ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org

Vol 11, Issuse.2 July 2020

An HMM-Based Speaker-Independent Isolated Speech Recognition System for the Tamil Language.

Mr.T Nagaraju Yadav , Mr.P Viswanatha Reddy , Mrs.B Jyothsna

Abstract:

For almost 50 years, scientists have worked toward the lofty goal of developing a computer that can interpret spoken speech. Do this, an Automatic Speech Recognition (ASR) system must be created. In this work, we zero in on a key role for the Indian language. Several distinct voice recognition technologies are under several fields of use. In this study, we provide a speaker-neutral Tamil isolated voice recognition system. Hidden Markov Model (HMM), the most adaptable and effective method for voice recognition to date, is used in this study's implementation. The HMM technique to voice recognition is natural and very dependable, making it useful in many contexts. High-quality word accuracy of 88% is provided by the HMM experiments for both the training and test utterances of the speakers. Word Error Rate (WER) is used to measure how well a translation system performs, and this analysis yields a WER of 0.88.

Keywords:*Isolated Speech Recognition; Tamil language, Sphinx4; Mel Frequency Cepstral Coefficients; Hidden Markov Model; Word Error Rate.*

Introduction

Humans have an innate ability to communicate via speech. Due to the low barrier to entry for learning to communicate verbally, most people have a natural ease with it. More people will adopt computers if they can communicate with them through voice rather than keyboards and mice. sufficient computers Very good. A computer may use a technique called automatic speech recognition (ASR) to transcribe what a user says into text while speaking into a microphone or phone [1]. The voice, speaker, and vocabulary choices available allow for the development of a wide variety of speech recognition systems. Diverse types of applications need different sets of tags, and so on. Most current studies [2] concentrate on improving voice recognition technology for Indian languages. This study creates a speaker-independent isolated voice recognition system for the Tamil language with a limited vocabularies [3]. Tamil is a Dravidian language spoken mostly in Sri Lanka and the Indian state of Tamil Nādu. It has official recognition in the countries of Sri Lanka and Singapore in addition to the Indian state of Tamil Nādu. As one of the world's most spoken languages, Tamil is understood by more than seventy-seven million people [3]. Therefore, an ASR system tailored to the Tamil language is an urgent need. This study use Sphinx4, which is based on HMMs, to analyse solitary speech that is not.Attributed to a specific speaker. It is a malleable, modular, and pluggable framework designed to encourage fresh approaches to fundamental studies of HMM recognition systems. Because of its adaptability and modularity, Sphinx4 is employed in this study.This paper will be structured as follows. The system overview is presented in Section 2, followed by discussions of the pre-processing stages in Section 3, the post-processing procedures in Section 4, the experimental findings in Section 5, and the performance assessment in Section 6. Section 7 provides a brief conclusion and further research.

Assitant Professor[1,2,3]
Department of CSE
Viswam Engineering College (VISM) Madanapalle-517325 Chittoor District, Andhra Pradesh, India,

System Overview

Isolated speech, linked speech, continuous speech, and spontaneous speech are the categories of utterance-based speech recognition. Isolated speech recognition [4] is easier than the others since the words have distinct boundaries and are often spoken clearly. When speaking into an isolated voice recognition system, as opposed to a continuous speech recognition system, the speaker must stop short between words. That does not imply you may only type in a single word; rather, it means you have to limit yourself to one sentence at a time before you have to pause [4]. There are typically two stages to a voice recognition system. Pre-processing and post-processing are the terms used to describe them. Features are extracted during pre-processing, and an acoustic model, phonetic lexicon, or pronunciation dictionary, and language modelling are created during postprocessing. The big picture of the ASR system is shown in figure1 below.

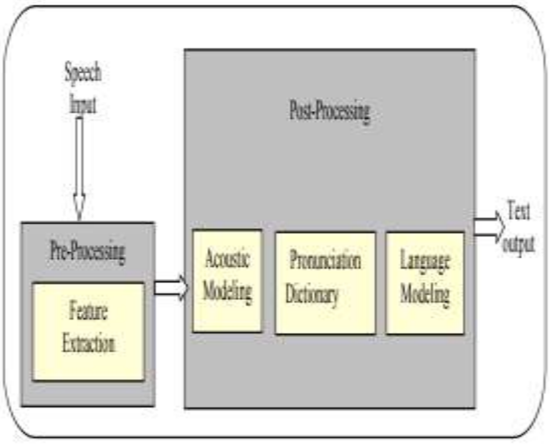


Figure 1 Steps involved in ASR system.

The first step involves converting the speech waveform into a new representation, such as a collection of vectors whose values indicate attributes or parameters. Acoustic models are constructed using the speech parameters that were extracted throughout the recording process. Terms from the Language Model are spoken in the Dictionary. Words are broken down into their component sounds using the Acoustic Model, which informs the pronunciations. The second piece is a language model, which in this case would offer the likelihood of word sequences. In the next paragraphs, you will find comprehensive details on the whole setup.

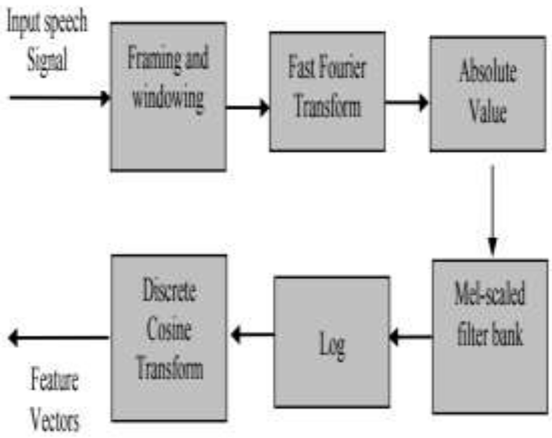
Pre-Processing:

The front-end or pre-processing of a speech recognizer is in charge of feature extraction from the speech waveform. The goal of feature extraction is to reduce the amount of data needed to do further analysis on a voice waveform. The Mel Cepstral Frequency The most blatant example of a feature set that is widely utilized in voice recognition is the multi-frequency, continuous-coefficients (MFCC) feature set. It comes closer than any other system I have seen to mimicking the human reaction. The Front-End of Sphinx-4 is commonly set up to generate MFCCs [5]. The MFCC vector is calculated from each frame using a technique based on short-term analysis [7]. Using Eq. (1), we can determine the MFCC.

$$Mel(f)= 2595*\log10(1+f/700)$$

The following figure 2 shows the steps involved in MFCC feature extraction [7]. These features are.

used for further process of post processing.



Post-Processing

Fundamentally, the problem of speech recognition can be stated as follows. When given with acoustic observation X = X1, X2…XNA, the goal is to find out the corresponding word sequence W = W1, W2…Wm that has the maximum posterior probability P(W|X) expressed using Bayes Theorem as shown in equation

(2).

$$W = \arg\max_{W} P(W/X) = \arg\max_{W} \frac{P(W)P(X/W)}{P(X)} \qquad \text{---------} (2)$$

Where P(W) is the chance that the word W will be said, and P(X|W) is the likelihood that X will be audibly observed when the word W is spoken. Class conditioned probability distribution P (X|W) is another name for it. P(X) is the typical likelihood that X will be seen. Equation (2)'s maximization may be seen as done with X held constant, finding the word W is as simple as maximizing the numerator.There are three essential aspects, sometimes known as post-processing components, which must be included in every ASR system if it is to successfully conduct speech recognition. Either the Pronunciation Dictionary Language Model or the Acoustic Model Phonetic Lexicon During the recognition procedure, these three models must cooperate with one another. Read on for a comprehensive breakdown of each.

Acoustic Models

For a speech recognition system to be effective, acoustic models must come primarily [4]. This step is crucial since it determines the overall efficiency and precision of the search. Establishing statistical models for the acoustics of human speech is often referred to as acoustic modelling of speech. sequences of feature vectors represented as a representation based on the voice waveform. In addition, "pronunciation modelling" is included, which explains how a sequence or multi-sequence of basic speech units (such phones or phonetic feature) is utilized to represent bigger speech units like words or phrases that are the target of speech recognition. In this study, we use a variant of HMM [6] that is used in the most popular acoustic models.

Pronunciation Dictionary

You may find information on how to properly pronounce words in a Pronunciation Dictionary, which is also known as a lexicon. All the terms that the voice recognition system needs to understand should be in the lexicon. Records in the pronunciation dictionary include nouns and their corresponding monophony successions [4]. There are often a variety of ways to say a single word. A pronunciation dictionary, like a speech corpus, is a specialized tool for a certain language. Therefore, a manual pronunciation dictionary has been built for the whole Tamil lexicon. This is an example of a Tamil pronunciation dictionary.

திருக்குறள்	thirukkuRaL
இணையம்	inNaiyam

Language Model

Word searches are limited by a linguistic model. It aids a speech recognizer in determining, independently of acoustics, how probable a word sequence is. It stands for what you already know about language and what you anticipate from words. It might be stated in terms of the words or word combinations that are allowed or how often they happen to be. Word sequences in text corpora are often used for training purposes. The language model's pronunciation dictionary must include all of the words in the model. Here is an example of the grammar we developed for solitary Tamil speech.

```
வார்த்தை = ( அமைப்பு | அழி | அனுப்பு | இணையம் | உலகச்செய்திகள் );
```

Experimental Results

Sphinx4 is used to create a small-vocabulary speaker-independent isolated voice recognizer for the Tamil language. When it comes to HMM-based speech recognition systems, Sphinx-4 is at the forefront. In order to do so, HMM-based speech recognition systems [10] estimate the probability of each phoneme at a given time. frames that are all part of the same voice transmission. The most probable order of phonemes is found by a search technique [8]. Phoneme sequences matching words in the lexicon are prioritized in this search, and those with the greatest overall probability are assigned to actual words in speech. Having a speech corpus [7, 8] is crucial for the creation of any ASR system. The lack of a pre-existing speech corpus necessitates the human creation of a Tamil one. A corpus of fifty discrete spoken utterances from ten different women is utilized for instruction. Each speaker's whole vocabulary is represented in the database five times over. A total of 2,500 words (50 times ten times 5) are used in this study.

Performance Evaluation

The word error rate is a common metric used to evaluate the accuracy of a voice recognition system. This rate is calculated by dividing the number of incorrectly categorized words by the total number of words that were subjected to testing. Independent of vocabulary quantity and noise, ASR studies have focused on achieving a recognition error of zero in real time. features of a particular speaker or accent. Out of a total of ten speakers, this study only puts four speakers' data to the test. On average, the algorithm can understand 44% of the words you feed it. Using equation (3), we can determine the WER [9].

$$\mathrm{WER} = \frac{\text{Number of words correctly recognized}}{\text{Total number of words}}$$

The analysis found an error rate of 0.88%. Vocabulary size affects the accuracy of speech recognition systems; smaller vocabularies are easier to work with, whereas larger vocabularies have a higher word mistake rate.

Conclusion

Researchers have hoped for years for the day when they might create technologies that would allow people to communicate with computers in a natural way. Creating an automatic speech recognition system may help with this. Isolated voice recognition system for the Tamil language with a short vocabulary is shown in this study. was created, and sphinx4 is being used to examine its effectiveness. HMM is used to implement the four major components of an ASR system: feature extraction, acoustic model, pronunciation dictionary, and language model. This study uses a database of 2,500 words, which yields an accuracy rate of 88%. The technique provides a low word mistake rate since the vocabulary size is minimal. In the not-too-distant future, speakers of Tamil will be able to put into practice and experience medium- or large-vocabulary isolated speech and continuous speech.

References

[1] Mr. R. Arun Thilak & Mrs. R. Madurai, "Speech Recognizer for Tamil Language", Tamil Internet 2004, Singapore.

[2] M. Chandrasekar, M. Panabaker, "Spoken TAMIL Character Recognition", in Electronic Journal Technical Acoustics (EJTA), ISSN 1819-2408, 2007.

[3] M. Chandrasekar, and M. Panabaker, "Tamil speech recognition: a complete model", Electronic Journal Technical Acoustics (EJTA), ISSN 1819-2408, 2008.

[4] Gailes RAŠKINIS, "Building Medium-Vocabulary Isolated-Word Lithuanian HMM Speech Recognition System", INFORMATICA, Vol. 14, No. 1, 75–84 75, 2008.

[5] Apoherm Charles and Devaraj, "Alai gal-A Tamil Speech Recognition", Tamil Internet 2004, Singapore.

[6] Mohammad A. M. Bashara, Raja Nailon, Rozita Zainuddin, Moustafa Eltife and Othman O. Khalifa, "Natural Speaker-Independent Arabic Speech Recognition System Based on Hidden Markov Models Using Sphinx Tools", International Conference on Computer and Communication Engineering (ICCCE 2010), 11-13 May 201), Kuala Lumpur, Malaysia.

[7] Rath navel, Kanupriya, A.S. MuthannaMur gavel, "Speech Recognition Model for Tamil Stops", Proceedings of the World Congress on Engineering 2007 Vol I WCE 2007, July 2 - 4, 2007, London, U.K.

[8] Saraswathi and TV Gateway, "Morpheme based language model for Tamil Speech Recognition system", The International Arab Journal of Information Technology, Vol.4, No.3, July 2007.

[9] H. Satori, M. Harti and N. Chen four "Arabic Speech Recognition System Based on Cushing", 1-4244-11 58-0/07/$25.OO © 2007 IEEE.

[10] R. Thangarajan, A.M. Natarajan, "Syllable Based Continuous Speech Recognition for Tamil", South Asian Language Review, VOL. XVIII. No. 1, January 2008.