The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages
29-31 August 2018, Gurugram, India

Hindi Speech Vowel Recognition Using Hidden Markov Model

Shobha Bhatt[1],*Dr.Amita Dev* [2], *Dr.Anurag Jain*[3]

[1] University School of Information and Communication Technology, Guru Gobind Singh Indraprastha University Delhi
[2]Indira Gandhi Delhi Technical University for Women,Delhi
[3] University School of Information and Communication Technology, Guru Gobind Singh Indraprastha University Delhi
bhattsho@gmail.com,amita_dev@hotmail.com,anurag@ipu.ac.in

Abstract

Vowel recognition is an important step for developing speech recognition system. The aim of this paper is to present Vowel recognition for Hindi language. Phoneme based speech recognition are widely used to overcome vocabulary constraint in developing large vocabulary continuous speech recognition system. Thus there is a need to explore vowel recognition related issues for obtaining better speech recognition results. Experiments were conducted using Connected word Hindi speech corpus for Speaker dependent mode using widely used Hidden Markov Model(HMM) based HTK Tool kit for both training and testing. Hindi Speech Corpus, made of 600 utterances spoken by 5 speakers, was used in this experiment. Mel Frequency Cepstral Coefficients (MFCCs) were used with 5 states monophone HMM model for feature extraction. Speech recognition process heavily depends on spoken language. Different Hindi speech characteristics were explored using formant analysis. Experimental results achieved as average vowel recognition scores of 77.12% for front vowels, 84.4% for middle vowels and 86% back vowels. Average vowel recognition score was achieved was 83.19%. Finally paper concludes with difficulties faced during system development and future development direction.

Index Terms: Vowel recognition, speech recognition, hidden markov model, MFCC

1. Introduction

Speech recognition is the process of converting spoken language in to the text. Speech recognition technologies can be applied to many fields such as voice operated devices, spoken dialog systems, automated dialing system.

Speech recognition using phones is very important because it overcomes the constraint of vocabulary size. Performance of Large Vocabulary speech recognition systems depend on how accurately vowels are recognized. Researchers have experimented to improve quality of phone recognition for developing improved speech recognition systems. Phone recognition can also be applied to speaker and language recognition, keyword spotting, music identification and translation systems [1].

In large vocabulary continuous speech recognitions sub-word models are used such as syllable base, phoneme base to overcome the constraints of vocabulary size. Then it is very important to decide which sub word units to be used and what are different characteristics of these sub words. Phonetic model plays important role for improving speech recognition results because of speech variability, different dialects [2].

This research work focuses on the need of vowel recognition for Hindi language. Hindi is one of the official languages of India. It is also spoken in other parts of the world. Hindi language is different acoustically and phonetically from English language. As the speech recognition systems advance from isolated word to continuous speech recognition system, complexity of the system increases. In this work vowel recognition from Connected word is explored. Connected word Hindi Speech Corpus with 5 state HMMs were used. Accuracy of speech recognition is dependent on spoken language; therefore there is a need to understand different issues related to spoken language. This includes knowledge of linguistic sounds, pronunciation variation, interpretation of the meaning of words, grammatical structure, possible meaning of sentences in the language. In order to understand language related issues different acoustic and phonetic characteristics of Hindi Speech are discussed for Hindi vowels using formant analysis explored. Twelve Mel Frequency Cepstral coefficients were extracted. Recognition scores were obtained for front, back and middle vowels.

Remaining part of the paper is prepared as follows: Section 2 describes HMM based speech recognition. Section 3 is about related work in the past. Section 4 explains Hindi Speech characteristics. Experimental set up is described in section 5. Section 6 is about results and discussions. Finally paper concludes with future direction and observations during the experimentation of this work.

2. HMM based Speech Recognition

Hidden Markov models (HMMs) are widely used statistical method for modeling time series data. HMM is a finite machine. HMMs are generated by defining probability distribution. HMMs are best suited for modeling variation in speech. Variability is modeled by associating probability density function with each state. The term used Hidden Markov Model indicates that only the output of random function related to each state is observed and underlying state cannot be seen. HMM based Speech recognition starts with preparation of speech corpus. Next step is pre-processing of the speech signal. The speech signal is converted into suitable representation. This process of suitable representation is called feature extraction. Mel Frequency Cepstral Coefficients (MFCCs) were used in this process. In the Next process acoustic models are generated. Hidden Markov Models are generated from extracted features. Language models are prepared. Using language and acoustic models, a most likely word sequence is calculated using dynamic algorithms [3-4].

10.21437/SLTU.2018-42

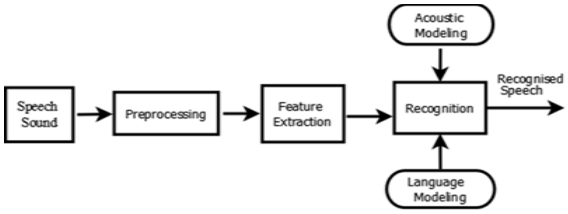

Fig1**:** *Basic block diagram of Speech Recognition System*

The speech signal is converted into observation sequence.

O = (o1, o2, o3…….on). The problem of phone based speech recognition is to find out most probable phone sequence given input speech observed sequence. In other words, to maximize the probability P(Ph/O). Bay's rule is used to solve the problem.

Ph* =arg Max (Ph)(P(Ph/O)

$$= \arg Max\ (Ph)(\ P\left(\frac{Ph}{O}\right) = \ argMax(Ph)(\frac{P(Ph)P\left(\frac{O}{Ph}\right)}{P(O)})$$

1. Here P(O) is fixed. It is the probability of observation sequence.
2. P(Ph) is calculated from language model
3. P(O/Ph) it is the probability of observing O given word sequence. It is calculated from the acoustic model [5-6].

3. Overview of related work

Researchers have conducted lot of research work in the past to explore vowel recognition. Lot of work has been done using vowel recognition for widely used TIMIT data base. In [7] researchers identified vowels from TIMIT database using temporal radial basis function. Experiments were performed using test and training data. The recognition accuracy reported is 98.06% for training data and 90.13% for testing data. In [8] experiments were performed to investigate the effect of different parameters, kernel tricks in vowel recognition. In [9] the authors implemented feed forward artificial network for recognition of vowels from TIMIT corpus. The recognition accuracy reported is 91.5%. In [10] authors presented vowel classification by analyzing dynamics of speech production in a reconstructed phase space using TIMIT data base. The experimental results reveal that results match with a classifier using MFCC. In [11] experiments were conducted by reducing the dimensionality of extracted features. To reduce dimensionality linear discriminant analysis and principal component analysis were used with Support vector machines. In [12] authors presented vowel classification by studying dynamics of speech production in a reconstructed phase space. The results match to the results produced using Mel Frequency Cepstral Coefficients (MFCCs). In[13] effort has been made to categorize Hindi Phoneme using TDNNs. Six neural networks were trained for broad Hindi phoneme classes. Recognition scores obtained for broad categories are 99% for vowel classes. In [14] authors investigated the effect of pre emphasis on Hindi vowel recognition using minimum distance classifier. It was observed that pre emphasis did not improve the recognition score. In [15] researchers proposed a method to improve consonant-vowel units for low bit rate coded speech. Experiments were conducted with samples from Telugu database using two levels one for vowel category and other for consonant category.

4. Hindi Speech Characteristic

Hindi is one of the official languages of India. It is widely spoken language in India. Phonemes are abstract linguistic unit in any language. Hindi language alphabets are broadly classified as Hindi consonants and vowels. Further division is made on the basis of production manner and articulation place namely vowels, semivowels, fricatives and stop consonants .Vowel set also contains diphthongs [16]. Hindi consonants differ from English. Stops and affricate use both voicing and aspiration. Aspiration, gemination, nasalization and retroflexives are different characteristics of Hindi. For aspiration generally suffix h is used to denote aspiration k versus kh consonants. For retroflex consonants suffix x is used to differentiate t and tx consonants. Nukta /bindu use suffix q. Nukta is a dot which is used below Hindi phonemes. Bindu is also a dot that is used above a phoneme. Nasalized vowels use suffix n to nasalize the vowel kaha versus kahan. Geminated sounds are labeled by repeating the single consonant pakaa and pakka [17].Here k,kh,x,tx,h are Hindi consonants and their equivalent IPA notations are given below in Table1.

Table 1*: Example Hindi consonants with their aspirated Retroflexive parts*

Label	Hindi	IPA Symbol
K	क	k
Kh	ख	k^h
t	त	t̪
tx	ट	ʈ

For Hindi Speech recognition previous work related to isolated, connected word, continuous speech has been experimented. Different acoustic models, different Gaussian mixture, context dependent, context independent and different states of hidden markov models have been included to improve Hindi Speech recognition[18-20]. Table2 shows Hindi vowel acoustic classification for indigenous vowels.

Table 2*: Hindi Vowel Acoustic classification* [12],[23]

Front		Middle		Back	
Hindi vowels	**IPA**	**Hindi Vowels**	**IPA**	**Hindi Vowels**	**IPA**
इ	ɪ	अ	ɑ	उ	ʊ
ई	i:	आ	ɑ:	ऊ	u:
ए	e:			ओ	o:
ऐ	ɛ:			औ	ɔ:

Table 3 shows nasalized vowels and breathy vowel. The first row in Table 2 shows all nasalized vowels and breathy counter part of vowel /ɑ/.

Table 3: *Hindi nasalized and breathy counterpart* [23]

IPA symbol					Hindi Vowels
ŋ̃	ɲ̃	ɳ̃	m̃	◌̃	अं
ɦ					अः

The given below Table 4 shows borrowed vowel from highly educated Sanskrit speaking people.

Table 4: *Hindi borrowed vowel*[23]

Hindi Vowels	IPA
ऋ	ɾɪ

4.1 Hindi vowel formant analysis

This section gives overview of formant analysis for better understanding of Hindi vowels. In some research work formants were used for vowel recognition. Formants are defined as resonating frequencies of vocal tract [21]. As per the frequency bands, formants for the vowels can be classified in to first formant F1, second formant F2 and third formant F3. Formants can be observed by dark bands in wide band spectrogram. Analysis of these formants for recognition of the vowels play very important role. Fig 2 depicts mean formant frequencies of Hindi vowels spoken by male child. These research findings support literature regarding importance of F1 and F2 for discrimination among vowels.

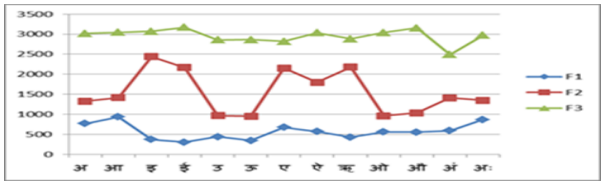

Fig 2: Mean Formant frequencies of Hindi Vowels spoken by male child

Fig 2 illustrates that F1 is high for middle vowel /आ/ and minimum for front vowel /ई/.F2 is high for front vowel /इ/ and minimum for back vowel /ऊ/. Further F3 can be used to determine the phonemic quality of the speech [21]. However this paper does not employ formant analysis in vowel recognition.

5. Experimental Setup

Experiments were performed using Connected word Hindi Speech Recognition system for speaker dependent mode using Hidden Markov Model based HTK tool kit version 3.4.1 in Windows environment [22]. Connected word Hindi speech corpus made of 600 utterances spoken by 5 speakers, was used in this experiment. All the recordings were done in room environment with wave surfer in wav file formats.For implementation purpose the following sub steps were performed. For acoustic modeling Monophones based 5 state HMMs were used. For feature extraction Mel Frequency cepstral coefficients were used. In Monophones based acoustic modeling only single phone is considered.

1. Hindi Speech corpus was divided into testing and training data and Speech wave files were converted in to Mel Frequency Cepstral Coefficients (MFCC) feature vectors. Lexicon created. Grammar was defined. Word nets were created using Hparse. Monophones based transcriptions were generated.
2. First monophones HMMs were generated with silence fixing and re-estimation Final model was prepared after re-estimation.
3. The test data was evaluated against final models.
4. Results were generated for analysis.

In the following section data preparation is the elaboration of step1. Step 2 is described under the heading training of HMMs Step3 is described under heading recognizing test data.

5.1 Data Preparation

Speech corpus was divided in to training and testing. Pronunciation dictionary was developed for the Hindi language sentences. Grammars definition files were created. HTK Tool kit provides tool for defining task grammar. The grammar defines allowable words, sequence of words or phones to be used. Word nets were created from grammar file for further use in recognition process. Mel Frequency Cepstral coefficients were extracted for feature vectors. Monophone transcriptions were generated. Table5 describes the details of MFCC features used.

Table 5*: Coding parameters for feature extraction MFCC*

S.No.	Parameter	Value of
1.	Features Extracted	MFCC
2.	Window length	25ms
3.	Frame periodicity	10ms
4.	Number of coefficients	12
5.	Window used	Hamming
6.	Pre-emphasis	0.97
7.	Filter Bank channels	26
8.	Cepstral liftering coefficients	22

5.2 Training HMMs

First set of monophone HMMs are generated. Initial sets of phone models are created. First a proto type model was defined. All HMMs have five states. Each HMM state was represented by single Gaussian having mean, variance and mixture weight. The HTK tool kit module HCompv generates flat start HMMs for all phones. Global speech mean and variance is calculated from MFCC features extracted from speech corpus. All initial phone models have same state mean and variance equal to global speech mean and variance. These HMMs were re-estimated till convergence to produce trained model for monophones.

5.3 Recognizing the test Data

In this step test data is matched to trained model for recognition. The speech wav files to be tested are first converted in to feature vectors (MFCCs) using HCopy tool. These feature vectors are used to recognize the speech using Hvite command.

6. Results and Discussion

Hindi vowels were recognized for speaker dependent mode using Connected word Hindi Speech data base. Recognition results scores were obtained with HResult tool of HTK.

Percent Correct = (N −D−S)/ N ×100

Percent Accuracy = (N −D−S−I)/ N ×100%

Here D is deletion error, I is insertion error and S is substitution error, N is total number of labels in reference transcription. Table 6 presents vowel percentage correct(%c) which is correctly recognized vowels in a row and percentage of incorrectly recognized vowels (%e) in the row as a percentage of total vowels. The vowels have average percentage correct score of 81.93%.

Table 6*: Vowel Recognition Score*

Vowels	IPA	[%c / %e]
आ	ɑ:	[68.8/0.8]
ऐ	e:	[72.5/0.2]
ए	ɛ:	[84.6/0.1]
इ	ɪ	[82.6/0.1]
ई	i:	[68.8/0.5]
औ	ɔ:	100
अ	ɑ	100
उ	ʊ	[76.1/0.2]
ओ	o:	[83.5/1.1]
ऊ	u:	[84.4/0.1]
Average		83.19

Table 7 depicts category wise average vowel correct (%c) score for front, middle and back. The average vowel recognition scores are 77.12% for front vowels, 84.4% for middle vowels and 86% back vowels. Highest score was achieved for back vowels. Lowest score was achieved with front vowels.

Table 7*: Hindi vowel recognition category wise*

SNo	Hindi Vowel recognition	Average %c
1.	Front(इ,ई,ए,ऐ)	77.12
2.	Middle(अ,आ)	84.4
3.	Back(उ,ऊ,ओ,औ))	86

7. Conclusions

In this work Hindi vowel recognition was investigated to understand and improve speech recognition process. Recognition system was developed using HMMs. Widely used HMM based tool kit was used for implementation. The system mainly divided in to data preparation, training and testing phase. Hindi Connected word speech corpus was used in this study. Results for Hindi vowels were generated and analyzed. Average recognition scores were evaluated and analyzed. Further subcategory wise recognition score were obtained. Further these research findings may be applied for the development of large vocabulary and Hindi continuous Speech recognition system. It is also recommended to use different dialects and pronunciation dictionaries to explore different aspects of phoneme recognition. Other research work may include use of different language models to explore effect of language modeling in vowel recognition.

8. Acknowledgements

The authors would like to acknowledge the Ministry of Electronics & Information Technology (MeitY), Government of India, for providing financial assistance for this research work through “Visvesvaraya Ph.D. Scheme for Electronics & IT”.

9. References

[1] Lopes, Carla, and Fernando Perdigao. "Phoneme recognition on the TIMIT database." Speech Technologies. InTech, 2011.

[2] Seide, Frank, and Nick JC Wang. "Phonetic modeling in the Philips Chinese continuous-speech recognition system." Proc. ISCSLP. Vol. 98. 1998.

[3] Poonam Bansal, Amita Dev &Shail Bala Jain(2008) Optimum HMM combined with vector Quantization for Hindi Speech word Recognition, IETE Journal of Research, 54:4, 239-243

[4] Kumar, Kuldeep, R. K. Aggarwal, and Ankita Jain. "A Hindi speech recognition system for connected words using HTK." International Journal of Computational Systems Engineering 1.1 (2012): 25-32.

[5] Sinha, Shweta, Shyam S. Agrawal, and Aruna Jain. "Continuous density Hidden Markov Model for context dependent Hindi speech recognition." Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on. IEEE, 2013.

[6] S. J. Young, G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. C. Woodland, The HTK Book (for HTK version 3.4), 2006. Available: http://htk.eng.cam.ac.uk/ docs/docs.shtml

[7] Guezouri, Mustapha, Larbi Mesbahi, and Abdelkader Benyettou. "Speech Recognition Oriented Vowel Classification Using Temporal Radial Basis Functions." arXiv preprint arXiv:0912.3917 (2009).

[8] Amami, Rimah, Dorra Ben Ayed, and Noureddine Ellouze. "Practical selection of SVM supervised parameters with different feature representations for vowel recognition." *arXiv preprint arXiv:1507.06020* (2015).

[9] Merkx, Peter, and Jadrian Miles. "Automatic vowel classification in speech." *An Artificial Neural Network Approach Using Cepstral Feature Analysis* (2005): 1-14.

[10] Liu, Xiaolin, Richard J. Povinelli, and Michael T. Johnson. "Vowel classification by global dynamic modeling." *ISCA Tutorial and Research Workshop on Non-Linear Speech Processing*. 2003.

[11] Wang, Xuechuan, and Kuldip K. Paliwal. "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition." *Pattern recognition* 36.10 (2003): 2429-2439

[12] Liu, Xiaolin, Richard J. Povinelli, and Michael T. Johnson. "Vowel classification by global dynamic modeling." *ISCA Tutorial and Research Workshop on Non-Linear Speech Processing*. 2003.

[13] Dev, Amita, S. S. Agrawal, and D. Roy Choudhury. "Categorization of Hindi phonemes by neural networks." AI & SOCIETY 17.3-4 (2003): 375-3

[14] Paliwal, Kuldip K. "Effect of preemphasis on vowel recognition performance." speech communication 3.1 (1984): 101-106.

[15] Vuppala, Anil Kumar, K. Sreenivasa Rao, and Saswat Chakrabarti. "Improved consonant–vowel recognition for low bit-rate coded speech." International Journal of Adaptive Control and Signal Processing 26.4 (2012): 333-349.

[16] Chandra, Mahesh. "Hindi Vowel Classification using QCN-PNCC Features." Indian Journal of Science and Technology 9.38 (2016).

[17] https://www.iitm.ac.in/donlab/tts/downloads/cls/cls_v2.1.6.pdf

[18] Pruthi, Tarun, Sameer Saksena, and Pradip K. Das. "Swaranjali: Isolated word recognition for Hindi language using VQ and HMM." International Conference on Multimedia Processing and Systems (ICMPS). 2000.

[19] Aggarwal, Rajesh Kumar, and Mayank Dave. "Integration of multiple acoustic and language models for improved Hindi speech recognition system." International Journal of Speech Technology 15.2 (2012): 165-180.

[20] Kumar, Mohit, Nitendra Rajput, and Ashish Verma. "A large-vocabulary continuous speech recognition system for Hindi." IBM journal of research and development 48.5.6 (2004): 703-715.

[21] Alotaibi, Yousef Ajami, and Amir Hussain. "Comparative analysis of Arabic vowels using formants and an automatic speech recognition system." International Journal of Signal Processing, Image Processing and Pattern Recognition 3.2 (2010): 11-22.

[22] Official site of HTK toolkit. Available: http://htk.eng.cam. ac.uk

[23] Kachru, Yamuna. Hindi. Vol. 12. John Benjamins Publishing, 2006.