International Journal of Computer Applications (0975 - 8887)
Volume 121 - No. 11, July 2015

A Survey Report on Speech Recognition System

Moirangthem Tiken Singh
Institute of Engineering and Technology
Dibrugarh University
Assam

Abdur Razzaq Fayjie
Institute of Engineering and Technology
Dibrugarh University
Assam

Biswajeet Kachari
Institute of Engineering and Technology
Dibrugarh University
Assam

ABSTRACT

Speech Recognition is the process of converting an acoustic waveform into text containing the similar information conveyed by speaker. This paper present a report on a Automatic Speech Recognition System (ASR) for different language under different accent. The paper describe the methods used and comparative study of the performance of every system so far developed. The study shows that Hidden Markov Model(HMM) as classifier and Mel Frequency Cepstral Coefficients(MFCC) as speech features are the most common technique used. And Moreover ASR implemented by using Hidden Markov Tool kit(HTK) are more efficient then the other systems implemented by using other tools

General Terms:

Automatic Speech Recognition, Different Language

Keywords

Hidden Markov Model (HMM), MFCC, Different Language Accent, Hidden Markov Tool kit(HTK)

1. INTRODUCTION

Automatic Speech Recognition is a field of Computer Science, aims to design computer system that can recognize human voice. After TITS, ASR came into existence with an IV system support. It takes an utterance of speech signal as input, captured by a microphone, a telephone etcând convert it into a text sequence as close as possible to spoken data[1].ASR was first introduced during 1950?s. The first attempt (during the 1950?s) to develop techniques for speech recognition, which were based on the direct conversion of speech signal into a sequence of phoneme-like units, failed. The first positive results of spoken word recognition came into existence in the 1970?s, when general pattern matching techniques were introduced[2]. ASR has attracted much attention over the last three decades and has witnessed dramatic improvement in the last decade. Today it has different areas of application like dictation, program controlling, automatic telephone call, weather report information system, travel information systems etcḂut its implementation is difficult due to the different speaking styles of human beings (i.e. the accents). Therefore the main aim of ASR today is to transform an input voice signal to its corresponding text output independent of speaker or device. This paper aims to present a review on methodology and results obtained during speech recognition by various researchers. A comparison is done based upon their recognition level and accuracy.

2. LITERATURE REVIEW

A. N. Mishra[3] and his team worked on automatic speech recognition on speaker independent connected digits with Revised perceptual linear prediction, Bark frequency cepstral coefficients and Mel-frequency cepstral coefficients with a clean dataset. Hidden Markov Model is implemented using HTK. A.N.Mishra again with Astik Biswas and Mahesh Chandra[4] designed a system for isolated digit recognition in Hindi. HMM is chosen as the classifier and MFCC algorithm for features extraction. They performed experiments using both HTK and Matlab, with both clean and noisy data. Ganesh S. Pawar and Sunil S. Morade[5] designed a digit recognition system for isolated English digits with a huge database of 50 speakers using HMM and MFCC algorithm. HTK is used for training and testing purposes. Maruti Limkar[6] works on a system for speech recognition with a proposed approach to speech recognition for isolated English digit using MFCC and DTW(Dynamic time wrapping) algorithm. Elitza Ivanova et. al. [7] worked on American and Chinese spoken English using HMM and HTK. Babita Saxsena and Charu Wahi[8] worked on Hindi digits recognition. They collected their data in natural noise environments. Mohit Dua et al. also worked on digit recognition with Punjabi language[1].

3. METHODOLOGY USED

A.N. Mishra[3] et al. used HTK to implement HMM for training and testing purposes for connected Hindi digits. Database was prepared by 40 speakers, 23 female speakers and 17 male speakers using cool edit software. For features extraction, MFCC, Δ MFCC along with RPLP, PLP and BFCC is used, where MFCC is done through HTK and all other features are extracted through Matlab and saved in HTK format. Analysis is done in both clean and noisy data. A.N. Mishra[4] et al. again performed some experiments using HTK for isolated Hindi digits. Using 35 speakers for training and 5 speakers for testing, 3500 features were extracted. Out of which 350 features were chosen, 12 MFCC coefficients were obtained for each frame from where only 13 MFCC coefficients were chosen for vector quantization. 10 HMM's were created for each digit. Ganesh S. Pawar and Sunil S. Morade[2] used HTK implementing HMM as classifier. They prepared the database with 50 speakers i.e. of 500 samples. 400 samples were used in training and 100 were used in testing. CUAVE(Clemson University Audio Visual Experiments) database was used for speaker independent environment. Maruti Limkar[6] et al. used MFCC algorithm

to extract features, implemented features vector matching for training purposes. Dynamic Time Warping is used as classifier rejecting the HMM. 100 samples were analysed and results were obtained accordingly. Elitza Ivanova[7] et al. worked on a database of MPEG-1 Audio Layer 3 (.mp3) samples of spoken English. 375 word passage were chosen for recording data. All .mp3 files were converted to .wav format. HMM is used as classifier for the study. Babita Saxena[8] used MFCC for features extraction of data collected from 10 speakers- 8 for training and 2 for testing. Database was prepared in noisy environment. HMM is used as the acoustic model here, with more than 61 context independent phonemes. Converting models to tri-phone models, Baum-Wells method is applied to obtain results. Mohit Dua[1] and his team worked on automatic speech recognition on Punjabi language using class room and open space environment. 115 distinct words were used and trained the system using HTK. MFCC and HMM were used in features extraction and acoustic model.

4. RESULTS

A. N. Mishra[3] discussed the efficiency of different features extraction algorithm mainly PLP, RPLP, BFCC and MF-PLP for both clean and noisy data using connected Hindi digit recognition system. The efficiency of algorithms for clean data, obtained during the working is shown in Fig.1. The performance is based on percentage of recognition.

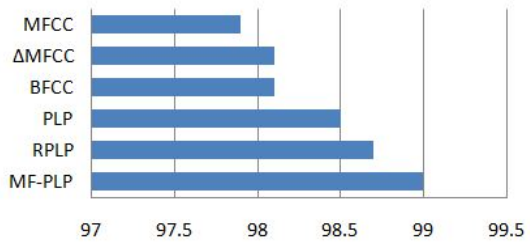

Fig. 1. Algorithm performance on clean data

On noisy environment, Mishra tested the system with Babble noise, White noise, Pink noise and F16 noise. He performed each test for three times with the SNR values 5 dB, 10 dB and 20 dB. The system is evaluated with constant characteristics, which are 5-HMM and 9-Gaussian Mixtures. Below 5 dB SNR ratio, the system did not worked properly and gives a result equal to clean data result. For noisy data, with all the respective noises, the system performed best with MF-PLP with 98% recognition rate. MFCC produce an output with 96-97% accuracy rate.
MF-PLP has shown best recognition performance compared to other features extraction techniques. It is due to the fact that it incorporates Mel-filter into a perceptual linear features extraction method. PLP result in recognition was better than MFCC because the signal was pre-emphasized by a simulated equal-loudness curve to match the frequency magnitude. RPLP features have also shown good results for clean as well as noisy data. This is due to the fact that it takes advantage of pre-emphasis filter, Mel scale filter bank along with linear prediction and cepstral analysis.[three-9]
A.N. Mishra[4] again performed experiments on a system with isolated Hindi digit in clean data environment. HMM is implemented using HTK and MFCC is used as features extraction algorithm. He used the same classifier and algorithm in Matlab and performs the same experiments. The results obtained was compared. Tab. 1 gives the comparison for the results for both the tools.

Table 1. Recognition comparison

Speaker No.	% Recognition (MATLAB)	% Recognition (HTK)
1	94	100
2	88	97
3	90	99
4	92	100
5	91	100

All the above results are obtained strictly in a clean environment. Thus MATLAB gives an average recognition of 91% and HTK gives an average recognition of 99.2%.
Ganesh S. Pawar and Sunil S. Morade[2] obtained a result with 95% recognition rate. For self recorded database, they obtained recognition rate as 80% in average. Fig. 2 represents the result for Ganesh S. Pawar and Sunil S. Morade.

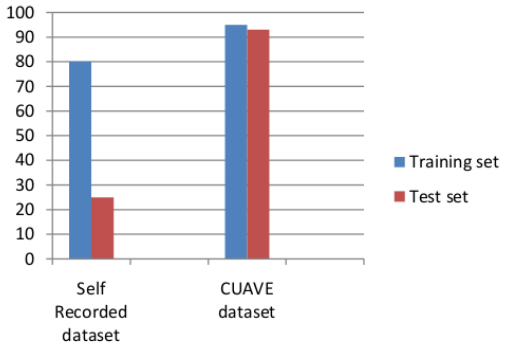

Fig. 2. Results obtained by Ganesh S. Pawar

Maruti Limkar[6] worked on automatic speech recognition by MFCC vectors to that provide an estimate of the vocal tract filter. Meanwhile DTW is used to detect the nearest recorded voice with appropriate constraint. Accuracy was emphasized rather than recognition. 95% accuracy was obtained using the method. Using 200 dataset for training and 70 for testing, Elitza Ivanova[7] obtained a result with accuracy rate 70-75%. The work is done using HTK and implementing HMM. Following Table gives the result for each digit with accuracy.

Table 2. Results obtained by Maruti Limkar

Word	Accuracy	Word	Accuracy
zero	80	five	100
one	95	six	80
two	80	seven	100
three	100	eight	100
four	90	nine	80

Babita Saxsena and Charu Wahi[8] worked on a digit speech recognition with 2 seen and 2 unseen speakers. Using HTK, implementing HMM and MFCC for features extraction, they obtained a result of word recognition equal to 85%. during the testing with unseen speaker. In speech recognition, recognition rate of 95.63% and 94.08% were obtained by Mohit Dua and his team in a classroom and open space environment respectively.

5. CONCLUSION

By comparison of all the work done by the respective researchers in speech recognition field, the following conclusion can be drawn.

Table 3. Comparison statistics

Mishra	HMM	HTK	C	99
Mishra	HMM	Matlab	C	91
Pawar	HMM	HTK	CUAVE	95
Pawar	HMM	HTK	R	80
Limkar	DTW	HTK	C	95
Elitza	HMM	HTK	M	75
Babita	HMM	HTK	Noisy	85
Mohit	HMM	HTK	CL	95
Mohit	HMM	HTK	OS	94

Here C, M, CL, OS is used for the word Clean, Mixed, Classroom, Open Space.

Most of people uses Hidden Markov Model as acoustic model. It is due to the fact that it provides better recognition and its efficiency is accepted universally. Maruti Limkar[6] is used Dynamic Time Warping which provides a n accuracy rate of 95%. But recognition rate must be emphasized compared to accuracy. As Ye-Yi Wang[9] et al. proves a good accuracy never indicates a good rate of recognition. In case of tools HTK is chosen over Matlab, due to its efficiency in implementing HMM, open source and better recognition rate. Most of tools are provided with HTK for easy speech recognition. Mel frequency cepstal coefficients are used to extract features where recognition level reduces by 1-2% compared to MF-PLP. But easy to understand and easy to use nature, researchers compromises the 1-2% and uses MFCC. Using HTK, MFCC algorithms can be implemented directly which generates the Mel-coefficients. In noisy environments the recognition level falls compared to clean database.

6. REFERENCES

[1] Mohit Dua et al., "Punjabi Automatic Speech Recognition Using HTK." in IJCSI International Journal of Computer Science Issues, IJCSI press, Mauritius, Vol. 1, Issue 4, No. 1, Jul. 2012.

[2] Ganesh S. Pawar, Sunil S. Morade, "Isolated English Language Digit Recognition Using Hidden Markov Model Tool kit," in International Journal of Advanced Research in Computer Science and Software Engineering, Jaunpur-222001, Uttar Pradesh, India, Vol. 4, Issue 6, June 2014.

[3] A. N. Mishra et al., "Robust Features for Connected Hindi Digits Recognition" in International Journal of Signal Processing, Image Processing and Pattern Recognition,Vol. 4, No. 2, June 2011.

[4] A.N. Mishra et al.,"Isolated Hindi Digits Recognition: A Comparative Study" in International Journal of Electronics and Communication Engineering, India, Vol. 3, No. 1, 2010, pp. 229-238.

[5] Ganesh S. Pawar, Sunil S. Morade, "Isolated English Language Digit Recognition Using Hidden Markov Model Tool kit." in International Journal of Advanced Research in Computer Science and Software Engineering, Jaunpur-222001, Uttar Pradesh, India, Vol. 4, Issue 6, June 2014.

[6] Maruti Limkar et al., "Isolated Digit Recognition Using MFCC AND DTW" in International Journal on Advanced Electrical and Electronics Engineering, Uttar Pradesh, India, Vol. 1, Issue 1, 2012.

[7] Elitza Ivanova et al., "Recognizing American and Chinese Spoken English Using Supervised Learning."

[8] Babita Saxena and Charu Wahi, "Hindi Digits Recognition System On Speech Data Collected in Natural Noise Environments." in David C. Wyld et al. (Eds) : CSITY, SIGPRO, DTMN - 2015.

[9] Ye-Yi Wang et al.,"Is word Error rate a good indicator for spoken language understanding accuracy" in IEEE Workshop on Automatic Speech Recognition and Understanding, St. Thomas, U.S. Virgin Islands, 2003.pp. 23?30, 2015.