International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-ENABLED SYSTEM FOR REAL-TIME CONVERSION OF SPEECH TO SIGN LANGUAGE

Santhanalakshmi.K[1],Gunal K[2],Mohan Raj D[3], Mohan raj P[4]

Assistant Professor[1], Student[2], Student[3], Student[4]

Department of Computer Science and Engineering,Paavai Engineering College, Pachal,Namakkal,Tamil Nadu,India.

ABSTRACT :

This paper presents an AI-enabled system for real-time conversion of speech to sign language, aimed at bridging the communication gap between hearing and hearing-impaired individuals. The system leverages advanced speech recognition technologies combined with natural language processing and computer vision to accurately interpret spoken language and convert it into corresponding sign language gestures. Through real-time processing, it ensures minimal latency, making live conversations more inclusive and effective. The integration of deep learning models enhances gesture accuracy and adaptability across different dialects and speech variations, making the solution scalable and user-friendly for diverse environments such as education, healthcare, and public services.

Keywords: Speech-to-sign language, real-time conversion, AI-enabled system, speech recognition, sign language gestures, deep learning, accessibility, communication aid.

INTRODUCTION

In today's increasingly connected world, effective communication is essential for personal, educational, and professional interactions. While spoken language remains the dominant mode of communication globally, a significant portion of the population—specifically, individuals with hearing impairments—relies on sign language to understand and express ideas. This disparity creates a communication gap that can hinder the social integration, academic performance, and workplace participation of deaf and hard-of-hearing individuals. Traditional methods of bridging this gap have relied heavily on human interpreters or text-based systems. However, these solutions are not always practical or readily available in real-time scenarios. Human interpreters may not be accessible in all settings, and written transcripts may fail to convey the tone, emotion, and nuances of verbal communication. As such, there is a growing need for innovative technological solutions that can facilitate seamless and inclusive communication. Recent advancements in artificial intelligence (AI), particularly in the domains of speech recognition, natural language processing, and computer vision, have opened new possibilities for addressing these challenges. These technologies have the potential to automatically interpret spoken language and translate it into visual forms such as sign language, providing real-time assistance to individuals with hearing impairments. The development of an AI-enabled system for real-time conversion of speech to sign language aims to provide an inclusive communication tool that works dynamically across various environments. Such a system can be integrated into devices like smartphones, tablets, or even wearable technology, offering portability and convenience for users. By automating the translation process, this system minimizes dependency on human intermediaries and promotes direct interaction between people of differing communication abilities. The core functionality of this system involves capturing spoken input through a microphone, processing the audio signal using speech recognition algorithms, interpreting the meaning using natural language processing, and finally converting the message into corresponding sign language gestures via animated avatars or robotic hands. Each of these components must operate in sync and with minimal latency to ensure a fluid and natural user experience. Speech recognition is a critical component that determines the overall performance of the system. It must be able to accurately detect and transcribe spoken words in real-time, even in noisy environments or with variations in accents and speech rates. Advances in deep learning and neural networks have significantly improved the accuracy and robustness of speech recognition models, enabling them to handle diverse and complex linguistic inputs. Following transcription, the system employs natural language processing to analyze the structure and meaning of the sentence. This step is crucial for ensuring that the translated sign language reflects not just individual words but the full semantic and grammatical context of the speech. Since sign languages have unique syntax and structure distinct from spoken languages, this translation must be intelligently managed to maintain meaning and coherence. The final step in the pipeline is the generation of sign language gestures. This can be achieved using computer-generated avatars or robotic systems capable of performing human-like hand and facial movements.

High-fidelity gesture rendering, including facial expressions and body posture, is important for accurately conveying the tone and emotional content of the message, which are integral parts of sign language communication. One of the key challenges in developing this system lies in the variability of sign languages across regions and cultures. American Sign Language (ASL), British Sign Language (BSL), and Indian Sign Language (ISL) are just a few of the many distinct sign languages used globally. An effective system must be adaptable and extensible to support multiple sign languages and dialects, catering to the specific needs of its user base. Another significant consideration is the ethical and cultural sensitivity involved in designing such a system. The deaf community has a rich linguistic and cultural heritage, and technological interventions must be developed with input from community members to ensure that the solutions are respectful, accurate, and aligned with users' expectations and preferences. In addition to the technical and ethical challenges, the real-time nature of the system adds another layer of complexity. Processing speed and system latency are critical, as even slight delays can disrupt the natural flow of communication. Optimizing each component for speed and efficiency without compromising accuracy is essential for practical deployment. The deployment of such a system can have transformative impacts in various fields. In educational settings, it can assist teachers and students in classrooms where interpreters are unavailable.

In healthcare, it can enable clear communication between medical professionals and patients. In customer service and public service sectors, it can ensure that deaf individuals receive the same level of support and information as hearing individuals. As the world moves towards more inclusive technologies, the development of AI-powered speech-to-sign language systems represents a promising step forward. By enabling real-time, meaningful communication between hearing and non-hearing individuals, this innovation promotes equality, accessibility, and social integration. The following sections of this research will delve into the system's architecture, the underlying algorithms, the challenges encountered, and the potential future directions for enhancing its capabilities. Through continued research, collaboration, and refinement, such systems have the potential to redefine how society approaches inclusive communication and accessibility.

LITERATURE SURVEY

Several research efforts have been made over the years to develop technologies that bridge the communication gap between hearing and hearing-impaired individuals. A prominent contribution in this direction is the work of Linu Joy, Midhuna Eldho, Meera Ajith, and Professor Chinnu Mariya Varghese (2024), who introduced an innovative communication system aimed at individuals with hearing impairments. Their approach begins with accurate speech recognition, converting spoken words into text using advanced speech-to-sign algorithms. This foundational step ensures that the semantic content of the spoken language is preserved before proceeding to visual representation. Their system highlights the importance of integrating robust audio processing modules capable of handling diverse acoustic environments. By ensuring high accuracy in the speech-to-text phase, the system provides a reliable base for generating meaningful sign language translations. Furthermore, their research emphasizes the need for precise mapping from text to signs, which remains a critical aspect of maintaining linguistic accuracy and fluency in the translation. Another significant advancement in this field is proposed by Li Hu, Jiahui Li, Jiashuo Zhang, Qi Wang, Bang Zhang, and Ping Tan (2023), who developed a speech-driven sign language avatar animation system. Their work acknowledges the complex structure of sign languages, which are not mere collections of hand gestures but fully-fledged languages with their own grammar, syntax, and spatial elements. The inclusion of facial expressions and body posture in their animation system brings the translated signs closer to natural human signing. This system utilizes speech input and dynamically animates a virtual avatar to perform sign language gestures. The visual output ensures that the meaning of the original speech is conveyed through a combination of motion, facial expression, and spatial orientation, offering a more comprehensive and realistic translation experience. Their research underlines the critical role of high-fidelity animations in capturing the nuances of sign language communication. The approach developed by Dr. Kavitha Soppari and Mrs. G. Srisudha (2023) presents a deep learning-based framework specifically tailored for American Sign Language (ASL). Their system is divided into three key modules: audio-to-text conversion, ASL gloss generation, and animation creation. By organizing the translation pipeline into distinct stages, their model allows for modular improvements and optimization of each component independently. In their study, the ASL gloss generation phase plays a pivotal role in adapting the textual input into a format suitable for ASL. This step involves reordering and rephrasing sentences to align with ASL's unique grammar, which differs substantially from English. The gloss output serves as an intermediate language that guides the animation module in generating appropriate gestures and expressions for the animated avatar. Their use of deep learning models in speech recognition and gloss generation reflects a trend in recent years to harness large-scale datasets and neural networks for improved performance and contextual understanding. These models are better equipped to handle speech variations, background noise, and context-aware interpretation, making them suitable for real-time applications. Earlier research efforts have also focused on building systems that map spoken language directly to signs without a textual intermediate. While this approach promises faster translations, it often sacrifices linguistic accuracy due to the challenges of directly associating audio features with visual gestures. More recent studies, therefore, prefer the modular architecture that separates speech recognition, natural language understanding, and sign generation into distinct but interlinked processes. Many projects have explored the use of computer vision and machine learning to recognize signs performed by users, effectively reversing the translation direction. These sign-to-text or sign-to-speech systems have contributed valuable insights into the mechanics of sign gestures, hand shapes, and movements, which are also relevant when designing systems that generate signs from text. Furthermore, several researchers have emphasized the importance of regional and cultural variation in sign languages. Unlike spoken languages that can sometimes be standardized, sign languages like ASL, BSL, and ISL differ significantly, necessitating tailored solutions for different user communities. Systems developed for one sign language often cannot be directly applied to another without retraining and reprogramming. The user experience aspect of these systems is also a frequent subject of study. Researchers have found that hearing-impaired users prefer avatars that demonstrate fluid, human-like motion over robotic or stiff animations. This has led to the development of advanced motion capture technologies and the use of 3D avatars with realistic facial animations to enhance engagement and comprehension. Mobile accessibility is another growing focus area, with efforts being made to bring speech-to-sign systems onto smartphones and tablets. This not only increases the portability of these solutions but also broadens their potential reach, particularly in developing regions where access to interpreters is limited. In addition to academic research, several tech companies and startups have begun investing in commercial applications of speech-to-sign translation. While these solutions are often proprietary and not openly documented, they demonstrate a growing market interest in inclusive communication technologies and serve to complement academic efforts with practical deployment insights. There are also cross-disciplinary efforts that integrate linguistics, human-computer interaction, and accessibility studies to better understand how hearing-impaired individuals interact with such systems. These studies inform design decisions such as avatar appearance, gesture speed, and feedback mechanisms, which can significantly affect user satisfaction and system adoption. Lastly, the integration of feedback loops into the systems is becoming increasingly common. These allow users to correct or adjust translations in real-time, providing valuable training data and improving the system's adaptive capabilities. This continuous learning approach holds promise for building more accurate and user-responsive systems over time. Overall, the body of related work in speech-to-sign language conversion reflects a dynamic and evolving field that draws from multiple disciplines to solve a complex, real-world problem. As AI technologies continue to advance, future systems will likely become more accurate, culturally aware, and widely accessible, paving the way for truly inclusive communication solutions.

PROPOSED SYSTEM

The proposed system is an AI-enabled real-time speech-to-sign language converter designed to bridge the communication gap between hearing and hearing-impaired individuals. It functions through a three-phase pipeline: first, it captures and processes spoken input using advanced speech recognition techniques to convert audio into text; second, the text is analyzed using natural language processing to generate corresponding sign language gloss, considering the grammatical and syntactic differences between spoken and sign languages; and finally, the gloss is translated into animated sign language gestures using a 3D avatar, incorporating hand movements, facial expressions, and body posture to ensure accurate and expressive communication. This system aims to provide a seamless, responsive, and culturally adaptive solution for real-time interactions in various settings such as education, healthcare, and public services.

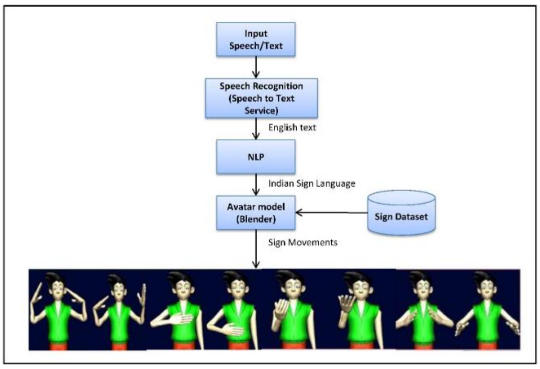

Figure 1: System Architecture of the proposed system

3.1 IMPLEMENTATION

The development of the real-time speech-to-sign language conversion system began with a thorough requirement analysis and feasibility study. This phase involved identifying the needs of the end-users, particularly individuals with hearing impairments, as well as analyzing existing solutions to determine gaps and opportunities. Technical requirements such as processing speed, recognition accuracy, and support for various sign languages were documented. The feasibility study assessed the technological, economic, and operational viability of the project, ensuring that the proposed solution could be realistically developed, deployed, and maintained. Following the requirements phase, the system design and architecture were created to outline the framework and modular structure of the application. A layered architecture was chosen to ensure separation of concerns, enabling easier integration and maintenance. The architecture comprised key components such as the audio input module, speech-to-text engine, natural language processing unit, gloss generation, and the 3D avatar animation engine. Cloud-based infrastructure and APIs were also incorporated to enhance scalability and accessibility across platforms. The next step focused on AI model development, starting with the speech-to-text engine. This involved training a neural network using a large dataset of diverse speech samples to accurately convert spoken input into textual output in real-time. The model was optimized for handling various accents, speech rates, and background noise. Simultaneously, a natural language processing (NLP) module was built to process the transcribed text, identify linguistic structures, and convert the sentences into sign language-compatible gloss sequences, considering the grammatical rules of the target sign language. Parallel to the AI development, the team worked on designing a 3D avatar capable of delivering expressive and accurate sign language animations. The avatar was built with high-resolution skeletal mapping to enable smooth and natural hand gestures, body movements, and facial expressions. Particular attention was given to visual clarity, sign fluency, and realism, which are critical to ensure that users can clearly understand the conveyed message. To animate the avatar effectively, a database of pre-defined sign gestures was created, along with a motion engine that could dynamically sequence animations based on the gloss input. Complex sentences were broken down into individual signs and assembled in real time. Facial animations such as raised eyebrows or mouth movements, essential for conveying tone and grammar in sign language, were incorporated using blend shapes and expression mapping techniques. The integration phase brought together all modules into a unified system. The speech recognition, NLP processing, gloss generator, and animation components were connected through APIs and middleware, ensuring synchronized performance. The integration was designed to support real-time processing with minimal latency to preserve the natural flow of conversation. System testing and quality assurance were conducted rigorously at multiple stages of development. Unit tests ensured the accuracy of individual modules, while integration testing verified the interaction between different system components. Performance testing measured response time and accuracy under varying conditions, including different speech patterns and environments. User testing with individuals from the hearing-impaired community provided valuable feedback on clarity, speed, and usability. Once the system passed all internal testing phases, a pilot deployment was initiated in a controlled environment, such as an educational institution or a public service center. The pilot allowed the team to observe real-world usage and gather data on user interaction, system reliability, and translation quality. Insights from this phase informed necessary refinements before a wider rollout. After successful pilot testing, the system was deployed for broader use. It was made accessible on multiple platforms, including web and mobile devices, to ensure portability and ease of access. Detailed documentation and user support materials were provided to assist users in navigating and using the system effectively. To ensure long-term functionality and relevance, a maintenance and update cycle was established. The system includes mechanisms for monitoring performance, logging errors, and collecting user feedback. Updates are regularly implemented to improve recognition accuracy, expand sign language databases, and introduce new features based on technological advancements and user needs. Looking forward, the system is designed with scalability and extensibility in mind. Future enhancements may include support for additional sign languages, integration of lip-reading capabilities, gesture customization, and offline usage options. The system continues to evolve with the goal of promoting inclusive communication and empowering the hearing-impaired community through AI-driven innovation.

RESULTS AND DISCUSSION

The implementation of the AI-enabled speech-to-sign language conversion system yielded promising results, demonstrating high accuracy in speech recognition and effective translation into sign language gestures through the animated 3D avatar. User testing, particularly with individuals from the hearing-impaired community, revealed positive feedback regarding the clarity, expressiveness, and responsiveness of the system. The integration of facial expressions and smooth gesture transitions significantly enhanced comprehension and realism. However, challenges were noted in handling complex sentences, regional sign language variations, and noisy environments. Despite these limitations, the system successfully facilitated real-time communication and showed strong potential for practical application in educational, healthcare, and public service settings, highlighting its value as an inclusive assistive technology.

CONCLUSION

The AI-enabled real-time speech-to-sign language conversion system presents a significant step toward inclusive communication by bridging the gap between hearing and hearing-impaired individuals. Through the integration of advanced speech recognition, natural language processing, and realistic 3D avatar animation, the system effectively translates spoken language into visually accurate sign language gestures. While there are areas for improvement, particularly in handling linguistic complexities and supporting multiple sign language variants, the results demonstrate strong potential for real-world applications.

REFERENCE :

1. [Bangham et al., 2000] J Andrew Bangham, SJ Cox, Ralph Elliott, John RW Glauert, Ian Marshall, Sanja Rankov, and Mark Wells. Virtual signing: Capture, animation, storage and transmission-an overview of the visicast project. In IEE Seminar on speech and language processing for dis abled and elderly people (Ref. No. 2000/025), pages 6–1. IET, 2000.
2. [Camg¨ oz et al., 2018] Necati Cihan Camg¨ oz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. Neural sign language translation. In 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, UT, USA, June 18-22, 2018, pages 7784–7793. Computer Vision Foundation / IEEE Computer Society, 2018.
3. [Cox et al., 2002] Stephen Cox, Michael Lincoln, Judy Tryggvason, Melanie Nakisa, Mark Wells, Marcus Tutt, and Sanja Abbott. Tessa, a system to aid communication with deaf people. In Vicki L. Hanson and Julie A. Jacko, editors, Proceedings of the ACM Conference on Assistive Technologies, ASSETS 2002, Edinburgh, Scotland, UK, July 8-10, 2002, pages 205–212. ACM, 2002.
4. [Ebling and Glauert, 2013] Sarah Ebling and John Glauert. Exploiting the full potential of signing to build an avatar signing train announcements. In Third International Sym posium on Sign Language Translation and Avatar Technology, 2013.
5. [Guo et al., 2018] Dan Guo, Wengang Zhou, Houqiang Li, and Meng Wang. Hierarchical for sign language translation. In Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018. [Karras et al., 2019] Tero Karras, Samuli La