Speech Recognition performance of Dental Nasal an Retroflex nasal Phonomes of Malayalam Language

Cini Kurian [#1]
[#]*Associate Professor, Al-Ameen College, Edathala,Aluva,Kerala, India*

Abstract *— Interaction with computer through a convenient and user-friendly interface has always been an important technological issue. Machine-oriented interfaces [1] restrict the computer usage to a minuscule fraction of the population, who are both computer literate and conversant with written English. Computers which can recognize speech in native languages enable common man to make use of the benefits of information technology. In this paper speech recognition performance of Dental Nasal and Retroflex Nasal phonemes of Malayalam language have been explored.*

Keywords *— Automatic Speech Recognition, Malayalam language*

I. INTRODUCTION

Speech recognition system keeps elderly, physically handicapped and 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. It has large application for use over telephone, including automated dialing, telephone directory assistance, spoken database querying for novice users, voice dictation systems like medical transcription applications, automatic voice translation into foreign languages etc. 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. Yet computers are still nowhere near the level of human performance at speech recognition, and it appears that further significant innovation requires serious research/studies.

Automatic speech recognition technology needs knowledge from multidisciplinary areas like Acoustics, Linguistics, Biology, Physiology, Cognitive Science, Intelligence, Artificial Intelligence, Electrical Engineering. Computer science, Digital signal processing Mathematics and Statistics.

For most of literate languages, phonemes and letters in their scripts have varying degrees of correspondence[13]. Since such a relationship exists, a major part of a speech technology deals with the correlation of script letters with time-varying spectral stretches in that language. Indian languages said to have more direct correlation between their sounds and letters. Such similarity gives a false impression of similarity of text-to-sound rule across these languages. A given letter which is parallel across various languages may have different degrees of divergence in its phonetic realization in these languages.

II. LITERATURE SURVEY OF MALAYALAM SPEECH RECOGNIZERS

Malayalam is one among the 22 languages spoken in India with about 38 million speakers. The language has 37 consonants and 16 vowels. There are different spoken forms in Malayalam although the literary dialect throughout Kerala is almost uniform

SYAMA [148] built an isolated word and speaker independent speech recognition system for Malayalam. Microsoft Visual Studio was used for compiling HTK and Active Perl as interpreter. Accuracy of this system is just 62%. Vimal Krishnan et.al[149] developed a small vocabulary (5 words) speech recognition using 4 types of wavelet for feature extraction and Artificial neural network technique (ANN) is used for classification and recognition purpose. By using this method they have achieved a recognition rate of 89%. Raji kumar et. al has presented recognition of the isolated question words from Malayalam speech query using DWT and ANN [150]. A recognition accuracy of 80% has been reported. A small vocabulary speech recognizer has been developed by Anuj Mohamed, et.al [151] using Hidden Markov Models and MFCC. The system has produced 94.67% word accuracy. Sonia Sunny et.al worked on the speech recognition of isolated 20 words of Malayalam language and reported results in three papers [152,153,154] i.e comparative study of two wavelet based feature extraction methods ; comparison of LPC and DWT for speech recognition ; and the accuracy of the speech recognizer with DWT and ANN

III. SPEECH RECOGNITION APPROACHES

Automatic speech recognition system is the method to transform or produce a sequence of text/messages from a speech signal. This process is called decoding. Speech signal is decoded and then converted into writing (e.g. dictation machine) or commands to be executed (e.g. hands free dialing). There are three classical approaches for speech recognition technology. They are, Acoustic-phonetic approach, pattern recognition approach and artificial intelligence approach[102].

i) Statistical Pattern Recognition approach

Under the pattern-recognition approach, the speech patterns are used directly without explicit feature determination and segmentation. There are two main steps under this approach: Training of speech patterns and recognition of patterns via pattern comparison. Speech knowledge is supplied into the system via the training procedure. Most of the current and modern ASR systems are based on the principles of statistical pattern recognition. As shown in Figure 3.1, speech recognition using statistical pattern recognition paradigm has four steps in addition to the two main steps mentioned above.

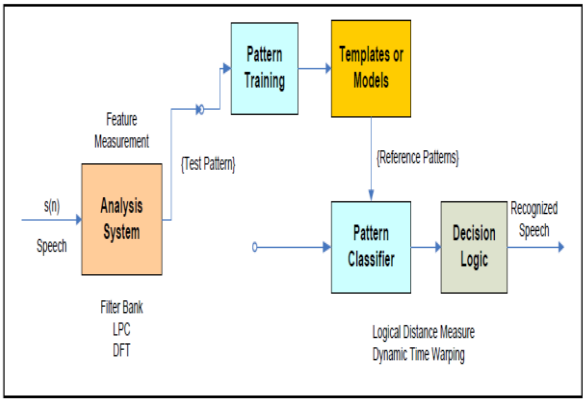

Figure 3.1: Diagram of statistical pattern speech recognition system (Rabiner & Juang, 1993)

Firstly, features are extracted from the input signal and represented into a form of features. There are a number of spectral analysis techniques, such as filter bank analyzer, Discrete Fourier Transform (DFT) analysis, Linear predictive coding analysis (LPC), Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Predictive Cepstral Coefficient (PLPCC) analysis [102]. Secondly, in pattern training one or more test patterns corresponding to speech sounds of the same class are used to create a pattern representatives of the features of that class. The resulting pattern, generally called reference pattern, derived from some type of averaging technique, or it can be a model that characterizes the statistics of the features of the reference pattern. This is followed by pattern classification process. Here, the unknown speech input is compared with reference pattern and a similarity (distance) between the training and testing pattern is computed. Lastly, decision logic is applied to decide which reference best matches the unknown test pattern. Based on Rabiner & Juang 's [102] research, pattern recognition in speech recognition has its strengths and weaknesses which are detailed as follows[102,168]

i. The system performance is sensitive to the amount of training data available for creating sound class reference patterns. Generally; the more the training data the higher the performance of the system..
ii. The reference patterns are sensitive to the speaking environment and transmission characteristics of the medium used to create the speech; this is because the speech spectral characteristics are affected by transmission and background noise.
iii. The computational load for both pattern training and pattern classification is generally linearly proportional to the number of patterns being trained or recognized; hence, computation of a large number of sound classes could and often become prohibitive.
iv. Because the system is insensitive to sound class, the basic techniques are applicable to a wide range of speech sound, including phrases, whole words, and sub-word units. A basic set of techniques developed for one sound class (e.g. words) can generally be directly applied to different sound classes (e.g., sub-word units) with little or no modifications to the algorithms.
v. It is relatively straightforward to incorporate syntactic (and even semantic) constraints directly into the pattern recognition structure, thereby improving recognition accuracy and reducing computation.

Hence, pattern recognition approach is the most common approach applied in most current ASR systems.

IV. Hidden Markov Model (HMM)

The Hidden Markov Models (HMMs) are widely used statistical tools in recognition system. It covers from isolated speech recognition to very large vocabulary unconstrained continuous speech recognition and speaker identification fields. Therefore most of the current speech recognitions are conducted based on Hidden Markov Model (HMMs). The Hidden Markov Model is a statistical model where the system is modelled as Markov process which can be represented as a state machine with unknown parameter through it [179]. The main important process is to determine the unknown parameter from the observable parameter. After

determining the parameter, it then used to perform further analysis. One of the examples of the Hidden Markov Model is normally use in pattern application. For example, Hidden Markov Model in pattern application is speech, signature, handwriting, gesture recognition and bioinformatics and genomics in medical.

V. SPEECH DATA BASE AND SYSTEM DEVELOPMENT

Speech data collected from the age group of 20 to 45 years keeping almost equal male and female ratio. Speakers were requested to read the word /sentences in a normal reading manner. . Speech was recorded in normal office environments. A headset which contains microphone with 70 Hz to 16000 Hz of frequency range is used for recording. The recording is done with 16 kHz sampling frequency quantized by 16 bit, using a tool named CoolEdit in Microsoft wave format.. 21 speakers (10 male and 11 female) spoke the words. Transcription file is created for each utterance of the speaker and a language dictionary is created for each word in the string. These are stored in separate files.

i). Creation of Pronunciation Dictionary

In pronunciation dictionary all words in the training data to be mapped onto the acoustic units which are defined in the phone list. Theoretically, creation of phonetic dictionary is just a mapping of grapheme to phoneme. But this alone would not work especially for a language like Malayalam as many phonemes pronounced differently in different contexts. For example, ഫ (ph'a) pronounced differently in ഫലം (/ph'alam/-fruit) and ഫാൻ /ph'aan'/ - fan) . ന (n1a and na - Nasal dental and Nasal alveolar) is pronounced differently even though the grapheme notation is same (eg. നനക്കുക(/n1anaykkuka/- watering). Hence for creating pronunciation dictionary, initially mapping have been completed for all grapheme into the corresponding phoneme units. Then some phonological rules have been applied manually and edited the dictionary. Multiple pronunciations are also incorporated in the dictionary.

VI. DENTAL NASAL

i) Design of database

For the analysis of dental nasal, the minimal pair of words have been designed as below. We have selected two minimal pairs with retroflex nasal. Since Alveolar nasal and dental nasal have the same writing script, these two phonemes were of greater interest for analysis. But unfortunately, we could not find any minimal pairs with these phonemes. Hence we could not carry out the speech recognition analysis. For these pairs, we restricted the work with the acoustic and auditory analysis.

a) Dental nasal vs. Retroflex nasal

- പനി , പണി (/pan1i/ - fever , pan'i - work)
- ആനി , ആണി (/aan1i/ – a female name , / aan'i / - nail)

b) Dental nasal vs. retroflex nasal

Since there is no minimal pairs in this category, the word /n1anavu' / for analysis

- നനവ (/n1anavu'/ - wetness)

VII. Dental nasal vs. Retroflex nasal – Speech Recognition performance

For evaluating the speech recognition performance, test data which contain total of 20 words spoken by 5 speakers is considered. The training data base has 80 words. The performance of the test data is depicted below (table 1.1). 20% of /n1i/ seems to be confusing with /n'i/ and 40% of /n'i/ confusing with /n1i/

Table 1.1Confusion matrix - speech recognition performance of Dental nasal vs. Retroflex nasal

	n1i	n'i	others	total
n1i	8	2	0	10
n'i	4	6	0	10
total	12	8	0	20

iii) Conclusion

There is a confusion between them in recognition of these two phonemes in the speech recognition performance. Hence dental nasal and retroflex nasal may mislead the speech recognizer because of the similarity in phonetic features.

Hence it is concluded that the dental nasal which is a unique phoneme of Malayalam language has similar features with that of retroflex nasal. But dental nasal and alvelor nasal shows divergence in auditory analysis. These factors have to be considered while preparing phonetic dictionaries for speech recognizers of Malayalam language.

REFERENCES

[1] Syama, R. and Mary Idicula, S., "*Speech Recognition for Malyalam Language*", 2008.

[2] Vimal Krishnan V.R Athulya Jayakumar Babu Anto.P - 2008 , Speech Recognition of Isolated Malayalam Words Using Wavelet Features and Artificial Neural Network -

[3] Raji Sukumar.A Firoz Shah.A Babu Anto.P -2010- Isolated question words Recognition from speech queries by Using artificial neural networks

[4] Sonia Sunny, David Peter S., and K Poulose Jacob A Comparative Study of Wavelet Based Feature Extraction Techniques in Recognizing Isolated Spoken Words," International Journal of Signal Processing Systems, Vol.1, No.1, pp. 49-53, June 2013.doi: 10.12720/ijsps.1.1.49-53

[5] Continuous Malayalam speech recognition using Hidden Markov Models.Anuj Mohamed, K. N. Ramachandran Nair Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India, September 16-17, 2010, Tamilnadu, India; 01/2010

[6] SUNNY, SONIA; DAVID PETER, S.; POULOSE JACOB, K. "RECOGNITION OF SPEECH SIGNALS: AN EXPERIMENTAL COMPARISON OF LINEAR PREDICTIVE CODING AND DISCRETE WAVELET TRANSFORMS"International Journal of Engineering Science & Technology;Apr2012, Vol. 4 Issue 4, p1594

[7] SUNNY, SONIA; DAVID PETER, S.; POULOSE JACOB, K Discrete Wavelet Transforms and Artificial Neural Networksfor Recognition of Isolated Spoken Words International Journal of Computer Applications (0975 – 8887) Volume 38– No.9, January 2012

[8] Rabiner, L. Juang, B. H., Yegnanarayana, B., "Fundamentals of Speech Recognition", Pearson Publishers, 2010.

[9] Felinek, "Statistical Methods for Speech recognition" MIT Press, cambridge assachusetts, USA, 1997

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