



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



Advances in Speech Emotion Recognition: A Comprehensive Review

M Sai Akash, S Praveena

I M. Tech, DECE, Department of ECE, MGIT, Hyderabad, India

Associate Professor, Department of ECE, MGIT, Hyderabad, India

ABSTRACT: Recognizing emotions from speech signals is a significant challenge in the field of speech technologies. Three major factors impact speech emotion recognition: the selection of an appropriate database, the choice of features for emotion detection, and the selection of a suitable classifier. This paper presents a review of existing research on speech emotion recognition, focusing on these critical aspects.

KEYWORDS: Speech Emotion Recognition, feature extraction, classifier

I. INTRODUCTION

The primary objective of speech emotion recognition is to identify various emotions conveyed through a speech signal. Automatic emotion detection becomes complex when the training and testing datasets originate from different domains due to variations in recording environments, languages, and speakers [1]. Despite advancements in artificial intelligence, achieving natural human-machine interaction remains a challenge, partly due to machines' limited understanding of human emotional states.

Speech emotion recognition, which focuses on identifying emotions from speech signals, has gained increasing attention. It is a complex task requiring effective emotional feature extraction, which continues to be an open research question [2, 9].

A speech emotion recognition system typically involves five stages: emotional speech input, feature extraction, feature selection, classification, and emotion output.

Identifying a set of significant emotions for classification is crucial in designing an effective emotion recognition system [2]. The primary emotions include anger, disgust, fear, joy, sadness, and surprise. Prosodic features such as pitch, energy, duration, formants, Mel Frequency Cepstrum Coefficients (MFCC), and Linear Prediction Cepstrum Coefficients (LPCC) serve as indicators of a speaker's emotional state [3, 4].

For example:

- **Anger:** High pitch mean and variance, and elevated energy levels.
- **Sadness:** Low pitch mean and variance, slow speech rate, and reduced high-frequency spectrum components.
- **Fear:** High pitch range and enhanced high-frequency spectral components.

The choice of classifier significantly influences emotion recognition accuracy. Classifiers are broadly categorized into linear (e.g., Naïve Bayes, linear SVM) and nonlinear (e.g., ANN, GMM, HMM, K-NN) types.

II. RELATED WORK

1. Deep Neural Networks for Feature Extraction:

Han et al. [5] employed deep neural networks (DNNs) to extract high-level features from raw data, generating utterance-level features from segment-level probabilities. Their method achieved a 20% relative accuracy improvement compared to state-of-the-art techniques.

2. Fourier Parameters and Harmony Features:

Wang et al. [6] introduced Fourier parameters derived from voice quality and harmony features, achieving substantial accuracy improvements across multiple databases. Combining these with MFCC further enhanced recognition rates.



3. **Transfer Learning Using PCANet:**

Huang et al. [7] proposed a feature transfer approach using PCANet, aligning domain-specific features with the target domain. Their method outperformed traditional machine learning and transfer learning techniques.

4. **Attention-Based RNNs:**

Mirsamadi et al. [8] developed a deep recurrent neural network with local attention to identify emotionally salient speech regions, achieving superior results on the IEMOCAP corpus.

5. **Prominence Features for Emotion Recognition:**

Jing et al. [9] introduced prominence features in conjunction with acoustic features, improving recognition rates by 6–6.2% across experiments. The features were validated using a dual-mode emotional speech database.

III. CONCLUSION

Key challenges in speech emotion recognition include selecting appropriate features and designing effective classifiers. Accurate classification depends on these components. Future work should focus on enhancing feature extraction methods and integrating advanced classifiers to improve system performance.

REFERENCES

- [1] Seyedmahdad Mirsamadi, Emad Barsoum, Cha Zhang, Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention.
- [2] Ayadi, M. S., Kamel, F., Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases," Pattern Recognition, 44, pp. 572–587, 2011.
- [3] A. Nogueiras, A. Moreno, A. Bonafonte, Jose B. Marino, "Speech Emotion Recognition Using Hidden Markov Model," Eurospeech, 2001.
- [4] P. Shen, Z. Changjun, X. Chen, "Automatic Speech Emotion Recognition Using Support Vector Machine," International Conference on Electronic and Mechanical Engineering and Information Technology, 2011.
- [5] Han, K., Yu, D., Tashev, I., "Speech Emotion Recognition Using Deep Neural Network and Extreme Learning Machine," 2014 ISCA.
- [6] Wang, K., An, N., Li, B. N., Zhang, Y., Li, L., "Speech Emotion Recognition Using Fourier Parameters," IEEE Transactions on Affective Computing, 2015.
- [7] Huang, Z., Xue, W., Mao, Q., Zhan, Y., "Unsupervised Domain Adaptation for Speech Emotion Recognition Using PCANet," Multimedia Tools and Applications, Springer.
- [8] Mirsamadi, S., Barsoum, E., Zhang, C., "Automatic Speech Emotion Recognition Using Recurrent Neural Networks with Local Attention," ICASSAP 2017 IEEE International Conference.
- [9] Jing, S., Mao, X., Chen, L., "Prominence Features: Effective Emotional Features for Speech Emotion Recognition," Digital Signal Processing, 2017.
- [10] Schuller, B., Batliner, A., Steidl, S., Seppi, D., "Recognizing Realistic Emotions and Affect in Speech: State of the Art and Lessons Learnt," Speech Communication, 53(9), pp. 1062–1087, 2011.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details