

Automatic Speech Recognition

M.Tech. Artificial Intelligence, Second Year, NMIMS

By,

Mayur Borkhatariya, Sr. Data Scientist, Kotak Securities

Audio Signal: (Automatic Speech Recognition)

An audio signal is a type of longitudinal vibration that occurs in the air.
When an event creates a disturbance or vibration in the air, like when you pluck a guitar string

Longitudinal vibration that produces vitality

Audio Signals are composed of Sound Waves

Sound Wave:

Vibration signal produces by moving energy

Parameters
↳ Amplitude

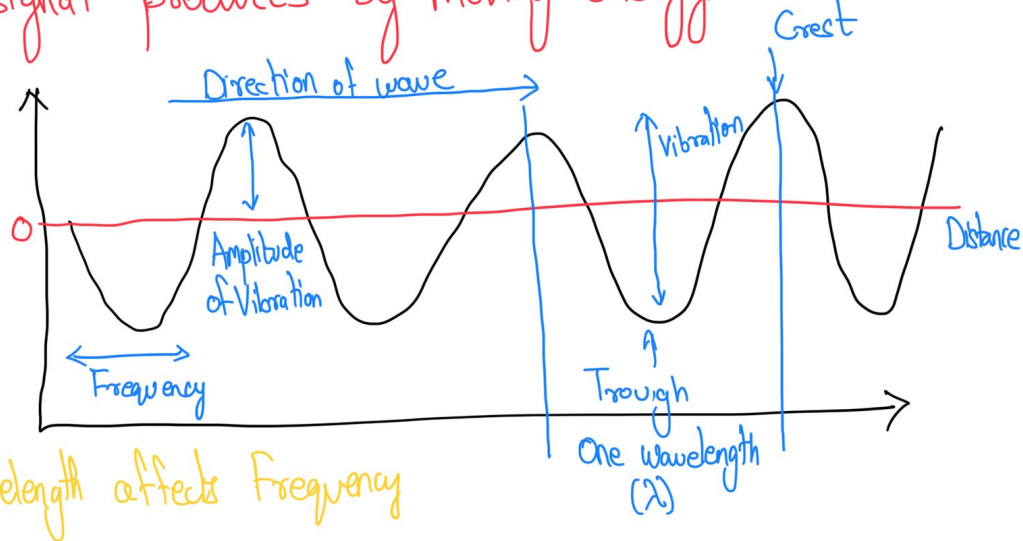
Crest and Trough

Wavelength

Cycle

Frequency

Parameters of
Audio Signals

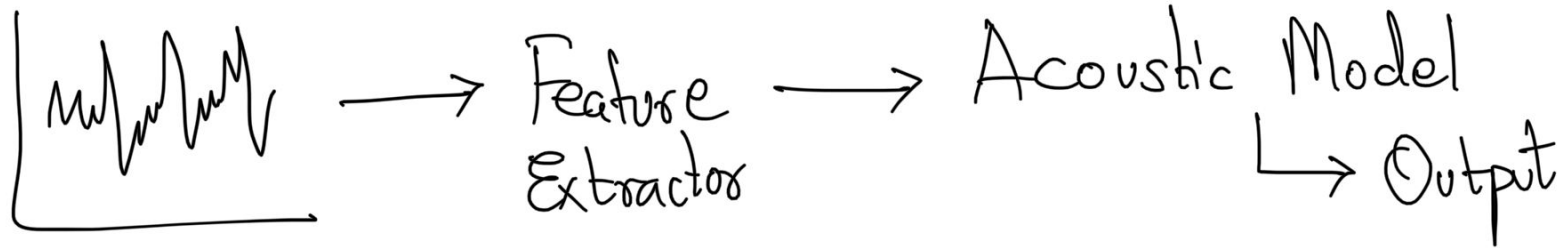
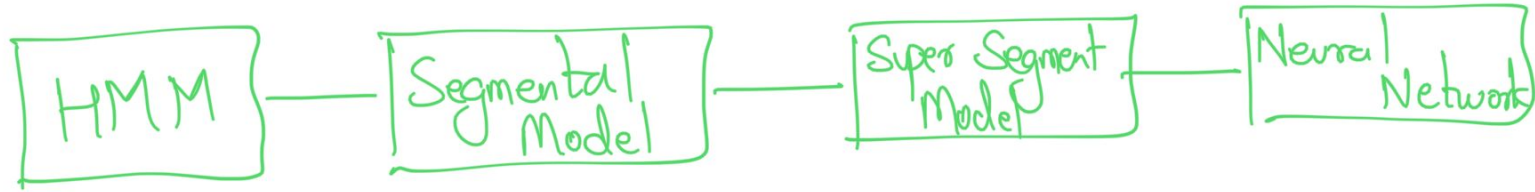


Acoustic Modelling

Identifies what is spoken based on Features

→ Statistical Representation of computed feature vectors

Different Acoustic Models



Language Modelling

Probability of Word Occur

- Performing pre-processing from the speech text
- Natural language processing
- Converting text into vectors (Word2Vec, TF-idf ...)
- Challenges: Phrases with different tones, Ambiguities, picking out correct words
- Ambiguities can be resolved by combining language pronunciation and acoustic model

Zero Crossing Rate:

Shift rate at which signal changes from positive to negative

Spectral Centroid & Roll off

→ Spectral Centroid is weighted mean average

$$\text{Centroid} = \frac{f(n) w(n)}{w(n)}$$

→ Spectral Roll off is weighted moving average

MFCC (Mel-frequency Cepstral Coefficients)

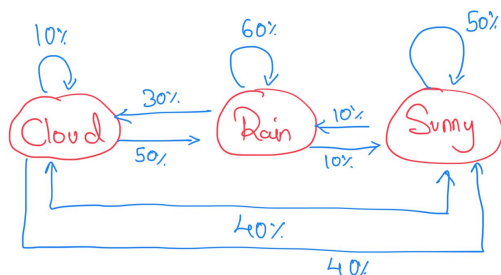
Mel Spectrogram

→ Spectrogram converted to Mel scale

- Widely used in deep learning
- Powerful tool to extract the feature from speech
- Process includes: Fourier Transform, discrete cosine transforms and overlapping windows
- It helps for classification problems such as genre classification, disease detection related to speech and etc.

Hidden Markov Model (HMM)
in speech processing

Example:



State Transition

	Cloud	Rain	Sunny
Cloud	10%	50%	40%
Rain	30%	60%	10%
Sunny	40%	10%	50%

} 100%

Viterbi Algorithm

Problem : Given today is Monday and it is sunny
What is the probability that Wednesday would be cloudy

	c_1	c_2	c_3
M	S	S	S
T	S ^{0.5}	R ^{0.1}	C ^{0.4}
W	C ^{0.4}	C ^{0.3}	C ^{0.1}

$$P(c_1) = 0.4 \times 0.5 = 0.2$$

$$P(c_2) = 0.1 \times 0.3 = 0.03$$

$$P(c_3) = 0.4 \times 0.1 = 0.04$$

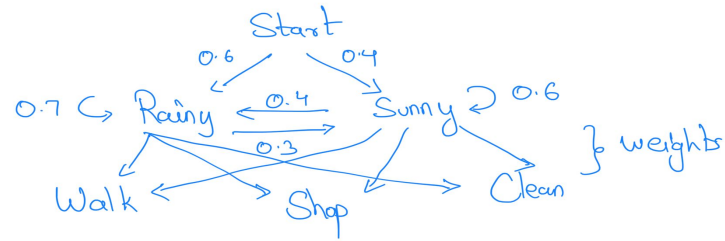
$$P(W \rightarrow C) = \frac{0.27}{0.27}$$

In Markov Chain

Initial Probability distribution \rightarrow Few states \rightarrow Transition Probability

In hidden Markov model

Initial Probability distribution \rightarrow Hidden States \rightarrow Transition Probability
Emission Probability \leftarrow Sequence of Observations \leftarrow



Application

\rightarrow Speech Recognition : To predict what my next word is?

\rightarrow Retail ; Travel ; Medical ; Marketing
(RNA-Seq)
gene regulation

References

https://www.cse.iitb.ac.in/~nirav06/i/HMM_Report.pdf

https://en.wikipedia.org/wiki/Spectral_centroid

<https://vitalflux.com/hidden-markov-models-concepts-explained-with-examples/>

<https://wiki.aalto.fi/display/ITSP/Zero-crossing+rate>

<https://hmmlearn.readthedocs.io/en/latest/index.html>

<https://www.youtube.com/watch?v=1-ldEjzEkYE>