# AI Based Virtual Assistant System for Deaf and Blind People

Saiarya Katkar[1], Chirag Dabhere[1], Anushka Alekari[1], Dr. R. P. Patil[2]

*[1,2] Department of Electronics and Telecommunication Engineering, Sinhgad College of Engineering, Vadgaon Pune, India*

***Abstract:*** *The major means of natural communication for the deaf is sign language. Sign Language (SL) is a subset of gestures or signs made of donuts, hands, arms, eyes, heads, faces, etc. Someone who relies entirely on these gesture SLs for communication will have communication problems if they try to talk to someone who doesn't understand SL. Each country has its own advanced SL. In India, this language is called "Indian Sign Language (ISL)". Our plan seeks to create an algorithm that reformulates the ISL into English. Using the ISL, this approach interprets hand motions into English. Reformulated gestures include shapes, foundations, and many expressions. The algorithm first performs data access, also performs gesture pre-processing to track hand movements using a combinatorial algorithm, and performs recognition using template matching. The research presented in this document also describes a simple and effective system for speech recognition. Speech is converted into corresponding text to generate figurative text and vice versa.*

***Keywords–*** *Classification, Conversion of text to speech, Feature Extraction, Pre-processing, Speech Recognition.*

## 1. Introduction

An intuitive, hands-free machine learning program that enables communication between people with special needs and average people without the use of translators. The applications that are now on the market solely assist with text translation from sign language to English. But we solve the problem so that ordinary people can communicate with disabled people in any language. It also solves the problem of sign language subtitles for web movies and movies, helping viewers understand the context [15-18]. The deaf community uses sign language, a natural language, for communication. A subset of movements and signs performed with fingers, hands, arms, eyes, etc. is called Sign Language (SL). Data collection is the first step of the algorithm, followed by pre-processing to track hand movements using a combinatorial algorithm and detection using template matching. Sign Language (SL) is a visual-spatial language [19-20].

The system is inaudible, making it difficult for people who are mute or deaf to communicate in the workplace. Also, traveling alone can be dangerous as you cannot hear cars, bicycles or other people approaching [22]. They have trouble communicating with themselves and are unable to respond to other normal people or adapt quickly to their surroundings. A visual language or method of communication known as sign language has a recorded history in the West dating back to the 17th century. Sign language is made up of conventional motions, imitation, numerical spelling, sign language, and hand postures to represent the alphabet. Using traditional gestures, mimicry, sign language, number spelling, and hand positions to represent the alphabet all make up sign language. A character may represent a concept or an entire statement [21]. The main goal is to intelligently provide speech and text output to deaf people via hand gesture sign language without using sensors.

## 2. Literature Survey

Review of the Systematic Literature: American Sign Language Interpreters Mario Halim, Andra Ardiansyah, Brandon Hitoyoshi, and Novita Hanafiah Although sign language recognition (SLR) is a somewhat popular subject of study, very few people actually use it in daily life because of the complexity and variety of resources needed [1]. ASL symbol T. Utheim, J. Havskov, M. Ozyazicioglu, J. Rodriguez, and E. Talavera (2014) integrated SEISAN: Between 2015 and 2020, the authors found 22 different study papers. All of the study papers that were chosen for analysis produced impressive results, though not flawless ones because each thesis has its own advantages and disadvantages [2]. System for translating text between sign languages: Document Review 15 India's Jalandhar DAV College's Lalit Goyal Punjabi University, Patiala, India's Vishal Goyal Because the grammar rules of sign language are not standardised, translating text into sign language is challenging. There are a variety of methods that have been used to translate text into sign language, where the input is text and the output is a pre-recorded video or computer-generated character (Avatar). This article examines studies done on text-to-sign language translation technology [3]. Photographers Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black used expressive body shots to create 3D hands, faces, and bodies from a single photograph. We create 3D models of the human body, hands, and facial expression from a monocular image to 2 simplify the study of human movements, interactions, and emotions. To do this, we train a new, unified 3D model of the human body called SMPL-X, which augments SMPL with fully articulated hands and facial expressions using tens of thousands of 3D images. It is difficult to learn how to directly regress SMPL-X parameters from photos without concatenated photographs and 3D facts on the ground. So, using SMPLify's methodology, we first estimate the 2D features before optimising the model parameters toto fit the features [4]. A camera-assisted text-reading framework was put up by T. Rubesh Kumar and C. Purnima to assist blind persons in reading text labels and product packaging from hand-held objects. The following are the primary innovations included in this archetype system: 1) A new motion-based algorithm to help blind users target objects by briefly shaking the object of interest; 2) A new automatic text localization algorithm to extract text areas from multiple text patterns and complex backgrounds; and 3) A wearable camera-based support frame to assist blind people in reading text from handheld objects [5]. Pooja Sharmaetal Proposed Blindness is the absence of vision as a result of physiological or neurological issues. The suggested technique is a creative and cost-effective solution for blind and visually impaired persons in third-world nations since it is quick, simple, and inexpensive [6]. People with hearing and speech impairments should use sign language as their primary form of communication, according to Sujay R (BE Student, Faculty of ECE, KS Institute of Technology). Poor people find it challenging to interact with others on a daily basis. Our goal is to create a system that will make it easier to communicate with stupid and deaf individuals [7]. Aruna Rao B P (Assistant Professor, Faculty of ECE, KS Institute of Technology) stated nonverbal communication between deaf and dumb people is called sign language. Deaf and mute people face many challenges in their daily lives. Going to places with different native languages became even more difficult for them [8]. To assist blind persons in reading the text-on-text labels, printed notes, and products, Nagaraja L et al presented a camera-based assisted reading technique. The proposed project involves extracting text from an image and converting a text-to-speech converter, a process that allows text to be read by blind people [9]. Mallapa D.Guravetal proposes that this project introduces a smart device that helps visually impaired people read printed text on paper efficiently and effectively. The proposed project uses the method of a camera-based assistive device that can be used by anyone to read text documents [10]. Kristin K. Liu, Martha L. Thurlow, Anastasia M. Press and Michael J. Dosedel This literature review describes research conducted between 2008 and 2018 informing the field about the use of STT tools by K-12 and postsecondary students with disabilities [11]. Nisha P1, Dr. J. Vijayakumar There is an image everywhere around us and we see images and read texts in our daily lives. In this study, the identification process was performed using OCR, Raspberry Pi, MAT lab and openCV library. The function, setup, and test outcomes of the gadget are described in this paper. This project involves taking photos, translating the text, and turning the text into audio [12]. Ayushi Trivedi, Navya Pant, Pinal Shah, Simran Sonik, and Supriya Agrawal are members of the NMIMS University's faculty of computer science in Mumbai, India. The majority of these programmes use features like audio and clear speech recognition, text-to-synthetic speech signals and text-to-speech conversion, language translation, and a variety of other services. In this review, we will observe different techniques and algorithms applied to achieve the mentioned features. [13]

## 3. Proposed Methodology

Artificial Intelligence based Virtual Assistant System for Deaf and Blind People system is modeled as shown in Figure 1.

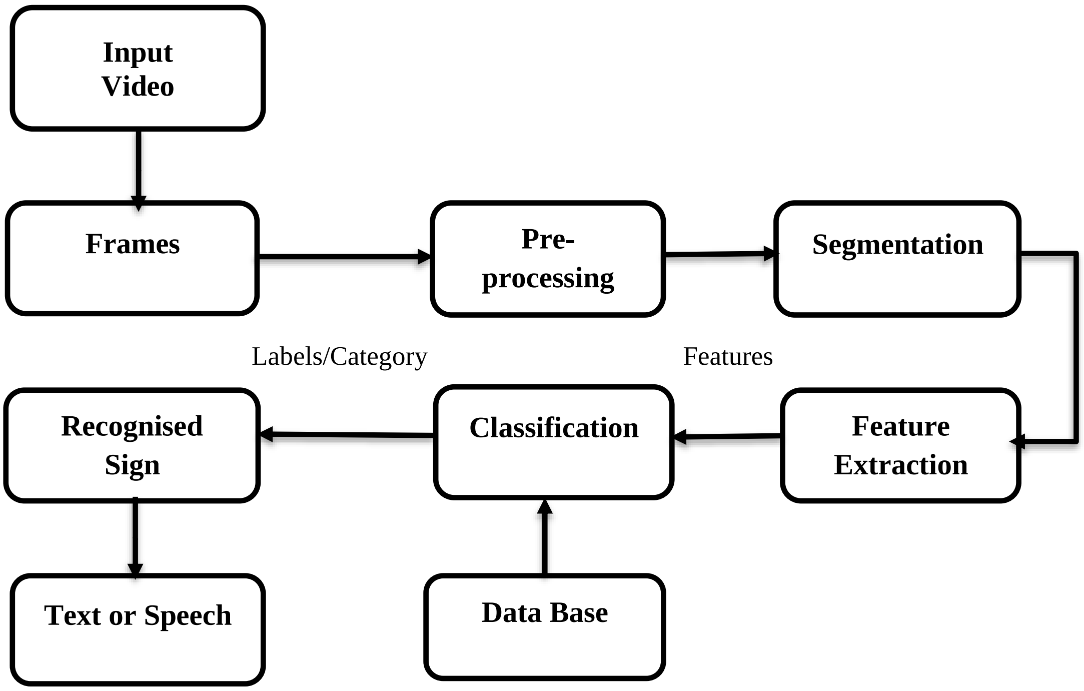


**Fig. 1 Proposed Model of Virtual Assistant System for Deaf and Blind People**

### 3.1 Sign extraction from video

**3.1.1 Preprocessing -** The image is prepared for feature detection and extraction in this stage. To maintain the uniformity of the proportions, the dimensions of all images are retained. In default option, captured video image is converted to HSV color space for image obtained with simple background [14]. Since the color of the skin tone is different from the background color, it is easily pulled out. Then, a test threshold is applied to the color calculation framework and filters out the flesh-colored pixels in the image. Furthermore, the image is binary, the blur is performed to remove noise, and the maximum contour is obtained from the result assuming that the contour with the largest area represents the hand. Errors are further eliminated by applying median filter and morphological operations.

**3.1.2 Segmentation -** Image segmentation is an important step in machine vision. The machine must manipulate the visual data segments for segment-specific processing to take place.

**3.1.3 Feature Extraction -** This stage entails establishing a codebook for the model, feature extraction, feature clustering, and graph construction, as well as creating a Bag of Visual Words (BOVW). The definition of the widely used Bag of Visual Words (BOVW) image classification model was taken from data retrieval and NLP (Natural Language Processing) Bag of Words (BOW). In it, we count the number of times each word occurs in a text, use the frequency of each word to get the keyword and generate a histogram of the word frequency. This idea is modified in such a way that instead of words, we use features of images as words. To build a vocabulary in which each image is represented as a histogram of acquired features, image descriptors and key-points are used. Then the category of another comparable image can be predicted from this histogram.

**3.1.4 Classification -** To classify input signals into distinct classes, a classifier in sign language recognition is required. During the training phase, the feature vector obtained from the training database is used to train the classifier. When the test input is presented, the trained classifier recognizes the sign type and displays or plays the appropriate text or sound. We move on to the classification step once the feature detection and extraction is complete. Support vector machines (SVMs) and convolutional neural networks(CNNs) are used for classification in this process. Due of its accuracy, convolutional neural networks (CNN) are the most widely used technique. The CNN determines which node this data belongs to and continues to learn from the training results. The more layers, the higher the accuracy, but it also becomes computationally expensive. A popular technique for addressing problems involving profit maximisation is Support Vector Machine (SVM). The solution of generating a decision boundary will be of great help in generalizing linear classifiers.

**2.2 Conversion of Speech to Text**

**3.2.1 Speech Recognition -** The ability of a machine or program to recognize words and phrases in spoken language and translate them into a machine-readable format is known as speech 1 recognition. The following criteria can be used to classify speech recognition systems:

**3.2.2 Loudspeaker -** All speakers have a different type of voice. Therefore, models are either designed for a particular take holder or for an independent stakeholder.

**3.2.3 Sound -** Speech recognition is also influenced by the speaker's speaking style. Some models are capable of identifying single utterances or several utterances interspersed with pauses.

**3.2.4 Vocabulary -** The size of the vocabulary has a significant impact on the system's complexity, performance, and accuracy.

**2.3 Conversion of Text to Speech**

**3.3.1 Text-to-speech** - Text-to-speech is a process that converts input text into digital audio, which is then spoken. First, the text is parsed, then it is processed and understood, and last it is turned into digital audio again.

**3.3.2 Word Processing -** The input text is parsed, normalized (deals with abbreviations and acronyms, and matches text) and transcribed into a phonetic or linguistic representation.

**3.3.3 Voice synthesis -** Some of the speech synthesis techniques are articulation synthesis, format synthesis, concatenation synthesis, etc.

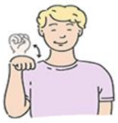
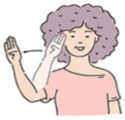
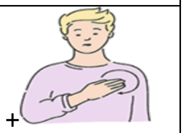
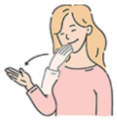
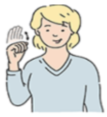
**2.4 Tools and Libraries**

**3.4.1 Tensor flow -** A free and open-source software library for machine learning and artificial intelligence is called TensorFlow. Although it can be used for many different things, deep neural network training and inference are where it really shines. Four high-level APIs for TensorFlow are based upon low-level APIs in a hierarchical structure. To develop and uncover new machine learning algorithms, researchers use low-level APIs. You will define, train, and make predictions using machine learning models in this course using the high-level API tf.keras. A TensorFlow alternative that is open source is tf.keras.

**3.4.2 Pyttsx3 -** A Python text-to-speech library is called Pyttsx3. It works offline and is compatible with Python 2 and 3 in contrast to competing libraries. The factory method pyttsx3.init() is called by an application to obtain a reference to the pyttsx3 file. It is a very user-friendly tool to turn typed text into speech. Two voices are supported by the pyttsx3 module, a male voice and a female voice, both powered by "sapi5" for Windows.

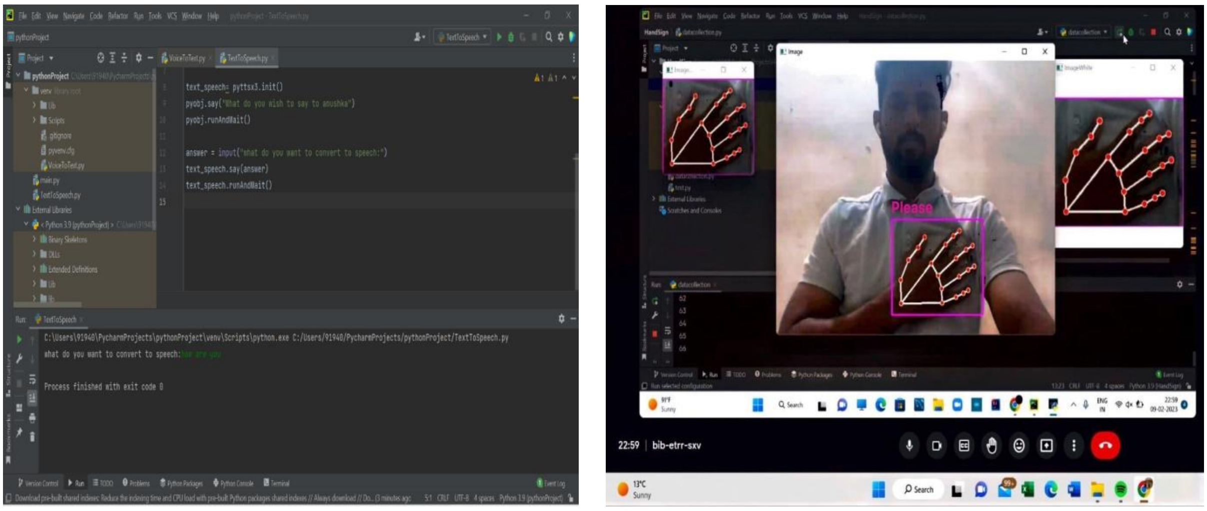
**3.4.3 PyAudio -** A cross-platform audio I/O library called PortAudio v19 has Python bindings available from PyAudio. Python may be used to play and record audio on a number of systems, including GNU/Linux, Microsoft Windows, and Apple macOS, with the help of PyAudio. The MIT License governs PyAudio's release.

**Table 1: Sign and Actions covered in proposed simulation**

| Sign | Action | Sign | Action | Sign | Action |
|---|---|---|---|---|---|
| Yes |  | Hello |  | Please |  |
| Thank You |  | Good Bye |  | | |

## 4. Results

Input provided to the system will include video input, image or voice note, this input language can be in English or sign two words at the same time. Depending on the need, give the output in the desired language or output type, whether audio or text. We will be able to collect up to ten words at a time to form a text message and read it. The end result of the project is an interface that allows people with disabilities to especially easily communicate with people with disabilities. Figure 2 shows the simulation results of proposed model.



```
text_speech= pyttsx3.init()
pyobj.say("What do you wish to say to anushka")
pyobj.runAndWait()

answer = input("what do you want to convert to speech:")
text_speech.say(answer)
text_speech.runAndWait()
```

```
C:\Users\91940\PycharmProjects\pythonProject\venv\Scripts\python.exe C:/Users/91940/PycharmProjects/pythonProject/TextToSpeech.py
what do you want to convert to speech:how are you

Process finished with exit code 0
```

**Fig. 2 Simulation of proposed model for "Please" word**

## 5. Conclusion

An easy-to-use hands-free machine learning program where the communication gap between people with special disabilities and normal people can be bridged without the need for a third-person interpreter. Currently available programmes only assist people in translating sign language into too-English text. But we solve the problem that even if a normal person wants to talk to a person with a disability in any language, it is still possible. Even the issue of subtitles in sign language helping viewers grasp the content of internet videos and movies is resolved. Sign language is the natural language that the deaf community uses to communicate. A subset of hand, arm, finger, eye, and other body motions are used to create signs in sign language (SL).

## REFERENCES

[1] Systematic Literature Review: American Sign Language Translator by Andra Ardiansyah, Brandon Hitoyoshi, Mario Halim Novita Hanafiah, Aswin Wibisurya

[2] ASL sign Integrated with SEISAN by T. Utheim, J. Havskov, M. Ozyazicioglu, J. Rodriguez, and E. Talavera (2014)

[3] Text to Sign Language Translation System::A Review of Literature by Lalit Goyal and Vishal Goyal.

[4] Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A., Tzionas, D. and Black, M.J., 2019. Expressive body capture: 3d hands, face, and body from a single image. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10975-10985).

[5] T. Rubesh Kumar Speech to text is obviously essential in today's society. Printed text is everywhere in the form of reports, receipts, bank statements.

[6] Pooja Sharmaetal Proposed Blindness is a state of lacking the visual perception due to physiological or neurological factors.In this proposed work by Pooja Sharma, Mrs. Shimi S. L. and Dr. S. Chatterji, a simple, cheap, friendly user, virtual eye will be designed and implemented to improve the mobility of both blind and visually impaired people in a specific area.

[7] Sujay R (B.E. Student, ECE Dept., K S Institute of Technology) Sign language is a basic means of communication for those with hearing and vocal disabilities.

[8] Aruna Rao B P (Assistant Professor, ECE Dept., K S Institute of Technology). This non-verbal communication between deaf and mute people is called Sign Language.

[9] Nagaraja L etal proposed that the method is a camera based assistive text reading to help blind person in reading the text present on the text labels, printed notes and products.

[10] Mallapa D.Guravetal proposed thatthis project presents a smart device that assists the visually impaired which effectively and efficiently reads paper-printed text.

[11] Kristin K. Liu, Martha L. Thurlow, Anastasia M. Press, and Michael J. Dosedel This literature review describes what research conducted between 2008 and 2018 tells the field about the use of STT tools by K-12 and post-secondary students with disabilities.

[12] Nisha P1, Dr.J.Vijayakumar 1student of Mphil, Department of Electronics &instrumentation, Bharathiar University, Coimbatore. 2Associate Professor & Head, Department of Electronics & instrumentation, Bharathiar University, Coimbatore.

[13] Ayushi Trivedi,Navya Pant, Pinal Shah,Simran Sonik and Supriya Agrawal DepartmentofComputerScience, NMIMS University,Mumbai,India. Corresponding Author:Navya Pant In present industry, communication is the key element to progress.

[14] Rohita Jagdale and Sanjeevani Shah, "Video resolution enhancement and quality assessment strategy", 3rd International Conference on Computing Communication Control and Automation (ICCUBEA-IEEE) 2017.

[15] Rohita Jagdale and Sanjeevani Shah, "A Novel Algorithm for Video Super-Resolution", Information and Communication Technology for Intelligent Systems Proceedings of ICTIS (Springer) 2018, Volume 1.

[16] Rohita Jagdale and Sanjeevani Shah, "Super Resolution Reconstruction of Low Resolution Video using Sparse Representation", 5th International Conference on Computing Communication Control and Automation (ICCUBEA-IEEE) 2019.

[17] Rohita Jagdale and Sanjeevani Shah, "Modified Rider Optimization-based V Channel Magnification for Enhanced Video Super Resolution", Int. Journal of Image and Graphics, doi.org/10.1142/S0219467821500030, world scientific. Vol. No. 21. No. 01, January 2021.

[18] Rohita Jagdale and Sanjeevani Shah, "Search Based Optimization approach for Video Super Resolution" Test Engineering and Management, ISSN:0193-4120 Page No. 2599 – 2613, November-December 2019.

[19] Rohita Jagdale and Sanjeevani Shah, "Optimally Balance trade of Video Super resolution and SNR using Gaussian Mixture Model with up-scaling", Test Engineering and Management, ISSN:0193-4120 Page No. 2108 – 2116, March-April 2020.

[20] Rohita H. Jagdale, and Sanjeevani K. Shah, "Video Super Resolution and Performance Enhancement of Mixture Mapping Model by

Deep Learning De-Noising" Published in International Journal of Test Engineering and Management, Volume-83, pp. 2108 - 2116, 18 March 2020, ISSN: 0193-4120.

[21] Rohita H. Jagdale, and Sanjeevani K. Shah, "Search Based Optimization Approach for Video Super Resolution" Published in International Journal of Test Engineering and Management, Volume-81 , pp. 2599 - 2613, 12 December 2019, ISSN: 0193-4120.

[22] Rohita H. Jagdale, and Sanjeevani K. Shah, "V - Channel magnification enabled by hybrid optimization algorithm: Enhancement of video super resolution" Gene Expr Patterns. 2022 Sep.