International Journal on Emerging Research Areas (IJERA) *ISSN:2230-9993*
DOI : 10.5281/zenodo.14714650

AI-Powered Assistive Communication Software for the Deaf

Honey Thomas
Dept. of Computer Science
Amal Jyothi College of Engineering,(Autonomous)
honeythomas2025@cs.ajce.in

Linna Benny
Dept. of Computer Science
Amal Jyothi College of Engineering,(Autonomous)
linnabenny2025@cs.ajce.in

Saya Nezrin
Dept. of Computer Science
Amal Jyothi College of Engineering,(Autonomous)
sayanezrin2025@cs.ajce.in

Navya Neethi S
Dept. of Computer Science
Amal Jyothi College of Engineering,(Autonomous)
navyaneethis2025@cs.ajce.in

Niya Joseph
Dept. of Computer Science
Amal Jyothi College of
Engineering,(Autonomous)
niyajoseph@amaljyothi.ac.in

Abstract- The development of AI-powered assistive communication software is crucial for improving interaction for individuals with speech difficulties or those who are deaf. Current Augmentative and Alter- native Communication (AAC) devices often experience latency, limiting the effectiveness of real-time communication. This paper explores advanced AI technologies such as real-time speech-to-text conversion, predictive text, and auto-complete functionalities, which are designed to reduce communication delays and enhance fluency. Addition- ally, features like text-to-speech synthesis and image-based communication tools foster more seamless interactions, while environmental sound alerts and sentiment analysis provide contextual awareness. The integration of adaptive learning enables personalized experiences, allowing the software to cater to the unique needs of each user. By leveraging these innovations, the pro- posed system aims to empower individuals with communication impairments, providing them with a more intuitive and independent way to engage with their surroundings and improve their quality of life.

Key Words: RNN - Recurrent Neural Network, CNN - Convolution Neural Network

I. Introduction

AI-powered assistive communication software is designed to enhance accessibility and interaction for deaf individuals by leveraging advanced AI and machine learning. It provides real-time speech-to-text conversion, enabling users to read spoken language instantly during conversations or public announcements. Additionally, it offers critical environmental sound alerts for sounds like doorbells and alarms, notifying users through visual cues and customizable vibrations. Context-aware phrase prediction aids in seamless conversation by suggesting relevant responses based on context, while flexible notification settings allow users to tailor alerts to suit different environments. Multi-device synchronization ensures accessibility across mobile and wearable devices, integrating with smart home systems for real-time alerts. To facilitate communication, a text-to-speech feature enables users to vocalize typed messages, empowering them to interact effectively in diverse settings. Altogether, this software fosters greater inclusivity and autonomy, helping deaf individuals navigate their surroundings with confidence.

II. Literature Survey

Recent research highlights the potential of AI-generated communication aids, such as “speech macros,” which serve as shortcuts to facilitate quicker interactions. These tools can significantly

reduce the physical and cognitive effort required for communication, making them beneficial for AAC users [1]. However, user feedback emphasizes the necessity for customization, ensuring that AI outputs are contextually appropriate and aligned with individual communication styles. Furthermore, the adaptability of AI systems plays a crucial role in enhancing AAC functionalities. By learning from user interactions, AI can improve aspects such as word prediction, speech recognition, and the generation of relevant phrases. This adaptability enables a more personalized communication experience, catering to the diverse needs of users [2]. Additionally, context-aware systems have emerged as innovative solutions that enhance multi-turn dialogue capabilities. By leveraging user input along- side contextual information, these systems allow for more natural and efficient communication, reducing the effort required from users.

Research on Environmental Sound Recognition (ESR) for Deaf and Hard of Hearing (DHH) users highlights key challenges and advancements. Studies emphasize the need for mobile-friendly ESR systems that operate efficiently on devices with limited processing power. Personalization is crucial, as existing systems often lack the ability to adapt to users' specific sound environments. Additionally, improving sound classification accuracy through larger datasets and advanced neural networks, like Convolutional Neural Networks (CNN), remains a priority. ESR systems for home use also show potential in providing real-time alerts for crit-ical sounds.

Paper	Model	Methods	Datasets	Technologies
Mobile Sound Recognition for the Deaf and Hard of Hearing[7]	Environmental Sound Recognition (ESR) using machine learning, personalized knowledge base	- Audio Feature Extraction - Classifiers - Spectrogram Visualization - Group Pertinence Index (GPI)	- 300 audio records across 30 sound classes	Platform: Android (on-device processing). - Programming Language: Java. - Real-time processing: All sound processing and classification occur on the mobile device.
ProtoSound: A Personalized and Scalable Sound Recognition System for Deaf and Hard-of-Hearing Users[8]	Prototypical Networks for few-shot classification, using nearest neighbour and Euclidean distance for sound recognition.	- Sound feature extraction - Open-set - Real-time on-device training and prediction.	Dataset 1: 4.5 hours of recordings, 22 sound classes, collected in 21 real-world locations.	- MobileNetV2 CNN architecture - PyTorch - Android devices.

Table 1: Comparison of Environmental Sound Recognition Models

Features	Model	Methods	Datasets	Technologies
NaturalSpeech: End-to-End Text-to-Speech Synthesis with Human-Level Quality [3]	RNN-based Deep Speech 2.	AMR parsing for summarization.	CNN/DailyMail corpus, speech data.	Deep Learning, AMR, GooglTTS
Real-Time Speech-To-Text / Text-To-Speech Converter With Automatic Text Summarizer using Natural Language Generation And Abstract Meaning Representation[4]	Deep Speech 2, AMR Parsing.	Speech recognition, text summarization, NLG.	CNN/DailyMail, AMR-based summarization data.	Deep learning, CTC, Batch Normalization.

Table 2: Comparison of Speech-to-Text Models

Paper	Model	Method	Dataset	Technologies
Contextualized End-to-End Speech Recognition with Contextual Phrase Prediction Network [5]	Contextualized CTC/AED, Transducer	Contextual phrase prediction network, Multi-head attention (MHA)	LibriSpeech, GigaSpeech	Deep Biasing, BLSTM, LayerNorm, SpecAugment
Deep Keyphrase Generation [6]	RNN, Copy-RNN	Encoder-Decoder framework with Copying Mechanism	Scientific Publications	Deep Learning, Recurrent Networks

Table 3: Comparison of Phrase Prediction Models

III. Methodology

This section outlines the methodology adopted for developing key functionalities in an assistive communication system: Environmental Sound Alerts and Enhanced Real-Time Speech-to-Text and Text-to-Speech Conversion. The system aims to improve accessibility for individuals with hearing and speech impairments, leveraging advanced machine learning models, real-time processing, and cross-platform integration. Specific datasets and APIs play a pivotal role in ensuring the accuracy and efficiency of sound detection, speech recognition, and speech synthesis tasks. These features are designed to be lightweight yet powerful, ensuring seamless operation across devices. Together, these components form an integrated solution to support independent communication and engagement in diverse settings.

A. Environmental Sound Alerts

A comprehensive, on-device sound recognition system that will perform sound processing—including feature extraction, classification, and notification—entirely on the mobile device. By eliminating cloud dependencies, the planned system is designed to provide uninterrupted, real-time sound alerts and address the open-set problem, accommodating unpredictable sounds in uncontrolled environments.The implementation approach includes creating a Personalized Knowledge Base (KB), which will allow users to record and classify sounds specific to their surroundings.Techniques for implementing the feature are informed by existing systems that use similar knowledge base structures, which enhance accuracy and provide

adaptability to new sounds as they are recorded by the user.To classify sounds, the planned system will leverage machine learning algorithms—such as Nearest Neighbour, Naive Bayes, Bayes Network, and Random Forests—for diverse audio classification capabilities. The system will also feature sound visualization via spectrograms to help users recognize when a sound has been captured or classified.

For difficult-to-recognize sounds (e.g., emergency alarms), the system will likely use prototypical networks, a few-shot learning model, which has been successful in recognizing rare or user-specific sounds in similar implementations. Our system is enabling the application to store specific sounds in the database and generate alerts or warnings when similar sounds are detected in the environment[9]. The method allows for classification based on minimal samples and could be particularly effective for user-recorded sounds. Implementing nearest neighbor classification with Euclidean distance metrics will help address open-set classification, improving the system's accuracy in identifying unknown sounds.The MobileNetV2 CNN as a foundation due to its lightweight nature, adaptability,real-time and low-latency sound processing directly on devices. By implementing on-device sound recognition, ensure privacy and efficiency by minimizing data dependency on external sources.Initial testing plans involve two datasets: ESC-50 and UrbanSound8k. The widely-used datasets provide diverse sound classes suitable for model training and benchmarking. ESC-50 contains environmental sounds across 50 categories, while UrbanSound8k offers urban sound recordings, both valuable for preliminary model accuracy testing and adjustment.The methods will be refined based on feedback and performance results in real-world testing with individuals who have hearing impairments, ensuring an accessible, privacy-focused solution for real-time environmental sound alerts.

B. Enhanced Real-Time Speech-to-Text and Text-to-Speech Conversion

Real-time speech-to-text conversion using the speech-to-text package, enabling users to capture and transcribe spoken input accurately. When listening mode is activated, the system initializes the speech recognition service, continuously interpreting spoken words into text while capturing a confidence score to help users gauge transcription accuracy. The transcribed text is stored in a variable, enabling easy access and manipulation, including a convenient copy-to-clipboard feature for quick sharing. This functionality highlights efficient state management and real-time feedback, emphasizing Flutter's suitability for assistive applications.

Text-to-speech (TTS) capabilities using the flutter-tts package, allowing users to convert text input into spoken language. The speak function initializes TTS settings by configuring language and pitch before vocalizing the provided text. Users input their desired text through a TextFormField and activate the TTS feature by pressing a button that triggers the speak function [10]. Additionally, the design includes a navigation option to return to the previous menu, improving app flow. This stream- lined implementation demonstrates how TTS can enhance accessibility in applications, facilitating audio feedback for improved user interaction. In addition to core TTS capabilities, the implementation enhances user engagement by providing audio feedback, which is particularly beneficial for individuals with visual impairments or reading difficulties. This integration of speech capabilities significantly improves accessibility, allowing users to access written information in an auditory for- mat, thereby enhancing their overall interaction with the application.

C. Phrase Prediction

The methodology developed by Rui Meng et al. employs a Recurrent Neural Network (RNN) Encoder- Decoder model with a copying mechanism to ad- dress challenges in keyphrase extraction, partic- ularly the prediction of both present and absent keyphrases. This model improves upon traditional extraction techniques by using deep learning to cap- ture the semantic meaning of text. It was trained and tested on several datasets, including Inspec, Krapivin, NUS, SemEval-2010, and KP20k, which consist of scientific publications and their author- assigned keyphrases. The CopyRNN model incor- porates an attention mechanism that allows it to dynamically focus on relevant portions of the input text, improving the generation of meaningful keyphrases. Furthermore, the copying mechanism enables the model to select important words from the source text and predict out-of-vocabulary or absent keyphrases, which traditional models, such as TF-IDF, TextRank, and even supervised modelslike KEA and Maui, fail to do.

In comparison to these models, CopyRNN performed significantly better, especially in handling absent keyphrases, achieving F1-scores of 0.34 (Top-5) and 0.32 (Top-10) on the KP20k dataset. The key to this improvement lies in the model's ability to infer keyphrases based on the context and semantic structure of the document, rather than relying solely on frequency or syntactic patterns. Traditional models are limited to extracting keyphrases that explicitly appear in the text, whereas CopyRNN can generate keyphrases that

summarize deeper meanings. This capability, combined with its attention and copying mechanisms, makes CopyRNN superior for tasks like keyphrase generation, text summarization, and information retrieval, as it can provide more accurate, contextually relevant predictions.

D. Context Aware Text Suggestion

"The Less I Type, the Better" [1] employs a user-centric methodology, conducting a study with 12 AAC users to evaluate AI-generated phrase sugges- tions using GPT-3 in various communication sce- narios. The study gathers qualitative feedback on how the system reduces typing effort and whether the suggestions align with the user's intent, tone, and communication style. It also focuses on fea- tures like speech macros and user control, ensuring the AI-assisted phrases feel personal. In contrast, KWickChat adopts a more technical methodology, using GPT-2 to implement a bag-of-keywords model for generating context-aware sentences. It incorporates conversation history and user per- sona to ensure coherent responses, and its effective- ness is measured through metrics like BLEU and Word Error Rate. Rather than direct user testing, KWickChat evaluates performance through envelope analysis and keystroke savings, emphasiz- ing sentence generation quality and efficiency for multi-turn dialogues.

IV. Result and Discussion

The assistive communication system aimed to enhance accessibility for individuals with hearing and speech impairments through machine learning-based Environmental Sound Alerts, speech-to-text (STT), and text-to-speech (TTS) functionalities. While the system performed well in controlled environments, its real-time effectiveness in noisy or unpredictable settings was inconsistent. The sound recognition component, designed to classify sounds using algorithms like Nearest Neighbour and Random Forests, showed high accuracy with personalized datasets but struggled in dynamic environments with unfamiliar or unexpected sounds. Although the system allowed users to create a personalized knowledge base (KB) of sounds, it did not generalize well to new or open-set sound categories, reducing its practicality in real-world applications. The STT component, which relied on the Google Web API, functioned effectively in low-noise envi- ronments but experienced latency and diminished accuracy in everyday scenarios with background noise or varied speech patterns, which hindered its real-time responsiveness. The TTS system, while capable of functioning offline, produced less natural speech, impacting its usability for seamless communication.

These challenges underscore the need for more adaptive and context-aware algorithms that can handle diverse soundscapes and improve real-time performance in uncontrolled environments. The reliance on cloud-based APIs for speech recognition introduced latency and privacy concerns, highlighting the importance of shifting towards more robust on-device processing to maintain performance and security. Improvements in sound recognition algorithms, particularly in their ability to handle open-set classification and adapt to unfamiliar environments, are crucial for enhancing the system's real-world functionality. Additionally, conducting extensive user testing across various real-life settings would provide valuable insights to refine the system's design, ensuring it better meets the needs of individuals with hearing and speech impairments. These refinements, combined with a focus on real-time adaptability and privacy, can significantly improve the system's effectiveness for daily use in a variety of environments.

V. Conclusion

This paper focuses on developing an AI-powered assistive communication software designed to enhance communication for individuals with hearing impairments. The software includes real-time speech-to-text conversion, text-to-speech synthesis, and environmental sound alerts. Additional features such as predictive text and phrase suggestions are implemented to improve the efficiency of communication. The software is designed to provide users with a personalized, seamless communication experience, enabling them to engage more effectively with their surroundings. The system leverages advanced AI technology to enhance independence and improve quality of life for individuals with hearingimpairments.

References

[1] Stephanie Valencia,Richard Cave,Krystal Kallarackal,Katie Seaver, Michael Terry,and Shaun K. Kane. 2023. "The less I type, the bet- ter": How AI Language Models can Enhance or Impede Communication for AAC Users. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems(CHI '23), April 23–28, 2023, Hamburg, Ger- many. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3581560

[2] Junxiao Shen, Boyin Yang, John Dudley, and Per Ola Kristensson. 2022. KWickChat: A Multi-Turn Dialogue System for AAC Using Context-Aware Sentence Generation by

Bag-of-Keywords. In 27th International Conference on Intelligent User Interfaces (IUI '22), March 22–25, 2022, Helsinki, Finland. ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3490099.3511145.

[3] "NaturalSpeech: End-to-End Text-to-Speech Synthesis With Human-Level Quality".Xu Tan , Senior Member, IEEE, Jiawei Chen , Haohe Liu , Jian Cong , Chen Zhang , Yanqing Liu ,Xi Wang , Yichong Leng , Yuanhao Yi , Lei He , Sheng Zhao , Senior Member, IEEE, TaoQin , Senior Member, IEEE, Frank Soong , Fel- low, IEEE, and Tie-Yan Liu , Fellow, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 46, NO. 6, JUNE 2024.

[4] "Real-Time Speech-To-Text / Text-To-Speech Converter with Automatic Text Summarizer using Natural Language Generation and Ab- stract Meaning Representation". P. Vijayaku- mar, Hemant Singh, Animesh Mohanty on In- ternational Journal of Engineering and Ad- vanced Technology (IJEAT) ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April, 2020. DOI: 10.35940/ijeat.D7911.049420

[5] Kaixun Huang, Ao Zhang, "Contextualized End-to-End Speech Recognition with Contextual Phrase Prediction Network," 2023.

[6] Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, Yu Chi, 'Deep Keyphrase Generation,' 2021

[7] Mobile Sound Recognition for the Deaf and Hard of Hearing LA Fanzeres, AS Vivacqua, LWP Biscainho arXiv preprint arXiv:1810.08707, 2018

[8] Protosound: A personalized and scalable sound recognition system for deaf and hard-of-hearing user's D Jain, K Huynh Anh Nguyen, S M. Goodman, R Grossman-Kahn, H Ngo, A Kusu- pati, R Du. Proceedings of the 2022 CHI Con- ference on Human Factors in Computing Sys- tems, 2022-dl.acm.org

[9] Ann Mary Babu, Anto K Thomas, Ashwin Se- bastian, Befin K Lalu, Dr. Jacob John, "As- sistive Technology for Deaf and Dumb", In- ternational Journal on Emerging Research Ar- eas (ISSN:2230-9993), vol.03, issue 01, 2023 doi:10.5281/zenodo.8210550

[10] Athulya Anilkumar, Abhinav V V, Aneeta Shajan, Anjana S Nair, Bini M Issac, Neenu R, "Image Descriptor for Visually Impaired", In- ternational Journal on Emerging Research Ar- eas (ISSN:2230-9993), vol.03, issue 01, 2023 doi:10.5281/zenodo.8210962