# Appendix - A Publications

## International Journals

[1] Meera M. Shah, Hiren R. Kavathiya. Voice Recognition for Gujarati Dialects: An in-depth Survey. International Journal of Computer Applications. 186, 5 ( Jan 2024), 1-4. DOI=10.5120/ijca2024923112

## International Conferences

[2] M. M. Shah and H. R. Kavathiya, “Exploring Speech Corpus for Voice Recognition in Gujarati: An In-depth Study,” *Educational Administration: Theory and Practice*, Jun. 2024, doi: 10.53555/kuey.v30i6(s).5343.

## Book Chapters

[3] M. M. Shah and H. R. Kavathiya, “Unveiling the Future: Exploring Conversational AI,” in *Studies in big data*, 2024, pp. 511–526. doi: 10.1007/978-3-031-52280-2_32.

## Patent

We have applied for the patent. (Patent on Process)

**Educational Administration: Theory and Practice**
2024, 30(6)(s), 26-33
ISSN: 2148-2403
https://kuey.net/
**Research Article**

# Exploring Speech Corpus For Voice Recognition In Gujarati: An In-Depth Study

Meera M. Shah[1*], Dr. Hiren R. Kavathiya[2]

[1*]Department of Computer Science, Atmiya University, Rajkot, meerashah1013@gmail.com
[2]Department of Computer Science, Atmiya University, Rajkot, kavathiyahr@gmail.com

***Citation:*** Meera M. Shah, et al, (2024) Exploring Speech Corpus For Voice Recognition In Gujarati: An In-Depth Study, *Educational Administration: Theory And Practice*, 30(6)(s), 26-33,
Doi: 10.53555/kuey.v30i6(S).5311

**ARTICLE INFO**

**ABSTRACT**

Automatic Speech Recognition (ASR) technology has gained significant importance in modern communication systems, enabling the conversion of spoken language into written text. This research paper presents an in-depth analysis of voice recognition in the context of the Gujarati language, a tonal and multilingual language with unique phonetic characteristics. The study focuses on a meticulously curated Gujarati speech corpus, comprising diverse speakers of various ages, genders, and regional backgrounds. The corpus is subjected to detailed acoustic analysis, exploring prosodic features and tonal variations inherent in the language. Through the development and evaluation of ASR models, this research investigates the challenges and opportunities posed by the Gujarati language's phonemic complexity and tonal nuances. The findings shed light on the impact of corpus characteristics, including speaker diversity and phonemic inventory, on ASR model performance. As the field of voice recognition continues to advance, this research contributes valuable insights into effective ASR model design and training strategies for tonal languages, specifically focusing on the linguistic and acoustic peculiarities of Gujarati. The outcomes of this study offer directions for further advancements in ASR technology and corpus analysis, addressing the challenges of accurately capturing the intricate linguistic features of tonal languages for robust voice recognition systems.

**Keywords:** Speech Corpus; Gujarati Dataset; Voice Recognition; Speaker Recognition; Gujarati Speech Corpora; ASR

## Introduction

In the realm of biometric technology, speaker recognition has gained considerable attention as a pivotal method for identifying and verifying individuals based on their unique vocal characteristics. While substantial research has been conducted in this domain across various languages, the complexities inherent in languages like Gujarati necessitate a focused investigation. Gujarati, a tonal and diverse language spoken by millions globally, presents distinct phonetic intricacies that have a profound impact on the development of effective speaker recognition systems.

**Roza Chojnacka,Jason Pelecanos, Quan Wang Ignacio Lopez Moreno (2021)** engineered a versatile multilingual model, known as "Gujarati SPEAKERSTEW," designed for the purpose of speaker identification. This model exhibits the remarkable capability to operate across a diverse dataset of 46 languages simultaneously. Notably, the model's proficiency extends to text-independent speaker recognition systems, achieving an impressive accuracy rate of 73% in Thai.

**Sivaram G, Samudravijaya K** have studied the impact of online speaker adaptation on the performance of a speaker independent, continuous speech recognition system for Hindi language dataset. The speaker recognition is executed using the Maximum Likelihood Linear Regression (MLLR) transformation approach. The MLLR transform based speaker adaptation technique is found to significantly improve the accuracy of the Hindi ASR system by 3%. After the experiment they have concluded that MLLR transform based speaker adaptation of Hindi speech models indeed decreases the recognition error by a factor of 0.19

**Pramod Mehra, Shashi Kant Verma(2022)** developed a multilingual model for speaker recognition.The model includes all the indian languages they used the MFCC method for feature extraction and built their own model for identifying the speaker. They also worked on emotions a person can have and a person can be angry while having a slow voice tone.The model provides 98.34% accuracy.

Despite having a lot of work done for speaker recognition for Gujarati language, the accuracy for specific recognition on different environments and different accents was found unexplored. As they are saying, each area has its own accents of people when we talk about speech and these accents need to be focused on for accurate results. In addition to Gujarati Framework's shortcomings and complexity, implementing speaker recognition systems has a vast area of research.

This research paper embarks on a comprehensive exploration of a speaker recognition corpus tailored specifically for the Gujarati language. By delving into the linguistic intricacies, tonal variations, and phonetic characteristics of Gujarati speech, this study seeks to provide insights into the development of robust and accurate speaker recognition systems. Through meticulous analysis of pertinent linguistic features, acoustic patterns, and strategic implementation of machine learning techniques, this research aims to contribute to the advancement of speaker recognition technology catered to Gujarati speakers.

## Literature Review

India is one of the foremost multilingual countries where multilingualism is ingrained and most people speak more than one language with more than 75 languages having more than one million speakers as per 2011 Census of India data(Choudhary, 2021)**.** For Indian languages, there are a tonne of speech corpora available, however Gujarati has a fairly limited amount of data. As a result, we provide comprehensive information on the various resources where Gujarati speech datasets can be found.

The creation and analysis of speech corpora for the Gujarati language have garnered increasing attention within the field of natural language processing and speech technology. These corpora serve as invaluable resources for training and evaluating automatic speech recognition (ASR) systems, enabling advancements in various applications such as voice assistants, language learning, and accessibility tools. The following literature review provides an overview of existing research and insights related to the development, characteristics, applications, and challenges of Gujarati speech corpora.

(Choudhary, 2021) In their work, they delve into the subject of. LDC-IL (Linguistic Data Consortium for Indian Languages), a project initiated under the Department of Higher Education, Ministry of Human Resource Development in the Government of India, introduced its inaugural set of speech corpora in 2019. This dataset comprises a comprehensive collection of 13 scheduled languages of India including Gujarati, meticulously gathered from diverse environments spanning the country. The data includes contributions from 5662 speakers representing various age groups and accumulates to a substantial 1552 hours of content. It's worth noting that this dataset is continually expanding as it undergoes refinement and preparation for public release. The corpus specific to each language is currently among the most extensive available resources.

Established in 2008, the LDC-IL, modeled after the LDC at the University of Pennsylvania, has dedicated over a decade to creating a variety of language resources, encompassing the development of speech corpora. LDC-IL operates as a government-funded project managed by CIIL (Central Institute of Indian Languages) in Mysuru. They referred in their paper that offers a comprehensive overview of the recently released raw speech corpora, which are now accessible for public use. These corpora have been made available to serve a wide range of purposes.

(Dalsaniya et al., 2020) introduces an innovative audio dataset comprising isolated Gujarati digits. This database encompasses recordings of digits spoken by 20 individuals from five distinct regions within Gujarat, captured in real-world environments. To the best of their knowledge, this marks the initial publicly accessible Gujarati spoken digits database. They proceed to employ a basic neural network classifier on statistical features extracted from the audio data. The performance of this classifier is assessed using the newly developed Gujarati database as well as an existing English language database of spoken digits. Additionally, they conduct cross-corpus experiments to assess the adaptability and generalization capabilities of their approach.
(Srivastava et al., 2018) organized a low-resource Automatic Speech Recognition (ASR) challenge focused on Indian languages as a part of the Interspeech 2018 event. They made available 50 hours of speech data with transcriptions for Tamil, Telugu, and Gujarati, collectively amounting to 150 hours. Participants were instructed to solely use the provided data for the challenge to maintain a low-resource scenario, though they had the freedom to address any aspect of speech recognition. The released dataset includes training and test

data for conversational and phrasal speech in Telugu, Tamil, and Gujarati. The package contains audio recordings alongside their corresponding transcripts. The data encompasses both recorded phrases and conversational speech, which were segmented into individual utterances and transcribed using the native script of each respective language.

(Bhogale et al., 2022) focus on generating speech datasets using publicly available resources from All India Radio (AIR). Due to inconsistencies in the data, they introduce a method to extract pairs of audio and text at the document level. This technique involves employing CTC-based Automatic Speech Recognition (ASR) models and the Needleman-Wunsch algorithm. By implementing this approach on the AIR archives, they produce the Shrutilipi dataset, comprising 6,457 hours of labeled audio covering 12 Indian languages.The study demonstrates that Shrutilipi possesses high-quality attributes and displays notable diversity in terms of speakers and content, differentiating it from other publicly accessible datasets. The effectiveness of the dataset is evaluated through various benchmarks in the context of ASR. This involves training on efficient models and showcasing resilience to background noise.

Kaggle offers an extensive collection of audio samples spanning 10 distinct languages, each lasting for a duration of five seconds. This dataset was curated using regional videos sourced from YouTube. The compilation encompasses diverse linguistic content, featuring languages such as Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Punjabi, Tamil, Telugu, and Urdu.

Research efforts to construct comprehensive Gujarati speech corpora have led to the collection of diverse and representative datasets. The Indian Institute of Technology (IIT) Bombay, for instance, has played a prominent role in creating a phonetically rich and regionally diverse speech corpus for Gujarati. This corpus involves recordings from various speakers, including those from different age groups, genders, and linguistic backgrounds.

(He et al., 2020) offer freely accessible, high-quality speech datasets for six Indian languages: Gujarati, Kannada, Malayalam, Marathi, Tamil, and Telugu. These languages are among the twenty-two official languages of India and are spoken by a total of 374 million native speakers. These datasets are primarily designed for text-to-speech (TTS) applications, such as developing multilingual voices or adapting voices to specific speakers or languages. The majority of the corpora contains over 2,000 recorded lines from both male and female native speakers of each language. The process of acquiring these corpora is outlined, and this methodology can be expanded to gather data for other languages of interest. The authors detail their experiments in creating a multilingual text-to-speech model by combining these corpora

(Yarra et al., 2019) make notable contributions by offering Indic TIMIT a linguistically diverse Indian English speech dataset. It comprises approximately 240 hours of recorded speech contributed by 80 individuals. Each participant has articulated a collection of 2342 stimuli, which corresponds to the items found in the TIMIT corpus. Moreover, a subset of the recordings within the corpus is accompanied by phoneme transcriptions. These transcriptions have been meticulously annotated by two linguists to accurately represent the speaker's pronunciation. In the realm of Indian speech corpora, Indic TIMIT stands out due to these distinctive attributes.

(Maity et al., 2012) undertook the endeavor of creating a multilingual speech corpus encompassing various Indian languages. The project, known as IITKGP-MLILSC (Indian Institute of Technology Kharagpur - Multi Lingual Indian Language Speech Corpus), aimed to develop a speech database for language identification tasks within the context of Indian languages. This speech database comprises recordings in 27 distinct Indian languages. Among these, 16 languages, which have a broad-speaking base, were specifically selected for evaluating the effectiveness of language identification. The analysis of the language identification system encompasses both speaker-dependent and speaker-independent scenarios.

In conclusion, the literature reviewed highlights the significance of Gujarati speech corpora as essential resources for advancing ASR technology and linguistic research. The collaborative efforts to develop comprehensive corpora, along with the exploration of phonetic, prosodic, and application-related aspects, contribute to the broader understanding of Gujarati speech and its integration into modern language technologies.

## Analysis of Gujarati Speech Corpus

The primary objective of the paper at hand is to underscore the significance of speech repositories, particularly in the context of speaker recognition for the Gujarati language.In the paper, a thorough and comprehensive analysis is presented, focusing on the phonetic, acoustic, and linguistic attributes that are intrinsic to a variety of Gujarati speech datasets. This analysis delves deeply into the unique characteristics that define each dataset, encompassing elements such as phonetic diversity, acoustic quality, and linguistic intricacies.

By presenting this comparative analysis, the paper aims to highlight the distinctions and similarities among the collected speech databases. This serves to underscore the diverse characteristics of the datasets and their implications for speaker recognition applications. Ultimately, the paper contributes to emphasizing the pivotal role of speech repositories in enhancing speaker recognition capabilities specifically tailored for the Gujarati language.

**Tabel 1.** Analysis of Gujarati Speech Corpus based on phonetic, acoustic, and linguistic attributes

| Speech Corpus | Duration (hh:mm) | # Speakers | Sampling Rate | Audio Segments | File Types |
|---|---|---|---|---|---|
| LDC-IL | 57:17 | 204 | 48 kHz | 25,712 | WAV |
| LDC-IL (Mono Recordings) | 64:44 | 233 | 16 kHz | 26,223 | WAV |
| FSDGG | ~00:32 | 20 | 44 kHz | 1,940 | WAV |
| Microsoft Speech Corpus | 50:00 | - | 44 KHz | 1,24,599 | WAV |
| Shrutilipi | 460:00 | - | 16 kHz | 270 | WAV |
| Indic TIMIT | ~10:75 | 63 | 48 kHz | 2,342 | WAV |
| IITKGP-MLLILSC | 00:49 | 13 | 16 kHz | - | WAV |
| OpenSource MultiSpeaker Speech Corpora | 07:89 | 36 | 48 kHz | 4,472 | WAV |

The table presented offers a comprehensive portrayal of the compiled speech databases and conducts a comparative analysis based on several key parameters. These parameters encompass various attributes, including durations of the audio recordings, sampling frequency rates, audio file types and sizes, audio segmentation methods, and the total number of speakers contributing to the databases.This table serves as a valuable resource for researchers and practitioners aiming to comprehend the unique features of various Gujarati speech datasets, facilitating informed decisions regarding their applicability and suitability for specific tasks within the field of speech processing and analysis.

Conducting an evaluation that considers gender, diverse age groups of speakers, and recording environments within the framework of speech databases entails investigating speech patterns, distinctive attributes, and variations between male and female speakers. Here's an outline of how such an assessment could be undertaken:

**Tabel 2.** Analysis of Gujarati Speech Corpus based on Gender, Age Group and recording Environment.

| Speech Corpus | Speakers | | Age Group | Environment |
|---|---|---|---|---|
| | Male | Female | | |
| LDC-IL | 108 | 96 | 18-60 | Controlled |
| LDC-IL (Mono Recordings) | 109 | 124 | 18-60 | Controlled |
| FSDGG | 14 | 6 | 18-60 | Controlled |
| Shrutilipi | - | - | 10-70 | Noisy |
| Indic TIMIT | 3 | 1 | 18-60 | Quite Noisy |
| IITKGP-MLLILSC | 7 | 6 | 18-60 | Controlled |
| OpenSource MultiSpeaker Speech Corpora | 18 | 18 | 10-70 | Noisy |

The table presented shows a comprehensive analysis of the Gujarati speech corpus, categorized by gender, age group, and recording environment. This analysis provides a detailed breakdown of the corpus's diverse attributes, considering factors such as speaker demographics, developmental stages, and recording conditions. The table serves as a valuable resource for researchers and practitioners, allowing them to discern trends, variations, and potential insights associated with gender-based, age-based, and environment-based differences within the Gujarati speech dataset.

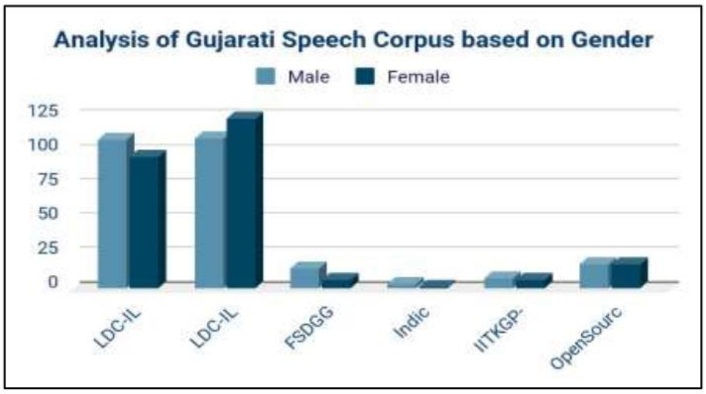



**Figure 1.** Analysis of Gujarati Speech Corpus based on Gender

The graph illustrates a comparison between the proportions of male and female speakers within the Gujarati speech corpus. This visual representation offers insights into the gender distribution present in the dataset, allowing viewers to discern the relative frequencies of male and female speakers.

The x-axis of the graph likely represents different categories or segments within the corpus, which could be specific datasets, recordings, or other relevant divisions. The y-axis denotes the percentage or count of speakers. Bars or data points corresponding to "male" and "female" are depicted side by side or on the same graph, each reflecting the respective ratio of speakers within the given category.The graph aids in understanding the gender balance within the Gujarati speech corpus, highlighting potential variations and disparities.

## Performance Evaluation

When embarking on the evaluation of diverse speech corpora, a crucial preliminary step involves comparing them across various parameters. This comparative analysis is essential for gauging the effectiveness and suitability of each dataset. Several key metrics are typically employed, such as accuracy yielded when the dataset is applied to a specific model, Mean Opinion Score (MOS), and Word Error Rate (WER). These metrics provide valuable insights into the corpus's performance, user perception, and linguistic fidelity.

Regarding the Shrutilipi corpus, When we tried to implement the dataset on our own modal and evaluate on the Gujarati benchmarks, The outcomes showed a consistent enhancement in Word Error Rate (WER) across all benchmarks. Specifically, the Average WER witnessed an improvement from 12.8% to 9.5%.

(He et al., 2020) Throughout the training process, they incorporated speaker identification (ID) as an input feature from the corpus. Subsequently, during the synthesis phase, they employed the best speaker as the conditioned speaker ID feature. Evaluations of the generated voice were conducted through the application of Mean Opinion Score (MOS). The MOS scores were calculated, accompanied by corresponding confidence interval statistics at a 95% confidence level. In the context of the Gujarati corpus, the MOS score for male speakers was $3.950 \pm 0.056$, while for female speakers, it was $4.269 \pm 0.047$.

(Yarra et al., 2019) In an initial exploration, they examine the advantages offered by the Indic TIMIT dataset. To do so, they undertake experiments within an Automatic Speech Recognition (ASR) framework. The analysis encompasses two distinct aspects. The first one centers around ASR performance, quantified through the measurement of Word Error Rate (WER). The second aspect focuses on the Phoneme Error Rate (PER), with consideration for the forced-alignment process. Drawing from both experiments, they deduce that the evaluated WER for Indic TIMIT stands at 15.02 and PER stands at 28.79 performing forced-alignment process.

Spectral characteristics are examined to investigate the presence of language-specific attributes within the Developed Gujarati speech Corpus IITKGP-MLILSC. Specifically, Mel-frequency cepstral coefficients (MFCCs) and linear predictive cepstral coefficients (LPCCs) are utilized to represent the spectral information. In order to capture language-specific nuances present in these features, Gaussian mixture models (GMMs) are developed. The evaluation of the language identification system is carried out within both speaker-dependent and speaker-independent scenarios. Notably, the recognition performance exhibits a rate of 96% in the speaker-dependent environment and 45% in the speaker-independent environment. These findings underscore the system's effectiveness in discerning language-specific information present in the spectral attributes across distinct recognition scenarios (Maity et al., 2012).

In the context of the Interspeech 2018 Low Resource Automatic Speech Recognition Challenge, the focus was on evaluating dataset performance using speech processing methodologies. To establish baselines, three Acoustic Models were developed: GMM-HMM, Karel's DNN, and TDNN. The evaluation process employed Word Error Rate (WER) measurements on the blind test set for each language. Upon analysis, the best-performing models for the Gujarati language showcased Word Error Rates of 14.06%, 14.70%, and 15.04% respectively. Across all cases, the most effective systems demonstrated a notable ability to surpass the performance of the TDNN baseline, signifying significant improvements in recognition accuracy (Srivastava et al., 2018).

The Free Spoken Digit Gujarati Dataset prompts the conclusion that, following cross-corpus experiments conducted by the research team, there exists substantial potential for enhancing the dataset's capabilities. This enhancement involves exploring diverse features and techniques to augment the system's ability to generalize effectively across various contexts (Dalsaniya et al., 2020).

In conclusion, the evaluation of the Gujarati speech corpus based on accuracy parameters has yielded valuable insights into its quality and applicability. Through meticulous analysis, we have discerned the accuracy levels of various models and methodologies applied to the corpus, shedding light on its performance in speech processing tasks.

The assessment of accuracy metrics, including Word Error Rate (WER), Mean Opinion Score (MOS), and other relevant measures, has provided a comprehensive understanding of the corpus's strengths and limitations. These accuracy-based evaluations have demonstrated the effectiveness of certain approaches, as evidenced by the improvement in recognition rates observed across experiments.

However, while accuracy evaluations offer substantial insights, it's important to recognize the broader context in which the corpus will be utilized. Future research endeavors should aim to strike a balance between accuracy and other factors like robustness, adaptability, and application-specific requirements.Through this evaluation, not only can the quality of the analysis be assessed, but also insights can be gleaned into the real-world implications of the findings.

## Future Directions

Voice recognition technology in Gujarati holds promising applications across various domains, contributing to improved accessibility, communication, and efficiency. Additionally, there are several exciting directions for future research and development in this field. The applications of voice recognition technology in Gujarati are numerous and varied, with potential benefits spanning communication, accessibility, and efficiency. Future directions involve the integration of advanced language models, handling challenges specific to Gujarati, and enhancing the sophistication of voice recognition systems. Continued research and development in these directions will pave the way for more natural and effective human-computer interactions in the Gujarati language.

## Conclusion

In this research paper, we embarked on a comprehensive journey through the analysis of a dedicated Gujarati speech corpus. The corpus, carefully curated from diverse sources, offered a rich tapestry of the Gujarati language in various registers, accents, and contexts. Through rigorous analysis, we explored the phonetic, acoustic, and linguistic characteristics inherent in the language's spoken form. Throughout our analysis, we acknowledged the challenges posed by dialectal variations, and speaker diversity. These hurdles highlight the complexity of accurately recognizing and interpreting Gujarati speech, urging the development of specialized models and techniques. The applications of this analysis are far-reaching. From developing accurate voice recognition systems that facilitate customer service and education to preserving cultural heritage and fostering accessibility, the impact on technology and society is profound.

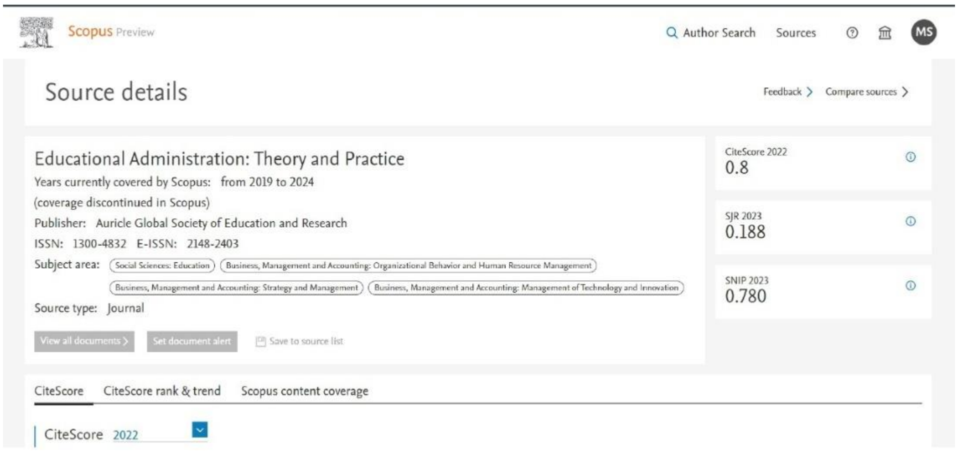
As we conclude this journey, it is evident that the analysis of the Gujarati speech corpus lays the groundwork for advancements in both linguistics and technology. It shapes the trajectory of voice recognition in Gujarati, paving the way for improved human-computer interaction, cross-cultural communication, and linguistic research. With continued collaboration, innovation, and adaptation, the future holds the promise of a more inclusive, technologically empowered, and linguistically enriched landscape for Gujarati speakers around the world.

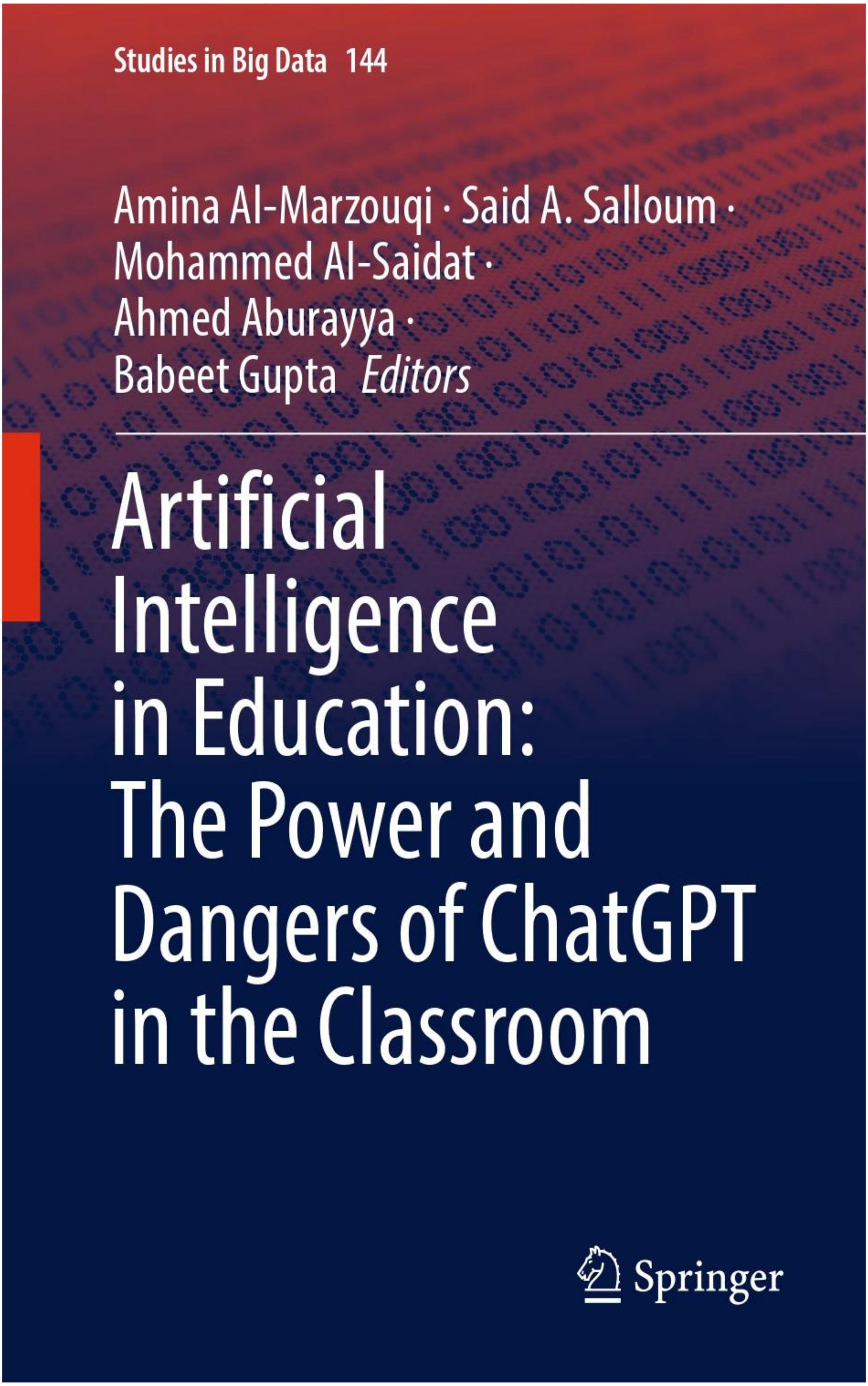
## References


1. C. Yarra, R. Aggarwal, A. Rajpal and P. K. Ghosh, "Indic TIMIT and Indic English lexicon: A speech database of Indian speakers using TIMIT stimuli and a lexicon from their mispronunciations," 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA), Cebu, Philippines, 2019, pp. 1-6, doi: 10.1109/O-COCOSDA46868.2019.9041230.
2. He, F., Chu, S. H. C., Kjartansson, O., Rivera, C. E., Katanova, A., Gutkin, A., Demirsahin, I., Johny, C. C., Jansche, M., Sarin, S., & Pipatsrisawat, K. (2020). Open-source multi-speaker speech Corpora for building Gujarati, Kannada, Malayalam, Marathi, Tamil and Telugu speech synthesis systems. Language Resources and Evaluation,6494–6503.http://dblp.uni-trier.de/db/conf/lrec/lrec2020.html#HeCKRKGDJJSP20
3. S. Maity, A. Kumar Vuppala, K. S. Rao and D. Nandi, "IITKGP-MLILSC speech database for language identification," 2012 National Conference on Communications (NCC), Kharagpur, India, 2012, pp. 1-5, doi: 10.1109/NCC.2012.6176831.

4. Bhogale, K. S., Raman, A., Javed, T., Doddapaneni, S., Kunchukuttan, A., Kumar, P., & Khapra, M. M. (2022). Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2208.12666
5. K. Bhogale et al., "Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10096933.
6. Srivastava, B.M.L., Sitaram, S., Kumar Mehta, R., Doss Mohan, K., Matani, P., Satpal, S., Bali, K., Srikanth, R., Nayak, N. (2018) Interspeech 2018 Low Resource Automatic Speech Recognition Challenge for Indian Languages. Proc. 6th Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU 2018), 11-14, doi: 10.21437/SLTU.2018-3
7. Dalsaniya, N., Mankad, S.H., Garg, S., Shrivastava, D. (2020). Development of a Novel Database in Gujarati Language for Spoken Digits Classification. In: Thampi, S., et al. Advances in Signal Processing and Intelligent Recognition Systems. SIRS 2019. Communications in Computer and Information Science, vol 1209. Springer, Singapore. https://doi.org/10.1007/978-981-15-4828-4_18
8. Choudhary, N., & Rao, D. G. (2020). The LDC-IL Speech Corpora. 2020 23rd Conference of the Oriental COCOSDA International Committee for the Co-Ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA). doi:10.1109/o-cocosda50338.2020.9295011
9. Ramamoorthy L., Narayan Kumar Choudhary, Mona Parakh, Hiren Gadhavi R, Solanki Mahesh kumar R, Rejitha K. S., Rajesha N., Manasa, G.., 2021. Gujarati Raw Speech Corpus. Central Institute of Indian Languages, Mysore.
10. Choudhary, Narayan, Rajesha N., Manasa G. & L. Ramamoorthy. 2019. "LDC-IL Raw Speech Corpora: An Overview" in Linguistic Resources for AI/NLP in Indian Languages. Central Institute of Indian Languages, Mysore. pp. 160-174.
11. https://www.kaggle.com/datasets/hbchaitanyabharadwaj/audio-dataset-with-10-indian-languages
12. 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). (n.d.). IEEE.
13. Aarti Bakshi, & Sunil kumar kopparapu. (2018). Spoken Indian Language Identification: a review of features and database.
14. Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2019). Common Voice: A Massively-Multilingual Speech Corpus. http://arxiv.org/abs/1912.06670
15. Basu, J., Khan, S., Roy, R., Basu, T. K., & Majumder, S. (2021). Multilingual Speech Corpus in Low-Resource Eastern and Northeastern Indian Languages for Speaker and Language Identification. Circuits, Systems, and Signal Processing, 40(10), 4986–5013. https://doi.org/10.1007/s00034-021-01704-x
16. Bhogale, K. S., Raman, A., Javed, T., Doddapaneni, S., Kunchukuttan, A., Kumar, P., & Khapra, M. M. (2022). Effectiveness of Mining Audio and Text Pairs from Public Data for Improving ASR Systems for Low-Resource Languages. http://arxiv.org/abs/2208.12666
17. Kothari Associate Professor, J., P Shah, S. M., Kumbharana, C., & Professor, A. (2015). A Phonetic Study for Constructing a Database of Gujarati Characters for Speech Synthesis of Gujarati Text. In International Journal of Computer Applications (Vol. 117, Issue 19).
18. kurian, C. (2014). A Review on Speech Corpus Development for Automatic Speech Recognition in Indian Languages.
19. Prahallad, K., Naresh Kumar, E., Keri, V., Rajendran, S., & Black, A. W. (n.d.). The IIIT-H Indic Speech Databases. http://www.isca-speech.org/archive
20. Shrishrimal, P. P., Deshmukh, R. R., & Waghmare, V. B. (2012a). Indian Language Speech Database: A Review. In International Journal of Computer Applications (Vol. 47, Issue 5).
21. Sztahó, D., Szaszák, G., & Beke, A. (n.d.-a). Deep learning methods in speaker recognition: a review.
22. Thampi, S. M., Rajesh, ·, Hegde, M., Krishnan, S., Mukhopadhyay, J., Chaudhary, V., Marques, O., Piramuthu, S., & Corchado, J. M. (n.d.). Revised Selected Papers Advances in Signal Processing and Intelligent Recognition Systems Communications in Computer and Information Science 1209. http://www.springer.com/series/7899
23. Kibria, S., Samin, A. M., Kobir, M. H., Rahman, M. S., Selim, M. R., & Iqbal, M. Z. (2022). Bangladeshi Bangla speech corpus for automatic speech recognition research. Speech Communication, 136, 84-97.
24. Raulji, J., & Saini, J. R. (2020). Bilingual Dictionary for Sanskrit—Gujarati MT Implementation. In ICT Analysis and Applications: Proceedings of ICT4SD 2019, Volume 2 (pp. 463-470). Springer Singapore.
25. Gujarathi, P. V., & Patil, S. R. (2021). Gaussian Filter-Based Speech Segmentation Algorithm for Gujarati Language. In Smart Computing Techniques and Applications: Proceedings of the Fourth International Conference on Smart Computing and Informatics, Volume 2 (pp. 747-756). Springer Singapore.
26. Aasofwala, N., Verma, S., & Patel, K. (2021, September). A Novel Speech to Sign Communication Model for Gujarati Language. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1-5). IEEE.

27. Vaishnav, Z. B., & Sajja, P. S. (2019). Knowledge-based approach for word sense disambiguation using genetic algorithm for Gujarati. In Information and Communication Technology for Intelligent Systems: Proceedings of ICTIS 2018, Volume 1 (pp. 485-494). Springer Singapore.
28. Srivastava, N., Mukhopadhyay, R., Prajwal, K. R., & Jawahar, C. V. (2020, May). Indicspeech: text-to-speech corpus for indian languages. In Proceedings of the 12th language resources and evaluation conference (pp. 6417-6422).
29. Basu, J., Khan, S., Roy, R., Saxena, B., Ganguly, D., Arora, S., ... & Agrawal, S. S. (2019, October). Indian languages corpus for speech recognition. In 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA) (pp. 1-6). IEEE.
30. Terhiija, V., Sarmah, P., & Vijaya, S. (2019). Development of Speech Corpus and Automatic Speech Recognition of Angami. 22nd Oriental-COCOSDA, Cebu, Philippines.
31. Siripragada, S., Philip, J., Namboodiri, V. P., & Jawahar, C. V. (2020). A multilingual parallel corpora collection effort for Indian languages. arXiv preprint arXiv:2007.07691
32. Shetty, V. M., & NJ, M. S. M. (2020, May). Improving the performance of transformer based low resource speech recognition for Indian languages. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 8279-8283). IEEE.


Studies in Big Data 144
Amina Al-Marzouqi · Said A. Salloum ·
Mohammed Al-Saidat ·
Ahmed Aburayya ·
Babeet Gupta Editors
Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom
Springer

# Unveiling the Future: Exploring Conversational AI

**Meera M. Shah and Hiren R. Kavathiya**

## 1 Introduction

In an era defined by rapid technological advancement, one of the most captivating frontiers in artificial intelligence (AI) is Conversational AI. This chapter embarks on an expedition into the depths of this captivating realm, unfurling its historical roots, current manifestations, and the tantalizing prospects it holds for the future. Conversational AI, with its transformative capabilities, has redefined the human–machine interface, ushering in novel ways of communication, commerce, and innovation.

The chapters to follow will unravel the intricate layers of Conversational AI, examining its evolutionary journey, dissecting its fundamental components, and exploring its diverse applications across industries. From the intricate dance of Natural Language Understanding (NLU) to the artistry of Natural Language Generation (NLG), this chapter will paint a comprehensive portrait of Conversational AI's inner workings.

Moreover, we will delve into the fascinating interplay of dialogue management and the critical role it plays in orchestrating seamless conversations between humans and machines. As the landscape of AI continues to evolve, Conversational AI stands as a pinnacle of human ingenuity, a domain where technology endeavors to bridge the gap between the synthetic and the human [1]. We will navigate the ethical currents that underpin this technological transformation, addressing concerns such as bias mitigation, data privacy, and the ever-pressing need for responsible AI development.

Furthermore, this chapter gazes toward the horizon, envisioning the potential trajectories that Conversational AI might take in the coming years. The concept

M. M. Shah (✉) · H. R. Kavathiya
Department of Computer Science, Atmiya University, Rajkot, India
e-mail: mira.shah@atmiyauni.ac.in

H. R. Kavathiya
e-mail: hiren.kavathiya@atmiyauni.ac.in

 511
A. Al-Marzouqi et al. (eds.), *Artificial Intelligence in Education: The Power and Dangers of ChatGPT in the Classroom*, Studies in Big Data 144,
https://doi.org/10.1007/978-3-031-52280-2_32

of conversational understanding, where AI systems navigate the ebb and flow of conversations with a human-like adeptness, emerges as an alluring destination. The fusion of multimodal interactions and emotional intelligence promises to elevate AI to unprecedented levels of sophistication, fundamentally altering the ways in which we interact with machines.

## 2 Evolution of Conversional AI

The roots of Conversational AI stretch back to the earliest aspirations of technology to decipher and engage with human language. The journey from rudimentary rule-based systems to the sophisticated neural networks of today has been marked by leaps of innovation, technological breakthroughs, and a deepening understanding of the intricacies of language [2].

### *2.1 Early Attempts and Rule-Based Systems (1950s-1990s)*

The inception of Conversational AI can be traced to the 1950s when computer scientists began exploring the possibilities of automating human language understanding. Early efforts involved rule-based systems that relied on predefined grammatical rules and a limited vocabulary to simulate conversations. ELIZA, created in the mid-1960s by Joseph Weizenbaum, was a groundbreaking example. Although primitive, ELIZA could engage in simple textual exchanges, laying the foundation for future advancements.

### *2.2 Statistical Approaches and Limited Context (2000s)*

The turn of the millennium witnessed a shift towards statistical methods, with probabilistic models enabling AI systems to predict the likelihood of words and phrases appearing in a given context. Yet, these systems struggled with understanding nuanced language and context shifts. The introduction of chatbots for customer support marked an early application, albeit with limited success due to their constrained capabilities.

### *2.3 Rise of Machine Learning and Neural Networks (2010s)*

The emergence of deep learning in the 2010s sparked a renaissance in Conversational AI. Neural networks, especially the Transformer architecture, revolutionized language understanding and generation. Models like Google's BERT and OpenAI's GPT series demonstrated unprecedented capabilities in capturing context, semantics, and subtleties in language. These models learned from massive datasets, empowering AI to produce more human-like responses and enabling broader applications.

### *2.4 Conversional Context and Personalization*

As AI systems became more proficient at understanding context, maintaining conversation flow, and generating coherent responses, they ventured into personalization. Conversational agents began adapting to user preferences, enhancing user experiences across various domains. Additionally, contextual understanding improved, allowing AI to comprehend references made earlier in conversations, mimicking the continuity of human discourse.

### *2.5 Multimodal Interactions and Beyond*

In recent years, the evolution of Conversational AI has taken a multimodal turn, embracing inputs beyond text, such as images, voice, and gestures. This expansion aims to mirror human communication, where contextual cues are derived from diverse sources. The integration of multimodal interactions amplifies the potential for more natural and immersive conversations.

The evolution of Conversational AI showcases a remarkable journey from its humble beginnings to its current prowess in understanding and generating human-like language. As we peer into the future, the ongoing fusion of AI with linguistic intricacies holds the potential to redefine human–machine communication, fostering an era where interactions are not just transactional but genuinely conversational [2]. This evolution is a testament to human ingenuity and an unwavering desire to bridge the gap between the synthetic and the human, propelling Conversational AI into the forefront of technological innovation.

## 3 Foundation of Conversional AI: NLP and Machine Learning

Conversational AI is built upon a strong foundation of Natural Language Processing (NLP) and Machine Learning (ML) techniques. This fusion of disciplines empowers machines to understand, generate, and engage in human-like conversations, bridging the gap between human communication and artificial intelligence.

## 3.1 Natural Language Processing (NLP)

At the heart of the transformative power of Conversational AI lies Natural Language Processing (NLP), the wizardry that bestows machines with the ability to understand, interpret, and generate human language. This chapter delves into the intricate realm of NLP within Conversational AI, uncovering the technologies, techniques, and advancements that enable machines to engage in dynamic and contextually relevant conversations. From syntax to sentiment analysis, from understanding context to generating coherent responses, NLP is the cornerstone that breathes life into human–machine discourse [3].

### 3.1.1 Linguistic Foundations: Decoding the Grammar of Conversations

The journey into NLP's role in Conversational AI begins with an exploration of the linguistic foundations. This section illuminates the components of language processing, from tokenization and part-of-speech tagging to parsing and syntactic analysis. The magic of syntactic trees and grammatical structures becomes evident as we unravel how NLP algorithms dissect and comprehend the grammatical intricacies of human language.

### 3.1.2 Beyond Words: Understanding Semantics and Context

Words are the building blocks of language, but meaning transcends mere vocabulary. Delving into semantics, this part delves into how NLP techniques capture the deeper meaning of words and their relationships in sentences. The emergence of word embeddings and distributional semantics enables machines to discern semantic similarities, paving the way for more nuanced comprehension and relevant responses.

### 3.1.3 Intent Recognition and Named Entity Recognition

The crux of effective Conversational AI lies in understanding user intent and extracting relevant information. This section uncovers the techniques of intent recognition and named entity recognition. Through the lens of NLP, machines learn to identify the goals and objectives behind user queries, as well as discerning crucial entities such as names, dates, and locations. This fosters context-aware conversations and enhances the overall user experience.

### 3.1.4 Contextual Understanding: The Key to Coherence

Context is the glue that binds meaningful conversations. This segment delves into how NLP equips AI systems to understand and maintain contextual continuity in conversations. Techniques like coreference resolution ensure that pronouns and references are properly linked, creating a coherent and fluid dialogue that mirrors human interaction.

### 3.1.5 Sentiment Analysis: Gauging Emotional Nuances

Human conversations are often infused with emotions that shape the tone and intent. Sentiment analysis, a pivotal aspect of NLP, enables AI systems to gauge emotional nuances within text. This part explores how machines identify sentiments like joy, anger, and sadness, allowing them to tailor responses that resonate with the user's emotional state.

### 3.1.6 Response Generation: The Art of Human-Like Replies

Generating responses that mimic human dialogue is the ultimate aim of Conversational AI. This section delves into the techniques of response generation, from rule-based approaches to the transformative power of neural language models. The advent of transformers has revolutionized response generation, enabling AI systems to produce contextually relevant, coherent, and contextually-aware replies.

### 3.1.7 Ethical Considerations in NLP

As NLP-driven Conversational AI becomes more integral to our lives, ethical considerations gain prominence. This part delves into the challenges of bias in training data, potential misinformation propagation, and the importance of responsible AI practices. The chapter also underscores the significance of transparency, fairness, and user consent in maintaining ethical integrity.

The marriage of NLP and Conversational AI is a symphony of linguistic understanding and technological innovation. From deciphering grammar to grasping context, from understanding emotions to generating human-like responses, NLP is the thread that weaves the magic of human–machine discourse. This chapter has illuminated the intricate workings of NLP within Conversational AI, highlighting its crucial role in shaping a future where machines converse with us seamlessly and intelligently.

## 3.2 Machine Learning in Conversional AI

Conversational AI's evolution has been deeply intertwined with the rise of Machine Learning (ML), a transformative force that empowers machines to learn from data and adapt their behavior. This chapter embarks on an illuminating journey through the realm of ML in Conversational AI, unveiling the algorithms, techniques, and innovations that enable AI systems to not only understand human language but also engage in dynamic and contextually relevant conversations.

### 3.2.1 Foundations of Machine Learning: Learning from Data

The bedrock of ML in Conversational AI lies in understanding how machines learn from data. This section introduces the fundamental concepts of supervised, unsupervised, and reinforcement learning, providing readers with insights into the underlying mechanisms that enable AI systems to make data-driven decisions and refine their responses over time.

### 3.2.2 Supervised Learning: The Path to Contextual Understanding

Supervised learning is a cornerstone of Conversational AI, enabling machines to recognize patterns in labeled data. This part delves into how ML algorithms, such as support vector machines and neural networks, are trained to understand context, intent, and sentiment within human language [4]. Through annotated datasets, AI systems navigate the nuances of dialogue and build a foundation for meaningful conversations.

### 3.2.3 Unsupervised Learning: Extracting Insights from Text

Conversations are rich sources of unstructured text, and unsupervised learning techniques shine in extracting valuable insights. This segment explores how algorithms like clustering and topic modeling unravel the hidden structures and themes within conversations. These techniques not only aid in organizing information but also provide AI systems with a deeper understanding of user preferences and interests [4].

### 3.2.4 Reinforcement Learning: Dynamic Interaction and Response Refinement

The art of conversation is a dynamic dance of responses and feedback. This part unveils the role of reinforcement learning, where AI agents learn through trial and error, receiving rewards for desirable responses and refining their behavior over time. From training dialogue agents in simulated environments to real-world interactions, reinforcement learning imbues Conversational AI with adaptability and continuous improvement [5].

### 3.2.5 End-To-End Training: Seamless Dialogue Generation

One of the transformative aspects of ML in Conversational AI is end-to-end training. This section explores how neural network architectures, such as sequence-to-sequence models, enable the training of systems that can generate human-like responses directly from input queries [2]. By minimizing the need for handcrafted rules, end-to-end models usher in a more fluid and natural conversational experience.

### 3.2.6 Transfer Learning and Pre-trained Models

The emergence of transfer learning has been a game-changer in Conversational AI. This segment delves into how pre-trained language models, like GPT-3, have reshaped the landscape by learning from vast amounts of text data [6]. These models encapsulate diverse linguistic patterns, allowing AI systems to engage in conversations spanning a wide spectrum of topics and styles.

### 3.2.7 Adversarial Training and Robustness

In the realm of Conversational AI, robustness against adversarial attacks is crucial. This part explores how adversarial training techniques bolster AI systems against manipulative inputs and ensure coherent responses even in the face of crafted perturbations. The quest for robustness enhances Conversational AI's ability to maintain meaningful interactions without succumbing to deliberate distortions.

Machine Learning is the driving force that propels Conversational AI into the realm of adaptive and intelligent interactions. From supervised learning's foundation in understanding context to reinforcement learning's dynamic responsiveness, the fusion of ML and Conversational AI is a symphony of data-driven ingenuity. This chapter has offered a glimpse into this synergy, emphasizing how ML infuses Conversational AI with the capability to evolve, learn, and engage in conversations that resonate with human understanding.

The foundations of Conversational AI, rooted in Natural Language Processing and Machine Learning, form the bedrock upon which intelligent conversations with machines are built. The marriage of linguistics and technology empowers machines to comprehend the intricacies of language, adapt to context, and generate human-like responses. As these foundational techniques continue to evolve, they propel Conversational AI toward unprecedented heights, ushering in an era where machines seamlessly converse with humans, enriching interactions across domains and driving the boundaries of technological possibility.

## 4 Applications of Conversational AI Across Industries

Conversational AI's transformative capabilities have transcended boundaries, infiltrating diverse industries and reshaping how businesses interact with their customers, clients, and stakeholders. From healthcare to finance and beyond, the technology's applications have been instrumental in enhancing user experiences, optimizing processes, and driving innovation. Here, we explore some of the prominent sectors where Conversational AI has made its indelible mark.

### *4.1 Customer Service and Support*

Customer service has undergone a paradigm shift with the introduction of Conversational AI. Chatbots and virtual assistants provide instant and round-the-clock support, answering common queries, troubleshooting issues, and guiding users through processes. These AI-driven systems not only reduce response times but also free up human agents to focus on more complex inquiries, thus improving overall customer satisfaction.

### *4.2 Healthcare and Telemedicine*

In the healthcare sector, Conversational AI has enabled telemedicine to thrive. Virtual health assistants offer preliminary diagnoses, schedule appointments, and provide patients with accurate medical information. These AI-powered companions have played a crucial role during the COVID-19 pandemic, offering reliable guidance while alleviating the burden on healthcare providers.

### *4.3 E-commerce and Retail*

Conversational AI has transformed the online shopping experience. Chatbots assist customers in finding products, offering personalized recommendations, and aiding in the purchasing process [7]. By understanding customer preferences and behavior, AI-driven systems enhance engagement, drive sales, and foster customer loyalty.

### 4.4 Finance and Banking

In the finance industry, Conversational AI has streamlined interactions between customers and financial institutions. Virtual assistants help with account inquiries, balance checks, fund transfers, and even investment advice. These systems ensure faster and more convenient transactions while maintaining a high level of data security.

### 4.5 Education and E-learning

Conversational AI has infiltrated the education sector, creating intelligent tutoring systems that provide personalized learning experiences. These systems adapt to students' learning styles and pace, offering real-time feedback and guidance. Conversational AI also plays a role in language learning, offering interactive practice and language immersion.

### 4.6 Entertainment and Media

Entertainment has embraced Conversational AI through interactive storytelling experiences, voice-controlled gaming, and personalized content recommendations. AI-driven chatbots can engage users in imaginative narratives, creating a dynamic and immersive form of entertainment.

### 4.7 Hospitality and Travel

In the hospitality industry, Conversational AI enhances guest experiences by offering information about accommodations, local attractions, and services. Hotels and travel companies deploy AI-powered concierge services to answer queries, arrange bookings, and provide travel recommendations, elevating the overall travel experience [8].

### 4.8 Human Resources and Recruitment

Conversational AI has revolutionized HR processes by automating candidate screening, scheduling interviews, and answering frequently asked questions. Virtual assistants efficiently handle administrative tasks, allowing HR professionals to focus on strategic initiatives and fostering a more efficient recruitment process.

### *4.9 Government and Public Services*

Governments have harnessed Conversational AI to provide citizens with information about public services, government programs, and policy updates. Virtual assistants address queries, guide citizens through bureaucratic processes, and offer a user-friendly means of accessing important information.

The proliferation of Conversational AI across various industries underscores its versatility and potential to reshape traditional business models. By automating routine tasks, personalizing interactions, and providing efficient customer support, Conversational AI enhances user experiences and fosters deeper engagement. As industries continue to embrace and innovate with this technology, the boundaries of what's achievable through AI-powered conversations continue to expand, opening doors to a future where machines seamlessly collaborate with humans across a myriad of sectors.

## 5 Challenges and Ethical Consideration

The evolution of Conversational AI is accompanied by a range of challenges and ethical considerations that require careful examination. While these technologies hold immense potential, it is crucial to navigate them responsibly to ensure that their deployment aligns with societal values, respects user rights, and avoids harm. Let's explore some of the significant challenges and ethical considerations in Conversational AI.

### *5.1 Challenges*

#### 5.1.1 Bias and Fairness

Addressing bias in Conversational AI is a paramount challenge. Biases present in training data can lead to discriminatory responses, perpetuating social inequalities. Ensuring diverse and representative datasets and employing debiasing techniques are essential to mitigate this challenge.

#### 5.1.2 Context and Understanding

Developing AI systems that understand and respond coherently to contextually nuanced conversations is challenging. Contextual understanding requires fine-tuning models to capture subtle shifts, maintaining meaningful conversations over time.

#### 5.1.3 Privacy and Data Security

The vast amount of personal data processed in Conversational AI raises concerns about user privacy. Striking a balance between providing personalized experiences and safeguarding user information is essential, involving robust data encryption and transparent data handling practices.

#### 5.1.4 Misinformation and Manipulation

Conversational AI can inadvertently propagate misinformation or be manipulated to spread false information. Ensuring that AI systems are equipped with fact-checking capabilities and are resistant to malicious manipulation is a critical challenge.

#### 5.1.5 Accountability and Liability

Determining who is accountable when AI systems generate incorrect or harmful responses is complex. Establishing clear lines of responsibility and liability in cases of negative outcomes is challenging yet necessary to ensure accountability.

### *5.2 Ethical Considerations*

#### 5.2.1 Transparency and Explainability

Transparency is key to building trust between users and AI systems. Conversational AI should provide clear explanations for its decisions and disclose its AI-driven nature, enabling users to understand and trust the technology's processes.

#### 5.2.2 User Consent and Control

Obtaining informed consent for data usage and interactions with AI systems is vital. Users should have control over their data and be able to opt-out or delete information collected during interactions.

#### 5.2.3 Human-Like Deception

As AI systems become more sophisticated, there's a concern that they might appear too human-like, potentially deceiving users into believing they are interacting with humans. Implementing mechanisms to clearly identify AI interactions can mitigate this concern.

#### 5.2.4 Impact on Human Relationships

Over Reliance on Conversational AI might impact human-to-human interactions. Striking a balance between AI assistance and genuine human communication is important to preserve meaningful relationships.

#### 5.2.5 Socio Economic Impact

The widespread adoption of Conversational AI could potentially impact job roles, particularly those involving customer service and support. Mitigating the potential negative socioeconomic consequences through upskilling and retraining initiatives is an ethical consideration.

Conversational AI holds immense promise for transforming various aspects of our lives, but its deployment must be guided by a thorough understanding of the challenges and ethical considerations it entails. By addressing bias, ensuring transparency, safeguarding privacy, and fostering accountability, we can build AI systems that contribute positively to society. Ethical foresight and responsible development are vital to harnessing the potential of Conversational AI while safeguarding human values and well-being [9].

## 6 The Future of Conversational AI: A Glimpse into Tomorrow

Conversational AI has embarked on an extraordinary journey, evolving from simple rule-based systems to complex neural networks capable of understanding and generating human-like language. As we peer into the future, the trajectory of Conversational AI reveals a landscape filled with possibilities that promise to redefine the ways we communicate, interact with technology, and experience the world. Here, we explore the exciting prospects that await Conversational AI in the years to come.

## 6.1 Multimodal Mastery

The future of Conversational AI is set to embrace a convergence of communication modes. Systems will seamlessly integrate text, speech, images, and even gestures, creating richer, more immersive interactions. Imagine instructing a virtual assistant using voice, showing it an image for context, and receiving a detailed response that takes into account various modalities of input.

## 6.2 Emotional Intelligence

As AI systems become more sophisticated, they will develop a nuanced understanding of human emotions. Future Conversational AI could recognize subtle emotional cues in speech and text, responding with empathy, compassion, and even humor. This emotional intelligence will enable AI to establish deeper connections and enhance user experiences.

## 6.3 Specialization and Expertise

Conversational AI will evolve from being general-purpose to domain-specific experts. Specialized virtual assistants could provide intricate knowledge and support in fields like medicine, law, finance, and more. This expertise will extend beyond basic interactions, offering in-depth insights and informed recommendations tailored to specific industries.

## 6.4 Contextual Continuity

Future Conversational AI systems will excel in maintaining conversations that span days, weeks, or even months. They will seamlessly remember past interactions, capture nuances in conversation flow, and provide continuity as if the conversation never ended. This prolonged context will contribute to more natural, engaging dialogues.

## 6.5 Ethical AI

The future of Conversational AI lies hand in hand with ethical considerations. AI developers and organizations will prioritize responsible AI development, ensuring transparency, fairness, and adherence to privacy regulations. Ethical guidelines will shape the design, deployment, and behavior of AI systems, fostering trust and minimizing harm.

### *6.6 Personal AI Companions*

Imagine having your own AI companion that understands your preferences, anticipates your needs, and helps manage your daily life. Future Conversational AI could evolve into personalized digital assistants that schedule appointments, manage tasks, curate content, and offer companionship.

### *6.7 Advancements in Learning*

AI models will become even more adept at learning from fewer examples and adapting to new tasks. This will expedite the development and deployment of Conversational AI, making it easier to create specialized conversational agents that cater to specific industries and user needs.

### *6.8 Global Language Accessibility*

Future Conversational AI models will be proficient in a multitude of languages and dialects, breaking down language barriers and fostering cross-cultural communication. This will enable people around the world to engage with AI systems in their native languages, making technology more inclusive.

The future of Conversational AI holds a world brimming with transformative potential. Through the integration of multimodal interactions, emotional intelligence, domain expertise, and ethical considerations, these systems are poised to revolutionize the way we communicate, work, learn, and live. As we navigate the uncharted waters of the AI frontier, responsible development and innovation will be essential to harness the vast benefits of Conversational AI while upholding the values that define our humanity.

## 7 Conclusion: Navigating the Conversational AI Frontier

The journey through the realms of Conversational AI has unveiled a landscape of innovation, challenges, and ethical considerations. From its modest beginnings in rule-based systems to the sophisticated neural networks of today, Conversational AI has transformed the way we interact with technology, enriching human experiences and reshaping industries across the globe.

As we reflect on the chapters traversed in this exploration, several key takeaways emerge. Conversational AI's foundations in Natural Language Processing and Machine Learning provide the building blocks for systems that can understand, generate, and manage human-like conversations. This fusion of linguistics and technology gives rise to machines that bridge the gap between the synthetic and the human, challenging us to find a harmonious coexistence.

Yet, this journey is not without its challenges. Bias, privacy concerns, and the potential for unintended consequences beckon for ethical considerations to steer the path forward. Developers, policymakers, and society at large are tasked with charting a course that upholds transparency, fairness, and accountability while harnessing the immense potential of these technologies.

As we peer into the future, Conversational AI beckons us toward a realm of unimaginable possibilities. Multimodal interactions, emotional intelligence, specialized expertise, and ethical AI stand as beacons guiding our aspirations. The evolution of Conversational AI is an embodiment of human ingenuity, stretching the boundaries of innovation and illuminating the intricate dance between human creativity and machine intelligence.

In closing, the chapters unveiled in this exploration only mark the beginning of the Conversational AI saga. The narrative continues, inviting us to engage with questions, challenges, and innovations that shape the way we communicate, connect, and coexist. As we journey forward, the lessons learned from the past and the ethical considerations pondered today will guide us toward a future where Conversational AI enriches our lives, respects our values, and propels us toward a more interconnected and intelligent world.

## References

1. J. Ford, V. Jain, K. Wadhwani, D.G. Gupta, AI advertising: an overview and guidelines. J. Bus. Res. **166**, 114124 (2023). https://doi.org/10.1016/j.jbusres.2023.114124
2. A. Haleem, M. Javaid, M.A. Qadri, R.P. Singh, R. Suman, Artificial intelligence (AI) applications for marketing: a literature-based study. Int. J. Intell. Netw. **3**, 119–132 (2022). https://doi.org/10.1016/j.ijin.2022.08.005
3. S.M. Cheema, S. Tariq, I.M. Pires, A natural language interface for automatic generation of data flow diagram using web extraction techniques. J. King Saud Univ.-Comput. Inf. Sci. **35**(2), 626–640 (2023). https://doi.org/10.1016/j.jksuci.2023.01.006
4. J. Chubb, S. Missaoui, S. Concannon, L. Maloney, J.A. Walker, Interactive storytelling for children: a case-study of design and development considerations for ethical conversational AI. Int. J. Child-Comput. Interact. **32**, 100403 (2022). https://doi.org/10.1016/j.ijcci.2021.100403
5. M. Schmitt, Automated machine learning: AI-driven decision making in business analytics. Intell. Syst. Appl. **18**, 200188 (2023). https://doi.org/10.1016/j.iswa.2023.200188
6. K. Barnova, M. Mikolasova, R. Kahankova, R. Jaros, A. Kawala-Sterniuk, V. Snášel, S. Mirjalili, M. Pelc, R. Martinek, Implementation of artificial intelligence and machine learning-based methods in brain-computer interaction. Comput. Biol. Med. **163**, 107135 (2023). https://doi.org/10.1016/j.compbiomed.2023.107135

7. T. Schachner, R. Keller, F. Wangenheim, Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J. Med. Internet Res. **22**(9), e20701 (2020)
8. J.L.Z. Montenegro, C.A. da Costa, R. da Rosa Righi, Survey of conversational agents in health. Expert Syst. Appl. **129**, 56–67 (2019)
9. F. Cichos, S.M. Landin, R. Pradip, Artificial intelligence (AI) enhanced nanomotors and active matter, in *Elsevier eBooks*, pp. 113–144 (2023). https://doi.org/10.1016/b978-0-323-85796-3.00005-6
10. R. Dazeley, P. Vamplew, C. Foale, C. Young, S. Aryal, F. Cruz, Levels of explainable artificial intelligence for human-aligned conversational explanations. Artif. Intell. **299**, 103525 (2021)
11. S. Hussain, O. Ameri Sianaki, N. Ababneh, A survey on conversational agents/chatbots classification and design techniques, in *Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications (WAINA-2019)*, vol. 33 (Springer International Publishing, 2019), pp. 946–956

# Voice Recognition for Gujarati Dialects: An in-depth Survey

**Research** · January 2024

DOI: 10.13140/RG.2.2.22120.85763



**2 authors**, including:

Meera Shah
Atmiya Institute Of Technology & Science

*International Journal of Computer Applications (0975 – 8887)*
*Volume 186 – No.5, January 2024*

# Voice Recognition for Gujarati Dialects: An in-depth Survey

Meera M. Shah
Department of Computer Science, Atmiya university, Rajkot, 360005, Gujarat

Hiren R. Kavathiya, PhD
Department of Computer Science, Atmiya university, Rajkot, 360005, Gujarat

## ABSTRACT
Voice recognition technology nowadays is gaining so much importance, and plenty of work has been done on it for different languages like English, Arabic, Hindi, Chinese, etc. But when we talk about a language like Gujarati, we find a particular lack of work. In this paper, we examined the process of voice recognition in Gujarati. The systematic literature review for voice recognition has been shown here. This paper mainly focuses on the problems that can be found in voice recognition systems for Gujarati.

## Keywords
Voice Recognition, Speech Processing, Gujrati, Feature Extraction, MFCC, HMM

## 1. INTRODUCTION
A user-friendly interface is provided by voice recognition systems to the user. Having it in natural languages will make it more beneficial. Voice recognition software allows people with impairments and those who are less at ease using machines due to lack of expertise or a language barrier to use technology. The user benefits from a convenient, hands-free environment thanks to voice recognition in native languages. The voice recognition approach is frequently used to address practical problems. The effectiveness and performance of speaker recognition systems are influenced by numerous factors. The task of creating an autonomous speaker recognition system is difficult because of the many grammatical conventions, noisy surroundings, and speaker pronunciations.

## 2. ANALYSIS OF RESEARCH WORK IN DIFFERENT LANGUAGES
The work done on speaker recognition for several languages in different situations, along with multiple feature extraction techniques, is summarized in the tables below.

**Table 1: Analysis of Voice Recognition in different languages**

| Author | Dataset/ Language | Performance | Major Findings |
|---|---|---|---|
| [2] | Real Speech Dataset English | 93.33% accuracy | Noisy Speech Signals Affects the performance |
| [3] | Voxceleb (English Language) | Speaker Identification 89.1% Accuracy & Speaker Verification EER 5.5% | Accuracy can be improved in terms of speech identification |
| [4] | Real Speech Dataset Hindi (10 M /5 F & 17 trails) | Text Independent : MFCC-VQ-77.64% / MFCC-GMM-86.27% Text Dependent : MFCC-VQ-85.49% / MFCC-GMM-94.12% | Accuracy can be improved in terms of text independent speech recognition |
| [5] | TIMIT (English) | 6.94% EER | used Low dimensional Feature vectors |
| [1] | English(Voxceleb Dataset) | 3.48% EER | — |
| [6] | Dataset: LDC-IL | 96.2% for Speech Identification | Dataset comprising single words & phrases of adults. |
| [7] | Manually Collected Dataset of 30 speakers (10 F & 20 M) | Accuracy rate is 1% higher than traditional MFCC+GMM approach | Accuracy can be improved in terms of new approach |
| [8] | VoxCeleb2(6000 speakers' dataset) | Obtain 3.48% Equal Error Rate. | Determine if two given uncontrolled utterances originate from the same speaker or not. |
| [9] | Fisher (English, Arabian, Chinese) 4000 Speakers & 343 Hours speech signal | 76.9% accuracy rate for individual voice segments, and 99.5% for each | Doubling the dataset can lead to accuracy improvement for the |

1

International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.5, January 2024

| | | | |
|---|---|---|---|
| | | speaker as a bundle. | BiLSTM model. |
| [10] | sepedi Home Language | Accuracy: MLP: 97% RF: 99.9% | It was observed that MLPs performed well on the given dataset, however, Auto-WEKA selected Random Forest as the best algorithm |
| [11] | Manually Collected Database in noisy Environment | 82% accuracy using MFCC model | In this paper, an automatic speech-speaker recognition system is implemented in a real time noisy environment. |
| [12] | THUYG-20 | EER is 4.01% | Speaker identification for short utterances using English manual and THYUG dataset. |
| [13] | Dataset: Arabic | 98.38% recognition rate | Dataset were recorded in office environment and only 5 fixed sentences are considered |
| [14] | 2 different corpora of British English (adult, children) | 17% and 31% relatively improvement over baseline | In this, the effectiveness of prosody-modification technique based on fuzzy classification of spectral bins is studied. |
| [15] | Dataset: Studio Recordings & Dialogs of Indian Language | Overall, 96.2% accuracy | ERIL is a multilingual emotion classifier, it is independent of any language |
| [16] | Manually Collected Audio-Visual dataset: 154 identities, 3 language annotations | Performance Degradation | Audio-Visual Speaker Recognition |
| [17] | Kaggle & Urdu Corpus (Regional Language of Pakistan) | vectors method demonstrates 80.4% FSR accuracy. With AC, it achieves 85.4% accuracy. With LI, its accuracy is 90.2%. Whereas by combining AC and LI it obtains 95.1% accuracy. | This new method is based on extracting accent and language information from short utterances. |
| [18] | Indian Languages | Accuracy Rate 98.34% | Identify not only frequency of voice but also textual features to improve accuracy I.e., a person can be angry with a slow voice. |
| [19] | English: Voxceleb1 (153516 audio files extracted from 1251 speakers) | 95% accuracy (Top - I accuracy rate for voxceleb1.) | —— |

## 3. ANALYSIS OF RELATED WORK IN GUJARATI LANGUAGE

A Systematic survey has been done for a speaker recognition system developed for Gujarati language. the summary of this survey discussed here:

A model has been purposed for automatic speaker identification in Gujarati. Model includes two major processes: voice verification of speakers and identification. At registration phase voice model generated and stored on smart card that will be further used for verification. There also listing of some verification errors which include stress, time varying, aging, sickness etc. model presents phases that include feature selection and measures and pattern matching. Two types of models are included stochastic and template. The proposed model is implemented using MARF (Modular Audio Recognition Framework) which includes a collection of algorithms of sound, speech and natural language processing[20].

An algorithm that worked in different 13 languages like Gujarati, Assam, Punjabi, etc. includes the speaker recognition model like HMM and CNN and has an accuracy of 95.21% in both the environment text-dependent & text-independent but found language mismatch more pronounced in speaker verification.

The Speaker-independent systems that accept telephony commands in Gujarati that used speaker independent 29 words for the experiment. Implementation is done using HMM based speech recognizer Sphinx4 toolkit. In the whole experiment 20 speakers are involved, 10 female and 10 males with the age criteria between 20 to 30 years. Total 29 words were used for testing as a command. During the experiment, Average accuracy for female speakers was 83.79%. Average accuracy for male speakers found 80%. The minimum accuracy of the system is 72.41% and after taking extra care the system can achieve highest accuracy up to 96.55%. Average accuracy rate of all the words found 83.62%[21].

A multilingual model including Gujarati SPEAKERSTEW for identifying a speaker. The model capable of working with different 46 languages at the same time. That model work for text-independent speaker recognition systems with 73% accuracy[22].

The impact of online speaker adaptation on the performance of a speaker independent, continuous speech recognition system for Hindi language also been listed as a part of speaker recognition system. The speaker recognition is executed using the Maximum Likelihood Linear Regression (MLLR) transformation approach. The MLLR transform based speaker adaptation technique is found to significantly improve the accuracy of the Hindi ASR system by 3%. After the experiment they have concluded that MLLR transform based speaker adaptation of Hindi speech models indeed decreases the recognition error by a factor of 0.19.

A multilingual model for speaker recognition that includes all the Indian languages they used the MFCC method for feature extraction and built their own model for identifying the speaker. Modal also worked for emotions recognition for a person like a person can be angry while having a slow voice tone. The model provides 98.34% accuracy [18].

Based on the analysis, it can be inferred that numerous models have been created to identify various speech parameters, such as emotions, voice, and pitch, across different languages. However, when considering models specifically designed for recognizing speakers in vernacular Gujarati dialects, a notable deficiency in accuracy becomes apparent.

Despite significant efforts in speaker recognition for the Gujarati language, there remains an unexplored territory regarding accuracy in specific recognition across different environments and accents. It is emphasized that each region has its unique speech accents, which demand focused attention for achieving precise results. Furthermore, the challenges posed by the limitations and complexities of the Gujarati Framework underscore the extensive research opportunities in implementing speaker recognition systems.

## 4. CONCLUSION & FUTURE ENHANCEMENT

This paper presents the research activities done in the area of Gujarati voice Recognition using different platforms and experimenting the same on various models along with different sample sizes. The activity of Voice recognition involves taking sound in the form of input and providing text, which exactly matches the sound. Speaker recognition process deals with variability in an individual's speech, range, pitch, accent, style of speaking etc. The previous model might provide a higher accuracy rate for multi languages but the model specifically worked for Gujarati language has not had that much of accuracy rate. There is therefore an opportunity to design a Voice recognition system for Gujarati.

## 5. ACKNOWLEDGMENTS

I would like to thank my guide Dr. Hiren R. Kavathiya and all my colleges that helped me in my work.

## 6. REFERENCES

[1] Xu, J., Wang, X., Xu, S., & Liu, W. (2020). Deep multi-metric learning for text-independent speaker verification. *Neurocomputing*, *410*,394–400.https://doi.org/10.1016/j.neucom.2020.06.045

[2] Devi, K. J., Singh, N. H., & Thongam, K. (2020). Automatic Speaker Recognition from Speech Signals Using Self Organizing Feature Map and Hybrid Neural Network. *Microprocessors and Microsystems*, *79*, 103264. https://doi.org/10.1016/j.micpro.2020.103264

[3] Bian, T., Chen, F., & Xu, L. (2019). Self-attention-based speaker recognition using Cluster-Range Loss. *Neurocomputing*, *368*, 59–68. https://doi.org/10.1016/j.neucom.2019.08.046

[4] Maurya, A., Kumar, D. P., & Agarwal, R. (2018). Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach. *Procedia Computer Science*, *125*, 880–887. https://doi.org/10.1016/j.procs.2017.12.112

[5] Kinnunen, T., Karpov, E., & Fränti, P. (2006). Real-time speaker identification and verification. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*(1), 277–288. https://doi.org/10.1109/tsa.2005.853206

[6] Gupta M, Singh RK, Singh S. G-Cocktail: An Algorithm to Address Cocktail Party Problem of Gujarati Language using CatBoost. Research Square; 2021. DOI: 10.21203/rs.3.rs-305722/v1.

[7] Patel, J. A., & Nandurbarkar, A. B. (2015). Development and Implementation of Algorithms for Speaker recognition for Gujarati Language. *International Research Journal of Engineering and Technology (IRJET)*.

[8] Xu, J., Wang, X., Xu, S., & Liu, W. (2020b). Deep multi-metric learning for text-independent speaker verification. *Neurocomputing*, *410*, 394–400. https://doi.org/10.1016/j.neucom.2020.06.045

[9] Hanifa, R. M., Isa, K., & Mohamad, S. (2021). A review on speaker recognition: Technology and challenges. *Computers & Electrical Engineering*, *90*, 107005. https://doi.org/10.1016/j.compeleceng.2021.107005

[10] Mokgonyane, T. B., Sefara, T. J., Modipa, T. I., Mogale, M. M., Manamela, M. J., & Manamela, P. J. (2019). Automatic Speaker Recognition System based on Machine Learning Algorithms. *2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA)*. https://doi.org/10.1109/robomech.2019.8704837

[11] Kakade, M. N., & Salunke, D. B. (2020). An Automatic Real Time Speech-Speaker Recognition System: A Real Time Approach. *Lecture Notes in Electrical Engineering*, 151–158. https://doi.org/10.1007/978-981-13-8715-9_19

[12] Tiwari, V., Hashmi, M. S., Keskar, A. G., & Shivaprakash, N. C. (2019). Speaker identification using multi-modal I-vector approach for varying length speech in voice interactive systems. *Cognitive Systems Research*, *57*, 66–77. https://doi.org/10.1016/j.cogsys.2018.09.028

[13] Ghoniem, R. M., & Shaalan, K. (2017). A Novel Arabic Text-independent Speaker Verification System based on

Fuzzy Hidden Markov Model. *Procedia Computer Science*, *117*, 274–286. https://doi.org/10.1016/j.procs.2017.10.119

[14] Shahnawazuddin, S., Adiga, N., Sai, B. T., Ahmad, W., & Kathania, H. K. (2019). Developing speaker independent ASR system using limited data through prosody modification based on fuzzy classification of spectral bins. *Digital Signal Processing*, *93*, 34–42. https://doi.org/10.1016/j.dsp.2019.06.015

[15] Mehra, P., & Jain, P. (2021). ERIL: An Algorithm for Emotion Recognition from Indian Languages Using Machine Learning. *Wireless Personal Communications*. https://doi.org/10.21203/rs.3.rs-449758/v1

[16] Nawaz, S., Saeed, M., Morerio, P., Mahmood, A., Gallo, I., Yousaf, M. H., & Del Bue, A. (2021). Cross-modal Speaker Verification and Recognition: A Multilingual Perspective. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvprw53098.2021.00184

[17] Saleem, S., Subhan, F., Naseer, N., Bais, A., & Imtiaz, A. (2020). Forensic speaker recognition: A new method based on extracting accent and language information from short utterances. *Forensic Science International: Digital Investigation*, *34*, 300982. https://doi.org/10.1016/j.fsidi.2020.300982

[18] Mehra, P., & Verma, S. B. (2022). BERIS: An mBERT-based Emotion Recognition Algorithm from Indian Speech. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *21*(5), 1–19. https://doi.org/10.1145/3517195

[19] Farsiani, S., Izadkhah, H., & Lotfi, S. (2022). An optimum end-to-end text-independent speaker identification system using convolutional neural networks. *Computers & Electrical Engineering*, *100*, 107882. https://doi.org/10.1016/j.compeleceng.2022.107882

[20] Patel H., Virparia P., - "Generic Model for Text Dependent Automatic Gujarati Speaker Recognition", International Journal of Emerging Trends & Technology in Computer Science, Vol. 1, Issue 3, September – October 2012

[21] Patel J., Patel P., and Virparia P., - "Voice Enabled Telephony Commands using Gujarati Speech Recognition", International Journal of Advanced Research in Computer Science and Software Engineering", Vol. 3, Issue 10, October 2013

[22] Chojnacka, R., Pelecanos, J., Wang, Q., Moreno, I.L. (2021) SpeakerStew: Scaling to Many Languages with a Triaged Multilingual Text-Dependent and Text-Independent Speaker Verification System. Proc. Interspeech 2021, 1064-1068, doi:10.21437/Interspeech.2021-646

[23] Purnima P., Bhatt S, - "Automatic Speech Recognition of Gujarati Digits using Dynamic Time Warping", International Journal of Engineering and Innovative Technology, Vol. 3, Issue 12, June 2014

[24] Rania M. Ghoniem, Khaled Shaalan, (2017), A Novel Arabic Text-independent Speaker Verification System based on Fuzzy Hidden Markov Model, Procedia Computer Science,Volume 117.

[25] Kharibam Jilenkumari Devi, Nangbam Herojit Singh, Khelchandra Thongam,(2017),Automatic Speaker Recognition from Speech Signals Using Self Organizing Feature Map and Hybrid Neural Network,Microprocessors and Microsystems,Volume 79,(2020).

[26] Tengyue Bian, Fangzhou Chen, Li Xu, (2019), Self-attention based speaker recognition using Cluster-Range Loss,Neurocomputing,Volume 368.

[27] Ankur Maurya, Divya Kumar, R.K. Agarwal, (2018), Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach,Procedia Computer Science,Volume 125.

[28] Shabnam Farsiani, Habib Izadkhah, Shahriar Lotfi, (2022), An optimum end-to-end text-independent speaker identification system using convolutional neural network,Computers and Electrical Engineering,Volume 100

[29] M. M. Kabir, M. F. Mridha, J. Shin, I. Jahan and A. Q. Ohi, "A Survey of Speaker Recognition: Fundamental Theories, Recognition Methods and Opportunities," in IEEE Access, vol. 9, 2021.

[30] Mohammad K. Nammous, Khalid Saeed, Paweł Kobojek,Using a small amount of text-independent speech data for a BiLSTM large-scale speaker identification approach, Journal of King Saud University - Computer and Information Sciences,Volume 34, Issue 3,2022.

[31] Sajid Saleem, Fazli Subhan, Noman Naseer, Abdul Bais, Ammara Imtiaz,Forensic speaker recognition: A new method based on extracting accent and language information from short utterances,Forensic Science International: Digital Investigation,Volume 34,2020.

[32] Monika Gupta, R K Singh, Sachin Singh et al. G-Cocktail: An Algorithm to Address Cocktail Party Problem of Gujarati Language using CatBoost, 17 March 2021.

[33] B. Pandey, A. Ranjan, R. Kumar and A. Shukla, "Multilingual speaker recognition using ANFIS," 2010 2nd International Conference on Signal Processing Systems, 2010.

[34] T. Kinnunen, E. Karpov and P. Franti, "Real-time speaker identification and verification," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 14, no. 1, pp. 277-288, Jan. 2006.

# Appendix-B Certificate of Plagiarism

DrillBit

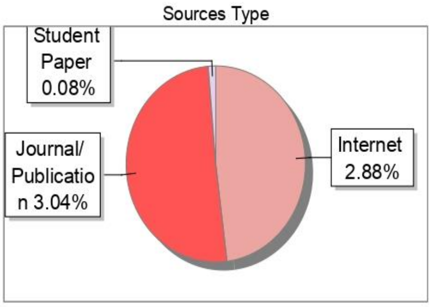
The Report is Generated by DrillBit Plagiarism Detection Software

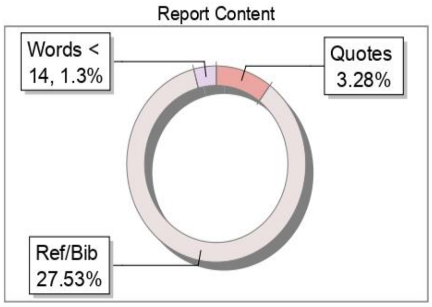
***Submission Information***

| | |
|---|---|
| Author Name | Meera Shah |
| Title | DEVELOPMENT OF A MODEL TO ANALYZE AND INTERPRET VERNACULAR VOICE RECOGNITION FOR GUJARATI DIALECTS |
| Paper/Submission ID | 2822846 |
| Submitted by | director.rit@atmiyauni.ac.in |
| Submission Date | 2024-12-20 16:07:48 |
| Total Pages, Total Words | 173, 46269 |
| Document type | Thesis |

***Result Information***

Similarity **6 %**




***Exclude Information***

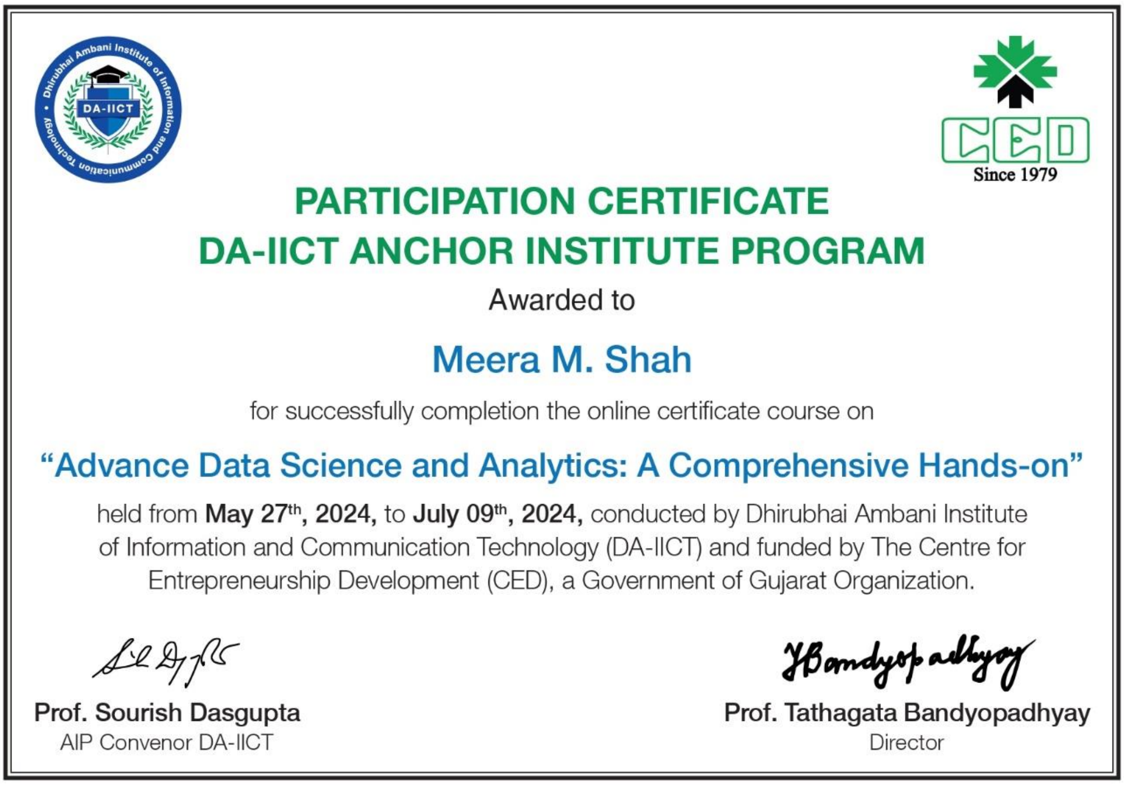
| | |
|---|---|
| Quotes | Not Excluded |
| References/Bibliography | Not Excluded |
| Source: Excluded < 14 Words | Not Excluded |
| Excluded Source | **0 %** |
| Excluded Phrases | Not Excluded |

***Database Selection***

| | |
|---|---|
| Language | English |
| Student Papers | Yes |
| Journals & publishers | Yes |
| Internet or Web | Yes |
| Institution Repository | Yes |

A Unique QR Code use to View/Download/Share Pdf File

# Other Certifications


Since 1979

PARTICIPATION CERTIFICATE
DA-IICT ANCHOR INSTITUTE PROGRAM

Awarded to

Meera M. Shah

for successfully completion the online certificate course on

"Advance Data Science and Analytics: A Comprehensive Hands-on"

held from **May 27th, 2024,** to **July 09th, 2024,** conducted by Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT) and funded by The Centre for Entrepreneurship Development (CED), a Government of Gujarat Organization.

Prof. Sourish Dasgupta
AIP Convenor DA-IICT

Prof. Tathagata Bandyopadhyay
Director


Stanford | ONLINE

COURSE CERTIFICATE

May 9, 2022

Meera Manishbhai Shah

has successfully completed

Machine Learning

an online non-credit course authorized by Stanford University and offered through Coursera

Associate Professor Andrew Ng
Computer Science Department
Stanford University

SOME ONLINE COURSES MAY DRAW ON MATERIAL FROM COURSES TAUGHT ON-CAMPUS BUT THEY ARE NOT EQUIVALENT TO ON-CAMPUS COURSES. THIS STATEMENT DOES NOT AFFIRM THAT THIS PARTICIPANT WAS ENROLLED AS A STUDENT AT STANFORD UNIVERSITY IN ANY WAY. IT DOES NOT CONFER A STANFORD UNIVERSITY GRADE, COURSE CREDIT OR DEGREE, AND IT DOES NOT VERIFY THE IDENTITY OF THE PARTICIPANT.

Verify at: https://coursera.org/verify/8KRBUALA7XTK
Coursera has confirmed the identity of this individual and their participation in the course.


Researcher Academy
researcheracademy.com
Certificate of Excellence
ELSEVIER
This certifies that
Meera Shah
has successfully completed every module within
Ethics
on Friday 07 October, 2022
Suzanne BeDell
Managing Director, Education
Reference & Continuity Books
Laura Hassink
Managing Director, Science,
Technology & Medical Journals


Springer
International Conference on Deep Learning and Visual Artificial Intelligence
(ICDLAI-2024)
Certificate of Presentation
Presented to
Meera M. Shah
of Atmiya University, Rajkot, Gujarat, India
For Presenting the Paper
"Advancements in Speaker Recognition: A Comprehensive Review of Techniques, Challenges, and Future Directions"
At the International Conference on Deep Learning and Visual Artificial Intelligence (ICDLAI-2024) on 16-17 March 2024 in Hybrid Mode. Organized by Bikaner Technical University, Bikaner in Association with Springer and Hari Krishan Research and Development Nexus.
Dr. aDiti Sharma
ConferenCe Chair
Dr. ViShal gaur
general Chair
Dr. SuDhir BharDwaj
Program Chair

Certificate

ATMIYA UNIVERSITY

Sponsored By:

This is to certify that Prof./Dr./Mr./Ms. Meera Manishbhai Shah

of Atmiya University has presented a paper titled

Exploring Speech Corpus for Voice Recognition in Gujarati: An In-depth Study during the

**2nd International Conference on**
**Emerging Trends & Contemporary Practices – ICETCP 2024"**

hosted by **Atmiya University, INDIA** and sponsored by **AICTE** & **CSIR**, on **February 9th-10th, 2024.**

**Dr. Divyang D. Vyas**
Registrar, Atmiya University

**Dr. Shiv Tripathi**
Vice-Chancellor, Atmiya University

**Partners & Supporting Organizations:**

Member:

ATMIYA UNIVERSITY

This

**CERTIFICATE**

is awarded to

*"Ms. Meera M. Shah, Atmiya University"*

for attending

**'International Conference on Emerging Trends & Contemporary Practices 2022',** organized by Atmiya University, India on 19th & 20th May, 2022, in association with University of Miyazaki (Japan), Keck Graduate Institute (USA), Lincoln University College (Malaysia) & Hemchandracharya North Gujarat University (India)

Dr. Divyang D. Vyas
Registrar, ATMIYA UNIVERSITY, India

Dr. Shiv Tripathi
Vice Chancellor, ATMIYA UNIVERSITY, India

Member University

Global Partners

Sri Balaji Society's

**Sri Balaji University, Pune (SBUP)**

Private University (Sri Balaji University, Pune Act 2019 created under Maharashtra Act No. XII of 2019)

Founder President & Chancellor - Late Prof. Dr. (Col) A. Balasubramanian

• Discipline • Dedication • Determination

AIMS ASSOCIATION OF INDIAN MANAGEMENT SCHOOLS

**MIMANSA'23**

2nd International Research Conference

## Certificate of Appreciation

This certificate is awarded to

PROF./DR./MR./MS. Meera Shah of Atmiya University who presented the paper titled A Systematic Survey on Voice Regnition for Gujrati Dialects at the **International Research Conference** on

**'INNOVATIVE PRACTICES FOR SUSTAINABLE BUSINESS'**

organised by Sri Balaji University, Pune, India.

during 17th & 18th March 2023.

| PROF. PARAMANANDHAN BALASUBRAMANIAN | PROF. (DR.) G.K. SHIRUDE | DR. SWETA BANERJEE | DR. LEENA DAM |
|---|---|---|---|
| PRO-CHANCELLOR, SRI BALAJI UNIVERSITY, PUNE (SBUP) PRESIDENT, SRI BALAJI SOCIETY | VICE CHANCELLOR, SRI BALAJI UNIVERSITY, PUNE (SBUP) | PROFESSOR, BITM CONVENER | PROFESSOR, BIMHRD CO-CONVENER |

G20 भारत 2023 INDIA

**NATIONAL CONFERENCE**

on

**"Recent Trends on Applied Statistics & Data Science"**

**In the Memory of 100th Birthday of Prof. M. N. Das**

Azadi Ka Amrit Mahotsav

Organized By

P.C. Mahalanobis Department of Statistics

SAURASHTRA UNIVERSITY, RAJKOT

## CERTIFICATE

This is to certify that Dr./Mr./Ms. **SHAH MEERA MANISHBHAI** of **ATMIYA UNIVERSITY** has participated in the National Conference held at P. C. Mahalanobis Department of Statistics, Saurashtra University, Rajkot - 360005, Gujarat on **28-29th January, 2023** and presented paper entitled **A REVIEW ON SPEAKER RECOGNITION FOR GUJARATI LANGUAGE.**

| Prof. G. C. Bhimani | Prof. D. K. Ghosh | Prof. K. H. Atkotiya | Dr. D. K. Rank | Dr. M. V. Savsani |
|---|---|---|---|---|
| Chief Petron | Organizing Chair | Convener | Organizing Secretaries | Organizing Secretaries |

IEEE Signal Processing Society GUJARAT CHAPTER | IEEE Gujarat Section | IEEE Signal Processing Society MEFGI Student Branch | Marwadi University Marwadi Chandarana Group | NAAC A+

## CERTIFICATE OF PARTICIPATION

This is to certify that

Mr/Ms Meera Shah

has successfully participated in the event

**NEW TRENDS IN SIGNAL PROCESSING (NetSiP 2024)**

organized by IEEE Signal Processing Society, Gujarat Chapter
held on 25-26 July, 2024.

We appreciate your valuable contribution and active participation in making this event a great success.

*In Association with*

IEEE Region 10 | IEEE INDIA COUNCIL

| | | |
|---|---|---|
| **Dr. MITA PAUNWALA**<br>Chair,<br>*IEEE SPS GS* | **Dr. MANISH KHARE**<br>**Dr. CHINTAN VARNAGAR**<br>Program Chair, NetSiP 2024<br>*IEEE SPS GS* | **Prof. CHAND**...<br>**Dr. SUNIL** ...<br>Organizing Chair, ...<br>*Marwadi Univer...* |

# CERTIFICATE

## OF PARTICIPATION

This is to certify that

**Meera Shah**

Assistant Professor

has participated in the International Lecture titled **"Research Problem Identification & Literature Review "** organized by Department of Research & Publications, A2Z EduLearningHub LLP on 15-05-2023.

A2Z EDULEARNINGHUB LLP PIN:685 595 UDUMBANNOOR

Director
A2Z EduLearningHub LLP, India

www.edulearninghub.com

Venue: Online
WXCD34A44

Researcher Academy On Campus

Certificate of Attendance

ELSEVIER

This certifies that

Meera Shah

has attended the following

APAC customer success hub webinar, Beyond Borders: How to promote interdisciplinary & international collaborative research to contribute to the achievement of the SDGs.

at APAC customer success hub webinar,

on Wednesday 01 February, 2023

Presented by JUNYA I Customer Consultant

Suzanne BeDell
Managing Director, Education Reference & Continuity Books

Laura Hassink
Managing Director, Science, Technology & Medical Journals

Five Days International workshop
on
Application of Machine Learning and Artificial Intelligence in Research and Innovations

REGIONAL COLLEGE
a unit of deepshikha group of colleges

(IWMLAI-2023)

Certificate of Participation

Ref. No. RCERT/IWMLAI/1768

This is certify that Miss. Meera Shah of Atmiya University has participated in the five days International workshop on "Application of Machine Learning and Artificial Intelligence in Research and Innovations " held from 2nd - 6th May 2023 at Regional College for Education Research & Technology, Jaipur.

Mr. Madhumay Sen
Coordinator, IWMLAI

Dr. Anshu Surana
Vice Chairman, RCERT

Dr. Pramod Sharma
Principal,RCERT

6TH INTERNATIONAL CONFERENCE ON
ARTIFICIAL INTELLIGENCE & SPEECH TECHNOLOGY
AIST-2024

CERTIFICATE OF APPRECIATION

This certificate is awarded to

Meera M. Shah from ATMIYA University, Rajkot

for presenting a Research Paper titled
Optimizing Speaker Recognition through Feature Extraction Techniques: A Focus on Gujarati Dialects

in the 6th AIST-2024 jointly organized by Centre of Excellence - AI and Department of IT, Indira Gandhi Delhi Technical University for Women (IGDTUW), Delhi held on 13th- 14th November, 2024.

Financial Sponsor
Ministry of Science and Technology
Anusandhan National Research Foundation

Publication Partner
Springer CCIS

Prof. Arun Sharma
Organizing Chair - AIST 2024
Dean Academic Affairs, IGDTUW

Prof. Ranjana Jha,
General Chair, AIST-2024
Vice-Chancellor, IGDTUW