Real time speech recognition for regional languages
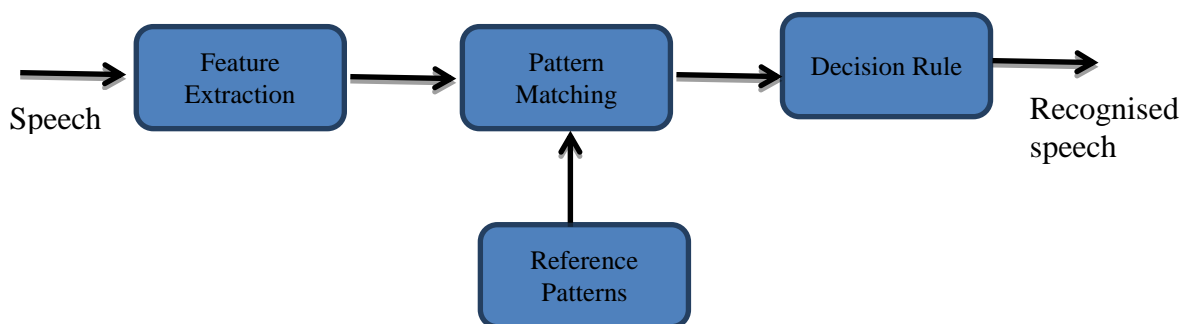
Problem statement

The Speech is most essential and primary mode of communication among all human being. Human beings have long been motivated to create computer system that can understand and talk like humans. Interactions with machine through speech technology is easier instead of other input devices like mouse and keyboard etc. The process of automatic recognition of speech by machines is known as Automatic Speech recognition.The speech is a signal of infinite information. There are different aspects related to speech like speech recognition, speech verification, speech synthesis, speaker recognition, speaker identification etc.The goal of speech recognition is to develop a system which takes spoken words as an input.

Background

Speech recognition has become a practical concept, which is now being implemented in different languages around the world andit is used in real-world human language applications, such as information recovery. The conversation or speech that is captured by a microphone or a telephone is converted from acoustic signal to a set of words in speech recognition.Various approaches and types of speech recognition systems came into existence in last five decades gradually. This evolution has led to a remarkable impact on the development of speech recognition systems for various languages worldwide. Languages, on which so far automatic speech recognition systems have been developed, are just a fraction of total around 7300 existing languages. Russian, Portuguese, Chinese, Vietnamese, Japan, Spanish, Filipino, Arabic, English, Bengali, Tamil, Malayalam, Sinhala, Hindi are prominent among them. English is the language for which maximum work for recognition is done.The system uses Mel-Frequency Cepstral Coefficients (MFCC) and Hidden Markov Model(HMM) for pattern training and feature extraction, respectively. The success of MFCC combined with their robust and cost-effective computation, turned them into a standard choice in speech recognition applications. And HMM provides a highly reliable way of recognizing speech. It concludes with the decision on feature direction for developing technique in human computer interface system using regional languages.

Methodology

The Automatic Speech Recognition (ASR) field of research is near about 60 years old. Many researchers have tried for advanced research and have successfully developed the particularsystems.The main objective of ASR systems is automatically obtaining the output string from input speech signals. Hidden Markov model (HMM) i.e., statistical model is used in ASR.The ASRis the system being modelled and assumed to be a Markov process with unknown parameters. The challenge is to determine the hidden parameters from the observable data.

Speech → Feature Extraction → Pattern Matching → Decision Rule → Recognised speech

Reference Patterns → Pattern Matching

Experimental design

Dataset
Initially, a standard Clemson University Audio Visual Experiments (CUAVE) dataset will used for experimentation and evaluation.Dataset will be the record of speeches of 1000 or more people in particular regional language.

Evaluation measure
Measures such as Accuracy, precision, and recall will be computed.
Software and hardware requirements
Python based Deep Learning libraries will be exploited for the development and experimentation of the project. Training of the data will be conducted on NVIDIA GPUs.