

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

DIA

International Journal of Innovative Research in SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 2, February 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

 \bigcirc





|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 2, February 2024

| DOI:10.15680/IJIRSET.2024.1302038 |

Emotion Recognition through Speech

Siddhi Yadav, Madeeha Zardi, Roshni Yadav, Bodhisatva Waghmare, Nilam Parmar

Thakur Polytechnic, College of Engineering, Thakur Complex, Kandivali (East), Mumbai, India

ABSTRACT: In this paper we overview emotional speech recognition having in mind three goals. The first goal is to provide an up-to-date record of the available emotional speech data collections. The number of emotional states, the language, the number of speakers, and the kind of speech are briefly addressed. The second goal is to present the most frequent acoustic features used for emotional speech recognition and to assess how the emotion affects them. Typical features are the pitch, the formants, the vocal tract cross-section areas, the mel-frequency the intensity of the speech signal, and the speech rate. The third goal is to review appropriate techniques in order to classify speech into emotional states.

KEYWORDS: Speech, Emotions, Frequency, Emotion Recognition

I. INTRODUCTION

Emotional speech recognition aims at automatically identifying the emotional or physical state of a human being from his or her voice. The emotional and physical states of a speaker are known as emotional aspects of speech and are included in the so-called paralinguistic aspects. Although the emotional state does not alter the linguistic content, it is an important factor in human communication, because it provides feedback information in many applications as it is outlined next. Speech emotion recognition (SER), a subdiscipline of affective computing has been around for more than two decades and has led to a considerable amount of published works A typical SER system can be considered as a collection of methodologies that isolate, extract and classify speech signals to detect emotions embedded in them



Fig. 1: Logo of Emotion Recognition

II. LITERATURE REVIEW

Emotion recognition through speech is a fascinating and rapidly evolving field that intersects the domains of psychology, linguistics, signal processing, and machine learning. In this literature review, we will explore key studies, methodologies, and findings related to emotion recognition through speech. Introduction to Emotion Recognition through Speech Initial studies in this field aimed at understanding the acoustic and linguistic features associated with different emotions conveyed through speech. Research often starts by categorizing emotions into basic categories such as happiness, sadness, anger, fear, surprise, and disgust Acoustic Features in Emotion Recognition Acoustic features play a crucial role in emotion detection. Pitch, intensity, duration, and formants are among the commonly studied acoustic parameters. Studies have shown that certain emotions are associated with distinct patterns in pitch and intensity. For instance, high pitch and intensity might be associated with excitement or anger, while low pitch and intensity could be linked to sadness or calmness.

e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |



|| Volume 13, Issue 2, February 2024 ||

DOI:10.15680/IJIRSET.2024.1302038



III. FEATURES

Emotion recognition through speech involves extracting and analysing acoustic features from spoken language to identify underlying emotions.

Here are key features in detail:

- 1. Pitch (Frequency): Emotions often influence the pitch of one's voice. High pitch can indicate excitement or fear, while low pitch may convey sadness or anger.
- 2. Intensity (Amplitude): The loudness or intensity of speech can reflect emotions. Increased intensity may represent anger, while decreased intensity could signal sadness or calmness.
- **3.** Duration: The length of pauses, speech segments, or phonemes can provide insights into emotional states. For instance, long pauses might indicate uncertainty or hesitation.
- 4. Speech Rate (Tempo): Emotions can affect how fast or slow someone speaks. Rapid speech may indicate excitement or anxiety, while slow speech may suggest sadness or contemplation.
- 5. Voice Quality (Timbre): Changes in the tonal quality or timbre of the voice can be indicative of emotional states. Stress or tension might alter the overall voice quality.
- 6. Prosody (Rhythm and Intonation): Emotions influence the rhythm and intonation of speech. Happy emotions might be associated with a more rhythmic and melodious tone, while sadness may result in a flatter intonation.
- 7. Formants (Resonance Frequencies): Resonance frequencies in speech can be influenced by emotions, affecting vowel sounds. Analysing formants can provide additional cues for emotion recognition.
- 8. Voice Modulation: Variability in pitch, intensity, and duration within a sentence or phrase can convey emotional nuances. Emotional speech often involves more dynamic voice modulation.



e-ISSN: 2319-8753, p-ISSN: 2347-6710 www.ijirset.com | Impact Factor: 8.423 | A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 2, February 2024

| DOI:10.15680/IJIRSET.2024.1302038 |

- 9. Articulation: Changes in speech articulation, such as speech rate and clarity of pronunciation, can provide information about emotional states. Stress or excitement may lead to less precise articulation.
- **10.** Silences and Pauses: Emotional speech patterns may include specific patterns of silences and pauses. Studying these temporal aspects can offer valuable information for emotion recognition.
- **11.** Emotional Prosody Patterns: Different emotions are associated with distinct prosody patterns, involving pitch, intensity, and duration variations. Analysing these patterns helps in identifying specific emotions.
- 12. Speech Content and Language Use: The choice of words, expressions, and language style can also contribute to emotion recognition. Certain emotions may be associated with specific lexical choices and linguistic patterns.
- **13.** Speech Quality:- Articulation: Emotional states can influence how clearly and precisely words are articulated. Stress or excitement might lead to rushed or less precise articulation, while calmness may result in clearer speech.
- 14. Speech Prosody:- Prosodic Features: Beyond pitch and rhythm, prosody involves variations in duration, loudness, and pitch contour. Rising or falling pitch patterns, changes in speech rate, and variations in loudness contribute to the overall prosody of emotional speech.
- 15. Non-Verbal Vocalizations:
 - Laughter, Crying, Sighs: Emotional expressions often extend beyond words to include non-verbal sounds. Laughter might indicate joy, while sighs can convey relief or resignation.
- 16. Formant Frequencies:
 - Resonance Patterns: Formants are resonant frequencies in the vocal tract. Changes in these frequencies can provide information about the emotional state. For example, higher formant frequencies may be associated with excitement.
- 17. Speech Energy Distribution:
 - Energy in Different Frequency Bands: Analysing how energy is distributed across various frequency bands in speech can reveal emotional cues. Certain emotions may be characterized by distinctive energy patterns in specific frequency ranges.

Combining these features through machine learning techniques, such as deep learning or traditional signal processing methods, allows systems to accurately recognize emotions in speech, contributing to applications in areas like human-computer interaction, sentiment analysis, and mental health assessment.





e-ISSN: 2319-8753, p-ISSN: 2347-6710 www.ijirset.com | Impact Factor: 8.423 | A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 2, February 2024

| DOI:10.15680/IJIRSET.2024.1302038 |

IV. TECHNOLOGIES USED

The methodology for emotion recognition through speech typically involves a multi-step process combining signal processing, feature extraction, and machine learning techniques. Here's a general overview:

1. Data Collection:

Gather a diverse dataset of audio recordings that capture a range of emotions. This dataset should include various speakers, contexts, and emotional expressions.

2. Preprocessing:

Clean the audio data to remove noise and ensure high-quality recordings. This step may also involve segmenting the audio into relevant portions corresponding to different emotional expressions.

3. Feature Extraction:

Utilize signal processing techniques to extract relevant features from the audio data. This includes extracting features such as pitch, intensity, speech rate, formant frequencies, and other acoustic and linguistic features discussed earlier. 4. Labelling:

Annotate the dataset with ground truth labels indicating the emotional state associated with each audio segment. This labelling is crucial for training and evaluating machine learning models.

5. Model Training:

Choose an appropriate machine learning model for emotion recognition. Common choices include support vector machines, decision trees, or neural networks. Train the model using the labelled dataset, with extracted features as input and labelled emotions as output.

6. Validation and Testing:

Evaluate the trained model on a separate set of data not used during training. This helps ensure the model generalizes well to new, unseen data. Adjust the model parameters if needed.

7. Optimization:

Fine-tune the model to improve its performance. This may involve adjusting hyperparameters, exploring different feature sets, or experimenting with more advanced machine learning architectures.

8. Integration:

Implement the trained model into the target application or system where emotion recognition is required. This could be integrated into realtime speech processing systems or used for analysing recorded audio.

9. Deployment:

Deploy the emotion recognition system in realworld scenarios, continuously monitoring and refining its performance based on user feedback and additional data.

10. Ethical Considerations:

Consider ethical implications surrounding privacy and consent, especially when dealing with sensitive emotional data. Ensure transparency in how the technology is used and provide users with clear information about the purpose and implications of emotion recognition.

Throughout this process, collaboration between experts in signal processing, machine learning, and domain specialists in linguistics and psychology is crucial for developing a robust emotion recognition system that accurately captures the subtleties of human emotional expression in speech.



JIRSET

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 2, February 2024

| DOI:10.15680/IJIRSET.2024.1302038 |

V. LIMITATIONS

_While emotion recognition through speech has promising applications, it comes with several limitations and challenges:

1. Subjectivity and Cultural Variability: - Emotions are subjective and can vary across individuals and cultures. What may be considered an emotional expression in one culture might not hold the same meaning in another, making it challenging to create universally applicable models.

2. Context Dependency:

- The meaning of speech and associated emotions heavily depends on context. The same words or tone of voice can convey different emotions based on the surrounding circumstances, making accurate emotion interpretation complex.

3. Ambiguity and Overlapping Emotions: - Speech often involves a mix of emotions, and emotions can overlap. Disentangling and accurately categorizing these overlapping emotional states pose challenges for current recognition systems.

4. Dynamic Nature of Emotions:

- Emotions are dynamic and can change rapidly. Current models might struggle to keep pace with real-time emotional shifts, impacting the accuracy of predictions.

5. Data Bias and Generalization: - Training datasets may not adequately represent the diversity of emotional expressions. Biases in training data can lead to models that perform well for certain demographics but poorly for others, affecting generalization to new and diverse populations.

6. Limited Expressiveness of Speech: - Speech might not fully capture the complexity of emotions. Non-verbal cues, facial expressions, and body language play crucial roles in emotional communication, and relying solely on speech might provide an incomplete picture.

7. Privacy Concerns:

- Analysing emotional states through speech raises privacy concerns. Users may be uncomfortable with systems that continuously analyse and interpret their emotional expressions, raising ethical questions about consent and data usage.

8. Emotion Deception:

- Individuals may intentionally modify their speech to convey or conceal certain emotions, challenging the accuracy of emotion recognition systems. This is particularly relevant in situations like job interviews or security contexts.

9. Ethical Considerations:

- The use of emotion recognition technology raises ethical questions, including potential misuse for surveillance, emotional manipulation, or decisions with significant consequences based on inferred emotional states.

10. Technical Challenges:

- Developing accurate models requires addressing technical challenges, such as the need for large and diverse datasets, ensuring real-time processing capabilities, and dealing with the computational complexity of feature extraction and model training.

Understanding and addressing these limitations is crucial for the responsible development and deployment of emotion recognition systems. Ongoing research and advancements in technology will likely contribute to mitigating some of these challenges in the future.

VI. CONCLUSION

In conclusion, emotion recognition through speech is a rapidly advancing field that leverages acoustic, linguistic, and prosodic features, along with machine learning models. While significant progress has been made, challenges like subjective labelling and cultural variations persist. Integrating multimodal approaches and applying natural language



e-ISSN: 2319-8753, p-ISSN: 2347-6710 www.ijirset.com | Impact Factor: 8.423 | A Monthly Peer Reviewed & Referred Journal |

Volume 13, Issue 2, February 2024

| DOI:10.15680/IJIRSET.2024.1302038 |

processing have enhanced accuracy. Ethical considerations and the responsible deployment of technology are emerging concerns. The potential applications, from human-computer interaction to mental health monitoring, highlight the significance of this research. Ongoing efforts are focused on improving model generalization, robustness, and the incorporation of contextual information for a more nuanced understanding of human emotions. In summary, the future holds promising developments in emotion recognition through speech, with potential breakthroughs and applications on the horizon.

REFERENCES

[1]https://www.sciencedirect.com/science/art/pii/

[2]Emotion recognition from speech signal | IEEE Conference Publication | IEEE Xplore

[3] Speech Emotion Recognition | Kaggle

[4] <u>Speech Emotion Recognition Papers With</u> Code

[5] <u>Speech Emotion Recognition - an overview</u> <u>ScienceDirect Topics</u>

[6] Speech emotion recognition using machine learning — A systematic review - ScienceDirect

[7] Speech Emotion Recognition Project using Machine Learning (projectpro.io) A Survey of Human Emotion

 Recognition Using Speech Signals: Current Trends and Future
 Perspectives | SpringerLink

 [8]
 (PDF) Automatic Emotion Recognition from Speech Signals: A Review (researchgate.net)

[9] Google Image Result









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com