.2 Detection of Face, Gaze, and Human

3.2.1 A single camera-based method for face and head-pose detection

Referring to the criteria of k-degree facing and face-to-face, I developed a monocular vision-based system to measure face-to-face. Recently researches using monocular vision to obtain 3Dposition have been widely studied (Royer, et al. 2007; Horprasert, et al 1997; Ponsa, et al. 2005). In this research, a monocular camera based routine Face API (by Seeing Machines) was used for developing the system. The system consists of two units: the image-processing unit and the judgment unit. In the image-processing unit, the system detects the face, estimates its position in 3D camera-coordinate system, and estimates the rotation of the face with Tait-Bryan angles (x-y-z convention of Euler angles). The judgment unit computes all the parameters provided by the image-processing unit in order to get the spatial details of the detected face's and an agent's perceptive zones. The agent is defined as a person who wears a camera on his head (e.g. using Google glass), or a humanoid robot that has a camera installed on the head (e.g. the robot NAO).

Figure 7 shows the spatial details of face detection. *F'* represents the center of agent's face which is predefined related to camera *O*. Point *F(Fx, Fy, Fz)* is the position of detected face. It has a normal line *Ln* to camera's *XOY*-plane. The head-orientation line *Lf* can be obtained by rotating *Ln* with Tait-Bryan angles. Since Face API provides both the coordinates of face position and the Tait-Bryan angles of face's pose, we could use a transformation matrix *M* to calculate the position of point *P(Px, Py, Pz)*, which represents the projection of head orientation on the camera *XOY*-plane:

$$P^T = MF^T \text{(T: transpose of matrix)}$$

$$\text{where } M = \begin{bmatrix} 1 & 0 & tan\beta/cos\alpha \\ 0 & 1 & tan\alpha \\ 0 & 0 & 0 \end{bmatrix}$$

Knowing the coordinates of *F'*, *F* and *P*, two angles θi and θj could be worked out:

$$cos\theta i = \frac{\overrightarrow{FP} \cdot \overrightarrow{FF'}}{\left|\overrightarrow{FP}\right| \cdot \left|\overrightarrow{FF'}\right|}$$

$$cos\theta j = \frac{\overrightarrow{F'F} \cdot \overrightarrow{F'Z'}}{\left|\overrightarrow{F'F}\right| \cdot \left|\overrightarrow{F'Z'}\right|}$$

$$where\ \overrightarrow{F'Z'} = (0,0,1)$$

These angles determine if the faces of the agent (A) and the detected person (B) are within the perceptive zone of each other. Knowing these angles, the system could give a final judgment with one of the four following patterns (Figure 8): (i) Face-to-Face, (ii) A facing B, (iii) B facing A, (iv) Not face-to-face or facing. The system has two modes. One is analyzing recorded videos, and the other one is analyzing video stream during real-time video capturing.

To evaluate the physical precision of the system, a set of reference data is needed. In this research, I used a human-head model of a regular face. A laser pointer was fixed on top of the head with the laser beam aligned to the model's head orientation (Rae and Ritter. 1998). I prepared a 30cm×30cm graph paper, fixed it on the wall and set a camera coinciding with the origin. As shown in Figure 9, I put the model on the ground with its bottom located on a normal line of the coordinate system in order to control the distance, then rotated the model so that the laser beam could hit some reference points on the graph paper. There were totally 16 reference points, the coordinates of which were multiple of ten. During hitting each point with laser beam, the system (connected with the camera) analyzed the model's spatial information and computed the position of *Pp* which represents laser beam's projection on the graph paper, then compared the computed coordinates of *Pp* to the reference value (the position hit by laser) for evaluating system's precision.

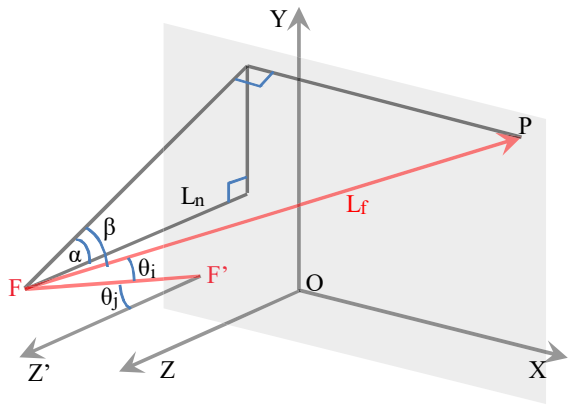

Figure 7. Spatial details of the detected and agent's face in the camera coordinate system.
Lf is the head orientation of the detected face *F*, while *F'Z'* shows that of the agent's face *F'*.

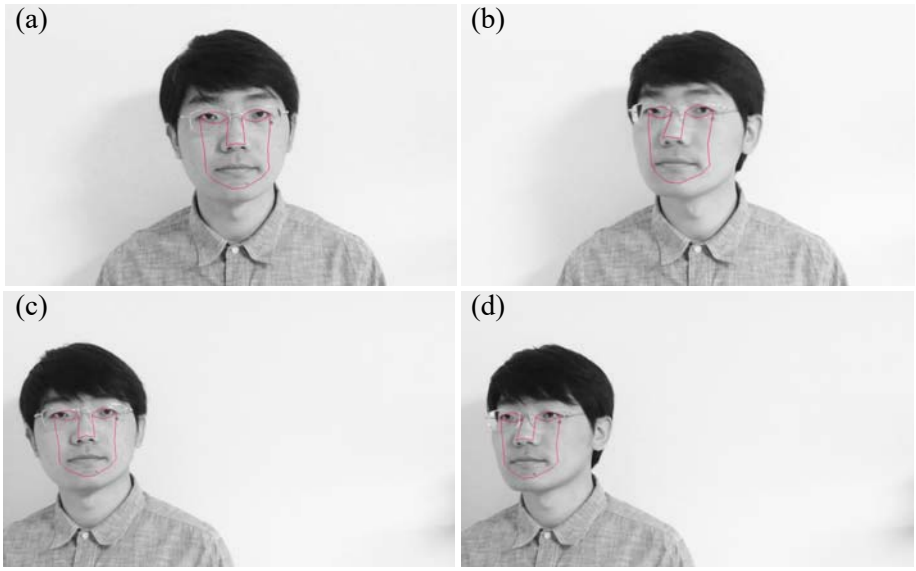

Figure 8. Face detection and judgments of facing and face-to-face.
(a): Face-to-face. (b): The agent facing the detected person. (c): The detected person facing the agent. (d): No facing or face-to-face.
In all cases, we assume that the camera is set on the agent's head and aligned with the agent's head orientation.

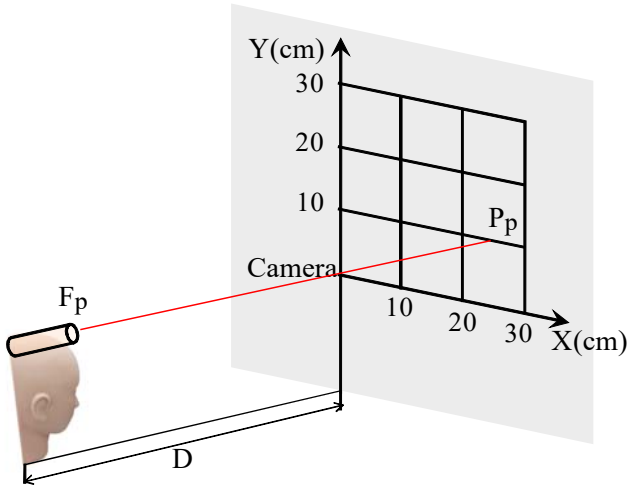

Figure 9. A projection test to evaluate precision of the camera-based system

The computation of *Pp* satisfies the following equations:

$$(F_p)^T = F^T + R(0, 9.5, 0)^T$$

$$\text{where } R = \begin{bmatrix} c_\beta c_\gamma & -c_\beta s_\gamma & s_\beta \\ c_\alpha s_\gamma + c_\gamma s_\alpha s_\beta & c_\alpha c_\gamma - s_\alpha s_\beta s_\gamma & -c_\beta s_\alpha \\ s_\alpha s_\gamma - c_\alpha c_\gamma s_\beta & c_\gamma s_\alpha + c_\alpha s_\beta s_\gamma & c_\alpha c_\beta \end{bmatrix}$$

In this equation, *Fp* is the position of the laser-head. When face was aligned with the camera-coordinate system, (0, 9.5, 0) represented the vector difference (unit: cm) between *Fp* and *F*. After rotating the face with Tait-Bryan angles, the difference changed according to a matrix *R*. In matrix *R*, s and c represent sine and cosine of the angle. Since the direction of the laser beam was parallel with head orientation, after getting *Fp*, we could use the transformation matrix *M* to calculate the laser's projection on camera's *XOY*-plane:

$$(P_p)^T = M(F_p)^T$$

The process of evaluation was done under a stable lighting condition. I set the distance of head model *D*={45cm, 75cm, 100cm, 120cm}, according to the proxemics defined by Hall (1966) and Lambert (2004). The authors summarized human's social normative behavior into spatial zones, and suggested the interval between 0.45m and 1.20m as common distance for mutual behavioral interaction. For each of the four distance I set, I tested all the reference points for 5 times (totally 16×5=80 trails), and calculated the mean absolute error *(MAE)* of computed coordinates on axis *X* and *Y* separately. A reference error *(RE)* was provided for each distance:

$$RE = D \cdot tan5°$$

The 5-degree angle is considered as the criterion for evaluating systems that estimate head-orientation (Murphy-Chutorian et al, 2009). An error that is less than 5 degree enables the system to satisfy a majority of applications. In this research, I combined this criterion with distance values to calculate *RE*. In Table I, I summarized all the *MAE* and *RE* according to each distance. The results revealed that the system has a fine precision in the tested distance range. According to different distances, the tested field-of-view was *FOV*={34°, 22°, 17°, 14°}. Since all the tested field-of-view could cover the 12-degree perceptive zone (12 degree: a basic threshold of k-degree face-to-face), it could be inferred that the system is qualified for being embedded into applications.

Table I. Precision of the camera-based system in a projection test

Distance	Reference error	Mean absolute error on Axis-Y	Mean absolute error on Axis-X
45cm	3.9cm	3.4cm	3.7cm
75cm	6.6cm	4.6cm	5.8cm
100cm	8.7cm	6.1cm	7.2cm
120cm	10.5cm	7.3cm	7.9cm

3.2.2 Using wearable device to obtain gaze information

To create a model of face-to-face that refers to gaze direction, in this research, I used SMI gaze tracker (Figure 10). The SMI gaze tracker is a glasses-shaped wearable device with a digital camera in the middle that captures the user's first-person view, and two infrared cameras that monitor two eye balls' movement. Based on the eye balls' movement, SMI gaze tracker calculates the user's gaze point in the first-person view (Figure 11(a)) according to each frame (60 frames/second) and records it.

Users' head-orientation during looking at others' face is what I would like to investigate in order to create the face-to-face model. To achieve this mission, I also used Face API. I combined the functions of SMI gaze tracker and Face API to build a software that automatically computed the angle between users' head-orientation and gaze-direction when the users were looking at others' face. Figure 11 shows some details of the process: First, SMI gaze tracker computes the 2D coordinates of gaze point in the first-person view, and Face API is used for face detection. During face detection, the system divides the first person view into several parts to improve the efficiency of Face API, the coordinates of face position (defined as middle of two eyes) in divided scene was then converted back to the coordinates in the full first-person view. Having the detected face, the software defines a circular facial region with center on the face position and with a radius that equals to 150% of the distance between face position and lower lip. The defined circular facial region is around 120% of the size of head. The software automatically checks if the gaze point is located within the facial region. If so, the status is recognized as looking at face. When the status is looking at face, the estimated 3D gaze vector by SMI gaze tracker at current moment will be extracted and used for computing its horizontal component. Note that SMI provides two gaze vectors according to each eye. The system first calculates two horizontal components, and uses the average direction of the two horizontal components as the binocular horizontal gaze vector.

Figure 10. SMI Gaze Tracker

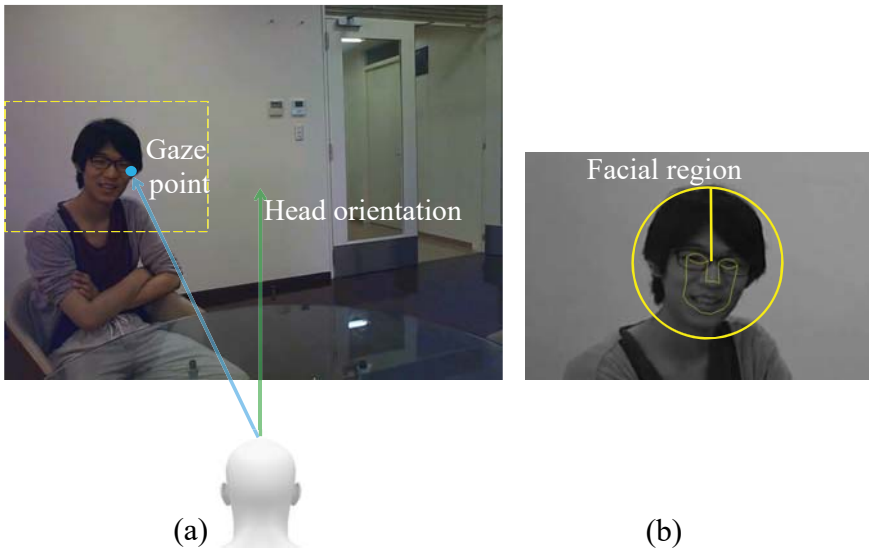

Figure 11. Gaze tracking and face detection
(a): First-person view with gaze point, provided by SMI gaze tracker. (b): Face detection and facial region based on first-person view, detected using Face API.

3.2.3 Human detection in public space

The previous methods are mainly used in personal zone (Hall, 1966, definition shown in Table II). In this research, I also investigated face-to-face in public space using a humanoid robot. In public space, face-to-face usually starts in social zone. Under such distance, facing a person can be generally defined as facing the person's body. In a public space (a hotel's lobby), I introduced human detection method to let the robot know the presence and position of people surrounding so that the robot could face them by rotating its head.

Table II. The human proxemics

Spatial zone	Distance range
Intimate zone	[0m, 0.45m)
Personal zone	[0.45m, 1.20m)
Social zone	[1.2m, 3.6m)
Public zone	[3.6m, ∞)

Microsoft Kinect was used for human detection. I programmed the Kinect sensor with OpenNI (online citation b) for getting presence and position information of the detected people. OpenNI first provided 2D coordinates of a person's center point (Figure 13), then transformed the 2D coordinates into 3D real-world coordinates with the depth information measured by Kinect sensor. The robot kept real-time communication with the Kinect system via TCP/IP to get position information of the detected people (Figure 12(b)). I hid the Kinect system in a black box, which was also used for putting the robot (Figure 12(a)). The box was designed with a small window and an embedded panel where we can put Kinect sensor. This was to keep Kinect able to work as usual, and to avoid attracting people's interest by Kinect rather than the robot. Once there was person detected, the robot would start its behavior to have interaction with the person.

I did a pilot study to check the accuracy of Kinect system's human detecting function: One coder stood behind the robot, and the robot started the process of human detecting and greeting. Within a 27-min trail, a total number of 20 people entered the social zone of the robot. The expected behavior by rotating the robot's head orientation towards the detected person was correctly done for all the 20 people.

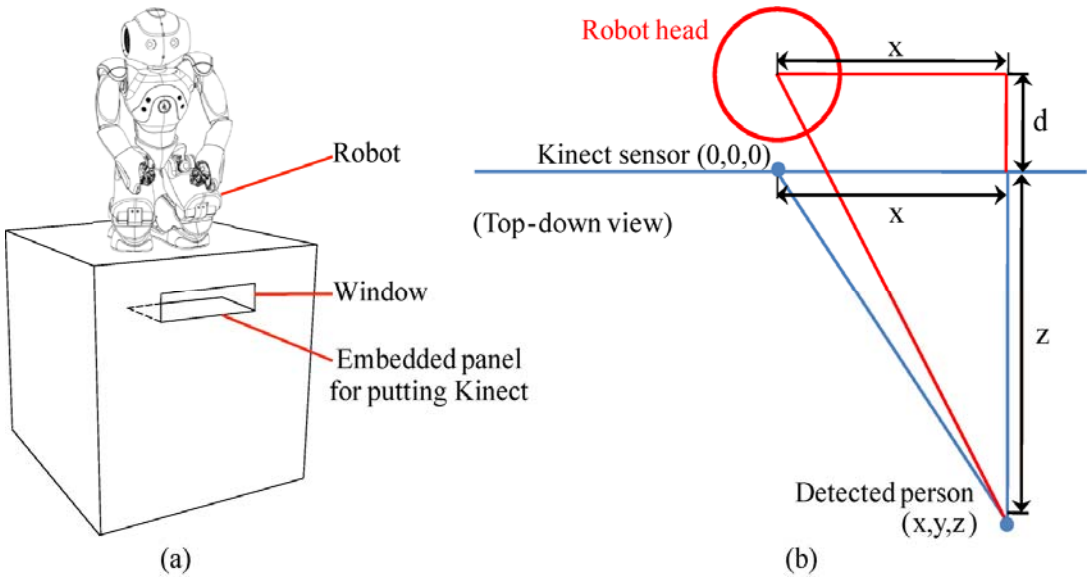

Figure 12. Robot system for human detection and tracking.
(a): Box with NAO and Kinect for setting up experiment. (b): Geometric model that robot used to track the detected person with head orientation.

Figure 13. Human detection in Kinect's 2D visual plane.
Person marked in green is the latest person who entered the space.

3.3 A Scenario-based Advanced Model

3.3.1 Experiments

To create a model of face-to-face, in this research, 18 subjects (13 male, 5 female, from 5 different countries, aged from 24 to 29) were asked to participate in 10 sessions of experiment. In each session, one subject was asked to wear the gaze tracker and there were other two confederates who did not wear any device on them. The 10 subjects who wore gaze tracker were different subjects, whereas each subject could participate in different sessions as confederate. During the experiment, the subjects were sitting around a round table (diameter: 1.1m). The seat arrangement was close to a regular triangle (Figure 14), and the seats were fixed during the experiment sessions. In each session, I conducted experiment under the two following scenarios:

- Non-instructed scenario: 10-minute free talk with topic decided by the three people.
- Instructed scenario: The subject was asked (orally, by an experiment instructor) to look at the left and then the right confederate's face; During such process, keeping looking at each confederate's face for at least 3 seconds; Repeating the process for 5 times.

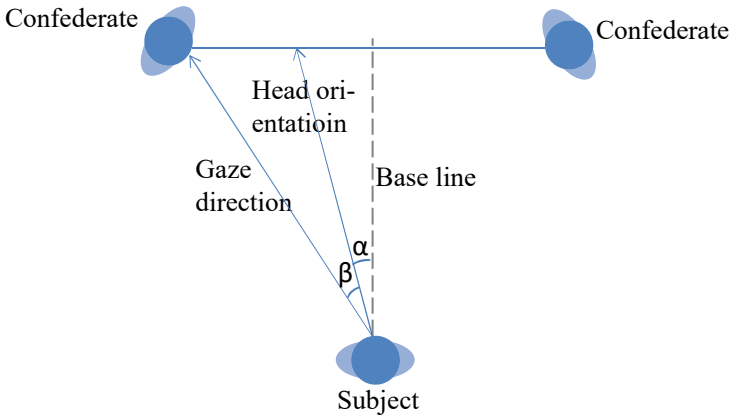

Figure 14. Geometric model of the experiment for modeling of face-to-face

I did calibration of gaze tracker for each subject before starting the experiments, and conducted a 1-minute test to check the gaze tracker's performance. During 1-minute test, an experiment instructor asked the subject to look at each confederate's face as well as some other objects in the environment (objects were not on the table), and checked the real-time gaze tracking via a tablet device. The tablet showed the subject's first person view with a moving mark that represented the gaze point. If the experiment instructor judged that the tracking was reliable, I started doing the experiments. Otherwise, I re-calibrated the gaze tracker and repeated the process.

For each of the 10 subjects, I first conducted the non-instructed scenario and then the instructed scenario. After all sessions, the experiment data including first-person view video and gaze information was analyzed by the developed software to generate the statistical analysis. As mentioned in section 3.2.2, what I calculated by using Face API and gaze tracker was the horizontal angle between head-orientation and gaze-direction when the subject was looking at others' face. According to the geometric model shown in Figure 14, angle beta is what we obtained from gaze tracker. Since the seats were arranged based on an regular triangle formation, we can refer to the following equation to estimate angle alpha.

$$\alpha + \beta = 30°$$

Angle alpha was defined - under the experiment context - as the subject's head orientation. Note that the baseline in Figure 14 might not be user's body orientation, but just a middle-line in the geometric model. Also note that throughout all sessions of experiment, it happened only with 0.2% of the time that the subject distributed his/her head-orientation out of the triangle range (the space among seats). From the first-person view, such irregular situation was observed like - for example - looking at left person's face but the person's face was on right half of the scene. I ignored that 0.2% data so that the equation for alpha and beta could be suitable for all analyzed data.

Since the target in this research is to investigate relation between head-orientation and looking-at-face, as the output of data analysis, I calculated the probability density of head orientation with resolution of 1 degree on the interval [-30°, 0°] in case of looking at the left confederate's face, and on [0°, 30°] in case of looking at the right confederate's face. Figure 15 is an example of head-orientation probability density. It is the case of one subject when he is looking at the right confederate's face during free-talk (non-instructed scenario). We

could see his head orientation distributed from 6 to 24 degree, and with a peak on 16 degree. Such probability density was generated for each of the 10 subjects according to different scenario: non-instructed and instructed scenario.

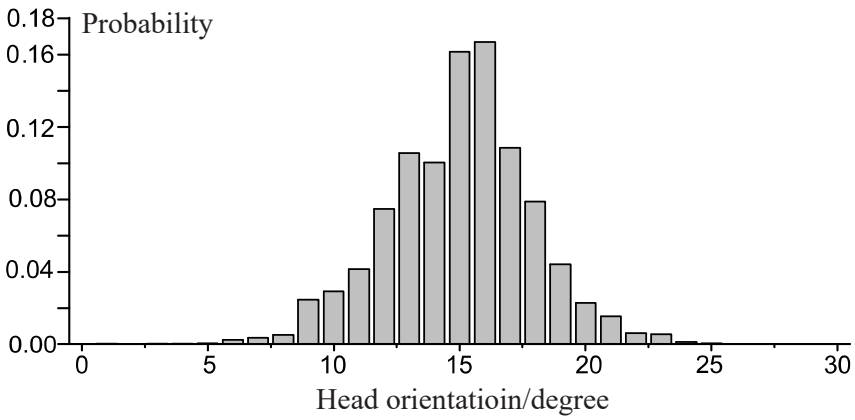

Figure 15. An example of probability density of head-orientation (one subject's case).

3.3.2 Features of head orientation

I analyzed the features of each subject's head orientation by looking into the distribution and its peak value, and categorized them into three types (Figure 16). Referring to the definition of k-degree face-to-face, subjects with Type-A are more sensitive in rotating head around a threshold (which is near the peak). Type-B indicates that the subject tend to fix more attention on others' face by doing almost direct facing. Type-C is like typical Gaussian distribution. Table III summarized the type of distribution and peak value for all subjects under different scenarios. In most of the cases, Type-C was observed. Based on the peak angles, to investigate correlation and significant difference among different experiment conditions, I calculated Pearson coefficient and conducted T-test to check the correlation and statistic significance between different scenarios. As a result, I found significant difference between non-instructed and instructed scenarios in case of looking at left confederate's face (t=2.95, T-test), similarly there was significant difference between two scenarios when subject looked at right confederate's face (t=3.43, T-test). For the comparisons between looking at left and right confederate under the same scenario, although there was no significant difference, the correlation was also weak (Left: R=0.25, Pearson coefficient, Right: R=0.23, Pearson coefficient).

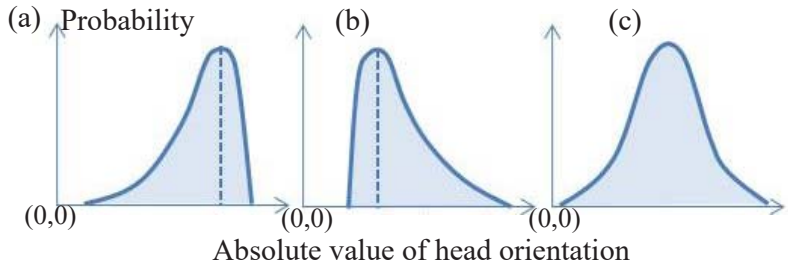

Figure 16 : Types of head-orientation distribution
(a): Type-A, P(angles larger than peak)≤15%, P stands for Probability. (b): Type-B, P(angles smaller than peak)≤15%. (c): Type-C, the distribution which is neither type-A nor -B.

Table III. Peaks and types of subjects' head-orientation distribution

		Looking at left confederate		Looking at right confederate	
Scenario	Subject	Type	Peak value	Type	Peak value
Non-instructed	1	C	10	C	14
	2	C	11	C	16
	3	C	18	C	18
	4	C	14	C	12
	5	C	22	C	20
	6	C	18	C	13
	7	C	21	C	15
	8	C	19	C	17
	9	C	17	A	11
	10	C	19	C	23
Instructed	1	C	18	A	14
	2	B	4	C	9
	3	C	9	C	10
	4	B	4	C	9
	5	A	15	C	5
	6	B	10	C	7
	7	C	10	C	4
	8	C	16	C	16
	9	C	11	B	10
	10	C	14	B	15

3.3.3 The Gaussians

Having 10 subjects' probability density charts of their head orientation, for each value in [-30°, 0°] and [0°, 30°], I calculated a mean probability among the 10 subjects, and used Gaussian Model to fit the results under 2×2 conditions (Figure 17). For all conditions, the fit process was successfully done, and an *Rnew* according to the following equations was computed:

$$Q = \sum(y - y^*)^2$$
$$R_{new} = 1 - (Q/\sum y^2)^{1/2}$$

where y is the observed and y^* is the fitted probability. For each subject under each scenario, size of the sample y was 31 on either left or right side. Therefore total size of sample y was 310 in each of the 2×2 conditions in Figure 17. Q is sum of square of deviations, *Rnew* is the determination of coefficient in non-linear regression equation. The computed *Rnew* according to each condition was 0.471 (left side, non-instructed scenario), 0.346 (right side, non-instructed scenario), 0.149 (left side, instructed scenario) and 0.215 (right side, instructed scenario). The *Rnew* under instructed scenario was slight. This is most probably because of the limited time when looking at face under instructed scenario, compared with having a 10-minute free talk. From the results, peaks of head orientation in the fitted Gaussian Models located on -14° (left side, non-instructed scenario), 15° (right side, non-instructed scenario), -18° (left side, instructed scenario) and 20° (right side, instructed scenario). The result indicated that: (i) Comparing with instructed scenario, when having free talk, the subjects used head movement with smaller angles to look at faces. (ii) Under each specific scenario, the peak of head orientation on the left and right sides were close to each other.

I also analyzed the probability density when subjects were not looking at face under the non-instructed scenario, and similarly used Gaussians distribution to model the average results among 10 subjects (Figure 18). Compared with looking at face, when not looking at face, the peak of Gaussians is lower, which inferred that when looking at face, subjects' head orientation is much more fixed (Figure 19).

3.3.4 K-degree face-to-face

Based on the concept of k-degree face-to-face, in this research, I calculated the minimum ranges of head-orientation that could cover more than 70%, 80%, and 90% probability in the fitted Gaussian Models (Table IV). Each range had one side on -30° or 30°, and the length of range was considered as the value of k. These values of k under different scenarios and conditions contributed to the model of k-degree face-to-face. For example, when subjects look at left confederate's face under non-instructed scenario, there is 90% probability that the subjects' head orientation was within a 23-degree perceptive zone. Taking 80% of probability as a base line, in case of non-instructed scenario, 20 degree could be accepted as threshold, while 15 degree was accepted as threshold for instructed scenario.

Table IV. Scenario-based thresholds of k-degree face-to-face

Scenario	Side	Probability	Minimum range	Threshold
Non-instructed	Left	>70%	[-30°, -11°]	19
		>80%	[-30°, -9°]	21
		>90%	[-30°, -7°]	23
	Right	>70%	[12°, 30°]	18
		>80%	[10°, 30°]	20
		>90%	[8°, 30°]	22
Instructed	Left	>70%	[-30°, -16°]	14
		>80%	[-30°, -14°]	16
		>90%	[-30°, -11°]	19
	Right	>70%	[17°, 30°]	13
		>80%	[16°, 30°]	14
		>90%	[13°, 30°]	17

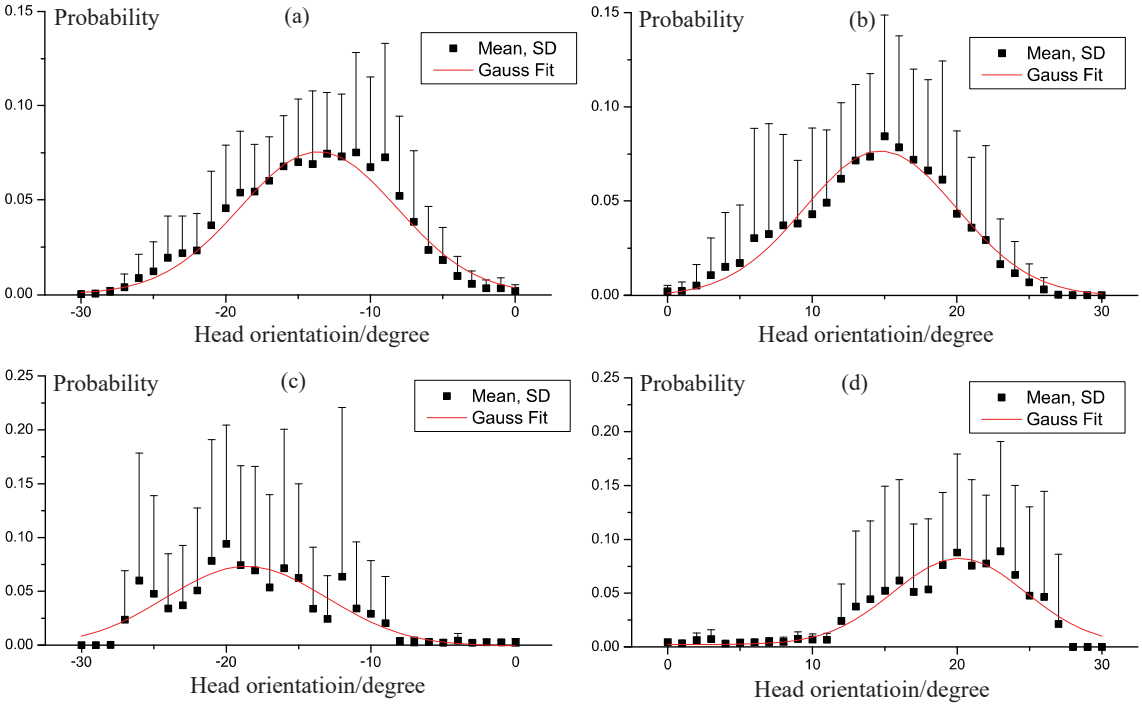

Figure 17. The Gaussians of head-orientation while looking at face
(a): Non-instructed scenario, looking at the left confederate's face. (b): Non-instructed scenario, looking at the right confederate's face. (c): Instructed scenario, looking at the left confederate's face. (d): Instructed scenario, looking at the right confederate's face.

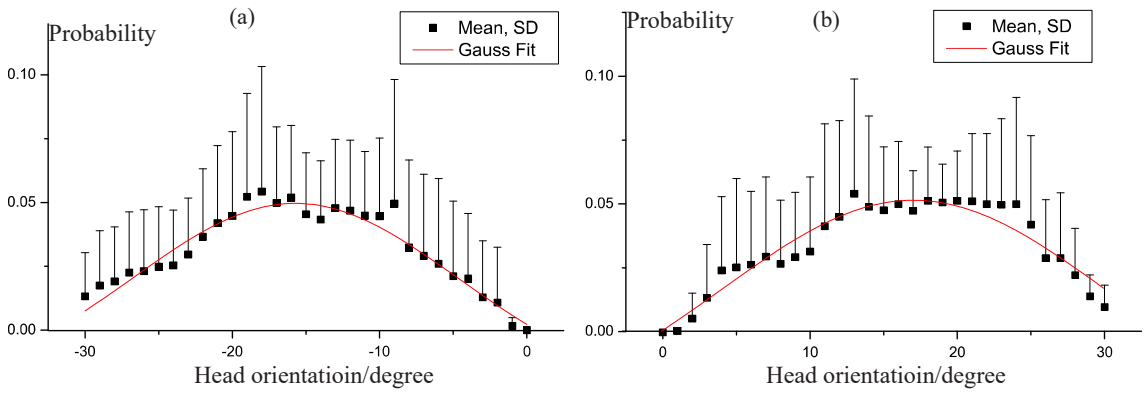

Figure 18. The Gaussians of head-orientation while NOT looking at face under non-instructed scenario
(a): Eye ball moving to the left. (b): Eye ball moving to the right

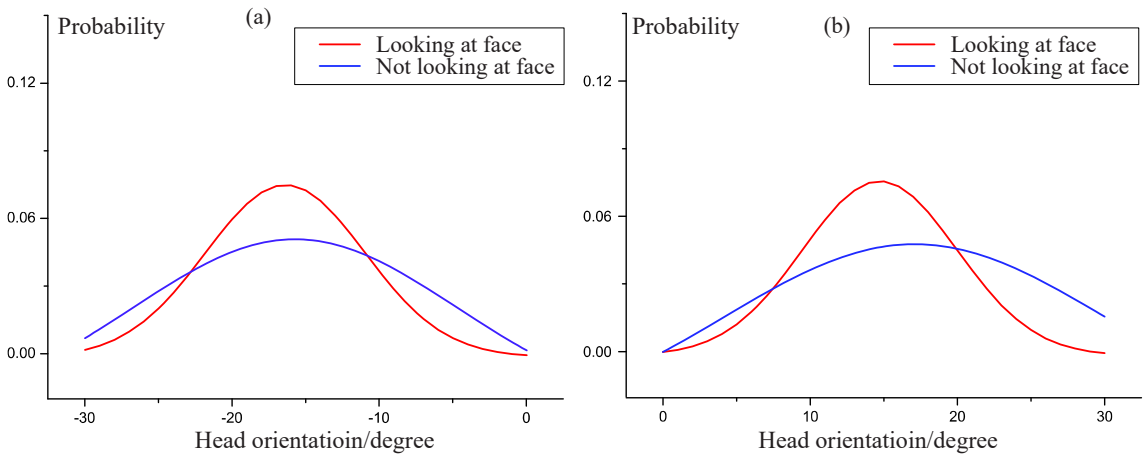

Figure 19. Comparison of Gaussians between looking and not looking at face
(a): Eye ball moving to the left. (b): Eye ball moving to the right

3.4 Performance Evaluation using a 360-degree Camera

Before conducting experiments with specific users (e.g. children with developed disorders, which will be introduced in chapter 4), in order to evaluate if the proposed method is reliable, I conducted a preliminary experiment by using 360 degree camera. 360-degree camera has a benefit on capturing multiple people from different perspectives. This provides an opportunity to measure face-to-face behavior. In the preliminary experiment, two subjects sit in face-to-face formation around a table where the camera is located (Figure 21), and randomly rotated their heads during a 3-minute session of recording. After the session, I corrected the 360-degree scene captured by the camera into several scenes with normal geometric layouts (from different perspectives, as shown in Figure 20). The image processing method for removing distortion of image is referring to a work by Mori et al (2006) and Ikeda (2013).

Having the corrected scenes for the two subjects, I developed a system by using face API to detect the 3D position and rotation of face in each of the two camera coordinate systems shown in Figure 21. The two camera coordinate systems have their axes aligned (however, axis-Y and -Z in each system are with opposite directions to those in another system). Based on the spatial relation, the system merged the two coordinate systems into one, and could make judgment on whether face-to-face happens based on the following criteria:

$$cos^{-1}\left(\left\|\overrightarrow{FaFb}\right\| \cdot \left\|\overrightarrow{Pa}\right\|\right) \leq 15°$$
$$cos^{-1}\left(\left\|\overrightarrow{FbFa}\right\| \cdot \left\|\overrightarrow{Pb}\right\|\right) \leq 15°$$

where *FaFb* is the vector that connects the two faces, *Pa* and *Pb* are the head orientations of two subjects. All these vectors are defined in the merged coordinate system. Since this experiment was under an instructed scenario, 15 degree was applied in the system as threshold for judgment of face-to-face.

One human coder was also asked to code the source video (36-degree video, one frame shown in Figure 22). Although in the source video the captured scene was distorted, since the beginning status of two subjects was face-to-face, the coder was able to refer to that status to make his judgment on whether face-to-face occurred. The coder was asked to look

into each frame of the video to check the subjects' behavior (10fps). The result was then compared to the system's coding. As a result, the observed agreement between human coder and system's coding was 0.73, which indicated that the system can be used instead of human coder for doing face-to-face analysis.

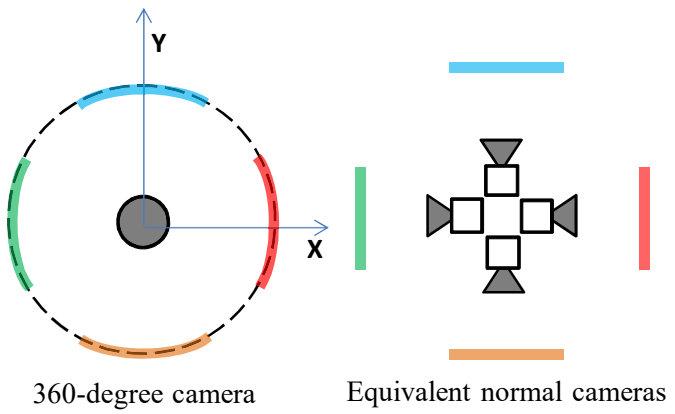

Figure 20. Peripheral view of 360-degree camera and its equivalent normal cameras

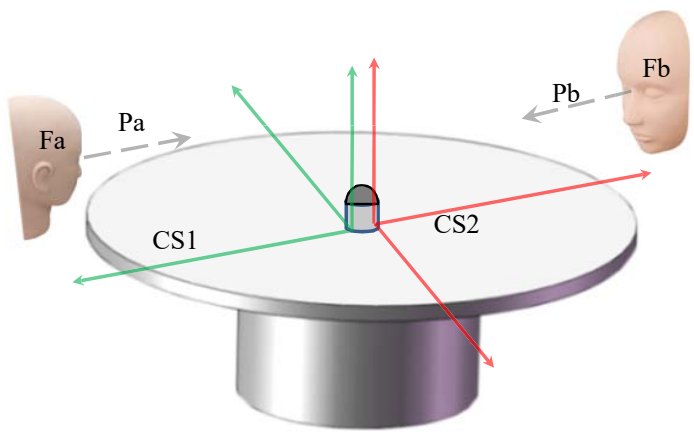

Figure 21. Geometric model of face-to-face detection using 360-degree camera

Figure 22. Image correction and face-detection of the system with 360-degree camera
(left): The source image from 360-degree camera; (center): The corrected images from different perspective; (right): Face detection on the corrected images.

Chapter 4
Applications to Human-Robot Interaction

4.1 Robot-Assisted Therapy

Recently computational behavior science has been addressing more attention. Especially for people with developed disorders such as children with autism spectrum disorder (ASD), quantifying their behavior in social interaction could provide data support for therapists to understand their state-of-health and to design therapeutic activities (Rajagopalan, 2013). More precisely, researchers have organized such research scope into three aspects (online citation c).

- **Capturing Behavioral Signals**
  Using current technologies such as RGB/IR camera, wearable sensors, etc, it is possible to capture various behavior signals. As to children with ASD, the most studied behavior is gaze (Nakano et al. 2010), and recently there are also researches in which the author focused on analyzing head movement of those children (Tsuji et al. 2016). However, instrumentation on the body of those children is still a challenging issue.
- **Measuring Behavioral Variables and Finding Patterns**
  Having the captured behavior signals, the next challenge is to make clear the features of behavior by looking into variables and patterns of the data. To achieve these targets, it is also important to synchronize the time line of different signals. For example, in Leekam et al's research (1998), to infer the orientation of a child's attention in activities with therapists, it is required to analyze multiple behavioral signals including gaze of the child, and head-turn behavior of the therapist within the overall context of the on-going interaction.

- **Understanding Dyadic Behaviors**
  After previous two stages, a possible way towards understanding of behavior is to analyze and quantify complex behavioral interactions between a child and another actor. Joint attention is a typical formation of dyadic behavior. The analysis of joint attention can be also considered as a basis for researches on other specific topics such as emotion and symbolic recognition, and language development (Mundy, 1995; Kasari, Freeman, and Paparella. 2006; Mundy, 1990).

Regarding Dyadic behavior, Robot-Assisted Therapy has been widely studied (Robins, et al. 2005, 2009; Lee et al. 2012; Ferrari et al 2009). Since usually social robots are developed with robot vision, it is good opportunity to embed the camera-based system that measures face-to-face into those robots, and to measure the features of face-to-face interaction between robots and children with ASD.

To introduce the system into ASD children's activities, I used humanoid robot NAO as agent. The technical method is referring to a previous work done by Hirokawa (2014), in which the author created a framework for controlling humanoid robot NAO remotely. By using this framework, I conducted experiments with ASD children. In the experiment, social behavior (e.g. nodding head, waving hand, giving presents) performed by NAO was used as stimuli to trigger affective response of those children in a 20-minute activity. The activity was done in a circle area set in a room (Figure 23). The camera installed on NAO's head was used to record video of the activity (frame rate: 7/second). Each video recorded the activity of one child.

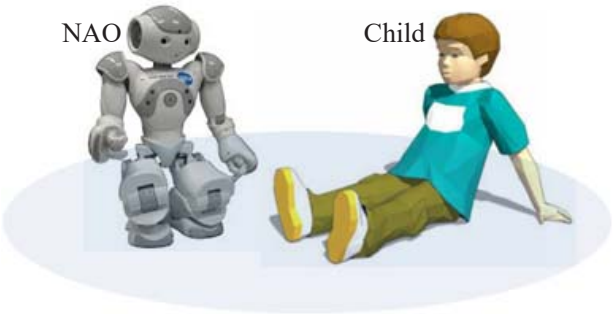

Figure 23: Robot-Assisted Activity for children with ASD

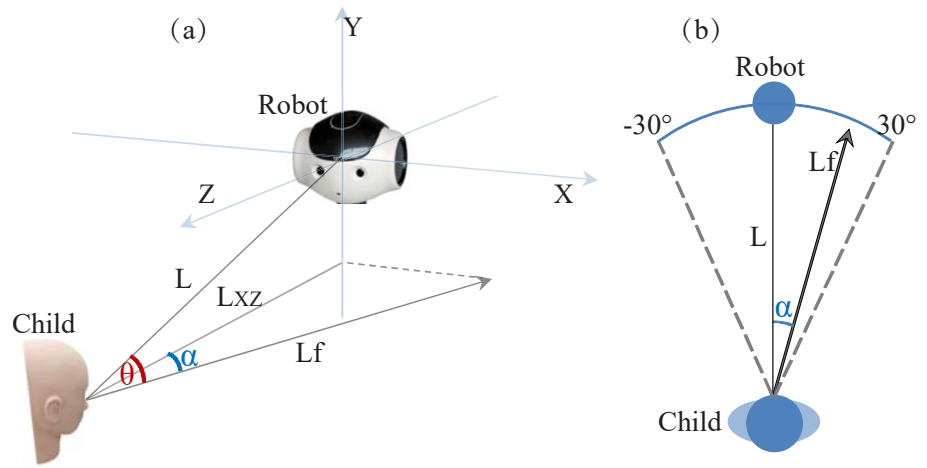

Figure 24: Spatial details of the Robot-Assisted Activity
(a): Spatial details in the camera coordinate system. *L*: link between child's and robot's faces, *Lxz*: projection of *L* on *XZ* plane, *Lf*: child's head orientation. (b): Top-down view

In order to analyze the children's behavior, I focused on two angles which are related to the children's head orientation in the activity, as shown in Figure 24: (i) Angle *Theta* is the offset angle from the child's complex head orientation (sum of horizontal and vertical orthogonal components) to the link that connects child and robot's face. Since k-degree face-to-face is defined based on a cone-shaped perceptive zone around head orientation, by comparing such angle with the perceptive zone's semi-vertex angle, it would be possible to judge if face-to-face happens or not. (ii) Angle *Alpha* is the orthogonal components of angle *Theta*. It is similar to the definition of head orientation in section 3.3. Since the robot was the only cue for children's attention in the activity, I assumed that such angle would have a Gaussian distribution, from which we could understand how those children used head orientation to facilitate interaction with the robot.

4.1.1 Judgment of k-degree face-to-face

Making judgment of face-to-face requires the comparison between angle *Theta* (Figure 24) and a certain threshold of k-degree face-to-face. This could be done by the camera-based automatic system. At the same time, it is also possible to ask human coders to make the judgment based on their spatial perception. One important thing is to check whether the system's judgment is similar to human coders' coding, which indicates how much the

system can be used to process behavior analysis instead of human. In this study, I randomly chose two children's sessions for checking the similarity between system and human coders' judgment. For each session, I first used the automatic system to analyze the video. During system's processing, each frame of the videos was marked with 0/1 (Table V). Similarly, I asked two human coders to check the videos and mark the frames with 0/1 that represent the judgment of face-to-face. In the coder's case, we assumed that the coder would perceive that the child was facing him if the child actually faced the camera. In this study, I only checked if the child was facing the camera (the robot's face), as in the activity basically the robot was always facing the child.

Table V. Rules for human coder and system's coding on the robot's vision

	Label	Meaning
Human coder	1	The coder feels that the child is facing the therapist (by considering only head-orientation)
Human coder	0	The coder does not feel that the child is facing the therapist (by considering only head-orientation)
System	1	The child is facing the camera
System	0	The child is not facing the camera, or face is not detected

In addition to measuring face-to-face, the system also provided the distance between the robot and the child in each frame when the child's face was successfully detected. Within a 20-minute activity, the range of detected distance was [22cm, 128cm] for the first child, and [29cm, 112cm] for the second one. I ignored the frames in which the distance was within 35cm before marking each frame (before both system' and human's coding). Such ignored frames occupied 0.8% and 0.4% in the two videos. After coding the videos, I compared the system's with the coders' results. Table VI shows the calculated Kappa coefficient between each pair of different coders (according to the coding results on each frame). Data revealed that two human coders and system gained moderate kappa to each other. Such results inferred that the criterion defined for measuring face-to-face in this research could lead to coders' corresponding perception. On the other hand, I compared the time spent on video analysis between system and human. The results (Table VII) indicated that system's computing speed is much higher than human's. Above all, it is revealed that the system can be used in robot-assisted activities for the ASD children, helping to improve the efficiency of measuring face-to-face.

Table VI. Kappa coefficient between human coder and system's coding on robot's vision

Session	Pair	Kappa
1	System v.s coder 1	0.50
1	System v.s coder 2	0.58
1	Coder 1 v.s 2	0.44
2	System v.s coder 1	0.54
2	System v.s coder 2	0.51
2	Coder 1 v.s 2	0.60

Table VII. Comparison of coding speed on robot's vision

Session	System	Coder 1	Coder 2
1	4'39"	>4h	>4h
2	4'55"	>4h	3h46'

I also summarized the probability density of the angle *Theta* for each session (Figure 25, Note that only the head orientation in the interval of [0°,30°] was considered.) From the results, several features can be observed in both sessions: (i) There was only one peak in the distribution; (ii) 12-degree was on the decline stage; (iii) Compared to some larger angles (e.g. 16 degree and larger), the probability of 12-degree was much higher. These features indicated that the children did use head orientation with a threshold near to 12-degree in responding to robot's social behavior, and the definition of k-degree face-to-face applies to those children under the experiment context.

4.1.2 Modeling of horizontal head movement

For totally 10 children who participated in the activity, the probability density of angle *Alpha* was analyzed, which indicates how those children use horizontal head movement to facilitate interaction with the robot. Similarly as the model in section 3.3, Gaussian distribution was used to fit the average probability density among the 10 children in [-30°, 30°]. From the result (Figure 26, *Rnew:* 0.77) we could see that: (i) Peak was located close to 0 degree, which suggests on the horizontal way, those children tended to move their head directly towards the robot's head and body. (ii) Horizontal head movement in [-11°, 11°] covered more than 80% of probability. Although in this experiment gaze information was not referred to, since the robot was the only cue for attention, it could be inferred that 11 degree might be a threshold of k-degree face-to-face under the robot-assisted activity.

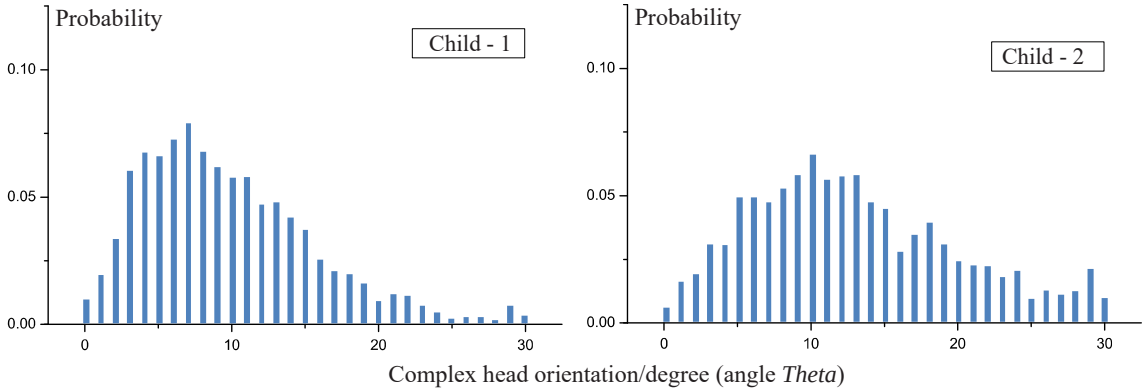

Figure 25. Probability density of two children's complex head-orientation in the robot-assisted activity

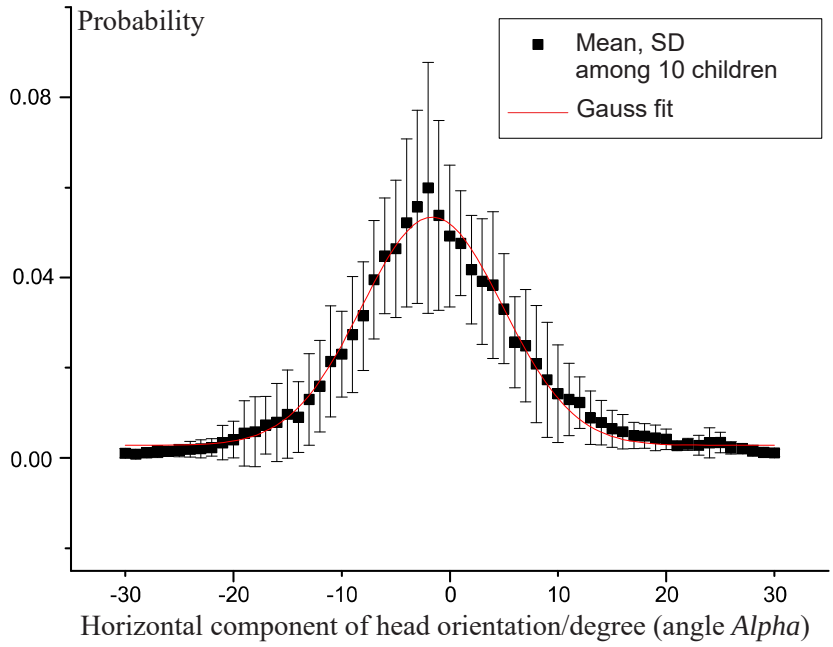

Figure 26. Gaussian of ten children's horizontal head-orientation in the robot-assisted activity

4.2 Social Robots in Public Space

As a supplementary study towards understanding face-to-face in social zone, I introduced robots in a hotel's public space. I prepared different robot behavior under different interaction styles, in order to investigate people's response to robots' behavior in those interactions. The experiments in this research were conducted in the hotel's lobby (Figure 27). The lobby is located on the 6th floor and near the front desk. Three elevators on the left side of the lobby are used as common entrances. All the guests were supposed to pass through the elevator hall from elevators to the front desk, during which they would encounter with the robot. Humanoid robot NAO was used for doing experiments. I enabled NAO with different behavior, and put it in the elevator hall. This offered an opportunity to study human-robot interaction under a public space context.

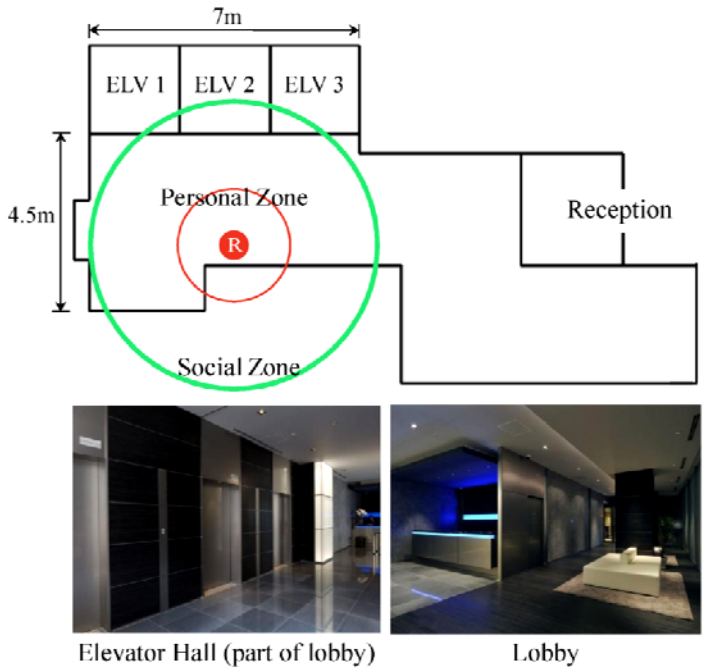

Figure 27. The hotel public space.
R represents the position where robot was located

4.2.1 Interaction styles

I designed two interaction styles for the human-robot interaction in the hotel public space (Table VIII). The first interaction style is based on direct body and verbal behavior. In public spaces, greeting is usually the best way to initiate communication (Duranti. 2014), and most of the communication follows greeting. In human-robot interaction, greeting is also considered as an important aspect (Lee, and Makatchev. 2009). In this research, I enabled a single robot with functions that facing and greeting individual guests. By doing this, the robot could directly address the hotel guests' attention, and we could see how the robot's greeting affect human's behavior under a public space context.

Table VIII. Interaction styles in the hotel public space

Style	Robot's body behavior	Robot's verbal behavior
Direct interaction	Facing and tracking the guest with head orientation	Greeting each guest
Indirect interaction	Arm gesture to supplement verbal behavior	Soliloquy about hotel's information

Another interaction style is indirect interaction, which could happen as a group behavior (Platon, Sabouret, and Honiden. 2005). One of the typical examples of having indirect interaction in public spaces is through public broadcast that attracts multiple people's attention concurrently and triggers human-human interaction. In this research, I used a single robot with soliloquy function. The content of soliloquy was about information of the hotel. The robots started soliloquy when it detected person in the elevator hall. During the soliloquy, the robots did not provide any speech directly aimed to individual guests.

4.2.2 Human behavior

A previous research suggested that the beginning period of interaction is crucial in human's social activities, and indicates how the interaction will be carried out (Schegloff, 1967). A further research using AIBO robot also showed that in human-robot interaction, the "first five seconds" of encounters between people and a robot has a significant effect on the users' further engagement (Pitsch et al. 2009). In this research, I also focused on the first few seconds of human-robot interaction, during which I extracted people's first response after they encountered with the robot, and compared those behaviors under different interaction styles.

A psychological research suggested that human's initial response to a robot is fast, unconscious and highly stimuli-driven (Aarts, and Dijksterhuis. 2000). It is important to investigate such kind of first-response of human during interaction with social robots. Another important feature of first-response is that such behavior is unaffected by group. Usually in public spaces, many customers appear in groups, which means information can be disseminated and passed from individual to individual to broaden its impact. Such group context is a sensitive factor that can affect people's response to an attention-cue (in this work it means the robot), while the first response is unaffected by group and could reflect people's response to the robot more purely.

During experiments, the guests' first response to the robot could be classified into four patterns (Figure 28). Those patterns reflect the levels of the guests' interest to the robots. Level I was regarded as low-interest response because the guest walked to the front desk without looking at the robot. Level II included people who noticed but were not particularly interested in the robot. Level III showed that people who were interested enough to stop for at least 2 seconds to look at the robot just when he noticed the robot. Finally, in level IV, the guests showed significant interest by approaching the robot without stopping when they noticed the robot. The guests on level IV approached the robot for more interaction such as touching and doing face-to-face with the robot.

4.2.3 Experiments and results

I conducted two experiments under different interaction styles. Each experiment started at 16:00 and ended at 18:00 on Saturday afternoon. The time period was chosen with reference to the guest-flow history data so that we could have at least 50 guests who went through the elevator hall during each experiment. Saturday afternoon was also a time when the hotel guests have more time to pay attention to the robot.

After each experiment, the guests' behavior was analyzed by watching the hotel's surveillance video. Two human coders classified the guests' behavior into the patterns I had defined. (There was no dissent between two coders during categorizing the behavior.) They started analyzing guests' behavior from the recorded time 16:00 in the surveillance video, and stopped after counting a total number of 50 guests for each experiment.

Figure 29 shows the results of first response for each of the three experiments. From the results we can see the distribution of the number of guests on each interest-level. For checking statistical significance, I conducted chi-square test of independence. I compared

each two experiments' results on each interest-level (0.05 as significance level). As a result, on level IV, significant difference was found between direct and indirect interaction. (χ^2=9.003, p<0.05, Yates' correction) It could be inferred that people's interest to robot's behavior under different interaction styles could be differentiated even in a short time (within 3 seconds for all the guests counted in the experiments to make their first response after encountering with the robot). Also, since level IV reflected people's particular interests in robot's behavior, we could understand from results of level IV that at a beginning period of interaction between guests and the robot, direct interaction is more attractive than indirect interaction. There was also significant difference between greeting's and soliloquy's cases on level I (χ^2=7.644, p<0.05), indicating that soliloquy might not be an effective way of starting human-robot interaction in the hotel public space.

Figure 28. Behavior patterns of first response.
(a): Level I:Walking to the front desk without looking at the robot. (b): Level II: Looking at the robot while walking to the front desk. (c): Level III: Stopping to look at the robot when he/she noticed the robot. (d): Level IV: Approaching the robot when he/she noticed the robot.

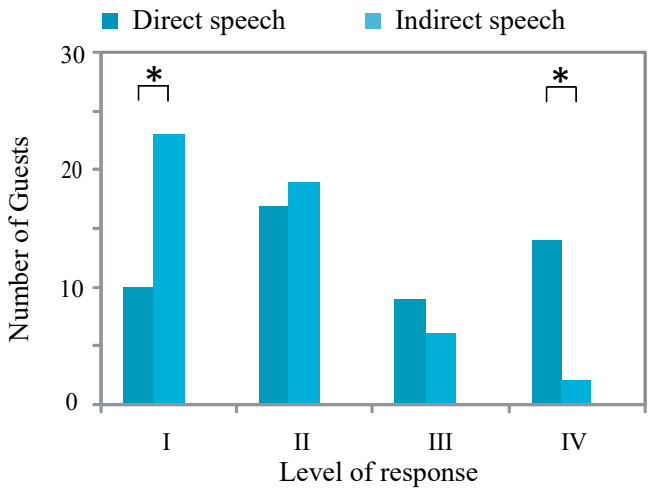

Figure 29. Human response towards direct and indirect interaction in the hotel public space

Chapter 5
Applications to Human-Human Interaction

5.1 Wearable Camera for Describing Therapy

Wearable device is recently addressing bloated inventories (Billinghurst and Starner, 1999). To introduce wearable systems into practical use, therapeutic as well as rehabilitation activities are considered as typical applications (Occhialini et al. 2004; Sung et al. 2005). Because of the fusion between computational behavior science and therapy, a brand-new research area - effective computing - has gained rapid growth (el Kaliouby et al, 2006). Similarly as using robot vision, it is also possible to apply the vision based automatic system with head-mounted camera, so that the face-to-face interaction between children with ASD and therapists could be measured and analyzed.

The videos (first-person view) used for data analysis is recorded by therapists from University of Keio, Japan, who conducted social activity with autistic children. In the activity, one therapist who wore the head-mounted camera used some toys (tablet, mini car, painting, etc) to interact with the children. During the interaction, the therapist sometimes guided the children to behave some body-gestures (raising hand, doing hi-five, etc). The duration for each child's session was controlled to no more than 20 minute. Similarly to the robot-assisted activity, I analyzed the children's behavior by looking into two angles as shown in Figure 30. One is the complex head orientation and the other is its horizontal orthogonal component.

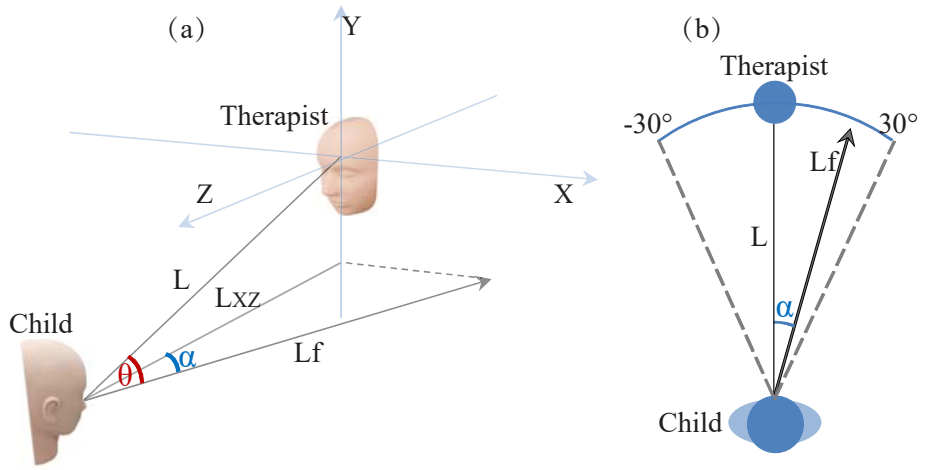

Figure 30: Spatial details in the activity between therapist and children with ASD
(a): Spatial details in the camera coordinate system. *L*: link between child's and therapist's faces, *Lxz*: projection of *L* on *XZ* plane, *Lf*: child's head orientation. (b): Top-down view

5.1.1 Judgment of k-degree face-to-face

Focusing on analysis of angle *Theta*, two sessions among the experiments were selected and analyzed using the camera-based system that measured face-to-face. Since the activity was conducted under a human-human instructed scenario, 15 degree was considered as one possible threshold for the system for the judgment of face-to-face. However, since the interaction was also under a one-v.s-one setting. A strict threshold - 12 degree - was also used for reference. The video of each session was processed twice by the system with different criteria of face-to-face. Regarding the human coder, one naive coder was asked to code the videos. To save the time for human coder's coding, the rules were changed from frame-by-frame to time-stamp based coding. Table IX summarized the detailed rules for human coder and the automatic system. After human coder's coding, the time-stamp based result was converted to frame-by-frame results (0/1) and compared with the system's coding. Since system uses two different criteria of face-to-face, for each session, the comparison between human and system was conducted twice. The results were shown in Table X. From the results, we could see that when system using 12 degree as threshold, the kappa between system and human coder's coding was higher.

Table IX. Rules for human coder and system's coding on therapist's vision

	Label	Meaning
Human coder	Time stamp	The coder feels that the child starts to face the therapist (by considering only head-orientation)
Human coder	Time stamp	The coder feels that the child finishes facing the therapist (by considering only head-orientation)
System	1	The child is facing the camera
System	0	The child is not facing the camera, or face is not detected

Table X. Kappa coefficient between human coder and system's coding on therapist's vision

Session	Pair	Kappa
1	coder vs system(12-degree)	0.44(moderate)
1	coder vs system(15-degree)	0.28(fair)
2	coder vs system(12-degree)	0.39(fair)
2	coder vs system(15-degree)	0.17(slight)

To verify the results, two kinds of probability density chart was generated: (i) angle *Theta* in [0°, 30°] over the whole session, as shown in Figure 31(a), (ii) angle *Theta* when human coder felt that the child was facing the therapist (Figure 31(b)). The differences indicated that angle *Theta* of the two children's sessions were distributed in a wide range, while the human coder judged a smaller range as face-to-face, with peak located at around 6 degree in both two sessions. Therefore, when system used 12 degree which is more close 6 degree, the kappa was higher. Most probably, since the children often bowed their head to look at the toys used in the activity, it brought error to the coder's spatial judgment on vertical way.

5.1.2 Modeling of horizontal head movement

To model the horizontal head movement, all 5 children's sessions were processed to generate probability density of angle *Alpha* in [-30°, 30°]. The average of those results was then fitted with Gaussian distribution. The result (Figure 32, *Rnew:* 0.87) indicated that: (i) Those children tended to shift their heads to the left side during the activity. (ii) Comparing with angle *Theta*, which distributed in a wider range, angle *Alpha* was more distributed around [-10°, 10°]. This could infer - as mentioned in 5.1.1 - that the coder's spatial perception on vertical way might have a large error. (iii) A range [-16°, 16°] was required to cover more than 80% of the probability, which was wider than that of the robot-assisted activity.

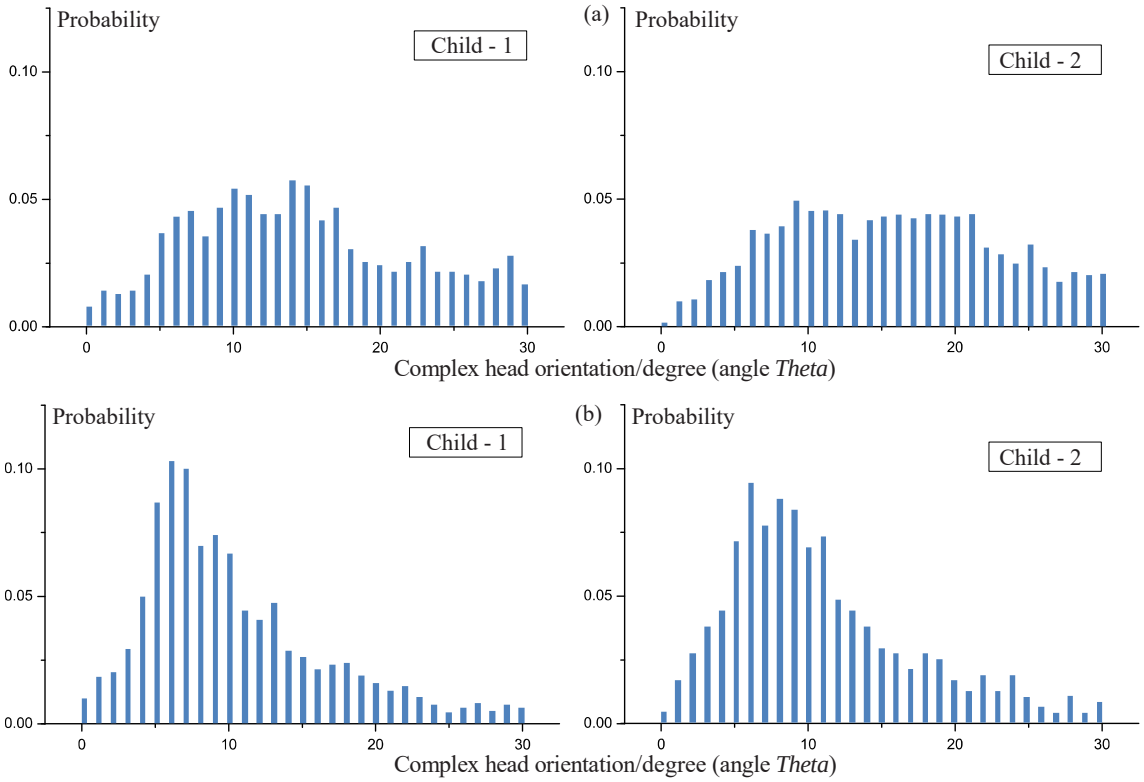

Figure 31. Probability density of two children's complex head-orientation in the activity with therapist
(a): in each whole session. (b): in case of face-to-face judged by human coder

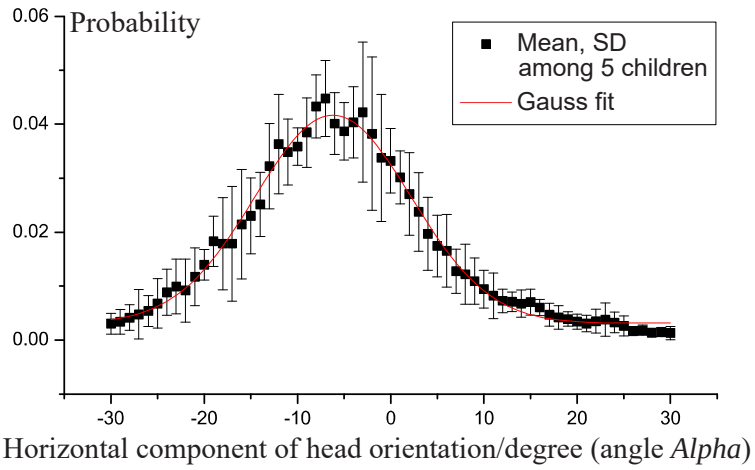

Figure 32. Gaussian of five children's horizontal head-orientation in the activity with therapist

5.2 Wearable Infrared Interface for Describing Face-to-Face Interaction

A head band-shaped wearable device (shown in Figure 33(a)) called FaceLooks is developed and used to apply the proposed model of face-to-face for application. The device has an IR emitter and receiver. When two people wear device and do face-to-face, each device will receive IR signal from the other one. The IR signal is programmed as sending the ID of each device, so that each device could know which one is facing it while receiving IR signal. A remote server manages all the events including one-way facing (one device receives another's IR signal) and face-to-face with timestamps and ID numbers. The radiant intensity of IR emitter on each device can be adjusted before/during its working. The half-angle of IR light changes according to different intensity. To set-up user experiments, I calibrated the system with a 20-degree threshold of IR half angle. I physically measured and prepared the situation of 20-degree facing between the two FaceLooks used in experiments, and adjusted the intensity of IR light on each device so that the IR communication just reached a stable status under 20-degree facing. The calibration was done under a 1.5-meter distance between two devices.

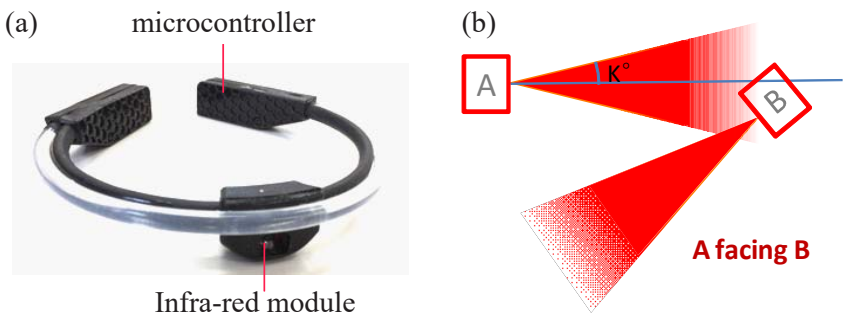

Figure 33. The device FaceLooks that measures facing and face-to-face
(a): Design of the device; (b): An example of detection of facing

I conducted a one-session experiment to verify the usefulness of the applied model as well as the developed device. Two subjects (S1, S2) wore both FaceLooks and SMI gaze tracker (Figure 34(a)). Two confederates (C1, C2) without wearing any device also participated in the session. Four people sit in F-formation (Figure 34(b)), which is simply the best way to enable each subject with an equal chance to interact with others (Kendon, 1976). The four people had a 10-minute free talk. During free talk, C1 and C2 could trigger S1 and S2's gaze move to left and right, and some objects on the table could trigger their gaze move up and down. I synchronized the time-line of FaceLooks with that of SMI gaze tracker, and recorded S1 and S2's face-to-face and one-way facing behavior, as well as their gaze movement.

Figure 34. Experiments for evaluation of FaceLooks

After the experiment, I extracted the 1/0 event of S1 and S2's face-to-face and one-way facing behavior from the server of FaceLooks. The sample rate was 30/second. Regarding gaze information, since two subjects wore SMI gaze tracker that blocked some features on the face, it was difficult to use Face API to analyze whether their gaze point was located in the other's facial region. I asked one human coder to code the first-person video recorded by gaze tracker, coding each frame (30 frames/second) with "1" if the subject's gaze point was within the other subject's head region (simply defined based on the size of head without extension), or with "0" if not in the region. Status marked with "1" was recognized as looking at face.

Figure 35 showed the events of facing and looking-at-face on the time-line. According to the results, 71.1% of "S1 looking at S2's face" happened during "S1 facing S2". On the other hand, 77.6% of "S2 looking at S1's face" was accompanied by "S2 facing S1". Regarding mutual behavior, 69.0% of "S1 and S2 looking at each other's face" happened at the moments when "S1 and S2 doing face-to-face". These values indicated that the face-to-face behavior - detected by using FaceLooks - could cover around 70% of the time when one subject was looking at another subject's face under the experiment context.

I also checked each subjects' head orientation during looking at the other subject's face (Figure 36(a)). Since gaze direction to the face can be aligned with the base line, angle alpha is numerically equal to the horizontal gaze vector. Figure 36(b, c) showed the probability density charts. S1's head orientation during looking at S2's face was distributed from -28° to 32°, with 92.4% in [-20°, 20°]. S2 had a narrower range from -23° to 18°, with 97.8% in [-20°, 20°]. I checked the probability within 20 degree because it is close to the half-angle of IR light I set for the device. Although the detected face-to-face did not reach the probability in [-20°, 20°] which is more than 90%, it was able to achieve a result around 70%.

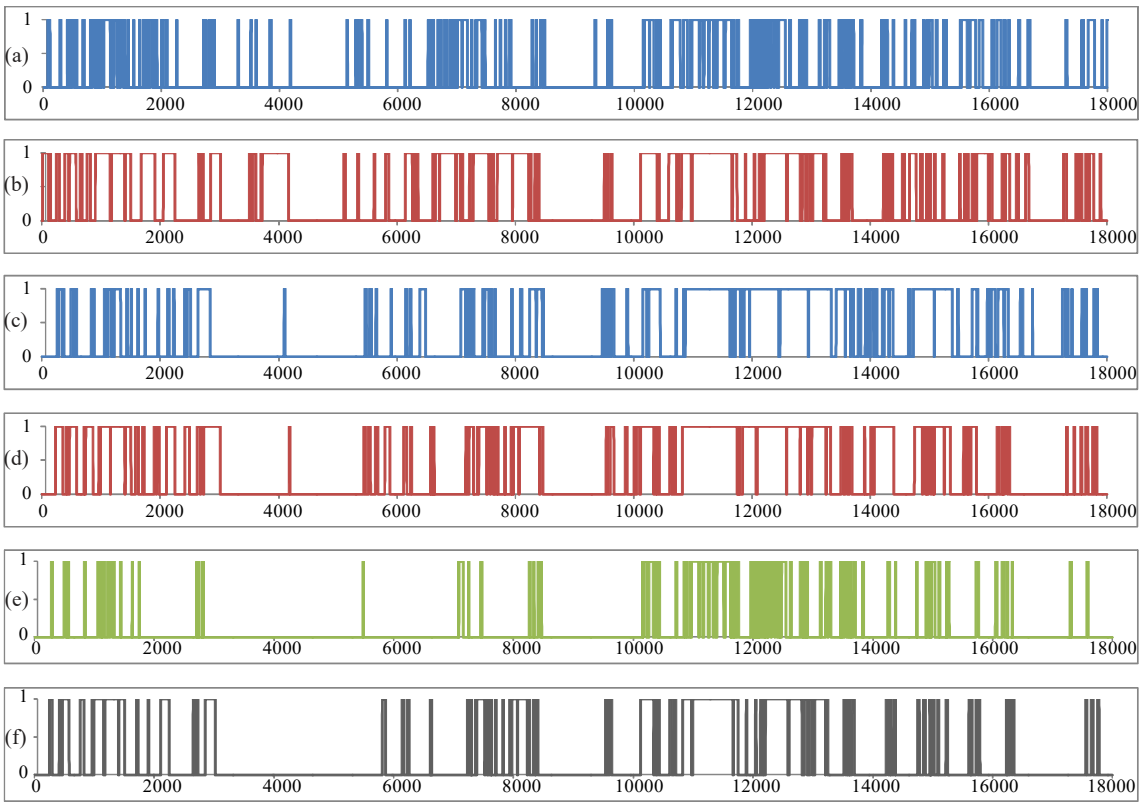

Figure 35. Events of eye-contact and facing/face-to-face in the experiment for evaluation of FaceLooks

(a): S1 looking at S2's face, by gaze tracker. (b): S1 facing S2, by FaceLooks. (c): S2 looking at S1's face, by gaze tracker. (d): S2 facing S1, by FaceLooks. (e): S1 and S2 looking at each other's face, by gaze tracker. (f): S1 and S2 doing face-to-face, by FaceLooks.

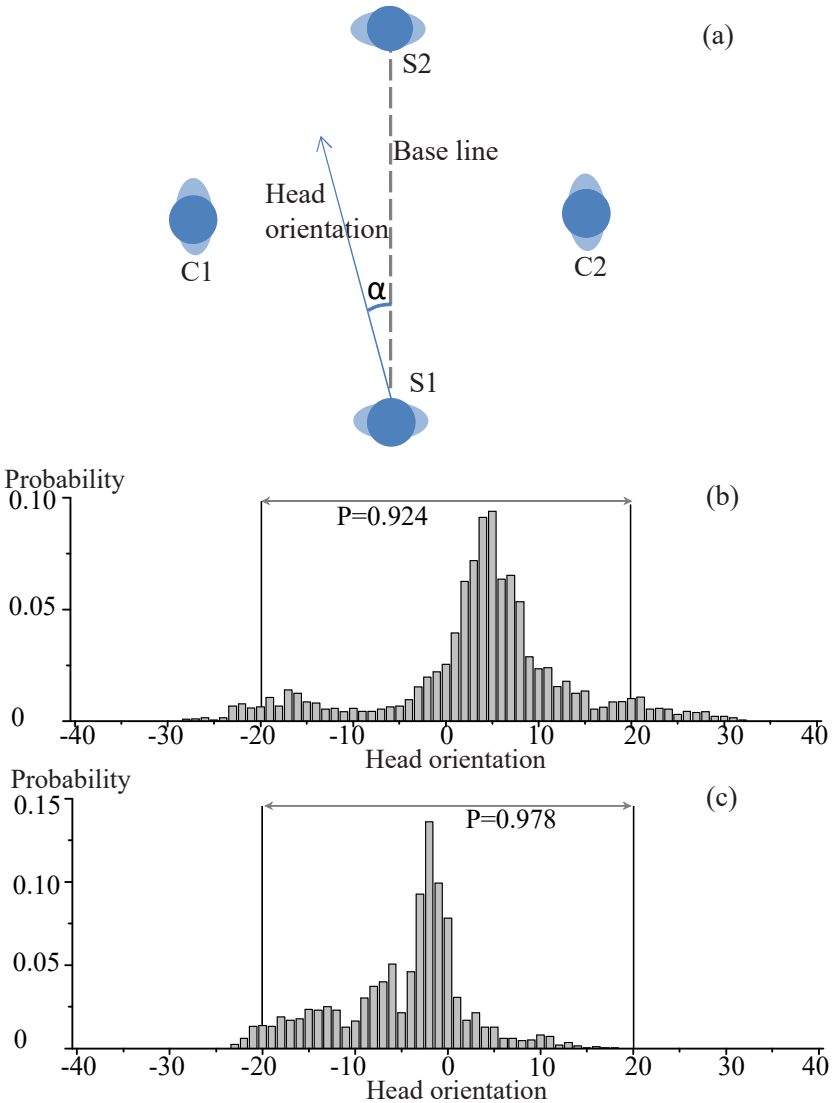

Figure 36. Probability density of head-orientation in the experiment for FaceLooks' evaluation (a): Description of head-orientation. (b, c) Probability density of S1 and S2's head orientation.

Chapter 6

Discussions

6.1 Experiments and Scenarios

6.1.1 Modeling of face-to-face

Under different scenarios, I gained different value of k for supporting k-degree face-to-face. In case of free talk, most probably, subjects tried keeping awareness towards all the people. Therefore even when they looked at and talked to a specific person, their head orientation was not too much close to that person, so that they could simply use gaze movement to shift attention to another person and talked to him/her. In contrast, in case of instructed scenario, subjects were asked to pay attention to specific persons, and their head orientation became closer to each specific person. The experiment results under human-human interaction is similar to a previous human-computer interaction research conducted by Fang et al (2013), in which the author investigated human subjects' head orientation in watching ultra high definition television. However, the findings under the two scenarios could only supported the theory of VFOA with small gaze shift (<30°). It is suggested that head tends to be stable under small gaze shift, whereas when gaze is shifted to peripherally area (>30°), gaze movement is more accompanied by head movement (Tweed, Glenn, and Vilis, 1995; Stahl. 1999). Regarding peripherally areas, the value of k in k-degree face-to-face might be different from the experiment results.

I separated the non-instructed and instructed scenario to create model of face-to-face. This is because those two scenarios might be used for different purpose. For example, non-instructed scenario could be planned in public spaces during people's social activities (e.g. ball games), whereas instructed scenario might be useful as a way of training social orienting

capability for children with developed disorders (e.g. children with ASD).

6.1.2 Experiments with autistic children

Experiments with children with ASD were conducted under different setting (human-robot interaction, human-human interaction). Throughout the experiments, I compared human coder's to the automatic system's judgment on face-to-face. Such comparison could indicate people's perception to face-to-face interaction from the coder's perspective. On the other hand, during the experiments for modeling of face-to-face, I investigated people's natural gaze-movement during fixing their attention on others' face or some pictures of face on the wall. The thresholds obtained in order to support k-degree face-to-face could be understood as from the actor's perspective. These two perspectives are interactive factors, as what we perceive could reflect what we perform in terms of body language.

In the robot-assisted activities, because of the close distance between the child and robot, and the one-v.s-one experiment setting, I hypothesize that human coders' criteria for judgment of face-to-face might be strict. Based on such hypothesis, 12 degree - a strict threshold of face-to-face - was used by the automatic system to process the experiment video. By comparing the system' to human coder's coding, moderate kappa coefficient was obtained, which suggested system and human coder had similar judgment. In the analysis of children's horizontal head movement, I found that the peak of Gaussian was located around 0 degree, indicating that those children tend to move their head directly towards the robot's face or body. Most probably, it is because the robot was the only cue for attention in the activity. Under the experiment context, the children do not need to keep awareness of other attention cues but could keep facing to a certain range surrounding the robot. In such case, human's subjective might not affect too much on their judgment of face-to-face when looking at the video, compared to the situation when children tend to use large angle of head orientation in the activity. What's more, from the Gaussian of horizontal head movement, the range [-11°, 11°] could cover more than 80% of probability. This is very close to the threshold we set for the automatic system (12 degree) to measure k-degree face-to-face.

In case of using first-person view camera on therapist's head to set-up activities, as shown in the probability density of two children's complex head orientation, the autistic children were more active than the robot's case by using larger angles of head orientation. This is because of the therapist put some toys on a table in front of the child to trigger the

affective response of those children, and the children often bow their heads to look at the toys. Such behavior might trigger more subjectivity on human coders' coding. Therefore, for those experiments I tried looking into the human coder's criteria of face-to-face. From Figure 31(b) we could understand that during the time when the human coder gave an answer Yes to judgment of face-to-face in the two sessions, the children's head orientation had a peak at around 6 degree and degree. Therefore, when applying 12 degree to the automatic system, the kappa value between system's and human coder's coding was higher than applying 15 degree to the system. Another important finding is that the 5 children who participated in the experiments tended to shift their heads to the left side with the peak located at around -6 degree (Figure 32). For therapists, such feature might help to understand the children's inner aspects as well as the effect of social activities.

Gaze information was not referred to in the experiments with autistic children, because it was difficult to ask them wear the heavy device that measures their gaze behavior. By looking at the experiment video, there were some cases that the autistic children faced but did not make eye-contact with the robot or therapists. This feature could refer to a previous study, in which the author investigate autistic children's gaze behavior and suggested they tended to shift their gaze-fixation from others' eyes to the mouth (Spezio et al. 2007). However, I counted such behavior as face-to-face, since such behavior provides chance for the children to have eye-contact with robots or therapists.

6.2 Modeling and Measurement

In this dissertation, the main contribution is "modeling of face-to-face", as I put "modeling" at the beginning of the title. Based on the experiments under different scenario, I proposed a method to model face-to-face by looking into the probability density map of the subjects' horizontal head orientation. The results showed that:

1) The means of horizontal head orientation of the 10 neurotypical subjects, 10 autistic children in robot-assisted activity, and 5 autistic children in activity with therapists, can be fitted with Gaussian distribution in each of the scenarios. This indicates that the subjects followed specific patterns to behave face-to-face interaction.
2) Face-to-face behavior has different features according to different scenarios. This could

be considered as a contextual factor that affect the relation between head orientation and visual focus of attention (VFOA), which contributes to the state-of-art of VFOA research field.

3) Based on the Gaussian distributions under each scenario, I calculated the ranges of head orientation that could cover 80% of the time when subjects' attention was fixed on others' face. The values are summarized in Table XI. Those values directly contributed to the design of the systems that measure face-to-face.

Table XI. Thresholds of k-degree face-to-face for neurotypical people and children with ASD

Subjects	Number of attention targets	Scenario	Threshold
Neurotypical people	2	Non-instructed	20
Neurotypical people	2	Instructed	15
Children with ASD	1	HRI	11
Children with ASD	1	HHI	16

Based on modeling's results, I proposed two approaches for measurement of face-to-face: camera-based approach and infra-red based approach. Both approaches could adjust the criteria for judgment of face-to-face according to different scenarios. What's more, camera-based approach can be used to create different systems such as robot-vision system, head-mounted camera system or 360-degree camera system. Each system has its certain benefits. For example, robot vision system was used for children with ASD, because those children have difficulty in human-human interaction; For children with ASD, head-mounted camera system was also used in the social activity between autistic children and therapists. In this case, therapists wore the camera in order to avoid instrumentation on the children's body. FaceLooks was developed based on IR communication. It is less dependent on lighting condition and speed of movement during face-to-face detection, compared to camera-based systems. However, it might have difficulty to use it for autistic children. Considering that the main contribution of this dissertation, which is modeling of face-to-face, the further evaluation of those systems is planned in other coming researches.

6.3 Technologies and Limitations

To model face-to-face by referring to gaze information, I used SMI gaze tracker in the research. In general, the machine learning-based approaches that recognize head orientation using camera systems also make inferences about gaze (Yang et al 2002; Lu et al. 2014; Zhang et al. 2015). I choose SMI gaze tracker because it is head mounted so that the sensors (IR cameras) are much more close to the eyes. This could help to make the estimated gaze region accurate to smaller area. Considering the size of face in subjects' field of view, it is required to make the gaze tracking accurate the face region. As mentioned, before the modeling of face-to-face, I first calibrated the gaze tracker, and ask each subjects to look at other's face in order to check the accuracy of gaze tracker particularly when looking at faces. Since the subject was not instructed with how to move head or gaze, the behavior that looking at face could be considered as natural behavior under the experiment setting.

However, the glasses-shaped device might be unnatural and affecting the social situation. Such issue might also exist for FaceLooks, and it is difficult to evaluate. As reference information, after I asked the two subjects who wear both SMI gaze tracker and FaceLooks whether they feel encouraged to look at face or tend to avoid looking at face. Both subjects answered that they did not feel the devices changes their natural behavior. What's more, SMI gaze tracker might potentially affect peripheral vision. To avoid such kind of unexpected issue as much as possible, during modeling face-to-face, the seats were arranged in regular triangle so that each subject could use at most 30-degree gaze movement to look at others. Actually since the subjects move not only eye-gaze but also head to reach the attention target, the gaze movement related to head orientation was even decreased. Under such experiment context, I considered that the gaze tracker would not affect the peripheral vision of the subjects.

6.4 Face-to-Face in Public Space

In the hotel public space, although a dominant factor in interaction design was the robots' different kinds of behavior, there was also face-to-face behavior that accompanied

the verbal interaction. In case of direct interaction style, the robot was programmed to face each guest then greet him/her, and many hotel guests stood to face the robot or approached the robot to have face-to-face in a close distance. Under the direct interaction style, face-to-face happened between human and the robot. On the other hand, in case of indirect interaction style, although the robot did not provide any facing behavior to the guests, it was observed that the guests tended to follow other guests' behavior that standing to look at the robots, which means, guests' behavior that facing the robot triggered other guests to face the robot. Whatever direct or indirect speech the robots provided, we could see face-to-face happened and accompanied the verbal interaction. However, in the hotel public space, the face-to-face behavior I analyzed was a general meaning of face-to-face in the guests' group dynamics. It extends the understanding of face-to-face from personal zone (0.4m-1.2m) to social zone (1.2m-3.6m).

6.5 Face-to-Face in terms of Inner Aspects and Multimodal Interaction

As mentioned in section 1.2, it is suggested by Ono et al (2012) that face-to-face behavior is correlated with mental health, however in Ono's work, the correlation could not indicate any changes of inner attitudes when having face-to-face. To fulfill this gap, it is required to have face-to-face interaction occur with particular perception in social interaction.

Face-to-face - as a fundamental interactive behavior - provides opportunity towards further interaction such as eye-contact, smile, getting close to each other, etc. To measure face-to-face together with other social behavior could contribute to the theory of multimodal interaction. For example, gaze with face-to-face and without face-to-face might have different meaning in social interaction. Measuring face-to-face is therefore with certain use as a practical way to create social imaging (Suzuki, 2015). This is a challenging topic for the future research.

Chapter 7

Conclusion

In this dissertation, I proposed a model of face-to-face interaction. The model refers to both geometric and perceptual information to describe face-to-face. Under each specific scenario, people's face-to-face behavior can be fitted with one Gaussian distribution. Referring to the proposed model, it is possible to select spatial thresholds of k-degree face-to-face, so that head orientation behaved during face-to-face could contribute a majority to visual focus of attention under specific scenarios. The model also provides theoretical support to creating different kinds of social activities.

The method used to model face-to-face - which looks into probability density of head orientation during looking at face - can be used in different scenarios to create statistical behavior analysis. Based on the proposed model, several automatic systems that measure face-to-face behavior were developed and used in different activities to describe social interaction.

Bibliography

Aarts Henk, and Ap Dijksterhuis. "Habits as knowledge structures: automaticity in goal-directed behavior." Journal of personality and social psychology 78.1 (2000): 53.

Anstis Stuart M., John W. Mayhew, and Tania Morley. "The perception of where a face or television 'portrait' is looking." The American journal of psychology 82.4 (1969): 474-489.

Ba Sileye O., and Jean-Marc Odobez. "Recognizing visual focus of attention from head pose in natural meetings." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 39.1 (2009): 16-33.

Billinghurst Mark, and Thad Starner. "Wearable devices: new ways to manage information." Computer 32.1 (1999): 57-64.

Blehar Mary C., Alicia F. Lieberman, and Mary D. Salter Ainsworth. "Early face-to-face interaction and its relation to later infant-mother attachment." Child development (1977): 182-194.

Bruce Allison, Illah Nourbakhsh, and Reid Simmons. "The role of expressiveness and attention in human-robot interaction." Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on. Vol. 4. IEEE, 2002.

Cattuto Ciro, et al. "Dynamics of person-to-person interactions from distributed RFID sensor networks." PloS one 5.7 (2010): e11596.

Chapple Eliot D., and Carleton S. Coon. "Principles of anthropology." the University of Michigan, 1942.

Cline Marvin G. "The perception of where a person is looking." The American journal of psychology 80.1 (1967): 41-50.

Cohn Jeffrey F., et al. "Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles." Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on. IEEE, 2004.

Dawson Geraldine, et al. "Children with autism fail to orient to naturally occurring social stimuli." Journal of autism and developmental disorders 28.6 (1998): 479-485.

Duranti Alessandro. "Universal and Culture-Specific Properties of Greetings." Journal of Linguistic Anthropology 7.1 (1997): 63-97.

El Kaliouby Rana, Rosalind Picard, and Simon Baron-Cohen. "Affective computing and autism." Annals of the New York Academy of Sciences 1093.1 (2006): 228-248.

Eyssel Friederike, et al. "'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism." Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. ACM, 2012.

Fang Yu, et al. "Eye position distribution depending on head orientation in watching Ultra High Definition Television." Asia-Pacific Conference on Vision (APCV2013). 2013.

Ferrari Ester, Ben Robins, and Kerstin Dautenhahn. "Therapeutic and educational objectives in robot assisted play for children with autism." RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, 2009.

Go Kentaro, and John M. Carroll. "Scenario-based task analysis." The handbook of task analysis for human-computer interaction (2004a): 117-133.

Go Kentaro, and John M. Carroll. "The blind men and the elephant: Views of scenario-based system design." interactions 11.6 (2004b): 44-53.

Hall Edward Twitchell. "The hidden dimension." (1966).

Hayashi Kotaro, et al. "Humanoid robots as a passive-social medium-a field experiment at a train station." Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on. IEEE, 2007.

Hirokawa Masakazu, et al. "A doll-type interface for real-time humanoid teleoperation in robot-assisted activity." Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction. ACM, 2014.

Horprasert Thanarat, Yaser Yacoob, and Larry S. Davis. "Computing 3D head orientation from a monocular image sequence." 25th Annual AIPR Workshop on Emerging Applications of Computer Vision. International Society for Optics and Photonics, 1997.

Iida Kazuki, and Kenji Suzuki. "Enhanced touch: a wearable device for social playware." Advances in Computer Entertainment Technology. 2011.

Ikeda Hiroo, Internal Report, NEC (2013)

Jarke Matthias, X. Tung Bui, and John M. Carroll. "Scenario management: An interdisciplinary approach." Requirements Engineering 3.3-4 (1998): 155-173.

Kanda Takayuki, et al. "An affective guide robot in a shopping mall." Proceedings of the 4th ACM/IEEE international conference on Human robot interaction. ACM, 2009.

Kasari Connie, Stephanny Freeman, and Tanya Paparella. "Joint attention and symbolic play in young children with autism: A randomized controlled intervention study." Journal of Child Psychology and Psychiatry 47.6 (2006): 611-620.

Keltner Dacher. "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame." Journal of Personality and Social Psychology 68.3 (1995): 441.

Kendon Adam, Richard M. Harris, and Mary R. Key, eds. "Organization of behavior in face-to-face interaction". Walter de Gruyter, 1975.

Kendon Adam. "The f-formation system: Spatial-orientational relations in face to face interaction." Man Environment Systems 6 (1976): 291-296.

Kraut Robert E., and Robert E. Johnston. "Social and emotional messages of smiling: An ethological approach." Journal of personality and social psychology 37.9 (1979): 1539.

Kret M. E., et al. "Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study." Neuroimage 54.2 (2011): 1755-1762.

Lambert David. Body language. HarperCollins, 2004.

Langton Stephen RH. "The mutual influence of gaze and head orientation in the analysis of social attention direction." The Quarterly Journal of Experimental Psychology: Section A 53.3 (2000): 825-845.

Lee Jaeryoung, et al. "Which robot features can stimulate better responses from children with autism in robot-assisted therapy?." International Journal of Advanced Robotic Systems 9 (2012).

Lee Min Kyung, and Maxim Makatchev. "How do people talk with a robot?: an analysis of human-robot dialogues in the real world." CHI'09 Extended Abstracts on Human Factors in Computing Systems. ACM, 2009.

Leekam Susan R., Emma Hunnisett, and Chris Moore. "Targets and Cues: Gaze - following in Children with Autism." Journal of Child Psychology and Psychiatry 39.7 (1998): 951-962.

Lu Feng, et al. "Learning gaze biases with head motion for head pose-free gaze estimation." Image and Vision Computing 32.3 (2014): 169-179.

Maruyama Kinya, and Mitsuo Endo. "The effect of face orientation upon apparent

direction of gaze." Tohoku Psychologica Folia (1983).

Mills Albert J., et al. Organizational behaviour in a global context. University of Toronto Press, 2006.

Mondloch Catherine J., Nicole L. Nelson, and Matthew Horner. "Asymmetries of influence: differential effects of body postures on perceptions of emotional facial expressions." PloS one 8.9 (2013): e73605.

Mori T., et al. "High Quality Image Correction Algorithm with Cubic Interpolation and its Implementation of Dedicated Hardware Engine for Fish-eye Lens." journal-institute of image electronics engineers of Japan 36.5 (2007): 680.

Mundy Peter, Marian Sigman, and Connie Kasari. "A longitudinal study of joint attention and language development in autistic children." Journal of Autism and developmental Disorders 20.1 (1990): 115-128.

Mundy Peter. "Joint attention and social-emotional approach behavior in children with autism." Development and Psychopathology 7.01 (1995): 63-82.

Murphy-Chutorian Erik, and Mohan Manubhai Trivedi. "Head pose estimation in computer vision: A survey." IEEE transactions on pattern analysis and machine intelligence 31.4 (2009): 607-626.

Nakano Tamami, et al. "Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour." Proceedings of the Royal Society of London B: Biological Sciences (2010): rspb20100587.

Nardi Bonnie A., and Steve Whittaker. "The place of face-to-face communication in distributed work." Distributed work (2002): 83-110.

Nisbett Richard E., and Timothy D. Wilson. "Telling more than we can know: Verbal reports on mental processes." Psychological review 84.3 (1977): 231.

Nuñez Eleuda, Kyohei Uchida, and Kenji Suzuki. "PEPITA: A Design of Robot Pet Interface for Promoting Interaction." International Conference on Social Robotics. Springer International Publishing, 2013.

Occhialini James, et al. "Wearable system for positive airway pressure therapy." U.S. Patent Application No. 10/982,958. (2004)

Online citation a: Beautycheck - average faces:
www.beautycheck.de

Online citation b: Openni: Open-source sdk for 3d sensors:
www.openni.org

Online citation c: Computational Behavioral Science - Georgia Tech: www.cbs.gatech.edu

Ono Eisuke, et al. "Fundamental deliberation on exploring mental health through social interaction pattern." Complex Medical Engineering (CME), 2012 ICME International Conference on. IEEE, 2012.

Otsuka Kazuhiro, Yoshinao Takemae, and Junji Yamato. "A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head directions, and utterances." Proceedings of the 7th international conference on Multimodal interfaces. ACM, 2005.

Otsuka Rieko, Kazuo Yano, and Nobuo Sato. "An organization topographic map for visualizing business hierarchical relationships." 2009 IEEE Pacific Visualization Symposium. IEEE, 2009.

Perrett David Ian, et al. "Organization and functions of cells responsive to faces in the temporal cortex [and discussion]." Philosophical Transactions of the Royal Society of London B: Biological Sciences 335.1273 (1992): 23-30.

Pitsch Karola, et al. "“The first five seconds”: Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI." RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, 2009.

Platon Eric, Nicolas Sabouret, and Shinichi Honiden. "Overhearing and direct interactions: Point of view of an active environment." International Workshop on Environments for Multi-Agent Systems. Springer Berlin Heidelberg, 2005.

Ponsa Daniel, et al. "3D vehicle sensor based on monocular vision." Proceedings. 2005 IEEE Intelligent Transportation Systems, 2005.. IEEE, 2005.

Poyatos Fernando. "Language and nonverbal systems in the structure of face-to-face interaction." Language & Communication 3.2 (1983): 129-140.

Rae Robert, and Helge J. Ritter. "Recognition of human head orientation based on artificial neural networks." IEEE Transactions on neural networks 9.2 (1998): 257-265.

Rajagopalan Shyam Sundar. "Computational behaviour modelling for autism diagnosis." Proceedings of the 15th ACM on International conference on multimodal interaction. ACM, 2013.

Robertson Neil, and Ian Reid. "Estimating gaze direction from low-resolution faces in video." European Conference on Computer Vision. Springer Berlin Heidelberg, 2006.

Robins Ben, et al. "Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?." Universal Access in the Information Society 4.2 (2005): 105-120.

Robins Ben, Kerstin Dautenhahn, and Paul Dickerson. "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot." Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on. IEEE, 2009.

Rocco Elena. "Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., 1998.

Royer Eric, et al. "Monocular vision for mobile robot localization and autonomous navigation." International Journal of Computer Vision 74.3 (2007): 237-260.

Sabelli Alessandra Maria, Takayuki Kanda, and Norihiro Hagita. "A conversational robot in an elderly care center: an ethnographic study." Proceedings of the 6th international conference on Human-robot interaction. ACM, 2011.

Schegloff Emanuel Abraham. The first five seconds: The order of conversational opening. University of California, 1967.

Schement Jorge Reina, and Brent D. Ruben, eds. "Between communication and information". Vol. 4. Transaction Publishers, 1993.

Schon Donald A. "Technology and change: The new Heraclitus." Seymour Lawrence, 1967.

Sheikhi Samira, and Jean-Marc Odobez. "Investigating the midline effect for visual focus of attention recognition." Proceedings of the 14th ACM international conference on Multimodal interaction. ACM, 2012.

Spezio Michael L., et al. "Analysis of face gaze in autism using "Bubbles"." Neuropsychologia 45.1 (2007): 144-151.

Stahl John S. "Amplitude of human head movements associated with horizontal saccades." Experimental brain research 126.1 (1999): 41-54.

Stiefelhagen Rainer, and Jie Zhu. "Head orientation and gaze direction in meetings." CHI'02 Extended Abstracts on Human Factors in Computing Systems. ACM, 2002.

Stiefelhagen Rainer, Jie Yang, and Alex Waibel. "Estimating focus of attention based on gaze and sound." Proceedings of the 2001 workshop on Perceptive user interfaces. ACM, 2001.

Stiefelhagen Rainer, Jie Yang, and Alex Waibel. "Modeling focus of attention for meeting

indexing based on multiple cues." IEEE Transactions on Neural Networks 13.4 (2002): 928-938.

Sullivan Larry E. "The SAGE glossary of the social and behavioral sciences." Sage, 2009.

Sung Michael, Carl Marci, and Alex Pentland. "Wearable feedback systems for rehabilitation." Journal of neuroengineering and rehabilitation 2.1 (2005): 1.

Suzuki Kenji. "Social imaging technology to identify and represent social behaviors." Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers. ACM, 2015.

Tanaka Fumihide, Aaron Cicourel, and Javier R. Movellan. "Socialization between toddlers and robots at an early childhood education center." Proceedings of the National Academy of Sciences 104.46 (2007): 17954-17958.

Thrun Sebastian, et al. "MINERVA: A second-generation museum tour-guide robot." Robotics and automation, 1999. Proceedings. 1999 IEEE international conference on. Vol. 3. IEEE, 1999.

Tsuji Airi, Soichiro Matsuda, and Kenji Suzuki. "Interpersonal Distance and Face-to-face Behavior During Therapeutic Activities for Children with ASD." International Conference on Computers Helping People with Special Needs. Springer International Publishing, 2016.

Tweed D., B. Glenn, and T. Vilis. "Eye-head coordination during large gaze shifts." Journal of neurophysiology 73.2 (1995): 766-779.

Van Breemen A. J. N., et al. "A user-interface robot for ambient intelligent environments." Proceedings of the 1st International Workshop on Advances in Service Robotics (ASER), Bardolino. 2003.

Watanabe Junji, et al. "Visual resonator: interface for interactive cocktail party phenomenon." CHI'06 Extended Abstracts on Human Factors in Computing Systems. ACM, 2006.

Yang Ruigang, and Zhengyou Zhang. "Eye gaze correction with stereovision for video-teleconferencing." European Conference on Computer Vision. Springer Berlin Heidelberg, 2002.

Zhang Xucong, et al. "Appearance-based gaze estimation in the wild." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

Publications

Journal Paper:

Yadong Pan, Haruka Okada, Toshiaki Uchiyama, Kenji Suzuki: "On the Reaction to Robot's Speech in a Hotel Public Space" International Journal of Social Robotics. Vol7(5), pp 911-920 (2015)

Proceedings in International Conferences:

Masakazu Hirokawa, Atsushi Funahashi, Yadong Pan, Yasushi Itoh, Kenji Suzuki: "Design of a Robotic Agent that Measures Smile and Facing Behavior of Children with Autism Spectrum Disorder" Proc. 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2016), pp 843-848 (2016)

Yadong Pan, Masakazu Hirokawa, Kenji Suzuki: "Measuring K-degree facial interaction between robot and children with autism spectrum disorders" Proc. 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2015), pp 48-53 (2015)

Yadong Pan, Haruka Okada, Toshiaki Uchiyama, Kenji Suzuki: "Direct and indirect social robot interactions in a hotel public space" Proc. 2013 IEEE International Conference on Robotics and Biomimetics (ROBIO2013), pp 1881-1886 (2013)

Yadong Pan, Haruka Okada, Toshiaki Uchiyama, Kenji Suzuki: "Listening to vs overhearing robots in a hotel public space" Proc. 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI2013) , pp 205-206 (2013)