# Real-Time Communication System Powered by AI for Specially Abled Literature Survey

**ABSTRACT:**

People affected by speech impairment can't communicate using hearing and speech, they rely on sign language for communication. Sign language is used among everybody who is speech impaired, but they find a hard time communicating with people which are non-signers (people aren't proficient in sign language). So, requirement of a sign language interpreter is a must for speech impaired people. This makes their informal and formal communication difficult. There has been favourable progress in the field of gesture recognition and motion recognition with current advancements in deep learning. The proposed system tries to do a real time translation of hand gestures into equivalent English text. This system takes hand gestures as input through video and translates it text which could be understood by a non-signer. There were similar researches done earlier with most of them focussing on just sign translation of English alphabets or just numbers. There will be use of CNN for classification of hand gestures. By deploying this system, the communication gap between signers and non-signers. This will make communication speech impaired people less cumbersome. There were many researches which helped us to establish the idea of Artificial Neural Networks for this project. Many models were available that detected only characters with an accuracy of around 86%. We also explored the Linear discriminant analysis (LDA) technique but didn't use it due its drawback to express complex data. The hardware implementation of the project was also read about. It has a cost and maintenance factor involved which we tried to eliminate in our project. We have achieved high accuracy of 96.5% in our model, with the feature of suggestions of words and formation of sentences, an idea which was not found in any of the researches.

**INTRODUCTION:**

Sign language is a predominant sign language Since the only disability D&M people have is communication related and they cannot use spoken languages hence the only way for them to communicate is through sign language. Communication is the process of exchange of thoughts and messages in various ways such as speech, signals, behaviour and visuals. Deaf and dumb(D&M) people make use of their hands to express different gestures to express their ideas with other people. Gestures are the nonverbally exchanged messages and these gestures are understood with vision. This nonverbal communication of deaf and dumb people is called sign language.

**Sign language is a visual language and consists of 3 major components:**

| **Finger spelling** | **Non manual features** | **Word level vocabulary** |
|---|---|---|
| Used to spell words letter by letter | Facial expression & tongue, mouth & body | Used for the majority of communication |

In our project we basically focus on producing a model which can recognise Finger spelling□based hand gestures in order to form a complete word by combining each.

**The gestures we aim to train are as given in the image below.**


**OBJECTIVE:**

**• To develop an application interface that interprets sign Language to Text in real time to assist deaf and dumb for communicating with them effectively eliminating the requirement of a translating individual.**

**• To devise a model to achieve the highest possible accuracy and least time consumption for prediction of symbols as compared to already existing models.**

**• To reduce the cost and develop an economical and user-friendly graphical user interface (GUI) application that requires minimal maintenance for conversion of sign to its corresponding text.**

**• To provide suggestions based on current word to eliminate the need of translating the full word, thereby improving accuracy and reducing time for sign to text conversion.**

**• To reduce the chances of spelling mistakes by suggesting correct spellings from English dictionary words close to the current word.**

**• To formulate characters, words and sentences with interpretation of symbols in Sign Language in a portable and real time application.**

**LITERATURE SURVEY:**

There are many researches on recognition of sign languages or finger spellings. We introspected various research papers that were available and came up with an interface that converts sign language to text, provides the feature of adding a word and suggestions based on the word being translated.In, they have proposed a system that recognizes dynamic hand gestures for English numbers (0-9) in real-time using Hidden Markov Model. HMM is highly dependent on the probability of the hidden states, hence there are more number of parameters to learn which is time consuming. The systemcontains two stages: Pre-processing for hand tracking and Classification to recognize gestures. Hidden Markov Model is used for the isolated and dynamic gesture recognition whose average recognition rates are 99.167% and 93.84% respectively. Hidden Markov Models (HMM) is used for the classification of the gestures. This model deals with dynamic aspects of gestures. Gestures are extracted from a sequence of video images by tracking the skin-colour blobs corresponding to the hand into a body– face space cantered on the face of the user. The goal is to recognize two classes of gestures: deictic and symbolic. The image is filtered using a fast look–up indexing table. After filtering, skin colour pixels are gathered into blobs. Pradumn Kumar and Upasana Dugal proposed an algorithm using TensorFlow based on Advanced Convolutional Neural Networks for identification of plants. This research paper motivated us to use Convolutional Neural Networks for identification of Sign Language symbols. Specially using CNN is a very trending procedure for Deep learning in computer point of view. ImageNet have produced a lot of expectation by giving exciting results. Here CNN takes the most challenging task for identification of plants by using their complete picture or any parts of that plants while others tackles one by one process like firstly, they take any specific organisms (flowers, leaves and bark etc.) then whole picture of organisms. In CNN there are some limitations like it is not better with very large sets of images or lack of explanatory power. So Advanced CNN will replace CNN because in Advanced CNN is small in size as compare to CNN for recognizing images. Here large models can be easily scale up and these models are small enough to train fast, by this we will get out new ideas and have a good chance for experiment on other methods also. The architecture of Advanced CNN is multi-layer consisting of alternate use of Convolution layers and nonlinearities. All these layers are followed by fully connected layers leading into a SoftMax classifier. This model gives a good accuracy results within few timeswhen we run on a GPU. presents a comprehensive review of deep learning and develops a categorization scheme to analyse the existing deep learning literature. It divides the deep learning algorithms into four categories according to the basic model they derived from: Convolutional Neural Networks, Restricted Boltzmann Machines, Autoencoder and Sparse Coding. The state-of-the-art approaches of the four classes are discussed and analysed in detail. For the applications in the computer vision domain, the paper mainly reports the advancements of CNN based schemes, as it is the most extensively utilized and most suitable for images. Most notably, some recent articles have reported inspiring advances showing that some CNN based algorithms have already exceeded the accuracy of human rates. Despite the promising results reported so far, there is significant room for further advances. For example, the underlying theoretical foundation does not yet explain under what conditions they will

perform well or outperform other approaches, and how to determine the optimal structure for a certain task. This paper describes these challenges and summarizes the new trends in designing and training deep neural networks, along with several directions that may be further explored in the future. Convolution neural network has long been used in the field of digital image processing and speech recognition, and has achieved great success. Before the convolutional neural network was proposed, both image processing and speech recognition were done by traditional machine learning algorithms. Although great results were achieved, it was difficult to make further breakthroughs, so CNN came into being. Currently, CNN for image processing and speech recognition are relatively mature. Both the theoretical research and the industrial application have been very successful, which has promoted CNN's leap-forward development. CNN's success of image processing and speech recognition has stimulated its research frenzy in natural language processing. The current CNN to handle natural language has been widely used, although some achievements have been made, the current effect is not very good. The purpose of is to give a clearer explanation of the structure of CNN. At the same time, give a brief summary and prospect of current CNN research in image processing, speech recognition and natural language processing. Results observed in the comparative study with other traditional methods suggest that CNN gives better accuracy and boosts the performance of the system due to unique features like shared weights and local connectivity.CNN is better than other deep learning methods in applications pertaining to computer vision and natural language processing because it mitigates most of the traditional problems. To reduce the problem of overfitting, we used the Dropout Technique as suggested in. Dropout is a technique for improving neural networks by reducing overfitting. Standard backpropagation learning builds up brittle co-adaptations that work for the training data but do not generalize to unseen data. Random dropout breaks up these coadaptation's by making the presence of any particular hidden unit unreliable. This technique was found to improve the performance of neural nets in a wide variety of application domains including object classification, digit recognition, speech recognition, document classification and analysis of computational biology data. This suggests that dropout is a general technique and is not specific to any domain. Dropout considerably improved the performance of standard neural nets on other data sets as well. This idea can be extended to Restricted Boltzmann Machines and other graphical models. The central idea of dropout is to take a large model that overfits easily and repeatedly sample and train smaller sub-models from it.Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur and Dr. Dayananda P in developed a finger spelling sign language translator is obtained which has an accuracy of 95%. They created a desktop application that uses a computer's webcam to capture a person signing gestures for ASL, and translate it into corresponding text and speech in real time. The translated sign language gesture will be acquired in text which is farther converted into audio. In this manner they are implementing a finger spelling sign language translator. To enable the detection of gestures, they used Convolutional neural network (CNN). This research paper provided us insight about the base model for our project.In the following paper, Support Vector Machine (SVM) was used as the machine learning method. They proposed a recognition method of fingerspelling in Japanese sign language, which uses classification tree based on pattern recognition and machine learning. Fingerspelling's of Japanese sign language are based on alphabets, and some are added according to Japanese Character, gestures, numbers, and

meanings. They constructed a classification tree for easily recognized finger spellings and also used machine learning for difficultly recognized ones. They achieved an accuracy of 86% using this model. Representing high-volume and high-order data is an essential problem, especially in machine learning field. The research paper developed a model using Linear discriminant analysis (LDA). The classical LDA, however, demands that input data should be represented by vector. Such a constraint is a significant drawback to express complex data. In this paper, they have proposed a convolutional 2D□LDA method for nonlinear dimensionality reduction. The difficult problem of optimization is solved by a clever equivalence of two objective functions. The proposed method employs a two-stage end-to□end CNN to realize dimensionality reduction. Effectiveness of such structure has been proved with two different networks. The convolutional 2D LDA method out- performs the classical LDA in all experiment settings. Another method is a hardware-based controller gesture recognition system that extracts hand gestures using Flex sensors for sensing the hand movements using a glove. The sensor glove design along with the tactile sensor helps in reducing the ambiguity in gestures and shows improved accuracy. The output from the microcontroller is the recognized text which is fed as input to the speech synthesizer. Arduino microcontroller processes the data for each particular gesture made. The system is trained for different voltage values for each letter. However, this method is not very economical and requires high maintenance from time to time.