

# Automatic Spoken Language Identification:

## Introduction.

1. Automatic language identification means identification of language which is spoken by the speaker.

2. spoken language recognition is far more challenging than text-based language.

3

a. text based. (solved problem)

b. speech signal based (active area of research)

4. spoken LID systems:----

a. Implicit System

b. Explicit System.

**5. Explicit System.:** This system needs different language recogniser to identify different languages. The system cannot recognise those languages whose recogniser are not present in the system

**6. Implicit LID systems:** This system do not need any recogniser. In other words, these systems require only the raw speech data along with the true identity of the language spoken.

**7. The problems faced during language identification:**

- Variation in speaker characteristics
- Variation in accents
- Variation in environment and channel characteristics

8. Language identification task involves three stages namely,

- feature extraction,
- modeling
- evaluation.

9. A system is needed with following features\_

- 
- computationally
- efficient.
- *acoustic-phonetic (same syllable must be differentiated by system)*

language identification is carried out using spectral features such as mel-frequency cepstral coefficients (MFCC).

Gaussian mixture models (GMMs): language specific information from extracted features is captured using this model.

2. maintaining male female ratio to build a balanced system

3. language model must be tested with different set of speaker to ensure the system robustness.

### **Feature Extraction:**

a. Apply the Hamming window divide the speech signal into sequence of frames with a frame size of 20 ms and a shift of 10 ms.

b. Compute magnitude spectrum for each windowed frame by applying DFT.

c. Mel spectrum is computed by passing the DFT signal through mel filter bank.

### **Development of Language Model**

a. GMM for each of the language are needed for language identification.

In the speech and speaker recognition the acoustic events are usually modeled by Gaussian probability density functions (PDFs), described by the mean vector and the covariance

matrix.

$$P(x_t|\Omega) = \sum_{i=1}^M w_i P_i(x_t)$$

$P_i(x_t)$  are the component densities.

where  $w_i$  are the weights and

## **MFCC Features**

### Pre-emphasis:

- preemphasis filter is used for this process. Transfer function of filter is  $H(z) = 1 - bz^{-1}$ .

pre-emphasis removes some of the glottal effects from the vocal

- tract parameters

acoustic energy radiates from the lips, this causes a roughly +6dB/octave boost to the spectrum. so this filter is used.

### Frame blocking and windowing:

- Hanning or Hamming windows are used.
- speech signal analysis must be done for a very short period where speech signal is stationary.

### ***Short-term spectral measurements are***

- typically carried out over 20 ms windows, and advanced every 10 ms.

### DFT spectrum:

- Each windowed frame is converted into magnitude spectrum by applying DFT.

### Mel-spectrum:

Mel-Spectrum is computed by passing the Fourier transformed signal through a set of band-pass filters known as mel-filter bank.

### Discrete Cosine Transform (DCT):

- The DCT is applied to the transformed mel frequency coefficients produces a set

of cepstral coefficients.

- It is represented on a log scale.

Dynamic MFCC features:

- The extra information about the temporal dynamics of signal is computing the first and second derivatives of cepstral coefficients.
- The first coefficient is called delta derivative and the second one is called delta delta derivative.

Performance % of Speaker independent LID per language is also obtained for the above test.

Gender dependent LID study has 3 parts :

1. male speakers are used for language modelling & language test.
2. female speakers are used for language modelling & language test.
3. cross gender LID study- has 2 parts.

a) male speakers for language modelling and female speaker for language test of all languages

b) female speakers are used for language modelling only and male speakers for language test.

For all above gender dependent test performance percentages are obtained .