

# Hidden Markov model-based Sign Language to Speech Conversion System in TAMIL

Aiswarya V, Naren Raju N, Johanan Joy Singh S, Nagarajan T, Vijayalakshmi P  
SSN College of Engineering  
Chennai

**Abstract**—*Quick-eared and articulately speaking people convey their ideas, thoughts, and experiences by vocally interacting with the people around them. The difficulty in achieving the same level of communication is high in the case of the deaf and mute population as they express their emotions through sign language. An ease of communication between the former and the latter is necessary to make the latter an integral part of the society. The aim of this work is to develop a system for recognizing the sign language, which will aid in making this necessity a reality. In the proposed work an accelerometer-gyroscope sensor-based hand gesture recognition module is developed to recognize different hand gestures that are converted to Tamil phrases and an HMM-based text-to-speech synthesizer is built to convert the corresponding text to synthetic speech.*

## I. INTRODUCTION

Speech and language are the primary media of communication for the human race which enables them to move about with their everyday activities with ease. However, the vocally and hearing-impaired population is deprived of this medium thus making it difficult for them to completely be a part of the dominant unimpaired society. The deaf and mute use sign language to express their emotions and views unlike their normal counterparts who are natural language speakers. Sign language chiefly uses hand movements, facial expressions and movements of the eyebrows and head to convey their intended message. The people who benefit from the usage of the sign language are the ones who can interpret the same whereas, it is highly difficult for a person who is not trained in the sign language to understand the same. Hence, a translator that is capable of interpreting different hand gestures/signs and to convert it to speech is necessary to make communication between an untrained unimpaired listener and impaired speaker feasible.

The sign language to text conversion module is basically a hand gesture recognition module. Hand-gesture recognition involves recognizing the dynamic gesture by comparing it with pre-defined metrics of the trained models. Hand-gesture recognition systems find their application in areas such as character-recognition, home automation, robotic arm controllers and much more. Hand gesture recognition systems use sensors to sense a dynamic hand movement and based on the types of sensors used, the gesture recognition systems are classified into vision-based gesture recognition systems and glove-based gesture recognition systems.

**Vision-based gesture recognition systems:** In a vision-based approach, the sensor used is a camera for capturing the image/video of the static/dynamic gesture performed and the

captured information is directed to the image processing unit which processes the images through different filtering and image processing techniques. Salient features for training are extracted from the images and then the gestures are recognized using various image recognition algorithms during the testing phase.

M. Hasan et al [2] proposed an USB camera-based machine learning approach using a k-NN algorithm where the nearest possible neighbors are grouped together and a Support Vector Machine (SVM) classifier is used for classifying 16 gestures. K. K. Dutta et al [3] proposed a double handed sign language translator using a Logitech web camera sensor with minimum Eigenvalue algorithm for translating English alphabets. M. Ahmed et al [5] proposed a novel technique of implementing a translator using a Kinect sensor for depth sensing of the whole body. Here, the sensor recognizes the gestures by calculating the distance between the spinal cord and joints. Though all the vision-based techniques are natural and are economical to an extent, its performance is strongly influenced by external factors such as lighting conditions and background color. Their immobility and complicated algorithms is a big turn-down as well.

**Glove-based gesture recognition system:** It involves using wearable sensors that can capture the physical movement of the hand. Some of the glove-based sensors used are copper glove-based sensors, flex sensors [6] [7], tactile sensors and MEMS [1] [4] sensors. The biggest advantage of using a glove-based sensor is its precise data collection ability and its portability. Algorithms such as the State Estimation algorithm and k-means clustering [8] are used in combination with glove-based sensors but have resulted in extreme time consumption and increased computational complexity. Therefore, in this paper, a robust, easily compatible and a mobile translator is proposed for facilitating effective communication between the deaf and mute and the hearing and vocally unimpaired population. The sign language to speech conversion system developed here consists of a sign language to text conversion module and a text-to-speech synthesis module.

The proposed system and its modules are described in section II. The process of gesture recognition and sign-language-to-text conversion followed by the process of text-to-speech synthesis is discussed in section III. Performance of the sensor-based gesture recognition system is analyzed, and the results are discussed in section IV.

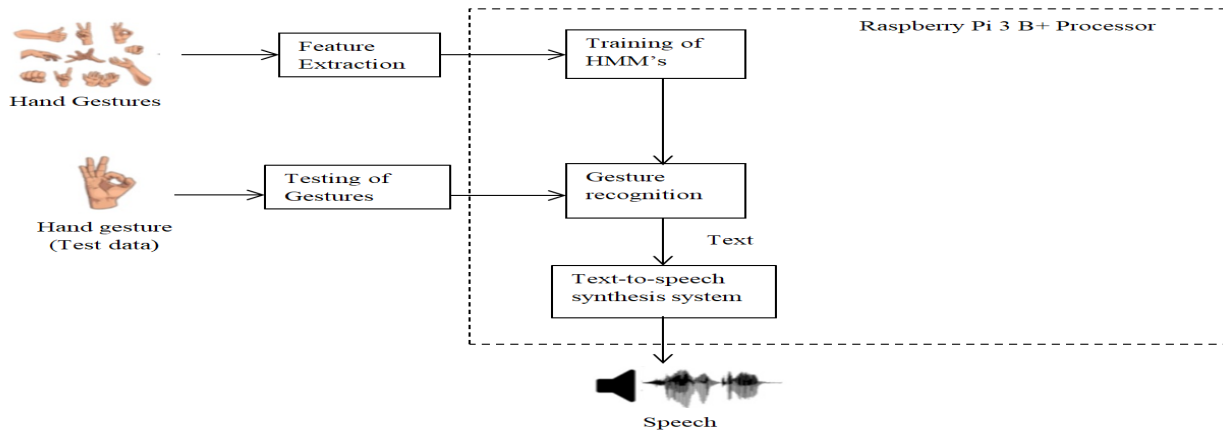


Fig.1 Sign Language to Speech Translation System

## II. PROPOSED SYSTEM AND ITS MODULES

This section addresses the implementation techniques required to build the proposed system. The proposed system is a hidden Markov model-based sign language to speech conversion system. It uses a glove-based approach consisting of a 6-axis MEMS sensor for sensing dynamic hand movements (refer to Fig.1) and is interfaced to the digital ports of a raspberry pi (3B+) that hosts the entire device.

The sensor 'MPU6050', is a 6-axis dual sensor device consisting of a gyroscope and an accelerometer. The sensed data is characterized by 6 feature vectors that are further used for training the models.

The translator operates in 3 modes mainly:

- **Training mode** – involves storing the features in the database and building models using hidden Markov modeling technique.
- **Testing phase** – involves comparing the generated features from the test data with the trained model and generating the corresponding text in Tamil.
- **Translation mode** – involves synthesizing speech in Tamil for the text corresponding to the gestures recognized.

### A. MPU6050 sensor

The MPU6050 (refer to Fig.2) sensor combines a 3-axis gyroscope and a 3-axis accelerometer on the same silicon die, together with an onboard Digital Motion Processor (DMP), which processes complex 6-axis motion fusion algorithms.



Fig.2. MPU6050

The 3-axis coordinates of the accelerometer capture the acceleration of the motion and the 3-axis coordinates of the gyroscope capture the rotation of the hand in a particular direction. These coordinates represent values corresponding to rotational shift and accelerated shift with respect to position.

### B. Raspberry Pi

A raspberry pi (refer to Fig.3) is a single board computer that is used for processing multiple tasks simultaneously. The chip's lightweight and low cost makes it affordable and portable. The pi has 40 General Purpose Input Output (GPIO) pins.

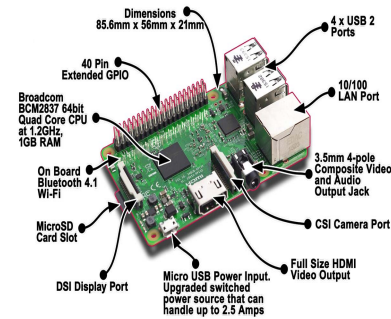


















Fig.3. Raspberry Pi

The pi has a processor of 1.2 GHz, 64/32-bit quad-core ARM Cortex-A53 and a memory of 1GB RAM at 900MHz. The pi is equipped with onboard Wi-Fi, an audio jack for playing the synthesized speech, 4 USB ports for supporting other input/output devices and a MicroSD card slot for storing the database. This single module is sufficient for training and testing thereby making it a suitable choice for the proposed system.

Using the above hardware, the proposed system is developed to take the hand-gesture as input, recognize the text corresponding to the gesture and in turn convert the recognized text into synthetic speech in Tamil thus resulting in an end-to-end communication system.

Table 1. Gestures in the conversational domain

 உங்களுக்கு வணக்கம்	 நான் சென்றுவருகி றன்	 தயவு செய்து உட்காருங்கள்	 எழுந்து நில்லுங்கள்
 என்னை மன்னிக்கவும்	 மிக்க நன்றி	 உங்களை பற்றி சொல்லுங்கள்.	 நான் வேறுபடுகிறேன்
 எப்படி இருக்கிறீர்கள்.	 நான் நன்றாக இருக்கிறேன்	 நான் சந்தோஷமாக இருக்கிறேன்	 நான் கவலையாக இருக்கிறேன்
 நேற்று நான் இங்கே வந்தேன்.	 நானை மீண்டும் சந்திப்போம்	 எனக்கு புரியவில்லை	 நீங்கள் அங்கே செல்லுங்கள்.

### III. SIGN LANGUAGE TO SPEECH CONVERSION

The proposed Sign Language to Speech Translation (SLATS) system discussed in earlier sections makes use of a hidden Markov model (HMM) based modules for sign language recognition system and text-to-speech conversion. The hidden Markov model is a statistical model in which the system is assumed to follow a Markovian process with finite hidden states. The advantage of hidden Markov modeling technique is its ability to determine the sequence of occurrence of states in a gesture by using the transitional probability metric. The most probable sequence is determined by the Viterbi Algorithm. HMM has been widely used in many applications, such as speech recognition, activity recognition from video, gene finding and gesture tracking. The extracted features corresponding to each gesture are used to train the hidden Markov models.

The total number of gestures chosen in this current work is 16 and 16 hidden Markov models are trained. HMMs are used to synthesize speech in the text-to-speech synthesis unit which is unlike existing sign-to-speech converters where pre-recorded audio files are played back. Hence, the proposed conversion system is capable of handling a large vocabulary using a text-to-speech synthesizing unit.

#### A. Sign language to text conversion

Text conversion involves mapping the recognized gestures to its corresponding TAMIL texts. The number of gestures chosen for training is 16 in the conversational domain. Each gesture lasts for 3 seconds and around 90 examples are used for training and 10 examples for testing. So, a total of 100 examples are collected from 3 individuals. Each of the 16 gestures depicts 16 different Tamil phrases. Sixteen models are trained, one for each gesture in the training phase and during the testing phase a dynamic gesture is sensed by the sensor and the gesture is recognized by comparing it with the models trained. The recognized gesture is mapped to a Tamil text.

#### B. HMM-based text-to-speech synthesis system

A text-to-speech (TTS) synthesis system converts any given text to its corresponding speech. The Tamil text obtained after recognition is fed to a text-to-speech synthesizer that provides a voice to the gesture made. This HMM-based TTS system also involves a training phase and a synthesis phase.

**Training phase:** Five hours of speech data is collected from a male speaker in a laboratory environment at a sampling rate of 16 KHz. Text from short stories forms the text corpus required for speech data collection. Feature vectors derived from the collected speech data is used to train the phoneme level HMMs.

**Synthesis phase:** In synthesis phase, each text sentence is converted to a pentaphone sequence. A sentence-level HMM is then generated by concatenating the appropriate phoneme-level HMMs. Specialized parameter generation algorithms are used to determine the spectral and excitation parameters. Finally, a vocal output characterized by the determined parameter values is synthesized with a source-filter model and is heard via a loudspeaker.

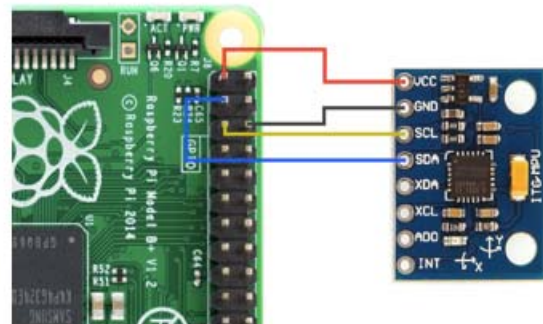


Fig.4. Interfacing MPU6050 with Raspberry Pi

The interfacing of MPU6050 with raspberry pi and the Sign Language to Speech conversion system is depicted in the Fig.4 and Fig.5 respectively.

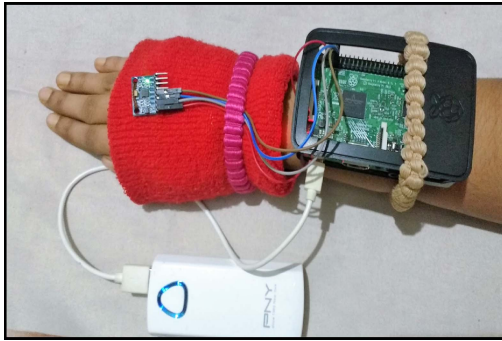


Fig.5. Sign Language to Speech Translator

#### IV. RESULTS AND ANALYSIS

The testing phase involves performing a gesture that captures the motion-related features and comparing it with the 16 logical HMMs. The output of testing is shown in Fig.6. This Tamil text is fed to a TTS synthesizer to produce a voice output.

```

=====***
Read 12 physical / 16 logical HMMs
Read lattice with 6 nodes / 7 arcs
Created network with 10 nodes / 11 links
File: gesture_recognition_htk/data/test/hi56.htk
உங்களுக்கு வணக்கம் == [30 frames]
=====***

```

Fig.6. Sign language to Text Conversion

In the proposed work the system is trained for a set of 16 gestures (Table 1) in the conversational domain. Each gesture was tested 10 times in real time and also with test datasets, and the accuracy of the recognition system was found to be **87.5% and 100%** respectively. The overall performance of the translator achieves an average score of **80-90%**. The performance of the proposed system during online testing depends on the orientation of the hand during its dynamic movement. Hence, the performance can be improved by placing the hand in the right orientation to avoid any misinterpretation of the gestures.

#### V. CONCLUSION

Communication between the deaf and mute and their normal counterparts is very difficult when the latter is not trained in sign language. Therefore, to overcome this problem, a sign language to speech conversion system is proposed. The conversion system is basically an HMM-based hand gesture recognition system that recognizes dynamic hand gestures and converts them to Tamil text. This text is fed to a text-to-speech synthesizer which gives the system a voice. The system developed follows a glove-based

approach which makes it a very portable, light weighing and power saving device. This work can be further developed by training with as many gestures as required.

#### References

- [1] B.D.Jadhav, Nipun Munot, Madhura Hambarde, Jueli Ashtikar "Hand Gesture Recognition to Speech Conversion in Regional Language" *IJCSN International Journal of Computer Science and Network*, Volume 4, Issue 1, February 2015, pp.161-166.
- [2] M. Hasan, T. H. Sajib and M. Dey, "A machine learning based approach for the detection and recognition of Bangla sign language," *2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec)*, Dhaka, 2016, pp.1-5.
- [3] K. K. Dutta, Satheesh Kumar Raju K, Anil Kumar G S and Sunny Arokia Swamy B, "Double handed Indian Sign Language to speech and text," *2015 Third International Conference on Image Information Processing (ICIIP)*, Wanknaghat, 2015, pp.374- 377.
- [4] Kiran R, "Digital Vocalizer System for Speech and Hearing impaired", in proc of *The International Journal of Advanced Research in computer and communication Engineering*, volume 4, Issue 5, May 2015, pp.81-84.
- [5] M. Ahmed, M. Idrees, Z. ul Abideen, R. Mumtaz and S. Khalique, "Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2," *2016 SAI Computing Conference (SAI)*, London, 2016, pp. 334-335.
- [6] P.Vijayalakshmi and M. Aarthi, "Sign language to speech conversion," *2016 International Conference on Recent Trends in Information Technology (ICRTIT)*, Chennai, 2016, pp. 1-6.
- [7] K.Abhijith Bhaskara, Anoop G.Nair, K Deepak Ram, Krishnan Ananthanarayanan and H.R. Nandhi vardhan, "Smart Gloves for hand gesture recognition" *2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*, 2016, pp.1-6.
- [8] T. H. S. Li, M. C. Kao and P. H. Kuo, "Recognition System for Home-Service-Related Sign Language Using Entropy-Based K-Means Algorithm and ABC-Based HMM," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.46, no.1, Jan.2016, pp.150-162
- [9] <https://github.com/raspberrypi>
- [10] <https://www.youtube.com/watch?v=ZqXnPcyIAL8>
- [11] [www.raspberrypi.org/documentation/configuration/raspi-config.md](http://www.raspberrypi.org/documentation/configuration/raspi-config.md)