Grenze International Journal of Engineering and Technology, Jan Issue

Audio and Text to Indian Sign Language Converter using NLP Techniques

Pranav Agwan[1], Sahil Kirpane[2] and Madhuri Sahu[3]
[1-3]Department of Artificial Intelligence and COE AIML, G.H. Raisoni College of Engineering, Nagpur, India
Email: agwanpranav123@gmail.com , sahilkirpane2002@gmail.com, madhuridec9@gmail.com

***Abstract*— *The* "Audio and Text to Indian Sign Language Converter" is a transformative technological endeavor aimed at bridging communication barriers between the dumb people and the general population. Leveraging advanced Natural Language Processing techniques such as text recognition and audio recognition, in conjunction with visual representation techniques such as 3d animation using Blender 3d tool, this project offers a dynamic solution for real-time translation of both text and audio inputs into engaging Indian Sign Language animations. This system is delivered through a user-friendly web application interface, granting users the ability to convert single words or sentences into 3D sign language animations. Beyond its immediate utility in facilitating communication, this system holds vast potential in diverse fields, including healthcare, sign language education, and research & development, marking it as a pivotal advancement in the realm of assistive technology and language accessibility.**

***Index Terms*— Natural Language Processing, Indian Sign Language, Blender 3D, Audio Recognition and Text Recognition**

I. INTRODUCTION

The field of technology has witnessed remarkable advancements in the realm of accessibility and communication. One such field is Natural Language Processing (NLP). NLP is one of the applications of Artificial Intelligence that gives computers or system the ability to interpret, manipulate and comprehend human language or natural language and respond accordingly. Some of NLP’s main applications are audio recognition and text recognition.
Audio recognition is the method of NLP in which a computer or a system can understand and interpret natural human language and respond according to the user’s given input. With the rise of technology and artificial intelligence (AI), Audio recognition has gained a lot of attention in the current era. Some of the applications of audio recognition are smart assistants such as Amazon’s Alexa and Apple’s Siri.
Text recognition is one of the fields of natural language processing (NLP) which is the process of identifying the written text from images, handwritten notes or scanned documents and converting them into machine-readable and editable text.
This project is the combination of all the above mentioned technologies i.e., conversion of the audio and text to Indian sign language animations. Indian Sign Language (ISL) is a hand sign language used for the purpose of communication with other people. They are generally used by dumb and deaf community. There are various platforms where a person can learn Indian sign language such as google, YouTube, etc. But you need to search a lot for a particular sign language on these platforms and you also need to search different websites for different

Grenze ID: 01.GIJET.10.1.442

signs. If we get a platform where we can find ISL in the form of single words or sentences based on our input then it will be a better option.
The main objective of our system is to create a platform which converts written words or audio input into visual representations (3D animations) through Indian Sign Language. The animations are made using the Blender 3D tool. This bridges the communication gap between spoken language and Indian Sign Language, enhancing inclusivity for the hearing-impaired community.
The significance of this project extends beyond mere communication facilitation. It marks a pivotal advancement in the realm of assistive technology and language accessibility. The transformative impact is not limited to the hearing-impaired community but has broader implications for society at large. The application of NLP techniques plays a crucial role in the accurate interpretation of both text and audio inputs, ensuring that the resulting sign language animations effectively convey the intended message. The utilization of Blender 3D as the animation tool adds a layer of depth to the visual representations. The dynamic nature of 3D animations enhances the expressiveness and clarity of the ISL translations, making them engaging and comprehensible.
The potential applications of this technology are diverse, ranging from everyday communication to specialized fields such as healthcare. In healthcare, the system can facilitate effective communication between healthcare providers and patients with hearing impairments, improving overall healthcare accessibility. Education is another domain where the system can make a significant impact. It serves as a valuable tool for learning Indian Sign Language, catering to both the hearing-impaired community and individuals interested in acquiring sign language skills. The real-time environment in which the work can be done further enhances its educational utility, providing an interactive and immersive learning experience.

II. Related Work

In this section, we have discussed several published papers relevant to the subject of this study. These research articles were sourced from a range of reputable publications available on online databases such as IEEE and ResearchGate. The majority of the studies we gathered are dated between 2015 and 2022.
Ref. [1] The authors proposed a human interface for communication between dumb and deaf community. They used Kinect 360 for motion data, Unity 3D for animation and created an Android app which achieved a 91% success rate with some speech to text limitations.
Ref. [2] The creators proposed a system that converts six different Indian languages into ISL motions. This system uses MFCC and GMM for speech recognition and uses LSTM for text translation which achieved over 80% accuracy in each regional language.
Ref. [3] focuses on a translation system that converts English audio to English text, analyzes it for structural grammar, applies Indian Sign Language rules, removes stop words, and transforms words to their root form. It utilizes a dictionary with videos for word lookup and selects synonyms when a term is not found.
Ref. [4] aims to create a 24 x 7 portable system for bidirectional translation between voice and sign language using a mobile app. This system is generally made for Argentinian Sign Language. Hand gestures are recognizedand converted to text or 3D animations. The system used RNN and CNN, achieving high accuracy rates of 97% for sign-to-speech and 95% for speech-to-sign conversion.
Ref. [5] focuses on creating software to bridge the educational resource gap for individuals with hearing impairments. The software converts text into motions, aiding both those with hearing loss and hearing individuals in understanding content through translation, with a focus on Bangla Sign Language.
Ref. [6] introduces a program that translates speech into Indian Sign Language (ISL) by first converting speech to text using an API and then applying Machine Learning to segment and interpret the text into sign Ref. representations. This process facilitates communication for individuals using ISL.
Ref. [7] presents an online platform designed for individuals with hearing impairment, featuring speech-to-text conversion, language processing, and Indian Sign Language (ISL) generation. The accuracy ratefor this model is 97.86%, enhancing accessibility for thoseusing ISL.
Ref. [8] utilizes central moments for feature extraction, valuable for image analysis, and HU's moments for shape analysis. This model uses Support Vector Machine (SVM)and neural networks. The model recognizes the sign language, offering robust techniques for pattern and objectrecognition.
Ref. [9] The Bag of Visual Words model (BOVW) is employed for Indian sign language alphabet and digit recognition. The model utilizes SURF features and different classification techniques. CNN and SVM are used for training of the model. This system offers effective methods for various applications, including object recognition and 3D reconstruction.

Ref. [10] takes a unique approach, utilizing Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) descriptors for feature extraction. It employs a K-Nearest Neighbor (K-NN) Classifier, a versatile machine learning algorithm commonly used in tasks such as pattern recognition and computer vision.
Ref. [11] The authors proposed a real time sign language recognition system. They used CNN on the Massey University Dataset. This system detects real time hand gestures and translates the American sign language into text for communication purposes. This model achieved 98.05% accuracy.
Ref. [12] gives a study for American Sign language (ASL) and gives a recognition system for hand gesture capturing using motion controllers. This system uses Backpropagation and multi-layer perceptron for training of themodel. This model achieved 96.15% accuracy.
Ref. [13] gives us the application of Recurrent Neural Networks (RNN) which recognizes Arabic Sign Language. This system generally recognizes static images of hand gestures using the RNN model. The accuracy rate for the system is 95%.
Ref. [14] shows the importance of sign languages and provides a sign language system for Indian Sign Language (ISL). This is a real time system which uses Viola-Jones algorithm. The creators develop an android application which recognizes gestures and converts it into text for communication. This system is very useful for dumb and deaf community.
Ref. [15] expresses the challenges faced by dumb and deaf people while communicating. The authors created a real time system which uses Hybrid-CNN to recognize hand gestures for Indian sign language and convert them into text and speech. The model also used Multi-Headed BERT for ensuring grammatical accuracy.

III. PROPOSED METHODOLOGY

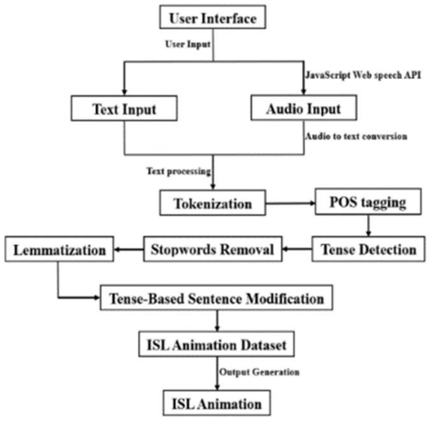

Figure 1. System Architecture

The system consists of a simple user interface where user need to sign up. You need to sign up only one time, then whenever you open the website, you will be redirected to our converter system. We have used Django for storing the information of the users in our database. Django is a high-level web framework for Python, which comes with its inbuilt object-relational mapping (ORM) system for database management. This ORM simplifies database interactions by allowing developers to use Python code instead of SQL queries. After signing up user will be redirected to the system. Django supports various database backends, including PostgreSQL, MySQL, SQLite, and Oracle.
On the interface the user can give both audio and text input. The audio input uses Java web speech API for recognizing the user's audio. Java Web Speech API is a framework that facilitates audio recognition in web applications. It allows developers to integrate speech recognition capabilities into their web projects, enabling users to interact with the application through spoken commands. This API leverages advanced speech processing algorithms to convert spoken language into text, providing a hands-free and intuitive user experience. In the backend, audio is converted to text. After the backend receives the text whether in the form of text or converted from audio, it processes it using NLP Techniques. NLP uses natural language tool kit (NLTK), a library of NLP used for performing text processing operations. The NLP techniques are as follows:

A. Tokenization

The input text is first converted into lower case for the purpose of uniform processing. After then we perform tokenization on the input sentence. Tokenization is the method of breaking the given sentence in the form of tokens or individual words. With the help of word_tokenize function we can tokenize input text. Using tokenization system can process the input easily.

B. Part-of-Speech (POS) Tagging

In a sentence, all individual words consist of grammatical categories such as nouns, pronouns, etc. For system to understand each part of the system in brief we apply POS tagging to the tokenized text. We use pos_tag function on the tokenized words to perform part-of-speech tagging. This function assigns a grammatical category such as verb, noun, adjective, etc. to each word in the sentence and the result is stored in a variable.

C. Tense Detection

ISL rules don't have different signs for different tense. Signs are same for every tense just there is some modification done to the signs based on what kind of tense the sentence consists of. We use tense detection for this purpose. In a given sentence the tense is generally detected from the verb present in it. POS tagging helps to generate tags for all types of verbs present in the sentence. If we have a future tense sentence then the POS tagging will give the tag "MD". If we have a present tense then the tags given are "VBP", "VBZ" and "VBG".
For past tense we have the tag "VBD" and "VBN". Based on these tags, system detects the tense of the sentence.

D. Stopwords removal

Stopwords are the words that individually don't have any meaning such as "is", "the", "and", etc. In the python code we define a list of stopwords. System removes these words while processing. In ISL, the signs of these stopwords are omitted while communicating. Even there are no sign available for these stopwords.

E. Lemmatization

The process of reducing any given word into its short form or root form is called lemmatization. For ex- "Waiting" will get reduced to "wait". The system iterates through the words and applies lemmatization using NLTK's WordNetLemmatizer function and reduces the words and also filters out the stopwords. The words are additionally lemmatized based on their POS tags.

F. Tense-Based Sentence Modification

As discussed before, that there are no separate signs for different tense in a sentence, we modify the input sentence to satisfy ISL rules. The sentence is modified as follows:

- If the tense is past and there's at least one past tense word, the system adds the word "Before" to the beginning of the sentence. For ex – If the input is "He was hungry", then the system will change it to "Before he hungry". The word "Was" is omitted as it is a stopword.
- If the tense is present and there's at least one present tense word, the system adds the word "Now" to the beginning of the sentence. For ex – If the input is "He is running", then the system will change it to "He is run". The word "Is" is later omitted as it is a stopword.
- If the tense is future and there's at least one future tense word, the system adds the word "Will" to the beginning of the sentence. For ex – If the input is "He will run", then the system will change it to "Will he run". Here "will" comes in the front.

After applying all the NLP techniques, the system searches for the 3D animations for the words in the dataset. The dataset is made using the Blender 3D tool. It is an open-source tool that helps user to design, model, animate and render three-dimensional content. It is used in fields such as animation, visual effects, etc.
If the corresponding animation of the processed text us found then the system renders it to the user on the main screen where the user can play and pause the animation. If an animation is not found, the system splits the words into individual letters or characters and the animation of each character is shown to the user. In this way users can get ISL animations for words and sentences.

IV. RESULT

The system correctly takes the input and renders the animation to the users. Following are some of the images presenting the user-friendly converter system. Figure 2 shows the user interface of our website. You can see the animation of the word "Hello". You need to push the "Click to Start" button. Then you will be redirected to the converter which is given in Figure 3. In the left portion you can see the audio and text input. On the right portion user will get the rendered animation.

Figure 2. User Interface

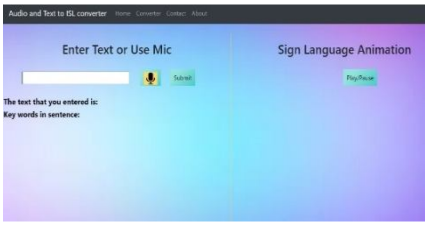

Figure 3. ISL Converter

Figure 4. Audio and Text Input

Figure 4 shows the close view of the audio and text input portion. Either you can give text input by clicking on the textbox or you can give audio input by clicking on the mic button.

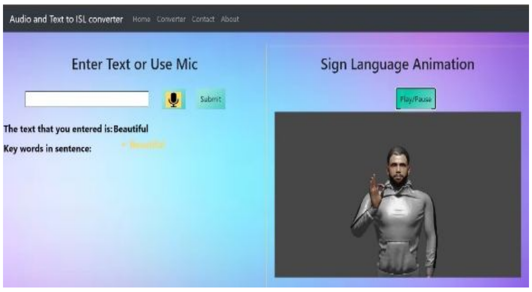

Figure 5. Animation for the word "Beautiful"

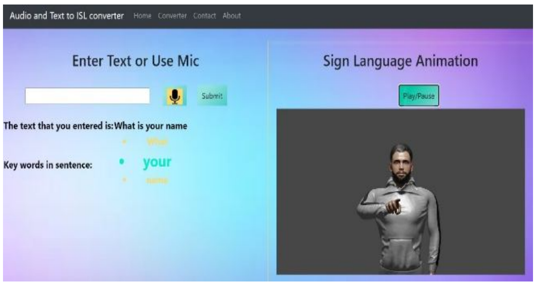

Figure 6. Animation for the sentence "What is your name"

Figure 5 shows animation for the single word "beautiful". Figure 6 shows animation for a whole sentence "What is your name". These were results for the normal words and sentences. Let's see for different tense.

Figure 7. Animation for the sentence "He came yesterday" Figure 8. Animation for the sentence "He is coming"

Figure 7 shows the animation for the sentence "He came yesterday". As discussed before in the methodology there are no different signs for different words in ISL. We modify the words based on the tense present in it. This is why the modification of "He came yesterday" becomes "Before He come yesterday".
Figure 8 shows the animation for the sentence "He is coming". Since this is a present continuous tense. So, the system will modify it and give the output as "Now he come". This is based on ISL rules.

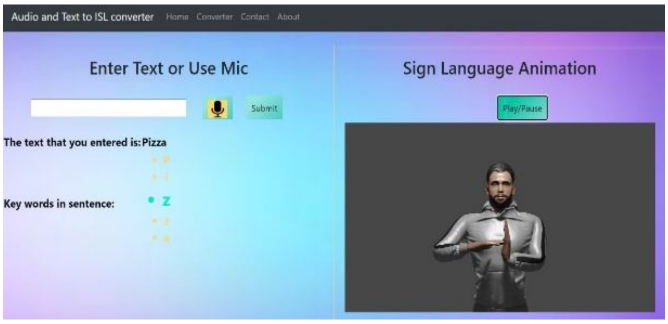

Figure 9. Animation of word not present in data

Figure 9 represents the animation of the word not present in the data. The word "pizza" is not present in the dataset so the system tokenizes it and renders animation of the individual characters 'p', 'i', 'z', 'z', 'a'. Here concludes all type of possibilities of input the user can give to the system. For now, since the dataset contains only limited words, the system is able to generate animations for the known word only. We can further make and add animation to the dataset to cover maximum words present in the ISL dictionary.

V. Conclusion And Discussion

We developed a simple website or system for Indian Sign Language (ISL), which takes text or audio as an input and then using NLP techniques it converts the sentences into its corresponding 3D animation in ISL. The work can be done in a real time environment. This system can be used by all types of people such as dumb and deaf community, normal people, etc, for communication and learning Indian sign language. Beyond its immediate utility in facilitating communication, this system holds vast potential in diverse fields, including healthcare, sign language education, and research and development, marking it as a pivotal advancement in the realm of assistive technology and language accessibility. Following is the future scope for the system:

- The word library can be expanded to include most of the commonly used words in English language as well as technical words and others which are available.
- Speech to text conversion can be made more accurate and text processing can be optimized using various NLP algorithms which are available.
- Website UI can be improved and new functionalities can be added.
- We can add various different languages from our side, so the user can give the input in their own language and can generate ISL corresponding to it.
- We can add an interface where user can show live sign language to the system and it will convert it into sentences.

References

[1] Sonawane, P., et al., 2021, "Speech To Indian Sign Language (ISL) Translation System," 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, doi: 10.1109/ICCCIS51004.2021.9397097.

[2] Peguda, J., et al., 2022, "Speech to Sign Language Translation for Indian Languages," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, doi: 10.1109/ICACCS54159.2022.9784996.

[3] Monga, H., et al., 2021, "Speech to Indian Sign Language Translator," Volume 39: Recent Trends in Intensive Computing, doi: 10.3233/APC210172.

[4] Kanvinde, A., et al., 2021, "Bidirectional Sign Language Translation," 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, doi: 10.1109/ICCICT50803.2021.9510146.

[5] B. Sarkar et al., 2009, "A Translator for Bangla Text to Sign Language," 2009 Annual IEEE India Conference, Ahmedabad, India, doi: 10.1109/INDCON.2009.5409449.

[6] Jadhav, K.; Gangdhar, S. and Ghanekar, V., 2021, "Speech to ISL (Indian Sign Language) Translator," International Research Journal of Engineering and Technology (IRJET) (Volume: 08 Issue: 04, Page 3696- 3698).

[7] Kulkarni, A., et al., 2021, "Speech to Indian Sign Language Translator," 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021), doi: https://doi.org/10.2991/ahis.k.210913.035.

[8] Rokade and Jadav, 2017, "Indian Sign Language Recognition System," International Journal of Engineering and Technology 9(3S):189- 196, doi: 10.21817/ijet/2017/v9i3/170903S030.

[9] Katoch, S.; Singh, V. and Tiwary, U. S., 2022, “Indian Sign Language recognition system using SURF with SVM and CNN,” Array 14 (2022): 100141, doi: https://doi.org/10.1016/j.array.2022.100141.
[10] Gupta, B.; Shukla, P. and Mittal, A., 2016, “K-nearest correlated neighbor classification for Indian sign language gesture recognition using feature fusion,” 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, doi: 10.1109/ICCCI.2016.7479951.
[11] M. Taskiran, M. Killioglu and N. Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning," 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, 2018, pp. 1-5, doi: 10.1109/TSP.2018.8441304.
[12] D. Naglot and M. Kulkarni, "Real time sign language recognition using the leap motion controller," 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2016, pp. 1-5, doi: 10.1109/INVENTIVE.2016.7830097.
[13] M. Maraqa and R. Abu-Zaiter, "Recognition of Arabic Sign Language (ArSL) using recurrent neural networks," 2008 First International Conference on the Applications of Digital Information and Web Technologies (ICADIWT), Ostrava, Czech Republic, 2008, pp. 478-481, doi: 10.1109/ICADIWT.2008.4664396.
[14] Swamy, Shanmukha, M. P. Chethan, and Mahantesh Gatwadi. "Indian sign language interpreter with android implementation." International Journal of Computer Applications 97.13 (2014).
[15] Gangadia, Disha, et al. "Indian sign language interpretation and sentence formation." 2020 IEEE Pune section international conference (PuneCon). IEEE, 2020.
[16] Venuthurumilli, V.P., Kargirwar, H., Shah, K., Sahu, M. “Mental Illness Prediction using Myers Briggs Personality Indicators” 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, 2023-June, pp. 89–94
[17] Bhave, R., Thakre, B.P., Kamble, V., Gogte, P., Bhagat, D. “BMSQABSE: Design of a Bioinspired Model to Improve Security & QoS Performance for Blockchain-Powered Attribute-based Searchable Encryption Applications” International Journal on Recent and Innovation Trends in Computing and Communication, 2023, 11, pp. 527–535
[18] Agarwal, M., Kotecha, A., Deolalikar, A., ...Yadav, R.K., Thomas, A. 2023 “Deep Learning Approaches for Plant Disease Detection: A Comparative Review” IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2023, 2023
[19] Dsouza, F., Bodade, A., Kolhe, H., Chaudhari, P., Madankar, M. 2023 "Optimizing MRC Tasks: Understanding and Resolving Ambiguities” 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing, PCEMS 2023, 2023
[20] Sahu, M., Dhawale, K., Bhagat, D., Wankkhede, C., Gajbhiye, D. “Convex Hull Algorithm based Virtual Mouse” 14th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2023, 2023, 2023-June, pp. 846–851
[21] Bobde, A., Narnaware, V., Tawari, S., Thomas, A. 2022 “Robust Time Series Forecasting via Time Differencing and Stacking” International Conference on Smart Generation Computing, Communication and Networking, SMART GENCON 2022, 2022
[22] Agarwal, M., Kalia, R., Bahel, V., Thomas, A. 2021 “AutoEval: A NLP Approach for Automatic Test Evaluation System” IEEE 4th International Conference on Computing, Power and Communication Technologies, GUCON 2021, 2021
[23] Bahel, V., Malewar, S., Thomas, A. “Student Interest Group Prediction using Clustering Analysis: An EDM approach” Proceedings of 2nd IEEE International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2021, 2021, pp. 481–484, 9410741
[24] Nirale, P., Madankar, M. 2021 “Analytical Study on IoT and Machine Learning based Grading and Sorting System for Fruits” International Conference on Computational Intelligence and Computing Applications, ICCICA 2021, 2021
[25] Madankar, M., Chandak, M.B., Chavhan, N. "Information Retrieval System and Machine Translation: A Review" Physics Procedia, 2016, 78, pp. 845–850