A PROCESS OF DEVELOPING AN ASR SYSTEM FOR ENGLISH AND TAMIL LANGUAGES

S.Santhi[1], S.Subhashini[2], M.Nithiyanandhan[3]

[1,2,3,]Assistantprofessor,Department Of English,AEC,Salem

Abstract

The fundamental way of interacting with people was speech. It is used to express our message, thoughts and feelings to others. Languages like English and Tamil was the process to convert speech into acoustic message. This paper highlights the instinctive recognition of speech for English and Tamil language. This instinctive process was known an Automatic Speech Recognition method. An important issue of Automatic Speech Recognition was analyzed in this study. For various under resourced languages, this Automatic Speech Recognition method was a correct starting point for everyone who interested in this topic. This research must be clear and many approaches and issues are applied to speech technique to analyze this method.

Keywords: Speech Recognition, language pronunciation, ASR, Language Probability, Speech and language resources acquisition

Research Highlights

An aim of this research was to create an efficient ASR system for English and Tamil languages. A factor of Automatic Speech Recognition was analyzed. The guidelines and strategies for gathering speech script information and text information were also examined. The gathered sources of Tamil and English were applied to construct basic purpose of Automatic Speech Recognition method.

1. Introduction

A study in ASR method through machines was finished for almost 4 generations. The growth of speech recognition apps provides not valuable participation to the research field over the past decades. But it was becoming mature in current years. ASR method performs well under speechless environment. Its accuracy rate of word was up to 99%. Comparing to the mouse or keyboard in the computer, speech has the great potential to have a good interface. [1] Many research and growth are finished in current generations in speech identifying in many languages. Therefore, speech recognition of English language was still considered as infancy. Present systems are quite variations in style of speech, sensitive to channel and environment. A no. of techniques for developing the strong Automatic Speech Recognition method has only limited achievements in degraded places. Non-traditional uses visual and acoustic characteristic speech level and strong enough to be removable in this application. The most believing source of speech data was visual speech. It is not affected by the audible noise.

Automatic Speech Recognition was applied in various fields like commercial, education, military, medical and telephony. Speech technology presents its contribution to support the handicapped people to work their internet and computer. These systems was differentiated into many types depend upon the utterance kinds, size of vocabulary, environment, channel and speaker design. The development of ASR was becoming more difficult. It is becoming more challenging because of its difference. [2] The kinds of speech recognition were shown in the figure 1.

Speech recognition process was isolated form the dependence of medium or small vocabulary speaker. They are very easy and simple to execute. They can easily succeed more perfection but there was an absence in providing flexibility which correlated to the independent speaker or adaptive speaker systems. But it was just opposite to the large vocabulary speaker. It was easily recognized the regular pattern of speech from a big group of people and their vocabulary. This type of process was more complex to execute and succeeds less perfection compared to dependent speaker separated ASR system.[3] these process provides flexibility, but when vocabulary and speaker develops high, the chaos in various pattern of speech also develops larger. The difference in environment, age, speech rate, style of speaking and channel makes the Automatic Speech Recognition process as more difficulty.[2] A lot of study efforts are placed into the practice to construct a strong Automatic speech Recognition process which works ignoring of this difference. An important aim of this paper was to develop a speech recognition process for identifying Tamil works which are spoken; Automatic Speech Recognition for Tamil was at the early years. Most of the research is depend on NN and HMM among the some attempts on Tamil language. Therefore the contribution of this research was involved GMM and learning through machine like decision tree and SVM algorithms.

2. RelatedWork

B. Bharathi et al. examined a DNN approach to construct independent speaker separated recognition of word process for Tamil[4]. Some steps are executed to develop the process such as no. of zero crossing, denoting using the wavelet transformation and removal of silence using the energy values. And next MFCC vector characteristics were separated and normalized by CMB to decrease the particular difference in various human beings. So here the researcher used SOM network to create length difference in mel frequency cepstral coefficients into fixed MFCC length. And finally it feeds neural network which was used to identify the words which were already spoken. This process has succeeded 95% in accuracy.

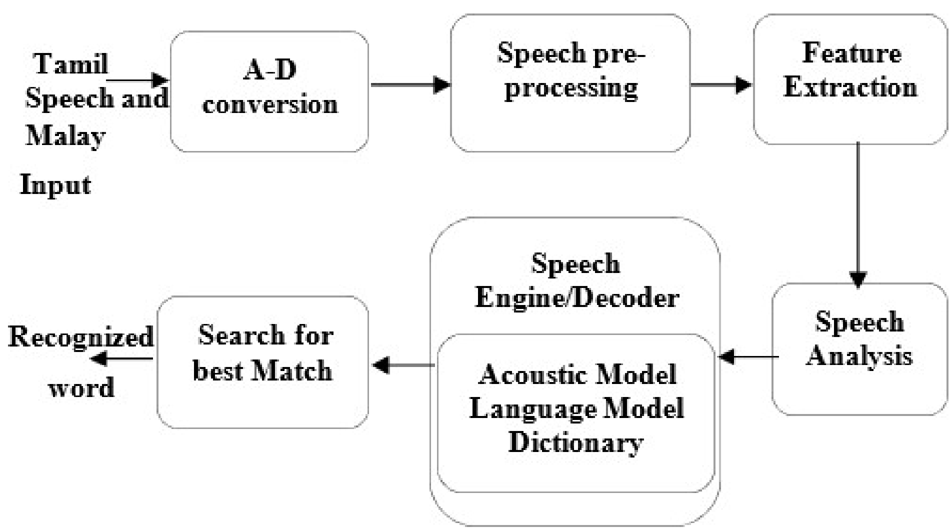
Sigappi etal. examined a method to recognize the preferred words spoken in the Tamil [5]. HMM and ANN network was chosen with the Mel Frequency Cepstral Coefficients. HMM was appointed to design the speech nature and auto associative neural networks to catch the deployment of vector characteristic. An analyzed model has given a path to investigate speech recognition areas for the language of Tamil. An information set have hundred railway station names in TN. Each and every station is stated 5 times by 23 speakers. Out of ten thousand samples, eight thousands samples are used for the purpose of training and remaining for the networks.

Rojathai et al., examined a speech recognition process for the language of Tamil [6]. FEBNN was trained with the characteristic of MFCC. This well trained feed forward back propagation neural network was used for investigating latest characteristics vectors isolated from the new signals. An information set have 10 words in Tamil from 10 people out of 8. These eight people information are examined for the process of training and remaining person's information was utilized for the process of testing. Sifappi et at proposed an ASR method by comparing it with HMM. Both of them strongly states that the rate of recognition of the examined work was succeed high perfection while comparing to the technique which exits. The specificity, accuracy and sensitivity of proposed work are 91%,91% and 89% and the existing process was achieved 65.32%, 76.56%, 87.9% for their research. A next topic narrates the fundamental steps belongs to the growth of Automatic Speech Recognition process.

Because of the heavy population, they analyzed a latest trend of market increment of Tamil and English languages. E.g. there are 58 million native speakers of Tamil and 78 million native speaker of English in 2008[3]. A young people are positive and willing to spend on the smart applications and gadgets in such developing markers. This paper aim to create more vocabulary for southeast Asian languages. These strategies were proceeding on various Asian Automatic Speech Recognition process for Tamil and English languages.

3. Basic Steps Involved in Developing an ASRSystem

The step involved in the growth of Automatic Speech Recognition was shown in the fig2.



system

Fig. 2. Steps Involved in the growth of an Automatic Speech Recognition process.

The 1st step was digitization which converts analog to digital and translates analog to digital signals in the computer. The second step was carried out to analyze the better bandwidth level and execute the input signal for the speech recognition process. The signal of speech consists of small frames to analyze the signal. Therefore this research adopted the MFCC technique for the better representation of audible signal. And finally the search engine will create a decoder or search for phonemes or words comparison with the design. It was trained and created using these vector characteristics.

4. Speech RecognitionTechniques

Researches for more than 55 years have a machine to understand fluently the speech which is spoken. Many techniques are examined and applied in different areas. The DPT was examined for word finding depend on the matching of template. Many researchers highlights the statistical data approach like GMM and HMM. So, it must have the description and knowledge of the current issues [7]. The technique of machine learning like DBN, ANN SVM are examined to place the HMM and GMM process.[8]

5. Creating an Automatic Speech Recognition process

5.1 GATHERING SPEECH CORPORA

Samples of speech are considered as better in having a correct speaker coverage, environment, vocabulary and recording channels to construct a basic purpose of Automatic Speech Recognition. The collection of corpora was shown in the figure 3

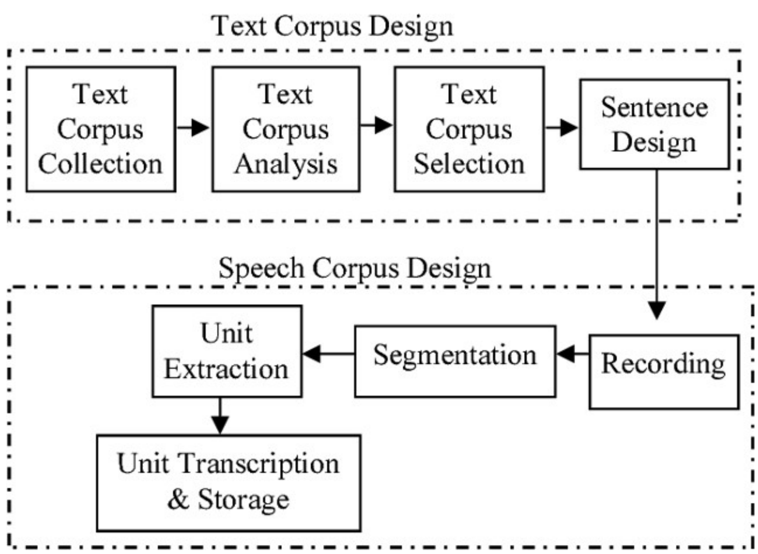


Fig 3 English Language corpora set

A. SpeakerCoverage

i) Only limited amount of speaker involves in this research.
ii) Men and women are equal in number.
iii) Sixteen to sixty year old people are appointed.

iv) They prefer only native speakers.

B. Vocabulary Size andContent

Prompts was modeled and given to everyone to record the speech from appointed speakers. The transcripts have a group of correct words or sentences which are rich in phonemes. It can be picked form various fields like twitter, news, chats and so on. There are twenty words for each sentence. The whole scripts must cover the word as much as possible.

C. Recording Channels andDevices

Recording channels are divided as wideband like mobile gadgets of 17kHertz and narrow band like telephony channel of 9kHertz. Sample speech was recorded through many devices to maximize the Automatic Speech Recognition capability. Eg. 19kHertz sample speech was recorded by mobile gadgets running windows, ios, and androidand so on.

D. Environments

An Analyzed environment is differentiated into two types they are noisy and quiet. The quite includes indoor, office and home environment with no echo sound. And noisy consists of restaurants, streets and shopping malls. A sample combination noted from various environments will develop the energy of the Automatic Speech Recognition process. This research highlights the gathering of information in quite place.

6. PRONUNCIATION DICTIONARY AND TEXT DATACOLLECTION

A. Pronunciation Dictionary: this dictionary was considered as the main elements in the Automatic Speech Recognition process. Some studies examined various kinds of phone for the same Lang to succeed in higher factor of real time or WER. This segment describes the dictionary used for Tamil and English.

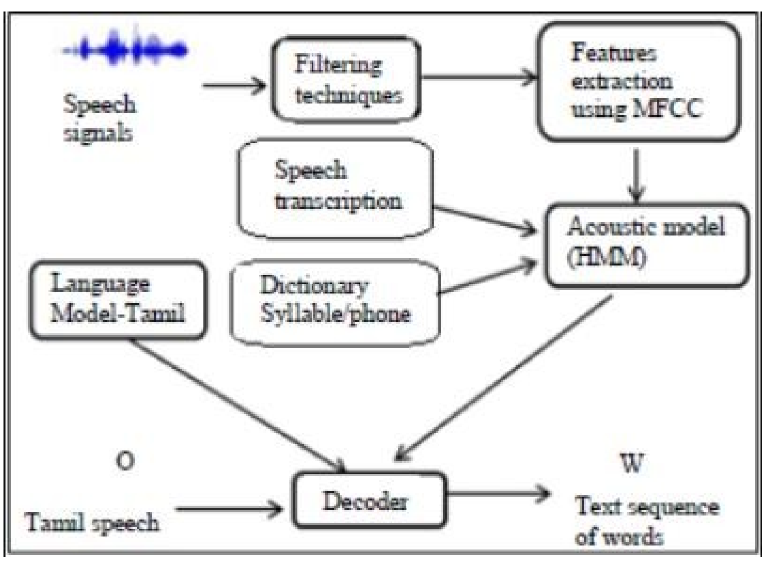


Fig 4 Automatic Speech Recognition system for Tamil Language

English was written with the help of scrip Latin. Its letters are depending upon the twenty six letters from Latin. The fundamental audible languages are straightforward and phonemes way

38 phonemes such as fricatives, vowels, nasals, semi-vowels and plosive.

Tamil Language involves to the Tai group of the Kadai language. It was a tonal language. The meaning of it phoneme follows IPa and have 45 phonemes such as nasals, consonants, diphthongs and vowels. Other than foreign and compound words, the words of Tamil are mono-syllabic. Many syllables are linked with tonal data. There are five tones thye are rising, mid, falling, low and high. It also involves tonal data, diphthongs and vowels in Tamil are classified to the tone classes which were denoted in the dictionary of Mandarin Pronunciation. A dictionary of Tamil has 159 tonal phonemes.

B. Text DataCollection

An important task in text information collection was to gather much information from various places as the Automatic Speech Recognition process focus for basic purpose apps. Everyone should follow the rules to crawl the internet to gather many corpora text.[17-18]

First pick a list of ten thousand words from the transcription training and operate over 10,000 of queries picking randomly any 2 words from the list for each language. Then send the problems to a no. of commercial engines which was searching and gather the Uniform Resource Locator. Then limit the engine of searching to the internet page form the ID domain to ignore the internet pages of other languages. (e.g. BahasaEnglishsia). Group of words are considered due to the some internet pages in Tamil and English languages. Some duplicate Uniform Resource Locator and non-Hyper Text Markup Language urls like Pdf, Jpg and Microsoft power point which are removed and the other hypertext markup languages pages are installed for data cleansing.

EXPERIMENTS

A. Information

In section 2, 850 hour speech information was gathered from the phone for each and every language. The origin speakers are appointed form the all over the world to cover various accents. The statistical data of second Corpora was shown in the tab1. Each corpus from 13,500 speeches is retained for investigating purpose. Pronunciation dictionary was generated and the entries of word were shown in the tab.1. An information web page was gathered form the internet.

.

TABLE I. STATISTICS DATA OF SPEECH CORPORA

	English	***Tamil***
Total No. of Speakers	900	799
Total No. of Utterances	~1.2 million	~1.1 million

Total Durations	799 hours	799 hours
Pronunciation Dictionary Sizes	~131,000	~267,000

B. ExperimentalSetup

An audible characteristics have 14 dimensional Mel-Frequency Cepstral Coefficients feature, 1D tone characteristic and its ấ,ấấandấấấ, in 58 dimensions. An audible design was a HMM and deep neural network design.

The HMM audible design has 36 mono-phones for English and 156 mono-phones for Tamil Language. The context dependent triple phones are designed for about 9000 tied states. The last design was well trained with deep neural network [1] and the top design was trained with the help of MMI technique. The deep neural network design has five invisible segments and each segment have 1087 joints. The MBR [19] was presented to refine the design of deep neural network. A toolkit of Kaldi was used to construct AM.

C. Experimental Results

The experimental analysis of Automatic Speech Recognition system was shown in the tab.2

The Deep Neural Network and Membrane Bio Reactor design of language English achieves 14.8% WER. This type of function was better than the other languages. It was very close to the process of English Automatic Speech Recognition Process. It was acceptable because English and Tamil are from the identical language family. Both languages have 30 mono-phones and more which are very less compared to other languages.

TABLE II. WER USING 3-GRAM WEB LM

	English	***Tamil***
MMI	19.8	45.9
DNN	16.1	79.7
DNN-sMBR	15.6	86.6

The Deep Neural Network and Membrane Bio Reactor design of Tamil succeeded 38.9% wireless network. It is larger than other languages. There should be some reasons. They are 1) Tamil was considered as tonal language. Each and every vowel sound was linked with five various tones. Therefore, five different mono-phones must be developed for a vowel sound. 2) The words of Tamil may be mono-syllable or multi-syllable depend upon the words.3) There was always mismatch in LM web and test information. To verify the AM Tamil, much work was carried out on the local language to decreases the impact of domain difference. The LM web was interpolated linearly with the help of transcripts. The WERs of various Am were decreased by 24.3% 28.9% which are shown in the tab.3

TABLE III. WORD ERROR RATES OF TAMIL ASR USING INTERPOLATED LM

	Tamil
MMI	24.3

7. CONCLUSIONS

Automatic Speech Recognition attracts various applications in the last 2 generations. This research recorded their hard work in creating Automatic Speech Recognition process especially for the people belong to the Southeast Asian Languages. This paper highlights the instinctive recognition of speech for English and Tamil language. This instinctive process was known an Automatic Speech Recognition method. An important issue of Automatic Speech Recognition was analyzed in this study. For various under resourced languages, this Automatic Speech Recognition method was a correct starting point for everyone who interested in this topic. Automatic Speech Recognition was applied in various fields like commercial, education, military, medical and telephony. Speech technology presents its contribution to support the handicapped people to work their internet and computer. These systems was differentiated into many types depend upon the utterance kinds, size of vocabulary, environment, channel and speaker design. other learning techniques like TDNM and LSTM was used in our AM for future research.

References:

1. Thamburaj, Kingston Pal, et al. "An Critical Analysis of Speech Recognition of Tamil and English Language Through Artificial Neural Network."(2021).
2. Shareef, S. R., and Y. F. Irhayim. "A Review: Isolated Arabic Words Recognition Using Artificial IntelligentTechniques."
3. Yu,Fuxun,etal."Masker:Adaptivemobilesecurityenhancementagainstautomaticspeechr ecognition in eavesdropping." *Proceedings of the 56th Annual Design Automation Conference 2019*.2019.
4. Nasereddin, Hebah HO, and Ayoub Abdel RahmanOmari. "Classification techniques for automatic speech recognition (ASR) algorithms used with real time speech translation." *2017 Computing Conference*. IEEE,2017.
5. Lokesh, S., et al. "An automatic tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map." *Neural Computing and Applications* 31.5 (2019):1521-1531.
6. Jeyalakshmi, C., and A. Revathi. "Efficient speech recognition system for hearing impaired children in classical Tamil language." *International Journal of Biomedical Engineering and Technology* 26.1 (2018):84-100.
7. Kalith, Ibralebbe Mohamed, et al. "Comparison of Syllable and Phoneme Modelling of Agglutinative Tamil Isolated Words in Speech Recognition." *Current Journal of Applied Science and Technology* (2018):1-10.
8. Manjutha, M., et al. "Automated speech recognition system–a literature review." *COMPUTATIONAL METHODS, COMMUNICATION TECHNIQUES AND INFORMATICS* (2017):205.
9. Khaw, Yen-Min Jasmina, Tien-Ping Tan, and Bali Ranaivo-Malançon. "Automatic Phoneme Identification for English Dialects." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9.2-9 (2017):85-94.
10. Kusumah,Riky,RudyHartanto,andRisanuriHidayat."Hybridautomaticspeechrecognitio

nmodelfor speech-to-text application in smartphones." *2019 International Conference on ICT for Smart Society (ICISS)*. Vol. 7. IEEE,2019.

11. Akila,A.,andE.Chandra."PerformanceenhancementofsyllablebasedTamilspeechrecognitionsystem using time normalization and rate of speech." *CSI transactions on ICT* 2.2 (2014):77-84.
12. Seman, Noraini, Zainab Abu Bakar, and Nordin Abu Bakar. "An evaluation of endpoint detection measures for English speech recognition of an isolated words." *2010 International Symposium on Information Technology*. Vol. 3. IEEE,2010.
13. Fook, C. Y., et al. "A review: English speech recognition and audio visual speech recognition." *2012 International Conference on Biomedical Engineering (ICoBE)*. IEEE,2012.