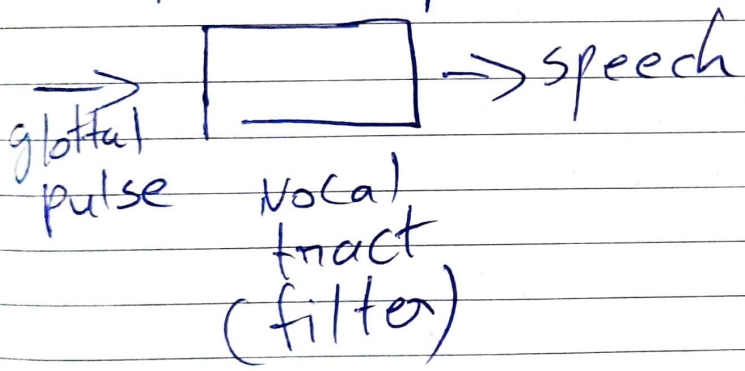
ASR:- (Automatic speech recognition)

⇒ The goal of ASR is to go through a audio file and extract text out of it.

⇒ To detect this we need to understand how speech is produced.



⇒ As speech is a non-stationary signal we can extract using cepstral analysis.

⇒ segment of speech → preemphasized signal → windowed signal → DFT → Mel spec ← log mel spec ← cepstrum

=

PRIORITY

APPOINTMENTS

IMPORTANT NOTES

The problem with cepstrum is that we can get complex numbers out of it.

so we use MFCC Features to get the low frequency region we want

It uses discrete cosine transform at the end of analysis instead of inverse fourier transform

13 -> dimention MFCC can capture most of the timbre.

Δ MFCC -> 13 dimention
ΔΔ MFCC -> 13 dimention

Total MFCC -> 39 dimention

Hidden Markov model:-

-> Markov property satisfaction:-

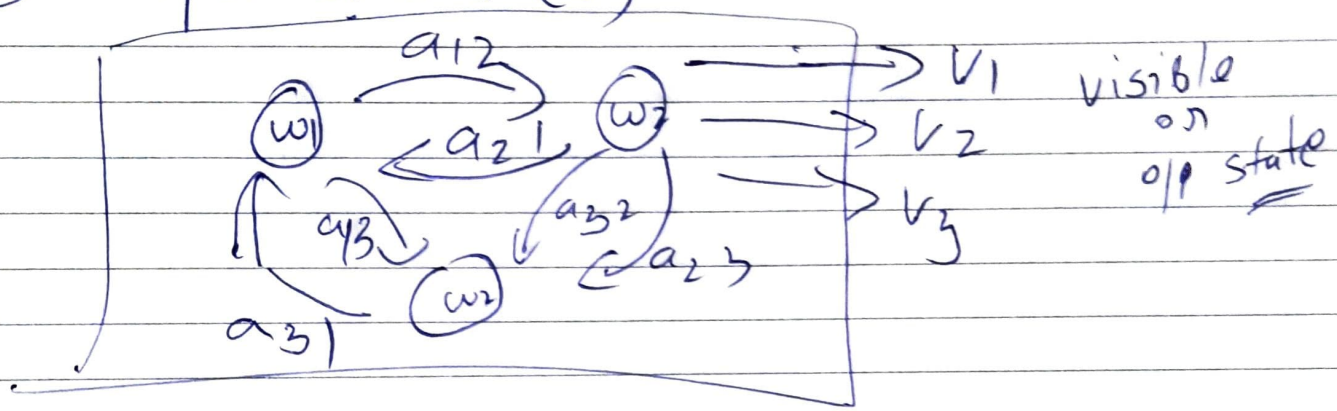
-> It is memory less
-> It takes only one state in consideration.

-> future prediction relies on present

State.

→ It is like finite state machine.

(1) Hidden states:- (w)
(2) output states (v)



Input → d

Phoneimes → /t/ /θ/ /d/

↳ In my model.

so it most closly
relateds to /d/

then it cheaks by ~~sim~~ using this state
what will be the next probable sound
It is most likely going to be a vowel
like dog, disk, dank etc.