Feature extraction

INTRODUCTION:

Speech is one of the ancient ways to express ourselves, and today's speech signals are also used in biometric recognition and machine communication techniques.

One of the theoretical aspects is that we can recognize speech directly from the digital wave, and with the great variation in the speech signal, it is better to extract features that reduce this contrast, and we can get rid of sources that produce some different information.

In this chapter, we will talk about Feature extraction, as it includes a lot of topics, and we will use LPC and MFCC and examples based on them.

Feature extraction:

-Feature extraction is the process of obtaining various features such as energy composition, tone and vocal tracts from a speech signal. Parameter conversion is the process of converting these features into signal parameters

-In speaker independent speech recognition, a premium is placed on extracting features that are somewhat invariant to changes in the speaker. So future extraction involves analysis of speech signal. Broadly the feature extraction techniques are classified as temporal analysis and spectral analysis technique. In temporal analysis the speech waveform itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis.

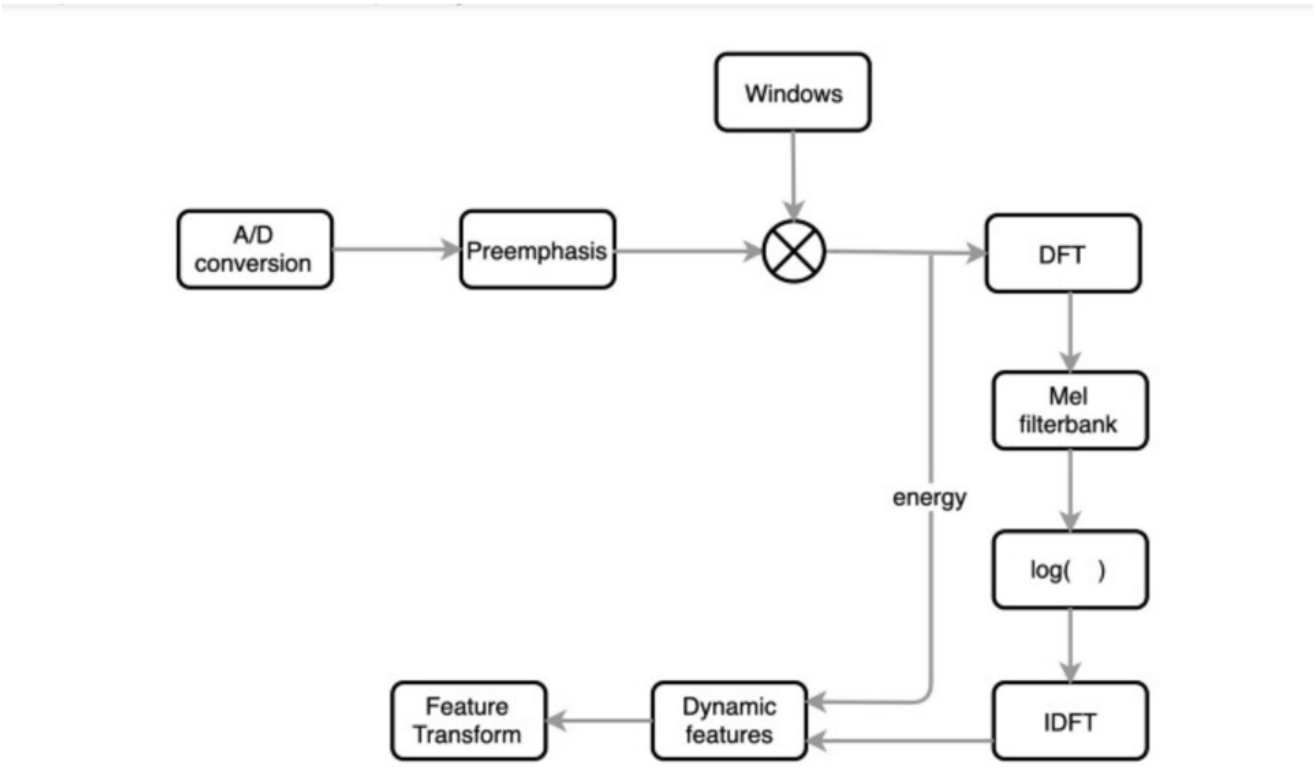
LINEAR PREDICTIVE CODING (LPC) :

LPC is one of the most powerful speech analysis techniques and is a useful method for encoding quality speech at a low bit rate. The basic idea behind linear predictive analysis is that a specific speech sample at the current time can be approximated as a linear combination of past speech samples.

1.Mel Frequency Cepstral Coefficients MFCC

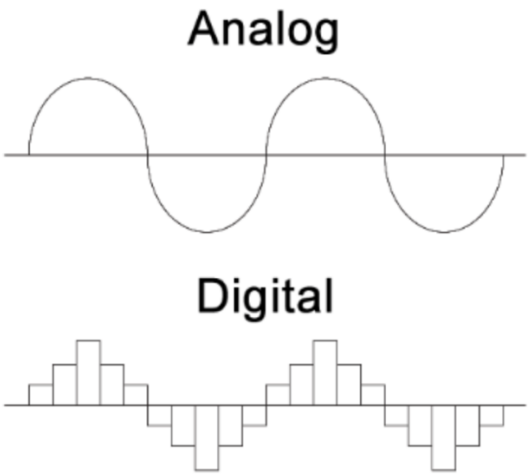
-The use of Mel Frequency Cepstral Coefficients can be considered as one of the standard method for feature extraction.

Extraction Mel-frequency cepstral coefficients (MFCC) from the audio recording signals.



1.1 A/D Conversion:

In this step, we will convert our audio signal from analog to digital format with a sampling frequency of 8kHz or 16kHz.



1.2 Preemphasis:

Preemphasis increases the magnitude of energy in the higher frequency. When we look at the frequency domain of the audio signal for the voiced segments like vowels, it is observed that the energy at a higher frequency is much lesser than the energy in lower frequencies. Boosting the energy in higher frequencies will improve the phone detection accuracy thereby improving the performance of the model.

1.3 Windowing:

The MFCC technique aims to develop the features from the audio signal which can be used for detecting the phones in the speech. But in the given audio signal there will be many phones, so we will break the audio signal into different segments with each segment having 25ms width and with the signal at 10ms apart

1.4 DFT (Discrete Fourier Transform):

Convert the signal from a field to a field (dft) for engineering signals, and the analysis is easier.

1.5 Mel-Filter Bank:

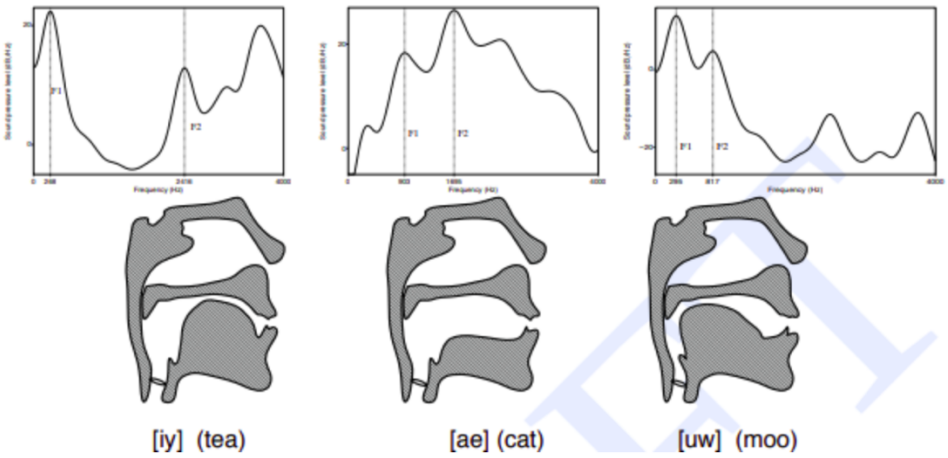
The way our ears will perceive the sound is different from how the machines will perceive the sound. Our ears have higher resolution at a lower frequency than at a higher frequency.

1.6 Applying Log:

Humans are less sensitive to change in audio signal energy at higher energy compared to lower energy. Log function also has a similar property, at a low value of input x gradient of log function will be higher but at high value of input gradient value is less. So we apply log to the output of Mel-filter to mimic the human hearing system.

1.7 IDFT:

Here he performs the inverse conversion of the output from the step before it, and we have to understand how the sound is produced by humans



Dynamic Features:

It computes derivatives by coefficients among audio signal samples and helps understand the occurrence of the transition.