



THE COMPARATIVE ANALYSIS OF DISTINCT DIALECTS OF HINDI SPEECH IN INDIA.

¹Anuja Y. Jadhav, ²Mr. Vikram K. Hankare, ³Ms. Sushama S. Desai

¹ student, ^{2,3} Assistant professor

^{1,2,3} Department of forensic science, Yashvantrao Chavan Institute of Science, Satara, India

Abstract: Hindi language is the most commonly used language in India. In India, 43.6% of people consider Hindi as their mother tongue. It has a major impact in the northern and eastern parts of India. Many variations in the Hindi language can be observed from region to region. This study includes the study of variations in the Hindi language and the influence of other dialects on the Hindi language. Dialect categorization of Hindi is the major complication that is scrutinized here. Each person has a distinctive way of speaking that affects discourse features. Data from six different Hindi dialects i.e., Marathi, Marwari, Sindhi, Gujarati, and Islamic is produced. Each language mentioned previously belongs to different regions. The same word is spoken in different elegance by the people of different regions. The various prosodic features are analyzed in the course of stress, rhythm, and pause by auditory analysis. The means frequency of each dialect was calculated in this study. Around 120+ samples were analyzed all across India. 25 to 30 samples of each dialect were analyzed. The conclusion shows that the various features and spectral analysis can together help to recognize the dialect or to which region the individual belongs. This study can help map the region of the unknown voice sample.

Keywords: Hindi Dialect, Frequency mean, Auditory Analysis, Prosodic features, Speech analyzer.

I. INTRODUCTION:

Dialect is the local variation of speech that is different from the standard language in terms of semantics, lexicon, and accent. the dialect can be distinguished into three different types, i.e. regional dialect, social class dialect, and occupational dialect. A regional dialect is a dialect of a specific geographical region. Social class dialect refers to the dialect that differs from one social class to another. Occupational dialect is the dialect that differs according to the occupation of people.

Forensic Science is the use of scientific methods and concepts in criminal justice, particularly concerning the gathering, examining, and examination of physical pieces of evidence. Tape authentication and Speaker Identification is one of the divisions of Forensic Science which work on the suspected voice samples. Linguistics is the study of language, including its components, essence, formation, and alterations in human speech. In this study, the dialect variation of Hindi speech is considered. Hindi is the official language of India. India is a diverse country. There are several different languages spoken all across the country. The mother tongue differs from region to region over here. As Hindi is widely spoken here, it has a mild impact on every regional language on it. So, the Hindi language also differs from region to region.

Some of the major dialects of Hindi are:

1. Braj Bhasha: This dialect is spoken in the region of Braj; which includes parts of Uttar Pradesh and Rajasthan. It is characterized by the use of many Sanskrit loanwords. Braj Bhasha is known for its melodic and lyrical qualities and is often used in devotional music and songs
2. Bhojpuri: Bhojpuri is the dialect of Hindi spoken in the Bihar and eastern Uttar Pradesh region of India. It has a unique vocabulary and grammar and is known for its use of folk music and storytelling. Bhojpuri has had a significant impact on popular culture, particularly in the film Industry.
3. Haryanvi: Haryanvi is the dialect of Hindi spoken in the northern state of Haryana. It has a unique vocabulary and semantics and is known for its straightforward style.

In this study the samples were collected from different regions of India; i.e. Uttar Pradesh, Maharashtra, Gujarat, Rajasthan, and Punjab. The samples collected were analyzed by the method of Auditory analysis and by using the software Speech Analyser. The frequencies differed slightly in every language. The frequency mean was calculated from the obtained data.

II. MATERIAL AND METHODS:

The first step was the collection of samples from various regions of India. A paragraph containing the same content was recorded. The samples were taken in the Hindi language. The duration of the recording varied according to their speed and tone. It ranged from 40sec – 1min. The samples were then heard several times and their analysis was done accordingly. The variation in utterances was noted. The different acoustics features and prosodic features were taken into consideration. The time taken to read the paragraph was noted. And the time mean was considered. A dialect variation in accent was noted in different samples. The frequency means were calculated for each region in Speech Analyser software. The Speech analyzer software is mostly used to calculate amplitude, frequency, and spectrographic analysis. 120+ samples were recorded. 25 to 30 samples of the above five regions were recorded. The analysis of the sample can give approximately 70 to 75% identification of the person's region.

III. RESULT AND CONCLUSION:

Auditory analysis of Stress/pressure on words:

1. Marwari: More pressure was observed when they pronounce the letters ja, cha and tha
2. Gujrati: The word desh is pronounced as des. Shree is pronounced as sree. They pronounce sh as ss. Gha is pronounced as ga. Jha is pronounced as ja.
3. Marathi: The ya, tha, and cha ja pronunciations have more stress than standard Hindi.
4. Islamic: The ending letters of the words have more pressure than other letters. Ka is pronounced as ki. Sh is pronounced as she.
5. Sindhi: More nasal utterances are observed in these people. The letter na and ya is pronounced by a nose in words such as kyon, sthan, nahi, and vyakti.

Table 1: Stress/Pressure on words.

Standard Hindi	Marwari	Gujarati	Marathi	Islamic	Sindhi
Kyon	Kyoon (yo was pronounced as yoo)	Kyon (the pronunciation was similar to standard Hindi)	Kyu (yon is pronounced as yu)	Kiyu (k is pronounced as ki)	Kyon (yo is more pronounced)
Vyakti	Vekti (ya is not pronounced)	Vakti (ya is pronounced to a small extent	Vyakti (more stress is observed on word ya)	Vyekti (more stress is observed on word ti)	Vyakti (the pronunciation was similar to standard Hindi).
Sthaan	Sthaan (more stress is observed on tha)	Sthaan (this word was pronounced the same as standard Hindi)	Sthhaan (more stress was observed on tha)	Sthaann (s was less pronounced than thaa and na)	Isthan (there was a pronunciation of I in this word)
Shree	Shree (ra is pronounced tp a very small extent)	Sree (sh was pronounced as sa)	Shhree (more stress on sha)	Sheree (she is more pronounced)	Sheree (sh is pronounced as she there is a slight pause between she and ree)
Chalaana	Chhalaana (more stress on cha)	Chhalaana (la is pronounced with more pressure)	Chalaanaa (la and na is more pronounced)	Chalanaa (naa was more pronounced)	Chalaana (cha , la, and ne all the three were said with pressure)

Mean of time duration required for speaking (Speed):

The time required for the standard Hindi language was 40 seconds. whereas the time required by other dialects is as follows;

Table 2: mean time duration (speed)

Sr No.	Dialect	Mean Time Duration (Speed)
1	Marwari	1 min 3 sec
2	Gujarati	1min 8 sec
3	Marathi	56 sec
4	Islamic	59 sec
5	Sindhi	52 sec

Frequency means:

Table 3: frequency mean of dialect

Sr no.	Dialect	Frequency mean of the dialect	
		Males	Females
1	Marwari	195 Hz	335 Hz
2	Gujarati	250 Hz	392 Hz
3	Marathi	213 Hz	372 Hz
4	Islamic	203.3 Hz	358.2 Hz
5	Sindhi	182Hz	320.5 Hz

The study states that there is variation in accent, speed, and pressure on words. There is an influence of their regional dialects on the Hindi language. Every dialect has its own variations. The frequencies of the dialect also differed. The Sindhi region people had the lowest frequency while the Gujarati people has the highest frequency mean. The speed also varied from one dialect to another. The mean duration of Sindhi required for Sindhi people was low as that of Gujarati people was high. This study can help map the geographical region of an individual.

This study will contribute to differentiating between the Hindi language according to their region. This analysis will help in recognizing, evaluating, understanding, and investigating the various suspected Hindi voice samples. This study will serve as a springboard for additional investigation into Hindi dialects and the sociolinguistic effects of those dialects.

Figure 1.1: frequency vs time (sec) graph of male

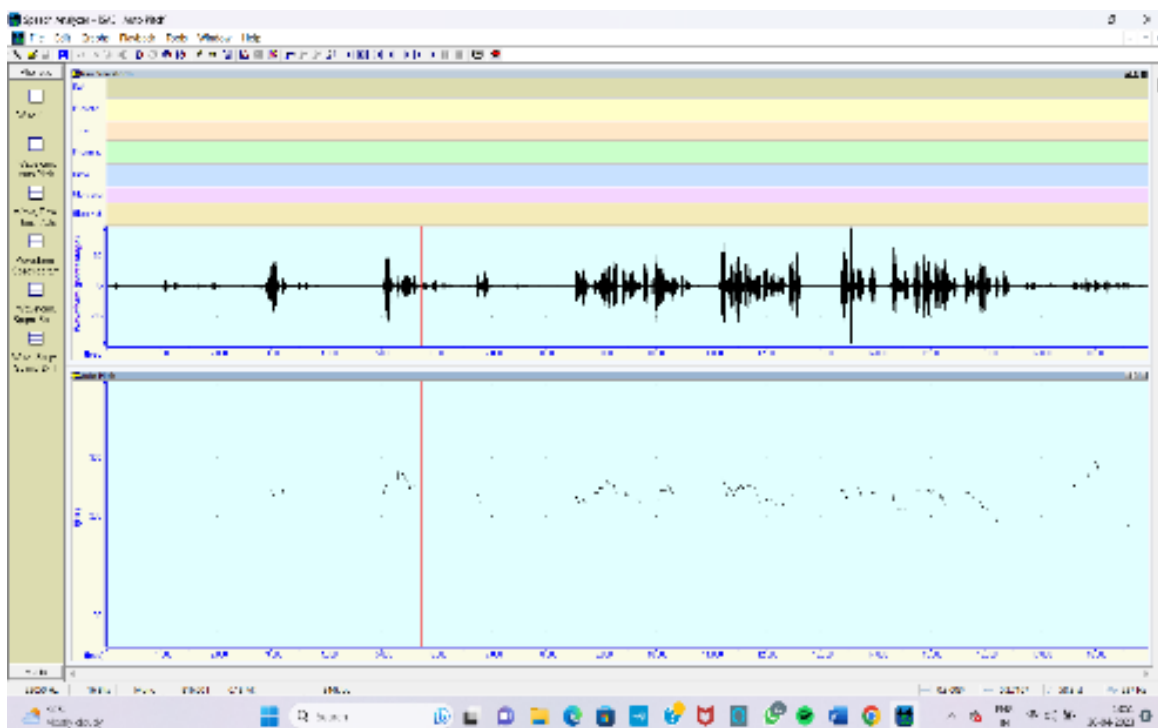
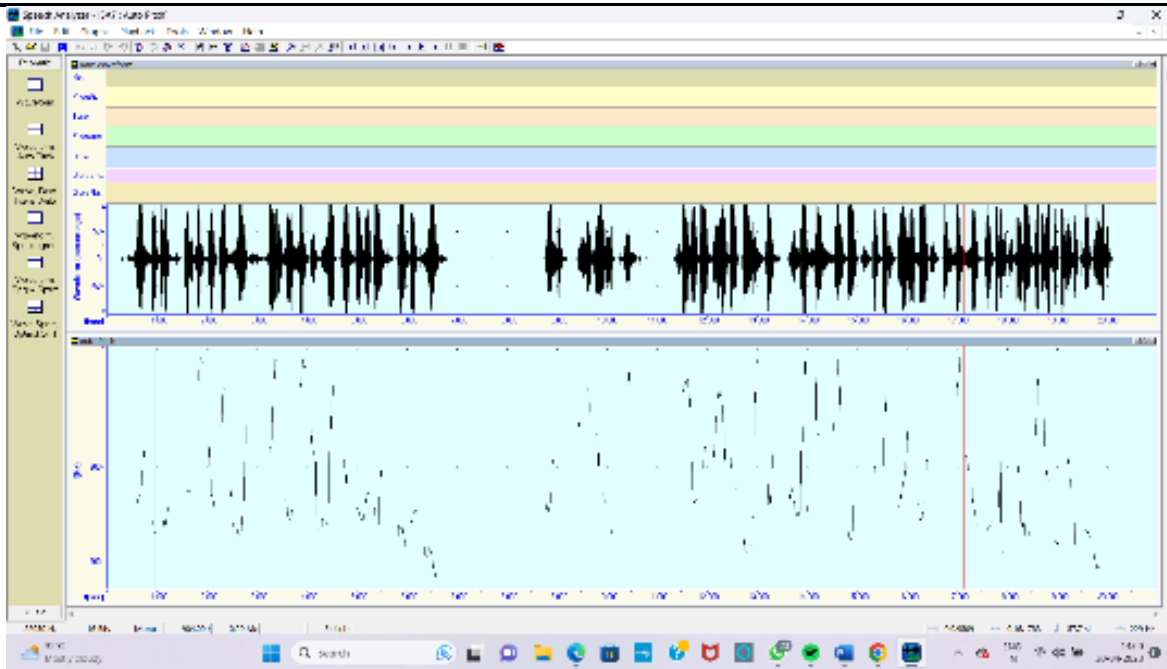


Fig. 1.2: frequency vs time (sec) graph of female



REFERENCES:

1. Shweta Sinha Analysis and Recognition of Dialects of Hindi Speech
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. 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
4. Diwakar Mishraa & Kalika Bali A COMPARATIVE PHONOLOGICAL STUDY OF THE DIALECTS OF HINDI
5. Aggarwal, R.K., Dave, M.: Integration of multiple acoustic and language models for improved Hindi speech recognition system. International Journal of Speech Technology (IJST) 15, 165–180 (2012).
6. Shweta Sinha, Aruna Jain, and Shyam S. Agrawal Speech Processing for Hindi Dialect Recognition
7. K Sreenivasa Rao, Sourav Nandy, and Shashidhar G Koolagudi IDENTIFICATION OF HINDI DIALECTS USING SPEECH