1. Introduction to Voice Recognition System

1.1 Introduction to Speech Processing System

A speech processing system can be defined as various integrated techniques and technologies for the analysis, interpretation, and understanding of the spoken language. By using appropriate algorithms along with signal processing methods and machine learning models, meaningful information is extracted from human speeches. The system can be employed for a number of different applications, including speech recognition, speaker recognition, speech synthesis, and speech enhancement.

Speech is among the most natural and efficient modes of communication in humans; it is a basic building block in human-computer interaction. The great growth in technology has brought speech processing systems to one of the most vital positions, by which a machine will be able to understand, interpret, and generate human speech. These systems encompass a number of techniques that are targeted at the analysis of speech signals with the aim of extracting useful information from them; hence, the possibility of easy interaction with machines.

This area has received tremendous impetus in these recent years from the domains of Artificial Intelligence and Machine Learning; hence, applications such as speech recognition, speaker identification, speech synthesis, and emotion detection have now become robust. These developments have opened the door for a wide range of applications, including speech-driven assistants, smart home systems, real-time transcription services, and even security applications involving voice authentication.

The latest emergent trends based on voice-activated interfaces require sophisticated speech processing methods that must work with high efficiency in diverse acoustic conditions and languages/dialects. Although there is great improvement on the processing of most widely spoken languages, developing systems with high performance accuracy for regional and less-studied languages continues to date. This issue is very relevant in a multilingual society like India, where linguistic diversity raises unique opportunities and challenges for speech technology.

1.2 Types of Speech Processing

A Speech Recognition System encompasses a wide range of applications as shown in Figure 1-1, which can be categorized as follows:

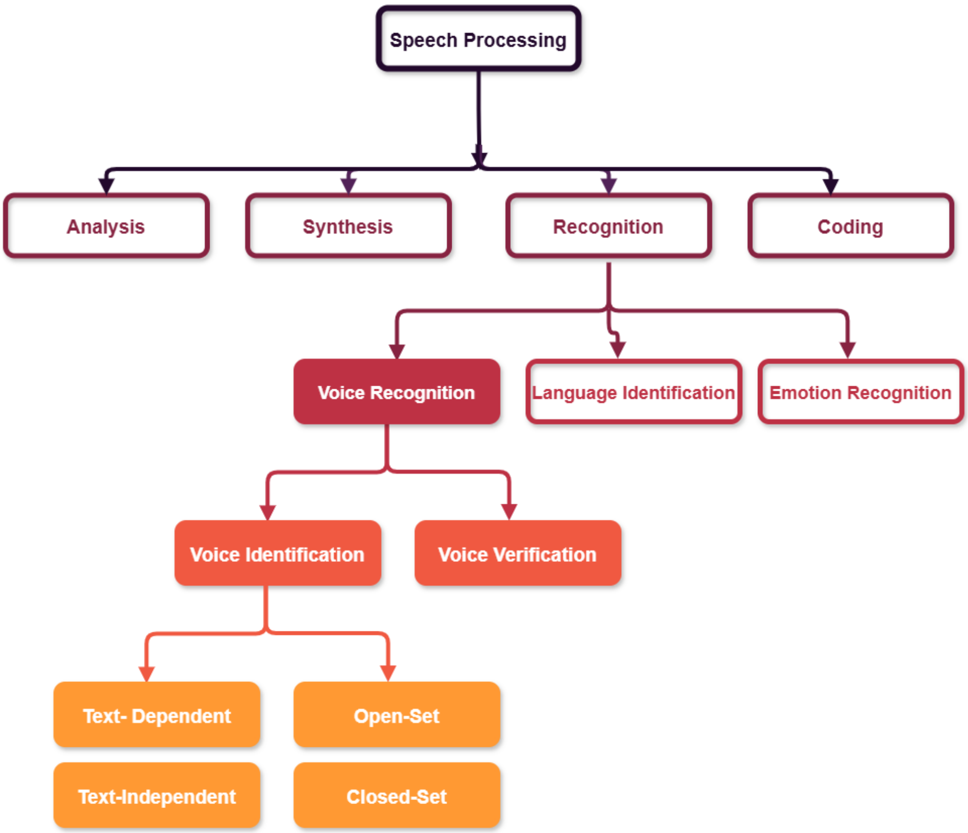

Figure 1- 1 Speech Processing Taxonomy

1.2.1 Speech Recognition

Speech Recognition, also referred to as Automatic Speech Recognition (ASR) or computer speech recognition, is the process of transforming spoken language into a sequence of textual words using algorithms embedded within computer software[1].

Speech recognition systems can be categorized into several distinct classes based on the types of utterances they are capable of recognizing. These categories include the following:

Isolated Words: Isolated word recognition systems typically require a period of silence before and after each utterance within the sample window. These systems are designed to process one word or utterance at a time, operating in distinct "listen" and "not-listen" states, where processing primarily occurs during pauses. Given this structure, the term "Isolated Utterance" might be a more accurate descriptor for this category, as it emphasizes the system’s ability to recognize discrete spoken segments rather than individual words [2].

Connected Words Systems (or, more accurately, "connected utterances") are similar to isolated word systems but provide greater flexibility by allowing individual utterances to be spoken in succession with minimal pauses between them. This enables the recognition of sequences of words that are smoothly connected, rather than strictly isolated by silences [2].

Continuous Speech: Continuous speech recognition systems enable users to speak in a near-natural manner while the system interprets the spoken content, functioning similarly to computer-based dictation. Developing these systems is particularly challenging, as they require sophisticated techniques to accurately identify word boundaries within a continuous flow of speech [2].

Spontaneous Speech: At its core, this refers to speech that is natural and unscripted. An ASR system capable of recognizing spontaneous speech must effectively manage various features of natural spoken language, such as words spoken in quick succession, filler sounds like "ums" and "ahs," as well as minor disfluencies such as stutters[2].

1.2.2 Voice Recognition

Voice recognition refers to a system designed to identify the individual who is speaking. It leverages distinctive features such as pitch, speaking style, and accent, which contribute to the unique characteristics of a speaker’s voice. Voice recognition involves identifying the identity of an unknown speaker by comparing their voice against stored voice profiles within a database. Voice recognition technology has been applied in various fields, including biometrics, security, and human-computer interaction[3].

Voice recognition can be classified into several categories based on its functionality and purpose as outlined below but the main two categories of the Voice Recognition technology involve Voice Identification and Voice Verification:

1.2.2.1 Voice Identification

Voice identification involves classifying an unknown voice, spoken anonymously, and determining which one of a predefined set of N reference speakers it belongs to[3]. It involves identifying an unknown speaker by comparing his voice against a database of known speakers. It basically focuses on the question, "**Who is speaking?**" It further becomes an important task in applications involving security systems, forensics, and also personalized services. As opposed to Voice Verification, which confirms an identity claimed by one, Voice identification is concerned with identifying a speaker from among a group of references.

1.2.2.1.1 Categories of Voice Identification

Identification of voice can be broadly classified based on the following dimensions:

Closed-Set vs. Open-Set Identification: In Closed-Set Identification, the speaker is known to be part of a predefined set of speakers, and the system will return a result based on the best match. That means that the speaker might not be in the reference database. The ability of Open-Set Identification allows the identification system to report "unknown" in case no match is found.

Text-Dependent vs. Text-Independent Identification: In text-dependent systems, it requires certain phrases or passcodes to be uttered for identification, they can offer higher accuracy under controlled conditions. What this really means is that text-Independent systems are capable of identifying speakers independent of the spoken content, which makes them more applicable to unconstrained real-world environments.

1.2.2.2 Voice Verification

Voice verification is the process of determining whether an unknown voice matches a specific reference speaker. The outcome of this task can be one of two possibilities:

either accepting the voice as belonging to the reference speaker's voice or rejecting it as an impostor[3]. Voice verification is a subcategory of the Voice Recognition process, which verifies that a speaker really is who he or she claims to be by taking into consideration whether the claimed identity matches his or her voice. It responds to the question, "**Is this person who they claim to be?**", and it oftentimes serves in systems like biometric security: giving voice-based authentication to banking or mobile devices.

Unlike Voice Identification, which identifies who, out of a set of known speakers, is speaking, Voice Verification is a binary decision against an acceptance threshold: accepting or rejecting the claimed identity.

1.2.2.2.1 Categories of Voice Verification

Text-Dependent vs. Text-Independent Identification: Voice Verification can be divided into two categories: text-dependent and text-independent systems. Text-dependent systems are those in which the speaker is required to utter a pre-defined phrase, such as a passphrase or password, which provides higher accuracy if done in a controlled environment. On the other hand, text-independent systems verify the speaker based on anything he says and are adapted for more dynamic applications, which are close to operational scenarios—for example, voice authentication during a phone call.

1.2.3 Speech Synthesis

Speech synthesis is a computer-based process that produces artificial human speech. It is the process of converting text or any other form of input into audible speech, whereby machines are enabled to convey information in a natural-sounding manner, resembling human speech. The process entails a series of successive stages, commencing with the analysis of the text for interpretation and conversion to phonetic or linguistic representation. This is followed by phonetic conversion, which means that the text is broken down into phonemes—that is, the basic sound units of a language. Then comes prosody generation: for natural-sounding speech, rhythm, intonation, and stress patterns are added. Finally, through waveform generation, the processed data is converted into an audible sound.

1.2.4 Speech Enhancement

Speech enhancement is the enhancement of speech signal quality and intelligibility, with noise, reverberation, and other forms of distortion reduced. It finds its most common use in applications where clarity of speech is of utmost importance, such as in telecommunications, hearing-aid devices, voice-operated systems, and audio recordings. Techniques used include noise suppression, echo cancellation, and filtering, aimed at delineating and enhancing the desired speech signal while minimizing undesired background interference. Modern methods of speech enhancement involve the use of several models, including machine learning and deep learning. These operate on real-time analysis and adaptation to different noise patterns to elicit clearer and more natural-sounding audio.

1.2.5 Emotion Recognition

Emotion is an internal factor that can cause variability in a speaker's vocal characteristics. Emotion recognition is a technique used to identify emotions from speech, which can include various distinct emotions such as disgust, surprise, fear, anger, sadness, happiness, calmness, and neutrality. Systems specifically designed to analyse and account for the impact of emotions on vocal traits are referred to as emotion recognition systems[4].

1.2.6 Language Recognition

Language recognition is the technique of determining the language included inside a specific speech sample. It comprises two primary phases: language identification and language verification. Language identification entails ascertaining the language spoken by the interlocutor. The verification stage confirms the accuracy of the recognized language by validating the outcomes of the identification procedure. This dual-step methodology facilitates the precise identification and verification of the language utilized in speech.

1.3 Gujarati Language and Dialect

Gujarati is an Indo-Aryan language predominantly spoken by the Gujarati community in the Indian state of Gujarat. It functions as the official language of Gujarat and the union territories of Dadra and Nagar Haveli and Daman and Diu. As of 2011, Gujarati is the sixth most spoken language in India, with 55.5 million native speakers, representing around 4.5% of the nation's population. In 2007, it ranked as the 26th most spoken language worldwide by native speakers [6].

Language is a communication system, encompassing both spoken and written forms, utilized by individuals, whereas a dialect is a distinct variant of a language prevalent in a given geographical location or section of a nation. Various factors, including age, education, and religion, affect language. Automatic Speech Recognition (ASR) is a method employed to decipher human speech by transforming analog data into digital representations. Languages frequently possess a profound history, exemplified by Gujarati, whose first literature may be traced to the 12th century. Gujarati is integral to the Indian business community, extensively utilized in trade and commerce. Of the 65.5 million Gujarati speakers globally, the predominant population is located in Gujarat, India.

Similar to other languages, Gujarati exhibits regional variations, with individuals possessing distinct accents and speech patterns. The dialects of Gujarati exhibit variations in pronunciation, accents, vocabulary, and phrases. Dialects are unique variants of Gujarati utilized by particular communities in different places, highlighting the necessity of comprehending these distinctions. Gujarati comprises eight principal dialects, along with several sub-dialects, distributed throughout the Gujarati-speaking areas. Known as Gojarati or Gujerati, it serves as the official language of Gujarat. The language derives its name from the Gujar or Gurjar people, which is thought to have established itself in the region during the 5th century C.E. Gujarati is a member of the Indo-Aryan branch of the Indo-European language family.

A dialect represents a distinguishable variety of a language spoken by a community and can be identified by characteristics such as phonemes, pronunciation, tonality, loudness, and nasality. Forensic Linguistics provides a scientific framework for identifying these dialectal differences. Linguistically, dialects are subcategories of a language, differentiated through grammar, vocabulary, and phonology. They are described as manners of speaking, speech characteristics, or language peculiar to a specific group. While dialects are often perceived as "sub-standard" rather than "non-standard" forms of language, linguists prefer the term "variety" to avoid this bias. However, dialects, including standard ones, can evoke social prejudice, covert prestige, ridicule, or even humour[7].

People from different geographical areas often exhibit variations in speech as shows in Figure 1-2, with socio-pragmatic and pragma-linguistic rules differing slightly. A well-known Gujarati saying captures this phenomenon: *"બાર ગાઉ એ બોલી બદલાય"* (At every twenty miles, a dialect changes). In Gujarat, three major regional dialects—**Kathiyawadi**, **Charotari**, and **Surati**—have emerged and are widely spoken. Other regional dialects include **Kutchhi**, **Saurashtri**, **Gamadia**, **Pachchimi**, and **Pattni**, spoken in various parts of India. For instance, Pachchimi is predominantly spoken in

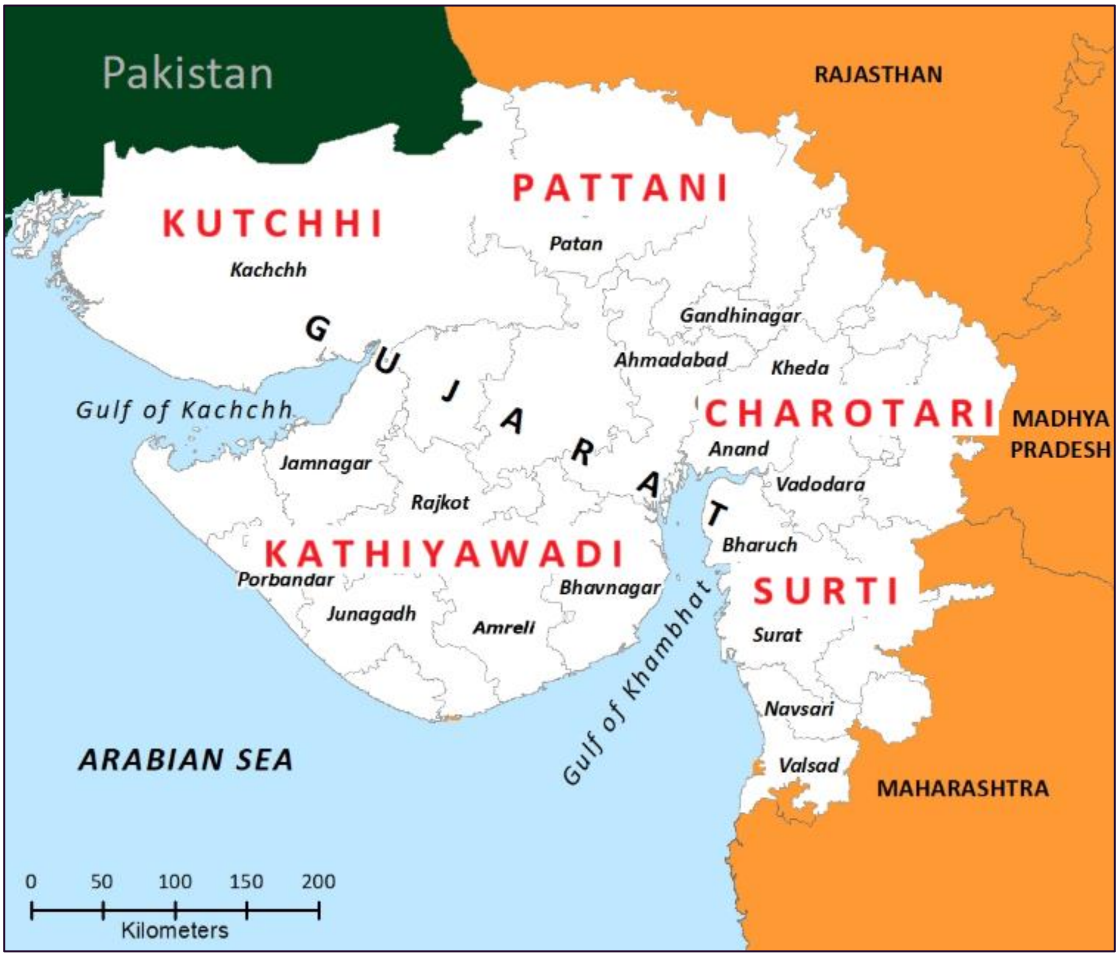

Figure 1- 2 Map showing Region wise Dialect Diversity in Gujarat western Gujarat, Kutchi in the Kutch district, and Patni in the northern regions of Gujarat[7].

A person's way of speaking can reveal their place of origin and, in some cases, their caste. In addition to regional differences, social variations in speech have also been observed, further highlighting the diversity within the Gujarati language. These dialects vary in pronunciation, vocabulary, and expressions, reflecting the cultural and geographical nuances of the region.

1.4 Statement of the problem

Speaker recognition technology, which involves identifying and verifying individuals based on their voice, has grown increasingly important in areas such as security, authentication, and forensic science. While extensive research has been conducted on widely spoken languages like English, Spanish, and Mandarin, the challenges associated with speaker recognition in languages with diverse dialects remain underexplored.

Gujarati is the Indo-Aryan language spoken in different dialects by the people of Gujarat and also its diaspora at large across the world. The variations in regional, cultural, and socio-economic factors have led to dialects that are very different in phonetics, prosody, and intonation. These varying factors pose a great challenge before speaker recognition systems, which makes dialect recognition in Gujarati quite an unexplored problem. A shortage in resources, standardized datasets, and strong speaker recognition models tailored for Gujarati dialects hampers the accuracy and efficiency of automatic speaker recognition systems. The diversity in Gujarati dialects complicates tasks like speaker verification and identification because traditional models, generally trained on very limited or nonrepresentative data, struggle to adapt to variations in accent, speech patterns, and regional linguistic features.

The paper proposes overcoming these difficulties by developing a speaker recognition system that can identify and verify speakers using different Gujarati dialects, independent of regional or socio-linguistic consideration. The ultimate objective is to enhance the reliability and usability of speaker recognition technology for Gujarati speakers, making a significant contribution to academic research and practical applications in voice biometrics, access control, and forensic analysis.

The primary motivation for exploring these dialects is that, while significant work has been done on speech processing systems for other Indian languages such as Hindi, Bengali, English, and Rajasthani, Gujarati remains relatively unexplored. Gujarat, with its vibrant community and status as a prominent commercial hub, plays a crucial role in contributing to India's GDP, making it essential to focus on developing linguistic technologies for this region.

This study shall aid the development of more efficient, robust, and inclusive systems for speaker recognition to handle linguistic variability due to regional dialects in a population speaking Gujarati. Further, results can then be generalized for other languages also with similar dialect complexities and wider applicability of research work in speaker recognition and voice-based biometric systems.

1.5 Need of Proposed Research

Among other state-of-the-art voice biometric systems, speaker recognition remains of vital interest in most applied domains these days, ranging from security over access control and forensic investigations to personal assistants. The effectiveness of such systems primarily depends on how linguistic diversity—an accent, pattern, and dialect of speech—can be dealt with. Whereas a number of works have been carried out with good speaker recognition for widely spoken languages, dialect complexities in less-studied languages, such as Gujarati, have hardly been considered.

Linguistic and Phonetic Diversity in Gujarati Dialects: Gujarati is spoken throughout the world by millions of people, though it is very varied even in its dialects. These dialects, which include Kathiyawadi, Saurashtra, Charotari, and Kutchhi, vary greatly in pronunciation, vocabulary, and intonation. These variations are influenced by profound geography, culture, and historical factors. Traditional speaker recognition systems, usually trained on standard accents, have so far failed to take into consideration such dialectal variations. This causes a great decline in the accuracy of identifying speakers from other regions. Without dialect-sensitive recognition, speaker models can very well fail at identifying or verifying individuals, especially when regional speech patterns differ greatly from those of the training set.

Poor Gujarati Dialectal Datasets: Lacking good-quality, large, and diverse datasets to represent all dialects of Gujarati is one of the big challenges to building decent speaker recognition systems. Existing datasets usually have narrow outlooks; they most often relate to a single variant, further even to a single type of speech, such as formal speech. Consequentially, the models trained with such data lack the generalization capability to handle the full spectrum of dialectal diversity. The basic requirement of the research is hence to build a comprehensive dataset of speakers of Gujarati from different regions, social strata, and demographics in order to develop more robust Speaker Recognition Systems.

Current Models Are Not Dialect-Sensitive: Most of the speaker recognition models which have been proposed so far, especially traditional machine learning or deep learning approaches, have been developed for well-studied languages like English. These models very often cannot perform well on languages which have significant internal dialectal variation since these models are insensitive to the phonetic, syntactic, or prosodic differences between dialects. This fact implies that there is a lot of variation in the phonemic inventory (sounds) and prosody (intonation, rhythm) across the various dialects of Gujarati. A mismatch like this in the training data and diverse speech characteristics makes reliable speaker recognition models hard to achieve without explicit consideration of dialectal variations in Gujarati speech.

Growing Dependency on Voice Biometrics: While voice-based biometrics and their use in automated systems for authentication and access are becoming increasingly applied, there is an emergent need for speaker recognition systems that are linguistically diverse. Applications such as mobile banking and services related to government and regional security systems are therefore rapidly moving towards voice authentication, as seen within India, especially in Gujarat. For this, the development of speaker recognition systems that can work well across dialects is very urgent, in an attempt to secure and ensure user identification.

Contribution to Forensic and Security Applications: Speaker identification may be of paramount importance in forensic linguistics and legal issues. Gujarat, being a state with a huge population with diverse dialects, is a place where voice-based forensic analysis is highly essential. However, the available tools and methods for speaker identification might not distinguish between speakers of different Gujarati dialects and may produce incorrect identification or verification. The development of robust models for speaker recognition in Gujarati dialects will, to a great extent, contribute toward forensic investigations dealing with criminal justice and national security where regional linguistic variation plays an important role.

Technological Inclusion and Equity: Gujarati is among the most spoken languages in India, and many native Gujarati speakers rely catastrophically on technology in helping everyday areas of education and health care, among other government-provided services. The big problem has always been that most speech recognition technologies have been executed for standard dialects, leaving out a bigger population that speaks regional dialects. It is expected that this research will contribute therefore to the development of speaker recognition models sensitive to dialectal variation, thus allowing all speakers irrespective of their region to avail different voice-based services with equal efficiency and reliability.

The proposed research, by addressing these challenges, therefore contributes toward developing more accurate, inclusive, dialect-aware speaker recognition systems that cater to the linguistic realities of Gujarati-speaking populations and afford greater reliability and accessibility with voice-based technologies.

1.6 Objectives of the Research

The aim of the research proposal is to investigate the intricacies associated with identifying the distinct voices within Gujarati dialects and to create a model capable of analysing and interpreting voice recognition specifically for vernacular Gujarati dialects. The developed model should be robust enough to address, to some extent, challenges concerned with various environmental noises, speaker variability, and threats against voice impersonation. The project described here will, with the introduction of different state-of-the-art machine learning techniques such as deep learning and feature extraction methods, make speaker recognition more robust, scalable, and efficient for real-world applications like biometric security, personalized voice assistants, and forensic voice analysis.

Major objective of the proposed study is to develop “A Model to Analyse & Interpret Vernacular Voice Recognition of Gujarati Dialects.”

Additional objectives include:

- The additional purpose of this study is to examine existing algorithms and identify the research gap in the domain of voice recognition, with a focus on Gujarati regional dialects.
- To produce a Gujarati dataset enriched with regional dialects, recorded in a controlled context from varied speakers across different age groups, with speaker-independent voice samples. This dataset can enhance the accuracy of the voice recognition system model.
- To identify the most effective feature extraction methods for extracting structural aspects from speech, as well as for capturing voice characteristics, prosody, and phonetics.
- To design a user-friendly interface for the proposed framework and developed model, enabling easy access and usage of the speaker recognition model tailored for regional Gujarati dialects like Kathiyawadi, Standard Gujarati, Surti and Kutchhi.
- Design and develop the components of the speaker recognition model, followed by testing and performance evaluation through an analysis of the recognition model's results.

1.7 Scope of the Study

The present work is devoted to developing a speaker recognition system that can identify and verify speakers through different dialects of Gujarati, which is a language with very significant internal variation. The acknowledged well-known dialects of Gujarati, spoken by millions in the world, embrace Kathiyawadi, Saurashtra, Charotari, Kutchhi, and a standardized variety used in formal or media contexts. The present research, therefore, aims at the design of a system that can withstand the burden of this dialectical difference, structural in distinct phonetic, prosodic, and syntactic features.

In essence, the knowledge of the impact these varieties have on recognition accuracy is important for dependable performance across a range of diverse Gujarati-speaking populations. Much emphasis will be placed on collecting a diverse dataset of various speakers of Gujarati, representing different regions, age groups, and social backgrounds. The resultant dataset will have representatives of different dialects; take some conversational and some formal speech to capture the richness of Gujarati. This data will form a basis for training and testing the models with the system so that it considers speech pattern variations, pronunciation, and intonation.

The research will investigate multiple Machine Learning techniques, including traditional methods like Gaussian Mixture Models (GMMs) [8], [9]and Hidden Markov Models (HMMs)[10], [11], alongside advanced deep learning approaches such as Convolutional Neural Networks (CNNs)[12], [13] and Recurrent Neural Networks (RNNs)[14]. By comparing these methods, the study aims to identify the most effective algorithms for dialect-sensitive speaker recognition. Feature extraction will also be emphasized, focusing on acoustic features like Mel-frequency cepstral coefficients (MFCCs) [15]and prosodic characteristics critical for distinguishing speakers from different dialects.

The performances of the systems will be measured regarding standard metrics such as accuracy, FAR, FRR, and EER. The overall recognition and verification performance of the system will be benchmarked, including cross-dialect challenges. This approach will indicate how well the model generalizes across diverse speech patterns and dialectal variations. Though cantered on Gujarati, the findings of this research could generalize into speaker recognition systems for any other language that presents a strong dialect diversity. This study can further give ways in which methodologies applied can be beneficial to linguistically complicated languages and hence widen the field of speaker recognition as a whole. Ethical considerations will also be foregrounded, with data collection being done in a manner consonant with informed consent, rigorous participant privacy, and anonymization measures.

1.8 Limitations of the Study

A limitation of this proposed model is that they have some functional and application limits. To begin with, this system's ability to understand and process all regional Gujarati dialects are difficult because of the extent of variation in regional Gujarati dialect and accents of these dialects so the current work focuses on Four regional dialects: Kathiyawadi, Kutchhi, Surti and Standard Gujarati. The current design already gives a solid structure but may need to be further adjusted for each dialect to be handled optimally. This would improve the system's reliability and trustworthiness a lot. Furthermore, currently only the model's ability to be retrained for existing speaker enrolment is allowed. But it can also recognize new voices, and a mechanism to seamlessly adjust for new users would greatly increase the system's flexibility and user centric quality. Finally, the system might give ambiguous results in noisy environments to accurately recognize the voice. That background noise can decrease the clarity of your audio input to decrease the accuracy of recognition. To make the system's overall effectiveness and usability in real world applications sensible, we therefore have to address these limitations using advanced algorithms and techniques.

1.9 Structure of the Research

Researcher has distributed entire work into six different chapters. Summary of the remaining chapters i.e. from chapter 2 to chapter 6 is as follow.

Chapter 2 Literature Review

This chapter presents a comprehensive literature review in the field of speech processing. Previous studies on speech recognition across various languages, including English, Hindi, and Bengali, have been examined. Additionally, research related to speaker recognition in different languages has been reviewed, with a specific focus on studies involving speaker recognition based on Gujarati. Furthermore, the research work utilizing Gujarati datasets has also been explored in detail.

Chapter 3 Study and Analysis of Gujarati voice for Stuctural Feature Extraction and classification

This chapter serves as an introduction to the Gujarati language and its numerous dialects, encompassing the characteristics of Gujarati speech, including diverse prosodic and phonetic traits. The investigation encompasses many Feature Extraction Techniques, from which we assess the most appropriate for the Gujarati dataset, specifically evaluating the optimal techniques for Gujarati dialects. The various categorization techniques were also analysed for the vernacular Gujarati dialect dataset.

Chapter 4 Proposed Framework & Voice Recognition Model for Vernacular Gujarati Dialects

This chapter provides a detailed overview of the system design. The proposed framework and algorithm of the suggested model have been discussed herein. The dataset pertaining to the generation process of vernacular Gujarati languages and its details will also be explored here. in conjunction with the feature extraction method to derive the structural characteristics. The proposed model and classification approaches of the model are presented.

Chapter 5 Component Development of Voice Recognition Model for Vernacular Gujarati Dialects

This chapter delineates the specifics of the Development Tools and Framework for the proposed system designed for speaker detection based on Gujarati dialects. The system architecture is illustrated as a flowchart to enhance comprehension of the task. The User Interface of the created method is detailed herein. Ultimately, the chapter discusses and delineates various assessment criteria to assess the constructed model based on Performance assessment. All challenges encountered during the investigation were discussed, along with the study's limitations. The final topic pertains to the analysis of outcomes derived from various parameters.

Chapter 6 Performance Analysis, Results, Conclusion and Future Scope for Extension of Research Work

The final chapter of this work presents the conclusion, indicating that all objectives established by the researcher have been met and fulfilled. What measures may be implemented to further advance this area of research, and what enhancements can be anticipated for future work.

References

[1] D. Yu and L. Deng, Automatic Speech Recognition. London: Springer London, 2015. doi: 10.1007/978-1-4471-5779-3.

[2] M. A. Anusuya and S. K. Katti, “Speech Recognition by Machine: A Review,” 2009. [Online]. Available: http://sites.google.com/site/ijcsis/

[3] R. Mohd Hanifa, K. Isa, and S. Mohamad, “A review on speaker recognition: Technology and challenges,” Computers and Electrical Engineering, vol. 90, Mar. 2021, doi: 10.1016/j.compeleceng.2021.107005.

[4] T. J. Sefara and T. B. Mokgonyane, “Emotional Speaker Recognition based on Machine and Deep Learning,” in 2020 2nd International Multidisciplinary Information Technology and Engineering Conference, IMITEC 2020, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/IMITEC50163.2020.9334138.

[5] E. Ambikairajah, L. Wang, B. Yin, and V. Sethu, “Language Identification: A Tutorial.”

[6] https://en.wikipedia.org/wiki/Gujarati_language

[7] Sakshi A. Patil, Gaurav A. Varade, and Vikram Hankare, “Tracing Gujarati Dialects Philogically and Sociolinguistically,” International Journal of Modern Developments in Engineering and Science, vol. 2, no. 5, May 2023, [Online]. Available: https://www.ijmdes.com

[8] Z. Liu, Z. Wu, T. Li, J. Li, and C. Shen, “GMM and CNN Hybrid Method for Short Utterance Speaker Recognition,” IEEE Trans Industr Inform, vol. 14, no. 7, pp. 3244–3252, Jul. 2018, doi: 10.1109/TII.2018.2799928.

[9] A. Maurya, D. Kumar, and R. K. Agarwal, “Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach,” in Procedia Computer Science, Elsevier B.V., 2018, pp. 880–887. doi: 10.1016/j.procs.2017.12.112.

[10] J. H. Tailor and D. B. Shah, “HMM-Based Lightweight Speech Recognition System for Gujarati Language,” in Lecture Notes in Networks and Systems, vol. 10, Springer, 2018, pp. 451–461. doi: 10.1007/978-981-10-3920-1_46.

[11] 1ms Puspa Machhar and M. Dipak Agrawal, “HMM Based Gujarati Tricky Words Recognition.” [Online]. Available: www.ijariie.com

[12] G. Costantini, V. Cesarini, and E. Brenna, “High-Level CNN and Machine Learning Methods for Speaker Recognition,” Sensors, vol. 23, no. 7, Apr. 2023, doi: 10.3390/s23073461.

[13] N. N. Prachi, F. M. Nahiyan, M. Habibullah, and R. Khan, “Deep Learning Based Speaker Recognition System with CNN and LSTM Techniques,” in 2022 International Conference on Interdisciplinary Research in Technology and Management, IRTM 2022 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/IRTM54583.2022.9791766.

[14] H. Sailor and H. Patil, “Neural Networks-based Automatic Speech Recognition for Agricultural Commodity in Gujarati Language,” International Speech Communication Association, Oct. 2018, pp. 162–166. doi: 10.21437/sltu.2018-34.

[15] M. Singh, “Speaker Identication using MFCC Feature Extraction ANN Classication Technique,” 2023, doi: 10.21203/rs.3.rs-2407488/v1.