IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DUAL MODE INTELLIGENT ASSISTANT

Miss. Shruti Uttam Desai*1, Miss. Prajakta Pratap Kasabe*2, Miss. Madhuri Ravsaheb Kadam*3, Miss. Pratiksha Sanajy Kamble*4, Prof. M.M.Wakarekar*5

*1, 2, 3, 4 Students, *5 Assistant Professor

Department of CSE (Data Science),

D. Y. Patil College of Engineering and Technology, India

Abstract: This study presents a Dual Mode Intelligent Assistant System designed to provide hands-free human-computer interaction using voice recognition and EMG-based neuro-control techniques. The proposed system enables users to control computer operations through two independent modes: Voice Mode and Neuro Mode. The Voice Mode allows users to perform tasks using speech commands, while the Neuro Mode utilizes facial Electromyography (EMG) signals, eye movement tracking, and blink detection for cursor control and navigation. The system supports various operations including mouse movement, clicking, scrolling, typing and application launching without the need for conventional input devices such as a keyboard or mouse. EMG signals acquired from facial muscle activity are processed using machine learning algorithms to improve classification accuracy and system stability. The integration of multimodal interaction methods enhances accessibility, usability and flexibility especially for users with limited physical mobility. The proposed system offers a low-cost, portable, and efficient assistive technology solution for advanced human-computer interaction applications.

Index Terms - EMG (Electroencephalography) signal, Brain Computer Interface (BCI), Voice Mode, Human Computer Interaction (HCI).

I. Introduction

Human-Computer Interaction (HCI) has changed a lot with the growth of intelligent technologies and smart computing systems. Traditional input devices like keyboards and mice are still commonly used, but they may not always be convenient for people with physical disabilities, limited hand movement, or users who require hands-free interaction. Because of this, researchers are focusing on developing alternative methods that make computer interaction easier, smarter, and more accessible.

Voice recognition technology has become one of the most popular ways to interact with computers. Modern voice assistants allow users to perform tasks such as opening applications, searching the internet, managing files, controlling media, and operating system functions through simple speech commands. Voice-based interaction improves user convenience and reduces the need for manual input devices. However, voice systems may not work effectively in noisy environments or for users with speech-related difficulties.

Another emerging approach is neuro-control using Electromyography (EMG) signals. EMG signals are generated by muscle activity and can be captured from facial movements such as eyebrow raises and blinking actions. These signals can be used to control cursor movement and clicking operations. Eye movement tracking and blink detection further enhance the usability of neuro-control systems, especially for physically challenged users.

This research presents a Dual Mode Intelligent Assistant System that combines both Voice Mode and Neuro Mode into a single platform. In Voice Mode, users can control the computer using speech commands, while Neuro Mode enables cursor control through facial EMG signals, eye movement tracking, and blink detection. By integrating both interaction methods, the proposed system provides greater flexibility, improved accessibility, and a more reliable hands-free user experience under different environmental and user conditions.

II. Literature Survey

The Zander T. O. and Kothe C.'s work in Towards Passive Brain-Computer Interfaces: Applying Brain-Computer Inter- face Technology to Human-Machine Systems in General [1] explores hybrid Brain-Computer Interface (BCI) systems that combine EEG with other input modalities to enhance interaction performance. The study demonstrates how merging brain signals with behavioral inputs can increase system accuracy and responsiveness. This aligns with proposed system EEG- based control mode, where brain activity is combined with voice commands for a more reliable dualmode assistant. One major disadvantage noted is that EEG signal acquisition is often sensitive to noise and user variability, which can impact recognition accuracy.

The Rabiner L. and Juang B.-H.'s work in Fundamentals of Speech Recognition [2] provides foundational knowledge on speech processing, including techniques like Hidden Markov Models (HMMs) and feature extraction methods such as Mel- Frequency Cepstral Coefficients (MFCCs). These methods form the theoretical basis for modern Automatic Speech Recognition (ASR) systems. Proposed system leverages these principles through modern APIs and deep learning models for real-time voice-based control. One major disadvantage mentioned is that traditional HMM-based systems are less robust to background noise, which can reduce performance in noisy environments.

The Krafka K. et al.'s work in Eye Tracking for Everyone [3] presents a cost- effective gaze estimation method using standard webcams, supported by a large-scale dataset and deep learning techniques. This approach enables precise eye- tracking without specialized hardware, making it highly rele- vant to proposed system eye- tracking module for cursor and navigation control. One major disadvantage identified is that accuracy can be affected by lighting conditions and camera quality, limiting consistent performance.

The Lin Z. et al.'s work in Robust Speech Recognition in Noisy Environments Using Deep Learning [4] focuses on improving ASR performance in challenging acoustic environ- ments. By using deep neural network models, the authors achieved significant noise-robustness in speech recognition systems. This is directly applicable to our voice control module, ensuring reliability even in public or noisy work- places. One major disadvantage is the increased computational demand of deep learning models, which can limit their use in resource- constrained devices.

The Mullen T. et al.'s work in Real-Time Neuroimaging and Cognitive State Estimation in Neuroadaptive Human- Computer Interaction [5] investigates real-time EEG signal processing for adaptive interfaces. The study shows how BCIs can infer user intent and adapt system behavior accordingly. This approach supports proposed system EEG-based action control for clicks, scrolling, and application launching. One major disadvantage noted is that continuous EEG monitoring can cause user fatigue over prolonged periods.

The studies reviewed shows different ways to improve human-computer interaction using audio and EEG signals. Zander and Kothe showed that combining EEG with other inputs improves interaction but suffers from noise. Rabiner and Juang explained speech recognition basics, while Graves and Lin improved accuracy and noise-resistance using deep learning. Krafka proposed webcam-based eye tracking, though limited by lighting. Mullen explored real-time EEG control but noted fatigue, and Gu¨rko¨k with Nijholt highlighted multimodal interaction as flexible but complex. These works inspire our dual-mode assistant combining voice and neuro-based control.

III. System Design

The proposed Dual Mode Intelligent Assistant System is designed to provide hands-free computer interaction through two independent operational modes: Voice Mode and Neuro Mode. The system architecture is divided into three major sections: Input Layer, Processing Layer, and Execution Layer.

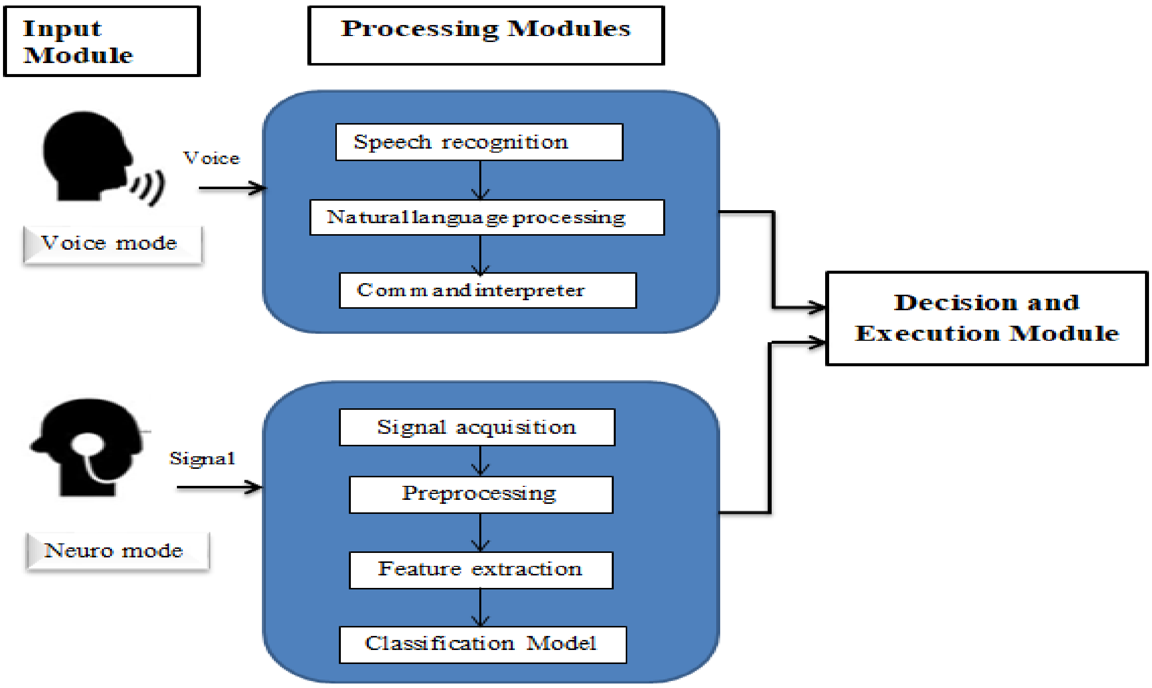

Fig. 1 Dual Mode Intelligent Assistant System Architecture

A. **The Input Layer** is responsible for receiving user commands either through voice input or facial EMG signals. Voice commands are captured using a microphone, while facial muscle activity is collected using surface electrodes connected to the BioAmp EXG Pill sensor and Arduino Uno.
B. **The Processing Layer** handles the analysis and interpretation of incoming data. In Voice Mode, speech signals are processed using speech recognition and natural language processing techniques to identify user commands. In Neuro Mode, EMG signals undergo preprocessing, feature extraction, and machine learning-based classification to detect cursor control actions.
C. **The Execution Layer** performs the final system operation based on the processed command. Depending on the detected action, the system executes tasks such as cursor movement, mouse clicking, scrolling, application launching, media control, or text typing.

The proposed architecture combines both interaction methods into a single intelligent framework, improving flexibility, accessibility, and overall user experience.

3.2 Hardware Components

Hardware Component	Purpose
Arduino Uno	EMG signal acquisition and serial communication
BioAmp EXG Pill	Amplification of facial EMG signals
Surface Electrodes	Collection of facial muscle activity
Microphone	Voice input acquisition

The BioAmp EXG Pill sensor is connected to the Arduino Uno to acquire EMG signals generated from facial muscle movements such as eyebrow raises and blinking actions. Surface electrodes are placed on the forehead region for signal collection. The microphone is used for capturing speech commands in Voice Mode

IV. METHODOLOGY

The implementation of the proposed system follows a step-by-step workflow for both Voice Mode and Neuro Mode to achieve real-time hands-free interaction.

A. Voice Mode Workflow

The Voice Mode begins by capturing user speech through a microphone. The speech recognition module converts the audio signal into text format. The recognized text is then analyzed to identify the intended command using command interpretation techniques. After successful recognition, the command is executed to perform operations such as application launching, text typing, system control and web search.

B. Neuro Mode Workflow

In Neuro Mode, facial EMG signals are acquired using surface electrodes placed above eyebrow. The BioAmp EXG Pill amplifies the acquired signals and the Arduino Uno transmits the signal data to the computer using serial communication.

The incoming EMG signals are preprocessed to remove unwanted noise and fluctuations. The processed signals are divided into small windows for feature extraction. Statistical features such as mean, standard deviation, maximum value and minimum value are extracted from each signal window.

A Random Forest machine learning model is used to classify the extracted features into predefined cursor actions. To improve system stability, techniques such as majority voting, noise filtering and action delay control are implemented. The classified outputs are finally converted into cursor movement and clicking operations using Python automation libraries.

C. System Integration

Both Voice Mode and Neuro Mode are integrated into a unified intelligent assistant framework. Users can interact with the system using either speech commands or neuro-control methods depending on their convenience and environmental conditions. The integration of both modes improves system adaptability and accessibility for different categories of users.

V. RESULT ANALYSIS

The proposed Dual Mode Intelligent Assistant System was successfully implemented and tested under real-time operating conditions. The system demonstrated effective hands-free computer interaction using both Voice Mode and Neuro Mode. Experimental evaluation was carried out to analyze system performance, command execution accuracy, response stability, and usability for both interaction methods.

The integration of voice recognition and EMG-based neuro-control improved system flexibility by allowing users to interact with the computer using either speech commands or facial muscle activity depending on environmental conditions and user convenience.

A. Voice Mode Performance

The Voice Mode was tested using multiple speech commands related to system operations, media control, application launching, web searching, and text input. The speech recognition module successfully converted user speech into executable commands in real time.

The system was able to perform operations such as:

- Opening applications
- Searching the internet
- File handling
- Text typing and formatting
- Basic system navigation

The Voice Mode showed fast response time and stable performance under normal environmental conditions with stable internet connectivity. The speech recognition engine accurately identified commands spoken in a clear voice with minimal delay during execution.

B. Neuro Mode Performance

The Neuro Mode was evaluated using facial muscle movements including eyebrow raise and blinking actions. Surface electrodes placed in the forehead region (above the eyebrow) successfully acquired EMG signals generated from facial muscle activity.

The acquired EMG signals were processed using preprocessing and feature extraction techniques before classification using a Random Forest machine learning model. The classified outputs were mapped to cursor movement and mouse control actions.

The Neuro Mode successfully performed:

- Cursor movement
- Mouse clicking
- Double clicking
- Navigation control

Prediction smoothing and noise filtering techniques improved interaction stability and reduced unintended cursor movements during real-time operation.

C. Classification Performance

The performance of the machine learning model used in Neuro Mode was evaluated using precision, recall, F1-score, and confusion matrix analysis. The Random Forest classifier achieved an overall accuracy of approximately 67.30% during testing.

Table I: Classification Performance of Neuro Mode

Class Label	Precision	Recall	F1-Score	Support
Neutral State	0.95	0.94	0.95	89
Mouse Click	0.53	0.72	0.61	69
Double Click	0.41	0.28	0.33	46
Cursor Movement	0.62	0.51	0.56	59

The results indicate that the model performed effectively for stable EMG signal patterns, particularly for the neutral state class. However, minor overlap was observed between blinking and cursor movement actions due to similarities in facial muscle signal characteristics.

D. Confusion Matrix Analysis

Fig. 2 shows the confusion matrix heatmap of the Random Forest model used in Neuro Mode. The heatmap represents how accurately the system classified different facial EMG signal patterns. The diagonal values indicate correctly classified actions, while the remaining values show incorrect predictions between similar facial movements. The results indicate that the system performed better for stable and clear facial actions, while some overlap occurred between similar cursor control actions due to similarities in EMG signal patterns.

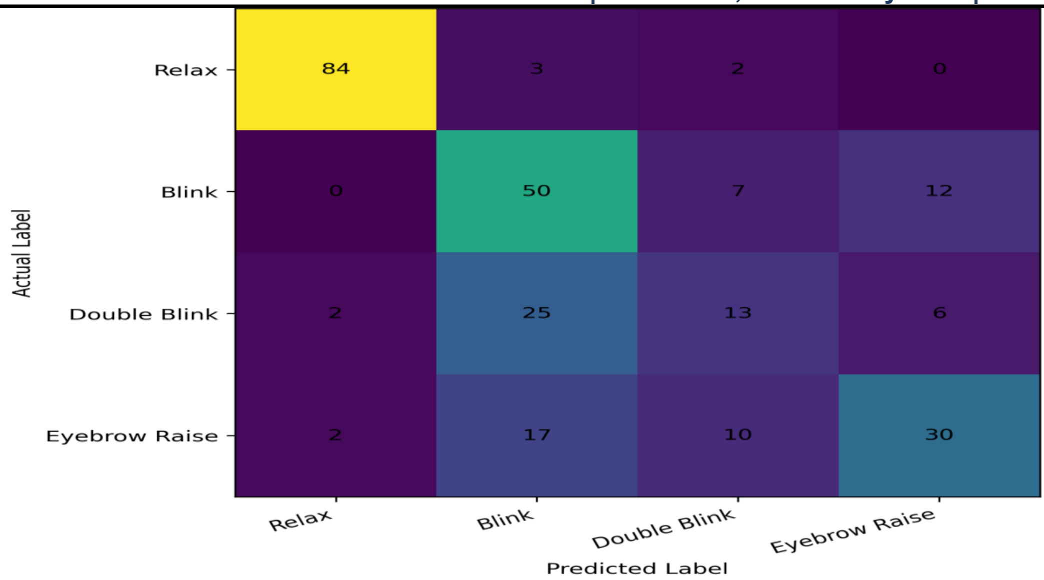

Fig. 2 Confusion Matrix of the Random Forest Classifier Used in Neuro Mode

VI. Conclusion

The proposed Dual Mode Intelligent Assistant System successfully demonstrated an effective hands-free human-computer interaction framework using both Voice Mode and Neuro Mode. The system integrated speech recognition, facial EMG signal processing and machine learning techniques to perform various computer operations without relying on conventional input devices such as keyboards and mice. The Voice Mode enabled users to execute commands related to application launching, file handling, web searching, and text input through speech recognition, while the Neuro Mode provided cursor control and navigation using facial muscle activity and blinking actions. Experimental analysis showed that the system achieved stable real-time performance with satisfactory classification accuracy using affordable hardware components and lightweight software frameworks. The integration of both interaction methods improved system flexibility, accessibility, and usability under different environmental and user conditions. The developed system provides a practical, portable and low-cost assistive technology solution for advanced intelligent human-computer interaction application

VII. Future Scope

The current system provides basic hands-free computer interaction using voice commands and EMG-based neuro-control. In future, the system can be improved by increasing cursor control capabilities, enhancing signal accuracy, and reducing response delay during real-time operation. Additional features such as wireless communication, smarter voice interaction, and improved eye tracking can make the system more efficient and user-friendly. The proposed framework can also be extended for applications in assistive technology, smart environments and advanced human-computer interaction systems.

References

[1] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2016.

[2] Richhariya, P., Kane, L., Chauhan, P., Nigam, M. (n.d.). A Survey: Movement of the HandMotion Trajectory for Dependent and Indepen- dent Recognition. IEEE Xplore.

[3] Zhang, S., Zhang, S. (2019). A Novel Human-3DTV Interaction System Based on FreeHand Gestures and a Touch Based Virtual Interface. IEEE Access, 7, 165961– 165974.

[4] Monteiro, P., Gonc¸alves, G., Coelho, H., Melo, M., Bessa, M. (2021). Hands- Free Interaction in Immersive Virtual Reality: A Systematic Review. IEEE Transactions on Visualization and Computer Graphics, 27(5), 2702–2717.

[5] Sancheti, K., Krishnan, S., Suhaas, A., Suresh, P. (2018). Hands-free Cursor Control using Intuitive Head Movements and Cheek Muscle Twitches. Proceedings of TENCON 2018 -2018 IEEE Region 10 Con- ference, Jeju, Korea. IEEE.

[6] Tripathy, S. (2020). Build Your Own Desktop Voice Assistant in Python. Analytics Vidhya.

[7] Kumar, A., Kumar, A., Kumar, P. (2021). Development of Voice Assistant using Python. International Journal of Research in Engineering and Science, 10(2), 43-50.

[8] Mukherjee, A. (2019). Survey on Virtual Assistant (Google Assistant, Siri, Cortana, Alexa). 4th International Symposium SIRS 2018, Revised Selected Papers, 65-74.

[9] Nallamothu, M., Mukkamala, R. (2019). A Study on Speech Recognition and Desktop Assistant using Python. Journal of Emerging Tech- nologies and Innovative Research, 6(4), 521-526.

[10] excellent style manual for science writers is “Speech recognition with flat direct models,” IEEE Journal of Selected Topics in Signal Processing, 2010.