International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Headspeak: Morse code Based Head Gesture To Speech Conversion Using Intel Realsense™ Technology

Rupam Das, Kuderu B. ShivaKumar

Abstract: ***Advancement in technology has ensured better and more intuitive Human Computer Interaction (HCI). However, technological leap has not been able to bridge the gap for individuals with disability, particularly people with different impairments. Work has been done towards translating various sign languages and lip reading, but not much work is identified towards head gesture recognition and head gesture to speech conversion. Paralyzed patients, with ability to only move their head and eye find it rather difficult to communicate and interact with machines. This work presents a novel and efficient technique to map restricted head movements to text format using face and emotion detection capabilities of Intel RealSense SDK using Morse Code Mapping. The work maps UP and DOWN position of the head to DASH and DOT symbols of Morse code for an EYE_BLINK event trigger and SMILE emotion to convert the sequence of Morse symbols to an English character. The new language called HeadSpeak effectively reduces the pain for generating head gestures through 3D face position based gesture detection.The system can produce an average of 19 characters par minute, gesture error rate of .03% and average 4.5 words per minute.***

Keywords: Intel RealSense™ Technology, Face Detection, Head coupled Perspective, Morse code.

I. INTRODUCTION

People suffering from moderate to severe motor disabilities [1] need assistive technology to enable them to communicate with surroundings and improve their way of life. Brain computer interface has been an effort to use patient's EEG signal to generate gestures for controlling devices and communicating important messages. Technologies like sEMG [2] offer an assistive technology for individuals with spinal cord disability to interact with a computer. Most of assistive technologies come with specialized hardware and software to offer assistance to patients through either direct or emulative methods. Visual Speech recognition [3] is one of the techniques that can offer a reasonable solution towards this direction.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Rupam Das, PhD in Computer Science from VTU, Belgaum.
Dr. K. B. Shivakumar Professor, Department of Telecommunication Engineering, Sri Siddhartha Institute of Technology, Tumkur.

Lip reading based speech recognition system can produce high accuracy if the head of the subject is steady over the capture period. Effect of unavoidable head movements and light variance can be compensated using suitable filtering techniques and region based features [4-5].
Bag of Keypoints are used widely [6-7, 28] for effective visual lip reading and eye gesture recognition. Another technique called EOG is also used successfully in order to interact with machines like computers and robots [8].

Most of these are computer aided assistive technologies and are command based which support accurate and nearly accurate recognition of a set of words and gestures [9]. In this juncture Multimodal expressive based on physical movement of fingers, hands, arms, head, face and or entire body is perceived as a definite gesture that is used to convey meaningful information in order to interact with the environment [9]. Speech recognition is one of thecomplementary technologies for Human Computer Interaction (HCI) which is evolved to acommercial grade application. Such recognition is divided into command and dictation modes, wherein the former is a gesture recognition system which recognizes a group of words with very high accuracy and the latter recognizes general words and sentences with lower accuracy. Accuracy of dictation mode can be improved with use and learning. Case study of speech recognition opens up an interesting prospect for more accurate recognition system. Speech recognition with limited number of commands is more accurately recognized and interpreted. Such a system can produce highly accurate result. Again, measuring the standard of a speech recognizer as a trustworthy system can strengthen various real life implementations [10].

Head part of the body is an important aspect in assistive technologies which serve a wide range of purposes for severely disabled people who are left with minimal motor abilities [11]. Both speech and head based Computer Aided Assistive Technology (CAAT) systems are being considered as a low cost viable and efficient HCI tool for such individuals. But the problem with eitherof them is limited set of gestures that they support.Therefore, there is a need fora paradigm shift of the approaches that are adopted to perceive such gestures.
This paper presents a unique method of mapping limited head gestures to complete English vocabulary using Morse Code Mapping (MCM). Morse codes are used as an effective tool in assisting and rehabilitation for the individual who is suffering from motor neuron disease.

Morse Codes have been considered as an effective tool in assistive and rehabilitation tool for individuals with motor neuron disease. An efficient communication tool can be developed for such disabled individuals using a time adaptive filtering and fuzzy based technique for Morse code [12].

The main aim of such tool is to detect the face. Highly accurate real time face detection techniques using cascades and Adaboost proposed by Voila-Jones [13] has opened up new horizons for facedetection based assistive technologies. Efficient face detection and segmentation with high frame rate can be used as an effective tool for head gesture recognition, eye tracking, gaze tracking, lip tracking and lip reading. However, Cascade based technique as proposed by Voila-Jones is considerably slows in detecting face as it searches for face in 2D images.

Utilizing depth data to segment the face from the rest of the background can significantly improve the detection speed and efficiency.

RealSense™ is an exciting new technology offered by Intel. This provides several features like 3D hand tracking, Hand Gesture Recognition (HGR), face tracking, facial keypoints marking, Speech Synthesis, Voice Recognition (Both Command as well as dictation), 3D scanning, object tracking and augmented reality. The SDK also offers a range of other features through a specialized 3D Creative Camera (now called a RealSense camera). Multimodal support of the SDK provides a wide range of possibilities for application development for individuals with disability. The SDK also supports 3D face tracking which increases the face tracking speed and accuracy significantly. In this work RealSense with 3D face tracking was considered for HeadSpeak.The contribution of the work is justified by analyzing few of the existing techniques of natural text inputs and HCI. Various techniques exist for gesture recognition that ranges from computer vision to specialize hardware supported methods. HGR, Body Gesture Recognition, Face Gesture Recognition, Eye Gaze, Eye Gesture Recognition and Visual Lip reading are gesture recognition based technique which has attained significant focus. A wearable solution for HCI using data glove technique is proposed in the past which offers a wearable glove with sensors that tracks the position of the fingers and further converts these positions into definite gestures and is transmitted to a PC using Bluetooth for execution of the task using these gestures. Such a glove enables the user to operate in 3D space providing better degree of freedom [15]. An alternative design to such a glove technique is proposed by Vitor F Pamplona et al. [16] which uses visualmarker along with a camera to track the position of the fingers accurately. This design was bulky which makes it unsuitable for patients with disability. A brighter prospect for gesture recognition in 3D space is provided by [15-16]. HGR [17,21] aims at detecting a) Sign language b) detecting specific gestures like CLICK, SWIPE, SHINK, ROTATE commands. Most of these techniques are divided into mainly four stages: hand detection, position of hand/finger acquisition, filtering and gesture interpretation. Hand detection is mainly categorized into marker based techniques [16]. Model based techniques [17-18], soft computing techniques by using classifiers like Neural Network [19] and Fuzzy [20].

It is observed that model based techniques provide quick and more efficient detection and recognition of gestures [17].

Comprehensive understanding of HGR [29-30] techniques provides a detailed overview of available choices and models for any gesture detection and recognition system. Also it is quite apparent that less number of gestures makes the system more accurate, robust and fast. Thus, fewer gestures in a gesture recognition system will be more accurate in comparison to a system that offers complex gestures. Building a system with fewer gestures and combining these gestures to recognize more complicated gestures can offer a more realistic solution.

Human face is more natural way of interaction than hands. For instance, simple acceptance and denial can be performed using head nods. Emotions like surprise and happiness can provide an important clue to the machine. For instance, when a command is interpreted correctly, user may affirm successful character detection by smiling. In addition, status of human mind and ailment in body parts can be determined by analyzing specific points on the face like the eyes, nose, location of the lips etc. [14]. As the number of possible gestures that can be generated by head movement alone is quite restricted, such a system can be more accurate than any hand gesture based system.

Louis-Philippe Morency [22] proposed a technique which accepts user input through head nods for dialog boxes using stereo based tracking and discriminative based classifier. This technique shows accurate recognition for head nod and head shake gestures as context and vision are combined. Hence, it is evident that visual clues when combined with head gesture recognition could be a very stable system. Further, speech recognition when combined with context based recognition [23] shows that the latter can be easily integrated into a multimodal environment resulting in high precision and accuracy. A framework is presented which combines prosody patterns of the voice and headgesture in a two stage patterns with HMM [27] which strengthens the correlation of head gesture with the context [24]. A real time application of Head Gesture Recognition [25-26] which controls the speed of a wheel chair through UP, DOWN, LEFT and RIGHT movement of the head is presented. Primarily this technique detects the face and then classifies the continuous variations as gesture.

Further template based feature tracking methods have gained much importance as these techniques use facial feature point to control the mouse on the screen. Selection of the best feature point improves the overall performance of a face based input techniques. After carefully reading and analyzing the existing literature on HCI, Vision based interaction and different text entry techniques, it is found that synthesizing mouse like movement to click on-screen buttons need several head displacement, which is difficult with current technology. Restricting the movement could mean more accurate detection. Detecting a Face gesture like LEFT, RIGHT, UP and DOWN is much more efficient than moving an on-screen cursor.

Therefore we propose a new research in this direction called HeadSpeak that helps patients with motor disabilities to efficiently enter text data with ease.

II. MATERIALS AND METHODS FOR PROPOSED WORK

A. Intel RealSense™ technology

Many computer vision libraries are available now for application development. OpenCV is one of the most popular libraries which support Voila-Jones [13] face detection and allows a cross platform application development.

Intel's RealSense™ is one of the most enthralling emerging platforms for intuitive multimodal HCI. This offers Face Detection, Hand Tracking, finger tracking, hand gestures, gesture driven events and triggers, live background separation, speech synthesis, voice recognition (both command as well as dictation

mode). Heart of this technology is a RealSense™ camera which includes color sensor, IR sensor and array of microphone. IR sensor produces depth stream, which is of gray scale in nature, wherein the objects near to the camera appear bright and fades to black as distance of the objects from the IR sensor increases.

Live 3D background separation is a mode which separates user background by projecting the depth stream over the color stream. This enables 3D face tracking. As background is effectively removed by projection, face search region is reduced which helps faster tracking. This SDK provides 3D face tracking with 3D coordinate of the face location where z-axis provides the distance with respect to camera. Further, bag of feature points over face that marks the eye, nose and mouth regions is also provided.

The platform also provides cross platform application development with web based applications and desktop applications. The software supports a wide range of programming languages starting from C++, C#, Java, Java Script to develop intuitive system with the potential of redefining HCI at the same time making the technology available to different devices and applications. Due to accurate tracking and gesture detection capabilities of the camera and the SDK, it is proposed to use RealSense™ in this proposed work.

Section 1 elaborates the importance of multi modal systems where HGR can be combined with face detection and facial expression to produce fascinating results.

B. Problem formulation

We envision that an individual with a motor disability can be provided with an integrated platform with depth sense camera and specialized software tohelp him with an assistive and rehabilitation framework.

Different facial gesture and head movement based input techniques are provided by[27-30]. According to the literature, the face tracking based text entry system can be designed with following general architecture.

1) Face Detection: Adaboost based technique proposed by Voila-Jones [13] is arguably the most popular face detection techniques used by several facial gesture driven systems. However, the problem with Voila-Jones method is that, a random displacement of the face area occurs due to background and motion noise. Such a noise is generally minimized by using appropriate background separation technique. But Adaboost based technique is also quite slow. A 30FPS stream is reduced to 8-10FPS when face tracking is enabled. Thus good 20FPS speed is compromised. Hence, there is a need for faster face detection technique for a more intuitive and responsive system.

2) Facial Feature Tracking: Literature argues that instead of using the entire face rectangle as a tracking boundary, part of the face can be used to track the features. Eye, Lips, Mouth, Nose are common points being considered for feature tracking. In a conventional setup, for instance, the Adaboost technique first detects the face boundary and then on this boundary feature points like eyes are searched using template matching or using cascade classifier. Therefore, even particular feature points also depend upon the effectiveness of the face tracking system. Deviation and distortion in the face boundary also result in misdetection of such feature points. Therefore, more innovative approach is needed for the facial feature extraction.

3) On-Screen Text Entry: There are twodistinct approaches for on-screen Text entry: code based text entry and on-screen cursor based text entry. Morse code driven technique for text entry by recognizing tongue gesture [34] is on the same line of design as that of the proposed work.

Several other techniques exist which uses on-screen visual keyboard and require the user to move a cursor with head movement over appropriate character key button and then emulate the click event. In click based systems, user does not need to remember any code and only needs to follow a visual clue on the screen. However, it is observed that the method is inconvenient as the amount of face displacement needed by the user to cover the entire key region is quite large and also accurately pointing the cursor on the keys is observed to be significantly challenging.

Tongue gesture based text entry is relatively simple and needs only LEFT and RIGHT primary gestures for generating DASH and DOT symbols respectively. It uses a timer to trigger the Morse code to character code mapping. The system converts a sequence of DASH and DOT symbols to English character after an ideal TIME_OUT period.UP and DOWN gestures are used for deleting a character and accepting the text. Using tongue reduces the possibility of extending the technique to any other application beyond Morse code based text entry. On the other hand, face displacement and facial feature point based technique suffers from text input accuracy.

A more robust system is needed for enabling a person with motor disability and limited head movement for text entry and subsequently converting the entered text to speech. Also the method should be adoptable enough for more complicated interactive systems such that the work can be used over a larger framework of rehabilitation and assistive technology.

Next subsection outlines the contribution of the work and justifies the use of every building block in this work.

C. *Contribution*

Section 2.1 provides details about RealSense™. The major contribution of this work is to utilizing 3D face tracking features offered by RealSense™ along with facial feature points. This is used to detect primary head gestures and basic emotions. Further the gestures and emotions are integrated into the design of an effective face based text entry system.

The first step is to obtain some primary gestures like: UP, DOWN, RIGHT, LEFT, LEFT_EYE_CLOSED, RIGHT_EYE_CLOSED, EYE_BLINK, and MOUTH_OPEN as shown in figure 1.

Figure 1: Face Detection and Facial Gesture Recognition

As number of gesture required to efficiently design a Morse code based text entry is limited to two,we can potentially use any pair of gestures to map them with Morse code. For example, we can use UP, DOWN gesture or LEFT, RIGHT gesture or SMILE, MOUTH_OPEN gesture pair in order to map them to dash and dot symbols of the Morse code. SMILE and MOUTH_OPEN gestures were found to be accurate, but detection time was 102ms and 119ms respectively which is little higher in comparison to UP and DOWN gestures which were found to be 69ms and 77ms respectively. Gesture detection time was calculated as the time difference between the startof the gesture generation at the time of gesture detection.Therefore, UP and DOWN gestures were adopted for two primary symbols DASH and DOT respectively. SMILE gesture is used to convert a sequence of symbols into a character using MCM and MOUTH_OPEN gesture is used to remove the last Morse code symbol. LONG_EYE_CLOSE gesture is used to speak out the entered text using SDK's speech synthesis engine. Hence, the primary contribution of the work can be summarized as

a) Detecting head postures from 3D face position data
b) Detecting primary emotion from facial points
c) Mapping the head postures and emotions to Morse code for designing Morse code based text entry system.

Table 1 provides in details, symbols and their use in the proposed work.

Table 1: Gesture and their use in HeadSpeak

Gesture	Usage
FACE_UP	- (DASH)
FACE_DOWN	. (DOT)
EMOTION_SMILE	Morse Code to Character Conversion
EMOTION_MOUTH_O	Delete last entered Morse Code Symbol
FACE_RIGHT	SPACE
FACE_LEFT	Delete last character
LONG_EYE_CLOSE	Speak out entered text
GESTURE_EYE_BLINI	Accept a Symbol

III. METHODOLOGY

Figure 2 shows the overall methodology. Face Detection, Gesture Identification, Emotion Detection and Morse Code Mapping (MCM) are the major building block of the system.

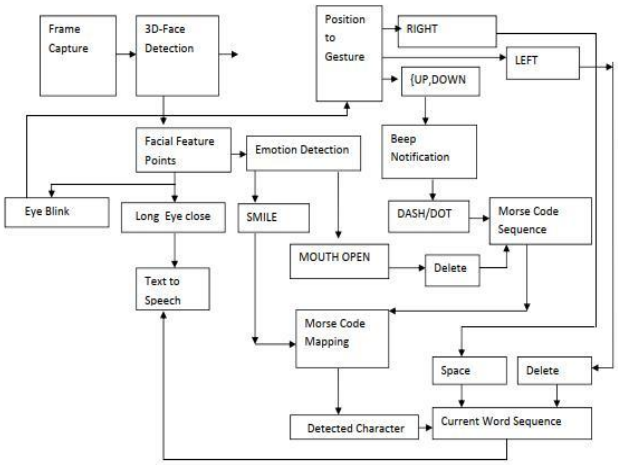

Figure 2: Overall Block Diagram of HeadSpeak

A. Face detection

Intel RealSense™ returns a depth stream through its IR sensor. The SDK also provides a means of 2D as well as 3D face tracking. 3D Face tracking projects thedepth stream over the RGB color stream to separate the background first. Face detection technique adopted by SDK is Adaboost based which significantly improves the efficiency of the detection due to less noise in the form of background information. 3D object tracking of the SDK also provides 3D deviation of any tracked object, thereby returning Yaw-Pitch-Roll of the object from the center of the mass.

When the head is straight and steady, the value of Yaw, Pitch and Roll is zero. The SDK also detects primary feature points on the face which includes a bag of points over eye. Nose and lips area is shown in figure 1. Once the face area is located, facial feature points and the posture are used to detect primary gestures.

B. Detecting facial gesture

In a 2D face tracking system such as Voila-Jones [13] method, gestures are recognized by continuously obtaining the neutral face location and monitoring its first order derivatives dx/dt and dy/dt[26]. Both derivatives are integrated over afive frame periods to calculate deviation in x and y direction, and are classified into UP, DOWN,RIGHT, LEFT gestures.

However, this method suffers from periodic distortion of face detection, which was mainly due to light intensity variation between the frames.Voila-Jones method though accurate does not take into consideration of the past data and frames. Hence, face location is noisy. This variation is sometimes a very significant constraint in efficient head gesture recognition. If dx/dt and dy/dt are updated in every frame without suitable filtering than a head gesture recognition system may become unstable[26]. Running Median filter is observed to be a good solution for the system[32]. However, median filter negates the face derivative on first two frames, increasing the detection time as well as the total deviation needed for detecting the gesture.

In order to overcome this drawback of slow detection rate and large head displacement resulting from 2D face tracking based head gesture detection system, 3D gesture recognition based on 3D displacement vector set of {Yaw, Pitch, Roll} is opted where gesture can be detected based on the current position of the head rather than depending upon the head displacement.

In this work, facial gestures are divided into two categories: Posture Detection and Expression Detection. Posture detection deals with obtaining the 3D displacement of head and detecting UP, DOWN, RIGHT and LEFT gesture from the posture. Emotion detection on the other hand analyses the facial feature points and detects three primary gestures namelyEYE_BLINK, SMILE and MOUTH_OPEN.

B1. Posture detection

Figure 3 demonstrates the variation of Yaw-Pitch-Roll with respect to facial movement.

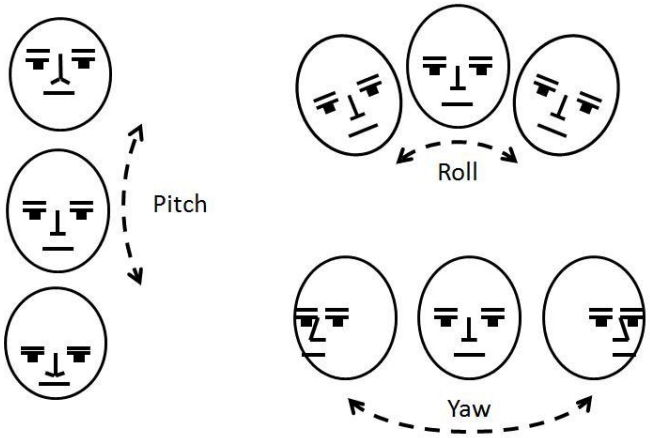

Figure 3: Yaw-Pitch-Roll of the Face (image curtsey msdn.microsoft.com

When face is straight, value of all the three deviations are zero. When the face posture is changed, signs of the Pitch and Yaw change. Roll varies depending upon the distance from the camera. Table 2 represents thesign of Yaw and Pitch depending upon head postures.

Table 2: Head Posture relationship with 3D displacement vectors

Posture	Change	Sign
UP	Pitch	+ve
DOWN	Pitch	-ve
RIGHT	Yaw	-ve
LEFT	Yaw	+ve

As the distance of the face from the camera is increased, Rollis increased. Roll is zero when the face is at the nearest detectable range from the camera. It is observed that it is highly difficult to maintain absolute zero values for vectorseven under steady face. Therefore we use a carefully selected threshold of |5| for detecting deviation. Also, the user cannot move his head up and down without any changes in right and left direction which means that there is no absolute displacement. Maximum displacement vector of {Yaw, Pitch, Roll}is detected. The vector is selected for gesture detection if absolute displacement is greater than 5|.If so, we obtain the sign of the vector and classify the posture depending upon table 2.

B2. Emotion detection

RealSense™ SDK provides corner feature points for eyebrow, both left and right eyes, nose and lip area. For SMILE and MOUTH_OPEN gestures, the feature points over lip area are used. The corner points of the boundary of a horizontal axis and boundary points of vertical axis as the major and minor axis are obtained and are shown in figure 4. When their ratio is less than 1, MOUTH_OPEN emotion is triggered and if greater than 5, SMILE emotion is detected. Similarly, eye closing is detected when the distance between four boundary points over each eye is minimum. Consecutive eye open and close triggers eye blink gesture.

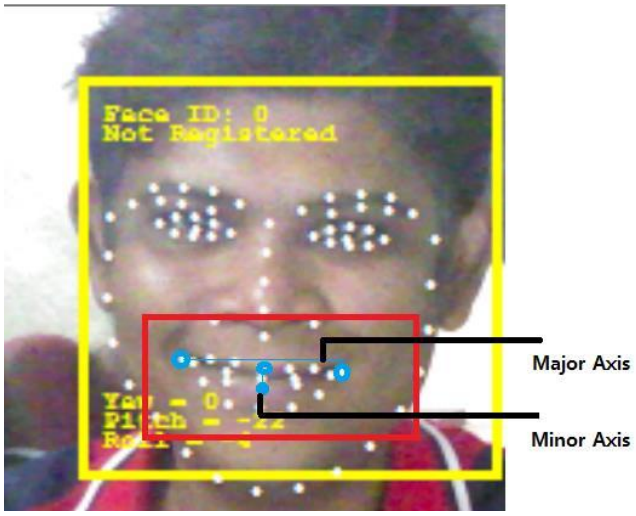

Figure 4: Major and Minor Axis of lip feature point

C. Morse code mapping (MCM)

The core objective of this work as stated above is to convert UP and DOWN head movements to English characters by Morse code mapping (MCM).
Morse code is a method of transmitting text information as a series of on-off tones, lights, or clicks that can be directly understood by a skilled listener or observer without special equipment[31].
They are essentially permutation of two symbols DASH(-) and DOTS(.) that can be used to represent alphabets and numerals. This is similar to variable length binary encoding technique. On an average 3.15 symbols are used to represent alphabets and 5 symbols are used to represent numerals. Morse Code Table is shown in figure 5. When Morse code is transmitted as a stream of sequences, decoder decodes the characters based on inter symbol and inter sequence time delay.

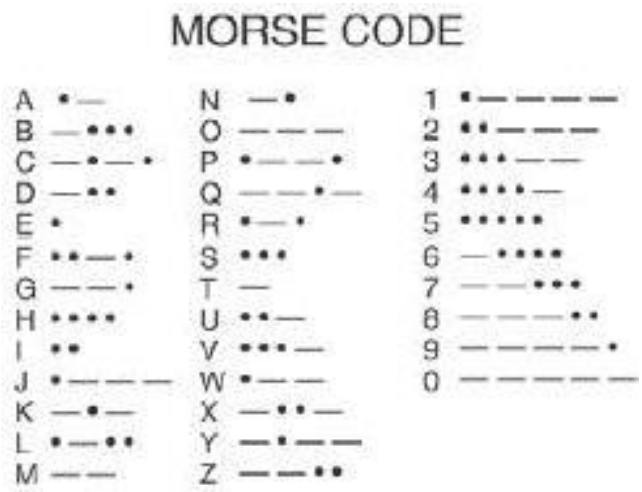

Figure 5: Morse code table for English Alphabet and Numerals

In the proposed system DASH with UP and DOT with DOWN gesture are mapped because when (**-.**) Symbol is presented in GUI for user hint for the characters. User can easily identify DASHas UP and DOT as DOWN by looking at the symbols.
One of the significant problems with traditional head gesture recognition system is that once a user moves his head up for generating UPmovement. The head has to come back. Hence, the next gesture willautomatically become DOWN. The similar problem occurs for DOWN gesture as next gesture would automatically be UP gesture.

Another drawback of derivative based traditional technique is that the user needs to start from a neutral position every time for detection of any gestures. This demands repetitive movement which results in unnecessary delay and causes extra pain in the neck.

However, as the proposed system estimates the gesture based on the sign of the vectors which in turn gets adjusted over the neutral face position, derivative is not needed for detection. This eliminates the drawback of dealing with the head posture needed to bring the head in the neutral position after generating a gesture. User can keep his head up and then blink his eyes twice in order to generate two consecutive DASH symbols.

Thus, the proposed technique overcomes the drawback of head movement based gesture recognition with head position based gesture detection. Hence to produce two consecutive DASH symbols. Userwill not need to take his head upfollowed by a down movement to neutralize the position and then repeat the same again. Instead a user can now keep his head up and blink twice for two consecutive DASH symbols.
For example, with the proposed technique when the user needs to enter '**J**', he can bring his head down once and then keep his head up for three subsequent detections. Therefore, our method significantly reduces the overall movement of head which becomes an important advantage.

It is also considered that an interpretation based mechanism using Dichotomatic Search [33] method can be used to predict the character as soon as symbols are entered. All possible characters that can be generated from current state of symbols can be displayed so that user can select the desirable character just by generating the SMILE emotion without generating the head posture. However, it is noticed that such a search based prediction is less efficient in comparison to detection of the character at the end of symbol sequence as continues searching resulted in slow frame rate. This is because there are numerous suggestions and user again needs a gesture to loop through the suggestions.
All the numerals are five symbols long. Therefore, numerals are omitted from the current work to test the system for suitably of alphabets.

IV. RESULT AND DISCUSSION

A, Participants

The proposed system is meant to assist patients with motor disabilities in their rehabilitation process. However, we could not get any hospitals that offered its patients to volunteer for the test. Therefore we could not provide any patient's aspect in the result section.
For testing the system, a group of twenty under graduate students in SSIT College, Tumkur, Karnataka, India is selected. Out the total participants, four were female and remaining were male students and all were under the age of 22. Participants volunteered to test four distinct methods which included proposed workand three other state of art to test them independently. Details of the tests are elaborated in consecutive subsection.

B. Apparatus

Five Lenovo All in Onedevice with Intel 4th generation i5Core with 8GB RAM running Windows 8.1 operating system, were used for the tests. Intel RealSense™ SDK is installed in each of the devices. RealSense™ camera is provided with each of the fivetest machines. The machines wereallocated to the participants in time slot basis with three hours being assigned to a participant in a session.

C. Design

We developed1) tongue based text entry system[34], 2) pointer based text entry through facial movementand 3) 2D facial gesture detection[26] based text entry system for comparison with the proposed system. It is important to design and build the state of the art rather than adopting the result directly from literature.

This is due to the fact that RealSense™ provides better frame rate and performance over OpenCV. Also, the machines used were relatively high configuration machines. Hence webelieved that comparing the result of a state of the art being tested in P4 machine with 1GB RAM would not provide performance distinction as a lot of performance improvement would be attributed to the system configuration. Designs of other three state-of-the-arts were not difficult as their design was in line with the proposed work.

We useda single GUI for all the four methods. The GUI developed is shown in figure 6.

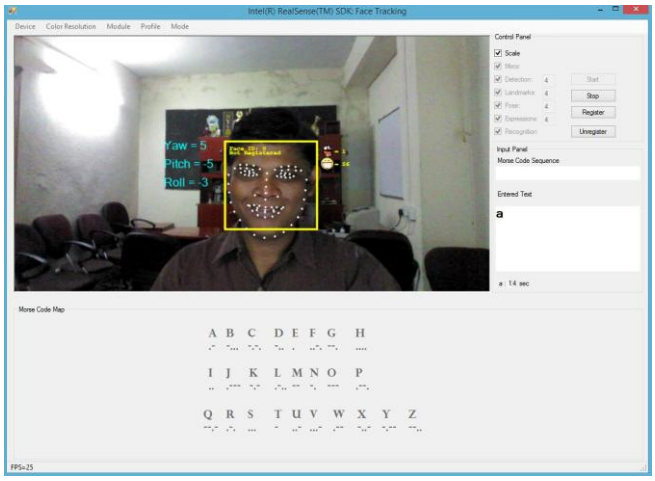

Figure 6: GUI of the proposed system

Figure 7 also demonstrates the head movement needed to detect character A.

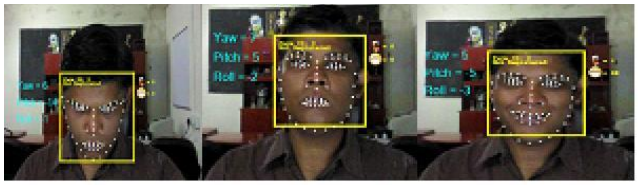

Figure 7: Sequence of movement for detecting character 'A'

D. Procedure

We conducted two distinct experiments. In the first experiment, we asked all the participants to enter the letters from A to Z- 10 times in all the above three techniques as well as the method proposed here. Users used the system once in a session.Users whoare allocated with the machine for three hour sessions were allowed to work in a particular method and a particular test. Therefore, in threehour sessions a user tested any of the four techniques with either of two tests: CharacterTest and sentence test. We provided the sentence *"The quick brown fox jumps over the lazy dog"* to all the participants and asked them to enter the sentence using all four techniques. Users used respective techniques to enter the sentence 10 times each with all four techniques.

E. User tests

We found that in every Morse code based techniques,detection time gradually reduces as the user tends to remember the sequence.

In the second method,i.e. in pointer based text entry system, the detection time increased significantly as the distance of the characters increased from neutral face location on a standard keyboard layout. We found that pointer based technique is difficult for text entry as significant amount of controlled head movement was needed for selecting the characters.

With the proposed method as well as first and third methodswe observed reduction in detection time after averaging four entries of each independent character. The generation timegradually converged to relatively static value.Weobtained the data from each user for each of the techniques which are presented in table 3.

Table 3: Recognition time in seconds for independent English characters

Symbol	Proposed	Cursor	2D Face	Tongue Code
A	1.79	6.83	3.32	4.01
B	3.798	2.1	6.11	7.35
C	4.5	4.4	6.11	7.86
D	3.9	6.1	4.71	5.92
E	0.66	5.6	1.9	2.1
F	4.063	4.4	6.11	7.41
G	3.14	1.4	4.83	6.2
H	2.17	1.1	6.03	7.5
I	1.31	3.2	3.5	3.9
J	4.5	1.6	6.92	7.3
K	4.06	3.2	4.53	5.5
L	5.79	4.6	7.1	7.93
M	2.06	2.4	3.02	4.22
N	2.53	1.8	3.11	3.82
O	3.53	8.7	5.2	6.2
P	4.92	10.3	7.1	8.2
Q	4.31	10.2	6	8.35
R	3.26	2.8	4.054	5.1
S	2	8.9	2.9	3.7
T	0.6	1.7	1.5	1.8
U	2.33	1.2	4	4.4
V	3.19	3.2	5.84	7.1
W	2.86	9.1	4.5	5.39
X	4.26	6.2	5.3	5.98
Y	5.19	0.9	6	7.33
Z	4.73	8.1	5.7	7.01
Average	3.2865769	4.61653846	4.8228462	5.83

It can be observed that the proposed technique achieve significant performance improvement over all the three other techniques. The proposed technique allows the user to keep his head in the current posture if the gestures are repeated. This increases the efficiency of the system multifold times.

All the other techniques needed a median filter to smooth the movement. We observed that the median filter of length 3 reduce the frame rate by about 4.2FpS which results in slow detection. Pointer based technique needed median filter of length 5 for stabilizing the distortion in face rectangle. Lip based technique was more efficient than 2D face detection based technique in terms of effort. Time delay is introduced in this method for character detection and space entry for words resulted in unwanted delay in detection.

Proposed technique detected DASH(-) quicker than DOT (.) BecauseUP posture was detected quicker than DOWN posture. This is because 3D face detection tends to lose the facial points in DOWN posture. Therefore, characters with more DASH were detected quicker in comparison to the characters with more DOTS.

We also observed that prolonged use of the system make it easier for the user to accurately generating the postures.

Having tested the performance of independent characters, wewanted to test the system performance for complete sentences. The result of the given sentence is presented in table 4.

Table 4: Analysis of Text Entry Performance for different techniques

Techniques	Average Time	Word per Minute	CPM
Proposed	138.3	4.9	18.69
Cursor Based	171.5	3.62	15.08
Tongue Morse Code	185.5	3.1	13.94
2D Face Gesture	216.6	2.72	11.94

Average time in seconds is measured as average sentence generation time of all the users. Words per minute are calculated by subtracting the total space generation time from sentence generation time and dividing it bythe number of words.Character per minute is calculated as total sentence generation time divided by number of characters including space. Surprisingly our results in all the techniques were little better than the results claimed by 'existing techniques). This might be due to better camera and processing capabilities of the SDK.

Character per minute(CPM) is considered as a standard performance measure for computer vision based tasks and text entry was superior in theproposed system in comparison to all other techniques.

F. Accuracy and Error Handling

Performance analysis is incomplete without discussing the accuracy of the system. The ultimate objective of the system is to develop headmovement based comprehensive text entry system. Therefore, rather than tracking the accuracy of gesture detection we measured the accuracy of the system by calculating total use of "delete character" and "delete symbol" events including all sessions. We define "Gesture Error Rate(GER)" as the percentage of the total deletion of symbols."Character Error Rate(CER)" is defined as the percentage of character deletion in sentence entering test.GER includes the delete operation performed on both character test as well as sentence test. Results are presented in Table 5.

Table 5: Error Rate Performance of the proposed system in comparison to state of the art.

Technique	GER	CER
Proposed	0.03	0.84
Cursor Based	1.2	3.7
Tongue Morse Code	0.09	0.91
2D Face Gesture	2.4	1.1

We found that 2D face detection causes more misdetection than 3D face tracking technique. GER was more for both 2D face tracking based techniques. Surprisingly cursor based technique had a very high CER. We found that it was difficult for users to steady the cursor over a key while synthesizing the key event. Morse code based techniques had less CER in comparison to cursor based technique. This strengthens the claim that symbol based text entry is more suitable in assistive and rehabilitation framework for text entry. Wealso observed that by distinguishing between symbol generation and the input method, significant performance gain is achieved in terms of reducing both GER and CER.

V. CONCLUSION AND FUTURE SCOPE

New paradigm in research for human computer interaction, rehabilitation and assistive support to individuals with lower body disability is need of the hour. Individuals with such disabilities can communicate only by generating gestures through head, lips, eyes. So far, no significant advancement has been observed in this direction barring a few isolated works. In this work,weproposed head gesture to speech conversion system aided by eye and lip based gestures. Result shows that the system is immensely stable. The efficiency of the system was observed to be better than other techniques in terms of both detection time as well as gesture generation effort. Proposed system tends to become more responsive and faster with increasing use. Prolonged use also increases the accuracy of the detection where use of deleting a symbol and a character reduces. The work can also be thought of a part of an integrated face driven HCI technique which could include several models to communicate with the environment like emulating mouse movement, click[35] and so on.

The system can be evolved into a comprehensive text entry system which may be used widely for rehabilitation of patients with motor disabilities. This can also be adopted as a unique tool in assistive technology.

REFERENCES

1. Cincotti, F., Mattia, D., Aloise, F. et al. "Non-invasive brain-computer interface system: Towards its application as assistive technology", 2008, Brain Res. Bull. 2008, Vol. 75, pp. 796–803
2. Changmok Choi, Youngjin Na, Byeongcheol Rim et al., "An SEMG computer interface using three myoelectric sites for proportional two-dimensional cursor motion control and clicking for individuals with spinal cord injuries", 2013, Medical Engineering & Physics, Vol. 35, Elsevier, pp. 777– 783

3 .Hassanat, A.B.A and Jassim, S., "Visual words for lip-reading", 2010, Proc. SPIE, Vol. 7708, pp. 279-302
4. Takeshi Saitoh, Ryosuke Konishi, "Real-Time Word Lip Reading System Based on Trajectory Feature", 2011, IEEE TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING, IEEE Trans 2011, Vol. 6, Wiley, pp. 289–291
5. Iain Matthews, Tim Cootes, J. Andrew Bangham, "Extraction of Visual Features for Lipreading", 2002, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, No. 2, pp. 198-213
6. Csurka, G., Dance, C., Fan, L. et al. "Visual categorization with bags of keypoints", In: ECCV'04 workshop on Statistical Learning in Computer Vision, Prague, May, 2004, pp. 59–74
7. Min Lin, Guoming Mo, "Eye gestures recognition technology in Human-computer Interaction", 4th International Conference on Biomedical Engineering and Informatics (BMEI), Shanghai, 2011, **INSPEC Accession Number:** 12436708
8. Anwesha Banerjee, Sumantra Chakraborty, Pratyusha Das et al,, "Single Channel Electrooculogram (Eog) Based Interface For Mobility Aid", IEEE Proceedings Of 4th International Conference On Intelligent Human Computer Interaction, Kharagpur, December, 2012, pp. 1-6.
9. Sushmita Mitra, Tinku Acharya, "Gesture Recognition: A Survey", 2007, IEEE Transactions on Systems, Man, and Cybernetics—Part C (Applications and Reviews), Vol. 37, No. 3, pp. 311-324
10. Hui Jiang, "Confidence Measures for Speech Recognition: A Survey", 2005, Speech Communication, Vol. 45, No. 4, Elsevier, pp. 455-470
11. Amer Al-Rahayfeh, Miad Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey", 2013, IEEE Journal of Translational Engineering in Health and Medicine Transaction, Vol. 1, Article Sequence Number: 2100212
12.M. C. Hsieh, C. H. Luo, C. W. Mao, "Unstable Morse code recognition with adaptive variable-ratio threshold prediction for physically disabled persons", 2000, IEEE Transactions on Rehabilitation Engineering : A Publication of the IEEE Engineering in Medicine and Biology Society, Vol. 8, No. 3, pp. 405-13
13. Paul Viola, Michael J. Jones, "Robust Real-Time Face Detection", 2004, International Journal of Computer Vision, Vol. 57, No. 2, pp. 137–154
14. T. R. Chandrashekhar, Dr. K B ShivaKumar, D R Shashikumar et al., "A Fuzzy Logic based Human Behavioural Analysis model using Facial Feature Extraction with Perceptual Computing", 2014, Vol. 3, No. 2, IJETTCS, pp. 36-41
15. Piyush Kumar, Jyoti Verma, Shitala Prasad, "Hand Data Glove: A Wearable Real-Time Device for Human-Computer Interaction", 2012, International Journal of Advanced Science and Technology, Vol. 43, pp. 15-26
16.Vitor F. Pamplona, Leandro A. F., Fernandes Jo˜ao, L. Prauchner et al., "The Image-Based Data Glove", 2008, Symposium on Virtual and Augmented Reality, pp. 204–211
17.Ankit Chaudhary, J. L. Raheja, Karen Das et al., "Intelligent Approaches to interact with Machines using HGR in Natural way: A Survey", 2011, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.1, pp. 122-133
18. Heung-Il Suk, Bong-Kee Sin,Seong-Whan Lee, "Recognizing Hand Gestures using Dynamic Bayesian Network", IEEE International Conference on Automatic Face & Gesture Recognition, Amsterdam, September, 2008, pp. 1-6
19. Yuelong Chuang, Ling Chen, Gangqiang Zhao et al., "Hand Posture Recognition and Tracking Based on Bag-of-Words for Human Robot Interaction", IEEE International Conference on Robotics and Automation, Shanghai, May, 2011, pp. 538-543
20. Nguyen Dang Binh, Toshiaki Ejima, "HGR Using Fuzzy Neural Network", ICGST Conference on Graphics, Vision and Image Processing, GVIP-05, Cairo, December, 2005, SN: P1150535210
21. Noor Adnan Ibraheem, Rafiqul Zaman Khan, "Survey on Various Gesture Recognition Technologies and Techniques", 2012, International Journal of Computer Applications, Vol. 50, No.7, pp. 38-44
22. Louis-Philippe Morency, Trevor Darrell, "Head Gesture Recognition in Intelligent Interfaces: The Role of Context in Improving Recognition", Proceedings of the 11th international conference on Intelligent user interfaces, Sydney, January, 2006, pp. 32-38
23. Louis-Philippe Morency, Candace Sidner, Christopher Lee et al., "Head gestures for perceptual interfaces: The role of context in improving recognition", 2007, Artificial Intelligence, Vol. 171, Elsevier, pp. 568–585
24. Mehmet Emre Sargin, Yucel Yemez, Engin Erzin et al., "Analysis of Head Gesture and Prosody Patterns for Prosody-Driven Head-Gesture Animation", 2008, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 30, No. 8, pp. 1330-1345
25. P. Jia, H. H. Hu, T. Lu et al., "Head gesture recognition for hands-free control of an intelligent wheelchair", 2007, Industrial Robot: An International Journal, Vol. 34, No. 1, pp. 60–68
26. Parimita Saikia, Karen Das, "Head Gesture Recognition using Optical Flow based Classification with Reinforcement of GMM based Background Subtraction", 2013, International Journal of Computer Applications, Vol. 65, No. 25, pp. 5-11
27. Nickel, K. and Stiefelhagen, R., "Pointing gesture recognition based on 3D-tracking of face, hands and head orientation". 2003, International Conference on Multimodal Interfaces, pp. 140-146
28. Nicholas Michael, Dimitris Metaxas, Carol Neidle, "Spatial and Temporal Pyramids for Grammatical Expression Recognition of American Sign Language", 2009, ASSETS-09, ACM, pp. 75-82
29. H. Wu, T. Shioyama, H. Kobayashi, "Spotting recognition of head gestures from color image series", 1998, Proceedings. Fourteenth International Conference on Pattern Recognition, IEEE, pp. 83–85
30. J.W. Davis, S. Vaks, "A perceptual user interface for recognizing head gesture acknowledgements", 2001, Proceedings of the 2001 workshop on Perceptive user interfaces, pp. 1-7
31. "Morse Code", http://en.wikipedia.org/wiki/Morse_code
32, Alexander Alekseychuk, "Hierarchical Recursive Running Median", 2012, IEEE International Conference on Image Processing, pp. 109-112
33. Sellmann, M., Kadioglu, S., "Dichotomic search protocols for constrained optimization", 2008, Fourteenth International Conference on Principles and Practice of Constraint Programming, Springer, pp. 251–265
34. Luis Ricardo Sapaico, Makoto Sato, "Analysis of Vision-based Text Entry using Morse Code generated by Tongue Gestures" , 2011, International Conference on Human System Interactions, HSI, pp. 158-164
35. J. Tua, H. Taob, T. Huang. "Face as mouse through visual face tracking", 2007, CVIU, Vol. 108, No. 1–2, pp. 35–40

AUTHORS PROFILE

Rupam Das Rupam Das received his BE degree in Electronics and Communication from VTU, Belgaum in 2002 and M. Tech degree in Computer Science and Engineering from VTU, Belgaum in 2012. He is founder and CEO of Integrated Ideas, a R&D firm in Gulbarga, Karnataka. He is also heading the R&D department of the company. He has created over 30 applications in RealSense technology and is awarded Pioneer and Trailblazer in IntelPerceptual and Intel RealSensetechnologies respectively. He is currently pursuing his PhD in Computer Science from VTU, Belgaum. His research area includes Image Processing, Pattern Recognition, Internet of Things, RealSense Technology, Augmented reality and Gesture driven techniques.

Dr. K. B. Shivakumar Dr. ShivaKumar K B received the BE degree in Electronics& Communication Engineering, ME degree in Electronics, MBA from Bangalore University, Bangalore and M Phil from Dravidian University, Kuppam. He obtained his Ph.D. in Information and Communication Technology from Fakir Mohan University, Balasore, Orissa. He has got 50 publications in International journals and conferences. He is currently working as Professor, Dept. of Telecommunication Engineering, Sri Siddhartha Institute of Technology, Tumkur. His research interests include Signal processing, Multi rate systems and filter banks and Steganography.