Doctoral Dissertation

Eye Behaviors Using an Android Robot for Naturalistic Human-Robot Interaction in Face-to-Face Communication

Akishige Yuguchi

June 25, 2021

Graduate School of Information Science
Nara Institute of Science and Technology, Japan

A Doctoral Dissertation
submitted to the Graduate School of Information Science,
Nara Institute of Science and Technology, Japan
in partial fulfillment of the requirements for the degree of
Doctor of ENGINEERING

Akishige Yuguchi

Thesis Committee:

Professor Takahiro Wada	(Supervisor)
Executive Director Tsukasa Ogasawara	(Co-supervisor)
Professor Satoshi Nakamura	(Co-supervisor)
Visiting Professor Jun Takamatsu	(Co-supervisor)
Visiting Associate Professor Gustavo Alfonso Garcia Ricardez	(Co-supervisor)
Assistant Professor Sung-Gwi Cho	(Co-supervisor)

Eye Behaviors Using an Android Robot for Naturalistic Human-Robot Interaction in Face-to-Face Communication*

Akishige Yuguchi

Abstract

This dissertation clarifies how eye behaviors using an android robot (hereinafter referred to an android) with a human-like appearance affect the human impression on Human-Robot Interaction. In particular, this work focuses on eye behaviors of a humanoid robot toward a human (*i.e.*, face-to-face situation), which are important elements in nonverbal communication.

Since precise control of android's eye behaviors is fundamental for the clarification of eye behaviors' effects, it is necessary to calibrate its eye-direction controller for establishing human-android eye contact. First, this dissertation proposes a gaze calibration method for an android by using only orientation relationships between the coordinate systems of a camera, a robot, an external sensor system, which relates an eye direction with an input control command in order to accurately adjust the android's gaze. Moreover, a method to evaluate the effectiveness of the gaze calibration is proposed. In the evaluation, the subjects' perception of the calibrated android's gaze with a human gaze was compared.

Next, this dissertation investigates which eye behaviors make what impressions on humans and clarifies which are the important factors for attractive eye behaviors. Hence, I evaluate the human impression of eye behaviors displayed by an android while talking to a human by comparing the motions generated by the imitation-based, the rule-based, and the combined approaches. Through a subjective evaluation, four findings were reached: 1) the imitation and rule-based

*Doctoral Dissertation, Graduate School of Information Science, Nara Institute of Science and Technology, June 25, 2021.

behaviors showed no difference in terms of human-likeness, 2) the 3-second eye contact obtained better scores regardless of the imitation or rule-based eye behavior, 3) the subjects might regard the long eyeblinks as voluntary eyeblinks, with the intention to break eye contact, and 4) female subjects preferred short eyeblinks rather than long ones and considered that short eyeblinks might be one of the keys to make eye contact more suitable, in contrast to male subjects who preferred long eyeblinks.

Finally, this dissertation clarifies how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening. The hypothesis is that if an android acting as a listener imitates eyeblinks and nodding of a human speaker in face-to-face communication, the android can make a human speaker perceive its listening behavior as attentive listening. First, I develop a real-time method to imitate human eyeblinks and nodding using an android. Next, I evaluate the subjective impression of the imitation by comparing to 1) rule-based eyeblink and nodding, which are simple duration-based motions, 2) eyeblink or nodding generated at breakpoints, and 3) combined eyeblink and nodding generated at breakpoints and/or in simple duration. Through a subjective evaluation, two findings were reached: a) eyeblink was effective as a cue of attentive listening behaviors, b) the imitation of both eyeblinks and nodding did not improve the perception of attentive listening.

Keywords:

Human-Robot Interaction, Nonverbal Behaviors, Gaze, Eyeblink, and Android Robots

Contents

List of Figures

List of Tables

1. Introduction

1.1 Background and Objective

Up to the present, many kinds of robots have been developed and researched. Although industrial robots which work in well-controlled environment (*e.g.*, a factory isolated from working spaces for humans) are more effective for executing tasks nowadays, there are still many challenges in human-centered environments.

Looking back on the history of robotics, *Gakutensoku* as the first robot in Japan was developed in 1928. It has a face and two arms, and can move its facial area and arms, to explore to design human-likeness and animacy on robots [1]. Since this robot had fascinated people in some exhibitions, it indicates that people have a desire to interact with robots.

When attempting to interact with robots, communication situations occur. Communication consists of both verbal and nonverbal communication. It is said that nonverbal communication occupies 60-70% of human communication [2]. Among cues of nonverbal communication, gaze plays an important role, as concluded by Kleinke [3]. Eye contact, which is one of the functions of gaze, has a role in daily-life communication of coordination, *e.g.*, to seek information, to regulate interaction with others, and to signal when someones turn to speak [4, 5, 6]. McCarthy *et al.* [7] investigated the eye movements during thinking. The direction of the eye movement for breaking eye contact is related to the mental state. Nakano and Kitazawa [8] showed that eyeblinks of a listener are synchronized with eyeblinks of a speaker at breakpoints in a conversation.

On the other hand, as a robot platform for the research on human-robot interactive communication, android robots (hereinafter referred to as androids) have been developed [9]. The appearance of androids is very similar to that of human beings, as exemplified in Figure 1. Compared to more machine-like robots, people expect androids to exhibit sophisticated communication traits [10] because of their human-like appearance. However, after we recognize that the actual communication of the androids is poor, people usually get very disappointed since the androids do not meet their expectations [11]. To demonstrate all the interaction abilities of androids, realizing the emulation of human-like communication is essential.

The main question of this dissertation is what kinds of eye behaviors let us sense a naturalistic impression. We assume that, if androids can realize naturalistic motions for communication, the impression of androids will be improved. Hence, this dissertation clarifies how to generate eye behaviors using an android with a human-like appearance to make humans perceive naturalistic impressions on Human-Robot Interaction through subjective experiments.

Since precise control of android's eye behaviors is fundamental for the clarification of eye behaviors' effects, it is necessary to calibrate its eye-direction controller for establishing human-android eye contact. First, this dissertation proposes a gaze calibration method for an android robot, which relates an eye direction with an input control command in order to accurately adjust the android robot's gaze.

Next, this dissertation investigates which eye behaviors makes what impressions on humans and clarify which are the important factors for attractive eye behaviors.

Finally, this dissertation clarifies how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening.

1.2 Definition and Approach

- **Eye Behavior**
  Eye behavior is defined as 1) the attempt to establish eye contact (referred to as *eye-contact bids* [12]), 2) the eye movement to break eye contact, and 3) the eyeblink including their frequency and duration.

- **Accurate Gaze Control**
  To generate eye behaviors, especially eye contact, it is necessary to shift an android's gaze to look at a human face. To do this, it is required for controlling the android's gaze to follow the planned trajectory. Hence, we need to estimate the relationship between an eye direction and an input control command. We call this estimation *gaze calibration*.

- **Human-like Motion Generation**
  We assume that, if androids can realize naturalistic motions for communication, the impression of androids will be improved. One of the approaches to realize human-like eye behaviors is the imitation of human motions. This

Figure 1. Android robot *Actroid-SIT* by Kokoro Company Ltd.

means to imitate the trajectories and timing of actions obtained from observation. The approach to realize designed eye behaviors is a rule-based motion generation. This means to generate motions with rules in which the motion primitives and timing are described. To describe the rules, we use the knowledge from psychology and cognitive science. These two approaches may be in a trade-off relationship; the imitation-based method may rely on the imitated person and the rule-based approach may lead to one of two different results: negatively affect the impression due to non-human-likeness or positively affect the impression due to the human-likeness.

- **Robot Platform**
  As a common robot platform in this dissertation, we employ the android *Act roid-SIT* produced by Kokoro Company, Ltd., shown in Figure 1 and 2. The surface of this android is covered by skin colored silicone This android has 42 degrees-of-freedom (DOFs) and is actuated by pneumatic actuators. Out of all DOFs, 13 DOFs are used for facial area movements. Especially, two DOFs (pitch and yaw axes) are used for eyeballs. The resolution of the actuator controller is 8 bits. The input value for each actuator controller is related to the actuator displacement, which covers a range of 256 steps. Thus, the appearance of this android can give humans a human-like impression.

1.3 Contributions

The contributions of this dissertation are threefold:

1. **Gaze Calibration for Human-Android Eye Contact**
   We proposed a gaze calibration method for human-android eye contact by using only orientation relationships between the coordinate systems of a camera, a robot, an external sensor system, which relates an eye direction with an input control command in order to shift an android's gaze accurately to gaze at a human, *i.e.*, make eye contact. (Chapter 3).

2. **Eye Behaviors for Naturalistic Talking**
   We investigated which eye behaviors make what impressions on humans

and clarified which are the important factors for attractive eye behaviors. Through a subjective evaluation, we reached four findings: 1) the imitation and rule-based behaviors showed no difference in terms of human-likeness, 2) the 3-second eye contact obtained better scores regardless of the imitation or rule-based eye behavior, 3) the subjects might regard the long eyeblinks as voluntary eyeblinks, with the intention to break eye contact, and 4) female subjects preferred short eyeblinks rather than long ones and considered that short eyeblinks might be one of the keys to make eye contact more suitable, in contrast to male subjects who preferred long eyeblinks. (Chapter 4).

3. **Eye Behaviors for Attentive Listening**
   We clarified how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening. Through a subjective evaluation, we reached two findings: 1) eyeblink was effective as a cue of attentive listening behaviors, 2) the imitation of both eyeblinks and nodding did not improve the perception of attentive listening. (Chapter 5).

1.4 Dissertation Layout

The rest of this dissertation is organized as follows:

Chapter 2:
This chapter summarizes the related works on eye movements and eye behaviors for human-robot communication using humanoid robots and android robots.

Chapter 3:
This chapter proposes a gaze calibration method for human-android eye contact, which relates an eye direction with an input control command in order to accurately adjust the android robot's gaze and also proposes and describes a method to evaluate the effectiveness of the calibration method.

Chapter 4:
This chapter describes eye behaviors for natural talking to a human. First, the imitation method of eye movements and eyeblink is developed. Second, the rule-based method of the generation of eye movements and eyeblink is

developed from existing knowledge of the psychology and cognitive research. Finally, the subjective experiment for the evaluation of the impression of the eye behaviors is conducted by comparing the imitation-based, the rule-based, and the combined approaches.

Chapter 5:

This chapter describes eye behaviors for attentive listening to a human. First, to clarify how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening, a real-time method to imitate human eyeblinks and nodding using an android is developed. Next, the subjective experiment for the evaluation of the impression of imitation is conducted by comparing the imitation to 1) the rule-based eyeblinks and nodding, which are simple duration-based motions, 2) eyeblink or nodding generated at breakpoints, and 3) combined eyeblink and nodding generated at breakpoints and/or in simple duration.

Chapter 6:

This chapter summarizes and concludes the dissertation. Additionally, some possible future work is presented.

Figure 2. Android robot *Actroid-SIT* changed its clothes

2. Related Works

2.1 Eye Movement Control

There are two problems to consider when generating the gaze behaviors for androids. The first problem is the planning of motions including the gaze. Kondo *et al.* [13] proposed a gaze motion planning method based on the convergence of the eyes and the ratio between the angle of the android's gaze and the angle of the android's head angle. Yamamori *et al.* [14] investigated the conditions for establishing eye contact when the eyes of an android present saccades around a stationary position.

The second problem is controlling the android's gaze to follow the planned trajectory. To solve this problem, we need to estimate the relationship between the gaze in the coordinate system of the android's head and the input values of the actuator controller in the android's eyes.

If cameras are built into the android's eyes, the calibration can be performed by moving markers to several fixed positions in displayed images, similarly to the method of calibration of human gaze tracking devices, *e.g.*, *Tobii Pro eye tracking*[1]. However, in the actual case of androids with built-in cameras [15, 16], these cameras cannot be used for controlling the android's gaze. Palinko *et al.* [15] have reported that it is impossible to track human gaze using these cameras because the cameras are located behind the plastic covers of the eyes, which blurs the image. Even in the latest androids such as *ERICA*[16], the cameras can be used only for human face tracking, and it is difficult to track small objects such as calibration markers.

As most androids do not have built-in cameras, we need to use different methods to add constraints between the orientation of the android's eyes and the gaze direction. Even *et al.* [17] proposed a gaze control method for an android to make eye contact with humans using a network of external sensors such as laser range finders and microphone arrays. They reported that calibration is required to achieve eye contact. They performed calibration between the human-perceived gaze angle around the yaw axis and the android's target gaze angle obtained by using the sensor network. However, they assumed that the relationship between

[1] Tobii Pro, https://www.tobiipro.com/

the target angle and the input values of the actuator controller was known *a priori*.

2.2 Eye Behaviors

Previous research has shown the effectiveness of an android displaying eye behavior. Minato *et al.* [18] evaluated human-likeness of an android by comparing gaze behaviors in human-android interaction to that in human-human interaction. Shimada *et al.* [19] evaluated the effect of eye contact between a human and an android. Tatsukawa *et al.* [20] investigated the synchronization of eyeblinks between an android and a human. Lala *et al.* [21] used the eye movement of an android to let a speaker know that the android is listening to a participant and to take turns in the conversation. Kondo *et al.* [22] used the gaze of a receptionist android to increase the number of people who voluntarily talk to the android. Luo *et al.* [23] attempted to express personality with an android by controlling the eye movements. Iwamoto *et al.* [24] found out the combination between the eye directions and the timing of eyeblinks as one of the subconscious behaviors that can become a deception cue when an android tells a lie in a simple game.

Other studies have shown the effectiveness of the robots' eye behaviors even without using androids, *e.g.*, using social humanoid robots with human-like eye mechanisms. Yoshikawa *et al.* [25] investigated the effect of the responsive eye movement and blinking behavior using a humanoid robot called *Robovie-R2*. Lehmann *et al.* [26] investigated the influence of eyeblinking behavior based on human physiological data using a humanoid robot called *iCub*. *Disney Research* developed an architecture that seeks not only to create gaze interactions from a technological standpoint, but also through the lens of character animation where the fidelity and believability of motion is paramount, using a humanoid bust [27]. Hardjasa *et al.* [28] found that cultural experience could have some significant effects on gaze behavior, particularly in averted gaze direction and frequency of gaze shifts or changes in gaze direction, using a small humanoid robot called *CommU*. In addition, to improve humanlikeness of humanoid robots' eye behaviors, Todo, an artist who focuses on gaze, especially eye contact, developed the animatronic humanoid robot *SEER* [29]. *SEER*, as an art piece, has fascinated people in various exhibitions[2].

To realize a human-like behavior, various approaches to imitate human behavior have been attempted. One of the methods is to model the relationship between stimuli from the outside and human reactions. Zaraki *et al.* [30] modeled the gaze movements using the observation of subjects and predict gaze movements from the visual stimuli. They also modeled the relationship between the frequency of eye movements and eyeblinks [31]. Some research methods proposed to construct gaze models with a probabilistic approach for robot heads to socially interact with humans. Hoffman *et al.* [32] proposed a probabilistic model of gaze directions during joint attention; the directions are achieved by a pan-tilt movement of a robot head. Duque-Domingo *et al.* [33] proposed a probabilistic model of head-eye movement ratio to control gaze directions for social interaction with multiple people. On the other hand, methods to imitate behaviors by playing back the observation with small editing are often used [22, 34, 35]. These imitation methods limit the applicable situations to those that are similar to the situations in the observation. However, the difficulty in explicitly modeling the human behaviors can be skipped.

[2] http://www.takayukitodo.com/#seer

3. Gaze Calibration for Eye Contact

In this chapter, we propose a gaze calibration method which relates an eye direction with an input control command, using a single camera as the gaze direction target with the android staring at the camera in various locations. The proposed method consists of two steps. First, as shown in Figure 3, we estimate the android's gaze direction from the android's head orientation and the camera orientation, only using the relationships between each coordinate system. Next, we model the relationship between the gaze direction and the input values of the actuator controller to create datasets to estimate the parameters of the modeled relationship.

We also propose a method for evaluating the effectiveness of the gaze calibration method. To do this, we compared subjects' perception of the calibrated android's gaze with a human gaze.

3.1 Calibration Method

Figure 4 shows an overview of our proposed calibration method. For gaze calibration, it is necessary to estimate the android's gaze directions using external sensors. We then perform gaze calibration by collecting the datasets of the relationships between an input value of the actuator controller and the corresponding gaze direction.

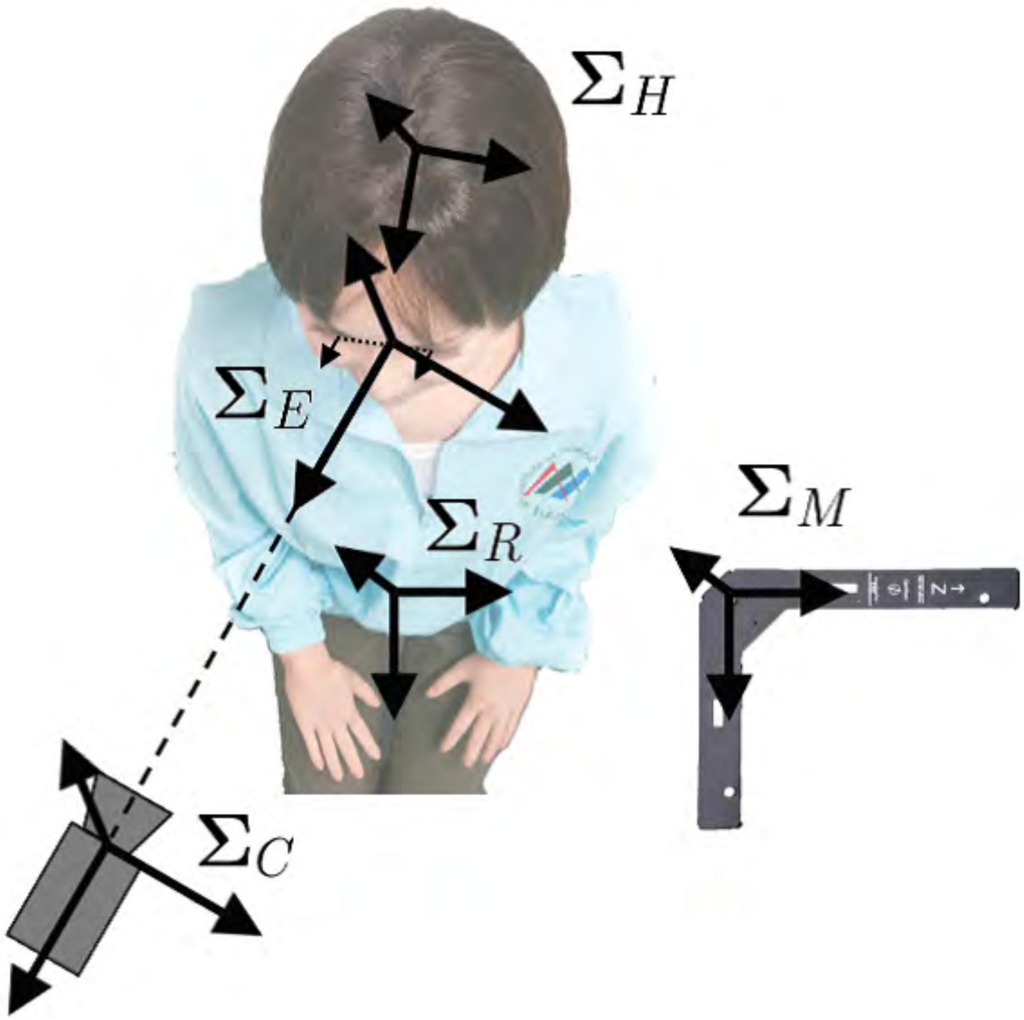

Figure 3. Relationships between the coordinate systems of the camera, the android's eyes, head, and base, and the motion capture system. The optical axis of the camera is aligned with the gaze direction of the android looking at the camera.

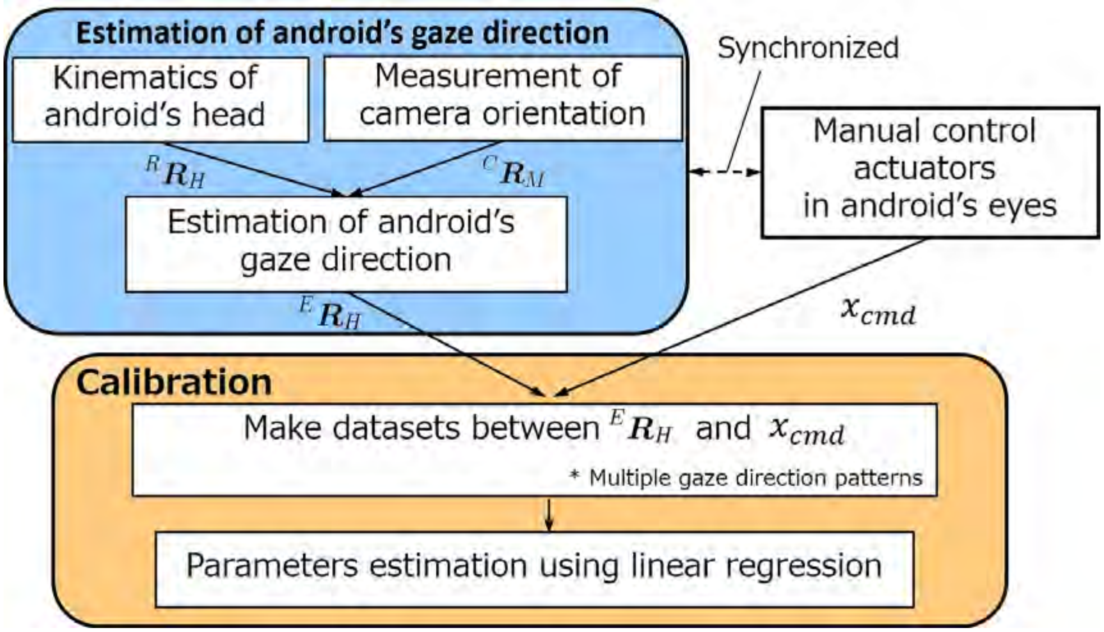

Figure 4. Overview of the proposed gaze calibration method.

3.1.1 Estimation of the Android's Gaze Direction

As shown in Figure 3, we assume the optical axis of the camera is aligned with the gaze direction. The coordinate systems of the android are defined as the body $\boldsymbol{\Sigma}_R$, the eyes $\boldsymbol{\Sigma}_E$, and the head $\boldsymbol{\Sigma}_H$, while the coordinate systems of the camera is $\boldsymbol{\Sigma}_C$, and the motion capture system is $\boldsymbol{\Sigma}_M$. The gaze direction has only two DOFs along the pitch axis (up/down) and the yaw axis (right/left) as we assume there are no roll movements of the eyes. With respect to the head orientation, then, the rotation matrix ${}^E\boldsymbol{R}_H$ of the android's gaze direction is defined as

$$ {}^E\boldsymbol{R}_H = \boldsymbol{R}_y(\theta_\beta)\boldsymbol{R}_z(\theta_\gamma). \tag{1}$$

The matrix $\boldsymbol{R}_y(\theta_\beta)$ is the rotation matrix of the pitch axis. The matrix $\boldsymbol{R}_z(\theta_\gamma)$ is the rotation matrix of the yaw axis.

The rotation matrix ${}^E\boldsymbol{R}_H$ from the coordinate system of the eyes to that of the head is expressed by

$$ {}^E\boldsymbol{R}_H = {}^E\boldsymbol{R}_C \, {}^C\boldsymbol{R}_M \, {}^M\boldsymbol{R}_R \, {}^R\boldsymbol{R}_H. \tag{2}$$

There are four rotation matrices from one coordinate system to another that we must define:

1. ${}^E\boldsymbol{R}_C$ from the coordinate system of the eyes to that of the camera,
2. ${}^C\boldsymbol{R}_M$ from the coordinate system of the camera to that of the motion capture system,
3. ${}^M\boldsymbol{R}_R$ from the coordinate system of the motion capture system to that of the android,
4. ${}^R\boldsymbol{R}_H$ from the coordinate system of the android to that of the head.

The matrix ${}^{R}\boldsymbol{R}_{H}$ is calculated from the android's forward kinematics. The matrix ${}^{C}\boldsymbol{R}_{M}$ is measured by the motion capture system and the matrix ${}^{M}\boldsymbol{R}_{R}$ is already known after we set up the measurement environment.

As mentioned, we assume that the camera optical axis is aligned with the gaze direction. When we set the coordinate system of the camera $\boldsymbol{\Sigma}_C$ so that the optical axis is a roll axis, the matrix ${}^{E}\boldsymbol{R}_{C}$ is expressed by

$$ {}^{E}\boldsymbol{R}_{C} = \boldsymbol{R}_{x}(\theta_\alpha), \tag{3} $$

where the matrix $\boldsymbol{R}_{x}(\theta_\alpha)$ is the rotation matrix around the roll axis.

Then, according to (2), the following relation can be obtained:

$$ \begin{aligned} {}^{C}\boldsymbol{R}_{E}\ {}^{E}\boldsymbol{R}_{H} &= \boldsymbol{R}_{x}(-\theta_\alpha)\boldsymbol{R}_{y}(\theta_\beta)\boldsymbol{R}_{z}(\theta_\gamma) \\ &= {}^{C}\boldsymbol{R}_{M}\ {}^{M}\boldsymbol{R}_{R}\ {}^{R}\boldsymbol{R}_{H}. \end{aligned} \tag{4} $$

We can solve the three values, $\theta_\alpha, \theta_\beta$, and θ_γ in (4) analytically because the right side of the equation in the second line is already known. Therefore, we can obtain the parameters, θ_β and θ_γ about the gaze direction.

3.1.2 Calibration Procedure

To estimate the gaze direction, we have to align the camera optical axis and the gaze. To achieve this, we first align the eye positions with the center of the camera lens as the camera optical axis goes through the center of the eye ball. Next, we control the gaze direction to look at the camera so that the gaze goes through the camera optical center. As a result we can align the gaze and the camera optical center.

The procedure for the calibration is as follows:

1. Setup the camera so that the android's eyes are centered.
2. Adjust the actuator input values so that the android stares at the camera by watching the image obtained from the camera.
3. Estimate the gaze direction using (4).
4. Record the final values of the actuators and the gaze direction into the datasets.
5. Repeat 1 to 4 in various positions.

3.1.3 Calibration from Datasets

The equation of a relationship between the gaze direction and the input values of the actuator controller is modeled as

$$\theta_\beta = a_1 x_{cmd_1} + b_1, \tag{5}$$

$$\theta_\gamma = a_2 x_{cmd_2} + b_2, \tag{6}$$

where θ_β and θ_γ are the pitch and yaw angles of the gaze direction, and x_{cmd_1} and x_{cmd_2} are the input values of the controller for the eyes's actuators. To calculate a_1, b_1, a_2, b_2, the parameters of the relationship, we use simple linear regression with the datasets between θ and x_{cmd}.

3.2 Implementation

For the implementation, we used the android *Actroid-SIT* as aforementioned in Chaper 1.2.

3.2.1 Parameter Estimation

As shown in Figure 5, we set up the measurement environment to estimate the android's gaze direction and make datasets for calibration. A camera with reflective markers as the gaze direction targets, shown in Figure 6 was placed in front of the android inside the range of a motion capture system. We measured the camera's orientation with the motion capture system *OptiTrack*[3] when the android was looking directly into the camera, as shown in Figure 7.

We generated datasets for eighteen different positions by adjusting the actuator input values for the eyes so that the android stared at the camera. The estimated regression lines expressed by

$$\theta_\beta = 0.104\ x_{cmd_1} - 8.95, \tag{7}$$

$$\theta_\gamma = 0.465\ x_{cmd_2} - 63.97. \tag{8}$$

Figure 8 and 9 are plots of the datasets and the regression lines of the eyes' pitch and yaw.

[3]OptiTrack, http://www.optitrack.com

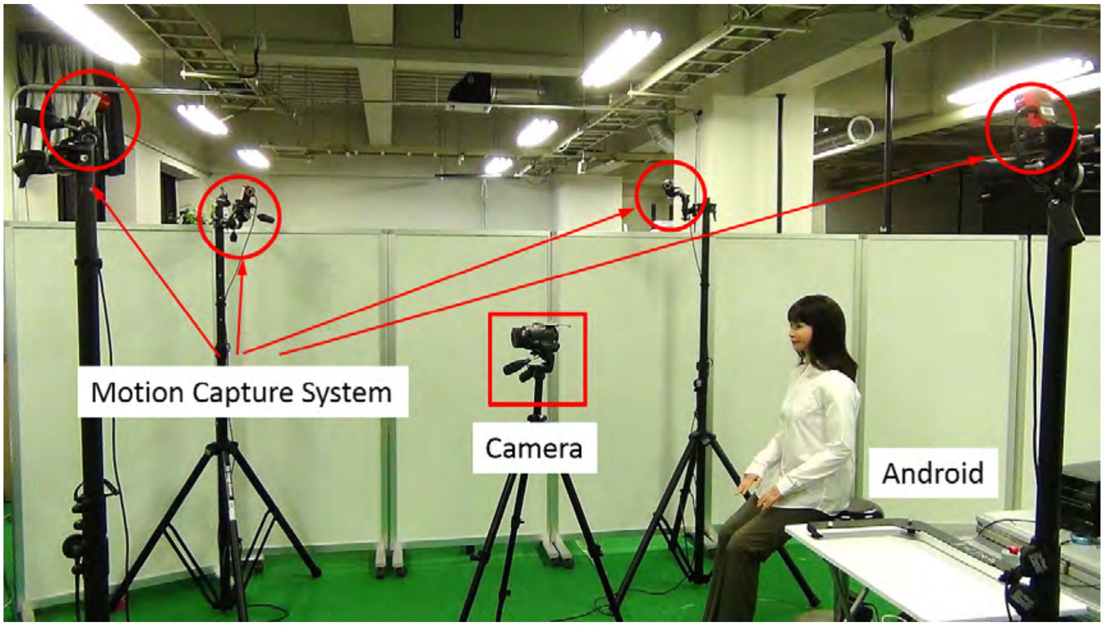

Figure 5. Calibration environment.

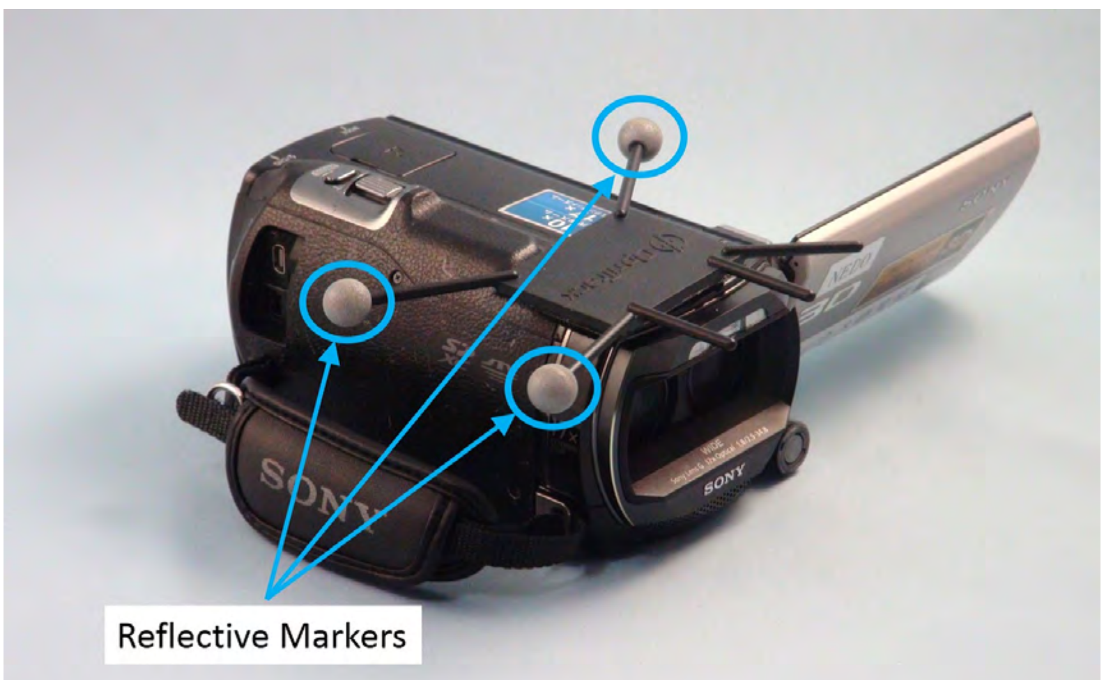

Figure 6. A camera with reflective markers as the gaze direction target.

Figure 7. The android is looking directly into the camera.

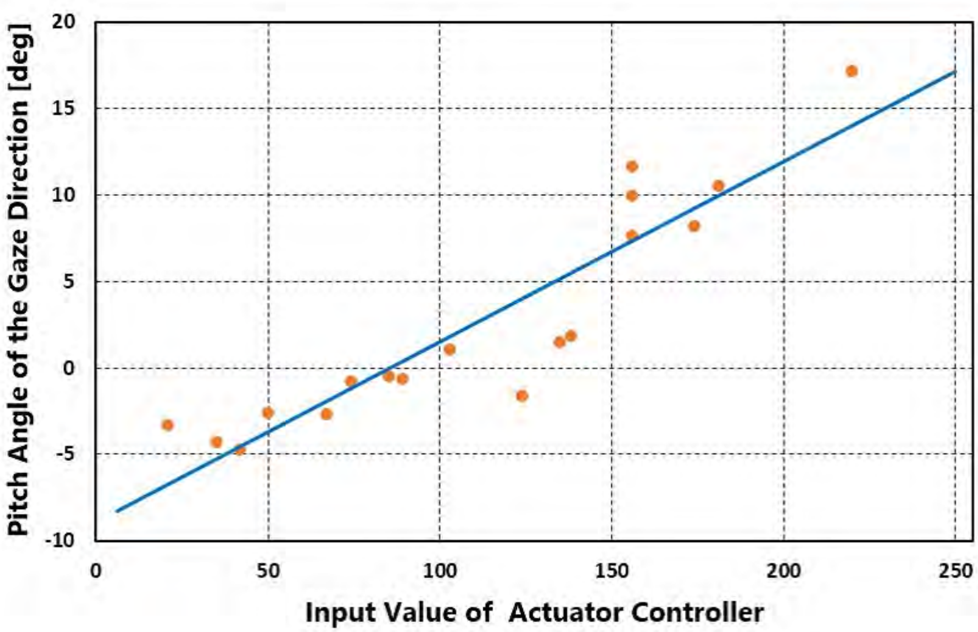

Figure 8. Calibration results of eyes' pitch. The points are the datasets. The line is the regression line.

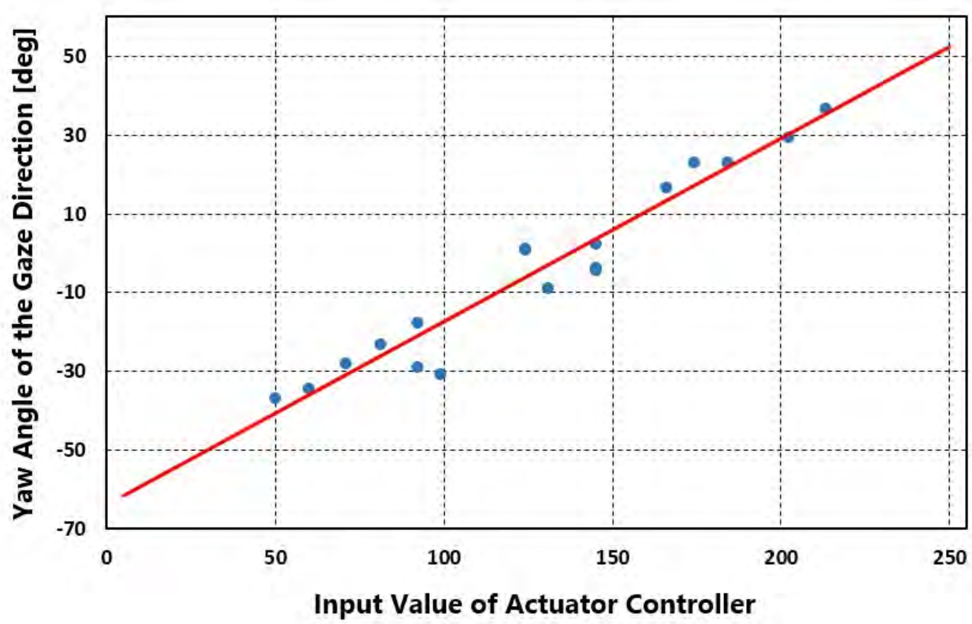

Figure 9. Calibration results of eyes' yaw. The points are the datasets. The line is the regression line.

Table 1. T-test results of linear regression.

	Eyes' pitch		Eyes' yaw	
	slope	intercept	slope	intercept
p-value	7.81e-08 ***	9.10e-06 ***	5.23e-11 ***	4.48e-11 ***

***: $p < 0.001$

Table 2. Coefficient of determination results.

Eyes' pitch	Eyes' yaw
0.84	0.94

To analyze the confidence of the regressions, we used the t-test and calculated coefficients of determination. Table 1 shows the t-test results. The results indicate a significant difference in the pitch and yaw of the eyes. Table 2 shows the coefficients of determination. These results indicate high scores for the coefficients of determination of both the pitch and yaw of the eyes. Therefore, we conclude that we have high confidence in the estimated regressions.

3.2.2 Gaze Behavior Generation

Gaze behavior can be generated using (7), (8) to determine the eye's pitch and yaw for any gaze direction. We tested the effectiveness of the proposed method by examining the differences between images from a fixed camera and images from the camera as the gaze direction target. Figure 10 and 11 show examples of the changes in the gaze when the eyes' pitch and yaw are changed in each direction. The upper row shows the images obtained from a camera fixed in front of the android. In this case, eye contact is achieved with the center image. The lower row shows images obtained from a camera in the direction of the gaze. In all of these images, eye contact is achieved. We conclude from this that the calibrated gaze can be directed in various directions.

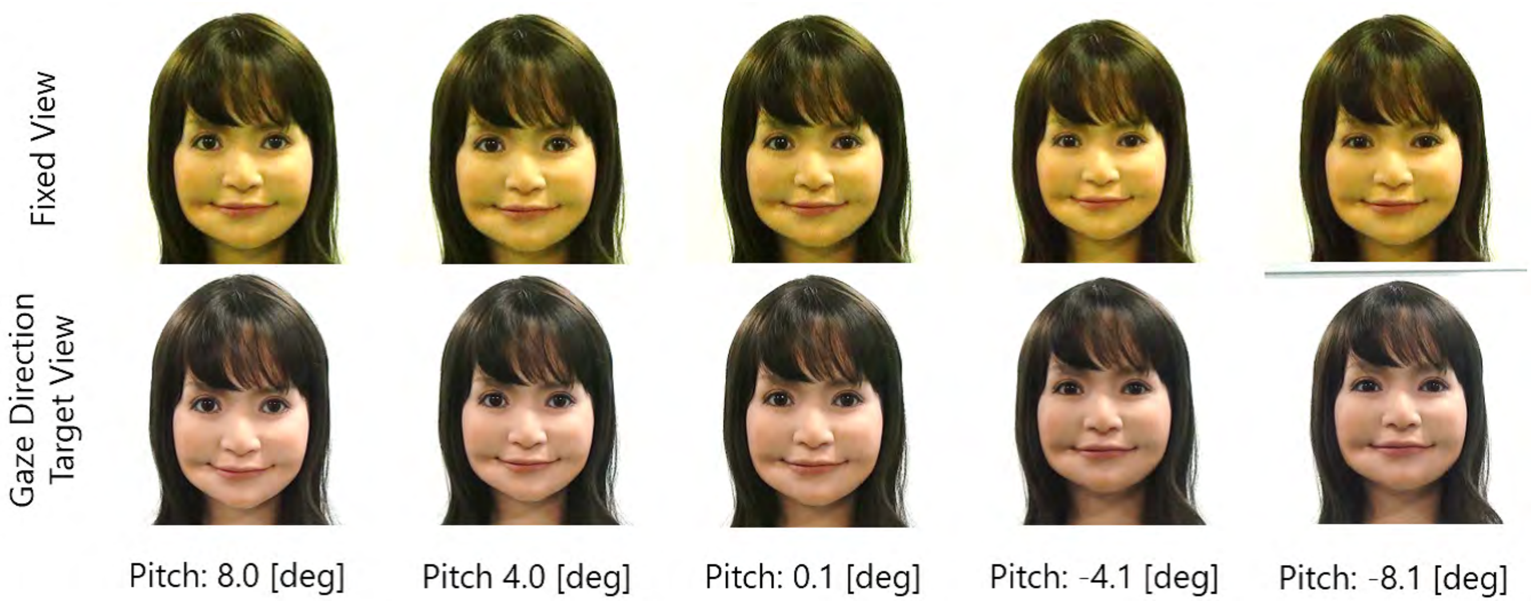

Figure 10. Results of gaze behavior generation in eyes' pitch.

Figure 11. Results of gaze behavior generation in eyes' yaw.

3.3 Calibration Evaluation

3.3.1 Comparison of Gaze Perception

We can assume that a human has the ability to achieve and recognize eye contact. If the calibration succeeds, the android should be able to achieve eye contact like a humans. Therefore, we hypothesize that if the calibration succeeds, if the proposed method works well, then the accuracy of the android's gaze will approach the accuracy of a human's gaze as perceived by human subjects.

We conduct an experiment with 20 subjects of the perception of the android gaze compared to a human gaze. The subjects' task was to indicate whether or not the gaze in a facial picture of an android or a human is directed at where the subject was sitting, *i.e.*, the android/human looking at the subject. Figure 12 shows the experimental environment. The pictures were simultaneously taken with five cameras fixed in the directions of yaw angles ($-10°$, $-5°$, $0°$, $5°$, and $10°$) at a distance of 1.5m from the android or human (one adult male). The android or human looked at one of the five cameras and repeated this five times. One of five pictures provided the android gaze or the human gaze. For this task, 25 pictures of the android or the human were obtained. These pictures were then shown to the subject one by one in random order. Due to the limitations of the cameras' size and space, we did not conduct a comparison experiment for the pitch angles.

Figure 13 shows the averages for accurately determining whether or not the android or human was looking at the camera in this experiment. The results of this experiment indicate no significant difference between the android and human averages as assessed by the t-test for the cases of $0°$, $\pm 5°$ and $10°$ in the yaw angle of the camera. However, the $-10°$ yaw angle of the camera did have a significant difference ($p < 0.001$).

To check these results, we used McNemar's test [36]. Table 3 shows the results of comparison between the android and human responses in true/false numbers of subject's answers in the experiment. Again, the results indicate no significant differences for the case of $\pm 5°$, and $10°$ in the yaw angle of the camera. However, both the $0°$ ($p < 0.05$) and $-10°$ ($p < 0.001$) did have the significant differences according to this test.

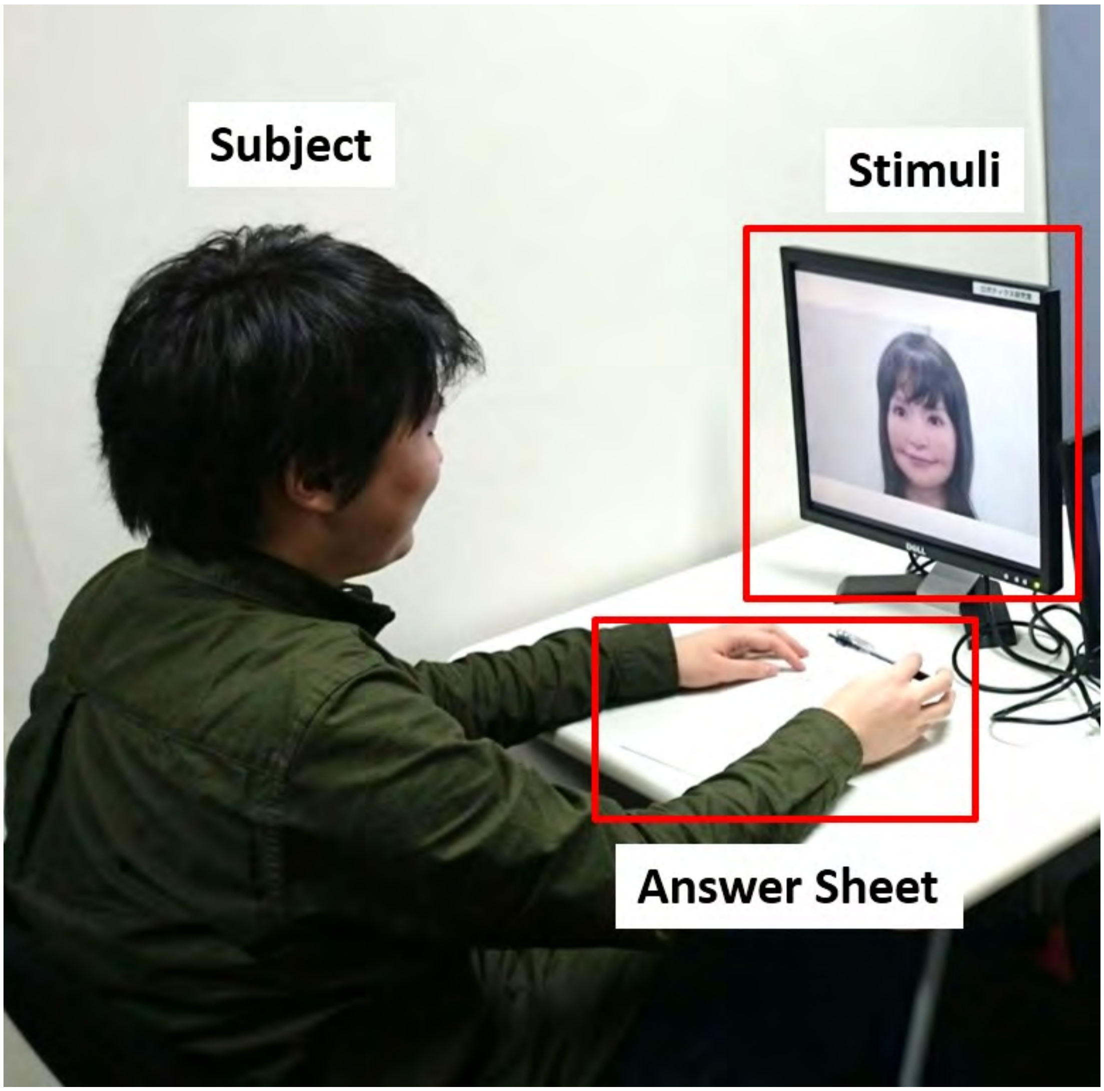

Figure 12. Experimental environment to compare the perception of an android's gaze to a human gaze.

Table 3. Significance of differences in true/false numbers for android and human of subjects' answers, using McNemar's test.

Yaw angle of camera [deg]	-10	-5	0	5	10
p-value	0.0001012 ***	0.4669	0.02905 *	0.0707	0.1317

*: $p < 0.05$, ***: $p < 0.001$

3.3.2 Discussion

The experimental results indicate that compared to human gaze control, the calibrated android's gaze control is equivalent in most cases. However, there were significant differences in the cases of 0° and $-10°$. Here we analyze the factors causing these differences.

In the case of 0° as can be seen in Figure 13, the perception for the human gaze was lower than for other angles. This may be due to the influence of individuality, such as the balance of a human face and eyes, unlike an android's artificial face and eyes, or the influence of sex, such as impressions of the facial contour or hairstyle.

In the case of $-10°$, however, we found that the android's gaze was not controlled accurately. Figure 14 shows pictures of the $-10°$ and $-5°$ android gazes'. Examining these pictures, we can see that there is a possibility that the gaze is not controlled within the range of $-10°$ to $-5°$ since eye contact was not achieved.

Therefore, to evaluate the hypothesis more thoroughly that if the calibration method succeeds, then human subjects will perceive android and human gazes at equivalent accuracy, we believe it is necessary to have a human facial picture that is the same sex as that of the android, to use more than one individual for the human facial pictures to have various faces for the human facial pictures, and to expand the diversity of subjects in the experiment. We should also expand the range of angles of the gaze in the pictures.

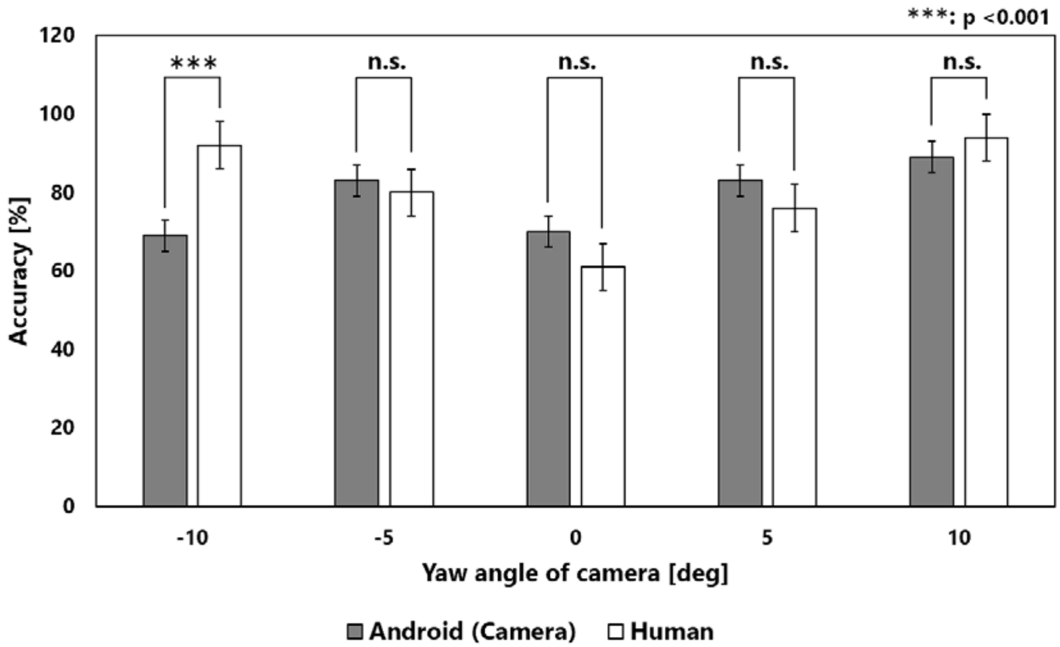

Figure 13. Comparison of the average accuracy of subjects' answers between android and human, t-test $p < 0.001$.

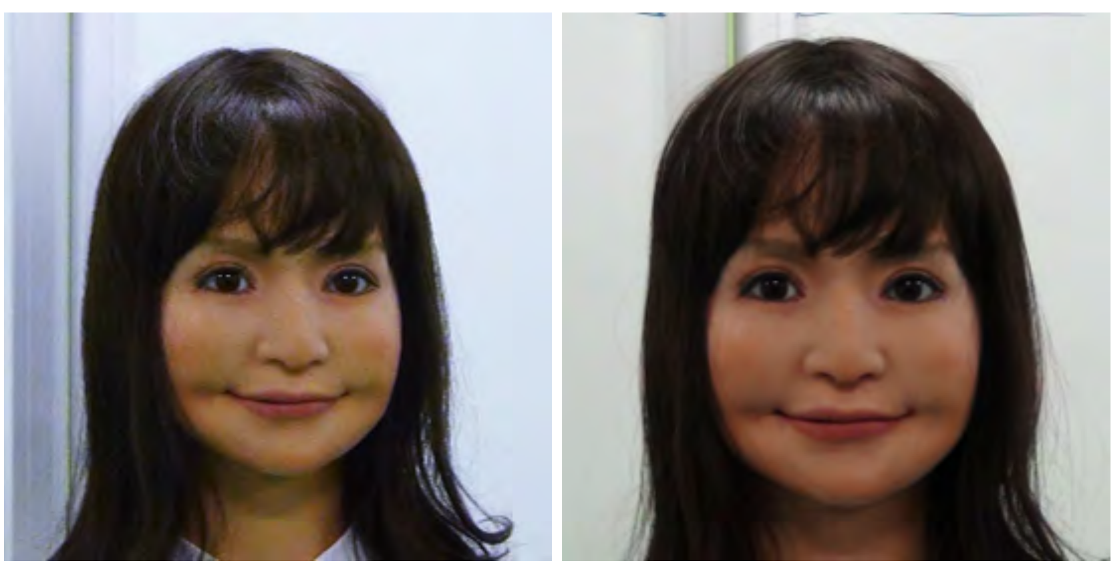

Figure 14. Examples of the gaze control in the case of -10° (left) and -5° (right).

4. Eye Behaviors for Naturalistic Talking

In this chapter, we clarify which eye behaviors make what impressions on humans when an android speaks. Thus, we evaluate the human impression of eye behaviors displayed by an android while talking to a human by comparing the motions generated by the two approaches. Concretely, in the imitation-based approach, the android imitates human eye behavior while explaining some research topics to a subject who acts as the listener. We develop a method to imitate the eye behavior observed using eye trackers worn by the human subject. In the rule-based approach, to generate the eye behavior of the android, we manipulate the eye direction, eye-contact duration, and eyeblinks based on the preference and regularity of human eye behavior found in psychology and cognitive research.

For the evaluation, we conducted two experiments with male and female subjects following the same protocol. We prepare seven patterns of android's eye behaviors: three for imitation-based and four for rule-based behaviors. We show subjects the eye behaviors in the random order and ask them to do a subjective evaluation.

4.1 Imitation-based Eye Behavior

To realize the imitation of human eye behavior by an android, we need to develop 1) a data acquisition method to collect human eye behaviors, and 2) an imitation-based method which uses the acquired data. In our scenario, the android imitates the eye behavior and mouth movements of a female speaker who explains her research topic in Japanese. Since the eye behavior is related to the context [8], the android replays the eye behavior with the recorded voice. Here, we control the eyes, eyelids, and mouth of the android.

4.1.1 Observation

We used two subjects to collect the behavior to be imitated: a female speaker and a female listener. The two subjects were sitting on opposite sides. The distance between the subjects was 80 cm. We only asked the female speaker to explain her research topic for one minute while the listener listened to her. Before starting, we explained what the eye tracker is to both of them and asked them not to pay

Figure 15. Image obtained from the eye tracker overlaid with a gaze point and the results of facial detection.

attention to the eye tracker. We also explained that it is necessary to calibrate the tracker before the observation. We did not ask the subjects to do anything else.

The speaker did not have much time to practice her explanation. Thus, while she talked, she needed to consider what she was about to say. We expect that the speaker subconsciously performs eye behaviors representing various mental states, such as thinking and making decisions about the contents of talking.

We used two *Pupil Labs* [37] eye trackers: the speaker and listener wore one each. Simultaneously, we recorded the voice of the speaker. From the speaker's tracker, we extract the speaker's eye behavior. From the listener's tracker, we estimate the mouth movements of the speaker.

Figure 15 shows an observation from the speaker's tracker. The tracker obtains an image from the speaker's view and records the gaze point (red point) on the image. As shown in Figure 16, the tracker observes the pupils using an IR camera that records the eye close-up in order to estimate gaze directions. The tracker also extracts the boundary of the pupil.

Figure 16. Capturing the pupil using an IR camera.

4.1.2 Extraction of Speaker's Eye Behavior

When the speaker looks at the listener's face, we regard that as the speaker establishing eye-contact bids. Thus, to detect eye-contact bids, we calculate the gaze positions of the speaker relative to the listener's face. We detect the face on the image using *OpenFace* [38]. Figure 15 also shows the results of detecting the face (the blue dots). We calculate the axis-aligned bounding box that covers the blue dots. If the gaze point is inside this bounding box, we determine that there is an eye-contact bid. As shown in Figure 17, to determine how an eye-contact bid is broken, we divide the areas outside the facial bounding box into eight areas. We denote the eye movement behavior to break the eye-contact bids using the names of the eight areas.

As shown in Figure 16, we use the extracted boundary of a pupil to estimate eyeblinks. From the boundary, we can calculate the size of a pupil on the image

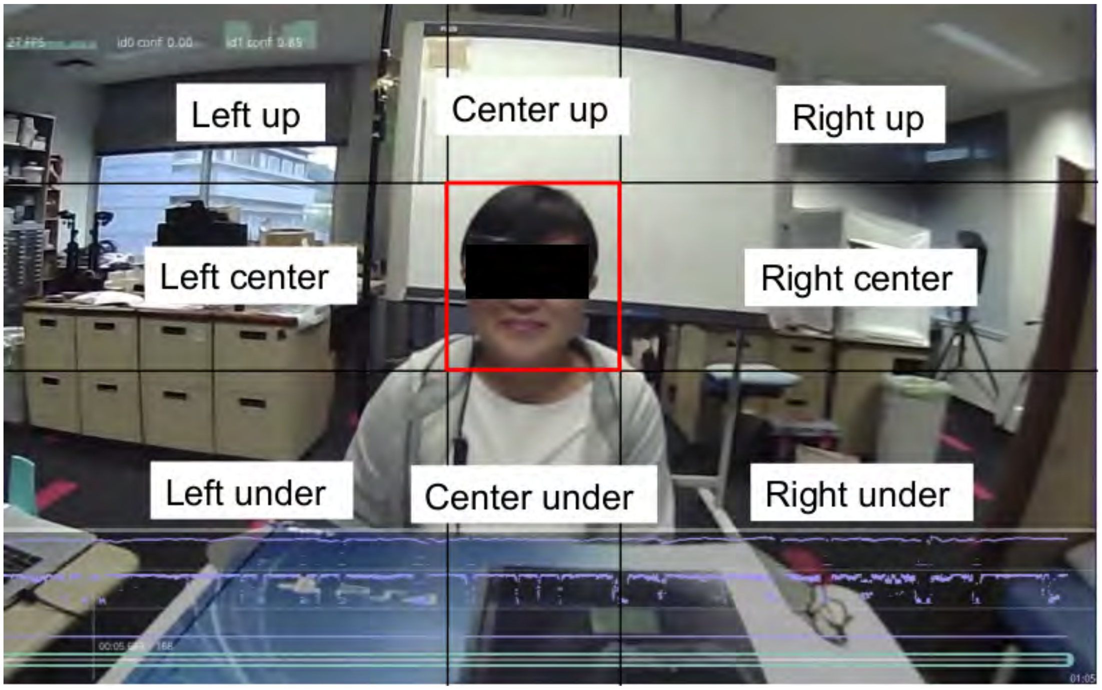

Figure 17. Face area and eight areas for representing eye movements.

coordinates. Since the size changes continuously, we obtain the maximum size from the whole video and consider as eyeblinks whenever the size is less than half of the maximum size.

4.1.3 Mouth Movement

To extract the mouth movements, we also apply *OpenFace* [38] to the listener's images and extract facial landmarks. From the landmarks, we can estimate the contour of the mouth. Figure 18 shows a zoomed-in area around the mouth. We obtain the axis-aligned bounding box of the mouth and then calculate the size of the box.

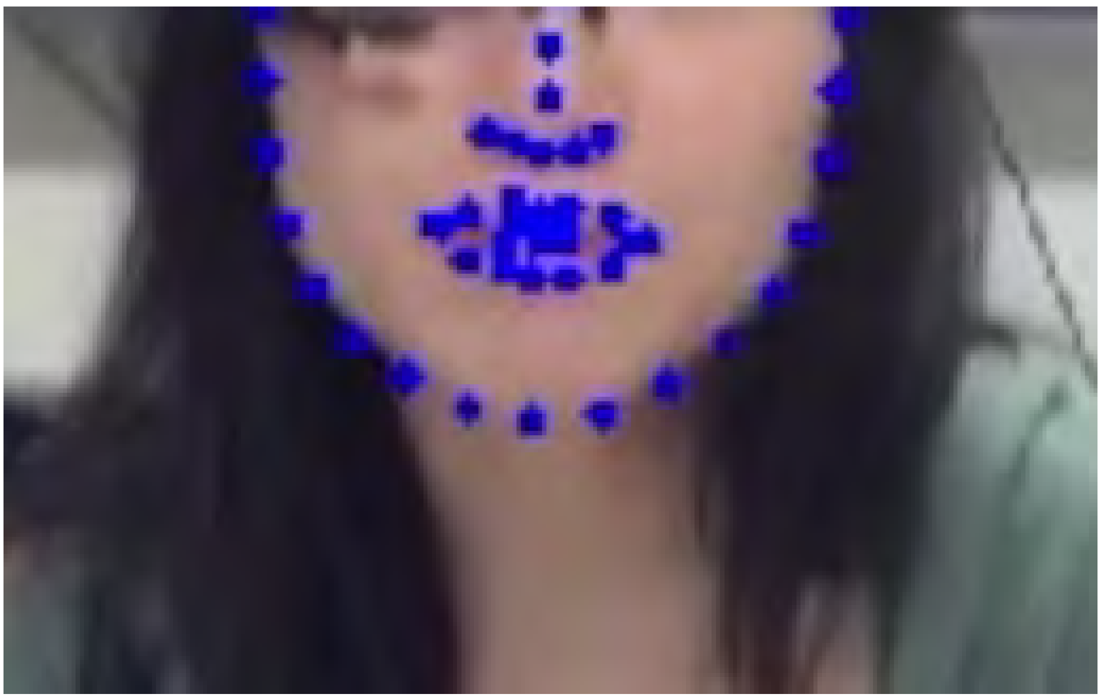

Figure 18. Zoom-in of the listener's image of the speaker's mouth overlaid with facial landmarks.

4.2 Rule-based Eye Behavior

To describe the rules for generating human-like eye behaviors using an android, we use knowledge from psychology and cognitive science research on humans, in which the preference and regularity regarding human eye behavior has been investigated. Binetti *et al.* [39] found out that the preferred period of eye-contact duration is around 3 seconds rather than shorter or longer by using data including a wide range of ages, cultures and personality types. As we mentioned in Chapter 1, the direction of the eye movement for breaking eye contact is related to the mental state through the investigation of the eye movements during thinking. Stern *et al.* [40] points out that eyeblinks are classified into voluntary and reflexive eyeblinks, and, in particular, that voluntary eyeblinks may represent a certain intention.

From the findings of the research works described above, we define the rules considering the following three aspects and then generate the behaviors of eye movement following these rules. First, an android breaks eye contact by changing eye direction. Second, an android repeats the 3-second eye contact. Third, an android uses two types of eyeblinks: short eyeblinks as reflexive eyeblinks, *i.e.*, close the eyes for 0.2 seconds, and long eyeblinks as voluntary eyeblinks, *i.e.*, close the eyes for 0.5 seconds.

4.3 Implementation

4.3.1 Setup

For the implementation, we used the android *Actroid-SIT* as aforementioned in Chapter 1.2. We used our proposed method as mentioned in Chapter 3.1 to relate the actuators' input to the eye direction to imitate the eye behavior of the speakers.

Figure 19 shows the experimental setup. To precisely establish eye-contact bids, we use a chin rest to fix the subject's face in front of the android's face. The chin rest also fixes the subject's face direction to the android's face and at 80 cm from the android, which are the same conditions as in the observation part. We can regard eye-contact bids as eye contact in this situation.

Since the android is actuated by pneumatic actuators, the air compressor

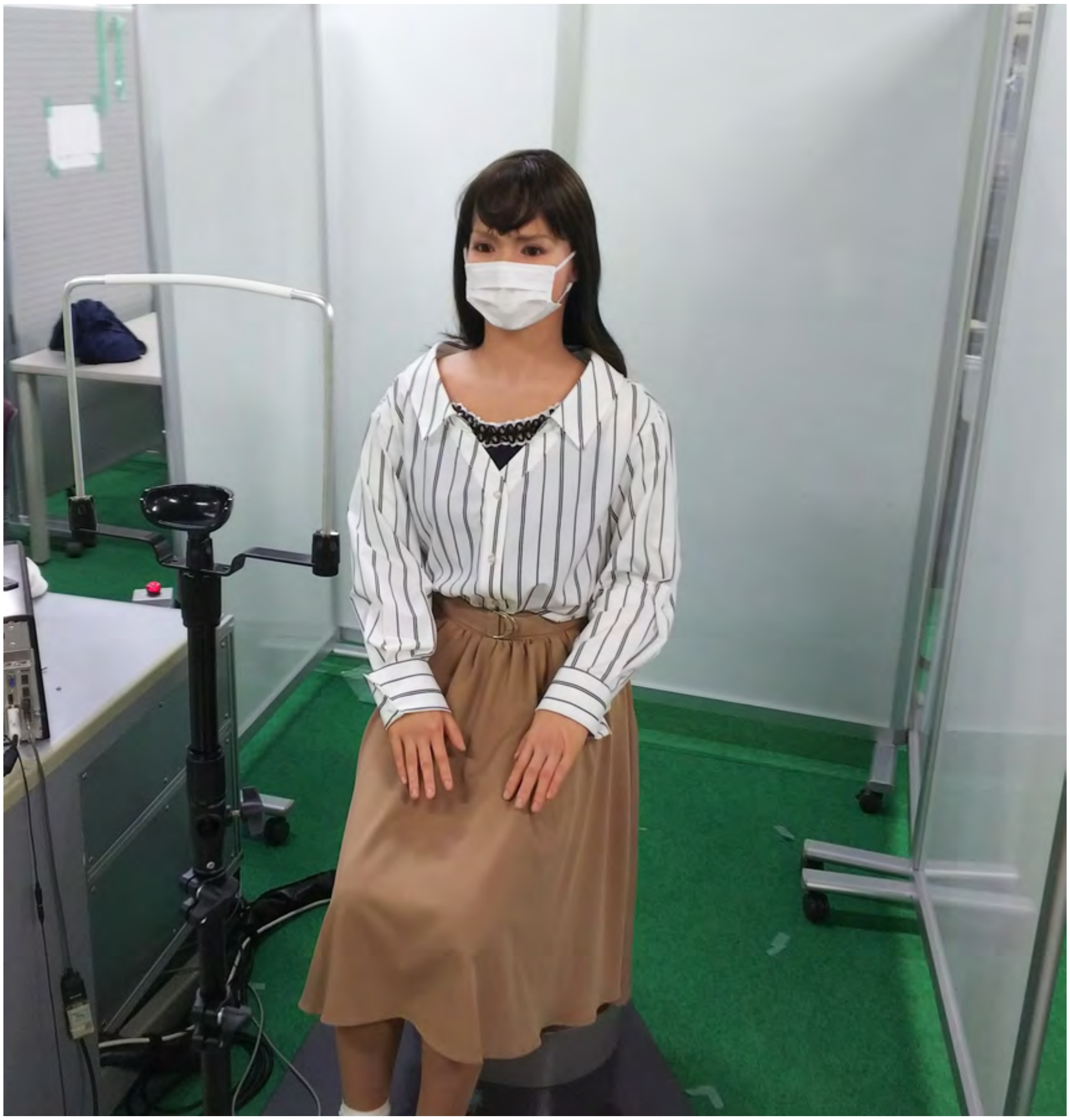

Figure 19. Experimental scene. We fix the subject's face using a chin rest.

that drives the actuators makes noise. To avoid that this affects the subjects' evaluation, we asked them to wear earphones and played the sound of the female speaker through them.

4.3.2 Eye Behavior Generation

With this android, eye directions, eyeblinks, and mouth movements are all independently controlled. As shown in Figure 20, to control the eye direction, we define nine eye directions and choose one of them. In the imitation case, we switch between the eye contact direction and the other eight directions. We also control the eyelids of the android following the timing of eyeblinks.

Since the android only has one DOF in the mouth to control the amount of opening, we can control the mouth using the estimated mouth sizes. However, due to the limitation of this DOF, it is difficult to open the mouth horizontally, such as the mouth shape used to pronounce i in Japanese. This may make subjects feel that the mouth movement is unnatural. Therefore, we decided that the android would wear a mask to avoid this, as shown in Figure 21. As the subject can perceive the mouth movement under the mask, the subject feels that the android is talking. Finally, the voice is replayed as it is, since the observed speaker is female.

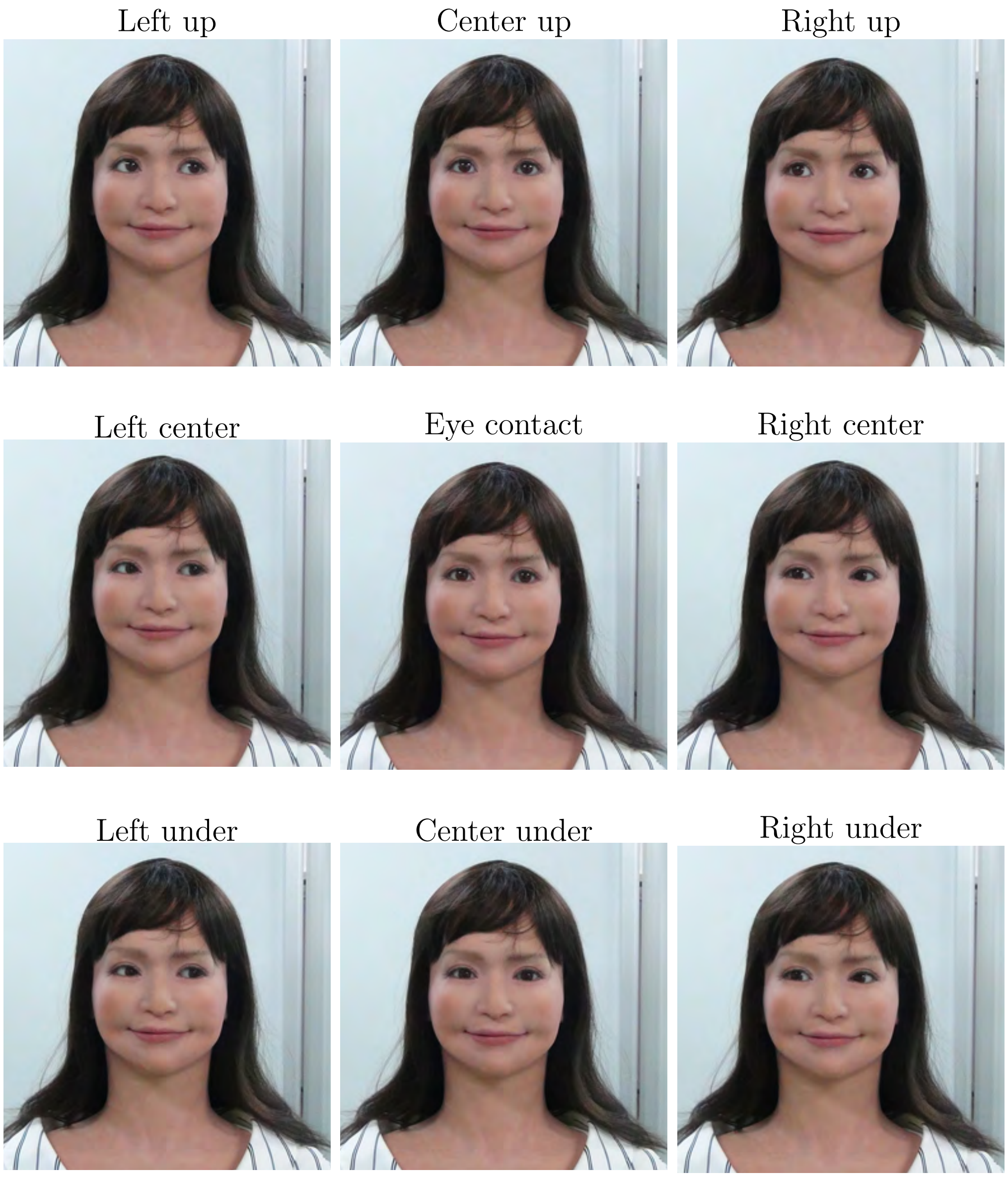

Figure 20. Nine eye directions.

Figure 21. Mask for hiding mouth.

Table 4. Eye behaviors for the comparative experiment.

Pattern	Timing of eye movements	Direction of eye movements	Eyeblinks
(1)	Imitation	Imitation	Imitation
(2)	Imitation	Right up or left up	Imitation
(3)	Imitation	Right up or left up	Short eyeblinks at breakpoints
(4)	Every 3 [s]	Right up or left up	No
(5)	No	No	Short eyeblinks every 3 [s]
(6)	No	No	Long eyeblinks every 3 [s]
(7)	No	No	No

4.4 Subjective Evaluation

4.4.1 Policy for Evaluation

We evaluated the impression of the generated eye behaviors on the android by comparing between the imitation-based motion generation approach and the rule-based approach. In the comparison, we only focused on eye behavior, with the mouth movements and voices kept always the same. We setup the comparisons considering the following four aspects. First, we evaluate the effects of the imitated eye behavior: if human-like motions are realized or not, and what impression it makes on humans. Second, we investigate the impression and duration of eye contact using questionnaires, as mentioned in Chapter 4.2. Third, we evaluate the effect of eyeblinks between short one and long one, as also mentioned in Chapter 4.2. Finally, we also investigate the impression difference of eye behaviors between male and female subjects.

4.4.2 Compared Behaviors

Table 4 shows the seven patterns of eye behavior including imitation, rule-based and their combinations. Pattern 1 is a complete imitation pattern, which is obtained from the situation introduced in Chapter 4.1. We used one set of the observed data in the experiment.

In Pattern 2, we edit eye movements from Pattern 1 by adding the rule-based approach. When the eye moves away from the eye contact, we change the original direction by randomly choosing either the left-up or right-up directions. Eye contact and eyeblinks keep the original duration and timing of the observed data. In Pattern 3, we edit the eye movements and timing of eyeblinks from the Pattern 1. First, we remove all eyeblinks and then manually insert eyeblinks at the breakpoints in the conversation. If the duration of eye contact becomes longer than 3 seconds, we insert eye movement (left-up or right-up directions) to break the eye contact. In other words, from Pattern 1 to 3, the amount of editing is increased. We do this to evaluate the imitation by controlling the level of imitation.

From Pattern 4 to 7, the behaviors are based on rules. Patterns 4, 5, and 6 repeat 3-second eye contact, and breaking it by changing eye direction, short eyeblinks, and long eyeblinks, respectively. Pattern 7 always keeps eye contact without eyeblinks.

4.4.3 Evaluation Method

We recruited 17 male Japanese subjects (mean age = 23.29; SD = 0.85) and eight female Japanese subjects (mean age = 24.00; SD = 1.31), all students of Nara Institute of Science and Technology, who were asked for their impression using a questionnaire. In the questionnaire, we asked the subjects to score the following five items (in Japanese) on a scale from 1 to 7 points (lower score is better):

(a): kind (1) - bad (7)

(b): adequate eye contact (1) - long eye contact (7)

(c): adequate eye contact (1) - short eye contact (7)

(d): attractive (1) - unattractive (7)

(e): humanlike (1) - non-humanlike (7)

Questions (b) and (c) directly ask the preference of eye-contact duration. This is because we would like to distinguish between adequate, short, and long

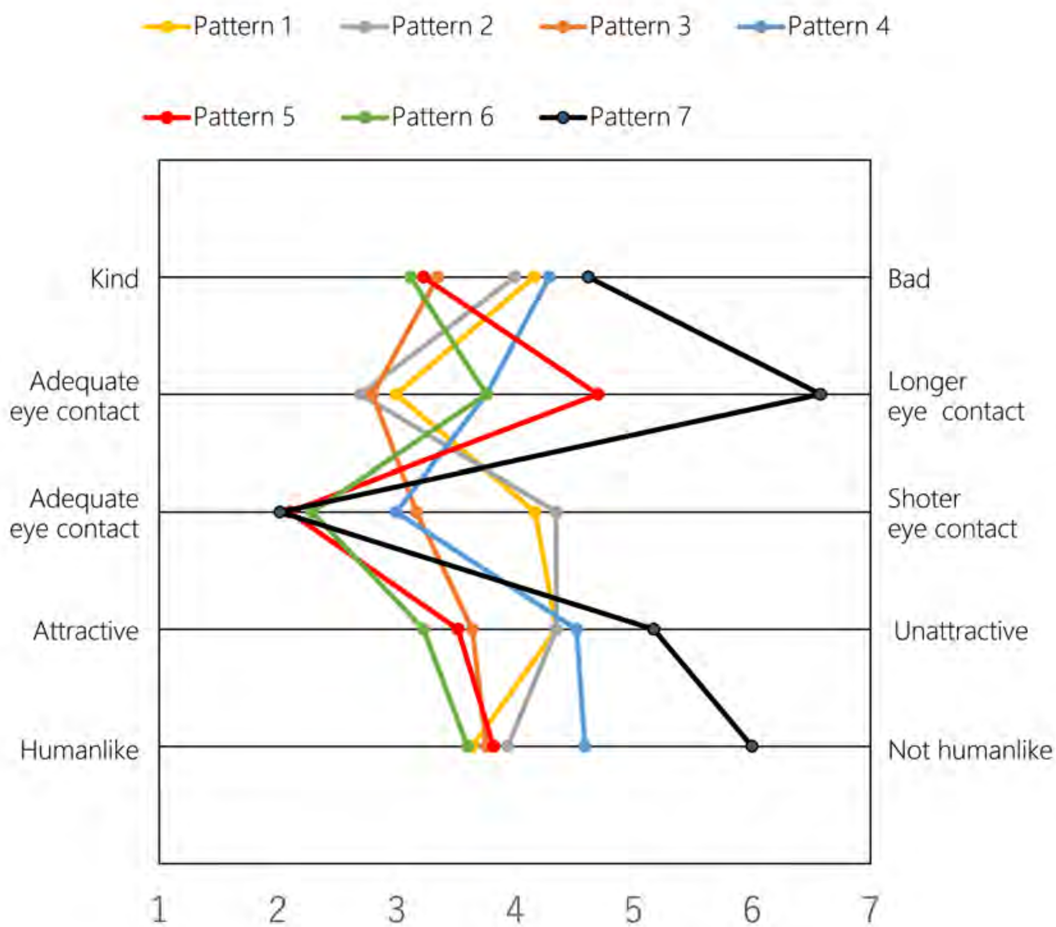

Figure 22. Experimental results of male subjects about the impression on the eye behaviors of the android from the 5-item questionnaire.

eye contact preferences. We intended that question (e) evaluates the quality of the generated android behaviors, while questions (a) and (d) evaluate the overall impression of the android behavior. We have expected that, as the amount of imitation decreases, the score in the human-like question becomes worse. In addition to the aforementioned questionnaire, we included a free-description type of question and asked the subjects to write about their impression of the android behaviors.

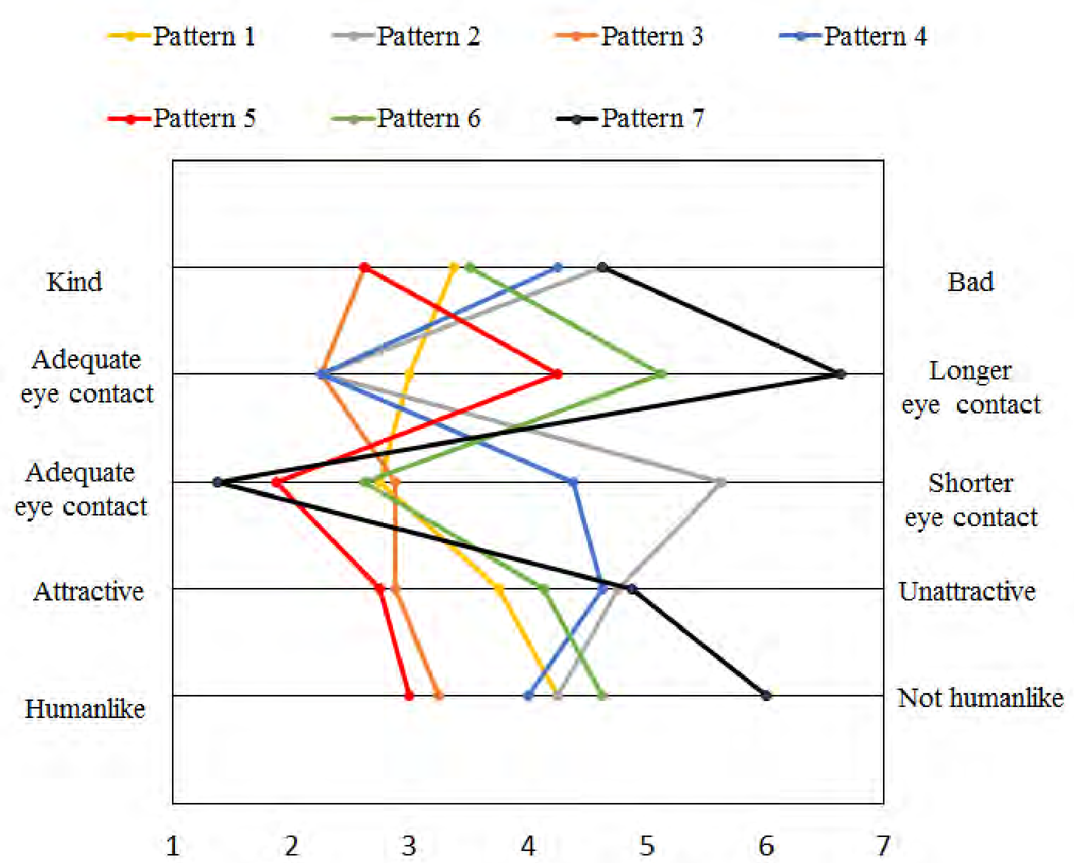

Figure 23. Experimental results of female subjects about the impression of the eye behaviors on the android from the 5-item questionnaire.

4.5 Result

4.5.1 Male Subjects

Figure 22 shows the average scores of 17 male Japanese subjects. The different colored lines show the results of the seven different patterns. In all items, except for short eye contact, Pattern 7 obtained the worst scores. In all patterns, the scores in the kind-bad item showed a similar tendency to those in the attractive item.

To analyze these results statistically, we performed a one-way ANOVA test and then performed false discovery rate (FDR) [41] to solve multiple testing problems after significant difference was revealed by the ANOVA test. In the humanlike item, the statistical analysis revealed significant differences ($p < 0.05$) between Pattern 7 and the rest of the patterns. We also found that there were no significant differences ($p > 0.1$) between the imitation and rule-based eye behavior patterns in the humanlike item.

In the adequate-long eye contact item, the statistical analysis revealed i) significant differences ($p < 0.05$) between Pattern 5 and Patterns 1, 2 and 3; ii) significant differences ($p < 0.01$) between Pattern 7 and the rest of the patterns; and, iii) significant difference tendency ($p < 0.1$) between Pattern 2 and the rest of the patterns except for Pattern 3. In the adequate-short eye contact item, the statistical analysis revealed iv) significant differences ($p < 0.05$) between Pattern 1 and the rest of the patterns except for Patterns 2 and 3, and v) significant differences ($p < 0.05$) between Pattern 2 and the rest of the patterns. Therefore, Pattern 5 gave the feeling of longer eye contact than Patterns 1, 2 and 3 to the male subjects. Moreover, the male subjects regarded Patterns 1 and 2 as having short eye-contact duration.

Compared to Pattern 6, the male subjects regarded Pattern 5 as having long eye-contact duration ($p < 0.1$) in the adequate-long eye contact item even though the duration was the same.

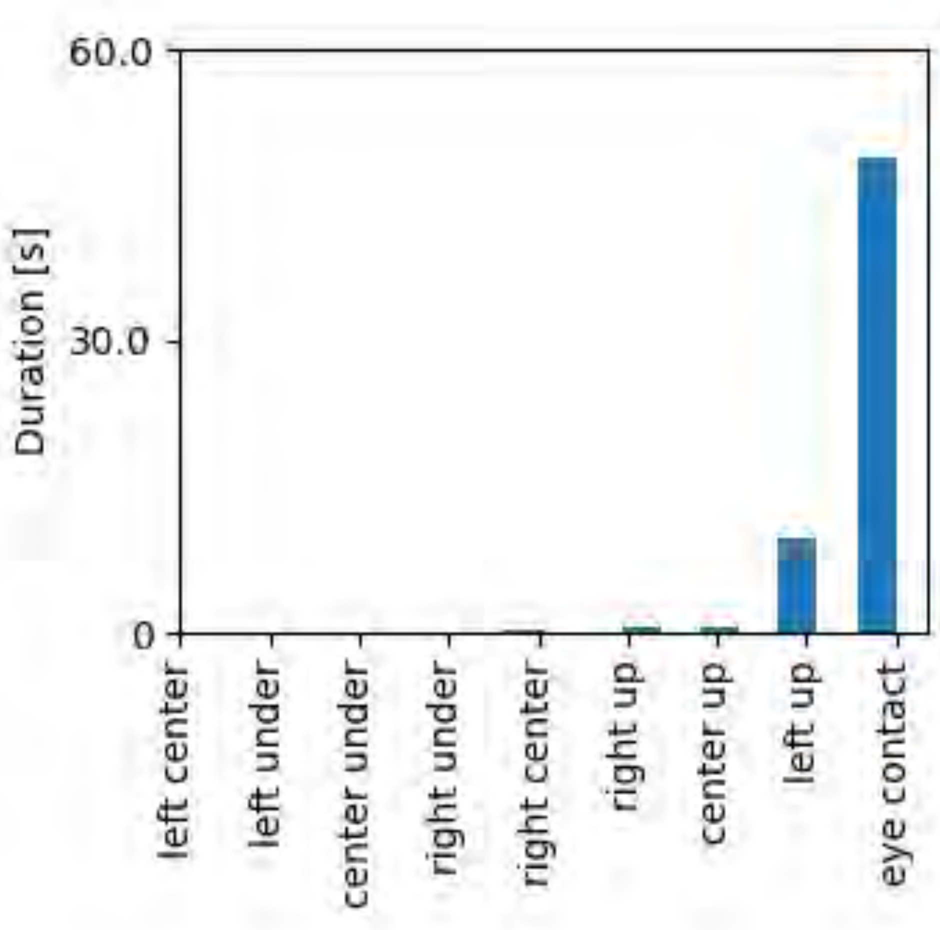

Figure 24. Histogram of eye-direction patterns obtained by observation. Pattern 1 is generated from this result.

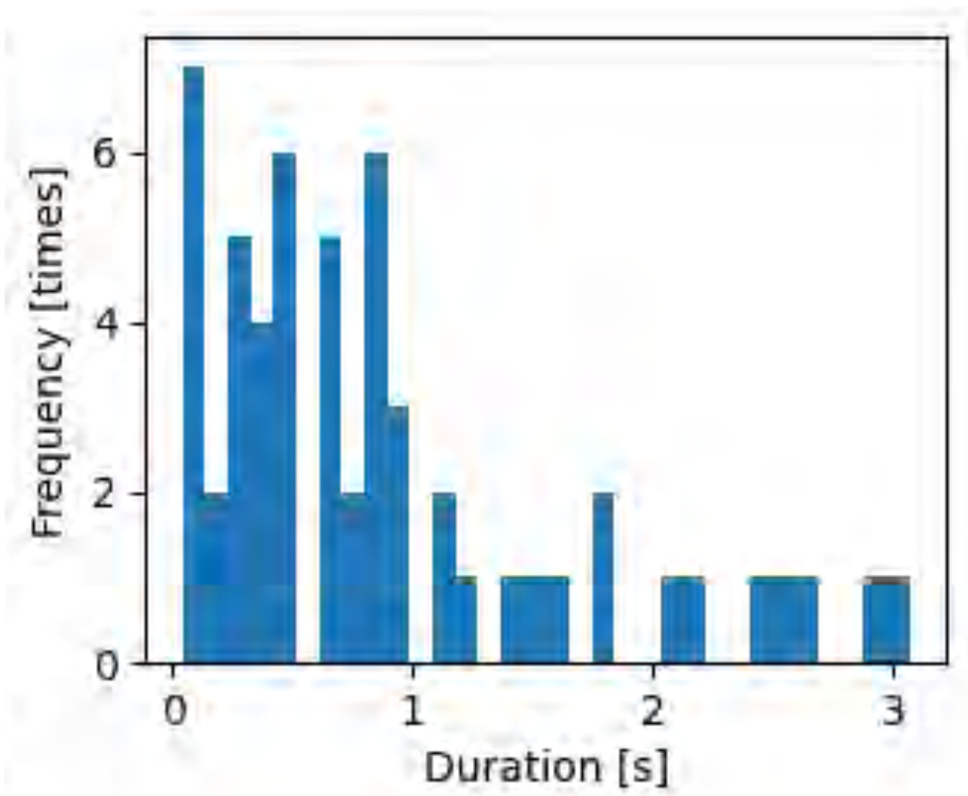

Figure 25. Histogram of eye-contact duration obtained by observation. Patterns 1, 2, and 3 are generated from this result.

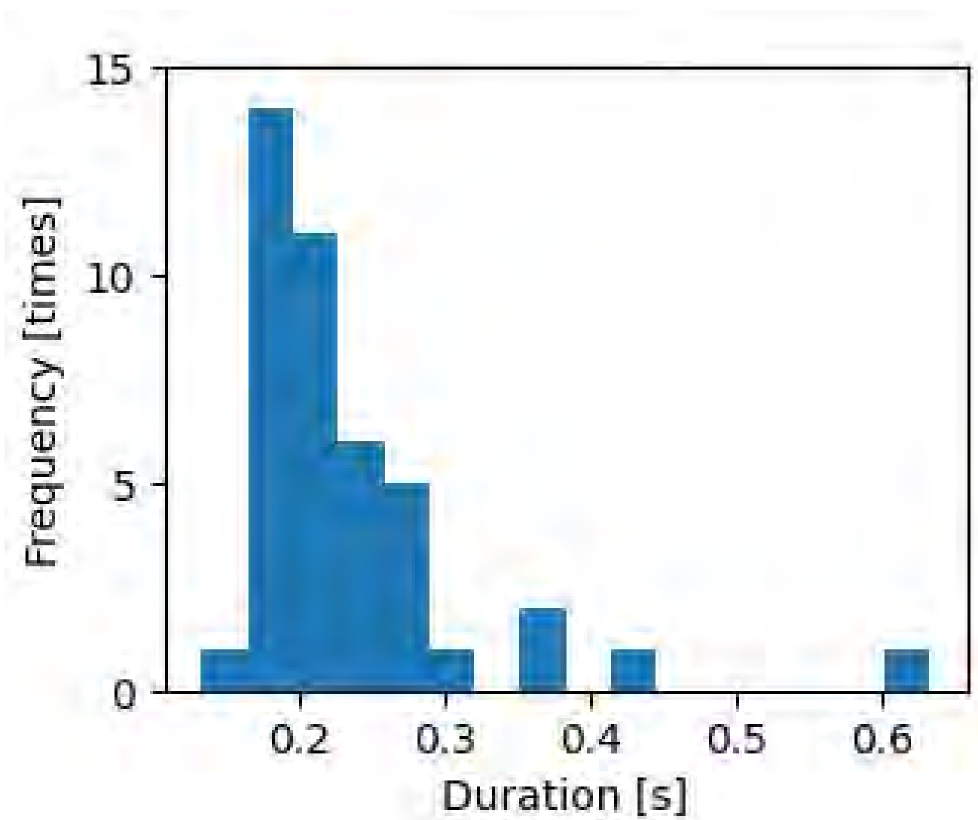

Figure 26. Histogram of the duration of closed eyes in eyeblinks obtained by observation. Patterns 1 and 2 are generated from this result.

4.5.2 Female Subjects

Figure 23 shows the average scores of eight female Japanese subjects. The different colored lines show the results of the seven different patterns. In all items except for the adequate-short eye contact item, Pattern 7 obtained the worst scores. In all patterns, the scores in the kind-bad item showed a similar tendency to those in the attractive item.

To analyze these results statistically, we performed a one-way ANOVA test and then performed false discovery rate (FDR) [41] to solve multiple testing problems after significant difference was revealed by the ANOVA tests as in Chapter 4.5.1.

In the humanlike item, the statistical analysis revealed significant difference tendency ($p < 0.1$) between Pattern 7 and the rest of patterns. We also found that there were no significant differences ($p > 0.1$) between the imitation and rule-based eye behavior patterns in the humanlike item.

In the adequate-long eye contact item, the statistical analysis revealed significant difference tendency ($p < 0.1$) between Pattern 7 and the rest of patterns, Pattern 6 and the rest of patterns except for Pattern 5, and Pattern 5 and the rest of patterns. In the adequate-shorter eye contact item, the statistical analysis revealed significant difference tendency ($p < 0.1$) between Pattern 1 and the rest of the patterns except for Pattern 3 and 6, and also revealed significant difference tendency ($p < 0.1$) between Pattern 2 and the rest of the patterns. Therefore, Pattern 5 and 6 gave the feeling of longer eye contact to the female subjects. Moreover, the female subjects regarded Pattern 5 and 6 as having longer eye-contact duration and Patterns 2 as having shorter eye-contact duration.

While not statistically significant ($p > 0.1$), compared to Pattern 5, the female subjects regarded Pattern 6 as having long eye-contact duration even though the duration was the same.

4.6 Discussion

4.6.1 Effect of Imitation

Comparing between the imitation and rule-based eye behavior, there were few differences in the humanlike item ($p > 0.1$). Compared to whole-body motion, eye movement is simple and has small numbers of DOFs. We think that is why the rule-based eye behavior is perceived as humanlike over the imitation eye behavior. Considering the results of Patterns 4 and 7, subjects did not perceive the behaviors as kindness in the patterns where there is neither movement nor eyeblinks.

While not statistically significant ($p > 0.1$) in most of items, there were fewer differences between Patterns 1 and 2. Figure 24 shows the histogram of eye-direction patterns obtained by observation. Pattern 1 (*i.e.*, the imitation of the eye directions) is generated from this result. The speaker that the android imitates explains her research topic while considering the contents. This leads to many left-up eye movements when breaking eye contacts. Changing left-up eye movement to right-up eye movement may not cause significant changes in the feeling of the subjects.

4.6.2 Feeling and Duration of Eye Contact

From the results, keeping 3-second eye contact obtained better overall scores. Especially, Pattern 5 obtained the best scores. In the imitation aspect, compared to Patterns 1 and 2, Pattern 3 tried to keep eye contact longer. In the adequate-short eye contact item, the statistical analysis revealed ($p < 0.05$) between Patterns 2 and 3.

Figure 25 shows the histogram of eye-contact duration obtained by observation. Patterns 1, 2, and 3 (*i.e.*, the imitation of eye contact) are generated from this result. We assume that both eye movements and eyeblinks break eye contact. In most cases, the speaker established eye contact for two seconds or less. We think that is the reason why subjects tend to regard Patterns 1 and 2 as having shorter eye contact. As a result, Pattern 3 obtained better scores than Patterns 1 and 2.

4.6.3 Effect of Eyeblinks

From the results, compared to Patterns 4 and 7, Patterns 5 and 6 obtained better scores in the kind-bad, attractive and humanlike items. In most cases, the statistical analysis revealed significant difference tendency ($p < 0.1$) except for the humanlike item. Hence, compared to the eye behaviors without eyeblinks, using eyeblinks brings positive impressions.

While not statistically significant ($p > 0.1$) in most of items, Figure 22 and 23 show slight differences in the scores between Patterns 5 and 6. The difference between the condition of the two patterns is only the duration of eyes closing in eyeblinks. Even though the eye-contact duration is the same in both patterns, the subjects tended to feel that the eye-contact duration is different with the two types of eyeblinks. The duration of eye closing might affect the impression of the eye-contact duration.

Figure 26 shows the histogram of the duration of closed eyes in eyeblinks obtained by observation. Patterns 1 and 2 (*i.e.*, the imitation of eyeblinks) are generated from this result. From Figure 26, in most eyeblinks, the eyes are closed for 0.3 seconds or less, *i.e.*, short eyeblinks. We think reflexive eyeblinks are represented by short eyeblinks and voluntary eyeblinks are considered much more as attempts to break eye contact.

4.6.4 Effect of Gender Difference of Subjects

From the results, we found that there was a difference between the male and female subjects. Concretely, regarding eyeblinks, compared to Pattern 6 (*i.e.*, the long eyeblinks), the female subjects answered that the scores of Patterns 3 and 5 (*i.e.*, the short eyeblinks) in all items except for the adequate-short eye contact item are better. By contrast, compared to Patterns 3 and 5 (*i.e.*, the short eyeblinks), the male subjects answered that the score of Pattern 6 (*i.e.*, the long eyeblinks) in all items except for the adequate-long and adequate-short eye contact items was better. Additionally, through the free-description type questionnaires, most of the female subjects mentioned that they noticed differences between all seven patterns regarding the frequency and/or duration or/and timing of the eyeblinks. In contrast, only some male subjects mentioned the eyeblinks, which means that they might have not given much importance to the differences in the eyeblinks.

Therefore, by comparing to the male subjects, the female subjects preferred short eyeblinks over long eyeblinks and regarded that short eyeblinks might be one of the keys to make eye contact more suitable than long eyeblinks, even though the eye-contact duration is the same in both eyeblinks.

5. Eye Behaviors for Attentive Listening

The hypothesis of this chapter is that if an android acting as a listener imitates eyeblinks and nodding of a human speaker in face-to-face communication, the android can make a human speaker perceive its listening behavior as attentive listening. Here, we clarify how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening. First, we develop a real-time method to imitate human eyeblinks and nodding using an android. Next, we evaluate the subjective impression of the imitation by comparing to 1) rule-based eyeblink and nodding, which are simple duration-based motions, 2) motions generated at breakpoints, and 3) combined motions.

5.1 Imitation

For an android to imitate humans' eyeblink and nodding, it is necessary to obtain the information of eyeblink and nodding from the movement on humans' facial area. To do this, there is a method in which a human speaker wears an eye tracker or a wearable device [42]. However, we would like to avoid to employ it in order to realize a natural face-to-face communication situation with an android. Therefore, we decided to use a camera placed outside an android to obtain information on a human's facial area. Hence, we employ *OpenFace 2.0* [43], which can get facial landmarks in real time to obtain the information.

5.1.1 Eyeblink Imitation

By manipulating open/close the eyelid of an android (*i.e.*, only one DOF in this case), we realize the generation of eyeblink motions. To obtain human eyeblink information, we use the number 45 of Action Units (AU45) which is an index of Facial Action Coding System [44] calculated from *OpenFace 2.0*. Setting a threshold as half the value of AU45, we can decide whether to open or close the eyelid. Finally, by converting the binarized output with the threshold to the input value of the controller for the eyelid's actuator, we achieve the eyeblink imitation motion.

5.1.2 Nodding Imitation

By moving up/down the neck of an android, we realize the generation of nodding motion. We manipulate only the pitch of the neck. To obtain human nodding information, we use the orientation of the head calculated by *OpenFace 2.0*. Concretely, we use the pitch angle of the head (*i.e.*, move up/down) in the camera coordinate system from the calculated head orientation. Then, we convert it to the input value of the controller for the neck's actuators with the method described in Appendix A. Here, we define the nodding motion as the motion where the neck moves in the range from the frontal pose to a looking-down pose because the neck does not move backward when nodding. Finally, by limiting the range of the input value of the controller for the neck's actuators to the range that makes the neck move between the frontal pose and the looking-down, we achieve the nodding imitation motion.

5.2 Implementation

For the implementation, we used the android *Actroid-SIT* as aforementioned in Chapter 1.2. To construct the software system, we employ the *Robot Operating System* (ROS)[4] as a software middleware and then use the ROS package of *OpenFace2.0*[5] and a USB web camera for *OpenFace2.0*.

5.2.1 Eyeblink Motion Generation

Figure 27 shows an example of the result of eyeblink motion generation. The human, as the imitation target, started with the eyes closed, and then opened them. The upper row of Figure 27 shows that the information of the eyes area can be obtained from the detected facial landmarks using *OpenFace 2.0*. The bottom row of Figure 27 shows that the android could start with the eyes closed, and then open them. Thus, we confirmed that the android can imitate the human's eyeblinks.

[4]ROS, `http://wiki.ros.org/`

[5]openface2_ros, `https://github.com/ditoec/openface2_ros`

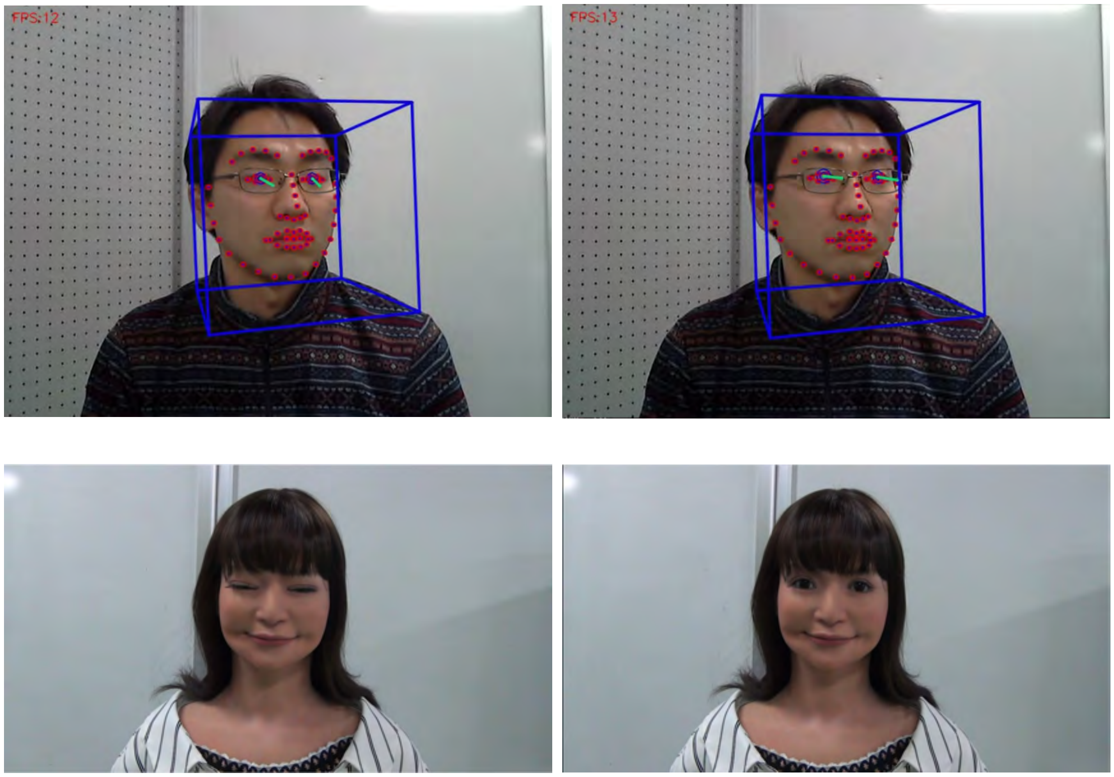

Figure 27. Example of the human eyeblinks imitation achieved by the developed method. The eyes close (left) and open (right).

5.2.2 Nodding Motion Generation

Figure 28 shows the example result of the nodding motion generation. First, the human as an imitation target started looking at the front, and then moved the head down. The upper row of Figure 28 shows that the head orientation can be obtained from the visualized 3D bounding box of the head area by *OpenFace 2.0*. First, the bottom row of Figure 28 shows that the android could start looking at the front, and then move the head down. Thus, we confirmed that the android can imitate the human's nodding.

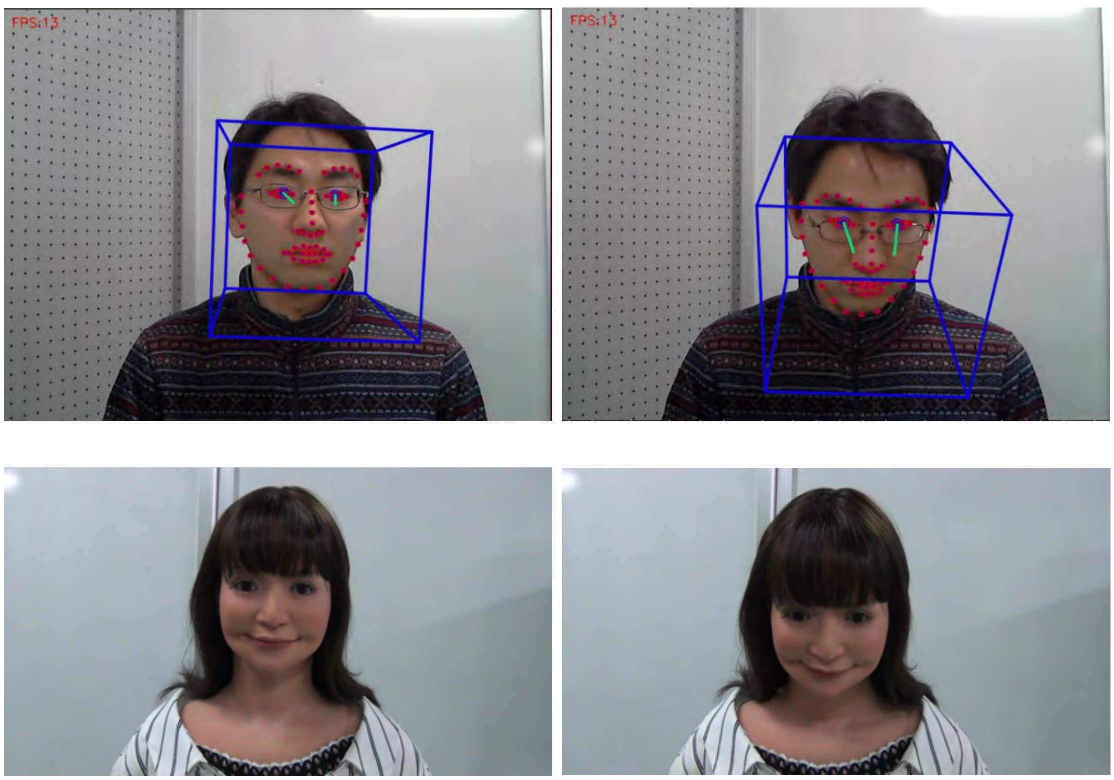

Figure 28. Example of the human nodding imitation. The head directs the front before starting to move down (left). The head finishes moving down (right).

5.2.3 Combined Motion Generation

Figure 29 shows the example result of the combined motion generation (*i.e.*, eyeblinks and nodding). Moreover, Figure 30 shows the plot on the actuator value of the same example result. First, the human as the imitation target started facing down (*i.e.*, nodding). Next, the human faced up and then blinked. Finally, the human stopped blinking and then faced down. The upper row of Figure 29 shows that the facial information can be obtained using *OpenFace 2.0*. The bottom row of Figure 29 shows that the android first could start facing down, next could face up and then blink, and finally, could stop blinking and then face down. Thus, we confirmed that the android can imitate the human's eyeblinks and nodding.

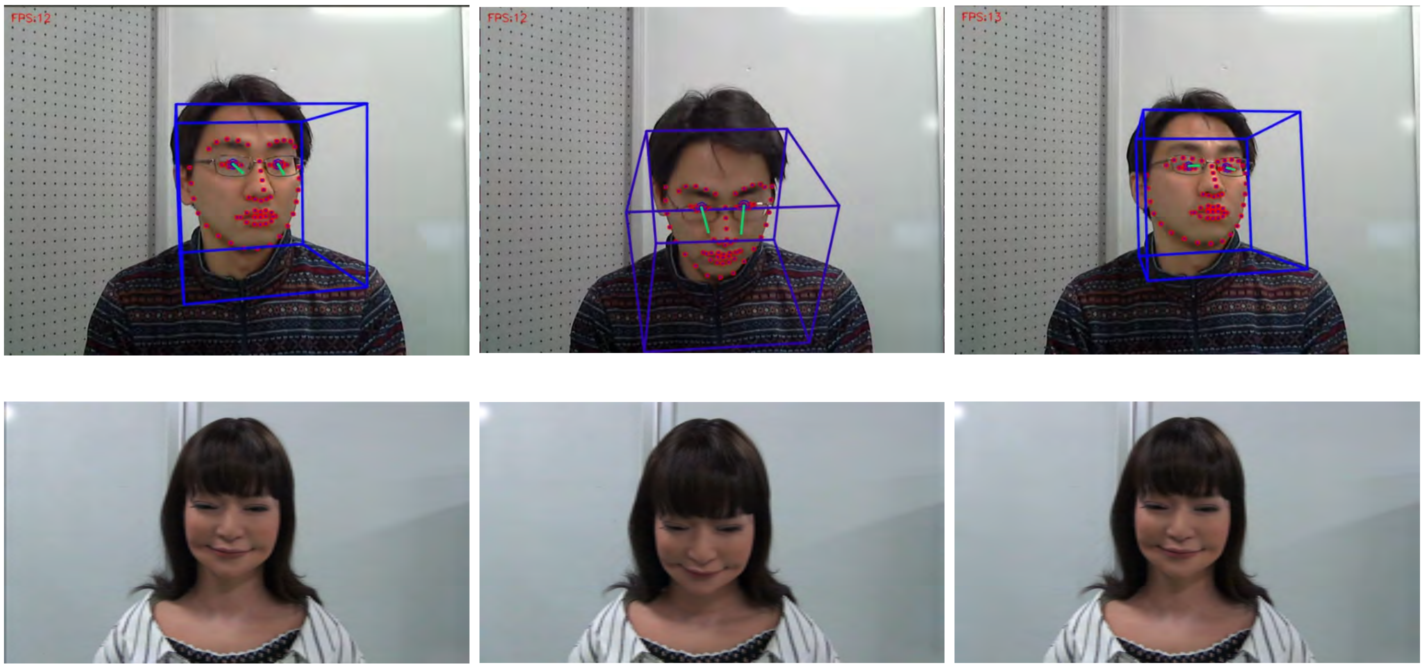

Figure 29. Example of the imitation of the combination of eyeblink and nodding. First, the android nods. Second, the android blinks. Third, the android nods again.

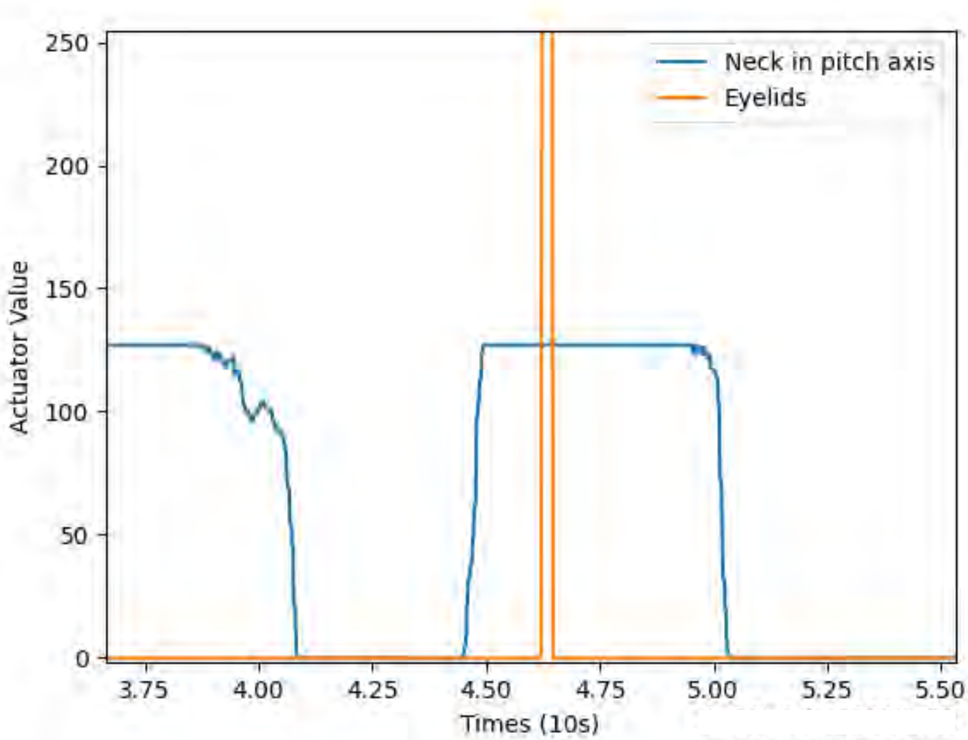

Figure 30. Actuator value of the example of the combined motion generation.

5.3 Subjective Evaluation

5.3.1 Hypotheses

For the evaluation, we hypothesize the following things:

1. Eyeblink is a necessary cue of attentive listening behaviors for a human's speaker.

2. Imitation of eyeblink and nodding makes a human speaker perceive the listening behavior as attentive listening.

We focus on the two cues (*i.e.*, eyeblink and nodding) of the nonverbal behaviors for attentive listening from the existing findings in the previous works of HRI research. Tatsukawa *et al.* found that the eyeblink of a human listener synchronizes to the eyeblink of an android speaker in a face-to-face situation [20]. Hence, we think that if the android in a listener role imitates the eyeblink of a human's speaker, it can give humans better impressions for attentive listening. Yoshikawa *et al.* [45] verified that the nodding motion when an android synchronizes it to the human's nodding by playing in a listener role is effective.

5.3.2 Experimental Setup

We asked a subject in a speaker role to sit at 1.0 m from the android as shown in Figure 31. The height of the subject's eyes was adjusted to the height of the android's eyes to establish the eye contact. Then, we asked the subject to speak in about one minute with the android and evaluate the impression on each motion pattern with a questionnaire.

For speaking to the android, we asked the subject to introduce him/herself because we think that it can be easy to speak the same topic repeatedly to all motion patterns. Before starting the experiment, we asked the subject to practice the talk a few times to fit it within one minute.

Furthermore, as we manipulate only the eyelids and neck of the android, we made the android wear a mask on its face in order to avoid the effect of the appearance of mouth movement and facial expression as the same case in Chapter 4.

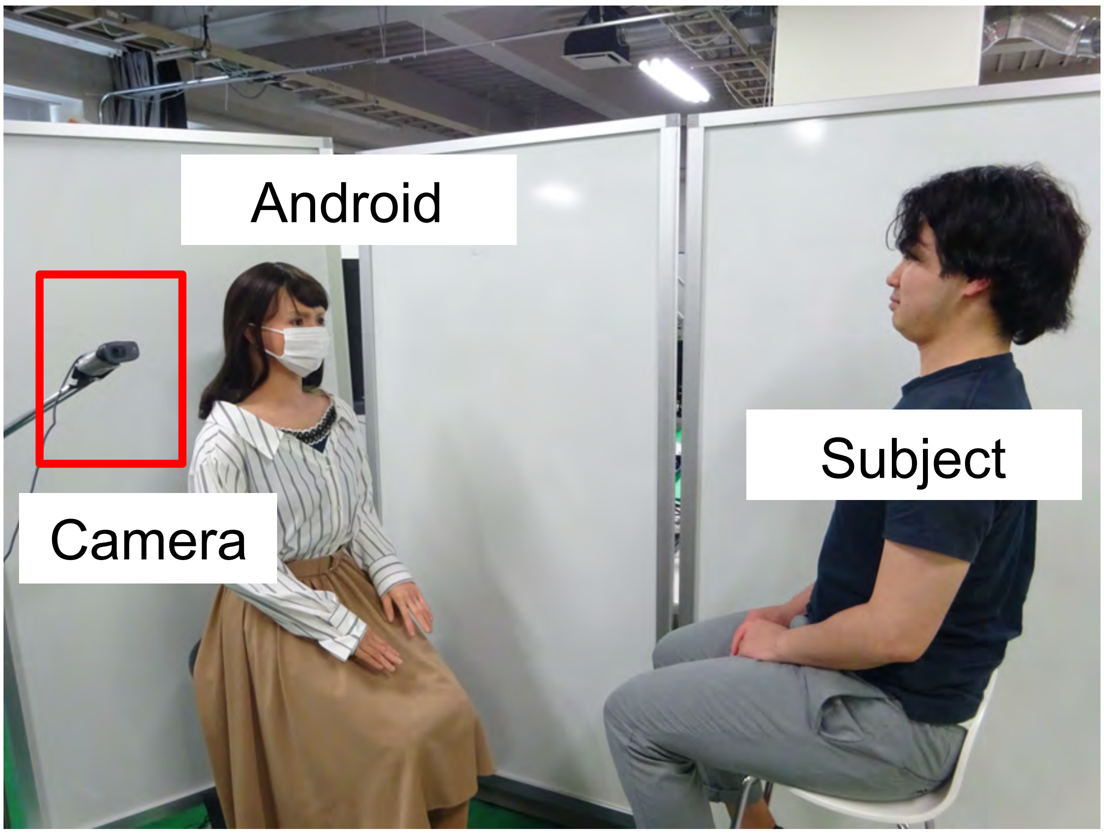

Figure 31. Experimental setup. A subject is sitting in front of the android which wears a mask, keeping the height of mutual gaze, *i.e.*, establishing eye contact.

Table 5. Patterns for the experiment.

		Eyeblink			
		Imitation	Simple	Nothing	At Breakpoint
Nodding	Imitation	Pattern 1	Pattern 2	Pattern 3	Pattern 4
	At Breakpoint	Pattern 5	Pattern 6	Pattern 7	-

5.3.3 Patterns

Our research focus is the two important cues, *i.e.*, eyeblinks and noddingh of the nonverbal behaviors for attentive listening.

For the competitors in the evaluation, in the first cue, *i.e.*, eyeblinks, we prepared three types of the motion patterns. First, the eyeblinks occured by the developed imitation method. Second, the eyeblink was generated every 3 seconds from the findings that 3-second eye contact is preferred by humans [39] and 3-second eye contact obtained better impressions when an android speaks to a human in face-to-face communication from Chapter 4, *i.e.*, the eyeblinks break eye contact. We call this motion type *simple duration-based eyeblinks.* Third, we used no motion whatsoever because it is necessary for the verfication of the first hypothesis (*i.e.*, the effect of eyeblinks). We call this motion type *no eyeblinks.* Fourth, the motion was generated at breakpoints of the speech from the finding that eyeblinks of a human occur at the end and during pauses in speech [8]. This motion was generated by the Wizard-of-Oz method [46] because the implementation of this competitive pattern to be easier to avoid employing automatic voice recognition tools. We called this motion type *eyeblinks at breakpoints.*

In the second cue, *i.e.*, nodding, we prepared two types of motion patterns. First, the motion was generated by the developed imitation method. Second, the motion was generated at breakpoints of the speech from the finding that the nodding of a human often occurs during pauses in speech [47]. This motion was generated by the Wizard-of-Oz method [46] because the same reason of the case of *eyeblinks at breakpoints.* We called this motion type *nodding at breakpoints.*

Thus, there were seven patterns in total for the evaluation, as shown in Table 5.

Table 6. Result of Godspeed Questionnaire.

	Patterns						
	#1	#2	#3	#4	#5	#6	#7
Anthropomorphism	**3.12**	**3.12**	2.68	2.92	**3.17**	2.92	2.62
Animacy	2.68	2.94	2.35	2.94	2.99	**3.21**	2.75
Likeability	2.75	3.00	2.73	**3.12**	3.00	**3.18**	3.02
Perceived Intelligence	2.78	2.92	2.65	**3.12**	2.87	**3.07**	2.87
Perceived Safety	2.91	**3.25**	**3.22**	**3.12**	**3.36**	**3.33**	**3.33**

5.3.4 Evaluation Method

To evaluate the human impression for each pattern, we asked the subjects to score with the *Godspeed Questionnaire* [48]. The *Godspeed Questionnaire* can have subjects score quickly the impression of artificial agents. This evaluates five impression items including a total of 24 paired adjectives on a scale from 1 to 5 points (higher score is better). In this evaluation, the first item *Anthropomorphism* and the second item *Animacy* are related to the effectiveness of the imitation. The third item *Likeability*, the fourth item *Perceived Intelligence*, and the fifth item *Perceived Safety* are related to the perception of the attentive listening.

5.4 Result

Twelve Japanese subjects (mean age = 24.25; SD = 2.38) who were recruited at Nara Institute of Science and Technology under the approval of its ethics committee. Table 6 shows the average scores of the results with the *Godspeed Questionnaire.*

5.4.1 Eyeblinks

From Table 6, compared to the patterns of no eyeblinks (*i.e.*, Patterns 3 and 7), the rest of patterns, *i.e.*, the motions include eyeblinks, obtained better overall scores in all items.

Specifically, compared to the imitation (*i.e.*, Patterns 1 and 5), *simple duration-*

based eyeblinks (*i.e.*, Patterns 2 and 6) and *eyeblinks at breakpoints* (*i.e.*, Pattern 4) obtained slightly better scores in the items *Likeability* and *Perceived Intelligence*.

5.4.2 Combined Eyeblinks and Nodding

From Table 6, the imitation motion of both eyeblinks and nodding (*i.e.*, Pattern 1) did not obtain better scores than the rest of patterns in all items except for the item *Anthropomorphism*.

On the other hand, the combination of simple eyeblinks and *nodding at breakpoints* (*i.e.*, Pattern 5) obtained better scores than the rest of patterns in all items except for the item *Anthropomorphism*.

Specifically, compared to the type of the nodding imitation (*i.e.*, Patterns 1, 2, and 3), the type of *nodding at breakpoints* (*i.e.*, Patterns 5, 6, and 7) obtained better scores in the items *Animacy*, *Likeability*, and *Perceived Intelligence*.

5.5 Discussion

5.5.1 Eyeblinks

From the result of Chapter 5.4.1, we found the eyeblink is more necessary than no eyeblinks for attentive listening because all types of eyeblinks except for the type of no eyeblinks gave better impression than the type of no eyeblinks.

In addition, the result indicated that *simple duration-based eyeblinks* and *eyeblinks at breakpoints* may improve the perception of attentive listening than the imitation of eyeblinks because these types obtained slightly better scores than the imitation in the items *Likeability* and *Perceived Intelligence* that being related to the perception of attentive listening.

Therefore, we conclude that the result proves the first hypothesis, *i.e.*, eyeblink is a necessary cue of attentive listening behaviors for a human's speaker.

5.5.2 Imitation of Both Eyeblinks and Nodding

From the result of Chapter 5.4.2, first, we found that the imitation of both eyeblinks and nodding gave lower impression for attentive listening. The cause might be that the android might imitate small head movements and then the generated nodding that subjects cannot recognize affected the human impression because the head moves down the bottom in the simple one, whereas the head follows a trajectory regardless the range in the imitation.

Second, we noted that the appropriate timing of reaction is possibly very important for attentive listening because the types of motions generated at breakpoints tended to obtain better scores.

Finally, we noted that the difference between the types of nodding patterns affected the impression because we think that the appearance of the nodding is more dynamic than the appearance of the eyeblinks.

Therefore, we conclude that the result does not prove the second hypothesis *i.e.*, imitation of eyeblinks and nodding makes a human speaker perceive the listening behavior as attentive listening.

6. Conclusion

6.1 Gaze Calibration for Eye Contact

This dissertation proposed a gaze calibration method to accurately control the direction of an android's gaze. First, we estimated the gaze direction using the constraint that the camera optical axis is aligned with the gaze direction. We also showed a method for this alignment. Second, we modeled the relationship between the gaze direction and the input values of the actuator controller using linear regression. Finally, we proposed a method to verify the effectiveness of the gaze calibration method. To do this, we compared the accuracy of subjects' perception of the gaze direction between pictures of an android's gaze and those of a human gaze.

The evaluation result indicated that the comparison between human gaze control and the calibrated android's gaze control is competitive in most cases. This verifies that the proposed calibration works well.

6.2 Eye Behaviors for Natural Talking

This dissertation investigated which eye behaviors using an android make what impressions on humans and clarify which are the important factors for attractive eye behaviors. Hence, we evaluated the human impression of eye behaviors displayed by an android while talking to a human by comparing motion generated following two approaches. The first approach is the imitation of human motions to realize human-like eye behaviors. To imitate human eye behavior, we develop a method to generate the eye behavior observed using eye trackers worn by human subjects. The other approach is a rule-based motion generation to realize designed eye behaviors. In the latter, we describe rules to manipulate the eye direction, eye-contact duration, and eyeblinks based on the preference and regularity of these features of human eye behavior revealed in psychology and cognitive research [7, 39, 40].

This dissertation conducted two experiments with male and female subjects to evaluate the subjective impression by comparing seven patterns of eye behaviors with an android, generated by editing the imitation parameters or the rule-based behavior. From the results, we concluded the following four findings. First, the imitation and rule-based behaviors show no difference in human-likeness, since eye movement is significantly simpler than the whole-body motion. Second, 3-second eye contact obtained better scores regardless of the imitation-based or rule-based eye behavior. Third, subjects might regard long eyeblink as voluntary eyeblink, with the intention to break eye contact much more noticeable than short eyeblink. Finally, the impression of eyeblink between male and female subjects is different. Compared to male subjects, female subjects preferred short eyeblink over long eyeblinks and regarded that short eyeblink might be one of the keys to make eye contact more suitable than long eyeblink, even though the eye-contact duration was the same in both eyeblinks.

6.3 Eye Behaviors for Attentive Listening

This dissertation clarified how to generate eyeblinks and nodding to make an android's listening behaviors be perceived as attentive listening. The hypothesis is that if an android acting as a listener imitates eyeblinks and nodding of a human speaker in face-to-face communication, the android can make a human speaker perceive its listening behavior as attentive listening. First, we developed a real-time method to imitate human's eyeblinks and nodding using an android by following the trajectory of each motion and tracking the human's facial area in real time. Second, we conducted the subject experiment to clarify if eyeblink is a necessary cue of attentive listening behaviors for a human's speaker and if imitation of eyeblinks and nodding makes a human speaker perceive attentive listening, by comparing the imitation to simple duration-based motions, the motions generated at breakpoints, and the motions generated in simple-duration and/or at breakpoints.

Through a subjective evaluation, we reached two findings: 1) eyeblink was effective as a cue of attentive listening behaviors, 2) the imitation of both eyeblinks and nodding did not improve the perception of attentive listening.

6.4 Future Work

6.4.1 Gaze Calibration

In the evaluation, we could not conduct a subjective experiment because it is difficult to take facial pictures from various views for stimuli due to the limitation of the size of cameras we used and space where the cameras were placed. Hence, we will consider how to evaluate the accuracy of the proposed calibration method for an android's eyes' pitch. Furthermore, we will try to combine our gaze calibration method with planning of head and body motions to establish good eye contact between an android and a human.

6.4.2 Eye Behaviors for Natural Talking

In the current implementation, an android only moved the eyes and mouth. To emulate natural conversation, it is better to also move the rest of the face, head and shoulders. To move the whole face, we are considering two main approaches. The first approach is that by imitating human facial movements with the extracted facial features, we can generate the face movements [49]. The second approach is that with the emotion recognition model from facial features, we can generate the facial movements [50]. To move the head, for example, we will have to combine gaze motion planning based on the head-eye combination such as [13] with the current implementation.

6.4.3 Eye Behaviors for Attentive Listening

We would like to conduct the same evaluation with elderly people as subjects to elucidate how to make them perceive the listening behavior as attentive listening because our research has a possibility to prevent lonely death which is a social problem in Japan [51]. Moreover, we would like to adopt automatic recognition of breakpoints of speech as exemplified in [52] by using voice recognition tools to realize full autonomous android that can generate nonverbal behaviors for attentive listening.

6.4.4 Perception of a Human's Gazed Object

It is necessary to handle with situation for turn taking (*i.e.*, switching a role of a speaker or a listener) to realize a naturalistic communication. In particular, it is important to estimate and understand a human's intentions. Hence, with a focus on gazed objects (*i.e.*, the objects humans are looking at), we also proposed a real-time gazed object identification method from a service robot's view by searching for an object along the estimated face direction, as shown in Figure 32. As future work, we would like to combine this method with the know to realize a naturalistic communication between a human and a humanoid robot.

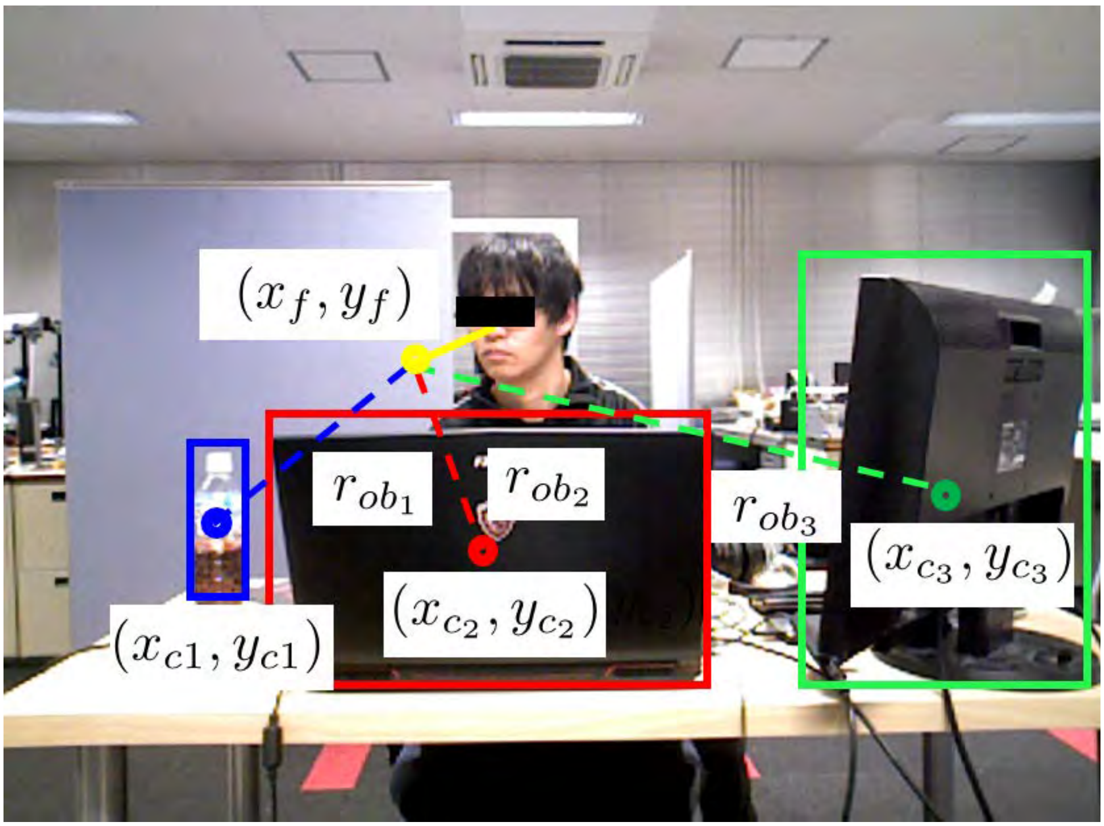

Figure 32. Overview of a gazed object identification by searching for an object along the estimated face direction.

Acknowledgments

First of all, I am deeply greatful to my actual main supervisor, a former Professor, Tsukasa Ogasawara (currently Executive Director at NAIST) who was the head of the Robotics Laboratory. He had accepted me as a student to and helped me to complete my endeavor, and given me many chances to broaden my perspective.

I am grateful to Professor Takahiro Wada, who is the head of the Human-Robotics Laboratory, to support me kindly to hold my Ph.D. defense, as my formal supervisor.

I am very grateful to Professor Satoshi Nakamura, who is the head of the Augmented Human Communication Laboratory. He gave me his meaningful comments for my dissertation and I strongly remember that he kindly gave me encouragement for my work when I met him at Karlsruhe Institute of Technology, Germany.

I am also very grateful to a former Associate Professor Jun Takamatsu (currently Visiting Professor). His insightful comments and questions from his immense knowledge and passion for Robotics taught me a lot to construct my research approach and style. Moreover, he provided me a chance (especially, the JST CREST project) to connect to some famous researchers to push my research to greater heights.

I am also very grateful to a former Assistant Professor Gustavo Alfonso Garcia Ricardez (currently Visiting Associate Professor). As my friendly and strong colleague, I could have fruitful experiences together. If not his help and passion, I could not publish my some research papers.

I am also grateful to Assistant Professor Sung-Gwi Cho to review my dissertation regardless of his main research area.

In addition, I am grateful to a former Assistant Professor, Ming Ding to give me a lot comments for my research and career (currently Designated Associate Professor at Nagoya University). Especially, he helped me a lot when I visited Carnegie Mellon University, USA.

I am also grateful to Associate Professor Atsushi Nakazawa, Kyoto University. He has invited me to build the team of the JST CREST project "Computational and Cognitive Neuroscientific Approaches for Understanding the Tender Care." He also helped me to design the conception of my studies about eye behaviors and gave me a chance to obtain my first research grant. Moreover, I also would like to thank the rest of the members of his CREST team members to give me some fruitful comments on my studies.

I am also grateful to Professor Tamim Asfour, who is the head of High Performance Humanoid Technologies (H^2T) Laboratory, Karlsruhe Institute of Technology (KIT), Germany, and the H^2T laboratory members. He had kindly accepted me for my study abroad. My research stay in Germany brought me many different perspectives. Moreover, I am grateful to the Japan Public-Private Partnership Student Study Abroad Program "TOBITATE! Young Ambassador Program" to realize my study abroad.

I am also grateful to Emeritus Professor Masahide Kaneko and Associate Professor Tomoaki Nakamura at the University of Electro-Communications (UEC). It was the first research activity experience for me under their supervision when I was an undergraduate student. Through this experience, I was brought under the spell of research activity.

I would like to thank my former research colleagues, Masahiro Iwamoto, Tetsuya Sano, Makoto Ikawa, and Takumi Nakamura. I was glad to achieve exciting research works together. Moreover, I would like to thank my former doctoral colleagues, Dr. Felix Wolf Hans Erich von Drigalski (currently Senior Researcher at OMRON SINIC X Corporation), Dr. Satoki Tsuichihara (currently Assistant Professor at University of Fukui), Dr. Lotfi El Hafi (currently Research Assistant Professor at Ritsumeikan University), Dr. Akio Noda (currently Professor at Osaka Institute of Technology), and Dr. Yuya Hakamata. They gave me a lot of advice for my research life.

I would like to thank the administrative staffs of the Robotics Laboratory, Michiyo Owaki and Miki Ikeda for their support with dedication for the office work sides of my research work.

I also would like thank the rest of all my former colleagues at NAIST and UEC to join me in my life all throughout.

Finally, I would like to thank my parents and brother from the bottom of my heart for their affection and kind understanding of my life.

Publication List

Refereed Journal Papers

1. **Akishige Yuguchi**, Tetsuya Sano, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, Atsushi Nakazawa, and Tsukasa Ogasawara, "Evaluating Imitation and Rule-based Behaviors of Eye Contact and Blinking Using an Android for Conversation," *Advanced Robotics*, 2021, doi: 10.1080/01691864.2021.1928544.

Refereed International Conference Papers

1. Tetsuya Sano, **Akishige Yuguchi**, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, Atsushi Nakazawa, and Tsukasa Ogasawara, "Evaluating Imitation of Human Eye Contact and Blinking Behavior Using an Android for Human-like Communication," in *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2019)*, pp. 1-6, New Delhi, India, October, 2019.

2. **Akishige Yuguchi**, Tomoaki Inoue, Gustavo Alfonso Garcia Ricardez, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "Real-Time Gazed Object Identification with a Variable Point of View Using a Mobile Service Robot," in *Proceedings of the 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2019)*, pp. 1-6, New Delhi, India, October, 2019.

3. **Akishige Yuguchi**, Gustavo Alfonso Garcia Ricardez, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "Gaze Calibration for Human-Android Eye Contact Using a Single Camera," in *Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO 2017)*, pp. 883-888, Macau SAR, China, December, 2017.

Unrefereed Conference Papers

1. **湯口 彰重**, GARCIA RICARDEZ Gustavo Alfonso, 高松 淳, 中澤 篤志, 小笠原 司, "より良い聞き手のためのアンドロイドによる瞬目と頷きの模倣効

果の検討,” 第 38 回日本ロボット学会学術講演会 (RSJ 2020), 212-02, オンライン, 2020 年 10 月.

2. **Akishige Yuguchi**, Jun Takamatsu, Atsushi Nakazawa, and Tsukasa Ogasawara, “Imitation of Human Eyeblinks and Nodding Using an Android Toward Attentive Listening,” in *Proceedings of the 32nd JSME Conference on Robotics and Mechatronics 2020 (ROBOMECH 2020)*, 1P2-E16, Kanazawa (online), Japan, May, 2020 (in Japanese).

3. 高松 淳, 豊島 健太, 佐野 哲也, **湯口 彰重**, 中澤 篤志, ガルシア リカルデス グスタボ アルフォンソ, 丁 明, 小笠原 司, “ロボットによる見る・触れる動作の模倣とそれを通じた評価,” 第 *37*回日本ロボット学会学術講演会 *(RSJ 2019)*, 1N3-05, 東京, 2019 年 9 月.

4. Tetsuya Sano, **Akishige Yuguchi**, Atsushi Nakazawa, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, and Tsukasa Ogasawara, “ Human-like Eye Movement Behaviors for Android Robots Based on Human Observation,” in *Proceedings of the 19th SICE System Integration Division Conference (SI 2018)*, pp. 815-819, Osaka, Japan, December, 2018 (in Japanese).

5. **湯口 彰重**, 丁 明, 高松 淳, 小笠原 司, “アンドロイドと人間の視線に対する知覚の比較に基づく眼球運動キャリブレーションの効果,” 第 *35*回日本ロボット学会学術講演会 *(RSJ 2017)*, 2L2-04, 川越, 2017 年 9 月.

6. **Akishige Yuguchi**, Gustavo Alfonso Garcia Ricardez, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, “Gaze Calibration Method for an Android to Make Eye Contact with a Human,” in *Proceedings of the 17th SICE System Integration Division Annual Conference (SI 2016)*, pp. 2449-2451, Sapporo, Japan, December, 2016 (in Japanese).

7. **Akishige Yuguchi**, Masahiro Yoshikawa, Jun Takamatsu, and Tsukasa Ogasawara, “Generating Upper Body Motions for an Android Using Kinect v2,” in *Proceedings of the 28th JSME Conference on Robotics and Mechatronics 2016 (ROBOMECH 2016)*, 1A1-15a4, Yokohama, Japan, June, 2016 (in Japanese)

Other Refereed International Conference Papers

1. Kiyu Nishida, **Akishige Yuguchi**, Kazuhiro Jo, Paul Modler, and Markus Noisternig, "Border: A Live Performance Based on Web AR and a Gesture-Controlled Virtual Instrument," in *Proceedings of the 19th International Conference on New Interfaces for Musical Expression (NIME 2019)*, pp. 43-46, Porto Alegre, Brazil, June, 2019.

2. Takumi Nakamura, **Akishige Yuguchi**, Maël Aubert, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, and Tsukasa Ogasawara, "Ontology Generation Using GUI and Simulation for Service Robots to Operate Home Appliances," in *Proceedings of the 3rd IEEE International Conference on Robotic Computing (IRC 2019)*, pp. 315-320, Naples, Italy, February, 2019.

3. Masahiro Iwamoto, **Akishige Yuguchi**, Masahiro Yoshikawa, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, and Tsukasa Ogasawara, "Human-like Subconscious Behaviors for an Android When Telling a Lie," in *Proceedings of the 27th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2018)*, pp. 390-395, Nanjing, China, August, 2018.

4. Makoto Ikawa, Etsuko Ueda, **Akishige Yuguchi**, Gustavo Alfonso Garcia Ricardez, Ming Ding, Jun Takamatsu and Tsukasa Ogasawara, "Quantification of Elegant Motion for Receptionist Android Robot," in *Proceedings of the 8th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management, Part of HCI International 2017*, pp. 435-446, Vancouver, Canada, July, 2017.

5. Gustavo Alfonso Garcia Ricardez, Lotfi El Hafi, Felix von Drigalski, Rodrigo Elizalde Zapata, Chika Shiogama, Kenta Toyoshima, Pedro Miguel Urigüen Eljuri, Marcus Gall, **Akishige Yuguchi**, Arnaud Delmotte, Viktor Gerhard Hoerig, Wataru Yamazaki, Seigo Okada, Yusuke Kato, Ryutaro Futakuchi, Kazuo Inoue, Katsuhiko Asai, Yasunao Okazaki, Masaki Yamamoto, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "Climbing on Giant's Shoulders: Newcomer's Road into the Amazon Robotics Challenge 2017," in *Proceedings of the Warehouse Picking Automation Workshop*

2017 (WPAW 2017) at ICRA 2017, pp. 1-4, Singapore, May, 2017.

Awards

1. **Outstanding Presentation Award**
   for Tetsuya Sano, **Akishige Yuguchi**, Atsushi Nakazawa, Gustavo Alfonso Garcia Ricardez, Jun Takamatsu, and Tsukasa Ogasawara, "*Human-like Eye Movement Behaviors for Android Robots Based on Human Observation*," in the 19th SICE System Integration Division Annual Conference (SI 2018), Osaka, Japan, December, 2018.

2. **Finalist of Best Paper in Biomimetics Award**
   for **Akishige Yuguchi**, Gustavo Alfonso Garcia Ricardez, Ming Ding, Jun Takamatsu, and Tsukasa Ogasawara, "*Gaze Calibration for Human-Android Eye Contact Using a Single Camera*," in the 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO 2017), Macau SAR, China, December 2017.

References

[1] Y. Frumer, "The short, strange life of the first friendly robot," IEEE Spectrum, 2020, https://spectrum.ieee.org/robotics/humanoids/the-short-strange-life-of-the-first-friendly-robot.

[2] A. Mehrabian, *Silent messages : implicit communication of emotions and attitudes*, 2nd ed. Belmont, CA: Wadsworth Pub. Co., 1981.

[3] C. L. Kleinke, "Gaze and eye contact: A research review," *Psychological Bulletin*, vol. 100, no. 1, pp. 78–100, 1986.

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

[5] A. Kendon, "Some functions of gaze direction in social interaction," *Acta Psychol*, vol. 26, pp. 22–63, 1967.

[6] R. Vertegaal, R. Slagter, G. Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proc. of the Annual SIGCHI Conference on Human Factors in Computing Systems (CHI 2001)*, 2001, pp. 301–308.

[7] A. McCarthy, K. Lee, S. Itakura, and D. W. Muir, "Cultural display rules drive eye gaze during thinking," *Journal of cross-cultural psychology*, vol. 37, no. 6, pp. 717–722, 2006.

[8] T. Nakano and S. Kitazawa, "Eyeblink entrainment at breakpoints of speech," *Experimental Brain Research*, vol. 205, pp. 577–581, 2010.

[9] S. Nishio and H. Ishiguro, "Android science research for bridging humans and robots," *The Journal of Institute of Electronics, Information and Communication Engineers*, vol. 91, no. 5, pp. 411–416, 2008, (in Japanese).

[10] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, "Can machines think? interaction and perspective taking with robots investigated via fmri," *PLOS ONE*, vol. 3, no. 7, 2008.

[11] T. Komatsu and S. Yamada, “Adaption gap hypothesis: How differences between users’ expected and perceived agent functions affect their subjective impression,” *Journal of Systemics, Cybernetics and Informatics*, vol. 9, no. 1, pp. 67–74, 2011.

[12] Z. Ye, Y. Li, Y. Liu, C. Bridges, A. Rozga, and J. M. Rehg, “Detecting bids for eye contact using a wearable camera,” in *Proc. of the 2015 11th IEEE Int. Conf. and Workshops on Automatic Face and Gesture Recognition (FG 2015)*, vol. 1, 2015, pp. 1–8.

[13] Y. Kondo, M. Kawamura, K. Takemura, J. Takamatsu, and T. Ogasawara, “Gaze motion planning for android robot,” in *Proc. of the 6th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI 2011)*, 2011, pp. 171–172.

[14] T. Yamamori, D. Sakamoto, S. Nishio, H. Ishiguro, and N. Hagita, “Inspection of the condition of eye contact with android,” in *Proc. of the 2007 Annual Human-Agent Interaction Symposium, 1B-2*, 2007, (in Japanese).

[15] O. Palinko, A. Sciutti, Y. Wakita, Y. Matsumoto, and G. Sandini, “If looks could kill: Humanoid robots play a gaze-based social game with humans,” in *Proc. of the 16th Annual IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids 2016)*, 2016, pp. 905–910.

[16] D. F. Glas, T. Minato, C. T. Ishi, T. Kawahara, and H. Ishiguro, “Erica: The erato intelligent conversational android,” in *Proc. of the 25th Annual IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN 2016)*, 2016, pp. 22–29.

[17] J. Even, C. T. Ishi, and H. Ishiguro, “Using sensor network for android gaze control,” in *Proc. of the 43rd Japanese Society for Artificial Intelligence Meeting of Special Interest Group on AI Challenges*, 2015, pp. 29–34.

[18] T. Minato, M. Shimada, S. Itakura, K. Lee, and H. Ishiguro, “Evaluating the human likeness of an android by comparing gaze behaviors elicited by the android and a person,” *Advanced Robotics*, vol. 20, no. 10, pp. 1147–1163, 2006.

[19] M. Shimada, Y. Yoshikawa, M. Asada, N. Saiwaki, and H. Ishiguro, "Effects of observing eye contact between a robot and another person," *International Journal of Social Robotics*, vol. 3, no. 2, pp. 143–154, 2011.

[20] K. Tatsukawa, T. Nakano, H. Ishiguro, and Y. Yoshikawa, "Eyeblink synchrony in multimodal human-android interaction," *Scientific reports*, vol. 6, p. 39718, 2016.

[21] D. Lala, P. Milhorat, K. Inoue, T. Zhao, and T. Kawahara, "Multimodal interaction with the autonomous android," in *Proc. of the 18th ACM Int. Conf. on Multimodal Interaction (ICMI 2016)*, 2016, pp. 417–418.

[22] Y. Kondo, K. Takemura, J. Takamatsu, and T. Ogasawara, "A gesture-centric android system for multi-party human-robot interaction," *Journal of Human-Robot Interaction*, vol. 2, no. 1, pp. 133–151, 2013.

[23] L. Luo, N. Koyama, K. Ogawa, and H. Ishiguro, "Robotic eyes that express personality," *Advanced Robotics*, vol. 33, no. 7-8, pp. 350–359, 2019.

[24] M. Iwamoto, A. Yuguchi, M. Yoshikawa, G. A. Garcia Ricardez, J. Takamatsu, and T. Ogasawara, "Human-like subconscious behaviors for an android when telling a lie," in *Proc. of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2018)*, 2018, pp. 390–395.

[25] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita, and T. Miyamoto, "The effects of responsive eye movement and blinking behavior in a communication robot," in *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, 2006, pp. 4564–4569.

[26] H. Lehmann, A. Roncone, U. Pattacini, and G. Metta, "Physiologically inspired blinking behavior for a humanoid robot," in *Proc. of the 8th International Conference on Social Robotics (ICSR 2016)*, 2016, pp. 83–93.

[27] M. K. Pan, S. Choi, J. Kennedy, K. McIntosh, D. C. Zamora, G. Niemeyer, J. Kim, A. Wieland, and D. Christensen, "Realistic and interactive robot gaze," in *Proc. of the 2020 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2020, pp. 11 072–11 078.

[28] L. Hardjasa and A. Nakazawa, "An examination of gaze during conversation for designing culture-based robot behavior," in *Proc. of the 12th Int. Conf. on Social Computing and Social Media, Held as Part of HCII 2020*, 2020, pp. 475–488.

[29] T. Todo, "Seer: Simulative emotional expression robot," in *Proc. of the ACM SIGGRAPH 2018 Emerging Technologies*, 2018, pp. 15:1–15:2.

[30] A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi, "Designing and evaluating a social gaze-control system for a humanoid robot," *IEEE Transactions on Human-Machine Systems*, vol. 44, no. 2, pp. 157–168, 2014.

[31] A. Zaraki, M. B. Dehkordi, D. Mazzei, and D. De Rossi, "An experimental eye-tracking study for the design of a context-dependent social robot blinking model," in *Proc. of the 2014 Int. Conf. on Biomimetic and Biohybrid Systems*. Springer International Publishing, 2014, pp. 356–366.

[32] M. W. Hoffman, D. B. Grimes, A. P. Shon, and R. P. Rao, "A probabilistic model of gaze imitation and shared attention," *Neural Networks*, vol. 19, no. 3, pp. 299–310, 2006.

[33] J. Duque-Domingo, J. Gómez-García-Bermejo, and E. Zalama, "Gaze control of a robotic head for realistic interaction with humans," *Frontiers in Neurorobotics*, vol. 14, no. 34, 2020.

[34] M. Gleicher, "Retargetting motion to new characters," in *Proc. of the 25th Annual Conf. on Computer Graphics and Interactive Techniques*, 1998, pp. 33–42.

[35] L. Kovar, M. Gleicher, and F. Pighin, "Motion graphs," *ACM Transaction on Graphics*, vol. 21, no. 3, pp. 473–482, 2002.

[36] Q. McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.

[37] M. Kassner, W. Patera, and A. Bulling, "Pupil: An open source platform for pervasive eye tracking and mobile gaze-based interaction," in *Proc. of*

the 2014 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing (UbiComp 2014), 2014, pp. 1151–1160.

[38] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Openface: An open source facial behavior analysis toolkit," in *Proc. of the 2016 IEEE Winter Conf. on Applications of Computer Vision (WACV2016)*, 2016, pp. 1–10.

[39] N. Binetti, C. Harrison, A. Coutrot, and I. Maresxhal, "Pupil dilation as an index of preferred mutual gaze duration," *Royal Society Open Science*, vol. 3, no. 7, p. 160086, 2016.

[40] J. A. Stern, L. C. Walrath, and R. Goldstein, "The endogenous eyeblink," *Psychnophysiology*, vol. 21, pp. 22–33, 1984.

[41] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal statistical society: series B (Methodological)*, vol. 57, no. 1, pp. 289–300, 1995.

[42] S. L. Rogers, C. P. Speelman, O. Guidetti, and M. Longmuir, "Using dual eye tracking to uncover personal gaze patterns during social interaction," *Scientific reports*, vol. 8, no. 1, pp. 1–9, 2018.

[43] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *Proc. of the 13th IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG 2018)*, 2018, pp. 59–66.

[44] P. Ekman, W. V. Friesen, and J. C. Hager, *Facial action coding system*, the manual on cd rom ed. A Human Face, Salt Lake City, 2002.

[45] M. Yoshikawa, Y. Matsumoto, M. Sumitani, and H. Ishiguro, "Development of an android robot for psychological support in medical and welfare fields," in *Proc. of the 2011 Annual IEEE Int. Conf. on Robotics and Biomimetics (ROBIO 2011)*, 2011, pp. 2378–2383.

[46] N. M. Fraser and G. N. Gilbert, "Simulating speech systems," *Computer Speech & Language*, vol. 5, no. 1, pp. 81–99, 1991.

[47] C. T. Ishi, H. Ishiguro, and N. Hagita, "Analysis of relationship between head motion events and speech in dialogue conversations," *Speech Communication*, vol. 57, pp. 233–243, 2014.

[48] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[49] E. Magtanong, A. Yamaguchi, K. Takemura, J. Takamatsu, and T. Ogasawara, "Inverse kinematics solver for android faces with elastic skin," in *Latest Advances in Robot Kinematics.* Springer, 2012, pp. 181–188.

[50] A. Majumder, L. Behera, and V. K. Subramanian, "Emotion recognition from geometric facial features using self-organizing map," *Pattern Recognition*, vol. 47, no. 3, pp. 1282–1293, 2014.

[51] N. Onishi, "A generation in japan faces a lonely death," The New York Times, 2017, https://www.nytimes.com/2017/11/30/world/asia/japan-lonely-deaths-the-end.html.

[52] K. Sakai, C. T. Ishi, T. Minato, and H. Ishiguro, "Online speech-driven head motion generating system and evaluation on a tele-operated robot," in *Proc. of the 24th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN 2015)*, 2015, pp. 529–534.

Appendix

A. Neck Kinematics of *Actroid-SIT* for Head Movements

The neck of the *Actroid-SIT* has four actuators. Two actuators of all them are aligned with the roll axis (*i.e.*, right/left down). Hence, we can assume that the neck has a total of three DOFs and then calculate the orientation of the head with the model expressed by

$$\theta_\alpha = \begin{cases} (x_{\mathrm{cmd}_1} - \delta_1)\ \theta_{lim,\alpha_1}/\delta_1 & (0 < x_{\mathrm{cmd}} < \delta_1), \\ \theta_{lim,\alpha_1} & (\text{otherwise}), \\ (\delta_2 - x_{\mathrm{cmd}_2})\ \theta_{lim,\alpha_2}/\delta_2 & (0 < x_{\mathrm{cmd}} < \delta_2), \\ \theta_{lim,\alpha_2} & (\text{otherwise}), \end{cases} \tag{9}$$

$$\theta_\beta = \begin{cases} (y_{\mathrm{cmd}} - y_{\mathrm{neutral}})\ \theta_{lim,\beta}/\delta & (0 < y_{\mathrm{cmd}} < \delta), \\ \theta_{lim,\beta} & (\text{otherwise}), \end{cases} \tag{10}$$

$$\theta_\gamma = \begin{cases} (z_{\mathrm{cmd}} - z_{\mathrm{neutral}})\ \theta_{lim,\gamma}/\delta & (0 < z_{\mathrm{cmd}} < \delta), \\ \theta_{lim,\gamma} & (\text{otherwise}). \end{cases} \tag{11}$$

where θ_α, θ_β, and θ_γ are the roll, pitch, and yaw angles of the head, x_{cmd}, y_{cmd}, and z_{cmd} are the input values of the controller for the neck's actuators, y_{neutral} and z_{neutral} are the neutral input values of the controller for the actuators, $\theta_{lim,\alpha}$, $\theta_{lim,\beta}$, and $\theta_{lim,\gamma}$ are the maximum angles of the actuators, and δ is the resolution of the controller.

B. Primary Experiment Regarding Eye Behaviors for Attentive Listening

B.1 Protocol

B.1.1 Experimental Design

From the existing findings from psychology and cognitive science, we focus on two cues of the nonverbal behaviors for attentive listening. The first cue is eyeblink. As described in Chapter 2, the eyeblink of a human in a listener role synchronizes the eyeblink of an android in a speaker role [20]. Hence, we think that if the android in a listener role imitates the eyeblink of a human in a speaker, it can give humans better impressions for attentive listening. The second cue is nodding. Yoshikawa *et al.* [45] verified that nodding motion when an android in a listener role is effective.

Therefore, in this experiment, we hypothesize as follows:

- Eyeblink is a cue of attentive listening behaviors for a human's talker
- Imitation of eyeblink and nodding makes a human's talker feel attentive listening

We ask a subject in a talker role to sit at 1.0 m from the android as shown in Figure 31 by fixing the height is the same as the android's gaze to establish the eye contact situation. Then, we ask the subject to talk in about one minute with the android and evaluate the impression on each competitor with a questionnaire. For talking with the android, we ask the subject to introduce himself/herself because we seem that it can be easy to talk to each competitor repeatedly. Before starting the experiment, we ask the subject to practice the talk a few times to fit it within one minute.

Furthermore, as we manipulate only the eyelids and neck of the android, it wears a mask on its face to avoid the effect of the impression on mouth movement and facial expression, as the same in Chapter 4.4.

Table 7. Patterns for the primary experiment.

		Eyeblink		
		Imitation	Simple	No
Nodding	Imitation	Pattern 1	Pattern 3	Pattern 5
	Simple	Pattern 2	Pattern 4	Pattern 6

B.1.2 Patterns

As aforementioned, we assume that there are two important cues (*i.e.*, eyeblink and nodding) of the nonverbal behaviors for attentive listening. For the evaluation, we consider the two types of conditions. The first type of condition is eyeblink. We decide to prepare for three types of eyeblinks as the imitation, the simple duration-based, and nothing. In the simple duration-based eyeblink, we decide that the eyeblink occurs every 3 seconds because 3-second eye contact is prefer to humans [39], as described in Chapter 4.

The second type of condition is nodding. We decide to prepare for three types of nodding as the imitation and the simple duration-based because nodding motions look more dynamic than eye behaviors, we decide to generate the conditions by combining with nodding motions. In the simple duration-based nodding, we decide that the nodding occurs every 3 seconds in a similar way to the eyeblink case. Thus, we combine with these types to generate the conditions, then the total conditions are six conditions, as shown in Table 7.

B.1.3 Evaluation Method

To evaluate the human impression for each condition, we asked the subjects to score the following five items (in Japanese) on a scale from 1 to 7 points (lower score is better):

(a): Kind (1) - Unkind (7)

(b): Attractive (1) - Unattractive (7)

(c): Humanlike (1) - Unhumanlike (7)

(d): Feel easy to talk (1) - Feel uneasy to talk (7)

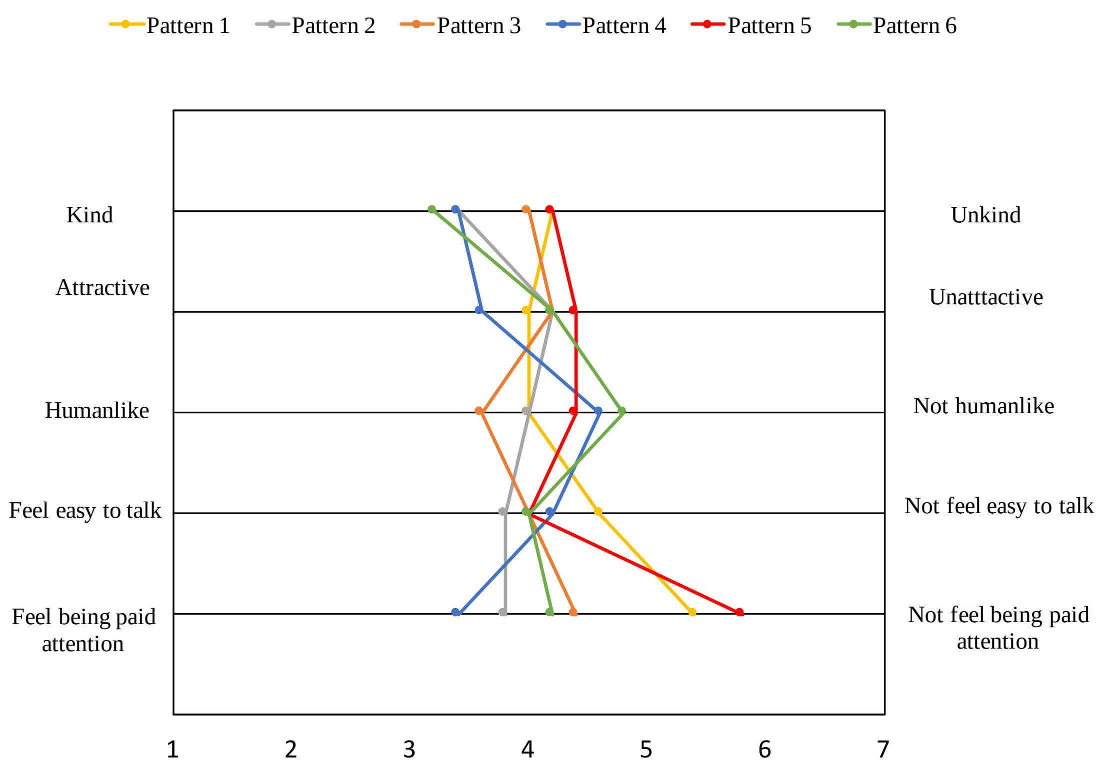

Figure 33. Experimental result of the subjective evaluation.

(e): Feel being paid attention (1) - Feel not being paid attention (7)

Items (a), (b), and (c) ask the impression of whole nonverbal behaviors. Items (d) and (e) directly ask the impression of the listener role.

B.2 Result

We recruited five male subjects (mean age: 23.0), all students of Nara Institute of Science and Technology under the approval from the ethics committee of the same institute. Figure 33 shows the graph plotting with the average scores of the subjects.

B.2.1 Impression of Eyeblink

From the result, compared to the imitation nodding conditions, in order of Patterns 3 (*i.e.*, the simple duration-based eyeblink), 5 (*i.e.*, no eyeblink), and 1 (*i.e.*, the imitation of eyeblink), the scores obtained better in the items (c) and (e).

Compared to the simple duration-based nodding conditions, the score in Pattern 4 obtained the best in items (b) and (e), and the score in Pattern 2 obtained the best in items (c) and (d).

Moreover, the score in Pattern 4 obtained the worse in all items except for item (a).

B.2.2 Impression of Combined Patterns

From the result, the score of Pattern 1 (*i.e.*, the imitation of eyeblink and nodding) did not obtain better in all items, especially it obtained the worst in items (d) and (e). However, the score of Pattern 2 (*i.e.*, the imitation of eyeblink with the simple duration based nodding) obtained better than Pattern 1's one in items (a), (d), and (e).

Compared Pattern 3 (*i.e.*, the imitation of nodding with the simple duration-based eyeblink) to Pattern 4 (*i.e.*, the simple duration-based nodding with the simple duration-based eyeblink), the similar tendency occurred in items (a), (b), and (e). However, the score of Pattern 4 obtained worse than Pattern 3's one in item (c).

Compared Pattern 5 (*i.e.*, the imitation of nodding with no eyeblink) to Pattern 6 (*i.e.*, the simple duration-based nodding with no eyeblink), the similar tendency occurred in items (a) and (e). Compared Pattern 6 to other Patterns, the score was the worst in item (c).

Moreover, compared the type of the simple duration-based nodding to the type of the imitation, the simple type obtained better in items (a) and (d).

B.3 Discussion

B.3.1 Eyeblink Effect

From the result, it was suggested that eyeblink gives better impressions than no eyeblink in almost all items. However, it was suggested that the simple duration-based eyeblink may make the human impression better than imitation of eyeblink in especially item (e) *i.e.*, feeling of being paid attention because the imitation of eyeblink did not affect better impressions. This cause might be that it is not easy technically to emulate to imitate short eyeblinks in the current system of the imitation motion generation.

Moreover, it was suggested that even Patterns in the same imitation of eyeblink, Patterns of nodding affect the human impression because the tendency between the scores of Patterns 1 and 2 is different significantly in items (d) and (e).

B.3.2 Effect of Combination of Eyeblink and Nodding

From the results, the both imitation of eyeblink and nodding did not affect better impressions. As aforementioned, Patterns of nodding affect the human impression of the eyeblink.

Hence, compared the simple duration-based nodding to the imitation of nodding, as described in Chapter B.2, the simple one affected the better impression in the aspect of the kindness and feeling of being paid attention. This cause might be that the android might imitate small head movements and then the generated nodding that subjects cannot recognize affected the human impression because the head moves down the bottom in the simple one, whereas the head follows a trajectory regardless the range in the imitation.