Alaigal-A Tamil Speech Recognition

A.P.Henry Charles[1] & G.Devaraj[2]
[1]<henry_charles45@yahoo.com>, [2]<g_devaraj1247@yahoo.co.in>
VelTech Engineering College, Chennai, India.

Abstract

The intent of this paper is to build an enhanced speech recognition in Tamil with speaker independent, device independent system with continuous recognition. Here the spoken sentence is represented as sequence of independent acoustic phonetic units. The system recognizes the spoken queries in the context of many applications as Web Browsing etc.

1. Introduction

In the present era of human computer interaction, the educationally under privileged and the rural communities of India are being deprived of technologies that pervade the growing interconnected web of computers and communications. One good solution for this problem would be computers interact with the common man in the language he is comfortable to communicate in. Indian population has a significant percentage of people who are educationally under-privileged. There are still quite a large number of areas where people do not have the capabilities of 3R's. The digital divide under such circumstances is constantly on a rise, where on one hand we claim that India is leading in IT and on the other hand, the advances we make are totally inaccessible by a large number of countrymen. Under such circumstances, we cannot expect rural/educationally under-privileged countrymen to use computers and IT products unless we remove the need of being literate, which exists as a barrier between them and computers.

Software application having speech and voice recognition abilities have a better chance to communicate with a large percentage of population which include educationally under-privileged, visually challenged and computer illiterates, if these applications can speak and understand the native language.

2. Continuous Speech Recognition

A continuous speech system operates on speech in which words are connected together, i.e. not separated by pauses. Continuous speech is more difficult to handle because of a variety of effects. First, it is difficult to find the start and end points of words. Another problem is “co-articulation”. The production of each phoneme is affected by the production of surrounding phonemes, and similarly the start and end of words are affected by the preceding and following words. The recognition is also affected by the rate of speech.

2.1 Tamil Continuous Speech Recognition

When compared to other languages, modeling a Tamil continuous speech recognition is not

.

so tedious because

· Each and every syllable is a character (in Tamil).

· Occurrence of single syllable in different positions of word can make different sound.

E.g. 'க' in **words like** கணினி **as ka,** சிங்கம் **as ga.**

· Phoneme Selection in Tamil becomes easy because of characteristics of Tamil grammar.

3. Architecture of Tamil Speech Recognition

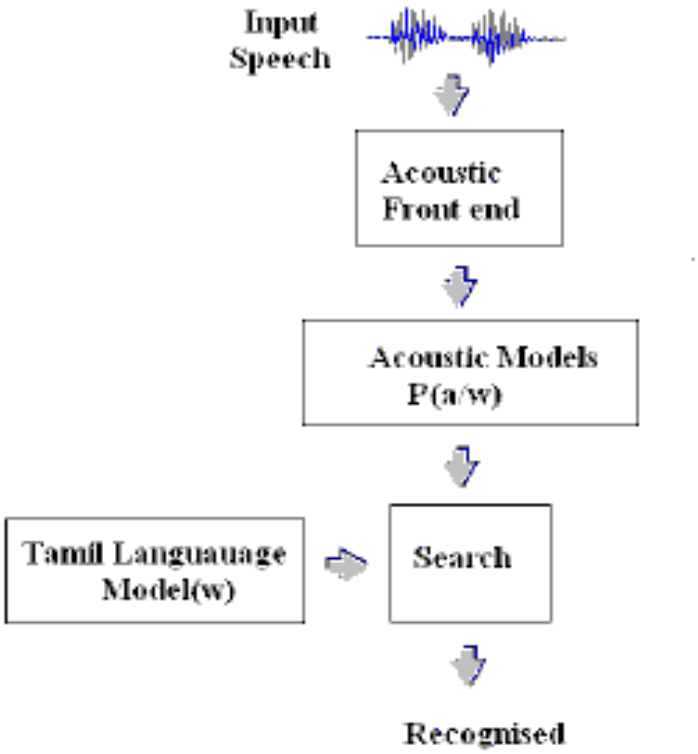

Figure 1. Architecture of a Tamil Recognition System

Initially the speech signal is converted into a sequence of feature vectors based on spectral and temporal measurements. Then the Acoustic models represent the sub-word unit's phonemes as a finite state machine in which, states model spectral structure and transition model temporal structure. The Language model predicts the next set of words, and controls. Finally Search system is crucial to the system, since many combinations of words must be investigated to find the most probable word sequence.

4. Operational Structure For System

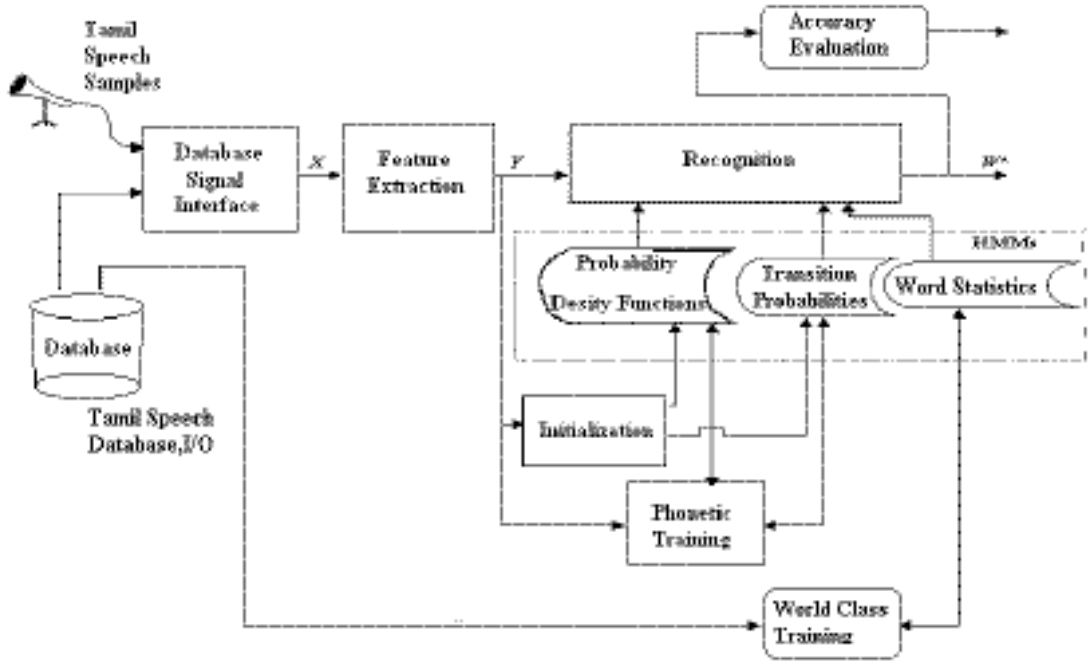

Figure 2. Operational Structure Of Tamil Continous Speech

4.1. Acoustic Modeling

Acoustic Modeling of Tamil speech recognition involves three process as Feature Extraction, HMM and Parameter Estimation

4.1.1. Feature Extraction

Process of incorporating knowledge of the nature of Tamil speech sounds in measurement of the features. Here we utilize rudimentary models of human perception.

Steps Involved are

· Measure features of Tamil speech samples 100 times per sec.Use a 25 msec window for frequency domain analysis.Include absolute energy and 12 spectral measurements.Produce time derivatives to model spectral change.

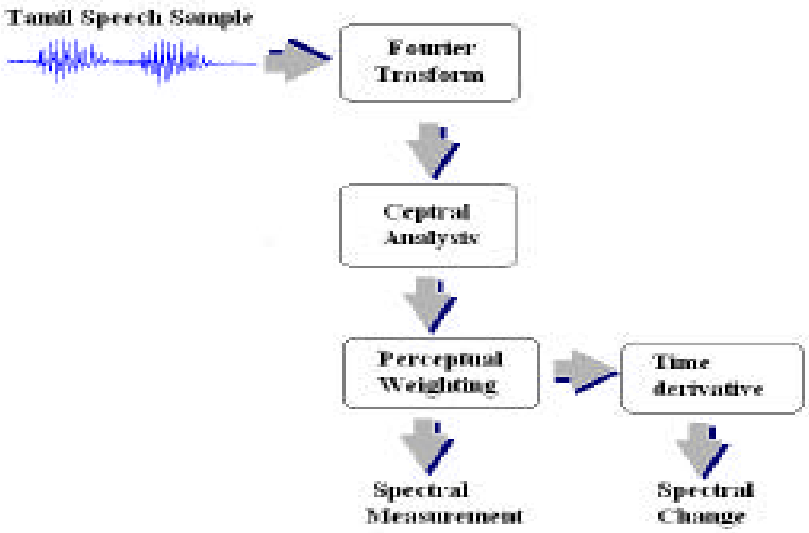

Figure 3.Acoustic Modeling-Feature Extraction

4.1.2. HMM in Acoustic Modeling of Tamil speech system

HMM is used to encode the temporal evolution of the extracted features. Gaussian distributions are used to measure variations in speaker, accent, and pronunciation. Phonetic model in Tamil are simple left-to-right structures with separation of phonemes by Tamil grammar specification. Skip states and multiple paths are also common features of this model. Sharing model parameter in Tamil language is a common strategy to reduce complexity.

4.1.3. Parameter Estimation

Closed-loop data-driven modeling is used to estimate parameter from a word-level transcription. Single Gaussian Estimation processes the word level transcription. These estimates are then splitted accordingly. Batch mode parameter updates are typically preferred for Continuous Speech recognition on Tamil. The decision-tree algorithm is used to optimize parameter sharing of the system.

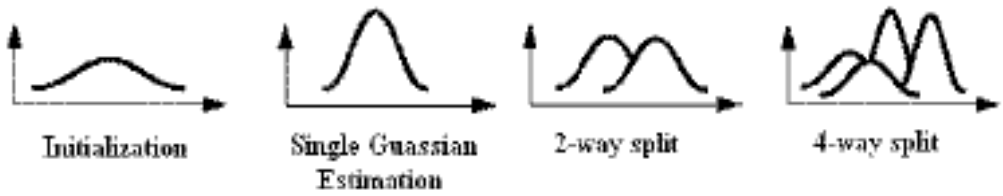

Figure 4.Parameter Estimation

4.2. Language Modeling

Unigram Models

Most Common- ஐ, தான் Rank 100- [illegible], [illegible] Least Common-[illegible]

.

Bigram Models

E.g.- நீ அறிவாய்

Trigram Models

E.g.- நீ அதை அறிவாயா ?

Tamil Language Processing on Continuous Speech Recognition

Speech recognition typically produces a word-level time-aligned annotation. Time alignments for other levels of information are also available. The Syntactic sentence is then separated accordingly to their phonetics, which is explained in detail in the diagram.

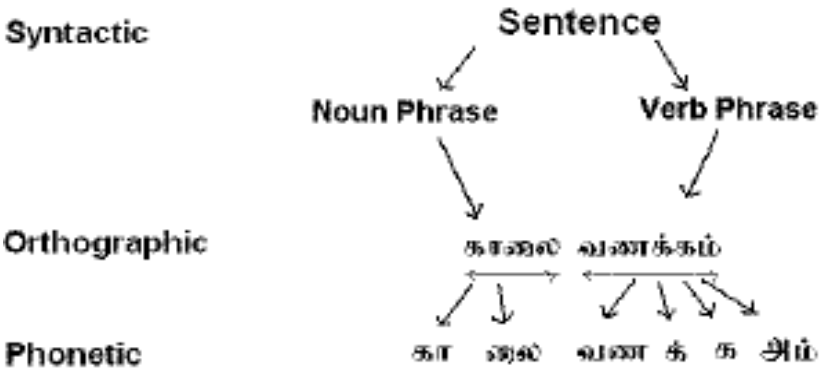

Figure 5.Integration of Natural language

4.3. Dynamic Programming Based Search

Dynamic programming is used to find the most probable path through the network. Beam search is used to control resources. Search is time synchronous and left-to-right. Arbitrary amounts of silence must be permitted between each word. Words are hypothesized many times with different start/stop times, which significantly increase search complexity.

Eg. Search for words அவன் பாடினான் can be done as explained in Fig 6.

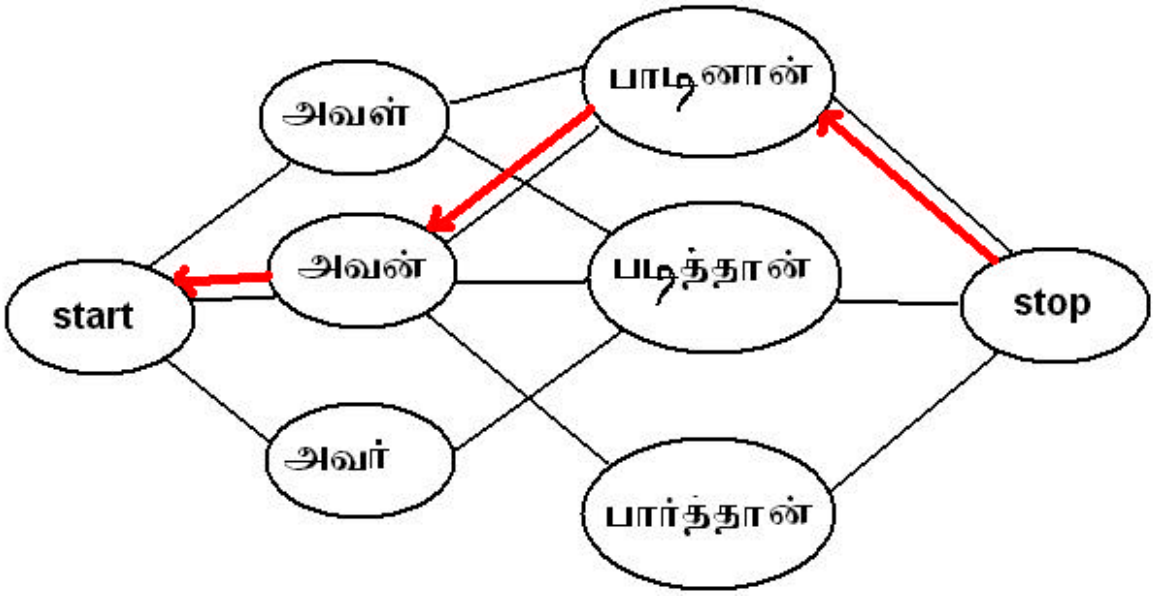

Figure 6.Dynamic programming based Search

5. Future Extensions

Robustness:

In a robust system, performance degrades gracefully (rather than catastrophically) as conditions become more different from those under which it was trained. Differences in channel characteristics and acoustic environment should receive particular attention.

Portability:

Portability refers to the goal of rapidly designing, developing and deploying systems for new applications. At present, systems tend to suffer significant degradation when moved to a new task. In order to return to peak performance, they must be trained on examples specific to the new task, which is both time-consuming and expensive.

Spontaneous Speech:

Systems that are deployed for real use must deal with a variety of spontaneous speech phenomena, such as filled pauses, false starts, hesitations, ungrammatical constructions and other common behaviours not found in read speech, development on this will make a better result on this area.

6. Conclusion

However, field of speech recognition technology has been completely revolutionized by natural language processing with the advent of continuous speech input. This type of input

capability allows Tamil users to do word processing, web-browsing applications on their own language. I hope **Alaigal** will act as a pillar to build the dreams of Tamil IT.

7. References

1. Spoken Language Generation and Understanding - edited by J. C. Simon, NATO Advanced Study Institutes Series
2. The Physics of Continuous Speech - D. B. Fry, Cambridge University Press, 1979.
3. Speech Recognition Principles - W. A. Ainsworth, Peter Peregrinus Ltd., 1988.
4. Reviewing Automatic Language Identification - Yeshwant K. Muthusamy, Etienne Barnard and Ronald A. Cole, IEEE Signal Processing Magazine, October 1994, p.33-41
5. Elements of Acoustic Phonetics - Peter Ladefoged, The University of Chicago Press, 1962.

.