# SPEECH TO SIGN LANGUAGE ANNOUNCEMENTS FOR DEAF AND DUMB IN TRANSPORT SERVICES

J.Shiyam Sunder,
CSE Dept,
Dr. Mahalingam College of Engineering and Technology,
Pollachi-642003.
TamilNadu, India.

M.Krishnapriya,
CSE Dept,
Dr. Mahalingam College of Engineering and Technology,
Pollachi-642003.
TamilNadu, India.

V.Ramya Sri,
CSE Dept,
Dr. Mahalingam College of Engineering and Technology,
Pollachi-642003.
TamilNadu, India.

***ABSTRACT:*** **Humans know each other well by conveying their ideas, thought and experiences to other. There are numerous ways to communicate ideas and best among them is the "speech". It will be injustice if we ignore those who are deprived of this invaluable gift: the deaf and dumb people. The better means of communication available to the deaf and dumb people is the use of "Sign Language". Using sign language they are limited to be exposed in a small world alone. Technology is best way to remove this hindrance and benefit these people to overcome difficulties. Here, speech and sign language are incorporated to develop the gesture based animated agent which was useful for the deaf and dumb people. The system is developed to translate speech to gesture in terms of sign language for the announcements in railway station. It consists of four modules: speech recognizer, semantic analysis, gesture sequence generation and gesture animation. For the speech recognizer uses sphinx API to convert the spoken speech input to text. Semantic analysis extracts main keyword from the speech input which in turns generates the action. Gesture sequence is used to create gesture library and associating gesture for the semantic concept. The 3D gesture animation plays the animation for the theme of spoken input which drives the sign language animation. The proposed system can be implemented in transport services i.e. railway station. The announcements made in railway station regarding train information will play the gesture animation for the announcement in railway with less effort and of low cost. It can be installed anywhere with less training.**

***Keywords:*** **Interpolation, 3D Avatar, Gesture**

## 1. INTRODUCTION

Communication is the important tool for the human being to convey one's thoughts to others. There are two way to communicate with others one is through speech and other through gesture actions. Speech is the vocalized form of communication based upon the syntactic combination of lexical and names that are drawn from very large set of vocabularies. A gesture is a form of non-verbal communication in which visible bodily actions communicate particular messages to others. The deaf and dumb people communicate only through gesture actions. There are nine billion dumb people in the world. Deafness is the impairment that worst affect communication among people. The communication between dumb and normal people is always a difficult task. One of the key aspects is adapting technologies to afford interaction between dumb and normal people by means of human-computer interaction system. People with disability and impairment constitute major part of our society. With the advent of science and technology, efforts are being made to develop certain systems to make them feel and act as normal human beings. Especially hearing impaired people interact easily through hand gestures or sign language. Human-Computer Interaction is emerging technology, used to create effective interaction between human and computers. The main advantage for choosing this system is to create a realistic and attractive animated gesture which enables visualization.

A gesture is non-verbal forms of communication basically physical actions are used by person to convey their message. Gestures are classified into two categories. Static gesture which depends only on the information about the flexure angles of fingers and dynamic hand gestures which depend not only on the finger flex angles but also on the hand trajectories and orientation. Language is tool for effective communication which clearly describes the thoughts of an individual. Sign language is the

language used by the deaf people for communication. Sign languages are full-fledged languages that possess grammar and lexicons just like any other spoken language. Deaf people use their hand and finger to project their thoughts to other.

Due to hearing impairment, people generally accept using sign language. Hearing impaired people face many difficulties in day to day activities of life. One of the difficulties they face is not capable to hear the announcement of train details in railway station and probability to miss their trains are more. We are contributing our idea to this area to relieve the sufferings of disabled peoples by introducing speech to sign language translated gesture announcement displays in transport services for easy understanding and to survive comfortably in public places without the help of others. It is also further considered as a teaching tool of sign language globally, when installed in public places for normal people to understand and learn sign language from gesture displays.

## 2. RELATED WORK

The communication between the deaf and normal people was always challenging task. The number of research was carried out to develop the efficient sign language translator. The speech to gesture translation architecture describe about the overall process of converting the speech to gesture. It described about the three modules in this architecture speech recognizer, translation rules and sign animation generation. The main advantage of this paper is it reduce 40% delay between the spoken utterance and animation sign sequence by modifying speech recognition system to report partial recognition for every 100ms.These partial recognition are translated into sign sequence does not need to wait until the end of the spoken utterance[1]. The Speech is most prominent & primary mode of Communication among of human being. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. Today Speech recognition is one of the promising technology .Speech recognition is recognize the speech and translate it into the text .Describe detailed process of converting the speech into text[2] .The speech recognition is also called as ASR. It converts the speech into text using analysis, feature extraction, modelling, testing. Speech analysis deal with suitable size for framing the segmentation for extraction of feature in next step .In this paper they described three methods such as segmentation analysis, sub -segment analysis and super-segmental analysis. Feature extraction is deriving descriptive features from the windowed and enhanced speech signal to enable a classification of sounds. In this paper many feature extraction techniques and algorithms are compared based upon their accuracy and speed. For feature extraction Mel-Frequency Cepstral Coefficient is efficient. Modeling is a technique used to generate the speaker model using speaker specific feature vector. This paper shown the HMM is efficient and more accurate than the other approach. Based on this our system speech recognition software is develop using HMM is chosen to meet the high accuracy. The adaptability is key factor for determining the performance and scope of any system. It describes a new version of a speech into sign language translation system with new tools and characteristics for increasing its adaptability to a new task or a new semantic domain [3]. This system is made up of a speech recognizer (for decoding the spoken utterance into a word sequence), a natural language translator (for converting a word sequence into a sequence of signs belonging to the sign language), and an animation module (for playing back the signs). Example-Based translation is a method of machine translation often characterized by its use of a bilingual corpus with parallel texts as its main knowledge base at run-time. Bilingual corpus is the dictionary that contains the list of words and their pronunciation. Based upon the speech input semantic analysis generates the example-based sentences. Explain the example-based translation approach and their advantage and their accuracy and speed [4]. Statistical-Based translation is a machine translation paradigm where translations are generated on the basis of statistical model whose parameters are derived from the analysis of bilingual text corpus. The statistical model is a class of mathematical model, which embodies a set of assumptions concerning the generation of some sample data and similar data from larger vocabulary [5]. Rule-based translation is based on linguistic information about source and target languages basically retrieved from word set and dictionaries covering the main semantic. Rule-based approach is used to translate the input text into rule-generation based on the RTN algorithm.

By comparing the process and computational speed [4] [5] [6].The rule-based approach is used to develop our system. The rule-based is work well for the small set of words. In future the statistical approach of machine translation will be used to develop the adaptable system. The statistical approach is work well for large set of words.

## 3. PROPOSED SYSTEM

The proposed system emerges with an idea to develop speech to sign language gestures driven

animation contributing mainly towards the deaf and dumb people. As, the communication is a main tool for expressing the thoughts of an individual and people with impairment to be treated as like normal in order to make them do their task by their own this system can be much more suitable when put in public areas like transport services especially in railway stations.

In the proposed system, speech to sign language announcements consists of four modules: speech recognition, semantic analyzer, gesture generation and gesture animation. In speech recognition module the software sphinx has been used to recognize the speech and convert it into text. For semantic analysis module we use pattern matching algorithm and rule-based approach to extract the main concept of the speech. In gesture generation module is to create gesture library for the list sign language used in the railway station and uses semantic concepts to associating them with the several gesture.

For gesture animation, the animated agent is developed and by using interpolation strategy for reducing the effort in gesture animation. This strategy makes the system to automatically generate all agent positions necessary for meaning full gesture animation. In our system, the complexity of gesture animation generation is reduced with the help of interpolation technique and we create gesture library for standard sign language which was used by world - wide so the system can be installed anywhere. The modeled system can be easily modified and further been extended to support any domains whichever is possible.

### 3.1SYSTEM OVERVIEW

The architecture proposed for the translation of speech into gestures for Deaf people. In this figure, the four main modules which needed in the translation process: speech recognition, gesture generation and gesture animation.

The first module is speech recognition converts speech utterances into text. For these modules, it uses the sphinx API. It is open source speech recognition tool that includes essential dictation and we have developed d control features. This module uses language and acoustic modules for pronunciation. This recognizer uses Hidden Markov Model

The semantic analysis module carries out the semantic evaluation of the text sentences and extracting the main concepts .the developed parser uses context-free grammar to extract the semantic concept from word sequence and generate the rule. The gesture generation module we create the gesture library for the sign language used in the railway and process the semantic analysis output and assign the sequence of gestures to the semantic concepts.

In this process, we consider four situations: one concept mapped onto a unique gesture, one concept generates several gestures, some concepts are mapped onto a unique gesture and several concepts generate several gestures.

In the gesture animation module, an animated agent signs the gesture sequence. This agent is a very simple representation of a human being but it permits the gestures of the sign language to be represented properly. The main problem with this agent is great effort needed to create an animation. Each animation needs at least 20 agent positions per second to guarantee a continuous movement. So, we develop the system automatically create significant percentage of agent position needed to generate animation. The system uses the main position andinterpolation strategies to create all agent position needed in continuous movement.

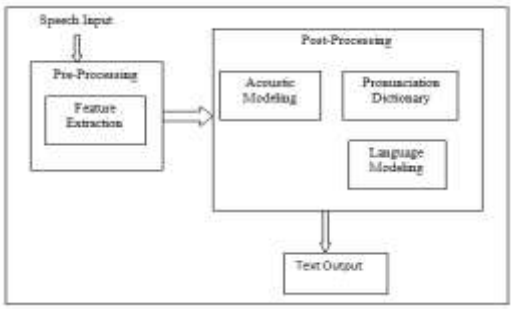
## 4. MODULE DESCRIPTION

The System consists of four modules: Speech recognition, Semantic analysis, Gesture generation, Gesture animation

### 4.1 Speech Recognition Module

This module uses the sphinx API for recognizing the speech and convert to text. Sphinx is the speech recognition software that includes essential dictation and command/control features. With this software, it is possible to control application using voice. The sphinx is developed by using efficient HMM model.

The application of the proposed is restricted only for railway services domain. In this, the vocabulary used by the sphinx recognizer must be adapted to this specific domain eliminating the ineffective words from the available set of huge dataset. Additionally, the acoustic and language models can be adapted to the speaker. These two facts allow us to obtain high word accuracy.



In order to recognize speech, the system usually consists of two phases. They are called pre-processing and post-processing. Pre-processing involves feature extraction and the post-processing stage comprises of building an acoustic model, phonetic lexicon or the renunciation dictionary and language modeling. Acoustic modeling of speech typically refers to the process of establishing statistical representations for the feature vector sequences computed from the speech waveform. It also encompasses "pronunciation modeling", which describes how a sequence or multi-sequences of fundamental speech units (such as phones or phonetic feature) are used to represent larger speech units such as words or phrases which are the object of speech recognition.

Pronunciation Dictionary also called as lexicon contains the detail about how the words are pronounced. The lexicon should contain all the words that the speech recognition engine need to recognize. The pronunciation dictionary consists of records containing words and associated homophones sequences. Some words have multiple pronunciations. A pronunciation dictionary is also a language specific resource. So, the pronunciation dictionary for every word in the vocabulary has been constructed manual for the English language. The sample set of words are add to the dictionary such as place, train names, A language model is used to restrict word search. It helps a speech recognizer figure out how likely a word sequence is, independent of the acoustics. It represents previous knowledge about language and the expectations at utterances. It can be expressed in terms of which words or word sequences are possible or how frequently they occur.

**4.2 Semantic Analysis Module**

Semantic analysis module is used to extract the main semantic concepts from the text sentences. In this module the parser is developed. The parser is designed for the development of simple, robust natural language parser.

The parses each input utterance into a sequence of one or more semantic frames. The parser is designed to set the frames and provide the grammar rules that specify the word string that can fill each slot in a frame. This information is used to map input word string onto a sequence of semantic frame. A frame represents some basic type of action or object for the final speech to sign language application. Slots in a frame represent information that is relevant to the action or object. Each slot has an associated context- free grammar that specifies word string patterns that match the slot. The grammar rules are compiled into Recursive Transition Networks(RTNs) and the slot name will be the root of the corresponding semantic parse tree . In this parser, it is possible to define new slots within other slots implementing a hierarchical structure.

The parsing process is a dynamic programming algorithm where grammars for slots are matched against a word string to produce a slot graph. The set of active frames defines a set of active slots each slot points to the root of an associated RTN. These networks are matched against the input word sequence by a top-down RTN chart-parsing algorithm. The parser proceeds from left to right in an attempt to match each slot network starting with each word of the input.

The recursive function is used that matches the word string. The function produces all matches for the network of starting at the word position and different endpoints. Each time a net match is attempted, this noted in the chart. Any time a net match is found, the chart is first checked to see if the match has been attempted before. When a slot match is found, it is added to the slot graph. Words are not skipped in matching slot, but words can be skipped between the matched slots. The resulting graph represents all sequence of words found in the sentences. The sequence of slots represents by the graph is then grouped into frames. This is done easily by assigning frame labels to the slot. The dynamic programming search produces the most complete, least fragmented parsed output. The processing speed of the parser is generally linear with the length of the input. The sample semantic output of the word set is:

Input: Good afternoon sir, your attention please Chennai Express from Madurai to Chennai will leave at 18:00 platform number 3

Output:

Greeting : Good Afternoon

Train Name : Chennai Express

Time : 18:00

Platform Number: 3

Place : Madurai -> Chennai

**4.3 Gesture Sequence Generation Module**

Gesture sequence generation the gesture libraries are created for the sign language. The semantic analysis output is a sequence of parsed slots: each parsed slot is considered a semantic concept. In this step, the gesture sequence generation processes the semantic analysis output to obtain the gestures to play the animated agent. In this process, describes four situations:

One semantic concept corresponds to a specific gesture. In this case, a semantic concept parses slot is directly mapped onto a specific gesture. The translation is simple and it consists of assigning one gesture to each semantic concept. This gesture can be a default translation, independent of the word

string, or can be different depending on the word string from which it is generated.

Several semantic concepts are mapped onto a unique gesture. The second situation appears when several concepts generate a unique gesture. This situation should be solved in the previous step. The solution is to unify the concepts in the parser grammar and to proceeds previous situation. As in the previous situation, the gesture being generated may or may not differ according to the slot content.

One semantic concept generates several gestures. This situation occurs when it is necessary to generate several gestures from a unique concept. This problem strongly justifies the need for the gesture sequence generation module. Similar to previous sections, the gesture sequence and its order can depend on the concept and its content, or just on the concept. This situation appears in many translation issues:

1. Verb-A verb concept generates a gesture related to the action proposed by the verb and some gesture provide information about the action tenses such as past, present or future, the action subject and gerund action
2. Complex Signs-Similar to the paraphrases, complex sign are made up of several gestures. Each of these gestures can be used independently but they are represented together in order to show another concept.
3. Date and Time- A date representation can be made with one or several gestures. The time generally requires several gestures for a full representation.
4. Plural Nouns -There are several ways of specifying an object in plural by repeating the gesture, introducing an adverbial gesture or representing the gesture with both hand.

Several semantic concepts generate several gestures. This is the most complicated situation appears when it is necessary to generate the several gestures from several concepts with certain relationships between them.

In our system, the cases presented are solved by mixing the strategies explained above. First, we group the different concepts under a unique concept structure, and then we apply similar strategies to generate a gesture sequence from a unique semantic concept structure. The characteristics of sign language used by deaf people have been extracted from [11] where obtained an extended and detailed description. Based on the standard sign language used by world wide the gesture library is created for the basic set of words used in the railway sector.

**4.4 Gesture Animation**

In order to represent the gesture sequence, we have developed animated agent. This agent is a simple representation of a human person is used to represent the gesture used in sign language. Around the world, several companies and research centers have developed animated agents for human-computer interfaces. The main problem with the agent is the great effort needed to build an animation. It is necessary to generate several agent positions per second in order to obtain a continuous movement. The focus of this project is to reduce drastically the effort in gesture design.

For representing the gestures, the system was developed a very simple animated agent by using the java programming. This agent is made up by combining rectangles, ellipse, and different sized lines.

An animation is generated automatically from a very small set of agent positions. The main target of this module is to generate an animation using a few positions as possible in order to reduce drastically effort of generating gesture animations. A typical gesture takes approximately 2s. This means that, considering 20 frames per second, we create 40 frames/agent positions for a typical gesture. In sign language, there are more than 5000 different gestures. Animation creation is no trivial task. To reduce effort we use strategy interpolation. The main idea of this is to define a small number of frames/positions and to generate the intermediate positions automatically. In order to design an interpolation, it is necessary to define two aspects: the trajectory and timing.

Final

Initial

We define the trajectory that any point of the agent body will follow when moving from the initial to final position Figure 6.The trajectory is specified by moving the mouse cursor. When trajectory defined, this trajectory can be assigned to a unique mobile point, a set of mobile points, or to all mobile points. No trajectory is assigned then static point of agent will be displayed. For the mobile points, the program generates a rectilinear one by default.

The second aspects to define the timing how fast the point passes through the different parts of the trajectory. The trajectory is a continuous line but the number of intermediate positions is small around 10.Because of this, the mobile point will be situated for each of the interpolated position in Figure 7. If no timing is specified, the program positions of the intermediate points equidistantly. The interpolated positions are created by the program combining the trajectory and timing associated to each point. The gesture animation stored in a file. To play a sequence of gestures, we carry out two actions: first we concatenate the gesture animations in order to produce a continuous movement, and seconds, to define the speed of play. When concatenating two

gestures, it is necessary to introduce new agent positions between the last position of a gesture and the first position of the next gesture. This action is important, in order to produce a continuous movement.

The final aspect is the speed of gesture sequence .This aspect is defined by the time between two consecutive agent position representations. To slow the gesture playing we increase the time and vice-versa. There is relationship between the duration of the utterance and duration of gesture sequence. The gesture sequence is around one and half and it is two times longer than the utterance duration. The speaking rate varies significantly for one utterance to another. This variation is applied in gesture sequence by reduce the time between positions or increased the time between the positions.

Thus now the animated agent is created for alphabets of English language and also for the words used in the railway station.

## 5. CONCLUSION

Main effort of this project is to provide a viable tool for people with impairments to have their day to day activities by their own without help of others in an effective way. Useful for deaf and dumb people, to communicate with others as like as normal people. System is designed in such a way that it converts the speech into sign language. Here, the railway services has been chosen as domain for speech driven sign language gestures, to facilitate the deaf people to know about the details of railway information. This includes arrival of train to the platform, departure, platform number, train number, train name are few examples.

It can also be used as a learning tool for the hearing impaired people. As an initial level it has been worked on railway domain in later part other possible areas can be identified and established with same. This can also be further be implemented for native languages after successful performance of current system

## REFERENCES

[1] M. Naor and A. Shamir, “Visual cryptography,” in Advances in Cryptology, vol. 950. New York, NY, USA: Springer-Verlag, pp. 1–12, 1995.

[2] G. Ateniese, C. Blundo, A. D. Santis, and D. R. Stinson, “Visual cryptography for General Access structures,”Inf. Compute., vol. 129, no. 2, pp. 86–106, 1996.

[3] A. Adhikari, T. K. Dutta, and B. Roy, “A new black and white visual cryptographic scheme for general access structures,” in Progress in Cryptology (LNCS), vol. 3348. New York, NY,pp. 86–106,2010.

[4] C. S. Hsu, S. F. Tu, and Y. C. Hou, “An optimization model for visual cryptography schemes with unexpanded shares,” in Foundations of Intelligent Systems(LNAI), vol. 4203. New York, NY, USA: SpringerVerlag, pp. 58–67, 2006.

[5] F. Liu, C. Wu, and X. Lin, “Step construction of visual cryptography schemes,” IEEE Trans. Inf. Forensics Security, vol. 5, no. 1, pp. 27–38, USA: Springer-Verlag, 2004, pp. 399–345, Mar. 2010.

[6] H. Koga, “A general formula of the (t,n)-threshold visual secret sharing scheme,” In Advances in Cryptology, Asia crypt. New York, NY, USA: Springer-Verlag, pp. 328–345, 2002.

[7] K. H. Lee and P. L. Chiu, “An extended visual cryptography algorithm for general access structures,”IEEE Trans. Inf. Forensics Security,vol.7,no. 1, pp. 219–229,Feb. 2012.

[8] P. L. Chiu and K. H. Lee, “A simulated annealing algorithm for general threshold visual Cryptography schemes,” IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 992–1001, Sep. 2014.

[9] L. A. MacPherson, “Grey level visual cryptography for general access structures,” M.S. Thesis, School of Comput. Sci., Univ. Waterloo, Waterloo, ON, Canada, 2002.

[10] J. Weir and W. Yan, “A comprehensive study of visual cryptography,” in Transactions on Data Hiding and Multimedia Security V (LNCS), vol. 6010. New York, NY, USA: Springer-Verlag, pp. 70–105, 2010.

[11] C. N. Yang, “New visual secret sharing schemes using probabilistic method,”Pattern Recognit. Lett., vol. 25, no. 4, pp. 481–494, 2004.

[12] R. Ito, H. Kuwakado, and H. Tanaka, “Image size invariant visual cryptography,” IEICE Trans. Fundam., vol. E82-A, no. 10,pp. 2172–2177, 1999-2011.

[13] F. Liu, C. K. Wu, and X. J. Lin, “A new definition of the contrast of visual cryptography Scheme,” Inf. Process. Lett. , vol. 110, no. 7,pp. 241–246, 2010.

[14] C. Blundo and A. De Santis, “Visual cryptography schemes with perfect reconstruction of Black pixels,”Comput. Graph. vol. 22, no. 4, pp. 449–455, 1998.

[15] T. Hofmeister, M. Krause, and H. U. Simon, “Contrast-optimal k out of n secret sharing schemes in visual cryptography, ”Theoretical Comput. Sci., vol. 240, no. 2, pp. 471–485, 2000.

[16] Y. W. Chow, W. Susilo, and D. S. Wong, “Enhancing the perceived visual quality of a size Invariant visual cryptography scheme,” in Information and Communications Security (LNCS), vol. 7618. New York, NY, USA: Springer-Verlag, pp. 10–21, 2012.

[17] H. B. Zhang, X. F. Wang, W. H. Cao, and Y. P. Huang, “Visual cryptography for general Access structure using pixel-block aware encoding,” J. Comput., vol. 3, no. 12, pp. 68–75, 2008.

[18] R. Ito, H. Kuwakado, and H. Tanaka, “Image size invariant visual cryptography,” IEICE Trans. Fundam. Electron., Commun., Comput. Sci.,vol. 82, pp. 2172–2177, Oct. 2013.

[19] C. S. Hsu, S. F. Tu, and Y. C. Hou, ―An Optimization Model for Visual Cryptography Schemes with Unexpanded Shares,‖ *Foundations of Intelligent Systems, LNAI*, vol. 4203, pp.58–67, 2014.

[20] Z. Zhou, G. R. Arce, and G. Di Crescenzo, “Halftone visual cryptography,” IEEE Trans. Image Process., vol. 15, no. 8, pp. 2441–2453, Aug. 2013 K. H. Lee and P. L. Chiu, “An extended visual cryptography algorithm for general access structures,” *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 1, pp. 219–229, Feb. 2012.