Automatic Speech Recognition

M.Tech. Artificial Intelligence, Second Year, NMIMS

By,

Bilal Hungund, Data Scientist, Halliburton

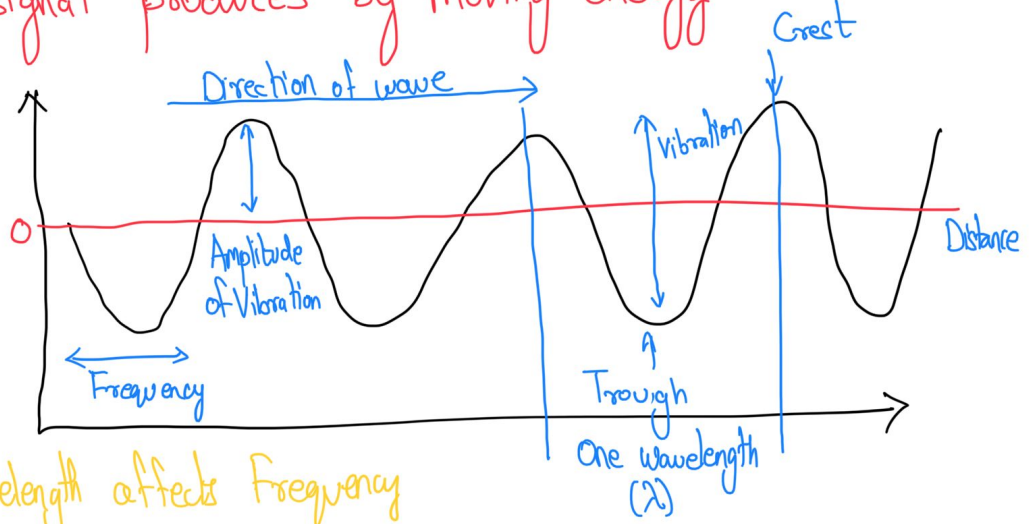
Audio Signal: (Automatic Speech Recognition)

Longitudinal vibration that produces vitality

Sound Wave:

Vibration signal produces by moving energy

Parameters
- Amplitude
- Crest and Trough
- Wavelength
- Cycle
- Frequency

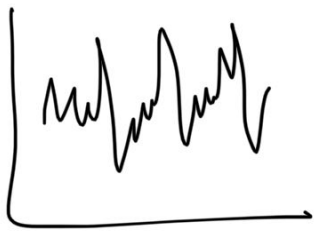


Wavelength affects Frequency

Acoustic Modelling

→ Statistical Representation of computed feature vector

HMM — Segmental Model — Super Segment Model — Neural Network

 → Feature Extractor → Acoustic Model → Output

Language Modelling

- 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 Ceptral Coefficients)

Mel Spectrogram

→ Spectrogram converted to Mel scale

→ Widely used in deep learning

→ Powerful tool to extract the feature from speech

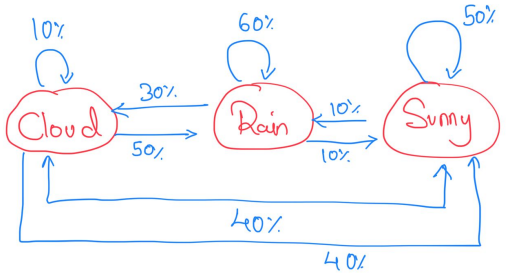
→ Process includes: Fourier Transform, discrete cosine transforms and overlapping windows

→ 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

	C1	C2	C3
M	S	S	S
T	$S^{0.5}$	$R^{0.1}$	$C^{0.4}$
W	$C^{0.4}$	$C^{0.3}$	$C^{0.1}$

$P(C1) = 0.4 \times 0.5 = 0.2$

$P(C2) = 0.1 \times 0.3 = 0.03$

$P(C3) = 0.4 \times 0.1 = 0.04$

$P(W \rightarrow C) = 0.27$

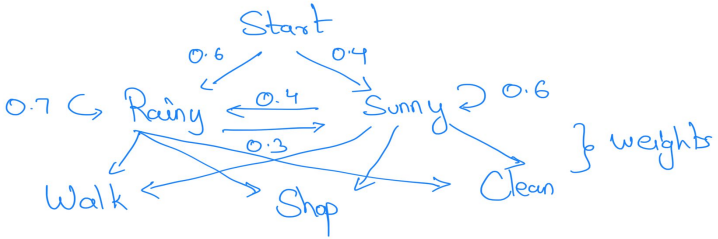
In Markov Chain

Initial Probability distribution → Few states → Transition Probability

In hidden Markov model

Inital Probability distribution → Hidden States → Transition Probability

Emission Probability ← Sequence of Observations



Application

→ Speech Recognition : To predict what my next word is?

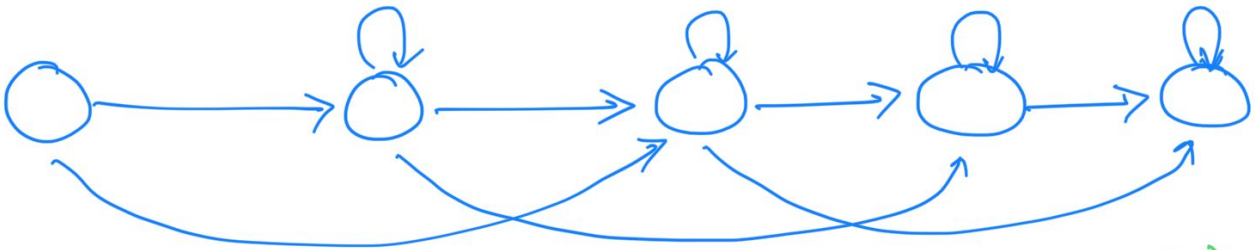
→ Retail ; Travel ; Medical (RNA-Seq) gene regulation ; Marketing

How HMM used in Speech Processing ?

Model Topologies ⟶ Fully Connected HMM

↓

Left - to - Right Transition



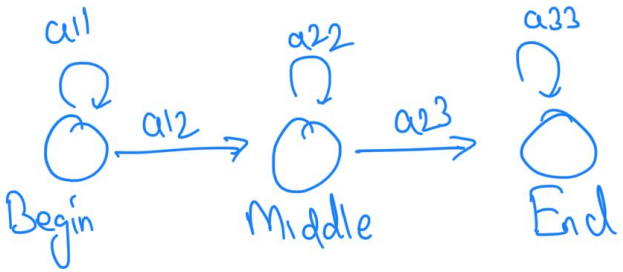
(No - backward possible)

Feature Extraction → Feature → Acoustic Model → Sequence of Phoneme → Dictionary → Language Construct

CAT – K A T
ONE - WAHN
TWO - TOO
→ Word-Phoneme Pair

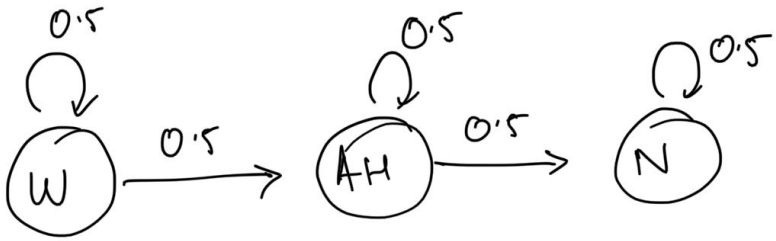
Language Construct → Recognised word

HMM for each word in vocabulary

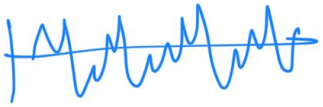


Car key
Khakee
} grammar Construct

→ Context Independent Phoneme HMM
→ Context Dependent Triphone HMM
→ Whole word HMM



Transcript One Four Two One



Lexicon

One	WAHN
Two	TOO
Four	FOAR

Feature Extraction
MFCC

Phoneme Sequences

WAHN FOAR TOO WAHN

Feature Vector

HMM

W → AH → N → F → OA → R ... → N

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