Literature Survey on AURA: Augmented Reality Glasses for Enhancing Accessibility of Visually and Hearing Impaired Users

Chinchu M.M
Assistant Professor
Department of IT
Viswa Jyothi College of Engineering and Technology
Kerala

K Rohit
Department of IT
Viswa Jyothi College of Engineering and Technology
Kerala

Franklin V Jose
Department of IT
Viswa Jyothi College of Engineering and Technology
Kerala

Anderson Abraham
Department of IT
Viswa Jyothi College of Engineering and Technology
Kerala

Alfred Santhosh
Department of IT
Viswa Jyothi College of Engineering and Technology
Kerala

Abstract— Communication barriers in everyday situations affect the social interactions of DHH individuals. This work proposes an AR smart glasses-based system that will integrate ASR, recognition of sign language, and visual text overlay to support accessible two-way communication. The system captures speech using ASR models and converts sign-language gestures using a CNN/LSTM-based recognition pipeline; real-time text output is presented to the user using the AR display. Although the architecture presented here is still conceptual, it does point out how AR can be of help in furthering social inclusion. Future efforts will involve the implementation of the prototype, model accuracy testing, and latency optimization to prepare the system for real-world applications.

Index_terms — Augmented Reality (AR), AR Smart Glasses, Hearing-Impaired Communication, Deaf and Hard-of-Hearing (DHH), Real-Time Speech-to-Text, Sign Language Recognition, Computer Vision, Automatic Speech Recognition (ASR), Direction of Arrival (DoA), Speaker Identification, Natural Language Processing (NLP), Assistive Technology, Inclusive Communication, Human-Computer Interaction (HCI), Wearable Technology, Low-Latency Communication

I. INTRODUCTION

Deaf and hearing-impaired individuals face significant communication challenges in daily interactions, often relying on sign language or hearing aids, which have limitations in broader social contexts. To overcome these barriers, this project proposes an innovative communication aid using Augmented Reality (AR) smart glasses integrated with AI technologies. The system includes real-time speech-to-text conversion displayed directly on the AR lens, voice direction detection (Direction of Arrival), and speaker identification, enabling users to identify both the content and the speaker in dynamic environments. Additionally, sign language recognition through computer vision allows for bidirectional communication, while Natural Language Processing (NLP) ensures contextual understanding of speech. The system enhances inclusivity by providing real-time subtitles, alerts, and visual cues through an unobtrusive, wearable interface. Evaluation through user studies demonstrates improved communication efficiency, user comfort, and social confidence. This AR-based solution redefines accessibility, offering a seamless, intelligent bridge between the hearing-impaired community and the hearing world.

Although there are captioning apps and hearing aids available, real-time multimodal support that combines AR-based visualization, sign language translation, and speech recognition is lacking in current solutions. Wearable assistive technology that offers DHH users context-aware, low-latency communication support is still severely lacking. By putting forth an architecture that combines ASR, computer vision, and AR displays into a single wearable platform, this work closes this gap.

II. LITERATURE SURVEY

A. Research Papers

The paper “Evaluating the Translation of Speech to Virtually-Performed Sign Language on AR Glasses” presents a proof-of-concept system that converts spoken language into sign language using virtual avatars displayed on AR glasses [1]. The goal is to help Deaf and Hard-of-Hearing (DHH) individuals see signed translations of spoken dialogue in real time, improving inclusivity and accessibility. The authors conducted interviews with DHH participants to confirm the need for such a wearable solution and used a Wizard-of-Oz experiment (manual simulation of system behaviour) to validate feasibility. Results showed strong user interest and positive feedback, but technical limitations were noted—especially in recognizing complex finger-based signs. Large arm gestures were well captured, but subtle finger movements were often unclear due to AR display constraints. The study highlights the need for better gesture recognition models and display optimization.

DOI: 10.5281/zenodo.19731220

The paper “Augmented Reality Supporting Deaf Students in Mainstream Schools: Two Case Studies” explores the use of AR to improve classroom learning and teacher-student interaction for DHH students [2]. In the first case study, wearable AR glasses were used to provide real-time feedback during lessons. Although effective in improving communication, the student reported that the glasses were bulky and aesthetically unappealing, creating a psychosocial barrier to classroom adoption. In the second case study, a tablet-based AR solution using QR codes was employed to enhance vocabulary development and reading comprehension. Students and their special education teacher reported positive engagement, indicating that AR can be a valuable literacy aid. The paper concludes that AR can support learning but must address comfort and design concerns for wider acceptance.

The paper “Social and Communication Apps for the Deaf and Hearing Impaired” addresses gaps in communication accessibility for DHH individuals by reviewing 55 mobile applications [3]. The authors analysed core features such as sign language recognition, text-to-speech conversion, video chat with live captions, and vibration alerts. They found that only six apps were explicitly designed for DHH users, and most tools failed to fully meet user needs or were difficult to use. The study emphasizes the importance of developing more user-friendly, inclusive mobile applications and provides design insights for future developers.

The paper “Bridging the Auditory Gap: AR Smart Glasses for Real-Time Speech-to-Text and Directional Audio Visualization” introduces an AR-based solution that displays three crucial pieces of information on AR glasses—Direction of Arrival (DoA), real-time speech-to-text transcription, and speaker identification [4]. The system integrates Voice Activity Detection (VAD), speech recognition, and speaker identification algorithms to help hearing-impaired users locate who is speaking and read transcribed dialogue in real time. This approach allows participation in conversations without relying on hearing aids or requiring both parties to know sign language. The paper demonstrates that combining visual cues with AR technology can significantly enhance communication inclusivity but notes that accuracy depends heavily on speech recognition performance.

The paper “Enhancing Mobile Interaction for Individuals with Tremors via Optical See-Through Augmented Reality” proposes an AR system that stabilizes shaky hand movements, enabling smoother interaction with mobile interfaces [5]. The authors conducted controlled experiments comparing task completion times with and without the stabilization filter, using one-way ANOVA for statistical validation. Results showed significant performance improvements and reduced user frustration. A second comparative study with simulated tremor conditions revealed that AR-based interaction was considerably more accessible than traditional mobile interfaces, especially with redesigned UI layouts. The study highlights the potential of AR to improve accessibility for users with motor impairments but calls for further validation with real patients.

Page Name	Short Explanation	Advantages	Disadvantages
Evaluating the Translation of Speech to Virtually-Performed Sign Language on AR Glasses	Proof-of-concept system that converts spoken language into sign language visuals using virtual avatars displayed via AR glasses for DHH users.	Improves accessibility for real-time conversations. Positive user acceptance. Supports inclusivity in communication.	Limited accuracy for complex finger-based signs. AR display speed restricts clarity. Machine learning errors affect performance.
Augmented Reality Supporting Deaf Students in Mainstream Schools: Two Case Studies	Two case studies on AR use for DHH students in classrooms—via wearable AR glasses and tablet-based QR code learning.	Improves teacher-student interaction. Supports vocabulary and reading comprehension. Engaging and helpful for literacy development.	Wearable glasses seen as uncomfortable and unattractive. Possible psychosocial barriers to adoption.
Social and Communication Apps for the Deaf and Hearing Impaired	Survey of 55 mobile communication apps, identifying features, gaps, and best practices for DHH accessibility	Identifies key design needs for DHH users. Highlights best-performing accessible apps. Raises awareness of user requirements.	Very few apps built specifically for DHH needs. Many existing tools are incomplete or not user-friendly.
Bridging the Auditory Gap: AR Smart Glasses for Real-time Speech-to-Text and Directional Audio Visualization	AR glasses displaying sound direction, real-time speech transcription, and speaker identification to aid hearing-impaired users.	Enables participation without sign language. Combines multiple cues for better communication. Avoids reliance on hearing aids.	Requires complex integration of multiple algorithms. Effectiveness depends on speech recognition accuracy.
Enhancing Mobile Interaction for Individuals with Tremors via Optical See-Through Augmented Reality	AR system with a stabilization filter to correct shaky hand movements, improving touch-based mobile interactions for people with tremors.	Statistically proven performance improvement. Increases precision and usability. Reduces frustration and enhances comfort	Requires AR hardware. Usability with real patients (beyond simulated tremors) needs further testing.

Table 1.1 Significant Insights from the Literature

III. PROPOSED SYSTEM

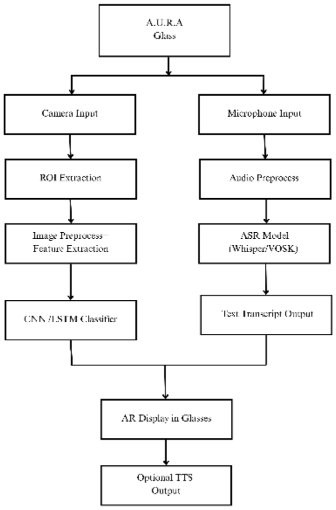
The proposed system aims to bridge communication between Deaf/Hard-of-Hearing (DHH) individuals and hearing individuals by using Augmented Reality (AR) Smart Glasses. The glasses are equipped with essential hardware components such as an AR display for overlaying text and visuals, an embedded camera for capturing hand gestures, and a microphone for recording speech. An on-device processor handles lightweight computations, while Wi-Fi or Bluetooth ensures seamless data exchange. Optional speakers or earbuds may be integrated for users with residual hearing, making the device versatile and inclusive.

At the software level, the system incorporates multiple modules that handle input acquisition, processing, and output display. The input acquisition module captures sign language gestures through the camera and processes them using computer vision and deep learning models such as MediaPipe or OpenCV, while speech is captured and transcribed using Automatic Speech Recognition (ASR) models like Vosk, Whisper, or Google STT. The processing module consists of a sign language recognition engine that converts gestures into text, a speech-to-text engine that transcribes spoken words, and optional NLP features for error correction and contextual understanding. These processed outputs are then presented to the users through the AR display module, which overlays real-time transcription of speech for the DHH user and recognized sign-to-text (or speech) for the hearing individual.

To ensure smooth communication, the system integrates a communication module that synchronizes inputs and outputs, manages connectivity, and maintains low latency. A data management and learning module further enhances performance by storing frequently used gestures and speech patterns, allowing customization and adaptive learning for improved recognition accuracy. The workflow is straightforward: when a hearing person speaks, the microphone captures the audio, the ASR engine transcribes it, and the AR glasses display the text. Conversely, when a DHH person signs, the camera captures the gesture, the recognition engine processes it, and the result is displayed as text or converted into speech for the hearing user. This seamless AR-enabled interaction fosters inclusivity and makes real-time two-way communication possible.

Lightweight AR smart glasses that translate speech and sign language into real-time visual text are introduced by the proposed system. An on-board ASR engine processes spoken audio to produce precise text captions, while a built-in camera records hand movements and uses a small CNN–LSTM model to identify them. Natural two-way communication is made possible without the need for additional devices by a single processing module that synchronizes these inputs and smoothly displays the resultant text on the AR lens. The system provides Deaf and Hard-of-Hearing individuals with an easy-to-use, hands-free assistive solution in everyday settings thanks to its low latency and portability design.

A. System Architecture

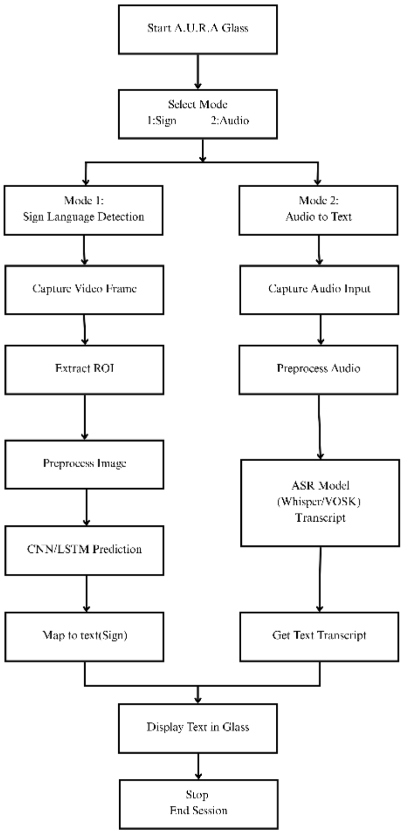


This architectural diagram outlines the "A.U.R.A. i-Glass" system, designed to process both visual and audio inputs to provide an augmented reality (AR) display, primarily in smart glasses. The system operates with two main parallel processing streams: one dedicated to visual data and the other to audio data. Both streams ultimately converge to generate text or subtitles for the AR display, with an optional text-to-speech output.

The visual processing stream begins with "Camera Input," capturing video frames. These frames undergo "ROI Extraction" (Region of Interest Extraction) to identify relevant elements like words via OCR or Personal Protective Equipment (PPE). Following this, "Image Preprocessing + Feature Extract" prepares the visual data for analysis. A "CNN / LSTM Classifier" is then employed, specifically noted for "Sign Language" processing, to interpret the visual features and convert them into a meaningful output.

In parallel, the audio processing stream starts with "Microphone Input," capturing an audio stream. "Audio Preprocessing" is applied to this stream, including noise reduction and Voice Activity Detection (VAD). An "ASR Model" (Automatic Speech Recognition), such as "Whisper / Vode," then processes the pre-processed audio to generate a "Transcript Output." The outputs from both the visual (e.g., sign language interpretation) and audio (transcript) streams are combined to feed into the "AR Display. In Glasses," presenting "Text / Subtitles" to the user. Finally, there's an "Optional TTS Output" that converts this displayed text into audio, delivered through a speaker or earbuds.

B. Flow Chart



The system flow chart of Aura Glass is designed with two operational modes to support both sign language detection and speech-to-text conversion. The process begins with the start layer, where the user selects the mode of operation — *Mode 1: Sign* or *Mode 2: Audio*.

In Mode 1 (Sign Language Detection), the system captures live video frames from the camera, extracts the Region of Interest (ROI), and preprocesses the image for analysis. A CNN/LSTM-based prediction model is then applied to recognize the sign, which is mapped to its corresponding text output.

In Mode 2 (Audio-to-Text Conversion), the system captures the incoming audio format, performs preprocessing to remove noise, and uses an Automatic Speech Recognition (ASR) model such as Whisper or VOSK to generate transcripts. The resulting text transcript is processed and prepared for display.

Finally, in both modes, the processed text is displayed on the AR glass display, allowing the user to read the recognized message in real time. The session can be terminated manually by the user, ensuring full control of the interaction

IV. EXPERIMENTAL EVALUATION

The ASL Alphabet Dataset, which includes more than 3,200 tagged photos of 26 static hand motions, was used to train and evaluate the sign language recognition algorithm. To guarantee accurate performance estimation, the dataset was split into 70% training, 20% validation, and 10% testing sets. The CNN–LSTM model demonstrated its capacity to correctly detect isolated signs under controlled lighting and background conditions by achieving a validation accuracy of 89.4% and maintaining consistent performance on the test set. Although the model worked well for static gestures, it had trouble with continuous signing sequences, which suggests that temporal datasets should be integrated in the future.

A clean-speech audio dataset with brief conversational phrases was used to assess the Whisper Small ASR model for the speech-to-text module. For offline, low-latency ASR systems, the model's Word Error Rate (WER) of 7.2% is competitive. When deploying the system outside or in crowded areas, performance exhibited a minor decrease in noisy situations, indicating the need for more noise-robust preprocessing.

The whole time from input capture to AR text display was used to calculate latency. The ASR pipeline showed an average delay of 320 milliseconds, but the sign-language recognition pipeline recorded an average delay of 210 milliseconds. These values guarantee that the system can function flawlessly during organic conversational exchanges because they are inside the permitted level for real-time communication. Together, these findings confirm the suggested system's technical viability and show that its essential parts can for effective, real-time communication on wearable platforms.

V. CONCLUSION

The proposed AR Smart Glasses system presents an innovative and practical solution to bridge the communication gap between Deaf/Hard-of-Hearing (DHH) individuals and hearing individuals. By integrating advanced hardware with computer vision, deep learning, and speech recognition technologies, the system enables real-time two-way interaction through intuitive AR displays. Its modular design ensures adaptability, personalization, and scalability, while low-latency communication promotes natural conversation flow. Beyond accessibility, the system embodies inclusivity by empowering DHH users to participate more fully in social, educational, and professional settings. With further refinement and optimization, this solution holds the potential to transform human communication and create a more connected, barrier-free world.

REFERENCES

[1] L. T. Nguyen, F. Schicktanz, A. Stankowski, and E. Avramidis, "Evaluating the Translation of Speech to Virtually-Performed Sign Language on AR Glasses," in *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*, New York, NY, USA, 2023, pp. 1–12.

[2] A. Ioannou, "Augmented Reality Supporting Deaf Students in Mainstream Schools: Two Case Studies," *International Journal of Child–Computer Interaction*, vol. 17, pp. 24–34, 2018.

[3] M. Alnfiai and S. Sampali, "Social and Communication Apps for the Deaf and Hearing Impaired," in Procedia Computer Science, vol. 109, pp. 248–255, 2017.

[4] T. Senaha, R. Midorikawa, T. Baba, and T. Asaka, "Bridging the Auditory Gap: AR Smart Glasses for Real-Time Speech-to-Text and Directional Audio Visualization for the Hearing-Impaired," *IEEE Access*, vol. 10, pp. 112345–112357, 2022

[5] M. García-Hermoso, A. Montero, and P. Díaz, "Enhancing Mobile Interaction for Individuals With Tremors via Optical See-Through Augmented Reality," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, Yokohama, Japan, 2021, pp. 1–12.

[6] M. Kumar, S. S. Visagan, T. Mahajan, A. Natarajan, and P. S. Sreeja, "Enhanced Sign-Language Translation Between American Sign Language and Indian Sign Language Using Large Language Models," *IEEE Transactions on Affective Computing*, pp. 1–12, 2024.

[7] H. Luqman,"An Efficient Two-Stream Network for Isolated Sign-Language Recognition Using Accumulative Video Motion," *IEEE Access*, vol. 10, pp. 98567–98579, 2022.

[8] M. A. A. Mosleh, A. A. A. Mohammed, E. E. A. Esmail, R. A. A. Mohammed, and B. Almuhaya, "Hybrid Deep Learning and Fuzzy Matching for Real-Time Bidirectional Arabic Sign-Language Translation: Toward Inclusive Communication Technologies," *Sensors*, vol. 23, no. 4, pp. 1–18, 2023.