

# AI-Based Multimodal Communication App for People with Disabilities

**Om Mane, Om Ambekar, Shruti Gharpankar, Sneha Kupate, Madhura Kadam, Prof. V. B. Koli**

Department of Computer Science and Engineering  
Ashokrao Mane Group of Institutions, Vathar Tarf Vadgaon  
Dr. Babasaheb Ambedkar Technological University, Lonere, India

**Abstract:** *People with speech and motor disabilities often face difficulties in expressing their needs and communicating effectively in daily life. Existing assistive communication systems are expensive, hardware dependent, and inaccessible for many users. This paper presents an AI-based multimodal communication application that combines text-to-speech conversion, webcam-based eye tracking, and an accessible user interface to support real-time communication. The system uses low-cost hardware and open-source technologies such as OpenCV and MediaPipe to provide an affordable and efficient assistive solution. The application offers typing mode, gaze-controlled interaction, customizable accessibility settings, and natural speech generation to improve user independence and communication efficiency. Experimental testing demonstrates reliable gaze detection, smooth text-to-speech output, and responsive interaction under standard conditions. The proposed system aims to promote inclusivity and improve the quality of life for individuals with disabilities.*

**Keywords:** Assistive Technology, Eye Tracking, Text-to-Speech, Accessibility, Artificial Intelligence, Human Computer Interaction, OpenCV, MediaPipe

## I. INTRODUCTION

Communication is one of the most essential human needs, allowing individuals to express emotions, ideas, and daily requirements. However, people suffering from speech and motor disabilities such as Amyotrophic Lateral Sclerosis (ALS), cerebral palsy, Parkinson's disease, and stroke often face major difficulties in communicating effectively.

These communication barriers can lead to social isolation, dependency on caregivers, and reduced participation in educational, professional, and social activities.

Existing assistive communication systems are often expensive, hardware-dependent, and difficult to access for users in low-resource environments. Many commercial systems require specialized eyetracking devices or dedicated hardware, making them unaffordable for a large number of disabled individuals. In addition, several existing applications support only a single communication method and lack flexibility for users with different levels of impairment.

To address these challenges, this paper proposes an AI-based multimodal communication application for people with disabilities. The system combines text-to-speech conversion, webcam-based eye tracking, and a user-friendly interface to provide accessible and real-time communication support. The application uses low-cost hardware such as standard webcams and laptops while integrating technologies including OpenCV, MediaPipe, and neural text-to-speech systems. The proposed system offers multiple communication modes including typing-based interaction and gaze-controlled virtual keyboard navigation. Accessibility features such as large buttons, high-contrast themes, customizable settings, and quick-access phrases improve usability for users with limited motor abilities. The application is designed to provide affordable, scalable, and inclusive communication assistance that enhances independence and improves the quality of life for people with disabilities.



**II. LITERATURE SURVEY**

Several assistive communication systems have been developed to support individuals with speech and mobility impairments.

Augmentative and Alternative Communication (AAC) devices such as Tobii Dynavox and Lingraphica provide communication support using eye tracking and text-to-speech technologies. Although these systems offer high accuracy and effective interaction, they are expensive and dependent on specialized hardware.

Mobile-based AAC applications such as Proloquo2Go and Avaz provide portable communication support through visual symbols and speech synthesis. However, many of these applications are subscription-based and mainly focus on speech impairments without supporting advanced gaze interaction or multimodal communication.

Eye-tracking systems including Gaze Speak and Eye Speak have demonstrated effective gaze-controlled communication. These systems use camera-based eye movement detection to help users select characters or phrases on a virtual keyboard. However, many existing solutions require infrared cameras or dedicated tracking hardware, increasing implementation cost and reducing accessibility.

Recent advancements in neural text-to-speech systems and artificial intelligence have improved speech synthesis quality and interaction efficiency. Technologies such as OpenCV, MediaPipe, and neural TTS models enable the development of low-cost and realtime assistive communication systems. The proposed system combines these technologies to provide an affordable, accessible, and multimodal communication platform for individuals with disabilities.

**III. SYSTEM ARCHITECTURE AND METHODOLOGY**

The architecture of the proposed communication application is divided into three major layers: user interface layer, application processing layer, and backend service layer.

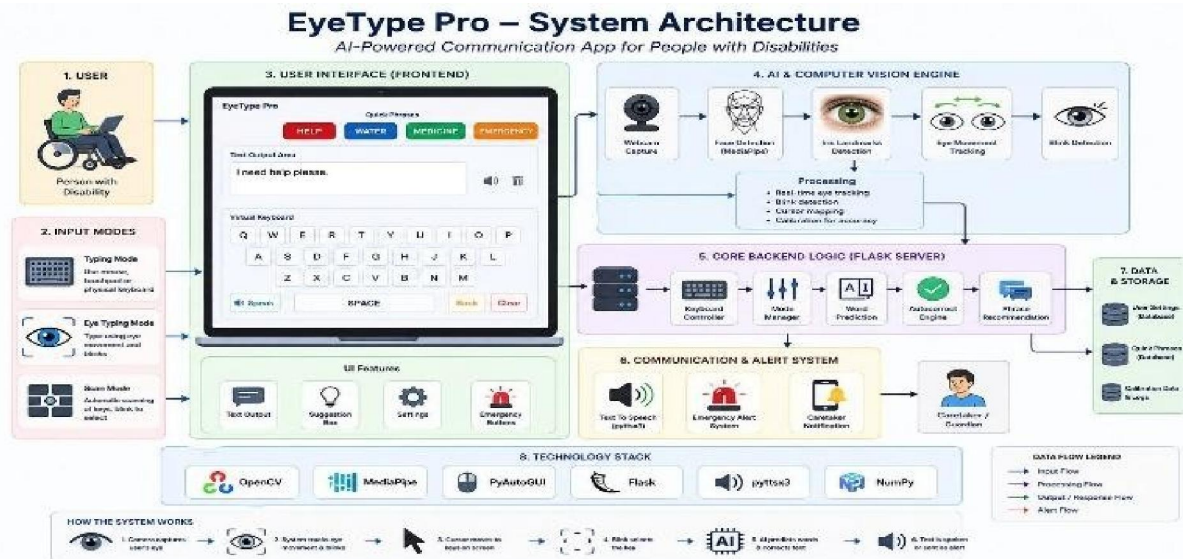


Fig. 1. System Architecture of Proposed Communication Application

The system is designed to provide real-time communication support through multimodal interaction methods including typing and eye tracking.

The user interface layer provides accessible interaction components such as typing mode, eye-tracking mode, virtual keyboard, text display area, and speech controls. The interface is designed using large buttons, high-contrast themes, and simplified navigation to improve usability for disabled users.



The application processing layer handles eye movement detection, text processing, and speech generation. The webcam continuously captures user eye movements, and OpenCV with MediaPipe libraries process gaze coordinates and facial landmarks. The detected gaze position is mapped to interface elements for virtual keyboard selection and message generation.

The text-to-speech module converts typed or selected text into natural speech output. The system also supports predefined phrases and customizable communication shortcuts for faster interaction. Backend services manage user settings, communication requests, and system responses to ensure smooth real-time operation. The complete workflow of the system begins with user input through typing or gaze interaction. The processed input is converted into text and then transformed into speech using the text-to-speech engine. The generated audio output is played through the device speakers to assist communication in real time.

#### IV. IMPLEMENTATION

The proposed communication system was implemented using open-source technologies and low-cost hardware components. The frontend interface was designed to provide an accessible and user-friendly environment for individuals with speech and motor disabilities. Large buttons, simplified controls, and highcontrast themes were incorporated to improve usability and interaction.



Fig. 2. User Interface of Communication Application

The eye-tracking module was developed using OpenCV and MediaPipe libraries. The webcam captures real-time video frames, and facial landmarks are extracted to estimate gaze direction and eye movement. The detected gaze coordinates are mapped to interface elements to allow users to interact with the virtual keyboard and communication controls without physical touch.

The text-to-speech module converts user-generated text into speech output using neural speech synthesis technology. Users can type messages manually or generate text through gaze-controlled selection. The generated speech output helps users communicate effectively with caregivers, family members, and other individuals.

The system was tested under normal lighting conditions using a standard webcam and laptop device. Experimental results demonstrated smooth communication flow, responsive gaze detection, and clear speech generation with minimal latency.

The proposed communication application was successfully tested for typing-based communication, eye-tracking interaction, and text-to-speech generation. The system demonstrated reliable performance under standard environmental conditions using low-cost hardware devices.



**V. RESULTS AND DISCUSSION**  
TABLE I: SYSTEM TEST RESULTS

Test Case	Expected Result	Status
Webcam Detection	Webcam initialized	Pass
Eye Tracking Calibration	Accurate gaze calibration	Pass
Text Input	Text displayed correctly	Pass
Text-to-Speech	Natural speech generated	Pass
UI Accessibility	High contrast visibility improved	Pass
System Performance	Stable operation without lag	Pass

The eye-tracking module accurately detected user gaze movement and allowed users to select interface elements using webcam-based interaction. The virtual keyboard and gazecontrolled selection mechanism enabled users with limited mobility to generate messages efficiently. Text-to-speech conversion produced clear and understandable speech output with low response delay.

System testing was performed for various functionalities including webcam detection, gaze calibration, typing input, speech generation, and interface accessibility. The experimental results showed stable system performance without significant lag or crashes during continuous operation.

The proposed solution provides an affordable and accessible assistive communication platform compared to expensive commercial assistive devices. The integration of multiple communication modes improves flexibility and usability for individuals with different levels of disability.

## VI. CONCLUSION

The proposed AI-based multimodal communication application provides an effective, accessible, and affordable communication solution for individuals with speech and motor disabilities. By integrating webcam-based eye tracking, text-to-speech conversion, and an accessible user interface, the system enables users to communicate more independently and efficiently.

The application successfully demonstrates the use of low-cost hardware and open-source technologies to build an inclusive assistive communication platform. Features such as gazecontrolled interaction, virtual keyboard navigation, customizable accessibility settings, and natural speech output improve usability and user experience for individuals with varying levels of disability.

Experimental testing showed reliable eyetracking performance, responsive interaction, and smooth speech generation under standard operating conditions. The proposed system reduces dependency on expensive assistive communication devices while improving accessibility and social inclusion for disabled users.

Overall, the project highlights the potential of artificial intelligence and computer vision technologies in developing meaningful assistive solutions that enhance communication, independence, and quality of life for people with disabilities.

## VII. FUTURE SCOPE

Future improvements to the proposed communication application can further enhance accessibility, accuracy, and user experience. Advanced deep-learning-based eye-tracking models can be integrated to improve gaze detection performance under varying lighting conditions and for users with limited eye movement.

The system can also be extended with offline text-to-speech and speech recognition models to reduce



dependency on internet connectivity. Support for multiple regional languages and customizable voice profiles can improve accessibility for diverse users.

Additional assistive technologies such as gesture recognition, brain-computer interface integration, and AI-based contextual message prediction can further improve communication efficiency. Future versions of the system may also include cloud synchronization, mobile platform optimization, and real-time video communication support for broader usability.

#### REFERENCES

- [1]. K. Chemnad and A. Othman, "Advancements in Arabic Text-to-Speech Systems: A 22 Year Literature Review," IEEE Access, vol. 11, pp. 1–22, 2023.
- [2]. H. Yamashita et al., "Fast Neural Speech Waveform Generative Models With Fully Connected Layer-Based Upsampling," IEEE Access, vol. 12, pp. 1–15, 2024.
- [3]. H. O. Edughele et al., "Eye-Tracking Assistive Technologies for Individuals With Amyotrophic Lateral Sclerosis," IEEE Access, vol. 10, pp. 1–20, 2022.
- [4]. Y. Zhang et al., "Uncertainty Inspired Early Autism Spectrum Disorder Screening via Contrastive Image-Viewing Paradigm," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 33, pp. 1–12, 2025.
- [5]. F. Jiang et al., "M4SC: An MLLM Based Multi-Modal, Multi-Task and Multi-User Semantic Communication System," IEEE Wireless Communications, vol. 32, no. 5, pp. 1–9, 2025.
- [6]. T. T. Win et al., "A Survey on Semantic Communications System Based on Multi-Modal Data," in Proc. 2024 International Conference on Information Networking (ICOIN), 2024, pp. 1–6.

