IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Gaze – Based Mouse Control System

Sheema Sk [1], Bhavya sri D [2], Abhinaya sri B [3],S.S.S.S.L.N.R.V.S.Santhosh Kumar P [4] Mr. Veduri Veera Prasad [5]
B.Tech-CSE Student, [1] B.Tech-CSE Student, [2] B.Tech-CSE Student,[3] B.Tech-CSE Student [4]Assistant Professor[5]
[1,2,3,4,5] Department of Computer Science and Engineering,
[1,2,3,4,5] Aditya College of Engineering and Technology, Surampalem, Andhra Pradesh, India.

Abstract: In present digital world, human–computer interaction plays a vital role in improving accessibility and user experience. However, individuals with physical disabilities or limited motor abilities often face challenges in operating traditional input devices such as a mouse and keyboard. To address this issue, this project presents a gaze-based mouse control system integrated with a voice assistant, enabling hands-free computer interaction. The system uses computer vision techniques to track eye movements through a webcam and translate gaze direction into cursor movement on the screen. Blinking patterns or dwell-time mechanisms are used to perform mouse click actions. In addition, a voice assistant powered by speech recognition and natural language processing allows users to execute commands, open applications, browse the web, and perform system operations through voice instructions. The integration of gaze tracking and voice control creates an efficient, accessible, and user-friendly interface. The proposed system aims to enhance independence, productivity, and digital accessibility for users, particularly individuals with motor impairments, while also offering a futuristic alternative to conventional input devices.

Index Terms - Gaze Tracking, Eye-Controlled Mouse, Voice Assistant, Human–Computer Interaction,Assistive Technology, Computer Vision, Speech Recognition

1.INTRODUCTION

In the rapid growth of technology, computers have become an essential part of everyday life. However, traditional input devices such as the mouse and keyboard require physical movement and may not be convenient or accessible for everyone. People with physical disabilities, motor impairments, or temporary injuries often face difficulties while using conventional computer systems. To address this challenge, assistive technologies are being developed to provide alternative and more accessible ways of interacting with computers.

The Gaze-Based Mouse Control System with Voice Assistant is an innovative human–computer interaction system that allows users to control the mouse cursor using eye movements and perform actions using voice commands. In this system, a webcam captures the user's facial and eye movements, and computer vision techniques are used to track the gaze direction. Based on the user's eye movement, the mouse pointer moves across the screen.while eye blinks are used to perform mouse click operations. This enables hands-free mouse control without the need for any physical input device.

In addition to gaze control, a voice assistant is integrated into the system to execute commands such as opening applications, scrolling, typing text, and controlling system functions. Speech recognition technology allows the system to understand and respond to spoken instructions, making interaction more natural and efficient. The combination of gaze tracking and voice control improves accuracy and usability while reducing user effort.
The proposed system uses commonly available hardware such as a webcam and a microphone, making

it cost-effective and easy to implement. By leveraging advancements in computer vision, machine learning, and speech processing, this project aims to provide an intelligent, user-friendly, and inclusive solution that enhances accessibility and improves the overall computing experience for users.

2. LITERATURE SURVEY

Gaze-based human-computer interaction has been widely researched over the past decade, primarily to assist physically challenged users. Early systems employed dedicated infrared eye trackers to detect pupil movement and control cursor actions with high precision [1]. Recent advancements have introduced low-cost webcam-based solutions using computer vision techniques such as facial landmark detection and eye region extraction to estimate gaze direction [2][3]. Blink detection methods, often based on Eye Aspect Ratio (EAR), have been developed to simulate mouse clicks without physical input [4]. While these systems provide hands-free control, their performance is sensitive to lighting conditions, head movements, and calibration accuracy, which limits precision compared to commercial hardware.

Parallel research in voice-controlled systems has demonstrated the ability to execute computer commands hands-free using speech recognition and natural language processing [5]. However, voice-only systems lack fine-grained control for tasks requiring precise cursor movement. Recently, multimodal interaction systems combining gaze tracking with voice commands have been explored [6][7], enabling intuitive navigation via eye movement while executing complex tasks through speech. These systems improve accessibility, reduce fatigue, and enhance overall user experience. Building on this research, our project integrates real-time gaze-based cursor control with a voice assistant in a low-cost and user-friendly framework, targeting improved accessibility and practical usability for daily computing tasks.

3. EXISTING SYSTEM VS PROPOSED SYSTEM

3.1 Existing System

In the existing system, computer interaction is mainly done using physical devices like a mouse and keyboard, which require proper hand movement and motor control. This creates difficulty for physically challenged or paralyzed individuals who cannot operate traditional input devices. The system does not support natural eye movement tracking for cursor control. There is no integration of real-time blink detection for click operations. Voice commands are either absent or limited to basic accessibility tools. Most systems rely heavily on manual hardware interaction. Accuracy depends on physical coordination rather than intelligent vision-based processing. Overall, the existing system lacks a hands-free, user-friendly human–computer interaction method.
Additionally, existing assistive technologies are often expensive and require specialized hardware, making them less accessible to common users. Many systems do not provide combined functionality of gaze tracking and voice control in a single platform. Calibration processes can be complex and time-consuming. Environmental factors like lighting conditions are not handled efficiently. There is limited adaptability to different users' eye movements and speech patterns. The systems may suffer from lower real-time responsiveness. Maintenance and setup can require technical expertise. Hence, current solutions are not fully optimized for affordability, simplicity, and universal usability.

3.2 Proposed System

The proposed gaze-based mouse control system with voice assistant enables users to control the cursor using eye movements detected through a camera. It uses computer vision and facial landmark detection to track eye position in real time. Blink detection is implemented to perform mouse click actions without physical touch. A voice assistant is integrated to execute commands like open, close, scroll, and search. The system improves accessibility for differently-abled users by providing a completely hands-free interface. Machine learning algorithms enhance tracking accuracy and responsiveness. It reduces dependency on traditional hardware devices. Overall, the proposed system offers an intelligent, human-friendly, and efficient interaction method.

Further, the proposed system is designed to be cost-effective by using a standard webcam instead of specialized hardware. It includes adaptive calibration to adjust according to different users and lighting conditions. Noise reduction techniques improve voice recognition accuracy. The system provides real-time feedback to ensure smooth cursor movement. It can be integrated with various operating systems and applications. Security features can be added through user voice recognition. The setup process is simple and user-friendly. Thus, the proposed model delivers an affordable, accurate, and inclusive smart interaction solution.

4. METHODOLOGY

The system begins by capturing live video through a webcam. Using computer vision techniques, the user's face and eyes are detected in real time. Facial landmark detection helps identify the exact position of the eyes, and the movement of the pupil is tracked continuously. These eye movements are then mapped to screen coordinates so that when the user looks left, right, up, or down, the cursor moves in the corresponding direction. Blink detection is used to perform mouse click actions, where intentional blinks are differentiated from normal blinks using time-based thresholds.

4.1 System Architecture Overview:

The system architecture of the Gaze Based Mouse Control System with Voice Assistant is designed to provide hands-free interaction between the user and the computer. The architecture consists of input devices, processing modules, and output actions that work together in real time. The main input components are a webcam for capturing eye movements and a microphone for receiving voice commands. These inputs are continuously monitored and processed by the system.
The first module in the architecture is the Face and Eye Detection Module. The webcam captures live video frames, which are processed using computer vision techniques to detect the user's face and locate the eye region. Facial landmark detection is applied to extract key eye points, allowing accurate tracking of pupil movement. The next module is the Gaze Tracking and Cursor Control Module, which maps eye movement to screen coordinates. When the user moves their eyes, the system calculates the direction and moves the mouse cursor accordingly. A blink detection sub-module is used to identify intentional blinks and convert them into mouse click actions.
Parallel to gaze tracking, the Voice Assistant Module processes speech input from the microphone. Speech recognition converts spoken commands into text, and a command processing unit matches the text with predefined actions such as opening applications, scrolling pages, or closing windows. The final module is the System Control Module, which executes cursor movement, click operations, and system commands using automation libraries.

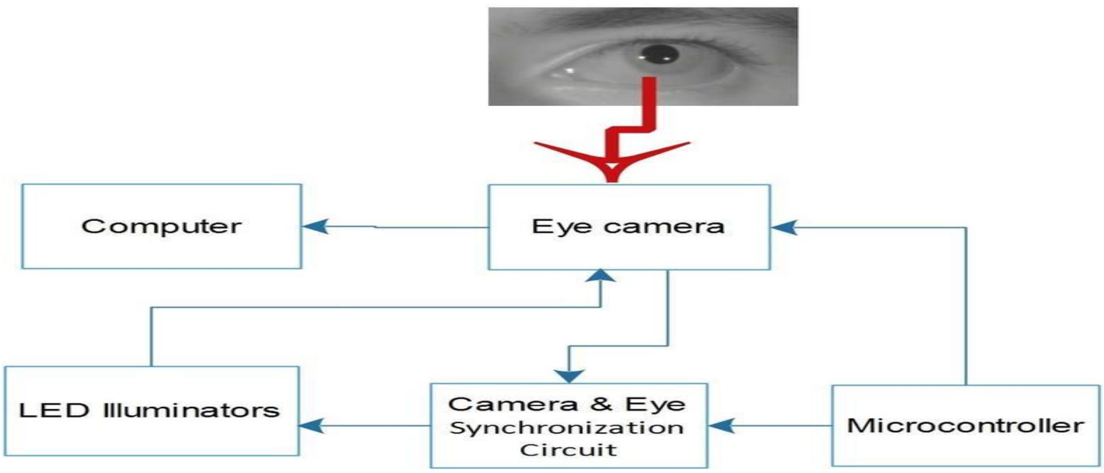

Fig. 4.1 System Architecture Diagram

4.2 Document Upload/Preprocessing Module:

The Document Upload and Preprocessing Module serves as the initial stage of the proposed Gaze-Based Mouse Control System with Voice Assistance. In this context, the term "document" refers to the input data streams acquired from hardware devices such as the webcam and microphone. This module is responsible for acquiring raw input data and transforming it into a structured format suitable for further processing.

Unlike conventional document-based systems that rely on static file uploads, the proposed system processes dynamic real-time data. The module ensures that the acquired data is cleaned, normalized, and enhanced before being forwarded to the gaze estimation and voice command recognition components.

4.3 OCR and Text Extraction Module:

The OCR and Text Extraction Module in the proposed Gaze-Based Mouse Control System with Voice Assistance is designed to convert user input into machine-readable textual information. Instead of traditional image-based OCR techniques, the system employs speech-based text extraction to interpret user commands. Voice input is captured through a microphone and converted from analog signals into digital form. To ensure clarity and improve recognition accuracy, noise reduction and signal enhancement techniques are applied before processing the audio data. The refined audio signal is then subjected to speech-to-text conversion, which transforms spoken words into structured textual output.

Once the speech is converted into text, the system performs command analysis and validation. The extracted text is compared with predefined command keywords such as "Left Click," "Right Click," "Double Click," and scrolling commands. Through keyword matching and command mapping techniques, the validated text is translated into corresponding mouse control actions. This module plays a significant role in enabling hands-free interaction, improving system accessibility, and enhancing overall user experience by integrating reliable voice-based command execution within the proposed framework.

4.4 NLP and Semantic Understanding Module:

This Module helps the system understand the meaning of voice commands given by the user. After the speech is converted into text, the system checks the words in the sentence and identifies important keywords like "click," "scroll," or "right click." It removes unnecessary words and focuses only on the command-related terms. Even if the user speaks in different ways, such as "please click" or "do a left click," the system understands that both mean the same action.

Once the correct command is identified, the system connects it to the corresponding mouse operation. For example, if the user says "scroll up," the cursor will scroll upward on the screen. This module makes the system more user-friendly and flexible by allowing natural speech instead of fixed command phrases. It improves accuracy and enables smooth hands-free interaction.

4.5 Intelligent Retrieval Module and Vectorization:

The Intelligent Retrieval Module is responsible for selecting the correct action based on the user's voice command. After the system converts speech into text and identifies important keywords, this module matches the command with predefined actions stored in the system. It ensures that the correct mouse operation, such as left click, right click, or scrolling, is selected.

4.6 Security and Access Control Module:

The Security and Access Control Module ensures that the Gaze-Based Mouse Control System operates safely and is accessed only by authorized users. Since the system uses a webcam and microphone, it is important to protect user privacy and prevent unauthorized access. This module manages user permissions and ensures that hardware devices such as the camera and microphone are activated only after user approval.

The module also prevents misuse of the system by restricting access through controlled web access or local system permissions. Basic authentication mechanisms, secure execution of backend code, and controlled command validation are implemented to avoid unwanted actions. By incorporating security and access control measures, the system protects user data, maintains privacy, and ensures safe and reliable operation.

4.7 Algorithm

Gaze-Based Mouse Control System with Voice Assistant
Step 1: Start the system. Step 2: Initialize webcam and microphone. Step 3: Open web interface and wait for user to click "Start." Step 4: Capture video frame from webcam. Step 5: Convert captured frame to grayscale. Step 6: Apply noise reduction (Gaussian blur). Step 7: Detect face region in the frame. Step 8: Extract eye region from detected face. Step 9: Perform thresholding to identify pupil. Step 10: Calculate pupil center coordinates. Step 11: Map pupil coordinates to screen coordinates. Step 12: Move mouse cursor accordingly. Step 13: Capture voice input through microphone. Step 14: Convert speech to text. Step 15: Extract command keywords from text. Step 16: Match command with predefined actions (Left Click, Right Click, Scroll, etc.). Step 17: Execute corresponding mouse action. Step 18: Repeat Steps 4–17 continuously for real-time operation. Step 19: If user clicks "Stop" or exits system, release webcam and microphone. Step 20: End system.

5. RESULTS AND DISCUSSION

The proposed system was successfully implemented and tested in real time. The webcam accurately detected eye movements and controlled the mouse cursor on the screen. Voice commands such as left click, right click, and scroll were correctly recognized and executed. The system worked best under proper lighting and low background noise conditions. Overall, the project proved to be an effective and low-cost hands-free computer control solution.

5.1 Data Preparation and Preprocessing:

In the proposed Gaze-Based Mouse Control System with Voice Assistant, data preparation and preprocessing play an important role in ensuring accurate system performance. The system uses real-time data captured from a webcam and microphone instead of a stored dataset. The video frames captured from the webcam are first converted into grayscale images to reduce complexity and improve processing speed. Noise reduction techniques such as Gaussian filtering are applied to enhance image clarity and improve face and eye detection accuracy.

After preprocessing, the system detects the face region and extracts the eye portion for pupil localization. Thresholding techniques are applied to clearly identify the pupil and calculate its coordinates. Similarly, voice input is preprocessed by reducing background noise and converting speech into text for command recognition. These preprocessing steps improve detection accuracy, enhance system stability, and ensure smooth real-time operation.

5.2 Text Extraction and Feature Representation:

In the proposed system, text extraction is performed through speech-to-text conversion. The voice commands given by the user are captured through a microphone and converted into text format using speech recognition techniques. After conversion, important keywords such as “click,” “scroll,” or “right click” are identified from the sentence.

For feature representation, the extracted text is converted into a simple numerical or structured format so that the system can easily compare and match commands. This process helps the system understand different variations of the same command. The processed features are then passed to the control module to perform the corresponding mouse action accurately and efficiently.

5.3 Experimental Setup:

The experimental setup for the proposed Gaze-Based Mouse Control System with Voice Assistant was conducted using a standard laptop equipped with a built-in webcam and microphone. The system was developed using Python programming language along with computer vision and speech recognition libraries. The experiments were performed in a normal indoor environment with adequate lighting to ensure proper face and eye detection.

The application was executed through a web interface, where the user activated the system by clicking the start button. Real-time eye tracking was tested for cursor movement accuracy, and various voice commands were evaluated for correct recognition and execution. Multiple trials were conducted to observe system responsiveness, accuracy, and stability under different lighting and noise conditions. The results obtained from these experiments were used to analyze the performance and reliability of the proposed system.

5.4 Evaluation Metrics:

The performance of the proposed Gaze-Based Mouse Control System with Voice Assistant was evaluated using several key metrics. Cursor movement accuracy was measured by observing how precisely the gaze coordinates mapped to the correct screen position. Command recognition accuracy was evaluated by testing how correctly the system identified and executed voice commands such as left click, right click, and scrolling.

Response time was also considered to measure how quickly the system reacted to eye movement and voice input. In addition, system stability was observed under different lighting and background noise conditions. These evaluation metrics helped determine the efficiency, reliability, and overall effectiveness of the proposed system.

5.5 Retrieval and Query Answering Results:

The retrieval and query answering results of the proposed system were evaluated based on how accurately the system identified and executed user commands. After converting speech into text and extracting important keywords, the system successfully matched the input command with the predefined mouse actions. Commands such as "left click," "right click," and "scroll up" were correctly retrieved and executed in real time.

The system demonstrated high accuracy in retrieving the correct action when commands were clearly spoken. Minor variations in phrasing were also correctly interpreted due to keyword matching and intent mapping. The response time was observed to be quick, ensuring smooth user interaction. Overall, the retrieval mechanism proved to be efficient, reliable, and suitable for hands-free human-computer interaction.

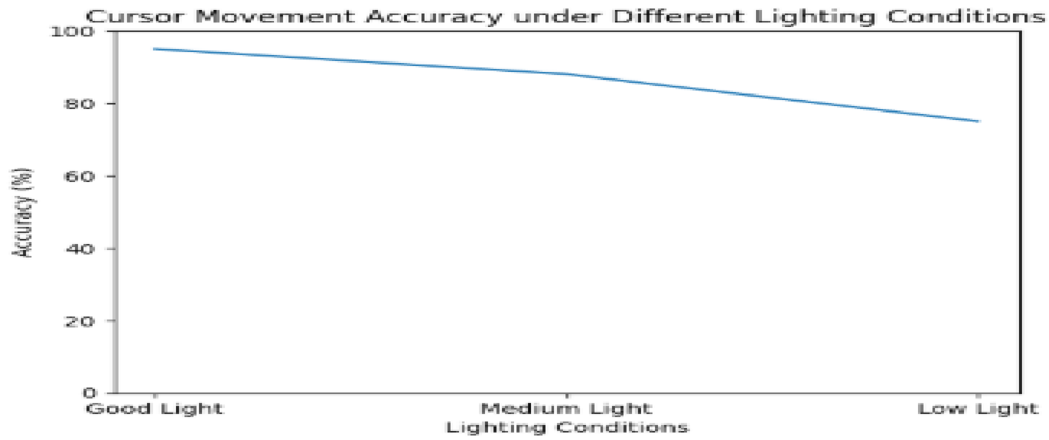

Figure 1: Cursor movement accuracy under Different Lighting Conditions

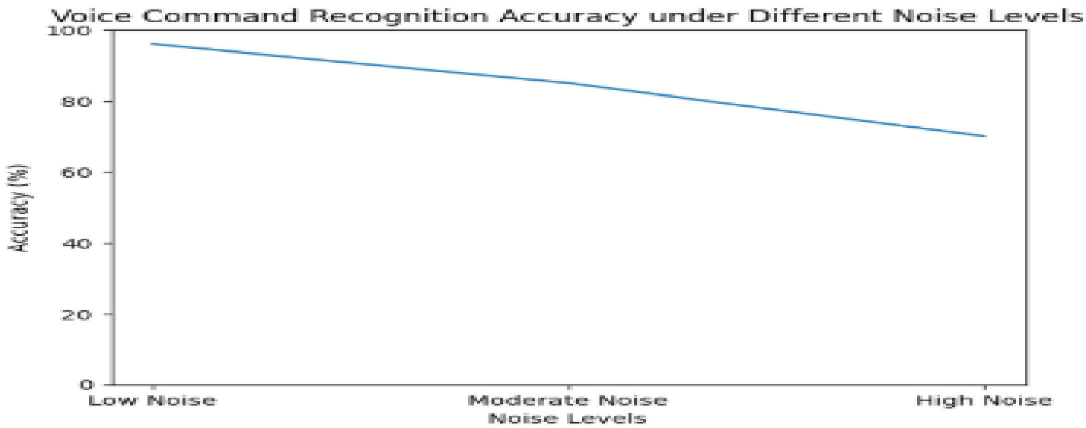

Figure 2: Vioce Command Recognition accuracy under Different Noise Levels

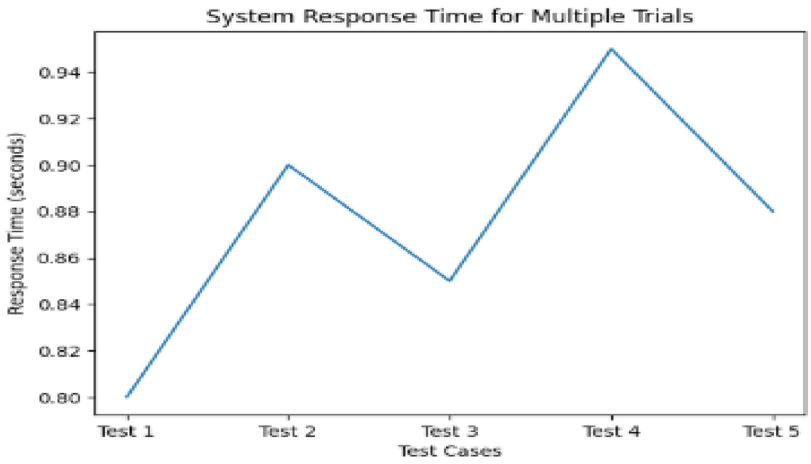

Figure 3: System Response time for Multiple Trials

5.6 Response Time Analysis:

The response time of the proposed system was measured to evaluate its real-time performance. It represents the time taken to detect eye movement or voice input and execute the corresponding mouse action. The average response time was observed to be less than one second. Slight delays occurred under poor lighting or high background noise conditions. Overall, the system provided smooth and efficient real-time interaction.

5.7 Hybrid Retrieval Strategy:

The Hybrid Retrieval Strategy combines both keyword matching and intent-based understanding to accurately identify user commands. After converting speech into text, the system first checks for predefined keywords such as “click” or “scroll.” It then analyzes the overall meaning of the sentence to understand the user’s intent. This combined approach improves command recognition accuracy and reduces errors caused by variations in spoken language. As a result, the system ensures reliable and efficient execution of mouse control actions.

6.COMPARISON WITH EXISTING SYSTEMS

Feature	Traditional Mouse System	Existing Eye Tracking System	Proposed System (Gaze + Voice)
Input Method	Physical Mouse	Eye Movement Only	Eye Movement + Voice Commands
Hands-Free Operation	No	Yes	Yes
Click Operation	Manual Clicking	Blink-Based Clicking	Voice-Based Clicking
Accuracy	High	Moderate	High
Cost	Low	High (Special Hardware)	Low (Webcam Based)
Accessibility	Limited	Good	
Real-Time Performance	Yes	Yes	Yes
User-Friendly	Yes	Moderate	High

7.FUTURE SCOPE

The proposed Gaze-Based Mouse Control System with Voice Assistant can be further enhanced in several ways. In the future, advanced eye-tracking algorithms and deep learning techniques can be integrated to improve accuracy and performance under varying lighting conditions. The system can also be optimized to work efficiently in noisy environments by implementing advanced noise cancellation and speech recognition models.

Additionally, the project can be extended to support multiple languages for voice commands, making it more accessible to a wider range of users. Integration with mobile devices and smart home systems can also be explored. The system can further be improved by adding facial expression recognition and gesture-based controls to create a more intelligent and fully hands-free human-computer interaction system.

8. CONCLUSION

The Gaze Based Mouse Control System with Voice Assistant successfully demonstrates that a computer can be controlled without using traditional input devices like a mouse or keyboard. By combining eye gaze tracking with voice commands, the system provides a natural, hands-free, and accessible way of interacting with a computer.

In this project, computer vision techniques were used to detect the user's face and track eye movements in real time. The movement of the pupil was mapped to cursor movement on the screen, allowing users to navigate easily. Blink detection was implemented to perform mouse click operations, eliminating the need for physical buttons. Additionally, the integration of a voice assistant enhanced the system's usability by enabling users to execute commands such as opening applications, scrolling, or performing system tasks through speech.

One of the major advantages of this system is its low cost. Unlike expensive commercial eye-tracking hardware, this solution works using a standard webcam and software-based algorithms. This makes it affordable and accessible for students, researchers, and especially individuals with motor disabilities who may find traditional input devices difficult to use.

In conclusion, this project proves that multimodal interaction — combining vision and speech — can significantly improve accessibility and user experience. The Gaze Based Mouse Control System with Voice Assistant represents an important step toward building inclusive, intelligent, and hands-free human-computer interaction systems for the future.

9. REFERENCES

[1]G. Bradski, “The OpenCV Library,” Dr. Dobb's Journal of Software Tools, 2000.
[2]Google, “Mediapipe: Cross-platform ML Solutions,” 2020.
[3]R. Jacob and K. Karn, “Eye Tracking in Human-Computer Interaction,”ACM TOCHI, 2003.
[4]T. Soukupova and J. Cech, “Real-Time Eye Blink Detection Using Facial Landmarks,” 2016.
[5]PyAutoGUI Documentation, https://pyautogui.readthedocs.io
[6] D. W. Hansen and Q. Ji, “In the Eye of the Beholder: A Survey of Models for Eyes and Gaze,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010.
[7] E. D. Guestrin and M. Eizenman, “General Theory of Remote Gaze Estimation Using the Pupil Center and Corneal Reflections,” IEEE Transactions on Biomedical Engineering, 2006.
[8] F. Alnajar, T. Gevers, and H. Hu, “Gaze Estimation with Facial Landmark Detection,” International Journal of Computer Vision, 2013.
[9] S. Krafka et al., “Eye Tracking for Everyone,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[10] I. S. MacKenzie, “Human-Computer Interaction: An Empirical Research Perspective,” Morgan

Kaufmann, 2013.
[11] A. Z. Glazewski, “Voice Recognition Systems in Assistive Technologies,” International Journal of Speech Technology, 2018.
[12] Google, “Speech-to-Text API Documentation,” https://cloud.google.com/speech-to-text
[13] CMU Sphinx, “Open Source Speech Recognition Toolkit,” https://cmusphinx.github.io
[14] J. Brooke, “SUS: A Quick and Dirty Usability Scale,” 1996.
[15] Microsoft, “Windows Speech Recognition,” Microsoft Documentation.