

Indian Speech Sentiment Recognition: A Review

Anjum Mansuri

GLS University, Ahmedabad, India

Dr. Krupa Mehta

Faculty of Computer Technology, GLS University, Ahmedabad, India

Abstract: Sentiment recognition is the addressed communication that has been gaining more interest from the public. Sentiment analysis using speech is the most popular approach on which people are working nowadays. In this review paper, we pinpoint Sentiment Analysis using speech recognition. Work in speech sentiment analysis using the English language in recent years is quite high. But for Indian regional languages Speech Sentiment recognition is still an emerging approach and progress is too slow. Gujarati is mostly used for speaking and writing purposes in Gujarat so Sentiment Analysis of Gujarati speech is a very valuable task. As we have seen, a lot of research has already been done in English and some Indian languages, but in this area, very little work is found in the Gujarati language compared to the English language. The main purpose behind this literature work is to construct such a reliable approach which is used mainly for Customer Care Call center application.

Keywords: Speech sentiment recognition, Gujarati speech, Indian Language

I. INTRODUCTION

Gujarati is affluent in morphology and has a complex structure of syntax, diversity of dialects in the form of pronunciation and grammar for most of the languages in India, the language takes on a different form every 50 kilometers. There are regional dialects that differ in some minor ways, slightly different from the original language. Some of them are listed below. In standard Gujarati, there are Saurashtra standards, Nagari, Bombay Gujarati, Patnuli. In Gramya Gujarati, Gramya, Surati, Anawla, Brathela, eastern broach Gujarati, Charotari, Patidari, Vadodari, Ahmedabad Gamadia, Patani Parsi. In kathiawadi category, there are jhalawadi, sorathi, holadi, gohilwadi, bhavnagari, and kakari tarimuki i.e. ghisadi [21]. This paper aims to review the best sentiment lexical resource for the Gujarati language, which can be effective for sentiment analysis of Gujaratispeech.

II. IMPORTANCE OF GUJARATI SPEECH SENTIMENT RECOGNITION

The Gujarati Language is the only language used in speaking and writing by most people in Gujarat. When the classification is an about Gujarati script, it becomes a complex task because Gujarati has complex morphology. Various tools and resources are developed for the English language and a few Indian languages. Gujarati is a resource-poor language for this task.

The motive of this paper is the importance of automatically recognizing sentiments from human speech. Research in understanding and modeling human sentiment, a topic that has been predominantly dealing with in psychology and linguistics, is increasingly attracting attention within the engineering community. A major inspiration comes from the desire to develop human-machine interfaces that are more adaptive and responsive to a user's behavior. The first investigation of speech sentiment recognition was conducted in 1972. This study sought general qualitative acoustic correlations of sentiment in a speech. In the mid 6 Study and Analysis of Speech Sentiment Recognition using Spectral, Prosodic and Hybrid Features 1980s, studies of sentiment recognition started to use the statistical properties of certain acoustic features. In recent years, sentiment recognition from speech has conspicuous applications in the speech processing systems, such as spoken tutoring systems, medical emergency domain to detect stress and pain, interactions with robots, computer games, call centers, etc. Extracting the sentiment from the short utterances of interactive voice

response (IVR) systems is another typical application. Also, sentiment recognition is the first step toward the implementation of sentimental speech recognition systems. Generally, a sentiment recognition system from speech has three main components: feature extraction unit, feature selection unit, and sentiment classification unit.[24] Some statistical properties of acoustic features have been used in the 1980s in feature extraction. In the 1990s, knottier features were employed. However, in recent years' research has been focused on finding reliable, informative features and combining influential classifiers that improve the performance of sentiment detection systems in real-life applications.

The Fig1: *Speech Emotion Recognition* depicts a simplified system utilized for speech-based emotion recognition[24] and includes the main components.

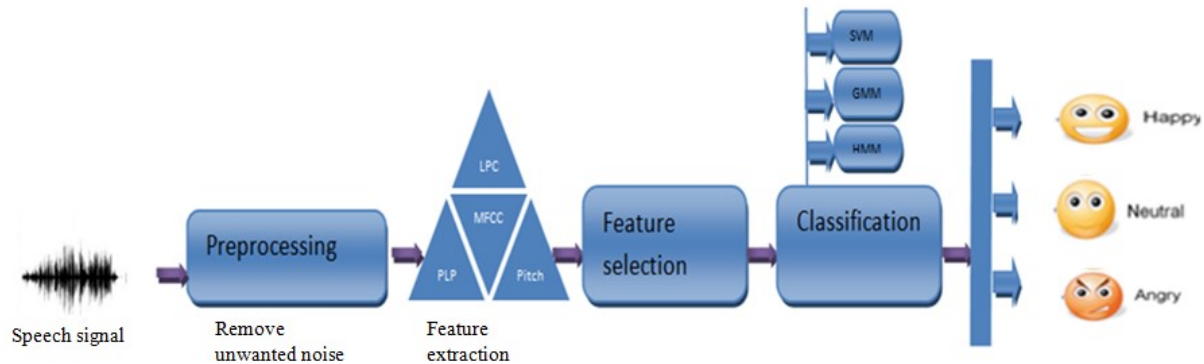


Figure 1: Speech Emotion Recognition

III. OBJECTIVES AND MOTIVATION

The main objective of this study is to recognize speech sentiments from the Gujarati language with high accuracy. The recognition accuracy depends on the selection of efficient feature extraction techniques and machine learning algorithms. Hence we planned to concentrate on pre-processing techniques, feature extraction techniques, and choosing an optimal pattern classifier for classifying the sentimental features. Since there is no previous work reported for Gujarati Speech Emotion Recognition using customer care call center corpus it is very much important to create a sentimentally well-balanced database for conducting meaningful experiments. Enormous advances have been made in speech sentiment recognition research. Many research results are reported for languages like Hindi, Telugu, Marathi, Kannada, Bangla, Assamese, Urdu, and Punjabi. In Indian languages, studies are active in Assamese, Bengali, Malayalam, and Kannada. However, in Gujarati, not many works were reported. So in-depth research works are highly essential in the Gujarati language towards speech sentiment recognition

IV. EMOTION SPEECH DATABASES

Primarily, three different classifications of databases have been studied in the literature for speech emotion recognition. These are acted, naturalistic, and induced.

A. Actor based speech database

Actor-based speech databases are most commonly used and are standardized databases. Using this kind of database, results can be compared effortlessly.

B. Natural speech database

A natural speech database is one that is completely naturally uttered. It is very useful in the case of real word emotion recognition. For example, call center conversation.

C. Induced speech database

Induced databases are near to natural databases. In this type of database, all emotions may or may not be existing. In

such types of databases, we create an artificial emotional situation so that we can speak in an emotional way.

For the speech emotion recognition system to produce optimum results, we need to have the most natural database which holds identical real-world emotions. There are standard databases available for German, Spanish, Korean, and English languages only. There is a need for the development of speech processing systems for Indian native and other low researched languages like Gujarati. The literature also suggests that the research for Indian languages is at an emerging stage and needs more attention from the researchers.

There are 22 official languages in India with over 720 dialects which can be useful in the development of an effective and versatile speech emotion recognition system with the ability to recognize emotions in varied speeches and languages. Attempts have been made to use Hindi and its various dialects for dialect recognition as well as emotion recognition. The database of Assamese, Dimasa, Bodo, Karbi, and Mishing all natives of Assam have also been created

Flow of the research paper: In section II, explain the Importance of Gujarati Speech Sentiment Recognition. Section III demonstrates the motivation and objective of this paper. Section IV represents Emotion Speech Databases Section V, we have created a table and listed content of paper and deep details about paper as well as Methods OR Techniques, Accuracy & Metrics. In section VI, we end up with Conclusion

V. LITERATURE REVIEW FOR INDIAN LANGUAGE

Here I have discussed some of the literature reviews on speech emotion recognition and the approaches they have focused on:

[1] According to this, in research work has designed an SER (Speech Emotion Recognition) model that depends on the GFCC algorithm to cite the feature sets based on the DCT and High pass Filter method. After that, the ALO algorithm is used to select the instances with the help of coverage and Fitness function. The novel MSVM algorithm is used to classify the emotion-based on the feature set and evaluate the performance metric such as accuracy rate etc. In proposed work using the MATLAB simulation tool and evaluating the maximum accuracy rate and mitigate the error rates as compared with the existing parameters.

[2] In this paper, speaker emotions are recognized using the data extracted from the speaker voice signal. Mel Frequency Cepstral Coefficient (MFCC) technique is used to recognize emotion of a speaker from their voice. The designed system was validated for Happy, sad and anger emotions and the efficiency was found to be about 80%.

[3] Researchers have focused more on classifiers. Instead feature extraction using time and frequency domain techniques, hybrid technique, will provide more accurate results. Researchers have used a single model of classifier like Radial Basis Function (RBF), Artificial Neural Network (ANN), Support vector machine (SVM), HMM. Exploring Hybrid models and investigating the performance of the same in emotion recognition system (ERS). Models derive IEEE 402228th ICCNT 2017 July 3-5, 2017, IIT Delhi, Delhi, India testimony with diverse perspective. Hybridization of testimony may improve system performance.

[4] In this research, a communication system with speech emotion recognition is proposed. The system pre-process speech with sound data enhancing methods in speech emotion recognition and transforms the sound into a spectrogram by MFCC (Mel Frequency Cepstral Coefficient). Then, GoogLeNet of CNN (Convolution Neural Network) is applied to recognize the five emotions, which are peace, happy, sad, angry and fear, and the top accuracy of recognition is 79.81%.

[5] This paper presents a study of distinguishing emotions by acoustic speech recognition (ASR) using K-means nearest neighbor (K-NN), a machine learning (ML) technique. The most significant paralinguistic information obtained from spectral features is presented by ASR i.e. by using Mel frequency cepstrum coefficient (MFCC). The emotions are classified using a K-NN algorithm over 2 separate models, based on the soft and high pitch voice. Model 1 and 2 achieved about 72.95% and 76.96% recognition rates respectively.

[6] In this study, researchers have analyzed the artificial emotional speech corpus by using Mel Frequency

Cepstral Coefficients (MFCCs). We have conducted the experiment on our own developed database of 100 utterances by the male actors and female actresses from Marathi movies. also successively trained and tested the data samples. The samples are recorded by using a high quality Software's at a sampling rate of 16 KHz. This paper elaborates the development of Artificial Marathi emotional speech which was developed using some Marathi Movies. It also describes recognition of Emotion from developed emotional speech databases by using MFCCs. The 12 cepstral coefficients have been extracted from all emotional speech signals and these coefficients are classified with the help of Fisher linear discriminant analysis.

[7] This paper focuses upon statistical techniques for reducing the dimensionality of data and then applying machine learning algorithms like KNN and NBC for identifying the emotion. After applying PCA & LDA it is required to apply KNN & NBC for predicting the emotional state of speech data sets then applied these techniques on test data sets, to get some satisfactory results.

[8] In this work the performance of Telugu ASR system is investigated in different emotive conditions. A significant degradation in the ASR system performance was observed in the presence of emotional speech. The performance was improved by adapting the emotional speech models to the existing ASR system. The importance of emotion recognition blocks at the front-end and adaptation of the ASR system towards different emotions was observed in this work.

[9] The paper gives the performance evaluation of emotion recognition for Tamil and Telugu languages using a linear, nonlinear and the combination of both these features. ANN and KNN classifiers were used for analysis. For all these cases considered, ANN gave better accuracy, precision and recall compared to KNN. Comparing linear and nonlinear features, the nonlinear feature i.e., Hurst parameter gave better performance measures for both the classifiers. For individual

[10] features, the highest accuracy, precision and recall were obtained while ANN classifier with Hurst feature was applied for Tamil language. In this case the values were 74.19%, 65.23% and 61.90% for accuracy, precision and recall respectively.

[11] In this paper, adopted a novel emotion recognition approach. To accomplish this, researchers have proposed an adaptive architecture for the artificial neural network to learn the multimodal fusion of speech features. Further, a hybrid PSO-FF algorithm was also accomplished to train the classifier. The performance of the proposed recognition model was analyzed by comparing it with the conventional methods. The developed approach has guaranteed the precise recognition of emotions from the speech signal. Moreover, it was observed that the average accuracy of the proposed feature model is 10.85%, which shows the efficiency of the proposed approach.

As we have discussed some literature reviews related to speech emotion recognition using Indian language. The table below contains 10 literature reviews on Indian languages.

Reference Paper	Authors	Year	Spoken Language	Sentiments	Methods OR Techniques	Category of Tokens	Accuracy & Metrics
[1]	Deepak Bharti	2020	French, Spanish, and English(US)	anger, sadness, contentment, amazement and fear.,	GFCC, MSVM, ALO, DCT	speech	79.48
[2]	Ming-Che Lee	2020	Chinese	Calm, Happy, Sad, Angry, Fearful, Disgust and Surprise	STFT, MFCC, CNN, deep-learning	speech	79.81%
[3]	Akshat Agrawal	2020	Hindi	happy, anger, surprise and sad alongside neutral speech	PCA, LDA, KNN, NBC	Speech	Satisfactory results.

[4]	Sujay G. Kakodkar	2018	Sanskrit	Happy, Sad, Anger, Afraid and Surprise.	Feature Extraction (MFCC), K-NN Classifier	speech	76%
[5]	Vishnu Vidyadhara Raju Vegesna	2018	Telugu	neutral speech	MFCC, GMM-HMM	speech	75.28%
[6]	Raviraj Vishwambhar Darekar	2018	Marathi	Happy, Sad, Anger, Surprise,neutral,fear	PSO-FF, FPR, FNR,NPV, FDR, F1Score and MCC	speech	10.85%
[7]	M.S. Likitha	2017	A data base has been generated with voices of 60 people with different emotions..	Happy,sad and anger	MFCC, FFT	speech	80%
[8]	Mr. Vaijanath. V. Yerigeri	2017	Marathi	stress emotion detection	(RBF), (ANN, SVM,HMM	speech	65%
[9]	S. Renjith	2017	Tamil and Telugu	anger, happiness and sadness	LPC, (LPCC, KNN, ANN	speech	Tamil ANN gave 75.27%, 65.07% and 63.01% of accuracy, precision and recall respectively. / Telugu ANN gave 76.41%, 64.00% and 64.49% of accuracy
[10]	Vishal B. Waghmare,	2014	Marathi	Happy, Sad and Angry	(MFCCs, LDA)	speech	12 sentiments out of 113.

Table. 1: speech sentiment recognition for Indian Languages

VI. CONCLUSION

Indo-Aryan languages are low resource languages due to lack of language resources like corpora set and standard language set. The complexity of SSR also includes digits and symbol representation. Some limitations involve accuracy, acceptance, and response in real-time.

This paper gives a basic overview of SER, the importance of the Gujarati language, and a literature review of speech sentiment recognition approaches. It includes the limitations of existing systems. It also states the problem statement with the objective and motivation to carry this research work.

REFERENCES

- [1] Bharti and P. Kukana, "A Hybrid Machine Learning Model for Emotion Recognition From Speech Signals," *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2020, pp. 491-496, doi:10.1109/ICOSEC49089.2020.9215376.
- [2] Lee, MC., Chiang, SY., Yeh, SC. et al. Study on emotion recognition and companion Chatbot using deep neural network. *Multimed Tools Appl* 79, 19629–19657 (2020). <https://doi.org/10.1007/s11042-020-08841-6>
- [3] Agrawal, A., & Jain, A. (2020). Speech emotion recognition of Hindi speech using statistical and machine learning techniques. *Journal of Interdisciplinary Mathematics*, 23(1), 311–319. doi:10.1080/09720502.2020.1721926
- [4] Sujay G. Kakodkar, Samarth Borkar (2018). "Speech Emotion Recognition of Sanskrit Language using Machine Learning" doi: <https://doi.org/10.1109/ijca.2018.917326>
- [5] Vegesna, V. V. R., Gurugubelli, K., & Vuppala, A. K. (2018). Application of Emotion Recognition and Modification for Emotional Telugu Speech Recognition. *Mobile Networks and Applications*. doi:10.1007/s11036-018-1052-9
- [6] B. Waghmare, V., R. Deshmukh, R., P. Shrishrimal, P., & B. Janvale, G. (2014). Development of Isolated Marathi Words Emotional Speech Database. *International Journal of Computer Applications*, 94(4), 19–22. <https://doi.org/10.5120/16331-5611>
- [7] M. S. Likitha, S. R. R. Gupta, K. Hasitha and A. U. Raju, "Speech based human emotion recognition using MFCC," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2017, pp. 2257-2260, doi: 10.1109/WiSPNET.2017.8300161.
- [8] V. V. Yerigeri and L. K. Ragha, "Marathi speech emotion detection: A retrospective analysis," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 2017, pp. 1-6, doi:10.1109/ICCCNT.2017.8203925.
- [9] Renjith, S., & Manju, K. G. (2017). Speech based emotion recognition in Tamil and Telugu using LPCC and Hurst parameters — A comparative study using KNN and ANN classifiers. 2017 International Conference on Circuit, Power and Computing Technologies (ICCPCT). doi:10.1109/iccpct.2017.8074220
- [10] Darekar, R. V., & Dhande, A. P. (2018). Emotion recognition from Marathi speech database using adaptive artificial neural network.
- [11] Biologically Inspired Cognitive Architectures, 23, 35–42. doi:10.1016/j.bica.2018.01.002
- [12] Pao T, Wang C. and Li Y., "A Study on the Search of the Most Discriminative Speech Features in the Speaker Dependent Speech Emotion Recognition." 2012 Fifth International Symposium on Parallel Architectures, Algorithms and Programming, Taipei, 2, no. 3, (2012): 157-162.
- [13] Tailor, Jinal. H. Shah, Dipti. B. Completed "The design and implementation of speech recognition algorithm for Gujarati language." in 2020, Department of Computer Science, Sardar Patel University, (2020)
- [14] B. Abraham, A. Davletcharova, S. Sugathan, and A. P. James, "Detection and analysis of emotion from speech signals," *International Symposium on Computer Vision and the Internet*, vol. 58, pp. 91-96, 2015.
- [15] S. Kakodkar and S. Borkar, "Acoustics Speech Processing of Sanskrit Language", *International Journal of Computer Applications*, vol. 180, pp. 27-32, 2018.
- [16] S. Foo, T. Nwe, and C. De Silva, "Speech emotion recognition using hidden Markov models", *Speech communication*, vol. 41, no. 4, pp. 603-623, 2003.
- [17] P. Y. Oudeyer, "The production and recognition of emotions in speech: features and algorithms", *International Journal of Human-Computer Studies*, vol. 59, pp. 157–183, 2003.
- [18] Patel, Dharmendra & Gohil, Lata. (2019). A Sentiment Analysis of Gujarati Text using Gujarati Sento Word Net. 8.2290-2292.
- Akçay, M. B., & Oğuz, K. (2019). Speech Emotion Recognition: Emotional Models, Databases, Features, Preprocessing Methods, Supporting Modalities, and Classifiers. *Speech Communication*, 2682. doi:10.1016/j.specom.2019.12.001
- [19] Z. Li, "A study on emotional feature analysis and recognition in speech signal," *Journal of China Institute of Communications*, vol. 21, no. 10, pp. 18–24, 2000. View at: [Google Scholar](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=ZLi1973)
- [20] G. Pengjuan and J. Dongmei, "Research on emotional speech recognition based on pitch," *Application Research of Computers*, vol. 24, no. 10, pp. 101–103, 2007. View at: [Google Scholar](https://scholar.google.com/citations?view_op=view_citation&hl=en&user=GPengjuan)
- [21] Cantero, Jessica, et al. "Gujarati Language Manual" *gujarati_language_manual*, 2010, [language manuals. weebly.com/uploads/4/8/5/3/4853169/gujarati_language_manual.pdf](https://www.weebly.com/uploads/4/8/5/3/4853169/gujarati_language_manual.pdf).
- [22] Mehta, B., Professor, B. R. A., & Assistant Professor. (n.d.). Gujarati poetry classification based on emotions using deep learning. Retrieved August 10, 2021, from Ijeast.com website: <https://www.ijeast.com/papers/358-362.Tesma601.IJEAST.pdf>
- [23] T. M. Wani, T. S. Gunawan, S. A. A. Qadri, M. Kartiwi and E. Ambikairajah, "A Comprehensive Review of Speech Emotion Recognition Systems," in *IEEE Access*, vol. 9, pp. 47795-47814, 2021, doi: 10.1109/ACCESS.2021.3068045.
- [24] Khalil, R. A., Jones, E., Babar, M. I., Jan, T., Zafar, M. H., & Alhussain, T. (2019). Speech emotion recognition using deep learning techniques: A review. *IEEE Access: Practical Innovations, Open Solutions*, 7, 117327–117345.