Journal of Forensic Sciences And Criminal Investigation
ISSN: 2476-1311

Juniper Publishers key to the Researchers

Research Article
Volume - 11 Issue 1 - October 2018
DOI: 10.19080/JFSCI.2018.11.555803

J Forensic Sci & Criminal Inves

Role of Acoustic Cues in Conveying Emotion in Speech

Jasdeep kaur[1], Kailash Juglan[1] and Vishal Sharma[2]

[1]*Department of physics, Lovely Professional University, India*

[2]*Institute of Forensic Science & Criminology (IFSC), Panjab University, India*

Submission: October 08, 2018; **Published:** October 17, 2018

***Corresponding author:** Kailash Juglan, Department of physics, Lovely Professional University, India, Email: kc.juglan@lpu.co.in

Abstract

Background: This paper presents role of acoustic cues for recognition of emotions in Punjabi language. An individual's emotional state as well as its linguistic configuration can be determined from his/her speech. From numerous experimental studies, we observed that for most Indian languages, both acoustic and articulatory facts played a crucial role for emotions recognition. To the best of our knowledge, this research will make a huge contribution in Punjabi language.

Methodology: A sample of 120 adult (58 males and 62 females) was taken for the experimental analysis. Speech sample consists of word tokens as each speaker repeated five Punjabi sentences with five different emotions. To extract Pitch and Intensity from the given sample, Recording and labelling of word tokens was done with help of PRAAT software.

Results: Through the scrutiny of these assessed parameters, the results are elucidated based upon contending patterns of pitch and intensity for different emotions. From experimental as well as graphical analysis, we conclude that value of pitch for normal emotions is highest in case of happiness (185.77 Hz) and lowest for sad (125.16 Hz). The value of intensity is highest in case of anger (91.8 dB) and lowest for fear (79.4 dB).

Conclusion: The study infers that in addition to vowel length, gap duration as well as voice onset time (VOT), pitch and intensity played a decisive role to detect emotions from speech sample. With this distinction we can also differentiate normal voice from disguised voice.

Keywords: Forensic science; Emotion recognition; Pitch; Intensity

Abbreviations: ASR: Automatic speech recognition; HMM: Hidden markov model; FSI: Forensic speaker identification; MMT: Micro muscle tremors; VSA: Voice stress analysis; STDEV: Standard deviation; GMM: Gaussian mixture model; SVM: Support vector machines; SFS: Sequential forward selection.

Introduction and Background

It is a well-known tendency that people generally disguised their voices to fleece criminology branch such that in extortion case, for kidnapping and even during emergency police calls. As the horizon of crime is increasing day by day so it is difficult to solve these cases by traditional methods. The Law should be at least one step before the crime, so forensic scientists are researching new discoveries every day. Every human being is different from other in terms of age, gender, weight and voice etc. so they response differently to different emotions. Principle of disguise defined as modification of voice of one person to sound differently or like another person [1-3]. For manual FSI, identification rate by normal voices can be degraded by the voice variations from great background noise, different transmission channels, illnesses, etc. Due to various applications of forensic science in our daily life, it has become strenuous to perceive the synergy among different emotions for a recognition system. Many of the behaviors felt or exhibited by humans can be detected by analysis of that person's voice [4-6]. These are:

a. Emotions (stress, anger, fear, sadness, depression, excitement, and happiness),

b. States induced by external conditions (ethanol intoxication and drugs),

c. Certain Intentional behaviors (deception and other intent), and

d. Health states (cold/flu and fatigue)

In the present study we focused on emotion recognition. Contemporary research in this field has covered the following major sectors as:

a) Emotional database.

b) Recognition of emotions.

c) Speech recognition.

Abundant studies have been carried out to determine acoustics cues for emotion recognition of various languages like Hindi [7], Arabic [8], Kannada [9], Tamil [10], Malayalam [11], Telugu [12], Bengali [13], Gujarati[14], and last but not least for Marathi [15]. With continuation to our work on voice stress analysis for detection of deception [16] in present study we have reported phonetic study for emotion recognition based on acoustic cues. Ample research has been done on segmentation of speech into smallest units [17-18] however the study for Punjabi language is still found in limited study. The eminence feature of the speech for recognizing emotion is to deal with prosody of speech which includes contours of pitch and intensity as well timing of utterance [19-21]. The rate of recognition of emotions from recognition systems has been raised from 70% to 90%. Speech signal gets affected by mechanical effects under a particular physiological constraint. As if nervous system gets excited then we can judge person is in fear or in mood of joy or anger, corresponding speech will be fast under high frequency range. On the other hand, BP and heart rate decreases for sad or bored emotion [22,23]. Looking at the scanty literature regarding detection of emotions in Punjabi language as compared to other languages, present study attempted to calculate pitch and intensity values for different sentences.

Materials and Methodology

Materials

Aim of our research is to study the effect of various suprasegmental parameters on emotions so that one can judge the values of pitch, intensity and formant frequency for different emotions. So to design the database, a sample of 120 adult (58 males and 62 females) was taken for the present experimental analysis. Out of which some are undergraduate while some are pursuing their master's degree having average age of about 20 years. The samples are natives of Punjab so they can read, write and understand Punjabi language well. To obtain an emotional database is a tedious job, because one should be very careful in order to get exact impression of happiness, sad, anger, fear and neutral emotions. We can also use the database that recorded earlier such as from movies, serials, telephone conversation and from recording of WhatsApp calling etc. As there is no ready to use database for such experimental analysis so we have to make database by ourselves however it has a limitation that speaker already knew that how to use this database into a particular emotion, this limitation can be overcome by putting a hidden microphone near them to analyze actual emotions. But this will become very cumbersome as we can have to wait for a very long time. So for present study we start with a prepared database, which were meaningful Punjabi sentences.

The most relevant factors in considering emotional database are:

a) Difference between acted and real world emotions.

b) Emotions are uttered by whom?

c) Unbalanced and balanced utterances.

d) Method of simulation of utterances.

From the inferences of various studies we can conclude that most of the databases share joy, sadness, anger, neutral [24]. The types of different emotional databases are:

I. **Type 1:** It is an acted emotional speech. They are obtained by examining an actor to speak with pre-defined emotions.

II. **Type 2:** These databases taken from real world examples. As natural human speech is unconstrained which produces all the emotions are real.

III. **Type 3:** This is computer-generated emotional speech, in which emotions are induced with self report. This elicited speech is neither neutral nor simulated.

Methodology

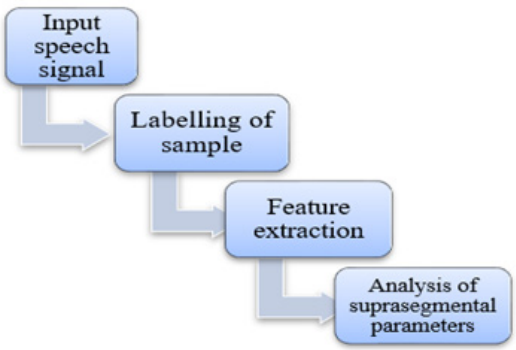

Figure 1: Scheme for Experimental Analysis.

Each speaker was asked to repeat the given sentences with different emotions (happiness, anger, fear. Sad, neutral). The sentences have been made in such a way that it can be spoken by speaker in all the emotions. These speech samples were recorded with a Sennheiser microphone with each emotion. This has a frequency range equivalent to human audible range with sensitivity of 110 dB. Recording was done with help of gold wave software. In which we have the facility of noise reduction as well as frequency modulation. To extract Pitch, Intensity and formants from the given sample, labelling of speech sample was done with help of PRAAT software. The schematic representation of proposed methodology is as following: Pitch can be defined as sound governed by the rate of vibrations producing. It is the fundamental frequency of vibration of vocal cords. It was calculated from the spectrograms of five different sentiments. Pitch is able to detect emotions from voice of person. So this suprasegmental parameter including intensity and formant frequencies were studied with help of PRAAT software.

How to cite this article: Jasdeep k, Kailash J, Vishal S. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci & Criminal Inves 2018; 11(1): 555803. DOI: 10.19080/JFSCI.2018.11.555803.

Variation in pitch values for different emotions shown in Table 1 for neutral, sad, fear, anger and happy emotions. Through the scrutiny of these assessed parameters, the results are elucidated based upon fluctuating values of pitch with time for different emotions (Figures 1 & 2).

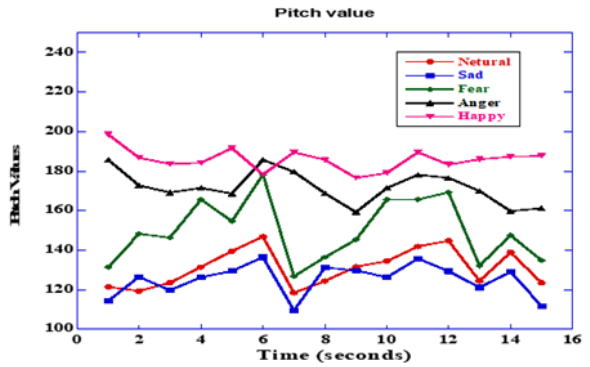

Figure 2: Pitch Values for Five Different Sentences with Five Emotions.

Table 1: Pitch values (neutral, sad, fear, anger and happiness).

speakers	Neutral	Sad	Fear	Anger	Happy
S 1	121.5	114.5	131.5	185.5	198.4
S 2	119.4	126.6	148.3	172.8	186.7
S 3	123.6	119.8	146.4	169.2	183.6
S 4	131.5	126.3	165.6	171.5	184.1
S 5	139.6	129.5	154.7	168.6	191.5
S6	146.8	136.4	178.3	185.7	178.2
S 7	118.5	109.6	126.8	179.5	189.4
S 8	124.4	131.3	136.4	168.8	185.7
S 9	131.6	129.8	145.3	159.2	176.6
S 10	134.5	126.3	165.6	171.5	179.1
S 11	141.9	135.8	165.7	178.1	189.5
S 12	144.8	129.4	169.3	176.6	183.2
S 13	124.5	121.2	132.1	169.9	185.9
S 14	138.9	129.1	147.5	159.8	187.3
S 15	123.5	111.8	134.9	161.2	187.6

From experimental as well as graphical analysis, we conclude that value of pitch for normal emotions is highest in case of happiness (185.77 Hz) and lowest for sad (125.16 Hz). Another important parameter is variance which measures how far each number in the set is from the mean.so that we may approximately hit the actual value for each emotion. For this we have calculated variance as well as standard deviation. A small variance (happy) indicates that pitch value tend to be very close to actual value and to each other however high value (fear) suggests that pitch values are far way spread from the mean (Figure 3). In addition to variance, we also tested the value of standard deviation. Standard deviation is a number which gives description about measurements spread out from mean. Following figure shows the variance and standard deviation for pitch values for each emotions. Figures 3 & 4 depicts that variance for fear emotion is very high for all speakers whereas sad emotion has very low value of variance. Since standard deviation was calculated by square root of variance, so consequently value of standard deviation for fear and sad emotion is high and low respectively.

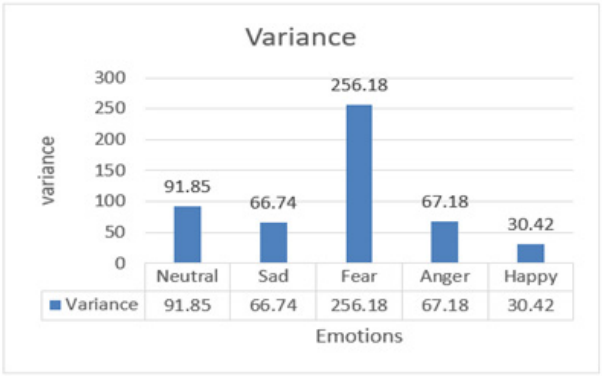

Figure 3: Variance for Neutral, Sadness, Fear, Anger and Happiness.

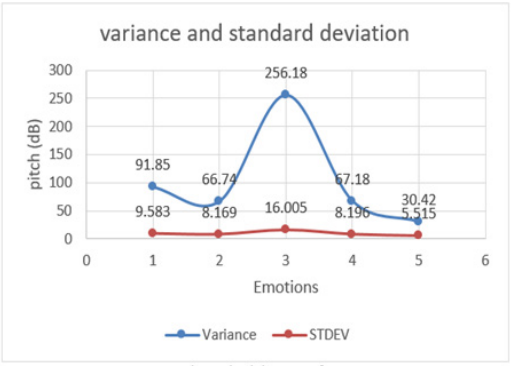

Figure 4: Variance and Standard Deviation for Emotions.

These results infer that when speaker try to speak any sentence with fear emotion they are having large variation of this emotion than normal (neutral) emotion. Heading towards next aim, that how pitch varies for each speaker in all emotions. We have taken mean of each sentence spoken by speaker in a particular emotion. Table 2 illustrates vale of pitch for different emotions (Figure 5). Since pitch is considered as frequency of vibration of vocal fold but intensity is measure of loudness of sound. So, to check effect of different emotions on intensity we measure the values of intensities for all speakers in average in terms of six sentences. Therefore, from above table we conclude that intensity is lowest in fear (79.4 dB) and attains highest value in anger (91.8 dB). These values of pitch and intensity are totally in line with previous studies. Figure 6 depicts that as intensity is measure of loudness so for anger emotion, intensity is highest (91.8 dB) and least for fear (79.4 dB). Analogous to pitch values, for intensity we also calculated the values of variance and standard deviation. From Figure 7, it can be concluded that both variance and standard deviation varies considerably for each emotion. Variance of intensity value for sad emotion is highest and so standard deviation is also attaining highest value.

How to cite this article: Jasdeep k, Kailash J, Vishal S. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci & Criminal Inves 2018; 11(1): 555803. DOI: 10.19080/JFSCI.2018.11.555803.

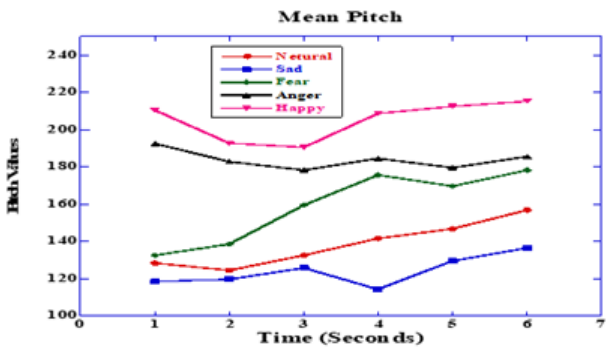

Figure 5: Pitch for Six Sentences.

Table 2: Pitch value for each sentence in different mode.

Sentences	Neutral	Sad	Fear	Anger	Happy
S1	128.3	118.5	132.5	192.5	210.4
S2	124.4	119.6	138.5	182.8	192.7
S3	132.6	125.8	159.5	178.2	190.5
S4	141.5	114.3	175.6	184.5	208.7
S5	146.8	129.5	169.7	179.6	212.5
S6	156.8	136.5	178.3	185.6	215.2

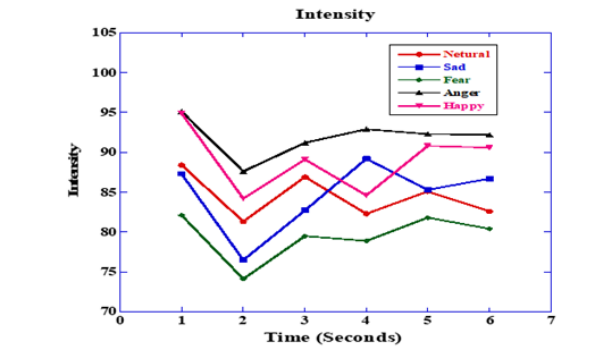

Figure 6: Intensity for Different Emotions.

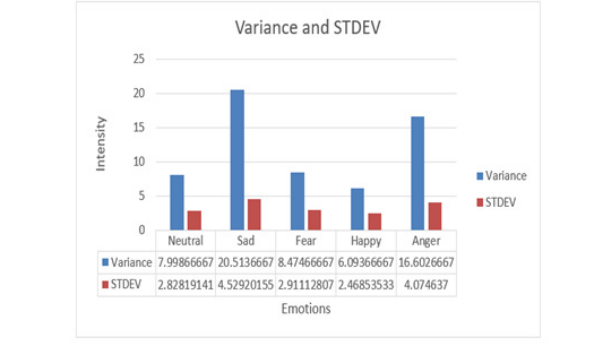

Figure 7: Variance and Standard Deviation for Intensities Corresponds to Different Emotions.

Discussion

Navas et al. [25] identified the emotions in Basque language in which emotion recognition was done by GMM. They made use of various prosodic features but later on they have chosen best six features. Kao and lee [26] have discussed pitch and power based feature for recognition of emotions in mandarin language by making use of SVM. They have used pitch and duration for recognition on basis of each syllable and word levels which yield 95% emotion recognition performance. Zhu [27] with help of SFS studied energy, duration and pitch based features for mandarin language. In which modular neural networks act as classifiers. With use of neural networks, Iliou [28] was successfully classified seven emotions of berlin emotional speech corpus. In which cases were considered as speaker independent cases. In order to assess the significance of energy, duration and pitch for emotion recognition, artificial neural networks were used. From review of literature [29] about present study, we can observe that in most of the cases prosodic features were dominated at utterance level such as minimum, mean, maximum and the standard deviation.

All these investigations with help of GMM, SVM and SFS and many more models points toward recognition based on emotions so, it is need of hour to scrutinize various parameters to analyze emotion recognition in Indian languages particular Punjabi, as best of my knowledge very few works has been done for Punjabi language. So, this motivate us to study suprasegmental parameters for Punjabi language. In addition to it, it is very important to study the contribution of each parameters for prosodic features extracted from syllables, word and hence from sentences. For present study, first we discussed types of databases which are emotionally strong. These sentences were chosen in such a way that may represent each emotion in proper way. Because for emotion recognition there is no ready to be database, so we have to make it manually corresponds to our requirement. So, to design the database, a sample of 120 adult (58 males and 62 females) was taken for the present experimental analysis. Out of which some are undergraduate while some are pursuing their master's degree having average age of about 20 years. The samples are natives of Punjab, so they can read, write and understand Punjabi language well. The reason behind taken of college going people is that it's easier to record their voices on repetition because this is not a single time activity, we may have to record over 10 times of a single speaker for proper course of action.

Furthermore, present study comprises pitch, intensity as a suprasegmental study for both male and female (Tables 1 & 3). One can make comparison for these two as well the relation of age with prosodic feature can also be studied. In addition to it, Value of formant frequency and speech rate for our native language is an open question for all. From graphical representation it is clear that there are different values of pitch and intensities for different emotions. Further there are another parameter which we studied is variance (Figure 3) and standard deviation to evaluate statistical analysis. Which measures how far each number in the set is from the mean.so that we may approximately hit the actual value for each emotion. A small variance (happy) indicates that pitch value tend to be very close to actual value and to each other

How to cite this article: Jasdeep k, Kailash J, Vishal S. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci & Criminal Inves 2018; 11(1): 555803. DOI: 10.19080/JFSCI.2018.11.555803.

however high value (fear) suggests that pitch values are far way spread from the mean. In addition to variance, we also tested the value of standard deviation. Standard deviation is a number which gives description about measurements spread out from mean.

Table 3: Values of Intensity for different emotions.

Sentences	Neutral	Sad	Fear	Anger	Happy
S1	88.4	87.3	82.1	95.1	94.9
S2	81.3	76.5	74.1	87.6	84.2
S3	86.9	82.7	79.5	91.2	89.1
S4	82.3	89.2	78.9	92.9	84.6
S5	85.1	85.3	81.8	92.3	90.8
S6	82.6	86.7	80.4	92.2	90.6

Conclusion

In the present study, after recording of speech signal and stages of processing of signal we extract various suprasegmental parameters. We have measured pitch values for single sentence in different emotions, for every different words of a single sentence and last but not least mean pitch value. In addition to pitch value we have also measured value of intensity. From measurements we can conclude that value of pitch for normal emotions as is highest in case of happiness and lowest for sad. Following trend followed by various emotions: Happy (185.77 Hz) > Anger (171.86 Hz)> Fear (149.8Hz)> Neutral (131 Hz)> Sad (125.16 Hz) The value of intensity for normal emotions is highest in case of anger and lowest for fear. Following trend followed by various emotions: Anger (91.8 dB) > Happy (89.03 dB) > Sad (84.61 dB) > Neutral (84.4 dB) > > Fear (79.4 dB)

So, we can conclude that in order to detect emotions from speech, suprasegmental parameters plays a crucial role. As the value of pitch and intensity varies for each emotion. It attains different value for happiness, anger, fear, neutral and sad emotions in Punjabi language.

References

1. Banse R, Sherer KR (1996) Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology 70(3): 614-636.
2. Burkhardt F, Sendlmeier W (2000) Verification of acoustical correlates of emotional speech using formant-synthesis. In Proceedings of the ISCA Workshop on Speech and Emotion.
3. Dua M, Kaur P, Saini P (2013) International journal of engineering trends and technology (IJETT): 44-52.
4. RL Diehl (1991) Phonetics 120-134.
5. Rabiner L, Juang B Fundamentals of speech recognition, ISBN-0130151572.
6. Siniscalchi SM, Reed J, Svendsen T, Lee CH (2013) Universal attribute characterization of spoken languages for automatic language recognition. Computer speech and language 27(1): 209-227.
7. Srivastava, Singh N, Vaish S (2013) Speech recognition for Hindi language. International journal of Engineering research and technology 2(4).
8. Walha R, Drira F, El Abed H, Alimi AM (2012) On developing an automatic speech recognition system for standard Arabic language. International journal of electrical and computer engineering 6(10).
9. Hegde S, Achary KK, Shetty S (2012) Isolated word recognition for kannada language using support vector machine. Springer-Verlag Berlin Heidelberg : 262-269.
10. Dharun VS, Karnan M (2012) Voice and speech recognition for Tamil words and numerals". International journal of modern engineering research (IJMER) 2(5): 3406-3414.
11. Kurian C, Balakrishnan K (2011) Automated transcription system for Malayalam language. International journal of computer applications 19(5).
12. Bhaskar PV, Rama Mohan Rao S, Gopi A (2012) HTK based Telugu speech recognition. International journal of advanced research in computer science and software engineering: 2(12).
13. Choudhury FN, Shamma TM, Rafiq U, shuvo HR (2016) Development of Bengali automatic speech recognizer and analysis of error pattern. International journal of scientific & engineering research 7(11).
14. Mistry KK, Shah SS (2016) Comparative study on feature extraction methods in speech recognition. International journal of innovative research in science, engineering and technology 5(10).
15. Jadhav SB, Ghorphade J, Yeolekar R (2014) Speech recognition in Marathi language on android o.s.", International journal of research in computer and communication technology 3(8).
16. Kaur J, Juglan K, Sharma V (2006) AIP Conf. Proc. 030022-1–030022-7.
17. Singh P, Lehal GS (2011) Text to speech synthesis system for Punjabi language, Communications in computer and information science, springer berlin 139: 302-303.
18. Kaur H, Bhatia R (2015) Speech recognition system for Punjabi language. International journal of advanced research in computer science and software engineering 5(8).
19. Hussain Q, Proctor M, Harvey M, Demuth K (2017) Acoustic characteristics of Punjabi retroflex and dental stops. Journal of acoustical society of America: 141(6): 4522-4542.
20. Bisio I, Lavagetto F, Marchese M, Sciarrone A (2013) Gender driven emotion recognition through speech signals in ambient intelligence applications. IEEE transactions on emerging topics in computing 1(2): 244-257.
21. Cheveign AD, Kawahara H (2002) YIN, A fundamental frequency estimator for speech and music. The Journal of acoustical society of America 111(4): 1917-1930.
22. Oudeyer P (2003) The production and recognition of emotions in speech: features and algorithm", Int. J. Human-computer studies 59(1): 157-183.
23. Rao KS, Koolagudi SG (2013) Robust emotion recognition using spectral and prosodic features. Springer New York Heidelberg Dordrecht London.
24. Scherer KR 2003) Vocal communication of emotion: A review of research paradigms". Speech Communication 40(1-2): 227-256.
25. Navas E, Hern'aez I, Luengo I (2005) Automatic emotion recognition using prosodic parameters. INTERSPEECH (Portugal), pp. 493-496.
26. Hao Kao Y, shan lee L (2006) Feature analysis for emotion recognition from mandarin speech considering the special characteristics of Chinese language. INTERSPEECJ-ICSLP, (Pittsburgh, Pennysylvania), pp. 1814-1817.
27. Zhu, Luo Q, (2007) Study on speech emotion recognition system in E-learning, Human computer interaction. Springer Verlag, pp. 544-552.

How to cite this article: Jasdeep k, Kailash J, Vishal S. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci & Criminal Inves 2018; 11(1): 555803. DOI: 10.19080/JFSCI.2018.11.555803.

28. Iliou T, Anagnostopoulos CN (2009) Statistical evaluation of speech features for emotion recognition. Digital telecommunication, pp. 121-126.

29. Murray IR, Amott JL, Rohwer EA (1996) Emotional stress in synthetic speech: Progress and future directions. Speech Communication 20(1-2): 86-91.

DOI: 10.19080/JFSCI.2018.11.555803

How to cite this article: Jasdeep k, Kailash J, Vishal S. Role of Acoustic Cues in Conveying Emotion in Speech. J Forensic Sci & Criminal Inves 2018; 11(1): 555803. DOI: 10.19080/JFSCI.2018.11.555803.