Analysis of Automatic Speech Recognition of Dental Plosive and Alveolar plosive Phonemes of Malayalam Language

Dr. Cini Kurian , Associate Professor, Al-Ameen College, Edathala, Aluva

Abstract : Speaking directly with the machine to achieve desired objectives, make usage of modern devices easier and convenient. Although may interactive software applications are available, the use these applications are limited due to language barriers. Development of speech recognition systems in local languages will help anyone to make use of the technological advancement of the speech recognition. In India, speech recognition systems have been developed for many indigenous languages, however very less work has been done in Malayalam Language. The objectives of this paper is to explore the speech recognition performance of dental plosive and retroflex plosive phonemes of Malayalam Language.

Keywords —Automatic speech Recognition, Malayalam Dental plosive , retroflex plosive

I. Introduction

Speech recognition technology is highly dependent on the spoken language. Phonetic and acoustic features of each phoneme in a language is the prime factor for speech recognition technology. In this paper we concentrate on the phonemes dental plosive and retroflex plosive[1] For speech recognition , we have used the famous statistical classifier the Hidden markov model [2] and the speech database has 32 selected minimal pairs spoken by 25 speakers. An analysis has been carried out to find out the resemblance of spectrogram , waveform , and format frequencies of the confusing pairs of the target word pairs. This is to find out whether the resemblance of these will lead to misclassification in speech recognition and whether there is any relation between phonetic study and their speech recognition performance

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. Phonetic realization[3] of all the phonemes will not be the same across languages. So in this work , we have tried to explore the phonemes of Malayalam language which has unique phonetic realization or has unique properties along with its speech recognition performance Speech is the most complex signal to deal with since several transformations occurring at semantic[4], linguistic, acoustic and articulator levels. In addition to this, the following factors make this area the most challenging.

- Context variability -: Some words having different meaning and usage may have same phonetic realizations. Few examples are: write vs. right, four vs. for. Since phonetic realization[5] is one of the key factors for speech recognizers, these types of words are of enormous challenge for speech recognizers
- Co-articulation Effect[6] -: This is caused by different acoustic realization for the same phoneme. e.g. Jeevs. peak. Here same phonemes /ee/ has different acoustic realization. This type of variability is difficult to model. Speaker Variability - This variability is due to variations in vocal tract size, vocal cord vibration, physical characteristic like age, sex etc. Speaking rate variation (words/minute) and emotional rate changes are other two factors which affect the speech quality/modeling.
- Environmental variability[6] is one of the most rigorous challenges since performance of speech recognition degrades due to mismatched conditions.
- Speech has undergone to many transformations, from the time it is delivered from mouth till it is converted to digital form. This factor affect the speech quality in a prominent way. Adverse conditions[7] ; and additive noise like fan, AC, another speakers voice and channel distortions are other factors which have to be taken into account while building ASR system.
- Automatic speech recognition technology needs knowledge from multidisciplinary areas like Acoustics, Linguistics, Biology, Physiology, Cognitive Science, Intelligence, Artificial Intelligence[8] Electrical Engineering. Computer science, Digital signal processing Mathematics and Statistics.

II. METHODOLOGIES USED

Phonetic Chart

Malayalam has 52 consonant phonemes, encompassing 7 places of articulation and 6 manners of articulation. 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[9]. 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. Review Stage

Alvelor Plosive റ്റ

Alveolar plosive is a characteristic of Malayalam and Assamees. Phonetic realization of alvelor plosive of English is different from that of Malayalam. Alevelor plosive have a good amount of friction at the time of release which sets them apart from their retroflex counter parts. It is the sole member of alvelor plosive, it is rather unusual because it is both aspirated and palatalized with no other counterpart in alvelor plosive category.

III. SPEECH RECOGNITION ANALYSIS

For conducting speech recognition performance of unique phonemes, we have conducted a special procedure. Initially, the words are recorded with carrier words like " njaan /the word / enn'u paranji' (i spoke the word /word/). This is to nullify the domination of language model in the speech recognition performance. Then the minimal pair counterpart also recorded with the same carrier words. Then we analyses the speech recognition performance . Here we can make a clear decision of whether these two words have been misclassified or not , thereby can make conclusion about the behavior of the phonemes under study. Three state HMM with phonemes in context dependent tied states with 8 gaussian mixture model and trigram language model in used for recognition.

Design of database : **Alveolar plosive** confuses with Retroflex plosives and dental plosive. We have selected 12 minimal pairs for the study (4 pairs with Retroflex plosive and 4 pairs with Dental plosive and 4 set of triplet with both of these). The following are the words that we have designed.

We have selected 4 minimal pairs for the study. The following are the words that we have designed.

a) **With Dental plosive**

- കുറ്റി , കുത്തി (/kut't'i/ - stump , / kuththi / - stabbed)
- പ റ്റി , പത്തി (/ pat't'i / - relating to , / paththi/ - hood of snake)
- കാറ്റ് , കാത്ത് (/ kaat't'u'/ - wind , /kaaththu'/ - to wait)
- (മറ്റ് , മത്ത്) (/ mat't'u'/ - other , / maththu‘ / - intoxication)

b) **With retroflex plosive and Dental plosive**

- കുറ്റി , കുട്ടി , കുത്തി (/kutt't'i/ - stump , / kutti/ - child , /kuththi/ - stabbed)
- ആറ്റ , ആട്ട , ആത്ത - (/aat’t’a/ - sparrow , /aatta/ - wheat flour, /aaththa /- custad apple)
- മറ്റ് , മത്ത് , മട്ട് (/ mat't'u'/ other , / maththu‘/ - intoxication , mattu’ measure

Alveolar Plosive vs Retroflex plosive vs Dental plosive:

The target tokens (patta, katta , paat't'a and kat't'a) were ready by 25 speakers, so that there are a total 100 tokens in the database. For training , we have selected data of 20 speakers . i.e 80 tokens and the remaining 20 tokens were kept for testing. The system has been tested with 20 tokens. Confusion matrix have been drawn for the target phonemes as in table minimal pairs evaluated for those tokens and a confusion matrix have been plotted .As discussed above a comparison with dental plosive token were also made. Here we have used a triplet (aatta, aat't'a and aaththa) for analyzing the performance. These tokens belong to Alveolar Plosive, Retroflex plosive and Dental Plosive categories . The training database has 75 token and the system is tested with 15 unknown tokens and the results are depicted below in table 1.10 % of alvelor plosive confuses with Dental Plosive and 10% with retroflex plosive. Retroflex plosive also confuses 10% with dental plosive .

Table 1.1 - Confusion matrix of speech recognition performance- tta vs t't'a vs ththa

	tta	t’t’a	ththa	total
tta	4	0	1	5
t’t’a	1	3	1	5
ththa	0	0	5	5
Total	5	3	7	15

IV. CONCLUSION

From the result of the confusion matrix , it can be seen that these three phomes Alveolar Plosive, Retroflex plosive and Dental Plosive) confuses in speech recognition performance . The alvelor plosive has pronunciation resemblance with both retroflex plosive and dental plosive. This analysis agree with speech recognition performance .i.e alveolar plosive confuses with retroflex plosive and dental plosive. Hence alvelor polosive which is a unique phoneme of the language, has acoustic and auditory properties similar to retroflex plosive and dental plosive

DOI : 10.35291/2454-9150.2019.0093

which make them confusing in speech recognition performance.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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.

[5] Itakura F (1975) Minimum prediction residual principle applied to speech recognition. IEEE Trans Acoustics Speech Signal Process ASSP 23:52–72

[6] [6] Miyatake, M. Sawai, H., & Shikano, K. (1990). Integrated Training for Spotting Japanese Phonemes Using Large Phonemic Time-Delay Neural Networks. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, 1990.

[7] Kimura, S. (1990). 100,000-Word Recognition Using Acoustic-Segment Networks. In Proc.IEEE International Conference on Acoustics, Speech, and Signal Processing.

[8] K.-F. Lee, Large-vocabulary speaker-independent continuous speech recognition: The Sphinx system, Ph.D. Thesis, Carnegie Mellon University, 1988.

[9] M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.

DOI : 10.35291/2454-9150.2019.0093