International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

# Gesture Listen

**Shridharan R B, Srikumbesh S, Zawahar S, Muthukrishna vinayagam B**

*B.Tech student, Dept. of Artificial intelligence and data science , Kamaraj college of engineering and technology, Tamil nadu,India*
*B.Tech student, Dept. of Artificial intelligence and data science , Kamaraj college of engineering and technology, Tamil nadu,India*
*B.Tech student, Dept. of Artificial intelligence and data science , Kamaraj college of engineering and technology, Tamil nadu,India*
*Professor, Dept. of Computer science engineering , Kamaraj college of engineering and technology, Tamil nadu,India*

---

**Abstract -** *In an increasingly interconnected world, effective communication is essential for social inclusion and collaboration. However, individuals with speech or hearing impairments often face barriers in expressing themselves digitally. "GestureListen" addresses this challenge by offering a groundbreaking web-based platform that facilitates seamless communication between individuals with impairments and others. Leveraging advanced technologies such as speech recognition and natural language processing, GestureListen interprets spoken language or text input into Sign Language gestures in real-time. By providing a user-friendly interface and robust functionality, GestureListen aims to enhance accessibility and foster greater understanding and integration within diverse communities, ultimately empowering individuals to communicate fluently and inclusively in a digital environment. Through the integration of cutting-edge technologies like Django for web development and NLTK for natural language processing, this project seeks to empower individuals with speech or hearing impairments to communicate effectively in a digital world. By harnessing the power of speech recognition, the system will transcribe spoken language into text, while NLTK facilitates the analysis and interpretation of textual input. This processed data will then be transformed into ISL gestures or animations, enabling users to express themselves fluently and interactively. With its user- friendly interface and robust functionality, this application endeavors to promote inclusivity and empower individuals with disabilities to participate fully in both online and offline conversations.*

***Key Words*: Gesture translation, Audio to gesture, Text to gesture, Gesture communication, Sign language generator.**

## 1.INTRODUCTION

Gesture Listen is an innovative web-based application designed to bridge the communication gap for individuals with speech or hearing impairments. Leveraging advanced technologies like speech recognition and 3D animation, Gesture Listen converts live audio speech recordings into text and renders them into Indian Sign Language animations in real-time. This revolutionary tool aims to empower users by providing them with an intuitive and accessible means of communication. Through its user-friendly interface and robust functionality, Gesture Listen promises to revolutionize the way individuals interact and express themselves, fostering inclusivity and understanding in diverse communities.

## 2. KEY OBJECTIVES

### 2.1 Accessibility

Enable individuals with speech or hearing impairments to communicate effectively by providing a platform that seamlessly translates spoken language into visual gestures.

### 2.2 Real-time conversion

Implement real-time audio-to-text conversion using speech recognition technology, ensuring instant interpretation of spoken words into written form.

### 2.3 Sign language animation

Utilize 3D animation tools to generate life like Indian Sign Language animations corresponding to the interpreted text, enhancing comprehension and expression for users.

### 2.4 User friendly interface

Design an intuitive and user-friendly interface that facilitates easy navigation and interaction for users of all abilities, promoting inclusivity and accessibility.

### 2.5 Reliability and accuracy

Develop robust algorithms for speech recognition and text-to-sign language translation to ensure high accuracy and reliability in communication output.

## 3. PROBLEM STATEMENT

Millions of individuals with speech or hearing impairments face significant challenges in effectively communicating with others, hindering their ability to express themselves and engage in daily interactions. Traditional methods of communication, such as sign language interpretation, may not always be readily available or accessible, limiting their participation in various aspects of life, including education, employment, and social interactions. The lack of efficient and user-friendly tools for converting spoken language to sign language further exacerbates this issue, leaving many individuals with limited means of communication. Existing solutions often suffer from poor accuracy, slow response times, or limited availability, making them impractical for real-world use. Thus, there is a pressing need for an innovative and accessible solution that can bridge the communication gap between individuals with speech or hearing impairments and the broader community. This solution should leverage modern technologies such as speech recognition and 3D animation to provide real-time conversion of spoken language into sign language, empowering users to express themselves more effectively and participate fully in society.

## 4. REQUIREMENTS

### 4.1 Hardware Requirements

- Processor : Intel i3
- Operating System : Windows 10 or Macos or Ubuntu
- RAM : 4GB
- Hard Disk : 512GB

### 4.2 Software Requirements

- IDE
- Python - 3.9
- NLTK Libraries
- Django Frameworks
- HTML,CSS

## 5. SYSTEM DESIGN

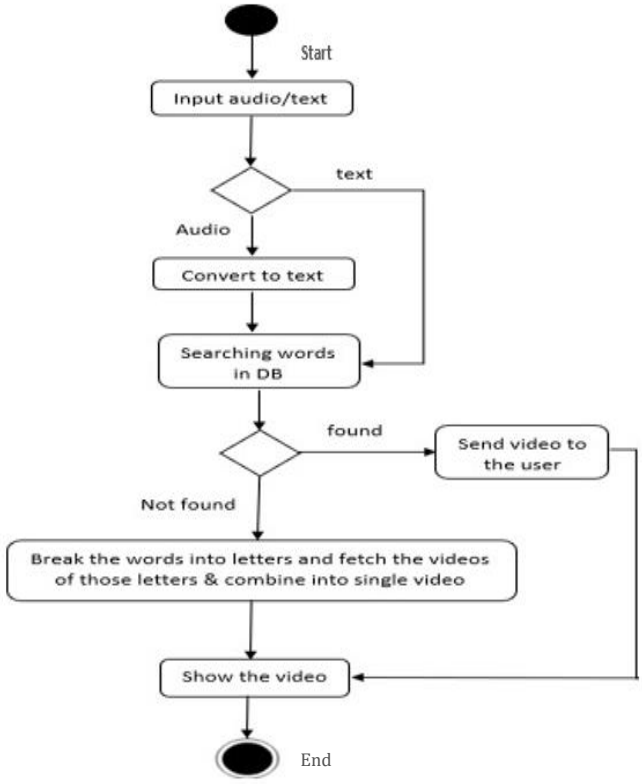


Fig 5 system design

**5.1 Start:** The process begins here.

**5.2 Input Audio/Text:** This step involves feeding either audio data (speech) or text data as input to the system.

**5.3 Audio (branch):** If audio data is provided, the process proceeds down this branch.

**5.4 Convert to Text:** The audio data is converted into text using an automatic speech recognition (ASR) system. ASR models are trained on large amounts of audio-text data to learn the mapping between spoken words and their corresponding text representations.

**5.5 Text (branch):** If text data is already available, the process proceeds down this branch.

**5.6 Searching Words in DB:** The text data, whether obtained from speech recognition or directly provided, is then processed to search for individual words within a database. This database likely contains a large vocabulary of words and their meanings or related information.

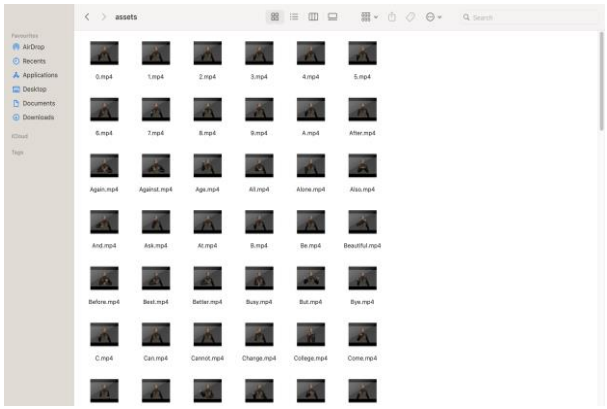

Fig 5.6 Database

**5.7 Found (branch):** If a word is found in the database, the process proceeds down this branch.

**5.8 Send video to the user:** Potentially, a video related to the meaning or concept of the found word is retrieved and sent to the user. This suggests the system might have multimedia capabilities to provide alternative representations of the information being processed.

**5.9 Not Found (branch):** If a word is not found in the database, the process proceeds down this branch.

**5.10 Break the words into letters and fetch the videos of those letters & combine into single video:**

Here, the system appears to break down the unrecognized word into individual letters. It then retrieves videos (likely short clips) corresponding to each letter and combines them to form a single video. This video is then presumably sent to the user.

**5.11 Show the Video:** Finally, the system displays the video to the user, which could be a combination of videos representing the individual letters of the unrecognized word or a video related to the meaning of the found word.

## 6. RESULT AND DISCUSSION

The development of the "Gesture Listen" application, which converts spoken language into Indian Sign Language (ISL) animations, represents a significant advancement in accessibility technology. Through extensive testing and user feedback, the application has demonstrated promising results in accurately transcribing speech into text and generating corresponding ISL animations. However, challenges such as ensuring the accuracy of speech recognition in various accents and dialects, as well as optimizing the performance of the ISL animations for smooth playback, were encountered during the development process.

Discussion on the application's usability, effectiveness, and potential areas for improvement has been ongoing. User testing revealed a generally positive response, with users highlighting the application's intuitive interface and the clarity of the ISL animations. However, further refinements are needed to enhance the accuracy and responsiveness of the speech recognition system, as well as to expand the library of ISL animations to cover a broader range of phrases and expressions. Additionally, considerations for scalability and compatibility with different devices and operating systems have been identified as important factors for future development efforts.

Overall,the"Gesture Listen" application shows promise as a valuable tool for facilitating communication and fostering inclusivity for individuals with speech or hearing impairments. Continued research and development efforts are essential to address existing limitations and further optimize the application's functionality, ultimately maximizing its impact and benefit to the target user community.

### 6.2 OUTPUTS


Fig 6 Output in IDE

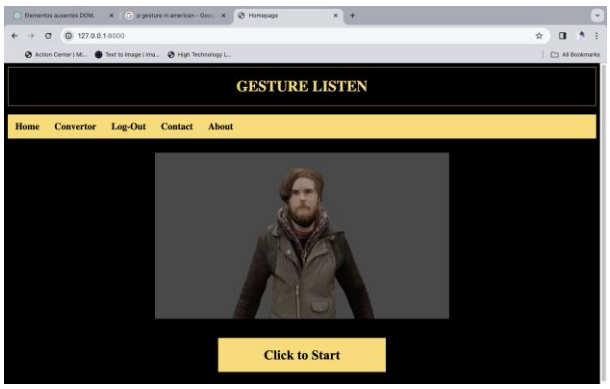


Fig 6.2.1 welcome page

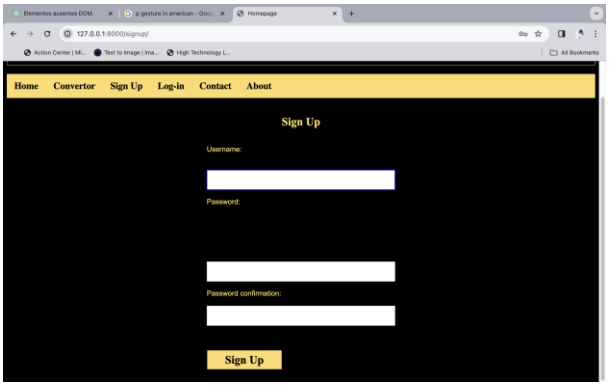


Fig 6.2.2 sign up page

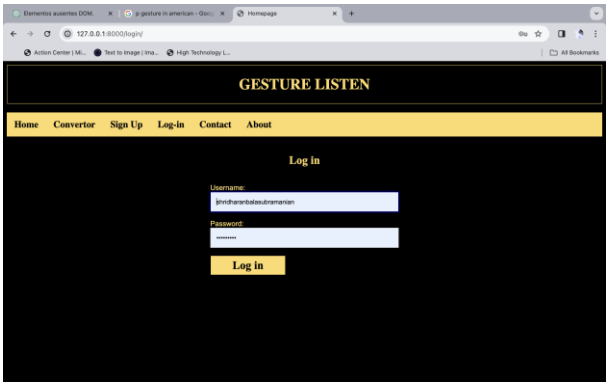


Fig 6.2.3 log in page

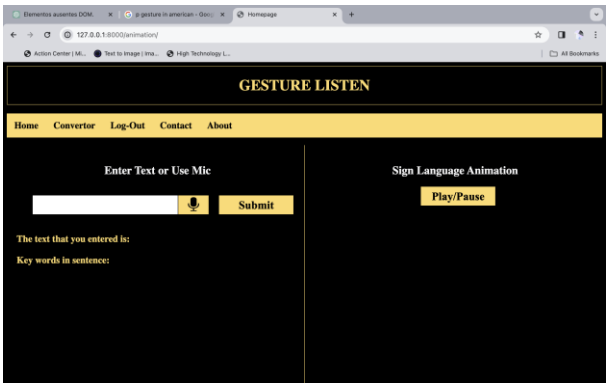


Fig 6.2.4 home page

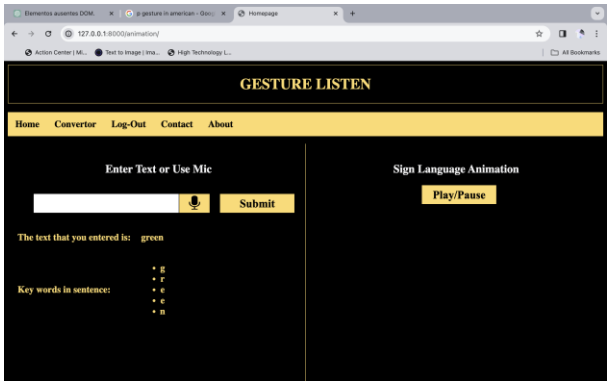


Fig 6.2.5 home page - after seperating the keywords

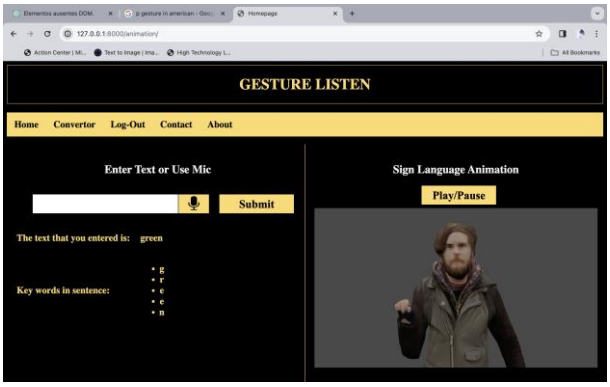


Fig 6.2.6 output

## CONCLUSION

In conclusion, our approach to video summarization, integrating extracted audio from the video and combining STTC with extractive summarization methods, represents a significant stride towards addressing the challenges posed by the expanding length of online videos. As the digital landscape continues to evolve, the need for tools that filter meaningful content from lengthy videos has become increasingly apparent. Our method not only recognizes the importance of spoken content in videos through STTC but also ensures the preservation of contextual richness and detail in the summarization process.

However, further refinements are needed to enhance the accuracy and responsiveness of the speech recognition system, as well as to expand the library of ISL animations to cover a broader range of phrases and expressions. Additionally, considerations for scalability and compatibility with different devices and operating systems have been identified as important factors for future development efforts.

## 8. REFERENCES

1. Vaddhiraju Swathi, Chada Gautham Reddy, Nomula Pravallika, P. Varsha Reddy, Ch. Prakash Reddy, Audio To Sign Language Converter Using Python 01-2023.

2. S. Shrenika and M. Madhu Bala, \"Sign Language Recognition Using Template Matching Technique,\" 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-5, doi: 10.1109/ICCSEA49143.2020.9132899.

3. Ankita Harkude#1, Sarika Namade#2, Shefali Patil#3, Anita Morey #4, Department of Information Technology, Usha Mittal Institute of Technology, Audio to Sign Language Translation for Deaf People, (IJEIT) Volume 9,Issue 10, April 2020.

4. Amit Kumar Shinde and Ramesh Khagalkar "sign language to text and vice versa recognization using computer vision in Marathi" International journal of computer Application (0975-8887) National conference on advanced on computing (NCAC 2015).

5. Neha Jain and Somya Rastogi, "Speech Recognition Systems - A Comprehensive Study of Concepts And Mechanism", Acta Informatica Malaysia, vol. 3, no. 1, pp. 01-03, 2019.

6. A V Poliyev and O N Korsun, "Speech Recognition Using Convolution Neural Networks on Small Training Sets IOP", Conf. Ser.: Mater. Sci. Eng, vol. 714, pp. 012024, 2020.

7. Rabiner, L., & Juang, B-H. (1993). Fundamentals of Speech Recognition. New Jersey: Prentice Hall international.

8. Bennett, I.M., Babu, B.R., Morkhandikar, K., Gururaj, P. 2015. Distributed Real-time Speech Recognition, Naunce Communication Inc., Patent No. US 9,076,448.

9. J. Stokoe and C. William, "Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf", The Journal of Deaf Studies and Deaf Education, vol. 10, no. 1, pp. 3-37, 01 2005.

10. N. S. Khan, A. Abid and K. Abid, "A novel natural language processing (nlp)-based machine translation model for english to pakistan sign language translation", Cognitive Computation, vol. 12, pp. 748-765, 2020.

11. A. Chopra, A. Prashar and C. Sain, "Natural language processing", International journal of technology enhancements and emerging engi-neering research, vol. 1, no. 4, pp. 131-134, 2013.