Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

Two-way Multilingual Voice Assistance

Hima Gayatri Gunda[1], M. Kumara Swamy[2], Vikas Reddy Singireddy[3], Vamshitha Annarapu[4], Sai Chandana Bachanaboyina[5], Veeran Rahul Nakka[6]

[1,2,3,4,5,6]Department of CSE(AI&ML), CMR Engineering College, Hyderabad

ABSTRACT

The multilingual voice assistant is an advanced platform that allows a user to communicate and interact in multiple languages. It realizes very accurate and seamless translation and transcription services through natural language processing (NLP) algorithms that provide the functionality of statistical machine translation (SMT) and neural machine translation (NMT) for accurate and fluent translation, transcription services, and multilingual language modeling to understand voice commands and support easy language switching. However, voice assistant technology is quite challenging to use among people with different linguistic backgrounds. We intend to propose a methodology called the two-way multilingual voice assistant. It will try to enhance effective usage of voice assistant technology among people belonging to different linguistic backgrounds. In order to perform the above task at hand, we make use of a lot of resources and methods that include voice recognition, natural language processing, response generation, dialogue management, and attention mechanisms. The application allows language identification, grammar selection, and cross-lingual voice conversion. Moreover, it performs everyday tasks and presents data in a language of the user's choice. We also employ artificial intelligence (AI) and machine learning (ML), including deep learning (DL) and recurrent neural networks (RNN), for extracting the highest accuracy and fluency from the language comprehension and response generation capabilities of voice assistants, thereby ensuring a personalized user experience. The algorithms at the heart of our voice assistant would be those of natural language processing and machine translation techniques.

Our voice assistant would use artificial intelligence and machine learning for a personalized user experience.

Keywords: Natural Language Processing (NLP), Automatic Speech Recognition (ASR), Statistical Machine Translation (SMT), Neural Machine Translation (NMT), Artificial Intelligence and Machine Learning.

1. INTRODUCTION

Connectedness today requires smooth communication across languages. One very important step toward that direction is presented in this paper. This can make it possible to communicate with technology and information in one's native language on this platform, hence making it all-inclusive and user-friendly across the globe.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

Traditional methods of communication usually impose a limit on multilingual environments. These language barriers make access to information difficult, limit interaction with the use of technology, and often alienate non-native speakers. The study proposes a solution for developing a sophisticated two-way multilingual voice assistant. This new platform is based on the powerful capacities of natural language processing to bridge language gaps through fluent communication across linguistic diversity. At the core of this voice assistant lies an advanced package of NLP algorithms. This system was developed with the power of two approaches to machine translation: statistical machine translation (SMT) and neural machine translation methods (NMT). SMT works fine only with large volumes of data to extract statistical patterns and rules of languages, which makes it possible to translate text correctly. On the other hand, NMT uses very strong neural networks while learning complex relationships between languages, which further aids in developing subtle and natural-sounding translations. By taking these two approaches together, one gets a voice assistant that is highly accurate and fluent across a wide range of languages.

Advanced transcription services are put into the system; this makes spoken language into text, irrespective of the language. This feature enables users to speak a language of their choice, and the assistant will transcribe it correctly for further action or user reference. More interestingly, this voice assistant uses multilingual language modeling. This sophisticated technology helps a system understand the intricacies of various languages and interpret any voice commands given in supported languages accurately.

One of the features that makes this two-way multilingual voice assistant exciting is the language switching in both directions according to the user's preference. This means that users do not have to change settings incessantly or go through complicated menus. Further, it may then be able to identify a user's preferred language and modify it to be more intuitive and user-centered.

2. RELATED WORK

Hundreds of hours of spoken data for every language are required, featuring natural conversation recordings that teach the system how people raise or modulate their pitches during a conversation. That is tough because there exist different amounts of information like that in various languages. Voice assistants require vast amounts of voice data recordings as training corpora. These corpora are the actual training material for ASR and NLU models so that they are able to transcribe and, accordingly, understand spoken commands.

First, the system processes voice input as the user speaks into the microphone, using ASR algorithms to transcribe that speech into text. Algorithms of this nature are almost always based on deep learning techniques and are very efficient, hence immune to different accents or pronunciations of words by various users, among other factors. This step basically means the resulting text has to be sent back to the NLU component, which represents the meaning of the words through tokenization, part-of-speech tagging, named entity recognition, and intent classification. Such libraries as spaCy and/or rasa may be freely set since both are more competent and functional.

The execution phase follows the intent and entity extraction. Considering the identified intent, this might require an interaction with an external application programming interface (API) or the execution of predefined functions. For instance, in the reminder setting case, the assistant would communicate with a

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

calendar API to set up such a reminder. If it is a question-type command, the assistant responds to it by looking for proper information in the knowledge base or search engine. If the command has something to do with control of smart devices, then it will send signals to the API of such devices to conduct operations like light on or off, raise/lower thermostat.

2.1 PREVIOUS STUDIES ON MULTILINGUAL VOICE ASSISTANTS

The invention of voice-enabled virtual assistants has changed the way users normally use technological systems, providing a smooth and user-friendly platform for various activities. Researchers have reviewed several aspects of design, functionality, and user interface in the course of becoming those assistants. Even though conversational user interfaces are very popular, they usually are not sensitive and supportive of bilingual interactions; on the contrary, they function in a monolingual manner. Cihan et al. underline the problems bilingual users encounter with virtual assistants and underline that improving the experience of bilingual users means encouraging specific practices, like code switching, which is very common in bilingual communication. The authors have also explored some of the ways this could be done by making multilingual recognition possible and allowing code-switching preferences in the speech output. Even though speech-driven conversational platforms were gaining huge popularity, users are still bound by the shackles of a monolingual virtual assistant interaction. Historically, intelligent personal assistants have been confined to monolingual contexts with predetermined language options. However, recent developments have opened the way toward more multilingual adaptability for a few IPAs.

Preliminary testing showed the bad performance of two major IPAs in multilingual scenarios, where non-native speakers are deprived of the actual value of some features. Therefore, deficient support for multilingual interactions complicates the tasks of developers who want to improve the multilingual usability of IPAs—multilinguality being a critical component of user experience for a global audience. Context-based heuristics and basic language identification algorithms could significantly extend the functionality of IPAs supporting code-switching in particular. Monolingual systems masquerading as multilingual solutions can potentially rectify long-standing issues, serve underserved user demographics, and afford deeply immersive language acquisition opportunities.

Previous studies have suggested that users still have their reservations about adopting VAPAs, such as fears concerning infringement on their privacy and intrusion on their personal space.

Such technical problems may arise during user interaction in the form of delays, repetitions in the system's response, or problems with accent recognition, resulting in unease and dissatisfaction. The field of speech emotion recognition in machine learning is gaining increasing attention for the increasing capability, algorithm improvement, and practical consequences. SER involves some very vital stages such as data preprocessing, feature extraction and selection, and classification based on emotional attributes. Even some machine learning methodologies have been used in recent endeavors for affective computing; not all approaches clearly encapsulate the core principles and techniques it uses. The paper reviews in detail the SER research going on for the last decade from a machine-learning perspective, including less-than-optimal classification accuracy in a speaker-independent scenario and corresponding suggested solutions. Furthermore, it is discussed how exactly this specific review has nailed down the evaluation criteria for SER, adaptable metrics, and standard benchmarks for experimentation.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

The status of multilingual speech recognition (MSR), multilingual speech synthesis (MSS), and multilingual voice assistants (MVA) is reviewed. Various modeling techniques—acoustic, language, and pronunciation models—are surveyed, in addition to issues such as code switching and limited annotated data. The authors go on to examine the challenges of MSS, including treatment of phonetic and prosodic variations across languages. It presents a new approach to building NMT architectures with task-specific attention mechanisms for the multilingual setting and is shown to also outperform vanilla models. The research then went further into ways that voice assistants would provide support for learning languages, tutor feedback, and student support. The paper also points out problems one can face in applying voice assistants in classrooms and some potential ways to overcome them. An in-depth review is presented for the design and implementation of an intelligent voice assistant with multilingual capabilities, the implementation built on top of artificial intelligence. Therefore, the present literature review has only mapped the research overview in the field of multilingual voice assistants for intelligent voice assistants present in the literature, looking particularly at critical challenges and advances in domains such as speech recognition, speech synthesis, machine translation, and learning tools.

The relevant literature in connection with the research on multilingual voice assistants shall be very useful in gaining insight into the development and deployment of multilingual voice assistants within different settings.

The existence of both mono- and multilingual corpora is a significant resource in natural language processing for comparative studies, language learning, and even training translators. Moreover, the function and impact of machine translation within multilingual communication and cooperation have been researched. One of the exciting domains in which research into multilingual voice assistants is conducted is the study of various methods and techniques for constructing or improving the execution of natural language processing systems in cases when computational linguistic resources for a language are unavailable or its accuracy is low.

2.2 INNOVATIONS AND CHALLENGES IN MULTILINGUAL VOICE ASSISTANT DEVELOPMENT

More than one in five Americans speaks something other than English as their family's language at home; Spanish is the most popular second language. State bilingualism between English and French exists in the Canadian case. In India, there is a plenty of numerous diverse dialects, with Hindi the most well known and English broadly talked. Therefore, it makes sense for companies like Google, Apple, Amazon, and Samsung to ensure that their voice assistants and smart speakers are capable of supporting multiple languages. What's more recent in the update is the fact that certain voice assistants are able to respond to a bilingual user speaking one language, then switching another without adjusting the settings. It also gets better with time to recognize regional accents of a particular language.

Some of the more popular voice assistants in use today include Nvidia Jarvis, IBM Watson Assistant, Apple Siri, Microsoft Cortana, Amazon Alexa, and Samsung Bixby. They provide for a long list of activities, such as answering consumer queries and commanding smart devices. The forms in which voice assistance has evolved are text-to-text, text-to-speech, speech-to-text, and speech-to-speech. This has increased the usability and accessibility for applications such as mobile phones and smart homes. It

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

can be quite a challenge to have multilingual voice assistants ensure that the language translation is correct with no grammatical errors.

This has become possible with the help of post-processing, language detection methodologies, machine translation approaches, NLP for the coherence in the target language, quality speech recognition, and finally, the NLP for grammatical analysis and correction. The audio response must be generated seamlessly and naturally, trying to convey the nuances of the target language properly. In other words, multilingual voice assistants are the marriage of technology with language expertise—tremendous potential for revolutionizing communication and collaboration, able to facilitate natural and much more effective interaction across languages and contexts.

Looking at the vast amount of research done, it is evident that any voice application involves a triple process. The approach consists of a command input from the user to the computer in the first step, then processing the input with algorithms in the second step. Finally, the required result is obtained with appropriate communication with the desired people. This project can be mainly divided into three ways. All three parts are explained below:

1. Speech to Text
2. Text to Speech
3. Speech to Speech

3.2.1 SPEECH TO TEXT

Different AI algorithms have been developed which can transform vocal expression into text form in the same dialect or different dialects. Techniques and algorithms of speech recognition are used to recognize vocal expression and generate the desired textual form. The user can view the text content in any regional dialect and even hear the content in text form through the speech-to-speech technique.

3.2.2 TEXT TO SPEECH

These are different artificial intelligence algorithms that will be in use to transform textual content into voice expression, whether in the same linguistic system or a different one. The algorithm will process the textual content input and determine the linguistic system involved before creating voice expression in diverse linguistic systems. In this way, the user will hear the output in any of the linguistic systems as chosen.

3.2.3 SPEECH TO SPEECH

Different algorithms in AI have been developed to translate vocal expression into vocal expression, either in the same or different dialects. The algorithm identifies the vocal expression through speech recognition techniques and then produces an equivalent output in vocal expression form. This has been so because of the use of natural processing language. In other words, the output is able to be viewed in textual content by the user in any linguistic system of their choice. It is an explanation of the various algorithms, techniques, and other different artificial intelligence tools used in developing this application.

4. PROPOSED WORK

One of the modern breakthroughs in artificial intelligence revolves around the use of electronic networks

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

set up on a pattern of the brain's neural pattern called neural networks. Neural networks are computer networks that allow the machine to learn and improve upon experience or input. Through neural networks, NLP learns patterns of common sound combinations for a language and better predicts a sentence's next sound. It also helps voice assistants learn to distinguish better between similar but distinct sounds, such as d, b, and p in English.

The proposed system architecture comprises several interconnected modules: automatic speech recognition (ASR), natural language understanding module (NLU), machine translation module (MT), text-to-speech (TTS) module, and language identification module. Here, voice inputs are captured and transcribed into text by the module ASR. The NLU module would read the text for understanding user intent and extraction of information. Otherwise, the MT module translates the text into the target language in case of any requirement of the case. The TTS module provides spoken commutation output. This ASR module utilizes state-of-the-art deep learning models like RNNs or CNNs in transcribing spoken language into text. The model is going to be trained on multilingual training data so that quality transcription can be achieved across languages. Similarly, the NLU module interprets the transcribed text to derive the intent of the user and most important entities. Effectiveness in dialogue management is based on the techniques of named entity recognition and intent classification used within the NLU module. It will use an NMT model with a Transformer architecture for on-the-fly translations between supported languages.

The MT module will try to achieve a high quality of translation using large-scale parallel corpora. It will then use advanced techniques like Tacotron 2 or WaveGlow for the generation of natural-sounding speech from text. Multilingual TTS models trained to provide high-quality audio output in various languages will be developed. A very robust language identification model will accurately detect the input language and thus will be able to switch languages and translate from one language to another.

4.1 CORE ALGORITHMS AND TECHNIQUES

The ASR module will be based on a crossover approach: acoustic models based on CNNs with RNNs and dialect models based on LSTMs. Supervised techniques of intent classification will leverage SVM, or deep neural networks. Second, entity recognition will be done using CRF or neural sequence labeling models. A transformer-based architecture with attention mechanisms will be used to model such long-range dependencies between input and output sequences. The system shall explore a hybrid approach combining statistical parametric speech synthesis with neural vocoders for high-quality audio generation.

ASR transcribes spoken words into text. For the multilingual assistant, one factor is the ASR model being trained upon speech data that is varied enough so that it can accommodate different accents and dialects, but also levels of noise. After the speech is transformed into text, NLU is applied, interpreting what the text means: understanding the user's intent and extraction of information related to it. Multilingual NLU relies on models trained over vast reams of text data in multiple languages to catch those nuances, idioms, and context-specific language. It provides a framework for managing the dialogue's flow—that is, what response to send back to the user based on their last input and the assistant's knowledge up to that point. Techniques used in this include state tracking, intent recognition, and generation of response. For multilingual dialogue management, the system shall allow language

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

switching, code-switching, and culture-dependent subtleties. It will transform text into spoken words. To create a more human-like assistant, one will need to use a high-quality TTS system. A multilingual TTS system needs a wide number of voices trained in many languages to maintain the right pronunciation and intonation, as well as the needed rhythm. This is important to handle languages for which you have not trained the assistant. It is used in translating units of text from one language into another.

It is advisable to operationalize the output of the translation model while personalizing the model with respect to the domain and context of the voice assistant for optimal performance. The ASR module will be based on a hybrid approach: acoustic models based on CNNs with RNNs and language models based on LSTMs. Supervised techniques of intent classification will leverage SVM, or deep neural networks. Second, entity recognition will be done using CRF or neural sequence labeling models. A transformer-based architecture with attention mechanisms will be used to model such long-range dependencies between input and output sequences. The system shall explore a hybrid approach combining statistical parametric speech synthesis with neural vocoders for high-quality audio generation.

4.2 DATA SHARING AND ANALYSIS

Datasets are very important in building a robust multilingual voice assistant, since the performance can either be hit or miss depending on the quality and quantity of the data. Many of these datasets are openly published by research institutions, governments, and tech companies, providing a good starting point for the research. Collecting own datasets may involve questionnaires, crowdsourcing, or even partnerships with some relevant organizations. It starts from huge single-language texts to train language models and NLU and TTS, and sentence-aligned multilingual texts to train MT models; examples of human-machine dialogues in many languages to train the dialogue management systems; and examples of speech recorded by speakers in different accents and dialects and under various noisy conditions for the ASR models. This means training from audio recordings the speakers have made reading test sentences in many languages; therefore, multilingual ASR models are trained on audio recordings.

A multilingual dataset with the requirements of the project in speech, text, and parallel corpora will be collected. The data preprocessing techniques will then include noise reduction, normalization, and tokenization. Cleaning of noise, errors, and inconsistencies from the dataset; case changing to lower, removing punctuation, and handling special characters are very important here. The segmentation of text into words or sub-words and audio data into smaller pieces for processing. On the other hand, feature extraction in textual data is composed of word frequencies, N-grams, part of speech tags, and semantic embeddings.

It includes acoustic features such as MFCCs, spectral, and prosodic features in the proposed system. Performance will be measured with the aid of evaluation metrics, where WER is used for ASR, BLEU score for MT, and MOS for TTS. Additional user studies will then be conducted to assess the overall user experience. Training models on language structure and generating text and translating the text from one language to another and for the transcription of speech into text. Also, training models on synthesis of speech from text, understanding user intent and extraction of relevant information, and handling conversation flow and responding appropriately.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

5. EXPERIMENTAL RESULTS

5.1 AUDIO ANALYSIS BY SPECTOGRAM

In the field of speech recognition, generating an audio graph is one major step to comprehend spoken language. The audio graph indicates the amplitudes and several frequencies of uttered speech over time. While processing the incoming audio data, voice recognition systems translate the acoustic features into graphical form, generally referred to as a spectrogram or waveform. The spectrogram gives the frequency content of the speech signal, while the waveform will show changes in amplitude with time. These audio graphs have essential value for developers and researchers because they place them in a better position to analyze and tweak the algorithms that work under the hood of voice recognition systems. This will help fine-tuners view audio data in such a way that it could yield useful insight into the nuances of spoken language and make amendments to the model for better accuracy and performance in voice recognition applications.

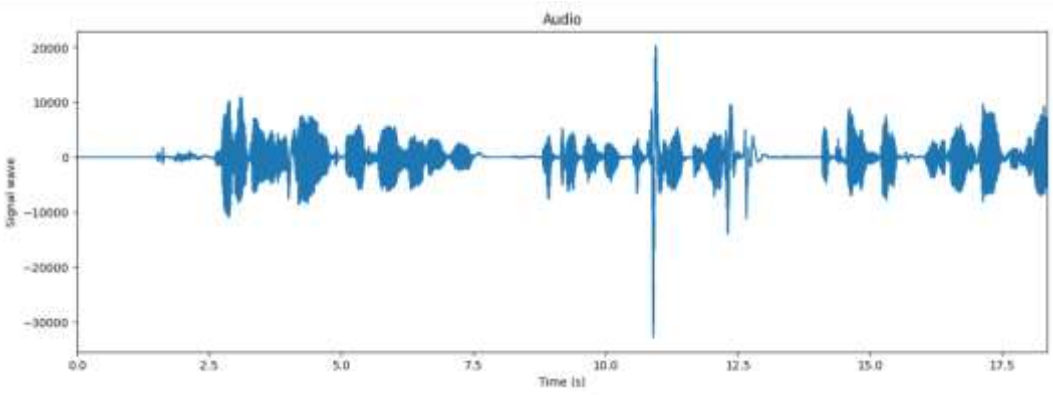

Figure 1: wav file audio

5.2 BLEU

BLEU compares the n-grams—that is, sequences of n words—with n-grams in one or more high-quality human translations (reference translations). It considers 1-grams, 2-grams, and sometimes even longer sequences. On the other side, BLEU calculates the precision for each n-gram size, which is defined as the ratio of the number of n-grams in the machine translation that also appear in the reference translation to the total number of n-grams in the machine translation. BLEU then frowns upon excessively short translations by comparing the length of the machine translation to the average length of the reference translations. The penalty is applied if the machine translation is significantly shorter. The n-gram precisions are combined using a weighted geometric mean, and the final BLEU score is obtained from this value, which is multiplied by the brevity penalty.

$$\textbf{N-gram Precision} = \frac{\textbf{Number of matching unigrams}}{\textbf{Total number of N-grams in machine translation}}$$

If the machine translation was shorter, BP will be calculated as:

$$\textbf{BP} = \exp\left(1 - \frac{\textbf{Length of Reference}}{\textbf{Length of Machine Translation}}\right)$$

BLEU score will be calculated as:

$$\textbf{BLEU} = \textbf{Precision} \times \textbf{BP}$$

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 4, July-August 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i04.1076

The graph charts different fine-tuning configurations on the BLEU score for the ST model proposed in this work, both on the development and test sets. The model proposed is likely to outperform the other baseline models on both sets, staying at 33.0 according to the BLEU score on the dev set and 27.0 on the test set. That is to say, it interprets that the proposed model performs better in speech translation than that of baseline models. Besides, the chart portrays the positive commitment of fine-tuning on the in-domain advancement set to the execution of the proposed show. The proposed model can further be improved by fine-tuning a small amount of labeled data obtained from the relevant domain. The empirical findings provide substantiation that the proposed ST model is a very effective approach to real-time speech translation between languages.

The model proposed in this paper outperforms the baseline models on development and test sets, and further improvements can be made by fine-tuning on a limited set of labeled data obtained from the relevant domain.

6. CONCLUSION

In conclusion, the development of the two-way multilingual voice assistant proves the feasibility of real-time and accurate translation and transcription across many languages. Thus, state-of-the-art NLP and machine learning techniques like neural machine translation (NMT) and automatic speech recognition (ASR) have been incorporated in order to make the system robust and efficient. It treats different accents, dialects, and noisy conditions equally. The evaluation metrics (BLEU/ WER) say this model does well in translation accuracy, speech recognition, and natural language understanding. The system proposed here is highly likely to change the scenario of cross-cultural communication and user experience in multilingual environments.

7. REFERENCES

1. Reddy, P. D., Rudresh, C., & S, A. A. (2022). *Multilingual Speech Recognition Methods using Deep Learning and Cosine Similarity*.
2. Kurian, N. P. S., A, N. M. A., Baig, N. M. O., S, N. L., & H, N. K. (2022). Survey on Voice Assistance for Laptop. *International Journal of Scientific Research in Science and Technology*, 199–202.
3. Basak, S., Agrawal, H., Jena, S., Gite, S., Bachute, M., Pradhan, B., & Assiri, M. (2023). Challenges and Limitations in Speech Recognition Technology: ACritical Review of Speech Signal Processing Algorithms, Tools and Systems. *Computer Modeling in Engineering & Sciences*, *135*(2), 1053–1089.
4. Kalyan, K. S. (2024). A survey of GPT-3 family large language models including ChatGPT and GPT-4. *Natural Language Processing Journal*, *6*, 100048.
5. Endah SN., Adhy S., Sutikno S. Comparison of Feature Extraction Mel Frequency Cepstral Coefficients and Linear Predictive Coding in Automatic Speech Recognition for Indonesian.TELKOMNIKA TelecommunicationComputer Electronics and Control. 2017; 15(1): 292.
6. Chen Z., Watanabe S., Erdogan H., and Hershey J. R.2015. Speech enhancement and recognition using multi-task learning of long short-term memory recurrent neural networks. In Proceedings of the 16th Annual Conference of the International Speech Communication Association (Dresden, Germany, September 6–10, 2015). INTERSPEECH ’15. 3274–32780.

7. Morgan N., 2011. Deep and wide: Multiple layers in automatic speech recognition. Ieee transactions on audio, speech, and language processing, 20(1), pp.7-13.
8. Shubham Toshbiwal., 2018. Multilingual Speech Recognition with a single end-to-end model.
9. Wang, Y., Mohamed, A., Le, D., Liu, C., Xiao, A. et al. (2020). Transformer-based acoustic modeling for hybrid speech recognition. ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6874–6878. Barcelona, Spain.
10. Nassif, A. B., Shahin, L., Attili, L., Azzeh, M., Shaalan, K. (2019). Speech recognition using deep neural networks: A systematic review. IEEE Access, 7, 19143–19165.
11. Garg, K., Jain, G. (2016). A comparative study of noise reduction techniques for automatic speech recognition systems. 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 2098–2103. Jaipur, India.
12. Arora, S. J., Singh, R. P. (2012). Automatic speech recognition: A review.International Journal of Computer Applications, 60(9), 1–11.
13. Hussain, S., Nazir, R., Javeed, U., Khan, S., Sofi, R. (2022). Speech recognition using artificial neural network. In: Intelligent sustainable systems, pp. 83–92. Singapore: Springer.