Voice to Sign Language Translation using Machine Learning Techniques

# Project Abstract

Sign language is a visual language that is used by deaf people as their mother tongue. Unlike acoustically conveyed sound patterns, sign language uses body language and manual communication to fluidly convey the thoughts of a person.

Due to considerable time required in learning the Sign Language, people find it difficult to communicate with these specially abled people, creating a communication gap.

This Audio to Sign Language converter is important and significant because :

1. To provide information access and services to deaf people in Indian sign language.
2. To develop a scalable project which can be extended to capture whole vocabulary of ISL through manual and non-manual signs.

It can be developed as a desktop or mobile application to enable specially abled people to communicate easily and effectively with others.

# OVERVIEW OF THE PROJECT

## Proposed Solution

Sign language is a visual language that is used by deaf people as their mother tongue. Unlike acoustically conveyed sound patterns, sign language uses body language and manual communication to fluidly convey the thoughts of a person.

Due to considerable time required in learning the Sign Language, people find it difficult to communicate with these specially abled people, creating a communication gap.

Thus, we propose an application which takes in live speech or audio recording as input, converts it into text and displays the relevant Indian Sign Language images or GIFs.

Front-end using EasyGui.

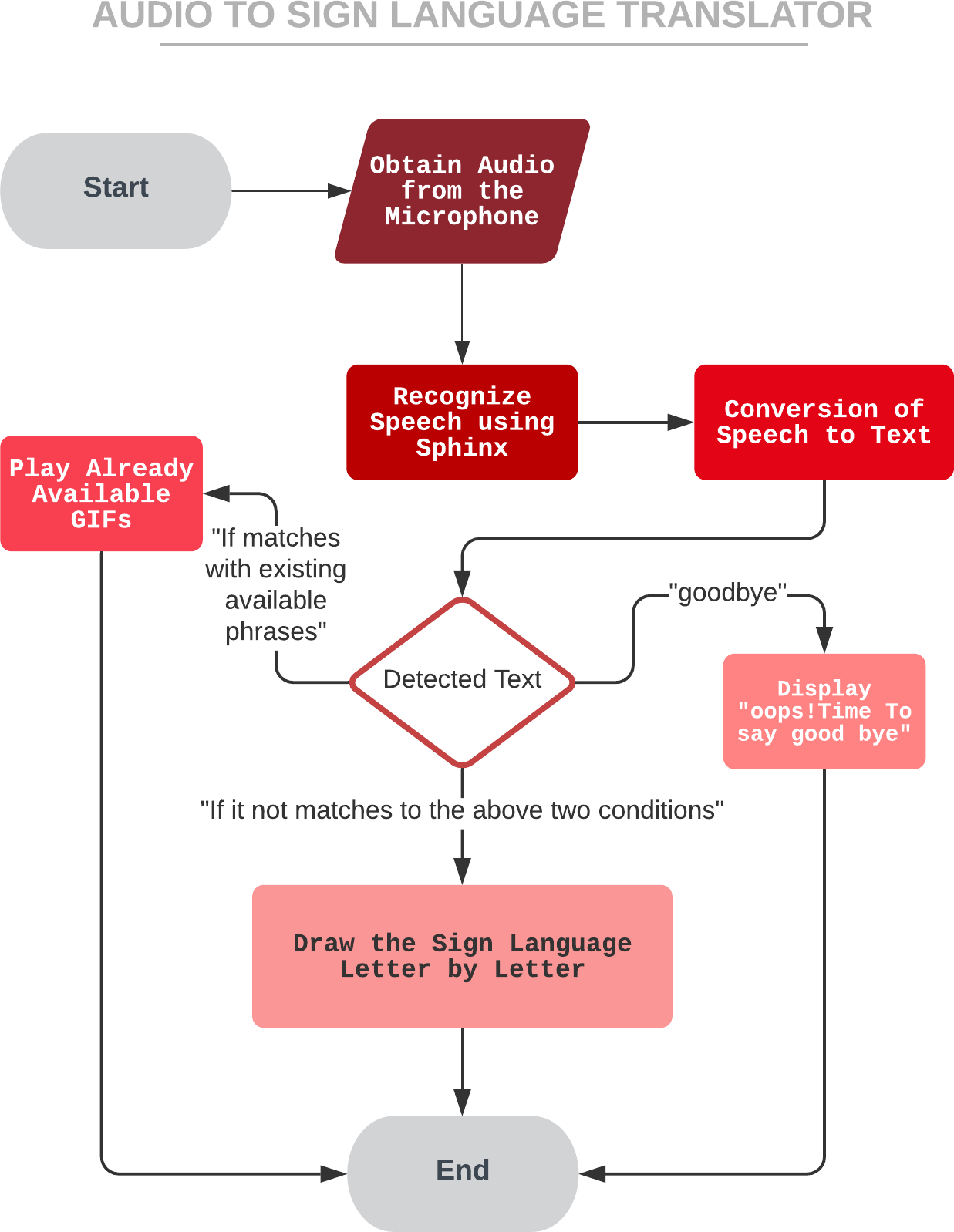
Speech as input through microphone using PyAudio. Text Pre-processing using NLP.

Dictionary based Machine Translation.

## Phases of the Project

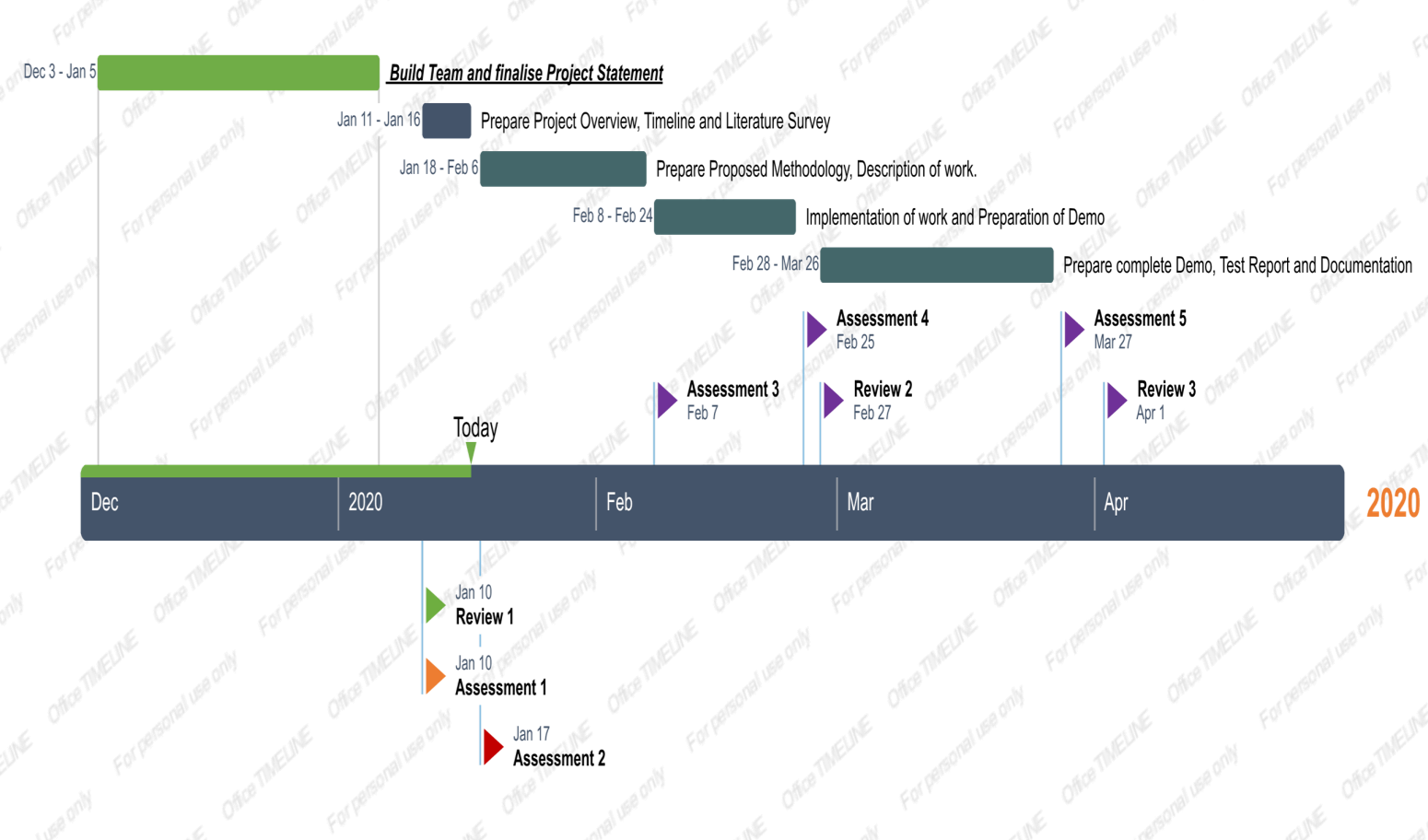
* + Obtain audio from the Microphone and perform Speech Recognition using sphinx.
  + Convert the speech recognized to text.
  + Display relevant images or GIFs as applicable.

# FLOWCHART



*Figure 1: Depicting the process of the Audio to Sign Language Translator*

# PROJECT PLAN



**LITERATURE REVIEW**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Ref. No.** | **Title of the paper** | **Methodology** | **Pros** | **Cons** |
| 1. | Real-Time Recognition of Indian Sign Language | The picture frames are resized to ensure a level playing field between all images. OpenCV (Computer Vision Open Source Library) is used for extraction and video classification of the apps | This FCM-based real-time sign language recognition system has achieved 75% accuracy in gesture labeling to identify the words of Indian Sign Language.The system developed is much better than other systems as it is capable of recognizing 40 words of ISL in real time. | In future work, more words are needed to be added to the system |
| 2. | Deaf Talk Using 3D Animated Sign Language - A Sign Language Interpreter using Microsoft’s Kinect v2 | The program was split into two different parts, the module speech to sign and the module sign to voice. Design takes place in controlled environments, keeping track of all the factors and possibilities in actual scenarios.  Achieved using the Continuous Gesture Builder from Microsoft Kinect V2 who is responsible for the identification and  training of gestures. | The sensitivity of the sign to speech module is 84 per cent. | The precision of English words in Asian accent is a bit difficult for device to understand. |
| 3. | Research on text sentiment  analysis based on CNNs and SVM | Convolutional Neural Networks (CNNs) model combined with SVM text sentiment analysis is used. | The text effective analysis model proposed in this paper based on CNNs and SVM will effectively improve the performance of  classifying text. | classification capability of fully connected classification layer is weak for nonlinear separable data. |
| 4. | Bengali Sign Language to Text Conversion using Artificial Neural  Network and  Support Vector Machine | Uses an optimum system comprising of artificial neural networks and support vector machine  (SVM). Microsoft Kinect is used to take the input. | The three basic features extracted from the image were contour using convex hull process, fingertip and our self-developed joint position to achieve a good and better result. | It can be observed that the accuracy would be nearly 100 if the tests were carried out on trained individuals. In fact, many of the hand signs for the different letter are very similar, often leading to confusion and thus incorrect tests of recognition. |
| 5. | Double Handed Indian Sign Language to Speech and Text | the double handed Indian Sign Language is captured as a series of images and it's processed with the | The experiment was carried out with bare hands, thus limitations by using data gloves is overcome. | Min Eigen value theorem is difficult to implement in matlab. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | help of MATLAB and then it's converted to speech and text. |  |  |
| 6. | Machine Learning Techniques for  Indian Sign Language Recognition | Indian sign language recognition using machine learning algorithm with thehelp of MATLAB. | For K-NN techniques with K=1, achieved 100% recognition rate whereas using back propagation technique achieved 94-96%. | The conjugate gradient approach is often used as an iterative algorithm, applicable to sparse structures that are too large for direct implementation or other similar methods, such as the Chelsey decomposition. |
| 7. | Semantic Analyses of Text to Translate to Russian Sign Language | Zardoz System, TEAM System and ViSiCAST System is being used. | It is proved that accounting semantic part in the translation process is a great benefit of having this property in the system | Failure to understand the semantic aspect of both the sound and sign languages |
| 8. | K-Nearest Correlated Neighbor Classification for  Indian Sign Language Gesture Recognition using Feature Fusion | The test image is first categorized into a single or double hand gesture followed by a fusion of SIFT and HOG descriptors  through K-Nearest Correlated Neighbours. | gives substantial accuracy for recognition of this gestures | At present, the technique is limited to detecting static movements. The creation of a model that also understands complex gestures would be an addition. |
| 9. | The Text Analysis and Processing of Thai Language Text to Speech  Conversion System | system is divided into front-end and back-end analysis module  speech synthesis module | accuracy rate is 97.83%, the outside set’s test accuracy rate is 97.61%. | Thai text normalization difficulties expressed mainly for two indexes: the first is the judgment of non-standard word recognition; the second is the ambiguity of the disambiguation carried out by non-standard word  processing. |
| 10. | Data Analysis Support by  Combining Data  Mining and Text Minin | A mining framework that can treat both numerical and textdata is proposed. | We can iterate data shrink and data analysis with both numerical and text analysis tools in the unique framework. | two systems are required because mining systems alone cannot treat both numerical and text data. |
| 11. | Moment Based Sign Language Recognition For Indian Languages | Two classification techniques (PNN & KNN) are used & performance is compared between both classifiers. | Using 7Hu moment techniques for feature extraction &  KNN classifier, Appx. 82% accuracy is achieved. Quality of  English Speech is very good because of inbuilt MATLAB function. | Can be used to translate and apply on the smartphone platform, due to its easy use, mobility and tremendous growth. |
| 12. | Bangladeshi Hand Sign Language Recognition from Video | Linear  Binary Pattern (LBP). For classifying the  hand signs SVM is used | In this method, signs that were created by one hand or two connected hands were recognized successfully. | System has faced troubles to detect multiple hands efficiently  if they are not connected. |
| 13. | A Cost Effective Design and Implementation of Arduino Based Sign | freeware software for design of the application and the design of the system | The use of Bluetooth as communication protocol has given a great enhancement in | Protocol that can be used is Wi-Fi or GSM for long distant  communication. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Language Interpreter | code in Arduino platform, and the former in MIT application inventor platform. | the design compared to the other wireless  communication  protocol like ZigBee or RF link. |  |
| 14. | Sign Language Recognition Using Modified Convolutional Neural Network Model | top-tier models in Action Recognition which is i3d inception this model is also a new Action Recognition model with very high accuracy | Good training accuracy | Very low validation accuracy.  Can do a lotmore things with this model, like freeze the layers, remove some inception module, remove the transfer learning, and change the fully connected  layer into another deep learning model. |
| 15. | Overview and Analysis of Existing Decisions of  Determining the Meaning of Text Documents | the concepts of semantic analysis were studied,  as well as its main components, which include a semantic model, keywords, stop words, water, and others | the importance of semantic analysis was  determined.  the semantic analysis has a high practical application, the definition of the meaning of text documents. | Work can be done on the analysis of texts for the presence in them of text borrowings and borrowing ideas, as well as in determining the authorship of the text,  taking into account its paraphrasing. |
| 16. | The Comparison of Some Hidden Markov Models for Sign Language Recognition | Gaussian Hidden  Markov Model (Gaussian HMM) Multinomial Hidden Markov Model (Multinomial HMM) Using edge detection and skin detection using the help of Contrast Adaptive Histogram Equalization  (CLAHE) for image enhancement. | for edge detection preprocessing Gaussian Hidden Markov Model got better accuracy than Multinomial Hidden Markov Model.  In skin detection preprocessing Multinomial Hidden Markov Model got better accuracy than Gaussian Hidden Markov Model. | The biggest flaw is in dataset, where the signers are using gloves, but this research is using skin recognition, making the features aren’t accurate. |
| 17. | A Unified  Framework for Tracking Based Text Detection and Recognition from Web Videos | generic Bayesian- based framework of Tracking based Text Detection And Recognition (T2DAR) text tracking,  tracking based text detection, and tracking based text  recognition. | A variety  of experimental results show that their framework has an impressive performance. | text tracking in complex videos (e.g., web videos) is still a challenging topic, and post-processing techniques of text tracking should be further investigated. |

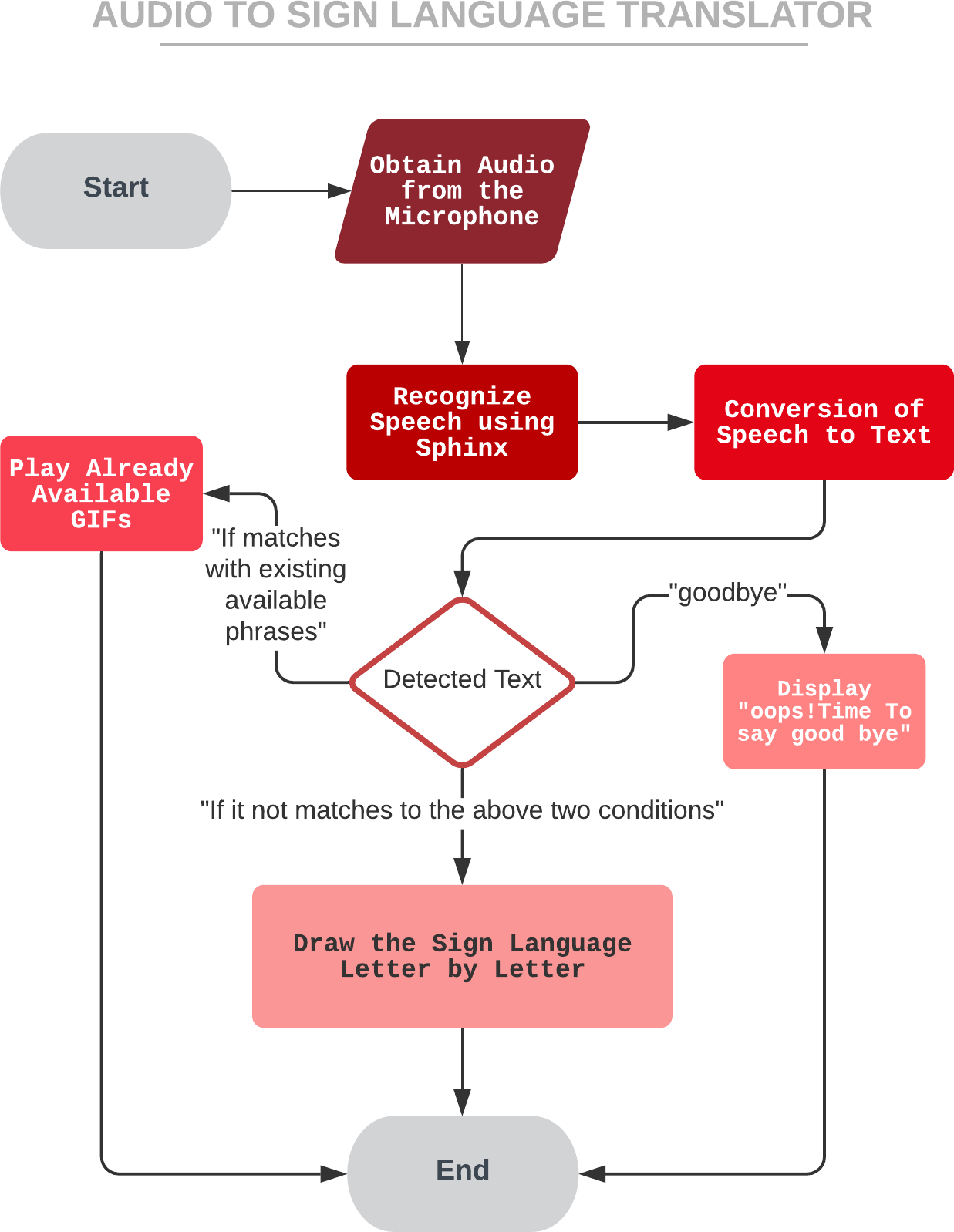
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | | | | |
|  | 18. | A Translator for American Sign Language to Text and Speech | AdaBoost and Haar- like classifiers  This paper proposed a system  that can automatically detect static hand signs of alphabets in American Sign Language (ASL). | The overall success rate is 98.7%.  The expected result of this system is able to correctly recognizing hand signs, converting these recognized signs  into letters, after that combining them into words, sentences  in textual format and then translating to audio format. | Lot of scope is there in this field. A Universal Language Translator can be made with this much efficiency. |
| 19. | Software Based Sign Language Converter | The cue symbols (ASL Gestures) to  speech using graphical programming (LabVIEW) | Reducing the  limitation of the existing system, thereby eliminating the need  for a translator.  This software-based approach  allows real-time translation with easy-to-use controlsfor the  user. | There are several Hardware and Software constraints which  decreases the  portability of the program/research. |
| 20. | Sign Language Spotting with a Threshold Model Based on  Conditional Random Fields. | Conditional random field (CRF) model | The proposed threshold model  with CRF is an excellent mechanism for distinguishing invocabulary signs and nonsign patterns. | Extending the proposed threshold model with CRFs to HCRFs and improving the sign language spotting performance for the  utterance data set. |
|  | | | | | |

**Proposed Methodology**

Sign language is a visual language that is used by deaf people as their mother tongue. Unlike acoustically conveyed sound patterns, sign language uses body language and manual communication to fluidly convey the thoughts of a person.

Due to considerable time required in learning the Sign Language, people find it difficult to communicate with these especially abled people, creating a communication gap.

Thus, we propose an application which takes in live speech or audio recording as input, converts it into text and displays the relevant Indian Sign Language images or GIFs.



*Figure 1: Depicting the process of the Audio to Sign Language Translator*

## Audio to Sign Language Translator

1. Start
2. Getting the Speech.
   1. Listen for 1 second and calibrate the energy threshold for ambient noise levels.
   2. Listen the Speech using Microphone.

Now the energy threshold is already set to a good value, and we can reliably catch speech right away.

1. Recognise the Speech.
2. Convert Speech to Text.
   1. Make the Text to lowercase for further manipulation.
3. Detected Text
   1. If “goodbye” then exit.
   2. Else if Detected Text in predefined Dictionary Words. Display respective GIFs of the Phrase.
   3. Else Count the Letters of the Word/Phrase.
      1. Display the Visual of the phrase with some delay of Actions.
   4. Continue all the steps from Step 3, and continue till the Speech Ends.
4. If Error in Step 2, That is if no Speech Detected then display error message “Could not listen”.

**Phases of the Project**

* + - Obtain audio from the Microphone and perform Speech Recognition using sphinx.
    - Convert the speech recognized to text.
    - Display relevant images or GIFs as applicable

# NOVELTY / INNOVATION

* Around 466 million people worldwide have disabling hearing loss (1), and 34 million of these are children.
* It is estimated that by 2050 over 900 million people will have disabling hearing loss.
* Hearing loss may result from genetic causes, complications at birth, certain infectious diseases, chronic ear infections, the use of particular drugs, exposure to excessive noise, and ageing.

- Source: World Health

Organization

As per the data from WHO it is estimated 900 million people will have Hearing Loss disability, after all this there is not much available technology to make communication with deaf people. We aim to provide a technology to ease the communication.

Generally, to have communication people need to either know the sign language or they need the human translator to make the communication possible.

To provide the solution we will make a software that takes an input in the form of speech and displays the respective Sign Language for the phrase.

**IMPLEMENTATION / DEMO**

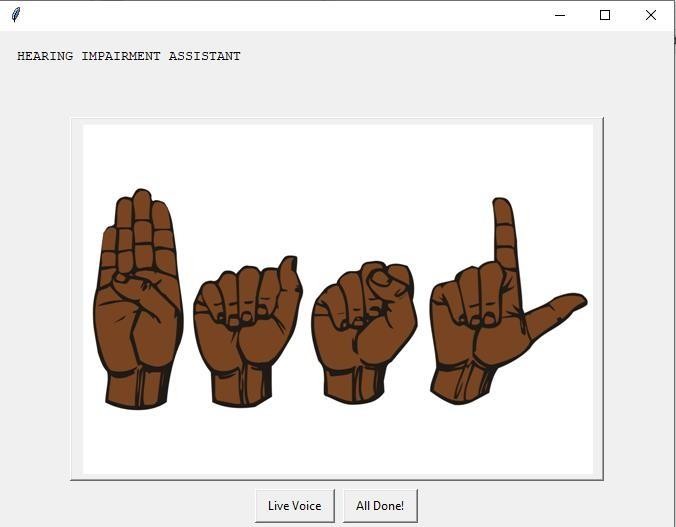
Sign language is a visual language that is used by deaf people as their mother tongue. Unlike acoustically conveyed sound patterns, sign language uses body language and manual communication to fluidly convey the thoughts of a person.

Due to considerable time required in learning the Sign Language, people find it difficult to communicate with these especially abled people, creating a communication gap.

Thus, we propose an application which takes in live speech or audio recording as input, converts it into text and displays the relevant Indian Sign Language images or GIFs.

# Use Cases

* Interface



This is the opening interface of the Hearing Impairment Assistant which gives us option to start the Application using two options:

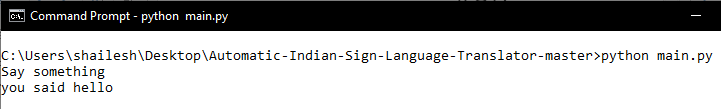
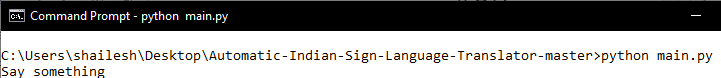
* + Live Voice

This allows user to start the interaction and get the resulted Indian Sign Language Visual from the given Speech.

* + All Done!

This option allows the user to exit the Program.

* Detecting Speech: Speech Recognition



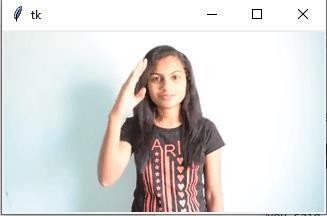
The application will ask to say something, then it will turn on the micro phone of the system to listen to the voice. The program will listen for a specific amount of time if no speech is detected then it will ask again to say something.

Here in this scenario we spoke the word *“hello”* and program recognized the word and thus converted it into text.

* Speech to Text Conversation / Translation

The application in the previous scenario listened to speech and detected the word *“hello”*. Speech recognized is then converted to text for further manipulation.

* Manipulation Text to Derive the Required Indian Sign Language.



In the previous scenario we saw the program successfully detected the word *“hello”* now the application matches the word initially with a database of GIFs which show the animation of the word or a phrase. But if the Phrase or word is not present in the database then Application will show the detected phrase in Indian Sign Language Visual letter by letter.