SILENT SOUND TECHNOLOGY

A Technical Seminar Report

Submitted by

MANAGARI SHIVANI

(17891A0435)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

VIGNAN INSTITUTE OF TECHNOLOGY AND SCIENCE
Near Ramoji FIlm City, Deshmukhi (V), Yadadri Bhuvanagiri Dist., Telangana - 508 284.
Approved by AICTE, New Delhi, Affiliated to JNTUH, Hyderabad
EAMCET CODE : VGNT PGECET CODE : VGNT1

2020-21

VIGNAN INSTITUTE OF TECHNOLOGY AND SCIENCE
Near Ramoji FIlm City, Deshmukhi (V), Yadadri Bhuvanagiri Dist., Telangana - 508 284.
Approved by AICTE, New Delhi, Affiliated to JNTUH, Hyderabad
EAMCET CODE : VGNT PGECET CODE : VGNT1

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that MANAGARI SHIVANI, 17891A0435 submitted the Technical Seminar Report titled “SILENT SOUND TECHNOLOGY” in partial fulfilment of the requirements for the award of degree of Bachelor of Technology in the Department of Electronics and Communication Engineering to Vignan Institute of Technology and Science.

Dr. P. A Harsha Vardhini — **Dr. N. Dinesh Kumar**

Tech. Sem Coordinator — **Head of the Department**

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of the task would be great but incomplete without the mention of the people who made it possible with their constant guidance and encouragement crowns all the efforts with the success.

I would like to express our heart-felt gratitude to our parents without whom, I would not have been privileged to achieve and fulfil our dreams.

I profoundly thank **Dr. N. DINESH KUMAR,** Head of the Department of Electronics and Communication Engineering who has been the source of inspiration and laid a great support to our work.

I would like to thank technical seminar coordinator **Dr P. A. HARSHA VARDHINI, Professor, Department of ECE** for her technical guidance, constant encouragement and support in completing the seminar and report.

Managari Shivani

17891A0435

INDEX

CONTENTS **PAGE NO**

LIST OF FIGURES AND TABLES

CONTENTS **PAGE NO**

ABSTRACT

Silent sound technology SST has being introduced to put end to noise pollution and help the people who have lost their voice and cannot speak on mobile phone. This device is developed at Karlsruhe institute of technology (KMT) and expected to be see in near future. This device will notice the lip movement, in the form of electrical impulse and transfer it to sound speech that can be understood. It will be useful for people who want to make a silent call by just receiving the electrical impulse from lips movement and neglect all other surrounding noise and convert it to sound speech at the receiver ends. It can be used for languages like English, German and French but it cannot be used for language like Chinese because a different tone means different meaning. It will be useful for secrete calling because the caller don't need to utter a word loudly just the lips movement. Silent sound technology (taking without talking) work base on two methods which are electromyography (EMG) and image processing.

CHAPTER – I

INTRODUCTION

Technology SST is a technology for devices that helps for communication purpose in the nasty environment. The uses of this technology are immense for people who are vocally challenged or have been rendered mute due to some accidents or others. Lip detection is a complex problem because of high variability range of lip shapes & colour. Lip-reading is an inference and inspired guesswork because of fast speech, poor pronunciation, bad lighting, faces turning away, hands over mouths, moustaches and beards etc. Lip Tracking is one of the biometric systems based on which a genuine system can be developed. With multiple levels of video processing, it's possible to obtain lip contour and location of key points in the subsequent frames is usually referred to as lip tracking. A large category of techniques referred to as model based, build a model of the lips and its configurations are described by a set of model parameters. Most of these techniques include tracking of the lip in sound speech may be with different accent & other facial parts consideration.

Our effort is to work on silent speech which means no sound is incurred; a device oriented package to design and implement for the purpose of lip reading that can recognize mandarin words, single sentence or even continuous sentences of the people of different regions in China country considering their non-speech accent and pronunciation by observing their every movement of the lip and facial expression.

Silent sound technology is the best solution for those people who lost their voice but wish to speak over phone .Silent Sound Technology aims to notice every movement of the lips and converted them into sounds, which helps people who are unable to speak, and allow people to make silent calls without disturbing others or sounds. The cell phone would detect the movements of mouth by measuring muscle activity, then convert this into speech so that the person on the other side of the call can hear .This new technology is very helpful for old people. Even we can tell our PIN number to a trusted friend or relative. At the other side, the listener can hear a clear voice. I.E. movements can be immediately converted into the language of the user's choice different toner language.

CHAPTER – II

HISTORY AND ORIGINATION

2.1 HISTORY

The Silent Sound technology is a technology that helps you to transmit information without using your vocal cords. Silent Sound technology is developed at the Karlsruhe Institute of Technology, Germany. This technology uses electromyography. It monitors tiny muscular movements that occur when we speak and converting them into electrical pulses that can then be turned into speech, without a sound uttered. (Kendra, 1995) It is very useful for those people who can't speak.

By using this technology they can easily interact with the other persons. The benefit of this technology is that the listener can hear voice clearly. This technology aims to notice lip movements & transform them into a computer generated sound that can be transmitted over a phone. Hence person on other end of phone receive the information in audio. The idea of interpreting silent speech electronically or with a computer has been around for a long time, and was popularized in the 1968 Stanley Kubrick science-fiction film ''2001-A Space Odyssey." (Brumberg, 2010)

In the 2010 CeBIT's "future park", a concept (Schultz, 2010) "Silent Sound" Technology demonstrated which aims to notice every movement of the lips and transform them into sounds, which could help people who lose voices to speak, and allow people to make silent calls without bothering others. Rather than making any sounds, your handset would decipher the movements your mouth makes by measuring muscle activity, then convert this into speech that the person on the other end of the call can hear. So, basically, it reads your lips. Thus people can talk on their cell phones at the crowded place without getting disturbed .

"We currently use electrodes which are glued to the skin. In the future, such electrodes might for example by incorporated into cell phones," said (Schultz, 2010) from the KIT. The technology opens up a host of applications, from helping people who have lost their voice due to illness or accident. The technology can also turn you into an instant polyglot. Because the electrical pulses are universal, they can be immediately transformed into the language of the user's choice. "Native speakers can silently utter a sentence in their language, and the receivers hear the translated sentence in their language. It appears as if the native speaker produced speech in a foreign language," said Schultz.

2.2 ORIGINATION

Humans are capable of producing and understanding whisperer speech in quiet environments at remarkably low signal levels. (Jorgensen C, 2010) Most people can also understand a few unspoken words by lip-reading The idea of interpreting silent speech electronically or with a computer has been around for a long time, and was popularized in the 1968 Stanley Kubrick science-fiction film ‘‘2001 – A Space Odyssey ” A major focal point was the DARPA Advanced Speech Encoding Program (ASE) of the early 2000’s, which funded research on low bit rate speech synthesis ‘‘with acceptable intelligibility, quality , and aural speaker recognisability in acoustically harsh environments”.

CHAPTER III

OVERVIEW OF SST

3.1 DEFINITION

It is a technology that helps you to transmit information without using your vocal cords. This technology aims to notice lip movements & transform them into a computer generated sound that can be transmitted over a phone. Hence person on other end of phone receives the information in audio.

Silent Speech technology enables speech communication to take place when an audible acoustic signal is unavailable. By acquiring sensor data from elements of the human speech production process – from the articulators, their neural pathways, or the brain itself – it produces a digital representation of speech which can be synthesized directly, interpreted as data, or routed into a communications network.

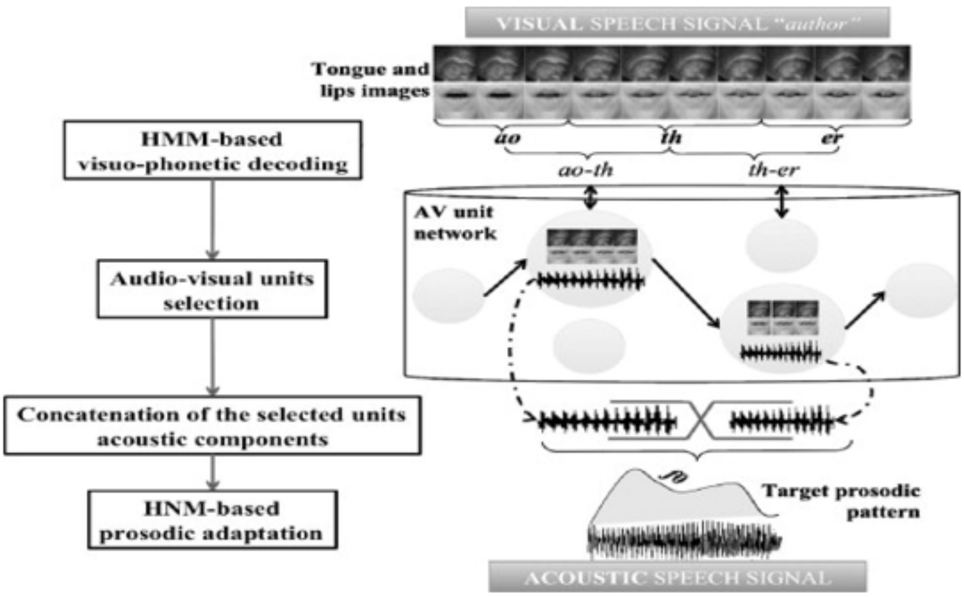


Fig 3.1: FLOWCHART OF SILENT SOUND TECHNOLOGY

3.2 DEPROCESS MODEL AND WORKING METHODOLOGY

To proceed with this research work, the Process Model assumed is Iterative Process Model since it is more adaptable for this work. Once the face detection and mouth region detection is achieved, speech analysis can be performed with the use of lip motion features strategies and emotional expression with the use of other facial parts. If efficiency with identification technique is not proper then the threshold value falls out of the defined unique index value and retrial has to be made. Those are one of the main reasons to choose the Iterative Process.

As the live video is captured by a high resolution camera, the video can be processed as normal or grayscale colour mode /or saved as Mpeg, Avi, Flv etc. for customization. Region of Interest (ROI) video is segmented from which Facial features like Mouth, nose & Eyes are detected.

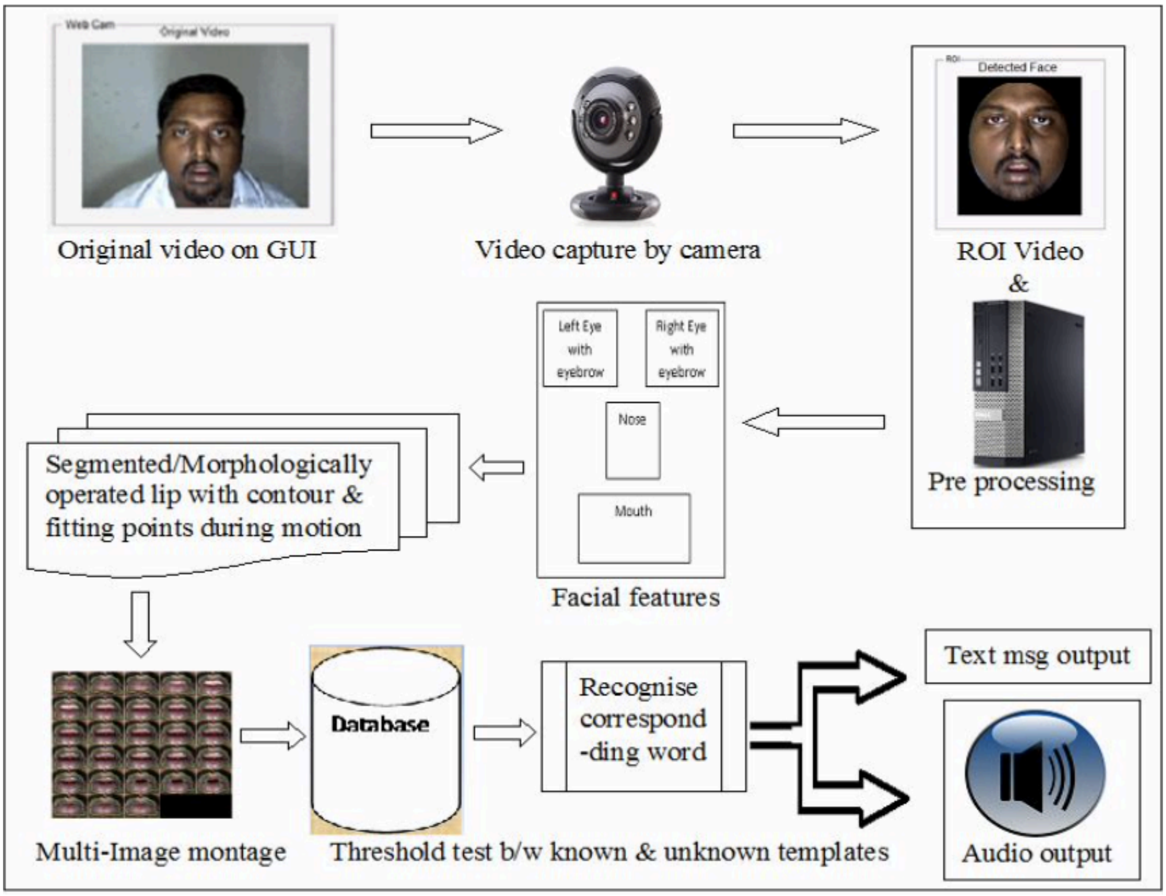


Fig 3.2: BLOCK DIAGRAM

As the lip contour is initiated accurately, the extracted lip contour is morphologically processed and corners are fitted by the key points like left/right and upper/lower corners with centroid. Obtaining the lip contour in subsequent frames and the lip movement is completely tracked. A multi frame montage in a single object image montage is built and a database is created for it and obtained feature like eyes and nose vector in the database. The unknown templates of a user are then compared with the existing templates in the database. If the frames pass the threshold test with the known & unknown templates, based on trained and tested index value/trials the user can receive a text output & later an audio output .

3.3 SKIN SEGMENTATION AND MORPHOLOGY

One the important step in face feature extraction process is skin segmentation. As we know human face varies from person to person, so the race, gender, age and other physical characteristics of an individual have to be considered. The partition is based on the color difference between the skin and non-skin regions . The efficiency of the colour segmentation of a human face depends on the colour space that is selected. We used the finding by Yang and Waibel who noted that the skin colours of different people are very closely grouped in the normalized R-G colour plane. So some of the seed pixels are taken from the face area obtained the normalized R-G spread and classified the pixels that were close enough to this spread as belonging to the face.

In this phase we detect face region from input video, extracting it into frames either in grayscale or colour mode.

To extract face region we perform lighting compensation on image, then extract skin region and remove all the noisy data from image region. Finding skin colour blocks from the image and then check face criterions of the image. In lighting compensation we normalize the intensity of the image, when extracting skin region then apply threshold for the chrominance and then select the pixels that are satisfying the threshold to find the colour blocks. The skin colour blocks are identified based on the measure properties of image regions in image. Height and width ratio is computed and minimal face dimension constraint is implemented. Crop the current region, existence and localization of face then compute vertical histogram . As a result of skin segmentation the interior holes that are created on the face region are morphologically processed.

Morphological image processing is a collection of techniques for digital image processing based on mathematical morphology. Since these techniques rely only on the relative ordering of pixel values, not on their numerical values, they are especially suited for processing of binary images and grayscale images based on two functions [2]:

Erosion- is an operation that thins or shrinks the objects in the binary image. Erosion are performed by IPT function imerode().

Dilation- is an operation that grows or thickens the objects in the binary image. Dilation is performed by IPT function imdialate().

For successful facial feature extraction the accurate computation of the contour points is very important. This helps in locating searching regions. Both neck and ears have the same colour as that of the face. Hence they are connected to the face region. Therefore we need to separate them so as to better locate the facial features. The Fig. 2 (a,b,c,d) demonstrates our results.

a. Original image b. Skin segmented image c. Binary image d. Active shape of face

Fig 3.3: MORPHOLOGICAL IMAGE AND FACE

The face detection from the captured video is based on the Active Shape Model (ASMs) [12]. The shape model is learned from a set of manually annotated shapes of faces (unobstructed frontal views) as follows: the first step is to align all shapes of the learning data to an arbitrary reference by a geometric transformation (rotation, translation and scaling). The second step is to calculate the average shape. These two steps are repeated until convergence by minimizing the average Euclidean distance between shape points. At the end of the process, the facial shape model is obtained by active contour analysis (ACA) of the average of the aligned shapes using computer vision toolbox.

faceDetector=vision.CascadeObjectDetector('FrontalFaceLBP'); videoFileReader = vision.VideoFileReader('path');

bwface = activecontour(grayface,mask,40,'Chan-Vese');

bw = activecontour(A,mask,method)

The above command helps to detect the face and specifies the active contour method used for segmentation, either 'Chan- Vese' or ‘edge'. Thus, we obtain the principal modes of shape variations.

To extract the facial features from a new image, first, face is detected using the MAT lab command based on Viola & Jones face detection algorithm [6, 7]. After that, the shape model is positioned on the face, and iteratively deformed until it sticks to the face of the image in the respective bounding boxes. The facial features can be located with high reliability in standard lighting conditions, frontal face position and classical facial expression [4]. Both the methods are feasible.

A. Contour Fitting Points Location

The points of interest also referred to as key points or contour fitting points widely used for lip reading and other applications like relationship with a particular structure. A corner can be defined as points for which there are two dominant and different edge directions in a local neighbourhood of the point. We know that lip detection is a complex problem because of the high variability range of lip shapes and colour. So the main advantages of a corner detector is its ability to detect the same corner in multiple similar images, under conditions of different lighting, translation, rotation and other transforms. We detect the centroid point of the bounding box then centroid Column and centroid Row using the Mat lab command:

centroidColumn = int32(centroid(1)); % "X" value centroidRow = int32(centroid(2)); "Y" value. middleRow = binaryImage(centroidRow, :); middleColumn = binaryImage(:, centroidColumn); centroidColumn, centroidRow – centroid point

The upper and lower key points are found as the intersection points between the minor axis line and the upper and lower lip boundary, respectively.

topRowY = find(middleColumn, 1, 'first'); centroidColumn, topRowY -this gives top

bottomRowY = find(middleColumn, 1, 'last'); centroidColumn, bottomRowY -this gives bottom

Unlike the upper and lower keypoints, which are precisely detected from the color segmented lip object, the mouth corners (left and right keypoints) are more difficult to detect because of their location in dark areas, where chromatic information is not visible. In order to detect them, we use the extreme left and right points of the lip object as starting points and search for corners in the proximity area using the below MAT lab command:

leftColumnX = find(middleRow, 1, 'first'); leftColumnX, centroidRow -this gives left rightColumnX= find(middleRow, 1, 'last'); rightColumnX, centroidRow -this gives right

The proximity areas, where corners are searched out are explained above. From the corners detected in each side, we choose the one with the smallest Euclidean distance from the corresponding extreme object points as the left and right keypoints, respectively. If no corners are detected in either side, the corner strength threshold is automatically reduced until at least one corner is detected. In the case where only on the one side a keypoint is found, the corresponding key point on the other side is assumed using symmetry towards the minor axis . Once the image has got the lip shape, we select lip as the biggest object inside the image. The contour of these detected local maxima pixels will be defined as the left/right corners of the mouth and top/bottom corners. Get the perimeter of the lip region and find the (x, y) of that perimeter. Colour coding is applied on the perimeter for ease of understanding the proper lip contour .

3.4 MONTAGE AND SIMULATION PROCEDURE

As the video file reader divides the live video into individual image frames, each frame is resized and saved in the MAT lab directory after checking for its existence. All disordered frames are sorted and each one is stitched in the ascending order to form a multi frame montage in a single object image montage. These montages are considered to be known templates provided a threshold value is calculated based on the coordinates x and y v/s time for every phoneme or a word from which an average or mean value is obtained. Thus stored features vector in the data base and threshold value comparison test is conducted between these known and unknown templates. If the frames pass the threshold test of average trained tested index value/trials the user can receive a text output & later an audio output.

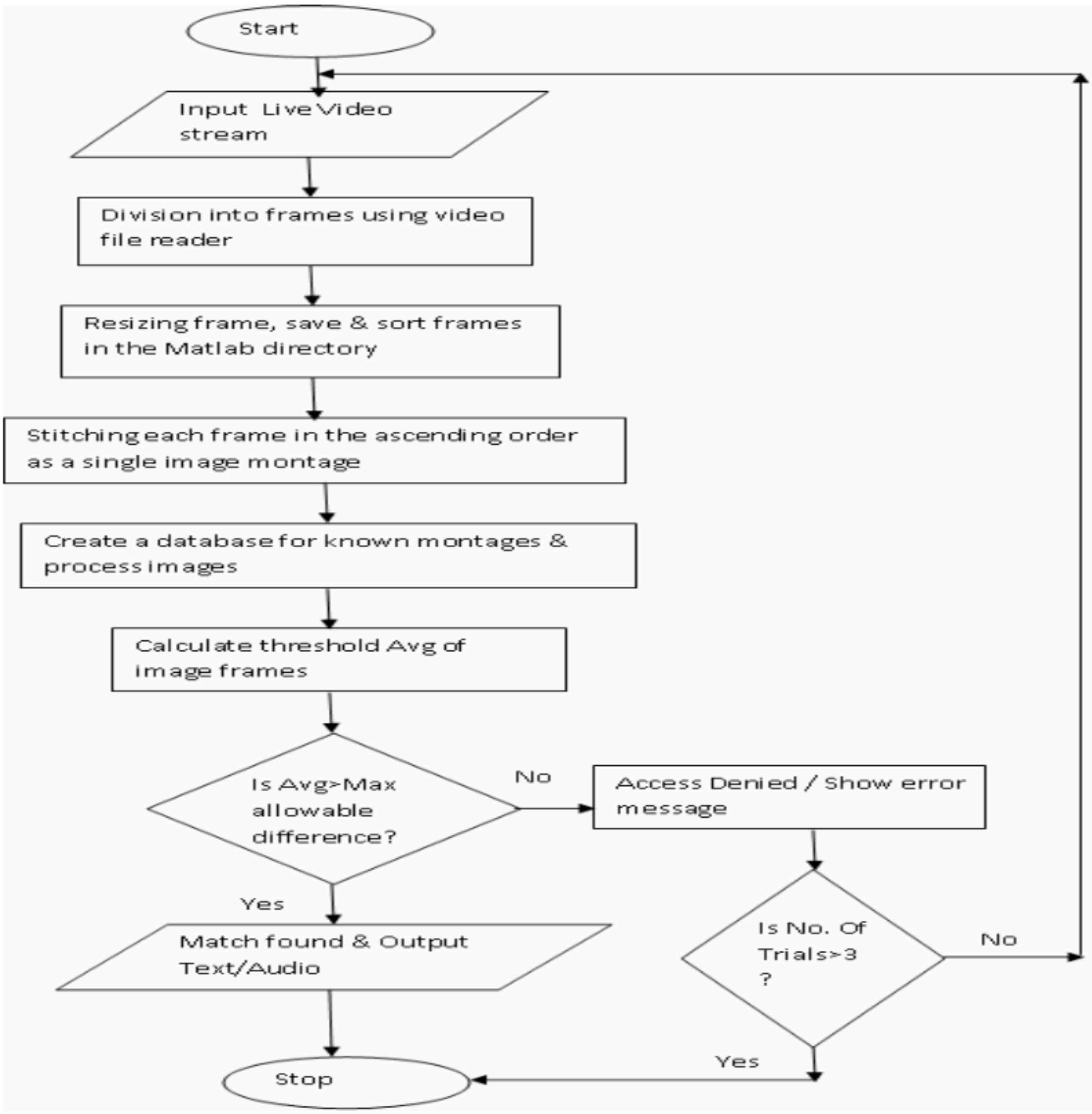


Fig 3.4: SIMULATION PROCEDURE OF THE MODEL

3.5 METHODS

Silent Sound Technology is processed through some ways or methods. They are –

1. Electromyography (EMG)
2. Image Processing

3.6 ELECTROMYOGRAPHY(EMG)

3.6.1 INTRODUCTION

The Silent Sound Technology uses electromyography, monitoring tiny muscular movements that occur when we speak. Monitored signals are converted into electrical pulses that can then be turned into speech, without a sound uttered. When we generally speak aloud, air passes through larynx & the tongue. Words are produced using articulator muscle in the mouth & jaw region. Recently it is proved that when we are about to speak due to reflex action of muscles the articulate muscle becomes active irrespective of whether air is passed through them or not. Even by saying words without producing sound weak electric currents are sent from brain to speech muscles. This phenomenon is called ELECTROMYOGRAPHS.

Fig 3.5: SENSORS ATTACHED TO THE FACE

An electromyograph detects the electrical potential generated by muscle cells when these cells are electrically or neurologically activated. The signals can be analysed to detect medical abnormalities, activation level, recruitment order or to analyse the biomechanics of human or animal movement.

- The Silent Sound Technology uses electromyography, monitoring tiny muscular movements that occur when we speak.
- Monitored signals are converted into electrical pulses that can then be turned into speech, without a sound uttered.
- Electromyography (EMG) is a technique for evaluating and recording the electrical activity produced by skeletal muscles.
- An electromyography detects the electrical potential generated by muscle cells, when these cells are electrically or neurologically activated.

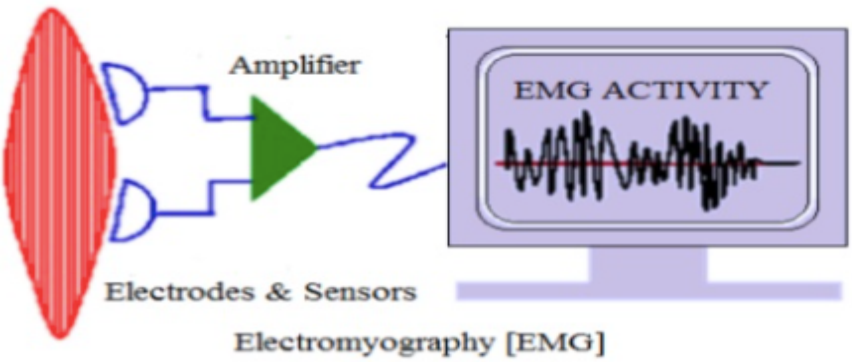


Fig 3.6: ELECTRO MORPHOGRAPHY SIGNAL GENERATION

3.6.2 ELECTRICAL CHARSTICSRACTE

The electrical source is the muscle membrane potential of about -90 mV. Measured EMG potentials range between less than 50 μV and up to 20 to 30 mV, depending on the muscle under observation. Typical repetition rate of muscle motor unit firing is about 7–20 Hz, depending on the size of the muscle (eye muscles versus seat (gluteal) muscles), previous axonal damage and other factors. Damage to motor units can be expected at ranges between 450 and 780 mV.

3.6.3 HISTORY

The first documented experiments dealing with EMG started with Francesco Redi's works in1666. Redi discovered a highly specialized muscle of the electric ray fish (Electric Eel)generated electricity. By 1773, Walsh had been able to demonstrate that the Eel fish's muscle tissue could generate a spark of electricity. In 1792, a publication entitled De Viribu Electrostatics in Motu Muscular Commentarius appeared, written by Luigi Galvani, in which the author demonstrated that electricity could initiate muscle contractions.

Six decades later, in1849, Dubois-Raymond discovered that it was also possible to record electrical activity during a voluntary muscle contraction. The first actual recording of this activity was made by Marey in 1890, who also introduced the term electromyography. In 1922, Gasser and Erlanger used an oscilloscope to show the electrical signals from muscles. Because of the stochastic nature of the myoelectric signal, only rough information could be obtained from its observation. The capability of detecting electromyographic signals improved steadily from the 1930s through the1950s, and researchers began to use improved electrodes more widely for the study of muscles. Clinical use of surface EMG (sEMG) for the treatment of more specific disorders began in the1960s. Hardyck and his researchers were the first (1966) practitioners to use sEMG. In the early 1980s, Cram and Steger introduced a clinical method for scanning a variety of muscles using an EMG sensing device. It is not until the middle of the 1980s that integration techniques in electrodes had sufficiently advanced to allow batch production of the required small and lightweight instrumentation and amplifiers. At present, a number of suitable amplifiers are commercially available. In the early1980s, cables that produced signals in the desired microvolt range became available. Recent research has resulted in a better understanding of the properties of surface EMG recording. Surface electromyography is increasingly used for recording from superficial muscles in clinical or kinesiological protocols, where intramuscular electrodes are used for investigating deep muscles or localized muscle activity. There are many applications for the use of EMG. EMG is used clinically for the diagnosis of neurological and neuromuscular problems. It is used diagnostically by gait laboratories and by clinicians trained in the use of biofeedback or ergonomic assessment. EMG is also used in many types of research laboratories, including those involved in biomechanics, motor control, neuromuscular physiology, movement disorders, postural control, and physical therapy.

3.6.4 PROCEDURE

There are two kinds of EMG in widespread use: surface EMG and intramuscular (needle and fine-wire) EMG. To perform intramuscular EMG, a needle electrode or a needle containing two fine-wire electrodes is inserted through the skin into the muscle tissue. A trained professional(such as a neurologist, physiatrist, or physical therapist) observes the electrical activity while inserting the electrode. The insertional activity provides valuable information about the state of the muscle and its innervating nerve. Normal muscles at rest make certain, normal electrical signals when the needle is inserted into them. Then the electrical activity when the muscle is at rest is studied. Abnormal spontaneous activity might indicate some nerve and/or muscle damage. Then the patient is asked to contract the muscle smoothly. The shape, size, and frequency of the resulting motor unit potentials are judged. Then the electrode is retracted a few millimetres, and again the activity is analysed until at least 10–20 units have been collected. Each electrode track gives only a very local picture of the activity of the whole muscle. Because skeletal muscles differ in the inner structure, the electrode has to be placed at various locations to obtain an accurate study.

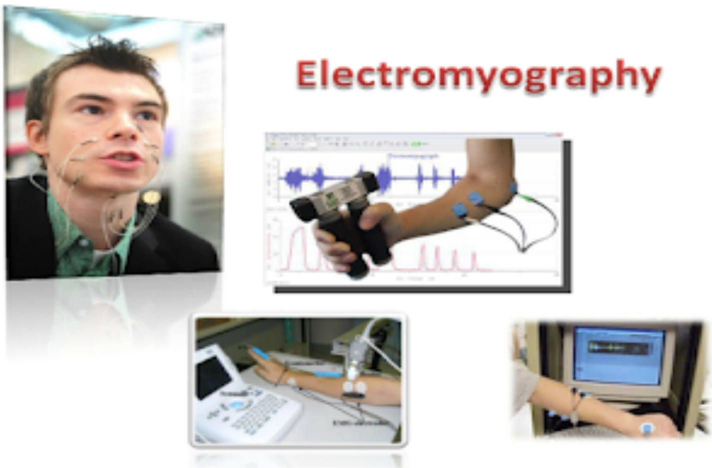


Fig 3.7: ELECTROMYOGRAPHIC SENSORS ATTACHED TO THE FACE RECORDS

As shown in Figure 3.3 the electromyography sensors attached to the face records the electric signals produced by the facial muscles, compare them with pre-recorded signal pattern of spoken words.

Intramuscular EMG may be considered too invasive or unnecessary in some cases. Instead, a surface electrode may be used to monitor the general picture of muscle activation, as opposed to the activity of only a few fibres as observed using an intramuscular EMG. This technique issued in a number of settings; for example, in the physiotherapy clinic, muscle activation is monitored using surface EMG and patients have an auditory or visual stimulus to help them know when they are activating the muscle (biofeedback).

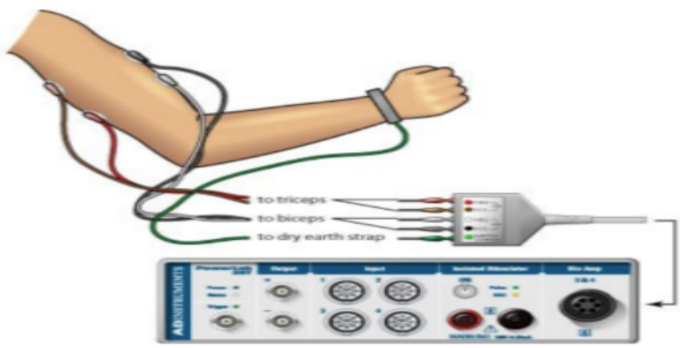


Fig 3.8: INTERFACING WITH ELECTROMYOGRAPHER AND BODY

A motor unit is defined as one motor neuron and all of the muscle fibres it innervates. When a motor unit fires, the impulse (called an action potential) is carried down the motor neuron to the muscle. The area where the nerve contacts the muscle is called the neuromuscular junction, or the motor end plate. After the action potential is transmitted across the neuromuscular junction, an action potential is elicited in all of the innervated muscle fibres of that particular motor unit. The sum of all this electrical activity is known as a motor unit action potential(MUAP). This electrophysiologic activity from multiple motor units is the signal typically evaluated during an EMG. The composition of the motor unit, the number of muscle fibres per motor unit, the metabolic type of muscle fibres and many other factors affect the shape of the motor unit potentials in the myogram. Nerve conduction testing is also often done at the same time as an EMG to diagnose neurological diseases. Some patients can find the procedure somewhat painful, whereas others experience only a small amount of discomfort when the needle is inserted. The muscle or muscles being tested may be slightly sore for a day or two after the procedure.

Normal results: Muscle tissue at rest is normally electrically inactive. After the electrical activity caused by their rotation of needle insertion subsides, the electromyograph should detect no abnormal spontaneous activity (i.e., a muscle at rest should be electrically silent, with the exception of the area of the neuromuscular junction, which is, under normal circumstances, very spontaneously active). When the muscle is voluntarily contracted, action potentials begin to appear. As the strength of the muscle contraction is increased, more and more muscle fibres produce action potentials. When the muscle is fully contracted, there should appear a disorderly group of action potentials of varying rates and amplitudes (a complete recruitment and interference pattern).Abnormal results: EMG is used to diagnose diseases that generally may be classified into one of the following categories:

neuropathies, neuromuscular junction diseases and myopathies. Neuropathic disease has the following defining EMG characteristics:

- An action potential amplitude that is twice normal due to the increased number of fibres per motor unit because of reinnervation of denervated fibres
- An increase in duration of the action potential
- A decrease in the number of motor units in the muscle (as found using motor unit number estimation techniques)

Myopathic disease has these defining EMG characteristics:

- A decrease in duration of the action potential

- A reduction in the area to amplitude ratio of the action potential
- A decrease in the number of motor units in the muscle (in extremely severe cases only)Because of the individuality of each patient and disease, some of these characteristics may not appear in every case.

3.6.5 WORKING OF ELECTROMYOGRAPHY

1. A needle containing two fine-wired electrodes is inserted through the skin into the muscle tissue.

2. Then the electrical activity when the muscle is at rest is observed. Each electrode track gives only a very local picture of the activity of the whole muscle.

3. Because skeletal muscles differ in the inner structure, the electrode has to be placed at various locations to obtain an accurate signal.

Thus by this way the speech can be communicated without sound the camera attached to the face.

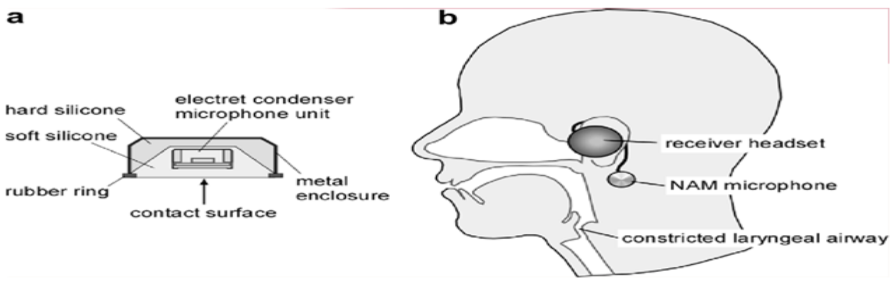


Fig 3.9: THE CAMERA ATTATCHED TO THE FACE

3.7 IMAGE PROCESSING

3.7.1 INTRODUCTION

In electrical engineering and computer science, image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or, a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it.

The simplest form of digital image processing converts the digital data tape into a film image with minimal corrections and calibrations.

Then large mainframe computers are employed for sophisticated interactive manipulation of the data. In the present context, overhead prospective are employed to analyze the picture.

Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard

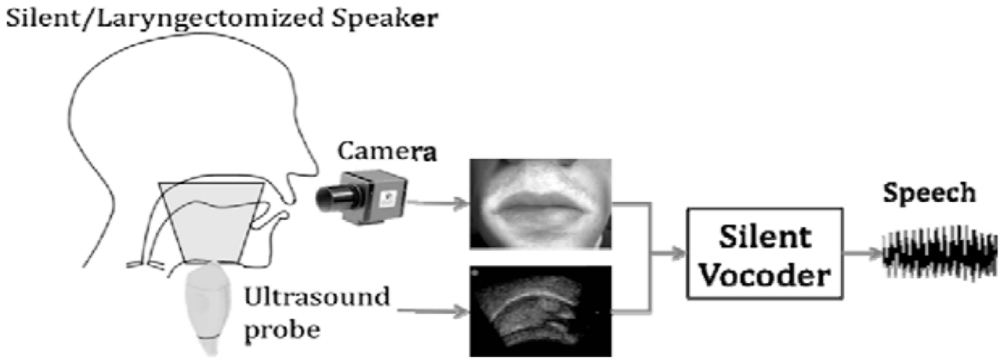


Fig 3.10: SST USING IMAGE PROCESSING

Signal-processing techniques to it. Figure 3.5 shows how image processing works in case of silent sound technology.

Analysis of remotely sensed data is done using various image processing techniques and methods that includes:

1. Analog image processing

2. Digital image processing

3.7.2 ANALOG PROCESSING TECHNIQUES:

- Analog processing techniques is applied to hard copy data such as photographs or printouts.
- It adopts certain elements of interpretation, such as primary element, spatial arrangement etc.,
- With the combination of multi-concept of examining remotely sensed data in multispectral, multitemporal, multiscale and in conjunction with multidisciplinary, allows us to make a verdict not only as to what an object is but also its importance.
- Apart from these it also includes optical photogrammetric techniques allowing for precise measurement of the height, width, location, etc. of an object. Analog processing techniques is applied to hard copy data such as photographs or printouts. Image analysis in visual techniques adopts certain elements of interpretation, which are as follow:

The use of these fundamental elements of depends not only on the area being studied, but the knowledge of the analyst has of the study area. For example the texture of an object is also very useful in distinguishing objects that may appear the same if the judging solely on tone (i.e., water and tree canopy, may have the same mean brightness values, but their texture is much different. Association is a very powerful image analysis tool when coupled with the general knowledge of the site. Thus we are adept at applying collateral data and personal knowledge to the task of image processing. With the combination of multi-concept of examining remotely sensed data in multispectral, multitemporal, multiscale and in conjunction with multidisciplinary, allows us to make a verdict not only as to what an object is but also its importance. Apart from these analog image processing techniques also includes optical photogrammetric techniques allowing for precise measurement of the height, width, location, etc. of an object.

ELEMENTS OF IMAGE INTREPRETATION	
Primary elements	Black and white tone
	Color tone
	Stereoscopic paralax
Spatial arrangement of tone and color	Size
	Shape
	Texture
	Pattern
Based on analysis of primary elements	Height
	Shadow
Contextual elements	Size
	association

Fig 3.11: ELEMENTS OF IMAGE INTERPRETATION

3.7.3 DIGITAL PROCESSING TECHNIQUE:

Digital Image Processing involves a collection of techniques for the manipulation of digital images by computers. Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modelled in the form of Multidimensional Systems. In a most generalized way, a digital image is an array of numbers depicting spatial distribution of a certain field parameters (such as reflectivity of EM radiation, emissivity, temperature or some geophysical or topographical elevation. Digital image consists of discrete picture elements called pixels. Associated with each pixel is a number represented as DN(Digital Number), that depicts the average radiance of relatively small area within a scene. The range of DN values being normally 0 to 255. The size of this area effects the reproduction of details within the scene. As the pixel size is reduced more scene detail is preserved Remote sensing images are recorded in digital forms and then processed by the computers to produce images for interpretation purposes. Images are available in two forms - photographic film form and digital form. Variations in the scene characteristics are represented as variations in brightness on photographic films. A particular part of scene reflecting more energy will appear bright while a different part of the same scene that reflecting less energy will appear black. Digital image consists of discrete picture elements called pixels. Associated with each pixel is a number represented as DN (Digital Number), that depicts the average radiance of relatively small area within a scene. The size of this area effects the reproduction of details within the scene. As the pixel size is reduced more scene detail is preserved in digital representation.

Image Resolution:

Resolution can be defined as "the ability of an imaging system to record fine details in a distinguishable manner". A working knowledge of resolution is essential for understanding both practical and conceptual details of remote sensing. Along with the actual positioning of spectral bands, they are of paramount importance in determining the suitability of remotely sensed data for a given applications. The major characteristics of imaging remote sensing instrument operating in the visible and infrared spectral region are described in terms as follow: Spectral resolution Radiometric resolution Spatial resolution Temporal resolution Spectral Resolution refers to the width of the spectral bands. As different material on the earth surface exhibit different spectral reflectances and emissivities. These spectral characteristics define the spectral position and spectral sensitivity in order to distinguish materials. There is a tradeoff between spectral resolution and signal to noise. The use of well -chosen and sufficiently numerous spectral bands is a necessity, therefore, if different targets are to be successfully identified on remotely sensed images.

Temporal resolution:

Temporal resolution refers to the frequency with which images of a given geographic location can be acquired. Satellites not only offer the best chances of frequent data coverage but also of regular coverage. The temporal resolution is determined by orbital characteristics and swath width, the width of the imaged area. Swath width is given by 2htan(FOV/2) where h is the altitude of the sensor, and FOV is the angular field of view of the sensor. It contain some flaws. To overcome the flaws and deficiencies in order to get the originality of the data, it needs to undergo several steps of processing.

Digital Image Processing undergoes three general steps:

1. Pre-processing

2. Display and enhancement

3. Information extraction

Fig 3.12: FLOW CHART OF DIGITAL IMAGE PROCESSING

Pre-Processing:

- Pre-processing consists of those operations that prepare data for subsequent analysis that attempts to correct or compensate for systematic errors.
- Then analyst may use feature extraction to reduce the dimensionality of the data.

- Thus feature extraction is the process of isolating the most useful components of the data for further study while discarding the less useful aspects.
- It reduces the number of variables that must be examined, thereby saving time and resources. Pre-processing consists of those operations that prepare data for subsequent analysis that attempts to correct or compensate for systematic errors. The digital imageries are subjected to several corrections such as geometric, radiometric and atmospheric, though all these correction might not be necessarily be applied in all cases. These errors are systematic andcan be removed before they reach the user. The investigator should decide which pre-processing techniques are relevant on the basis of the nature of the information to beextracted from remotely sensed data. After pre-processing is complete, the analyst may usefeature extraction to reduce the dimensionality of the data. Thus feature extraction is theprocess of isolating the most useful components of the data for further study whilediscarding the less useful aspects (errors, noise etc). Feature extraction reduces the numberof variables that must be examined, thereby saving time and resources.

Image Enhancement:

- Improves the interpretability of the image by increasing apparent contrast among various features in the scene.
- The enhancement techniques depend upon two factors mainly
- The digital data (i.e. with spectral bands and resolution)
- The objectives of interpretation
- Common enhancements include image reduction, image rectification, image magnification, contrast adjustments, principal component analysis texture transformation and so on. Image Enhancement operations are carried out to improve the interpretability of the image by increasing apparent contrast among various features in the scene.
- The enhancement techniques depend upon two factors mainly
- The digital data (i.e. with spectral bands and resolution)
- The objectives of interpretation

Information Extraction:

- In Information Extraction the remotely sensed data is subjected to quantitative analysis to assign individual pixels to specific classes. It is then classified.

- It is necessary to evaluate its accuracy by comparing the categories on the classified images with the areas of known identity on the ground.
- The final result of the analysis consists of maps (or images), data and a report. Then these are converted to corresponding signals. Information Extraction is the last step toward the final output of the image analysis. After pre-processing and image enhancement the remotely sensed data is subjected to quantitative analysis to assign individual pixels to specific classes. Classification of the image is based on the known and unknown identity to classify the remainder of the image consisting of those pixels of unknown identity. After classification is complete, itis necessary to evaluate its accuracy by comparing the categories on the classified images with the areas of known identity on the ground. The final result of the analysis consists of maps (or images), data and a report. These three components of the result provide the user with full information concerning the source data, the method of analysis and the outcome and its reliability.

Pre-Processing of the Remotely Sensed Images When remotely sensed data is received from the imaging sensors on the satellite platforms it contains flaws and deficiencies. Pre-processing refers to those operations that are preliminary to the main analysis. Pre-processing includes a wide range of operations from the very simple to extremes of abstractness and complexity. These categorized as follow:

1. Feature Extraction
2. Radiometric Corrections
3. Geometric Corrections
4. Atmospheric Correction

The techniques involved in removal of unwanted and distracting elements such as image/system noise, atmospheric interference and sensor motion from an image data occurred due to limitations in the sensing of signal digitization, or data recording or transmission process. Removal of these effects from the digital data are said to be “restored" to their correct or original condition, although we can, of course never know what are the correct values might be and must always remember that attempts to correct data what may themselves introduce errors. Thus image restoration includes the efforts to correct for both radiometric and geometric errors.

3.8 FEATURES OF SILENT SOUND TECHNOLOGY

Some of the features of silent sound technology are –

1. Native speakers can silently utter a sentence in their language, and the receivers can hear the translated sentence in their language. It appears as if the native speaker produced speech in a foreign language. The translation technology works to languages like English, French and German, except Chinese, where different tones can hold many different meanings.

2. Allow people to make silent calls without bothering others.

3. The Technology opens up a host of application such as mentioned below

4. Helping people who have lost their voice due to illness or accident.

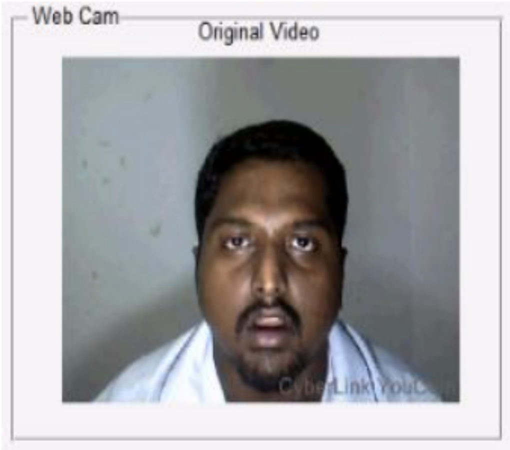
5. Telling a trusted friend your PIN number over the phone without anyone eavesdropping — assuming no lip-readers are around.

6. Silent Sound Techniques is applied in Military for communicating secret/confidential matters to others.

CHAPTER IV

RESULTS

In this section, some of the initial results obtained in each phase are shown.



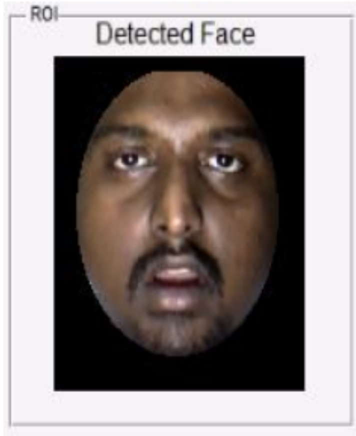
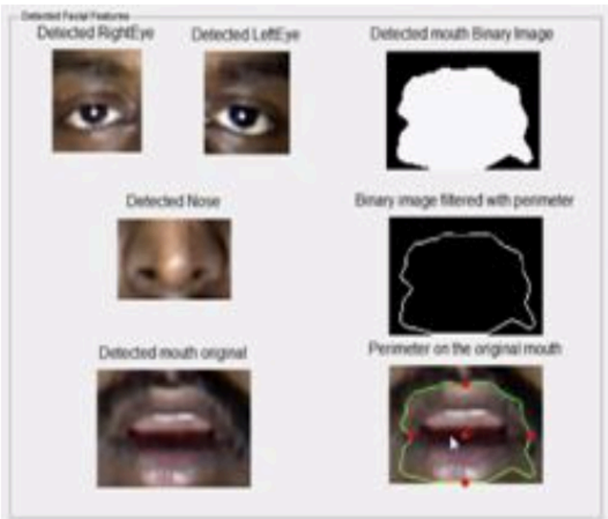


Fig 4.1: LIVE VIDEO CAPTURED BY WEBCAM.

Fig 4.2: REGION OF INTREST LIVE VIDEO FOR SILENT 'NIHAO'



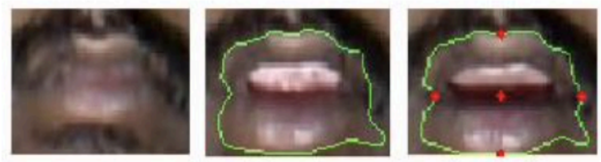


Fig 4.3: FACIAL FEATURES DETECTED LIVE VIDEO FOR SILENT 'NIHAO'.

Fig 4.4: LIP DURING MOTION WITH PERIMETER CONTOUR AND KEY POINTS.

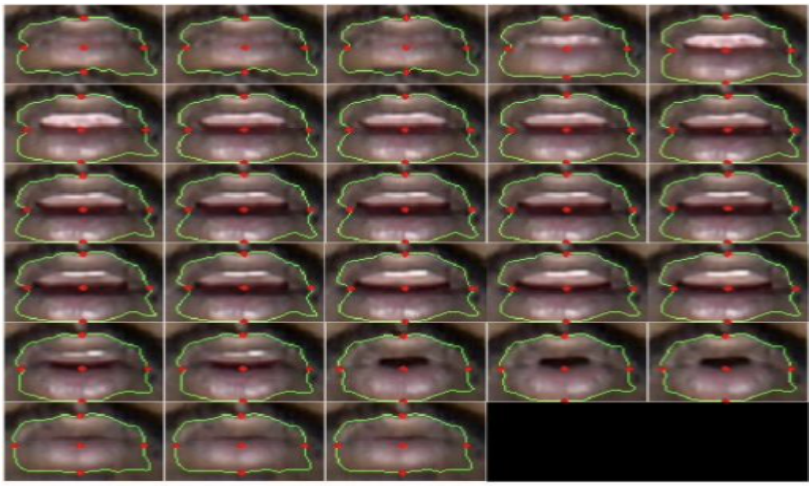


Fig 4.5: MULTI FRAME IMAGE(28 frames) MONTAGE IN A SINGLE OBJECT IMAGE MONTAGE FOR SILENT 'NIHAO'.

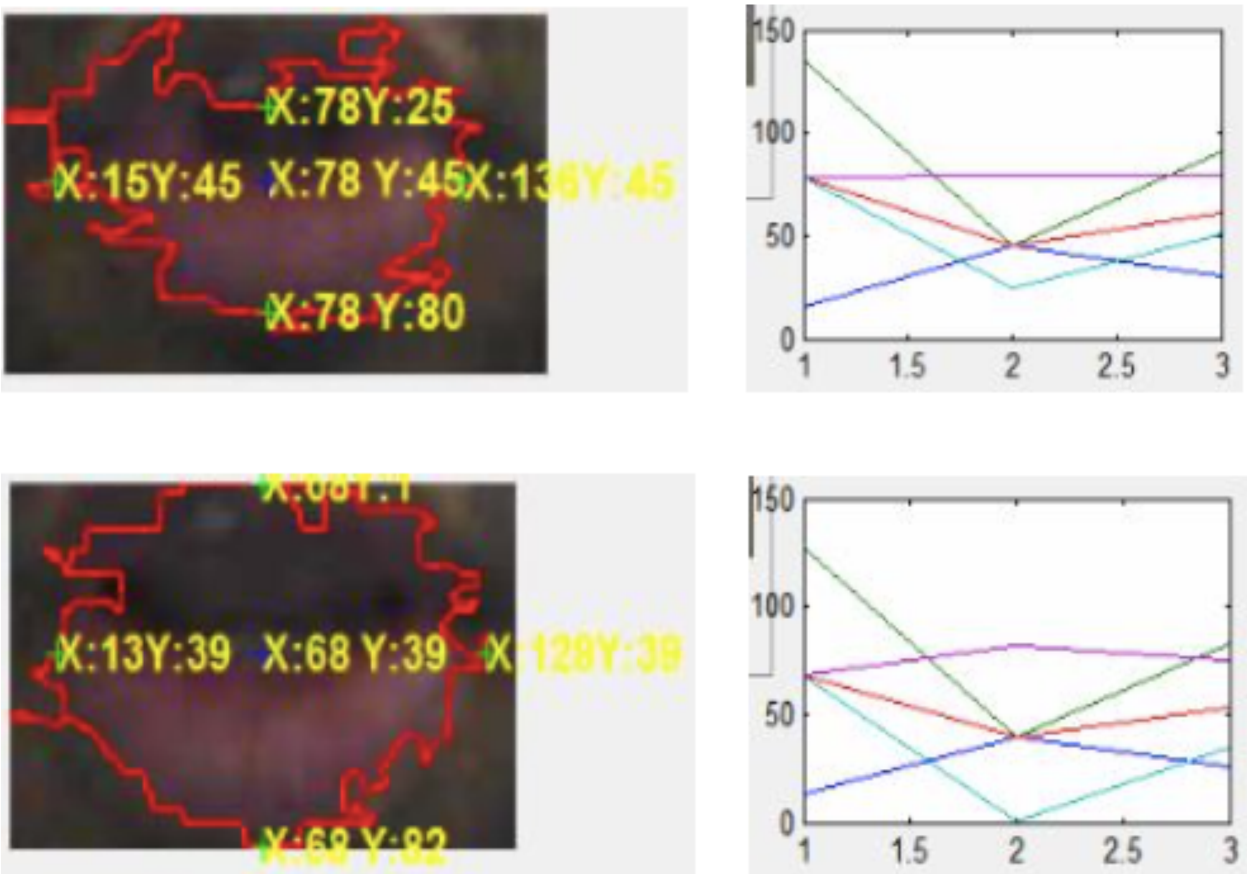


Fig 4.6: THRESHOLD ANALYSIS BASED ON COORINATES X,Y V/S TIME.

CHAPTER V
ADVANTAGES AND DISADVANTAGES OF SST

5.1 ADVANTAGES:

1. Helping people who have lost their voice due to illness or accident.
2. We can make silent calls even if we are standing in crowded place.
3. Very useful for sharing confidential information like secret PIN number on phone at public place.
4. This technology helps to transmit information without using vocal cords. Who are suffering from Aphasia (Speaking Disorder).
5. Allow the peoples to make silent calls without bothering others..
6. Very good technology for noise cancellation technique.

5.2 DISADVATAGES:

1. Translation into majority of languages but for languages such as Chinese different tone holds different meaning, facial movements being the same. Hence this technology is difficult to apply in such situations.
2. From security point of view recognizing who you are talking to gets complicated.
3. Even differentiating between people and emotions cannot be done. This means you will always feel you are talking to a robot.
4. This device presently needs nine leads to be attached to our face which is quite impractical to make it usable.

CHAPTER VI

APPLICATIONS OF SST

1. Silent Sound Technology is applied in military for communicating secret confidential matters to others.
2. As we know in Space there is no media for sound to travel therefore this technology can be best utilized by astronauts.
3. Helping people who have lost their voice due to illness or accident. Hence this technology can be best utilized by dumb peoples.
4. Since the electrical signals are universal they can be translated into any language. Native speaker can translate it before sending it to other side. Hence it can be converted into any languages of choice currently being German, English & French.
5. To tell a secret PIN no. , or credit card no. on the phone now be easy as there is no one eavesdrop anymore.

CHAPTER VII

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

Pronunciation of a silent word varies from person to person especially this difference makes the lip-read for Chinese language mandarin highly personalized. The software is being trained based on the lip structure, complexion and features of the lip area. The maximum allowable difference in the threshold value depends on environment and lighting factor. Minimizing the threshold values further enhances the level of security. Inter-disciplinary applications of this lip-reading technique helps for communication where language disorder raises, providing easier mode of communication for people with speech disabilities by converting the identified lip movements directly to speech. Emotional expressions can be correlated and synchronized to gain accuracy and adaptable dynamic. Since many of the systems are still preliminary, it would not make sense to justify the system comparing with speech recognition score or synthesis quality at this stage and hence initial results obtained are shown in the previous section.

Silent sound Technology is one of the recent invention in the field of information technology. Engineers claim that the device is working with 99 percent efficiency. Silent Sound technology aims to notice every movements of the lip and converted them into sounds, which could help people who lose voices to speak, and allow people to make silent calls without disturbing others. Rather than making any sounds, mobile would decipher the movements of mouth by measuring muscle activity and then convert this into speech that the person on the other end of the call can hear. So basically it reads lips. It will be one of the innovation and useful technology and in near future this technology will be used in day to day life.

7.2 FUTURE SCOPE

In future it is possible that electrodes will be incorporated into mobile phones without having electrodes hanging all around face. Nano technology will be a historical step towards making the device handle. When this software integrated onto mobile oriented or hand-held devices and public system machines, silent sound technology could prove to be much more secure, convenient and user-friendly technique for user authentication. Engineers claim that devices work on SST is not still accurate, so this would be an attempt to improve the accuracy.

1. Silent sound technology gives way to a bright future to speech recognition technology from simple voice commands to memorandum dictated over the phone all this is fairly possible in noisy public places.
2. Without having electrodes hanging all around your face, these electrodes will be incorporated into cell phones.
3. It may have features like lip reading based on image recognition & processing rather than electromyography.
4. Nano technology will be a mentionable step towards making the device handy.

With all of the millions of phones in circulation, there is great potential for increasing earnings by saving 'lost calls' - telephone calls that go unanswered or uninitiated because the user is in a situation in which he or she cannot speak – not just in business meetings, but everyday situations. According to research, these 'lost calls' are worth $20 billion per year worldwide. For the cellular operator, these are potential earnings that are currently being left on the table.

REFERENCES

1. https://www.ijirae.com/images/downloads/vol1issue1/MEC10078.March14.18.pdf
2. https://www.slideshare.net/lohitdalal1008jain/silent-sound-technology-final-report