Grenze International Journal of Computer Theory and Engineering , June 2017

Acoustic and Spectral Analysis of Kannada Emotional Speech

Geethashree A[1] and Dr. D.J Ravi[2]
[1]Associate. Professor, Vidyavardhaka college of engineering, Mysore
Visvesvaraya Technological University, Belagavi
Email: geethashree.a@vvce.ac.in
[2]Professor, Vidyavardhaka college of engineering, Mysore
Visvesvaraya Technological University, Belagavi
Email: ravidj@vvce.ac.in

***Abstract*—Emotion plays a significant role in communication. This work deals with an analysis of the sets of Kannada emotional sentences uttered by two speakers (one Male and one female). The data set used in this analysis is from KES (Kannada Emotional Speech) database, which consists of neutral, happy, sadness, anger and fear emotional sentences. The emotional analysis of acoustic features on KES database is carried out to build an effective emotion recognition and emotion conversion system. The features used of analysis of emotional speech are pitch, duriation, intensity, formats and spectral energy at sentence level. Results shown that prosodic and spectral parameters play significant role in expression of emotions. The acoustic values are in correspondence with the general knowledge for male and female voices. The analysis result obtained will be used for emotional speech conversion.**

***Index Terms*— KES, Spectral energy and acoustic.**

I. INTRODUCTION

One of the natural form of communication amoung human being is spoken language. Many work has been done on emotion conversion and emotion recognition in English, Hindi and other languages [8,10,11,14,17]. As India is mutilingual country, there is need to study the emotional aspects in Kannada speech. The KES database is developed in Kannada for analyzing the emotions present in speech. The review on analysis of emotional speech in different language is found in [6,18]. Natural speech carry two types of information, i.e. linguistic information and parelinguistic information such as speakers emotional state.

This work presents the analysis of acoustic parameters of emotional speech in Kannada language using KES database. The change of acoustic paramerters, prosody correlates such as speech rate, intensity difference, pitch contours and pausing structure, will communicate different emotion through voice [2,3]. Formant parameters [2], spectral energy distribution [3], and spectral noise [2] are significant speech correlates of emotion, which can be analysised by spectral analysis. As emotional expressions is gender-dependent [4], the gender features are also taken into consideration.

The analysis of acoustic and spectral parameters of emotional speech in Kannada language using KES database is dicussed in the paper. Section II describes the details of KES database and sentences used for analysis.The details of acoustic and spectral analysis is described in Section III and IV respectively. The result are presented and discussed in section V. Finally conclusions are stated.

Grenze ID: 01.GIJCTE.3.2.7`

II. KANNADA EMOTIONAL SPEECH (KES) DATABASE

An emotionally rich KES database is used for anlysis. The database consist of 4 primary emotion happy, sadness, anger and fear with neutral, uttered by two native actors one male and one female. Both the speakers were given 20 emotional kannada sentences, which can be reproduced in all the 4 emotions. Each speakers was asked to record the sentences with full emotion in an acoustic room with sampling rate of 44.1KHz and 16-bit precision with mono channel. The sentences used for emotional analysis are shown in Table 1. The sentences used for the analysis of acoustic parameters are labled as S1, S2 and S3.

TABLE I. SENTENCES FOR KES DATABASE USED FOR ACOUSTIC ANALYSIS

Lable	KANNADA (English)
S1	ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು? I realized, what is the reason for your fear?
S2	ಜೀವ ಜಗತ್ತು ವಿಸ್ತಾರವಾದದ್ದು Bio diversity is vast.
S3	ನೀವೊಬ್ಬರೆ ನನ್ನನ್ನು ಒಬ್ಬ ಮನುಷ್ಯನಂತೆ ಕಂಡು ವತರ್ಸಿದವರು You are the only one, who treated me like human being.

III. ACOUSTIC ANALYSIS

To achive the goal of emotion conversion and emotion recognition, the correlation of emotion in acoustic characteristics of speech signal has to be investigated. Although many research has been done in this area [1,3,5,9,15,16], not much is done in kannada language. From the litrature of emotion recognition, the most crucial aspects are those related to prosody: the pitch (or F0) contour, the intensity contour, and the timing of utterances. Prosody is the stress, rhythm and intonation of speech. The combination of duration of speech segment, variation of energy during speech and pitch of speech is Prosody. They give added sense to the spoken words which has significance in natural speech. The prosodic features used for analysing the emotional speech are intonation, pitch, intensity, duriation and formats.

A. Pitch

The pitch signal has information about emotion, as it depends on the tension of the vocal folds and the subglottal air pressure. The pitch signal is produced from the vibration of the vocal folds. The time elapsed between two successive vocal fold openings is called pitch period T, while the vibration rate of the vocal folds is the fundamental frequency of the phonation F0 or pitch frequency. The pitch of female voice ranges from 150-300, male voice ranges from 50- 200 and for child voice ranges from 200-400. For the extraction of Pitch values from speech PRAAT tool is used.

B. Duriation

It is one of the prosody parameter of speech segments which as immense significance in production of emotional speech. The duration of Phoneme are influenced by phones, phonetic identity and level of stress or position in the word or in the sentence. Word duration also have massive importance in the sentence.

C. Intensity

Intensity refers to the amount of energy flowing per unit time through unit area perpendicular to the direction of propagation. Few studies on the use of intensity in speech prosody exist [13]. The intensity pattern is highly related with fundamental frequency, it goes up in proportion to fundamental frequency [1]. [14] found noticeable increased energy envelop in anger speech. Happiness showed similar characteristics reported by [6]. Sadness was associated with decreased intensity. [15, 16].

D. Formants

Formants are resonance frequencies producted in the vocal tract and changes their center frequency and bandwidth during speech. [4] explained anger produced vowels with more open vocal tract and form greater mean of first format frequency than neutral speech.

IV. Spectral Analysis of speech

The purpose of spectral analysis is to find how acoustic energy is distributed across frequency. There are two methods for spectral analysis: fast Fourier transform (FFT) and Linear prediction(LPC). FFT finds the energy distribution in the actual speech sound, whereas LPC estimates the vocal tract filter that shaped that speech.
The spectral features are obtained by converting the time domain signal into the frequency domain using fourier transforms. Modifying prosodic features like pitch contour modifies only the fundamental frequency. It does not change the spectral shape of the signal, which may also play a major role in emotion perception. Therefore, in this work, we analyze how spectral shape affects emotions.
The time domain speech signal x[n] is converted to frequency domain speech signal X(k) in MATLAB uses a fast algorithm called the fast Fourier transform (FFT). X(k) is complex, and hence we plot both its magnitude and phase with respect to k, or ω. The magnitude plot gives us information about the energy content of different frequencies. This is also called magnitude spectrum. The square of this FFT magnitude, $|X(\omega)|^2$ when plotted against ω, gives the energy density spectrum (also called spectral energy).

V. Results and Discussion

For analysis of procidic parameters PRAAT software is been used and for spectral analysis MATLAB as been used. As 5 emotions are used for analysis, each emotion is represented by the starting letter (N-Neutral, H- Happy, S-Sadness, A-Anger and F-Fear).

A. Pitch

The Pitch values are extrated using PRAAT with the step size of 0.01 second and floor and ceiling pitch frquency of 75 – 500 Hz. The pitch values of the three sentences listed in Table 1 are extracted for both the speakers (male and Female).
The mean and standard deviation of pitch is extracted and tabulated in Table 2 and Table 3 respectivly.

TABLE II. Mean Pitch values of Male and Female Speckers for Different Emotions

Mean Pitch (Hz)						
		N	**H**	**S**	**A**	**F**
Female	**S1**	221	286	256	412	276
	S2	217	238	289	303	284
	S3	201	241	275	470	263
Male	**S1**	112.3	178.5	105.3	207.4	124.2
	S2	114.3	171.2	108.7	195.7	118.5
	S3	125.3	154.3	105.1	208.5	124.5

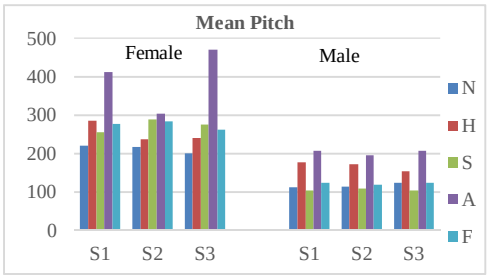

Fig. 1 Bar graph of Mean Pitch values of Male and Female speaker in different emotions

Fig 1 bar graph of mean pitch values of both the speakers for different emotion clearly indicate that the pitch values of female is greater than male. The variation of pitch pattern is dependent on the emotion expressed by the speaker. The mean pitch of anger is more when compared to all other emotional sentences for both female and male speaker. The mean pitch value of happy is higher than sadness, neutral and fear in female speaker, where as a significant difference is observed with male speaker. For sadness and fear emotion, the mean pitch value is almost same for all the three sentences in both the speakers. Anger is characterized by high mean pitch in contrast to sadness and fear which yields lower mean pitch.

Fig 2 bar graph of standard deviation of pitch values in both the speaker for different emotion clearly show that Anger in male speaker and happy in female speaker has higher standard deviation and wide range. Sadness and fear is having the lower standard deviation in both the speakers.

TABLE III. STANDARD DEVIATION OF PITCH VALUES IN MALE AND FEMALE SPECKERS FOR DIFFERENT EMOTIONS

Standard Deviation of Pitch(Hz)						
		N	H	S	A	F
Female	S1	28	36	30	91	27
	S2	34	53	33	65	45
	S3	30	29	30	55	32
Male	S1	15.7	69.9	7.6	41.5	15.8
	S2	20.6	30.8	11.3	28.8	11
	S3	18.5	20.4	8.9	28.5	9.9

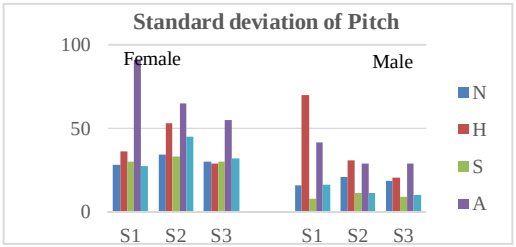

Fig. 2 Bar graph of standard Deviation of pitch values for Male and Female Speaker in different emotions

B. Duration

Analysis of duration can be performed at different levels from segment to sentence level, speaking rate and rhythm. Studies in other languages [9] noted anger and happiness have higher articulation rate, sadness and fear with reduced speaking rate. Table 4 and fig 3 show that duration of sadness and fear is more as compared to anger and neutral for both speakers.

TABLE IV. DURATION OF MALE AND FEMALE SPECKERS IN DIFFERENT EMOTIONS FOR DIFFERENT SENTENCES

Duration						
		N	**H**	**S**	**A**	**F**
Female	**S1**	2.5	2.8	2.9	2.6	2.9
	S2	1.7	2.1	2.3	1.7	2.5
	S3	2.9	2.8	3.3	2.5	3.3
Male	**S1**	2.4	2.5	2.7	2.1	3
	S2	2.1	1.8	2.2	1.6	1.9
	S3	2.7	2.8	3	2.2	3

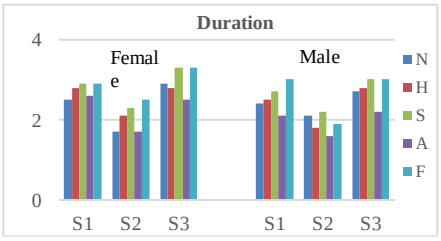

Fig. 3 Bar graph of Duration of male and Female Speaker in different emotion

C. Intensity

Table 5 and Fig 4 show the characteristics of intensity pattern for three different sentences for different emotions. The intensity of male speaker is high compared to female speaker. Happy and anger are

characterized by high intensity, where as sadness, fear and neutral is characterized by low intensity in male speaker. Anger and fear is characterized by high intensity with neutral having low intensity for female speaker.

TABLE V. MEAN INTENSITY OF MALE AND FEMALE SPECKERS IN DIFFERENT EMOTION

Mean Intensity						
		N	**H**	**S**	**A**	**F**
Female	S1	76.5	77	78.7	77.8	76.7
	S2	72.5	73.4	72.8	76.2	75
	S3	75.4	76.1	76.1	78.4	78.6
Male	S1	76.5	80.7	76.6	79.2	76.6
	S2	76.4	79.9	76	78.7	77.3
	S3	79.8	81.2	78.3	80.4	77

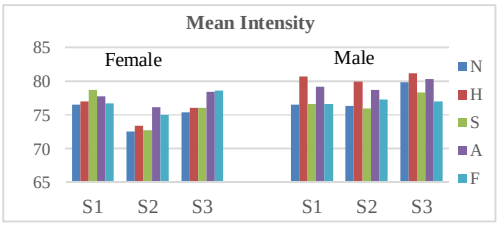

Fig. 4 Bar graph of Mean Intensity of Male and Female Speaker for different emotions

D. Pitch and intensity contours

In linguistics, the pitch counter of a sound is a curve that tracks the observed pitch of the sound over time. Many studies in different language have clamed that specific intonation patterns reflect specific emotions[9]. In this paper the intonation pattern is analysis for different emotions for kannada language, as not much work is done in this area. Fig 5 shows the variation of Pitch and intensity contour of female and male speaker for different emotions, where (A - Neutral, B – Happy, C – Sadness, D – Anger, E - Fear). [1] have found that intensity contributes to emotion production, and it should not be simply taken as a control paramerter. Numerous studies as shown that sound intensity is one of the important parameter to convey emotion. The pitch contour of neutral, sadness and fear are almost flat, when compared to happy and anger. In happy the pitch contour rises at the end of the sentence. There is a deep rise-fall, rise-fall contour in anger.

E. Formant

Table 6 and Fig 6 shows the first 3 formant frequency of the first sentence in Table 1 for both male and female spearkes. The anger and fear sentence has greater mean first, second and third formant frequncy than neutral sentence in both speaker.
Happy and sadness sentences has least mean first, second and third formant frequency.

F. Spectral Analysis

The spectral shape of the signal also play a major role in emotion. The analysis is performed on 10 emotional sentences. The spectrum of the entire semtence is calculated. Fig. 7 shows the magnitude spectra of a sentence for five different emotions. The spectrum of neutral speech is quite flat compared to the spectra of happy, sad and anger. The spectrum of neutral and fear is almost similar. The energy is negligible at high frequencies for all the emotions. Sad speech has higher magnitude and energy in low frequency than other emotional speech. Happy and anger has lower magnitude and energy in low frequncy when compared to other emotions. For frequencies higher than 1000 Hz, sad and fear speech seems to be quite same. Frequencies lower than 800 Hz may be useful to distinguish between sad, happy and fear as sad speech contains more energy followed by fear and happy in this range. Anger has a significant magnitude and energy component above 1000 Hz frequency. Almost similar patterns was observed for other 10 sentences that was analyzed for each emotion. This gives us a slight intuition that the frequency components in the lower frequency bands may contain some emotional cues. Fig 8 shows the energy Density Specutrum of male speaker for the sentence 'ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?'

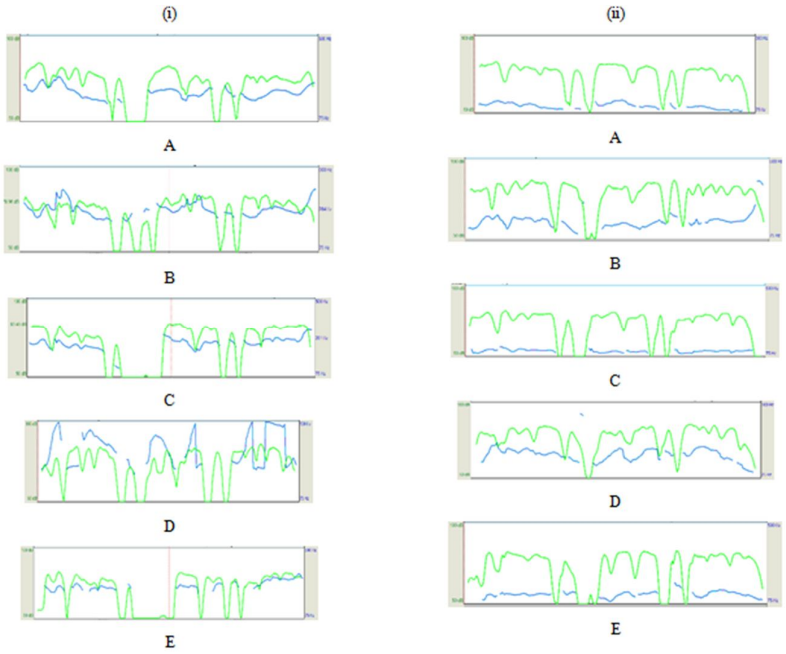

Fig. 5 Pitch and Intensity contour of sentence in Kannada "*ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?*".

(i) Female Speaker (ii) Male Speaker

*where (A - Neutral, B – Happy, C – Sadness, D – Anger, E - Fear)

TABLE VI. MEAN FORMANTS OF F1, F2 & F3 OF MALE AND FEMALE SPEAKER FOR THE SENTENCE "ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?"

Mean Formants						
		N	**H**	**S**	**A**	**F**
Female	**F1**	572	696	554	913	750
	F2	1804	1781	1772	1956	1936
	F3	2929	2854	2929	3012	3086
Male	**F1**	631	615	717	809	836
	F2	1811	1711	1907	1911	2080
	F3	3121	2780	3200	2943	3199

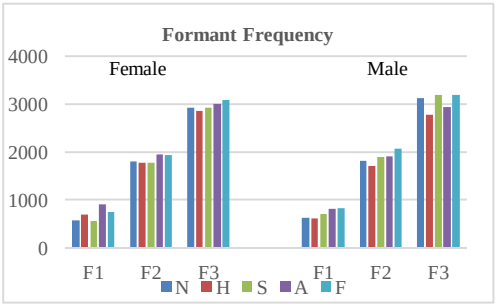

Fig. 6 Bar graph of Mean Formant of F1, F2 and F3 of Male and Female Speaker for the sentence "ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?"

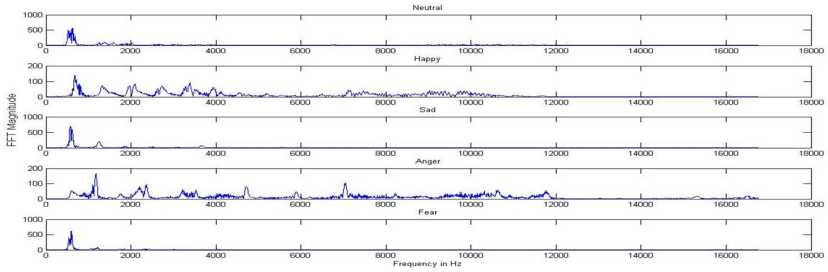

Fig. 7 Magnitude spectra of the sentence "ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?" in different emotion (Male speaker)

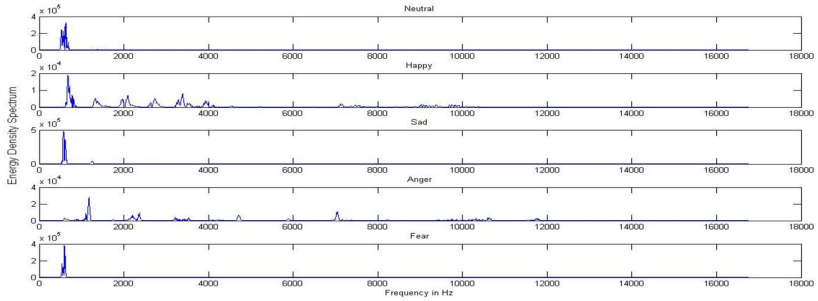

Fig. 8 Spectral Energy of the sentence "ಅನುಗ್ರಹವಾಯಿತು, ನಿಮ್ಮ ಭಯಕ್ಕೆ ಕಾರಣವೇನು?" in different emotion (Male speaker)

VI. CONCULSION

The acoustic and spectral parameters of speech associated with two speaker (one male and one female) with human emotions of Happy, Sadness, Anger and fear are analysed in relation to neutral speech. The study indicates that in emotional sentences in state

1. Happy: "Mean pitch and standard deviation of pitch increases, duriation and intensity increases slightly, formant F1 varies randomly, F2 and F3 decreases, spectral magnitude and energy density funtion is very much less than neutral".
2. Sadness: "Mean pitch and standard deviation in female speaker increases slightly and in male speaker it decreases, duriation of sadness increases, intensity is almost same, formants F1 and F2 in female decreases and in male it increases, F3 is almost same, spectrum magnitude and energy density function is greater than neutral".
3. Anger: "Mean pitch, standard deviation and intensity increases by a large value, duriation decreases, formants F1 and F2 increases, F3 in female decreses and in male increases, spectral magnitude and energy density funtion is very much lesser and concentration of the spectrum is slightly in higher frequency (approximately around 1KHz) than neutral".
4. Fear: "Mean pitch, duriation and intensity is slightly higher, standard deviation of pitch slightly increases in female speaker and slightly decreases in male speaker, formants F1 and F2 increases, F3 slightly increases, spectral magnitude and energy density funtion is lesser than sadness but greater than neutral".

The results of our analysis cannot be generalised as we have analysed sentences uttered by only one male and one female speaker.

REFERENCES

[1]. Noad, J.E.H., Whiteside, S.P. and Green, P.D. "A macroscopic analysis of an emotional corpus", Eurospeech'97, pp. 517-520, 1997

[2]. K. R. Scherer, "Vocal communication of emotion: A review of research paradigms," Speech communication, vol. 40, no. 1-2, pp. 227–256, 2003.

[3]. R. Banse and K. R. Scherer, “Acoustic profiles in vocal emotion expression.” Journal of personality and social psychology, vol. 70, no. 3, pp. 614–36, March 1996
[4]. J. Pribil1, A. Pribilova, J. Matousek, “Comparison of Formant Features of Male and Female Emotional Speech in Czech and Slovak”, elektronika ir elektrotechnika, issn 1392-1215, VOL. 19, NO. 8, 2013
[5]. V. Auberge, N. Audibert, and A. Rilliard, “Acoustic morphology of expressive speech: What about contours?”, Proceedings of Speech Prosody 2004.
[6]. Murray, I.R. and Arnott, J.L. "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", JASA Vol. 93(2), pp. 1097-1108, 1996
[7]. S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, IITKGP-SESC: Speech Database for Emotion Analysis. Communications in Computer and Information Science, JIIT University, Noida, India: Springer, issn: 1865-0929 ed., August 17-19 2009.
[8]. Kramer E., (1963). “Judgment of Personal Characteristics and Emotions from Nonverbal Properties of Speech”. Psychological Bulletin, 60.
[9]. Vroomen, J., Collier, R. and Mozziconacci, S. “Duration and intonation in emotional speech", Eurospeech 93, pp. 577-580, 1993
[10].R.Cowie., E.Douglas-cowie.,N.Tsapatsoulis, G.Votsis, S.Kollias, W.Fellenz, & J.Taylar, “Emotion recognition in human-computer interaction”, IEEE Signal Processing Magazine, 2001
[11].Takashi & Norman D. Cook, “Identifying Emotion in Speech Prosody Using Acoustical Cues of Harmony”, INTERSPEECH, ISCA, DBLP (2004).
[12].Sendlmeier, W.F., Kienast M. and Paeschke, A. “F0 contours in Emotional Speech.” Technische University, Berlin, Proc. ICPhS, 1999.
[13].Mozziconacci, S.J.L and Hermes D.J. “Role of Intonational Patterns in Conveying Emotion in Speech.” ICPhS 1999, 1999 - Citeseer.
[14].Rong J, Li G, Chen Y-P P. “Acoustic feature selection for automatic emotion recognition from speech”. J InfProcManag, 2009.
[15].D.J.Ravi and Sudarshan Patilkulkarni, “Kannada Text to Speech Synthesis Systems: Emotion Analysis” international conference on natural language processing (ICON-2009).
[16].Sushma Bahuguna1, Y. P. Raiwani. “A Study Of Acoustic Features Pattern Of Emotion Expression For Hindi Speech” international journal of computer engineering & technology (ijcet) measurement science review, Volume 10, No. 3, 201072.
[17].Shashidhar G. Koolagud, RaoSreenivasaKrothapalli “Two stage emotion recognition based on speaking rate” Received: 16 November 2010 / Accepted: 2 December 2010 / Published online: 11 December 2010 © Springer Science+Business Media, LLC 2010.
[18].Shashidhar G. Koolagudi, K. SreenivasaRao “Emotion recognition from speech: a review” Received: 7 July 2011 / Accepted: 17 December 2011 / Published online: 4 January 2012 © Springer Science+Business Media, LLC 2011.
[19].Syed Abbas Ali, SitwatZehar, Mohsin Khan & Faisal Wahab, “Development and Analysis of Speech Emotion Corpus using Prosodic Features for Cross Linguistics” International Journal of Scientific & Engineering Research, vol-4, Issue 1, Jan-2013, ISSN 2229-5518.
[20].Pallavi J, Geethashree A,”Emotion Analysis and Evaluation of Kannada Speech Database,” IJECET, vol 5, issue 8. August 2014, pp. 160-170, 2001
[21].http://www.fon.hum.uva.nl/praat/manual.html.