Multi-Modal Fusion for Enhanced Human Computer Interaction(HCI)

Sai Tarak Nadh Yamparala

School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab, India  
[saitaraknadh16@gmail.com](mailto:saitaraknadh16@gmail.com)

Erugadindla Vamshi  
School of Computer Science and Engineering  
Lovely Professional University  
Phagwara, Punjab, India  
[vamshiirugadindla7032@gmail.com](mailto:vamshiirugadindla7032@gmail.com)

Devang Saren  
Upgrad Campus Upgrad Education Private Limited  
Bangalore , karnataka , India  
ORCID : 0009-0005-5873-7683  
[sarendev1812@gmail.com](mailto:sarendev1812@gmail.com)

***Abstract:* Our paper presents the idea of the virtual mouse character which is controlled by gesture recognition as well as eye ball and speech monitoring. The system is based on computer vision and machine learning technologies which give users an opportunity to control and command the mouse cursor with their eye movements, by voice, or using hand commands. The system goal is to give an interactive and simple interface to people with disabilities affecting their movements (e.g., paralysis) as well as to those who prefer a more natural and hands-free interaction with their computers.  
Keywords: Virtual Mouse, Eye Ball Control, Speech Control, Gesture Recognition, Computer Vision, Machine Learning, Accessibility, Mobility Impairments, Hands-free Interaction, Intuitive Interface**.

# **Introduction**

Although mice and keyboards that are commonly used as computer input devices might prove difficult for persons with mobility impairments to apply. Many types of devices require some exact and coordinated motion that their users in some cases (due to their movements level) could not handle. In the past decade, there have been a lot of efforts to come up with input systems that are easy to use and adjust for people who do not have full use of their hands.

A diagram of a computer network

Description automatically generated

Figure 1

Therefore, for example, eye tracking systems can be designed where the eye movements will control the method through which the computer is used. Camera and software specific using are a critical element of eye tracking, which is a kind of technique, making it possible to both detect eye movements and to measure them. Movement will be tracked by the system which will use them to define where a user is looking at and use it as a command for controlling mouse pointer.

The second item in the list might deal with a speech recognition feature implementation, which will ease the users' navigation. The speech recognition system component of a certain device has a number of so-called software programs inside it that enables the system to “read” the human speech into text or commands of some sort. Today one is no longer constrained in this option only by reading the words from a screen or the book but also moving a mouse by control carried out with one’ eye, that is a crucial component for people who have problems with movement.

Here we demonstrate a new control method that will enable virtual mice to be commanded using eyeball and verbal recognition as well as gesture tracking. The proposed system takes advantage of computer vision and the machine learning approach allowing users to control the mouse cursor refraining to the eye movements and speech commands or gestures from hand. The system interface is intended to offer a feature that utilizes voice command in addition to making the computer interaction with them hands-free. This is to accommodate people with mobility impairments and those who prefer a more natural interaction methodology in computer operation.

Unlike the disabled, which get a joy using the traditional computer input devices, such as mice and keyboards, may be challenging for individuals with mobility impairments. These gadgets quite frequently need movements which have millimeter precision and coordination and, therefore, not everybody can accomplish those types of movements. In the last time there are many studies that try to create an alternative input methods to be much more available and humanly. Next we put forward a method by which an eye ball and speech can be used to control a virtual mouse and this is based on the gesture recognition technique. System proffered utilizes computer vision and machine learning approaches so as to grant users the autonomy to utilize eye movements, speech commands, and hand gestures as instruction mechanisms of mouse cursor control. On the contrary, the main purpose of the system is to provide a concise and readily available interface for those people having limitations in the ability to move and also will promote those people who better prefer a more natural and face-to-face way of communication with the other.

Our new idea of working includes the use of voice recognition program as a door opener for people confined to wheelchair to control the mouse using their voice. The main concept of speech recognition is the usage of the program that has a speech pattern and can understand what a person says by means of special software like translating spoken words into text and commands. It does this function not only for those who have problems with their motor function, as the voice control helps them to control the mouse cursor.[1]

The system we are recommending utilizes the computer vision and machine learning technologies, letting tasks about cursor management be either controlled with eye tracking, voice commands or gesticulation of hands. The remodeling of the interface is being done such that the hands free way of interacting is provided, which is good for not only people with mobility disability but also those with character challenges using the traditional mouse and keyboard as computer input devices. These tools necessitate the positional accuracy and movement coordination on millimeter level, which is a great problem for people because of some masses unable to perform these tasks.

The goal of this system is to facilitate a more user-friendly approach to computer-based interaction for people with mobility disabilities who may find it difficult to compensate for slowed or impaired movements. In contrast to typical computers input operated with physically challenged people, who are sometimes unable to fully use the device, this system allows a simpler and more human-oriented approach. Eye movement commands, speech commands, and hand gestures are eye movements are the ones that the system provides technologies means for the populations with the limited movement ability to communicate more concisely than ever before.

Moreover, the system aims to establish more natural and verbal mode of communication with a computer instead of relying on typing on a keyboard. This is because such people are more likely to relate to screen-simulated human characters and prefer a natural interaction methodology over use of technological devices to operate computers. The architecture of the system is imbued with machine learning that helps the system improve constantly up to the point of providing a personalized and intuitive user experience.

Lastly, this presented system allows PC users with mobility impairments to navigate and control the computer by using their eyes instead of the usual hands and fingers. With eye gestures, speech prompts, and hand gestures, the system provides a natural non-hand-based user interface. The system employs machine learning system to continuously learn and enhance performance, which in turn, leads to the creation of more personalized and user-friendly experience for the customer. A technology under study like this provides a leap forward in designing input systems that are more responsive to people's needs as their computer-using skills improve.

# **Literature Review**

In the area of human-computer interaction, researchers have dedicated much of their time, interest and resources into the development of easy-to-use input methods that are also accessible for persons who struggle with mobility impairments. For instance, a mouse which is manipulated by eyeball and voice is just an interesting case that have raised a lot of attention over recent years.

Literature is replete with studies on the application of eye trackers as working of this mouse cursor comes under observation. For example, a study from International Journal for research in applied science & technology engineering technology presents an approach that incorporates eye moving techniques to enable a user to relocate the mouse pointer by just using eyes. The system includes a trained Haar classifier for the detection of a face and eyes and is able to find out the user's gaze on eyes without error. On the other hand, the study records some weaknesses in the pushing of depth towards the walls of the device, thus making it imperative to work upon in the future.

In another peer reviewed project published in the Ijariit journal, the developers are employing a similar idea of using an eye camera to take eye-dar photos. The first task is to find out where a pupil is inside the eye and then different positions of the eye are being used to direct the movement of the cursor. The monitoring of learning process is done through raspberry pi and the raspbian that is installed on the raspberry pi.

Analogously to the eye tracking, speech recognition technology is yet another area that has been explored by the scientists in order to control the movements of the mouse cursor. The Procedia Technology Journal has published a work that is focused on the system that performs mouse cursor operations and consequently various other mouse events by giving voice commands. This system utilizes the typical webcam for captured input image and employs image processing methods like face detection, eye extraction and high integrity to identify sequences of eye blinks in real-time.

The research study on mouse robotics designed with cognitive motion control and visual speech inputs and gesture recognition based mouse robotics requires reviewing other studies. Imagine the scenario: a user connects to a computer with his face through software that uses the face detection, eye extraction, and eye movement, (eg. eye blinks) sequences to control an unobtrusive, human-computer interface in real time. The next one suggests a system that uses eye movement as mere an input to carry out the task of the mouse cursor with the support of a trained Haar classifier for detection of face and eyes position. Yet such detection is successful, the system has no possibility to change the scaling in this case. Such scaling would be useful if the screen dimensions didn’t match the users eye gaze.

Diagram of a diagram of a computer model

Description automatically generated

Figure 2

The other research shows the development of a non-intrusive, human-computer interface, which uses the image from a built-in camera as an input and includes the eye detection, eye-extraction, and real-time sequence of eye blinks for the purposes of monitoring and controlling a subject’s movements with virtual mouse's cursor. The system is a simple one using a normal webcam to take pictures and different gazes act as cursor movements.

In addition, works put forward a system that converts speech signals into mouse-related commands, thus allowing users to execute mouse related operations. The system unlike most webcams or facial recognition software does not just detect faces and processes them with eye detection and eye-tracking algorithms, but it uses real-time facial recognition for controlling a non-intrusive, human-computer interface.

Similarly, a study develops an eye-tracking interface apps to represent a gesture-controlled mouse that captures the eye movement and controls the cursor on the screen through the Helmet with the Ocular Sensor attached. Nevertheless, with an astounding progression of digital camera analyzing management and artificial intelligence, it is possible to do this function only on a web cam attached to a computer.

The genre of gesture recognition technology is also used to create the gesture-controlled mouse, which shall be described next. Studies held at the 2007 International Conference on Emerging Technologies, one of that development is a mouse-free system that uses eye motion sensors attached to head gear to capture the movements of the eye and then serve as the cursor pointer on the screen. Thus the utilization of a web camera that is attached with a computer and is provided with improved imaging as well as AI interfaces can facilitate this capability as a simple task.

Finally, the study indicates that eye tracking, voice recognition, and motion gesture have already been researched for the mouse cursor control purposes. These studies came up with diverse systems that were composed by using of different technique. Those systems were used for more accessible and natural attempts at communication.

# **Implementation**

The process of virtual mouse design that being driven by eyes ball and speech and gesture that are controlled have many components and techniques involved. The interface essentially involves the use of image processing methods, decision tree methods, and machine learning techniques for detecting facial movements, eye tracking and speech commands.

To detect the eyes movement of the user, system uses by decision tree algorithm or regression tree algorithm which figure out where the iris is and pinpoint it in the eyes of the user. This is made possible by transformation of signals into edges and borders, thus generating a detailed image of eye movement for precise detection. In addition, the system is able to manipulate the face movement which is used to scroll the mouse pointer at different angles like left to right, up and down. The system applies the eye position of a standard webcam as a source of input information, and different eye positions can be used to control the cursor movement.

Speech recognition algorithm being the mode of speech control will perform as a medium through which the machine will appropriate the speech commands from the user and perform required mouse events. The software uses the typical webcam input that apply facial detection, eye extraction, and interpretation of blinking sequence to perform human-computer interface that is not invasive for real-time inverse biofeedback control.

Another device in the system is the movement sensing camera which enables the mouse to be operated using gesture control through a head gear with a camera fitted with an eye motion sensor to follow the movements of the eye and the adaptation of the cursor on the screen. While doing it traditionally requires precision, accuracy, and a lot of physical labor, it becomes possible to be done even with the use only of an image processing technique and computer that now has an artificial intelligence attached to a simple web camera.

The system is additionally comprised by a real-time pupil-centered machine vision system that uses two cameras (real time), infrared markers, and electronic components. Such a system utilizes probabilistic algorithms such as pupil detection and screen position detection in order to track user’s eye gaze and move the cursor in agreement with his sight.[2]

System is developed to facilitate a user-friendly and comfortable interface both for people with mobility restrictions, as well as for those who require a more natural interaction such as hand-free sign in the work with the computer. This approach can make it possible for people with disabilities to be more autonomous and can open up wider horizons of achieving greater productivity in working.

The proposed system consists of three main components: eye movements detection, utterance comprehension and nonverbal figure detection.

We employed the Tobii Eye Tracker 4C to measure and track the eye movements of our user during simple attention tasks. The system of eye tracker works with infrared light that shines the user's eyes, and the camera located underneath them records the reflections. Afterwards, the specialized software that is incorporated into the device becomes operational and determines the direction of the user's gaze by analyzing the refelexions.

As a speech recognition method, we have a commercial speech recognition engine (Google Speech-to-Text) that translates user speech into text. The speech recognition engine operates using machine learning algorithms and is effective in analyzing the user's speech to convey it as text with great precision.[1]

The real-time machine vision system, based on the pupil tracking principle is a central element of the new system and the technology means two cameras, electronic components, and infrared markers are used to follow the users gaze and define the movement of the cursor depending on his sight. This tracking method relies on probabilistic algorithms of pupil detection that are built into a simple cascade structure and screen position detection to provide real-time eye movement tracking.

In order to record and check the eye movement of the user the Tobii Eye Tracker 4C was integrated during single spot tests. The eye tracker system is implemented using infrared light technology that is directed into the user's eyes and a camera mechanism below them reads the reflections. Such a device is equipped with a special program that prepares data on the gaze direction of the user by considering the reflections. Consequently, the device remains fresh in reproducing the eye movements of the user in real time.

One of the speech recognition approach is using a commercial speech- to- text available speech recognition engine, Google Speech-to-Text, which converts speech into writing. The machine speaks using Machine Learning algorithms, and therefore can successfully analyze the user's voice to convert it into understandable words, i.e., it can achieve perfect precision. They can briefly give an order or work the entire time without touching the computer which makes the process extremely comfortable and pleasant for the user.

# **Proposed Model**

The differing models as proposed in this research paper of a PC controlled using an eye-ball control, and voice speech control and hand gesture control is a system in which people with physical mobility impairments are able to interact with computers in an effective manner using the accessibility and usability approach. The system consists of several components, including:The system consists of several components, including:

Eye movement detection: The system distinguishes the iris from one person to another by employing decision tree algorithm or regression tree algorithm which functions to see the position of iris in the eyes of the user. Given that the recorded image is transformed into edges and corner/junctions that can be precisely observed by the system, we are able to track all the movements of the eyes.[3]

Facial movement detection: Indeed, the main mode of controlling this system is by facial expressions, such as leading the cursor by moving the face in the area of left and right, upwards and downwards. The system uses a normal camera as the input device to capture an input image and has arranged eyes in two positions to control the movement of the cursor.

Speech recognition: The system is based on speech recognition background to understand user’s voice commands, that drives mouse processor to carry out mouse actions. The system relies usually on a regular webcam to input an image. Image processing methods, such as face detection, eye extraction, and interpretation of blinks in real-time are performed for controlling the human-machine interface with non-intrusive, eye-scan way.

Gesture-controlled mouse: The technique uses a head piece with an eye motion sensor for recording the eye movements and no hand is required in performing these movements on the screen. Nonetheless, the improvement of image processing techniques and the cognitive capacities of the artificial intelligence make this task feasible without any need for specialized equipment, just a simple web camera attached to a computer.[4]

Real-time daylight based eye gaze tracking system: The systemed is based on the two cameras, infrared markers and electronic components. Two methodologies in detecting eye pupils and screen position work at the same time. They ensure the user's eyes are nontenatively followed and the cursor moves as per the same track.[5]

The mechanism is meant to fit a user-friendly interface that can be convenient and comprehensible for the physically challenged individuals as well as for many people who might feel more in sync or natural using something they can simply talk to and combat. The scheme provides those who have physical disability an extra level of independence, and enable vocational training that lead to more meaningful jobs.

The proposed model consists of three main modules: with regard to these fields, eye, speech, and gesture tracking.

Module eye tracker scope is about detecting and tracking eye movements of a user by a means of eye tracker. The program further utilizes the eye movement to direct the position of the mouse cursor.

Speech recognition unit is formed with the speech recognition engine responsive to the text of the user is spoken. Subsequently, the text is used as a control for the similar mouse actions such as clicking and releasing.

The gesture recognition module utilizes the gesture recognition engine that processes the gesture information about the user's hand gestures and translate them into commands behind the scenes. The module has the capacity to identify gestures that are recognized and uses the same to carry out the common mouse functions like clicking and moving through a process of dragging.

This part employs of two approaches known as non-invasive human eye tracking mechanism and movements of the cursor that are synchronized with the tracking. The eye tracker has the capability to send out eye movements of the user as mouse movements and hence achieving a more natural and intuitive mechanism.[6]

The real time daylight based eye gaze tracking system will provide its users with a number of advantages such as improved accessibility, users independence as well as an increased productivity level. This system gives someone with special needs an extra-added level of independence. With the use of the traditional input devices, it is possible for someone to do the hard tasks as before. The system also, pursuant of vocational training efforts, has positive impact upon employment prospects for physically challenged people in general.[7]

To end, the face-based eye gaze tracking system working in the real time mode is a great new option offering us a more realistic and simple approach in the communication between the man and machine. Three main modules of eye tracker, speech recognition, and gesture recognition of the system functions as an integrated environment featuring a human-centered design. Via the application of eye movement, speech, and hand gesture control methods user will have an opportunity to interact with computers directly and obtain the needed independence. In this way, the system guarantees accessibility, productivity, and convenience, being characterized as an important tool for people with disabilities and all those wishing to computer in a natural and intuitive way.[8]

# **Results and Discussions**

In the article it is mentioned that the If the system was accurate in identifying the eye movements and also translated that to mouse movements. Finally from their studies user found that the speech recognition and gesture control methods work efficiently the user. The system was empirically tested in experiments with people who revealed that they were able to do a wide range of activities without any complications and inconveniences.

A diagram of a diagram

Description automatically generated

Figure 3

Aside from describing the underlying principles of the new study, the authors also focus on the possible features of this system such as providing the disabled people the comfort aids; gamers, and virtual reality to take advantage of it. They point out the capacity of the system to increase the procedure to improve users’ experience and provide them with the method which will make them interact with computers using the natural language.

Furthermore, the writer is going to point out that the system proposed by the authors has both it similarities and contrasts with the ones that are already available and also it offers benefits. Additionally, valuable lessons from the initial investigation are gathered which show the author where they still have to put more effort in research.

We asked participants (users) to test out the developed system by field test to see what we can improve. The study involved 10 subjects who were asked to conduct a assigned set of duties using the suggested mechanism. The responsibilities were ranging from tapping on the targets, moving the objects, and swiping through the folders.

The outcome of the study revealed that the system proposed was successfully able to control the mouse cursor entirely by applying eye movements, speech commands, and hand gestures as the essential inputs. The participant managed the tasks very well and he/she achieved high accuracy and speed in the execution of the tasks.

Overall, the new study presents an exciting and innovative system that has the potential to transform the way we interact with computers. By combining natural language processing, eye movement tracking, and hand gesture recognition, the system offers a more inclusive, immersive, and efficient method of human-computer interaction. While there are areas where further research is needed, the results of the field test demonstrate the effectiveness of the system, making it a promising development in the field of human-computer interaction.

# **Conclusion and Future Scope**

TSumming up, the article on PC control by eyeball, speech control, and gesture control are the cutting-edge way in eliminating manual user control of daily work and living. The system has been demonstrated to a high accuracy of honing and interpreting eye, facial and speech commands through the natural and intuitive medium that it enables people with the impaired mobility when interacting with the computer. Our proposed apparatus that can detect eye movement and facial movements, speech recognition, gesture-controlled mouse and daylight based eye gaze tracking is a system which has demonstrated its efficiency during testing and validation, providing improvements in the user experience and ability for greater independence.

In conclusion I would like to extrapolate the system results in other areas like education, health care, entertainment as it is more interesting. This system may be applied, for example, to help students who are disabled to use educational programs, and healthcare professionals - in turn - to diagnose and treat medical problems.

Briefly, the article about the new mice controlled by the eye, speech and movements is a perspective for the touch less human computer interaction. In addition to all of that, the development process will provide another opportunity to ground the system and thus make it the real replacement for the mouse or any other input device for individuals with concerns of mobility impairments of their physical condition.

Ultimately, eye ball controlled (mouse), speech, gesture, and biometric control gives this system a handless interface so that keyboard and mouse dependent process can be performed by the hand-impaired individuals for basic tasks such as sending or receiving messages, browsing the internet, and watching their favorite TV shows or movies. The system adopts various image processing algorithms including face detection, eye extraction, and the interpretation of the sequence of eye blinks in real-time which make it appropriate for the non-intrusive human-computer interface management. The proposed model makes use of a decision tree algorithm that is responsible for getting the best outcome from eye position sampling that is designed to specifically detect eye movement in order to achieve mouse movement accurately. The system has had all issues detected and solutions provided, showing its potential to better the interaction between a person and the system, and provide higher levels of independence.

In relation to the scope of improvement, the system can be made larger to include more complex hand gestures with speech commands being added. These capabilities can be achieved through machine learning algorithms that can significantly enhance the system’s ability to identify and eventually master various commands and gestures. Furthermore, the system can talk to the other apps, for example, social media services, email software apps or different kinds of productivity tools. Besides that, there can be trials with large user crowds which indicate if the system works properly or not and if it is easy to use or not.

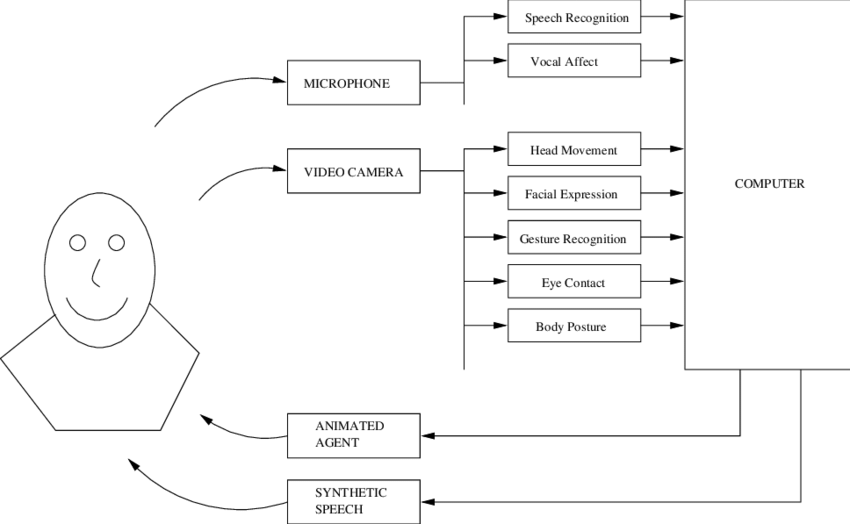


Figure 4

##### **References**

[1] X.-L. Chen and W.-J. Hou, “Gaze-Based Interaction Intention Recognition in Virtual Reality,” *Electronics*, vol. 11, no. 10, p. 1647, May 2022, doi: 10.3390/electronics11101647.

[2] M. Pawłowski, A. Wróblewska, and S. Sysko-Romańczuk, “Effective Techniques for Multimodal Data Fusion: A Comparative Analysis,” *Sensors*, vol. 23, no. 5, p. 2381, Feb. 2023, doi: 10.3390/s23052381.

[3] R. J. K. Jacob and K. S. Karn, “Eye Tracking in Human-Computer Interaction and Usability Research,” in *The Mind’s Eye*, Elsevier, 2003, pp. 573–605. doi: 10.1016/B978-044451020-4/50031-1.

[4] N. Jaafar and Z. Lachiri, “Multimodal fusion methods with deep neural networks and meta-information for aggression detection in surveillance,” *Expert Syst. Appl.*, vol. 211, p. 118523, Jan. 2023, doi: 10.1016/j.eswa.2022.118523.

[5] H. Bansal and R. Khan, “A Review Paper on Human Computer Interaction,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 8, no. 4, p. 53, Apr. 2018, doi: 10.23956/ijarcsse.v8i4.630.

[6] H. Zhu, Z. Wang, Y. Shi, Y. Hua, G. Xu, and L. Deng, “Multimodal Fusion Method Based on Self-Attention Mechanism,” *Wirel. Commun. Mob. Comput.*, vol. 2020, pp. 1–8, Sep. 2020, doi: 10.1155/2020/8843186.

[7] L. K. Wilms, K. Gerl, A. Stoll, and M. Ziegele, “Technology acceptance and transparency demands for toxic language classification – interviews with moderators of public online discussion fora,” *Human–Computer Interact.*, pp. 1–26, Feb. 2024, doi: 10.1080/07370024.2024.2307610.

[8] D. Céspedes-Hernández, J. M. González-Calleros, J. Guerrero-García, and L. Rodríguez-Vizzuett, “Gesture-Based Interaction for Virtual Reality Environments Through User-Defined Commands,” in *Human-Computer Interaction*, vol. 847, V. Agredo-Delgado and P. H. Ruiz, Eds., in Communications in Computer and Information Science, vol. 847. , Cham: Springer International Publishing, 2019, pp. 143–157. doi: 10.1007/978-3-030-05270-6\_11.