SPOKEN LANGUAGE IDENTIFICATION

Guide
Dr. K. G. Srinivasa

Adarsh Patil
Akshay V Joshi
Harsh K C
Pramod N

What is a Spoken Language Identification ?

- *Spoken Language Identification is the process of recognition of the language spoken in an utterance.*

Problem Statement

- Build an optimum corpora
- Identifying features for an acoustic model
- Employ learning techniques for multi-class classification
- Identify language

What are the characteristics of an ideal Language Identification System ?

Characteristics

- Accurate language identification
- Unaffected by noise
- Speaker independent:
  - Gender
  - Accent
  - Pronunciations

Approaches to Spoken Language Identification system

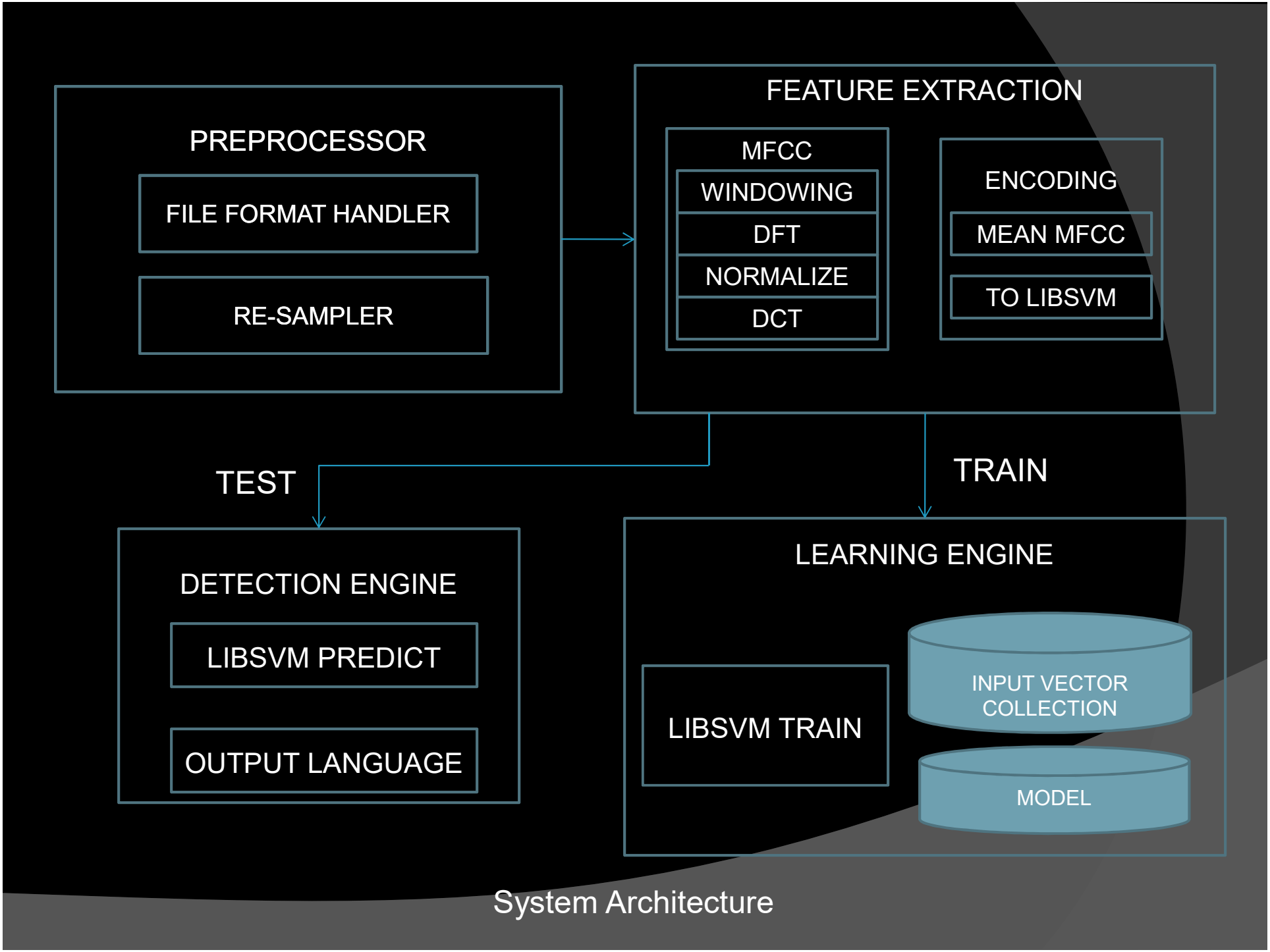
Approaches

- Prosodic
  - Rhythm, stress, and intonation of speech
  - Syllable length, loudness, pitch, and the formant frequencies of speech sounds.
- Phonotactic
  - Analysis at the phoneme or syllable level
  - Rules that govern permissible sequence of phonemes in speech signals.
- Acoustic
  - The power spectrum of a signal is indicative of acoustic information in speech.
  - Speech signal characteristics

The Corpora

- Unavailability of standard dataset
- Custom Corpora : Derived from podcasts and online audio books
- Uniform duration of 10 seconds
- Semi spontaneous and casual human adapted interaction

System Architecture



System Architecture

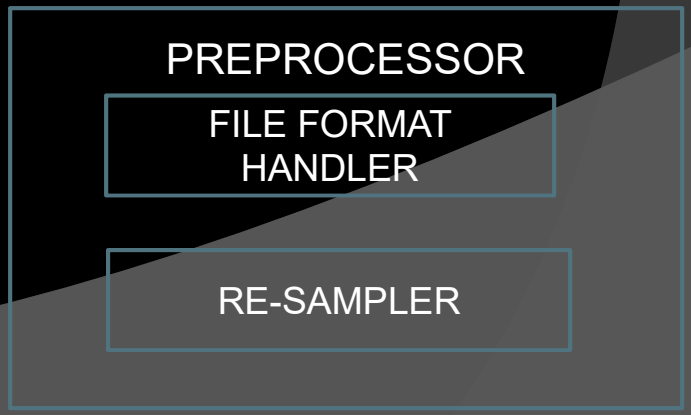
Pre-processing

- File format handling

  Converting all the input samples to WAV format

- Re-sampling

  re-sampling all the input sample to 44.1KHz

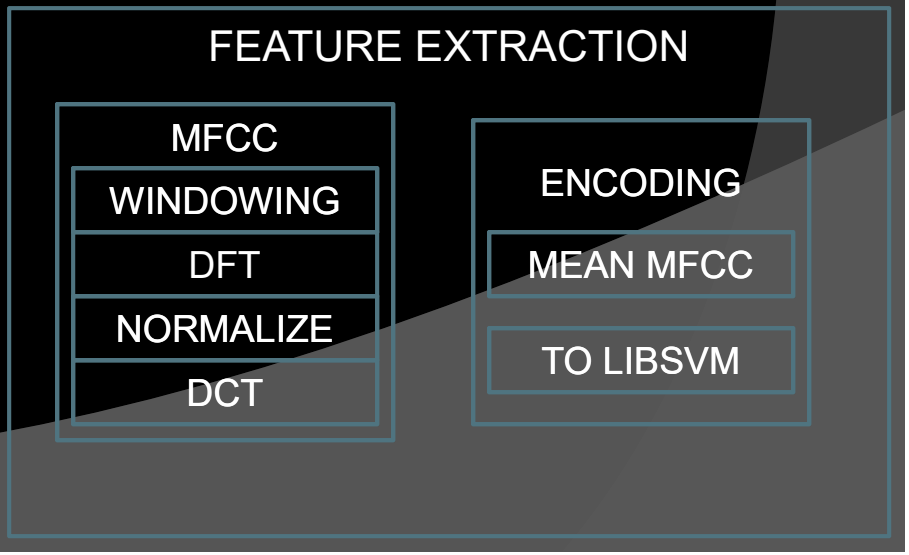


Acoustic model - MFCC

Mel Frequency Cepstral Coefficient

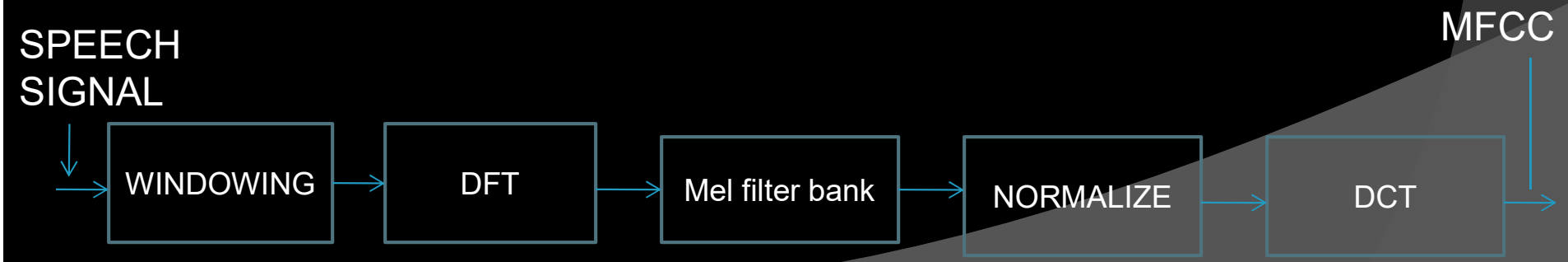
- A representation of the short term power spectrum of a sound

[-12.4518 , 2.30226 , 0.862566 , 0.585914 , 0.387642 , 0.591962 , -0.0487 , 0.206065 , 0.163443 , 0.254867 , 0.031689 , 0.042887 , 0.115736 , 0.025846 , 0.136068 , -0.02038 , 0.04294 , 0.082779 , 0.01792 , 0.025043]



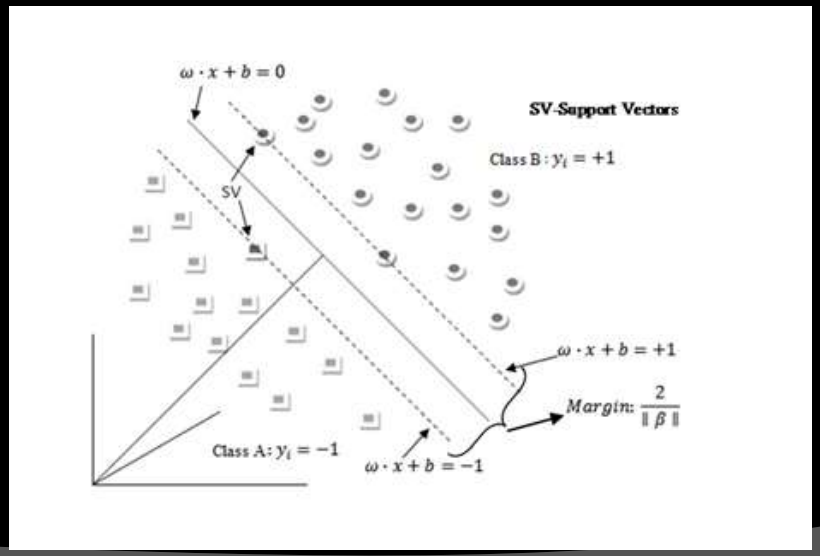
MFCC Computation

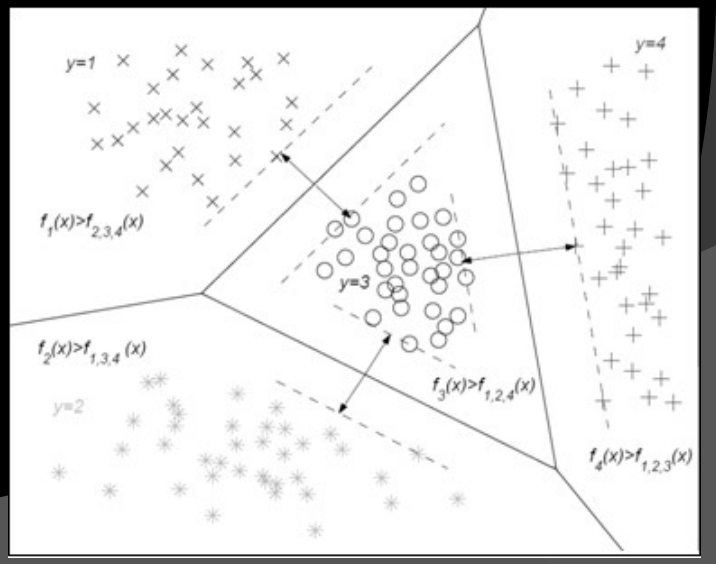
- Take the Fourier transform of a windowed excerpt of a signal
- Map the powers of the spectrum obtained above onto the Mel scale
- Normalize the powers at each of the Mel frequencies
- Take the discrete cosine transform of the list of Mel log powers
- The MFCCs are the amplitudes of the resulting spectrum



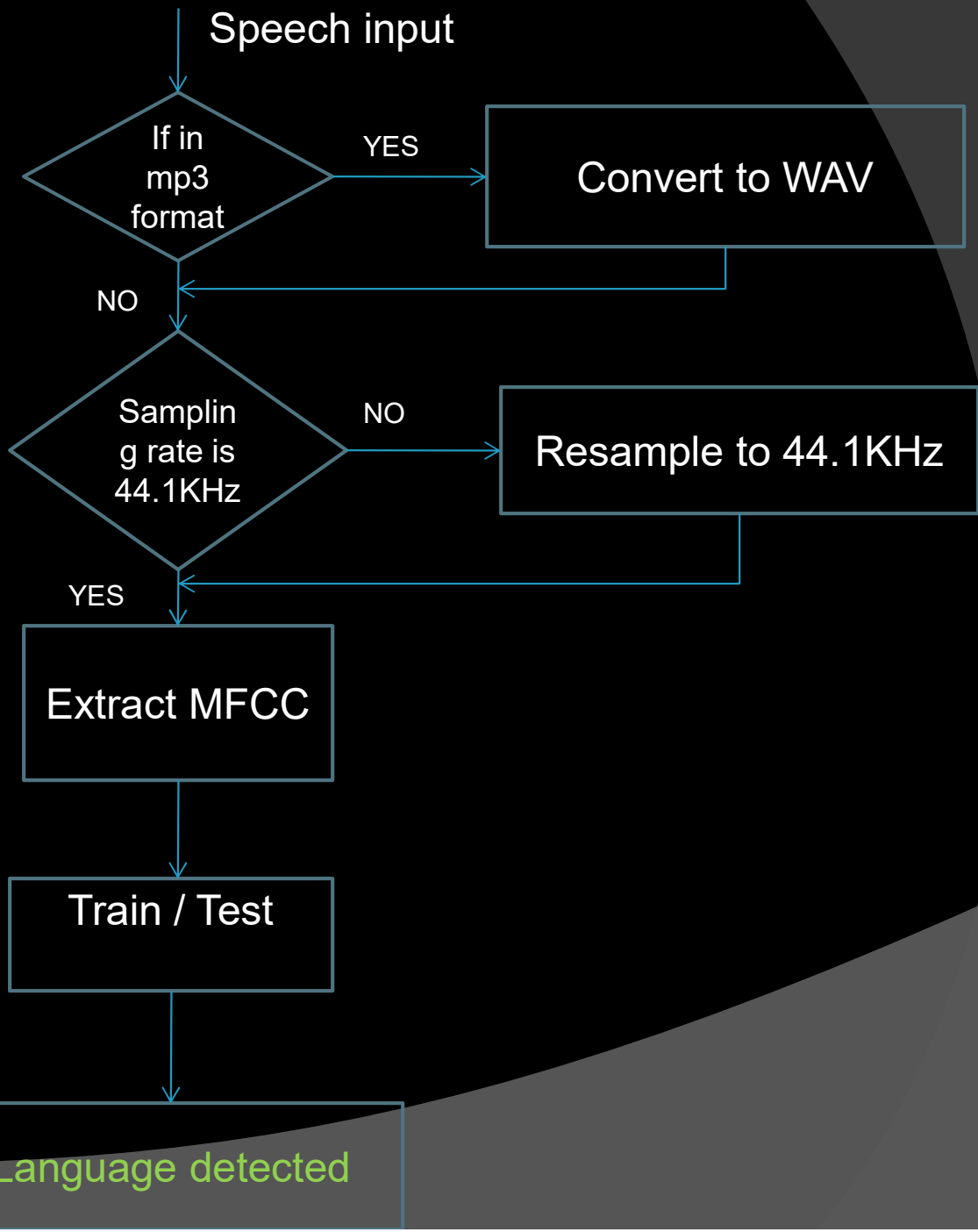
Learning Engine

- Support Vector Machines:
  - Uni-model multiclass classifier
  - Equation of the hyperplane: $\omega.X + \beta = 0$





FLOW CHART


Speech input
If in mp3 format
YES
Convert to WAV
NO
Samplin g rate is 44.1KHz
NO
Resample to 44.1KHz
YES
Extract MFCC
Train / Test
Language detected

Performance Analysis

Trivial Experiment

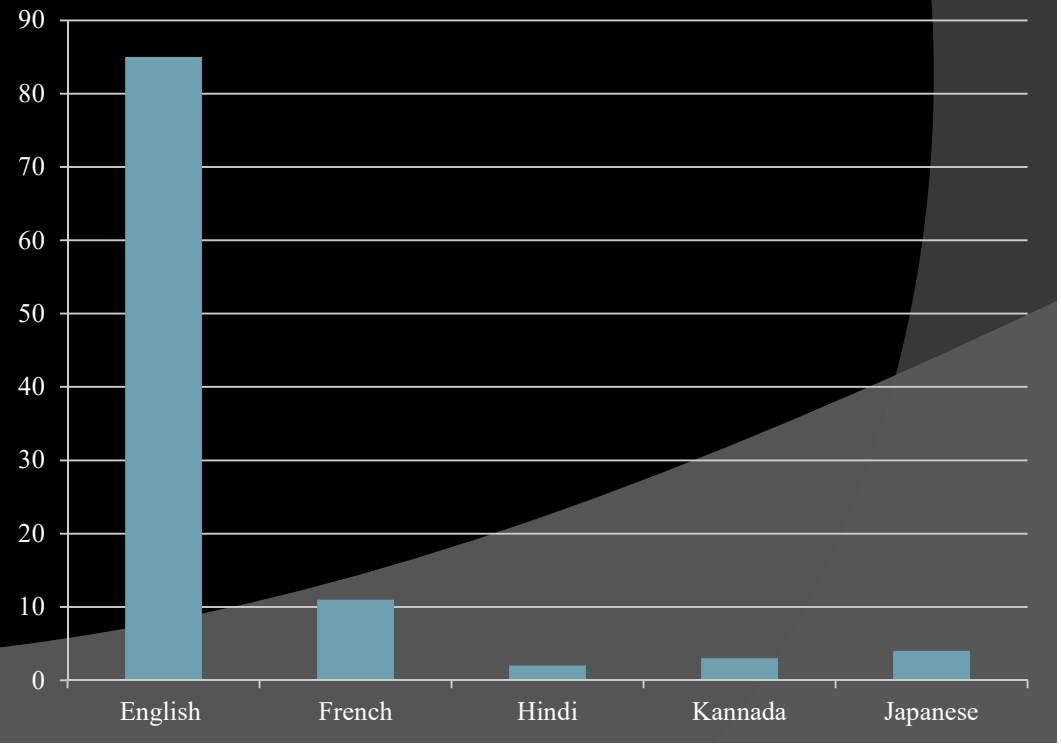
- The entire training set is used for testing

- Number of samples:
  - English: 1093
  - French : 1069
  - Hindi : 853
  - Japanese : 539
  - Kannada : 868

	Eng	Fr	Hin	Kan	Jap
Eng	98.558	0.0108	0.0027	0	0
Fr	0.0935	97.0065	0	0.0935	2.8
Hin	7.735	0.351	91.79	0.1172	0
Kan	2.9935	0.4608	0.1152	96.42	0
Jap	0	1.29	0	0.371	98.3302

Performance Analysis

- Test with open source multi lingual speech corpus – VoxForge
- Number of English Speech samples : 105
- Samples classified as :
  English : 85
  French : 11
  Hindi : 2
  Kannada : 3
  Japanese : 4
- **Accuracy : 80.95 %**



Challenges

- Standard multi lingual speech corpus
- Identifying features
- Differentiating accents and dialects
- Eliminating noise
- Scaling system to high throughput classification

Applications

- In Spoken language translators
- In tourist information systems (Airports)
- In customer support switchboard (call routing)
- Military applications (spy applications)

Future Enhancement

- Introduce incremental learning (feedback)
- Increase the number of languages supported
- Make the corpora robust
- Enhance the feature space
- A real time mobile App

Thank you