

Bachelor Thesis



**Czech
Technical
University
in Prague**

F3

**Faculty of Electrical Engineering
Department of Cybernetics**

Naturalistic Control of Eyes of a Humanoid Robot during an Interactive Game

Daria Mikhaylovskaya

**Supervisor: Mgr. Matěj Hoffmann, Ph.D.
Supervisor–specialist: Ing. Jakub Rozlivek
Study program: Open Informatics
May 2024**

I. Personal and study details

Student's name: **Mikhaylovskaya Daria** Personal ID number: **507405**
Faculty / Institute: **Faculty of Electrical Engineering**
Department / Institute: **Department of Cybernetics**
Study program: **Open Informatics**
Specialisation: **Artificial Intelligence and Computer Science**

II. Bachelor's thesis details

Bachelor's thesis title in English:

Naturalistic Control of Eyes of a Humanoid Robot during an Interactive Game

Bachelor's thesis title in Czech:

P írozené ovládání o í humanoidního robota b hem interaktivní hry

Guidelines:

The goal of the project is to develop control strategies for the eyes of a humanoid while playing an interactive game with a human (e.g., <https://youtu.be/gw8JB-1R3bs>). The main objective is for the eye movements to be perceived as naturalistic by the human (e.g., (Kompatsiari et al., 2021; Lehmann et al., 2017; Stanton & Stevens, 2017)). A secondary objective is to aid safety of the interaction in close physical proximity (Docekal et al., 2022).

1. Familiarize yourself with the iCub humanoid robot and its vision system (the cameras in the eyes with pan, tilt, and vergence) and RGB-D camera mounted on its head.
2. Implement a gaze controller for the iCub using eyes only - no neck joints. (Roncone et al., 2016) can serve as a starting point.
3. Detect a human in the camera image (e.g., using (Bazarevsky et al., 2019) or (Xu et al., 2022)) and gaze into her eyes.
4. Develop a gaze controller for the robot eyes in the context of a card game, considering the following objectives:
 - a. Naturalistic gaze control as perceived by the human participants (e.g., switching looking into the participants' eye and at the cards on the table),
 - b. Aiding safety of the interaction by providing approximate position of the human body parts when they are not in view of the RGB-D camera.
5. Evaluate the human-likeness of selected gaze control strategies on a sample of human participants (e.g., using (Bartneck et al., 2009)).

Bibliography / sources:

- [1] Bartneck, C., Kuli, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1, 71–81.
- [2] Bazarevsky, V., Kartynnik, Y., Vakunov, A., Raveendran, K., & Grundmann, M. (2019). BlazeFace: Sub-millisecond neural face detection on mobile gpus. *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*.
- [3] Docekal, J., Rozlivek, J., Matas, J., & Hoffmann, M. (2022). Human keypoint detection for close proximity human-robot interaction. *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2022)*.
- [4] Kompatsiari, K., Ciardo, F., Tikhanoff, V., Metta, G., & Wykowska, A. (2021). It's in the eyes: The engaging role of eye contact in HRI. *International Journal of Social Robotics*, 13, 525–535.
- [5] Lehmann, H., Keller, I., Ahmadzadeh, R., & Broz, F. (2017). Naturalistic Conversational Gaze Control for Humanoid Robots—A First Step. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyszel, & H. He (Eds.), *Social Robotics* (pp. 526–535). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_52
- [6] Roncone, A., Pattacini, U., Metta, G., & Natale, L. (2016). A Cartesian 6-DoF Gaze Controller for Humanoid Robots. *Robotics: Science and Systems*, 2016.
- [7] Stanton, C. J., & Stevens, C. J. (2017). Don't stare at me: The impact of a humanoid robot's gaze upon trust during a cooperative human–robot visual task. *International Journal of Social Robotics*, 9, 745–753.
- [8] Xu, Y., Zhang, J., Zhang, Q., & Tao, D. (2022). ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *Advances in Neural Information Processing Systems*.

Name and workplace of bachelor's thesis supervisor:

doc. Mgr. Mat j Hoffmann, Ph.D. Vision for Robotics and Autonomous Systems FEE

Name and workplace of second bachelor's thesis supervisor or consultant:

Ing. Jakub Rozlivek Vision for Robotics and Autonomous Systems FEE

Date of bachelor's thesis assignment: **15.01.2024** Deadline for bachelor thesis submission: **24.05.2024**

Assignment valid until: **21.09.2025**

doc. Mgr. Mat j Hoffmann, Ph.D.
Supervisor's signature

prof. Dr. Ing. Jan Kybic
Head of department's signature

prof. Mgr. Petr Páta, Ph.D.
Dean's signature

III. Assignment receipt

The student acknowledges that the bachelor's thesis is an individual work. The student must produce her thesis without the assistance of others, with the exception of provided consultations. Within the bachelor's thesis, the author must state the names of consultants and include a list of references.

Date of assignment receipt

Student's signature

Acknowledgements

I would like to express my deep gratitude to my thesis supervisor, Mgr. Matěj Hoffmann, Ph.D., for overseeing my research and guiding me in writing my thesis. His guidance and the particular expertise he provided over the past three years of my employment in his laboratory have been extremely valuable. I am extremely grateful for his consistently calm and supportive presence in all situations, which greatly contributed to my ability to navigate challenges and maintain focus on my research objectives.

I would also like to express my gratitude to my Supervisor Specialist, Ing. Jakub Rozlivek, for his exceptional patience and the significant amount of time he contributed to my thesis work. His assistance in all possible situations, along with his profound expertise in this field, have been crucial to my work. I am especially grateful for his remarkable ability to answer my questions quickly, anytime of day or night, which has been extremely helpful. Jakub has consistently provided exceptional support and kindness throughout this journey. His influence on my work has been invaluable.

Furthermore, I extend my thanks to RNDr. Petr Štěpán, Ph.D., for his invaluable support and for sharing his expertise, which has significantly influenced the direction of my professional journey. I am grateful for his role as a supportive professor and teacher, whose guidance has greatly contributed to my academic and professional growth.

Lastly, I wish to express a special thanks to the Czech Technical University in Prague for providing the necessary facilities and a conducive environment for my academic pursuits.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

I declare that I used AI tools to assist with translating specific sentences from my native language to English and to help rephrase my work in a style more suitable for academic work. The list of the specific tools I have used is in Appendix A.

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských prací.

Prohlašuji, že jsem použila AI nástroje k pomoci s překladem konkrétních vět z mého rodného jazyka do angličtiny a k přeformulování mé práce do stylu vhodnějšího pro akademickou práci. Seznam použitých nástrojů je uveden v příloze A.

V Praze dne 24. května 2024

.....
Daria Mikhaylovskaya

Abstract

This study contributes to the field of humanoid robotics by creating a gaze control system for the iCub robot. The system is designed to be used during an interactive card game and aims to achieve interactions that are both naturalistic and safe. The study uses the robot's vision system to replicate human eye activity by concentrating on the participant's eyes and game objects alternately.

The experiments involving human participants identified that the accuracy and consistency of gaze, the frequency of blinking, and the fluidity of eye movements are crucial factors in enhancing the perception of a robot's gaze as natural.

We worked on the development of the eye controller that is able to track humans in real-time. Additionally, we added features like blinking, small amplitude random body motions, and mouth movements that contribute to the naturalistic behavior of the robot.

Keywords: Humanoid Robotics, Human-Robot Interaction (HRI), Gaze Control, Naturalistic Eye Movements, Keypoint detection, iCub Robot, Safety in Human-Robot Interaction, Godspeed questionnaire, Real-Time Gaze Adjustment

Supervisor: Mgr. Matěj Hoffmann, Ph.D.

Abstrakt

Tato studie přispívá do oblasti humanoidní robotiky vytvořením systému ovládní očí pro robota iCub. Systém je navržen pro použití během interaktivní karetní hry a jeho cílem je dosáhnout interakcí, které jsou jak přirozené, tak bezpečné. Studie využívá zrakový systém robota k replikaci lidské oční aktivity tím, že střídavě zaměřuje pozornost na oči účastníka a herní objekty.

Experimenty s lidskými účastníky ukázaly, že přesnost a konzistence pohledu, frekvence mrkání a plynulost pohybů očí jsou klíčovými faktory při zlepšování vnímání robotického pohledu jako přirozeného.

Pracovali jsme na vývoji očního kontroléru, který je schopen sledovat lidi v reálném čase. Kromě toho jsme přidali funkce jako mrkání, malé náhodné pohyby těla a pohyby úst, které přispívají k chování robota, aby bylo vnímáno jako přirozené.

Klíčová slova: Humanoidní robotika, Interakce mezi člověkem a robotem (HRI), Ovládní pohledu, Naturalistické pohyby očí, Detekce klíčových bodů na těle, Robot iCub, Bezpečnost v interakci mezi člověkem a robotem, Dotazník Godspeed, Úprava pohledu v reálném čase

Překlad názvu: Přirozené ovládní očí humanoidního robota během interaktivní hry

Contents

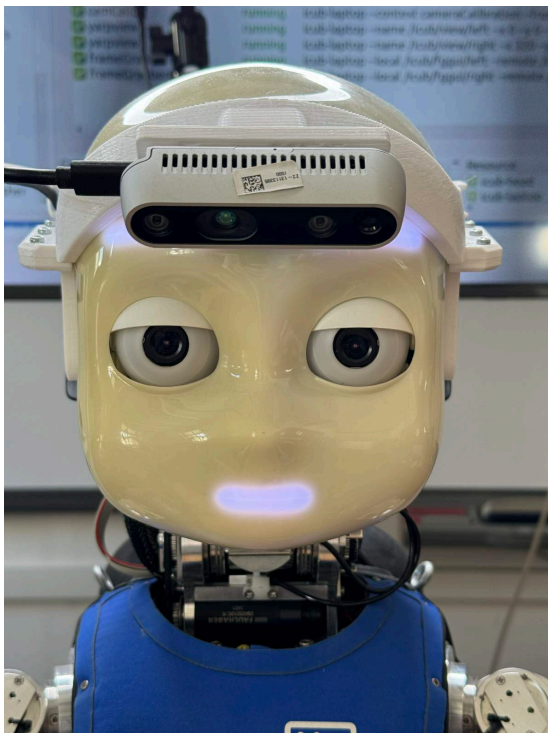
1 Introduction	1	3.5.2 Estimating Distance to Human Body Parts	17
1.1 Motivation	1	3.5.3 Connection with Eyes Controller	18
1.2 Goals	2	3.6 Experiment Script Overview . . .	19
1.3 Thesis Structure	3	3.7 Questionnaires for experiment participants	21
2 Related work	5	4 Experiments and Results	23
2.1 Related Work	5	4.1 Distance Estimation Experiment	24
2.1.1 Gaze in Human-Robot Interaction	5	4.2 Human Tracking Capability	25
2.1.2 Human detection methods . . .	6	4.3 Pilot Experimental Social Study	27
2.1.3 Psychological aspects in HRI .	6	4.3.1 Scenario	27
2.2 Thesis Contribution	7	4.3.2 Gaze strategies	28
3 Materials and Methods	9	4.3.3 Evaluation	29
3.1 Experimental Setup	9	5 Discussion, Conclusion and Future Work	37
3.2 Software and Programming	9	5.1 Conclusion	37
3.3 Eyes control	10	5.2 Discussion	37
3.4 Eyes Controller Development . . .	11	5.3 Future work	38
3.4.1 Initialization	11	Bibliography	39
3.4.2 Runtime Operation	12	A List of AI tools used in the work	43
3.4.3 Conversion from Pixel to Joint Coordinates	13	B The Godspeed questionnaire used after the each interaction with robot	45
3.4.4 Smoothing Eye Movements . .	14	C The Final questionnaire used after the all three interactions with robot	53
3.4.5 Blinking	16	D Pilot social study scenario of interaction with robot	59
3.5 Keypoint Detection for Interaction	17		
3.5.1 Keypoint detection	17		

Chapter 1

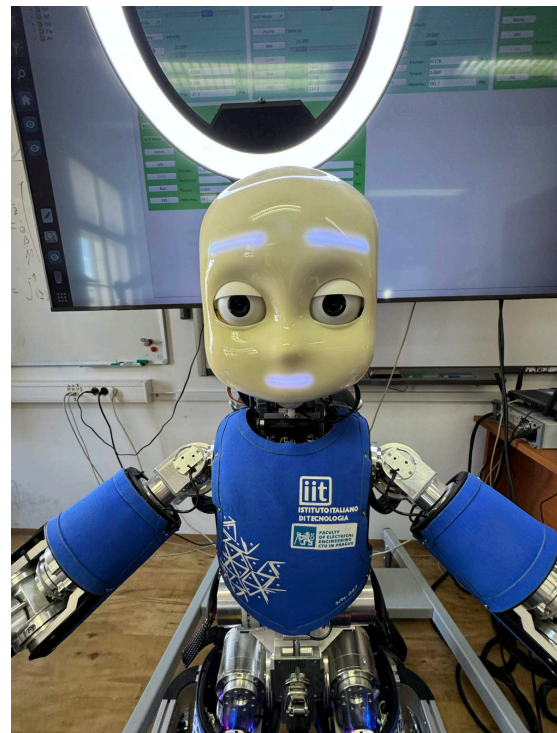
Introduction

1.1 Motivation

The growing presence of humanoid robots in everyday environments highlights the need of ensuring that their interactions are seen as both safe and pleasant by humans. As technology advances, there is a growing necessity to efficiently upgrade existing robots or their parts. Rather than creating new robots, a more practical strategy is to upgrade current models with additional sensors (see Fig. 1.1). However, this raises concerns regarding whether these robots will continue to be viewed as safe and approachable and how these changes may impact the authenticity of their interactions.



(a)



(b)

Figure 1.1: The iCub robot with (a) and without (b) any additional sensors.

Integrating modern sensors improves functionality, but also brings difficulties in preserving

the humanoid’s authentic appearance and behavior. Ideally, robots would have fully integrated vision systems from the start, but this is often not the case. Therefore, our approach is to improve detection by adding an extra camera (see Fig. 1.2) and optimizing the current visual systems. The main goal of this thesis is to minimize the disruption caused by new technologies on human-robot interaction.



Figure 1.2: Additional RGB-D camera installed on the iCub’s head.

1.2 Goals

This work focuses on developing an advanced eyes controller for the iCub humanoid robot. The goal is to improve the robot’s interactions with humans during social and interactive games, making them more naturalistic and safe. An important feature is the detection of the human hand and the estimation of the distance between the hand and the robot. Thanks to this feature, the gaze controller can adapt its strategies in real time, taking into account proximity and human actions. Through targeted social experiments, we plan to evaluate several different gaze strategies to see whether they are perceived as natural and human-like in interactions, effectively communicating with people in a comfortable and engaging manner.

■ 1.3 Thesis Structure

The thesis is structured as follows. Chapter 1 introduces the area of humanoid robotics and emphasizes the importance of developing a natural and safe human-robot interaction. Chapter 2 reviews the existing research, organizes different concepts, and places this work within the field. In Chapter 3, the materials and methods are explained, providing information on the setup, hardware, and development of the gaze control system. In Chapter 4, the experiments conducted and their results are presented, offering an analysis of the effectiveness of the implemented gaze strategies. Each chapter builds on the previous one, ensuring a logical progression of the research. Chapter 5 discusses and concludes the work done.

Chapter 2

Related work

2.1 Related Work

This section provides a review of the literature discussing human-robot interaction (HRI), specifically related to possible gaze control strategies, the psychological implications of human-robot interactions, and safety considerations.

2.1.1 Gaze in Human-Robot Interaction

The study of gaze behavior on robots has been extensively investigated, and early research has shown the essential role of gaze in promoting natural and realistic interactions between humans and robots. The research conducted by Argyle and Cook [1] on the psychology of interpersonal interaction established the foundation to understand the impact of gaze behavior on the effectiveness of communication, which has been a fundamental aspect in the field of Human-Robot Interaction (HRI). Breazeal [2] applied this psychological framework to robotics, investigating the effects of gaze and other non-verbal signals on the development of socially engaged robots.

Researchers such as Mutlu et al. [3] greatly improved the complexity of gaze processes. They studied how gaze signals may be used by social robots to indicate intentions, leading to increased human comfort and engagement. This research emphasized the importance of including gaze signals that mimic human interactions in order to enhance the robot's social presence and effectiveness in communicating.

Roncone et al. [4] presented a Cartesian 6-DoF gaze controller for humanoid robots. The architecture of the system enables accurate manipulation of both neck and eye motions, combining functionalities such as gaze stabilization and quick saccadic movements to accurately track a three-dimensional fixation point. The saccadic movements play an essential part, allowing the robot to imitate the rapid eye movements observed in human gaze behavior. This feature greatly improves the lifelike appearance and the ability to interact of humanoid robots.

Alshakhs et al. [5] have made recent progress by introducing a new algebraic inverse kinematics method to improve the control of gaze in humanoid robots. They achieved this using the cascading structure of the neck and eye motions. Mishra and Skantze [6] proposed a planning-based framework to automate gaze behavior in social robots. Their work suggests a future where gaze control is not just reactive, but also seamlessly connected

It enables better communication and coordination in dyadic interactions by mimicking human gaze behavior.

Additionally, Stanton and Stevens [19] conducted a study to examine how a humanoid robot’s gaze affects trust in a cooperative visual task. The results showed that acceptable gaze behaviors had a substantial positive influence on participants’ confidence in robots. Haefflinger et al. [20] studied the impact of independent control of head and eye movements on the naturalness of gaze.

Furthermore, Koller et al. [21] examined the effectiveness of gaze aversion and its influence on user experience in conversational settings, specifically focusing on the effects of nonhuman-inspired gaze timings. Shintani et al. [22] conducted a study of gaze behaviors based on roles in multi-party conversations, which improved comprehension of gaze aversions and their influence on turn-taking and participation.

■ 2.2 Thesis Contribution

This thesis presents the development of an eyes controller for the iCub humanoid robot. The source code for this project is available for public access on GitLab [23]. Additionally, demonstration videos showcasing the functionality of the gaze controller can be found at the following link [24]. The controller’s main objective is to combine two main gaze functions: to look at the desired location and to be perceived as naturalistic during the communication with humans. We integrated an eyes controller that uses state-of-the-art facial keypoint detectors to identify human faces or hands. It obtains pixel coordinates of the gaze target, then transforms these into joint coordinates to perform gaze movements. In addition to the eyes movement, we prepared the blinking feature for the iCub robot.

We implemented a method to estimate the distance between the parts of the human body and the camera. This allows the robot to continuously track the positions of human hands, even when they are not clearly visible to the RGB-D camera. This feature can improve the safety and smoothness of interactions, particularly in close proximity, where the entire human body is not visible.

Finally, we evaluated several different gaze strategies for a human-robot interaction through a pilot experimental study. For this goal, we developed an application that combines the movements of the eyes based on human tracking with the blinking, mouth movements and the ready-made module for the slight random body movements. The experimental study was conducted online and offline. The results have shown that the smart gaze strategies, where the gaze contact is established, are more comfortable for the participants than the random gaze strategy.

Chapter 3

Materials and Methods

This section provides an overview of the technical aspects and methodologies used in this thesis. The experimental setup is described in Section 3.1 and the specific technologies employed for creating the gaze control system in Section 3.2. The architecture of the eyes controller is described in Section 3.4. The keypoint detection used for the controller is described in Section 3.5. The social study script is described in Section 3.6 and evaluation of human-robot interaction scenarios in Section 3.7.

3.1 Experimental Setup

The iCub humanoid robot [25], which is central to our research, is a highly sophisticated platform specifically built for conducting studies on cognitive development and human-robot interaction (HRI). The iCub, with its human-like appearance and child-like proportions, is designed for engaging in naturalistic real-time interactions. The robot contains more than 50 degrees of freedom, including articulated hands, arms, legs, and a head capable of complex and synchronized motions that mimic human gestures (see Fig. 3.1).

The iCub is equipped with PointGrey Dragonfly 2 cameras in each of its eyes with a default resolution of 320 x 240 pixels and the framerate 30 fps. They can also run in the mode with a double resolution (640 x 480 pixels). The higher resolution was used for our experiments to make human keypoint detection more reliable. The motors in iCub's eyes enable precise control of pan, tilt, and vergence motions to move the gaze over a broad range of directions (see Fig. 3.2). The vergence mimics stereo focusing by moving the pan joints of the eyes in the opposite direction. As we do not employ stereo vision in our work, we only use the other two movements for the eyes—pan and tilt. This simplifies the setup and helps to avoid any strange-looking eye movements.

3.2 Software and Programming

YARP (Yet Another Robot Platform) is a middleware used to manage communication between the iCub robot and various software applications. It manages the integration of sensory data and motor commands, ensuring that the robot's visual capabilities and other functions are synchronized effectively (see Fig. 3.3). Programming for the robot's operations is primarily done in C++ for performance-critical tasks, while Python is used for fast development of target generation programs and easy incorporation of human detection networks.

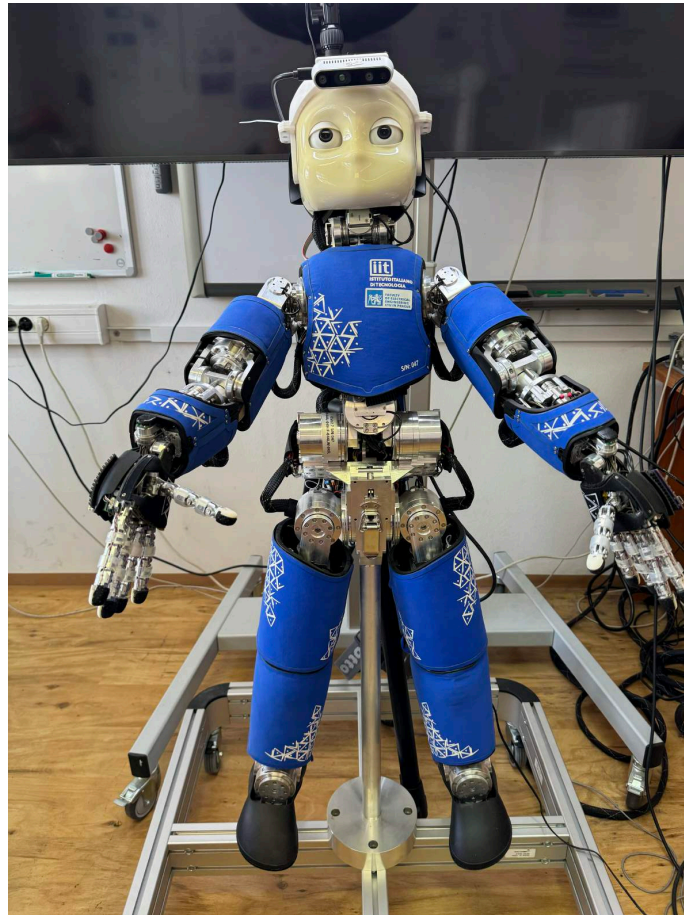


Figure 3.1: The iCub humanoid robot used in the experiments.

■ 3.3 Eyes control

The eyes control process starts with the iCub’s visual system collecting camera frames in real time. The frames are immediately transferred through a YARP port to a Python-based control system. The images are reviewed, and the robot’s next movements are decided based on the visual information gathered.

Once a decision is made, the specific motions that need to be taken are transmitted to a C++ application using a separate YARP port. The program is responsible for computing the motions required for the actions. The process begins by obtaining the current state of the robot, including the joint positions retrieved from the motor encoders. Then, based on the current state of the robot, the program calculates joint coordinates for the next eye movement.

The last stage of the process involves sending the movement instructions back to the robot. The commands, which precisely define the angles for the motors, are immediately transferred to the robot’s motors by YARP ports. After receiving these instructions, the

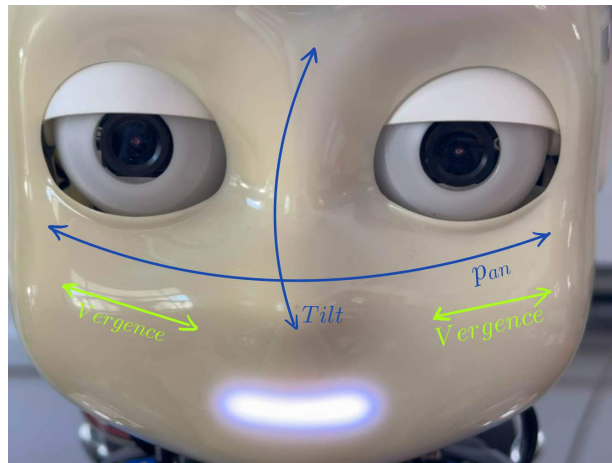


Figure 3.2: Illustration of the iCub’s vision system detailing the camera positions and eyes movements.

robot performs them. The entire process is described in Fig. 3.9. In the following, we describe the individual parts in more detail.

3.4 Eyes Controller Development

For the actual control of the eyes, we implemented an eyes controller. The robot’s gaze direction is determined by an external input such as analyzing real-time camera frames. The controller uses a state machine that manages the transition between several operating modes—waiting, reaching, and idle.

3.4.1 Initialization

The initialization of the eyes controller starts in the function that prepares the necessary system components and communication channels for operation. To handle inputs and outputs, we use YARP ports.

At first, the eyes controller initializes the system and prepares it for dynamic gaze tracking. It retrieves physical limits for eye joints (see Table 3.1) to ensure safe movements, creates instances to manage the neck and eyes, and releases any restrictive joints in the torso and head. In addition, it sets up data structures to handle joint positions.

Joint	Range (degrees)
Eyes Tilt	-29 to 29
Eyes Version (Pan)	-29 to 29
Eyes Vergence	0 to 44
Eyelids	-4 to 64

Table 3.1: Joint limits for the eyes.

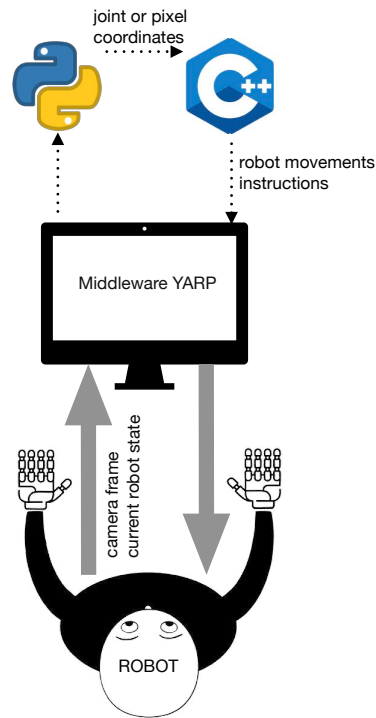


Figure 3.3: Flow diagram illustrating the communication and data processing pathways involved in programming the iCub robot movements using YARP.

The initialization process also includes preparing the drivers that control and communicate with the robot’s hardware, in particular the robot’s face, head, and torso. The joint limits for the face and head are retrieved, defining the operational range for eye and eyelid movements to ensure that they remain within mechanically safe values.

■ 3.4.2 Runtime Operation

The core operating logic of the eyes controller is incorporated within the `run()` function, which is executed continually in a loop as part of the thread’s lifetime. At the beginning of every execution cycle, the `run()` function checks for new input data from the YARP communication port. There are three types of the input data:

- **Blink Command:** If a single data item is received, it triggers a blinking action. The controller sets a flag (`must_blink`) to initiate the eyelid movement.
- **Joint Coordinates:** If two data items are received, they are interpreted as the new target joint coordinates for the eyes. This updates the target position and changes the state to `STATE_REACH`, indicating that the eyes should move to reach the new target.
- **Pixel Coordinates:** If three data items are received, we detect it as pixel coordinates. These pixel coordinates consist of two numbers for the x and y positions, with an

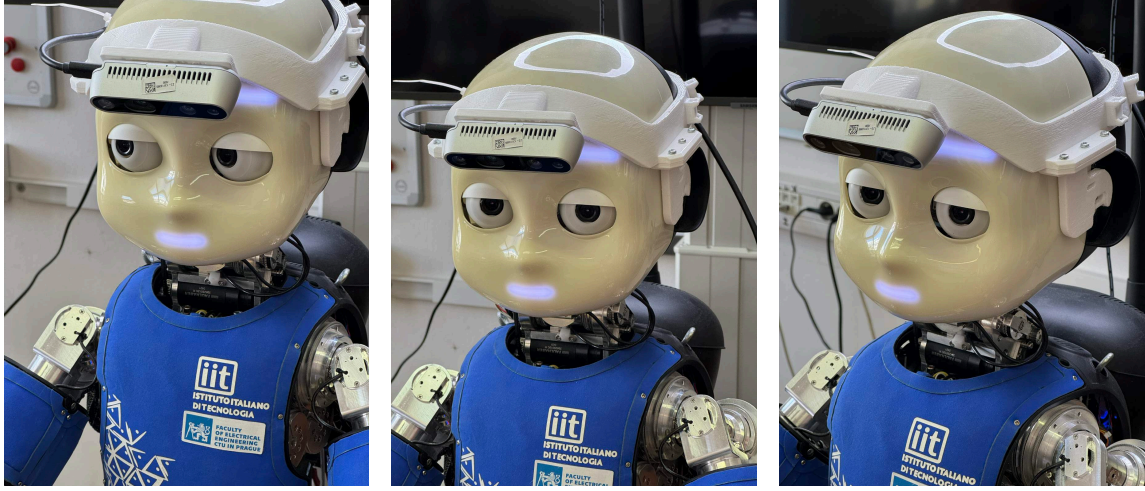


Figure 3.4: iCub robot eye movements.

additional 0 added for design purposes to easily distinguish them from blinking signals or joint coordinates received through the YARP port. These are processed to convert them into joint coordinates using the `setNewTarget()` function, which also updates the target position.

Depending on the input and current conditions, the function then updates the state of the controller.

3.4.3 Conversion from Pixel to Joint Coordinates

The camera calibration matrix \mathbf{K} is defined with focal lengths f_x and f_y , and principal point coordinates c_x and c_y . For a camera with resolution parameters, the matrix is given by:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.1)$$

where f_x , f_y are the focal lengths along the x and y axes, and c_x , c_y represent the optical center of the camera.

Given pixel coordinates (x_p, y_p) , the displacement from the center of the image frame (x_c, y_c) is calculated as:

$$\Delta x = x_c - x_p, \quad \Delta y = y_c - y_p \quad (3.2)$$

Using the inverse of the camera calibration matrix \mathbf{K}^{-1} , the displacement vectors are transformed to obtain normalized coordinates in the camera frame. z coordinate is expected to be 1, since the tracking object is expected to be 1 meter away from the robot:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K}^{-1} \cdot \begin{bmatrix} \Delta x \\ \Delta y \\ 1 \end{bmatrix} \quad (3.3)$$

The angular displacement required for the robot's eyes to align with the target is computed using the arctangent function $\text{atan2}()$:

$$\theta_p = \text{atan2}(y, z), \quad \theta_t = \text{atan2}(x, z) \quad (3.4)$$

where θ_p and θ_t are the angles for pan and tilt movements respectively (see Fig. 3.5). To set the same direction of pan movement as in the iCub robot, we will use $-\theta_p$ instead of θ_p .

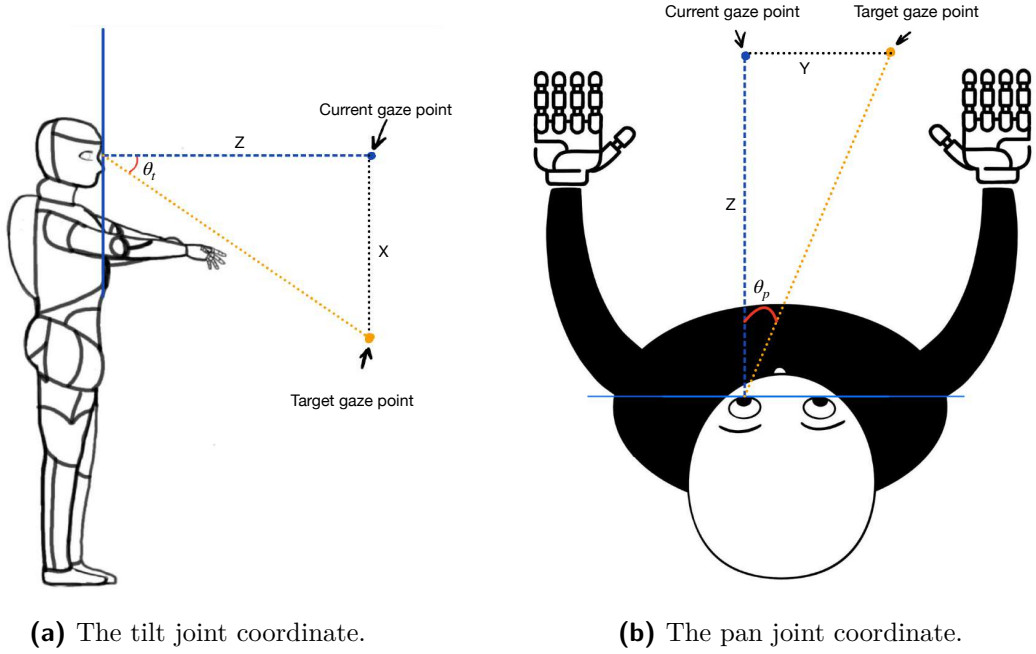


Figure 3.5: The illustration of tilt and pan joint coordinate computation.

The angular displacements obtained in radians are converted to degrees to match the input of the robot motor systems. After converting to degrees, these angles are further adjusted to fit within the mechanical limits of the robot's eye joints:

$$\theta_p = \max(\min(\theta_p, x_u), x_l) \quad (3.5)$$

$$\theta_t = \max(\min(\theta_t, y_u), y_l) \quad (3.6)$$

Here, x_l, x_u, y_l, y_u denote the maximum and minimum degrees the joints can safely rotate to achieve the desired eye movement, ensuring that the movements stay within safe operational bounds. Once the angles within the joint limits have been adjusted and verified, the robot's control system is configured with the new target absolute positions.

3.4.4 Smoothing Eye Movements

Once the new target joint positions are calculated, the trajectory has to be sampled to minimize sudden or jerky movements that may occur when the eyes attempt to adjust the focus immediately to the target (see Fig. 3.4). In this work, we use uniform sampling. The procedure for computing the incremental step is as follows:

1. It first retrieves the current angular positions of the eyes, which are represented by the pan and tilt joint coordinates, respectively.
2. The function then calculates the difference between the new target joint coordinates and the initial angles. This difference represents the total angular distance that needs to be covered.
3. A `timestep` of 0.02 seconds is defined according to the camera frequency of 30 fps to keep the human detection precise, along with a total time of 1.5 seconds to complete the movement, providing smooth eye movements.
4. The step increments are calculated by dividing the total difference by the number of steps derived from the `time` and `timestep`.

$$\text{step} = \left(\frac{\text{diff} \cdot \text{timestep}}{\text{time}} \right)$$

The `target_changed` flag is set to `false`, indicating that the target has been processed and the steps to reach it are set. Then, the current position of each eye joint (pan and tilt) is checked against its respective target position. If the absolute difference between the current and target position for any axis exceeds a threshold (1 degree in this case), it indicates that further adjustment is needed.

The function then incrementally adjusts the eye position by adding a predefined step size (`step[0]` for pan, `step[1]` for tilt) to the current position.

After each adjustment in the periodic loop, the function checks if the eye positions have reached the target within a threshold of 1 degree. If both axes have reached the position, the function sets the state of the controller to `STATE_IDLE`, indicating that no further adjustments are needed, and the target has been successfully acquired. By smoothing the transition between gaze points, the function helps prevent sudden eye movements, which can be perceived as unnatural by human observers. For further details of the implementation see Algorithm 1.

Algorithm 1 Eye Position Adjustment and State Update

```

1: procedure ADJUSTEYEPOSITION
2:   while true do ▷ Periodic loop
3:     Read target pixel coordinates
4:     if Received three data items then
5:       if data[2] == 0 then ▷ Pixel coordinates
6:         Convert pixel coordinates to joint coordinates using setNewTarget()
7:         Update the target position
8:       else if data[2] == 1 then ▷ Blink signal
9:         Execute blinking routine
10:      else ▷ Joint coordinates
11:        Update joint target directly
12:      end if
13:    end if
14:    Compute the difference between current and target joint positions
15:    Apply adjustment step to move eyes closer to target position
16:    if  $\text{abs}(\text{current\_position}[0] - \text{target\_position}[0]) \leq 1$  AND  $\text{abs}(\text{current\_position}[1]$ 
-  $\text{target\_position}[1]) \leq 1$  then
17:      Set controller state to STATE_IDLE
18:    end if
19:  end while
20: end procedure

```

3.4.5 Blinking

We additionally implemented the blinking movement for the robot. The algorithm (see Algorithm 2) smoothly adjusts the eyelid positions in small, incremental steps from open to closed and back to open, ensuring natural and realistic blinking within safe joint limits.

Algorithm 2 Perform Blinking Movement

```

1: function PERFORM_BLINKING_MOVEMENT
2:   Initialize joint positions array for blinking movement
3:   Calculate the entire array of absolute joint positions for closing and opening eyelid
4:   for each joint position in the array do
5:     Move eyelid to the current joint position
6:     Check if current joint position is within the retrieved joint limits
7:     if within limits then
8:       Continue to the next joint position
9:     else
10:      Adjust joint position to be within safe range
11:    end if
12:  end for
13:  Blinking movement performed successfully
14: end function

```

3.5 Keypoint Detection for Interaction

Keypoint detection is essential for enabling the robot to track and respond to human movements, maintaining eye contact, following hand gestures, and enhancing natural human-robot interactions.

3.5.1 Keypoint detection

For human detection, we use the MediaPipe Holistic [26] library. This library combines face, body, and hand landmark detection using three different ML models. The output of the algorithm is an array with the pixel coordinates of body keypoints. These keypoints include:

- 33 pose landmarks (see Fig. 3.6) from the body, such as elbows and knees, which is done using a convolutional neural network similar to MobileNetV2 [27] and is optimized for on-device, real-time fitness applications. This variant of the BlazePose [28] model uses GHUM [29], a 3D human shape modeling pipeline, to estimate whole three-dimensional body position of an individual in images or videos [30]
- 468 face landmarks [8], which cover features from the forehead to the chin. It [31] employs a lightweight feature extraction network similar to MobileNetV1/V2 [32] and Single Shot MultiBox Detector [33]
- 21 landmarks for each hand (see Fig. 3.7), showing details like finger joints, which also uses GHUM [29] to predict 3D coordinates based on the 2D point projections [34].

In total, MediaPipe can detect and track 543 landmarks in real time. It also provides the estimation of 3D coordinates of each of the keypoint which we will discuss in the following sections.

3.5.2 Estimating Distance to Human Body Parts

We noticed that while MediaPipe gives us 3D coordinates for body points, these often include incorrect distance measurements. Because of this, we decided to use other method to better estimate how far human hands are from the robot. We use the 2D pixel coordinates of keypoints identified by MediaPipe. Then we estimate the real distance to these points from the robot.

First, we calculate the distance between two points on the hand, between the wrist and the tip of the index finger, using simply using the Euclidean distance formula:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.7)$$

Where $(x_1, y_1), (x_2, y_2)$ are the pixel coordinates of the index finger and the wrist. We then use a pre-calibrated polynomial regression model [35] to convert pixel distance into real-world distance in centimeters.

$$d_r = A \cdot (d_p)^2 + B \cdot (d_p) + C \quad (3.8)$$

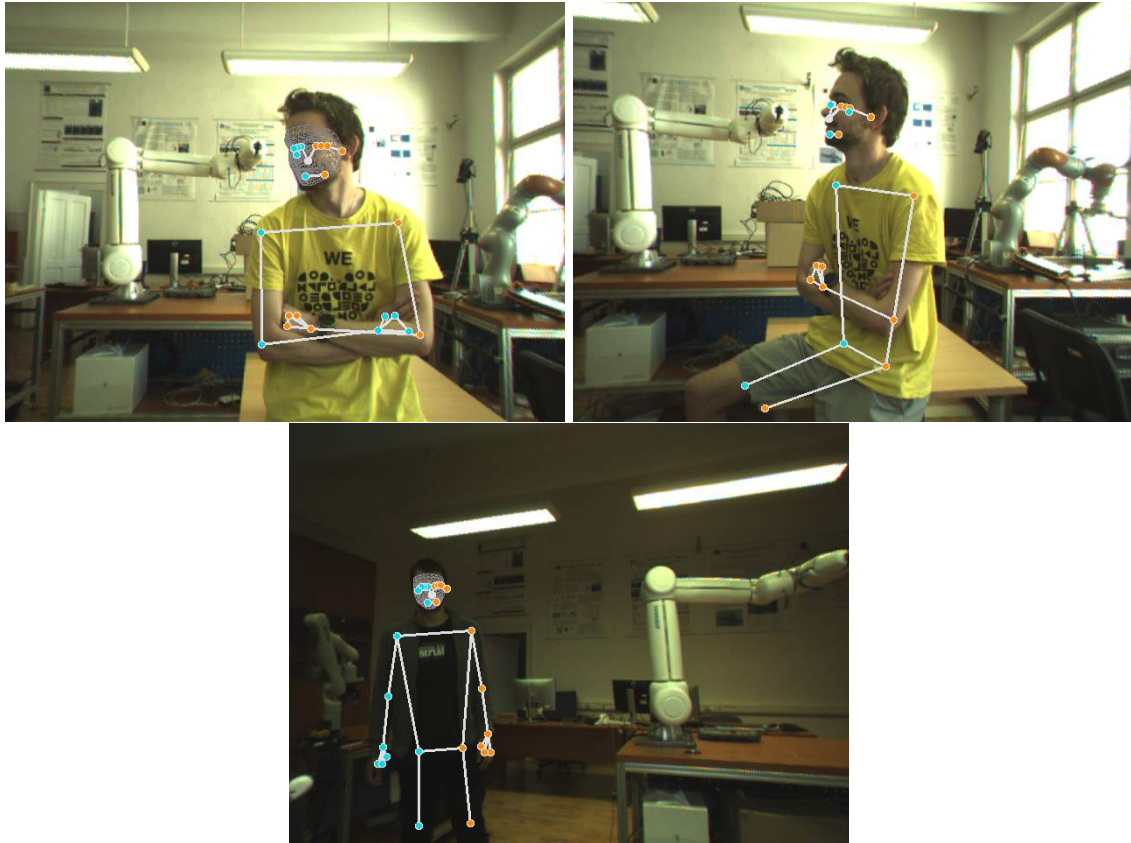


Figure 3.6: Pose and face landmarks used in the algorithm.

Here, d_r is the real-world distance, d_p is the pixel distance, and A , B , and C are the polynomial regression coefficients. These coefficients are obtained by fitting a polynomial to the data using the function `np.polyfit`. This function takes pixel distances \vec{x} and their corresponding real-world distances in centimeters \vec{y} and returns the coefficients of a second-degree polynomial that best fits this data.

This way, we get a more accurate measure of how far the hands are from the robot, which helps in making our robot’s interactions with people safer in the situations where human body parts are not detected in the RGB-D camera.

■ 3.5.3 Connection with Eyes Controller

To enable the robot to interact naturally with its environment, precise coordination between visual inputs and motor responses is needed. The coordinates of the desired keypoints are sent to the gaze controller through a YARP port, enabling the robot to adjust its gaze accordingly. Our gaze controller converts these pixel coordinates into joint coordinates. The robot then moves its eyes to the specific keypoints. For instance, to establish eye contact, the robot actually looks at the person’s nose tip.

When the robot cannot see a face but can see the body, it focuses on a point between the

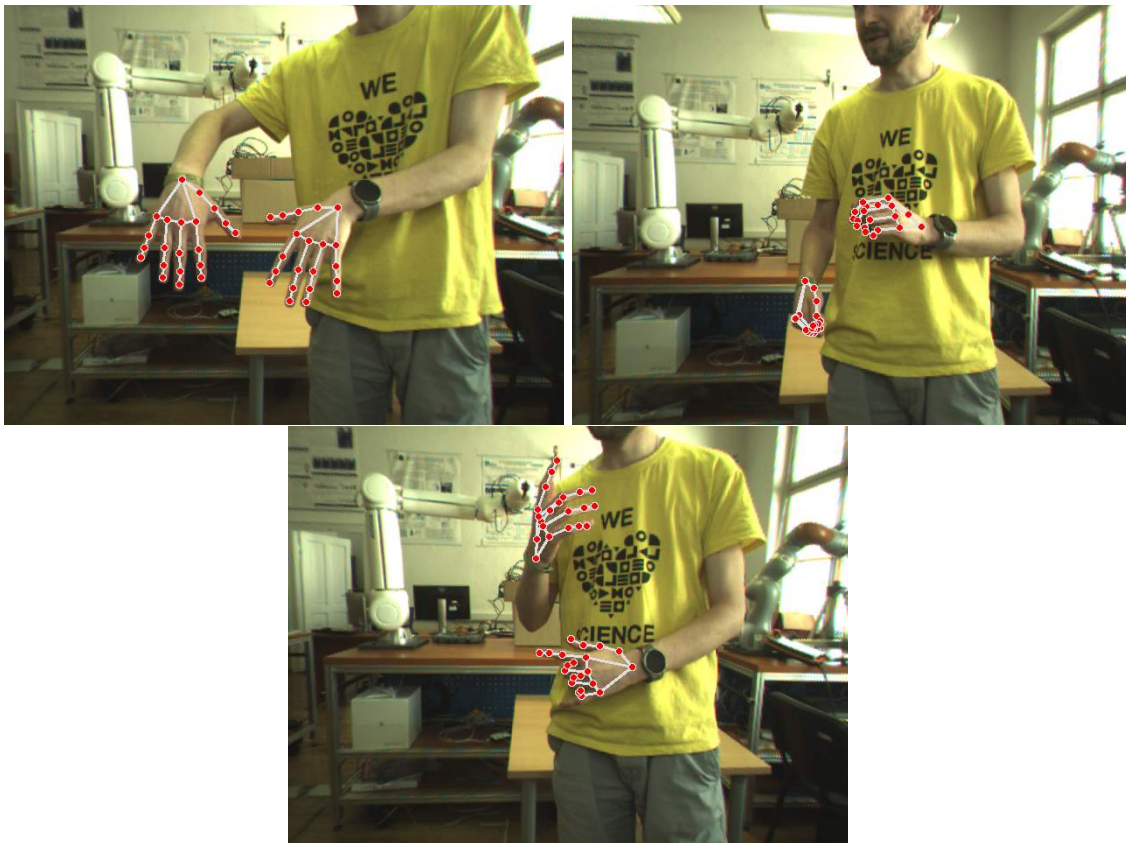


Figure 3.7: Hands landmarks used in the algorithm.

human shoulders. This helps the robot know where the person is facing or moving. If the robot needs to follow the hands but they are not clear, it uses landmarks from the whole body to find and follow the wrists. This way, even if the hands are not easy to see, the robot can still guess where they are and adapt the gaze by looking at the position of the arms.

3.6 Experiment Script Overview

For the social study experiment, we prepared a Python script, which manages the robot's interactions during experiments. It handles several different functions and sends commands to the eyes controller for further processing. The functionalities are separated into modules to make their adjustments and tuning easier. The communication between the script and the eyes controller is provided via YARP ports. The functionalities are:

- **Human detection:** the MediaPipe [26] library is used to detect and track human faces, hands, and body keypoints (see Section 3.5). The script gets pixel coordinates of the body parts and sends them to the eyes controller (see Section 3.4).
- **Blinking:** The script contains a function that sends the blink command to the controller. According to [15], the average blinking rate during the conversation must

be 23.3 b/min with inter-eye blink interval(IEBI): 2.3 ± 2 s to maintain naturalistic blinking behavior, 15% of blinks are double blinks. Blinks also must be at the onset and offset of the speech. Each blink should be divided into three phases with different speeds (attack, sustain, decay).

- Random body movements:** We incorporated an iCub Breather [36] module that produces small random torso movements to make the robot more natural during the interaction. The parameters were used as follows: `iCubBreather -part torso -autoStart -noiseStd 3 -refSpeeds 2`
- Eyes control:** The targets for the eyes are sent to the eyes controller in the format of pixel coordinates as in human tracking (mentioned above) or as desired joint coordinates. The desired coordinates can be either random (robot is looking around with no objective) or predefined to look at specific objects (e.g., the table).
- Speech:** For the purposes of the pilot study, we used a text-to-speech tool [37] to generate a sound that made it seem like the robot was talking. Moreover, we used Facial expressions module [38] to control the lights around the robot’s mouth to mimic opening (shocked face emotion) and closing mouth (neutral face emotion) during the speech (see Fig. 3.8). As the sound is played in a separate thread, we had to move the face expressions to a separate script which runs in parallel with the main one. The separation of scripts enables to let the mouth “moving” synchronously with the speech while other tasks, which can be computationally exhaustive, are happening—such as the human detection.

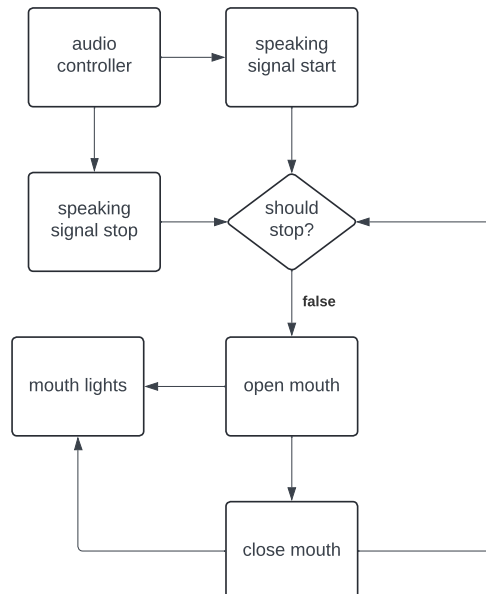


Figure 3.8: The diagram showing the mouth movements implementation.

■ 3.7 Questionnaires for experiment participants

For the pilot social study, we prepared two different questionnaires for each participant. As the aim of the study is to compare several gaze strategies during a human-robot interaction scenario, we used the Godspeed questionnaire [39] to evaluate the experience of the participant after each interaction session. The questionnaire can be found in Appendix B.

The second questionnaire (filled after final scenario) consists of a series of custom questions to gather detailed feedback on specific aspects of the robot’s behavior, such as eye movements, blinking, and mouth movements. These questions were aimed at identifying which behaviors were noticed and appreciated, and which were considered unnatural. We evaluated whether participants could find differences between interaction conditions and how the robot’s gaze influenced their sense of connection and engagement. A complete list of these additional questions is provided in Appendix C.

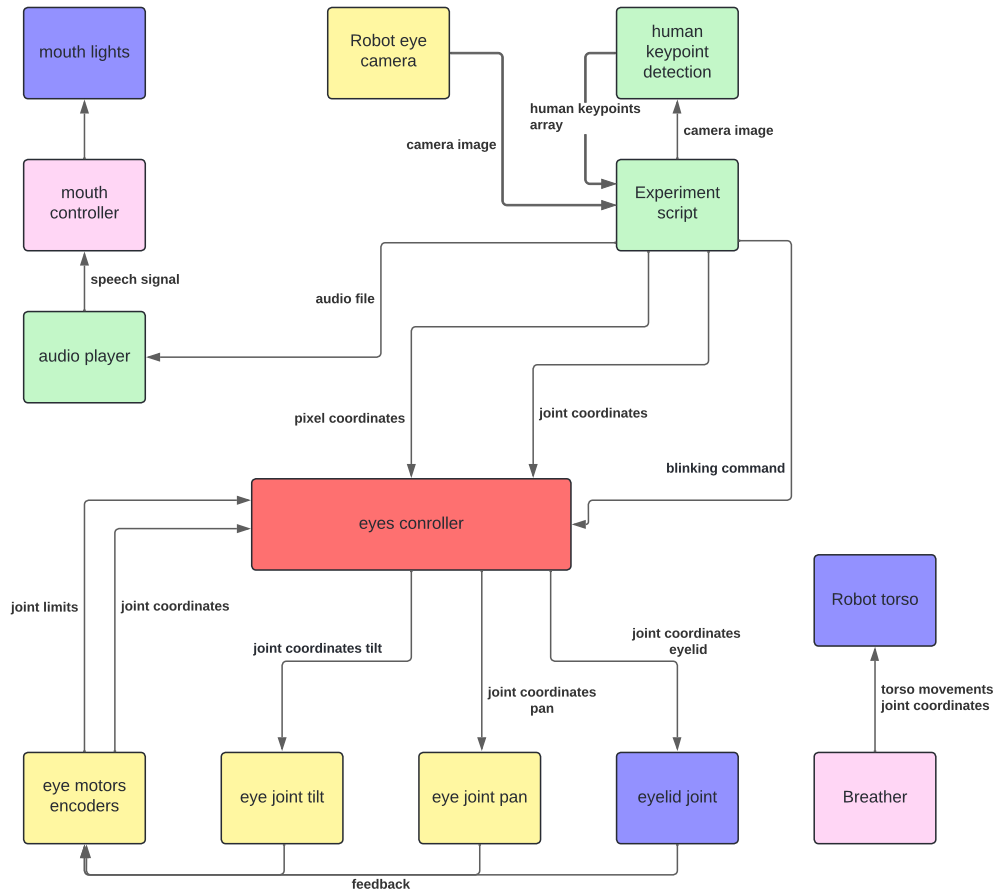


Figure 3.9: Diagram illustrating the overall system architecture. Yellow blocks represent the eye hardware, red block indicates the eye controller, violet blocks are for other robot hardware, green blocks depict the experiment script, and pink blocks represent other software controllers.

Chapter 4

Experiments and Results

We conducted three different types of experiments to evaluate the quality of the eyes controller. Distance Estimation Experiment described in 4.1, Human Tracking Capability described in 4.2 and Pilot Experimental Social Study described in 4.3

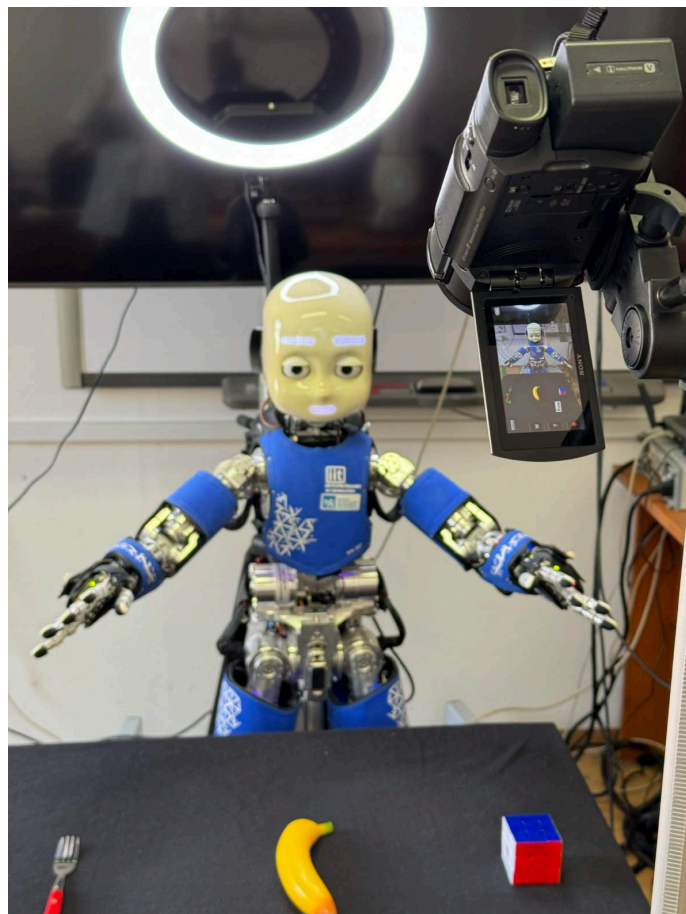


Figure 4.1: Social study experimental setup.

4.1 Distance Estimation Experiment

During this experiment, a participant moved their hand along a pre-set trajectory towards and away from the robot's eyes to evaluate the accuracy of our distance estimation method compared to MediaPipe's outputs. The hand was moved within a specific range of distances from 40 cm to 120 cm. While we did not have exact ground-truth data for the distance, we observed that MediaPipe's distance estimations did not reflect reality, often showing significant differences from the real estimations or random values (see Fig. 4.2). In contrast, our method provided more consistent and reliable estimations.

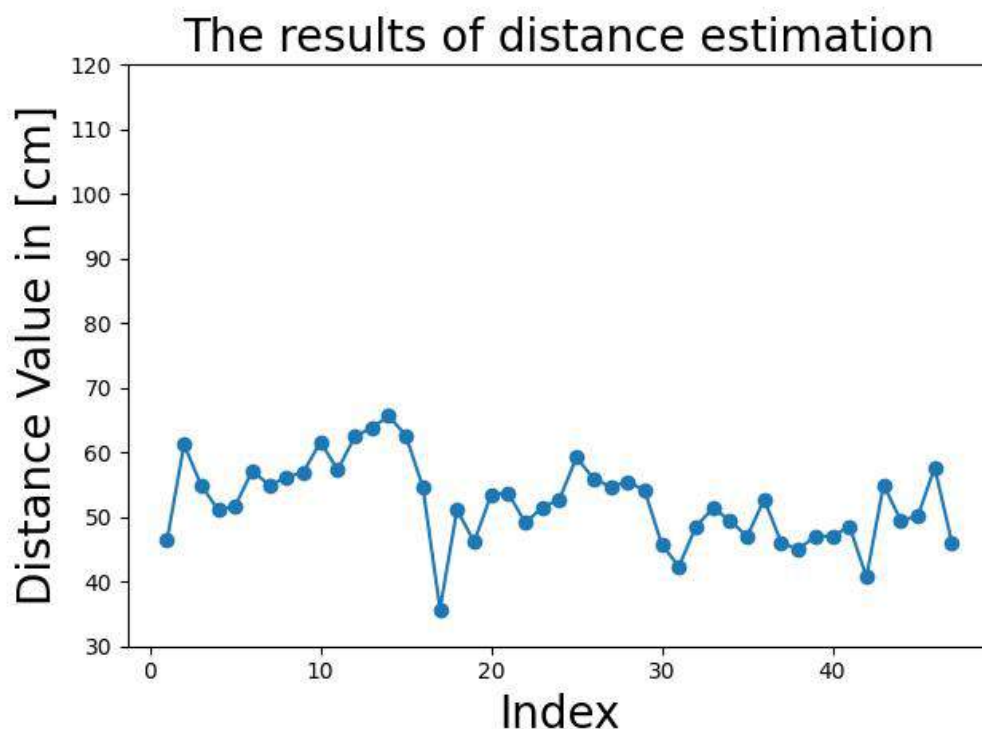


Figure 4.2: The results of distance estimation experiment with MediaPipe library.

Our method, while not perfect, provided better accuracy for distances further from the robot (up to 40 centimeters). For distances beyond this, our system did not give the exact distance but could still indicate that the human body part, in our case hand, was approaching. This level of accuracy is sufficient for our application. Exact distance measurement is less critical than recognizing proximity to the robot because at that moment we expect human body part to appear in the RGB-D camera image, where we can detect precise distance to the body parts.

The reason our method does not always return the correct exact estimation for close distances might be because MediaPipe keypoint detection, which is used in our method (see Section 3.5.2), cannot detect the whole hand when it is too close to the camera.

We further supported it with a graph that plots the estimated distance against the hand

area visible in the camera frame. This graph clearly showed that as the hand appears larger (moving closer to the robot), our method shows that the hand is indeed closer (see Fig. 4.3).

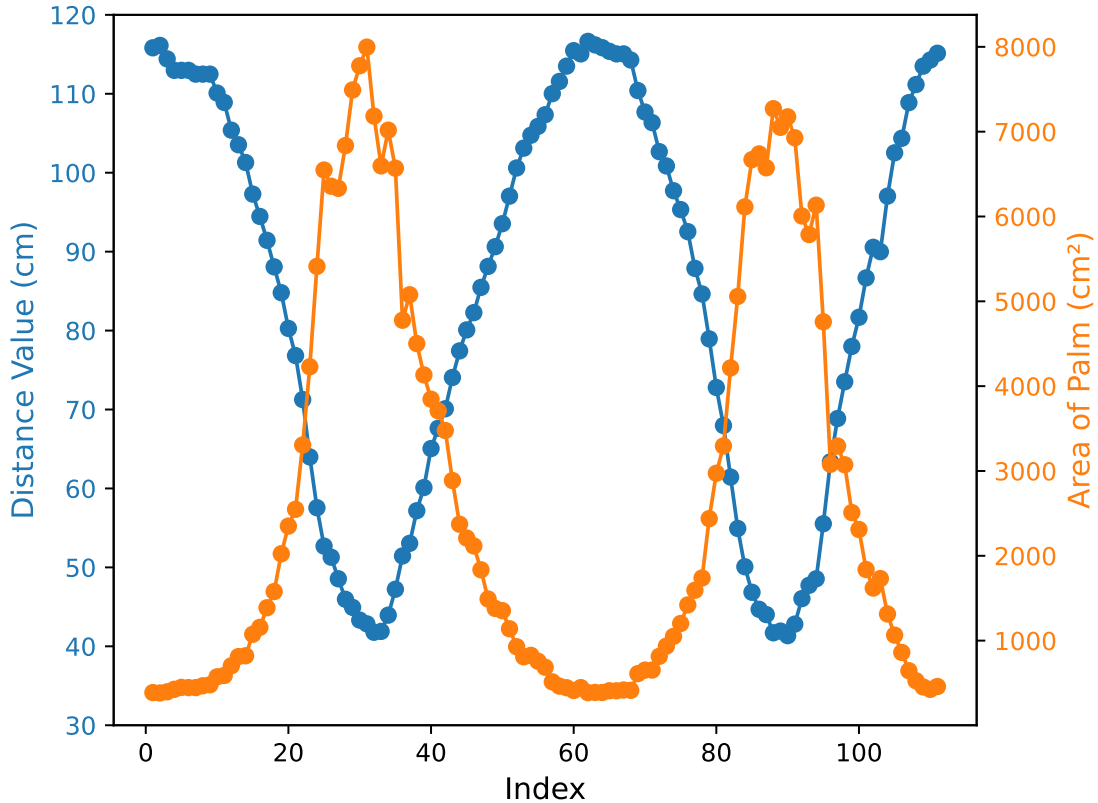


Figure 4.3: The results of distance estimation experiment with our method, showing the distance and area of palm on the current camera frame.

4.2 Human Tracking Capability

The second experiment demonstrated the robot's ability to track a human with its eyes (see Fig. 4.4). The participant was asked to walk in various directions in front of the robot in four different scenarios:

1. The robot followed the participant's head without any additional movements (see Fig. 4.5).
2. The robot followed the participant's head while the Breather module was active, adding small, random torso movements (see Fig. 4.6).
3. The robot followed the movements of the participant's hands without disruption (see Fig. 4.7).
4. Similar to the second scenario, but with the robot tracking the participant's hands while the Breather module was active (see Fig. 4.8).

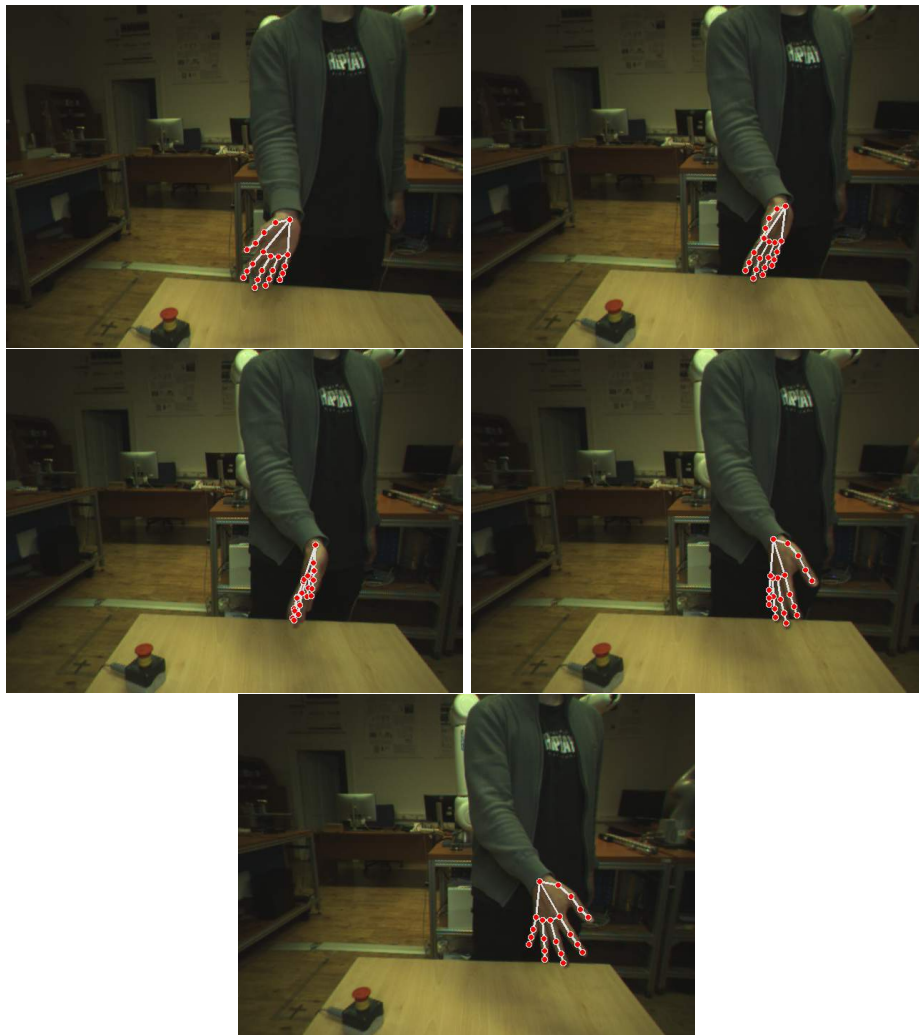


Figure 4.4: The hand tracking scenario with hand detections shown.

For each scenario, data regarding the robot's eye movements were logged. We analyzed these data and plotted graphs that illustrate how effectively the robot could maintain focus on the participant during the experiment.

The graphs show that the orange and blue curves are almost identical across all scenarios. This similarity indicates that the robot's eye movements are smooth and it accurately performs the planned movements. The green points, however, are more scattered and noisy. This noise is due to the instability in keypoint detection, which sometimes returns incorrect values for the human's position. When the graph shows a straight line, it means that the robot could not detect any human at that moment.

Despite the noise in the green points, the robot's movements generally follow the trends of the human's position. The robot's gaze tracking remains stable, demonstrating its capability to maintain focus on the participant even despite additional movements when Breather module activated.

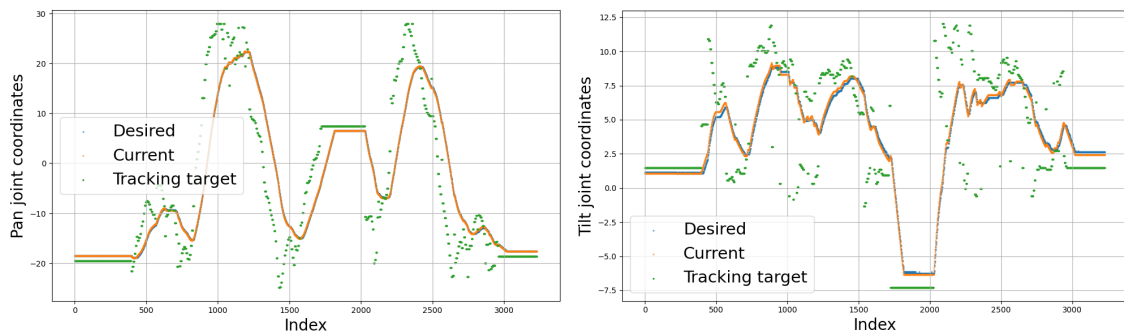


Figure 4.5: Scenario 1: Robot tracks human head without disruption. On these graphs orange points represent the current joint positions of the robot’s eyes, blue points show the desired joint positions for the robot’s eyes, green points indicate the coordinates of the tracking target, which in this case, is the human participant(head or hands).

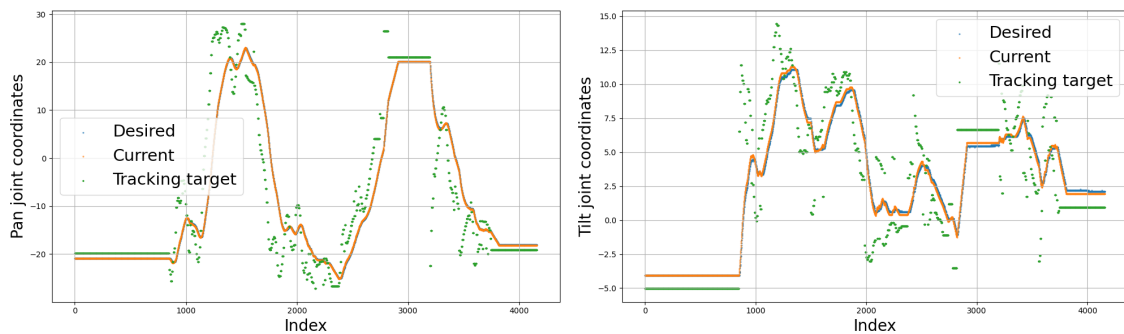


Figure 4.6: Scenario 2: Robot tracks human head with slight disruption.

4.3 Pilot Experimental Social Study

For evaluating our eyes controller in a human-robot interaction, we prepared an interaction scenario and conducted a pilot social study with several participants interacting with the real-robot (i.e., offline part as shown in Fig. 4.9) and several other participants observing the first-person video (i.e., online part as shown in Fig. 4.10).

The offline part of the experiment involved five male participants aged between 25 and 32 years, who rated their experience with robots from 2 to 5 on a scale where 5 indicates extensive experience. The group for the online experiment included four men and one woman aged between 20 and 45 years, most of whom had minimal experience with robots (rating 1).

4.3.1 Scenario

The experiment begins with the robot attempting to find the participant, who is standing at the start location a few meters away from the robot. If necessary, the robot performs random eye movements within the limited area until the participant is detected. The robot makes eye contact and greets the participant. The robot then uses his voice and eyes to direct the participant to approach the table in front of the robot. The robot verifies the

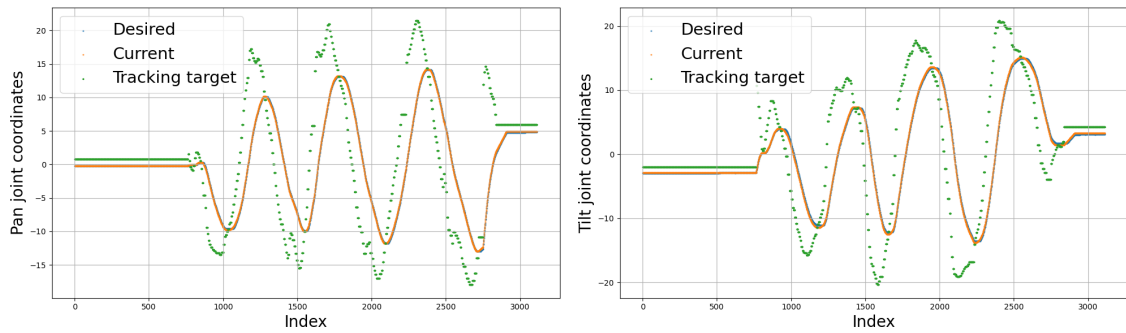


Figure 4.7: Scenario 3: Robot tracks human hand without disruption.

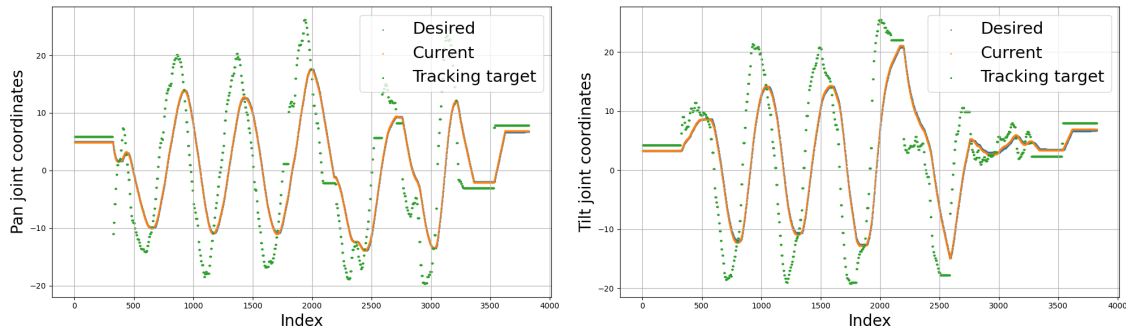


Figure 4.8: Scenario 4: Robot tracks human hand with slight disruption.

participant’s position using motor encoders and camera images, ensuring the participant is directly in front of him. If the pan-joint coordinate of the current eyes position is not in the center (or its small surroundings), but the human is detected on the current camera image, which means that the participant is not right in front of the robot, the robot asks the participant to adjust their position.

Once the human stands in front of the robot, the robot introduces the three objects on the table: a fork, a plastic banana model, and a Rubik’s cube from the YCB dataset [40] (see Fig. 4.11). After that, the robot invites the participant to play an interactive game, asking them to pick up and show the specific object. At the end of each session, the robot thanks the participant and asks for feedback.

The interaction detailed in the Appendix D is repeated three times with different gaze behaviors. After each interaction, the participants completed the Godspeed questionnaire (see Appendix B). After the last interaction, they answered our custom questionnaire comparing the scenarios (see Appendix C).

4.3.2 Gaze strategies

Three distinct gaze behaviors were tested:

- Random movement throughout the scenario, which will be marked on the graphs as Random.

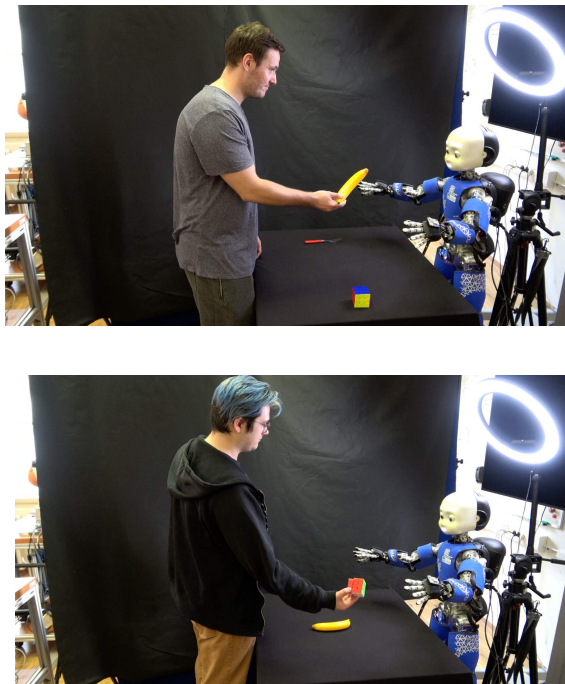


Figure 4.9: Offline part of the experiment.

- Smart alternating gaze based on interaction context, such as switching between eye contact, object focus, or hand following, which will be marked on the graphs as Alternatively Gazing.
- Consistent gaze following, focusing on the human face during introductions and the hand during interactive tasks, which will be marked on the graphs as Gaze on Human.

In all scenarios, the Breather module simulated subtle torso movements for realistic interaction. In addition, blinking was activated and mouth movements were imitated during the speech. There were blinks at the ends of each phrase, as well as additional random blinking during speaking and the rest of the interaction.

■ 4.3.3 Evaluation

According to the Friebe et al. [41] gaze behaviors perceived in a same way during the in-person interaction and virtual communication. This means that we can evaluate the results of the study not only separately, but also in a mixed conditions where data are analyzed across both online and offline experiments together.

The study involved 10 participants, 5 in an offline setting and 5 in an online setting. Each participant experienced three different interaction conditions, and after each condition, they completed a Godspeed questionnaire. This resulted in a total of 30 Godspeed questionnaire responses (15 from offline participants and 15 from online participants). Additionally, after

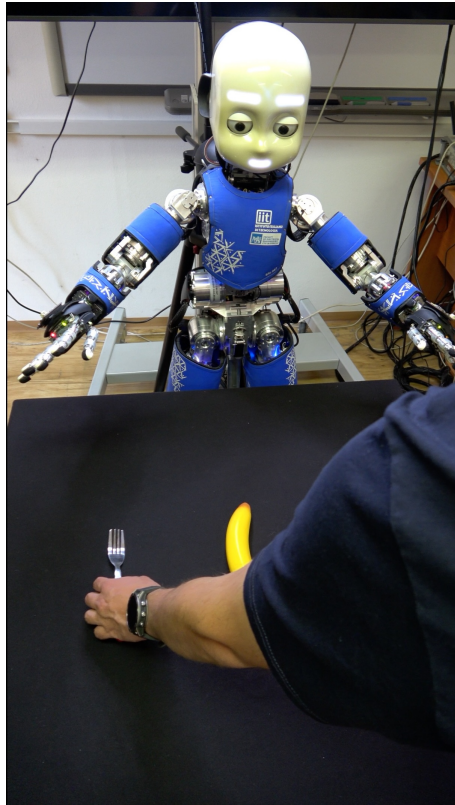


Figure 4.10: Online part of the experiment.

completing all three interaction conditions, participants filled out a final questionnaire, resulting in 10 final responses (5 from each setting).

We analyzed the data across all five Godspeed categories: Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety. Results were segmented by gaze behavior, and box plots were used for visual comparison (see Fig. 4.12). The smart alternating strategy was perceived as the most natural and effective (see Fig. 4.13). Compared to random movements, strategies that involved consistent eye contact and responsive gaze adjustments were rated higher for natural interaction and comfort (see Fig. 4.14).

According to Fig. 4.15, the participants found the small random body movements and blinking are the most unnatural behaviors. The blinking was particularly unnatural due to the strange sounds and slow movements, which were necessary because of hardware limitations. These issues are discussed further in the Discussion section (see 5.2). On the other hand, Fig. 4.16 shows that eye movements were ranked as the most likable and appropriate for the situation between all robot behaviors. The results demonstrate that the smart alternating gaze strategy, combined with robust eye and mouth movements, significantly enhances the naturalness and comfort of human-robot interactions (see Fig. 4.17).

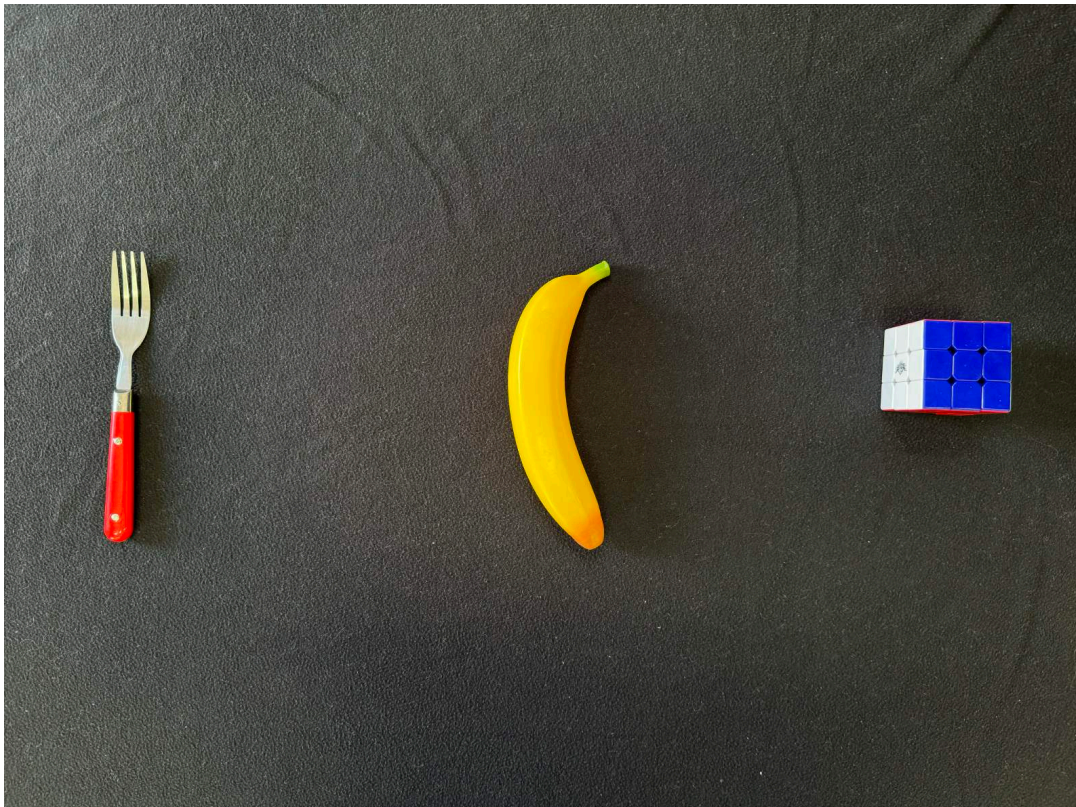


Figure 4.11: A fork, a plastic banana model, and a Rubik's cube from the YCB dataset used in the experiment.

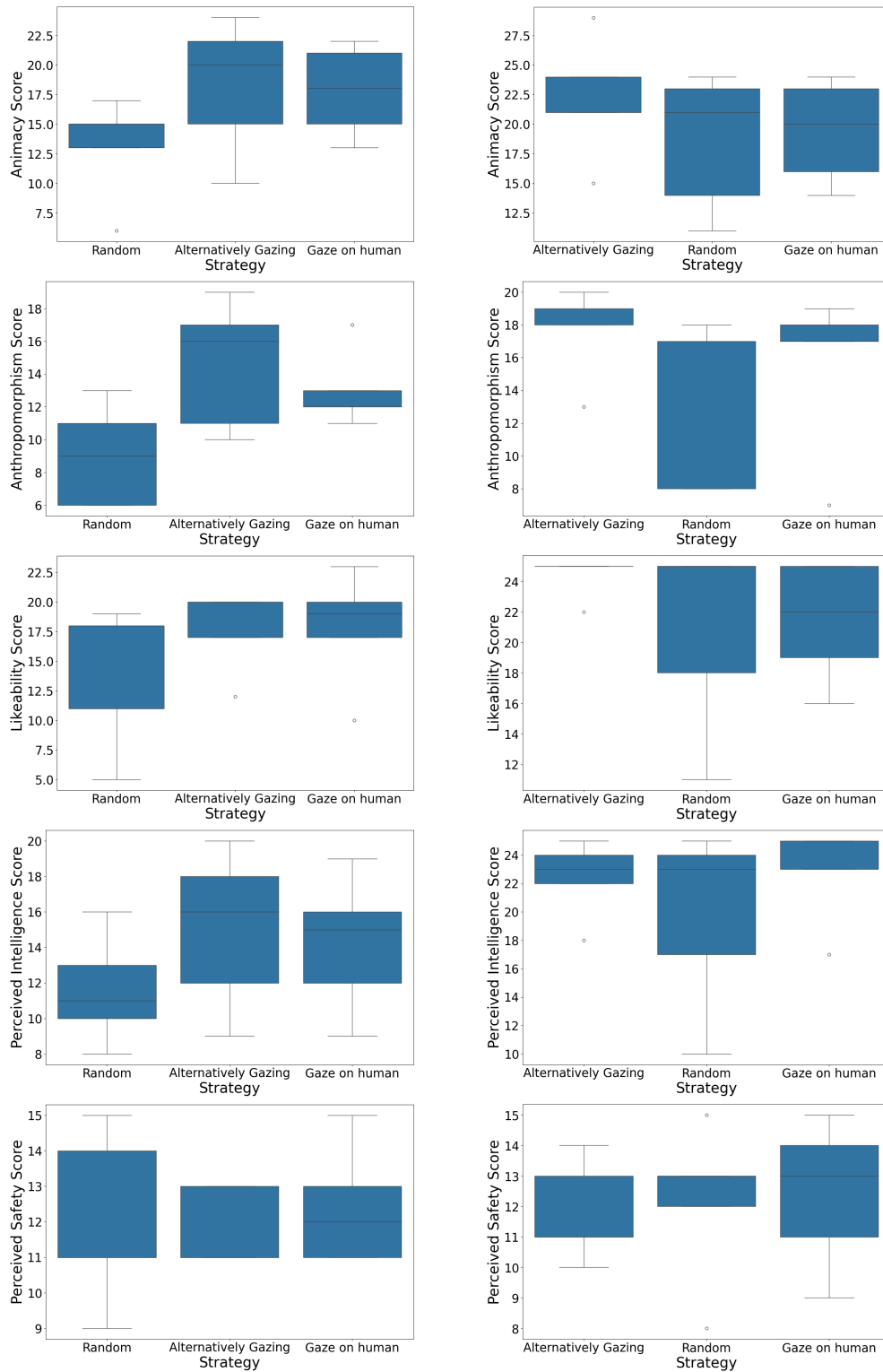


Figure 4.12: Results of the Godspeed questionnaire. The left column shows the offline experiment results, and the right column shows the online experiment results. Each of the boxes represents Random, Alternatively Gazing and Gaze on human strategy.

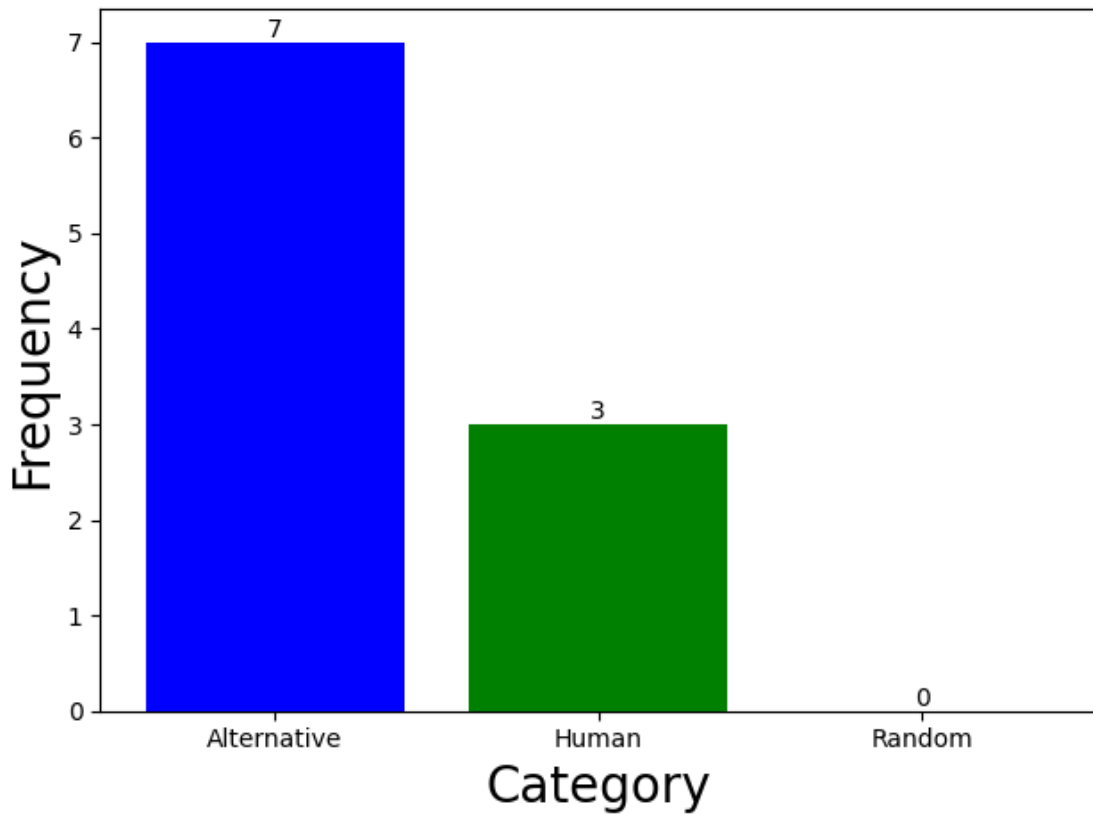


Figure 4.13: In which condition you felt the gaze behaviour of the robot was most appropriate?

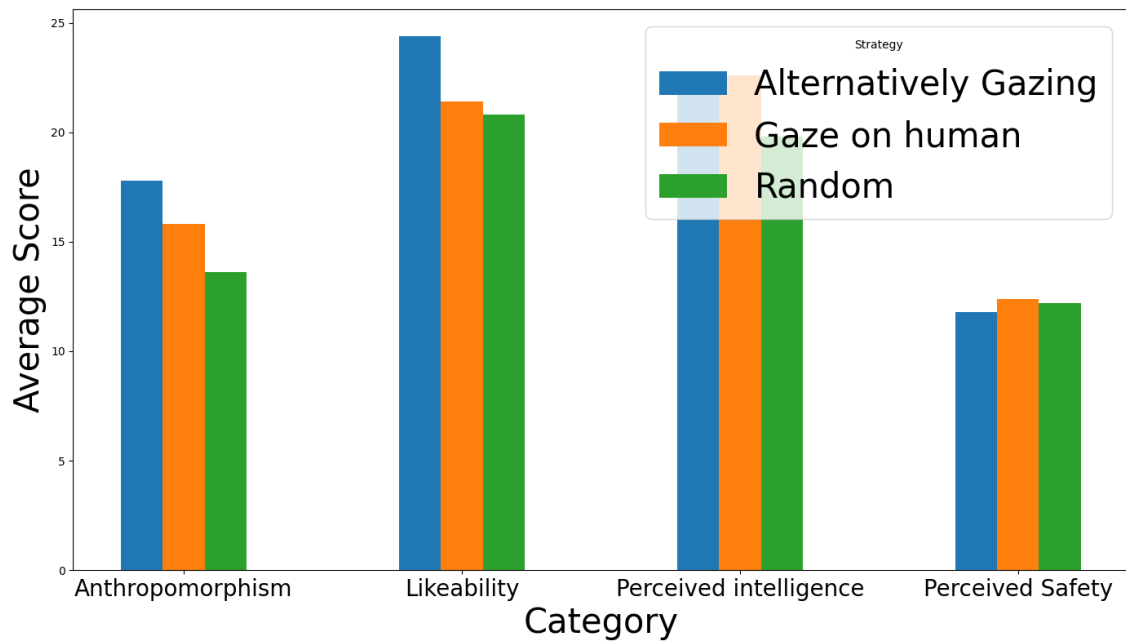
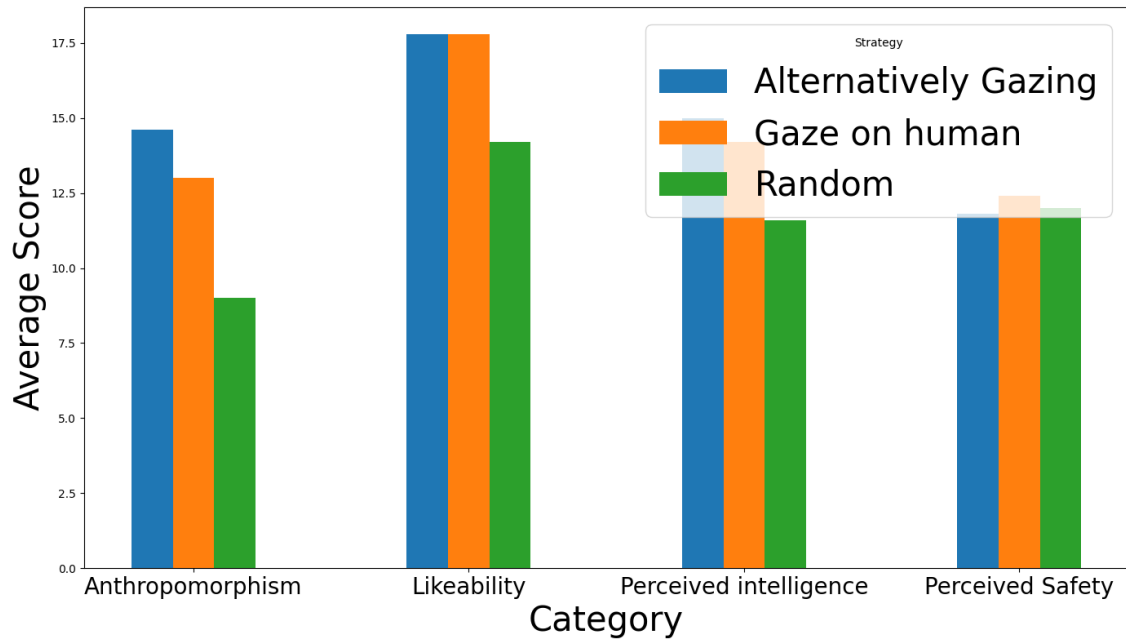


Figure 4.14: Comparison of Strategy Performance Across Categories

Which of the following robot behaviours you didn't like/ find unnatural?	Total
Eye movements	1
Blinking	4
Mouth movements	1
Small random body movements	5

Figure 4.15: Which of the following robot behaviours you didn't like/ find unnatural?

Which of the following robot behaviours did you like/ find appropriate for the situation?	Total
Eye movements	10
Blinking	7
Mouth movements	7
Small random body movements	5

Figure 4.16: Which of the following robot behaviours did you like/ find appropriate for the situation?

4. Experiments and Results

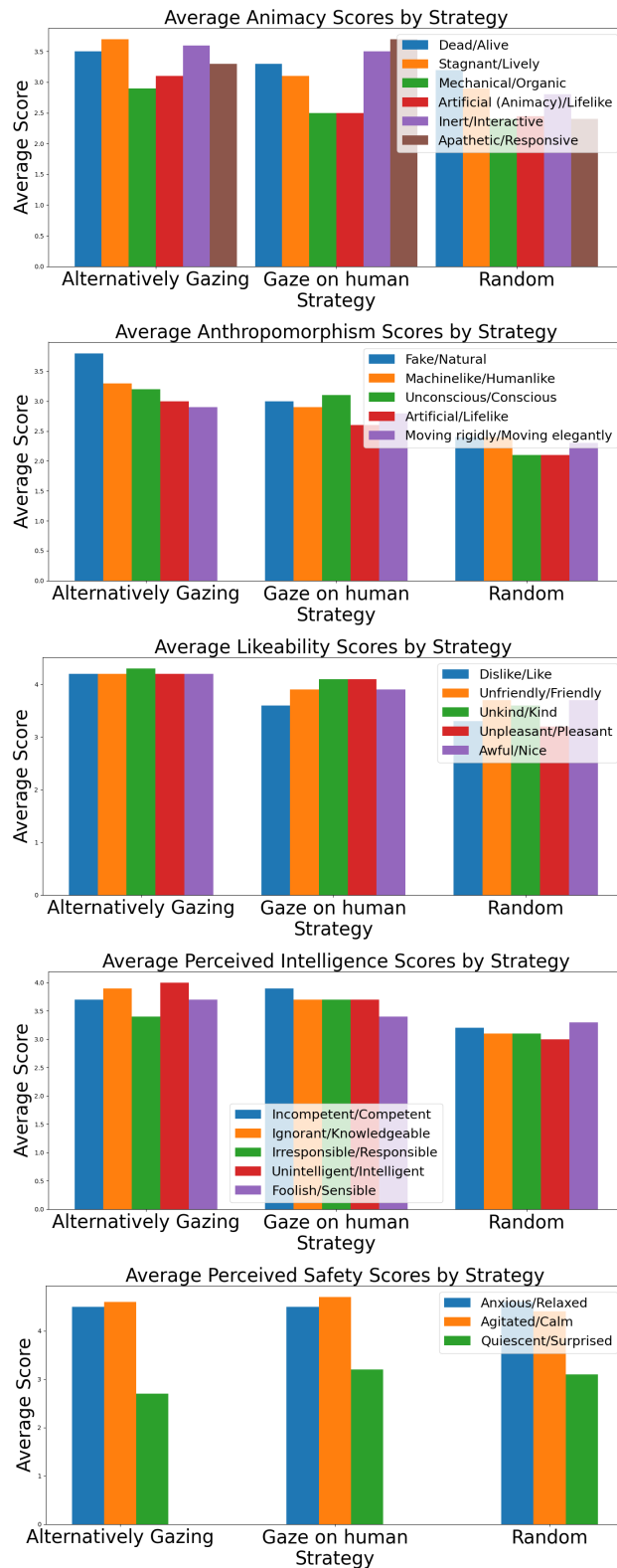


Figure 4.17: The overall results of Godspeed questionnaire from both online and offline part of the experiment by subscales.

Chapter 5

Discussion, Conclusion and Future Work

5.1 Conclusion

This thesis presents the development of an eyes controller for the humanoid robot iCub. The controller enables the robot to track human movements and respond with natural-looking gaze behaviors. We tested the controller's ability to track human in two scenarios, demonstrating its capability to follow human face and hand movements. For detailed results, refer to Fig. 4.5.

To improve safety in human-robot interactions, we implemented a method to estimate the distance between the robot and a human during interactions. We evaluated this method in an experiment in which the human hand was tracked and it achieved better results than the MediaPipe library method (see Fig. 4.3). Although the distance estimation is approximate, it can improve collision avoidance methods, contributing to safer interactions.

We put together a full system for the robot that includes mouth movements, audio from the speakers, small random torso movements, blinking, and eye movements. The solution includes both Python and C++ components. The Python part handles fast development of target generation and easy incorporation of human detection networks, while the C++ part manages the precise motor control. This integration allowed us to implement and test different gaze control strategies, from random movements to smart behaviors that consider the interaction context.

We conducted a pilot social study to evaluate these gaze strategies. Our results show that smart strategies, where eye contact was established and the eyes moved according to the current context of the interaction, were seen as more natural and effective. These methods made people feel more comfortable and connected with the robot during the interaction (see Fig. 4.13).

5.2 Discussion

In this work, we faced many challenges. Initially, we planned to test the controller in the context of an interactive card game. However, we realized that the game was cognitively challenging so that the participants would not pay attention to the robot's gaze. They would be too focused on the game. The next idea of the scenario made people bored as there was not much interaction and they started ignoring the robot's gaze. So, we ended up using a very simple and quick interaction scenario.



Bibliography

- [1] M. Argyle, M. Cook, and D. Cramer, “Gaze and Mutual Gaze,” *The British Journal of Psychiatry*, vol. 165, no. 6, pp. 848–850, 1994.
- [2] C. Breazeal, “Emotion and sociable humanoid robots,” *International Journal of Human-Computer Studies*, vol. 59, no. 1-2, pp. 119–155, 2003.
- [3] B. Mutlu, J. Forlizzi, and J. Hodgins, “A Storytelling Robot: Modeling and Evaluation of Human-like Gaze Behavior,” in *2006 6th IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2006, pp. 518–523.
- [4] A. Roncone, U. Pattacini, G. Metta, and L. Natale, “A Cartesian 6-DoF Gaze Controller for Humanoid Robots.” in *Robotics: science and systems*, vol. 2016, 2016.
- [5] A. N. Alshakhs, M. F. Mysorewala, A.-W. A. Saif, and K. Alshehri, “A Novel Algebraic Inverse Kinematics Based Approach to Gaze Control in Humanoid Robots,” *IEEE Access*, vol. 11, pp. 50 350–50 363, 2023.
- [6] C. Mishra and G. Skantze, “Knowing Where to Look: A Planning-based Architecture to Automate the Gaze Behavior of Social Robots,” in *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2022, pp. 1201–1208.
- [7] P. Viola and M. J. Jones, “Robust Real-Time Face Detection,” *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2004.
- [8] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, “BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs,” *arXiv preprint arXiv:1907.05047*, 2019.
- [9] Y. Xu, J. Zhang, Q. Zhang, and D. Tao, “ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 38 571–38 584, 2022.
- [10] J. Docekal, J. Rozlivek, J. Matas, and M. Hoffmann, “Human Keypoint Detection for Close Proximity Human-Robot Interaction,” in *2022 IEEE-RAS 21st International Conference on Humanoid Robots (Humanoids)*. IEEE, 2022, pp. 450–457.
- [11] S. Gu, A. Kshirsagar, Y. Du, G. Chen, J. Peters, and A. Knoll, “A human-centered safe robot reinforcement learning framework with interactive behaviors,” *Frontiers in Neurorobotics*, vol. 17, 2023.

- [12] K. Kompatsiari, F. Ciardo, V. Tikhanoff, G. Metta, and A. Wykowska, “It’s in the Eyes: The Engaging Role of Eye Contact in HRI,” *International Journal of Social Robotics*, vol. 13, pp. 525–535, 2021.
- [13] H. Lehmann, I. Keller, R. Ahmadzadeh, and F. Broz, “Naturalistic Conversational Gaze Control for Humanoid Robots - A First Step,” in *Social Robotics: 9th International Conference, ICSR 2017, Tsukuba, Japan, November 22-24, 2017, Proceedings 9*. Springer, 2017, pp. 526–535.
- [14] G. Briggs, M. Chita-Tegmark, E. Krause, W. Bridewell, P. Bello, and M. Scheutz, “A Novel Architectural Method for Producing Dynamic Gaze Behavior in Human-Robot Interactions,” in *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2022, pp. 383–392.
- [15] H. Lehmann, A. Roncone, U. Pattacini, and G. Metta, “Physiologically Inspired Blinking Behavior for a Humanoid Robot,” in *Social Robotics: 8th International Conference, ICSR 2016, Kansas City, MO, USA, November 1-3, 2016 Proceedings 8*. Springer, 2016, pp. 83–93.
- [16] H. Admoni and B. Scassellati, “Social eye gaze in human-robot interaction: a review,” *Journal of Human-Robot Interaction*, vol. 6, no. 1, pp. 25–63, 2017.
- [17] 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 *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2016, pp. 905–910.
- [18] M. Raković, N. F. Duarte, J. Marques, A. Billard, and J. Santos-Victor, “The Gaze Dialogue Model: Nonverbal Communication in HHI and HRI,” *IEEE Transactions on Cybernetics*, 2022.
- [19] C. J. Stanton and C. J. Stevens, “Don’t Stare at Me: The Impact of a Humanoid Robot’s Gaze upon Trust During a Cooperative Human–Robot Visual Task,” *International Journal of Social Robotics*, vol. 9, pp. 745–753, 2017.
- [20] L. Haeflinger, F. Elisei, S. Gerber, B. Bouchot, J.-P. Vigne, and G. Bailly, “On the Benefit of Independent Control of Head and Eye Movements of a Social Robot for Multiparty Human-Robot Interaction,” in *International Conference on Human-Computer Interaction*. Springer, 2023, pp. 450–466.
- [21] M. Koller, A. Weiss, M. Hirschmanner, and M. Vincze, “Robotic gaze and human views: A systematic exploration of robotic gaze aversion and its effects on human behaviors and attitudes,” *Frontiers in Robotics and AI*, vol. 10, p. 1062714, 2023.
- [22] T. Shintani, C. T. Ishi, and H. Ishiguro, “Analysis of Role-Based Gaze Behaviors and Gaze Aversions, and Implementation of Robot’s Gaze Control for Multi-party Dialogue,” in *Proceedings of the 9th International Conference on Human-Agent Interaction*, 2021, pp. 332–336.
- [23] D. Mikhaylovskaya, “The project code,” <https://gitlab.fel.cvut.cz/body-schema/icub/icub-eyes-control-naturalistic>, accessed: May 24, 2024.

- [24] —, “The project videos,” <https://drive.google.com/drive/u/1/folders/1hDS0SNNN4-EYgnJmElAe3FvTaqDR9nbE>, accessed: May 24, 2024.
- [25] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor *et al.*, “The iCub humanoid robot: An open-systems platform for research in cognitive development,” *Neural Networks*, vol. 23, no. 8-9, pp. 1125–1134, 2010.
- [26] Google AI, “MediaPipe Holistic,” <https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/holistic.md>, accessed: May 24, 2024.
- [27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
- [28] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device Real-time Body Pose tracking,” *arXiv preprint arXiv:2006.10204*, 2020.
- [29] H. Xu, E. G. Bazavan, A. Zanfir, W. T. Freeman, R. Sukthankar, and C. Sminchisescu, “GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 6184–6193.
- [30] Google AI, “Pose landmark detection,” https://ai.google.dev/edge/mediapipe/solutions/vision/pose_landmarker/index, accessed: May 24, 2024.
- [31] —, “Face detection,” https://ai.google.dev/edge/mediapipe/solutions/vision/face_detector/index, accessed: May 24, 2024.
- [32] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv:1704.04861*, 2017.
- [33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single Shot MultiBox Detector,” in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
- [34] Google AI, “Hand landmarks detection,” https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker/index, accessed: May 24, 2024.
- [35] M. A. Mhamdi, “Polynomial Regression Model for the Hand Distance Estimation,” <https://github.com/MohamedAlaouiMhamdi/Hand-Distance-Measurement/blob/main/HandDistanceMeasurement.py>, accessed: May 24, 2024.
- [36] A. Roncone, “The iCubBreather,” <https://robotology.github.io/funny-things/module/iCubBreather.html>, accessed: May 24, 2024.
- [37] Narakeet, “Child Text to Speech,” <https://www.narakeet.com/create/text-to-speech-child-voice-online.html>, accessed: May 24, 2024.

- [38] N. G. Lorenzo Natale, Alex Bernardino, “The iCub faceExpressions,” https://robotology.github.io/robotology-documentation/doc/html/group___icub___faceExpressions.html, accessed: May 24, 2024.
- [39] 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, pp. 71–81, 2009.
- [40] B. Calli, A. Singh, J. Bruce, A. Walsman, K. Konolige, S. Srinivasa, P. Abbeel, and A. M. Dollar, “Yale-CMU-Berkeley dataset for robotic manipulation research,” *The International Journal of Robotics Research*, vol. 36, no. 3, pp. 261–268, 2017.
- [41] K. Friebe, S. Samporová, K. Malinovská, and M. Hoffmann, “Gaze Cueing and the Role of Presence in Human-Robot Interaction,” in *International Conference on Social Robotics*. Springer, 2022, pp. 402–414.



Appendix A

List of AI tools used in the work

- grammarly, Writefull - Grammar insights
- DeepL - Translation between languages
- ChatGPT - Paraphrasing text into more academic language




Appendix B

The Godspeed questionnaire used after the each interaction with robot

All five subscales (Anthropomorphism, Animacy, Likeability, Perceived intelligence, and Perceived Safety) were used for the study.

After the each interaction



 Not shared

* Indicates required question

Code *

CR

GH

GA

ID *

Your answer

Anthropomorphism

Please rate your impression of the robot on these scales:

Question

Fake 1 2 3 4 5 Natural

Machinelike 1 2 3 4 5 Humanlike

Unconscious 1 2 3 4 5 Conscious

Artificial 1 2 3 4 5 Lifelike

Moving rigidly 1 2 3 4 5 Moving elegantly

Animacy

Please rate your expression of the robot on these scales:

Dead 1 2 3 4 5 Alive

1 2 3 4 5
Stagnant ○ ○ ○ ○ ○ Lively

1 2 3 4 5
Mechanical ○ ○ ○ ○ ○ Organic

1 2 3 4 5
Artificial ○ ○ ○ ○ ○ Lifelike

1 2 3 4 5
Inert ○ ○ ○ ○ ○ Interactive

1 2 3 4 5
Apathetic ○ ○ ○ ○ ○ Responsive

Likeability

Please rate your impression of the robot on these scales:

1 2 3 4 5
Dislike ○ ○ ○ ○ ○ Like

1 2 3 4 5
Unfriendly ○ ○ ○ ○ ○ Friendly

1 2 3 4 5
Unkind ○ ○ ○ ○ ○ Kind

1 2 3 4 5
Unpleasant ○ ○ ○ ○ ○ Pleasant

1 2 3 4 5
Awful ○ ○ ○ ○ ○ Nice

Perceived intelligence

Please rate your impression of the robot on these scales:

Incompetent 1 2 3 4 5 Competent

Ignorant 1 2 3 4 5 Knowledgeable

Irresponsible 1 2 3 4 5 Responsible

Unintelligent 1 2 3 4 5 Intelligent

Foolish 1 2 3 4 5 Sensible

Perceived Safety

Please rate your emotional state on these scales:

	1	2	3	4	5	
Anxious	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Relaxed

	1	2	3	4	5	
Agitated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Calm

	1	2	3	4	5	
Quiescent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Surprised

Submit

[Clear form](#)

Never submit passwords through Google Forms.

This form was created inside of Faculty of Electrical Engineering, Czech Technical University in Prague. [Report Abuse](#)

Google Forms



Appendix C

The Final questionnaire used after the all three interactions with robot

Participants had to fill in this questionnaire after all three interactions with the robot.

The first two questions were pre-filled. The Code question means the order of gaze behaviors that were observed by the participant encoded by three unique digits from 1 to 3. ID means the unique ID of each participant encoded by two letters.

Final questionnaire

Please answer the questions as precisely as possible. If you feel you would like to give additional comments or suggestions, please use the last question.



 Not shared

* Indicates required question

Code *

Your answer

ID *

Your answer

Which of the following robot behaviours have you noticed?

- Eye movements
- Blinking
- Mouth movements
- Small random body movements

Which of the following robot behaviours did you like/ find appropriate for the situation?

- Eye movements
- Blinking
- Mouth movements
- Small random body movements

Which of the following robot behaviours you didn't like/ find unnatural?

- Eye movements
- Blinking
- Mouth movements
- Small random body movements

Did you notice a difference between the three different conditions?

- Yes
- No

If YES, please briefly describe what you noticed.

Your answer

In which condition you felt the gaze behaviour of the robot was most appropriate?

1

2

3

Have you noticed that robot was trying to find you with his eyes sometimes, when it lost sight of you?

Yes

No

Did you notice that sometimes robot was alternating between looking at the table and at you?

Yes

No

If YES, what do you think was the purpose of these alternating eye movements?

Your answer

Have you noticed that robot was following you with his eyes in some of the conditions?

Yes

No

If YES, what do you think the robot wanted to "express" when he was following you with eyes?

Your answer

Have you noticed that the robot was looking at your hands when you were pointing to the objects?

Yes

No

How did the robot's gaze influence your feeling of connection or engagement with the robot during the experiment?

Not at all 1 2 3 4 5 Very much

How natural did the interaction with the robot feel in terms of its gaze behaviour?

Not at all natural 1 2 3 4 5 Very natural

Any other comments or observations about your experience with the robot's gaze during the experiment?

Your answer

Appendix D

Pilot social study scenario of interaction with robot

At the beginning of the experiment each participant got this printed text with instructions:

“Welcome, and thank you for participating in this study. Today, you’ll interact with a humanoid robot across three sessions, each followed by a brief questionnaire and starts on the same place(between the co-bots).

Robot will instruct you during the interaction.

During the sessions, please pay close attention to the robot’s facial movements.

When it uses command like ‘show me,’ you’ll need to grab an item and show it to the robot.

After all sessions, you’ll complete a final form comparing your experiences.

Please ask any questions now, if anything is unclear.

If you’re ready, we’ll begin with the first session.

Thank you again for your participation.”

The robot then greeted the participant: *“Hello! And welcome to our experiment!”*

If robot could not find the participant in its field of view it used the following instructions:

“I can’t see your face, please take a sit in front of me.”

Followed by: *“Good! Now I can see you clearly, thank you.”*

After that robot continued with the basic scenario:

“I’m excited to engage with you today. Please stand in front of me and let’s get started!”

“Let’s take a look at these objects on the table.”

“Here we have a three objects: ”

