A Study of the Effect of Emotions and Software on Prosodic Features on Spoken Utterances in Urdu Language

Syed Abbas Ali
Email: saaj.research@gmail.com

Maria Andleeb
Department of Computer Engineering, UIT, Karachi, Pakistan.
Email: maria_l3augi@hotmail.com

Danish Ur Rehman
Email: dani_khan@hotmail.com

Abstract—Speech emotions have potential to provide valuable source of information which can lead us toward human perception and decision making process. This paper analyzes the variation and effect on prosodic features (Formant and Pitch) of female and male speakers in two different emotions (angry and neutral) and softwares (PRAAT and MATLAB) in Urdu language using two ways ANOVA testing. The objective of this paper is to determine the significant effect of emotions and softwares on prosodic features (Pitch and Formant) using recorded speech emotion of both male and female voices of same age group in Urdu language. Experimental results of two-way ANOVA testing considerably show that emotions have effect on pitch and formant both in male and female voice unlike software.

Index Terms—Prosodic Features, Speech Emotion, Urdu Language, Two-way ANOVA Testing.

I. INTRODUCTION AND RELATED WORK

One of the essential tasks of the speech emotion signal analysis is to estimate the prosodic features to convey the psychological state of speaker to others by spoken utterances. Prosodic features are used in daily life conversation as a significant vocal correlates of emotion for recognizing and discriminating emotion from utterances of speakers [1]. Emotions like angry, neutral, sadness, happiness are commonly used to recognize emotions from utterances spoken by speaker [2]. Two prosodic features (Pitch and Format) are considered in this experiment to observe the effect of emotions and software on spoken utterances. Autocorrelation method based comparative analysis of pitch using MATLAB was presented in [3]. Another approach of time domain based algorithm for formant estimation in classifying voice and unvoiced speech signal were presented in [4]. In this approach, a spoken utterances database is used to test the performance of proposed time domain based algorithm in comparison with PRAAT modified autocorrelation method. Another approach for frequency analysis of Urdu number was proposed in [5] comprises of FFT algorithm in MATLAB to analyze the data. They found the high correlation among frequency contents of the same word, spoken by many different speakers. To alter the attractiveness rating the manipulation of voice pitch has been shown in [6]. This paper concluded that men prefer higher pitched women’s voices at very high frequencies, whereas women do not prefer low pitched men’s voices at very low frequencies. To estimate the continuous fundamental frequency of speech signals, a different approach was proposed in [7]. The algorithm used in this approach was based on the IFS (frequency estimation) technique. Another idea presented in [8] were based on two MATLAB function namely Generate spectrogram.m and Matrix to Sound.m. It is resolved that the spectra of the sound corresponding to time can be computed using the generate time versus frequency.m matlab files. Likewise an improved method to emotion recognition by pitch is proposed in [9]. The concept of Bayesian classifier is applied and it precisely explained the concept of non-zero pitch. In [10], effect of speech act and tone on rhythm was investigated to compare the mean difference between the PVI (Pair wise Variability Index) values for intensity of different speech acts using One-way ANOVA testing and results concluded that the comparison did not show any significant differences between the two pairs, tone1-tone2, and tone2-tone3. A novel data-driven approach was proposed in [11] to analyze syllable-sized tone-pitch contour similarly in a corpus of Beijing opera (381 arias). The ANOVA statistic modelling and machine learning methods are applied to 30-point pitch contour vectors. Similarly JRPD with a stable stimuli pattern was measured to determine the psychometric functions in [12]. A large standard
A Study of the Effect of Emotions and Software on Prosodic Features on Spoken Utterances in Urdu Language

II. RECORDING SPECIFICATIONS AND CORPUS COLLECTION

The data evaluated in this study were collected from the university students of age group 18-19 years, and consist of one Urdu language sentence that is suitable to be uttered with two emotions (Angry and Neutral). In this initial work, the samples of emotional speech were randomly taken from daily life personnel and non-actors with an objective of estimating the recorded speech emotion utterances for real-world implementation. The corpus recording has been performed in Urdu language based on ITU recommendations with following specifications: SNR≥45dB, bit rate of 24120 bps. Microsoft Windows 7 built-in sound recorders has been used to record the entire speech emotion utterances from speakers in Urdu language with following recording format: 48 KHz sampling rate with microphone impedance and sensitivity of 2.4ohm and 56dB±2.5dB respectively, cable length is 1.8m and stereo type of 3.5mm. Carrier sentence in this experiment was based on well-known desiderata, according to following considerations 1) should be easy to analyze, 2) semantically neutral, 3) dependency on presented situation and 4) should have analogous meaning for each languages. Based on the earlier results [23], the selected sentence was: “It is not your problem”. The sentence is spoken in two different emotions (Neutral and Angry) in Urdu Language: بی ہ ی جا کا موضع کایبی ہے

III. RESEARCH METHODOLOGY

The presented experimental framework for analyze the effect of emotions and software’s on prosodic features is divided into three stages:

1. Analysis methods for Pitch and Formant
2. Software’s for parameters Estimation
3. Two-way Analysis of Variance

A. Analysis of Pitch and Formant

The fundamental frequency of excitation source is defined as the pitch. Pitch is loosely related to the log of the frequency. As the frequency gets double, the perceived pitch increases about an octave. The pitch can be analyzed in frequency and time domain. Time domain methods of pitch estimation are ZCR (zero crossing rate), autocorrelation method etc. Frequency domain methods include cepstral analysis, maximum likelihood estimator etc. We use cepstral analysis to estimate the mean pitch value. Formant models the time evolution of speech signal and is considered as the frequency characteristic and also the prosodic feature. Several studies have been attempted to use spectral analysis to extract accent features in accent recognition such as filter bank analysis[24], cepstral analysis such as Mel-frequency Cepstral coefficient[25] etc. In this experiments, linear predictive coding coefficient method and cepstral analysis were used for prosodic parameters estimation.

B. Software’s for prosodic features Estimation

Two software’s were considered to analyze the effect of emotions and software’s (PRAAT and MATLAB) on prosodic features (Pitch and Formant) using recorded speech emotion of both male and female voices in Urdu language. Speech emotion corpus consists of utterances of one Urdu sentence uttered in angry and neutral emotions by male and female speakers. The reason of
using PRAAT is to overcome the instability of speech signal. PRAAT assumes that the speech signal stable by taking into account only the small fragments of it. PRAAT is used to analyze the mean pitch of the recorded samples. In MATLAB, the pitch value is estimated by using the cepstral analysis using following steps:

- The wav files are first read in MATLAB using \[ x, fs \] = wavread('filename.wav);
- The audio speech signal is plotted.
- Windowing of the speech signal is performed.
- Fourier transform of the windowed signal is taken using:
  \[
  Y = \text{fft}(x);
  \]
- The spectrogram is plotted.
- Cepstrum coefficient is formed which is the DFT spectrum.
- The cepstrum signal is plotted.
- Conversion of cepstrum value from s-domain to Hertz.

Linear Prediction coding coefficient method [26] is used to estimate the formant frequency values by taking following steps:

- The wav files are first read in MATLAB using \[ x, fs \] = wavread('filename.wav);
- Resampling of audio signal using:
  \[
  x = \text{resample}(x, 10000, fs);
  \]
- The signal is then passed to the linear predictive filter (LPC) with coefficients:
  \[
  ncoeff = 2 + fs/1000;
  \]
- Copying of coefficient values in the matrix i.e. \( A \):
  \[
  A = \text{Lpc}(x, ncoeff);
  \]
- Plotting of frequency response:
  \[
  [h, f] = \text{freqz}(1, A, 512, fs);
  \]
- Formant frequencies are then found by root solving:
  \[
  r = \text{roots}(A);
  \]

C. Two-way ANOVA Testing

Two-way Analysis of Variance (ANOVA) testing is used to determine the significant effect of emotions and software on the pitch and formant values. The influence of two different independent variables on one dependent variable is examined by using Two way Analysis of Variance (ANOVA). It not only examined the main effect of each independent variable but also the interaction between them if exist. The general model of equation of Two way ANOVA is represented in (1):

\[
 y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}
\]

Where,

\[
i = 1, 2, \ldots, n, j = 1, 2, \ldots, m
\]

\[
y = \text{Independent factor, consider as (pitch and formant)}
\]

\[
\mu = \text{Constant value}
\]

\[
\alpha_i = \text{Main effect of first factor, consider as (emotion_neutral, emotion_anxious)}
\]

\[
\beta_j = \text{Main effect of second factor, consider as (software_PRAAT, software_MATLAB)}
\]

\[
(\alpha\beta)_{ij} = \text{Two factor interaction effect.}
\]

The following preliminaries are defined in two way ANOVA testing for pitch and formant value:

- 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.
- 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.
- Adjusted Mean Squares (Adj MS): dividing the sum of squares value by corresponding degree of freedom yields the means 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 ANOVA table for pitch and formant (P-value and F-value) of both female and male speakers are shown in Table 1., Table 2., Table 3. and Table 4. respectively.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>1</td>
<td>19357</td>
<td>19357</td>
<td>7.87</td>
<td>0.021</td>
</tr>
<tr>
<td>Softwares</td>
<td>1</td>
<td>4507</td>
<td>4507</td>
<td>1.83</td>
<td>0.209</td>
</tr>
<tr>
<td>ERROR</td>
<td>9</td>
<td>22137</td>
<td>2460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack-of-fit</td>
<td>1</td>
<td>5066</td>
<td>5066</td>
<td>2.37</td>
<td>0.162</td>
</tr>
<tr>
<td>Pure Error</td>
<td>8</td>
<td>17071</td>
<td>2134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>46001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2. ANOVA Table for Male Speakers Pitch

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>1</td>
<td>7475.8</td>
<td>7475.8</td>
<td>5.17</td>
<td>0.049</td>
</tr>
<tr>
<td>Softwares</td>
<td>1</td>
<td>1444.0</td>
<td>1444.0</td>
<td>1.00</td>
<td>0.344</td>
</tr>
<tr>
<td>ERROR</td>
<td>9</td>
<td>13017.2</td>
<td>1446.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack-of-fit</td>
<td>1</td>
<td>103.9</td>
<td>103.9</td>
<td>0.06</td>
<td>0.806</td>
</tr>
<tr>
<td>Pure Error</td>
<td>8</td>
<td>12913.3</td>
<td>1614.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>21937.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Analysis of Variance**

Table 3. ANOVA Table for Female Speakers’ Formant

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>1</td>
<td>61456</td>
<td>61456</td>
<td>5.99</td>
<td>0.037</td>
</tr>
<tr>
<td>Softwares</td>
<td>1</td>
<td>51418</td>
<td>51418</td>
<td>5.01</td>
<td>0.052</td>
</tr>
<tr>
<td>ERROR</td>
<td>9</td>
<td>92413</td>
<td>10268</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack-of-fit</td>
<td>1</td>
<td>17898</td>
<td>17898</td>
<td>1.92</td>
<td>0.203</td>
</tr>
<tr>
<td>Pure Error</td>
<td>8</td>
<td>74514</td>
<td>9314</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>205286</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. ANOVA Table for Male Speakers’ Formant

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>DF</th>
<th>Adj SS</th>
<th>Adj MS</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>1</td>
<td>38306.2</td>
<td>38306.2</td>
<td>6.16</td>
<td>0.035</td>
</tr>
<tr>
<td>Softwares</td>
<td>1</td>
<td>910.8</td>
<td>910.8</td>
<td>0.15</td>
<td>0.711</td>
</tr>
<tr>
<td>ERROR</td>
<td>9</td>
<td>55968.2</td>
<td>6218.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lack-of-fit</td>
<td>1</td>
<td>3014.4</td>
<td>3014.4</td>
<td>0.46</td>
<td>0.519</td>
</tr>
<tr>
<td>Pure Error</td>
<td>8</td>
<td>52953.7</td>
<td>6619.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>11</td>
<td>95185.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IV. EXPERIMENTAL RESULTS & DISCUSSION

The pitch and formant values of male and female speakers as extracted from PRAAT and MATLAB in neutral and angry emotions are shown in Table 5, Table 6, Table 7 and Table 8 respectively. The regression equations of pitch value for both female and male speakers in neutral and angry emotions are represented in (6) and (7) respectively:

- Pitch = 317.4 + 40.2 Emotions_Angry - 40.2 Emotions_Neutral + 19.4 Softwares_Matlab - 19.4 Softwares_Praat
  \[(6)\]
- Pitch = 171.1 + 25.0 Emotions_Angry - 25.0 Emotions_Neutral + 11.0 Softwares_Matlab - 11.0 Softwares_Praat
  \[(7)\]

Table 5 and Table 6 provide the comparative analysis of pitch with two different emotions: Angry and Neutral when extracted from two different softwares. The mean values of pitch and the variation among them in term of speech emotion samples is presented in above tables. The comparison of the speech emotion samples has been made on the basis of the mean values. Experimental results show following observations:

- It is quite clear that in neutral emotions, the pitch values of both male and females are less than as compared to the pitch values in angry emotions.
- The male speakers pitch values are less than the female speakers pitch values in both emotions.
- The mean pitch values of male speakers have average range of 115Hz-195Hz in neutral emotion and 125Hz-290Hz in angry emotion.
- The mean pitch values of female speakers have average range of 225Hz-350Hz in neutral emotion and 290Hz-490Hz in angry emotion.

Similarly, the regression equations of Formant value for both female and male speakers in neutral and angry emotions are represented in (8) and (9) respectively:

- Formant = 879.9 + 71.6 Emotions_Angry - 71.6 Emotions_Neutral + 65.5 Softwares_Matlab-65.5 Softwares_Praat
  \[(8)\]
- Formant = 879.9 + 71.6 Emotions_Angry - 71.6 Emotions_Neutral + 65.5 Softwares_Matlab-65.5 Softwares_Praat
  \[(9)\]
Table 7. Mean Formant Values of Female Speakers in Neutral and Angry Emotions

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Neutral (Hz)</th>
<th>Angry (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRAAT</td>
<td>MATLAB</td>
</tr>
<tr>
<td>Speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>768.24</td>
<td>884.6</td>
</tr>
<tr>
<td>Angry</td>
<td>890.78</td>
<td>1019.5</td>
</tr>
<tr>
<td>Speaker</td>
<td>853.94</td>
<td>883</td>
</tr>
<tr>
<td></td>
<td>931.05</td>
<td>1184.2</td>
</tr>
<tr>
<td>Speaker</td>
<td>722.39</td>
<td>738.0</td>
</tr>
<tr>
<td></td>
<td>720.40</td>
<td>963.0</td>
</tr>
</tbody>
</table>

Table 8. Mean Formant Values of Male Speakers in Neutral and Angry Emotions

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Neutral (Hz)</th>
<th>Angry (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PRAAT</td>
<td>MATLAB</td>
</tr>
<tr>
<td>Speaker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>728.23</td>
<td>872.8</td>
</tr>
<tr>
<td>Angry</td>
<td>847.37</td>
<td>930</td>
</tr>
<tr>
<td>Speaker</td>
<td>759.17</td>
<td>638.7</td>
</tr>
<tr>
<td></td>
<td>850.253</td>
<td>799.4</td>
</tr>
<tr>
<td>Speaker</td>
<td>619.53</td>
<td>742.8</td>
</tr>
<tr>
<td></td>
<td>843.40</td>
<td>768.8</td>
</tr>
</tbody>
</table>

Table 7 and Table 8 provide the comparative analysis of first formant with two different emotions: Angry and Neutral when extracted from two different softwares. The mean values of formant and the variation among them in term of speech emotion samples is presented in above tables. The mean values of speech emotion samples are used to compare the samples of speech emotions. The following observation has been obtained from experiments:

- It is quite clear that in neutral emotions, the formant values of both male and females are less than as compared to the formant values in angry emotions.
- The first formant values of female speakers’ lies range 800Hz-1200Hz.
- The first formant values of male speakers’ lies range 700Hz-1000Hz.

A. Pictorial view of Pitch in PRAAT and MATLAB

To evaluate the experimental results in neutral emotion, Fig.1 and Fig.2 provide the pictorial view of the prosodic feature (Pitch) of the sentence “It is not your problem” in Urdu language using PRAAT. Figures shown that the peak of pitch speech signal of both female and male speakers is in the range defined above. The PRAAT and MATLAB software was used to observe the mean value of pitch in Urdu languages with two different emotions.

Similarly, Fig.3 and Fig.4 shows the MATLAB results of pitch analysis of both male and female speakers’ speech samples in the same sentence. Pitch was estimated using the cepstral method.
A Study of the Effect of Emotions and Software on Prosodic Features on Spoken Utterances in Urdu Language

ANOVAs were carried out with 5% level of significance. Both the male and female pitch show following observations:

1) Emotions has a significant effect on pitch of human voice (P<0.05).
2) Software didn’t has any significant effect on pitch of human voice (P>0.05).
3) The interaction of emotion and software also has no significant effect on pitch of human voice (P>0.05).

B. Main Effect Plot of Pitch and Formant using Two Way Anova:

The main effect plots of emotions and software of both female and male speakers pitch are presented in Fig.7 and Fig.8 respectively:

1) For female speakers the mean pitch value in neutral emotion is 276 Hz and in angry tone is 359 Hz.
2) For male speakers the mean pitch value in neutral tone is 149 Hz and in angry tone is 165 Hz.
Both the male and female first formant results show that:

1) Emotions has a significant effect on formant of both male and females (P<0.05)
2) Software didn’t has any significant effect on formant of male voices (P>0.05) but for the female speech samples software nearly effect the formant as P-value corresponds to 0.052 which is exceeding the significance level only by 0.2%.

![Fig.9](image9.jpg)

Fig.9. Main Effect Plot for Female Formant Values of Two Way ANOVA.

![Fig.10](image10.jpg)

Fig.10. Main Effect Plot for Male Formant Values of Two-Way ANOVA.

The main effect plots of emotions and software of both female and male speakers’ first formant are shown in Fig.9 and Fig.10 respectively:

1) For female speakers the mean formant value in neutral emotion is 808.36 Hz and in angry tone is 951.48 Hz.
2) For male speakers the mean formant value in neutral tone is 726.87 Hz and in angry tone is 839.87 Hz.

V. CONCLUSION

This paper presented the two way ANOVA testing to analyze the effect of the two factors i.e. emotions and software’s on prosodic features (Pitch and Formant) of male and female speech emotion samples recorded in Urdu language. Demonstrative experiments have been conducted using two different emotions (angry and neutral) and softwares (PRAAT and MATLAB) to analyze the influence of software’s and emotions on prosodic features (Formant and Pitch) using recorded speech emotion of both male and female voices in Urdu language. Experimental results provide evidence from Two-way ANOVA table that emotions have significant effect on prosodic features (Pitch and Formant) unlike the software and the interaction of the two factors, whether it is extracted from PRAAT or MATLAB the angry and neutral pitch values have approximately ±50 Hz variation. From future research prospective, authors are focusing that Two-way ANOVA testing can be observed by pursuing more emotions and analyze their impact on other prosodic features.

REFERENCES


Authors’ Profiles

Syed Abbas Ali received his B.E. and M.Engg. degrees in Computer Systems & Electrical Engineering from the N.E.D University of Engineering & Technology, Pakistan, in 1999 and 2002 respectively. Also, he received Ph.D. degree in Computer Science from N.E.D University of Engineering & Technology, Pakistan, in 2015. Since 2005, he has been working as an Assistant Professor in Department of Computer & Information Systems Engineering, Faculty of N.E.D University. His current research interests include Machine Learning & Speech Recognition, Speech Emotion Recognition and Computational Intelligence.

Maria Andleeb Siddiqui received her B.E degree in Telecommunication Engineering in 2010 from NED University of Engineering and Technology, Karachi, Pakistan. She received M.E degree in Computer and Information system engineering in 2013 from NED University of Engineering and Technology. Currently she is enrolled in PhD in the same University. Her research area is speech emotion recognition. She is lecturer in Usman Institute of Technology.

Danish Ur Rehman Khan received his B.E. degree in Computer Systems Engineering in 2000 from NED University, Karachi, Pakistan. He received his first M.E. degree in Electrical Engineering with majors in Telecommunications (by coursework) in 2007 and second M.E. degree in Computer and Information Systems Engineering with majors in Computer Networks and Performance Evaluation (ISP) in 2012. He is also accepted for PhD in NED in 2014. Current research interests include Human Behavior and Speech Recognition.

How to cite this paper: Syed Abbas Ali, Maria Andleeb, Danish ur Rehman,”A Study of the Effect of Emotions and Software on Prosodic Features on Spoken Utterances in Urdu Language”, International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.8, No.4, pp.46-53, 2016.DOI: 10.5815/ijigsp.2016.04.06

Copyright © 2016 MECS