

# Harnessing AI to Generate Indian Sign Language from Natural Speech and Text for Digital Inclusion and Accessibility

Parul Yadav<sup>1</sup>, Puneet Sharma<sup>2</sup>, Pooja Khanna<sup>3</sup>, Mahima Chawla<sup>4</sup>, Rishi Jain<sup>5</sup>, Laiba Noor<sup>6</sup>  
Computer Science and Engineering Dept. Institute of Engineering and Technology, Lucknow, UP 226021, India<sup>1,4,5,6</sup>  
School of Computer Science and AI, SR University, Warangal, Telangana, 506371, India<sup>2</sup>  
Amity School of Engineering and Tech. Amity University, Lucknow, UP 226028, India<sup>3</sup>

**Abstract**—Sign language is the fundamental mode of communication for those who are deaf and mute, as well as for individuals with hearing impairments. Regrettably, there has been a dearth of research on Indian Sign Language, primarily due to the lack of adequate grammar and regional variations in such language. Consequently, research in this area has been limited. The primary objective of our research is to develop a sophisticated speech/ text-to-Indian sign language conversion system that employs advanced 3D modeling techniques to display sign language motions. Our research is motivated by our desire to promote effective communication between hearing and hearing-impaired individuals in India. The proposed model integrates Automatic Speech Recognition (ASR) technology, which effectively transforms spoken words into text, and leverages 3D modeling techniques to generate corresponding sign language motions. We have conducted a comprehensive study of the grammar of Indian Sign Language, which includes identifying sentence structure and signs that represent the tense of the subject. It is noteworthy that the sentence structure of Indian Sign Language follows the Subject-Object-Verb sequence, in contrast to spoken language, which follows the Subject-Verb-Object structure. To enhance user experience as well as digital inclusion and accessibility, the research incorporates user-friendly and simple interfaces that allow individuals to interact effortlessly with the system intuitively. The model/ system is equipped to receive speech input through a microphone/ text and provide immediate feedback through 3D-modeled videos that display the generated sign language gestures and has achieved 99.2% accuracy. Our main goal is to promote digital inclusion and improve accessibility and enhance the user experience.

**Keywords**—Sign language generation; automatic speech recognition; speech-to-indian sign language; indian sign language; digital inclusion and accessibility

## I. INTRODUCTION

The impact of information and communication technology (ICT) on human life has been immense [1]. It has revolutionized the way people conduct business, learn, travel, and communicate. Given the transformative power of ICT, it can be a valuable tool to aid the deaf and dumb community in overcoming communication obstacles [1] and bridging the gap between digital content availability and accessibility for them [2].

Hearing loss is a condition that affects an individual's ability to hear, with a hearing threshold of 20 dB or higher in both ears being considered normal [3]. It may impact one or both

ears, and its severity can range from mild to profound. The underlying causes of hearing loss include congenital or early-onset childhood hearing loss, ongoing middle ear infections, noise-induced hearing loss, age-related hearing loss, and the use of ototoxic medications that damage the inner ear. According to research, there are approximately 63 million people worldwide who experience some form of auditory impairment [4]. In India, the Indian Sign Language Research and Training Center (ISLRTC) estimates that roughly five million people are deaf or hard of hearing [5]. For many of these individuals, sign language is their primary mode of communication. However, the cost of training in sign language institutions in India can range from a few thousand to several tens of thousands of rupees, making it challenging for a significant portion of the population, including family members, acquaintances, and professionals who work with the deaf community, to learn sign language as a secondary form of communication. Unfortunately, available data suggests that only 20% of the deaf population in India have access to formal education in sign language [2]. Although sign language is a valuable tool for individuals with hearing disabilities, its limitations can hinder effective communication. These challenges include:

- **Limited Accessibility:** Sign language is not universally understood and varies across different regions and cultures, making it less accessible [6].
- **Language Barriers:** Individuals with hearing disabilities may experience challenges in communicating with those who are not well-versed in sign language. This can result in barriers to effective communication between sign language users and non-sign language users.
- **Time and Cost:** Learning sign language can be a time-consuming and costly process, deterring some individuals from pursuing it [7].
- **Dependence on Interpreters:** Relying on interpreters for communication can also limit independence and autonomy [7].
- **Limited Resources:** There is a shortage of qualified sign language interpreters in many areas, such as rural or low-income communities.

In order to improve the accessibility and efficacy of sign language for people with hearing disabilities, it is crucial to

TABLE I. COMPARISON OF EXISTING WORK

Author	Problem addresses	Methodology	Outcome
Abhisek Mishra et al.[5]	No. of hearing disables in India.	2011 census data	Around 5 mn. people have hearing disability
Sulabha M Naik et al.[4]	No. of hearing disables in India.	Rehabilitation Council of India Act,1992	Approximately 63 mn. people in India suffer from serious hearing loss.
<b>Speech to Text Conversion</b>			
Muhammad Yasir et al.[9]	Speech Recognition using web speech API	MFCC and HMM.	Avg. accuracy- 96.63% (Indonesian lang.)-82.78% (English lang.)
Santosh K. Gaikwad et al.[13]	Speech Recognition Techniques	MFCC and GMM or HMM	MFCC and HMM are best for speech recognition
S. Rajeswari et al.[8]	Speech to text conversion	Feature extraction, acoustic models ,language models and algorithms.	Voice-based e-mail system for blind.
<b>Text to Sign Generation</b>			
Navroz Kaur Kahlon et al.[7]	Text to sign language conversion	Machine Translation tech.-RBMT, EBMT	Text can be converted to sign.
ALAN CONWAY et al.[10]	Challenges of English language conversion to sign	Central blackboard control structure and Doll Control language	English to Sign system.
GARY TONGE et al.[11]	English to British Sign Language (BSL)	Human sign interpreter, Bones Animation Format and SIGML	Television broadcast System in UK.
Matthew P. Huenerfauth et al.[12]	Different sign generation system	ViSiCAST, ASL workbench, Team etc.	Symbolic representation limits Expressiveness.
<b>Tokenization and 3D modelling</b>			
Sabrina J. Mielke et. al. [14]	Explanation on Tokenization.	Studies on words, characters and sub-words	No perfect method of handling tokenization.
Zeeshan Bhatti et. al [16].	Creating 3D Animated movie.	Adobe Software	A 3D movie generated
<b>ALL</b>			
Lalit Goyal et al.[18]	Explanation on text to sign generation.	Studies on words, characters and sub-words	constructed a synthetic dictionary.
Krunal et al.[19].	converting text to sign and Creating 3D Animated signs	words and sentences	A text to sign model.
Sugandhi et al.[20]	Text to Sign Conversion	HamNoSys and SiGML	Corpus of 2950 words and 1000 sentences.
Kullami et al.[21].	converting text to sign and Creating 3D Animated signs	words and sentences	A text to sign model

employ cutting-edge and inventive technologies to overcome the challenges at hand. The other modern-day issue is; the concept of **digital inclusion** which entails removing barriers that hinder people from accessing digital content. These barriers may take the form of inaccessible systems, inadequate knowledge or skills, limited access to digital devices or materials, or other factors. As such, it is crucial to promote seamless communication between individuals with hearing impairments and those who may not comprehend sign language. Hence, **Digital accessibility** which pertains to the degree to which digital content can be easily accessed and utilized by all users, including those with visual or hearing impairments who may necessitate supplementary access provisions is crucial. This research strives to accomplish the following primary objectives:

- To close the communication divide between individuals who communicate through spoken language and those who use Indian Sign Language (ISL) by providing a system that can convert spoken language into ISL.
- To guarantee that spoken language is translated into ISL with precision and clarity, without any distortion or misinterpretation of the intended meaning.
- To establish an online platform that allows the general public to learn sign language and precisely translate spoken words into sign language.
- To develop a user-friendly interface that enables both sign language users and non-sign language users to communicate effortlessly, without encountering significant technical obstacles.

This research paper proposes and implements a novel model named as *Listen* which transforms spoken words/ text into sign language gestures and focuses on speech-to-hand-sign generation and designed with the intention to enhance communication for those with hearing disabilities. By enabling real-time translation of spoken or text input into sign language, *Listen* empowers individuals who rely on this form of communication. The output is presented through animated visuals for optimal accessibility. *Listen* model is able to achieve the accuracy of 99.21% when rigorously tested with different inputs.

The novelty of our model is its impressive capability of

comprehending the connection between English and Indian sign language sentences through the use of advanced Natural Language Processing (NLP) techniques. This involves the mapping of words to their corresponding hand signs, enabling the generation of sign language gestures in response to spoken input. Despite the intricacy of sign language, regional variations, and the requirement for accurate speech recognition, this research work has made significant progress in developing a robust system for generating hand signs with better accuracy than the existing models to the best of our knowledge.

The research paper is structured as follows: Firstly, it presents a comprehensive overview of the existing related works in the field in Section II. Following that, it discusses the methodology proposed for implementing the proposed model in Section III. Subsequently, Section IV exhibits the results and visualizations of the model. Lastly, in Section V, it provides a summary of the research work done and outlines the future work.

## II. RELATED WORKS

Numerous studies have shed light on the prevalence of auditory impairments within the Indian population. Furthermore, there have been commendable accomplishments in the deployment of speech recognition systems in meetings and email assistance [8], as well as the translation of text into sign language through various techniques. Additionally, hand gestures have been utilized for virtual mouse control. These endeavors have produced a broad spectrum of results, ranging from statistical data on disability to the development of useful applications and systems for communication and interaction.

The review highlights the specific areas that require attention. For example, speech recognition systems necessitate distinct enunciation [9], while sign language generation has lexicon constraints [6]. Avatars currently lack genuine emotional expression [7], and natural language processing encounters challenges with tokenization techniques. Furthermore, virtual mouse systems need enhancement in precision and efficiency. Papers [10], [11], [12], [13], [14], [15], [16], [17] revolve around the above stated techniques which is explained in Table I. Followed by Lalit Goyal et al.[18] in 2016 constructed a synthetic dictionary that classified ISL words. These words were then translated into the HamNoSys (Hamburg Notation System) writing notation for sign language, which was translated into SiGML (Signing Gesture Markup Language) to

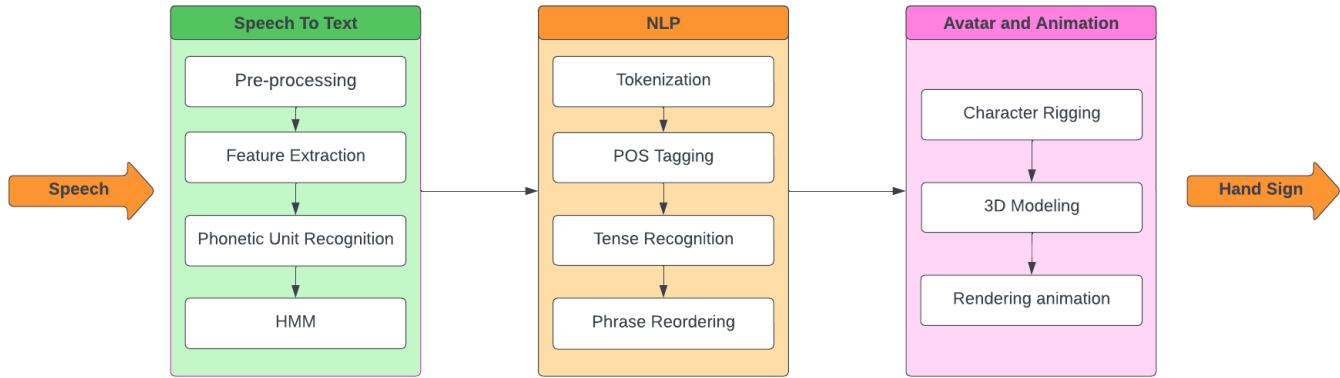


Fig. 1. Process flow of model.

produce a synthetic animation of the sign using a computer-generated cartoon with an overall accuracy of 94.35%. Krupal et al.[19] presented a method that uses speech recognition to turn an English voice dictation into text which bagged the highest accuracy of 94.53%. Followed by Sugandhi et al.[20] describes a system that generates Sign language based on Indian Sign Language grammar. It includes elements like an ISL parser, the Hamburg Notation System, the Signing Gesture Mark-up Language, and 3D avatar animation and achieved an accuracy of 96.67%. In order to translate English speech to Indian Sign Language with a 97.86% accuracy rate, Kulkarni et al.[21] presented an online platform that accepts voice as input and outputs a series of films showing the matching sign language. Hence, In the realm of assistive technologies for individuals with disabilities, a multitude of studies have explored various challenges and potential solutions. Table I provides a comprehensive compilation of related works, covering topics such as education and employment opportunities for the hearing impaired, rehabilitation options for deaf children, speech-to-text conversion systems, sign language generation, and virtual mouse systems utilizing hand gestures.

The Table I includes details regarding the authors, the issues they addressed, the methodologies employed, and the resulting outcomes. These studies employed a range of approaches, including data analysis, speech recognition techniques, language models, and 3D animation, to overcome specific challenges faced by individuals with disabilities. Additionally, the limitations of these studies have been provided in this section which is followed by Section III that explains our proposed model, its working, and corresponding components to address the objectives at hand.

### III. METHODOLOGY

Effective communication is vital to human society, and spoken language is the primary tool we use to convey our ideas and beliefs [9]. Crafting a reliable platform for converting speech into sign language requires a multitude of intricate steps, as elucidated in Section II. Bearing this in mind, we are pleased to present our innovative approach for deploying our solution, which we have aptly named *listen*. The *Listen* system consists of three distinct subsystems: Speech-To-Text pre-processing, Natural Language Processing (NLP), and the

avatar and animation. Fig. 1 provides an illustration of each subsystem's components. The model's process flow diagram, also shown in the figure, begins with the input Speech and proceeds through the Speech-To-Text pre-processing subsystem (explained in Subsection III-A), the NLP subsystem (explained in Subsection III-B), and ultimately the avatar and animation (explained in Subsection III-C), which displays the Hand sign corresponding to the input Speech. The first subsection of the model is explained in the following subsection.

#### A. Speech to Text

Speech recognition is the technology that converts spoken words into digital text, making it easily editable, storable, and shareable. The process involves analyzing and interpreting the acoustic and linguistic features of speech [8], such as words and phrases, to extract useful information. This is done by taking audio data input and performing recognition using a model. A labeled dataset is used to map sound vibrations to different letters, but some letters can be challenging to recognize due to similar pronunciations. In such cases, the probability of how commonly two letters appear together is used to determine the most likely recognition. Example:- "called" and "pan" Notice how 'a' produces different sounds in both words. In such cases, we use likelihood probability of how commonly two letters appear together.

The Fig. 2 shows the entire process of Speech recognition used in the research. The steps include pre-processing, feature extraction, phonetic unit recognition, and hidden Markov model all of which are explained below.

1) *Pre-processing*: During the course of this research, the pre-processing stage was implemented to partition the input into smaller frame sizes. In our model, each frame size is composed of 0.025 seconds of audio, equivalent to the duration of a single or a few phonemes, which are the basic units of speech, and typically have a duration in the range of tens to hundreds of milliseconds. This duration is ideal for capturing the unique characteristics of phonemes, which typically last for tens to hundreds of milliseconds. It was determined that only one English phoneme is typically pronounced within this time period. The pre-processed output is then forwarded to the feature extraction subsection.

## SPEECH RECOGNITION PROCESS

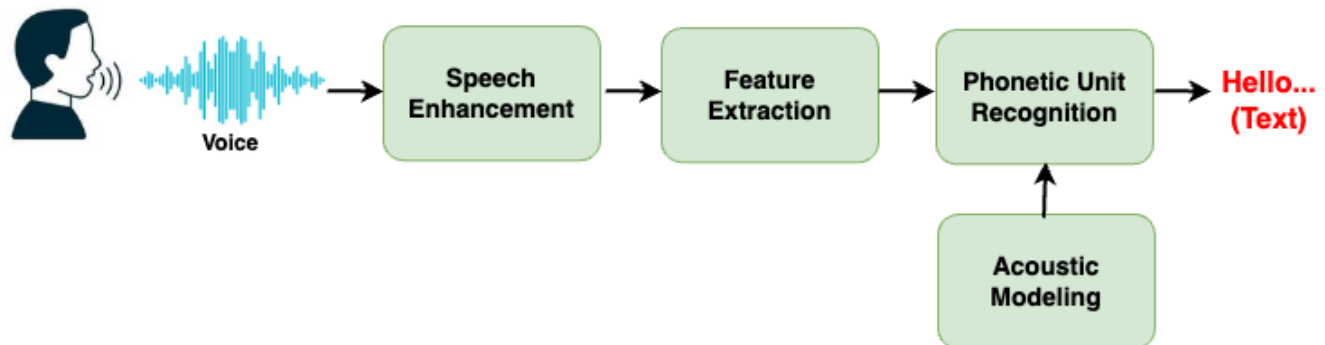


Fig. 2. Speech recognition.

2) *Feature Extraction*: In the feature extraction phase, each frame audio frequency feature is extracted. There are various techniques for feature extraction, some of them are Short-time Fourier Transform (STFT) [13], Mel Frequency Cepstral Coefficients [9], Chroma feature [22], and Spectral Contrast [23]. Our *Listen* model leverages **Mel Frequency Cepstral Coefficients** (MFCC) to accurately capture audio features. This technique mimics the human auditory system by utilizing a filter bank that simulates the Mel scale to approximate the magnitude spectrum produced by the STFT. Through a multi-step process, we transform the audio signal by first computing the logarithm of filterbank energies and then applying Discrete Cosine Transform (DCT) to obtain a concise representation of the signal. These features are then segmented into frames of equal size, with each frame representing a distinct phoneme. These frames are then passed onto the next component of the model, known as the **Phonetic Unit Recognition** (PUR) module.

3) *Phonetic Unit Recognition*: Once the features are extracted the phoneme is predicted by the Pre-trained acoustic model, which finds the most similar match of a feature of the current frame to existing labeled features in the data set followed by the **Hidden Markov Model**(HMM).

The HMMs are trained using a labeled dataset that contains speech recordings along with corresponding transcriptions or phonetic annotations. The training process involves estimating the parameters of the HMM, including the state transition probabilities. The state transition probability in simple terms represents the probability of one phoneme occurring with another phoneme. Hidden Markov Models (HMMs) are commonly used to model the temporal dependencies in speech signals. HMMs consist of several matrices that define the model's parameters [9] [13]. The outputted text is then fed to the Subsection III-B for conversion of text using Natural Language Processing(NLP) which is explained next.

### B. Natural Language Processing (NLP)

The ultimate aim of Natural Language Processing (NLP) is to enable machines to comprehend human language in

a way that is similar to how humans understand it. This involves the creation of intricate algorithms and models that allow computers to analyze, comprehend, and generate human language in a meaningful and contextual manner. NLP encompasses various techniques and tasks, including text tokenization, removal of stop words, part-of-speech (POS) tagging and phrase reordering. In this section, we will delve deeper into these techniques which are as follows:

1) *Tokenization*: Tokenization is the process of breaking down text into smaller units, or tokens, in NLP. These tokens can be words, phrases, or even individual characters, depending on the task at hand. Tokenization is a crucial step in NLP, as it helps to reduce the complexity of text, making it easier for machines to understand and analyze. This technique is used in various NLP tasks, such as sentiment analysis, text classification, and machine translation [14]. It can be performed using various libraries and tools available in Python programming language such as NLTK [15] and spaCy [24]. NLTK is used for Tokenization in our model whose working is shown in Fig. 3. Fig. 3 explains how the tokenization process works. First, the function removes any leading or trailing white space characters from the input string. Next, it splits the text into sentences using a pre-trained sentence tokenizer. This tokenizer is designed to identify the boundaries between sentences, such as periods, question marks, and exclamation points. The function then tokenizes each sentence into individual words using a pre-trained word tokenizer for example: Boundaries: white space. Finally, the function returns a list of tokens, where each token is a separate word in the text. The separated words are then inputted into the next section for further processing.

2) *Removal of Stop Words*: In the field of Natural Language Processing (NLP), the removal of stop words plays a crucial role in the pre-processing of text data. This step involves the cleaning and preparation of text data before further analysis. Stop words refer to the commonly used words in a language that do not contribute significantly to the overall

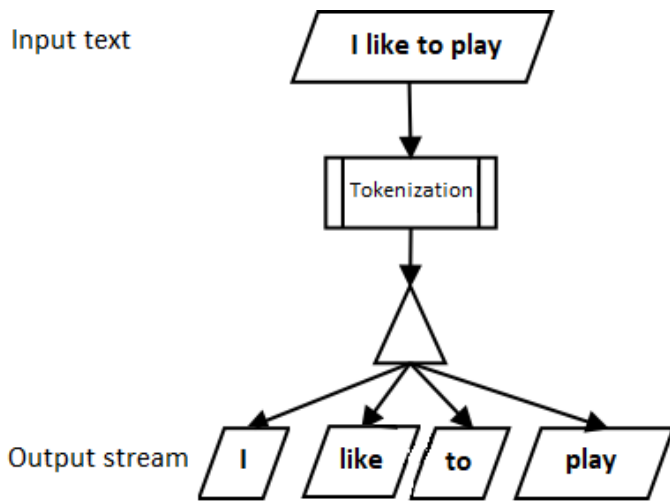


Fig. 3. Tokenization.

meaning of a sentence. Words like “a,” “an,” “the,” “is,” and “and” are examples of stop words, and their exclusion can significantly reduce noise and improve efficiency in data analysis. Other examples of stop words include “mightn’t,” “re,” “wasn’t,” “wouldn’t,” “being,” “were,” “isn’t,” “needn’t,” “don’t,” “nor,” “aren’t,” “as,” “didn’t,” “should’ve,” and “be”, among others, which are generally omitted during NLP analysis. After, Tokenization, the model obtains a list of stop

TABLE II. POS TAG EXAMPLE LIST

Word	POS Tag	Word Class
She	PRP	pronoun, personal
went	VBD	verb, past tense
to	TO	“to”
the	DT	determiner
market	NN	noun (singular)
.	.	punctuation mark

words specific to the language or domain. The model then iterates through the tokens and compares each word against the stop word list. If a word is found in the list is a stop word, remove it from the token list. After which they are fed to the POS tagging phase [6].

3) *Part-of-Speech (POS) Tagging*: It is also known as grammatical tagging, is the process of assigning grammatical information, such as nouns, verbs, adjectives, etc., to individual words in a given text as shown in the Table II. Natural Language Processing (NLP) relies heavily on one fundamental task that serves as a foundation for many downstream tasks. Specifically, the task in question is essential for syntactic parsing, information extraction, and machine translation. Fig. 4 explains how POS Tagging will assign grammatical information in a sentence for example: I love to read poems.

In NLTK (Natural Language Toolkit), a popular Python library for natural language processing, you can perform POS tagging using the ‘pos\_tag’ function. The ‘pos\_tag’ function is designed to accept an input list of tokens and generate a corresponding list of tuples. Each tuple contains a word and

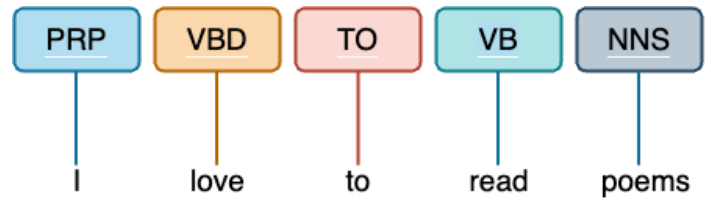


Fig. 4. POS Tagging.

TABLE III. IMPLEMENTATION DETAILS

Language	Python
Tool	Blender (3D modelling and animation)
Libraries	NLTK, spaCy
Back End	Django
Front End	HTML/ CSS, Javascript

its corresponding POS tag which can then be utilized in phrase reordering.

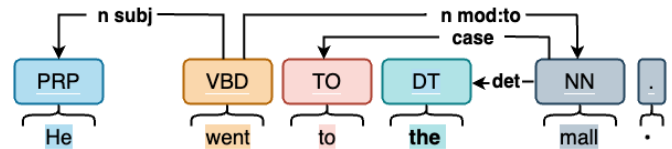


Fig. 5. Phrase reordering.

4) *Phrase Reordering*: Research has revealed that Indian Sign Language exhibits a distinct sentence structure when compared to English. Specifically, it adopts a Time-Subject-Object-Verb format, as opposed to the Subject-Verb-Object structure found in English [6]. Therefore, to effectively convey a phrase in sign language, it must first be restructured accordingly. For example, the sentence “He went to the mall” would require reordering, as illustrated in Fig. 5. Following this restructuring, we must consider the tense of the sentence and incorporate time-reference words such as Now, Before, and Later for present, past, and future tenses respectively. Once the input data has undergone speech-to-text pre-processing and NLP, it is then ready for the avatar and animation stages.

### C. Avatar and Animation

To produce 3D videos, one typically needs access to specialized software, such as Adobe [16]. However, we prefer to utilize a free and open-source alternative known as Blender [25] for our 3D avatar and animation modeling. Blender is a powerful 3D creation suite that boasts an array of features for modeling, animation, rendering, and video editing. It is a favored tool among designers, artists, and animators for creating stunning visual content, including 3D models, animations, simulations, and visual effects. The techniques that we used for avatar creation and animation includes: **Character rigging**.

which is used for making sure the Sign Language is accurately performed by the Avatar we need to give it a human-like bone structure. We designed a structure as shown in Fig. 6.

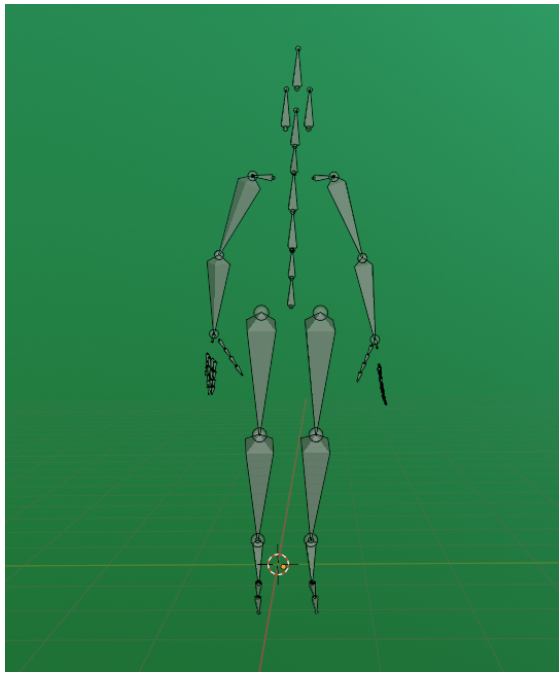


Fig. 6. Character rigging.

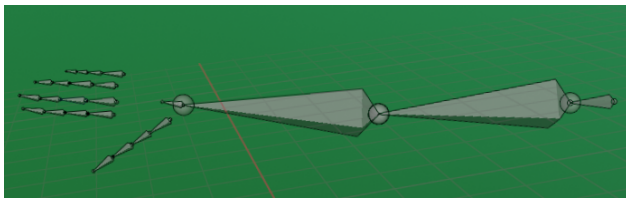


Fig. 7. Armatures.

The armatures in Fig. 7 provide our avatar with the necessary bone structure for movement. Accurate hand movements are essential for effectively conveying the inputted message, which is why meticulous attention has been given to hand rigging. In total, each hand comprises 46 armatures, including four for each finger, one for the wrist, elbow, and shoulder. Finally, the avatar is rendered in 3D to complete the process. After character rigging we jumped into 3D modeling of the obtained output from the preceding step.

In the concluding phase of **3D Modeling**, we initiated the process by fabricating a fundamental mesh for our avatar by taking advantage of Blender's modeling tools. We implemented an array of techniques, including extrusion, scaling,



Fig. 8. Avatar (Female).



Fig. 9. Avatar (Male).

and sculpting, to refine the mesh into the desired form, meticulously focusing on details such as facial features, body proportions, and clothing. As we progressed with the modeling procedure, we referred to reference images or concept art to steer our work. The ultimate outcome was a breathtaking portrayal of the female and male avatars, as displayed in Fig. 8 and Fig. 9.

TABLE IV. TENSE RECOGNITION RESULTS

English Sentence	Tense Identified	Actual Tense
She is reading a book.	Present Continuous	Present Continuous
They will arrive tomorrow.	Future Simple	Future Simple
I have eaten dinner.	Present Perfect	Present Perfect
He plays the guitar.	Present Simple	Present Simple
We had studied for the test.	Past Perfect	Past Perfect
You should go to bed early.	Present Simple	Present Simple
The party was fantastic!	Past Simple	Past Simple
It is raining outside.	Present Continuous	Present Continuous
She will have finished by then.	Future Perfect	Future Perfect

#### IV. DISCUSSION: RESULTS AND ANALYSIS

Throughout our diligent research, we carefully monitored the outcomes at various stages to guarantee precise conversion from English to Indian Sign Language. Our comprehensive results include a detailed step-by-step analysis for each phase, commencing with the speech recognition process. During this stage, the input consists of audio speech, which we process to transform it into text using our sophisticated speech recognition algorithm. We implemented the model in Python.

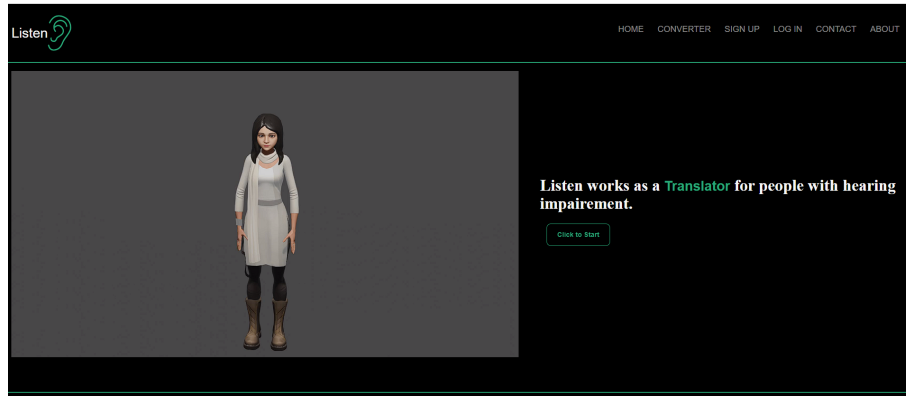


Fig. 10. Home page of Listen.

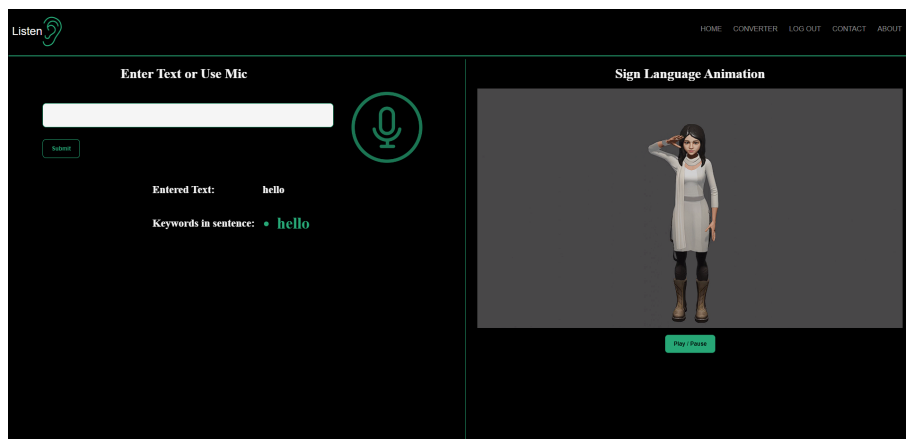


Fig. 11. Hello by Listen.

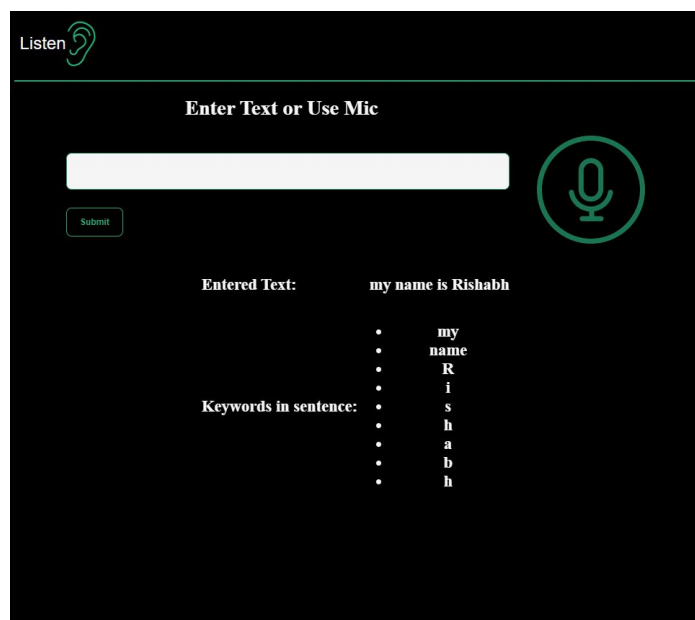


Fig. 12. Input text and keywords identification.

TABLE V. ISL SENTENCE RESULTS

English Sentence	Sentence after reordering	ISL Sentence
she has 10 books	she 10 books has	she 10 books
go to sleep	sleep go to	sleep go
I shall go after you	I after you shall go	later me after you go
I like my college	I my college like	me my college like
you are watching a movie	you movie are watching	now you movie watch
Rahul can change it	Rahul can it change	later Rahul can it change
show me your hands	me your hands show	me your hands show
Amit went alone	Amit alone went	before Amit alone go
define computer	define computer	define computer
Ram gave a flower to Sita	Ram flower gave Sita	before Ram flower give Sita
She made coffee for Shashwat	she coffee Shashwat made	before she coffee Shashwat make
we are waiting for Shashwat	we Shashwat are waiting	now we Shashwat wait

Implementation details are given in Table III which shows environment used to implement the different components of our *Listen* model. A snapshot of the home page of the model is shown in Fig. 10. Furthermore, we have included the outcomes of our conversion process from English text to Indian Sign Language sentence structure, along with a side-by-side comparison of our output and the actual sign language sentence structure as shown in Fig. 11. Additionally, we have incorporated a screenshot of the final output page, which showcases captivating hand sign animation videos. The following subsections include the results obtained of the speech-to-text conversion, tense recognition and phrase reordering followed by 3D modeling of the obtained outputs. Detailed explanations of the subsections are as follows:

#### A. Speech to Text

This section includes the snapshot of the result when the word ‘my name is Rishabh’ is pronounced and is added in Fig. 12. The entered text and keyword in the sentence both are extracted.

#### B. Tense Recognition and Phrase Reordering

This phase deals with the tense recognition and phrase reordering of the preprocessed input data. First we are identifying the tense of the sentences using POS tags this is required for adding a timeline to ISL sentences. Some examples of Tense Recognition results are shown in Table IV.

Once the tense of the sentence is recognized next step is to reorder the sentence according to the sentence structure of Indian Sign Language. The snapshot of the data set tested is shown in Table V. Note that sentences with present continuous/ past indefinite/ future tenses etc. are treated differently since we need to add ‘Now’ ‘Before’ or ‘Later’ at the beginning of such sentences for correct ISL interpretation.

#### C. 3D Animations

ISL sentence is sent to the 3D model and sign language animations were obtained for female and male avatar. Screenshot for the animation of word ‘hello’ is shown in Fig. 11.

Images/ screenshots at a particular moment of a word in 3D animations of the female and male avatars for 10 test cases (5 test cases for each avatar) are shown Table VI and Table VII.

Our proposed model, *Listen*, has made significant strides in the recognition of tense and operates with great efficiency in the following areas:

- Tense recognition
- It does not rely on pre-designed notations, such as the Hamburg Notation System (HamNoSys) [17].
- It is specifically designed to work with the Indian Sign Language, which has not been extensively studied in the past.

As a result of our research, we are able to successfully achieve our objectives. We compared the performance of our model with existing models as shown in Table VIII. The accuracy of our *Listen* model is of 99.21% and sensitivity is 98.73% which is better than the existing models in the line of research.

## V. CONCLUSION AND FUTURE WORK

In conclusion, speech-to-sign research has the potential to revolutionize communication and interaction for those with hearing impairments. By bridging the gap between the hearing communities, it promotes digital accessibility and inclusion. We achieved optimal results by implementing a robust methodology that involved data collection, pre-processing, sign language recognition, text pre-processing, text-to-sign language translation, 3D modelling of animations and evaluation of model. The system achieved accurate results that were adaptable, user-friendly, accessible, and scalable, with a diverse dataset on which the proposed model *listen* was trained. Our research focuses on converting English audio to Indian Sign Language (ISL), which presents unique challenges due to the absence of specific grammatical rules. Our model has yet to accomplish the task of animating facial expressions to denote negative and interrogative sentences. We plan to include ISL for phrases in the next phase, along with non-manual components for the sentence as a whole. Our goal is to improve



TABLE VI. IMAGES OF AVATAR (FEMALE) FOR TEST CASES











Sentence	Word	Result
She has 10 books.	10	
Go to sleep.	sleep	
I shall go after you.	after	
I like my college.	my	
You are watching a movie.	you	

TABLE VII. IMAGES OF AVATAR (MALE) FOR TEST CASES

Sentence	Word	Result
Rahul can change it.	change	
Show me your hands.	hands	
Amit went alone.	alone	
Define computer.	computer	
Thank you.	thank you	

the quality of life for individuals with hearing impairments and promote inclusiveness in society.

REFERENCES

[1] Kaur, S. and Singh, M., 2015, September. Indian Sign Language animation generation system. In 2015 1st International Conference on Next Generation Computing Technologies (NGCT) (pp. 909-914). IEEE.

[2] Lexdis: Accessible Technology For Learning, ATBar, <https://www.lexdis.org.uk/digital-accessibility/what-is-digital-accessibility-and-inclusion/>, accessed on 18 August 2023.

[3] World Health Organizations, Deafness and hearing loss, <https://www.who.int/health-topics/hearing-loss>, accessed on 05 August 2023.

[4] Kahlon, N.K. and Singh, W., 2023. Machine translation from text to sign language: a systematic review. Universal Access in the Information Society, 22(1), pp.1-35.

[5] Sugandhi, Parateek Kumar, and Sanmeet Kaur. Sign language generation system based on indian sign language grammar. ACM Transactions on

TABLE VIII. COMPARISON FOR PERFORMANCE

S. No.	Paper	Accuracy	Sensitivity
1	Lalit Goyal et al. [18]	94.35%	92.89%
2	Krunal et al. [19]	94.53%	93.45%
3	Sugandhi et al. [20]	96.67%	94.02%
4	Kulkarni et al. [21]	97.86%	94.56%
5	<b>Proposed Model</b>	<b>99.21%</b>	<b>98.73%</b>

Asian and Low-Resource Language Information Processing (TALLIP), 19(4):1–26, 2020.

[6] M Naik Sulabha, S Naik Mahendra, and Sharma Akriti. Rehabilitation of hearing impaired children in india-an update. Online Journal of Otolaryngology, 3(1):20, 2013.

[7] Abhisek Mishra, Anu N Nagarkar, and Nitin M Nagarkar. Challenges in education and employment for hearing impaired in india. Journal of Disability Management and Special Education, 1(1):35, 2018.

- [8] S Rajeswari and J Karthika. Voice based email assistance for visually impaired—a comprehensive review. 2023.
- [9] Muhammad Yasir, Marlinec NK Nababan, Yonata Laia, Windania Purba, Asaziduhu Gea, et al. Web-based automation speech-to-text application using audio recording for meeting speech. In Journal of physics: conference series, volume 1230, page 012081. IOP Publishing, 2019.
- [10] Santosh K Gaikwad, Bharti W Gawali, and Pravin Yannawar. A review on speech recognition technique. International Journal of Computer Applications, 10(3):16–24, 2010.
- [11] Tony Veale, Alan Conway, and Bróna Collins. The challenges of cross-modal translation: English-to-sign-language translation in the zardoz system. Machine Translation, pages 81–106, 1998.
- [12] MICHELE Wakefield. Visicast. Final Report, page 97, 2002.
- [13] Shah, A., Kattel, M., Nepal, A. and Shrestha, D., 2019. Chroma feature extraction. Chroma Feature Extraction using Fourier Transform.
- [14] Industrial-Strength Natural Language Processing, spaCY, <https://spacy.io/>, accessed on 30 august 2023.
- [15] Blender, Blender 3.6 LTS, <https://www.blender.org/>, accessed on 30 august 2023.
- [16] Sabrina J Mielke, Zaid Alyafeai, Elizabeth Salesky, Colin Raffel, Manan Dey, Matthias Gallé, Arun Raja, Chenglei Si, Wilson Y Lee, Benoît Sagot, et al. Between words and characters: A brief history of open-vocabulary modeling and tokenization in nlp. arXiv preprint arXiv:2112.10508, 2021.
- [17] Hanke, T., 2004, May. HamNoSys-representing sign language data in language resources and language processing contexts. In LREC (Vol. 4, pp. 1-6).
- [18] Goyal, L., Goyal, V., Development of Indian Sign Language Dictionary using Synthetic Animations, Indian Journal of Science and Technology, 9(32), 2016. <https://doi.org/10.17485/ijst/2016/v9i32/129404>.
- [19] K. Saija, S. Sangeetha and V. Shah, WordNet Based Sign Language Machine Translation: from English Voice to ISL Gloss, 2019 IEEE 16th India Council International Conference (INDICON), Rajkot, India, 2019, pp. 1-4, doi: 10.1109/INDICON47234.2019.9029074.
- [20] Sugandhi, Kumar, P. and Kaur, S., 2020. Sign language generation system based on Indian sign language grammar. ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 19(4), pp.1-26.
- [21] Kulkarni, A., Kariyal, A.V., Dhanush, V. and Singh, P.N., 2021, September. Speech to indian sign language translator. In 3rd International Conference on Integrated Intelligent Computing Communication & Security (ICIIC 2021) (pp. 278-285). Atlantis Press.
- [22] Kwarteng, P. and Chavez, A., 1989. Extracting spectral contrast in Landsat Thematic Mapper image data using selective principal component analysis. Photogramm. Eng. Remote Sens, 55(1), pp.339-348.
- [23] Natural Language Toolkit, NLTK Project, created with Sphinx and NLTK Theme, <https://www.nltk.org/>, accessed on 30 august 2023.
- [24] Matthew P Huenerfauth. American sign language natural language generation and machine translation systems. Technical report, Technical Report, computer and information sciences, University of Pennsylvania, 2003.
- [25] Zeeshan Bhatti, Ahsan Abro, Abdul Rehman Gillal, and Mostafa Karbasi. Be-educated: Multimedia learning through 3d animation. arXiv preprint arXiv:1802.06852, 2018.