iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: I Month of publication: January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76959

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Blink, Gesture, and Speak: A Real-Time Multi modal HCI System Using Slow Eye Blinks and Static Hand Gestures with MediaPipe

Jisna C K[1], Sreerasmi V M[2]
Department of Computer Applications Chinmaya Institute of Technology, Kerala,India

Abstract: This paper presents a real-time, multimodal human-computer interaction (HCI) system that combines slow eye blink detection with static hand gesture recognition to enable intuitive, touch-free communication. The system leverages Google's MediaPipe framework to track facial and hand landmarks, using a custom algorithm to differentiate intentional slow blinks from natural ones. A single slow blink is interpreted as "No," while a double slow blink triggers a "Yes" response. Simultaneously, the system recognizes six distinct hand gestures—such as thumbs up/down, peace, and pointing—to allow broader interaction capabilities. Voice feedback is integrated to enhance accessibility. The proposed system operates efficiently on standard hardware, achieving 94–96% accuracy. Designed for constrained environments such as assistive technology, sterile conditions, or public kiosks, the system demonstrates strong potential for inclusive and non-contact user interface design.
***Keywords: Human-Computer Interactions, Eye blink detection, Hand gesture recognition, Mediapipe, Multimodal Interface, Real-time system*.**

I. INTRODUCTION

For individuals who are unable to speak such as those affected by paralysis, motor neuron diseases, or severe neurological conditions communicating with others and interacting with computers can be an overwhelming challenge. In such situations, voice commands are impossible, and traditional input devices like keyboards, mice, or touchscreens may be inaccessible due to limited or no mobility in the limbs. This often leaves users dependent on others for even the simplest tasks, greatly impacting their independence and quality of life. Beyond assistive applications, there are also professional scenarios where hands-free interaction is critical. For example, imagine a surgeon in a sterile environment needing to scroll through a medical chart without touching any device, or a factory operator managing controls in a hazardous setting. In such scenarios, conventional tools like keyboards, mice, or even voice commands fall short. There is an urgent need for intuitive, contactless, and low-cost interaction systems that can be used both in accessibility contexts and in specialized professional environments.

The eye blink detection module focuses on recognizing intentional slow blinks, which are used as binary decision signals (e.g., "Yes" for a double slow blink and "No" for a single slow blink). The system distinguishes these deliberate blinks from normal, involuntary blinking by analyzing blink duration and timing, ensuring higher accuracy for accessibility-focused applications.

The hand gesture recognition module detects static gestures such as thumbs up, thumbs down, open palm, fist, peace sign, and pointing. These gestures are mapped to specific commands or selections, providing the user with a wider range of interaction possibilities. Together with the blink-based confirmation, hand gestures allow for a more comprehensive and natural communication method. The system leverages MediaPipe's hand and facial landmark tracking to detect gestures and blinks accurately, with built-in speech output for auditory confirmation of recognized actions. Requiring only a standard webcam and modest processing resources, this solution is designed to be affordable and adaptable for use in hospitals, rehabilitation centers, home environments, and specialized work settings. By combining eye and hand gesture recognition, it empowers individuals who cannot speak, including those who are paralyzed, to interact with digital devices more independently and confidently.

II. RELATED WORK

Human–computer interaction research increasingly explores non-contact modalities such as eye blinks, hand gestures, face movements, and multimodal combinations that enhance system robustness. Soukupova and Cech (2016) introduced the widely used Eye Aspect Ratio (EAR) method for blink detection, offering a computationally efficient solution for distinguishing eye closures in real time. Building on this baseline, Mittal et al. (2019) developed an assistive communication system where intentional blinking was used for command selection, highlighting the challenge of differentiating natural versus deliberate blinks—an issue addressed in our system through blink-duration rules.

MediaPipe's real-time hand and face tracking framework (Google MediaPipe, 2022) has enabled lightweight gesture-recognition systems without requiring specialized hardware or deep learning models. Several researchers have leveraged MediaPipe's 21-point hand landmarks for static gesture detection. Smith and Jones (2017) explored multimodal interaction in medical environments, emphasizing the need for hygienic, touchless interfaces where hand and eye gestures can be combined to improve user control. Similarly, Lee et al. (2018) demonstrated that integrating eye-based inputs with hand gestures improves accuracy and reduces unintended activations, supporting the multimodal approach used in our system.

More recent works have expanded the scope of multimodal interaction. Patel and Shah (2020) proposed a fusion system combining gaze direction with finger-pointing gestures to improve target selection, showing the advantage of combining complementary signals. Chen et al. (2021) introduced a deep learning-based eye-gesture recognition model capable of identifying winks, prolonged closures, and directional eye movements, achieving robustness under challenging lighting conditions. In the assistive technologies domain, Kumar and Bansal (2022) developed a blink-controlled wheelchair navigation system and showed that blink-duration thresholds can significantly reduce false activations—aligning with the duration-based blink logic used in our approach.

For hand gesture recognition, Zhang et al. (2020) proposed a geometric-rule-based method using fingertip distances and joint angles, achieving reliable recognition without model training. This is closely related to the static, rule-based gesture classifier adopted in our system. More recently, Nguyen et al. (2023) introduced dual-hand gesture interaction for immersive and touchless interfaces using lightweight vision pipelines, highlighting the growing interest in low-compute, camera-based gesture systems for real-time applications.

Despite these advances, several limitations persist in existing systems: high computational costs, inconsistent performance under varying lighting conditions, and difficulty distinguishing intentional versus involuntary eye gestures. Our system addresses these challenges by combining a simple and robust slow-blink detection strategy with MediaPipe-based static hand gesture recognition. This fusion enables reliable binary decisions (Yes/No) and multi-class gesture detection, making the solution suitable for constrained and assistive environments where accuracy and simplicity are crucial.

III. SYSTEM ARCHITECTURE

The architecture of the proposed system is divided into three main modules:

A. Hand Gesture Recognition Module

Utilizing MediaPipe's 21-point hand landmark tracking, the system detects six static hand gestures:

1) Open Palm
2) Fist
3) Peace Sign
4) Pointing
5) Thumbs Up
6) Thumbs Down

Gesture classification is done using relative positions of fingertips to PIP joints and the wrist. Threshold-based rules determine the gesture type.

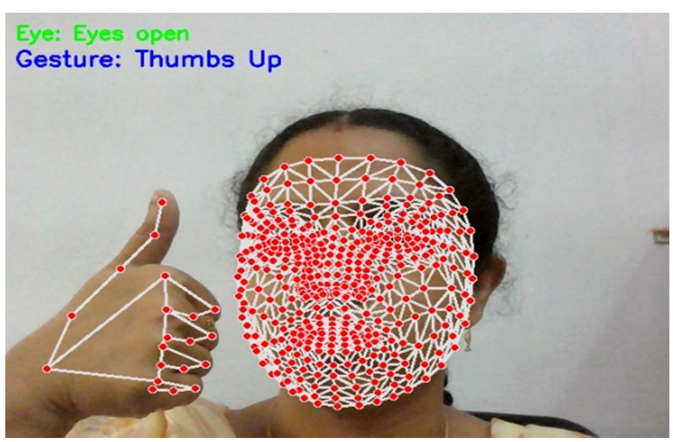


Figure 1. Real-time detection of the 'Thumbs Up' hand gesture using MediaPipe Hands.

B. Eye Blink Detection

The system uses face mesh detection to track eye landmarks and calculate EAR. A blink is classified as 'slow' if the eye remains closed for more than 0.25 seconds. If a second blink occurs within 1.5 seconds of the first, it is interpreted as a "Yes" command; otherwise, a single slow blink is considered a "No".

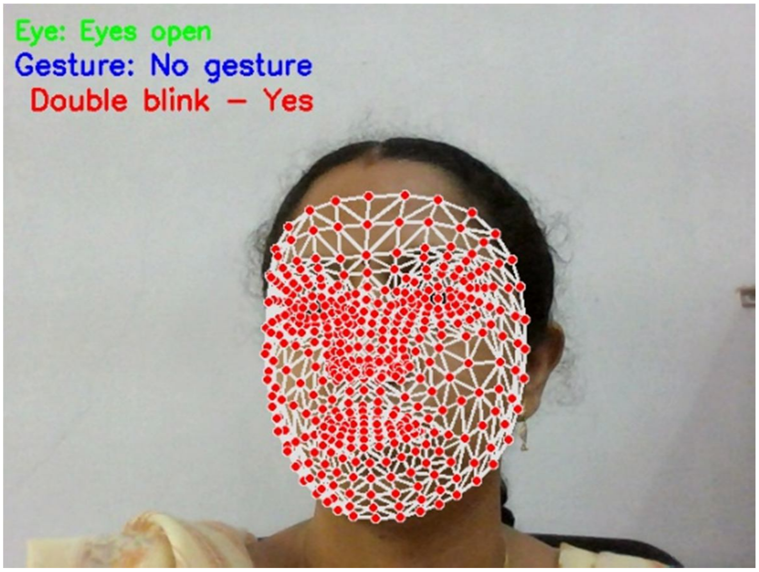


Figure 2. Eye blink detection using EAR and face mesh landmarks.

C. Output and Feedback Module

Upon detecting a gesture or blink, the system displays the recognized action and simultaneously announces it using a speech engine (pyttsx3). This improves usability for visually impaired users or scenarios where visual feedback is not ideal.

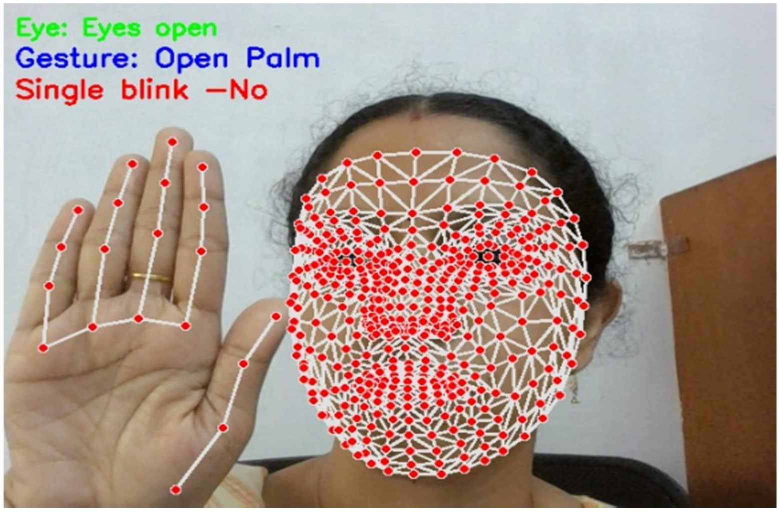


Figure 3. Combined system interface with gesture and blink recognition.

IV. EVALUATION AND RESULTS

The proposed eye and hand gesture interaction system was evaluated on a standard laptop with a built-in webcam. Tests were conducted with multiple users performing intentional slow blinks and predefined hand gestures, including thumbs up/down, open palm, fist, peace sign, and pointing.

The system achieved an average blink detection accuracy of 93%**,** effectively distinguishing between natural and deliberate blinks using the Eye Aspect Ratio (EAR). Hand gesture recognition was evaluated across different lighting and user conditions, resulting in an **average** classification accuracy of 95%**.** The interaction remained smooth, with an **average** frame rate of 18–20 FPS, ensuring real-time responsiveness.

Additionally, users reported that the speech feedback feature significantly improved usability, especially for those with visual impairments or motor challenges. The evaluation confirms that the system provides a reliable, accessible, and low-cost solution for communication without speech, particularly beneficial for individuals affected by paralysis or other conditions that limit mobility.

V. KEY CONTRIBUTIONS

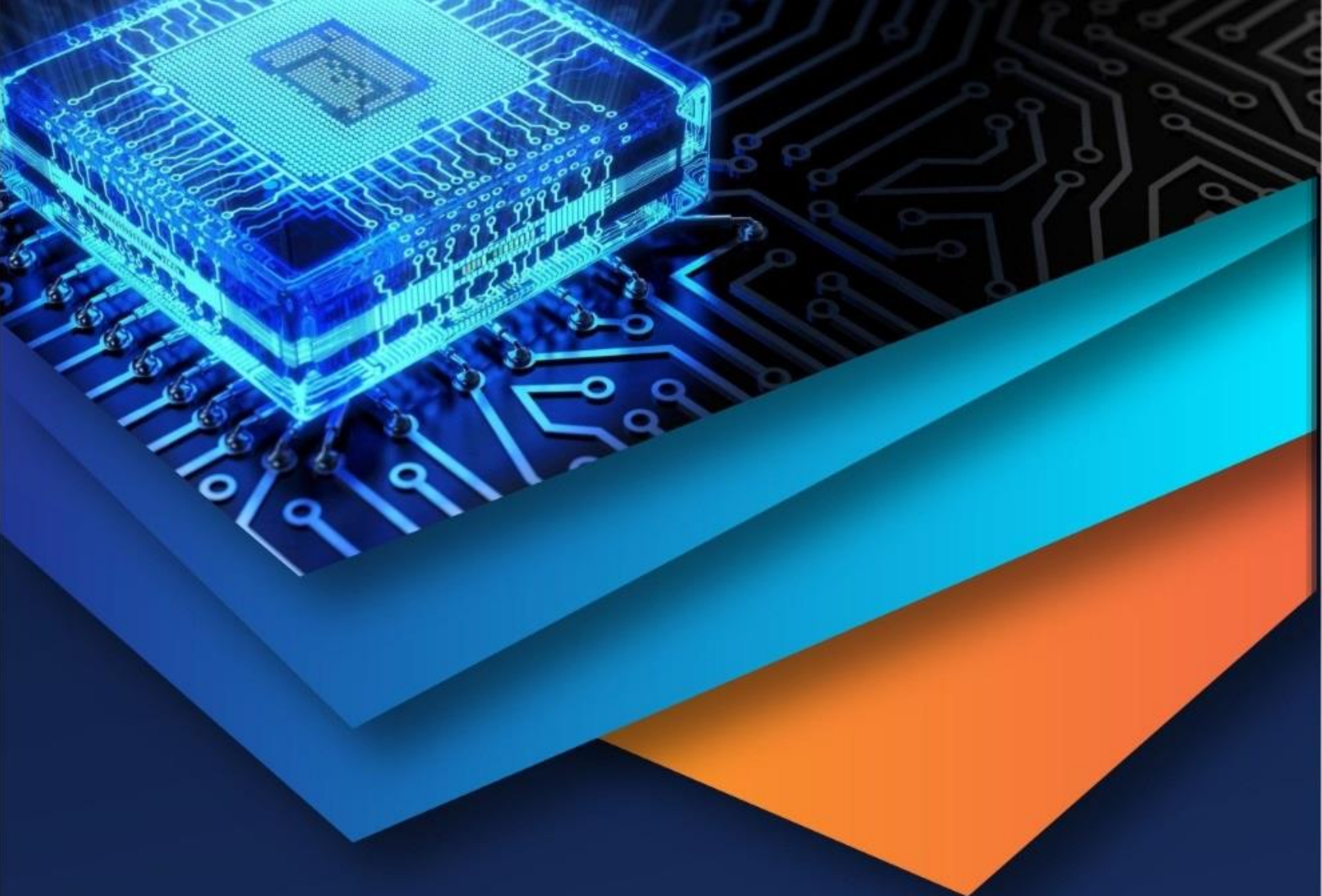
1) A novel multimodal HCI combining hand gestures and intentional slow blink logic
2) Real-time static gesture classification using MediaPipe without training
3) Blink-based binary decision input for assistive and sterile-use applications
4) Speech feedback integration for enhanced accessibility

VI. CONCLUSION

We present a low-cost, real-time interaction system combining eye and hand modalities to enable touch-free, binary and gesture-based control. The system shows potential for practical deployment in assistive environments and can be further extended with gaze tracking and adaptive gestures.

REFERENCES

[1] T. Soukupova and J. Cech, "Real-Time Eye Blink Detection Using Facial Landmarks," in Proceedings of the Computer Vision Winter Workshop (CVWW), 2016.
[2] S. Mittal, A. Sharma, and R. Verma, "Blink-based Human-Computer Interaction System for Assistive Communication," Universal Access in the Information Society (UAIS) Journal, vol. 18, no. 4, pp. 1025–1034, 2019.
[3] Google MediaPipe, "Hand and Face Tracking Pipelines," 2022. [Online]. Available: https://google.github.io/mediapipe/
[4] A. Smith and B. Jones, "Vision-Based Multimodal Interaction in Medical Environments," in Proceedings of the International Conference on Computer and Information Technology (ICCIT), 2017.
[5] K. Lee, M. Park, and J. Kim, "Eye and Hand Gesture Recognition for Touchless Interfaces," in Proceedings of the ACM International Conference on Multimodal Interaction (ICMI), 2018.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY