International Journal of Scientific Research in Computer Science and Engineering
Research Paper Volume-3, Issue-5 ISSN: 2320-7639

Analysis and Recognition of Dialects of Hindi Speech

Shweta Sinha

Research Scholar (CS), Birla Institute of Technology, Mesra, Ranchi
meshweta_7@rediffmail.com
Available online at www.isroset.org

Received: 12/Aug/2015 Revised: 28/Aug/2015 Accepted: 15/Sep/2015 Published: 30/Oct/2015

Abstract- Every Individual has some unique speaking style that influences his/her speech characteristics. A major reason for this variability is caused by speaker's accent due to his native dialect. Prior knowledge of speaker's accent can help improve the performance of any speech recognizer. In this study, the problem of dialect classification of the spoken utterances in Hindi is considered. A database of four Hindi dialects; Khariboli, Bhojpuri, Haryanvi and Bagheli is created. Six hundred isolated words from Indian travel domain are recorded from 12 male and 8 female speakers of each dialect. In total, forty eight thousand utterances are taken into consideration. Spectral and prosodic features of C_1VC_2 syllable structure are extracted. Mel Frequency Cepstral Coefficients are computed as spectral feature. Vowel characteristics are measured in terms of formant frequency and duration. These features of different dialects are compared and analyzed with respect to Khariboli, the standard Hindi dialect. Sequence of fundamental frequency is evaluated to study the acoustic features associated with lexical tone. A multi layer feed forward neural network is implemented to show the sufficiency of these features for dialect classification. Experimental results shows that spectral and prosodic features combined together can give 82% recognition of dialect**.**

Index Terms: Dialect Recognition, Feature Analysis, Spectral and Prosodic Feature

1. INTRODUCTION

Current research in the area of speech recognition does not only concentrate on the correct evaluation of the linguistic information embodied in the speech signal, it also works towards identifying variations naturally present in speech. Several variability due to speaker related characteristics can be observed that influences the success of speech recognition system. Next to gender, dialect used by speaker is one of the major factor that influences the result of an ASR[1].Characterization of the effect of dialects, together with related techniques to achieve ASR robustness is a major research topic.

Dialect of a given language is a pattern of pronunciation or vocabulary of a language used by community of native speakers belonging to same geographical area. A language when used by people from different regions can be analyzed to see the usage of words with different lexis and even if they speak some standard form of the word the difference in spectral properties of sound produced can be observed [2]. Every individual develops a style of speaking at an early age. This style is dependent upon his native language or the dialect he speaks. When speaking some other language or even the standard form of his native language the speaker will carry traits of this style into it. This style may be influenced by sociolinguistic or regional environment of the speaker and impacts into speaking variations referred as accent. Different accents give rise to several differences in the realization of an utterance. Incorporating knowledge about dialects in pronunciation dictionary [3], acoustic training [4] can increase efficiency of an ASR.

Few studies have been carried out for automatic dialect identification. Linear discriminants successfully classify dialects of American English using average duration and average cepstrum of phonemes [5]. Arslan et al. [6] proposed algorithm for English accent classification using mel-cepstrum coefficients and energy as speech features. J C Wells [7] in his study emphasized that accent variation does not only stretch out in phonetic characteristics but also in prosodic characteristics of speakers. Influences of acoustic, prosodic and contextual information on dialects were emphasized by Kumpf et al. [8] in their work on Australian English accent classification. Huang et al. [1] studied the difference among English dialects and proposed Gaussian Mixture Model(GMM) for identification of dialects. Mahnoosh et al. [9] showed efficiency of pitch pattern for dialect recognition on Arabic dialects. A few studies on the analysis of dialects of Indian language have been carried in recent past [10]. Most of these are based on phonological approach. Some accent based classification approach for Hindi based on acoustic characteristics has been proposed in recent times [11,12]. These work uses speech samples from non native Hindi speakers and the system performance is also not noteworthy.

In this study our focus is on estimating parameters for dialect specific information at segmental and supra segmental level and exploring spectral and prosodic features for identification of dialects of Hindi Language. Four major

dialects of Hindi; Khari Boli (KB), Bhojpuri (BP), Haryanvi (HR) and Bagheli (BG) have been considered for studying these differences in the spoken utterance. The features so obtained are further used as input to a feed forward neural network to show their sufficiency for mechanical dialect classification. The paper is arranged as : section 2 describes speech material and database , section three discusses the features and their analysis for different dialects, Section four presents the implementation of FFNN for dialect classification. Next section discusses the results and observations and the last section presents the conclusion.

2. SPEECH MATERIAL AND DATABASE

Hindi is the first official language of India. It is spoken by more than 41% of Indian population. People in different Hindi speaking regions speak different forms of Hindi. There are about 200 different varieties of spoken Hindi with speaker population ranging from hundreds in one dialect to millions in another. These varieties of Hindi, known as its dialect are divided into eastern and western Hindi, which are further divided into various dialects. For the purpose of this experiment Khariboli (KB) which is spoken in Uttar Pradesh, Delhi, some parts of Uttarakhand and Himachal Pradesh and Haryanvi(HR) which is spoken by people of Haryana, Punjab, parts of Rajasthan and Uttar Pradesh are selected from the western dialects. From the eastern dialects Bagheli (BG), which is spoken in central India and Bhojpuri (BP) which is spoken by people of eastern Uttar Pradesh, Bihar and Jharkhand is selected for this study.

The text data is based upon words from travel domain. Six hundred words from Indian travel and tourism domain is selected and the text is prepared in standard Hindi using Devnagri script. The informants for this experiment have been selected from a uniform area of regional dialect and are from the age group 18 to 51 years with minimum higher secondary education. Several informants were recorded and finally 20 speakers (12 male and 8 female) from each of the dialects have been considered based on the score of perception test conducted to measure the influence of other native language on the speakers delivering style. All recordings have been done in office environment at 16khz. In total, forty eight thousand utterances have been used for this research. Syllables are considered as the basic phonetic unit for processing of Indian Languages. In Hindi language syllables of CV and C_1VC_2 structure have maximum frequency. In order to find the dialectal influence on spectral and prosodic features at segmental and supra segmental level, syllables of these two forms are taken into consideration.

For the spectral features MFCCs are extracted. These coefficients are obtained by reducing the frequency information of speech signal into values that emulate separate critical bands in the basilar membrane of the ear. Speech signal were divided into short frames of 25ms with a frame shift of 10ms. To obtain the information about amount of energy at each frequency band Discrete Fourier Transform (DFT) of windowed signal is obtained. A bank of 26 filters is applied to collect energy from each frequency band. Ten of these filters are spaced linearly below 1000Hz and remaining are spread logarithmically above 1000Hz. Twelve MFCCs, one zero coefficient and one energy coefficient totaling to fourteen coefficients are extracted from each frame. At the supra segmental level, syllable duration and sequence of fundamental frequency gives the prosodic information. The length of the segment is usually dependent on the duration of the vowel included in the segment. The recorded speech signals from different dialects are analyzed to obtain features distinct amongst them.

3. SPEECH FEATURE EXTRACTION AND ANALYSIS

Speaking style of any person can be characterized using phonetic features like style, phonation, dynamics of loudness and flow of speech. Based on these features linguistic characteristics of speakers can be compared. To compare the dialectal influence on individual speaking style auditory phonetic analysis and spectrographic analysis of recorded samples for all the dialects have been done. Syllables of C_1VC_2 structure have been extracted to analyze the samples. Khariboli is considered as the standard dialect of Hindi. All analysis has been done with reference to the Khariboli speakers.

3.1. Auditory Phonetic Analysis

The auditory phonetic analysis of speech of speech samples shows that Bhojpuri speakers often replace ‘sh’ sound with ‘s’ and ‘v’ sound with ‘b’. Also, ‘j’ and ‘z’ are interchangeably used by many speakers of this dialect. Haryanvi speakers have the tendency to replace cluster word with vowels inserted into it. Bagheli speakers show some similarity in speaking style with Bhojpuri speakers. They often replace ‘ph’ sound with ‘f’ sound. Also some times long vowels are shortened while speaking stressed syllables. The speaking utterance duration for Bagheli and Bhojpuri speakers are long as compared to Haryanvi and Khariboli.

3.2. Spectrographic Analysis

Vowel and consonants are small segments of speech that together form syllables. These syllables in turn make utterances. Syllables are considered as basic processing unit for Indian languages. Specific features that are superimposed on the speech utterance are known as supra segmental features. Spectrographic analysis of speech samples have been done to analyze the supra segmental features of spoken utterances in different dialect.

3.2.1. Segment Duration

Language can be discriminated on the basis of the length of the segment. The length is mainly concerned with vowel duration used in the segment [13]. Hindi Syllables can be formed using any of the 10 vowels. Mean vowel duration of these 10 vowels when spoken by speakers of four different dialects is presented in figure 1. The time taken by speakers of other three dialects for producing long vowel sounds is much less as compared to that of Khariboli speakers. It can be further observed that Haryanvi speakers have faster speaking rate than others. Bagheli as well as Bhojpuri speakers have the tendency to stretch the vowel sounds. The graph clearly shows that vowel qualities of these dialects are distinct from each other.

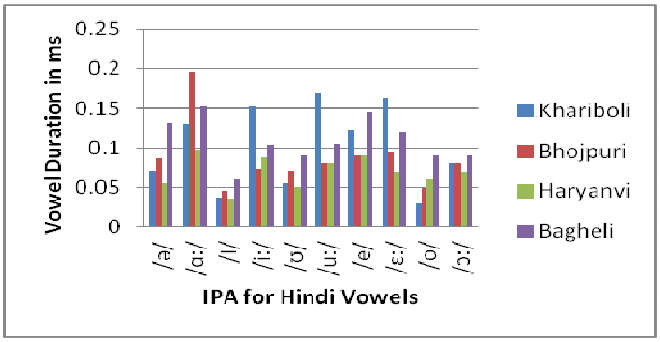


Figure 1: Comparative vowel duration of Hindi dialects

3.2.2. Formant Analysis

Vowels are sound units produced by voiced excitement of vocal tract. Acoustically vowels can be classified by formant pattern, spectrum, duration and shape of vocal tract. Formant frequency of vowels depends on the tongue height and tongue position. Figure 2, Figure 3, Figure 4 and figure 5 illustrates the formant frequencies representing vowel quality of all the four dialects for male speakers. The plot represents that Bhojpuri speakers have greater frontal open articulators as compared to Khariboli speakers. For long vowels most of the speakers show deviation from central tendency. Haryanvi speakers show deviation from central tendency for close frontal sounds. Bagheli speakers' formants show that they have some unique tendency and are different from Khariboli speakers. Most of the sounds are close to that of Bhojpuri speakers.

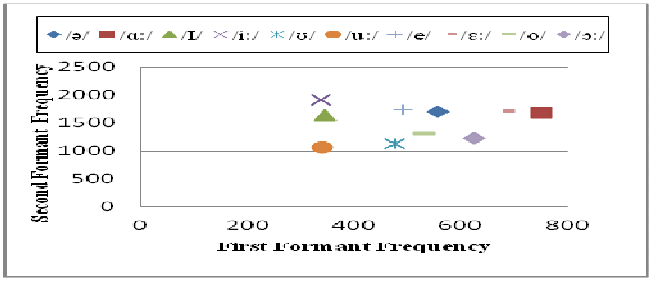


Figure 2: *Vowel quadrilateral for Khariboli male speakers*

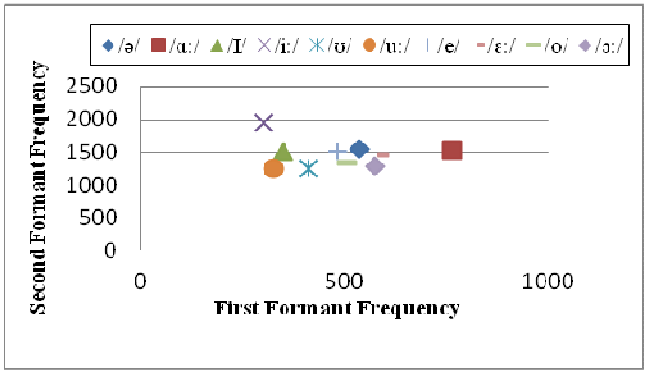


Figure 3: Plot of F1 and F2 for male Bhojpuri speakers

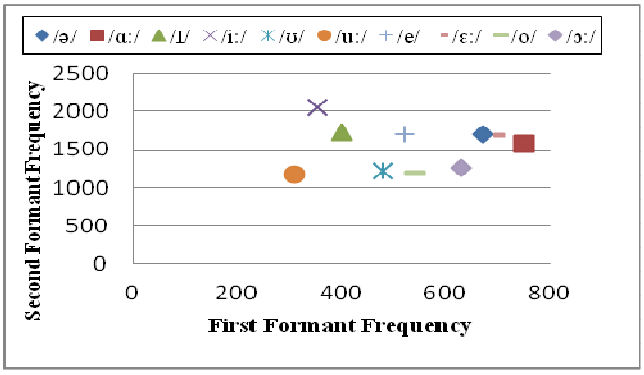


Figure 4: Plot of F1 and F2 for male Haryanvi speakers

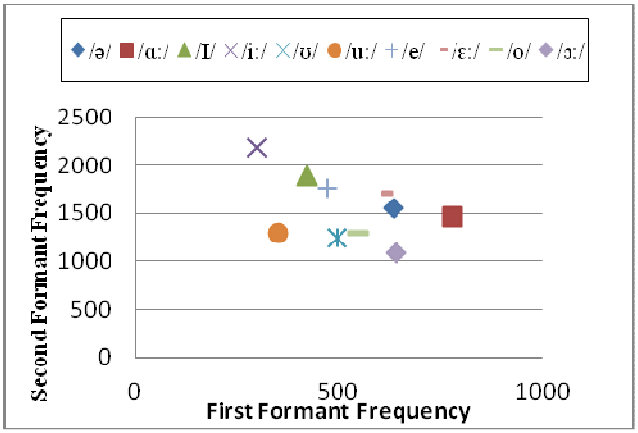


Figure 5: Plot of F1 and F2 for male Bagheli speakers

3.2.3. Lexical Tone

Intonation plays an important role to acquire distinct dialect accent. So study of tone becomes an important feature for dialect classification. Variation in tone is characterized by low and high variation in pitch contour over the pronunciation duration. Figure 6 to Figure 8 represents the tonal variation of male speakers of each dialect for the syllable /jheel/.

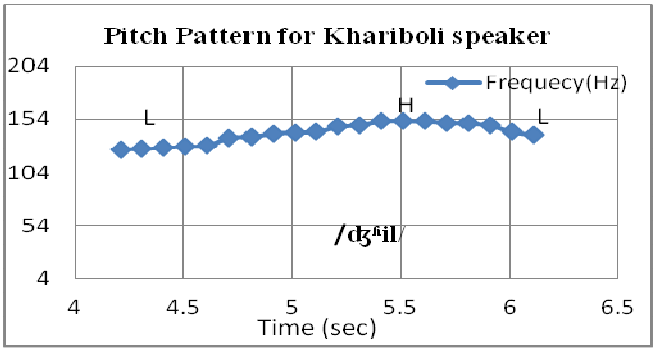
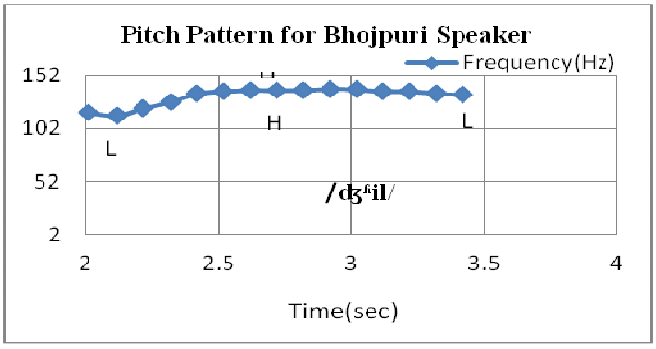


Figure 6: Lexical tone characteristics for male Khariboli speaker



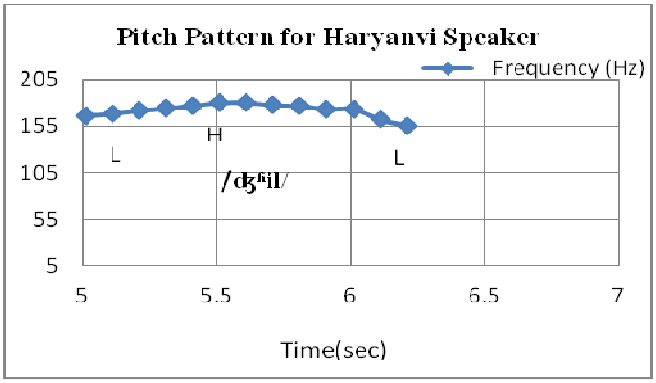


Figure 7: Lexical Tone plot for Bhojpuri and Haryanvi Speakers

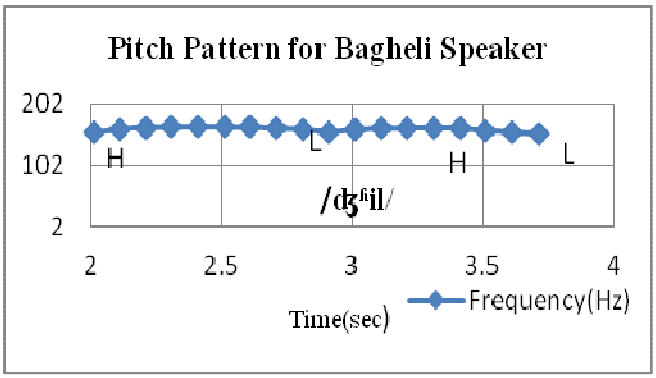


Figure 8: Lexical tone plot for Bagheli speakers

4. AUTOMATIC DIALECT CLASSIFICATION

A two layer feed forward neural network (FFNN) is expected to model behavior of any complex phenomena by establishing functional relationship between input and output vectors of the network [14]. With its learning ability it is subsequently able to classify different aspects of those behaviors. Its simplicity and reduced complexity makes it a suitable candidate for evaluating sufficiency of extracted speech features for dialect recognition. The input layer is fed by the feature vectors extracted from the syllables of C_1VC_2 and CV types. The activation function used at the hidden layer is the sigmoid function, tansig(). the learning rate of the system is fixed at 0.01. The network is trained with scaled conjugate back propagation algorithm. The choice of number of hidden layer neurons is arbitrary.

4.1. Feature Set

For dialect classification spectral and prosodic features have been extracted from the syllables of C_1VC_2 and CV structure from initial and final position of the utterance. For the spectral features MFCCs (14 coefficients) are extracted. For the prosodic features, syllable duration and pitch contour are extracted from the speech signal. The duration of syllables is measured in milliseconds. These values are normalized and stored as input features for the network. Two extra neurons, one for representing the normalized duration of syllables and second, positional value of syllable; 1 or 0 for initial and final position respectively. Pitch contours are extracted from the given utterance using autocorrelation method. The size of syllables varies based upon the dialect as well as the position, but FFNN requires fixed size of input in all iteration. The number of pitch components has to be fixed. Based upon inspection of the utterances under consideration fifteen pitch values are finalized for this work. Depending upon length of input utterances frame size is modified to keep the number of parameters fixed for each frame. Finally, fifteen features representing tonal characteristics are selected.

The recorded utterances under consideration belongs to both male and female speakers. A feature unit is used to represent the voice as belonging to male or female; 1 or 0 respectively. In totality thirty two features are used for representing the overall information about the input utterances.

5. RESULTS AND OBSERVATIONS

As the data representing each of the four dialects was sufficient for carrying this experiment, we have developed a single FFNN that accepts spectral features, pitch contour values, gender information, syllable duration and its position in spoken utterance as input. The system is trained 8 male speaker and 4 female speaker's data from each dialect. The rest of the data from each dialect, 4 male and 4 female speakers are used for testing the system. Based upon input features the trained network classifies input utterances

as one of the dialects: Khariboli (KB), Bhojpuri (BP), Haryanvi (HR) and Bagheli (BG). To study the influence of spectral and prosodic features on dialect recognition system the experiment is executed in two phases.

The system is trained with 14 MFCC features. Syllable duration and pitch contour values were also used for training the network. Inclusion of prosodic features significantly increased the recognition score to 82%. Table 2 represents the system performance when trained with both spectral and prosodic features together.

Table 2. *Performance of dialect recognition system*

% Recognition				
Dialect	KB	BP	HR	BG
KB	82	0	9	9
BP	4	87	1	8
HR	4	6	88	2
BG	10	7	2	81

6. CONCLUSION

Auditory and spectrographic study of Hindi speech is performed on the recorded data of four different dialects. Spectrographic analysis of speech signal shows the influence of dialects on the spoken utterances. Vowel duration is highly influenced by the dialect of the speakers. Vowel qualities of different dialects differ, however long and frontal vowels show significant differences. Tonal characteristics of speakers from different dialect also differ. These features provides significant clue for recognition of dialect. A feed forward neural network has been implemented to show that these features suffice the need of a dialect recognizer. Spectral feature set gives a recognition score of 71% which is further improved to 82% using these prosodic and spectral features. However increasing the size of database and study of some more dialects of Hindi will further enhance the system performance.

7. References

[1] Huang, R., Hansen, J. H. L. and Angkititrakul P., " Dialect /Accent Classification using Unrestricted Audio", IEEE Trans. Audio, Speech and Language Proc.15(2):453-464, **2007**.

[2] Sinha, S., Agrawal S. S. and Jain A., "Dialectal Influences on Acoustic Duration of Hindi Phonemes", in Proceeding of Oriental-COCOSDA, India, **2013**.

[3] Huang C., Chen T. and Chang E., "Accent Issues in Large Vocabulary Continuous Speech Recognition", in International Journal of Speech Technology (7):141-153,**2004**.

[4] Diakoloukas, V., Digalakis V., Neumeyer L., and Kala J., "Development of Dialect Specific speech recognizers using adaptation Methods" in Proc. IEEE International Conf. Acoustic, Speech, Signal Processing, Munich, Germany: (2),1455-1458,**1997**.

[5] Miller, D. R. and Trischitta, J., "Statistical Dialect Classification Based on Mean Phonetic Features", in Proc. IEEE International Conf. Spoken Language Processing, Philadelphia (4): 2025- 2027, **1996**.

[6] Arslan, L.M. and Hansen, J.H.L., " Language Accent Classification in American English" in Speech Communication (18): 353-367,Elsevier, **1996**.

[7] Wells, J. C., Accent of English, Vol. 2; Cambridge University Press, **1982**.

[8] Kumpf, K. and King, R.W., "Foreign Speaker Accent Classification using Phoneme Dependent Accent Discrimination Models and Comparison with Human Perception Benchmarks" ,in Proc. EUROSPEECH, Rhodes, Greece: 2323-2326, **1997**.

[9] Meharbani, M., Boril, H. and Hansen, J. H. L., " Dialect Distance Assessment Method Based on Comparison of Pitch Pattern Statistical Models" in Proc. IEEE International Conf. Acoustic, Speech, Signal Processing, Dallas, 5158-5161, **2010**.

[10] Mishra, D. and Bali, K., " A Comparative Phonological Study of the Dialects of Hindi" in Proc. International Congress of Phonetic Sciences XVII, Hong-Kong :1390-1393,**2011**

[11] Rao, K. S., Nandy, S., Koolagudi, S. G., " Identification of Hindi Dialect Using Speech", in Proc. 14th World Multi-Conf. on Systemics,Cybernetics and Informatics, Orlando, **2010**.

[12] Gaikwad, S., Gawali, B. and Kale, K. V., "Accent Recognition for Indian English Using Acoustic Feature Approach", International Journal of Computer Applications 63(7): 25-32,**2013**.

[13] Kulshreshtha, M. and Mathur, R., Dialect Accent feature for Establishing Speaker Identity: A case study, Springer Briefs in Electrical and Computer Engineering, **2012**

[14] Haykin S(**2002**), Neural Networks: A comprehensive foundation, Pearson education Asia, Inc, New Delhi.

