Sharika
Malayalam Speech Recognition System

Shyam.k
MES College of Engineering,Kuttipuram
shyam.karanattu@gmail.com

ABSTRACT

Sharika - an Automated Speech Recognition(ASR) system is an initial initiative in malayalam speech recognition.ASR in its wider perspective is still a dream project,in malayalam.Though the speech corpus required for building the models are ready built,we cannot make use of those corpus as those are given under non-sharable licence.Since ASR in malayalam in its true sense is a herculean time consuming project,here our endeavour is to make a pilot project limiting the word count by confining the ASR to a particular task namely voice control for desktop commands.

1.INTRODUCTION

Among the most ancient human communication systems, verbal form of communication is the most scientific mode.When we go deep in to the human history,we can find that vocal communication was effective far before written methods.In this most modern age also one cannot underestimate the effectiveness of speech as a mode of communication.

If our computer can occupy the place of a stenographer,read news according to our choice,can teach students in their own mother tongue,it will be a major break through.We leave the practical applications of speech recognition to the imagination of the readers as it is very vast.This effort becomes invaluable for those with physical inabilities.Our endeavour here is to minimize the digital divide by making computer applications more user friendly-just by spelling it! that too in our mother tongue.

2.MALAYALAM

Malayalam belongs to the Dravidian family of languages and is one of the four major languages of this family with a rich literary tradtion. Malayalam is the official language of the state of kerala ,India.Its spoken by around 37million people.Indian language at large and malayalam in particular are very phonetic in nature.The structure of alphabets in malayalam is arranged based on the method of sound production.As we pass through the consonant list ,we can see that its so arranged that the noice source of consonant sound, moves from inner mouth to lips.

According to KeralaPaniniyam the alphabets of malayalam consists of 16vowels and 37 consonants,though this has been to slight modifications through ages.The list of vowels is made of 4 short vowels and 3 long vowels in the division of samanakhshara,2 short vowels and 4 long vowels in the division of sandhyakshara and two more as അം[aum] and അ:[ah].The figure below shows an illustration on how the

formant frequencies vary for some of the vowels

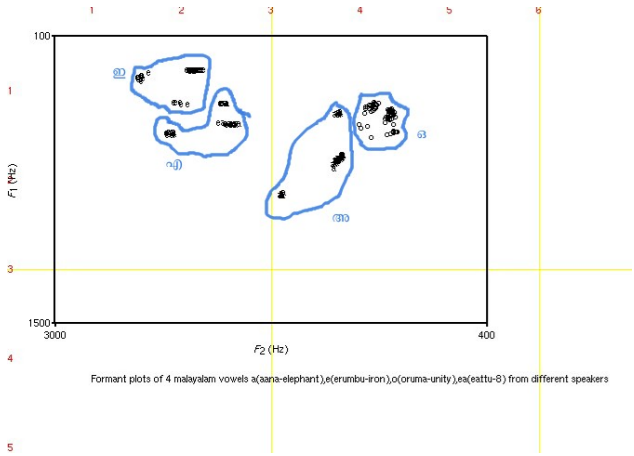

Illustration 1: graph shows first formant versus second formant frequency.It shows how formant frequencies change for different malayalam vowels for different speakers

The division of consonant sounds are even more interesting.Consonant sounds are produced by partial or complete closure of upper vocal tract.So taking the famous analogy of method of speech production with electrical networks,we have the noice sources changing their place along the vocal tract,as the consonant changes.Interestingly in malayalam, they are divided on the same basis of the place where the flow of air is affected.

2.SPEECH RECOGNITION

This project uses the Sphinx engines developed by Carnegie Mellon University (CMU). SPHINX is one of the best and most versatile recognition systems in the world today.Sphinx is having a BSD derived license so people can share the software and use it as they wish.But propreitary software authors can also use sphinx to make propreitary software due to the famous liberal attitude of BSD license. Sphinx has many versions which meets different environments in which the system is employed.Sphinx train is the acoustical model trainer used for this project and I am using sphinx2 as the decoder.The sphinx train is a statistical acoustic model trainer using Hidden Markov Models (HMM). Due to inavailability of a speech corpora, the project is designed to have a wrapper around sphinxtrain so that people can easily train the acoustic models for sharika and use those models to do the desktop control.So A new user first trains his model for sharika and then uses the system.This will greatly improve the accuracy of the model as a near-perfect user independent model itself requires long-term project of collection of speech corpora and tuning the models to perfection.

The desktop control part is developed over the GNOME libraries so that the gnome users can control their window navigation and general desktop control using this system.Initially I developed a simple 8 command system which is now enhanced to above 50 commands.

The sphinx decoder requires two things to perform decoding.

a)Acoustic model and

b)Language Model

2.a)Building the Acoustic Model

Acoustic model is developed using sphinxtrain package.Sphinx train has the training algorithms implemented so that we can configure the system and give the required data i.e. both speech and other configuration files such as dictionary, phonelist and transcripts to build the models in the way we want.The task of defining the phonemes is much simple,thanks to the phonetic nature of malayalam language. Majority of the lettters themselves form phonemes.We can remember the fact that almost all written malayalam words unlike english words can only be read in one way which shows how phonetic they are.As the project requires users to train their own models,the speech database made to train the models uses only my sound.Phone list,dictionary file and transcripts specific to this project are made.

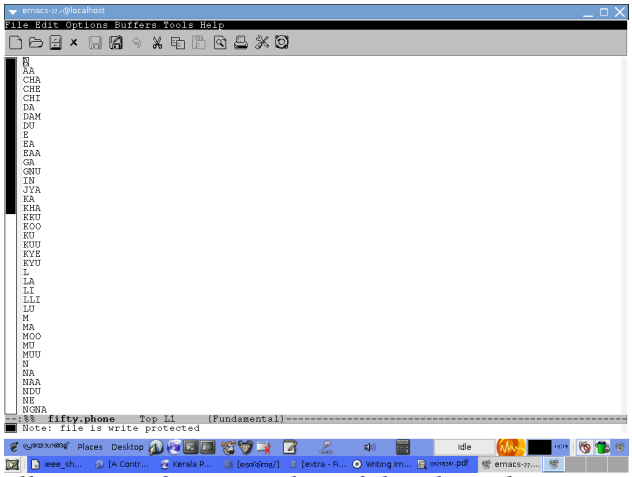

Illustration 2: screenshot of the phone list

Users are expected to read out the transcripts as they do the training as per the trainer program.

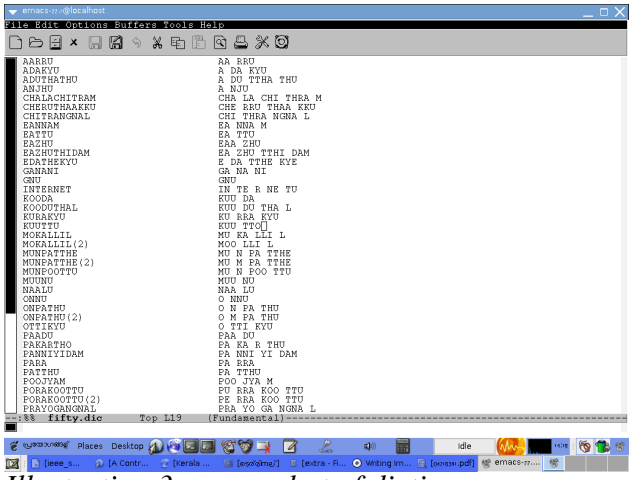

Illustration 3: screenshot of dictionary

As the user completes audio recording ,the recorded sounds are processed to get the feature vectors extracted.Mel Frequency Cepstral Coefficients(MFCC) are by far the favorite feature vector for ASR systems.Here we are trying to model the response of ear and have many narrow filter banks below 1000Hz and gradually wider bands above;using mel scale to define the width.Thus a frequency wrapped version of cepstrum is used.Feature vector sequence is now used to train acoustic models.

2.b)Building the language model

Being a project with small word count, I used Finite State Grammer as the language model and an fsg file is handwritten for the purpose.A screenshot of which is given in the illustration below.

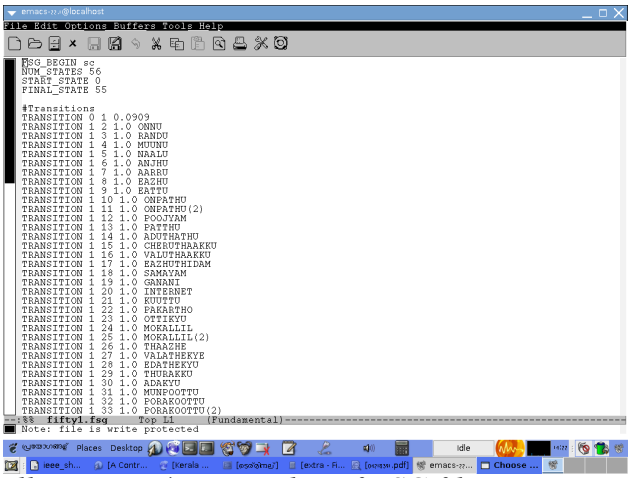

Illustration 4: screenshot of FSG file

These models are given to sphinx decoder to do the decoding.Sphinx decoder APIs are used along with the libraries related to desktop environment to perform live decoding. The trained models being user dependent,also as the no: of words are low, is having very high performance and is working well even for male voices similar to that of mine.

CONCLUSION AND FUTURE WORK

ASR systems are very complex and is a very time consuming project.Members of Swathanthra Malayalam Computing are now doing the ground works of building a Free Text Corpora for malayalam which will be in a sharable license.Building a speech corpora is even bigger activity and requiring more time and effort.But once built,those models can be adapted and used for enormous number of applications.As already stated, Speech recognition programs are having varities of applications and will become a revolutionary breakthrough when we consider the low percentage of literacy of our country.A speech-to-speech application which interactes with the user by accepting speech and giving out speech could be one of the best use of speech recognition systems,where the user gets the maximum convinence.

With world-class Free Software ASR systems like Sphinx which is built and through contribution from people all over the world,the task of building speech recognition programs are becoming more and more simpler.Still the bigger task remain for us - the making of a user independent acoustical and language model for malayalam,

ACKNOWLEDGEMENT

I would like to acknowledge the continuous support and help from my parents and friends and the members of SMC especially Santhosh Thottingal for supporting me through the development of the project.I would also like to say abt the help of Dr.Sathidevi of NITC during the initial stages of project.

REFERENCE

1.L.Rabiner., A Tutorial on Hidden Markov models and
Selected Applications in Speech Recognition , Proc. Of
IEEE, Vol. 77 No. 2, 1989.
2."Fundamentals of Speech Recognition" by Lawrence Rabiner and Biing-Hwang Juang
3."Speech And Audio Signal processing,processing and perception of speech and music" by Ben Gold and Nelson Morgan
4."Digital processing of speech signals" by L.R.Rabiner,R.W.Schafer
5."Speech Signal processing Principles and Practice" by Thomas.F.Quatieri
6.keralapaniniyam and common malayalam grammar books for the detailed reference of malayalam language.
7.cmusphinx.org which is the central link to the complete Sphinx project at CMU.
8.sphinx.subwiki.com is a newly created wiki.