



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

9940 572 462

6381 907 438

[ijirset@gmail.com](mailto:ijirset@gmail.com)

[www.ijirset.com](http://www.ijirset.com)

# Multilingual Speech Recognition - Challenges and Progress

Akshayaa Darshini. D<sup>1</sup>, Dhana Devi. D<sup>2</sup>, Shreenidhi.V.P<sup>3</sup>, Sri Shreya. N<sup>4</sup>,  
Mahimaa.V<sup>5</sup>, Angelin Christula. J<sup>6</sup>

UG Students, Department of Computer Science and Engineering, Velammal Engineering College, Chennai,  
Tamil Nadu, India <sup>123456</sup>

**ABSTRACT:** Multilingual speech recognition is becoming increasingly important in today's globalized world, where people speak different languages and dialects. This paper discusses the challenges and progress in multilingual speech recognition, highlighting the importance of developing accurate and reliable speech recognition systems that can recognize speech in multiple languages and dialects. By addressing these challenges, we can develop speech recognition systems that are more accessible and useful to people speaking different languages and dialects. The paper also highlights the advancements in deep learning, transfer learning, and multilingual corpora that have contributed to the progress in multilingual speech recognition. The paper concludes by discussing the importance of continued research and development in this field to overcome the remaining challenges and improve the accuracy and accessibility of multilingual speech recognition systems.

## I. INTRODUCTION

Speech recognition technology has made significant progress in recent years, enabling machines to accurately recognize and transcribe spoken language into text. However, as the world becomes increasingly interconnected, there is a growing need for speech recognition systems to be able to accurately recognize and transcribe speech in multiple languages and dialects. Multilingual speech recognition has become an important area of research, with the potential to break down language barriers and enable effective communication across cultures and languages. In this paper, we will explore the challenges and progress in multilingual speech recognition, discussing the advancements in machine learning and artificial intelligence that have led to significant improvements in accuracy, as well as the remaining challenges that need to be overcome.

## II. LITERATURE REVIEW

Firstly, the Audrey system, created by Bell Laboratories in 1952, was one of the earliest speech recognition systems and could recognize spoken digits when spoken by a single voice. IBM later developed the Shoebox in 1962, which was capable of understanding and responding to 16 English words, using a simple acoustic model.[1] Multilingual speech recognition has a rich history, dating back to the early days of speech technology. In the 1950s researchers began experimenting with pattern recognition techniques and early computers. The development of hidden Markov models in the 1970s and neural networks in the 1990s led to significant improvements in speech recognition accuracy. The use of language models that handle multiple languages, and advances in machine learning and artificial intelligence in the 2010s led to the development of various new applications such as virtual assistants and speech-enabled devices. Today, multilingual speech recognition is widely used and research continues to advance in accuracy and the range of recognized languages.

## III. CURRENT STATE AND ITS PROGRESS

**Deep Learning:** Speech recognition algorithms can be implemented in a traditional way using statistical algorithms or by using deep learning techniques such as neural networks to convert speech into text. **Transfer Learning:** Transfer Learning is particularly useful for low-resource languages, where there may be limited data available for training speech recognition models.

**Natural language processing:** While NLP isn't necessarily a specific algorithm used in speech recognition, it is the area of artificial intelligence that focuses on the interaction between humans and machines through language through speech and text.



**Multilingual Corpora:** The availability of multilingual corpora has increased in recent years, making it easier to develop speech recognition systems that can recognize speech in multiple languages.

**Code-Switching:** Code-switching (CS) refers to the process by that languages alter within a speech signal and leads to confusion of languages for automatic speech recognition (ASR).[2]

**Speaker Adaptation:** Speaker adaptation is the process of adapting a speech recognition model to a particular speaker's voice. Researchers have developed techniques to adapt speech recognition models to different speaking styles, accents, and dialects.

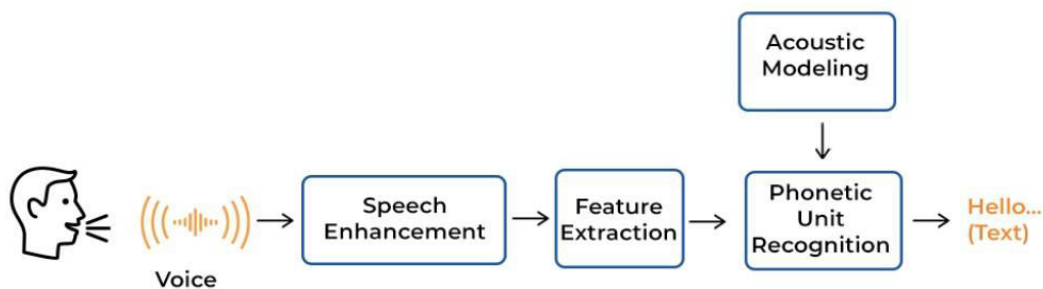
**Neural networks:** Neural networks are most important in the field of speech recognition. They are used to recognize the sound and distinguish them. [3]

**End-to-End Speech Recognition:** End-to-end speech recognition is a technique where speech is directly mapped to text without the need for intermediate steps such as phoneme recognition. This approach has shown promising results for multilingual speech recognition.

**Speaker Diarization:** Speaker Diarization algorithms identify and segment speech by speaker identity. This helps programs better distinguish individuals in a conversation and is frequently applied at call centers distinguishing customers and sales agents.[5]

**Hidden Markov Models:** HMM is the most important method used to recognize patterns in speech. It provides a secured mathematical foundation when compared to the knowledge-based template based. [3]

#### IV. FRAMEWORK



A framework for multilingual speech recognition typically involves a pipeline of several components. The first component is acoustic modeling, which involves training models to recognize the acoustic characteristics of different languages. The next component is language modeling, which involves training models to predict the probability of different words and phrases occurring in a given language. The final component is decoding, which involves using acoustic and language models to transcribe spoken language into written text[4]. In addition to these components, a language identification component may be used to automatically detect the language being spoken. The framework may also include preprocessing steps such as audio normalization, noise reduction, and speaker diarization to improve recognition accuracy.

The accuracy of the framework is dependent on the effectiveness of the machine learning algorithms used to train the models and the quality of the trained data. Its ultimate goal is to enable effective communication among individuals who speak different languages and access information across language barriers

#### V. CHALLENGES

**Acoustic Variability:** Acoustic modeling is essential for accurately recognizing speech, but the patterns and features that need to be analyzed vary significantly across different languages. For example, some languages have more tonal variations than others, while some languages may use different types of phonemes that are not present in other languages.

**Language Modeling:** LMs are designed to analyze the context and grammar of the spoken language, which can improve the accuracy of speech recognition. However, the context and grammar of different languages are often vastly different, making it challenging to develop a single language model that can accurately recognize speech in multiple languages.

**Data Collection and Resource Availability:** Multilingual speech recognition systems require large amounts of data to train and improve accuracy. However, data collection can be difficult and expensive, particularly for low-resource languages.



**Domain Adaptation:** The language and speaking styles used in different domains can vary significantly, making it difficult to develop speech recognition systems that are accurate across multiple domains. **Language Mixture:** Individuals may use multiple languages in the same sentence or utterance. This can create confusion for speech recognition systems, as it may be challenging to distinguish between different languages and dialects. **Speaker Variation:** Speaker Diarization is the process of dividing audio into segments belonging to each individual speaker. It helps AI and humans understand who is saying what throughout the conversation[5].

## VI. APPLICATIONS

**Customer service:** Multilingual speech recognition technology can be used in customer service centers to provide support in different languages. This can help businesses cater to a wider customer base and improve their customer experience.

**Language learning:** Multilingual speech recognition technology can be used in language learning apps to help learners improve their pronunciation and fluency in different languages.

**Translation:** Multilingual speech recognition technology can be used in real-time translation apps, allowing people to communicate with each other in different languages without the need for a human translator.

**Healthcare:** Multilingual speech recognition technology can be used in healthcare settings to help patients communicate with healthcare providers who may not speak their language.

**Education:** Multilingual speech recognition technology can be used in educational settings to transcribe lectures and discussions in multiple languages, making learning more accessible for non-native speakers. **Legal:** Multilingual speech recognition technology can be used in legal settings to transcribe court proceedings and depositions in multiple languages.

**Voice Assistants:** Multilingual speech recognition technology can be used in voice assistants to allow users to interact with their devices in different languages. This can help companies expand their market reach and improve user experience.

## VII. SCOPE FOR THE FUTURE

Multilingual speech recognition technology has been evolving rapidly in recent years, and its future looks promising. With the advancement of artificial intelligence and machine learning algorithms, it is now possible to create speech recognition systems that can recognize and understand multiple languages accurately.

The field of multilingual speech recognition is advancing rapidly and researchers are exploring various ways to improve the technology. Some of the key areas of focus include increasing accuracy, expanding language coverage, developing real-time translation systems, integrating speech recognition with other modalities, and identifying emotional cues in speech. In addition, future advancements may include developing systems for multilingual speaker identification, improving contextual understanding, enabling multilingual sentiment analysis, and creating personalized recognition systems. These improvements could lead to better virtual assistants, enhanced accessibility for people with disabilities, more accurate voice-controlled devices, Speech, and a better understanding of public opinion and global sentiment.

Overall, the future of speech recognition technology is full of exciting possibilities. From improved accuracy to new applications in healthcare and smart homes, speech recognition will continue to play an increasingly important role in our lives.

## VIII. CONCLUSION

This paper gives a review of Multilingual Speech recognition. The area of Multilingual Speech recognition is continuously changing and developing. This technology is capable of making it possible to communicate with everyone. In the future, there are vast possibilities to improve the area of speech recognition technology which provides better services for disabled persons. It can provide a secure surrounding to our system through voice authentication. The correctness of multilingual speech recognition and the quality of speech will be improved thus making communication so easy and dependable for everybody. Future systems must be more efficient and capable compared to traditional systems.

## REFERENCES

[1] <https://sonix.ai/history-of-speech-recognition#:~:text=The%20first%20speech%20recognition%20systems,to%2016%20words%20in%20English>



**Organized by**

Department of CSE, Velammal Engineering College, Chennai, India

- [2]X. Yue, G. Lee, E. Yılmaz, F. Deng, and H. Li, "End-to-End Code Switching ASR for Low-Resourced Language Pairs", IEEE,2019
- [3]Mr. Hardik Dudhrejia, Mr. Sanket Shah, "Speech Recognition using Neural Networks", Department of Computer Science and Engineering, 2018.
- [4]Fuliang Weng, Harry Bratt, Leonardo Neumeyer, and Andreas, "A study of Multilingual Speech Recognition", Speech Technology And Research Laboratory.
- [5] L. El Shafey, H. Soltan, and I. Shafran, "Joint Speech recognition and Speaker Diarization via Sequence Transduction",2019.



**SJIF: 8.423**



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INTERNATIONAL CENTRE



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details