ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

Real Time Conversion of English Audio to American Sign Language and Vice Versa

Manasa Mandapati[1]

[1]School of Computer Engineering and Applications DY Patil International University
Pune, India

Publication Date: 2025/09/15

Abstract: The hearing-impaired people form an ample community with required needs that technologists have started to address. There is no device to date that can convert audio to sign language and sign language to audio in real-time. This problem can be handled using an interpreter system for speech to sign language to translate English speech to American Sign Language video in real time[2]. Similarly the sign language is translated to speech using a device with arduino board and flex sensors. Gestures made by the wearer are detected using sensors, and as per the pre-defined conditions for numerous values generated by sensors, corresponding messages were sent to the Android device using Global System for Mobile (GSM) [4], which will convert these text messages to speech.

Keywords: *Sign Language, Speech Recognition, Flex Sensors, Gestures, Speech to Text, Text to Speech.*

How to Cite: Manasa Mandapati (2025) Real Time Conversion of English Audio to American Sign Language and Vice Versa. *International Journal of Innovative Science and Research Technology*, 10(9), 494-499. https://doi.org/10.38124/ijisrt/25sep099

I. INTRODUCTION

The commercial market is still trying to reduce the gap between deaf and non-deaf communities and the main motive for developing this device is to help hearing impaired people to improve their quality of life via speech to sign and sign to speech translation. The Sphinx is manipulated as the recognition engine for the speech to sign language interpreter system (SSLIS) and for translation, signed English manual is employed as a manual parallel to English. Figure 1 gives the basic structure of SSLIS. Speech is taken as input through microphone and translated to text using some speech recognition engine. The recognized text will given as input to ASL database which will look matching on word basis. The database contains prerecorded video signs where every basic word has a video clip. If there is any match then the equivalent ASL translation is going to be displayed as per signed English manual, otherwise the word is going to be finger spelled. The final output contains the Recognized text as well as ASL translation.

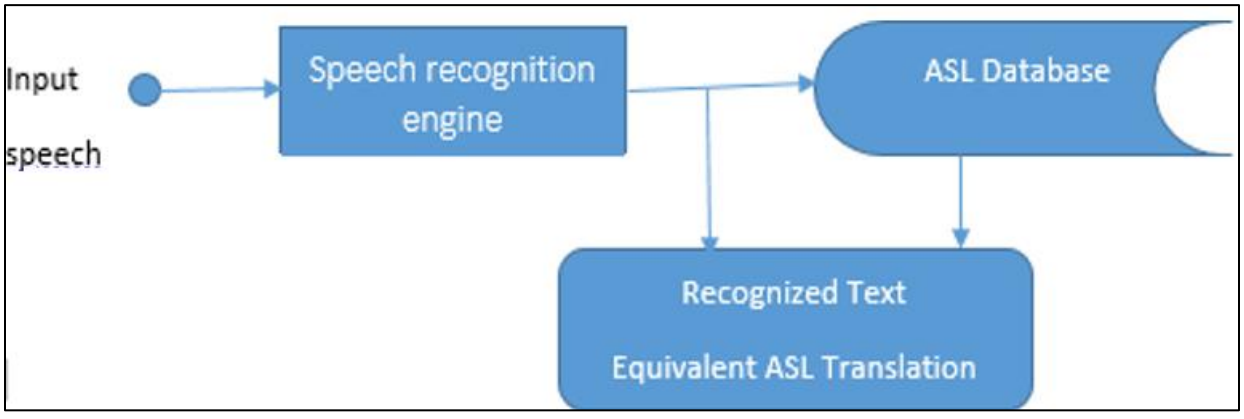

Fig. 1. SSILS Structure

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

II. SENSORS

Figure 2 gives the flex sensors circuit diagram. Gestures are detected by these flex sensors. For this purpose the flex sensors are attached to the gloves of the hearing impaired person. The change of flex sensors in various angles gives corresponding resistance value. At first, the change in voltage according to the bend is measured and secondly this voltage is converted to current so that based on flex sensor movement, the current is also generated[6].

III. METHODOLOGY

A. Automated Speech Recognition

Automated speech recognition (ASR) is the process of converting audio signal to a textual message. A large vocabulary, speaker independent and continuous speech recognizer is used here. Fig 3 shows a large vocabulary continuous speech recognizer (LVCSR) structure[9].

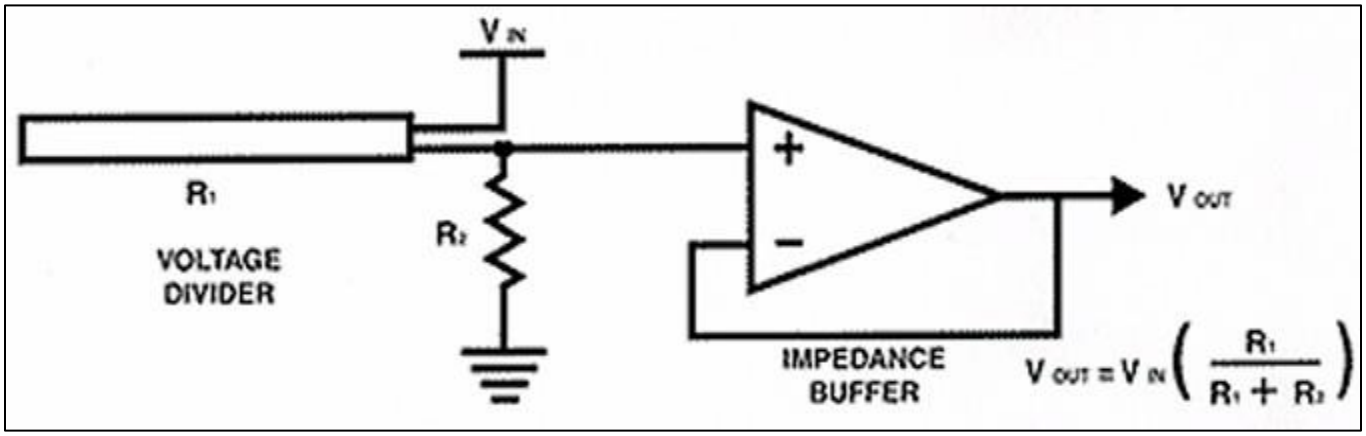

Fig. 2. Flex Sensor Circuit Diagram

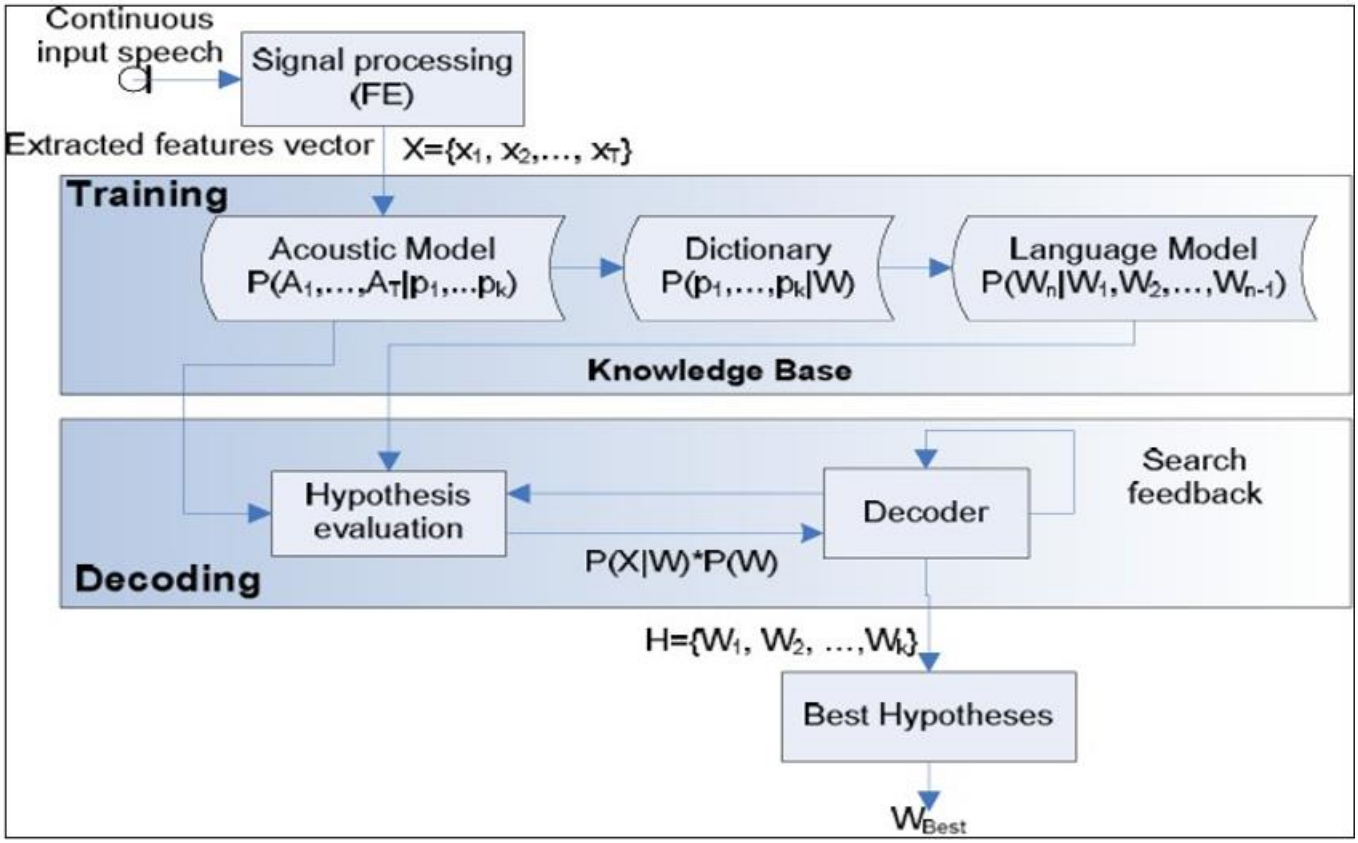

Fig. 3. Simplified Structure of LVCSR

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

B. The Sphinx 3.5

Sphinx [6]is a speech recognizer engine which was used for developing the SSLIS. It is a collection of tools and resources for building speech recognizers. 3.5 is the version. Latest versions can also be used.

C. ASL in SSLIS demonstration

The database of ASL contains 2600 prerecorded video clips each representing sign of each basic word of English vocabulary and single handed American manual alphabet. The word will be checked to find out whether it is a basic word. If it is a basic word then it will retrieve the prerecorded video clip from ASL. If the word is non basic then it will be gestured by applying it to American Manual alphabet database. The final output could be prerecorded videos and gestures.

D. Arduino Uno

Arduino Uno is a micro-controller board which has 14 digital pins. Source code is added to arduino uno using Arduino software (IDE) and it uploads the code to micro-controller. sensors, GSM module and LCD display are connected to micro- controller. All these components are connected to arduino using the pins as shown in fig 4.

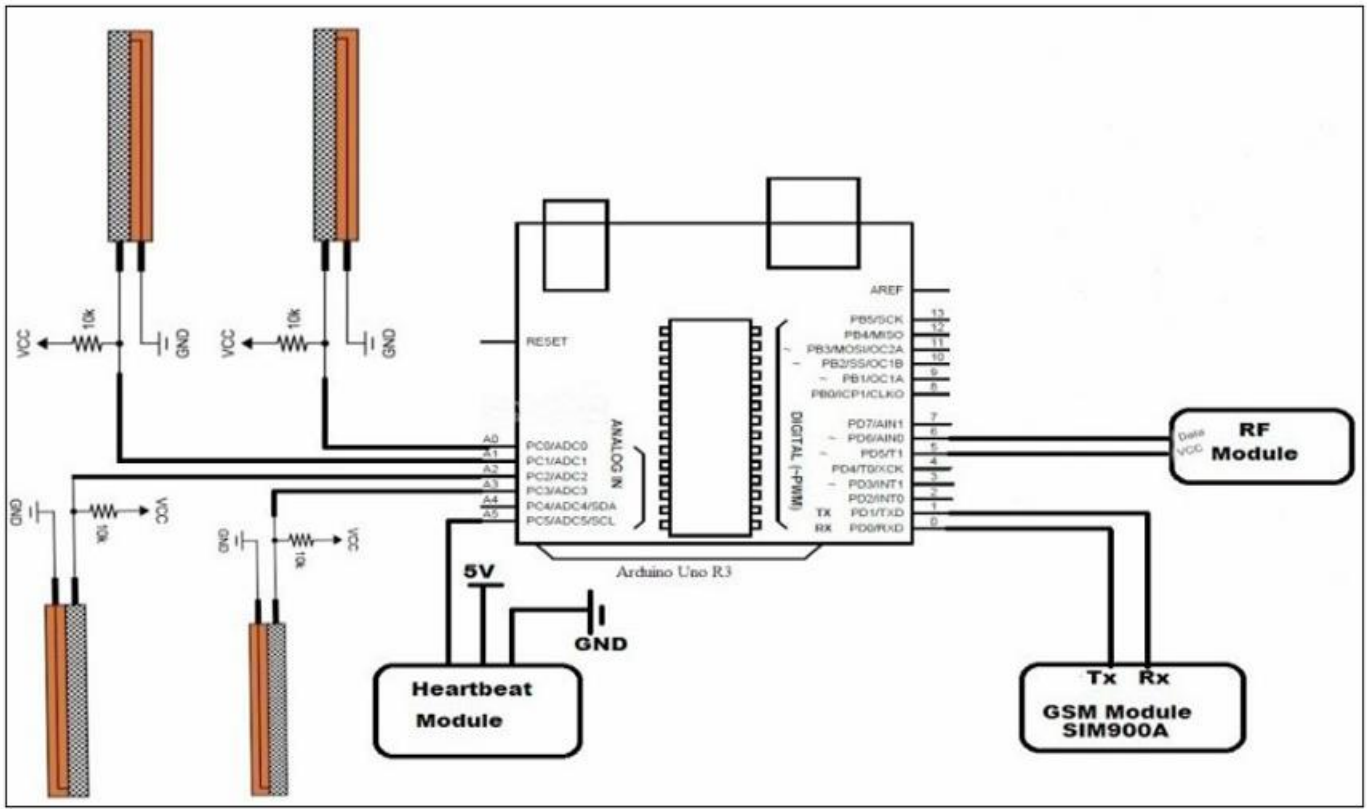

Fig. 4. Circuit Diagram of Sign to Speech Recognition

E. GSM Module

Values are sent to cloud via GSM. For controlling the SMS, data transfer and remote control we use Global System for Mobile. The values are plotted in thing speak cloud itself [12]. If the pulse sensor value exceeds certain condition then the message can be generated from hearing impaired person's mobile device. Text to speech conversion can be done by developing an android application. The text message obtained and be out using a speaker.

IV. GRAPHS

A dictionary and language model are generated and then a text set following their context has been recorded and are applied to live pretend program. The parameter of interest was expected to be tuned and word error rate (WER) is calculated Turing parameters pertaining to pruning behavior

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

V. RESULTS

The Sphinx is manipulated as the recognition engine for the speech to sign language interpreter system (SSLIS) and for translation, signed English manual is a manual parallel to English[9].Speech is taken as input through microphone and then translated to text through some speech recognition engine. The recognized text will given as input to ASL database which will look matching on word basis. The database contains prerecorded video signs where every basic word has a video clip. If there is any match then the equivalent ASL translation is going to be displayed as per signed English manual, otherwise the word is going to be finger spelled. The final output contains the Recognized text as well as ASL translation.

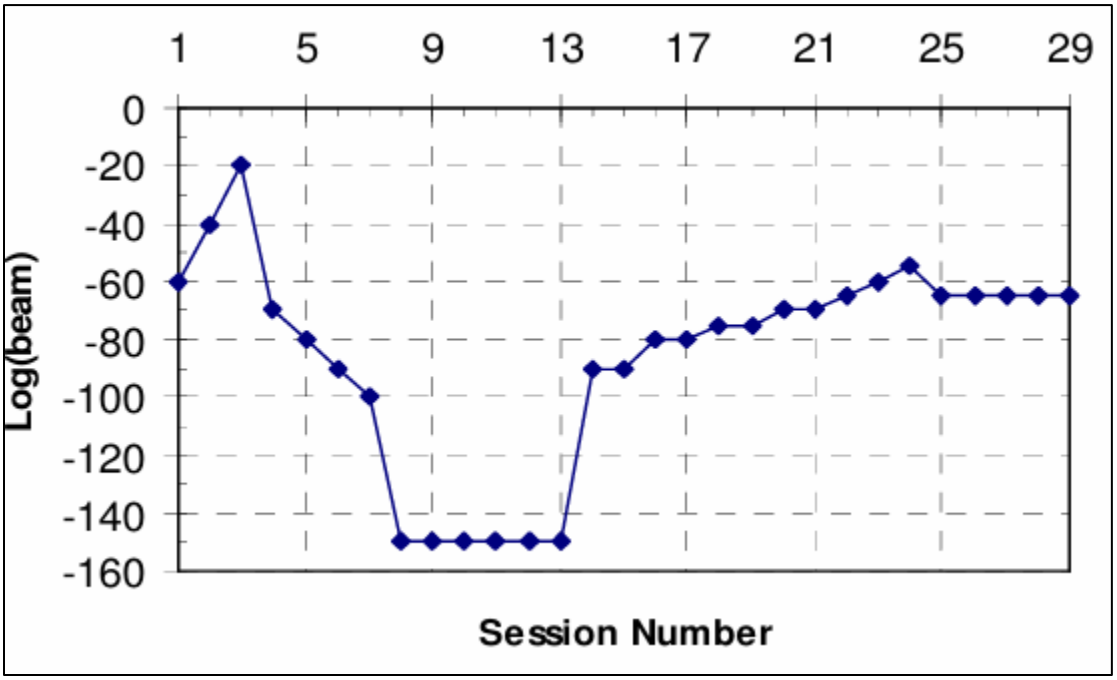

Fig. 5. Log(Beam) vs Session Number

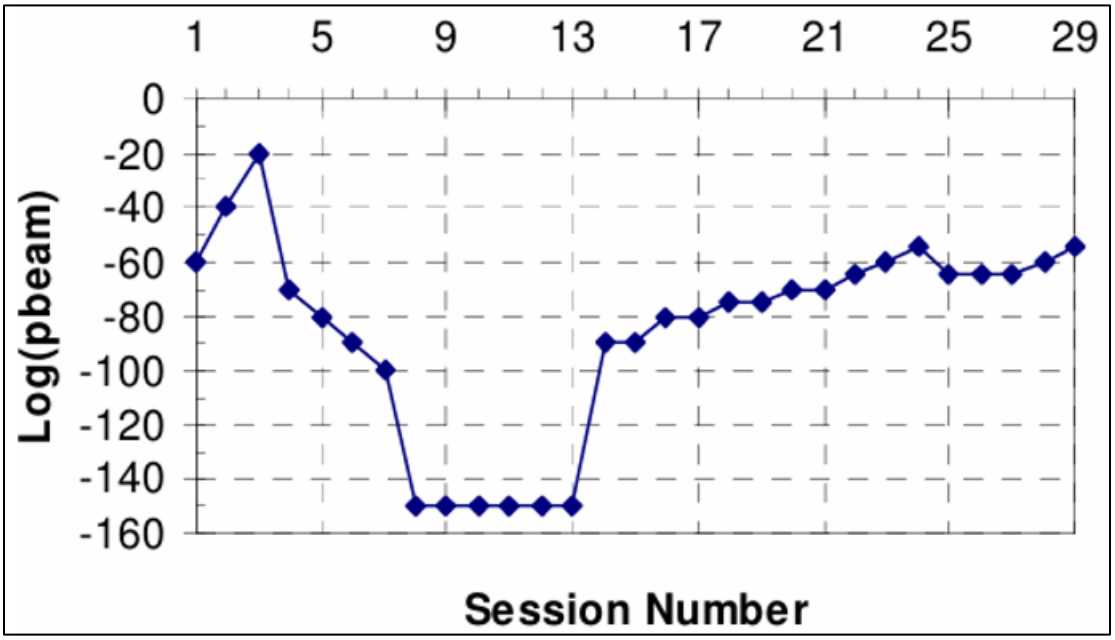

Fig. 6. Log(Pbeam) vs Session Number

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

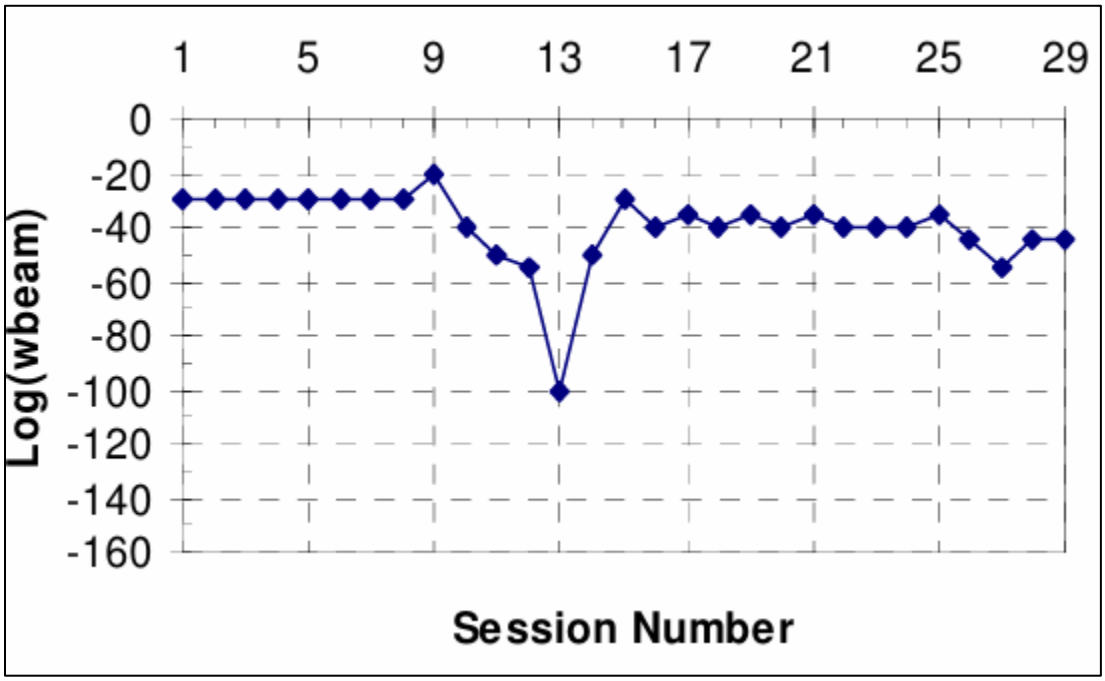

Fig. 7. Log (Wbeam) vs Session Number

VI. CONCLUSION

Association of various needs of hearing impaired people is done using flex sensors and predicted their needs using back propagation neural network. The input English speech is converted to equivalent sign video and the gestures made by the hearing impaired is converted to speech. The prediction can be made accurate by adjusting the weights in the neural network model. Along with real time conversion of speech to sign language and sign language to speech, prediction of hearing impaired persons can also be done using this model and future enhancements can be done if the person can evaluate the prediction if it is true or not. These results can be again used as input for training the neural networks for future predictions.

REFERENCES

[1]. American Sign Language Video Dictionary and Inflection Guide. (2000). [CD-ROM]. New York: US. National Technical Institute for the Deaf, Rochester Institute of technology. ISBN: 0-9720942-0-2

[2]. ASL University. Finferspelling: Introduction. http://www.lifeprint.com/asl101/fingerspelling/fingerspelling.

[3]. Baker, J.K. (1975). The DRAGON System-An Overview. IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-23(1). pp.24-29.

[4]. Becchetti, C., Ricotti, L. R. (1999). Speech Recognition Theory and C++ Implementation. England: Wiley.

[5]. Bornstein, H., Saulnier, K.L. Hamilton, L.B. (1992). The Comprehensive Signed English Dictionary (Sixth printing). USA: Washington DC, The Signed English series. Clerc Books, Gallaudet University Pres.

[6]. Gouveˆa, E. The CMU Sphinx Group Open Source Speech Recognition Engines. http://www.speech.cs.cmu.edu/sphinx/

[7]. Harrington, T. (July, 2004). Statistics: Deaf Population of the US. http://library.gallaudet.edu/dr/faq-statistics-deafus.html

[8]. Huang, X., Acero, A., Hon, H-W., Reddy, R. (2001). Spoken Language processing, a Guide to Theory, Algorithm and System Development. Prentice Hall PTR

[9]. Hwang, Mei-Yuh. (1993). Subphonetic Acoustic Modeling for Speaker Independent Continuous Speech Recognition. Ph.D. thesis, Computer Science Department, Carnegie Mellon University. Tech Report No. CMU-CS-93-230

[10]. iCommunicator TM pricing (2003). http://www.myicommunicator.com/?action=pricing

[11]. Jelinek, F. (Apr. 1976). Continuous Speech Recognition by Statistical Methods. Proceedings of the IEEE, Vol. 64, No. 4. pp. 532-556.

[12]. Ravishankar, M. (May 1996) Efficient Algorithms for Speech Recognition. Ph.D. dissertation, Carnegie Mellon University. Tech Report. No. CMU-CS-96 143.

[13]. Ravishankar, M. K. (2004). Sphinx-3 s3.X Decoder (X=5). Sphinx Speech Group. School of Computer Science, CMU. http://cmusphinx.sourceforge.net/sphinx3/

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep099

[14]. Rosenfeld, R. The CMU Statistical Language Modeling (SLM) Toolkit, http://www.speech.cs.cmu.edu/SLM*i n f o.ktml*