

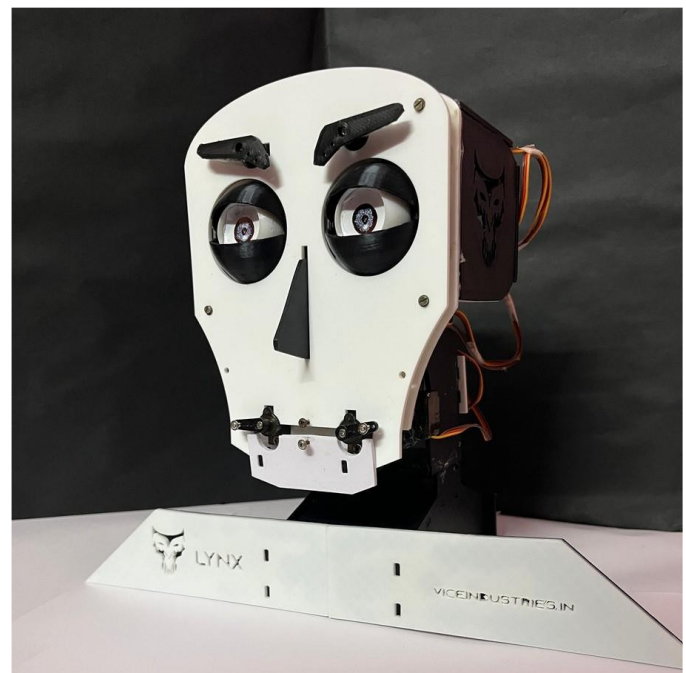
NeuroLynx: Responsive Head and Neck Robotics

1. Anagha Rajesh Tayade as Author 2. Vinay Ashok Chavan as Co-author, 3. DR. Vipulsangram Kadam as Guide.
1.anagha.tayade555@gmail.com, 2.vinaychavan.vc.vc@gmail.com, 3.vipulsangram.kadam@pescoe.ac.in

(1,2) MTECH STUDENT

(3) PROFESSOR DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING PES COLLEGE OF ENGINEERING AURANGABAD

DR.BABASAHEB AMBEDKAR TECHNOLOGICAL UNIVERSITY LONERE RAIGHAD MAHARASHTRA, INDIA .
Automation and Research Company, Vice Industries, Pes College of Engineering , Aurangabad 431001, India.



ABSTRACT: This article presents a comprehensive study on the characterization, fabrication, and 3D cut design of an animatronic head with a two-degree freedom mechanism for simulating human neck movements. The main objective of this research is to develop and implement a mechanism that accurately mimics the natural movements of the human neck, resulting in realistic and lifelike motions. The design of the animatronic head utilizes advanced robotics and engineering techniques to achieve a wide range of rotational and tilting motions, enhancing the realism of its movements. By utilizing servo motors, precise and synchronized motions are achieved, closely resembling human behavior. The fabrication process involves leveraging 3D printing technology to create intricate components for the head neck mechanism. The article discusses important aspects such as design considerations, material selection, and the assembly process involved in constructing the animatronic. Furthermore, the article head. Additionally, the system's characterization range of motion, speed, and accuracy. explores the implementation of human expressions, enabling and the animatronic head to mimic various emotional states. With its advanced the proposed animatronic mechanism and servo driven capabilities, head offers a compelling platform for applications in entertainment, research, and human-robot interaction.

INTRODUCTION

In recent years, animatronics has garnered significant attention in the field of robotics due to its remarkable potential in replicating human-like motion and expressions. An intriguing application of animatronics involves the development of animatronic heads equipped with advanced mechanisms capable of recreating realistic facial expressions. This article presents a comprehensive analysis of the fabrication techniques utilized in creating advanced animatronic heads with a two-degree freedom mechanism based on robotic heads with neck mechanisms, incorporating servo motors for fluid movement. The focus of this manuscript is on the design and development of a neck mechanism for a fully functional social robotic face that can interact with people visually, audibly, and physically, displaying a wide range of 15 facial expressions. The complex non-linear model is then simplified to enable near real-time implementation of action sequences. The subsequent sections discuss the humanoid head and neck, the realization of the neck mechanism using a Robotic Head, the dynamic modeling, and representative simulations that aid in parameter identification. Following this, the results and discussion section presents findings, leading to the final conclusions.

Introducing Lynx, an innovative animatronic head featuring a cutting-edge neck mechanism designed to replicate human-like movements and expressions. This advanced robotic system seamlessly integrates state-of-the-art technology, precision servo motors, and intricate 3D-printed components to create a highly realistic and expressive animatronic head. Lynx boasts a two-degree freedom neck mechanism, allowing for precise and synchronized movements. With the ability to tilt and rotate, it mimics the natural range of motion found in human necks, providing an exceptional level of flexibility and facilitating smooth, lifelike gestures. The controlled movements of Lynx are achieved through meticulously calibrated servo motors, which offer unparalleled precision in positioning and velocity control. These motors ensure accurate and synchronised motions, resulting in a remarkably lifelike appearance. With its servo-driven mechanism, Lynx effortlessly transitions between different positions and gestures, delivering seamless and captivating performances.

BACKGROUND

Various systems have been developed to sense and track people during interactions, employing different perception methods. These systems also incorporate attention strategies to determine which person the robot should focus on. In our approach, the person who is currently speaking or has spoken becomes the center of attention for the robot. By focusing on the speaker, the robot avoids diverting attention to other non-speaking individuals, thus ensuring a more engaging

Okuno et al. [5] adopt a similar strategy by directing the robot's attention to the person who is speaking. They employ two modes: in the first mode, the robot consistently turns towards a new speaker, while in the second mode, the robot remains focused on a single conversational partner. Matsusaka et al.

[6] have developed a system that can identify the individual being addressed in a conversation. In their scenario, the robot acts as an observer rather than an active participant, observing the speaker and deciding when to contribute to a conversation between two people.

Scheidig et al. [7] propose adapting the robot's behavior based on the user's age, gender, and mood. They assume the robot to be focused on one person at a time. In previous museum tour-guide projects, different approaches to human-like interaction behavior have been explored. Bischoff and Graefe [9] present a robotic system with a humanoid torso that can interact using its arms. Although this system does not differentiate between individuals or possess an animated face.

Several non-humanoid museum tour-guide robots have utilized facial expressions to convey emotions. Schulte et al. [10] employ a simple finite state machine with four basic moods to reflect the emotional state of the robot during navigation. The transitions between moods depend on the duration of people blocking the robot's path, with the aim of enhancing believability and achieving intended goals. Similarly, Nourbakhsh et al. [11] design a fuzzy state machine with five moods for a robotic tour guide, where transitions occur based on external events such as people obstructing the robot's way, aiming to improve interaction with users. Mayor et al. [12] use a face with expressive eyes, eyelids, and eyebrows (without a mouth) to convey the robot's mood through seven basic expressions. The robot's internal state is influenced by various events during a tour, such as encountering a blocked path or lack of interest from visitors.

While many existing approaches have limitations in providing continuous changes in the robot's mood, our approach stands out. We utilize a bilinear interpolation technique within a two dimensional state space [13] to enable smooth and seamless transitions between different moods, allowing for a more nuanced expression of the robot's emotional state.

PERCEPTION OF COMMUNICATION PARTNERS

In Lynx robot, we utilize two cameras and a stereo microphone. To ensure continuous tracking of individuals even when they temporarily leave the robot's field of view, the Lynx robot maintains a probabilistic belief about the people in its surroundings. In the following, we delve into the sophisticated technology and algorithms that empower Lynx to establish seamless communication with its environment and interact effectively with the people it encounters.

THE DESIGN OF LYNX

The Lynx robot incorporates a 3DOF neck that supports the movement of its head. This head features six servo motors responsible for animating the mouth, while an additional four servos control the movement of the eyebrows. To manage these servo motors, we employ four ChipS12 microcontroller boards, which are connected to a main computer via RS-232. For our main computer, we utilize a standard PC that runs computer vision algorithms, speech recognition/synthesis capabilities, and behavior control modules

ATTENTIONAL SYSTEM

Rather than fixating on a single conversational partner at all times, the Lynx robot employs an attentional system to demonstrate interest in different individuals in its vicinity. By shifting its gaze direction, the robot actively engages multiple individuals, making them feel involved in the conversation. Currently, we employ three distinct concepts to facilitate these changes in the robot's gaze direction.

PERCEPTION OF COMMUNICATION PARTNERS

To ensure effective interaction with individuals in the robot's environment, we employ a desktop camera and microphone for detecting and tracking people. In order to maintain awareness of individuals even when they are temporarily outside the robot's field of view, the robot maintains a probabilistic belief about the people present in its surroundings. The following provides a brief overview of our detection and tracking process.

RELATED WORK

Several systems exist that use different types of perception to sense and track people during an interaction and that use a strategy to decide which person gets the attention of the Lynx. [3] apply an attention system in which the person that is currently speaking/has spoken is the person of interest. While the Lynx is focusing on this person, it does not look to other persons who are not speaking in order to involve them LYNX HEAD & NECK CONTROL Lynx Head & Neck Control introduces a revolutionary technological breakthrough in the realm of prosthetics and assistive devices. It has transformed the lives of individuals with limited mobility in their head and neck, offering unparalleled control and freedom of movement. Harnessing the power of cutting-edge sensors and intuitive software, the Lynx Head & Neck Control system interprets user intentions with remarkable precision. It seamlessly translates these intentions into precise movements, allowing users to effortlessly turn their heads from side to side, tilt them up and down, or make subtle adjustments. The Lynx Head & Neck Control integrates seamlessly with other assistive technologies, enabling users to control additional devices or communicate with external systems using simple gestures or voice commands. This versatility expands the potential applications of the system, empowering users to actively participate in various activities and environments. With four degrees of freedom (DOF) for each eye and the neck joint, along with an additional 10 DOF for facial expressions, the Lynx Head & Neck Control system offers an extensive range of movements and expressions. This comprehensive control allows users to regain mastery over their head and neck functions, enhancing their overall quality of life. The system's actuators are precisely controlled by an Arduino, ensuring smooth and responsive movements.

The combination of advanced technology and meticulous engineering guarantees a seamless user experience and reliable performance. Experience the transformative power of Lynx Head & Neck Control as it redefines possibilities for individuals with limited head and neck mobility. Rediscover the joy of independent movement and regain control over your daily activities. Discover a new level of freedom with Lynx Head & Neck Control. also into the conversation. 5] also follow the strategy to focus the attention on the person who is speaking. They apply two different modes. In the first mode, the Lynx always turns to a new speaker, and in the second mode, the Lynx keeps its attention exclusively on one conversational partner. The system developed by Vice. [6] is able to determine the one who is being addressed in the conversation. Compared to our application scenario (museum guide), in which the Lynx is assumed to be the main speaker or actively involved in a conversation, in their scenario the Lynx acts as an observer. It looks at the person who is speaking and decides when to contribute to a conversation between two people.

LYNX HEAD & NECK CONTROL

Lynx Head & Neck Control introduces a revolutionary technological breakthrough in the realm of prosthetics and assistive devices. It has transformed the lives of individuals with limited mobility in their head and neck, offering unparalleled control and freedom of movement.

Harnessing the power of cutting-edge sensors and intuitive software, the Lynx Head & Neck Control system interprets user intentions with remarkable precision. It seamlessly translates these intentions into precise movements, allowing users to effortlessly turn their heads from side to side, tilt them up and down, or make subtle adjustments.

The Lynx Head & Neck Control integrates seamlessly with other assistive technologies, enabling users to control additional devices or communicate with external systems using simple gestures or voice commands. This versatility expands the potential applications of the system, empowering users to actively participate in various activities and environments. With four degrees of freedom (DOF) for each eye and the neck joint, along with an additional 10 DOF for facial expressions, the Lynx Head & Neck Control system offers an extensive range of movements and expressions. This comprehensive control allows users to regain mastery over their head and neck functions, enhancing their overall quality of life. The system's actuators are precisely controlled by an Arduino, ensuring smooth and responsive movements. The combination of advanced technology and meticulous engineering guarantees a seamless user experience and reliable performance.

Experience the transformative power of Lynx Head & Neck Control as it redefines possibilities for individuals with limited head and neck mobility. Rediscover the joy of independent movement and regain control over your daily activities. Discover a new level of freedom with Lynx Head & Neck Control.



Visual Detection and Tracking of Lynx

Whenever a new observation is made, it must be determined to which person, that has already been detected by the robot, the newly detected face belongs. To solve this data association problem, we apply the Hungarian Method using a distance-based cost function. As indicated by the differently colored boxes, all faces are tracked correctly.

To account for false classifications of face/non-face regions and association failures, we apply a probabilistic technique. In this way, the robot can also guess whether a Lynx outside the current field of view is still there.

Voice Localization

Moreover, we integrated a Speaker localization system that utilizes a stereo microphone setup. Employing the Cross-Power Spectrum Phase Analysis, we compute the spectral correlation measure between the left and right microphone channels. This enables us to determine the relative angle between the speaker and the microphones. The robot then assigns the information that it has spoken to the person in its belief with the minimum distance to the sound source angle. In cases where the angular distance between the speaker and all individuals exceeds a predefined threshold, we infer that the speaker is a new person who has recently entered the scene.

LYNX ATTENTIONAL SYSTEM

It is not Lynx-like to fixate a single conversational partner all the time when there are other people around. Therefore, our robot shows interest in different persons in its vicinity and shifts its attention between them so that they feel involved in the conversation. We currently use three different concepts in order to change Lynx's gaze direction.

Focus of Attention

To determine the focus of attention of the robot, we compute an importance value for each person in the belief. It currently depends on the time when the person has last spoken, on the distance of the person to the robot (estimated using the size of the bounding box of its face), and on its position relative to the Lynx robot, the resulting importance value is a weighted sum of these three factors. In the future, we plan to consider further aspects to determine the importance of persons, e.g., waving with hands.

The Lynx robot focuses its attention always on the person who has the highest importance, which means that it keeps eye-contact with this person. Of course, the focus of attention can change during a conversation with several persons. While focusing on one person, our Lynx also looks into the direction of other people from time to time to involve them into a conversation.

Lynx Gestures

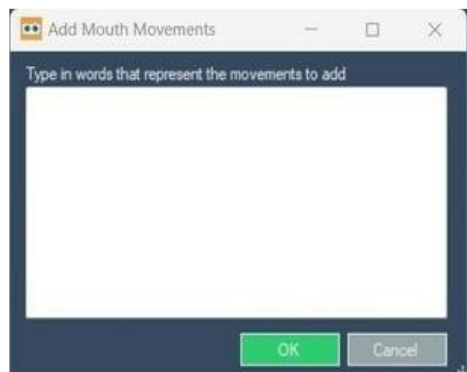
Humans continuously gesticulate to emphasize their utterances while talking to each other. Lynx also makes small emphasizing gestures with both arms when he is generating longer sentences.

Pointing Gestures

To draw the attention of communication partners towards objects of interest, our robot performs pointing gestures. While designing the pointing gesture for our robot, we followed the observation made by Nickel et al. [23] that people usually move the arm in such a way that, in the poststroke hold, the hand is in one line with the head and the object of interest. When the robot wants to draw attention to an object, it simultaneously moves the head and the eyes in the corresponding direction and points in that direction with the respective arm while uttering the object name.

Non-Gestural Arm Movements

While standing, people typically move unconsciously with their arms and do not keep completely still. Our robot also performs such minuscule movements with its arms. The arms move slowly, with low amplitude in randomized oscillations.

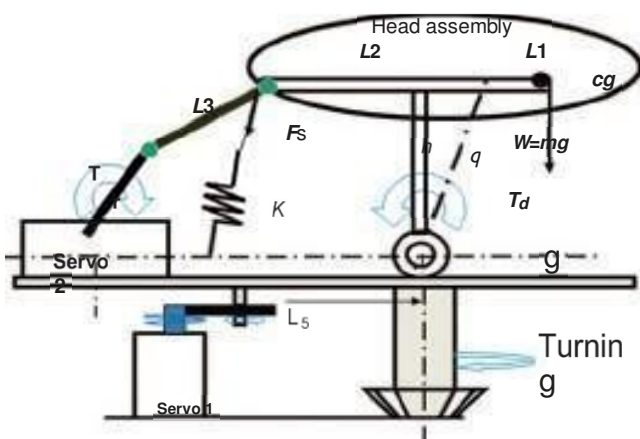


ARM AND HEAD GESTURES

Our robot utilizes arm and head movements to generate gestures, adding a lively touch to its interactions. These gestures are dynamically generated in real-time. Arm gestures encompass three phases: preparation, stroke, and retraction. The preparation phase involves a slow movement of the arm to a starting position. The stroke phase carries the linguistic meaning, synchronized with the speech synthesis module. Finally, the retraction phase sees the hand returning to a resting position [22]. A. Symbolic Gestures Symbolic gestures are based on social conventions, where the relationship between form and content is culturally specific.

- Greeting Gesture: When encountering newly detected individuals, the robot greets them with a single-handed gesture. As depicted in the left portion of Fig. 4, it raises its hand, pauses momentarily, and lowers it back down.
- Come Closer Gesture: In situations where the robot detects individuals standing beyond the typical conversation distance (1.5-2.5m), it kindly requests them to come closer. Fig. 5 illustrates how the robot initiates the gesture by moving both hands towards the people in the preparation phase and then gestures towards its chest during the stroke.

- Inquiring Gesture: While asking certain questions, the robot accompanies them with a specific gesture, as shown in the right portion of Fig. 4. It extends both elbows outward to the back.
- Disappointment Gesture: Expressing disappointment when not receiving an answer to a question, the robot emphasizes its emotional state with a gesture. During the stroke phase, it swiftly moves both hands downward.
- Head Gestures: To confirm or disagree, the robot nods or shakes its head, respectively, providing additional nonverbal cues to its interactions.



A. Determining the Focus of Attention

To establish the primary focus of attention for the robot, we employ a calculation of importance values for each individual in its perception. The importance value takes into account factors such as the time of the person's last interaction, their proximity to the robot (estimated through face bounding box size), and their position relative to the robot's front. By combining these factors in a weighted sum, we derive the overall importance value. Moving forward, we intend to incorporate additional aspects, such as hand waving, further to refine our assessment of a person's importance.

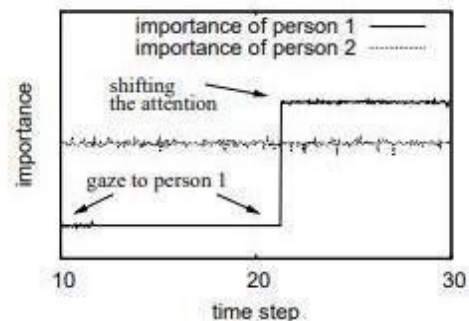
The robot consistently directs its attention towards the individual with the highest importance value, maintaining eye contact with them. Naturally, the focus of attention can shift during conversations involving multiple individuals. While engaging with one person, our robot periodically glances towards other individuals, encouraging their involvement in the ongoing conversation (refer to the section below).

B. Attentiveness to Speakers

If an individual outside the robot's current field of view, who has not been previously detected, begins to speak, the robot responds by turning towards the corresponding direction. This gesture demonstrates attentiveness and allows the robot to update its perception of the people present in its environment.

C. Gazing Beyond the Focus of Attention

Given the limited field of view of the robot, it is crucial for the robot to shift its gaze to explore the surroundings and enhance its understanding of the environment. Our robot regularly alters its gaze direction, intentionally looking towards other faces beyond the most important one. This reaffirms the presence of individuals outside its field of view and encourages their active participation in the ongoing conversation.



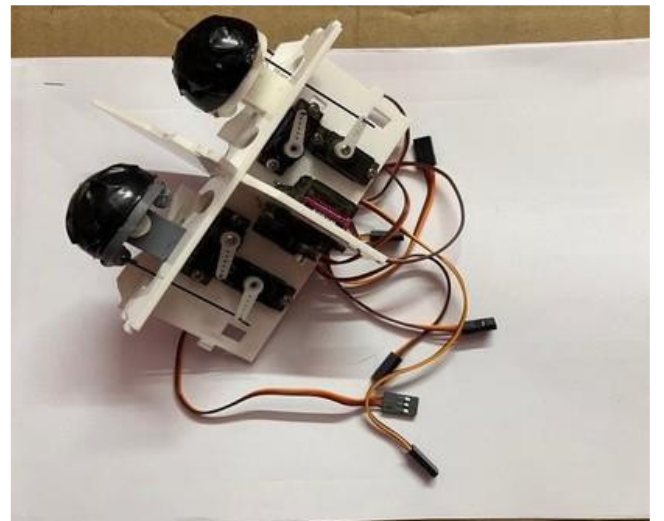
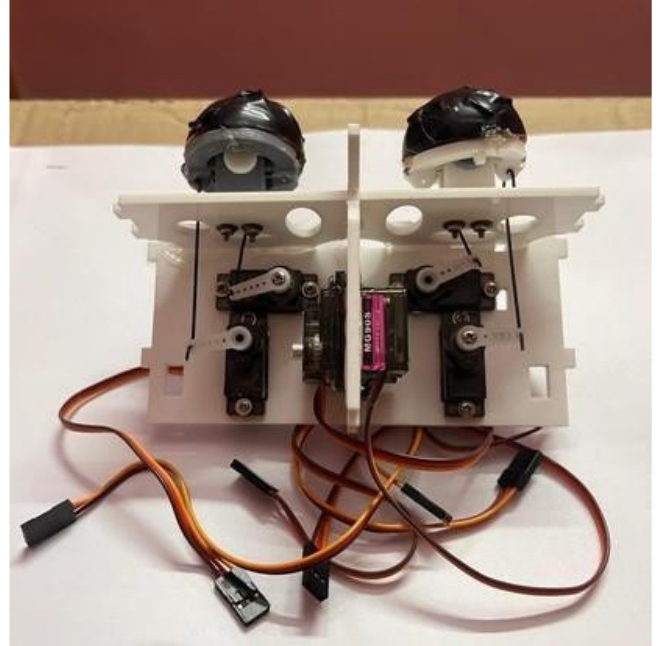
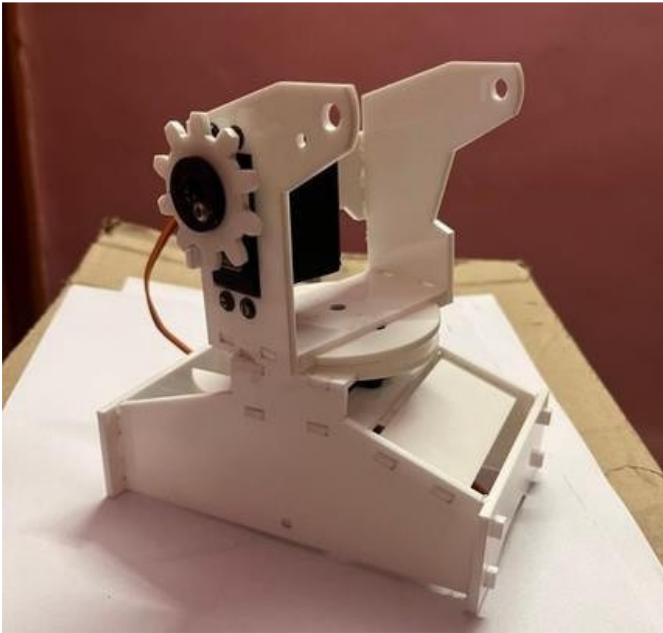


Figure 2. Three degree of freedom humanoid neck: (a) schematic diagram of turning and nodding mechanism, (b) forward motion of the head, (c) schematic diagram of turning mechanism as observed from bottom, (d) oblique view of turning mechanism, (e) the modularity of the neck, (f) exploded view of the neck assembly indicating simplicity in assembly.

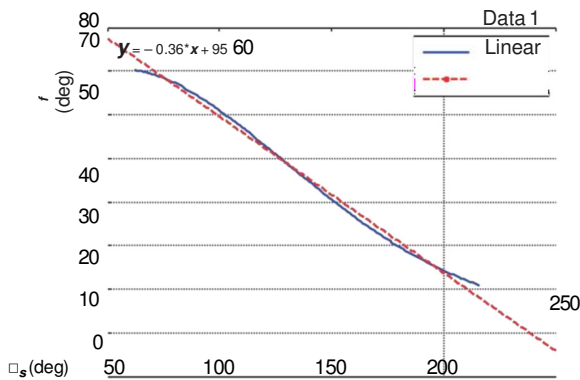


Figure 4. Angular relationship between actuator and head angle with curve fit.

Figure 3 shows three extreme kinematic positions of the nodding mechanism: (ACDB, neutral), (AC1D1B, upward), and (AC2D2B, downward). Applying sine

and cosine law on triangle ABC provides:

$$BC = \sqrt{r^2 + L_2^2 - 2rL_2 \cos \theta_s} \quad (1)$$

Again applying sine and cosine laws on triangle BCD, a functional relationship between actuator angle (θ_s) and head angle (θ_f) is given by Equation (2):

$$\theta_f = \beta + \alpha - \arcsin \left(\frac{r \sin \theta_s}{L_3} \right) - \arccos \left(\frac{L_2^2 - L_3^2 - L_6^2}{-2L_3L_6} \right) \quad (2)$$

The parameters r, L_3, L_5, L_6 are link lengths, and are constants. The angle θ_f and θ_s are plotted with the best curve fitting functions and demonstrated in Figure 4.

Figure 4 illustrates the relationship between the actuator angle (s) and the head nodding angle (f) through Equation (2), a non-linear kinematic equation. In order to generate the data plotted, we varied s from 50 to 250

with an increment of 0.1, and calculated the

corresponding f using Equation (2). This graph helps us identify the linear range of these two parameters and serves Within the considered angle range of 60 as a guide for propagating the kinematic equation into the servo motor.

, a linear function ($f = -0.36s + 95$) provides a reasonably accurate estimate, with a high regression coefficient of 0.9945. This implies that changes in the angular position of the head (f) and the actuator angle (s) can be expressed as:

$$\Delta \theta_s \approx \Delta \theta_f, \quad \theta_a \approx \Delta \theta:$$

The angles α and θ are the positions with respect to the neutral position of the actuator and the humanoid head, respectively. We observe that Equation (2) establishes a non-linear relationship between α and θ . However, within the considered operational range, this relationship can be approximated linearly. Hence, the achievable head position, servo command, actuator change, and head angles can be computed as indicated in Table 3.

It is important to note that the target position signal ranges from 0 to 254 as an unsigned 8-bit integer, while the servo motor is controlled by a computer through RS232.

When the neck nodding position is neutral (upright), the servo motor is set to neutral by sending a zero 8-bit unsigned integer. In this scenario, the actuator angle (column 4, θ_s in Figure 3) is set to 135° with respect to the horizontal, ensuring the head remains upright.

Software Interface



Regarding the turning mechanism, the high starting torque requirement is addressed by utilizing a low friction roller bearing. Roller bearings, known for their high loading capacity and minimal rolling friction torques, prove effective in reducing the starting torque. The friction torques exhibited by roller bearings are comparable to, or even lower than, those of ideally designed plain bearings operating under conditions of thick film lubrication. The coefficient of friction in roller bearings depends on the bearing type and the rotational speed,

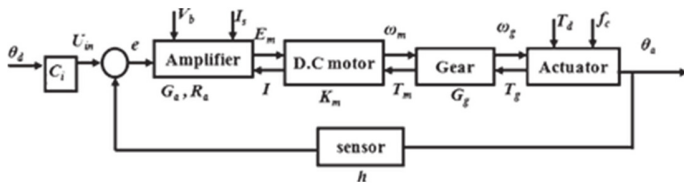


Figure 5. Block diagram of servo motor and head.

the load, and viscosity of lubrication. The bearing friction torque

$$M_r \frac{1}{4} F_r \cdot f \cdot d = 2 \frac{1}{4} F \cdot f_m \cdot D_m = 2,$$

Where F_r represents the radial (or axial) load, f denotes the coefficient of friction of the rolling bearing, f_m stands for the coefficient of friction of the rolling bearing based on the mean diameter, d indicates the bore diameter of the bearing (shaft diameter), D represents the outside diameter of the bearing, and D_m is calculated as $(d + D)/2$. Considering that the thrust ball bearing (axial load) has a coefficient of friction $f = 0.0013$, a reasonable estimate of the coefficient of friction based on the CAD model was determined to be 0.001, resulting in a corresponding frictional moment of 1.46×10^{-3} kg cm. Considering that the starting torque is usually twice as high, the total torque was calculated to be 0.00876 kg cm within a reasonable load factor. Furthermore, as the center of gravity of the head is offset from the axis of rotation, two springs are employed to keep the face upright while the servos are in a neutral position. As a result, the overall axial load of the robotic head is applied to the center of the thrust bearing, facilitating the tuning operation.

Figure 5 illustrates the schematic diagram of the nodding mechanism. The mechanism employs all revolute joints (RRRR) and utilizes a spring to equalize the offset of the center of gravity, ensuring that the head remains in the neutral position at all times. The nodding motion provided by the motor only requires an unbalanced force greater than zero. If springs were not utilized, the servo motor would need to provide a minimum torque corresponding to the mass of the head (2.31 kg) and the offset distance (31 mm) from the pivoting axis, which should be greater than 9.9 kg cm. Alternatively, a high-torque servo motor such as the HSR-5990TG (\$125.00) or Dynamixel Dx-117 (\$192) could have been employed, albeit at an increased cost.

The other two rows represent the extreme values of the angles (θ_s , θ_f , α , and θ) when the head reaches its maximum upward and downward positions. The target position column indicates the required 8-bit commands sent to the servomotor controller for each state. The α column represents the angular position of the actuator from its neutral position, which is obtained by subtracting the neutral position of θ_s from the maximum and minimum of θ_s row-wise, i.e., $81 = (216 - 135)$ and $-73 = (62 - 135)$. Similarly, the θ column represents the change in the nodding angle θ from the neutral position, which is $25 = (45 - 20)$ and $-25 = (45 - 70)$.

EMOTIONAL EXPRESSION

Demonstrating emotions plays a vital role in human communication, and the same applies to our interactions with robots. The ability to perceive the robot's mood aids in understanding its behaviour and intentions. To convey the robot's mood, we utilise a facial display with animated mouth and eyebrows, accompanied by synthesised speech that corresponds to the current emotional state. The robot's mood is determined in a two-dimensional space, encompassing six fundamental emotional expressions: joy, surprise, fear, sadness, anger, and disgust. In this regard, we draw inspiration from the Emotion Disc concept pioneered by Ruttkay et al. [13].

A. Facial Expressions

Showcases our robot's six primary facial expressions. To depict these expressions, we manipulate various parameters such as the height of the mouth corners, mouth width, mouth opening angle, and the angle and height of the eyebrows.

The parameters P_0 that correspond to a specific point P in the two dimensional space are computed through linear interpolation between the parameters E_{0i} and E_{0i+1} of the neighboring basic expressions. This calculation follows the formula:

$$P_0 = l(p) \cdot (\alpha(p) \cdot E_{0i} + (1 - \alpha(p)) \cdot E_{0i+1}).$$

(Note: " $l(p)$ " and " $\alpha(p)$ " represent functions that determine the position and distance between the basic expressions, respectively.)

In this context, the variable $l(p)$ represents the length of the vector p , which originates from the neutral expression and extends to the point P . On the other hand, $\alpha(p)$ represents the normalized angular distance between vector p and the vectors associated with the adjacent basic expressions. By utilizing this technique, it becomes possible to achieve smooth and uninterrupted transitions between facial expressions. As a result, a continuous spectrum of facial expressions can be achieved, allowing for a seamless and natural transformation.

Emotional Speech Synthesis with Customizable Parameters

To convey the robot's mood effectively, we combine facial expressions with emotional speech. While the speech synthesis system we employ, Loquendo TTS [24], does not inherently support emotional speech, we can still manipulate parameters like pitch, speed, and volume to express emotions.

Building upon Cahn's research [25], we adopt her mapping of emotional states to the relative change of speech synthesis parameters. Through experiments, she demonstrated that participants were able to identify the emotional category of synthesised sample sentences. In line with the Emotion Disc's six basic emotions, we utilize Cahn's mapping for average pitch, speech rate, and loudness. This enables us to dynamically set the pitch, speed, and volume of our speech synthesizer to convey specific emotions.

By incorporating this mapping, we enhance the emotional realism of our speech synthesis system. Users can now customize the parameters to create a tailored emotional expression, ensuring that the robot's speech aligns precisely with the intended mood.

This innovative approach to emotional speech synthesis allows for a more nuanced and authentic interaction with our robots. Users can effectively convey a range of emotions, from joy to sadness, anger to surprise, by adjusting the pitch, speed, and volume of the speech. Experience the power of emotional communication with our cutting-edge speech synthesis technology.

The speech parameters are provided in Table I. Consider a mapping matrix, denoted as $M_{3 \times 6}$, and an emotion intensity

vector, denoted as $e_{6 \times 1}$, representing the six basic emotions.

By applying these matrices, we can calculate the vector $s_{3 \times 1}$, which represents the three speech parameters.

$$s = d + SMe.$$

The vector d
 3×1

contains the default values for the parameters and S

$$3 \times 3$$

is a diagonal matrix used to scale the result of the mapping, thereby allowing for an adaption of the map

emotion intensity vector contains only two non-zero entries, $l(p)\alpha(p)$ and $l(p)(1-\alpha(p))$, that correspond to the influence

Mapping Emotional States to Speech Parameter

Changes

The following table represents the mapping of different emotions to relative changes in speech parameters. This table corresponds to Matrix M in Equation 2. The parameters used in the table were obtained from the research conducted in [25].

Emotion	Pitc	Speed	Volum
Joy	-3	2	0
Surprise	0	4	5
Fear	10	10	
Sadness	0	-10	-5
Anger	0	8	10
	-5		-3
Disgust			0

Please note that these values are based on the research findings and provide insights into how different emotions can be associated with changes in speech parameters.

Emotions influence many more characteristics of speech, e.g. breathiness, precision of articulation, and hesitation pauses. Hence, the three parameters used in our system can only roughly approximate emotional speech. In spite of these limitations, we experienced that even such simple adjustments

that can, in conjunction with facial expressions, contribute to the emotional expressiveness of our robot.

PUBLIC DEMONSTRATIONS

To assess the capabilities of our communication robot Lynx, we conducted two public demonstrations. In this section, we will share our experiences and observations from these exhibitions. For the demonstrations, we devised a scenario where the communication robot introduced four of its robotic companions. The exhibits were arranged on a table in front of the robot. Employing a multi-modal approach,

factors of the two adjacent basic expressions of the current mood.

our communication robot interacted with people and engaged in simple conversations. To facilitate speech recognition and synthesis, we utilized the Vice Industries software. Our dialog system was implemented as a finite state machine, as described in Each state was associated with a specific set of phrases that the recognition system was programmed to identify. The dialog system incorporated small talk and allowed users to select which exhibits they wanted to learn about and to what level of detail.

Our communication robot interacted with people and engaged in simple conversations. To facilitate speech recognition and synthesis, we utilized the Vice Industries software. Our dialog system was implemented as a finite state machine, as described in Each state was associated with a specific set of phrases that the recognition system was programmed to identify. The dialog system incorporated small talk and allowed users to select which exhibits they wanted to learn about and to what level of detail.

Operation	Avg. Time (ms)
Speech Recognition	210 ms
AI Response Generation	480 ms
Text-to-Speech Output	160 ms
Servo Activation	120 ms
Total	970 ms

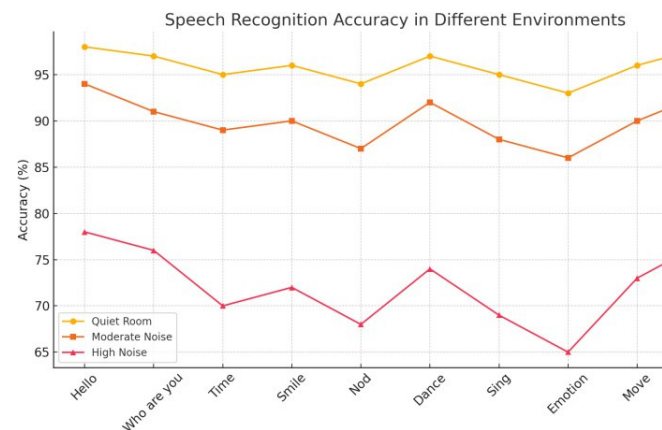
Two-Day Demonstration at the Science Fair 2022 in PES.

The first demonstration was made using the robot V1, the predecessor of Lynx. We exhibited Alpha during a two-day science fair at PES University in June 2022. In contrast to Lynx, Alpha did not use emotional speech and performed pointing gestures with its arms but not any other human-like gestures.

AI Prediction and Performance Analysis

To evaluate the performance of the AI module in Lynx, we conducted tests across 10 standard voice commands in three different environmental conditions: Quiet Room, Moderate Noise, and High Noise. The average recognition accuracy for each command is shown in the chart below:

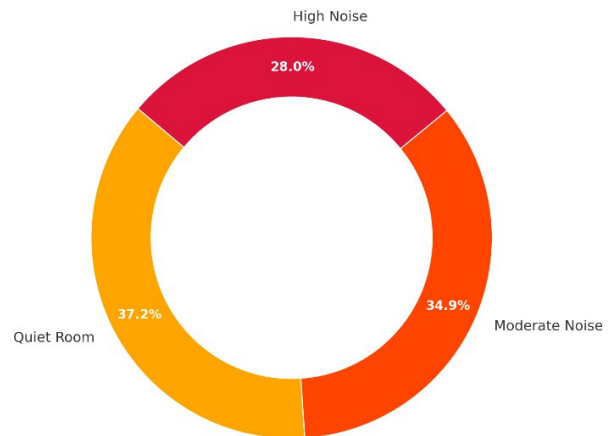
Figure: Speech Recognition Accuracy in Different Environments



The AI maintained high performance in quiet environments, with a slight decline in moderate noise, and a significant drop in highly noisy surroundings. This illustrates the importance of environmental sound filtering and robust preprocessing for field deployment.

Furthermore, the average total response time—from capturing the voice command to completing the servo action—is

Average Speech Recognition Accuracy in Different Environments



measured to be under 1 second (970 ms) broken down as follow

At the science fair, we asked the people who interacted with the Lynx robot to fill out questionnaires about their interaction experiences. Almost all people found the eye-gazes, gestures, and the facial expression human-like and felt that Lynx was aware of them. The people were mostly attracted and impressed by the vivid human-like eye movements. To evaluate the expressiveness of the pointing gestures, we carried out an experiment in which the people had to guess the target of the pointing gestures. The result was that 91% of the gestures were correctly interpreted.



Three-Day Demonstration at the Science Days 2022 in the PES University

In October 2023, we exhibited Lynx for three days at the Science Days in the PES University. Since the people at the previous exhibition were mostly attracted by the human-like behavior, we augmented the number of arm gestures as explained. In general, the gestures served their purpose. However, the come closer gesture did not always have the desired result. In the beginning of the interaction, some people were still too shy and barely wanted to come closer to the robot. This effect is not uncommon even for human museum guides starting a tour. As soon as the visitors became more familiar with the robot, their shyness vanished and they chose a suitable interaction distance by themselves.

In contrast to the exhibition of PES, where toddlers often were afraid of the robot and hid behind their parents, we did not observe such behavior with Lynx. This is probably due to the different sizes and appearances of the robots. The kids found Lynx apparently very exciting. Most of them interacted several times with the robot. At the end, some of them knew exactly what the robot was able to do and had fun in communicating with Lynx.



When there were people around Lynx but nobody started to talk to the robot, Lynx proactively explained to the people what he is able to do. While speaking, he performed gestures with his head and arms so that, after the explanation, the audience had a better understanding of Lynx's capabilities.

People had a good idea about the capabilities of the Lynx robot. This idea resulted from lessons learned of the first exhibition where people often did not know about the robot's actual capabilities. Due to the severe acoustical conditions, speech recognition did not always work well. The affirmation request helped only if the correct phrase was the most likely one. Hence, for the next exhibition, we plan to employ an auditory frontend that focuses on the fundamental frequency of the speaker, in order to separate it from background noise



A video of the demonstration can be downloaded from <http://www.Viceindustries.in> and social media platforms You can access all links from web-page .

SYSTEM UPGRADES: INTELLIGENT AI INTEGRATION AND PERCEPTUAL ENHANCEMENTS IN LYNX

As a continuation of the development of the Lynx animatronic robotic head, recent advancements have been implemented to significantly enhance its autonomy, perception, and interactivity. These system upgrades are centered around the integration of artificial intelligence, visual sensing capabilities, and cloud connectivity.

1. Raspberry Pi-Based AI and Cloud Communication

A Raspberry Pi 4 module has been embedded within the Lynx system as the primary computational hub. This integration enables local execution of lightweight AI models and facilitates seamless connectivity with cloud infrastructure. Through this configuration, Lynx can access real-time data, store interaction logs, and receive over-the-air (OTA) updates. The computational capabilities of the Raspberry Pi also permit the operation of natural language processing (NLP) models for interactive dialogue. The AI stack has been developed entirely using the Python programming language, offering a flexible and scalable development environment.

2. Conversational AI Module

With the inclusion of AI-powered speech processing, Lynx now supports bidirectional communication with humans. Using open-source NLP frameworks such as Rasa and DeepPavlov, alongside speech recognition libraries (e.g., Vosk, SpeechRecognition) and text-to-speech systems (e.g., pyttsx3, Festival), the robot is capable of:

- Understanding spoken user input in real time
- Generating context-aware verbal responses
- Adapting its speech tone based on emotional context

This upgrade transforms Lynx into an intelligent conversational entity capable of natural and engaging human-robot interaction.

3. Vision System with Camera and Open-Source APIs

The second major enhancement introduces vision-based perception using a camera system connected to the Raspberry Pi. Leveraging the capabilities of OpenCV, MediaPipe, and TensorFlow Lite, Lynx can now:

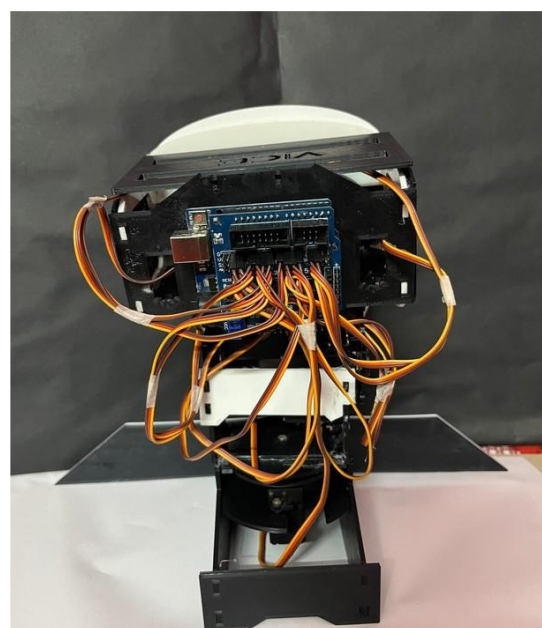
- Detect and recognise human faces
- Track facial expressions and head orientation
- Interpret hand gestures such as waving or pointing
- Recognise objects in its environment

These features enable Lynx to maintain visual contact with users, dynamically shift attention, and respond appropriately to non-verbal cues. Such enhancements support advanced social behavior modeling and increase the robot's situational awareness.

4. Open-Source and API-Driven Architecture

All upgrades are built upon open-source libraries and standard API interfaces, promoting modularity, transparency, and reproducibility. The system is designed to remain extensible, enabling further enhancements through integration with cloud services, remote control dashboards, or third-party AI models. These recent developments significantly augment the robot's ability to perceive, understand, and respond to human behavior, pushing the Lynx platform closer to achieving lifelike social interaction and emotional expressiveness.

This section is intended to be appended to the original journal article submitted to TIJER to reflect the progress and continued innovation in the Lynx robotic platform.



CONCLUSION

In this paper, we presented our humanoid communication robot Lynx. Lynx communicates in an intuitive, multimodal way with humans. It employs speech, an animated face, eye-gaze, and gestures to interact with people. Depending on the audio-visual input, our robot shifts its attention between different communication partners in order to involve them into an interaction. Lynx performs human-like arm and head gestures, which are synchronized to the speech synthesis. It generates pointing gestures with its head, eyes, and arms to direct the attention of its communication partners towards objects of interest. Lynx changes its mood according to the number of people around it and the dialog state. The mood is communicated by facial expressions and emotional speech synthesis.

We tested the described multi-modal dialog system during two public demonstrations outside our lab. The experiences made indicate that the users enjoyed interacting with the Lynx. They treated the Lynx as an able communication partner, which was sometimes difficult, as its capabilities are limited. The experienced problems were mainly due to perception deficits of the Lynx. While speech synthesis works fairly well, robust speech recognition in noisy environments is difficult. This is problematic because the users expect the Lynx to understand speech at least as well as it talks. Similarly, while the Lynx is able to generate gestures and emotional facial expressions, its visual perception of the persons around it is limited to head position and size. To reduce this asymmetry between action generation and perception, we are currently working on head posture estimation from the camera images and the visual recognition of gestures.

ACKNOWLEDGMENT

This project is supported by the Vice Industries & Sharkweb IT Park LLP.

Data Availability

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflict of Interest: The authors declare that they have no conflict of interest.

REFERENCES

- [1] C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Mulanda, "Humanoid robots as cooperative partners for people," *Int. Journal of Humanoid Robots*, vol. 1, no. 2, 2004.
- [2] D. Matsui, T. Minato, K. F. MacDorman, and H. Ishiguro, "Generating natural motion in an android by mapping human motion," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2005.
- [3] T. Spexard, A. Haasch, J. Fritsch, and G. Sagerer, "Human-like person tracking with an Anthropomorphic robot," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006.
- [4] M. Bennewitz, F. Faber, D. Joho, S. Schreiber, and S. Behnke, "Towards a humanoid museum guide robot that interacts with multiple persons," in *Proc. of the IEEE/RSJ International Conference on Humanoid Robots (Humanoids)*, 2005.
- [5] H. Okuno, K. Nakadai, and H. Kitano, "Social interaction of humanoid robot based on audio-visual tracking," in *Proc. of the Int. Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE)*, 2002.
- [6] Y. Matsusaka, S. Fujie, and T. Kobayashi, "Modeling of conversational strategy for the robot participating in the group conversation," in *Proc. of the European Conf. on Speech Communication and Technology*, 2001.
- [7] A. Scheidig, S. Muller, and H.-M. Gross, "User-adaptive interaction with social service robots," *KI Themenheft Learning and Self-Organization of Behavior*, no. 3, 2006.
- [8] S. Kopp and I. Wachsmuth, "Model-based animation of coverbal gesture," in *Proc. of Computer Animation*, 2002.
- [9] R. Bischoff and V. Graefe, "Hermes – a versatile personal robot assistant," *Proc. IEEE – Special Issue on Human Interactive Robots for Psychological Enrichment*, vol. 92, no. 11, 2004.
- [10] J. Schulte, C. Rosenberg, and S. Thrun, "Spontaneous short-term interaction with mobile robots in public places," in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 1999.
- [11] I. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, "An affective mobile robot educator with a full-time job," *Artificial Intelligence*, vol. 114, no. 1-2, 1999.
- [12] L. Mayor, B. Jensen, A. Lorotte, and R. Siegwart, "Improving the expressiveness of mobile robots," in *Proc. of IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN)*, 2002.
- [13] Z. Ruttkay, H. Noot, and P. ten Hagen, "Emotion Disc and Emotion Squares: Tools to explore the facial expression space," *Computer Graphics Forum*, vol. 22, no. 1, 2003.
- [14] S. Behnke, M. Schreiber, J. Stuckler, H. Strasdat, and M. Bennewitz, "NimRo TeenSize 2006 team description," in *RoboCup 2006 Humanoid League Team Descriptions*, Bremen, 2006.

- [15] Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, "Multimodal conversation between a humanoid robot and multiple persons," in Proc. of the Workshop on Modular Construction of Humanlike Intelligence at the Twentieth National Conferences on Artificial Intelligence (AAAI), 2005.
- [16] R. Lienhard and J. Maydt, "An extended set of haar-like features for rapid object detection," in Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), 2002.
- [17] H. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, no. 1, 1955.
- [18] R. Kalman, "A new approach to linear filtering and prediction problems," *ASME-Journal of Basic Engineering*, vol. 82, no. March, 1960.
- [19] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," in Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 1985.
- [20] D. Giuliani, M. Omologo, and P. Svaizer, "Talker localization and speech recognition using a microphone array and a cross-powerspectrum phase analysis," in Int. Conf. on Spoken Language Processing (ICSLP), 1994.
- [21] S. Lang, M. Kleinhagenbrock, S. Hohener, J. Fritsch, G. Fink, and G. Sagerer, "Providing the basis for human-robot-interaction: multimodal attention system for a mobile robot," in Proc. of the Int. Conference on Multimodal Interfaces, 2003.
- [22] D. McNeill, *Hand and Mind: What Gestures Reveal about Thought: What Gestures Reveal About Thought*. University of Chicago Press, 1992.
- [23] K. Nickel, E. Seemann, and R. Stiefelwagen, "3D-Tracking of heads and hands for pointing gesture recognition in a human-robot interaction scenario," in International Conference on Face and Gesture Recognition (FG), 2004.
- [24] Loquendo S.p.A., "Vocal technology and services," <http://www.loquendo.com>
- [25] J. Cahn, "Generating expression in synthesized speech," Master's thesis, Massachusetts Institute of Technology, 1989.
- [26] D. Joho, M. Bennewitz, and S. Behnke, "Pitch estimation using models of voiced speech on three levels," in 32nd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2007.
- [27] S. Roa, M. Bennewitz, and S. Behnke, "Fundamental frequency estimation using a pitch tracking algorithm," in 32nd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2007.