RN SHETTY TRUST®
RNS INSTITUTE OF TECHNOLOGY
Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi
Approved By AICTE, New Delhi. Accredited by NAAC 'A+' Grade
Channasandra, Dr. Vishnuvardhan Road, Bengaluru - 560 098
Ph:(080)28611880,28611881 URL: www.rnsit.ac.in
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

MINI PROJECT SYNOPSIS

Subject Name	**Mini Project Work**	**Subject Code**	**BCSP586**
Student Name	Deepthi DR	**USN**	1RN23CS068
	Dhruthi M Sathish		1RN23CS074
	Isiri H S		1RN23CS090
	Likitha M N		1RN23CS111
Domain	Multimodal Assistive Computing	**Group No:**	B22
Project Title	Eye-Tracking and Voice-Activated Interface: Real-Time Interaction with Alexa/Gemini Integration		
Guide Name	Dr. Bhavanishankar S		

ABSTRACT:

In today’s digital age, seamless interaction between humans and computers is essential.
Traditional input devices like the mouse and keyboard can be inaccessible for individuals with physical disabilities, limiting their ability to interact with computers effectively. This creates barriers in communication, productivity, and access to digital resources. To overcome these challenges, alternative control methods such as eye-tracking and voice recognition offer promising solutions. This project proposes an innovative solution that integrates eye-tracking for cursor control and voice recognition for executing commands, enabling.
This project aims to develop an eye-controlled mouse system enhanced with voice command capabilities, allowing users to perform cursor movements and execute commands without physical contact.
This project proposes an innovative solution that integrates eye-tracking for cursor control and voice recognition for executing commands, enabling seamless performance.
The solution aims to enhance digital accessibility for users with mobility impairments and also finds relevance in hands-free computing environments, gaming, and smart interfaces, making it relevant beyond assistive technology.

OBJECTIVES:

- To develop a hands-free computer control system using eye-tracking and voice recognition.
- To assist individuals with physical disabilities in accessing digital devices more independently.
- To integrate computer vision techniques (e.g., OpenCV) for accurate eye movement detection.
- To implement voice command functionality using speech recognition libraries for system control.
- To explore real-time gaze tracking for precise cursor movement and selection.
- To create a user-friendly interface that enables smooth navigation and command execution.
- To enhance accessibility and promote inclusive design of technology.
- To explore potential applications in gaming, virtual reality, and remote control systems.

INTRODUCTION ABOUT THE DOMAIN:

Domain Overview:

The proposed system belongs to the evolving field of **Human–Computer Interaction (HCI)**, with a strong focus on **multimodal interaction**—where multiple input methods are combined to create a more natural, efficient, and inclusive user experience.

At its core, the system integrates **eye-tracking technology** for precise cursor control with **voice recognition** for executing commands, creating a seamless, hands-free interface. This positions the concept firmly within **Assistive Technology** for accessibility, while also appealing to broader productivity, gaming, and extended reality (XR) domains.

Primary Research and Innovation Areas

1. **Human–Computer Interaction (HCI)** – The design, evaluation, and implementation of interactive computing systems for human use. Eye gaze input is already a known modality, but its combination with high-accuracy voice commands and real-time conversational AI sets this apart.

2. **Assistive Technology** – Supporting users with motor impairments or conditions like ALS, cerebral palsy, or spinal cord injury, where traditional mouse and keyboard input are challenging.

3. **Multimodal Interaction** – Integration of gaze tracking for pointing and speech for action execution creates a more fluid interface than either modality alone.

4. **Edge AI and Embedded Systems** – Running gaze detection, wake-word spotting, and basic NLP locally reduces latency, ensuring responsiveness without relying entirely on the cloud.

5. **Conversational AI Integration** – By connecting to APIs like Alexa or Gemini, the interface can go beyond simple commands (“click”, “scroll”) to intelligent dialogue, contextual searches, and application control.

Proposed System Capabilities

- **Eye-Guided Cursor Movement**: Infrared or camera-based tracking to follow the user's gaze, replacing the need for hand-operated devices.
- **Voice-Driven Actions**: Simple commands ("click", "drag", "scroll down") and complex tasks ("open my email and compose a message to John").
- **AI-Enhanced Interaction**: Integration with Alexa or Gemini enables context-aware queries ("Summarize this document I'm viewing"), app automation, and cross-platform coordination.
- **Customizable Command Mapping**: Users can define unique voice commands for personal workflows.
- **Adaptive Calibration**: The system learns over time, adjusting gaze sensitivity and speech recognition to the individual user.

Potential Applications Beyond Accessibility

- **Healthcare**: Allowing surgeons or radiologists to interact with on-screen data without touching surfaces.
- **Creative Industries**: Designers, video editors, or musicians controlling tools without breaking creative flow.
- **Industrial Environments**: Workers in gloves or hazardous zones operating machines hands-free.
- **Gaming and XR**: More immersive control in virtual and augmented reality environments.

Market and Societal Impact
This technology addresses a **dual market**:

- **Accessibility-first**: Millions of individuals globally require alternative input methods.
- **Productivity and Innovation**: Even able-bodied users can benefit from a faster, more intuitive way to control systems in high-demand workflows.

REQUIREMENT SPECIFICATION

1. Programming Languages

- **Python 3.10+** – Primary language for eye-tracking algorithms, voice recognition, and AI integration.
- **Tailwind CSS** – For styling the frontend with a utility-first approach for quicker and more efficient design.
- **React** – For building the **dynamic frontend** of the web app. It allows smooth rendering and state management, and will be integrated with the **Flask backend**.
- **Flask** – Lightweight **Python web framework** used to build the **backend** of the app, handling API calls, ML integration, and database management.
- **Socket.io (for Flask)** – For **real-time communication** between the frontend and backend (e.g., enabling or disabling specific features like **eye control** or voice recognition).

2. Eye Tracking Libraries & Tools

- **OpenCV –** For image capture, gaze estimation, and pupil detection.
- **Mediapipe** – Google's lightweight framework for real-time face and eye landmark detection.
- **PyGaze** – Specialized Python libraries for cursor mapping from eye movements.

3. Voice Recognition & NLP Tools

- **SpeechRecognition (Python) –** Simple speech-to-text engine with support for Google Cloud Speech API.
- **Vosk API** – Offline speech recognition to reduce latency.
- **Wake Word Engine (Porcupine / Snowboy)** – For always-on voice command activation without excessive power usage.

4. AI Integration APIs

- **Google Gemini API** – For advanced conversational AI and context-aware responses.
- Alexa API -
  Reason: Allows natural, real-time communication with the system and smart device integration.

5. Development Tools

- **Visual Studio Code** – Primary code editor with debugging and Git integration.
- **Git** – Version control for managing collaborative development.
- **Gemini API -** Intractability between the user and the interface.

6. Optional GUI Frameworks

- **Tkinter / PyQt5** – For local configuration and calibration UI.
- **Electron.js** – For a cross-platform desktop control application.

7. Testing & Deployment Tools

- **PyInstaller** – To package the Python code into an executable for end-users.
- **Docker –** For creating portable deployment environments.

BIBLIOGRAPHY

1. Majaranta, P., & Bulling, A. (2014). Eye tracking and eye-based human–computer interaction. In R. B. Goldberg & A. G. Andrew (Eds.), Advances in Physiological Computing (pp. 39–65). Springer. https://doi.org/10.1007/978-1-4471-6392-3_3

2. Wobbrock, J. O., & Kientz, J. A. (2016). Research contributions in human–computer interaction. Interactions, 23(3), 38–44. https://doi.org/10.1145/2907069

3. Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). Human–Computer Interaction (3rd ed.). Pearson Education.

4. Patel, D., & Patel, A. (2019). Voice Recognition Systems: A Review. International Journal of Computer Applications, 177(28), 1–5. https://doi.org/10.5120/ijca2019919392

5. Sharma, G., & Rajoria, S. (2021). Voice-controlled and eye-tracking-based computer navigation system for physically challenged individuals. International Journal of Advanced Computer Science and Applications (IJACSA), 12(4), 421–428. https://doi.org/10.14569/IJACSA.2021.0120450

6. Gajos, K. Z., Reinecke, K., & Herrmann, C. (2017). Accessibility and assistive technology. In A. Sears & J. A. Jacko (Eds.), The Human-Computer Interaction Handbook (pp. 901–922). CRC Press.

7. Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). Eye tracking: A comprehensive guide to methods and measures. Oxford University Press.

8. Rabiner, L. R., & Juang, B. H. (1993). Fundamentals of speech recognition. Prentice-Hall.

9. Bigham, J. P., Ladner, R. E., & Borodin, Y. (2011). The design of human-powered access technology. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1333–1342). https://doi.org/10.1145/1978942.1979139

10. IEEE. (2020). IEEE Standard for Accessible and Inclusive User Interfaces. IEEE Std 2885-2020. https://doi.org/10.1109/IEEESTD.2020.9090210

Signature of Guide **Mini Project Coordinator**