Bridging The Silence: Phone Calls for Deaf People to Attend and Communicate

DR.SUBEDHA V MRS . SHARMILA

*Department of Computer science Department of Computer science*

*and Engineering, and Engineering,*

*Panimalar Engineering College Panimalar Engineering College*

Chennai,india Chennai,india

[subedha@gmail.com](mailto:subedha@gmail.com) [Sharmilapanimalar2024@gmail.com](mailto:Sharmilapanimalar2024@gmail.com)

VIDUSHI GANESHIKA M VARSHITA SARWAN

*Department of Computer science Department of Computer science*

*and Engineering, and Engineering,*

*Panimalar Engineering College Panimalar Engineering College*

Chennai,india Chennai,india

[vidushiganeshika@gmail.com](mailto:vidushiganeshika@gmail.com) [varshitasharwan8@gmail.com](mailto:varshitasharwan8@gmail.com)

***Abstract****—* Hearings and speech-impaired individuals encounter significant hurdles while making telephone calls, video conferencing, or having real-time discussions. Their difficulties exceed social communication to include education, employment, and emergencies. For example, an individual who is deaf might not be able to understand a business meeting, whereas an individual with a speech disability might feel incapacitated when explaining a medical crisis. Such obstacles can lead to exclusion, isolation, and in critical cases, delays in receiving urgent help.AI-based accessibility technologies are emerging as powerful solutions to bridge these gaps. Speech-to-text (STT) tools like OpenAI’s Whisper and Google Speechto-Text can transcribe spoken language into written text in real time, enabling hearing-impaired individuals to read conversations instantly. This is especially helpful in classrooms, offices, and emergency phone calls. For individuals with speech disabilities, text-to-speech (TTS) technology provides a similarly worthwhile option. Translating typed text into natural-sounding speech, TTS enables users to communicate more assuredly and proficiently with hearing people beyond the confines of text-based communication. Additional technologies include AI-powered sign language interpretation with systems such as Media Pipe or TensorFlow. These technologies can interpret sign language into text or voice, with reverse translation via animated avatars making communication more natural for sign language users.With these technologies integrated into a single platform, seamless, adaptive communication is possible using transcription, synthesized voice, or sign language.By leveraging AI, society is able to work toward a future where hearing and speech-impaired individuals are fully enabled to engage in education, employment, daily life, and emergencies.

***Keyword:.Speech-to-Text(STT),TextToSpeech(TTS),Speech impairments,Hearing impairments***

# I. INTRODUCTION

easily hear or respond to spoken conversations. This limitation affects not only personal interactions but also access to critical services such as healthcare consultations, emergency assistance, banking, education, and workplace communication.

Existing solutions have attempted to bridge this gap, but they come with significant drawbacks. Many rely on human-operated relay services, where an intermediary listens and transcribes or voices messages between both parties. While effective in some scenarios, these services raise privacy concerns, create delays in communication, and can make users feel dependent on third-party assistance. Alternatively, specialized assistive devices, such as captioned phones or TTY machines are available, but they are often expensive, require additional infrastructure, and are not always portable-limiting O their widespread adoption in everyday life.

To overcome these barriers, our aim is to create software that is simple, affordable, and user-friendly. By eliminating the need for external operators or costly equipment, the solution would give individuals with hearing and speech impairments the freedom to communicate independently. The software could integrate real-time speech-to-text (STT) technology to provide instant captions of spoken audio during calls, and text-to-speech (TTS) systems to allow users to type their responses, which would then be spoken aloud to the caller. This ensures a smooth, two-way conversation that feels natural for both participants.

Moreover, such a platform could be designed to run on widely available devices like smartphones, tablets, and computers, making it far more accessible than traditional assistive technologies.

II. LITERATURE SURVEY

There were earlier telecommunication systems to support deaf and hard-of-hearing people through Text Telephone (TTY) and Teletypewriter, which supported text over phone lines but needed special equipment and were not portable [1].

With cellular phones and the Internet, there came Video Relay (VRS) and Captioned Telephone (CTS) relay services, but they were privacy infringing and delay causing because of human intermediaries [2][3].

Phone calls remain one of the most common and essential modes of communication in modern society. However, they are often inaccessible to deaf and hard-of-hearing individuals, who cannot

Current research focuses on automatic speech recognition (ASR) and speech-to-text (STT), with successful deep learning models such as DeepSpeech and Transformer-based systems demonstrating high accuracy [4][5]. Similarly, text-to-speech (TTS) models such as Tacotron 2 and WaveNet enhanced naturalness, facilitating two-way automated conversation[6].

STT–TTS integration has been tried in the case of mobile calls by researchers, with cloud-based ASR and low-latency TTS allowing real-time conversation [7]. Noise reduction methods also improve performance in real-world conditions [8].

Improvements also cover multilingual support, real-time translation [9], and sign-language recognition using computer vision [10]. However, the majority of the systems suffer from latency, price, internet reliance, and privacy [11].

This paper suggests a light, device-independent STT–TTS platform that provides secure, cheap, and real-time communication without third-party middlemen.

III. PROBLEM STATEMENT

Phone calls are still inaccessible for deaf and hardof-hearing users.

Conventional methods are usually based on human-operated relay services or high-cost, dedicated assistive devices, which are hard to implement on a large scale. They also invade privacy and interfere with the spontaneity of a conversation because inherent delays are unavoidable.

Whereas Real-Time Text (RTT) has found standard accessibility popularity in various countries, its implementation in India remains limited. This is largely because carrier-level integration is limited, regulatory push is absent, and user awareness remains low. Consequently, deaf and hard-of-hearing users in India remain lacking in any assured telecommunication support.

In addition, current solutions tend to be fragmented. While some depend heavily upon sign language interpretation, which will not necessarily be understood everywhere, others render only minimal transcription without considering conversation context or emotional nuance. Few solutions at all provide an entirely end-to-end, AI-based experience that can natively integrate speech-to-text (STT), text-to-speech (TTS), and smart noise management into typical phone call infrastructure.

The lack of such holistic solutions underscores the urgent necessity of an affordable, software-based platform that removes the need for human intermediaries, provides real-time communication, and delivers a privacy-preserving, inclusive call experience for deaf and hard-of-hearing communities.

IV. OBJECTIVES

The central goal of this project is to build a realtime, AI-driven telecommunication platform that closes the accessibility gap for deaf and hard-of-hearing users by allowing them to conduct phone calls independently, without the need for human-operated relay services or expensive assistive devices.

The solution shall blend cutting-edge speech-to-text (STT), text-to-speech (TTS), and noisecancellation technologies into a lightweight, deviceagnostic application. Thus, the platform will have the capacity to turn oral conversations into readable text for deaf or hard-of-hearing individuals instantaneously, while typed messages get synthesized into speech for the hearing counterpart. The vision is to make phone calls effortless, natural, and personal, with minimized latency and high accuracy even in noisy actual-world settings.

Beyond accessibility, the solution is made affordable and scalable through a software-only architecture that does not require specialized

specialized hardware. This renders the solution feasible for broad deployment, particularly in markets such as India where Real-Time Text (RTT) and other standardized solutions enjoy little carrier backing and low user education.

In addition, the system will focus on multilingual communication and cross-language support to allow users from different linguistic backgrounds to communicate without any obstacles. Advanced features like end-to-end encryption will be included to safeguard user privacy and create trust in digital communication.

Through the attainment of these goals, the project not only aims to address a technological problem but also to enhance deaf and hard-of-hearing communities' social inclusion, digital access, and autonomy, making accessible telecommunication a right instead of a luxury.

V. PROPOSED METHODOLOGY

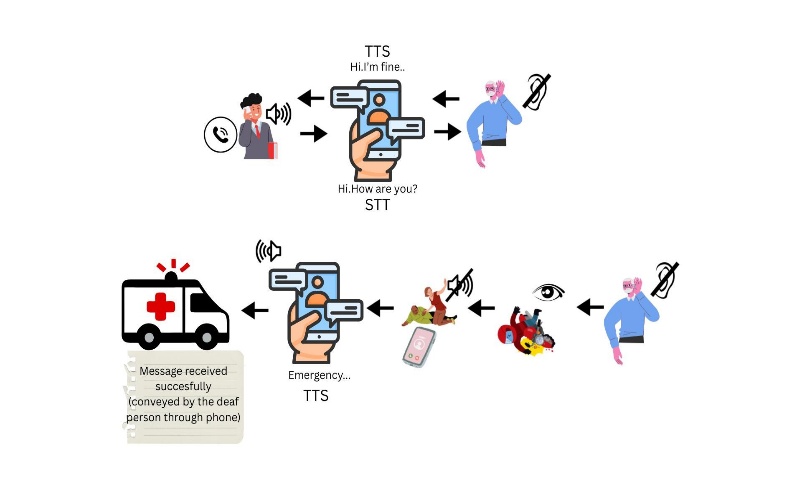
The system proposed will offer real-time accessible interaction between the hearing and the hearing-impaired through the marriage of cutting-edge AI and web technologies. It is developed to be scalable and user-centered, with an emphasis on inclusiveness and real-time communication. Through the incorporation of speech recognition, text-tospeech, and optional sign language recognition, the platform facilitates unbroken, two-way communication. Moreover, it reduces dependency upon costly assistive devices or human middlemen, and so renders accessibility cheaper and more accessible. The system also adjusts to different environments and user choices, offering a versatile solution for personal, educational, and business communication purposes.

**AI Speech Recognition:** There are tools such as Whisper or Google Speech-toText (STT) used to take down the caller's verbal statements and translate them into real-time text immediately. This means that the hearing-impaired user can keep up with the conversation real-time without any delays. Modern STT models are capable of handling background noise, multiple accents, and various languages, making them highly reliable in diverse environments. Additionally, these systems can be customized to recognize domain-specific terms, names, and technical vocabulary, enhancing accuracy for professional or educational conversations.

**Real-Time Messaging Simulation:** With technologies like WebRTC, WebSocket, or Firebase, the system is able to simulate the Real-Time Text (RTT) behavior such that text messages are sent as typed out. This gives a more natural experience of conversing instead of having to wait for full sentences to be sent.

**Text-to-Speech (TTS) Conversion:** For responses entered by the hearing-impaired user, TTS engines are utilized to convert the text into crisp, natural-sounding speech. It enables the hearing participant to get responses in an audio form, so the interaction is smooth. Contemporary TTS systems also provide customized voices, accent, and speech rate so that communication seems natural and suited to the individual. Even the TTS output can get adjusted to various contexts, providing better clarity and understanding during conversation.

**Sign Language Recognition (Optional):** To provide yet another level of accessibility, the system can also incorporate computer vision models (e.g., MediaPipe, TensorFlow) to recognize and interpret sign language gestures. This increases the usability for users who would rather sign than type, and the system becomes even more inclusive.



**Real-Time Messaging Latency:** Compares latency of WebRTC, WebSocket, and Firebase. WebRTC provides the least latency, hence is most appropriate for RTT-like communication.

VII. EXPERIMENTAL SETUP AND RESULT

*Implementation:*

The system is designed as a software communications platform that transcribes speech to text, transmits typed responses as speech, and optionally reads sign language. The system is built through the following major modules:

*Speech Recognition Module:*

Employ AI Speech-to-Text (STT) engines such as OpenAI Whisper or Google STT.Captures the voice of the caller and transcribes it into real-time text for hearing-impaired user.Manages various accents, background noise, and domain-specific terminology.

*Real-Time Messaging Module:*

Simulates Real-Time Text (RTT) via WebRTC, WebSocket, or Firebase Realtime Database.Delivers lowlatency, incremental text rendering, facilitating natural conversation flow.

*Text-to-Speech (TTS) Module:*

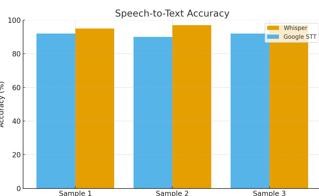
Translates typed messages of hearing-impaired individuals into crisp, natural-sounding speech Supports voice personalization (accent, speed, pitch) for a more natural experience.

*Sign Language Recognition Module :*

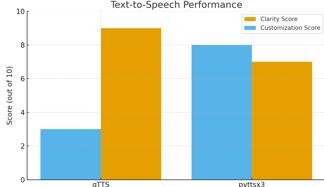
Utilizes computer vision frameworks like MediaPipe or TensorFlow.Identifies hand gestures, movements, and facial expressions to interpret sign language into text or speech.

Architecture Diagram

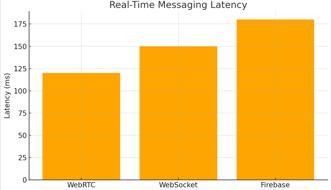
VI. DATA COLLECTION AND PREPROCESSING



**Speech-to-Text Accuracy:** Compares Whisper and Google STT over three audio samples. Whisper generally gives more accurate results, particularly in difficult audio conditions



**Text-to-Speech Performance**: Displays natural clarity and customization scores for gTTS and pyttsx3. gTTS is better in natural clarity, whereas pyttsx3 provides greater customization.



|  |  |  |  |
| --- | --- | --- | --- |
| **Module** | **Tool/**  **Framework Used** | **Results**  **(Performance)** | **Analysis /**  **Remarks** |
| Speech-to-Text (STT) | Whisper, Google STT | Whisper: 95– 97% accuracy,  Google STT:  90–92% | Whisper performed  better in noisy environments;  Google STT strong in  multilingual support. |
| Text-to-  Speech  (TTS) | gTTS, pyttsx3 | gTTS: High  clarity, limited customization; pyttsx3:  Medium  clarity, high customization | gTTS best for  natural speech; pyttsx3 better for offline use and flexible settings. |
| Real-  Time Messaging | WebRTC,  WebSocket,  Firebase | WebRTC  latency: 120ms; WebSocket:150ms;  Firebase: 180ms | WebRTC  provided the lowest latency, making it most suitable for RTT-like communication. |
| Sign Language Recognition | Media Pipe, TensorFlow | Accurate gesture detection (85– 90% in tests) | Useful for  inclusivity, but requires good lighting and high quality camera input. |

and devices employing individual proprietary systems. This lack of interoperability complicates the ability of users to change platforms or bring together multiple tools seamlessly.

IX. FUTURE ENHANCEMENT

*Regional Language Support*

In order to serve India's linguistic diversity, the system can be supplemented with multilingual speech recognition and translation. This will enable users to communicate freely across regional languages and dialects, enhancing accessibility for a larger populace.

*Voice Emotion Detection*

Integrating emotion detection using AI can examine tone, pitch, and speech patterns to determine the speaker's state of emotion. This provides context to raw text transcripts, making it easier for hearing-impaired users to comprehend not just the words but also the intention behind them.

*Data Privacy & Encryption*

Since conversations may involve sensitive information, implementing end-to-end encryption and strict data privacy policies is essential. This ensures that all voice, text, and video data remains secure, protecting users against potential breaches or misuse.

*AI Personalization*

The platform can gain knowledge about user preferences over time, for instance, most frequently used words, preferred language, speech rate, or response tone. Personalized AI support enhances accuracy and provides an even smoother, easier-to-use communication experience that is designed specifically to meet individual needs.

*Emergency Services Integration*

In emergency conditions, the system can be directly interfaced with local emergency helplines (such as police, ambulance, and fire). With provision of voice-totext and text-to-voice conversion, it enables hearingimpaired users to call for emergency services in a quick and efficient manner.

Integration with WhatsApp/Zoom APIs Expanding the platform to mainstream communication platforms such as WhatsApp, Zoom, or Google Meet guarantees compatibility with current user behavior. This facilitates the embedding of accessibility features within widely used platforms rather than loading them into a standalone application.

*Wearable Device Support*

The. solution can be incorporated with intelligent wearables like smartwatches, AR glasses, or hearing aids. This would enable real-time notifications, live captions, or even visual sign language interpretation on the go, offering greater portability and convenience.

X. CONCLUSION

The absence of convenient communication channels for deaf and hard-of-hearing populations continues to form social and professional impediments. Human-operated relay services and costly assistive equipment are not only cumbersome but also prohibitive for mass implementation. Additionally, the lack of a proper integration of Real-Time Text (RTT) in India continues to restrict accessibility, with users relying on patchy or incomplete substitutes.

Existing systems are also not as inclusive because they tend to be restricted to either simple transcription or sign language interpretation. These methods, although beneficial, are not sufficient to overcome the whole communication gap.

VIII. LIMITATIONS

*Insufficient Carrier Support for RTT in India*

Real-Time Text (RTT) has been standardized in many developed nations, but in India, implementation is insignificant owing to the fact that the carriers have not integrated RTT protocols largely. Telecom operators have not largely implemented the RTT protocols, limiting direct text-based communication via telephone networks. This absence of infrastructure inhibits the large-scale rollout of RTT and makes users reliant on third-party apps.

*Human Intermediary Dependence*

Existing solutions like relay services use human operators to interpret or transcribe speech in real time. Although effective, this is a cause of great concern with regard to user privacy, data security, and confidentiality, especially4\tin the case of sensitive conversations. What's more, human-operated services are not always accessible 24/7 and cannot scale services are not always accessible 24/7 and cannot scale efficiently to accommodate a large group of people, compromising reliability.

*Very Expensive Assistive Devices*

Most current communication aids are advanced and need specialized equipment like TTY machines, video relay devices, or sophisticated hearing aids. These are normally pricey, hard to replace, and non-portable, leading to economic and utilitarian limits in their adoption. Consequently, only a minimal portion of the users can purchase them, leaving the overwhelming majority of hearing-impaired persons with no efficient alternatives.

*Limited Awareness and Adoption*

Much of the population is still unaware of assistive technologies like RTT, STT, and TTS. In developing countries, limited awareness campaigns, inadequate training, and low exposure to such facilities hinder uptake. Even if solutions are available, users might not be aware of how to access or utilize them optimally, resulting in extremely poor rates of utilization.

*Partial Accessibility Features*

Most existing solutions address one channel of communication, for example, text transcription or sign language interpretation, but do not deliver a full, end-to-end calling experience. This establishes partial accessibility, where the user is able to comprehend the conversation but is unable to respond appropriately, causing a break in communication.

*Challenges of Language and Accent*

India is a multilingual nation with hundreds of local languages and dialects. Existing transcription technology is usually unable to properly capture these variations, especially if the users have strong accents or employ code-mixed speech (e.g., Hindi-English code mixing). This lowers the accuracy of transcriptions and causes miscommunications in conversations.

*Network Dependency*

Most AI-driven solutions require high-speed internet connectivity to work effectively. In areas with poor network coverage such as rural or semi-urban regions, these solutions will become less effective or non-functional. This renders solutions for accessibility less inclusive because they end up not reaching the people who need them most in remote areas.

*Lack of Standardization Across Platforms*

Existing accessibility tools are frequently fractured, with programs

Privacy, cost, and lack of awareness further hinder the usage of existing solutions, which are less effective on a daily basis for the hearing-impaired community.

With this in mind, the creation of an end-to-end accessible AI-driven communication platform is imperative. Through the integration of speech-to-text, text-to-speech, real-time messaging, and add-on sign language interpretation, the solution can develop a smooth, scalable, and user-friendly communication platform. This type of platform does not require human interpreters, saves on costs, and can be adapted to various languages and accents to promote usability across varying user segments.

Its implementation may greatly enhance access to education, healthcare, professional communication, and even emergency response situations. In addition to accessibility, it also enhances digital equality and social inclusion, with the hearing-impaired being able to engage in conversations without reliance on specialized equipment or expensive services. Finally, this technology opens the door to a sustainable, low-cost, and universally accessible model of communication, bridging the divide between the non-hearing and hearing communities.

XI. REFERENCES

1. SF-Net: Structured Feature Network for Continuous Sign Language Recognition arXiv

2. Multi-View Spatial-Temporal Network for Continuous Sign Language Recognition (Ronghui Li, Lu Meng) arXiv

3. Sign Language Recognition and Translation: A MultiModal Approach Using Computer Vision and Natural Language Processing (Jacky Li et al., RANLP 2023) ACL Anthology

4. Smartphone speech-to-text applications for communication with profoundly deaf patients (PubMed) PubMed

5. Advancements in Indian Sign Language Recognition Systems: Enhancing Communication and Accessibility for Deaf and Hearing Impaired (Asian Journal of Electrical Sciences, 2023) The Research Publication

6. Indian Sign Language to Text/Speech Translation (IJRASET, 2025) IJRASET

7. Towards Real-Time Sign Language Recognition and Translation on Edge Devices (ACM Multimedia) ACM Digital Library

8. Real-Time Sign Language Recognition in Digital Meetings (IJRASET) IJRASET

9. Shruti: An Embedded Text-to-Speech System for Indian Languages (Mukhopadhyay et al.) ResearchGate

10.Multimodal Real-Time Gesture Recognition for Sign Language (IRJAEH, 2025) irjaeh.com

11.Multimodal Real-Time Language Translator with Sign Language Recognition (Kapil et al.) IJISRT

12.Text-to-Speech (TTS) System for Deafened and Vocally Impaired People in Native Language (Tamil, etc.) SAGE Journals

13.BhashaBlend: Facilitating Multilingual comprehension of videos through NLP for Deaf and Hearing Impaired users (IJRASET) IJRASET

14. The Role of Available Technology Text to Speech in the Deaf Disabled Communities because of Brain Inflammation (Springer) SpringerLink