

Emotion-Based Speech Analysis for Disaster Response and Crisis Management

Rafa Siddiqua ID:23166023

Rabea Akhter ID: 23366029

Samia Mostafa Ishan ID: 23273011

Introduction: Speech Emotion Recognition (SER) stands as a cornerstone in the landscape of Human-Computer Interaction (HCI) and soft computing, playing a pivotal role in deciphering emotions conveyed through speech. Its applications span across diverse fields, encompassing call centers, robotics, healthcare, and the expansive realm of digital marketing. SER's significance lies in its ability to decode emotions, fostering more intuitive and natural interactions between humans and technology. In practical applications, SER's implications are far-reaching. For instance, within call centers, understanding emotional cues during conversations is crucial for delivering effective customer service. In robotics, recognizing human emotions enables more empathetic interactions between machines and individuals. Additionally, within healthcare, SER holds promise in mood assessment, patient care, and even early diagnosis of certain conditions. However, despite its broad applicability, SER faces challenges, particularly in achieving high accuracy when discerning emotions within natural speech. Deciphering nuanced emotional cues embedded in speech patterns, intonations, and cadences remains a complex task. Addressing these challenges, ongoing research in SER endeavors to develop systems capable of not only recognizing diverse emotions but also operating independently of the speaker and language while navigating noisy environments effectively. This review paper aims to comprehensively explore the multifaceted domain of SER, examining its foundational components and methodologies. It delves into the fundamentals, exploring emotional speech databases and elucidating the myriad of acoustic features and their combinations used in SER research. Moreover, this review critically evaluates advancements and limitations in both speaker-independent and language-independent SER systems. By dissecting methodologies, trends, and challenges encountered in the field, this paper aims to provide a comparative perspective on the current state of SER research. Structured across six distinct sections, each intricately addressing various facets of SER, this review endeavors to serve as a seminal resource, distilling comprehensive research, and contributing to a deeper understanding of this dynamic field.

Literature Review: In [1], the authors Sung-Woo Byun and Seok-Pil Lee constructed a Korean emotional speech database for speech emotion analysis. In [2], the authors proposed an approach based on emotional perception, which designs an implicit emotional attribute classification. The authors of [3] used two unidirectional LSTM layers for text recognition and fully connected layers are used for acoustic

emotion recognition, which are then merged to produce the predicted emotion categories. The authors of [4] used a technique in their project, where it classifies emotions into five different categories. Different approaches for developing speech recognition, which are language and speaker independent, are briefly discussed by the author of [5].

Proposed Methodology:

Emotion-based speech analysis for disaster response and crisis management typically involves a combination of algorithms and techniques from natural language processing (NLP), machine learning, and signal processing.

First of all, one needs to convert their speech to text for better understanding of different situations. To convert speech to text one can use deep speech technique or Google speech to text technique. Also, for emotion recognition one can use supervised machine learning model where different algorithms like support vector, gradient boosting or random forest will be used to classify in order to detect the emotions in any type of speech all these models also help one to train datasets. Along with that, the author also proposed to use different NLP techniques such as lexicon based techniques in order to analyze all different sentiment from any text to figure out whether it is positive or negative. In this research, one is also encouraged to use different speech signal processing formats like Gaussian Mixture where one can understand different features of speech like intensity or speech, which will help one to analyze the depth of any situation and try to provide help as early as possible.

For future, researchers also proposed to combine multiple classifier voting and bagging classifiers to improve the emotion classification and their accuracy.

Reference:

- [1] Byun, S.-W.; Lee, S.-P. A Study on a Speech Emotion Recognition System with Effective Acoustic Features Using Deep Learning Algorithms. *Appl. Sci.* 2021, *11*, 1890. <https://doi.org/10.3390/app11041890>
- [2] Liu, G., Cai, S. & Wang, C. Speech emotion recognition based on emotion perception. *J AUDIO SPEECH MUSIC PROC.* 2023, *22* (2023). <https://doi.org/10.1186/s13636-023-00289-4>
- [3] B. T. Atmaja, K. Shirai and M. Akagi, "Speech Emotion Recognition Using Speech Feature and Word Embedding," 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Lanzhou, China, 2019, pp. 519-523, doi: 10.1109/APSIPAASC47483.2019.9023098.
- [4] G. Deshmukh, A. Gaonkar, G. Golwalkar and S. Kulkarni, "Speech based Emotion Recognition using Machine Learning," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 812-817, doi: 10.1109/ICCMC.2019.8819858.

[5] Thakur, A., Dhull, S. (2021). Speech Emotion Recognition: A Review. In: Hura, G.S., Singh, A.K., Siong Hoe, L. (eds) *Advances in Communication and Computational Technology. ICACCT 2019. Lecture Notes in Electrical Engineering*, vol 668. Springer, Singapore. https://doi.org/10.1007/978-981-15-5341-7_61