Automatic Speech Recognition of Fricative Class Phonemes in Malayalam Language

Cini Kurian

Associate Professor, Al-Ameen College, Edathala, Aluva

Abstract - Speech recognition system keeps elderly, physically challenged especially blind people closer to the Information technology revolution. Speech recognition benefits a lot in manufacturing and control applications where hands or eyes are otherwise occupied. Speech enabled applications in public areas such as; railways, airport and tourist information centers might serve customers with answers to their spoken query. Such tantalizing applications have motivated research in automatic speech recognition(ASR) since 1950's. Great progress has been made so far, especially since 1970's, using a series of engineered approaches that include template matching, knowledge engineering, and statistical modeling [1]. Yet computers are still nowhere near the level of human performance at speech recognition, and it appears that further significant innovation requires serious research. In this paper we discuss the results of the fricative phonetic class wise speech recognition performance of Malayalam language

Keywords: Automatic speech recognition , Malayalam, Plosive class words.

1. INTRODUCTION

In speech technology and phonetics , speech signal is the basic material for study and analysis. Speech signal has spectral characteristics which differ dynamically within a stretch of a few milliseconds. These temporarily varying spectral chunks can be grouped into abstract classes called allophones in phonetic study. But since in most of the languages, phonemes and letters have a varying degree of correspondence, the allophones are mapped into its higher classes called phonemes. Indian languages have more direct relation between sounds and letters[2]. The correlation of the phonemes and letters , especially , in the Indian languages, make the speech research fairly easy compared to English. But in India, these concept cannot be similar for all languages. Many of the phonemes has similar realization[3]. Phonetic realization of all the phonemes will not be the same across languages. Some language has some peculiar phonemes and some phonemes has different realization.

In this paper we discuss the results of the phonetic class wise speech recognition performance of Malayalam language. The words have been collected in such a way that, each word contains at least any of the phonemes of each class. Maximum care has been taken in the selection of words, so that all phonemes of the particular class occurred in all the word positions.

Commonly, in the selection of text corpus for speech recognition, some phonemes never likely to occur and the words which includes such phonemes will never be recognized properly[4]. Therefore special care to be taken to include all phonemes of a language in maximum word positions(Start, end , middle). Our ultimate aim in this work is to develop a absolute speech recognizer for Malayalam Language[5]. Towards this objective we have chosen words in the above manner to include all phonemes of the language in all co-articulatory positions[6]. Thus this work will resolve co-articulation problem to some extent which is a superior challenge for a speech recognizer.

2. PHONETIC CHART

Malayalam has 52 consonant phonemes, encompassing 7 places of articulation and 6 manners of articulation, as shown in Table 6.1 below. In terms of manner of articulation, plosives are the most complicated, for they demonstrate a five-way distinction in bilabials, dentals, alveolar-palatals, retroflex, and velars. A bilabial plosive, for example, is either voiceless or voiced. Within voiceless bilabial plosives, a further distinction is made between aspirated and unaspirated ones whereas for voiced bilabial plosives the distinction is between modal-voiced and breathy-voiced ones. The same five-way distinction is also found in dental, alveolo-palatal, retroflex, and velar plosives. In terms of place of articulation, on the other hand, alveolars are the most complex because they involve all manners of articulation except for affricate[6]. Phonetic chart as presented by Kumari, 1972 [7] for Malayalam language is given in table 1.

For all speech sounds, the basic source of power is the respiratory system pushing air out of the lungs. Sounds produced when the vocal cords are vibrating are said to be voiced, where the sound produced when the vocal cords are apart are said to be voiceless[8]. The shape and size of the vocal tract is a very important factor in the production of speech. The parts of the vocal tract such as the tongue and the lips that can be used to form sounds are called articulators. The movements of the tongue and lips interacting with the roof of the mouth(palate) and the pharynx are part of the articulatory process[9].

Table 1. Phonetic chart of Malayalam

		Labial		Dental		Alveolar		Retroflex		Palatal		Velar		Glottal
		voiced	unvoiced	voiced	Unvoiced	voiced	unvoiced	voiced	Unvoiced	voiced	unvoiced	voiced	unvoiced	
Stop	un aspirated	പ p	ബ b	ത t	ദ dh	ററ t't'		ട d	ഡ d'	ച ch	ജ j	ക k	ഗ g	
	aspirated	ഫp h	ഭ bh	ഥ th	ധ h			ഠ d'h	ഢ h'	ഛ c h'	ഝ jh	ഖ k h	ഘ gh	
Nasals		മ m		ന n		ന n1		ണ ṇ'		ഞ ñ		ങ ṅ		
Fricative		ഫ f		സ s				ഷ sh		ശ s				ഹ h
Lateral						ല l		ള l'		ഴ zh				
rhotic						ര r'		റ r						

3. FRICATIVE CLASS

Fricative sounds are produced when the articulators are brought so closely together so that the sounds are accomplished by audible friction[10].

The four fricatives of Malayalam and their allphonic distribution are shown below:

[s] Alveolar voiceless fricative occurs initially and medially.

E.g. /samam'/ 'equal' , /vistaar'am' / 'area'

[sh] Retroflex voiceless fricative occurs initially and medially.

E.g. /s'haNmukham'/ 'a person name name' , /vish'am' / 'poison'

[š] palatal voiceless fricative occurs initially and medially.

E.g /shakti/ 'strength' , /vasham' / 'side' , /vishvam' / 'world'

[h] Velar voiceless fricative occurs medially after nasals.

E.g. /simham'/ 'lion'

[h] Velar voiceless fricative occurs initially and medially.

E.g. /haar'am' / 'garland' . /de'ham/ 'body'

4. DATABASE DESIGN

We have collected words in such a way that each phoneme should occur in initial or , medial or final positions in the word. In all the positions the phonemes are succeeded by the maximum possible vowels. Position wise listing of number of words of each phoneme is detailed in table 1. As an example for the phoneme /p/ , in the initial position, we have the words as shown below.

Table 2: Number of words of each phoneme (position wise listing)

	Number of words with phonomes in			total
	Initial position	Medial position	Final position	
glottal fricative ഹ	9	6		15
Dental fricative സ	12	11		23
retroflex fricative ഷ	4	11		15
palatel fricatives ശ	11	6		17
Total				70

6. METHODOLOGIES USED

Speech recognizer for all the above classes of words separately. Semi continuous, Context dependent tied state HMM's with 3 state per HMM[1] and 8 Gaussian per state were used for modeling. MFCC[11] were used for feature extraction and trigram models used for language modeling. In each phonetic class wise recognizer, speech corpus contain 25 speakers' data, out of which training is performed by 20 speakers' data and testing by 5 speakers' data. The experiment was conducted using 5 fold validation test as explained in section and the results were analyzed using the performance metric WER[9] using sclite from NIST[12].

7. RESULTS AND DISCUSSIONS OF SPEECH RECOGNITION OF FRICATIVE CLASS WORDS

The words in this category consists of 205 words which include 26 labial voiced stop words 20 labial unvoiced

stop words , 20 dental voiced stop words, 20 dental unvoiced stop words, 23 velar voiced stop words, 19 velar unvoiced stop words, 21 retroflex unvoiced stop words, 20 palatal voice stop words, and 22 palatal unvoiced stop words.

The speech performance results of these words in training and testing modules are as shown in table 3. The average accuracy obtained for test data is 82%.

Table 3: Speech Recognition results words having fricative class phonemes

Sl.No	Training %	Testing %
1	93	80.15
2	94.28	83.78
3	94.12	83.5
4	93.14	80.79
Average	93.635	82.055

7. SUMMARY

In this paper we have presented our effort of developing a automatic speech recognizer for fricative class of phonemes. This work can be treated as a contribution to the domain of Isolated word recognizer for Malayalam language, since a total of 70 words have been identified and recognized .

REFERENCES

[1] Furui, S., “50 Years of Progress in Speech and Speaker Recognition Research Identification”, In ECTI Transformations n Computer and Information Technology, vol. 1, no. 2, 2003

[2] Sorin Dusan and Larry R. Rabiner, “On integrating insights from human speech perception into automatic speech recognition,” in Proceedings of INTERSPEECH 2005, Lisbon, 2005.

[3] HILL, D. R. (1971). Man-machine interaction using speech. In Advances in Computers, 11. Eds F. L. Alt, M. Rubinoff & M. C. Yovitts, pp. 165-230. New York: Academic Press.

[4] Furui, S., “50 Years of Progress in Speech and Speaker Recognition Research Identification”, In ECTI Transformations on Computer and Information Technology, vol. 1, no. 2, 2003

[5] Balaji. V., K. Rajamohan, R. Rajasekarapandy, S. Senthilkumaran,"Towards a knowledge system for sustainable food security: The information village experiment in Pondicherry," in IT Experience in India : Bridging the Digital Divide Kenneth Keniston and Deepak Kumar, eds., New Delhi, Sage,2004.

[6] Madhuresh Singhal et al. ‘Developing Information Technology Solutions in Indian Languages: Pros and Cons’. At 1st International CALIBER: Mapping Technology on Libraries and People, 13-15 Feb. Ahmadabad, India, pages 655-666, 2003.

[7] Namboothiri, E.V.N. 2002. Bhashavinjaneeyam. Calicut: Poorna Publications

[8] Punnoose, R. (2010). An Auditory and Acoustic Study of Liquids in Malayalam. Ph.D. Thesis, Newcastle University, Newcastle, UK

[9] G. Doddington, (1989), "Phonetically Sensitive Discriminants for Improved Speech Rec.", Proc. IEEE Int Conf. Acoustics. Speech and Sig. Proc., ICASSP-89, pp. 556-559, Glasgow, Scot- land.

[10] L.,R Rabiner, "A tutorial on Hidden Markov model and selected application in speech recognition", Pro.IEEE,7(2):257-286, February 1998

[11] Hwang, M. Y. (1993). Sub-phonetic acoustic modeling for speaker-independent continuous speech recognition. Ph.D. Thesis. Carnegie Mellon University.

[12] C.H Lee, L.R Rabinar , R. Pieraccini, and J.G Wilpon , " Accoustic Modelling for Large Vocabulary speech Recognition", Computer speech and Language , 4:127-165,1990 Vocabulary speech Recognition ", Computer speech and Language , 4:127-165,1990.