Facial Emotion Recognition and Eye-tracking based Expressive Communication Framework: Review and Recommendations

Pradeep Kumar Kaushik
Assistant Professor
Department of CSE
JB Institue of technology, Dehradun

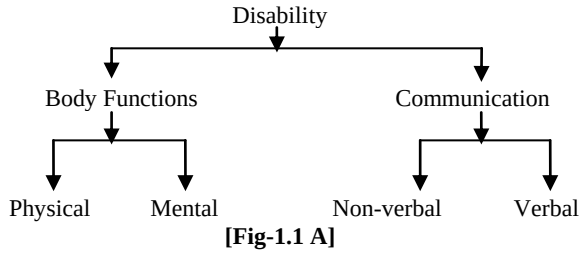
Shivam Pandey
Assistant Professor
Department of CSE
JB Institue of technology, Dehradun

Sushil Singh Rauthan
Assistant Professor
Department of CSE
JB Institue of technology, Dehradun

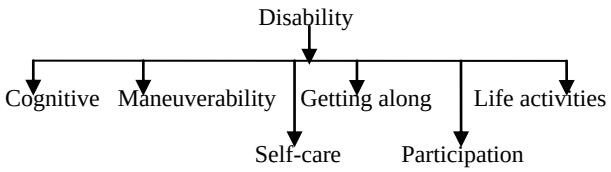
ABSTRACT
This paper presents a review of the application of facial emotion recognition and eye tracking technology for mentally fit intensive speech, disabled persons. It also presents a review of the augmentative and alternative communication (AAC), already developed emotion recognition systems or frameworks and some current speech disability support devices selected from the internet. Employing these techniques, the design of simplex-based automatic real-time expressive communication system to facilitate effective interpretation of thoughts of speech-disabled is advocated in this paper. While several devices are available, some with eye-tracking capability, but there is still a need for more versatile, cost-effective device or framework and the challenge is to achieve speech like a human from AAC based device.

Keywords
Facial Emotion Recognition, Augmentative and alternative communication, Eye-tracking technology, Speech disability

1. INTRODUCTION
Disability is a combined term, including impairments, activity restrictions, and participation limitations. Impairment is a problem in body action or structure; an activity restriction is trouble confronted by a person in performing a task; while a participation limitation is a problem faced by a person in involvement in life circumstances [1]. Disability can be categorized into different categories.



[Fig-1.1 A]



[**Fig-1.1 A, B** -Types of disability][2]

Physical disability is a limitation on a person's physical act or motion. While speech disability is a communication disorder such as voice impairment, language impairment, impaired articulation, and stuttering. Both adversely affect an individual's educational, social and professional performance. More than 10^3 million people in the world have some type of disability. Disabled persons are more likely to be jobless. The employment rate for disabled persons (44%) is slightly over half that for general persons (75%) in the group of Economic Cooperation and Development countries [3]. There is more research scope in speech disability than physical disability. Surbhi Rathi et al. [4] highlight the problem of communication faced by deaf and dumb, the communication system is selected for exploration. It has been seen since the '90s that research regarding speech and other disability is going on; augmentative and alternative communication is the best solution for speech disability. Pat Mirenda et al. [5] reviewed the issues referred to the communication of augmentation for persons with intensive mental disabilities. Barry M. Prizant et al. [6] presented a huge expansion of speech, language and communication disorders in kids related to psychiatric and brain health for sentimental and behavioral issues. The gained knowledge can be used in the development of support devices for communication. Additionally, emotion recognition has been an essential input for effective human-computer interaction. Emotion can be defined as an attentive experience characterized by fast brain activity and a definite intensity of pleasure or displeasure [7]. Emotion can be recognized through facial expressions, voice, and physiological signals. But facial expressions are the rich source of emotions.

2. FACIAL EMOTION RECOGNITION
Facial emotion recognition is the procedure of identifying human emotion, from facial expressions [7]. Facial emotion recognition is the emerging application region of human-computer interaction and computer vision. Modern years have witnessed an increasing attraction in systems able to interpreting facial emotions and responding to them. These systems can be used for designing human-computer interfaces, online contents, videos, advertisements, and games. Salwa Said et al. [8] proposed the latest technique for facial expression identification based on wavelet network classifier; it permits the trace of 6 key emotions: sadness, anger, fear, hatred, joy, and surprise. This paper discussed the possible application in the detection of driver emotional state, smart television by facial controls. Emotion recognition has also application in the intelligent communication systems, intelligent wheelchair, navigation path system, etc. for physically, visually and other disability support frameworks. Emotion recognition is important for a speech generating

device so that the output speech contains the user's emotions. Priya Saha et al. [9] presented a paper to trace the region of interest (ROI) of expressive facial images. Face recognition and facial emotion recognition are active research areas, an ROI (triangular area covers from left eye to right eye and lips on the face) plays a crucial role in them because major facial expressions always expressed by the muscles of this area. Normally emotions are specified into seven: anger, disgust, fear, joy, sad, startled and neutral. Jyoti Kumari et al. [10] presented a survey paper, the paper analyses the different facial emotion recognition techniques normally based on geometry and appearance. Ma Xiaoxi et al. [11] presented a paper and compared the experiment outcomes of various prediction frameworks and concluded that the excellent performance of occurrence detection of AUs is got by sentimental facial classification framework with SVM. Hence, emotion recognition has huge importance in a speech generating device or in the area of augmentative and alternative communication.

The relation between facial emotion and speech production

Speech is the biometric property of a human, emotion can be recognized from speech. Mohan Ghai et al. [12] presented a paper and main aim is to identify sentiments in speech and categorize them in 7 sentiment classes - neutral, sadness, boredom, hatred, anxiety, happiness, and anger. The presented viewpoint is based upon the Mel Frequency Cepstral Coefficients (MFCC) and energy of the speech signals as feature inputs and uses Berlin database. The paper concluded that the insertion of energy as a feature including other thirteen MFCC features lead to the good evaluation of the emotion present in the speech. But we analyze that 13 MFCC features extracted from human voice samples can be inserted in synthetic speech after recognition of facial emotion of the speech-disabled user. Hence the reverse process can generate synthetic expressive speech because the absence of expressivity is the major problem in the present speech generating devices or disability support frameworks.

Problem with facial emotion recognition

Roshan Jameel et al. [13] presented a survey paper that analyzed the few issues in the area of facial expression recognition and has discussed a few techniques used to the aim of facial expression recognition. The problem being faced currently in facial expression recognition is the absence of the databases having the natural facial expression and preparing a similar database is the big challenge. For this, the approach of using secret cameras will not act effectively. Search out the labeling data for investigation, this practice is a challenge too; creating a fully genuine database is very hard but the semi-genuine database is relatively easy. So, this paper focuses on creating a database and sets out some of the important problems during the creation of a database.

Classification performance of Facial emotion recognition

Various classifiers (Table 2.1) can be used to classify the captured emotion.

Table 2.1:-classification performance of popular classifiers, employed for the task of facial emotion recognition

Classifier	Average Classification Accuracy	Reference
Patch-based Gabor	92.93-94.48%	[14]
KNN	94.8-97.63%	[15]
MLP	90%	[16]
3-NN	96%	[16]
ICA and HMM	81.67%	[17]
LDP and HMM	82.08%	[17]
LD^2 BP and HMM	84.17%	[17]

3. EYE TRACKING TECHNOLOGY

Eye tracking technology measures the pupil movement of the eye relative to the head. Eye tracking is the easiest human-computer interface. It has application in the easy pilot interface, disability support framework, assistive technology, E-learning, Car assistive systems, Iris recognition and Field of view estimation. A disabled user can easily operate it by eye motion. Eye tracking technology has been used in commercial products: Eagle eyes, Blink, Smi, Tobii.

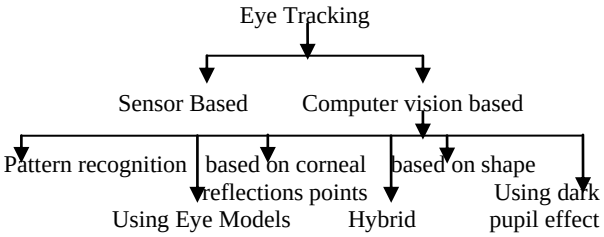


Fig: 3.1-Types of Eye-tracking techniques

Pramodini A. Pune et al. [18] presented a study of eye tracking technology its type and application. Amer Al-Rahayfeh et al. [19] presented a review paper on eye tracking and head movement detection; this is an active exploration field in human-computer interaction.

4. AUGMENTATIVE AND ALTERNATIVE COMMUNICATION (AAC)

AAC is the communication manner used for replacement of speech or writing for those with impairments in the composition of spoken or written language. AAC is the best solution for speech disability. So Many devices are available to facilitate the AAC, but many of them are not expressive (output speech is synthetic). Shaun K. Kane et al. [20] introduced a paper and conducted interviews with 7 individuals with Amyotrophic Lateral Sclerosis and their fellows, concentrating on how AAC use affected their communication and personal expression. Alexander Fiannaca et al.[21] discussed a method voice setting for improved expressivity in AAC. Basically, conveying emotions in speech is expressivity, in this paper expressive keyboard and voice setting editor is used for fast insertion of expressivity in speech. Expressive Keyboard allows the insertion of emoji and punctuation into text for fast expressivity in speech.

5. RELATED WORKS

The Augmentative and alternative communication is the active research field, some of the already developed devices or frameworks are selected as related works.

Christos D. Katsis et al. [22] presented a unified framework based on physiological signals; the paper described a methodology, named Intrepid system which provides a solution for the observation of patients with worry disorders during therapy. The limitation of the Intrepid system is that it can capture only five emotions from physiological signals. The system can't differentiate between internal and external emotion. Therefore an integrated approach of physiological and facial emotion recognition is needed in an emotion recognition framework.

M.T. Quazi et al. [23] demonstrated a smart sensing system detects human emotions based on bodily parameters, the amplified and filtered signals from the sensors are then processed by a microcontroller and transmitted wirelessly using Zigbee technology, received signal then displayed on a computer, the partially developed system has demonstrated nice outcomes in observing the bodily parameters. The only drawback of the discussed system is that it can visualize the effect of only four emotions happy, sad, angry and neutral.

Eun-Hye Jang et al. [24] presented a paper; it illustrated the differences between boredom, pain, and surprise by using bodily signals. The research in the paper could dedicate to enhance the sensibility of human feeling states and the basic bodily system. By revealing human's emotional reply features from different groups, the findings could dedicate to anthropology. But the limitation of this paper, bodily signals can only recognize inner feelings.

Kathawut Rojanasaroch et al. [25] presented an easy, cheap communication aid device for a consult with a doctor by deaf-dumb, the system includes the steps: selection of images, transform the image to the pattern, comparison and pair with words, form sentences, send a corresponding sentence to display. The device act as an interpreter between deaf-dumb and the doctor. The device dedicated to communication with a doctor by deaf –dumb during illness. The communication mode is simplex because device operates in one way and the sender is deaf –dumb and receiver is a doctor.

TobiiDynavox [26] introduced the new TobiiDynavox I-Series+ speech generating device along with Communicator 5 software. Trained eye tracking users can easily point over letters to form words they want to write. This has the ability to boost communication speed prominently for users because slow speed is the drawback of speech generating devices. I-Series+ devices with touch screen and alternative eye tracking capabilities provide communication such as SMS, social networking, electronic mail, phone calls, etc. The letter or word displays on the screen of TobiiDynavox I-Series+ with associated pictures of that word so that the user can clearly understand them (e.g. Hand word with a picture of a hand). However, the output speech is not expressive from this device and expressive speech is that which can convey emotion.

Mingmin Zhao et al. [27] presented an emotion recognition system using the wireless signal as EQ-Radio. It acts by emitting a radio frequency signal and obtaining its reflections off an individual's body. But such a system can be implemented over the human body for a small time, for example- x-ray. The human body can't bear radio signal continuously because radio waves harm the human body. So where we need continuous emotion recognition this system is not appropriate.

Elena Simion [28] presented a paper on augmentative and alternative communication (AAC) and discussed that whenever a kid's speech does not develop generally or he cannot communicate efficiently through speech due to pre-existing circumstances, it is suggested to receive AAC support as soon as possible.

Janice Light and David McNaughton [29] presented an article on augmentative and alternative communication (AAC) and discussed that since these early days of AAC the area has observed dramatic changes:

i. Changes in the population that uses AAC.

ii. Changes in the scope of communication needs that must be considered.

iii. Changes in the AAC systems that are available.

iv. Finally, changes in expectations for participation by persons who use AAC.

Jeff Sigafoos [30] presented an article on augmentative and alternative communication (AAC) and discussed that AAC refers to a field of exploration and clinical specialization that spans several disciplines, containing assistive technology, psychology, rehabilitation, special education, and speech-language pathology. AAC exploration has included a broad range of subjects; from developing recent and more effective methods of executing AAC, to examine the effects of AAC use, to evaluating innovative assistive communication technologies. Clinical practice in AAC is largely related to enabling people with limited speech and language to communicate using one or more options for speech. AAC frameworks are also often used to enhance unintelligible speech. The research revealed that people with Autism Spectrum Disorder or intellectual disability might be able to self-decide which of many AAC alternatives should be incorporated into their AAC interference.

6. FINDINGS

Augmentative and alternative communication has identified as an active research field and it is beneficial for disabled people so that they can live a better life. The problem faced by the disabled in terms of assessment of communication needs, communication opportunity: cognitive/linguistic and capability assessment: sensory/perceptual has identified.

But, in the present scenario, the major problems faced by speech disabled person are:-

1. Price of the device: - Many supporting devices are available for speech disability but the cost is high that can't be bearable for a disabled person.

2. Unavailability of Natural Speech:-The speech generated from the devices is not natural, a disabled person feel uncomfortable with synthetic speech.

3. Limitation of words and phrases:-Many devices have the limitation of words and phrases; a disabled person can express only a few thoughts with limited words e.g. Go Talk Pocket, Logan Prox Talker, Mega Bee Eye Pointing Communication Tablet.

Communication is very important for human life and this knowledge can be utilized in the development of perfect

devices for speech-disabled persons. Here are some products available (Table 6.1) to support speech disability and used as an alternative communication tool by speech disabled person

Table: 6.1:-Different speech generating devices

Year	Company/ Author	Device/ Framework	Price	Key Features	Drawback
2012	Attainment Company	Go Talk Pocket[31]	$ 189.00	The small device can store 30 messages in five levels, on six buttons.	Limitation of words and sentences, output voice is not expressive and it is prerecorded.
2012	Logan Tech	Logan Prox Talker[32]	$ 3149.00	1. It enables independent verbal picture communication using radio frequency identification (RFID) technology. 2. Put your sound tag card on any one of the five buttons and push to get output voice. A male or female voice can select as the output.	This product has a limitation of words and this device makes the voice of prerecorded RFID tags but cannot produce sentences.
2013	Logan Tech	Mega Bee Eye Pointing Communication Tablet[33]	$ 1595.00	Dual screens accommodate 20 characters in two lines.	The device has a limitation of words and sentences; the output speech is not expressive.
2014	Beamz Interactive Inc.	Beamz Music System[34]	$ 600.00	1. W-shaped tabletop optical control device with several laser beams. 2. Beamz system connects to a computer via USB. 3. Six lasers trigger and 2 buttons controlled triggers activate up to 64 freely controlled orders of musical notes or events.	This is a musical system used for entertainment of disabled persons, not a speech-generating device.
2015	TobiiDynavox	TobiiDynavox I-Series+[35]	£ 8,967.00	By using eye-tracking technology this can support communication such as email, text messaging, social networking, phone calls.	The output speech is not expressive from this device and expressive speech is that which can convey emotion.

6.2 Proposed Framework

The problem of speech disability has identified, augmentative and alternative communication has recognized as a good solution. Some of the current AAC support devices or frameworks have been reviewed; the problem of high cost, expressivity, and limitation of words and phrases has been identified. So, the objective of the proposed framework (named as aawaj) is to overcome the problem of expressivity, high-cost, and limitation of words and phrases. Because present speech generating devices is less expressive and the cost is very high that is not bearable by a disabled person. The proposed framework is divided into three modules: facial emotion recognition, symbol selection and text to speech conversion. The limitation of the proposed system is that it can't work effectively for the users those can't express facial expressions.

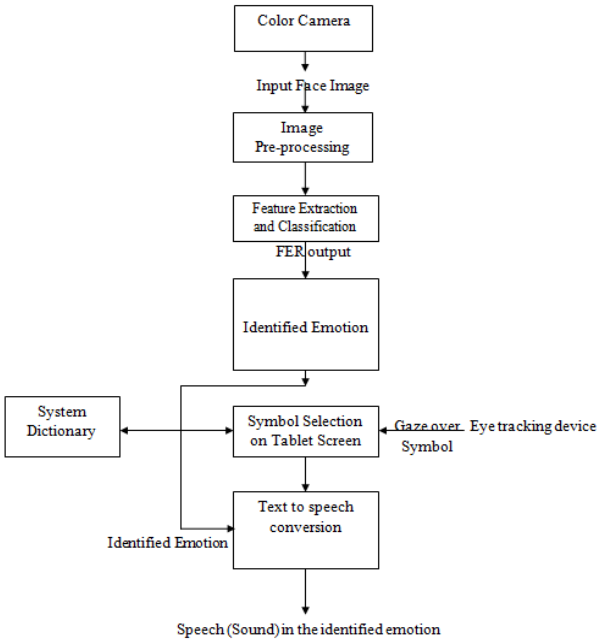
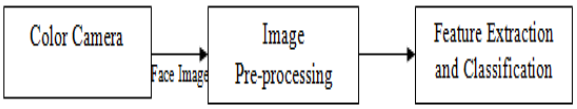


Fig6.3:-Architecture of proposed framework

Module A(Facial emotion recognition)- Module A recognizes eight basic emotions- Contempt, Fear, Happy, Sad, Surprise, Disgust, neutral and anger. G.J. Edwards et al. [36] exhibited a rapid, durable scheme of interpretation of face images using an Active Appearance Model (AAM); it is a repetitive process for fitting an AAM to face images. An AAM includes a computation model of shape and grey level appearance which can generalize to approximately any face.



2022922490

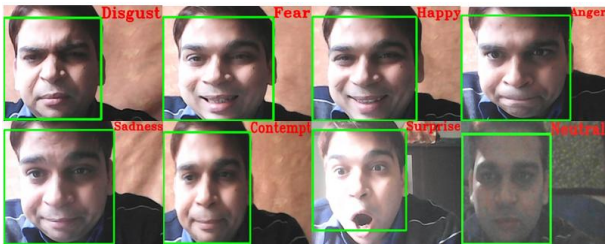


Fig6.5:-Module A- recognition of eight basic emotions

Steps of Module A:-

1. Find the facial area using active appearance modeling (AAM)
2. The image conversion into a Gabor magnitude representation, using a bank of Gabor filters at 8 orientations and 5 spatial frequencies (4:16 pixels per cycle at 1/2 octave steps)
3. Mark the Facial points: Feature selection algorithm
4. Feature extraction: Action unit detection algorithm
5. Classification: Support vector machine

Module B(Symbol Selection)- Disabled user selects the target symbol using the onscreen keyboard, and form the sentences, which he want to speak.

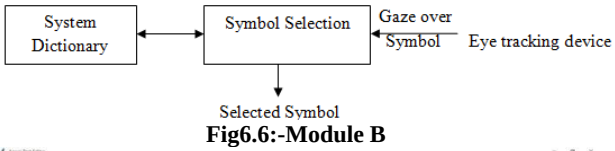


Fig6.6:-Module B

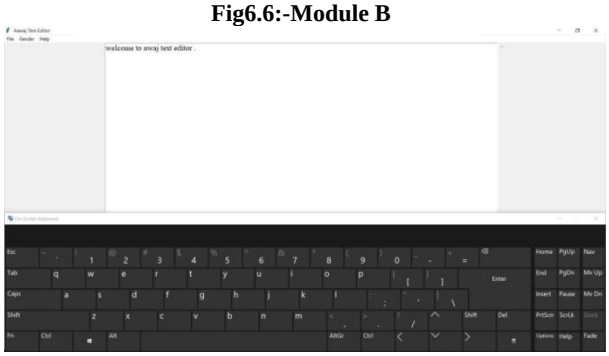


Fig6.7:-Module B- text editor

Steps of Module B:-
1. Display alphabets A to Z digits 0-9 and another symbol over the onscreen keyboard.
2. Select letter, digit, and another symbol using an eye tracking device.
3. Form the sentences.
4. Send the sentences to module C for text to speech conversion.

Module C(Text to speech Conversion)-Module C convert text to speech in the form of identified emotion. The output voice can be selected as a male voice or female voice. The voiced speech of a general adult male has a basic frequency from 85 to 180 Hz and that of a general adult female from 165 to 255 Hz.

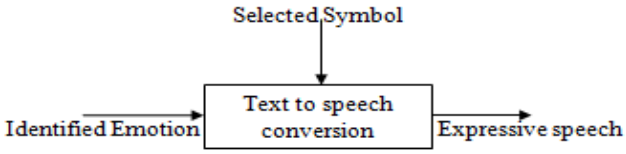


Fig6.8:-Module C

Steps of Module C:-
1. Standard Text Editor
2. Text Normalizer
3. Grammar Dictionary
4. Letter-to-phoneme rules
5. Prosody rules
6. Phonetic rules
7. Emotion Insertion
8. Voice generation
9. Interrupt driver
10. Output hardware

The preliminary test-
As an experiment to test and verify the validity of the partially developed proposed communication aid framework, a preliminary test was done without the eye-tracking device. The test includes naturalness test, intelligibility test, accuracy test, and comprehensibility test. The outcomes of the test are given below.

a. Naturalness Test: - Mean opinion score and Degraded Mean Opinion Score are calculated to test and verify the naturalness of the output speech.

i. Mean Opinion Score (MOS):-

MOS is the easiest manner to evaluate the quality of output speech. MOS gives a numerical sign of the quality of the output speech. In this manner, the main concern of evaluator should be on the naturalness of synthetic speech. Naturalness means that the voice is indistinguishable from human speech and it is identical to human voice [37]. The naturalness aspect of synthesized speech will have to understand by the evaluator, after hearing to the output voice.MOS is the average of the scores given by all the evaluators.

MOS	Quality of speech
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Calculation of MOS: Test Data for MOS should contain a short paragraph that included all the possible variations primely: symbols, numerals, abbreviations, English words written in Latin and Indian scripts, and various types of sentences such as negative, declarative and exclamatory. Sentences selected from various fields such as stories, news, sports, etc.

$$MOS=\frac{\sum_{j=1}^{M}\left(\frac{\sum_{i=1}^{N} Sij}{N}\right)}{M}$$

Si=Score of the ith evaluator
N=Total number of Evaluators
M=Total number of Sentences
J= Index of sentence

Evaluator	Score Given																			
	S1		S2		S3		S4		S5		S6		S7		S8		S9		S10	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Evaluator 1	3	4	4	3	2	3	4	4	4	4	3	3	4	3	5	4	3	2	3	3
Evaluator 2	4	4	2	3	2	2	4	3	5	4	3	3	4	4	4	4	4	4	5	5
Evaluator 3	5	3	3	5	4	4	4	5	5	5	4	4	5	5	3	3	2	2	5	5
Evaluator 4	4	3	4	4	3	3	3	4	3	3	3	3	4	4	3	3	3	4	4	4
Evaluator 5	5	3	4	2	4	4	4	5	5	5	3	4	5	5	4	4	5	5	5	4
Evaluator 6	4	3	4	4	3	4	3	4	3	5	2	3	5	5	4	4	4	4	5	4
Evaluator 7	4	4	4	4	4	4	3	4	4	4	3	3	5	5	3	3	4	4	3	4
Evaluator 8	4	2	4	3	4	5	3	4	5	4	3	3	5	4	3	3	4	4	5	4
Evaluator 9	3	3	3	2	3	3	4	4	3	4	3	4	4	5	3	3	2	3	3	4
Evaluator 10	5	3	4	3	3	4	5	4	4	4	4	3	5	5	3	4	3	4	5	4
Total	**41**	**32**	**36**	**33**	**32**	**36**	**37**	**41**	**41**	**42**	**31**	**33**	**46**	**45**	**35**	**35**	**34**	**36**	**43**	**41**

[Table 6.9:-MOS for naturalness]

Maximum Score=5
For Male voice, MOS Score = 3.76
MOS Score percentage = 75.2
For Female Voice, MOS Score = 3.74
MOS Score percentage = 74.8

ii. Degraded Mean Opinion Score (DMOS):- DMOS score is used to evaluate the naturalness aspect of the output voice by comparing the natural and synthetic speech. In a DMOS manner, evaluators need to listen to synthetic as well as a natural speech in random order, without having earlier information about the type of speech i.e. synthetic or natural. The objective of this manner is to evaluate the speech in terms of naturalness. Mean of scores given to natural and

synthesized speech separately by each evaluator will be computed [38].

$$\text{Standardized score of synthesized to natural} = \frac{\text{synthetic speech score}}{\text{natural speech score}} * 5$$

Score of Naturalness

5	Framework sound like Human
4	Robotic sound but reading properly
3	Reading sentences with less broken words in a robotic style
2	Approximately every word broken
1	Very intolerable

Table 6.10:-DMOS

Evaluator	Score Given			
	Synthetic		Natural	
	M	F	M	F
Evaluator 1	3.5	3.3	4	5
Evaluator 2	3.7	3.6	5	4
Evaluator 3	4	4.1	5	5
Evaluator 4	3.4	3.5	4	5
Evaluator 5	4.4	4.1	5	4
Evaluator 6	3.7	4	4	5
Evaluator 7	3.7	3.9	5	4
Evaluator 8	4	3.6	4	5
Evaluator 9	3.1	3.5	5	5
Evaluator 10	4.1	3.8	5	5
Total	**37.6**	**37.4**	**46**	**47**

Normalized score of synthesized to natural,

Male Voice =4.086
Female Voice=3.97

b. Intelligibility Test

Intelligibility test is one of the crucial factors affecting speech quality. We can compute intelligibility of speech by MOS and WER.

i. Mean Opinion Score (MOS):-

The Intelligibility refers to the precision with which every word is pronounced so that the general listener can understand the meaning of the spoken word or phrase. In this manner, the attention of evaluator should be on the intelligibility of synthetic speech.

$$\text{MOS} = \frac{\sum_{j=1}^{M} \left(\frac{\sum_{i=1}^{N} Sij}{N} \right)}{M}$$

Si=Score of the ith evaluator
N= Total number of Evaluators
M=Total number of Sentences
J=Index of Sentence

Evaluator	Score Given																			
	S1		S2		S3		S4		S5		S6		S7		S8		S9		S10	
	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F
Evaluator 1	4	4	3	3	3	3	4	3	4	3	3	3	4	3	3	3	4	4	4	4
Evaluator 2	4	4	3	3	3	3	3	3	4	4	3	3	4	4	4	4	4	4	4	4
Evaluator 3	4	4	5	5	4	4	4	4	5	5	4	4	5	5	2	2	4	4	5	5
Evaluator 4	4	4	3	3	3	3	4	4	4	4	3	4	4	4	3	2	3	4	4	4
Evaluator 5	4	4	5	5	4	5	4	4	5	5	3	4	5	5	4	4	5	5	5	4
Evaluator 6	5	5	4	4	3	3	3	4	5	4	3	3	5	5	4	4	5	5	5	5
Evaluator 7	4	4	4	4	3	3	3	3	4	4	3	4	5	5	4	4	4	4	5	5
Evaluator 8	4	4	4	3	4	4	5	4	4	4	4	4	5	5	3	3	5	5	5	5
Evaluator 9	3	3	3	3	3	4	4	4	4	4	3	3	4	5	3	2	4	4	4	4
Evaluator 10	4	3	4	4	5	4	4	4	5	5	4	3	5	4	4	3	5	5	5	4
Total	**40**	**39**	**38**	**37**	**35**	**36**	**38**	**37**	**44**	**42**	**33**	**35**	**46**	**45**	**34**	**31**	**43**	**44**	**46**	**44**

[Table 6.11:-MOS for intelligibility]

Maximum Score=5
For Male voice, MOS Score = 3.97,
MOS Score percentage = 79.4
For Female Voice, MOS Score = 3.9,
MOS Score percentage = 78

ii. Word Error Rate (WER):

For this test, sentences which are semantically not understandable (SNU) but are formed in such a way that they are grammatical should be used. Word length of SNU sentences should not exceed by 7; because, possibilities are that evaluators might forget them. SNU is used to judge the intelligibility because it becomes hard for listeners to predict unheard information. After hearing to a sentence, the evaluator will have to write whatever they heard, even if they don't understand the meaning. While computing WER, writing mistakes should be avoided [39].

For every evaluator,

$$\text{WER} = \frac{(A+B+C)}{D}$$

Where,

A is the no. of changes
B is the no. of deletions
C is the no. of insertions
D is the total no. of words in the Sentence

Evaluator	WER										
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Total
Evaluator 1	0.14	0	0.14	0.125	0.14	0	0	0.33	0	0.5	1.375
Evaluator 2	0.28	0.33	0.42	0.25	0.42	0	0	0.66	0	1	3.36
Evaluator 3	0	0.16	0.42	0	0	0	0	0.33	0	0	0.91
Evaluator 4	0.14	0.5	0.14	0	0	0	0.66	0.66	0	0.5	2.6
Evaluator 5	0.28	0.16	0.14	0	0	0	0	0.33	0	0	0.91
Evaluator 6	0.28	0.5	0.14	0	0.14	0	0	0.33	0	0.5	1.89
Evaluator 7	0.28	0.33	0.14	0	0	0	0	0.33	0	0.5	1.58
Evaluator 8	0.28	0.66	0.28	0	0	0	0	0.33	1	0.5	3.05
Evaluator 9	0.28	0	0.14	0	0.14	0	0	0.66	0	0.5	1.72
Evaluator 10	0.28	0.16	0.28	0	0	0	0	0	0	1	1.72

[Table 6.12:-WER]

Average WER =1.9115

c.Accuracy Test

For accuracy computation suitable selection of test data is important. All such data whose required output is well defined can be selected for the accuracy test. Classification of test data is as follows:

Table 6.13:-Test Data

Test Data	Performance	Score
a) Number Handling i. Digits (Phone no. +91-9456249716) ii. Fractions (1/2) iii. Numbers (9999) iv. Numerals (1st, XI)	Excellent	5
b) Date formats (MM/DD/YY)	Excellent	5
c) Acronyms (prof., Dr., Rs. ,etc.)	Bad	1
d) Abbreviations (DRDO, PWD,SBI)	Excellent	5
e) Names(Pradeep Kumar Kaushik)	Good	4
f) Addresses(E-135,Janakpuri ,Sahibabad ,Ghaziabad)	Good	4
g) Homographs(redress-regress)	Excellent	5
h) Punctuations and Brackets (, ; " " – [],(),{})	Bad	1
i) Special Symbols ($, @, %)	Excellent	5

Score	Quality of Speech
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

$$\text{Accuracy}=\frac{Sum\ of\ Score}{Total\ of\ Maximum\ \ Score}\ *100$$

$$\text{Accuracy}=\frac{35}{45}*100=77.77\%$$

d. Comprehensibility Test(CT)

Intelligibility test focuses only on the identification of every word without a focus on the meaning of the sentence. CT evaluates the complete sentence. CT carried out when the framework achieves the intelligibility up to admissible level else it is meaningless to carry out CT for the unintelligent framework; as intelligibility has a strong effect on comprehension [40].

In the CT, the evaluator will be asked to hear a paragraph and based on that some questions will be asked. Questions should be framed in such a way that whether the evaluator has understood the paragraph listen or not can be observed.
A two-point scale (0, 1) is suggested for the CT.

Scale	Evaluation
0	Incorrect answer
1	Correct answer

Table 6.14:- Comprehensibility test

Evaluator	Score Given
Evaluator 1	1
Evaluator 2	1
Evaluator 3	1
Evaluator 4	1
Evaluator 5	0
Evaluator 6	0
Evaluator 7	1
Evaluator 8	0
Evaluator 9	1
Evaluator 10	1

Comprehensibility test Result:-70%

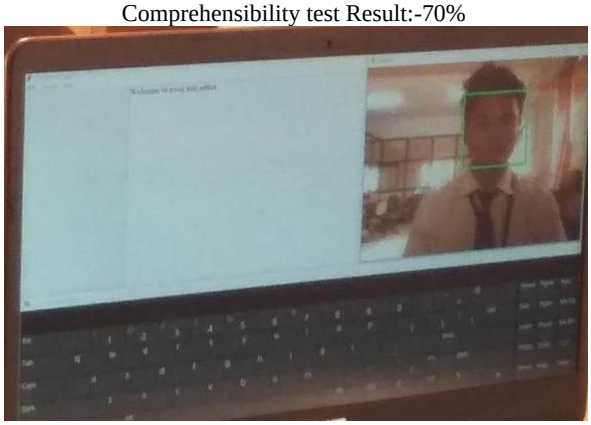


Fig 6.15:- The preliminary test

7. CONCLUSION

In the presented work, the type of disability has discussed and speech disability selected as the centroid of research; augmentative and alternative communication has found as a good solution to speech disability. An important application of facial emotion recognition and eye tracking technology has been identified in speech disability support frameworks. Several devices are available and some with the eye tracking capability but the cost is very high and these are less expressive or inexpressive. So, lack of expressive device is found but our aim is to find a real-time expressive communication framework dedicated to the speech disabled person and different already developed devices have reviewed and Simplex mode is best suited for communication, because the system will be used by speech disabled person to express

their feelings, and that will be one-directional communication. Therefore, there is a need for an expressive communication system for a speech-disabled person in simplex mode. That's why the design of the proposed framework is prepared in this paper. The preliminary test is performed in the lab to ensure the performance of the proposed framework; the above results in different metrics have shown the performance is 70 percent or more.

8. FUTURE WORK

The cost of the present speech generating devices is very high and output speech is not expressive, both identified as a major problem, so it is recommended that low cost and expressive speech generating device or framework should be developed in future. In the future, this is going to be a significant development until a full proof and cost-effective system for effectively interpreting the thoughts of speech disabled can be developed. In the future, the performance of the proposed framework will be increased and the complete system will be tested with an eye-tracking device.

9. REFERENCES

[1] "World Health Organization-Health topics-Disabilities" [Online] Available: https://www.who.int/topics/disabilities/en/

[2] "International Classification of Functioning, Disability and Health" [Online] Available: https://www.who.int/classifications/icf/en/

[3] "World report on disability" [Online] Available: https://en.wikipedia.org/wiki/World_report_on_disability

[4] Surbhi Rathi, Ujwalla Gawande, "Development of full-duplex communication system for deaf and dumb people", 7th International conference on cloud computing, data science, and engineering-Confluence, IEEE 2017

[5] Pet Mirenda, Teresa lacono, "Communication Options for Persons with Severe and Profound Disabilities: State of the Art and Future Directions", sage journals, Vol.15, issue: 1, pages: 3-21, March 1990, DOI: https://doi.org/10.1177/154079699001500102

[6] Barry M. prizant, Lisa R. Audet, Grace M. Burke, Lauren J. Hummel, Suzanne R. Maher, Geraldine theadore, "Communication Disorders and Emotional/Behavioral Disorders in Children and Adolescents", Journal of speech and hearing disorders, May 1990, Vol.55,179-192, DOI:10.1044/jshd.5502.179

[7] "Emotion Recognition" [Online] Available: https://en.wikipedia.org/wiki/Emotion_recognition

[8] Salwa Said, Olfa Jemai, Mourad Zaied, Chokri Ben Amar, "Wavelet Networks for Facial Emotion Recognition", 15th International Conference on Intelligent Systems Design and Applications (ISDA), IEEE 2015, DOI:10.1109/ISDA.2015.7489242

[9] Priya Saha, Debotosh Bhattacharjee, Barin Kumar De, MitaNasipuri, "An Approach to Detect the Region of Interest of Expressive Face Images", International Conference on Information and Communication Technologies (ICICT 2014), Elsevier Procedia Computer Science 46(2015)17391746, DOI:10.1016/j.procs.2015.02.123

[10] Jyoti Kumari, R.Rajesh, K.M Pooja, "Facial expression recognition: A survey", Second International Symposium on Computer Vision and the Internet (VisionNet'15), Elsevier Procedia Computer Science 58(2015), 486-491

[11] Ma Xiaoxi, Lin Weisi, Huang Dongyan, Dong Minghui, Haizhou Li, "Facial emotion recognition", 2nd International Conference on Signal and Image Processing (ICSIP), IEEE, Aug 2017 DOI: 10.1109/SIPROCESS.2017.8124509

[12] Mohan Ghai, Shamit Lal, Shivam Duggal, Shrey Manik, "Emotion recognition on speech signals using machine learning", International Conference on Big Data Analytics and Computational Intelligence (ICBDAC),2017, **DOI:** 10.1109/ICBDACI.2017.8070805

[13] Roshan Jameel, Abhishek Singhal, Abhay Bansal, "A Comprehensive study on facial expressions recognition techniques", 6th International Conference –Cloud System and Big data Engineering (Confluence), IEEE2016, DOI:10.1109/CONFLUENCE.2016.7508167

[14] Ligang Zhang, Dian Tjondronegoro, "Facial expression recognition using facial movement features ", IEEE Transactions on Affective Computing, Vol.2, Issue No.4,2011, **DOI:** 10.1109/T-AFFC.2011.13

[15] Shoaib Kamal, Dr. Farrukh Sayeed, Mohammed Rafeeq, "Facial Emotion Recognition for Human-Computer Interactions using hybrid feature extraction technique", International Conference on Data Mining and Advanced Computing (SAPIENCE), IEEE, March 2016, **DOI:** 10.1109/SAPIENCE.2016.7684129

[16] Pawel Tarnowski, Marcin Kolodziej, Andrzej Majkowski, Remigiusz J. Rak, "Emotion recognition using facial expressions", International Conference on Computational Science, ICCS 2017,12-14 June 2017, Zurich Switzerland, DOI: https://doi.org/10.1016/j.procs.2017.05.025

[17] Md. Zia Uddin, Weria Khaksar, Jim Torresen," Facial expression recognition using salient features and convolutional neural network", Vol. 5, IEEE Access, Nov 2017, DOI: 10.1109/ACCESS.2017.2777003

[18] Pramodini A. Punde, Mukti E. Jadhav, Ramesh R. Manza, "A study of eye tracking technology and its applications", 1st International Conference on Intelligent Systems and Information Management (ICISIM), IEEE, Oct 2017, DOI: 10.1109/ICISIM.2017.8122153

[19] Amer Al-Rahayfeh, Miad Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey", IEEE Journal of Translational Engineering in Health and Medicine (Volume: 1), DOI:10.1109/JTEHM.2013.2289879

[20] Shaun K. Kane, Meredith Ringel Morris, Ann Paradiso, Jon Campbell, "At times avuncular and cantankerous, with the reflexes of a mongoose: Understanding Self-Expression through Augmentative and Alternative Communication Devices", Proceedings of CSCW 2017, ACM, DOI: http://dx.doi.org/10.1145/2998181.2998284

[21] Alexander Fiannaca, John Campbell, Ann Paradiso, Meredith Ringel Morris, "Voice setting: Voice Authoring Uis for Improved Expressivity in Augmentative Communication", Proceedings of CHI

2018,ACM,DOI:https://doi.org/10.1145/3173574.3173857

[22] Christos D. Katsis, Nikolaos S. Katersidis, Dimitrios I. Fotiadis, "An integrated system based on physiological signals for the assessment of affective states in patients with anxiety disorders", Biomedical Signal Processing and Control 6(2011)261-268, Elsevier, DOI:10.1016/j.bspc.2010.12.001

[23] M.T.Quazi, S.C. Mukhopadhyay, N.K.Suryadevara, Y.M. Huang, "Towards the Smart Sensors Based Human Emotion Recognition", International Instrumentation and Measurement Technology Conference (12 MTC), IEEE 2012, DOI: 10.1109/12MTC.2012.6229646

[24] Eun-Hye Jang, Byoung-Jun Park, Mi-SookPark, SngHyeobKim, Jin Han Sohn, "Analysis of physiological signals for recognition of boredom, pain and surprise emotions", Journal of Physiological Anthropology (2015)34:25

[25] Kathawut Rojanasaroch, Terravisit Laohapensaeng, "Communication Aid Device for Illness Deaf-Mute", 12th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), IEEE2015, DOI:10.1109/ECTICon.2015.7207127

[26] "TobiiDynavox I-Series+" [online] Available: https://www.tobii.com/group/news-media/press-releases/tobii-dynavox-launches-the-new-i-series-and-communicator-5-for-even-more-efficient-communication/

[27] Mingmin Zhao, Fadel Adib, Dina Katabi, "Emotion recognition using wireless signals", MobiCom'16 Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, ACM, Oct 2016

[28] Elena Simion, "Augmentative and alternative communication - support for people with severe speech disorders", Procedia – Social and Behavioral Sciences, Vol. 128, April 2014, Pages 77-81, DOI: https://doi.org/10.1016/j.sbspro.2014.03.121

[29] Janice Light and David McNaughton (2012), "The Changing Face of Augmentative and Alternative Communication: Past, Present, and Future Challenges", Augmentative and Alternative Communication, 28:4, 197-204, DOI: 10.3109/07434618.2012.737024

[30] Jeff Sigafoos, Mark F. O. Reilly, Giulio E. Lancioni, Dean Sutherland, "Augmentative and Alternative Communication for Individuals with Autism Spectrum Disorder and Intellectual Disability", Springer International Publishing, Jan 2014, Online ISSN: 2196-2987, DOI: https://doi.org/10.1007/s40474-013-0007-x

[31] "Attainment Go Talk Pocket" [Online] Available: https://www.enablemart.com/attainment-gotalk-pocket

[32] "Logan Prox Talker" [Online] Available: https://logantech.com/products/proxtalker

[33] "MegaBee Eye Pointing Communication Tablet" [Online] Available: https://logantech.com/products/megabee

[34] "Beamz Interactive Music System" [Online] Available: https://www.inclusivetlc.com/beamz-interactive-music-system

[35] "Tobii Dynavox I-Series+" [Online] Available: https://www.tobiidynavox.com/devices/eye-gaze-devices/i-15-with-communicator/

[36] G.J. Edwards, C.J. Taylor, T.F. Cootes, "Interpreting face images using active appearance models", Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition,1998, DOI:10.1109/AFGR.1998.670965

[37] "ITU-T Rec. P 800 (08/96) Methods for subjective determination of transmission quality" [Online] Available: https://www.itu.int/rec/T-REC-P.800-199608-I/en

[38] Viswanathan Mahesh and Viswanathan Madhubalan (2005), "Measuring speech quality for text-to-speech systems: Development and assessment of a modified mean opinion score (MOS) scale", Computer Speech and Language-Elsevier, 19(1): 55-83, DOI:10.1016/j.csl.2003.12.001.

[39] Black Alan W. and Tokuda Keiichi, "The Blizzard Challenge 2005: Evaluating corpus-based speech synthesis on common datasets", Conference: INTERSPEECH 2005,77-80 - Eurospeech, 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4-8, 2005

[40] Yu Yun Chang, "Evaluation of TTS systems in intelligibility and comprehension tasks", ROCLING '11 Proceedings of the 23rd Conference on Computational Linguistics and Speech Processing Pages 64-78, Taipei, China-September 08-09, 2011