International Journal of Modern Trends in Engineering and Research
www.ijmter.com
e-ISSN No.:2349-9745, Date: 2-4 July, 2015

Isolated Digit Recogntion using Mathematical Tool based on Hidden Markov Model

Ganesh S Pawar[1],Sunil S Morade[2]
[1]E &TC Engineering, SNJB's KBJ COE College, Chandwad,ganesh7pawar@gmail.com
[2]E &TC Engineering, KKW IEER, Nashik, ssm.eltx@gmail.com

Abstract-The work presented here describes the use of Mathematical tool like HTK (Hidden Markov Model Toolkit) for recognition of English isolated digits. Two different databases are used here which comprise of audio recordings for the English language isolated digits. Digits used are ZERO to NINE. First database is a standard CUAVE database (36 speakers, each spoken 10 words) and other database is self-recorded sample files of students of Engineering. The system has been implemented for speaker dependent and independent way, corresponding comparison has been made. For every individual digit, separate HMM has been created. It has been trained many times and tested against every other digit to check the every possibility of recognition/ duplication. Further this system can be used for continuous digit recognition also and one can go for sentence recognition using database like TIMIT.

Keywords- Confusion Matrix, CUAVE, HMM, HTK, TIMIT.

I. INTRODUCTION

The techniques of speech recognition involves speech analysis, taking important properties from the signal i.e. Feature extraction. Then signal representation in the required format which can be used for the recognition or testing. It also makes use of some pattern recognition techniques or classifier. This idea or technology involves various processes and systems to be used to reach up to the final level. From the previous 5-6 decades, the attempts are going on to design an accurate speech recognizer. Some got success, and some shown a way in the direction of further research. The last decade has witnessed dramatic improvement in speech recognition technology, to the extent that high performance algorithms and systems are becoming available. The first attempts (during the 1950s) to develop techniques in ASR, which were based on the direct conversion of speech signal into a sequence of phoneme-like units, failed [1]. In this system, a single speaker was involved. It was purely speaker dependent system. The first positive results of spoken word recognition came into existence in the 1970s, when general pattern matching techniques were introduced. R Kumar [2] implemented an experimental, speaker dependent, real time, isolated word recognizer for the regional language like Punjabi andalso extended that to compare the performance of system for small vocabulary isolated words using the Hidden Markov Model (HMM) and Dynamic Time Warping (DTW) technique. Further some have developed a connected words speech recognition system for regional language like Hindi. The system was trained to recognize any sequence of words selected from vocabulary of 102 words and used HTK [3]. It got good success rate to recognize the words. There are various methods of feature extraction, among which we are using MFCC (Mel Frequency Cepstrum Coefficients) technique due to its advantages over other methods and familiarity in the Hidden Markov Model Tool Kit i.e. HTK [4].

There are different approaches for recognition or classification based on different methods like Template based, Statistics based, Learning based, Knowledge based and Artificial Intelligence based. The HMM is popular Mathematical especially statistical tool for modeling a wide range of time series data [5]. The reason for the evolution of speech recognition technology, hence improved is that it has a lot of applications in many aspects of our daily life, for example, telephone applications, applications for the physically handicapped and illiterates and many others in the area of computer science. Researchers are contributing to solutions of these problems of the society [6]. The aim of this work is therefore to design and train a speech recognition system which will recognize the speech signals of isolated digits of English language using HTK with MFCC as the feature of extraction with two different databases for speaker dependentand independent approach.

II. METHODOLOGY USING HTK

HTK is one of the most widely used tools for speech recognition research and teaching-learning. The HTK is a portable toolkit for building and manipulating hidden Markov models. The HMM Toolkit was originated in Machine Intelligence Laboratory in the Cambridge University Engineering Department. A currently available stable release is version HTK 3.4.1. This tool can be used with either Windows or Linux platform. Both versions are available as free on the official website for htk i.e. http://htk.eng.cam.ac.uk. The methodology used is nothing but the use of modeling technique like HMM with special tool like HTK. It is primarily used for speech recognition research although it has been used for numerous other applications including research into speech synthesis, character recognition and DNA sequencing. HMM is very powerful mathematical tool for modeling time series. HMM provides efficient algorithms for state and parameter estimation, and it automatically performs dynamic time warping of signals that are locally stretched. A natural extension of Markov chain is HMM, the extension where the internal states are hidden and any state produces observable symbols or observable evidences [7].

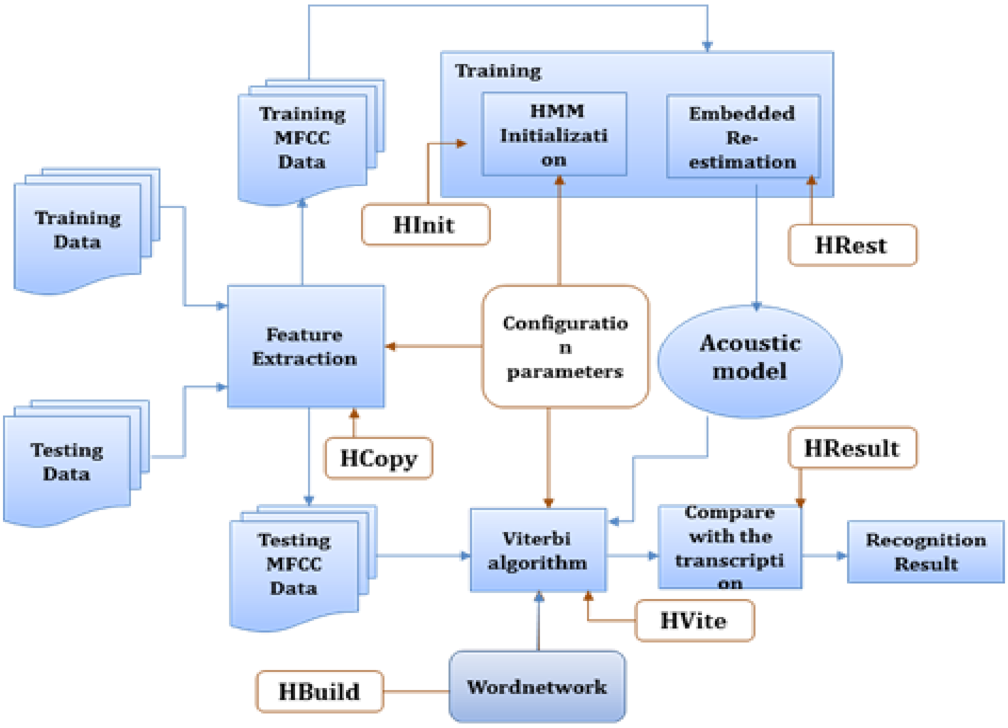

Figure 1. System realization using HTK

The complete diagram of the method used has been shown in the figure 1 above. It shows the different tools used at every different stage of training with necessary input and output routes to various other blocks of the system. Initially data preparation tools of HTK (like HParse, HCopy) are used to prepare the language model, dictionary and word network. Then parameter estimation tools of HTK (HInit) are used to define the HMMs of every digit and then to initialize the model. Further training tools like HRest are used for re-estimation to have proper and robust training. Finally recognition tool like HVite are used to get recognition results and confusion matrix for the given dataset. The performance of the speech recognition system can be checked by using HTK tool HResults.

III. RESULTS

The comparison of the results for CUAVE and self-recorded dataset has been given along with confusion matrix generated. The database used here is a CUAVE database. CUAVE (Clemson University Audio Visual Experiments) was recorded by E.K. Patterson of Department of Electrical and Computer Engineering, Clemson University, US [15]. The database was recorded in an isolated sound booth. This database is a speaker-independent database consisting of connected and continuous digits spoken in different situations. It contains mixture of speaker with white and black skin. The speakers consist of Males, Females from different age group. The system has been tested for speaker dependent system and yields maximum accuracy for CUAVE dataset and moderate accuracy for the own dataset. Also if we go for speaker independent system, where speakers involved in the training were different than those involved during testing. Here the recognition drops to some extent. Still we have up to 86% recognition for CUAVE dataset. The system is relatively successful, as it can identify the spoken digit at an accuracy of 96% for speaker dependent approach, which is relatively high. Further it is observed that for speaker independent approach, accuracy drops. It is due to the variations in the uttered speech by speakers, who were not involved in training, nature of utterance, environmental conditions, difference between speakers due to age, sex, accent etc.

Table 1. Comparison of Recognition rate for different databases

Database →	CUAVE		Self Recorded Database	
Recognition %	Training Set	Testing Set	Training Set	Testing Set
Sp. Independent	96	93	80	25
Sp. Dependent	91	86	71	22

3.1. Confusion Matrix

A confusion matrix has also been generated which tells us about the recognition rate for every individual digit checked against all other digit and to itself. It was observed that digit 4 and digit 6 gives up to 100% recognition. Digit 1 gets confused with digit 9 with confusion rate 12.50%. Digit 0 gets confused with digit 2, digit7 gets confused with 9. Also digit 9 gets confused with digit 1, 5.

IV. CONCLUSION

The experiments/test carried out showed that a higher level of accuracy can be achieved if the language model was designed for limited dictionary and trained the word model with a large set of speech data from the user. The system was tested using testing corpus data and the system scored up to 96% word recognition for speaker dependent approach and up to 86% for speaker independent approach. For the speaker independent approach; environmental conditions, speakers invariability, way of utterance should be considered and then accordingly system can be made more robust by more training the HMMs to give proper recognition. The work is however not all conclusive as ithas catered for only an Isolated Digit Speech data. For the work presented, digit-pronunciation was limited to the English language only. The model could be further developed to incorporate digits from other languages; most preferably the local languages of the clients. Furthermore, the dictionary size could be increased using alphabets, so the large test data could be generated and trained. For sentence recognition, TIMIT like database can be used. The system can be enhanced to a larger vocabulary including alphabets and commonly used words.

The authors would like to thank the experts who have contributed anything towards the development of the work and its smooth implementation as mentioned in the above sections. Not to forget the Staff members of SNJB's KBJ COE, Chandwad who have provided necessary environment and infrastructure for our work.

References

[1] R.Klevansand, R. Rodman, "Voice Recognition", Artech House, Boston, London 1997.

[2] R. Kumar, "Comparison of HMM and DTW for Isolated Word Recognition of Punjabi Language", In Proceedings of Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications, Sao Paulo, Brazil. Vol. 6419 of LNCS, pp. 244-252, Springer Verlag, November 8-11, 2010.

[3] K. Kumar, R. K. Aggrawal, A Jain, "A Hindi speech recognition system for connected words using HTK", International Journal of Computational Systems Engineering, Vol. 1, No. 1, 2012.

[4] Santosh K Gaikwad, Bharti W Gawali, Pravin Yannawar, "A Review on Speech Recognition Techniques", International Journal of Computer Applications (0975-8887), Vol. 10, No. 3, November 2010.

[5] Ibrahim patel, Dr. Y shrinivas rao, "Speech recognition using HMM with MFCC – an analysis using frequency spectral decomposition technique", Signal and Image Processing: An International Journal (SIPIJ), Vol. 1, No. 2, December 2010.

[6] Rabiner L.R., S.E. Levinson, "Isolated and connected word recognition - Theory and selected applications", IEEE Trans. COM-29, pp.621-629, 1981.

[7] Young S., G. Evermann, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V.Valtchev, P. Woodland, "The HTK Book", 2002 (Retrieved Jan 2, 2013) from: http://htk.eng.cam.ac.uk.

[8] Roux, J.C., Botha, E.C., Du Preez, J.A., "Developing a Multilingual Telephone Based Information System in African Languages", Proceedings of the 2nd International Language Resources and Evaluation Conference, Athens, Greece : ELRA (2),975-980, 2000.

[9] Juang B, Rabiner L, "Hidden Markov Models for speech recognition", Technometrics, 33 (1991), 251-272.

[10] Wei Han, Cheong-Fat Chan, Chiu-Sing Choy and Kong-Pang Pun – "An Efficient MFCC Extraction Method in Speech Recognition", Department of Electronics Engineering, The Chinese University of Hong Kong, Hong, IEEE – ISCAS, 2006.

[11] Dipmoy Gupta, RadhaMounima C., Navya Manjunath, Manoj P.B., "Isolated word speech recognition using VQ", International Journal of Advanced Research in Computer science and Software Engineering, Vol. 2, Issue 5, ISSN: 2277 128X, May 2012.

[12] Kritika Nimje, Madhu Shandilya, "Automatic isolated digit recognition system: an approach using HMM", Journal of Scientific and Industrial Research, Vol.70, pp. 270-272, April 2011.

[13] Mohit Dua, R. K. aggarwal, Virender Kadyan, Shelza Dua, "Punjabi Automatic Speech Recognition using HTK", IJCSI, Vol. 9, Issue 4, No. 1, July 2012.

[14] www.myfit.edu/~vkepuska/HTK/HTK-basic-tutorial.pdf

[15] E. K. Patterson, S. Gurbuz, Z. tufekci, and J. N. Gowdy, "CUAVE: A New Audio-visual Database for Multimodal Human Computer Interface Research", Clemson University, USA.