Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter

Anushri Deshmukh, Aryan Machindar, Sahil Lale, Prahshant Kasambe
Electronics Engineering Department, Sardar Patel Institute of Technology
Mumbai University, Mumbai India
anushri.deshmukh@spit.ac.in

2024 27th International Symposium on Wireless Personal Multimedia Communications (WPMC) | 979-8-3503-9231-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/WPMC63271.2024.10863135

***Abstract*— Whether it's conversing, playing computer games, attending a meeting, or participating in a video conference, deaf individuals constantly miss out on the enjoyment that other people may have. Since not every local speaks sign language, communicating with them is the hardest issue. The goal of our research is to develop a communication system for the deaf by converting words or speech into sign language. The system records voice, translates it into text, and displays Indian Sign Language actions application serves a dual function by not only translating English speech into textual representation but also into visual formats with appropriate Sign language counterpart of the word/alphabet. The uniqueness of its predictive analysis model lies in its ability to process frames over a 5-second duration, generating cumulative sentences that are synthesized into video, enhancing accessibility for a diverse user base. Furthermore, the SignLingo extends its inclusivity by providing a valuable service to individuals with visual impairments. The synergistic combination of advanced machine learning techniques with an intuitive and user-centric interface positions the sign language conversion as a pioneering initiative in creating truly inclusive communication spaces.**

Keywords — Audio Speech Recognition, Indian Sign Language (ISL), Real Time Translation, Text to Sign Conversion, Sign Language Animation, Speech to Text, Natural Language Processing (NLP), Web Speech API

I. INTRODUCTION

This research digs deep into a ground-breaking program based on an advanced system that harnesses the power of Natural Language Processing (NLP) [1]. Sign language plays a vital role in most parts of the world in reducing the communication barrier between people who are deaf or have a speech impairment. Although various implementations for converting text to sign language have been examined using English or any other foreign language, no effective models have been found to convert from text to Indian Sign Language (ISL) [2]. While great efforts have been made to advance other countries' sign languages, such as American Sign Language and British Sign Language, ISL has mainly remained under-researched.

Our effort is concentrated on ISL for the very purpose of gathering the pertinent videos for every word. In addition to the treatment of language duplication and parts of speech, substitutions are considered for those words that neither exist in the ISL dictionary nor have similar meanings. The translation factor also increases in the case of translation between spoken languages and sign language because the rules are different, and so is the grammar.

One laudable effort of this multilayered project is in the form of the application 'SignLingo,' an online app that is designed carefully to cater to the various communication needs of an auditory disabled person. This state-of-the-art concept works bi-directionally in the sense that it translates Indian Sign Language into the text as well as audio. Incorporated with ReactJS and Flask, the app promises a user-friendly interface and seamless integration for sign language conversion functionalities to be used with ease. The whole package stands out with the concept of breaking the barriers of communication in a walk toward the introduction of new horizons of inclusivity.

Our motivation for this research emerged from the numerous restrictions faced by previous systems, including their dependency on predetermined vocabularies, their inability to handle a wide range of accents, and their difficulties with real-time processing. Most significantly, they don't support ISL, which is still not well studied. By utilising voice recognition APIs [3], NLP approaches, and extensive ISL datasets, our study tackles these problems and offers a real-time, accurate, and user-friendly solution that can handle unfamiliar words and provide tailored feedback.

Further, the 'SignLingo' can convert audio/text input to video format, which will have broader outreach as it will act as a mode for the visually impaired to access written information. This platform is set up with an in-built web application that includes tools of speech-to-text and text-to-video converters to highlight the importance of the whole system [4]. Coming with the merger of capabilities, a single platform will be presented to facilitate the interaction of the deaf with ordinarily vocal individuals and, further the visually impaired. This research aims to outline what could be the potential applications of this breakthrough technology and chart a course for the future evolution of this technology. Ultimately, it aims to foster and give power to speech-impaired individuals to have more effective communication in society.

II. LITERATURE REVIEW

The field of automated conversational system application in converting speech into sign language has been adequately researched in an attempt to bridge the gap that exists between the deaf or non-verbal community and the rest of the population. This section aims to discuss a few of those studies, prominent in their ways, and how they can help one recover and make sense of the research being done on the SignLingo.

979-8-3503-9231-9/24/$31.00 ©2024 IEEE

Bharti et al. (2019) developed automatic software using Google API and Natural Language Processing to render spoken languages into sign languages [5]. The system by Bharti et al. works as follows: it first recognizes the speech, converts it to text, tokenizes it, and further matches it with a library of visual sign language videos. It concatenates that video to display the final combined video. How Bharti et al. perform is on the line of improving the communication process, as the visual outputs, in terms of spoken words, will bear a higher accuracy than the previous methods. This is particularly relevant in showing the potential of NLP and video concatenation techniques for real-world applications, which are really critical in the components of our SignLingo system. Khalid et al. (2006) developed a Sign Language Interpreter System (SSLIS) that translates spoken English into American Sign Language as a live video [6]. Using the Sphinx 3.5 speech recognition engine with the syntax of Signed English (SE) allows SSLIS to maintain English grammar to a point where more straightforward understanding is achievable. It also includes a detailed operational rule, parameter optimization, and system accuracy measurement, clearly stating its effectiveness. The ideation for this project finds that linguistic syntax needs to be upheld to be meaningful to any user, and our project maintains the same. The translated ISL needs to have a coherent structure that is preserved.

Reddy et al., 2022 have done a study on spoken Telugu translation which included a technique that involved turning spoken Telugu into animations in sign language [7]. The system captures the spoken Telugu and uses the Google API for the speech to text conversion, then compares the results to a predefined GIF dataset. The research examined the speech recognition capability of the four popular deep learning networks RNN, LSTM, Bi-LSTM and GRU that were trained on a Telugu multi speaker dataset. The Google API-based system showed promising results, with accuracy and recall scores of 94% and a word error rate of 4.1%. This displays how to effectively translate sign language using predefined visual datasets, which we have included in our system to guarantee high precision and dependability. Ashok et al. 2022 developed a user-friendly web application Speech to sign translator that converts speech/text to sign language animation and also reduces the amount of effort spent on communication. The system consists of both speech and text conversion using natural language processing (NLP) and machine learning techniques. Input speech is captured through a microphone then it is translated to text through Speech recognition using Mel-frequency cepstral coefficients (MFCC) features. The recognized text preprocessed using NLTK and resultant words looks for a word match in ASL dataset. The database contains a certain number of pre-recorded video animation signs using 3D avatar where mainly there is one video clip per each basic word. If a match occurred for a word, then the equivalent ASL translation will be displayed corresponding to that word. Otherwise, the word will be fingerspelled. This work will focus on closing the gap between time and effort in communication between disabled persons and the public at large.

The work of Kulkarni et al. in the use of 3D avatars in their system is also inspiring since it was more inspiring for the development to come up with exciting and accurate sign language animations between two-way interactions of the used convolutional neural network-based model [9]. The model should be able to translate hand postures into corresponding text and speech and, conversely, to convert text input into images of sign language. It can produce pictures of signs in real-time for conveying messages, hence no requirement for a human sign language translator. SignLingo has led to overcoming the limitations in bidirectional communication, and the system can thus be a much-required assistant in our project, the prototype of which will be much more interactive and accessible. These studies, taken together highlight some of the diverse methodologies and technological advancements undertaken in automatic speech-to-sign language translation. They all emphasize the need for solid speech recognition engines, the application of NLP in text processing, and the use of visual datasets in creating interactive, truthful representations of sign languages. Our work, built on these platforms, intends to make a complete system that enhances the accuracy, user interface, and real-time application while translating spoken language to sign language for the benefit of the Indian Sign Language community.

III. System Design

A. Data Collection Block Diagram -

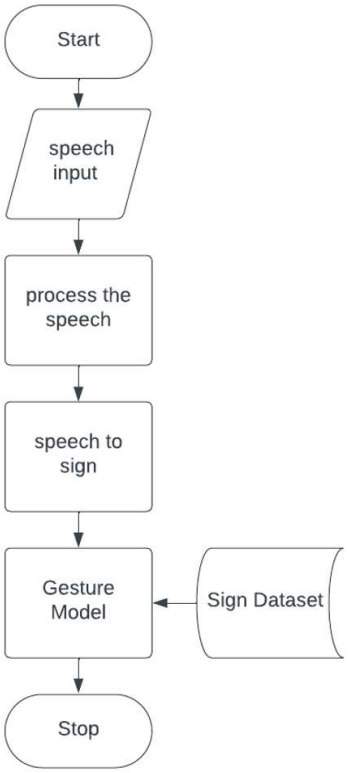

Fig – 1 Data Collection

B. Frontend Block Diagram -

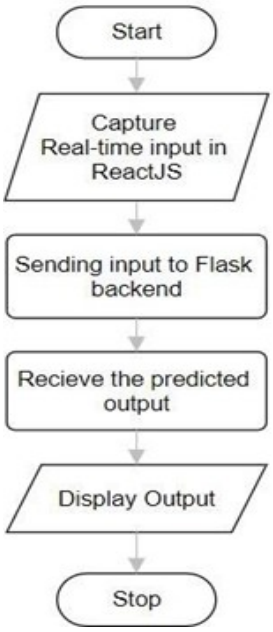

Fig – 2 Frontend

C. Backend Block Diagram -

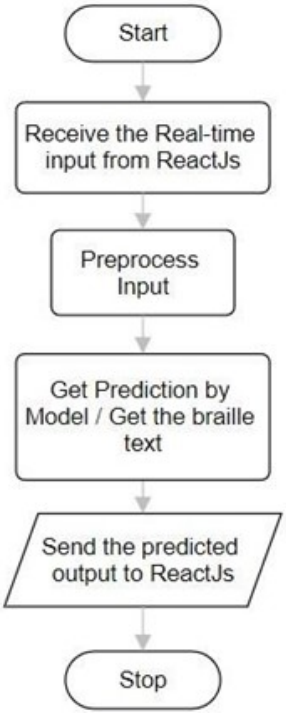

Fig – 3 Backend

The system is designed to process speech input, converting it into text and mapping words to a comprehensive database while having a backup mechanism for unrecognized words. The software system design have two primary functionalities: speech-to-text conversion and mapping words to a database consisting of associated images. The system accommodates both speech inputs, which are converted into text, and can also take direct text inputs. These inputs are then matched to the database where words are associated with images. The speech-to-text component uses speech recognition APIs or libraries to transcribe spoken words into textual format [10]. It ensures accuracy in converting spoken the words, handling various accents, and linguistic variations to produce correct text representations.

The database structure maps words with their corresponding images. This could involve various data modelling approaches, storing words as keys and linking them to their respective image files or URLs. The database should be designed for efficient storage and retrieval of word-image associations. Upon receiving input (either through speech-to-text conversion or direct text entry), the system searches the database for matching words. If a word is found, the associated video is retrieved and displayed, creating a visual representation of the word. In cases where a word isn't found in the database, the system uses backup mode. It displays the word letter by letter, providing an interactive interface for users. Simultaneously, it can showcase the image mapping process, revealing images associated with each individual letter or word to offer users a visual understanding of the missing word. The flow involves initial input processing, matching words to the database, and displaying associated images or, in the absence of matches, revealing words letter by letter through respective letter image. Users can engage with the system, potentially providing missing word-image associations to enrich the database.

Technologically, the system integrates speech recognition APIs or libraries for accurate conversion, a database management system for efficient word-image storage and retrieval, and programming languages for backend functionality and user interface development. This system aims to provide a comprehensive experience by combining speech recognition, text processing, and visual representation of words through their respective images.

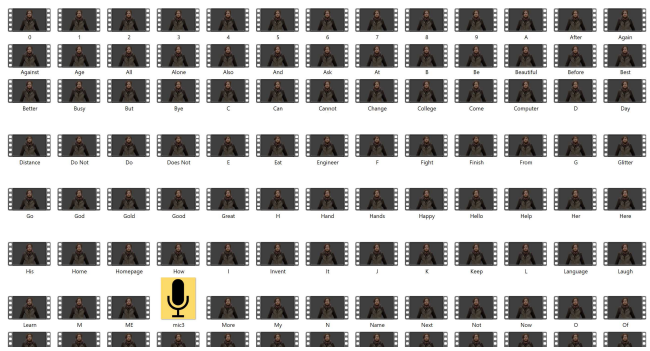

Fig 4. Screenshot of Database

Figure 4 illustrates the structure and contents of the database folder dedicated to storing videos of animated characters performing sign language gestures. The folder is organized systematically to facilitate efficient management and retrieval of these videos for the "Audio Speech To Sign Language Converter" project.

The database folder's organised structure allows for the efficient capacity and easy access of Indian Sign Language (ISL) animated videos, which is essential for real-time speech to sign conversion [11]. The method empowers user engagement and availability in sign dialect communication by improving the accessibility of ISL gestures and encouraging smooth association with the backend database administration framework.

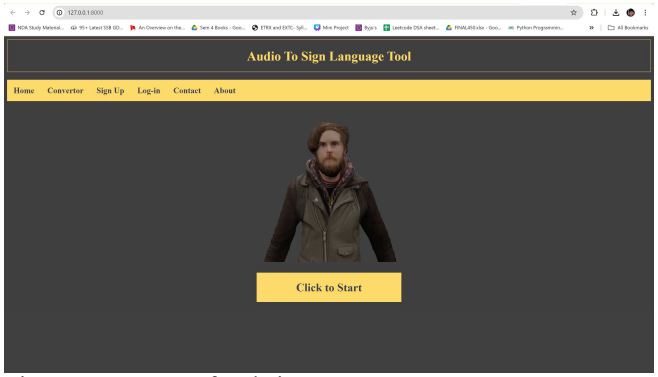

Fig 5. Homepage of website

Figure 5 shows the homepage of the web application featuring a top navigation bar with options for sign-up, login, contact, and about sections as well as our uniquely interesting animated character which is intended to make a clear user-friendly interface for users to interact with the application.

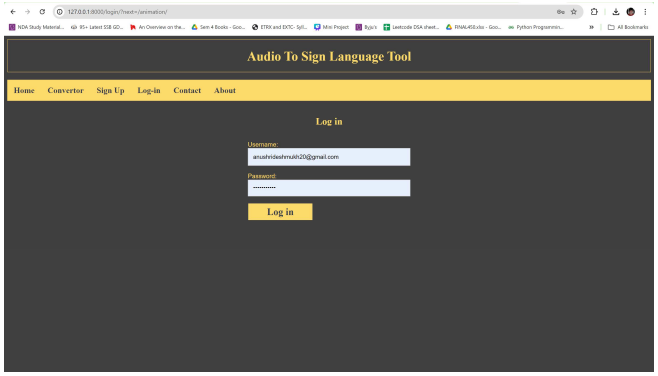

Fig 6. Login Page

Figure 6 shows the login page of the application which is aimed to perceive the user's login accreditations and personalise the user encounter within the program. The login page is important in ensuring secure access to program functionlity and individual settings. It progresses client security and information security whereas making it easier to switch to the Audio to Sign Language Converter's dynamic interface.

IV. RESULTS AND IMPLICATIONS

Our research introduces an 'Audio Speech to Sign Language Converter' application that easily consolidates discourse acknowledgement, keyword identification, and animated sign dialect creation for the hearing challenged. Utilizing the JavaScript Web Speech API, the framework precisely changes over live audio speech into text, laying the foundation for subsequent processing. The application employs Natural Language Processing(NLP) such as NLTK(Natural Language Toolkit) and WordNet lexical database to extricate critical semantic components or keywords from deciphered speech. These keyword play a important role in deciding the gestures appearing in sign language.

The animation of sign language gestures is achieved using Blender 3D, where each identified keyword triggers a corresponding sign language gesture. This integration guarantees that the animated character reliably reflects the spoken content in real time.

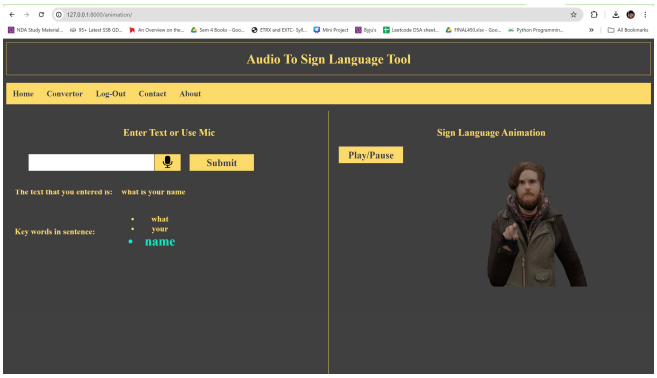

Fig 7. Screenshots of Results

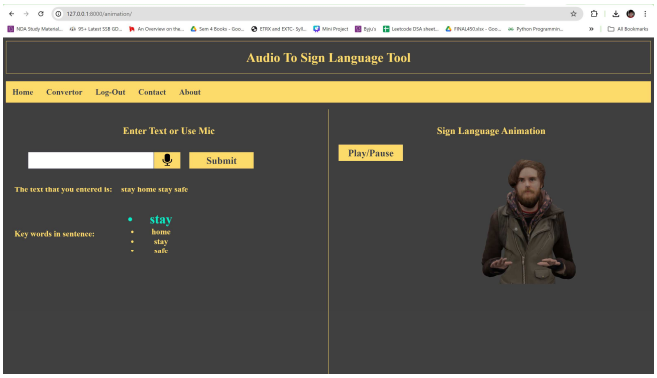

Fig 8. Screenshots of Results

Figure 7 and 8 illustrates the application interface which captures speech inputs, displaying recognized text with highlighted keywords, whereas concurrently presenting animated sign language gestures. This visual representation demonstrates the practical application of our system, which proficiently bridges the communication gap for the hearing impaired.

V. CONCLUSION AND FUTURE WORK

The 'Audio Speech to Sign Language Converter' is a vital step towards seamless communication between the hearing and hearing-impaired communities. The research gives a solid approach to interpret spoken words into Indian Sign Language(ISL) in real time by using cutting edge innovation such as Natural Language Processing(NLP) , 3D animation and voice recognition APIs. The system's user-friendly interface makes it accessible and usable for wide range of users by incorporating speech-to-text conversion, ISL grammatical standards, and animated sign creation.

In spite of its results, the research encountered a number of challenges, including ensuring that voice recognition was precise within the middle of encompassing clamor and a few lingos. To improve dependency, noise-cancelling strategies and thorough testing were utilized [12]. Besides, the constrained accessibility of ISL database required the creation of unique animations, complicating the method of creating natural and phonetically redress ISL animations. Iterative testing and input from a different assortment of

users were moreover required for building a natural and accessible interface that ensured utilization.

Addressing to these challenges and improving the applicant's functionality will be primary goals of future work. An essential goal is to progress voice recognition precision by using progressively complex machine learning approaches and differing datasets. Versatile learning algorithms, which adjust recognition to particular users, will improve performance. The innovation will be enhanced further by collaborating with linguists and sign language experts to extend the ISL animation video collection and investigating motion capture technologies for more realistic animations.

Future modifications will incorporate language support to reach a wide range of people, as well as customizable user interfaces with themes and accessibility options to improve accessibility and user satisfaction. Artificial intelligence-based relevant awareness and real-time interpretation will upgrade dynamic communication [13]. Moreover, integrating support for additional sign languages will transform the application into flexible tool that can be used anywhere in the world. There are moreover exciting opportunities to improve the application's adequacy and reach by creating a portable mobile application and collaboration with other programs.

In conclusion, this research not only fills a critical void in the community of hearing impaired, but it moreover lays a foundation for future progresses in assistive innovation. The objective of seamless communication for all can be accomplished by enhancing and expanding this research.

REFERENCES

[1] M. Sen and R. Rajkumar, "Fostering Inclusive Communication: A Tool Integrating Machine Translation, NLP, and Audio-to-Sign-Language Conversion for the Deaf," 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE), Bangalore, India, 2024, pp. 1-6

[2] N. Chandarana, S. Manjucha, P. Chogale, N. Chhajed, M. G. Tolani and M. R. M. Edinburgh, "Indian Sign Language Recognition with Conversion to Bilingual Text and Audio," 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA), Mumbai, India, 2023, pp. 1-7

[3] P. Yadav, T. Tugnait and S. K. Dubey, "Analysis of Personalized AI Assistant with Facial Recognition and Voice Representation," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 57-64

[4] A. D. Shetty, J. Shetty, K. K, Rakshitha and S. S. B, "Real-Time Translation of Sign Language for Speech Impaired," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 570-575

[5] R. Bharti, S. Yadav, S. Gupta, and R. B., "Automated Speech to Sign Language Conversion using Google API and NLP," Proceedings of the International Conference on Advances in Electronics, Electrical & Computational Intelligence (ICAEEC), Allahabad, India, Jun. 2019, pp. 1-9

[6] Khalid El-Darymli, O. O. Khalifa, and H. Enemosah, "Speech to Sign Language Interpreter System (SSLIS)," in Proceedings of the IEEE International Conference of Computer and Communication Engineering (ICCCE'06), Kuala Lumpur, Malaysia, 2006, pp. 1-6

[7] B. R. Reddy, D. S. T. Reddy, S. P. M. C., and S. Vekkot, "Creation of GIF dataset and implementation of a speech-to-sign language translator in Telugu," in Proceedings of the 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), Nov. 2022, pp. 1-6.

[8] N. Ashok, A. K. S., N. B. P., S. Suresh, and S. K. Suresh, "Speech to Sign Translator," International Journal of Creative Research Thoughts (IJCRT), vol. 10, no. 4, pp. 1-11, 2022

[9] Kulkarni, Nilesh, Ishan Misra, Shubham Tulsiani, and Abhinav Gupta. "3d-relnet: Joint object and relational network for 3d prediction." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2212-2221. 2019.

[10] H. Mohyuddin and D. Kwak, "Automatic Speech Recognition in Diverse English Accents," 2023 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2023, pp. 714-718

[11] J. P, G. B. A, H. A and K. G, "Real-Time Hand Sign Language Translation: Text and Speech Conversion," 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), Kollam, India, 2024, pp. 488-493

[12] M. Al-Qurishi, T. Khalid and R. Souissi, "Deep Learning for Sign Language Recognition: Current Techniques, Benchmarks, and Open Issues," in IEEE Access, vol. 9, pp. 126917-126951, 2021

[13] Y. A. Mohamed, A. Khanan, M. Bashir, A. H. H. M. Mohamed, M. A. E. Adiel and M. A. Elsadig, "The Impact of Artificial Intelligence on Language Translation: A Review," in IEEE Access, vol. 12, pp. 25553-25579, 2024