**Speech Emotion Recognition**

**PROJECT SYNOPSIS**

OF MAJOR PROJECT

Submitted in partial fulfillment of the requirement for the degree of

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING**

SUBMITTED BY

**AMRITA SINGH 2000290100019**

**ANSHIKA GUPTA 2000290100025**

**DIVYA GARG 2000290100058**

**2024 Batch**

(August 2022)

Under the supervision of

(Teacher)

**Dr. Sanjiv Sharma**



## **KIET Group of Institutions, Delhi-NCR,**

## **Ghaziabad (UP)**

## **Department of Computer Science and Engineering**

**Table of contents**

|  |  |
| --- | --- |
| **Content** | **Page no.** |
| INTRODUCTION | 3 |
| RATIONALE | 4 |
| OBJECTIVES | 5 |
| LITERATURE REVIEW | 6 |
| FEASIBILITY STUDY | 7 |
| METHODOLOGY | 8 |
| FACILITIES REQUIRED FOR PROJECT | 9 |
| EXPECTED OUTCOME | 10 |
| REFERENCES | 11 |

**INTRODUCTION**

Speech is a way of communicating our feelings, thoughts, and ideas to other people. Speeches provide us the opportunity to interact with others and build relationships. Understanding the speech's topic and the speaker's feelings is equally crucial for the audience. A person can easily tell the speaker's emotions by closely listening to his or her speech and provide appropriate feedback.

Multiple precise speech reinforcements that are regularly practiced using voice-based data. The effectiveness of voice in use is important. According to a recent analysis, by 2022, just voice commands will be necessary for around 12% of all user applications to function as intended. It is crucial to distinguish the speech signal in both examples of voice communications, whether they are bidirectional or monodirectional. One such reinforcement that controls a number of its functions by voice-based management is self-driving cars.

The voice-based smart machines are in high demand across a broad spectrum of applications. In a voice-based system, a computer agent must fully understand the human's speech perception in order to effectively understand the commands issued to it. Speaker identification, speech recognition, and speech emotion detection make up the three sections of this study, which is titled "Speech Processing."

Speech Emotion Recognition is claiming to achieve between the additional parts due to its complexity. Moreover, the representation of a sensible computer system lacks the system to impersonate human response. A fascinating character individual to humans is the capability to reconstruct communications based on the emotional nature of the addresser and the witness. Speech emotion detection built as a classification obstacle determined using numerous ML algorithms. This report emphasizes more on the Speech Emotion Recognition system. Recognizing the emotional nature of the user appears with a significant lead in this application.

The emotion conveyed in the user's expression of the decision can adjust to many emergency aspects of the vehicle, taking into account emergency situations in which the user may not be able to deliver a spoken order. Call centers offer a far more basic use of speech emotion detection, allowing for the efficient assignment of automatic voice suggestions to customer care agents for further examination.

**RATIONALE**

As human beings speech is amongst the most natural way to express ourselves. We depend so much on it that we recognize its importance when resorting to other communication forms like emails and text messages where we often use emojis to express the emotions associated with the messages. As emotions play a vital role in communication, the detection and analysis of the same is of vital importance in today’s digital world of remote communication. Emotion detection is a challenging task, because emotions are subjective. There is no common consensus on how to measure or categorize them. We define a SER system as a collection of methodologies that process and classify speech signals to detect emotions embedded in them. Such a system can find use in a wide variety of application areas like interactive voice based-assistant or caller-agent conversation analysis. In this study we attempt to detect underlying emotions in recorded speech by analysing the acoustic features of the audio data of recordings.

There are three classes of features in a speech namely, the lexical features (the vocabulary used), the visual features (the expressions the speaker makes) and the acoustic features (sound properties like pitch, tone, jitter, etc.). The problem of speech emotion recognition can be solved by analyzing one or more of these features. Choosing to follow the lexical features would require a transcript of the speech which would further require an additional step of text extraction from speech if one wants to predict emotions from real-time audio. Similarly, going forward with analyzing visual features would require the excess to the video of the conversations which might not be feasible in every case while the analysis on the acoustic features can be done in real-time while the conversation is taking place as we’d just need the audio data for accomplishing our task. Hence, we choose to analyse the acoustic features in this work. Furthermore, the representation of emotions can be done in two ways:

* **Discrete Classification:** Classifying emotions in discrete labels like anger, happiness, boredom, etc.
* **Dimensional Representation:** Representing emotions with dimensions such as Valence (on a negative to positive scale), Activation or Energy (on a low to high scale) and Dominance (on an active to passive scale).

Both these approaches have their pros and cons. The dimensional approach is more elaborate and gives more context to prediction but it is harder to implement and there is a lack of annotated audio data in a dimensional format. The discrete classification is more straightforward and easier to implement but it lacks the context of the prediction that dimensional representation provides. We have used the discrete classification approach in the current study for lack of dimensionally annotated data in the public domain.

**OBJECTIVES**

To build a model to recognize emotions from speech using python’s **librosa** library.

* **Data Source**

For this project, we will be using the **RAVDESS**dataset which is the abbreviated form of Ryerson Audio-Visual Database of Emotional Speech and Song dataset. This dataset has **7356** files rated by **247** individuals 10 times on emotional validity, intensity, and genuineness.

* **Libraries Used**

The main library that we will be using is **Librosa.**Apart from that we will also be using**Soundfile**and**Pyaudio. Librosa**is a Python library for analyzing audio and music. It has a flatter package layout, standardizes interfaces and names, backwards compatibility, modular functions, and readable code.

* **Features Used In This Study**

From the Audio data we have extracted **three**key features which have been used in this study, namely,**MFCC** (Mel Frequency Cepstral Coefficients), **Mel Spectrogram**and **Chroma.** Librosa was used in their extraction.

**MFCC:**MFCC was by far the most researched about and utilized feature in this dataset. It represents the short-term power spectrum of a sound.

**Mel Spectrogram:**This is just a spectrogram that depicts amplitude which is mapped on a Mel scale.

**Chroma:**A Chroma vector is typically a 12-element feature vector indicating how much energy of each pitch class is present in the signal in a standard chromatic scale.

* **Extracting The Features**

We simply define a function to extract the MFCC, Chroma, and Mel features from a sound file. This function takes 4 parameters- the file name and three Boolean parameters for the three features. Open the sound file with Sound File library. Read from it and call it **X.** Also, get the sample rate. If chroma is True, get the **Short-Time Fourier Transform** of **X.**

**LITERATURE REVIEW**

Over the last years a lot of researches have been made to extract the emotion of human speech some of the studies include:

s. Cao et al. [1] proposed a ranking SVM method for synthesize information about emotion recognition to solve the problem of binary classification.

Chen et al. [2] aimed to improve speech emotion recognition in speaker-independent with three level speech emotion recognition method.

Nwe et al. [3] proposed a new system for emotion classification of utterance signals. The system employed a short time log frequency power coefficients (LFPC) and discrete HMM to characterize the speech signals and classifier respectively.

T Wu et al. [4] proposed a new modulation spectral features (MSFs) human speech emotion recognition.

Rong et al. [5] presented an ensemble random forest to trees (ERFTrees) method with ahigh number of features for emotion recognition without referring any language or linguistic information remains an unclosed problem.

Wu et al. [6] proposed a fusion-based method for speech emotion recognition by employing multiple classifier and acoustic-prosodic (AP) features and semantic labels(SLs).

Narayanan [7] proposed domain-specific emotion recognition by utilizing speech signals from call center application. Detecting negative and non-negative emotion (e.g. anger and happy) are the main focus of this research.

Yang & Lugger [8] presented a novel set of harmony features for speech emotion recognition. These features are relying on psychoacoustic perception from music theory.

Albornoz et al. [9] investigate a new spectral feature in order to determine emotions and to characterize groups.

Lee et al. [10] represent a hierarchical computational structure to identify emotions.

Lee et al. [11-12] proposed hierarchical structure for binary decision tree in emotion recognition fields.

Yeh et al. [13] proposed a segment based method for recognition of emotion in Mandarin speech

**Feasibility Study**

Children with autism spectrum disorder have difficulty in understanding the emotional and mental states from the facial expressions of the people they interact. The inability to understand other people’s emotions will hinder their interpersonal communication. Though many facial emotion recognition algorithms have been proposed in the literature, they are mainly intended for processing by a personal computer, which limits their usability in on-the-move applications where portability is desired. The portability of the system will ensure ease of use and real-time emotion recognition and that will aid for immediate feedback while communicating with caretakers. Principal component analysis (PCA) has been identified as the least complex feature extraction algorithm to be implemented in hardware. In this paper, we present a detailed study of the implementation of serial and parallel implementation of PCA in order to identify the most feasible method for realization of a portable emotion detector for autistic children. The proposed emotion recognizer architectures are implemented on Virtex 7 XC7VX330T FFG1761-3 FPGA. We achieved 82.3 % detection accuracy for a word length of 8 bits.

Recognizing and responding to facial emotions is a key feature of interpersonal communication and a significant modulator of social behavior. In typical children, recognition of emotional facial expressions is an early developing social skill. Failure of these fundamental early emotion recognition skills would have profound consequences for a child’s social development, cutting the child off from learning about other people’s feelings and responses People, especially children, diagnosed with diseases such as autism, Alzheimer’s disease and/or Parkinson’s disease often lack or have impairments in some set of representation abilities such that they have difficulties operating in our highly complex social environment [[5](https://link.springer.com/article/10.1007/s11517-015-1346-z" \l "ref-CR5" \o "Happe F, Briskman J (2001) Exploring the cognitive phenotype of Autism: weak ), [9](https://link.springer.com/article/10.1007/s11517-015-1346-z" \l "ref-CR9" \o "LoPresti EF, Bodine C (2008) Assistive technology for cognition [Understanding the needs of person with disabilities]. Eng Med Biol Mag 27(2):29–39)]. Children suffering from pervasive developmental disorders (PDD), such as Asperger’s disorder [[9](https://link.springer.com/article/10.1007/s11517-015-1346-z" \l "ref-CR9" \o "LoPresti EF, Bodine C (2008) Assistive technology for cognition [Understanding the needs of person with disabilities]. Eng Med Biol Mag 27(2):29–39)], concentrate more on component parts rather than the wholesome features [[5](https://link.springer.com/article/10.1007/s11517-015-1346-z" \l "ref-CR5" \o "Happe F, Briskman J (2001) Exploring the cognitive phenotype of Autism: weak )]. As a result, they have difficulties in tasks such as envisioning another’s state of mind in social behavior or imaging future states to plan a task [[6](https://link.springer.com/article/10.1007/s11517-015-1346-z" \l "ref-CR6" \o "Hart M (2005) Autism/Excel study, ASSETS 2005. In: The 7th international ACM SIGACCESS conference on computers and accessibility, New York, pp. 136–141)]. Such difficulties in empathy underlie their social communication difficulties that form a core of the diagnosis. Emotion expression is an essential part of human interaction. Emotions are universal means of communication which can be expressed nonverbally without any language constraints. They are recognized through facial expressions, voice tones, speech and physiological signals. The speech signals provide the context for the expressed emotions but cannot be used as a universal indicator for recognizing them as they are dependent on the language content. The physiological signals such as blood pressure, body temperature and force feedback are relatively more accurate and universal indicators of emotions, but they require user-aware and intrusive methods for collecting the data.

**METHODOLOGY**

**Existing System:**

Given that the possible computational sources were confined, and only a small database of emotionally identified speech examples was available, the initial aim was to be more inclined towards a computationally productive approach that would work with a bit of training data set. These restrictions are pretty standard and may be addressed by the appliance of pre-trained networks and transfer learning. Since the bulk of existing pre-trained networks are designed for image classification, the SER problem had to be re-defined as a picture classification task to use these networks to speech. To realize this, labelled speech samples were buffered into short-time blocks. Every block was measured in a spectrogram array of spectral amplitude and later converted its format into RGB, then passed into a pre-trained Convolutional Neural Network Model. After comprehensive training, now the Convolutional Neural Network Model can be able to judge different emotions. Now every speech is going through a comprehensive process of Speech to Visualized Picture conversion. Within the experiments presented here, the SER performance was tested using two different sampling frequencies (sixteen and eight kHz) and, therefore, the μ-low companding procedure.

**Proposed System:**

The speech emotion detection system is performed as a Machine Learning (ML) model. The steps of operation are similar to any other ML project, with supplementary fine- tuning systems to make the model function adequately. The fundamental action is data collection, which is of prime importance. The model being generated will acquire from the data contributed to it and all the conclusions and decisions that a progressed model will produce is supervised data. The secondary action, called as feature engineering, is a combination of various machine learning assignments that are performed over the gathered data. These systems approach the various data description and data quality problems. The third step is often explored the essence of an ML project where an algorithmic based prototype is generated. This model uses an ML algorithm to determine about the data and instruct itself to react to any new data it is exhibited to. The ultimate step is to estimate the functioning of the built model. Very frequently, developers replicate the steps of generating a model and estimating it to analyze the performance of various algorithms. Measuring outcomes help to choose the suitable ML algorithm most appropriate to the predicament.

**Dataset:**

English Language Dataset, namely the Torronto Emotional Speech Set (TESS) was taken into consideration. This is a dataset which consists of 200 target words and were spoken by two women, one younger and the other older, and the phrases were recorded in order to portray the following seven emotions happy, sad, angry, disgust, fear, surprise and neutral state. There are 2800 files in total, in this dataset. The phrase spoken by them was “Say the word \_\_\_”. Both women have thresholds within normal range and the audio quality is very high. The authors are Kate Dupuis and M. Kathleen Pichora Fuller.

**Facilities Required**

* Emotional Speech Databases

There are three types of databases specifically designed for speech emotion recognition, simulated, semi-natural, and natural speech collections. The simulated datasets are created by trained speakers reading the same text with different emotions [54]. Semi-natural collections are made by asking people or actors to read a scenario containing different emotions. Moreover, natural datasets are extracted from TV shows, YouTube videos, call centers, and such, and then labeled the emotions by human listeners [54]. Simulated data sets such as EMO-DB (German) [55], DES (Danish) [56], RAVDESS [57], TESS [58], and CREMA-D [59] are standardized collections of emotions, which makes comparing results very easy. Although their numbers of distinct emotions are significant, as they have synthesized emotions, they tend to have overfitted models around emotions slightly different than what is happening in day-to-day conversations. Semi-natural collections of emotions include IEMOCAP [60], Belfast [61], and NIMITEK [62]. This group has the advantage of being very similar to the natural utterances of speech. However, even though they are based on scenarios and the speech is happening in a contextual setting, they are artificially created emotions

* Emotion Recognition Methods
* Traditional Methods
* . Hidden Markov Models (HMM)
* Support Vector Machine (SVM)
* Neural Networks and Deep Learning
* Artificial Neural Networks
* Deep Learning
* Convolutional Neural Networks
* LSTM Networks

**EXPECTED OUTCOMES**

Through this project, we confirmed how we will leverage Machine mastering to acquire the underlying emotion from speech audio information and a few insights at the human expression of emotion via voice. This machine may be hired in a whole lot of setups like Call Centre for lawsuits or marketing, in voice-primarily based totally digital assistants or chatbots, in linguistic research, etc.

**A few feasible steps that may be carried out to make the fashions extra strong and correct are the following**

* A correct implementation of the tempo of the talking may be explored to test if it may solve a number of the deficiencies of the model.
* Figuring out a manner to clean random silence from the audio clip.
* Exploring different acoustic capabilities of sound records to test their applicability with inside the area of speech emotion recognition. These capabilities may want to virtually be a few proposed extensions of MFCC like RAS-MFCC or they may be different capabilities absolutely like LPCC, PLP or Harmonic cepstrum.
* Following lexical capabilities primarily based totally method closer to SER and the use of an ensemble of the lexical and acoustic models. This will enhance the accuracy of the machine due to the fact in a few instances the expression of emotion is contextual in preference to vocal.
* Adding extra records extent both with the aid of using different augmentation strategies like time-transferring or dashing up/slowing down the audio or virtually locating extra annotated audio clips.

**REFERENCES**

1. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking:

Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–

202, Jan. 2015.

2. L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and

classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012

3. T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden

Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.

4. S. Wu, T. H. Falk, and W.-Y. Chan, “Automatic speech emotion recognition using

modulation spectral features,” Speech Commun., vol. 53, no. 5, pp. 768–785, May 2011.

5. J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion

recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.

6. C.-H. Wu and W.-B. Liang, “Emotion Recognition of Affective Speech Based on Multiple

Classifiers Using Acoustic-Prosodic Information and Semantic Labels,” IEEE Trans. Affect.

Comput., vol. 2, no. 1, pp. 10–21, Jan. 2011.

7. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio

Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

8. B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony

features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010

9. E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using

hierarchical classifiers,” Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.

10. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Speech Commun., vol. 53, no. 9–10, pp. 1162–

1171, Nov. 2011.

11. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Interspeech, vol. 53, pp. 320–323, 2009.

12. S. Bjorn, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 Emotion Challenge,” 2009.

13. J.-H. Yeh, T.-L. Pao, C.-Y. Lin, Y.-W. Tsai, and Y.-T. Chen, “Segment-based emotion

recognition from continuous Mandarin Chinese speech,” Comput. Human Behav., vol. 27,

no. 5, pp. 1545–1552, Sep. 2011.

14. W. Dai, D. Han, Y. Dai, and D. Xu, “Emotion Recognition and Affective Computing on

Vocal Social Media,” Inf. Manag., Feb. 2015.

15. M. M. H. El Ayadi, M. S. Kamel, and F. Karray, “Speech Emotion Recognition using

Gaussian Mixture Vector Autoregressive Models,” in 2007 IEEE International Conference

on Acoustics, Speech and Signal Processing - ICASSP ’07, 2007, vol. 4, pp. IV–957–IV–

960.

25

1. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking:

Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–

202, Jan. 2015.

2. L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and

classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012

3. T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden

Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.

4. S. Wu, T. H. Falk, and W.-Y. Chan, “Automatic speech emotion recognition using

modulation spectral features,” Speech Commun., vol. 53, no. 5, pp. 768–785, May 2011.

5. J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion

recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.

6. C.-H. Wu and W.-B. Liang, “Emotion Recognition of Affective Speech Based on Multiple

Classifiers Using Acoustic-Prosodic Information and Semantic Labels,” IEEE Trans. Affect.

Comput., vol. 2, no. 1, pp. 10–21, Jan. 2011.

7. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio

Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

8. B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony

features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010

9. E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using

hierarchical classifiers,” Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.

10. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Speech Commun., vol. 53, no. 9–10, pp. 1162–

1171, Nov. 2011.

11. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Interspeech, vol. 53, pp. 320–323, 2009.

12. S. Bjorn, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 Emotion Challenge,” 2009.

13. J.-H. Yeh, T.-L. Pao, C.-Y. Lin, Y.-W. Tsai, and Y.-T. Chen, “Segment-based emotion

recognition from continuous Mandarin Chinese speech,” Comput. Human Behav., vol. 27,

no. 5, pp. 1545–1552, Sep. 2011.

14. W. Dai, D. Han, Y. Dai, and D. Xu, “Emotion Recognition and Affective Computing on

Vocal Social Media,” Inf. Manag., Feb. 2015.

15. M. M. H. El Ayadi, M. S. Kamel, and F. Karray, “Speech Emotion Recognition using

Gaussian Mixture Vector Autoregressive Models,” in 2007 IEEE International Conference

on Acoustics, Speech and Signal Processing - ICASSP ’07, 2007, vol. 4, pp. IV–957–IV–

960.

25

1. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking:

Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–

202, Jan. 2015.

2. L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and

classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012

3. T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden

Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.

4. S. Wu, T. H. Falk, and W.-Y. Chan, “Automatic speech emotion recognition using

modulation spectral features,” Speech Commun., vol. 53, no. 5, pp. 768–785, May 2011.

5. J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion

recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.

6. C.-H. Wu and W.-B. Liang, “Emotion Recognition of Affective Speech Based on Multiple

Classifiers Using Acoustic-Prosodic Information and Semantic Labels,” IEEE Trans. Affect.

Comput., vol. 2, no. 1, pp. 10–21, Jan. 2011.

7. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio

Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.

8. B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony

features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010

9. E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using

hierarchical classifiers,” Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.

10. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Speech Commun., vol. 53, no. 9–10, pp. 1162–

1171, Nov. 2011.

11. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a

hierarchical binary decision tree approach,” Interspeech, vol. 53, pp. 320–323, 2009.

12. S. Bjorn, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 Emotion Challenge,” 2009.

13. J.-H. Yeh, T.-L. Pao, C.-Y. Lin, Y.-W. Tsai, and Y.-T. Chen, “Segment-based emotion

recognition from continuous Mandarin Chinese speech,” Comput. Human Behav., vol. 27,

no. 5, pp. 1545–1552, Sep. 2011.

14. W. Dai, D. Han, Y. Dai, and D. Xu, “Emotion Recognition and Affective Computing on

Vocal Social Media,” Inf. Manag., Feb. 2015.

15. M. M. H. El Ayadi, M. S. Kamel, and F. Karray, “Speech Emotion Recognition using

Gaussian Mixture Vector Autoregressive Models,” in 2007 IEEE International Conference

on Acoustics, Speech and Signal Processing - ICASSP ’07, 2007, vol. 4, pp. IV–957–IV–

960.

25

1. H. Cao, R. Verma, and A. Nenkova, “Speaker-sensitive emotion recognition via ranking: Studies on acted and spontaneous speech,” Comput. Speech Lang., vol. 28, no. 1, pp. 186–202, Jan. 2015.
2. L. Chen, X. Mao, Y. Xue, and L. L. Cheng, “Speech emotion recognition: Features and classification models,” Digit. Signal Process., vol. 22, no. 6, pp. 1154–1160, Dec. 2012
3. T. L. Nwe, S. W. Foo, and L. C. De Silva, “Speech emotion recognition using hidden Markov models,” Speech Commun., vol. 41, no. 4, pp. 603–623, Nov. 2003.
4. S. Wu, T. H. Falk, and W.-Y. Chan, “Automatic speech emotion recognition using modulation spectral features,” Speech Commun., vol. 53, no. 5, pp. 768–785, May 2011.
5. J. Rong, G. Li, and Y.-P. P. Chen, “Acoustic feature selection for automatic emotion recognition from speech,” Inf. Process. Manag., vol. 45, no. 3, pp. 315–328, May 2009.
6. C.-H. Wu and W.-B. Liang, “Emotion Recognition of Affective Speech Based on Multiple Classifiers Using Acoustic-Prosodic Information and Semantic Labels,” IEEE Trans. Affect. Comput., vol. 2, no. 1, pp. 10–21, Jan. 2011.
7. S. S. Narayanan, “Toward detecting emotions in spoken dialogs,” IEEE Trans. Speech Audio Process., vol. 13, no. 2, pp. 293–303, Mar. 2005.
8. B. Yang and M. Lugger, “Emotion recognition from speech signals using new harmony features,” Signal Processing, vol. 90, no. 5, pp. 1415–1423, May 2010
9. E. M. Albornoz, D. H. Milone, and H. L. Rufiner, “Spoken emotion recognition using hierarchical classifiers,” Comput. Speech Lang., vol. 25, no. 3, pp. 556–570, Jul. 2011.
10. C.-C. Lee, E. Mower, C. Busso, S. Lee, and S. Narayanan, “Emotion recognition using a hierarchical binary decision tree approach,” Speech Commun., vol. 53, no. 9–10, pp. 1162–1171, Nov. 2011.
11. S. Bjorn, S. Steidl, and A. Batliner, “The INTERSPEECH 2009 Emotion Challenge,” 2009.
12. W. Dai, D. Han, Y. Dai, and D. Xu, “Emotion Recognition and Affective Computing on Vocal Social Media,” Inf. Manag., Feb. 2015.
13. M. M. H. El Ayadi, M. S. Kamel, and F. Karray, “Speech Emotion Recognition using Gaussian Mixture Vector Autoregressive Models,” in 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07, 2007, vol. 4, pp. IV–957–IV–960.
14. E. Mower, M. J. Mataric, and S. Narayanan, ``A framework for automatic human emotion classification using emotion profiles,'' IEEE Trans. Audio, Speech, Language Process., vol.19, no. 5, pp. 1057-1070, Jul. 2011
15. M. Sidorov, S. Ultes, and A. Schmitt, ``Emotions are a personal thing: Towards speaker- adaptive emotion recognition,'' in Proc. IEEE Int. Conf.
16. M. Sidorov, S. Ultes, and A. Schmitt, ``Emotions are a personal thing: Towards speaker-
17. adaptive emotion recognition,'' in Proc. IEEE Int. Conf
18. E. Mower, M. J. Mataric, and S. Narayanan, ``A framework for automatic human emotion
19. classification using emotion profiles,'' IEEE Trans. Audio, Speech, Language Process., vol.
20. 19, no. 5, pp. 1057-1070, Jul. 2011

32.B. Schuller, A. Batliner, S. Steidl, and D. Seppi, ``Recognising realistic emotions and affect in

speech: State of the art and lessons learnt from the rst challenge,'' Speech Commun., vol. 53,

nos. 9-10,pp. 10621087, Nov./Dec. 2011.

33. M. Sidorov, S. Ultes, and A. Schmitt, ``Emotions are a personal thing: Towards speaker-

adaptive emotion recognition,'' in Proc. IEEE Int. Conf

32.B. Schuller, A. Batliner, S. Steidl, and D. Seppi, ``Recognising realistic emotions and affect in

speech: State of the art and lessons learnt from the rst challenge,'' Speech Commun., vol. 53,

nos. 9-10,pp. 10621087, Nov./Dec. 2011.

33. M. Sidorov, S. Ultes, and A. Schmitt, ``Emotions are a personal thing: Towards speaker-

adaptive emotion recognition,'' in Proc. IEEE Int. Conf