Educational Administration: Theory and Practice
2024, 30(6)(s), 168-175
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), 168-175
Doi: 10.53555/kuey.v30i6(S).5343

ARTICLEINO

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

I. INTRODUCTION

This template 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.

II. 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.

III. 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.

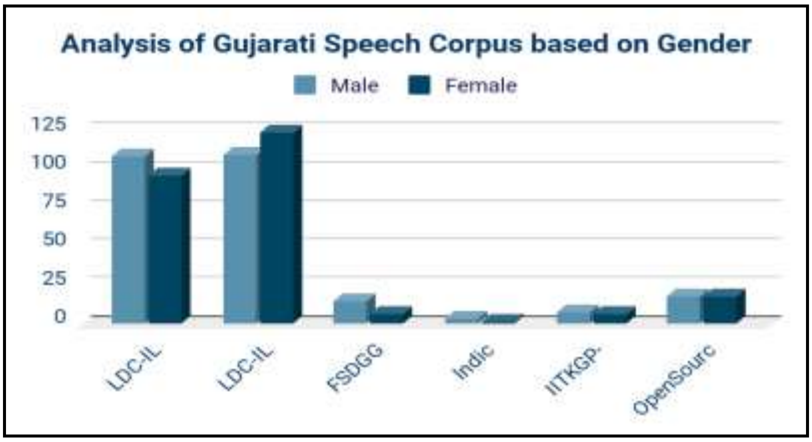

Figure 1. Analysis of Gujarati Speech Corpus based on Gender

IV. 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±0.056, while for female speakers, it was 4.269±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.

V. 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.

VI. 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.