International Journal of Information Processing, 7(1), 73-80, 2013
ISSN : 0973-8215
IK International Publishing House Pvt. Ltd., New Delhi, India

Analysis of Isolated Word Recognition for Kannada Language using Pattern Recognition Approach

Sarika Hegde[a], K K Achary[b], Surendra Shetty[c]

[a]Department of Computer Apllications, NMAMIT, Nitte, Udupi District, Karnataka, India-574110, Contact: sarika.hegde@yahoo.in

[b]Department of Statistics Mangalore University, Mangalagangothri, Mangalore Karnataka, India-574199, Contact: kka@mangaloreunivarsity.ac.in

[c]Department of Computer Apllications, NMAMIT, Nitte, Udupi District, Karnataka, India-574110, Contact: hsshetty4u@yahoo.com.

The speech recognition can be done with two approaches. In the first approach called as Isolated Word Recognition (IWR), the problem is to identify each word as an individual unit. The second approach is Continuous Speech Recognition (CSR), where speech must be broken into smaller units for identification. We have developed an Isolated Word Recognition (IWR) technique for identification of spoken words for the database created by recording the words in Kannada Language. Support Vector Machine (SVM) algorithm is used for designing the classifier model. We have also analyzed the variation in the performance of classification for words ending with same phonetics. We found that the classification accuracy using SVM with Mel-Frequency Cesptrum Coefficients (MFCC) is good and accuracy has an affect due to the similar phonetic sounds in different words.

Keywords : Isolated Word Recognition, Kannada Language, Speech Recognition, Support Vector Machine.

1. INTRODUCTION

The Automatic Speech Recognition problems have been solved using pattern recognition technique in most of the cases. In pattern recognition approach for speech recognition, speech patterns are used directly without explicit feature determination (like in acoustic) and segmentation. The method has mainly two steps: training of speech pattern and recognition of speech pattern via pattern comparison. The concept is that the enough versions of a pattern to be recognized are included in the training pattern. Machine learns the acoustic properties of speech class that are reliable and repeatable across all training tokens of the pattern. The process of classification of a given speech signal into a particular class has four steps: feature measurement, pattern training, pattern classification and decision logic [1].

2. PREVIOUS WORKS

The ASR problem can be researched in many ways like, how the speech data can be enhanced for improving the quality of speech, reducing the level of noise in the speech signals, detecting the presence/absence of voice(VAD) or how signal can be segmented into different units; (the unit may be phone, syllable or word *etc.*,). Some of the major works relating to noise and speech-end point detection can be seen in [2],[3],[4],[5]. Many authors have considered phone / triphone model as the basic unit of speech and segmentation of speech and recognition is done based on that [6]. Isolated Word Recognition (IWR) is an area of ASR where the input speech is an isolated word unit in contrary to a Continuous Speech Recognition where continuous speech sentence is uttered. There are also significant research done on segmenting speech into syllabic like models

signal assign target label from 1 to 10 that corresponds to the ten words.

4. EXPERIMENTAL ANALYSIS

The frame size for computing the log-energy values is 10ms and for computing the MFCC feature is 30ms. Each of the file is converted into sequence of nearly 50 number of feature vectors totally constituting around 5000 number of records. We have conducted two experiments for the speech recognition. In the first experiment, the entire database as it is considered for the evaluation. In second kind of experiment, we leave a trailing end of each file *i.e.*, around 15 frames and then the evaluation is applied. In Hold-out method the database is divided into two parts in a particular ratio and then one of the parts is used as training set and another as testing set. We have used a ratio 1:3 for training and testing set respectively. We have already explained the technique of training and testing in the previous section.

For the first type of experiment conduced, the Figure 4 shows the accuracy for the result of testing the 100 files with the trained SVM model. The first column gives the word uttered with the meaning given in both *English* and *Kannada*. The second column gives the total number of files for each word. We have done ten iterations and the average class recognition accuracy is 78.4%. In each of the iteration, database is divided into training and testing set randomly.

We observed that each of the word in the database that we considered is ending with same phonetic sound, /u. For analyzing the effect on classification accuracy of words due to this, we conducted another experiment. In this experiment we used the same feature extraction techniques as in first experiment, but the trailing 15 frames of each file are discarded. By doing this, we created a new database which has no feature vectors corresponding to the last part of the file. The experiment is repeated for *ten* times and average classification accuracy computed is 97.29%. The Figure 5 shows the results of classifying for one of the iteration with 99%. We can observe that the classification accuracy is increased and so the same ending phonetic sound has an effect in the classification accuracy for Isolated Word Recognition.

5. CONCLUSIONS

The speech recognition problem has application in many diverse areas. Due to the complex nature of the signals the ASR problem has not been solved to its full extent. Especially Indian languages need more focus so that it can be easily accommodated into the existing ASR system. The database creation for Indian languages spoken words is challenging job. We have demonstrated that the technique SVM combined with MFCC feature extraction, systematic method of noise reduction and speech end point detection yields generating classifier model that could classify the considered words with a good accuracy rate. We can test the same model with a large database and analyze the results. Due to similarity in some of the phonetics sounds in words, the classification accuracy may be affected. This problem can be analyzed further by dividing the word again into sub-word units. We can also consider the continuous speech database in the next step, in which case we should for efficient segmentation techniques.

REFERENCES

1. Rabiner L and Juang B H. Fundamentals of Speech Recognition, *Prentice Hall PTR, ISBN:0-13-015157-2*. NY, USA, 1993.
2. Jongseo Sohn, Nam Soo Kim and Wonyong Sung. A Statistical Model-based Voice Activity Detection, *IEEE Signal Processing Letters*, 6(1):1–3, Jan 1999.
3. Lori F Lamel, Lawrence R Rabiner, Aaron E Rosenberg and Jay G Wilpon. An Improved End Point Detector for Isolated Speech Recognition, *IEEE Transactions On Acoustics, Speech and Signal Processing*, Assp-29(4): 777–785, August 1981.
4. Main G Rahim and Biing-Hwang Juang. Signal Bias Removal by Maximum Likelihood Estimation for Robust Telephone Speech Recognition, *IEEE Transactions on Speech and Audio Processing*, 4, 1996.

WORD	Number of Files	Correctly Recognized
One(*"Ondu"*)	10	8
Two(*"Eradu"*)	10	7
Three(*"Mooru"*)	10	9
Four(*"Naalku"*)	10	10
Five(*"Aidu"*)	10	9
Six(*"Aaru"*)	10	8
Seven(*"Elu"*)	10	7
Eight(*"Entu"*)	10	9
Nine(*"Ombattu"*)	10	5
Ten(*"Hattu"*)	10	8
	100	80
Classification Accuracy --- 80%		

Figure 4. The Results of Classifying *ten Kannada* Words

WORD	Number of Files	Correctly Recognized
One(*"Ondu"*)	10	9
Two(*"Eradu"*)	10	10
Three(*"Mooru"*)	10	10
Four(*"Naalku"*)	10	10
Five(*"Aidu"*)	10	10
Six(*"Aaru"*)	10	10
Seven(*"Elu"*)	10	10
Eight(*"Entu"*)	10	10
Nine(*"Ombattu"*)	10	10
Ten(*"Hattu"*)	10	10
Total Files	**100**	**99**
Classification Accuracy is --- 99%		

Figure 5. The Results of Classifying Ten Kannada Words after Discarding the Trailing 15 Frames from Each File

5. Yariv Ephraim. Statistical Model Based Speech Enhancement Systems, *Proceedings of the IEEE*, 80(10):1526–1555, ISSN 0018-9219, October 1992.
6. Sarah Hawkins. Contribution of Fine Phonetic Detail to Speech Understanding, *Proceedings of the 15th International Congress of Phonetic Sciences* pages 293–296, 2003.
7. James R Glass. A Probabilistic Framework for Segment Based Speech Recognition, Computer, Speech and Language, 17(3), ISSN:0885-2308, pages 137–152, 2003.
8. Lakshmi A, Hema A Murthy. Syllable Based Continuous Speech Recognizer for Tamil, *Proceedings of International conference on Spoken Language , INTERSPEECH ICSLP*, pages 1878–1881, September 17-21, Pittsburgh, Pennsylvania, 2006.
9. R Thangarajan and A M Natarajan. Syllable Based Continuous Speech Recognition for Tamil, *South Asian Language Review*, XVIII(1):72–85, January 2008.
10. Steven B Davis and Paul Mermelstien. Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuous Speech Recognition, *IEEE Transactions on Acoustics, Speech and Signal Processing*, Assp-28(4):357–365, August 1980.
11. Jen-Tzung Chien a, Koichi Shinoda B and Sadaoki Furui. Predictive Minimum Bayes Risk Classification for Robust Speech Recognition, *INTERSPEECH*, pages 1062–1065, 2007.
12. Li X, Jiang H C-J. Large Margin HMM for Speech Recognition, *IEEE Trans. On Audio, Speech and Language Processing*, 14(5):1584–1595, 2006.
13. Scott Axelrod and Beno It Maison. Combination of Hmm with Dtw for Speech Recognition, *Proceedings of Internation Conference On Acoustics, Speech And Signal Processing (ICASSP 2004)*, Pages 173–176, 2004.
14. Xin He, Xian-Zhong Zhou. Audio Classification by Hybrid Support Vector Machine/Hidden Markov Model, *World Journal of Modeling and Simulation*, 1(1):56–59, 2005.
15. Frederick Jelinek. Continuous Speech Recognition by Statistical Methods, *Proceedings of the IEEE*, 64(4), pages 532-556, April 1976.