

Gesture Technology for Deaf and Dumb

¹Prof. Pavan Kumar E, ²Lithin C T, ³Balaraju K B, ⁴Nanjundeshwara B L, ⁵Mahesh Ashok Mankar

Electronics and Communication Engineering,
Sai Vidya Institute of Technology, Bengaluru, India

Abstract: This paper is to perceive different hand motion and order different hand developments used to ease correspondence of hard of hearing and dumb. The hand development for signal recognition can be static or dynamic. The hand motion are caught utilizing a camera, the perceived hand motion are handled utilizing different available algorithms, for example, multivariate fluffy choice tree, hidden Markov model (HMM), support vector machine (SVM). The motions are caught under legitimate illumination. The caught motion are checked for occlusion and finger close motion to distinguish and perceive legitimate motion and disregard some other discontinuous motion. At last, the characterized motions are contrasted with preparing and testing dataset with test yield.

This paper is additionally gone for giving discourse to content conversion, speech is the essential method for correspondence, in case of hard of hearing, where it is hard to analyze lip reading. The sound information is taken from microphone and is compared with the google dataset. The coordinated outcome is shown as content output on the display.

Index Terms— *Gesture, Speech recognition, Computer vision.*

I. INTRODUCTION

Body language is a type of non-vocal communication. Hand motion, lip development, eye developments are different types of non-verbal communication. Gesture is the technique for communicating or passing on data by the movement of the body. Gestures can be of two sorts either static or dynamic signal.

Signal acknowledgment is a multi-disciplinary approach with PC vision, pattern acknowledgment, movement investigation and machine learning essentially. Hand gesture methods are of two sorts (a) vision based system and (b) non vision based strategy.

In vision based system different calculations are utilized to foresee position of hand, direction and other natural factors, for example, lighting conditions, hand blocking and skin shading which are significant piece of the recognition. In the non-vision based method, a glove with worked in sensors is utilized to perceive the signal.

Gesture is a blend of three process (a) Extraction, (b) Estimation and arrangement, (c) Recognition.

Starting procedure of extraction is the segmentation in which input information is cut into different areas. Skin shading is additionally a piece of segmentation process. Different instruments utilized for division process are gaussian model and gaussian mixture model. State of hand and highlights are evaluated with hand framework or fingertip position. With portioned and evaluated inputs, different grouping algorithms are utilized for signal acknowledgment, for example, hidden Markov model (HMM), finite state machine (FSM) and so forth.

The data in the computerized world is available just for a couple of who can read and comprehend the language. The innovation can give arrangement by interfacing the substance to mass and trade data crosswise over various individuals talking diverse dialects, here we additionally include discourse to content change which take contribution as discourse from the amplifier and the yield is shown as content.

Speech processing includes the investigation of Speech signals and numerous strategies used to process them. Various applications support Speech processing. Some of them incorporate discourse coding, Speech recognition, and speaker acknowledgment advances. Among different available applications, Speech recognition acknowledgment is the suitable one. The motivation behind Speech processing is to change over acoustic marked from the receiver to a word or a sentence. Keeping in mind the end goal is to extricate and decide the data passed on by a Speech processing, electronic circuits are utilized.

II. GESTURES

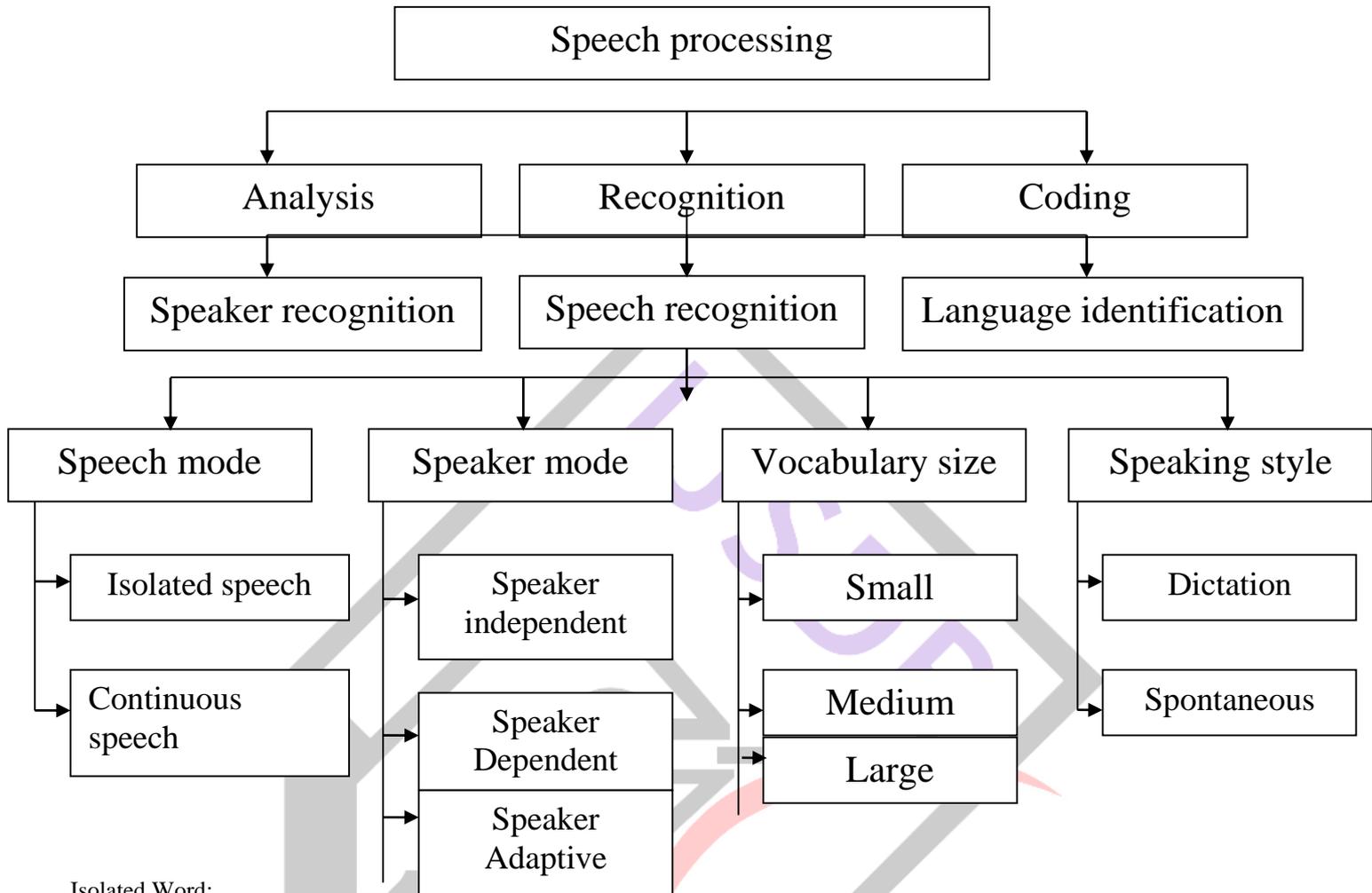
The motion contribution to the framework is acquired from image, videos and datasets. The input motions might be either 2 dimensional (plane), 3 dimensional (free form), static or dynamic gestures. Static signals are configuration dependent. For example, position of hands.

The dynamic signals are movement dependent. Static hand motion can be expressed as a blend of hand location, point, bearing and arrangement at the moment of time. Static motions are not changed with moment of time.

Dynamic hand signal can be expressed as an arrangement of ceaseless motions including development of hand. In a video grouping each casing demonstrate stance and whole succession shows a motion.

III. TYPES OF SPEECH

The system for speech recognition can be classified into different classes based on type of utterance recognized.



Isolated Word:

Detached word perceives every articulation to have calm on both sides of test windows. It acknowledges Single words or a single expression at once .This is having "Listen and Non- Listen state". Detached articulation may be better name of this class.

Connected Word:

Associated word framework is like detached words however permit to separate or separate sound to run together with least stop between them.

Continuous discourse:

Persistent discourse recognizers enables client to talk almost naturally, while the PC decide the content. Recognizer proceeds with discourse abilities that are absolute most difficult to make since they use remarkable sound and unique strategy to determine utterance limits.

Spontaneous discourse:

At initial level, it can be considered as a discourse that is natural sounding but not practiced. An ASR System with Spontaneous speech capacity is to have the capacity to deal with an alternate words and assortment of normal speech feature, for example, words being run together.

IV. VARIOUS SPEAKER MODELS

Every speaker has their remarkable voice in light of their identity

1) Speaker autonomous model

It is intended for variety of speakers. It has ability to perceive substantial number of people. This kind of framework is troublesome, costly and is less accurate. But it is very adaptable.

2) Speaker subordinate model

These are intended for an individual speaker. It is anything but difficult to develop, cheap and more accurate. These are less adaptable contrasted with speaker autonomous framework.

Kinds of Vocabulary

The measure of vocabulary influences the complexity, processing, performance and accuracy of the framework.

Kinds of vocabulary can be delegated as

- a) Small vocabulary – couple of ten words.
- b) Medium vocabulary – couple of hundred words.
- c) Large vocabulary - couple of thousand words.

Natural fluctuation, speaker style, sex, age, rate of discourse influences multifaceted nature of framework.

V. Gesture Capturing device

In order to catch the info signals, we make utilization of Raspberry-Pi camera module V1. This camera has a determination of 5 Megapixels supporting various video modes, for example, 1080p30, 720p60 and 640*480p60/90. It has Omnivision OV5647 sensor. This module bolsters outline rate upto 120 fps.

VI. Gesture Recognition Algorithm

HMM is a standout amongst the most broadly acknowledged apparatus for time-arrangement examination as this technique is fit for showing worldly connections among various models and tests likewise to division and characterization. This capacity is broadly utilized for preparing and impedance by an alternate variety of HMM.

HMM is considered as a standout amongst the most prevalent approaches for dynamic hand motion acknowledgment. In HMM, covered up factors guarantee diverse states and variations from these state advances are **unmistakable**. The yield of HMM additionally comprises of shrouded data of state succession, and however each state is **not discernible**, potential **yield** has probability dispersion.

VII. Experimentation dataset and Verification

Sign Language

One of the prominent standard datasets for hand development acknowledgments are hand communications via gestures. Human communication via gestures is one of the helpful technique for conveying to hearing tested individual and hard of hearing individuals. Gesture based communication is one of the visual dialects conveyed utilizing lip signs, facial articulations, hands and arm activities. Gesture based communication comprehension and recognizable proof process includes in unequivocally deciphering the sign activities and developments to discourse/voice or printed messages.

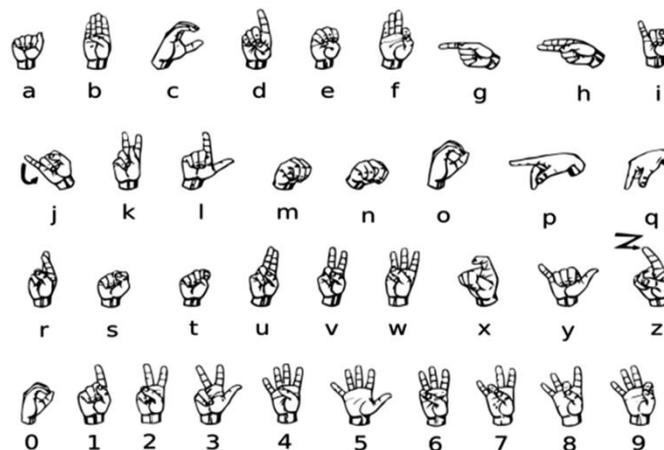


Fig. 2: Letter mapped for various gestures

Fig. 2 demonstrates the mapping of letters in order and numbers to special hand signals in American gesture based communication

Gesture based communications are not worldwide but rather are local particular offering ascend to various structures. For example, American communication through signing, Indian gesture based communication, Chinese gesture based communication, German communication through signing etc .

All the 26 letter sets in English vocabulary can be given a kind hand motions. Acknowledgment forms include, removing the important highlights from these signals and translate the particular implications lastly create tor instant messages.

COLOR TIP DATASET

Color Tip dataset is one people in general datasets utilized for hand motion acknowledgment and fingertip distinguishing proof. These datasets comprise of set of accounts and explanations caught from kinect sensor camera. In this dataset, signals are recorded from subjects wearing shaded gloves. Each motion changes in introduction and interpretation with many-sided quality and testing factor contingent upon the measure of intra-gesture fluctuation.

Audio to text system implementation

The implementation procedure accomplishes discourse to content by utilizing Raspberry pi and Raspbian picture introduced on raspberry pi. This is the essential advance to comprehend the summons of Linux for changing over discourse to content conversion. The Raspberry pi is a ultra-ease, deck-of-cards measured linux PC. It is controlled by an altered from Debian Linux enhanced for the ARM engineering. It has two models display A and B. The Model B has 512 MB RAM, BCM2385 ARM11, 700MHz System on chip processor. It has USB sound card that interfaces with pi display since it has just sound yield jack thus by utilizing sound card we can include voice information into pi.

Requirements for this execution:

1. The Raspbian module.
2. A web association.
3. The solid card drivers for headset.

VIII. **Automatic Speech Recognition System**

ASR frameworks work in two stages. Initial, a preparation stage, amid which the framework takes in the reference designs speaking to the diverse discourse sound (e.g. phrases, words, telephones) that constitutes the vocabulary of the application. Each reference is found out from talked cases and put away either as formats got by some averaging technique or models that portray the measurable properties of pattern. Second, a perceiving stage, amid which an obscure info design, is recognized by thinking about the arrangement of references.

SPEECH RECOGNITION TECHNIQUES:

The objective of discourse acknowledgment is for a machine to have the capacity to “hear”, “comprehend” and “follow up on” talked data. The most punctual discourse acknowledgment frameworks were first endeavored in the mid 1950s at Bell Laboratories. The objective of programmed speaker acknowledgment is to break down, remove portray and perceive data about the speaker character. The speaker acknowledgment framework might be seen as working in a four stages

- a. Examination
- b. Feature extraction
- c. Modeling
- d. Testing

[a]. Speech Examination:

In speech examination method speech information contains distinctive demonstrates a speaker character. This incorporates speaker particular data because of vocal tract, excitation source and conduct include. The physical structure also, measurement of vocal tract and excitation source are one of a kind for every speaker. This uniqueness is inserted in the discourse motion amid discourse creation and can be utilized for speaker utilized for speaker acknowledgment.

[b]. Highlight Extraction Technique:

Highlight Extraction is the most essential grammatical form acknowledgment since it assumes a vital part to isolate one discourse from other. Since each discourse has diverse individual qualities installed in utterances, these qualities can be separated from an extensive variety of highlight extraction procedures proposed and effectively misused for discourse acknowledgment assignment. In any case, separated element should meet a few criteria while managing the discourse flag such as:

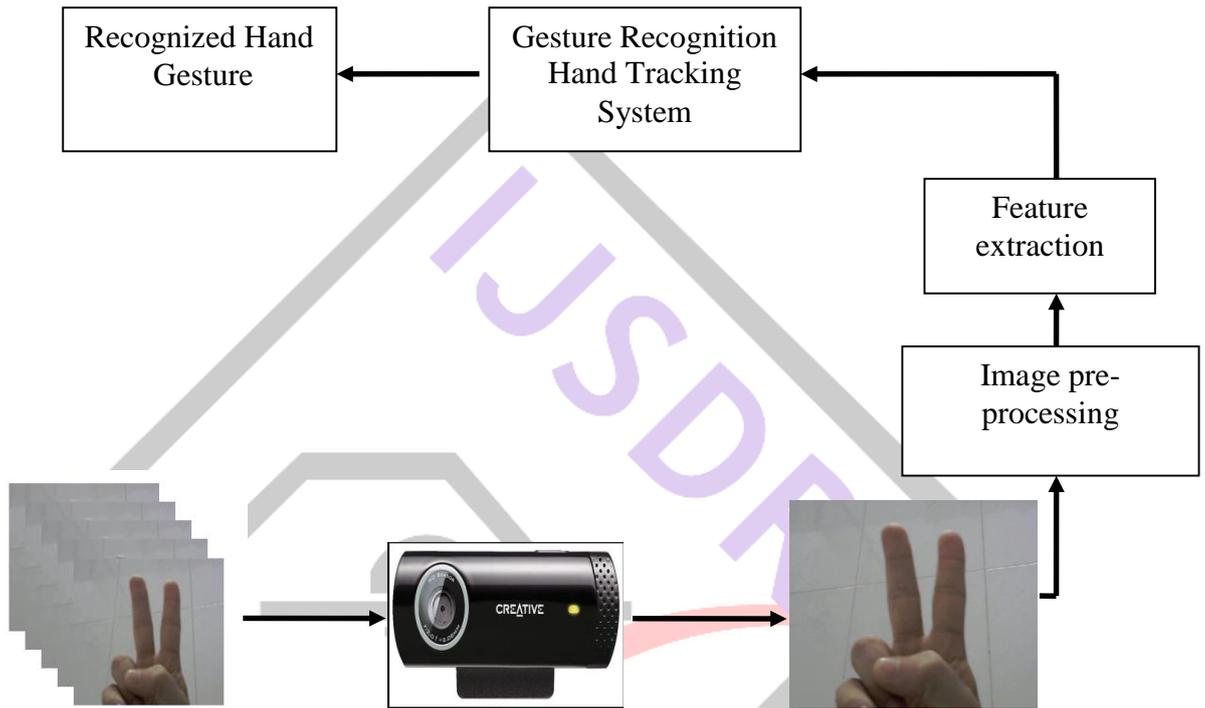
- a. Easy to quantify removed discourse highlights.
- b. It ought not to be defenseless to mimicry.
- c. It should demonstrate little vacillation starting with one talking condition then onto the next.

- d. It ought to be steady after some time.
- e. It ought to happen every now and again and normally in discourse.

IX. System Operation

Operation Of Hand Gesture Recognition System

Hand gesture recognition is the process of conveying information through motion of the body. Since we are using dynamic gesture recognition technique, this system intakes live video as input. The video is captured with the raspberry Pi camera module. User interface window has a boundary box, the frame captured within boundary box is segmented. In preprocessing phase the segmented frame is converted into grey scale. Further classification algorithm such as HMM is used to classify gestures based on the training dataset. Based on the classified gesture output text is displayed on the screen.



Operation of Speech to text conversion

The audio input is collected from the microphone. Here Speech recognition algorithm is used. Port audio is responsible for audio collection which is obtained in Pyaudio library file. Using the listen function the input audio is compared to the Google dataset. The matched output is displayed in the text form.

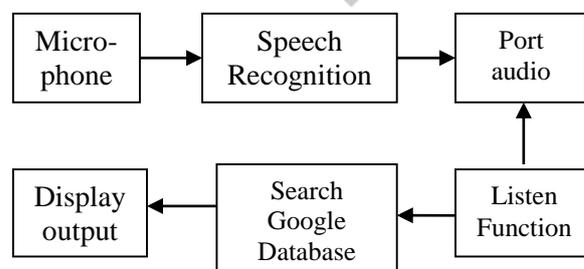


Fig. 4: Methodology for Speech recognition

X. VALIDATION

In this proto-type model, Hand Gesture and Speech recognition through computer vision technology. Firstly we have implemented Hand gesture recognition using cv2 libraries in which we have plotted the contours of fingers and is displayed on console.

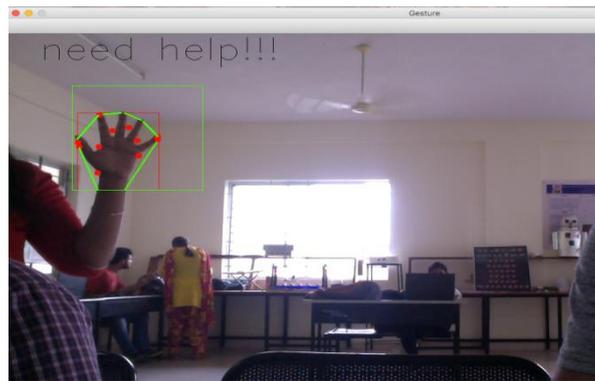


Fig. 5: Hand contour recognition

Fig. 5 displays the output of contour which displays “need help”. Fig. 6 shows the exact gradients of contour with points. In second phase we have displayed implementation of audio signal from Mic to text output. Speech recognition uses lot of libraries that recognize each and every word and sentence and is given to Audio port. The Audio port compares the input to the Google database and displays the output on the console as shown in Fig. 7.

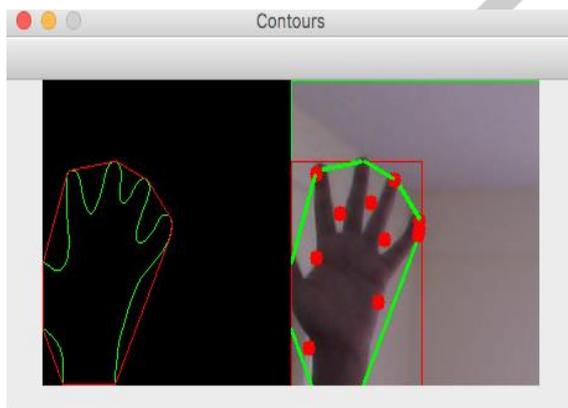


Fig. 6: Gradient points of recognized Hand

```
Python 2.7.14 Shell
File Edit Shell Debug Options Window Help
Python 2.7.14 (vs2.7.14:84471938ed, Sep 16 2017, 20:19:30) [MSC v.1800 32 bit (Intel)] on win32
Type "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
You said: Namaskara
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
Google Speech Recognition could not understand audio
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
You said: Kannada swagatha suswagatha
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
Google Speech Recognition could not understand audio
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
You said: Sai Vidya Apna Suswagatham
>>>
===== RESTART: C:/Python27/speech recognition.py =====
Say something!
You said: Sai Vidya Mein apka swagat hai
>>> |
```

Fig. 7: Output of Speech recognition

REFERENCES

- [1] Srinivas K, Manoj Kumar Gopal, “Study Of Hand Gesture Recognition And Classification”, *Asian Journal of Pharmaceutical and Clinical research*, 2017.
- [2] Kumbharana, Chandresh K, “Speech Pattern Recognition for Speech to Text Conversion”, *thesis PhD*, Saurashtra University, 2007.
- [3] Prachi Khilari, Prof. Bhope V. P, “Implementation of Speech to Text Conversion”, *Vol. 4, IJIRSET*, 2015.
- [4] Nuzhat Atiqua Nafis, Md. Safaet Hossain, “Speech to Text Conversion in Real-time”, *International Journal of Innovation and Scientific Research*, 2015