IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

Acoustic Vowel Parameters Based Dialect Classification for Punjabi Speech

H. Kaur[1*], M. K. Gill[2]

[1,2]Dept. of Computer Science and Engineering, Guru Nanak Dev Engineering College, Punjab, India

[*]*Corresponding Author: harsimrankaur2110@gmail.com, Tel.: 9555570002*

DOI: https://doi.org/10.26438/ijcse/v7i7.165175 | Available online at: www.ijcseonline.org

Accepted: 22/Jul/2019, Published: 31/Jul/2019

***Abstract*—** In this paper, the Acoustic Vowel Parameters which is based upon the Dialect Classification for Punjabi Speech is provided. The information from the formant's dynamics F1, F2, and F3 was analyzed and further used in the process. The sound was evaluated with open source software PRAAT for the acoustic assessment. Multilingual speakers having age between twenty to thirty years has been selected for recording from Malwai and Doaba dialects of Punjab. The data of the total 20 people from Doaba region and 20 from Malwai region in which 7 females and 13 males has been taken for each dialect. In the proposed work MATLAB platform is used. First all the training dataset for the Doabi and the Malwai sound files were collected. The total training set has 140 sound files and the testing file has 19 sound files. Various parameters were analyzed in the training process. These parameters are Duration of the Sound file, Pitch, and Formants (F1, F2, and F3). The Formants (F1, F2, and F3) values were analyzed through PRAAT also. The formants are evaluated using the LPC method in MATLAB. The classifier used in the work was LDA and has classified the input sound file as per Doabi and Malwai sound file. The overall accuracy achieved in the system is 94.44%.

***Keywords*—** Acoustic, Punjabi, Formants, classification, training, testing, LDA (Linear Discriminant Analysis), and MATLAB.

I. INTRODUCTION

Punjabi seems to be the most commonly spoken language in two major areas of the world. In East Punjab, it is considered as the state language and in the West Punjab, which is also known as Pakistan, the Punjabi language is considered as the most widely spoken language. There is no particular figure found for the Punjabi speakers but it is found that more than 100 million people speak Punjabi in all the three parts of the world. As per the Language department of Punjab, it was found that there is a total of 12 dialects of Punjabi and other 28 dialects. All this data is according to the Linguistic Department of Punjabi University, Patiala. Other main dialects of Punjabi are Majhi, Doabi, Malwai, and Puwadhi [1]. Punjabi is considered as an important member of the Indo-Aryan language family. The Punjabi language is being written in Gurmukhi content in Indian Punjab, though in Shahmukhi content in Pakistani Punjab. One of the all the more fascinating actualities about the Punjabi language is that where it is numerically the most generally spoken, in Pakistani Punjab, it is not really composed by any means. Below figure shows the Origin of Punjabi language.

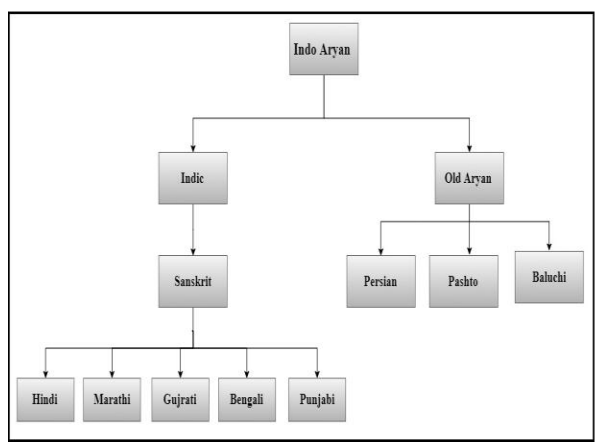


Figure 1. Origin of Punjabi language

Punjabi is regularly written in East Punjab in the Gurmukhi content. It is additionally conceivable to compose the language in the Persian content regularly alluded to as Shahmukhi in this unique situation [1]. Below shows vowels of Punjabi and their sounds.

Vowel	Consonant with vowel sign	Sound	symbol	International Phonetic Alphabet	Pronunciation
ਅ	ਸ	sa	a	ə	*a* in about
ਆ	ਸਾ	sā	ā	a	*a* in part
ਇ	ਸਿ	si	i	ɪ	*i* in it
ਈ	ਸੀ	sī	ī	iː	*ee* in see
ਉ	ਸੁ	su	u	ʊ	*u* in put
ਊ	ਸੂ	sū	ū	u	*oo* in food
ਏ	ਸੇ	sē	ē	e	*a* in cake
ਐ	ਸੈ	sai	ai	æ	*a* in man
ੋ	ਸੋ	sō	ō	o	*o* in show
ੌ	ਸੌ	sau	au	ɔ	*o* in bought

Figure 2. Vowels of Punjabi and Their Sound

1.1 Phoneme

A Punjabi phoneme can be characterized as a base sound unit of a language by which the significance might be separated. The phonemes are the components which remain conversely with one another in the phonological arrangement of that language. As it were, a phoneme in a language is characterized distinctly as far as its distinction from different phonemes of a similar language.

1.2 Formant Analysis

There are a few formants, each at an alternate recurrence, around one in each 1000Hz band. Each configuration relates to reverberation in the vocal tract. We recognize one dimension from another by the distinctions in their hints. As indicated by Lageforged (2006), every vowel has three formants, for example, three suggestion pitches. The first formant is contrarily identified with vowel stature. The second formant is identified with the level of the backness of a vowel. Vowels can be found in a wideband spectrogram as dim groups.

Vowel ratings were performed by frequencies of formant measured (F1, F2, F3). This is consistent with other tests for the classification of vowels. By evaluating the central frequencies of the reduced resonances at the sound signal known as the formant frequencies, bowel quality can be measured with appropriate accuracy and validity [7]. The center frequency of the smallest vowels, called the first forming frequency or F1, coincides tightly with the articulated and/or perceptive dimension of the vowel height. The F1 values range from 200 Hertz (Hz) for a strong vowel/i/ to about 800 Hz for a low vowel/a / for a male average tone. The second frequency of the formant (or F2) represents the maximum location to restrict it during vowel manufacturing, i.e. the forefront vs. the back dimension, so F2 values vary from approximately 2200 Hz for front / h / to approximately 600 Hz for sounds (including approximate, fragments, and trills, all at multiple articulation sites).

1.3 Classification

For the classification, we have three types of discriminative classifiers. These are LDA, FDA, and C5.0. All these classifiers seem to be opposed to the other generative models. The generative model includes Naive Bayes and HMM.

1.3.1 LDA

First, LDA is a classifier very comparable to regression with multiple responses. It allows the assessment of a binary dependent variable with ongoing as well as categorical predictors. In particular, the predictors seek to discover a linear framework that can distinguish two or more groups. LDA is based on the Bayesian probability, the highest probability and involves the normal distribution of information. Nexim Dehak and other research showed that when used to minimize the i- dimensions of acoustic characteristics in state-of-the-art i architectures for speakers, LDA can deliver excellent classification results. Linear Discriminant Analysis is a method used in machine learning and pattern classification as a preprocessing phase. The main goal of the techniques for dimensional reduction is to reduce the dimensions through the removal of the redundant and dependent characteristics by transforming the features from a larger space in a lesser space.

The objective of Linear Discriminant Analysis is to project the functions into a reduced room in a higher dimension. This is possible in three steps: The first phase is the calculation of the separability between the distinct classes, also known as an interclass variance.

1.3.2 FDA

FDA is a model of classification based on a combination of linear regressions that utilizes ideal scoring to convert the reaction variable, to make the information a better form of linear separation. For the classification of categorical factors, the FDA uses nonparametric methods. Thus in contrast to LDA, it does not involve the usual distribution of the information. Due to the failure to distribute all the forecasters of this research, the FDA will provide a better rating precision as LDA.

1.3.3 C5.0 Classifier

An interface to the models C5.0 is included in the C5.0 package. The two primary methods are a fundamental model based on trees and a model based on rules.C5.0 is a Quinlan (1993) classification algorithm. It evaluates class variables, like the dialect, based on several predictors. C5.0 creates a choice tree and provides a range of characteristics that can show the contribution in the classification of each acoustic characteristic. In particular, it remedies the information and uses predictors that can best divide the information in more sophisticated classifications. The most normalized information gain is the predictor chosen for the choice.

Typically, the variation or impurity of the information is interpreted in each split. If not, enough information is left to divide, the algorithm stops. Lastly, C5.0 offers tree and rule models (for the earliest C4.5 iterations, accents, and classifications of strained and unstressed African operators using C5.0) [2].

1.4 Speech Recognition Techniques

Speech signaling and processing are now used in many apps, whether it is a Google Speech Recognizer or a Smart Robot, a Call Center Automation, etc., speech obtaining, manipulating, storing, transferring, and displaying of Speech Signals can be performed in two ways. In the production of these signals, the methods of speech recognition play a key part, the first being the LPC, a digitized way to encode the analog signal. For anticipating a determined value from the prior value earned, this technique utilizes a linear function. LPC will be used to extract the function of the voice signal and to analyze speech and re-synthesis. LPC consists of subsequent stages such as pre-accent, frame-blocks, which divide the result of the last stage into blocks, a window which reduces the interruption for each frame at the beginning and the end, of each frame. LPCC is the next method for voice recognition and assumes that the identity of the produced voice signal is governed by the articulation form. The objective of the feature/attribute recovery is to generate voice signals using discreet signal measures. LPC extracts the LPCC coefficients and translates these into cepstral signals. The LPCC is commonly used to depict voice, signal quality and signal format parameters such as pitch length. It provides better efficiency and is easy to enforce, relatively more reliable and concerning LPC technology.

Rest of the paper is organized as follows, Section II contains the Related Work, Section III contains Problem formulation, Section IV contain the Proposed Methodology with flow chart, section V explain the results and discussions of the proposed work, and Section VI concludes the entire work and its future scope.

II. Related Work

Themistocleous (2017), have provided the overall classification model of two Modern Greek dialects. These are Cypriot Greek and Athenian Greek. They have used the information from the formant’s dynamics F1, F2, F3, and F4. In the work, the large set of vowels from a total of 45 speakers from Cypriot Greek and Athenian Greek was used. All these four formant frequencies were then measured at various time points. These values are then modeled with the help of second-degree polynomial. In the work, they have performed the classification using three classifiers. These are LDA, FDA, and C5.0 [1].

Mittal & Kaur (2016), have implemented the phonetic level speech recognition system. This system is for the Punjabi language as it is considered as the most prosodic language. The implementation was done with the help of HTK (Hidden Markov Toolkit). The data is collected in reading speech mode for nine hours. In the next step, the data is prepared which includes the hmm list, dictionary files, and grammar. They have divided the overall data and separate the 75% data for training and the other 25% data for testing. The overall results show that work accuracy has increased from 49.9% to 53.3%. The overall speech is transcribed to the ASCII characters with the help of the IPA chart. All the previous work is done in this same field but they are not using the Punjabi Language [2].

Phull & Kaur (2016), have described the section which is related to speech recognition. The major focus of their work is on the accent variations for the English language. Their main emphasis was on the analysis of the vowels. For comparison, they have considered the four types of accents in Indian English. These include the NI (North Indian), EI (East Indian), SI (South Indian), and WI (West Indian). The overall investigation is done with the help of PRAAT Software with the input as the Indian English dataset. The analysis of the Formant is performed using the two stages. These are Formant mean analysis and Formant space analysis. They have shown the triangle plots which helps in getting the required information for the occurrence of the vowels [3].

Lexuin (2015), have analyzed the acoustic characteristics of the various dialect of the Tibetan capital Lhasa. They have performed the work using the acoustic phonetics which includes three types of data which include the vowel formants frequency, fundamental frequency time, and acoustic vowel chart. Their study has found that the female student pronunciation mouth is bigger, the tongue seems to be relatively back, and lips seem to be relatively high as compared to that of males. As compared to the length of the vowels, the pronunciation of ‘i’ vowel seems to be longer, and ‘u’ as the shorter. The last one is that five vowels have shown a positive correlation with F1. In this, the ‘a’ is having the longest which is further followed by ‘o’, ‘e’, and ‘i’ [4].

Yue & Wei (2016), have described the Study of acoustic characteristics in Chinese whispered speech is essential for speech and speech recognition. In this article, whispered speech features are implemented and Chinese whispered speech's acoustic features are addressed. There is no fundamental frequency in whispered speech, thus extracting and analyzing other features such as formant length and frequency. Experiments with six easy Chinese whispered vowels demonstrate that format length and frequency can be used as the primary acoustic features in the Chinese whispered recognition. The study demonstrates that in Chinese whispered vowels the first, second, and third frequency of formant is upward change 1.46 times, 1.08 times, and 0.98 times, respectively, compared to normal. The bandwidths reach 3.98, 1.28 and 0.95 times respectively. Whispered vowel durations are 7–33% longer than ordinary syllables. Thus, the frequency of formant in Chinese whispered syllables is greater than in ordinary Chinese vowels [5].

Liu & Liang (2016), described Vowels are considered to play a significant part in the detection of subdialects. How they vary in tonal language among sub-dialects has not yet been examined. The current research explored acoustic correlates of vowels from three subdialects, such as vowel patterns, formant values, formant trajectories, vowel-inherent spectral change (VISC) and HNR. The findings of this acoustic research verified that even in tone language, vowels still played a major part in the detection of subdialects. Vowel's intrinsic spectral shift showed significant and systematic distinctions among these subdialects, which could prove to be an important source of acoustic identification for subdialects. Friction can also be a characteristic of vowels to differentiate themselves in other sub-dialects from their variants. However, there are indeed tonal differences across these sub-dialectal areas. They would, therefore, prefer to construct models for manufacturing and perception with weighted vowels and tones in further research [6].

Singh & Dutta (2011), analyzed the Punjabi vowel phonemes. It demonstrates a formal assessment of Punjabi vowels generated as a first language by Punjab speakers. Ten English vowel phonemes were compared and analyzed. Vowel productions of Punjabi were registered by a group of male speakers. The vowel formant frequencies and vowel length were carried out with acoustic-phonetic analyzes. The analysis shows quite clearly that the two languages have distinct phonemes of the vowels. The corresponding frequencies of the formant are distinct. First, Punjabi's vowel is distinctive for two languages characterizing vowels and differs from English, depending on the formant frequency. Secondly, three vowels vary in each language used for evaluation in our assessment. However, owing to the exposure to other non-native languages, a shift in the accent situation can be made [7].

III. PROBLEM FORMULATION

Compared to other languages, there is a lack of Punjabi vowel assessment of dialects. Over 100 million people worldwide speak Punjabi, but Punjabi has not yet increased to the status of a strong language. Sometimes, in another Punjabi dialect, the speakers of one Punjabi dialect are unable to comprehend some phrases of expression. The issue is due to the distinct pronunciation in distinct dialects of a specific Punjabi term. These distinct Punjabi word pronunciations function as a barrier of communication when two people from distinct areas of Punjabi interact. Therefore, owing to variation in vowels, there is a need to understand the variety in the significance of Punjabi language. In evaluating human psychology, dialects play a significant role. Acoustic voice signal analysis demonstrates the dialect's impact on the spoken utterances. Duration of the vowel is extremely affected by speaker's dialect. Vowel characteristics differ from distinct dialects. There are also distinct tonal characteristics of speakers from distinct dialects. These characteristics provide an important clue for dialect recognition and acoustic analysis. In addition, this research was not performed for distinct dialects in the Punjabi language.

IV. PROPOSED METHODOLOGY

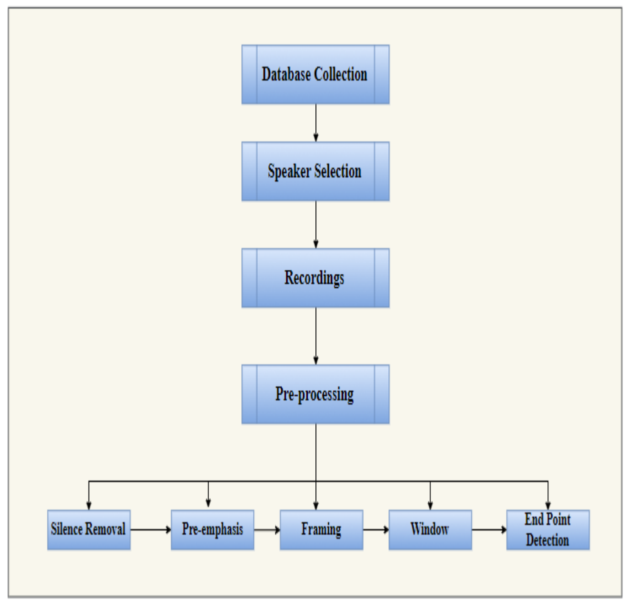


Figure 3. Flow diagram of Pre-Processing

4.1 Data Collection
In this thesis, the acoustic features of Punjabi vowels can be extracted for Doabi and Malwai dialect. Firstly, simple isolated words are selected from Punjabi Malawi language that contains non nasalized vowels. The words are selected in such a manner that individual phoneme can be segregated easily. The lists of the Malwai words are taken from the Malwai Shabadkosh published by Manmandir Singh Punjabi University Patiala. The list of the Doabi words are taken from Desh Duwaba and Dharat Doabe di by Dr. Karamjit Singh (Ex Professor of Kurushetra University).

4.2 Speaker Selection
Multilingual speakers have been selected for recording from Malwai region and Doaba of Punjab. Age of the selected

speakers is between twenty to thirty years. Speakers spoke Punjabi as a first language and had studied Punjabi since primary school. They all are university educated. The total 20 people from Doaba region and 20 from Malwa region was taken. From this, there are total 7 females and 13 males.

4.3 Wave File Recording

Wave files are recorded at a sampling frequency of 44000Hz. While recording wav files number of bits i.e. 8- or 16-bit size is defined by the user which is used to determine how much information is stored in a file. The user also provides duration in seconds like from 2 secs to 8 secs for which file will be recorded. In this mono recording is done for analysis. A corpus of Malwai words is recorded by selected speakers from Malwai dialect and corpus of Doabi words is recorded by selected speakers from Doaba dialect. Recorded files are saved with (*.wav) extensions. Then these files are digitized on a Dell computer using PRAAT software. All the recordings have been done through the LOGIC PRO X Software. The recordings were made in the studio as per the regions like Doaba and Malwa. For removal of noise, the equipment pop filter is used. The mic used for recording Samson C01. The overall System which is used in the work is MAC operating system.

4.4 Pre-Processing of Recorded Samples

The aim of the pre-processing process is to alter the voice signal to make it more appropriate for the assessment of the extraction function. Pre-processing involves the detection of noise, pre-emphasis and speech activation. Pre-processing is the basic signal processing applied to enhance the efficiency of feature extraction algorithms before extracting characteristics from the voice signal. The goal of pre-processing is to modify the speech signal in order to make it more suitable for the analysis of the feature extraction. Pre-processing involves the detection of noise, pre-emphasis and speech activation. Figure 3.2 shows pre-processing of speech signal.

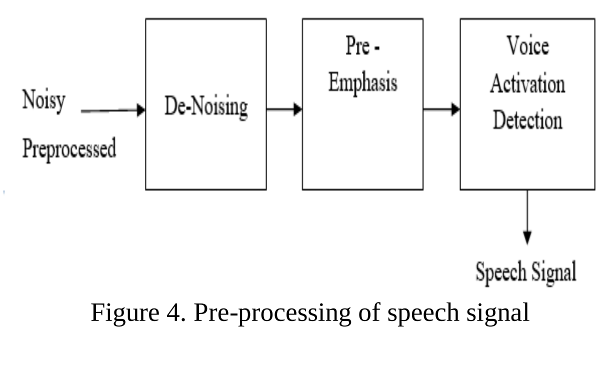


Figure 4. Pre-processing of speech signal

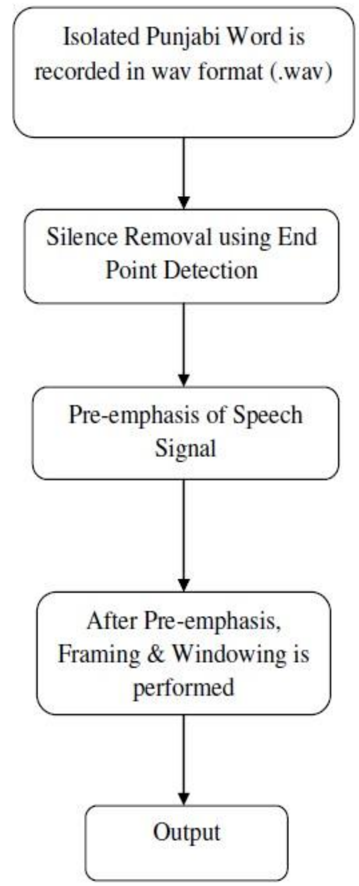


Figure 5. Flow diagram of Pre-Processing of Recorded Wav File

After the recording of wav files, separation of important speech data from recorded wav files are difficult. Pre-processing can be considered as the fundamental signal processing which is applied before the extraction of the features from the speech signal. This is done for the purpose of enhancing the performance of feature extraction algorithms. Pre-processing of sample data consist of the following steps shown in below figure.

4.5 Acoustic Analysis

The sound was evaluated with open source software PRAAT for the acoustic assessment. The keywords were positioned and manually segmented. Vowel forms (F1, F2, F3, and F4) were included in the readings. For the extraction of vowel formants, PRAATs conventional LPC-based technique can be used. In this, various Formants are analyzed for the particular sound file. This will have all the formants at every time duration of the sound file. The user can choose any time duration formants as per their requirements. There is options for viewing all the formants in the same page. The user can also save the values of the formants as per in their required format.

4.6 Overall Methodology

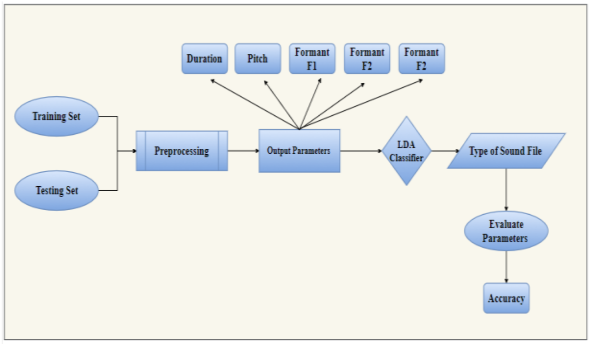


Figure 6. Flow diagram of Overall work

In this, all the work is performed using the MATLAB platform. The description for the overall process is as below:

4.6.1 Training Set

First all the training dataset for the Doabi and the Malwai sound files were collected. These sound files are then passed to the MATLAB. The total training set has 140 sound files. It contains the total 70 Doabi and 70 Malwai sound files. In the training set, there are two folders for the training set which are named as “Doabi” and “Malwai”. All the sound files got trained in the loop; the loop will go till the length of the directory folder.

4.6.2 Testing Set

In the testing set, there are total 19 sound files. The testing set sound files are different from the training set sound file. This means that the new sound files are used in the testing set. All the data from the training set is distinct from the testing set. In this, the Doabi sound files are 9 in numbers and Malwai as 10 in numbers.

4.6.3 Output Parameters

In this, various parameters were analyzed in the training process. These parameters are Duration of the Sound file, Pitch, and Formants (F1, F2, and F3). The analysis of the Formants (F1, F2, and F3) is done through PRAAT also. But the formants are evaluated using the LPC method in MATLAB. The Duration of the sound file is in seconds, and Pitch of the sound file is in Hz. The system will then extract all the parameters (Duration of the Sound file, Pitch, and Formants (F1, F2, and F3)) from the input testing file.

4.6.4 LDA Classifier

All the resultant output is then passed to the LDA Classifier. LDA is a classifier very comparable to regression with multiple responses. It allows the assessment of a binary dependent variable with ongoing as well as categorical predictors. In particular, the predictors seek to discover a linear framework that can distinguish two or more groups. LDA is based on the Bayesian probability, the highest probability and involves the normal distribution of information. The classifier will classify the input sound file as per the Doabi and Malwai sound file. This will help in generating the result and will help in finding whether the input sound file is from the Doabi region or the Malwai region. After this, all the testing sound files are then passed to the system. All the data of the testing sound files were extracted and then stored in the mat file. This will help in evaluating the overall accuracy of the system.

Relevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory).

V. RESULTS AND DISCUSSION

5.1 Testing results

Figure 7. Recording of an audio wave file

The above figure shows the wave file that has been recorded having 44000Hz sampling rate and 8 bit per sample for a few seconds is recorded. The screenshot of the system while recording wave file shown in the above figure.

5.2 Experiment Results

The overall work is implemented using the PRAAT Software and MATLAB Software.

5.2.1 Pre-Processing of Recorded Wave File

In this section experiment results were discussed after using multiple test audio wav files for isolated Punjabi words with the help of MATLAB software. Preprocessing can be considered as the fundamental signal processing which is applied before the extraction of the features from the speech signal. This is done for the purpose of enhancing the performance of feature extraction algorithms. In this work, the pre-processing of a speech signal is done using MATLAB. The criteria used here is the detection of the endpoint using the zero-crossing rate technique. The output we get after endpoint detection is used for framing and windowing. The rate of sign changes along a signal is called Zero Crossing Rate. It is simple to measure the frequency content of a signal. It is a measure of the number of times in a given signal, the amplitude passes through a value zero.

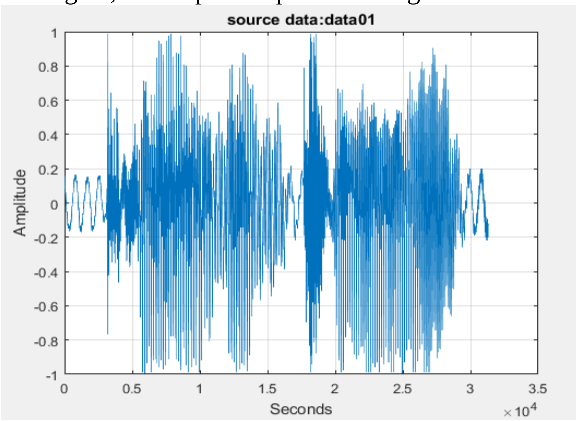


Figure 8. Representation of speech signal of the original wave file of word ‘n?Aseh’

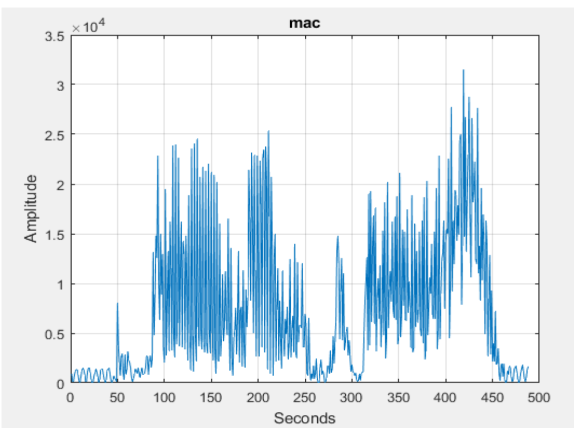


Figure 9. Representation of the actual speech signal of word ‘n?Aseh’

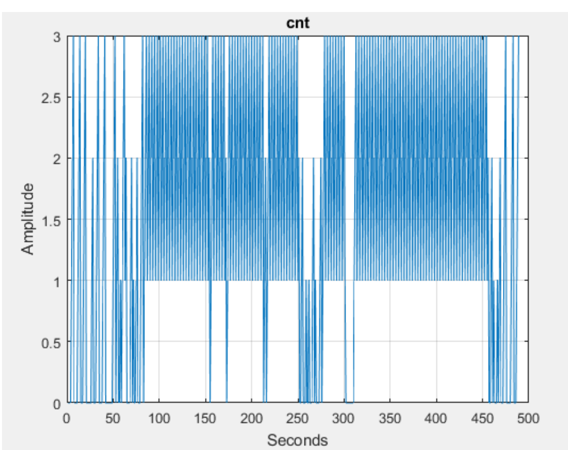


Figure 10. Representation of actual buffering points of an input speech ‘n?Aseh’

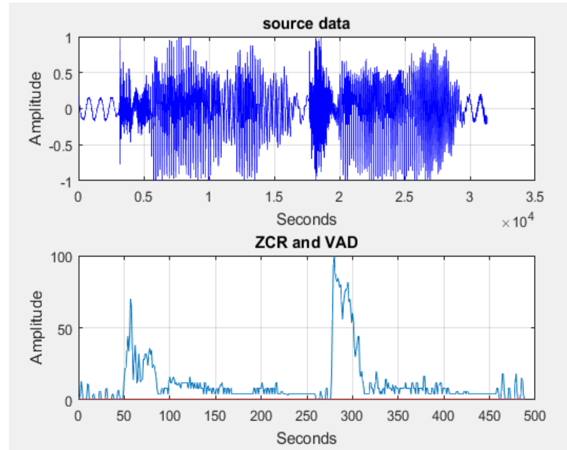


Figure 11. Representation of end points and comparison with the input speech signal

5.2.2 Acoustic Analysis

The sound was evaluated with open source software PRAAT for the acoustic assessment. The keywords were positioned and manually segmented. Vowel forms (F1, F2, F3, and F4) were included in the readings. For the extraction of vowel formants, PRAAT’s conventional LPC-based technique can be used. The sound shape and the spectrogram are displayed on top and bottom respectively when the Editor window is opened, and the cursor allows you to choose between them and measure. The menus above allow you to display, conceal and provide comprehensive queries to various acoustic data (e.g. formants, pitch, and intensity).

Step 1: Sound Signal selection

Figure below shows the Sound signal selection in PRAAT software

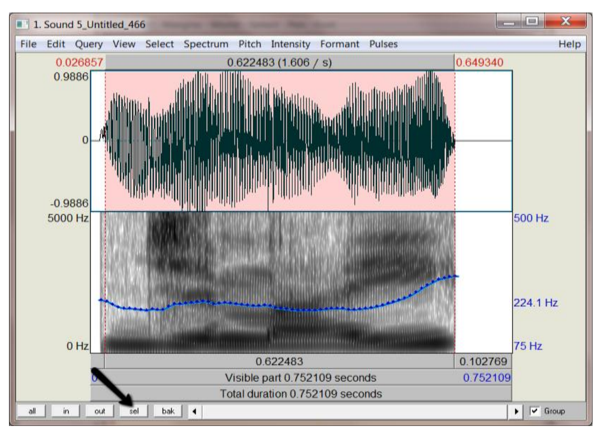


Figure 12. Sound signal selection in PRAAT software

In this step, the loaded sound signal is selected as per our requirement. The selection of the useful signal and then remove the extra signal from the overall signal has been done.

Step 2: Find the Formants

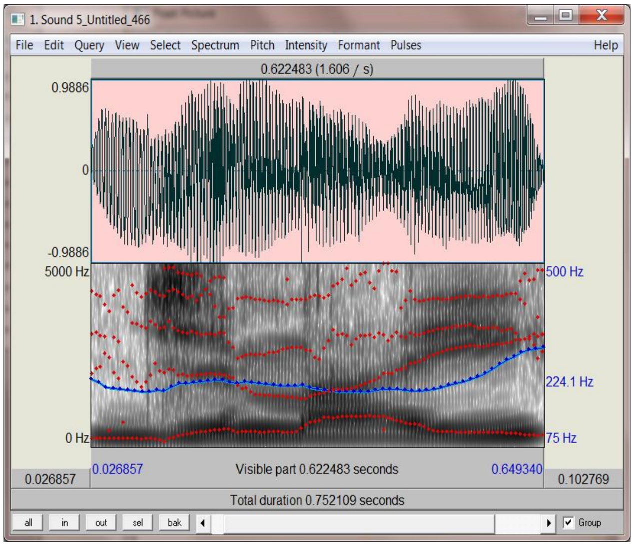
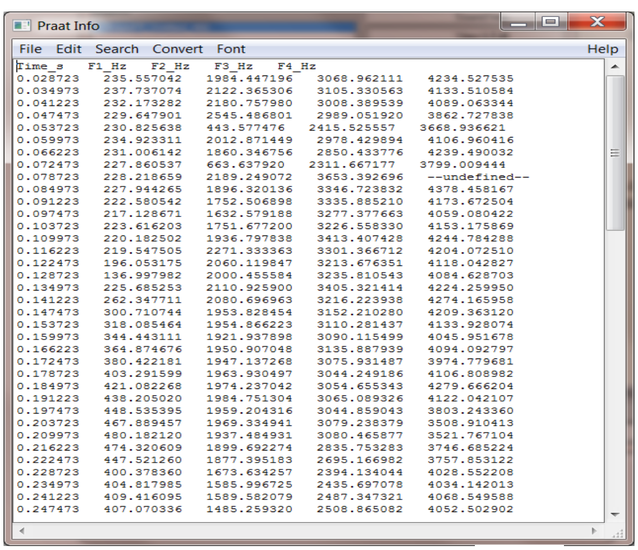


Figure 13. Red Dotted Formants

In this step, it can be analyzed the formants of the sound signal. This will help in representing all the formants of the signal. Below figure shows Red Dotted Formants of sound signal.

Step 3: List of the Formants

This will represent the list of the formants. In this, it will have all the values of the formants which include the F1, F2, F3, and F4 values at all the time. The below represents the F1, F2, F3, and F4 frequency formants of the sound signal. Below figure show the list of Formants generated by PRAAT software.



Praat Info

File Edit Search Convert Font Help

Time_s	F1_Hz	F2_Hz	F3_Hz	F4_Hz
0.028723	235.557042	1984.447196	3068.962111	4234.527535
0.034973	237.737074	2122.365306	3105.330563	4133.510584
0.041223	232.173282	2180.757980	3008.389539	4089.063344
0.047473	229.647901	2545.486801	2989.051920	3862.727838
0.053723	230.825638	443.577476	2415.525557	3668.936621
0.059973	234.923311	2012.871449	2978.429894	4106.960416
0.066223	231.006142	1860.346756	2850.433776	4239.490032
0.072473	227.860537	663.637920	2311.667177	3799.009444
0.078723	228.218659	2189.249072	3653.392696	--undefined--
0.084973	227.944265	1896.320136	3346.723832	4378.458167
0.091223	222.580542	1752.506898	3335.885210	4173.672504
0.097473	217.128671	1632.579188	3277.377663	4059.080422
0.103723	223.616203	1751.677200	3226.558330	4153.175869
0.109973	220.182502	1936.797838	3413.407428	4244.784288
0.116223	219.547505	2271.333363	3301.366712	4204.072510
0.122473	196.053175	2060.119847	3213.676351	4118.042827
0.128723	136.997982	2000.455584	3235.810543	4084.628703
0.134973	225.685253	2110.925900	3405.321414	4224.259950
0.141223	262.347711	2080.696963	3216.223938	4274.165958
0.147473	300.710744	1953.828454	3152.210280	4209.363120
0.153723	318.085464	1954.866223	3110.281437	4133.928074
0.159973	344.443111	1921.937898	3090.115499	4045.951678
0.166223	364.874676	1950.907048	3135.887939	4094.092797
0.172473	380.422181	1947.137268	3075.931487	3974.779681
0.178723	403.291599	1963.930497	3044.249186	4106.808982
0.184973	421.082268	1974.237042	3054.655343	4279.666204
0.191223	438.205020	1984.751304	3065.089326	4122.042107
0.197473	448.535395	1959.204316	3044.859043	3803.243360
0.203723	467.889457	1969.334941	3079.238379	3508.910413
0.209973	480.182120	1937.484931	3080.465877	3521.767104
0.216223	474.320609	1899.692274	2835.753283	3746.685224
0.222473	447.521260	1877.395183	2695.166982	3757.853122
0.228723	400.378360	1673.634257	2394.134044	4028.552208
0.234973	404.817985	1585.996725	2435.697078	4034.142013
0.241223	409.416095	1589.582079	2487.347321	4068.549588
0.247473	407.070336	1485.259320	2508.865082	4052.502902

Figure 14. List of Formants of word ‘n?Aseh’

5.2.3 Classification

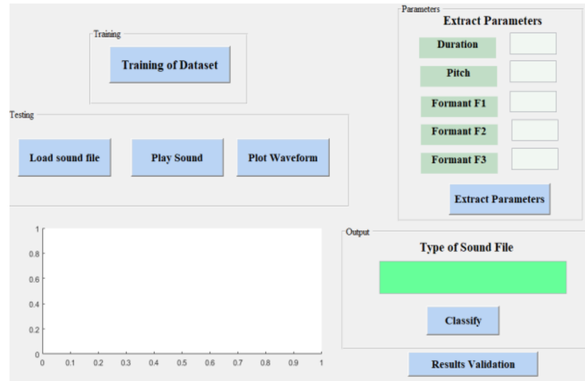


Figure 15. Main GUI

All the work is implemented in the software MATLAB. It has total 140 training sound files which includes 70 sound files from Doabi and 70 sound files from Malwai. All the parameters like Duration of the sound file, Pitch, and the Formants (F1, F2, and F3). All the data is the stored to one mat file. In this, the user will choose the input sound file as per their requirements. The sound will be played for the user and its plot is also there. The plot represents the actual waveform of the overall sound file. One right above part in the GUI is for the parameters extraction part. In this, it consists of various parameters like Duration, Pitch, and all the Formant values. The calculation of the Formants is done using the LPC method. The below shown figure is the main GUI which helps in evaluating the type of the Input sound file.

The below figure represents the training of the dataset. In this, the training of the Doabi Sound files is done as per the extracted parameters. There are total 70 training sound files for the Doabi dialect.

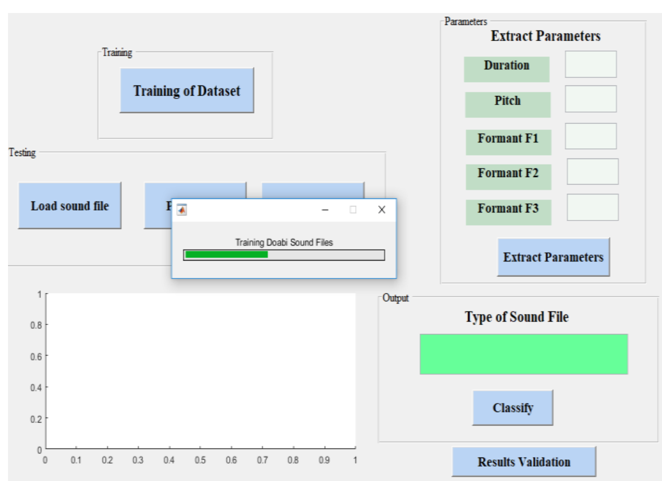


Figure 16. Training of Doabi Sound Files

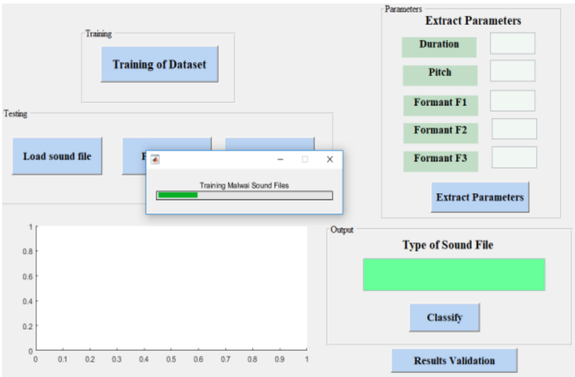


Figure 17. Training of Malwai Files

The above figure represents the training of the dataset. In this, the training of the Malwai Sound files is done as per the extracted parameters. There are total 70 training sound files for the Malwai dialect.

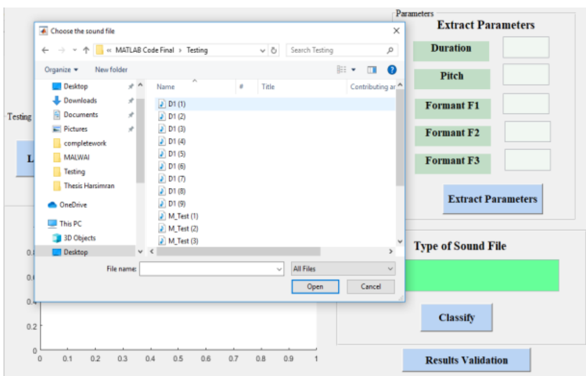


Figure 18. Load the Input Sound file of word ‘n?Aseh’

The above figure shows the loading of the sound file. This will help the user in choosing the sound file from particular folder. The user can choose any sound file as per their requirements.

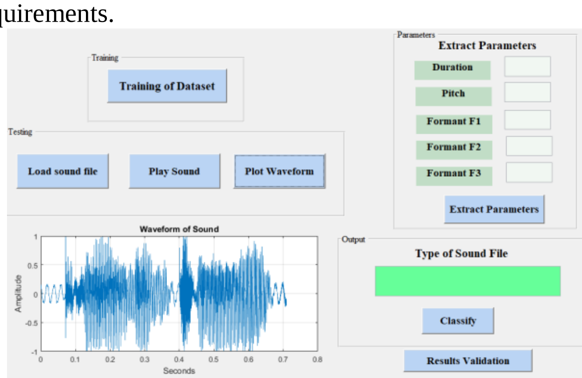


Figure 19. Plot Waveform of word ‘n?Aseh’

In this figure, the waveform and the sound of the audio file is represented. The Play sound will help in playing the audio file which was chosen by the user. The waveform of the audio file is representing the plot with the Amplitude and Duration.

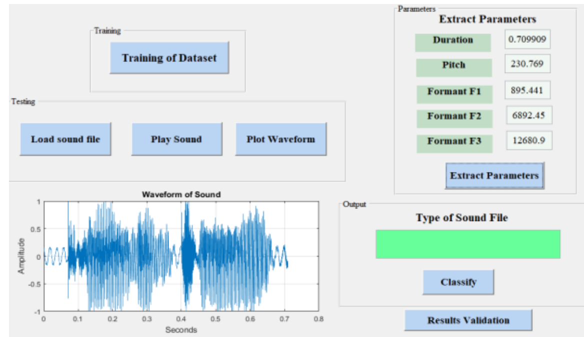


Figure 20. Extract Parameters of word ‘n?Aseh’

The above figure depicts the extraction of parameters. The parameters were extracted are Duration, Pitch, Formant F1, Formant F2, and Formant F3.

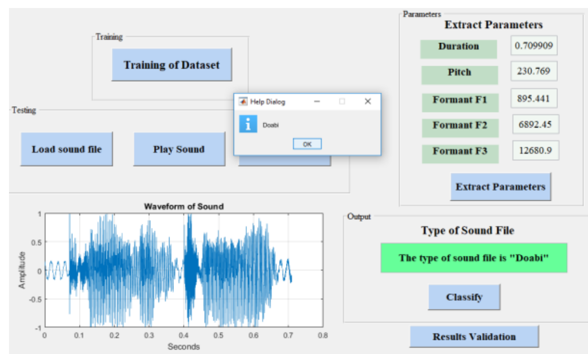


Figure 21. Classification result

The above figure shows the Classification results and have shown that the file which is chosen by the user is of Malwai.

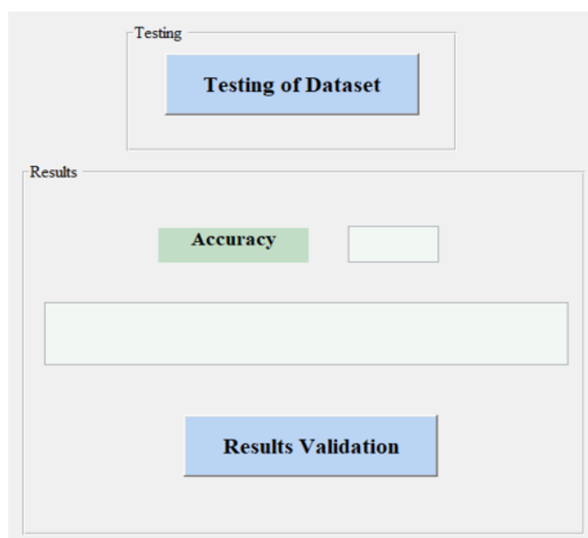


Figure 22. Result Validation

The figure 22 which is shown above is the main window for the Result Validations. This window is performing the testing of the dataset. The testing dataset has the total 19 sound files. In this, the Doabi sound files are 9 in numbers and Malwai as 10 in numbers

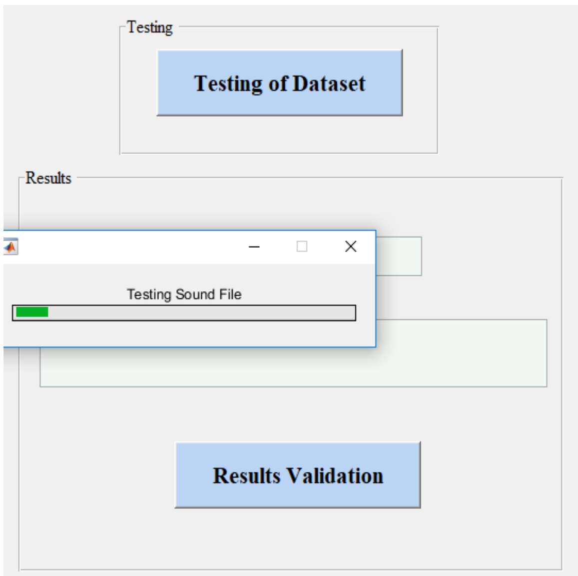


Figure 23. Testing of Sound Files

The above figure represents the process of testing the sound files. In this, there are total 19 testing sound files. From this, 9 belongs to the Doaba region and 10 belongs to the Malwa region. All these testing files are processed in one single loop. The loop will move till the length of the files in that particular folder.

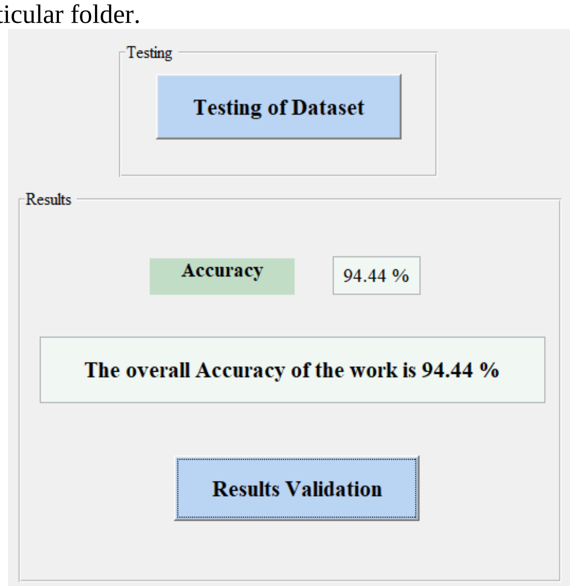


Figure 24. Accuracy Results

The above figure shown is the final window for the work. This represents the overall accuracy of the work. The accuracy of the overall work was calculated as per the Ground Truth and the predicted values. The overall accuracy of the work which is achieved 94.44%.

The Overall accuracy of the system was evaluated using the below mathematical formula:

$$\text{Accuracy} = \frac{\text{Results of the Predicted Labels are matched with Ground Truth values}}{\text{Total numbers of the Ground Truth data}} * 100$$

Here,

Ground truth value are the input labels of the testing image.

Predicted value are the resultant values from the classifier.

VI. CONCLUSION AND FUTURE SCOPE

In this thesis, the acoustic features of Punjabi vowels can be extracted for Doabi and Malwai dialect. Multilingual speakers have been selected for recording from Malwai region and Doaba of Punjab. The age of the selected speakers is between twenty to thirty years. It consists of total 20 people from Doaba region and 20 from Malwa region. From this, there are total 7 females and 13 males. Wave files are recorded at a sampling frequency of 44000Hz. The objective in the pre-processing is to modify the speech signal so that it will be more suitable for the feature extraction analysis. The sound was evaluated with open source software PRAAT for the acoustic assessment. The keywords were positioned and manually segmented. Vowel forms (F1, F2, F3, and F4) were included in the readings. For the extraction of vowel formants, PRAATs conventional LPC-based technique can be used. But the formants are further extracted with the help of MATLAB because it is providing the better results

In MATLAB, first of all the training dataset for the Doabi and the Malwai sound files were collected. These sound files are then passed to the MATLAB. The total training set has 140 sound files. Various parameters were analyzed in the training process. These parameters are Duration of the Sound file, Pitch, and Formants (F1, F2, and F3). The Formants (F1, F2, and F3) were analyzed through PRAAT also. The formants are evaluated using the LPC method in MATLAB.

All the resultant output is then passed to the LDA Classifier. The classifier will classify the input sound file as per the Doabi and Malwai sound file. This will help in generating the result and will help in finding whether the input sound file is from the Doabi region or the Malwai region. All the data of the testing sound files were extracted and then stored in the mat file. The overall accuracy of the system which we have achieved is of 94.44%.

For this work, the future scope can be as below:

- The dataset for the recordings can be increased to more speakers.
- We can use more classifiers in MATLAB for the classification
- The Hybrid classification methodology can also be implemented in the future for the classification.

REFERENCES

[1] C. Themistocleous, "*Dialect classification using vowel acoustic parameters*", Elsevier Speech Commun., Vol. 92, Issue.5, pp. 13–22, 2017.

[2] S. Mittal and R. Kaur, "*Implementation of Phonetic Level Speech Recognition System for Punjabi Language*", IEEE, 2016.

[3] D. Kaur and G.B. Kumar, "*Vowel Analysis for Indian English*", In the Proceedings of the 2016 6th International Conference On Advances In Computing & Communications, ICACC, 6-8 September 2016, Cochin, India, Vol. 93, pp.533–538, 2016.

[4] Y. Liuxin, "*The five main vowels acoustic analysis of Tibetan Lhasa dialect*", In the Proceedings of the 2015 Fifth Int. Conf. Instrum. Meas. Comput. Commun. Control, pp.0–4, 2015.

[5] Z. Yue and L. Wei, "*Study of the formant and duration in Chinese whispered vowel speech*", Elsevier Applied Acoustics, Vol. 114, pp.240–243, 2016.

[6] H. Liu and J. Liang, "*Vowels as acoustic cues for sub-dialect identification in Chinese*", IEEE, Vol. 1, pp.3–7, 2016.

[7] P. Singh "*Formant Analysis of Punjabi non-nasalized Vowel Phonemes*", In the Proceedings of the International Conference on Computational Intelligence and Communication Systems, India, pp. 382–387, 2011.

[8] O. Kalasnhnik and J. Fletcher, "*An acoustic study of vowel contrasts in North Indian English*",ICphS XVI, No.8, pp.953–956, 2007.

[9] P. Martiand and S. P. V. S. W. Beet, "*Analysis of ten vowel sounds across gender and regional/cultural accent*", pp. 2231–2234, 2016.

[10] H. Kaur and R. Bhatia, "*Speech Recognition System for Punjabi Language*", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 5, Issue 8, August 2015.

[11] V. Gupta and G.S. Lehal, "*Automatic Keywords Extraction for Punjabi Language*", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 3, 2017.

[12] W. Ghai and N. Singh, "*Continuous Speech Recognition for Punjabi Language*", International Journal of Computer Applications, Vol. 72, No.14, pp. 0975 – 8887, 2013.

Authors Profile

Ms. H. Kaur studied Bachelor of Technology in Information Technology from Ludhiana Group of Colleges, Chaukiman, Ludhiana (Punjab) in 2017. She is currently pursuing Master of Technology in Computer Science & engineering from Guru Nanak Dev Engineering College, Ludhiana (Punjab). Her main research focuses on Acoustic Vowel Parameters Based Dialect Classification for Punjabi Speech.

Ms. M.K. Gill received a Bachelor degree in Computer Science Engineering in 2006 and a Master degree in Computer Science & Engineering in 2009. She is with Guru Nanak Dev Engineering College, Ludhiana (Punjab) India as Assistant Professor, in CSE department since 2008- 2019. Her research area is natural language processing and speech recognition