Voice and speech recognition in Tamil language

Conference Paper · February 2017

DOI: 10.1109/ICCCT2.2017.7972293

4 authors, including:

Kiran Ramasamy
Carleton University
2 PUBLICATIONS **13** CITATIONS
SEE PROFILE

T. Subha
Sri Sairam Engineering college
14 PUBLICATIONS **128** CITATIONS
SEE PROFILE

VOICE AND SPEECH RECOGNITION IN TAMIL LANGUAGE

R.KIRAN[1], K.NIVEDHA[2], S.PAVITHRA DEVI[3]
Department of information Technology, Sri Sai Ram Engineering College,
Chennai-44, Tamil Nadu.
[1]kiranramasamy95@gmail.com
[2]niveakalathi26@gmail.com
[3]pavithu29@gmail.com

SUBHA.T
Associate professor,
Department of information Technology, Sri Sai Ram Engineering College,
Chennai-44, Tamil Nadu.
subharajan@gmail.com

ABSTRACT:

Speech Recognition is one of the most important and popular advancements in technology that provides human-device (computers, smart phones, etc.,) interaction that helps in handling modern devices in an easier and convenient manner. It is also known as "automatic speech recognition "(ASR," computer speech recognition or just "speech to text" (STT) Speech recognition (SR) is the process of translation of spoken words into text. There has not been much consideration for Tamil language to be used in voice and speech recognition in smart phones. For native users Tamil voice recognitions and speech would provide more flexibility in smart phone experience. Also people who have only been used to their native language Tamil, would feel easier to use the speech recognition system in their smart phones if provided in Tamil. There will be no more difficulty in usage of phones for local users and there is no need for any learning to use the smart phone. Our intention is to create a voice and speech recognition system in smart phones that recognizes voice and captures the speech data in Tamil and stores and converts the captured speech as text in Tamil language itself. This can be used in voice dialing, sending SMS by saying out the message and the captured message is sent to the recipient in Tamil. This is popularly called as Automatic Speech Recognition (ASR) system. Dynamic Time Warping (DTW), Statistical Pattern Matching techniques, Machine Learning techniques such as Neural Networks (NN), Support Vector Machine (SVM), and Decision Trees (DT) are some of the template pattern matching techniques for ASR system development. Statistical pattern matching techniques is further classified into Hidden Markov Model (HMM) and Gaussian Mixture Models (GMM) techniques, among which HMM is the most popular technique. For this ASR system, HMM technique is used to achieve highest word recognition accuracy. In our case, it offered 86.5% and approximately 92% accuracy, during training and testing process respectively.

***Keywords*-Voice, Speech, Recognition, Tamil, HMM, Smartphone.**

1. INTRODUCTION:

The term speech recognition or speech identification refers to identifying the words uttered by the speaker, rather than identifying the speaker. Speech recognition applications include voice user interfaces such as voice dialing (e.g. "Call Office"), call routing (e.g. "I would like to make a conference call"), search ,simple data entry (e.g., entering a password), speech-to-text processing(can be used for sending text messages).ASR applications are becoming common and useful in this day and age as many of the modern devices are designed and produced user-friendly for the convenience of general public. Speaking/communicating directly with the machine to achieve desired objectives make usage of modern devices easier and convenient. Though there are many interactive existing software applications for speech recognition, the use of these applications are limited due to language barriers. Hence development of speech recognition systems in local languages will help anyone to make use of this technological advancement. Therefore, our application recognizes Tamil words from speaker through a microphone and converts it to text in Tamil language itself. This process start from developing a phonetic dictionary for Tamil language and vocabulary files, which is further used to develop acoustic and language models for training the recognition of speech.

2. SYSTEM IN PREVALENCE:

There are many existing systems for voice and speech recognitions that are used worldwide in form of software and mobile applications. Some of the popular speech recognition systems are Dragon natural speaking, speech logger, talk text, etc. Most commonly used voice and speech recognition software in smart phones is Google now, Apple Siri and Windows Cortana. These existing systems are available only in few languages and have not been used in any of the Indian languages. Thus this hinders the native users to make use of the technical advancement of ASR systems.

3. DISADVANTAGES OF THE EXISTING SYSTEMS:

i. The use of these applications is limited due to language a barrier and there is no flexibility for native users.

ii. Lack of development in speech recognition systems in local languages has hindered Indian smart phone users to make use of this technological advancement who feel difficulty to use these systems in a foreign language rather than their own language.

4. SYSTEM OVERVIEW:

Our intention is to create a voice and speech recognition system in smart phone that recognizes voice and captures the speech data in Tamil and stores them and converts the captured speech as text in Tamil language itself. This can be used in voice dialing, sending SMS by saying out the message and the captured message is sent to the recipient in Tamil. There has not been much consideration for Tamil language to be used in voice and speech recognition in smart phones. For native users Tamil voice and speech recognition would provide more flexibility to use smart phone. Tamil is one of the most widely spoken languages of the world with more than 99 million speakers. Hence, there is an urgent need for the system, so that people can easily interact in Tamil language. Hence development of speech recognition systems in Tamil language will help native users to make use of this technological advancement in a beneficial way.

5. HMM - A SPEECH RECOGNITION TECHNIQUE:

HMM is a popular statistical pattern matching approach for speech recognition that achieves highest accuracy among other speech recognition techniques. It is based on statistical acoustic and language model and is the dominant technique for speech recognition. Automatic learning procedures are used by HMM to model the variations of speech. They are simple networks that generate speech, using a number of states for each model.

The parameters of the model are,

- Means
- State Transition Probabilities
- Mixture Weights
- Variances

These parameters characterize the state output distributions. It will have a different output distribution for each word or each phoneme. A HMM for a sequence of words or phonemes is generated by concatenating the individual trained HMM for the separate words and phonemes. The decoding process works by starting at upper left corner of trellis and generate observations according to permissible transitions and output probabilities. A HMM is defined by the output and transition probabilities .This algorithm not only computes the likelihood of single path but also the overall likelihood of observation string as sum of overall paths in trellis is computed. There are three major problems in developing an ASR system. They are addressed by HMM and the solutions are given by the following algorithms:

i. *Forward algorithm*-Computes the overall likelihood of generating strings of observations from HMM.

ii. *Viterbi algorithm*-Decodes the most likely state sequence/best path from HMM.

iii. *Forward-Backward algorithm/Baum-Welch*-Learns parameters (output and transition probabilities) of HMM from data.

6. HMM TECHNIQUE - ADVANTAGES:

i. HMM scales well. Hence it is Extendable.

ii. The complexity and duration of recognition process is extremely reduced for training large vocabulary.

iii. The prevalent model structure is Continuous mixture density HMM that achieves best results in accuracy.

iv. Keyword spotting, speech understanding and machine translation are some of the broadened applications of HMM.

v. Applications related to non-Speech are emerging as well.

vi. It is easier to train large vocabulary for recognition process.

.

7. ADVANTAGES OF PROPOSED ASR SYSTEM:

i. Easier, faster, better usage of phones for Tamil smart phone users.

ii. Improves the standards of Tamil language in the smart phone world.

iii. Wide usage of Tamil in speech recognition systems increases the prevalence and creates awareness of the language among people.

iv. Easier for both sender and receiver in case of text messages and the recognition system in mobile phone converts the voice data in Tamil in to Tamil text itself.

v. No more difficulty in usage of phones for local users and there is no need for any learning to use the smart phone.

8. STRENGTH OF HIDDEN MARKOV MODEL:

The strength of the HMM method lies in two broad areas:

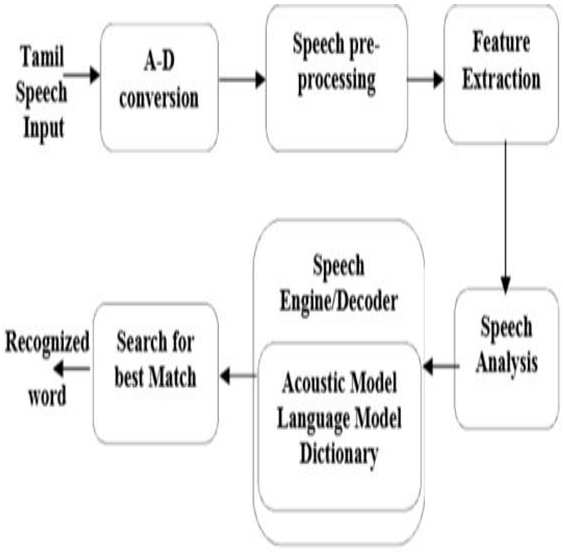
(1) Mathematical Framework

(2) Implementation Structure.

Mathematical Framework: HMM uses the statistical methodology and for related problems it provides solutions in a straight forward manner.

Implementation Structure: HMM helps in dealing with various sophisticated speech-recognition tasks as it is flexible. It is also easy to implement.

9. SYSTEM FLOW MODEL:



Analog to digital conversion is the first step in the development of ASR system where analog signal is translated to digital signal.

After this, preprocessing occurs that checks for better frequency levels to produce optimum speech input signal.

Then feature vectors are extracted which is an important task to achieve high accuracy in recognition.

Next, analysis of the extracted feature vectors is carried out.

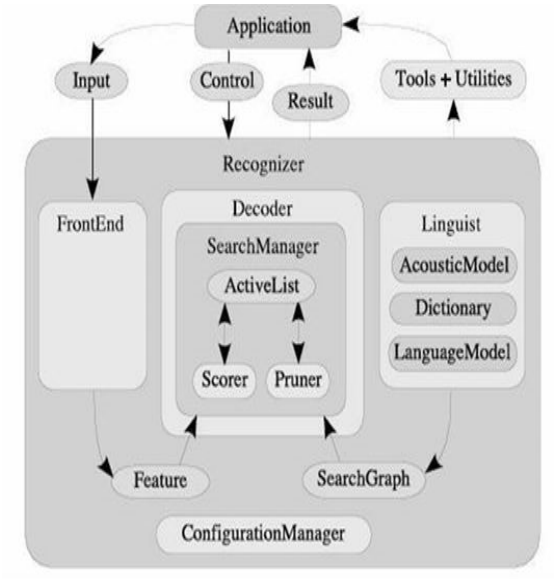
The main components comprised by the Speech engine/Decoder are:

- Acoustic model
- Language model
- Dictionary

The dictionary file consists of the phonetically transliterated version of the added Tamil words.

In the Speech engine/Decoder, the extracted features are scored with the phonemes, generated from the acoustic and language models for best match and that is returned as the recognized output.

10. SYSTEM ARCHITECTURE (CMU SPHINX):



There are three primary modules in the Sphinx-4 framework:

- Front End
- Linguist
- Decoder

Front End - The Front End takes one or more input signals and parameterizes them into a sequence of *Features.*

Linguist - The Linguist translates any type of Standard language model, along with pronunciation information from the *Dictionary* and structural information from one or more sets of *Acoustic Models*, into a *Search Graph*.

Decoder - The *Search Manager* in the Decoder uses the Features from the Front End and the Search Graph from the Linguist to perform the actual decoding and generates *Results*.

11. CONCLUSION:

Thus the voice and speech recognition system in Tamil language when implemented would be of great use to native users who feel difficulty with foreign languages that are default in their smart phones. This enables all types of smart phone users to easily use the technical advancements in their phones in their own indigenous language.

REFERENCES:

[1] Ms. Vimala C., V. Radha**, Isolated Speech Recognition System For Tamil Language Using Statistical Pattern Matching And Machine Learning Techniques,** *Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore.*

[2] Cini Kurian, **A Survey on Speech Recognition in Indian Languages** , *Department of Computer Science , Al- Amine College, Edathala, Aluva, Kerala-683561.*

[3]Willie Walker, Paul Lamere, Philip Kwok, Bhiksha Raj, Rita Singh, Evandro Gouvea, Peter Wolf, Joe Woelfel, **Sphinx-4: A Flexible Open Source Framework for Speech Recognition.**

[4] http://cmusphinx.sourceforge.net/wiki/tutorialdict

[5]https://www.medium.com/@imsreenadh

[6]http://cmusphinx.sourceforge.net/wiki/phonemerecognition?s

[7]https://www.quora.com/How-would-you-phonetically-translate-the-following-Tamil-Words-pronunciation-in-English

[8] www.anunaadam.com

[9]www.kandupidi.com/editor

[10]http://dhvani.sourceforge.net/doc/tamil.html

[11]http://manual.audacityteam.org/index.html

[12]Dr.Thennarasu, S.,**Tamil Pronouncing Dictionary: A promising trend in Language Technology Development,***Department of Linguistics,Central University of Kerala.*

[13]http://www.omniglot.com/writing/tamil.htm