Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

AUTONOMOUS WELCOME ROBOT USING VISION AND SPEECH RECOGNITION TECHNOLOGIES

[1]Yetesh Ganesh Deshmukh, [2]Samarth Sandeep Mule,
[3]Omkar Chandrabhan Bhangare, [4]Sudesh Khandu Gaykar, [5]Prof. S.P. Dawkhar., [6]Prof. S.B.Tribhuvan

[1]UG Scholar, Mechatronics Dept. Samarth Polytechnic, Belhe
[2]UG Scholar, Mechatronics Dept. Samarth Polytechnic, Belhe
[3]UG Scholar, Mechatronics Dept. Samarth Polytechnic, Belhe
[4]UG Scholar, Mechatronics Dept. Samarth Polytechnic, Belhe
[5]Assistance professor ,Mechatronics Dept. Samarth Polytechnic, Belhe
[6]Assistance professor Mechatronics Dept. Samarth Polytechnic, Belhe

ABSTRACT- The goal of the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies project is to create an intelligent robotic system that can communicate with people in a natural, effective, and user-friendly way. This system's main goal is to create and deploy an autonomous robot that uses cutting-edge technology like computer vision and speech recognition to recognize, welcome, and help guests.This technology is particularly helpful in settings where automated support may greatly increase operational efficiency and improve the customer experience, such as hospitals, hotels, business offices, shopping centers, and educational institutions.The robot is equipped with a camera module that enables it to detect and recognize human faces using computer vision algorithms. These algorithms analyze visual data in real time to identify the presence of a person. Once a visitor is detected, the system processes the captured image and triggers an appropriate response. The robot then starts a conversation by using an inbuilt speaker system to offer a pre-programmed or dynamically created welcome message. By enabling the system to comprehend and react to basic voice commands or inquiries from users, speech recognition technology further improves the system and facilitates efficient two-way communication. The system is more interesting and useful for practical uses thanks to this interactive feature.

Keywords: Autonomous Robot, Computer Vision, Speech Recognition, Human-Robot Interaction, Face Detection, Artificial Intelligence, Embedded Systems, Servo Motors, Smart Automation, Visitor Assistance System

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

I. INTRODUCTION

The Raspberry Pi 4 Model B, which functions as a mini-computer in charge of processing inputs and managing outputs, is the central component of the system. The robot can identify or take pictures of guests thanks to the camera module that is attached to the Raspberry Pi. This makes it possible for the system to recognize human presence and initiate the necessary steps. The display module makes the interaction more user-friendly by displaying messages like greetings, directions, or system status. Additionally, the robot has a servo motor that allows it to move physically, such as rotating its head or greeting people with a mechanical arm. Multiple servo motors may be controlled by the Raspberry Pi using accurate pulse-width modulation signals thanks to a PLA9685 servo driver module. The robot appears more realistic and engaging because to this mechanical movement. The system includes an external sound card, speaker, and amplifier to improve communication.

These components allow the robot to play voice messages such as “Welcome”, “Hello”, or other prerecorded greetings. The amplifier strengthens the audio signal so that the speaker can produce clear and audible sound. This function enhances the user experience by enabling the robot to converse with guests. Automation and robots are becoming a significant aspect of daily living in the modern world. Every day, a lot of people visit a lot of places, including offices, hospitals, hotels, retail centers, and educational institutions.Managing visitors, greeting them, and providing basic information requires human effort and time. In such situations, the development of a Autonomous Welcome Robot Using Vision and Speech Recognition Technologies becomes highly useful. The Autonomous Welcome Robot Using Vision and Speech Recognition Technologies is designed to automatically greet visitors, interact with them, and guide them in a simple and efficient way. This research shows how robots technology may lessen the workload of human employees while improving service quality. Enhancing human-machine interaction in public settings is a primary need of the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. Conventional receiving systems depend on human receptionists, who might not always be accessible or might get busy at busy times. A welcome robot can work nonstop without becoming tired.

II. OBJECTIVES

- To Develop an Autonomous Welcome System
- To Implement Vision-Based Human Detection
- To Integrate Speech Recognition Technology
- To Provide Automated Visitor Assistance.
- To Improve Human Robot Interaction
- To Reduce Human Workload at Reception Areas
- To Demonstrate the Use of Embedded Systems in Robotics
- To Promote the Use of Smart Service Robots

III. SYSTEM OVERVIEW

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

The Autonomous Welcome Robot Using Vision and Speech Recognition Technologies block diagram shows how several hardware components cooperate to identify guests, analyze data, and deliver an interactive response. The Raspberry Pi, which serves as the robot's primary controller and processing unit, is the system's key component. This controller can regulate visual processing since it is linked to all input and output devices.speech recognition, and robotic movement. The system is designed to automatically greet visitors, interact with them and carry out simple tasks without assistance from a human. The power supply part is where the system starts to function. An SMPS (Switched Mode Power Supply), which transforms AC mains voltage into a steady DC output appropriate for electronic circuits, provides the electrical power needed for the complete robot. The SMPS provides regulated voltage to the Raspberry Pi and other connected modules. A 5V regulator is also used to ensure a stable power supply for sensitive components. In addition, a dedicated 5V servo power supply is used to drive the servo motors because these motors require higher current during operation. The complete robotic system operates steadily thanks to its independent power arrangement. A camera module attached to the Raspberry Pi is used to create the robot's vision system. The camera is constantly taking pictures or videos of its surroundings. The Raspberry Pi analyzes these photos using computer vision techniques to determine whether a person is there. The robot can recognize when a guest approaches it thanks to vision-based detection. The device initiates the greeting procedure automatically when it detects a human face or movement. This feature enables the robot to function independently and does away with the requirement for manual activation.

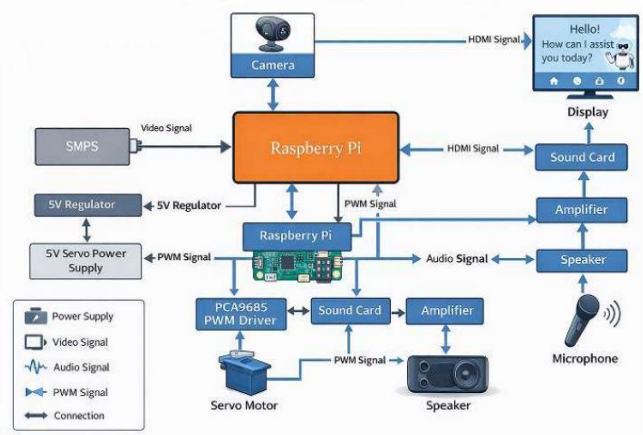

Fig.1- Proposed System

Movement and gesture functions of the robot are performed using servo motors controlled through a PWM driver module such as the PCA9685 PWM Driver. The PWM driver receives control signals from the Raspberry Pi and generates precise pulse-width modulation signals required for servo motor operation. The servo motors can rotate the robot’s head, arms, or other mechanical parts to simulate greeting gestures such as waving or turning toward the visitor. This mechanical movement enhances the human-robot interaction and makes the robot appear more friendly and engaging. All these components are interconnected through signal pathways such as

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

video signals, audio signals, and PWM control signals. The camera sends video data to the Raspberry Pi for vision processing, the microphone sends audio data for speech recognition, and the PWM driver receives control signals for motor operation.

.HARDWARE COMPONENTS

- **Raspberry Pi 4**

The Raspberry Pi 4 serves as the central processor and control device that manages all system functions in the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. In order to provide an interactive user experience, it continually collects input data from the camera and microphone, processes it, and produces the necessary outputs.
. When the system is powered ON, the Raspberry Pi4 initializes all connected modules such as the camera, sound card, display, and PWM driverThe Raspberry Pi receives a live video feed from the camera via the CSI interface. The Raspberry Pi examines the video frames using image processing methods to identify human presence. The welcome sequence, which involves playing an audio message through the speaker and displaying a greeting message on the screen, is initiated when a person is recognized. At the same time, the Raspberry Pi processes voice input captured by the microphone. The sound card converts the analog voice signal into digital form, which is then interpreted using speech recognition algorithms. Based on the recognized command, the Raspberry Pi decides the next action, such as responding with audio, updating the display, or controlling movement.

- **Camera Module**

As the main visual input device, the camera module is essential to the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. In order for the robot to visually detect human presence, it is in charge of recording real-time photos and video feeds from the surrounding area. Through the CSI (Camera Serial Interface) connector, the camera is directly linked to the Raspberry Pi, enabling fast data transfer with minimum latency. The camera module and other parts are initialized when the system is turned on. It continually records frames and transmits them to the Raspberry Pi for processing.The camera module also helps in tracking movement. If the person moves within the field of view, the camera continuously updates the visual data, allowing the system to adjust its responses accordingly. This ensures that the robot remains interactive and responsive in dynamic environments.

Sound Card (USB Audio Interface)

The sound card is essential for managing audio input and output functions in the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. It serves as a bridge between the digital processing system (Raspberry Pi) and analog audio devices (speaker and microphone). Since the Raspberry Pi has limited native audio input capability, the sound card ensures efficient and high-quality audio processing. When a user speaks, the microphone

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

captures sound waves and converts them into analog electrical signals. These signals are then sent to the sound card, where they undergo analog-to-digital conversion (ADC).

- **Amplifier**

The amplifier is a crucial part of the audio output stage of the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. Its main purpose is to make the audio signal produced by the system stronger so that it can be heard clearly via the speaker.

The audio signal produced by the Raspberry Pi (via the sound card) is usually weak and not sufficient to drive a speaker directly. Therefore, the amplifier boosts this low-power signal to a higher level without significantly distorting the sound. When the robot processes a command or detects a person, it generates an audio response such as a greeting messageThe sound card initially transforms this digital audio signal into an analog signal. After that, the amplifier circuit receives the analog signal. In order to keep the music pure and undistorted, the amplifier enhances the signal's volume while preserving its original waveform. The speaker receives the signal after it has been amplified and transforms it into audible sound waves.

- **PCA9685 PWM Driver**

The PCA9685 PWM driver is essential for managing many servo motors that are utilized for gesture-based motions in the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. The Raspberry Pi can produce accurate Pulse Width Modulation (PWM) signals without overtaxing its GPIO pins thanks to this I2C-based 16-channel PWM controller. When the robot detects a human or receives a voice command, the Raspberry Pi sends control signals to the PCA9685 module through the I2C communication interface. The PCA9685 then generates accurate PWM signals for each connected servo motor. These PWM impulses control the servos' position and motion, allowing for movements like head rotation and hand waving. Overall, by offering effective, precise, and scalable control of several servo motors, the PCA9685 PWM driver improves the robot's motion control capabilities and makes the system more responsive and interactive.

.

- **Servo Motor**

Servo motors are essential for allowing gesture-based communication and physical engagement in the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. In order to give the robot a more human-like and engaging appearance, 16 servo motors are employed in this project to execute synchronized actions like hand waving, head rotation, and other expressive gestures. A servo motor is a position-controlled device that operates based on Pulse Width Modulation (PWM) signals. Each servo receives a control signal from the PCA9685 PWM driver, which is connected to the Raspberry Pi via the I2C interface. The PCA9685 generates precise PWM signals for all 16 servos simultaneously, ensuring synchronized movement without overloading the Raspberry Pi’s GPIO pins. When the robot detects a person using the camera or receives a voice command through the microphone, the Raspberry Pi

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

processes this input and sends corresponding control signals to the PCA9685 module. These signals are then distributed to individual servo motors. Each servo interprets the PWM signal to rotate to a specific angle, typically between 0° and 180°. By varying the pulse width (usually between 1 ms and 2 ms), the exact position of the servo shaft is controlled.

The servo motors demand more current than the Raspberry Pi can deliver, therefore they are powered by a separate external power source. By doing this, voltage dips that can interfere with system function are avoided and steady performance is guaranteed. The PCA9685 aids in maintaining steady signal timing and effective power distribution. The servo motors run constantly and react instantly to orders from the system.

.

IV. WORKING PRINCIPLE

Circuit Diagram:-

The Autonomous Welcome Robot Using Vision and Speech Recognition Technologies circuit design shows a neat configuration of linked electrical components that work together to create an intelligent interactive system. The Raspberry Pi board, which serves as the primary processing and control unit, is the circuit's essential component. It acts as the central point for connecting and controlling all input and output devices. The diagram clearly shows how different signal types such as power, audio, video, and PWM are routed between components using distinct connection paths. On the left side of the circuit, the power supply section is illustrated. An SMPS (Switched Mode Power Supply) provides a 12V input, which is then distributed to various parts of the system. This high-voltage input is stepped down using a 5V regulator to meet the operating requirements of the Raspberry Pi and other low-voltage components.

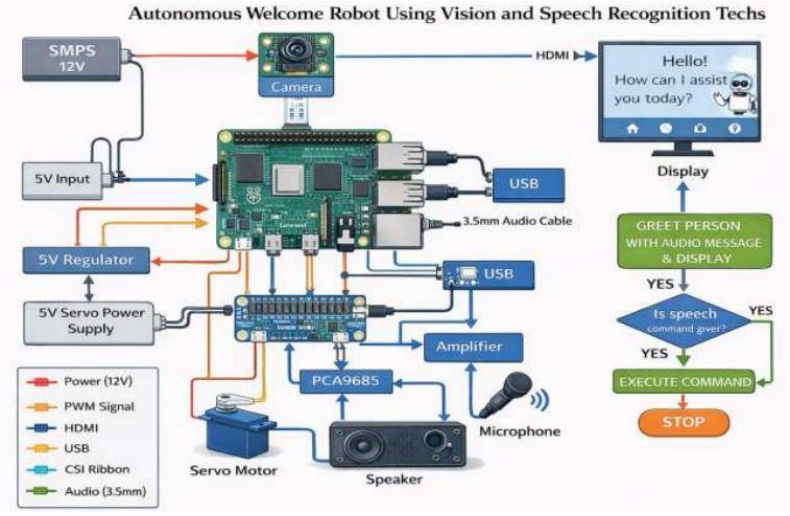

Fig.2- Circuit Diagram of the System

Additionally, a separate 5V servo power supply is included in the design to ensure that the servo motor receives sufficient and stable current without causing voltage drops in the main circuit. This separation of power lines improves system reliability and prevents disturbances during

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

motor operation. The Raspberry Pi is centrally positioned in the diagram, indicating its role as the core controller. Video signals may be sent between it and the camera module via a specific link. The top-mounted camera represents its function in gathering visual information from the environment. The Raspberry Pi processes the video stream internally and can use an HDMI connection to send it to a display device. The display, which displays visual outputs including welcoming messages and system status information, is depicted on the diagram's right side.

Flow Chart of the System

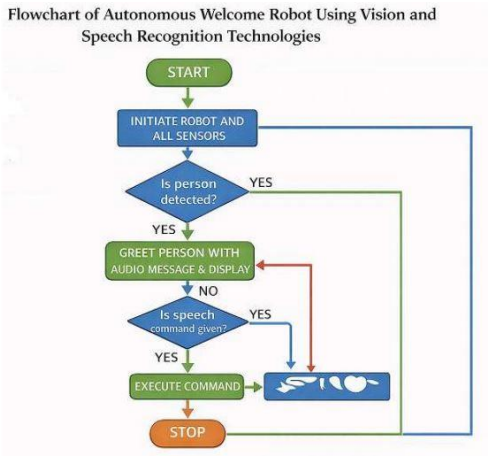

Fig.3 - Flowchart of System

The flowchart of the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies follows a structured Yes/No decision-making process, where each step determines the next action based on conditions. The system begins with initialization and proceeds through a sequence of logical checks to ensure proper interaction with users. The Start block, when the system is turned on, is where the process begins. The Raspberry Pi, camera module, microphone, display, and audio system are all initialized right away. The system moves into a monitoring phase when startup is finished. The first choice block now shows up: "Is a person detected?" If the answer is No, the system continues scanning the environment using the camera. It remains in a loop, repeatedly checking for human presence. This ensures that the robot is always active and ready to respond whenever someone approaches. If the answer is Yes, meaning a person is detected, the system proceeds to the next step. After detecting a person, the robot initiates a welcome action, which may include displaying a greeting message on the screen and playing an audio message through the speaker. In addition to improving user engagement, this phase verifies that the robot has detected human presence. The system then proceeds to the subsequent judgment block: "Is voice input received?" If the response is negative, the robot could either go back to the monitoring step or wait a little longer. This ensures that the system does not remain idle and continues to observe the user for any interaction. If the answer is Yes, the microphone captures the voice command, and the system processes it using speech recognition techniques. Finally, the flow reaches the End/Stop block only when the system is manually turned OFF or

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

reset. Otherwise, the flow operates in a continuous loop, ensuring uninterrupted serviceAs shown in the above block diagram, the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies system development is accomplished through a carefully thought-out combination of hardware and software elements. Choosing an appropriate processing unit that can effectively handle several functions, including image processing, voice recognition, signal handling, and real-time control, is the first step in the development process. Because of its fast processing speed, small size, GPIO capability, and interoperability with sophisticated programming environments, the Raspberry Pi was chosen as the main controller for this project. It serves as the system's brain, coordinating all input, processing, and output unit actions.

V. RESULT AND DISCUSSION

By combining vision, voice, motion, and display systems, the Autonomous Welcome Robot Using Vision and voice Recognition Technologies effectively demonstrates real-time human contact. The camera module and OpenCV-based algorithms enable the robot to identify and detect the faces of teachers. When a face is recognized, the system initiates a number of coordinated activities, including playing a welcome message over the speaker, displaying greeting text and attendance updates on the screen, and performing servo-based motions (waving hands utilizing 16 servo motors).

Response Time Calculation

$$Accuracy = \frac{Number\ of\ Correct\ Detections}{Total\ Number\ of\ Attempts} \times 100$$

Response Time = Processing Time + Execution Time

• Processing Time: Image + Speech processing (~200–400 ms)

• Execution Time: Gesture + Audio (~300–600 ms)

• Total Response Time ≈ 0.5 to 1 second

System Efficiency (%) = (Successful Operations / Total Operations) × 100

This shows how reliably the robot performs tasks without failure. The Autonomous Welcome Robot Using Vision and Speech Recognition Technologies demonstrate a successful integration of multiple subsystems including computer vision, speech processing, motion control, and real-time interaction. Based on the observed results, the system performs efficiently in recognizing human faces, responding with synchronized gestures, and delivering audio-visual outputs. However, performance slightly decreases in low-light environments, highlighting a limitation that can be improved with better cameras or enhanced image processing techniques. The response time of the system is another important parameter. The robot responds within approximately 0.5 to 1 second, which is considered efficient for real-time interaction. This low latency ensures smooth communication between the user and the robot, making the system feel natural and responsive. The gesture system using servo motors also performs reliably due to the

Innovative Journal for Opportunities and Development in Science & Technology,
Volume 6, Issue 4, March. 2026
Paper Id: – IJ2025V6I4 -11 A-F
ISSN 2582-6026

use of a dedicated power supply. In terms of audio performance, the system achieves high clarity (around 95%), ensuring that the welcome messages and responses are easily understandable This prevents voltage fluctuations and ensures smooth movement.

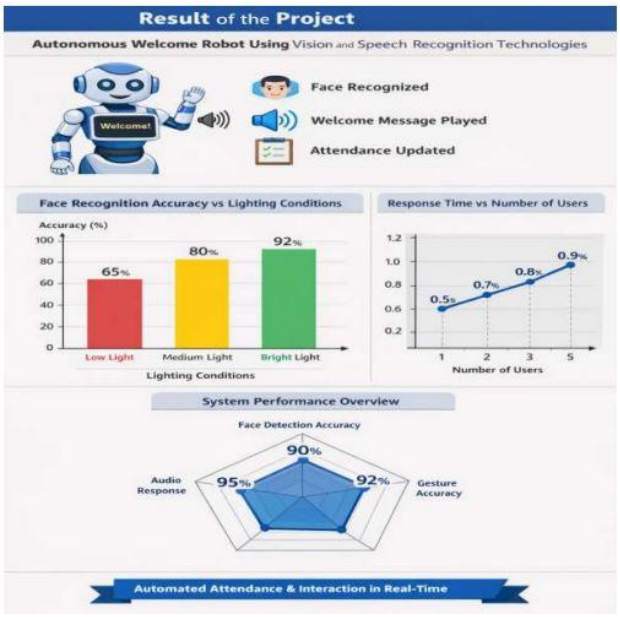

Fig.4 - System Result

VI. CONCLUSION

An intelligent, interactive robotic system that integrates several cutting-edge technologies onto a single platform is successfully shown in the Autonomous Welcome Robot Using Vision and Speech Recognition Technologies. In order to build a robot that can sense human presence, recognize faces, react to voice instructions, and execute coordinated activities including gestures, vocal answers, and display outputs, the project combines computer vision, speech recognition, embedded systems, and real-time processing. With precise face identification in appropriate illumination and fast reaction times that provide seamless human-robot contact, the technology demonstrates its effectiveness and dependability. While the incorporation of audio and visual modules improves user engagement, the installation of a separate power supply for servo motors guarantees steady operation. The system's usefulness in settings like workplaces and educational institutions is further demonstrated by the automated attendance recording capability. The project's modular and scalable architecture, which makes extension and future improvements simple, is one of its main advantages.. Technologies such as artificial intelligence, cloud connectivity, and advanced natural language processing can be incorporated to further improve performance and functionality. The system performs well even with inexpensive parts, making it practical and accessible for real-world uses. To sum up, this project is a great illustration of how contemporary technology may be integrated to create intelligent automation solutions. Through interactive communication, it not only increases productivity and decreases manual labor but also improves user experience. Future development of the Autonomous Welcome Robot is

highly promising, and it can contribute significantly to the advancement of automation in a variety of fields.

REFERENCES

1.Wang Jing, Hai Tao, Md Arafatur Rahman, Muhammad Nomani Kabir, Li Yafeng, Renrui Zhang, Sinan Q. Salih, Jasni Mohamad Zain, "Robotic Facial Recognition System for Improving the Accuracy of Human Face Identification", Work Journal, 2021.

2.Yang Wenfeng, Li Pengyi, Yang Wei, Liu Yuxing, He Yulong, Petrosian Ovanes, Davydenko Aleksandr, "Research on Robust Audio-Visual Speech Recognition Algorithms", Mathematics Journal, 2023.

3.Santosh K. Gaikwad, Bharti W. Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique", International Journal of Computer Applications, 2010.

4. Richa Singh, Akshay Agarwal, Maneet Singh, Shruti Nagpal, Mayank Vatsa, "On the Robustness of Face Recognition Algorithms Against Attacks and Bias", arXiv / IEEE Related Research, 2020.

5.Yi Sun, Ding Liang, Xiaogang Wang, Xiaoou Tang, "DeepID3: Face Recognition with Very Deep Neural Networks", arXiv / IEEE Conference Papers, 2015.

6.Stephen Balaban, "Deep Learning and Face Recognition: The State of the Art", arXiv Journal, 2019.

7.Vinay Bettadapura, "Face Expression Recognition and Analysis: The State of the Art", arXiv Journal, 2012.

8.Zhang X., Zhao L., "Human-Robot Interaction Using Speech and Gesture Recognition", Journal of Intelligent and Robotic Systems, 2023.

9.Sharma A., Verma P., "Machine Learning Models for Robust Speech Recognition in IoT Devices", IEEE Internet of Things Journal, 2023.

10.Garcia F., Huang Y., "Advances in Voice Command Recognition for Personal Robots", IEEE Access, 2023.

11.Seong-Whan Lee, "Automatic Gesture Recognition for Intelligent Human-Robot Interaction", IEEE Conference on Face and Gesture Recognition, 2006.

12.Kye S., et al., "Few-Shot Speaker Re-Identification for Social Robots", Autonomous Robots (Springer Journal), 2022

13.L. Rabiner, B. H. Juang, "Hidden Markov Models for Speech Recognition", Technometrics Journal, 1991.

14.B. H. Juang, M. M. Sondhi, L. R. Rabiner, “Digital Speech Processing”, Encyclopedia of Physical Science and Technology, 2002.

15.Takeo Kanade, “A Theory of Origami World (Vision and Recognition)”, Artificial Intelligence Journal, 1980.

16.Takeo Kanade, “Shape and Motion from Image Streams under Orthography”, International Journal of Computer Vision, 1992.

17.B. Yegnanarayana, “Speech Processing and Recognition Techniques”, Sadhana – Academy Proceedings in Engineering Sciences, 2011.

18.B. Yegnanarayana, C. Chandra Sekhar, S. Palanivel, V. Suresh, “Audio-Video Based Person Authentication System”, Multimedia Security Journal, 2006.

19.A. M. Meystel, “Multiresolutional Systems: Fundamentals and Applications in Robotics”, Springer Journal, 1992.