Evaluation of the Impact of Emotion in Assamese Emotional Speech Signals

Somi Kolita
Research scholar, School of Computing Sciences ,The Assam Kaziranga, University, Jorhat, Assam
Purnendu Bikash Acharjee
Assistant Professor, School of Computing Sciences ,The Assam Kaziranga, University, Jorhat, Assam

Abstract— Speech is a kind of communicational source that allows people to exchange information. Emotional speech is another valuable means through which people learn to convey their inner sentiments. Emotion is induced by certain circumstances, and emotion identification from human speech is regarded as a critical subject in the area of human-computer interfaces (HCIs). This study presents the findings of a fundamental analysis of prosodic features, as well as a comparison of prosodic features of different emotions in Assamese emotional speech signals on the basis of auditory perceptions through speakers of different gender. This research analyzes the pitch and formant of different informants in different emotions such as angry and normal, using two different software with reference to Assamese speech signals for observing the variation between different genders of speakers using Two-way ANOVA testing. For the analysis, have used nearly 300 data samples of Assamese speech signals recorded by 10 equal of male and female speakers of age group 18-30years. The major goal of this study is to evaluate the emotional effect with reference to sex and software influence extracted features from Assamese emotive speech of various genders in a variety of settings. The results of ANOVA testing show that emotion and sex have a substantial impact on both retrieved features in recorded voices using different software.

Keywords—Speech; Emotion; Prosodic features; Analysis Of Variance;

1. Introduction

Speech is a natural communicational mode between individuals. Exchanging information trough speech perceptibly plays a extremely major role in our daily lives auditory changes which are created contained by the spoken speech express a variety of emotions of the speaker. One most important task of the analysis of emotional speech signal is to calculate the prosodic features to suggest the mental state of presenter to others by spoken utterances. Emotions such as angry, neutral, sad, pleased, and surprise are commonly used to identify emotions from the speaker's spoken speech [1]. In this experiment, two intonation-based variables called Pitch and Formant are used to assess the emotional impact of speech signals. [2] showed how to use the Autocorrelation approach for a comparative examination of pitch using MATLAB. Another method is to use formant frequency estimates in conjunction with a time domain based algorithm to detect voice and unvoiced speech signals, as described in [3]. In this scenario, a speech data is used to compare the results of PRAAT's modified autocorrelation approach based on a time domain algorithm. In another paper, they analyzed on frequency comparison of Urdu speech signal based on FFT algorithm in MATLAB. They got the high correlation between frequency contents of the same word, which spoken by different speakers. Prediction of gender differentiation of the speakers by different parameters of speech analysis is the main aim included in [6].For the analysis they implemented MATLAB programming by calculating pitch using auto-correlation method. Another paper gives the concept of non zero pitch using Bayesian classifier [7]. In [8], they used the algorithm of IFS method for estimate the continuous pitch contour of speech signals. In this current work, discussed about the emotional effect between the speech signals of different gender and also identify comparison between two software which used for analyzing the features extracted from spoken utterances. This paper structure is given as follows. The second section mentioned recording specifications and collection of emotional speech corpus. Methodology of current analysis is discussed in section 3.Section 4 includes the Analysis of Variance testing. The “results and summery” of the experiments are present in section 5.At last, conclusions presents in section 6.

2. Collection of Speech Corpus

For this analysis, the vowel, consonants and words sounds of the Assamese language for both female and male speakers have been collected from different areas of Jorhat district of different age group 18-30 years, which consist of Assamese language. Samples consisted of acted speeches of eight short emotionally biased-sentences in Assamese languages which randomly taken from our day-to-day life. Speakers repeated these sentences three times with three different emotional styles- Angry, Normal and Surprise. In this early work, the main objective is to estimate the recorded sample of emotional speech for real world implementation. The speech stimulus “মই ঘৰত যামা“ was recorded in Assamese language by Windows built-in audio recorder and processed it by audacity at sampling frequency of 48KHz is converted to .wav files by Wave Surfer. The speech is super segmented and framed at phonemic level.

3. Methodology of This Analysis

The experimental framework of current task for analyze the consequence of emotions and software's on into national features is alienated into following steps:

1. Analysing the methods for estimating pitch frequency and Formant Frequency.

2. Software's using for intonational feature estimation

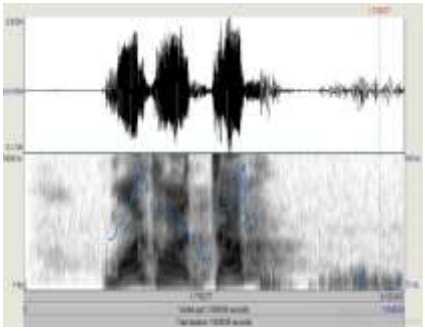
3. Using Two-way ANOVA

3.1. Methods For Estimating Pitch And Formant:

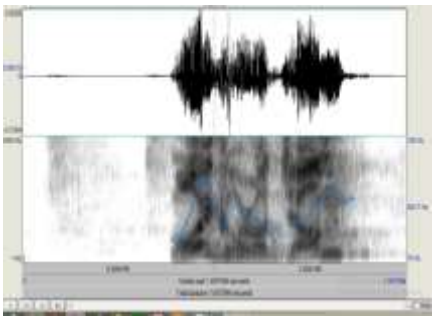
Pitch represents the fundamental frequency of the vibrating movement in the glottis and is associated to its physical behaviors. When vocal fold made taut, frequency increases. Pitch is heavily connected to the log of the frequency. Pitch analysis was categorized into two parts, time domain representation and frequency domain representation. The Time domain representation consists of Autocorrelation method and Zero crossing rate, AMDF(Average magnitude Differential functions etc. Frequency domain techniques of pitch estimations are cepstral analysis, maximum likelihood estimator etc. We considered cepstral analysis to estimating the value of mean pitch. Formant frequency is an important parameter for calculating the phonetic contents of speech signals. Estimating formant frequency is typically a slight complex than pitch estimation for the reason that the formant are the properties of vocal tract system and need to be inferred from the speech signal relatively just measure it. One of the foremost methods generally used for estimating formant frequency is based on modelling the speech signal. Such analysis is identified as Source filter model. In case of estimation of the formant frequency we are concerned on the frequency of resonances. To get the best matching system, a system called Linear Prediction Coding(LPC) system is used, where the signal is considered in a way as if it were generated by a signal of minimum energy being passed.

3.2. Software's For Prosodic Features Estimation

The Mathematical software MATLAB and speech processing software PRAAT are used for analyzing the emotional effect on prosodic features (Pitch and Formant) using recorded speech emotion of both male and female voices in Assamese language. Emotional speech corpus includes the utterance of one Assamese sentence which uttered by both male and female speakers in three emotions Angry, Normal and Surprise. Using PRAAT to rise above the shakiness of speech signal. PRAAT considers so as to the speech signal stable by take into consideration only the small fragments of it[5]. We analyze the mean pitch of the recorded samples using PRAAT. Following figure 1 represents the pictorial view of pitch of the sentence ―It is not your problem in Assamese language using PRAAT for (a) Female and (b) Male in Angry emotion. Another figure 2 provide the pictorial representation of formant values of same sentence for (a)Female and (b)male in Angry emotion.

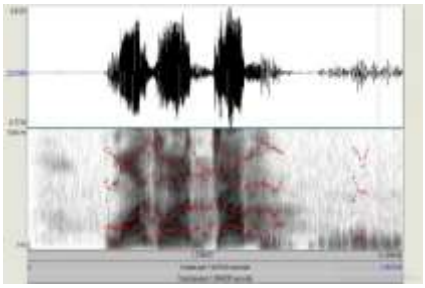


(a)

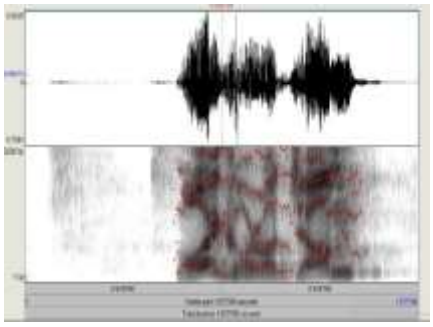


(b)

Fig 1:Pitch representation of Assamese sentence in Angry emotion for (a)Female and (b)Male



(a)



(b)

Fig 2:Formant representation of Assamese sentence in Angry emotion for (a)Female and (b)Male

In MATLAB, using the cepstral analysis, the pitch estimation is done by the following steps:

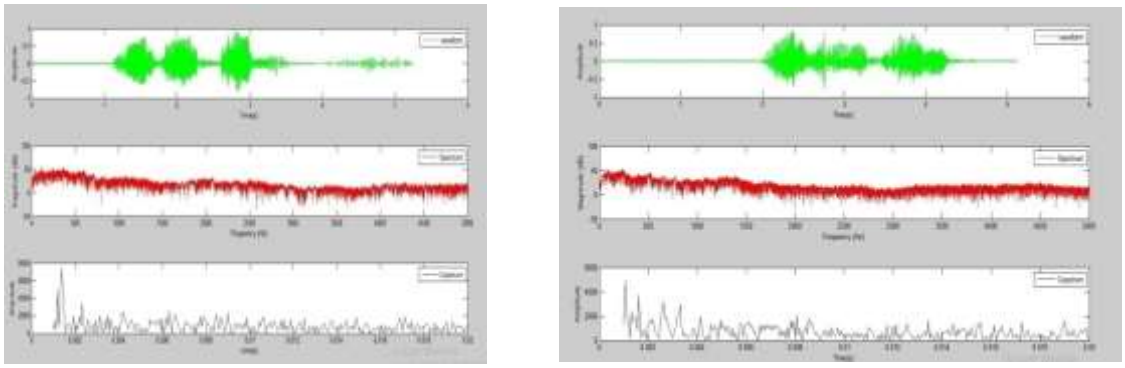
- The sample wav files of the selected sentence are firstly read in MATLAB using the coding

[x,fs]=wavread(filename.wav);

- Then plotted the audio speech signal. Windowing of the speech signal is processed
- After Windowing Fourier, transform is taken using y = fft(x);
- Plot the spectrogram

Creating the DFT Cepstrum coefficient and plotted

Following figure 3 represents the pictorial view of pitch estimation using MATLAB for (a) Female and (b) Male speaker.



(a) (b)

Fig 3:Pictoral representation of pitch estimation using MATLAB for Assamese sentence in Angry emotion for (a)Female and (b)Male

Using the method of Linear Prediction coding coefficient [6] for estimating the values of formant frequency by the following steps:

a) At first we read the speech signal using following coding in MATLAB.

$$[x, fs] = wavread\ (filename.wav);$$

(b)Repeat the sampling of speech signal using:

$$x = resample(x, 10000, fs);$$

(c)The speech signal is after that passed to the linear predictive filter (LPC) with coefficients:

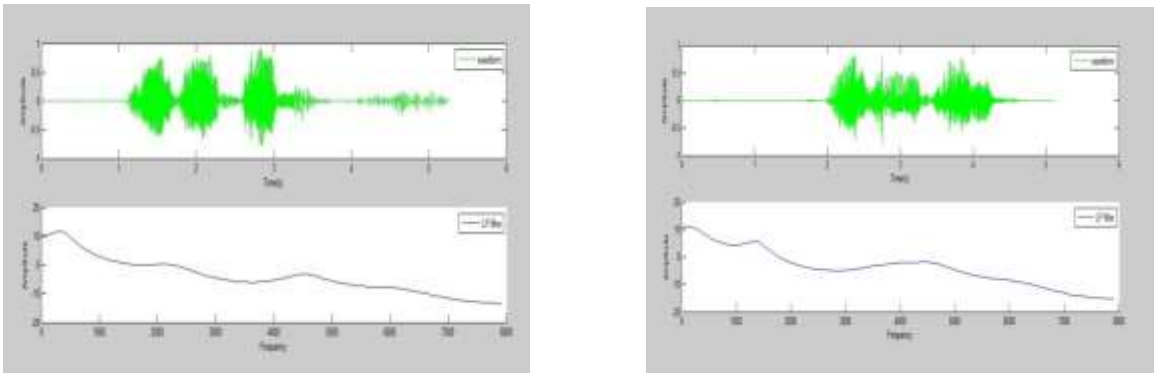
$$ncoef = 2 + fs/1000;$$

(d)The matrix A consists of coefficient values: $A = Lpc(x, ncoeff);$

(e)Plot the frequency response: $[h, f] = freqz(1, A, 450, fs);$

(f) By the root solving we got the Formant frequencies: $r = roots(A);$

After the above programming is done we get following pictorial representation in Figure 4 of both (a) Female and (b) Male



(a) (b)

Fig 4:Pictoral representation of formant estimation using MATLAB for Assamese sentence in Angry emotion for (a)Female and (b)Male

4.Two-way ANOVA Testing

ANOVA presents a statistical tool that can be used to analyzing the significant differentiation between the mean of different categories more than two. A two-way ANOVA is used to estimating the huge impact of intonations and software on the pitch and formant values. The affect of one of a kind impartial variables on one established variable is tested through the use of Two way Analysis of Variance (ANOVA).It now no longer most effective tested the principle effect of every independent variable however additionally the interaction between them if exist.

The general equation of Two way ANOVA testing is presented in following:

$$z = \mu + x_i + y_j + (xy)_{ij}$$

Where

$$i = 1, 2 \ldots \ldots \ldots \ldots \ldots \ldots . n, = 1, 2, \ldots \ldots \ldots \ldots \ldots . m$$

z is a Independent factor, such as (pitch and formant),μ is a Constant value,

x_i is significance between first factor, consider as (emotion_angry, emotion_neutral, emotion_surprise) y_j is the Main effect of second factor, consider as(software_PRAAT, software_MATLAB)

$(xy)_{ij}$= Interaction effect of Two factor

The subsequent basics are defined for pitch and formant value in two way ANOVA testing:

- Degree of freedom (DF): The number of values in the final calculation of ANOVA that are free to vary is called degree of freedom. It is the number by which a dynamic system can move in independent way without violating any constraint imposed on it[5].
- Adjusted Sum of Squares (Adj SS): It is the measure of variation from the mean. In ANOVA, the total variation that can be attributed to various factors is expressed by sum of squares[5].
- Adjusted Mean Squares (Adj MS): Dividing the sum of squares value by corresponding degree of freedom yields the mean squares value. It is used to determine whether factors have significant effect or not.
- F-value: The ratio of the two mean squares is called F-value. This value is close to 1 most of the time, if the null hypothesis is true. It determines whether the particular factor is significant or not.
- P-value: Each F-value corresponds to a particular P value. The cut off point for the alpha level of significance is usually represented by the critical value i.e. 0.05 and it is the p-value associated with certain F-statistics. If P-value is less than 0.05 then factor has significant effect.

The following table represented the two way Anova analysis for pitch and Formant of Male Speaker in Table (1&3) and Feale in Table(2&4) respectively. In these tables we gives the P (Significant) value and F value for mentioned features of both speakers correspondingly.

Table 1. Two Way ANOVA for Pitch of male Speakers

Dependent Variable: pitch					
Source	Sum of Squares (Type III)	df	Mean Square	Value of F	Value of P
Corrected Model	24363.008	3	8121.003	5.626	.023
Intercept	480317.54	1	480317.54	332.72	.000
Emotion	21784.26	1	21784.254	15.090	.005
software	529.371	1	529.371	.367	.562
Emotion * software	2049.383	1	2049.383	1.420	.268
Error	11548.62	8	1443.577		
Total	516229.17	12			
Corrected Total	35911.627	11			

Table 2. Two Way ANOVA for Pitch of Female Speakers

Dependent Variable: pitch					
Source	Sum of Squares (Type III)	df	Mean Square	Value of F	Value of P
Corrected Model	17967.896[a]	3	5989.299	2.550	.129
Intercept	784621.076	1	784621.076	334.069	.000
Emotion	15340.540	1	15340.540	6.532	.034
software	2605.413	1	2605.413	1.109	.323
Emotion * software	21.943	1	21.943	.009	.925
Error	18789.431	8	2348.679		
Total	821378.403	12			
Corrected Total	36757.327	11			

Table 3. Two Way ANOVA for Formant of male Speakers

Dependent Variable:formant					
Source	Sum of Squares (Type III)	df	Mean Square	Value of F	Value of P
Corrected Model	71115.127	3	23705.042	2.079	.181
Intercept	7222529.43	1	7222529.430	633.332	.000
Emotion	67750.894	1	67750.894	5.941	.041
software	268.155	1	268.155	.024	.882
Emotion * software	3096.079	1	3096.079	.271	.616
Error	91232.190	8	11404.024		
Total	7384876.74	12			
Corrected Total	162347.318	11			

Table 4. Two Way ANOVA for Formant of Female Speaker

Dependent Variable:formant					
Source	Sum of Squares (Type III)	df	Mean Square	Value of F	Value of P
Corrected Model	34151.344	3	11383.781	3.585	.066
Intercept	5013525.67	1	5013525.67	1.579	.000
Emotion	34108.731	1	34108.731	10.741	.011
software	3.244	1	3.244	.001	.975
Emotion * software	39.369	1	39.369	.012	.914
Error	25404.238	8	3175.530		
Total	5073081.25	12			
Corrected Total	59555.582	11			

5. Results And Summery

The prosodic feature (pitch and formant)values of male and female speakers are extracted from Assamese speech samples in Angry and Normal emotions with the help of PRAAT and MATLAB are mentioned in Table.5, Table.6,Table.7 and Table.8 respectively.

Table 5. Values of Mean Pitch for Male Speakers in Anger and Normal Emotions.

SPEAKER	Emotion	Software1	Software2
Male1	Anger	269.08133	264.141
	Normal	151.7473	162.244
Male2	Anger	189.9485	251.956
	Normal	152.3278	249.451
Male3	Anger	209.8589	321.053
	Normal	187.5819	141.403

Table 6. Values of Mean Pitch for Female Speakers in Anger and Normal Emotions.

SPEAKER	Emotion	Software1	Software2
Female1	Anger	287.310	344.444
	Normal	231.443	285.714
Female2	Anger	236.927	321.053
	Normal	201.880	257.143
Female3	Anger	301.880	457.143
	Normal	186.381	257.143

Table 7. Values of Mean Formant for Male Speakers in Anger and Normal Emotions.

SPEAKER	Emotion	Software1	Software2
Male1	Anger	641.272553	670.762975
	Normal	550.481851	517.897901
Male2	Anger	722.392275	697.292753
	Normal	660.170169	670.244182
Male3	Anger	728.392888	737.989493
	Normal	572.388700	587.150821

Table 8. Values of Mean Formant for Male Speakers in Anger and Normal Emotions.

SPEAKER	Emotion	Software1	Software2
Female1	Anger	886.632331	930.42163
	Normal	759.700177	884.43876
Female2	Anger	835.900457	684.97936
	Normal	607.153700	672.78190
Female3	Anger	864.314027	903.43341
	Normal	672.781905	607.15370

The above table 5 and 6 gives a comparative study of pitch between the two gender and also between two different types of emotion. The mean pitch values and their variations with reference to Assamese emotional speech samples are given in the above tables. The experimental observation and comparison between different emotions are mentioned in below.

- It is relatively understandable that the mean pitch values of every male and female speaker in Normal emotion are less than the mean pitch values in angry emotions with respect to time for same Speech sample.
- In case of two emotions, pitch values of male speakers are less than the pitch values of female speakers.
- The average range of mean pitch values of male speakers is141Hz-249Hz in Normal emotion and 189Hz-321Hz in Angry emotion.
- The average range of mean pitch values of Female speakers are186Hz-285Hz in Normal emotion and 236Hz-457Hz in Angry emotion.
- According to ANOVA analysis, based on the table 1 and 2, observed that emotions has the significant effect on the pitch of Male and Female Voices because P value is less than .05.
- In this case, also observed that Software and interaction of emotion and software has not any effect of significance on pitch of both voices (P>0.05).
- We also applied two way Anova testing for gender determination, and observed sex has highly significant effect on pitch in both the emotions (For Angry P value=.049 and for normal P=.005) and female speaker has high pitch than male Speaker. Following Fig 5 gives the estimated mean of pitch as follows:

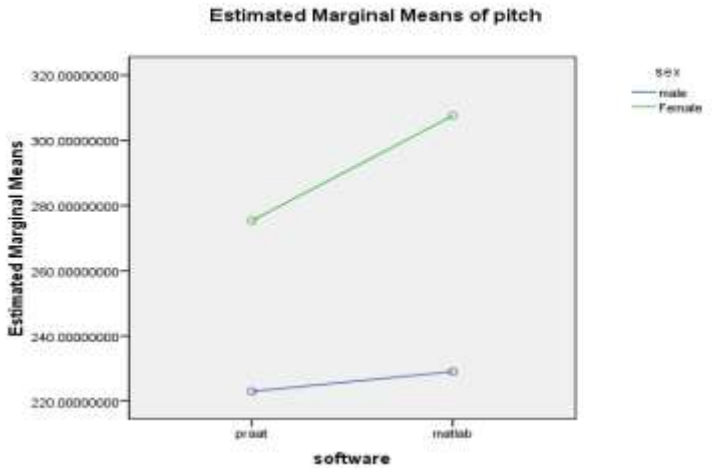


Fig 5: Estimated mean effect of pitch

The another table 7 and 8 provides the comparison of the mean formant values between different gender and emotion with respect to Assamese speech samples which extracted using two software (1)PRAAT and (2)MATLAB. Based on mean form formant values, obtained some observations from the experiments mentioned as follows:

- The Formant Values of Male and Female speakers in Angry emotion usually greater than that of the formant values in Normal emotion.
- The average mean range of first formant of female speakers' lies between 600Hz-1000Hz and the male speaker's range lies between 500Hz-900Hz.
- Based on formant analysis, observed that formant values of Female speaker with respect to both emotions are highly increases than the formant values of Male speakers.

- In Analysis of Variance testing, based on formant dependency, it also has significant effect between the emotional groups.
- From Table 3 and 4, shows that software has no Significant effect on formant of both speakers. Because P>0.05.
- We also applied two way ANOVA testing for gender determination, and observed sex has highly significant effect on Formant in both the emotions (For Angry P value=.009 and for normal P=.025) and female speaker has high pitch than male Speaker. Following Fig 6 and 7 gives the estimated mean effect of formant for both the speakers as follows:

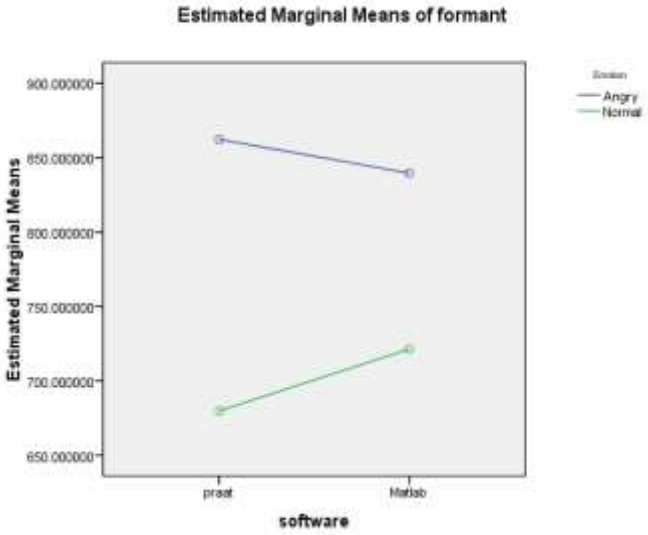


Fig 6: Estimated mean effect of formant for Female

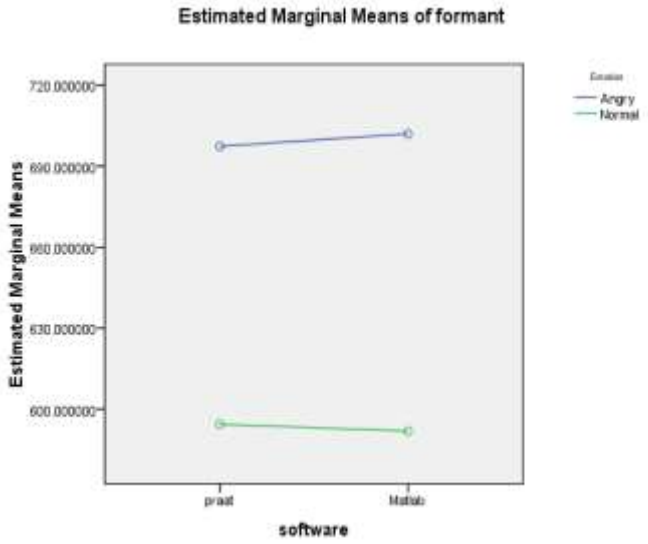


Fig 7: Estimated mean effect of formant for male

6. Conclusion

In this Paper, at we extracting two intonational features with the help of two softwares(1)PRAAT and (2)MATLAB, from the Assamese emotional speech signals. Applying Two way ANOVA testing method for analysing the effect of two emotions(Angry and Normal) and software's which on extracting these features of speakers of different gender. Also, tested the effect of Sex on the features of both speakers for gender identification. Experimental results based on ANOVA analysis tables, got significant effect between the emotions on extracting features dissimilar the software and the variation of the pitch value which is extracting using PRAAT and MATLAB is just about 50Hz.For future research eventual, authors may focus that the factors of ANOVA testing and can be experiment by pursuing various emotions and observe their impact on different intonational features.

References

[1] P. Ekman et al "An argument for basic emotions", Cognition and Emotion, Vol. 6, 1992.

[2] B.L. Jena et al "Gender Classification by Pitch Analysis", International Journal on Advanced Computer Theory and Engineering, Vol 1, 2012.

[3] H. Boril et al "Direct Time Domain Fundamental Frequency estimation of Speech in noisy condition" in Proceedings of European Signal Processing conference, 2004.

[4] S.G. Koolagudi et al "emotion recognition using speech features", Springer Science & Business Media New York ,2013

[5] Syed Abbas Ali et al, "A Study of the Effect of Emotions and Software on Prosodic Features on Spoken Utterances in Urdu Language", I.J. Image, Graphics and Signal Processing, 2016.

[6] Bhagya Laxmi Jena et al, "Gender Classification by Pitch Analysis", International journal of advance computer Theory and Engineering,2012.

[7] K.H. Hyun et al, "Improvement of emotion recognition by Bayesian classifier using non-zero-pitch concept", IEEE International Workshop on Robot and Human Interactive Communication,2005.

[8] Azam Beg et al "Frequency Analysis of Spoken Urdu Numbers using MATLAB and Simulink", PAF KIET Journal of Engineering & Sciences, vol.1, 2007.

[9] L. Qin, H. Yang et al, "Estimation of Continuous Fundamental Frequency Of Speech Signals" in proceedings of Speech Science and Technology Conference, Perth, 1994.

[10]WU Zhiyong et al, "Modeling pitch contour of Chinese Mandarin Sentence with Penta model", National Natural Science foundation of china, 2011

[11] U. Baumann et al,"Psychometric function of jittered rate pitch discrimination", Elsevier, pp 47-54, 10th May 2014.

[12] Davis S., and Mermelstein P., "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. ASSP 28, pp.357-366, 1980