# Human Computer Interaction Using Eye and Speech: The Hybrid Approach

**Hansaraj S. Wankhede, S. A. Chhabria, R. V. Dharaskar**

***Abstract - The physically impaired users cannot handle the traditional input devices such as keyboard, mouse etc. the alternate for this category of users must be available. Speech is another promising technology to achieve this goal. The first approach researches estimation of eye gaze point as pointing device. The second approach researches speech recognition as an input and the third approach deals with the hybrid approach for the combination of both.***

***The aim of ongoing research is to develop an application to replace a computer mouse for a people with physical impairment. The application is based on an eye gaze estimation algorithm and assumes that the camera and the head position are fixed. The system after successful development will able to interact user with specific application.***

***Keywords- Eye Gaze Point, Disable users, Speech Recognition, Hybrid approach etc.***

## I. INTRODUCTION

As the use of computer is increasing day by day, we cannot consider our life without computer. The internet technology plays a very important role to update our knowledge so it is very crucial part of our career. Unfortunately, the use of computer is limited to only those who can handle the input devices such as keyboard and mouse. Though the technologies are changing very rapidly the human computer interaction methodology does not provide a solution for those peoples who are suffering from the motor disability. So the physically challenged peoples are away from the use of computers. Therefore it is very necessary to take part in the research in the human computer interaction field and found solution how it would become possible to interact the user with computer in another way.

## II. SIMILAR WORK

Eye tracking is somewhat unusual as a field, in that it has been the subject of intense research for decades, usability and cost efficiency to become a widespread means of human-computer interaction.

The techniques used for the eye tracking are as follows:

First is a biological measurement technique called an Electro-Oculogram (EOG). The device consists of pairs of electrodes attached around the eye (often either right and left or top and bottom). Inside of the eye is an area called the retina, which carries an electric charge gradient.

When eye rotates, this charge gradient produces a potential difference between opposite sides of the eye, which can be detected by the electrons. Unfortunately, this signal is easily corrupted and tends to drift, making accurate detection difficult.

The second method uses plain visible-light cameras and computer-vision techniques to extract details about the position of various interesting features. The growth of the computer vision field in the last ten to fifteen years has led to a multitude of techniques that are capable of performing such analysis. One benefit of this method is that it doesn't rely on characteristics that are extremely specific to the eye (e.g. retinal charge gradients or infrared pupil reflection), and can be tailored to other features of more complex interactions.

The third method is the "Dark Pupil/Light Pupil" technique using infrared light. Under infrared illumination, the pupil becomes very white, almost the exact opposite of its visual-spectrum appearance. By capturing both the dark and light pupil images, the high contrast (which is mostly localized to the pupil) can be used via image subtraction to evaluate the pupil location with very high accuracy. [4][5] [6].

## III. HARDWARE SETUP

Proposed solution was made from the parts available in the market. Construction of eye gaze tracking system are modeled on articles [01,02,03,04] and the tests carried out during the construction. The head mounted system consists of modified web camera, Cap, IR LED, Negative film etc. Construction of system can be divided into three stages: the creation of the capture module, mounting hardware and the creation of infrared illumination. The list of items that were used during the construction of Head mounted system are illustrated in table 1.

| Part name | Quantity |
|---|---|
| Webcam iBall Night Vision | 1 |
| Cap | 1 |
| IR LED | 6 |
| Negative film | 20 cm |
| Pin | 2 |

Table 1 : Hardware for Head mounted tracking system.


Figure 1: Elements of Head mounted Eye tracking system
A) Cap, B) iBall Web Camera, C) IR - LED, D) Film

**Manuscript received on May, 2013.**

**Hansaraj S. Wankhede,** M. Tech. (IV Sem Pursuing) GHRCE, Nagpur, India.

**Ms. S. A. Chhabria,** Head IT Deptt. GHRCE, Nagpur, India.

**Dr. R. V. Dharaskar,** Director MPGI, Nanded, India.

The capture module is responsible for providing an image of the eye to the computer. It was created with iball night vision webcam. The software uses algorithms based on the image obtained in infrared light. Available webcams work in the visible spectrum. It is necessary to modify the camera and mount a suitable filter that allows capturing images in infrared light. The first step in modify a webcam is a complete disassembly of the outer casing. At the back of the camera is a screw, unscrew it and then release normal LEDs. The key operation of the capture module is to transfer images took in the infrared to a computer. In the camera lens should be installed filter that stops the rays of visible light and transmits infrared rays. It allows webcam to capture images in a way similar to the human eye. The filter was removed by undermining its banks by thin knife. In place of it is added infrared filter. Professional IR filters are relatively expensive. Foundation to create head mounted eye tracking system to gaze tracking was minimize the cost according to needs. Infrared filter in construction of the capture module was created from negative film.

This academicals solution seems not very professional however, brings the intended results. Most relevant piece is at beginning of the film frame just before the first photo shoot and is uniformly black. Used to color film the black and white film cannot achieve the desired effect. Square was cut of the film portion of the size such as disassembled filter approximately 5mm x 5mm and then placed in the recess of the lens. So the modified lens is mounted back on the chip. Capture module is ready.

## IV. SYSTEM ARCHITECTURE

The system architecture consists a different modules as shown in figure 2. Each module performs the specific task, The Eye tracking module uses various image processing techniques and the tracking algorithm for estimating the direction of the pupil. The Speech Recognition module recognize the spoken words and compare them according to the specified grammar.

The estimated values obtained from the Eye tracking module and the Speech recognition module are given to the Fusion module as an input. The output of the Fusion module will be generated depending upon the type of fusion the output will be estimated and further it is implemented on the application.

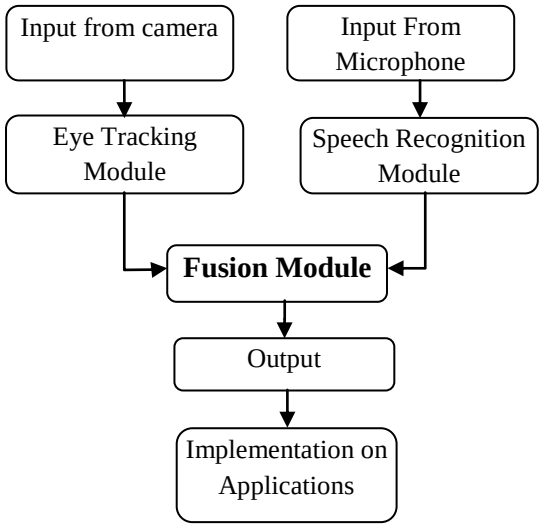



Figure 2 : Fusion of Eye tracking and Speech Recognition

## V. METHODOLOGY

### A. *Design & Implementation of Eye Tracking Module*

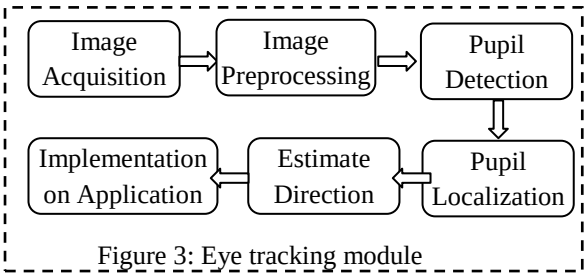



Figure 3: Eye tracking module

Eye tracking is a technology which has many applications in various research fields, notably psychology, medicine and computer science/engineering. Eye tracking enables system awareness of a person's gaze in relation to a shared environment - giving an indication of the person's focus of attention. The environment may be real or virtual.

Eye tracking systems broadly fall into two categories - head mounted and remote.

The head mounted eye trackers mostly use reflected light to track the eyes [11][12]. Camera suspended from a contraption mounted on the head capture video of the eye. The eye position is determined by shining a light source at the eyeball and measuring the distance between the light reflection and a feature of the eye (e.g. pupil). Head mounted trackers are accurate but can be intrusive and feel unnatural to wear, although component miniaturisation has extended their utility.

Remote eye trackers use multiple fixed cameras in the environment fixating on the face. The remote cameras are sensitive enough to capture images of the eyes and head position [5][10]. Advances in image processing and computational power have made remote tracking more popular, although the quality of the tracking is inferior to head-mounted tracking.

The primary information extracted from eye trackers is a person's eye position in relation their field of view.

In this project, real-time eye tracking is achieved with various processes[7][8][10][12].

i. *Step 1 - Image Acquisition:*
The Eye image was acquired through the iball night vision web camera.

ii. *Step 2 – Image Pre-processing:*
After image acquisition pre-processing is required to convert the acquired image into greyscale and further binary image.

iii. *Step 3 - Eye and Pupil Detection:*
The upper and lower threshold was applied to extract the pupil from the eye image.

iv. *Step 4 - Pupil Localisation:*
The new Region of interest (ROI) is considered for the further processing to detect location of the pupil in real time.

v. *Step 5 - Estimation of Direction :*
The current direction of the tracking module is estimated on the basis of the location of the pupil at the particular place in the ROI of the eye image.

### B. *Speech Recognition Module*

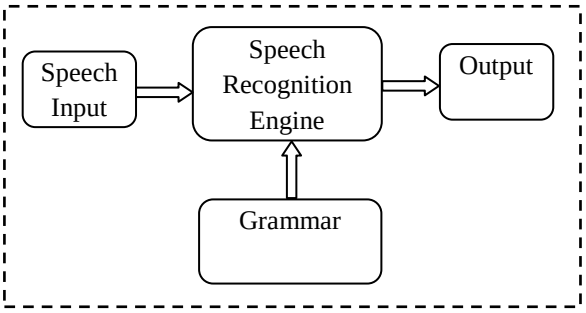


Figure 4 : Speech Recognition Module

Speech is a natural mode of communication for people. We learn all the relevant skills during early childhood, without instruction, and we continue to rely on speech communication throughout our lives.

The accuracy of any system - can vary along the following dimensions:

- *Vocabulary size and confusability.*

As a general rule, it is easy to discriminate among a small set of words, but error rates naturally increase as the vocabulary size grows. For example, the 10 digits "zero" to "nine" can be recognized essentially perfectly , but vocabulary sizes of 200, 5000, or 100000 may have error rates of 3%, 7%, or 45%. On the other hand, even a small vocabulary can be hard to recognize if it contains confusable words. For example, the 26 letters of the English alphabet (treated as 26 "words") are very difficult to discriminate because they contain so many confusable words (most notoriously, the E-set: "B, C, D, E, G, P, T, V, Z"); an 8% error rate is considered good for this vocabulary

- *Speaker dependence vs. independence.*

By definition, a speaker dependent system is intended for use by a single speaker, but a speaker independent system is intended for use by any speaker. Speaker independence is difficult to achieve because a system's parameters become tuned to the speaker(s) that it was trained on, and these parameters tend to be highly speaker-specific.

- *Isolated, discontinuous, or continuous speech.*

Isolated speech means single words; discontinuous speech means full sentences in which words are artificially separated by silence; and continuous speech means naturally spoken sentences. Isolated and discontinuous speech recognition is relatively easy because word boundaries are detectable and the words tend to be cleanly pronounced.

- *Read vs. spontaneous speech.*

Systems can be evaluated on speech that is either read from prepared scripts, or speech that is spoken spontaneously. Spontaneous speech is vastly more difficult, because it tends to be peppered with disfluencies like "uh" and "um", false starts, incomplete sentences, stuttering, coughing, and laughter; and moreover, the vocabulary is essentially unlimited, so the system must be able to deal intelligently with unknown words.

- *Adverse conditions.*

A system's performance can also be degraded by a range of adverse conditions. These include environmental noise (e.g., noise in a car or a factory); acoustical distortions (e. g, echoes, room acoustics); different microphones (e.g., close-speaking, or telephone); limited frequency bandwidth (in telephone transmission); and altered speaking manner (shouting, whining, speaking quickly, etc.).

### *C. Fusion of Eye Gaze Point and Speech Recognition*

Fusion is the process of joining two or more things together to form a single entity.

Fusion processes are stated as follows:

- *Levels of fusion.*

One of the earliest considerations is to decide what strategy to follow when fusing multiple modalities. The most widely used strategy is to fuse the information at the feature level, which is also known as early fusion. The other approach is decision level fusion or late fusion [16] which fuses multiple modalities in the semantic space. A combination of these approaches is also practiced as the hybrid fusion approach [16].

- *How to fuse?*

There are several methods that are used in fusing different modalities. These methods are particularly suitable under different settings. The discussion also includes how the fusion process utilizes the feature and decision level correlation among the modalities, and how the contextual and the confidence information influences the overall fusion process [16].

- *When to fuse?*

The time when the fusion should take place is an important consideration in the multimodal fusion process. Certain characteristics of media, such as varying data capture rates and processing time of the media, poses challenges on how to synchronize the overall process of fusion. Often this has been addressed by performing the multimedia analysis tasks (such as event detection) over a timeline [16].

A timeline refers to a measurable span of time with information denoted at designated points. The timeline-based accomplishment of a task requires identification of designated points at which fusion of data or information should take place. Due to the asynchrony and diversity among streams and due to the fact that different analysis tasks are performed at different granularity levels in time, the identification of these designated points, i.e. when the fusion should take place, is a challenging issue [16].

- *What to fuse?*

The different modalities used in a fusion process may provide complementary or contradictory information and therefore knowing which modalities are contributing towards accomplishing an analysis task needs to be understood. This is also related to finding the optimal number of media streams [16] or feature sets required to accomplish an analysis task under the specified constraints. If the most suitable subset is unavailable, can one use alternate streams without much loss of cost-effectiveness and confidence?

### *D. Neural Network for Fusion*

We are using Neural Network approach for data level fusion of eye gaze point and speech. we are using both AND/OR conditions for the fusion. AND type fusion can be used in a situation where both the input must be true viz. applications used for security purposes. OR type fusion can be used to interact disable users with the computer system. The users can operate the application through eye or speech.

The Perceptron model used for the fusion of eye gaze point and speech.

## VI. APPLICATION

### *A. Experimental Setup*

Figure shows the experimental arrangement of the proposed model

Figure 5 : Experimental Setup

### B. *Image Acquisition*

Eye image is captured by the modified web camera(iBall night vision webcam). The web camera is fixed on the cap as shown in the Figure 6.

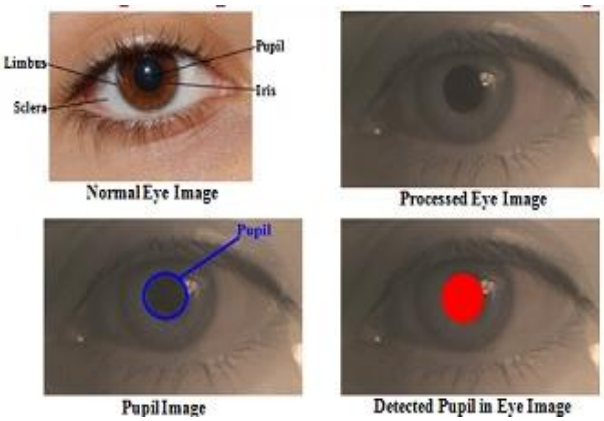



Figure 6 : Image pre-processing

### C. *Eye Tracking*

The pupil is detected as shown in the figure. The tracking is monitored and estimated by Gaze point estimation algorithm. In the proposed system voting scheme is used to estimate the localization of the pupil.

The ROI(Region of Interest) is considered for the estimation of the location of the pupil.[8][9][11] The Gaze directions such as Left, Right, Up and Down is determined on the basis of the change in position of the pupil in real time as shown in figure.

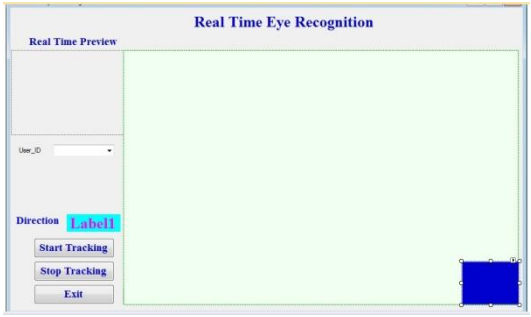



Figure 7 : Eye Recognition Module

### D. *Speech Recognition*

The speech recognition module is used to change the position of the object on the basis of the speech command. we have used Microsoft speech recognition engine to implement the speech recognition.

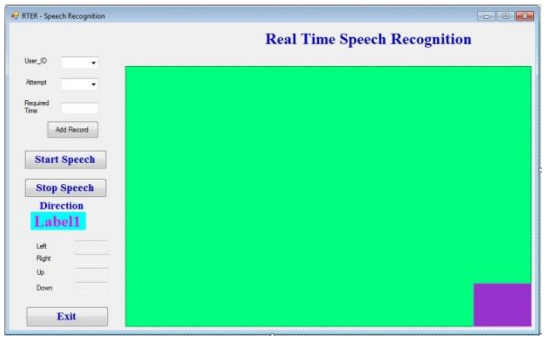



Figure 8 : Speech Recognition Module

### E. *Fusion of Eye gaze Point and Speech*

We have fused the output obtained from the eye Gaze Point Estimation module and Speech Recognition module using AND Level fusion and OR Level Fusion.

| TRUTH TABLE FOR AND LEVEL FUSION | | | TRUTH TABLE FOR AND LEVEL FUSION | | |
|---|---|---|---|---|---|
| EYE I/P | SPEECH I/P | FUSION O/P | EYE I/P | SPEECH I/P | FUSION O/P |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 1 |
| 1 | 0 | 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 | 1 |

Table 2: Truth Table for AND/OR Fusion

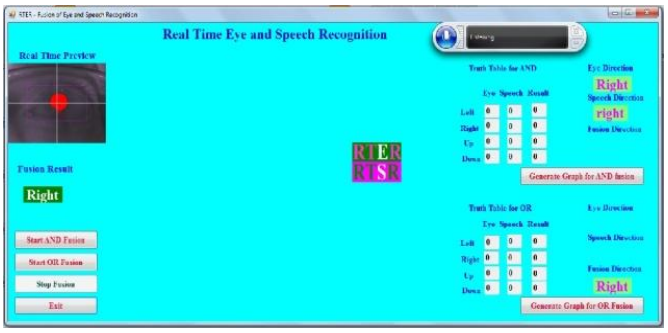
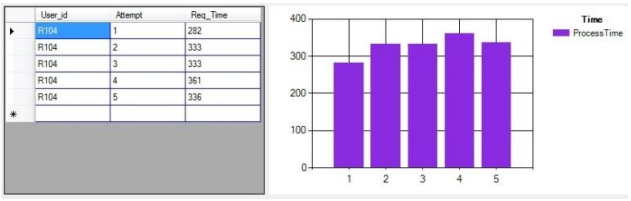



Figure 9: Fusion of Eye gaze point and Speech Recognition

From the above truth table it is cleared that for AND Level fusion both Eye Gaze Point and Speech must be same to move the object in a specific direction such as Left, Right, Up, Down. On the other hand for OR Level fusion only either the Input is required.




Graph 1 : User's Attempt wise Performance


Graph 2 : Average of all User Performance

## VII. CONCLUSION AND FUTURE WORK

This paper Demonstrate the various techniques used for the implementation of the fusion of Eye gaze point and speech in real time. The required hardware and modification in the camera according to the requirement for the pupil detection and localization. The complete model tested for the required results. The results were obtained successfully in real time.

We have performed different test for individual result of Eye gaze point and speech recognition separately as well as combination of both according to AND Fusion and OR fusion and obtained the results successfully.

Our Future work will be the development of The multimodal system that would be available at low cost, easy to use, and more accurate than the previous one.

## REFERENCES

[1] Jason S.Babcock and Jeff B. Pelz. "*Building a lightweight eyetracking headgear*." Eye Tracking Research & Application, Texas, 2004.

[2] J. Babcock, J. Pelz and J. Peak. "*The Wearable Eyetracker: A Tool for the Study of High-level Visual Tasks*". February 2003

[3] Li, D., Babcock, J., Parkhurst, D. J. openEyes: "*A low-cost head-mounted eye-tracking solution*". Proceedings of the ACM Eye Tracking Research and Applications Symposium 2006.

[4] Javier San Agustin, Henrik Skovsgaard, John Paulin Hansen, Dan Witzner Hansen. "*Low-Cost Gaze Interaction: Ready to Deliver the Promises*". CHI 2009, Boston, Massachusetts, USA.

[5] Michał Kowalik, *"How to build low cost eye tracking glasses for head mounted system "*, September 2010.

[6] Masrullizam Mat Ibrahim, John J Soraghan, Lykourgos Petropoulakis, *"Non Rigid Eye Movement Tracking and Eye State Quantification"*, IEEE,2012.

[7] Mehrube Mehrubeoglu, Linh Manh Pham, Hung Thieu Le, Ramchander Muddu, and Dongseok Ryu, "Real-Time Eye Tracking Using a Smart Camera", IEEE,2012.

[8] Peter M. Corcoran, Florin Nanu, Stefan Petrescu,and Petronel Bigioi, "Real-Time Eye Gaze Tracking for Gaming Design and Consumer Electronics Systems", IEEE,2012.

[9] Chiao-Wen Kao, Bor-Jiunn Hwang, Che-Wei Yang, Kuo-Chin Fan, Chin-Pan Huang, "A Novel with Low Complexity Gaze Point Estimation Algorithm", IMECS, March 14 – 16, 2012, Hong Kong.

[10] Minoru Nakayam , Yuko Hayashi , "Prediction of Recall Accuracy in a Contextual Understanding Task Using Eye Movement Features", IEEE, 2011.

[11] Corey Holland, Oleg Komogortsev Department of Computer Science Texas State University – San Marcos, "Eye Tracking on Unmodified Common Tablets: Challenges and Solutions", ACM Symposium on Eye Tracking Research & Applications (ETRA 2012)

[12] Vazquez L. J. G, Minor M. A., Sossa A. J. H. , "Low Cast Human Computer Interface voluntary Eye Movement as communication system for disable people with limited movement.", IEEE, 2011.

[13] Dan Hartescu, Andreas Oikonomou, "Gaze Tracking As a Game Input Interface", The 16 th International Conference on Computer Games ,CGAMES 2011.

[14] Jae Won Bang, Eui Chul Lee, Kang Ryoung Park, "New Computer Inteface Combining Gaze Tracking and Brainwave Measurements", IEEE, 2011.

[15] Giancarlo Lannizzotto, Francesco La Rosa, "Competitive Combination of Multiple Eye Detection and Tracking Technique", IEEE,2010.

[16] Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, Mohan S. Kankanhalli, " Multimodal fusion for multimedia analysis: a survey" , Multimedia Systems, Springer-Verlag 2010.

**Hansaraj S Wankhede** received his B. E. degree in Computer Engineering from Umrer College of Engineering, Nagpur University, Maharashtra, India in 2010. And pursuing M.Tech. ( Final Year) in Computer Science and Engineering from G.H. Raisoni College of Engineering Nagpur, Maharashtra, India.

His research interests include Image Processing, Pupil Detection and Localization Algorithm, and Gesture Recognition. At present she is engaged in Real Time Eye Recognition technique in Human Computer Interaction for Gesture Recognition under the guidance of Prof. S. A. Chhabria.