ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Accent Recognition using MFCC and LPC with Acoustic Features

Reena H. Chaudhari, Kavita Waghmare, Bharti W. Gawali

Research Student, Dept of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India

Research Student, Dept of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India

Professor, Dept of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India

ABSTRACT: India is a multilingual country where 29 individual languages as having more than 1 million native speakers. Hindi & English are the official languages of the Republic of India. Hindi is the most widely spoken language in India. Every language has its own sound structure, grammar syntax and intonation pattern, which makes it unique. Speech is a vocalized form of human language and it is a primary means of communication between people. The accent is a significant component of any speech. An accent is the way that particular person or group of people pronounce words or sounds. It typically differs in the tone of voice, pronunciation and distinction of vowels, consonants, stress and prosody. The purpose of this research is to determine the impact of Hindi language on other Indian languages like Marathi, Marwadi and Urdu Language with accent and some other features. It uses Mel Frequency Cepstral Coefficients (MFCC), Linear Productive Coding (LPC) to extract speech features from four different language groups, fundamental frequency Formant (F0) and energy feature vectors are used to examine the difference between language groups. This experiment tested on a database having 10 Hindi sentences which are let out by native speakers of Hindi, Marwadi, Marathi and Urdu. The observation from experiment indicates that both techniques give accurate and identical outcome. F0 and Energy parameter are founded effective in Urdu Dataset.

KEYWORDS: Language, Accent Identification, Hindi, Marwadi, Marathi, Urdu, Mel Frequency Cepstral Coefficients, Linear Productive Coding, F0, Energy.

I. INTRODUCTION

Language is a system of communication, consisting of sounds, words, and grammar rules which are linguistics of language. Linguistics is the scientific study of language. India having variety of languages. The major language families are Indo-Aryan and Dravidian has spoken by 73% and 24% respectively. This study focus on Hindi, Marathi, Marwadi and Urdu languages. All these languages come from Indo-Aryan group. Hindi literacy is written in the Devanagari script, has been strongly influenced by Sanskrit. Urdu is historically associated with the Muslims of the region of India. Urdu is closely connected to and mutually intelligible with Hindi, though a lot of Urdu vocabulary comes from Persian and Arabic, while Hindi contains more vocabulary from Sanskrit. Their distinction is most marked in terms of writing systems. Urdu uses a modified form of the Perso-Arabic script, while Hindi uses Devanagari [1]. Marwadi is the regional language of Rajasthan. Marwadi sounds similar to Hindi as it shares 50- 65% lexical similarity. It has more cognate words with Hindi. Marwadi language having quite same grammar structure of the Hindi language. Its primary sentence structure is SOV (subject–object–verb). Most of the pronouns and interrogatives used in Marwadi language are distinct from those used in Hindi. Marathi is the regional language of Maharashtra. There are regional variations in its pronunciation and by accent one would mark the region. Marathi grammar is mostly based on Sanskrit and Pali. Approximately 60% or more of the nouns in Marathi are derived directly from Sanskrit. Also, Marathi shares a considerable amount of words with Hindi. Unlike Hindi but like Sanskrit, Marathi has not 2 but 3 genders: masculine, feminine and neutral. [1, 2]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

The accent is an attribute of a speaker's speech through a systematic variation in pronunciation patterns due to different ethnic linguistic and cultural background. The direct effects of speech pattern can be determined from two major facets of linguistics which are prosodic features that are patterns of intonation and lexical stress and the other is, phonemic effects such as substitution, deletion and addition when mapping phonemes from ones first language to the second language when addressing. The accent is one of the most influential sources of speech variability other than gender, emotion states, health condition, speaking rate, and henceforth will always be a challenge for a machine to reproduce the human ability to separate words [3]. An accent does not include lexical differences, grammar, or other linguistic aspects. The prosodic features such as intonation, stress, rhythm, speaking rate is also considered as an accent, but acoustic features in pronunciation are normally more reliable to measure than the latter [4].

Past study shows that accent identification in speech increases recognition rate, but success rates vary among languages. The earlier researchers achieved a reduction of word error rate by 67% to 73% when accent knowledge was included in the system [3]. There are many techniques applied for feature extraction of speech signal. Mel Frequency Cepstral Coefficient (MFCC) and Linear Productive Coding (LPC) are most commonly used. MFCC based on the human peripheral auditory system and cannot perceive frequency over 1KHz [5, 6]. LPC analyses the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz [7]. MFCC considers the nature of speech while it extracts the features, while the LPC predicts the future features based on previous features [8]. Acoustic features like energy, First fundamental formant frequency F0 (pitch) play an important role in accent.
The rest of this paper is organized as follows. Section II describes about speech corpus database. Section III and IV explain techniques used to experiment. Section V discusses experimental analysis. The result of experiments showing in Section VI. And lastly, conclusion describe by Section VII.

II. RELATED WORK

In the past decade, much study has been performed in the area of speech recognition for accent identification. The [3] describes Malaysian English accent identification. Where the system uses a standard implementation of features extracted using linear predictive coding (LPC), formant and log energy feature vectors. The accent identification of a speaker is predicated using K-Nearest Neighbors (KNN) classifier. This system gives 94.2% accuracy when features are fused as LPC, formats, and log energy. The [4] gives the investigation of acoustic characteristics of ethnically diverse accents in Malaysian English across genders. The system is developed using Linear Predictive coding for features and formant extraction. In addition, an analysis of variance (ANOVA) is also applied to investigate the accent according gender. The investigation was found that males and females differ in terms of all formants scores that correlate to accents in great details using two-way and one-way analysis of variance and the plots of normal fit of individual formant. In [12], they are trying to improve accuracy of speech recognition by accent with the experimental approach of acoustic speech feature for Marathi and Arabic accents for English speaking. The detailed study of acoustic correlates the accent using formant frequency, energy and pitch characteristics. The experimental results indicate that the fifth formant frequency found to be really effective for accent recognition. In [13], developed Speech Recognition system for Hindi language to recognize isolated word. Hidden Markov Model (HMM) is used to train and recognize the speech that uses MFCC to extract the features from the speech. The experimental results show that the overall accuracy of the presented system is 94.63%.

III. SPEECH CORPUS

For this preliminary survey, a speech corpus was compiled from 18 male and 12 female speakers. The speakers were from four different groups who mother tongue is Hindi. Marathi, Marwadi and Urdu. There are 900 samples of speech signals collected from these groups. We have formed 10 sentences in Hindi which will enable us to detect accent of the speakers from different language background. The sentence, structure and semantics of speech are a very important factor. Therefore, we have taken sentences from well-known novels such as Godan, Gaban and History Books, etc. which are syntactically and semantically correct. The Speech sample was recorded on visi-pitch is a designed to operate with generic sound devices for computers. Each sentence was uttered for 3 times each. The speech sample was recorded in the afternoon session at the same location in order to minimize channel effects. The age groups of speakers were between 20 and 35. All speakers were from the department of CS and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

The Computerized Speech Lab (CSL) is a speech and signal processing computer workstation of software and hardware used for research and clinical speech therapy.

The Computerized Speech Lab suite of software covers speech analysis, teaching, research, voice measurement, clinical feedback, acoustic phonetics, and forensic work. KayPENTAX, a division of PENTAX medical, offers two CSL models, 4500 and 4150B. The latest generation CSL hardware is an input/output recording device for a PC. CSL, is an integrated hardware and software product, and as such offers input signal-to-noise performance typically 20-30 dB superior to generic, plug-in sound cards [9].

Table1 shows the technical specification of concentrated parameter for database creation:

Sr.no	Parameter	Specification
1.	Sampling frequency	16Khz
2.	Distance from microphone	10cm
3.	Environment	Office
4.	Temperature	33.4
5.	Channel	Single
6.	Gender	Male:16 Female:12

Table2 shows the Hindi sentences used for Accent identification.

Sentence 1	प्रश्नकर्ता दिव्यवक्ता से अनेक प्रश्न पूछता था ।
Sentence 2	विश्व में सभ्यता का विकास ऊपर वर्णित सभ्यताओं से ही हुआ ।
Sentence 3	साम्राज्यवाद के कारण लोकतंत्र की समाप्ती हुई |
Sentence 4	इस जंगल में अधिकांश वृक्ष शाल के है |
Sentence 5	आज सब निरव है |
Sentence 6	दाताओं ने दान बंद कर दिया है।
Sentence 7	उनके सामने अकाल का भीषण रूप है।
Sentence 8	बंगाल में बर-बर कोहराम मच गया |
Sentence 9	भयानक गर्मी से पृथ्वी अग्निमय हो रही थी |
Sentence 10	उनके मन में भय का संचार हो रहा था |

IV. FEATURE EXTRACTION TECHNIQUES

Feature extraction step identifies the components of the audio signal that are good for identifying the linguistic content and discarding all the other remaining content which carries information like background noise, emotions, etc.. It extracts the features of the speech signal. There are many features extraction techniques. Here we are using Mel Frequency Cepstral Coefficient (MFCC) and Linear Predictive Coding (LPC) techniques.

1) Mel Frequency Cepstral Coefficient:

Mel Frequency Cepstral Coefficient is most prevailing and dominant feature extraction method to extract spectral features. It is based on frequency domain using the Mel scale which is based on the human ear scale [10].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

MFCC is an audio feature extraction technique which extracts parameters of the speech similar to the ones that are applied by humans to hear speech, while at the same time, deemphasizes all other data. The speech signal is first divided into time frames consisting of an arbitrary number of samples. In most systems overlapping of the frames is used to smooth transition from frame to frame. Each time frame is then windowed with Hamming window to eliminate discontinuities at the edges. The filter coefficients w (n) of a Hamming window of length n is computed according to the formula:

$$W(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \le n \le N-1.$$

= 0, Otherwise

Where N is the total number of sample and n is current sample. After the windowing, Fast Fourier Transformation (FFT) is calculated for each frame to extract frequency components of a signal in the time-domain. FFT is used to speed up the processing. The logarithmic Mel-Scaled filter bank is applied to the Fourier transformed frame. This scale is approximately linear up to 1 kHz, and logarithmic at greater frequencies. The relation between frequency of speech and Mel scale can be established as:

Frequency (Mel Scaled) = [2595log (1+f (Hz) /700]

MFCCs use Mel-scale filter bank where the higher frequency filters have greater bandwidth than the lower frequency filters, but their temporal resolutions are the same. The last step is to calculate Discrete Cosine Transformation (DCT) of the outputs from the filter bank. DCT ranges coefficients according to significance [8].

2) Linear Predictive Coding:

Linear prediction is a mathematical operation which provides an estimation of the current sample of a discrete signal as a linear combination of several previous samples. The prediction error, i.e. the difference between the predicted and actual value is called the residual [10]. It estimates the formants. The process of removing the formants is called inverse filtering, and the remaining signal is called the residue. This equation is called a linear predictor and hence it is called as linear predictive coding [7].

s (n)=-(A)* X(n-1) –A(3)* X(n-2)- A(N+1)* X(n-N)

n is the order of the prediction filter polynomial, a = [1 a(2) ... a(p+1)]. If n is unspecified, LPC uses as a default n = length (x) -1. If x is a matrix containing a separate signal in each column, LPC returns a model estimate for each column in the rows of a matrix and a column vector of prediction error variances. The n is the order of the prediction filter polynomial, a = [1 a(2) ... a(p+1)]. If n is unspecified, LPC uses as a default n = length(x)-1.

If x is a matrix containing a separate signal in each column, LPC returns a model estimate for each column in the rows of a matrix and a column vector of prediction error variances. The length of n must be less than or equal to the length of x [11].

V. ACOUSTIC FEATURES

Acoustic features can be defined in terms of the physical properties of the speech sound relevant to the feature.

a) Fundamental Frequency (F0):

The fundamental frequency (F0) in speech contains various types of non-linguistic information such as the speaker's identity, emotion and level of attention. They also indicate intonation in pitch accent languages. During the production of voice segment regular excitation of the vocal tract produces basically fundamental energy (F0) and it multiple combination. In the tonal language of speech allows expressing an emotion as well as accent information of speaker background [12].

b) Energy Characteristic:

Signal energy is seen as the strength or power or voice volume. It is an important feature that can show the differences of speaking style and structure of two different languages [12].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

VI. EXPERIMENTAL ANALYSIS

The experiment is performed by calculating MFCCs, LPC coefficients along with acoustic features F0 and energy values. All experiments perform with the help of Matlab and CSL software. For Accent Identification, statistical function is applied. Table 3,4,5,6 shows the calculated Mean and standard deviation of MFCC, LPC, Fundamental Frequency and Energy coefficients respectively. Euclidean Distance is applied on the result for analysis.

The following Table 3 shows mean and standard deviation of the Mel Frequency Cepstral Coefficient features. By calculating Euclidean distance from Hindi language, getting that, standard deviation of MFCCs of Marwadi language is closer to Hindi language MFCCs. Then Urdu is closer, and Marathi is far way from Hindi Language.

Table3: Mean and Standard Deviation of Mel Frequency Cepstral Coefficients

	Hindi		Marwadi		Marathi		Urdu	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Female Speaker1	4.99	1.63	2.87	2.34	1.68	2.37	4.41	1.78
Male speaker2	3.78	1.62	4.48	1.36	4.70	1.58	4.91	1.43
Male speaker3	5.04	1.70	4.29	1.48	4.16	1.24	3.99	1.12
Male speaker4	4.50	1.40	5.36	1.54	1.46	2.26	4.90	1.49
Average	4.58	**1.59**	4.25	**1.68**	3.00	**1.86**	4.55	**1.46**

Table 4 describes mean and standard deviation of Linear Predictive Coding Coefficients. Here we got the same result as MFCCs coefficients. The Marwadi language is closer to Hindi language. Then Urdu is closer and again Marathi is diverging from Hindi language.

Table4: Mean and Standard Deviation of Linear Productive Coding

	Hindi		Marwadi		Marathi		Urdu	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Female Speaker1	-0.08	1.71	-0.08	2.37	-0.08	2.95	-0.08	2.11
Male speaker2	-0.08	2.69	-0.08	2.10	0.93	1.71	-0.08	2.98
Male speaker3	-0.08	1.77	-0.08	1.73	-0.08	1.51	-0.08	1.72
Male speaker4	-0.08	1.53	-0.08	1.62	-0.08	2.12	-0.08	0.44
Average	-0.08	**1.93**	-0.08	**1.95**	0.17	**2.07**	-0.08	**1.81**

Table 5 gives mean and standard deviation of first fundamental frequency F0. Here F0 found higher in Urdu language then in Marathi and lastly in Marwadi with respect to Hindi Language.

Table5: Mean and Standard Deviation of Fundamental Frequency (F0)

	Hindi		Marwadi		Marathi		Urdu	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Female Speaker1	160.44	97.56	181.11	98.59	186.26	105.63	142.90	96.70
Male speaker2	117.83	63.41	106.18	60.45	119.26	60.92	123.06	63.38
Male speaker3	145.74	68.56	114.11	61.69	104.53	58.84	117.10	63.62
Male speaker4	103.28	55.07	109.23	61.82	97.23	62.24	98.56	55.57
Average	131.82	**71.15**	127.66	**70.64**	126.82	**71.91**	120.41	**69.82**

Table 6 illustrates the mean and standard deviation of basic energy of speech. Results show that energy level is greater in Urdu language, then Marathi language and found less in Marwadi than Urdu and Marathi with respect to Hindi language.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Table6: Mean and Standard Deviation of Energy Contour

	Hindi		Marwadi		Marathi		Urdu	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Female Speaker1	52.73	9.11	60.78	8.86	63.12	9.60	55.20	9.55
Male speaker2	61.28	9.45	71.82	10.59	60.47	8.71	57.88	10.21
Male speaker3	60.59	7.82	59.17	9.14	51.94	7.11	68.15	10.98
Male speaker4	54.69	8.62	54.68	8.61	57.75	9.00	62.23	10.61
Average	57.32	**8.75**	61.61	**9.30**	58.32	**8.61**	60.86	**10.34**

VII. RESULT

The experiment resulted in following observation:

1. In accent specific information retrieval Mel Frequency Cepstral Coefficients feature and Linear Predictive Coefficients were analyzed.
2. It is observed that in both techniques the Marwadi language speakers resemble more to Hindi accents. Even next closet class to Hindi is an Urdu whereas the Marathi is quite far from Hindi.
3. The fundamental frequency is high in the Urdu dataset, after that in Marathi and lastly in Marwadi.
4. The Energy level is high in Urdu Dataset, then in Marwadi dataset and lastly in Marathi.

VIII. CONCLUSION

The observation shows that the Marwadi language is closer to Hindi language, and then next language is Urdu. Marathi accent is more diverging from Hindi accent This is because Marwadi language having more lexical similarity with Hindi language and Marathi language, accent wise more diverging from Hindi language as Marathi speaker pronounce the Hindi word "पृथ्वी" as "Pruthvi" where Hindi speaker as "Prithvi". This study also shows the comparative study of MFCC and LPC techniques with Fundamental Frequency and Energy Contour. It turns out that there is no any remarkable difference between the results of both techniques. Basic energy and fundamental frequency found to be high in the Urdu Accent Dataset.

REFERENCES

1. (2015) The Omniglot website [online] . http://www.omniglot.com/writing/urdu.htm. Viewed 9 March 2015.
2. (2015) The Wikipedia website [online]. http://en.wikipedia.org/wiki/Languages_of_India . Viewed 9 March 2015.
3. Yusnita M.A., Paulraj M.P., Sazali Yaacob ,Shahriman Abu Bakar, A.Saidatul5. "Malaysian English Accents Identification using LPC and Formant Analysis", IEEE International Conference on Control System, Computing and Engineering 2011
4. Yusnita M. A., Paulraj M. P.b, Sazali Yaacobb, Nor Fadzilah M.a, Shahriman A. B., "Acoustic Analysis of Formants across Genders and Ethnical Accents in Malaysian English using ANOVA", International Conference On DESIGN AND MANUFACTURING, IConDM 2013.
5. Gaikwad S., Gawali B. Yannawar P., Mehrotra S., "Feature extraction using fusion MFCC for continuous marathi speech recognition", India Conference (INDICON), 2011 Annual IEEE.
6. Gawali B.W, Gaikwad S.,Yannawar P.,Mehrotra S. C., "Marathi Isolated Word Recognition System using MFCC and DTW Features", ACEEE Int. J. on Information Technology, Vol. 01, No. 01, Mar 2011.
7. Mehta L.R. ,Mahajan S.P. ,Dabhade A. S. , "COMPARATIVE STUDY OF MFCC AND LPC FOR MARATHI ISOLATED WORD RECOGNITION SYSTEM", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 6, June 2013
8. Dave N., "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", www.ijaret.org Volume 1, Issue VI, July 2013.
9. (2015) The Wikipedia website [online]. http://en.wikipedia.org/wiki/Computerized_Speech_Lab viewed 12 March 2015
10. Das B. P.,Parekh R., " Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", International Journal of Modern Engineering Research (IJMER) www.ijmer.com Vol.2, Issue.3, May-June 2012
11. Shinde R. B., Pawar V. P., " Isolated Word Recognition System based on LPC and DTW Technique", International Journal of Computer Applications (0975 – 8887) Volume 59– No.6, December 2012

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

12. Gaikwad S.,Gawali B.,Kale K.V., "Accent Recognition for Indian English using Acoustic Feature Approach", International Journal of Computer Applications (0975 – 8887) Volume 63– No.7, February 2013
13. Kumar K., Aggarwal R. K., "HINDI SPEECH RECOGNITION SYSTEM USING HTK", International Journal of Computing and Business Research ISSN (Online): 2229-6166 Volume 2 Issue 2 May 2011

BIOGRAPHY

Reena Hiralal Chaudhari is a M. Phil Research student in the Department of Computer Science and Information Technology of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, and Maharashtra, India.

Kavita Waghmare is a M. Phil Research student in the Department of Computer Science and Information Technology of Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, and Maharashtra, India.

Prof. Bharti Gawali is presently working as Professor in Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, MS, India. She is a Fellow and Hony. Secretary of IETE Aurangabad Center. She is Member of IEEE, IACSIT and ACM. She is reviewer and editor of several journals and Conferences at national & international level. She has organized several technical workshops and conferences. She sanctioned two major research projects. She got DST Fast Track Young Scientist Award and Shikshak Pratibha Puraskar . Her areas of specialization are Speech Processing, HCI, Pattern Recognition, Brain Computer Interface, Networking, Data mining, Biometric, Neuroscience and GIS etc.