Full Name	**Nationality**	**Address**
Vishwakarma Institute of Technology	Indian	Survey No. 3/4, Kondhwa (Budruk) Pune – 411048, Maharashtra (India)

Full Name (Including middle name) (Min. two faculty name)	**Nationality**	**VIT Address (Start with dept. name)**	**Mail ID**	**Phone No.**
Sakshi Vijay Khutwad	IN	Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Technology, Pune	sakshi.22420181@viit.ac.in	9527582709
Onkar Abasaheb Sathe	IN	Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Technology,	onkar.22420257@viit.ac.in	7588270392

		Pune		
Kritika Chandrashekhar Damahe	IN	Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Technology, Pune	kritika.22311693@viit.ac.in	9579990719
Sanket Somnath Mahajan	IN	Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Technology, Pune	sanket.22420197@viit.ac.in	7588334366
Dr. Parikshit Mahalle	IN	Department of Artificial Intelligence and data science Engineering, Vishwakarma Institute of Technology, Pune	parikshit.mahalle@vit.edu	9822416316

Gitanjali Bhimrao Yadav	IN	Department of Artificial Intelligence and data science Engineering, Vishwakarma Institute of Technology, Pune	gitanjali.yadav@viit.ac.in	9762763363
Swapnil K. Shinde	IN	Department of Artificial Intelligence and Data Science Engineering, Vishwakarma Institute of Technology, Pune	Swapnil.shinde@viit.ac.in	8329675265
Dr. Bipin Sule	IN	Department of Engineering, Sciences (Computer Prg) and Humanities, Vishwakarma Institute of Technology, Pune	bipin.sule@vit.edu	9822225577
Dr. Datta Takale	IN	Department of	dattatray.takale@viit.ac.in	7709091013

		Computer Engineering Vishwakarma Institute of Information Technology, Pune		
Dr. Ganesh Dongre	IN	Department of Mechanical Engineering, Vishwakarma Institute of Technology, Pune	ganesh.dongre@vit.edu	9822445831

1. **Title of the invention:**
   A Systematic Approach for Automated Speech to ISL Sign Language Translation Using Virtual Avatars.

2. **Technical field of the invention:**

The technological area where this new invention applies is related to assistive communication systems based on artificial intelligence for differently-abled people, particularly those who are hearing-impaired. More specifically, this invention is a client-server system and method, utilizing visual representations of sign language to facilitate communication between people with hearing impairment and those who do not use sign language as a means of communication. This invention addresses a common communication dilemma often encountered in real-life environments, such as homes, working environments, learning environments, and public zones, where communication is done verbally by one party, who is a sign language speaker, and a hearing-impaired person uses sign language. As a means to eliminate this communication barrier, a client computer system, such as a microphone, is established to capture oral audio inputs through an interface, specifically a mobile application and transmit them to a processing server. Accordingly, natural language processing is applied to the converted text so that the generated sign language grammar is followed instead of spoken language grammars. A generated sequence of gloss is translated with the help of a standardised sign language notation system to a gesture representation in order to provide an accurate description of all hand configurations, orientations, spatial positions, and movements in relation to the signs. Thereafter, the gesture representation feeds into a three-dimensional animated avatar that uses sign language movements, body position, and optionally expressions to convey visually the meaning of spoken communication. With this in mind, the invention would fall squarely within the class of AI-driven assistive technologies and speech-to-sign language translation systems designed for the purpose of enhancing inclusiveness and accessibility for people with hearing impairments.

3. Prior art:

Different systems and methodologies have been put forward, which intend to counter the challenge of communicating for people with hearing impairments and non-users of sign language, the idea being to translate verbal or written language to representations of the signs. The existing prior art for the task can be classified into rule-based translation systems, SMT solutions, systems for translation of words to signs based on deep learning, and the use of avatars for the animation of the signs.

Initially, there has been imitation based on rules for conversion from speech to signs or from text to signs, which first identify the input speech to text by speech recognition techniques and later rely on linguistic rules to transform this text to signs according to their predefined patterns [1], [2], [8]. In general, there has been removal of stop words, stemming/lemmatization, followed by rearranging the words based on an imitation of the grammatical structure of sign language for conversion from text to signs. However, some of the limitations for such imitation rules based on conversion from text to signs lie in following strict grammatical rules, a fixed vocabulary, and inefficient processing of new/complex words.

There were methods in statistical machine translation to handle vocabulary and syntax discrepancies between spoken languages and sign languages. The techniques in statistical machine translation view the process of translation as a probability task in light of parallel corpora of spoken language and sign language gloss. Phrase-based statistical machine translation models, IBM alignment models, and GIZA++ were employed to complete the process of translating a sentence in English into a sequence of glosses in a sign language. Although SMT is more flexible than rule-based machine translation, its disadvantage is the necessity to use a lot of annotated examples and the possibility to generate a sequence of glosses with grammatical violations.

Some research has recommended lexical-semantic methods to overcome the limitation of the vocabulary offered by sign language dictionaries using WordNet. In the system, if there is an absence of a word relating to the spoken language in the sign language dictionary, semantically closely associated words are identified using WordNet relationships (synonyms, hypernyms, and hyponyms), which are then substituted during the construction of the dictionary glosses [12]. Although these approaches offer good improvements for both coverage and semantics, the improvements can be at the word level and not at the sentence level.

In the process, there are systems that are specifically working on the animation and visualization of sign language. The systems accept the HamNoSys notation for formal representation of hand shapes, positions, orientations, movements, and spatial locations to produce images within the Signing Gesture Markup Language (SiGML) for animation via 3D avatars [9]. The present systems that exist presently, such as eSIGN Editor and the JA SiGML Player, have been designed for American Sign Language/UK Sign Language to produce HamNoSys manually. There are systems based on speech to sign that include speech recognition preprocesses and natural language processing for avatar animation based on word recognition from avatars that display videos/basic animations [1], [8], [11]. Though the systems are designed to produce real-time systems, grammatical correctness is often absent alongside restricted vocabularies/semantic substitution abilities besides lack of standard annotation.

Despite the progresses made, there is no existing system that integrates completely with respect to

(i) effective speech-to-text processing

(ii) Context-aware gloss generation employing the benefits of LLM-based linguistic comprehension & WordNet-based semantic mapping, and

(iii) The use of HamNoSys-to-SiGML translation in a common client-server system for generating avatars through standardized sign

Distinction from the Present Invention

In contrast to the foregoing prior art solutions, the proposed invention introduces an integrated AI-based assistive communication system comprising the following:

- Speech Recognition for Real-time Voice Input through Mobile-friendly UI.
- Text to gloss conversion employs the hybrid technique that makes use of large language models for contextual comprehension combined with WordNet-based semantic mapping to deal with vocabulary representations.
- Representation of signing in a standardized way using HamNoSys notation and SiGML-based animation of an avatar for visualizing signs in 3D space.
- By integrating semantic-aware gloss creation, formal sign notation, and rendering with avatars, the solution developed in the present patent tackles the drawbacks of non-adaptive rule- or algorithm-based sign systems, statistical models that are data intensive, and animation solutions that are not integrated with sign communication.

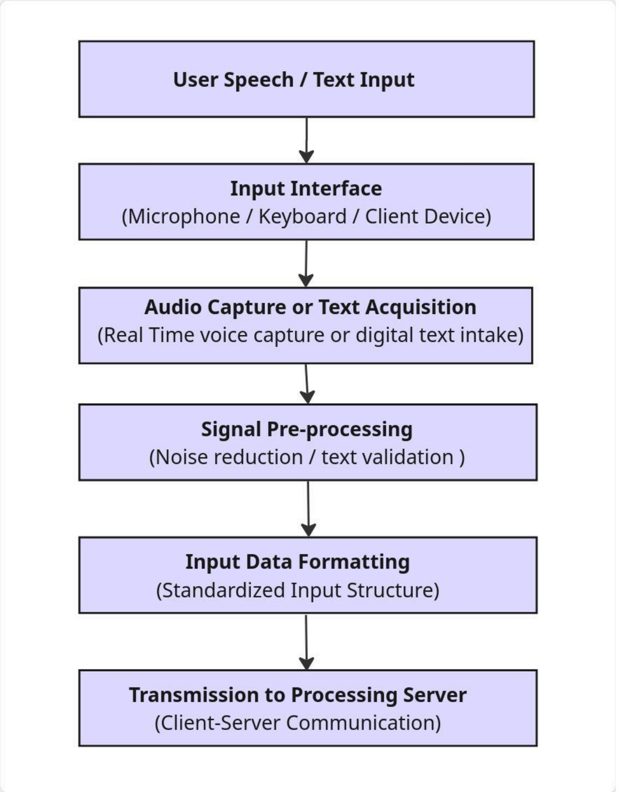
4. Synopsis:

The objective of the current assistive tool is to facilitate hearing-impaired people in grasping oral communication. The process involved in this tool is to record oral words in real time and process them by using artificial intelligence-based language understanding methods to produce sign language, which is organized in sign language grammar instead of voice. The processed result is displayed in a three-dimensional animated avatar, making oral communication easier to grasp for hearing-impaired people. This system can be used in mobile applications and provides online functionality, and thus it can be used in public, educational, and social settings. This system makes the process of communication easier for hearing-impaired people by providing a single solution for oral to sign language communication instead of relying on multiple translation aids.

5. Brief Description of the Assistive Communication System for Sign Language Translation:

Real-time rendering of sign language will provide the ability to comprehend spoken communication; hence, the system will allow highly improved accessibility and inclusivity for people with hearing impairments.

1. **Input Module**:



User Speech / Text Input
Input Interface
(Microphone / Keyboard / Client Device)
Audio Capture or Text Acquisition
(Real Time voice capture or digital text intake)
Signal Pre-processing
(Noise reduction / text validation)
Input Data Formatting
(Standardized Input Structure)
Transmission to Processing Server
(Client-Server Communication)

Figure 1. Illustrates a flow diagram of the speech module.

The input module of this system is configured to take spoken language, typically through a microphone or textual input via a client device - a smartphone, computer, or embedded system.

- For spoken input, the audio signal is captured in real-time.
- While in the case of text input, the system receives typed or digitally available text directly.
- This module ensures a more reliable acquisition of linguistic input from non-sign language users in everyday communication environments.

2. **Speech-to-Text Conversion Module**

This audio input is transmitted to the processing server, where the audio is processed using speech algorithms that result in the audio signal being translated into machine-readable text.

- Noise processing or language normalization is necessary for improved accuracy in transcription.
- The output of this stage is the text formed from the sentences uttered by consumers.
- This makes possible subsequent linguistic processing irrespective of the original audio format.

3. Natural Language Processing and Gloss Generation Module

This generated text is then processed by using natural language processing, which is adapted for translating sign language. The grammar of natural spoken language is broken down and recreated in grammatical forms suited to sign languages. This is the most critical step because sign languages have a syntactical and semantic structure that is completely different from that of spoken languages. In this system, the output preserves grammatical rules in sign languages without translating word for word.

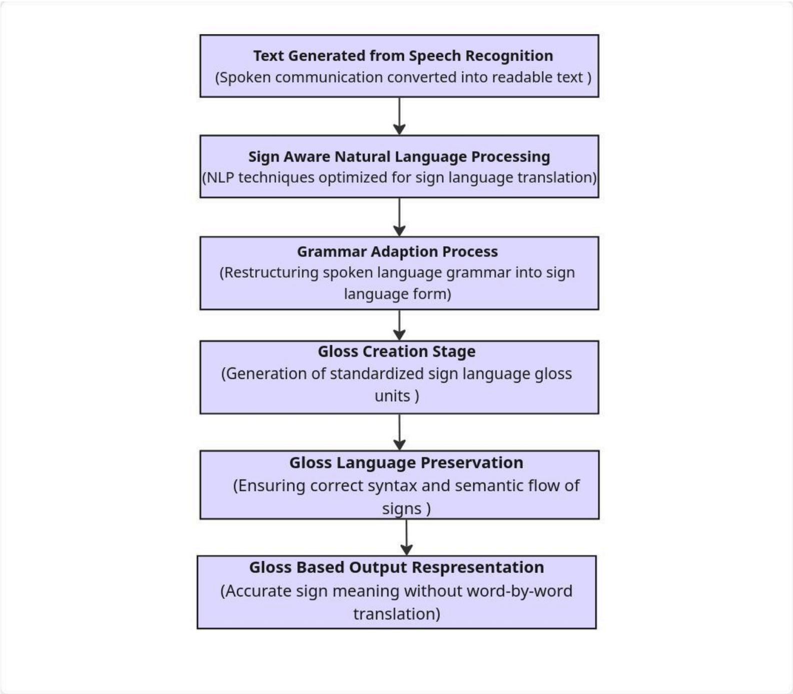

Text Generated from Speech Recognition
(Spoken communication converted into readable text)
Sign Aware Natural Language Processing
(NLP techniques optimized for sign language translation)
Grammar Adaption Process
(Restructuring spoken language grammar into sign language form)
Gloss Creation Stage
(Generation of standardized sign language gloss units)
Gloss Language Preservation
(Ensuring correct syntax and semantic flow of signs)
Gloss Based Output Respresentation
(Accurate sign meaning without word-by-word translation)

Figure 2. Illustrates a flow diagram of the Gloss generation module.

4. Sign Language Graphics Generation Module (SiGML and HamNoSys Based)

The system will have a sign language graphics module. It takes the processed sign to create visual sign language output by using HamNoSys (Hyposys) and SiGML. The module will realize accurate, expressive sign graphics, not altering spoken language grammar. The focus is on a straight, visual way of communicating.

The system, during this stage, converts the sign into standard HamNoSys codes and then into SiGML instructions. These instructions specify exactly how each sign shall appear, including:

- Hand shape and finger layout
- Hand direction and where it sits in space
- Direction, route, and sequence of movement
- Facial expressions and other non-manual features

These SiGML directions are followed by the system in order to display the graphics of sign language via a virtual avatar or a rendering engine. Output will be smooth, natural, linguistically accurate, clear, consistent, scalable sign animations for use in real-time translation and accessibility.

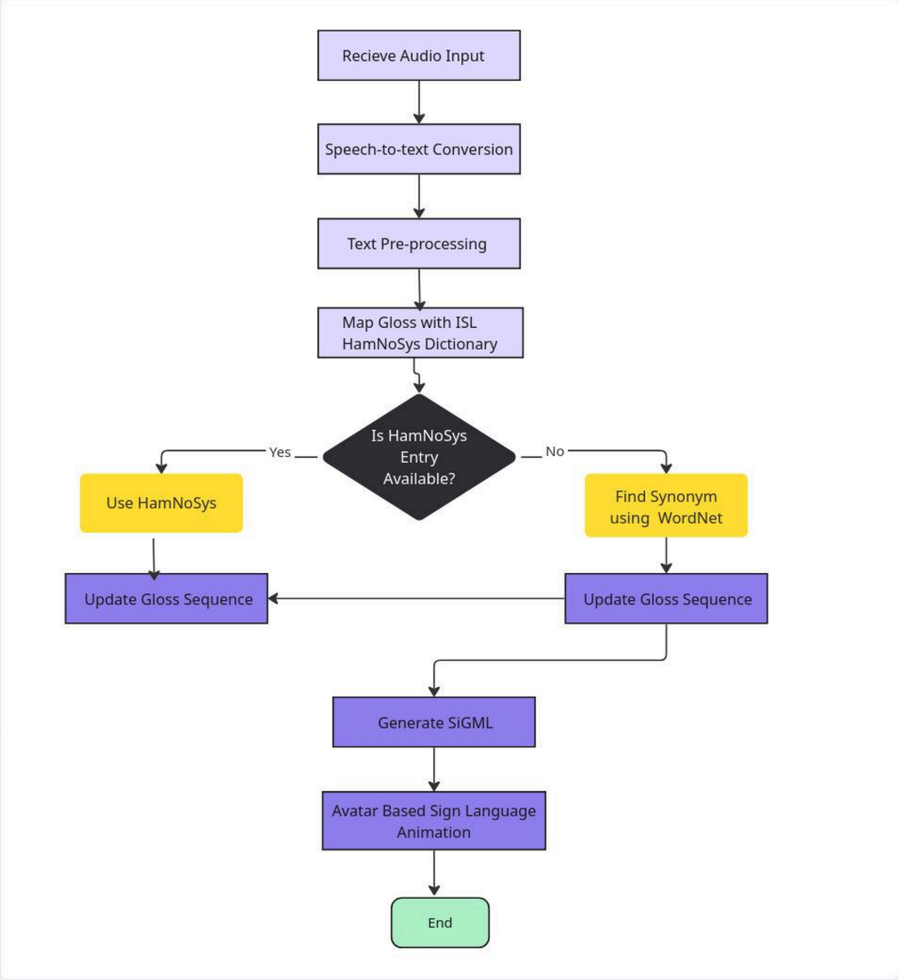

Recieve Audio Input
Speech-to-text Conversion
Text Pre-processing
Map Gloss with ISL HamNoSys Dictionary
Is HamNoSys Entry Available?
Yes
No
Use HamNoSys
Find Synonym using WordNet
Update Gloss Sequence
Update Gloss Sequence
Generate SiGML
Avatar Based Sign Language Animation
End

Figure 3. Illustrates a flow diagram HamNoSys-based Sign Language Graphics Generation Module.

I. Receive audio input:

The events start with the system receiving audio speech from a user through a microphone or an audio device. The speech may be live or a pre-recorded sound, which forms the prime source of the translation process.

II. Conversion of speech to text

The audio is then processed via a speech recognition component for converting the language used in speech into written form. This step converts audio into text form for processing by the computer.

III. Pre-process the text

Text cleaning for better accuracy include:

- Removing filler words and noise
- Lowercasing, punctuation, and numeric values
- Tokenizing and doing basic linguistic clean-up

The objective is to have clean and organized text that is ready for mapping to sign language.

IV. Map to ISL HamNoSys glosses

The preprocessed text is then linked to sign language glosses using the Indian Sign Language HamNoSys Dictionary. Each word or phrase is searched against the signs in the dictionary representation. This connects words in spoken language to signs in sign language.

V. Check if an entry in HamNoSys exists

It verifies whether a direct representation through HamNoSys exists for this given gloss. This ensures a stable process and prevents a translation failure.

VI. Are there transcriptions?

Yes →If an entry in the HamNoSys notation system exists, it chooses the code in HamNoSys precisely to define the shape of the hand, the motion, the orientation, and non-manual information for a proper language with visible visualization.

No →If there is no entry in HamNoSys, then the system relies on WordNet to identify a semantic synonym as a fallback mechanism. The synonym is also verified by the HamNoSys dictionary to ensure meaning is maintained, which remains sign language-friendly.

VII. Generate SiGML

In SiGML (Signing Gesture Markup Language), gloss sequences are converted into instructions. Machine-readable animation commands are one of SiGML's features. Accurate timing and control over space Facial expressions and hand gestures should be accurately depicted. This transforms language into visual instructions that can be executed.

VIII. Sign Language Animation Using Avatars

After that, the SiGML is sent to a rendering engine or virtual avatar that will produce real-time sign language animations. The signs appear natural as a result of the avatar's fluid hand movements, precise facial expressions, and appropriate timing.

IX. End

The translation is finished with a sign language animation that is accurate, expressive, and clear.

5. Gesture Mapping and Sign Notation Module

It has a module for sign notation and gesture mapping. The refined gloss output is converted to either HamNoSys, a standard sign notation format. This notation acts as a dependable transitional stage between language processing and visual animation, offering a symbolic explanation of how each sign is created.

HamNoSys codes outline important aspects of signing, such as:

- Finger positions and hand shapes
- The orientation of the hands and their spatial location
- The movement's trajectory, timing, and flow
- Non-manual markers that are optional, like facial expressions

After that, these HamNoSys codes are transformed into SiGML instructions, which provide a clear and machine-readable description of sign gestures for rendering and animation.

6. Animation Module for Virtual Avatars

A 3D virtual avatar is controlled by SiGML instructions from an earlier module. The resulting avatar displays visually accurate and time-synchronized sign language gestures, giving the impression of actual human signing rather than merely simple motions.

Coordinated hand gestures, body alignment, and optional facial expressions are all used in the animation to convey meaning clearly. The system maintains seamless transitions, positions the hands and body appropriately, and creates realistic motion flow across signs using SiGML-based control.

7. Module for Display and User Interaction

The user's smartphone then displays the animated sign language. Depending on how it is configured and the processing power, such a system can operate in real time or very nearly real time. There are quick-access features in this section:

- A virtual avatar displaying animated sign language
- rendering in real-time or almost real-time
- Pause, replay, and resume buttons for the signs
- Rapid adjustments to playback speed to facilitate comprehension

Operates on mobile devices and external environments like homes, offices, schools, and public areas can all benefit from the design. By offering simple controls and an understandable visual presentation, it facilitates spoken communication in an accessible sign language format for those who are hard of hearing.

Data Flow Diagram

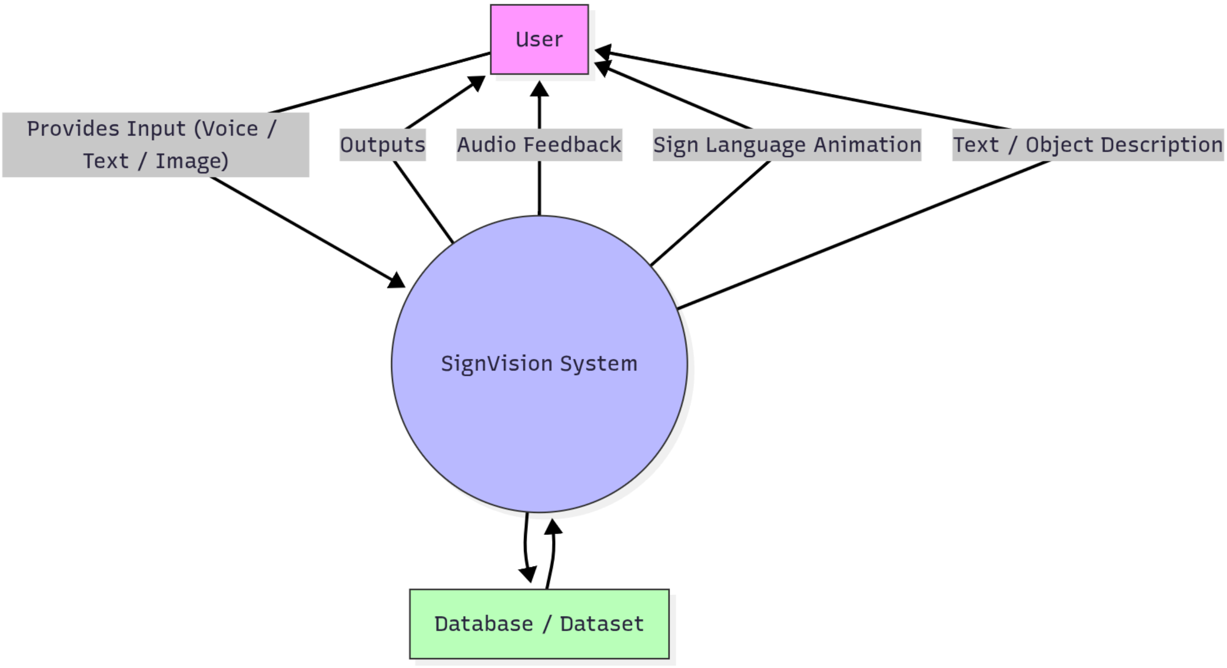


Figure 4. Illustrates a data flow diagram of an assistive communication system showing user input, system processing, database interaction, and multimodal output generation..

The data flow diagram shows how the assistive communication system—interacts with a user and a backend data repository. A user provides one or more forms of input, such as text, spoken audio, or image-based input, to the system while it is in operation. After receiving the input data, the SignVision System uses artificial intelligence-based modules for speech recognition, natural language processing, and semantic interpretation.

The System accesses a database or dataset to obtain linguistic resources, sign language gloss mappings, and standardised sign language notation data required for translation and visualisation. Based on the processed input, the system generates one or more output modalities, including textual or object-based descriptions, optional audio feedback, and sign language animation presented via a virtual avatar. These outputs are returned to the user, enabling bidirectional and multimodal communication. The data flow illustration highlights the centralised role of the SignVision System in coordinating input acquisition, data processing, database interaction, and output generation.

6. System Architecture:

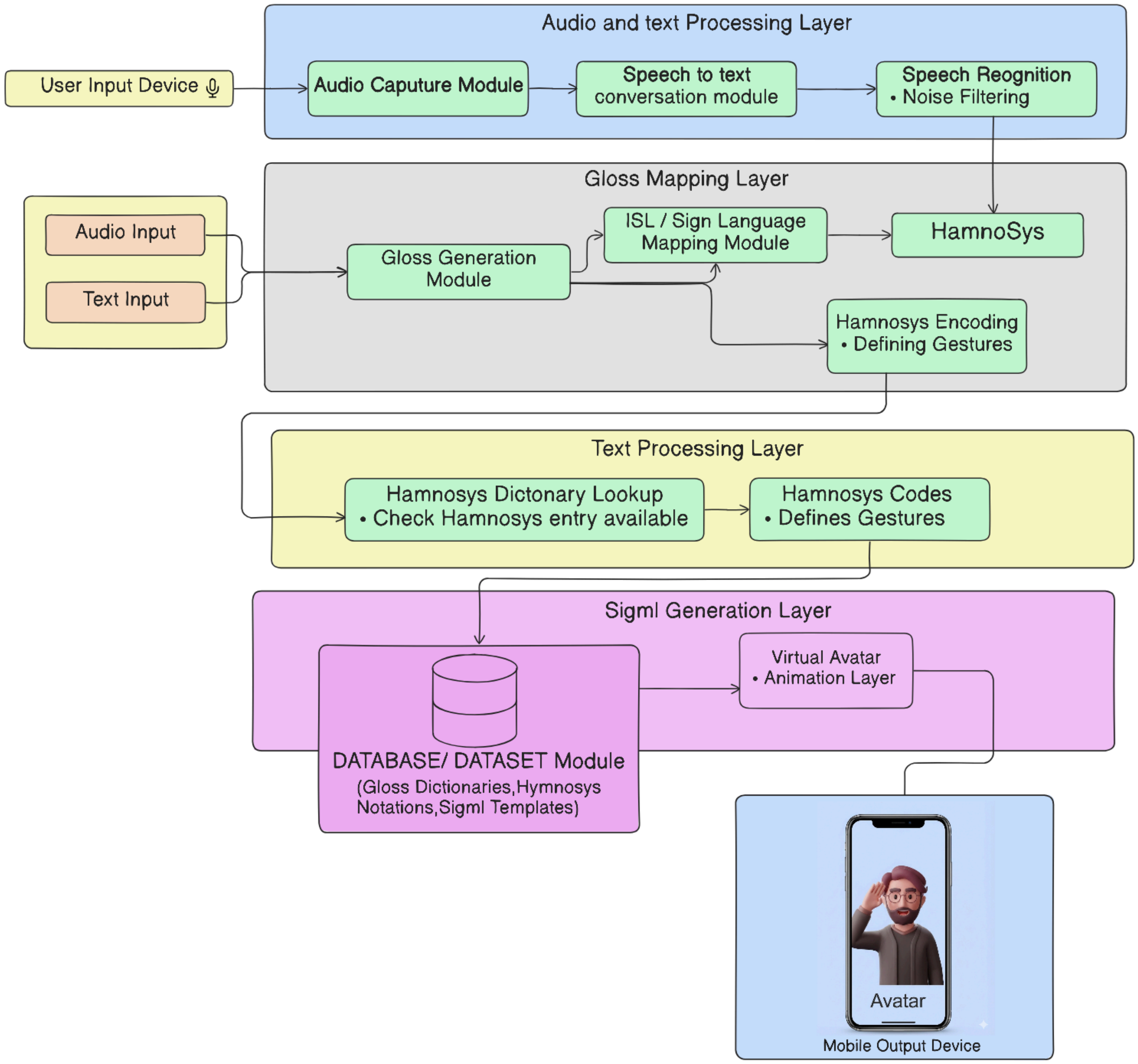


Figure 5. The System Architecture.

Users can input text or speech into the system, which is first recorded and processed to eliminate noise and transform speech into legible text. After that, this text is polished and converted into sign language glosses, which are then mapped to Indian Sign Language using HamNoSys to specify facial expressions, hand shapes, movements, and orientations. The system encodes full gestures based on the sign representations that are available. These encoded signs are stored in a database with linguistic and gesture data after being transformed into SiGML instructions. Lastly, a virtual avatar deciphers these commands and presents expressive, natural, and understandable sign language animations, allowing users to communicate in an accessible manner.

Below is a more thorough explanation of each component:

- User Input: Enables speech or text input from the user.
- Audio Capture Module: Uses an audio interface or microphone to record spoken input.
- Speech Recognition & Noise Filtering Module: Enhances audio clarity and eliminates background noise for precise recognition.
- Speech-to-Text Conversion Module: Produces text from a processed speech signal.
- The first processing layer that establishes standards is the Audio and Text Processing Layer.
- The Gloss Generation Module converts processed text into gloss units that can be represented in sign language.
- Glosses are mapped to corresponding Indian Sign Language (ISL) representations using the ISL/Sign Language Mapping Module.
- Standardised HamNoSys notations for hand shapes, orientations, movements, and non-manual characteristics are provided by the HamNoSys Module.
- HamNoSys Encoding (Defining Gestures): For accurate sign specification, each sign gesture is encoded using HamNoSys symbols.
- The HamNoSys Dictionary Lookup Module determines whether HamNoSys entries are available for every gloss in the dictionary.
- HamNoSys Codes (Defines Gestures): Contains verified HamNoSys codes that explain full sign gestures.
- Text Processing Layer: Manages HamNoSys representation validation, lookup, and verification.
- Gloss dictionaries, HamNoSys notations, and pre-made SiGML templates are stored in the database/dataset module.

- SiGML Generation Layer: Creates SiGML instructions for animation rendering from HamNoSys-coded gestures.
- Virtual Avatar Animation Layer: Uses a virtual avatar to interpret SiGML instructions and produce animated sign language output.
- The final sign language gestures are visually performed by the avatar.

In order to facilitate clear and accessible communication, the system takes speech or text as input, translates it into Indian Sign Language using gloss mapping, HamNoSys notation, and SiGML instructions, and then displays the result through a virtual avatar.

7. Claims

1. System Claim

An audio input unit that records spoken speech, a processing unit that transforms the speech into grammatically structured sign language representations, and a visual display unit that displays the representations using an animated sign language avatar make up this computer-implemented assistive communication system for those with hearing loss users.

2. Method Claim

A technique for improving communication with people who are hard of hearing that involves recording spoken words, turning them into text, converting the text into sign language grammar, and then using an animated avatar to visually represent the content as sign language.

3. Focused on Accessibility

An assistive communication system that automatically converts spoken input into visually understandable sign language representations so that hearing-impaired users can comprehend spoken communication from non-sign language users.

4. **Communication in Real Time**

Without the need for human interpreters, a real-time speech-to-sign language translation system enables direct communication between hearing and hearing-impaired users.

8. Inventive step of your invention:

Some of the current assistive communication systems for people with hearing impairments rely on static sign representations or speech-to-text transcription, which frequently fall short of communicating the expressive meaning and natural structure of sign language. The current invention presents an AI-powered assistive communication system that uses an animated avatar to convert spoken language into visually expressive sign language representations in real time. In terms of technology, the invention combines language comprehension, real-time speech processing, and avatar-based sign rendering. It was created especially to close the communication gap between hearing-impaired people and non-sign language users in everyday situations.

1. **Real-Time Language Processing and Speech Recording:**
   The suggested invention records live spoken audio and instantly transforms it into structured textual representations, in contrast to traditional systems that directly display spoken content as text. Instead of using spoken language syntax, the processed text is further examined and rearranged to adhere to sign language grammar. The generated sign representations are guaranteed to be natural and easily understood by users with hearing impairments thanks to this linguistic restructuring.

2. **Sign-Grammar-Based Translation and Representation:**
   Processed text is transformed into standardised sign language representations by the invention's sign-centric translation mechanism. The system overcomes a significant drawback of current assistive solutions by producing sign sequences that accurately reflect sign language grammar, spatial positioning, and semantic flow rather than depending on fixed sign libraries or word-by-word mappings.

3. **Avatar-Driven Expressive Sign Communication:**
   Using a three-dimensional animated avatar to visually convey sign language is a crucially creative feature. The avatar conveys the meaning and emotional context of the spoken input through precise hand gestures, body movements, and optional facial expressions. Compared to static images, text captions, or recorded gestures, this dynamic and expressive representation offers a more organic communication experience.

The integration of real-time speech understanding, sign-grammar-aware translation, and avatar-based visual expression within a single framework thus represents the innovative step of the current system. This combination represents a subtle improvement over current speech-to-text and basic sign language translation technologies and allows hearing-impaired people to communicate in a natural, expressive, and inclusive manner.

9. Abstract:

The goal of the current invention is to create an intelligent assistive communicative system that makes it easier for people with hearing impairments to interact in public and everyday situations. In real time, the system records spoken language and translates it into English text. The text is then further processed using language understanding techniques to conform to proper sign language grammar instead of spoken language structure. The final text is then converted into sign language representations and represented visually by an animated, three-dimensional avatar. In order to accurately convey the meaning and emotional context of spoken communication, the avatar would mimic precise hand gestures, body movements, and expressive cues. These features are combined into a single mobile framework that facilitates online use and smooth communication between non-sign language users and hearing-impaired users. Presenting spoken content as expressive, avatar-based sign language, as opposed to static text or symbols, lessens the need for generalised tools for translation and gives people with hearing impairments a more accessible, inclusive, and natural communication experience.

10. References:

[1] Shirisha, K., Deeksith, E. M., Madhava, M. S. S. S., Surendhar, S., & Karthikeyan, R. (2025). Signbridge-Audio to Sign Language Translator. IEEE, 1–5. https://doi.org/10.1109/sceecs64059.2025.10940434

[2] Amin, M., Hefny, H., & Mohammed, A. (2021). Sign Language Gloss Translation using Deep Learning Models. International Journal of Advanced Computer Science and Applications, 12(11). https://doi.org/10.14569/ijacsa.2021.0121178

[3] Othman, A., & Jemni, M. (2014). A novel approach for translating English statements to American sign language gloss. In Lecture notes in computer science (pp. 431–438). https://doi.org/10.1007/978-3-319-08599-9_65

[4] Sanaullah, M., Ahmad, B., Kashif, M., Safdar, T., Hassan, M., Hasan, M. H., & Aziz, N. (2021). A Real-Time automatic translation of text to sign language. Computers, Materials & Continua/Computers, Materials & Continua (Print), 70(2), 2471–2488. https://doi.org/10.32604/cmc.2022.019420

[5] Baltatzis, V., Potamias, R. A., Ververas, E., Sun, G., Deng, J., & Zafeiriou, S. (2024). Neural Sign Actors: A diffusion model for 3D sign language production from text. IEEE, 1985–1995. https://doi.org/10.1109/cvpr52733.2024.00194

[6] Othman, A., & Jemni, M. (2011). Statistical sign language machine translation: from English written text to American sign language gloss. arXiv preprint arXiv:1112.0168.

[7] Fayyazsanavi, P., Anastasopoulos, A., & Kosecka, J. (2024). Gloss2Text: Sign language gloss translation using LLMS and semantically aware label smoothing (pp. 16162–16171). https://doi.org/10.18653/v1/2024.findings-emnlp.947

[8] Deshmukh, A., Machindar, A., Lale, S., & Kasambe, P. (2024). Enhancing Communication for the Hearing Impaired: A Real-Time Speech to Sign Language Converter. IEEE, 1–5. https://doi.org/10.1109/wpmc63271.2024.10863135

[9] Kaur, K., & Kumar, P. (2016). HAMNOSYS to SIGML Conversion System for sign language Automation. Procedia Computer Science, 89, 794–803. https://doi.org/10.1016/j.procs.2016.06.063

[10] Boobal, A., Reddy, C. C. K., Reddy, C. A., & Rohith, C. B. V. S. (2024, December).

Real-Time Sign Language and Audio Conversion Using AI. In 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS) (pp. 1-6). IEEE.

[11] Saija, K., Sangeetha, S., & Shah, V. (2019, December). Wordnet based sign language machine translation: from english voice to isl gloss. In 2019 IEEE 16th India council international conference (INDICON) (pp. 1-4). IEEE.