



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 2)

Available online at: www.ijariit.com

Feature extraction in emotion recognition: An analysis of emotion using Praat

Kunal Kapoor

navdeep1967@gmail.com

Chitkara University Institute of Engineering and
Technology, Patiala, Punjab

Lokesh Kumar

lokesh171439.cse@chitkara.edu.in

Chitkara University Institute of Engineering and
Technology, Patiala, Punjab

Kshitiz Sagar Verma

kshitizsv@gmail.com

Chitkara University Institute of Engineering and
Technology, Patiala, Punjab

Kunal Sehgal

sehgalkunal18@gmail.com

Chitkara University Institute of Engineering and
Technology, Patiala, Punjab

ABSTRACT

Machine learning is being used to help robots to detect human emotions. Speech is an important tool to recognize emotion. It has been observed after reading different research papers from different researchers that pitch and intensity are major features to detect the emotion of sound. Consequently, the present research study is oriented towards discussing aspects pertaining to emotion recognition using speech. This paper presents the illustration of emotions that can be recognized by the speech of a person. Future research will be focused on enhancing features like energy, frequency, and amplitude of a speech.

Keywords— Artificial Intelligence, Machine learning, Speech recognition, Speech features, Speech intensity, Pitch

1. INTRODUCTION

Machine learning is the study of computer programs and algorithms that allows the user to perform a specific task without using the exact command [1, 2]. It is a sub-divided subject of artificial intelligence that is inclined towards making computer intelligent. [3] As for a computer to be called intelligent, it has to learn.

Machine learning provides a system the power to stimulate human learning. It focuses on gathering the data and using it automatically for developing its programs and to learn for themselves. [3, 4].

There are different techniques used by a machine for learning that is presented below:-

1. Supervised machine learning
2. Unsupervised machine learning [5]

Supervised machine learning is the task of providing a machine with the data that can help the machine to predict future instances. Machine algorithm analyses the data and produces a function used for mapping input for the observed output. [6]

Unsupervised machine learning uses a machine algorithm that draws a conclusion on the data that is not classified without any guidance or instructions. [7]

The learning for human emotion detection is enforced into a machine by extricating various features of speech, collecting and storing it in the form of a database. Further, the data is used by the machine algorithms for emotions detection and making required changes in the database automatically that can improve the accuracy of the machine. Hence making the machine better and better with use. The organization of paper inculcates as follows in section 3 discusses different techniques that are used for emotion detection. Section 4 consults with the application of speech recognition. Section 5 discepts about implementation description. Section 6 explains the feature extraction of speech and emotion detection. Section 7 tells about the conclusion of results that were observed.

2. TECHNIQUES FOR EMOTION DETECTION

Human emotions can be detected by various techniques such as Facial expression recognition, Words recognition and Speech recognition. Out of the three techniques, the present research work is focused on mechanisms for speech recognition. Speech

recognition technique is concentrated on converting human voice into a digital format that can be used by the computer to extract different features of speech. [8]

Several techniques are represented in figure 1.

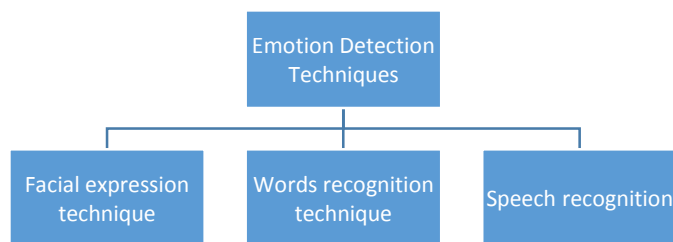


Fig. 1: Classification of Emotion Detection Techniques

2.1 Technique of Speaker Adaptation and Normalization

Different speakers in the world have different speeches that are easily recognizable by humans. This means that the human brain normalizes the speech processes. Speaker normalization creates a speaker space that is normalized and maps the human voice so that further differences can be minimized between the speakers. [8]

2.2 Words Positioning Technology

Keywords are the most important part of this technology. Keywords are related to the characteristics of language. The morpheme orientation has a great significance in sentence integrity because they outline sentence structure. Key morpheme acts as an iron structure for a building and increases sentence strength. Words positioning is the setting of keyword for the detection of a person's emotion through speech. [8]

2.3 Speech features extraction technology

Speech of human consists of many features like pitch, intensity, formats, frequency, and energy. The human voice is converted to digital signals by computer and these features are extracted through different software's. Features of speech changes with human emotions. Hence these features allow us to distinguish between the different human emotions. The research study presented in this paper further explicates in concept in a later section.

2.4 Related search technology

After acquiring the basic morpheme of the speech. Searches are made to find a relation with the morphemes and sense their group. Using this co-relation to meaning groups are determined that have rules know as grammar. Speech recognition uses grammar as rules to determine emotions. [8] The next section briefly discusses the applications of speech recognition.

3. APPLICATION OF SPEECH RECOGNITION

3.1 Application in Automobiles

While driving driver should not be distracted by the other operation like picking up phone calls, changing music, etc. To overcome such issues and to prevent accidents, speech recognition systems are being used to perform such tasks on commands given by the user.

3.2 Application in telephone communication

Voice dialing feature is present in every mobile phone that allows making hands-free calling.

3.3 Artificial assistant

Artificial assistance in the present is available on every mobile. This feature allows the user to make an appointment with doctors, type the messages and perform other activities. Siri is one of the examples of artificial assistance.

3.4 Application in Education

Speech recognition technology plays a great role in improving human speaking skills as it enables communication between human and a computer; consequently, computer programs find their mistakes in pronunciation.

The next section proceeds with the study of feature extraction from speech.

4. FEATURES EXTRACTION FOR EMOTION DETECTION

For recognition of emotion, it is imperative is to extract correct and meaningful features of speech. Human speech has two types of features, namely, linguistic features and acoustic features. The research study conducted here focuses on acoustic features [9]. Features that can be extracted and used for human detection are:-

- a) Pitch
- b) Intensity

4.1 Pitch

Pitch is the property of sound that helps us to detect how "low" or "high" the sound is related to musical melodies [9]. The two emotions Anger and Sadness have a direct linkage to sound as they are dependent upon the value for pitch i.e. Anger emotion is detected in case of low pitch and Sadness emotion is detected if the value of pitch is high.[10]

4.2 Intensity

The intensity of the sound is the power carried by a sound wave per unit area. The loudness of the sound depends on the intensity of sound. Higher the intensity louder the sound. Anger emotion is detected with high intensity and low pitch.

5. IMPLEMENTATION DESCRIPTION

The research study conducted here has obtained its results by verification of data practically. For the realization of feature extraction, some speech files inculcating the sentences depicting various kinds of emotions has been downloaded from the internet from the source presented in [11,12]. Subsequently, Praat software has been used for feature extraction for parameters Pitch and Intensity at specific points. (Praat is computer software for analysis of a speech that reads WAV files for extraction of features). Further, each feature was separately extracted and readings were noted. The generated results have been presented in the section below.

6. RESULTS

Emotion is the mental state of human being beings associated with feelings and thoughts. Emotion cannot be restricted with scientific definition as it varies from person to person.

Emotion is the best way to identify persons feeling. [13]. Based on the data collected and evaluated in the form of sound files on Praat software respectively, following values for different emotions were generated:

6.1 Anger

Anger is a vigorous emotion characterized by feelings of hatred, frustration, and conflict with others. When people are angry they tend to speak in a loud voice. As per data fed into the software, the values generated for Anger emotion are:

The maximum value of pitch 390.697 HZ

The minimum value of pitch 138.968 HZ

The maximum value of intensity 85.044 DB

The minimum value of intensity 22.289 DB

Based on the values, the data has been graphically presented in figure 2:

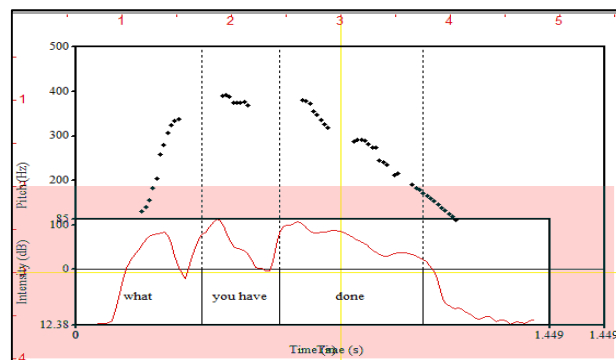


Fig. 2: Graph representing pitch and intensity of angry sound

6.2 Calmness

Calmness is a mental state of mind free from hatred, hostility, agitation, and revenge. The values generated are presented as :

The maximum value of pitch 448.949 HZ

The minimum value of pitch 104.903 HZ

The maximum value of intensity 64.57 DB

The minimum value of intensity 47.39 DB

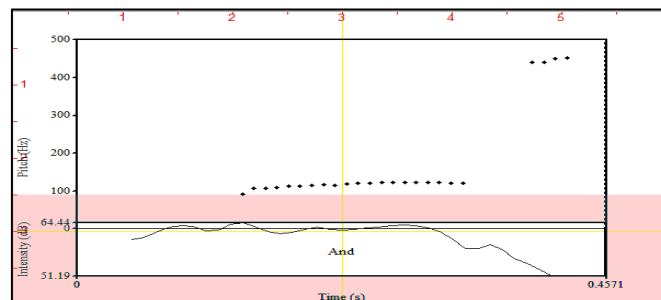


Fig. 3: Graph of pitch and intensity of calmness emotion

6.3 Surprise

The surprise is human emotion originally described by Eckman. The surprise is characterized by physiological astonish response. This emotion can be positive, negative or neutral. The values generated are presented as under:

The maximum value of pitch 320.87 HZ

The minimum value of pitch 126.043 HZ

The maximum value of intensity 70.109 DB

The minimum value of intensity 52.99 DB

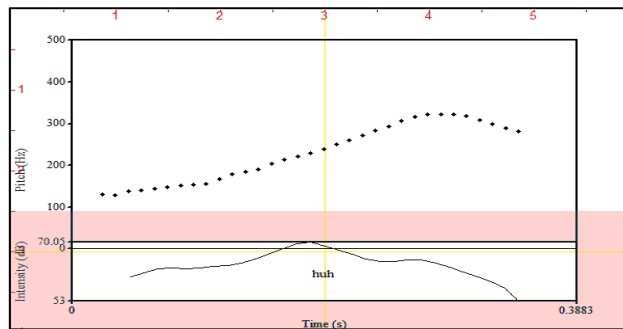


Fig. 4: Graph representing surprise emotion intensity and pitch

6.4 Sadness

Sadness is defined as a fugacious emotional state characterized by a feeling of disappointment, grief and dampened the mood. The values generated are presented as under:

The maximum value of pitch 486.53 HZ

The minimum value of pitch 97.49 HZ

The maximum value of intensity 68.33 DB

The minimum value of intensity 60.16 DB

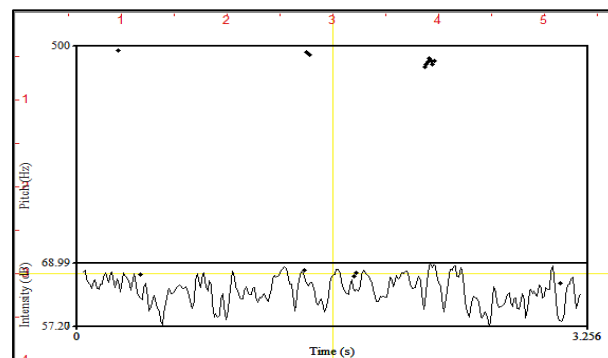


Fig. 5: Pitch and intensity graph for sadness

7. CONCLUSION

This paper explores methodologies for recognition of emotion, based on different speech features that can be used by different programmers for human-robot interaction. Prosodic features were taken into account like pitch and intensity for emotions, namely, sadness, angry, calmness and surprise using Praat Software.

The experimented technique is based on first extracting the pitch contour of the speech followed by the analysis of intensity. After experimentation, it was inferred that taking observations of calmness emotion as base, anger and surprise emotion has a lower pitch than that of base emotion. Besides, Sadness has a higher pitch.

The intensity of anger, the surprise is greater than that of calmness and sadness has the approximate same intensity as that of calmness.

One of the challenges faced through the research is the difficulty to create a worldwide database for emotional speech and observing the results for each.

8. REFERENCES

- [1] Youhao Yu, Research on Speech Recognition Technology and its Application, International Conference on Computer Science and Electronics Engineering, IEEE Computer Society, pp. 306-309, 2012.
- [2] https://en.wikipedia.org/wiki/Machine_learning
- [3] WANG Hua, ZHOU Lijuan, MA Cuiqin, A Brief Review of Machine Learning and its Application, IEEE Computer Society, pp 1-4.
- [4] <https://www.expertsystem.com/machine-learning-definition>
- [5] <https://www.expertsystem.com/machine-learning-definition>
- [6] https://en.wikipedia.org/wiki/Supervised_learning
- [7] <https://whatis.techtarget.com/definition/unsupervised-learning>
- [8] Youhao Yu, Research on Speech Recognition Technology and its Application, International Conference on Computer Science and Electronics Engineering, IEEE Computer Society, pp. 306-309, 2012.
- [9] Mohammad Rabiei , Alessandro Gasparetto, A Methodology for Recognition of Emotion-Based on Speech Analysis, for Applications to Human-Robot Interaction. An Explore Study, DOI 10.2478/pjbr-2014-0001 pp.1-11,2014
- [10] https://en.wikipedia.org/wiki/Emotional_prosody
- [11] <http://soundbible.com>
- [12] <https://www.soundsnap.com>
- [13] <https://en.wikipedia.org>