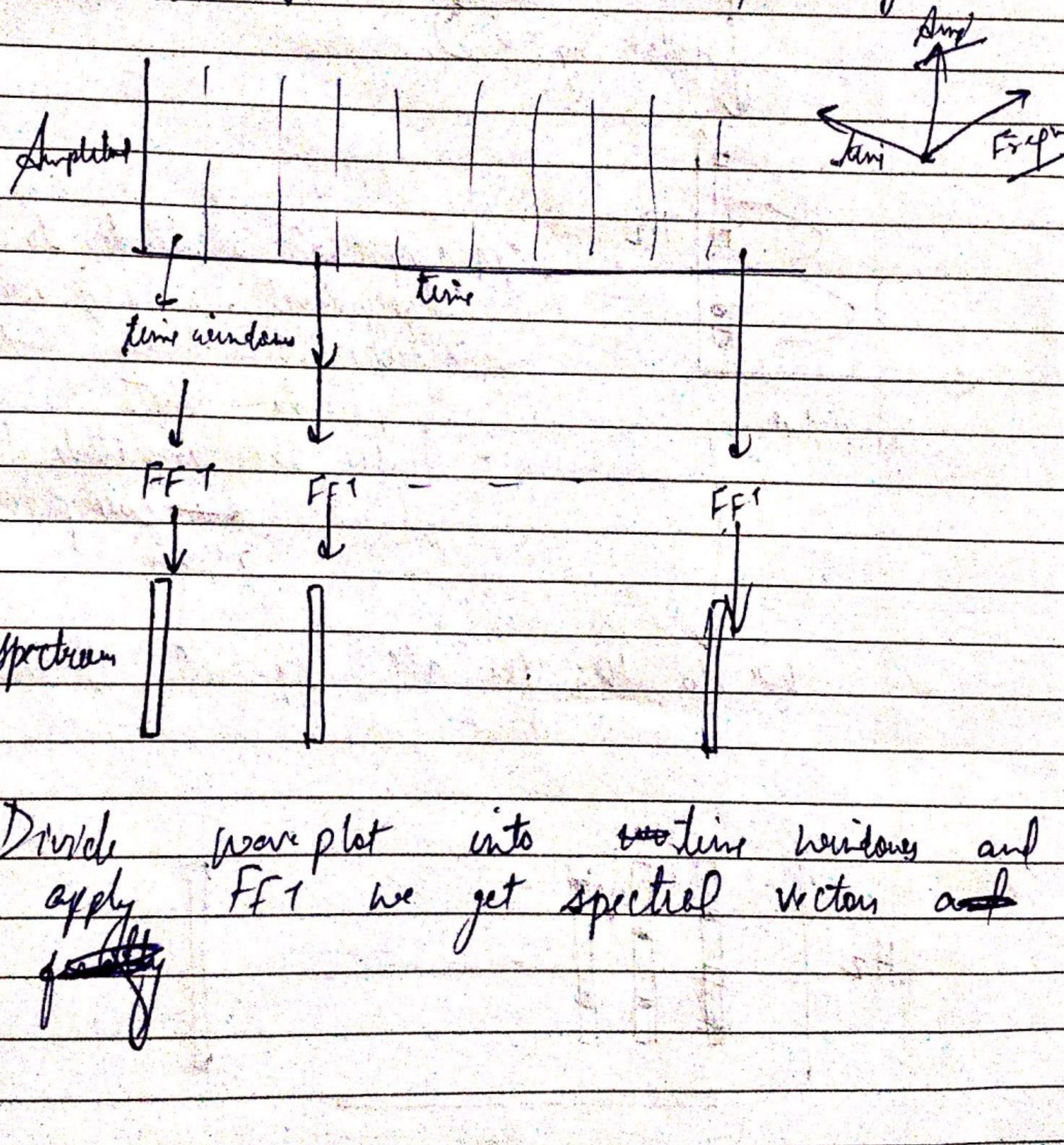
Date Page No.

Topics

Spectrogram
Cepstrum
Mel-Frequency Analysis
Mel-Frequency cepstral coefficients

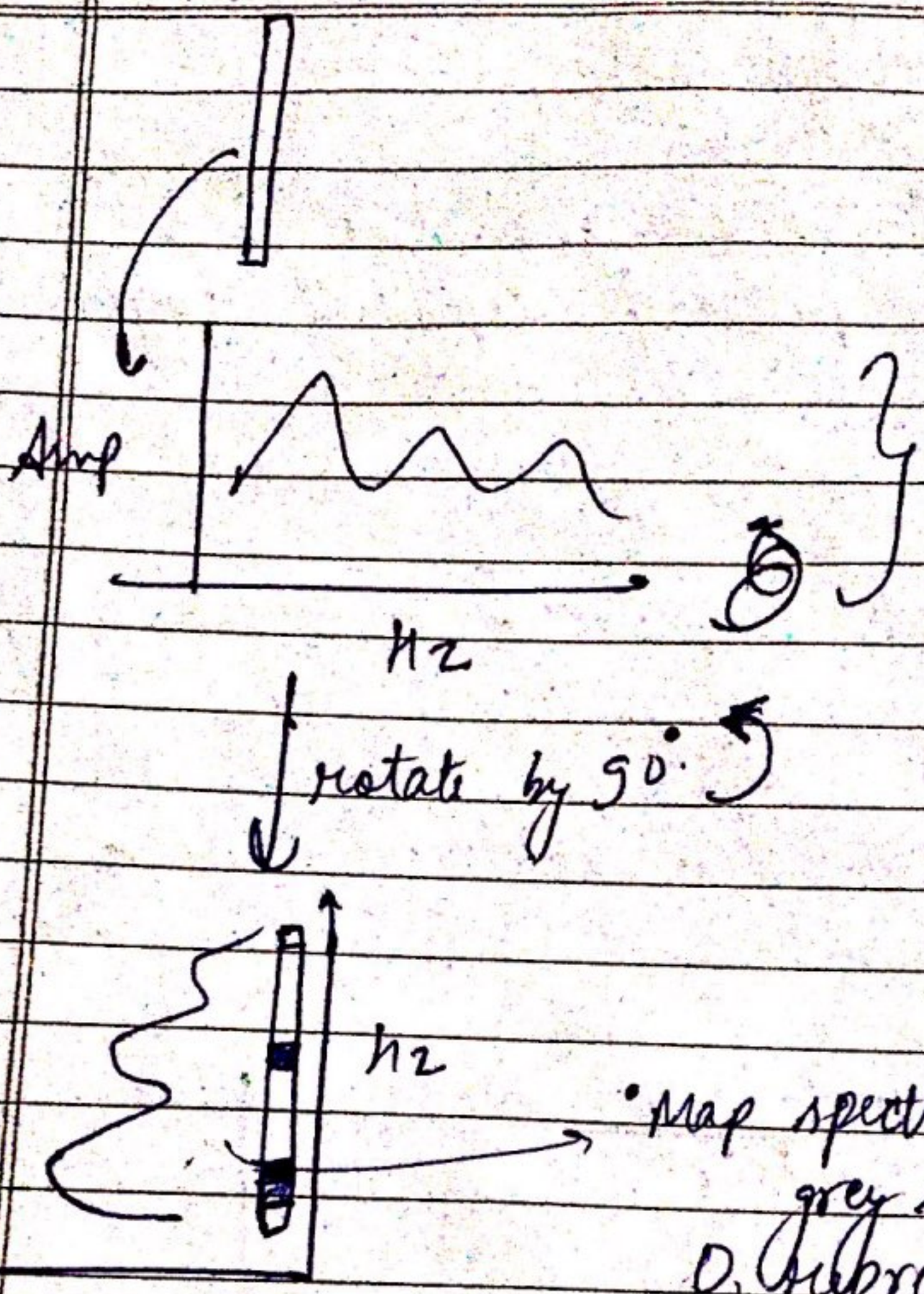
Spectrogram

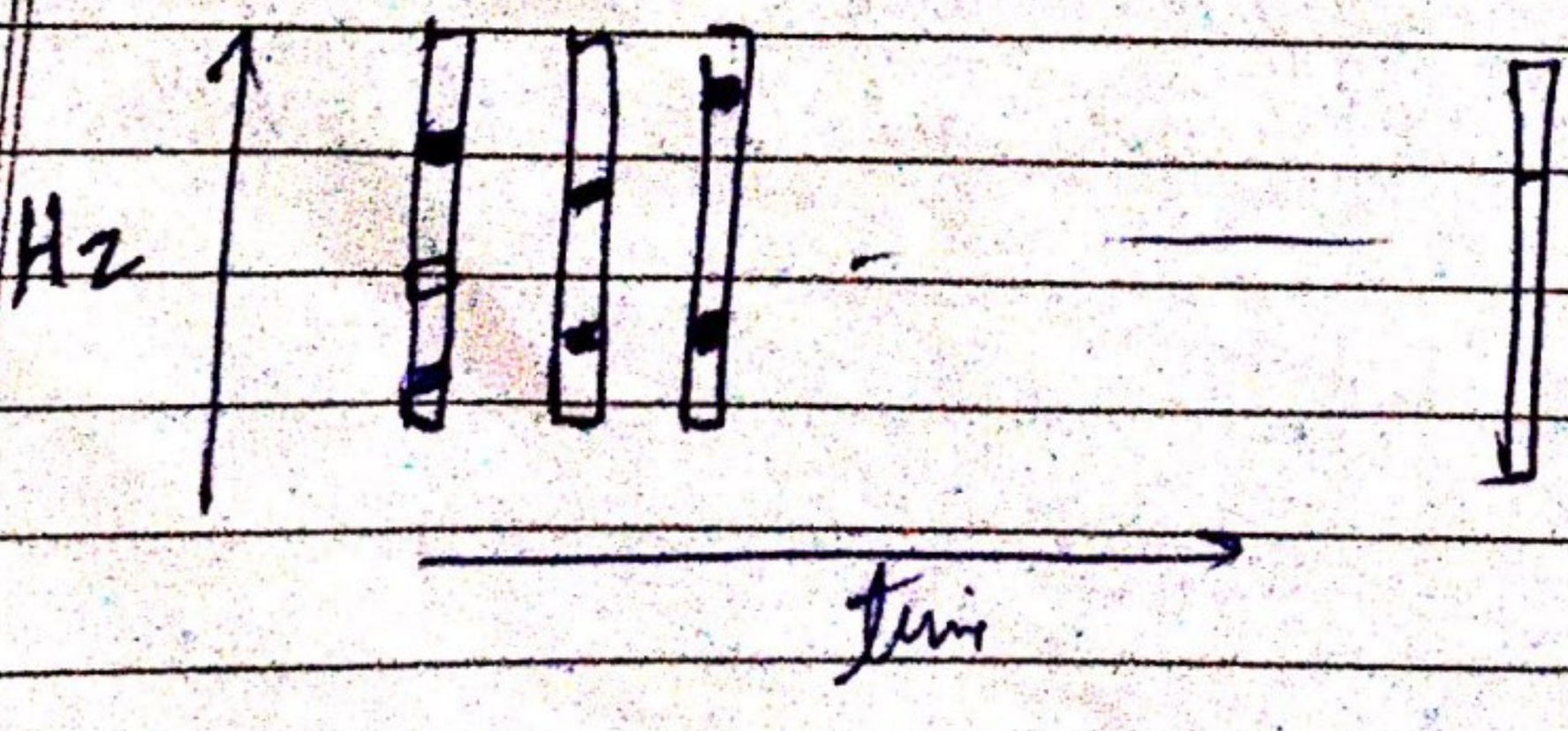
Speech signal represented as sequence of spectral vectors



⇒ Divide wave plot into ~~two~~ time windows and apply FFT we get spectral vectors ~~of~~ of
~~probability~~

Date / / Page No.





(Time vs Frequency representation of speech signal is referred to as spectrogram)

WHY SPECTROGRAMS

(speech sounds)

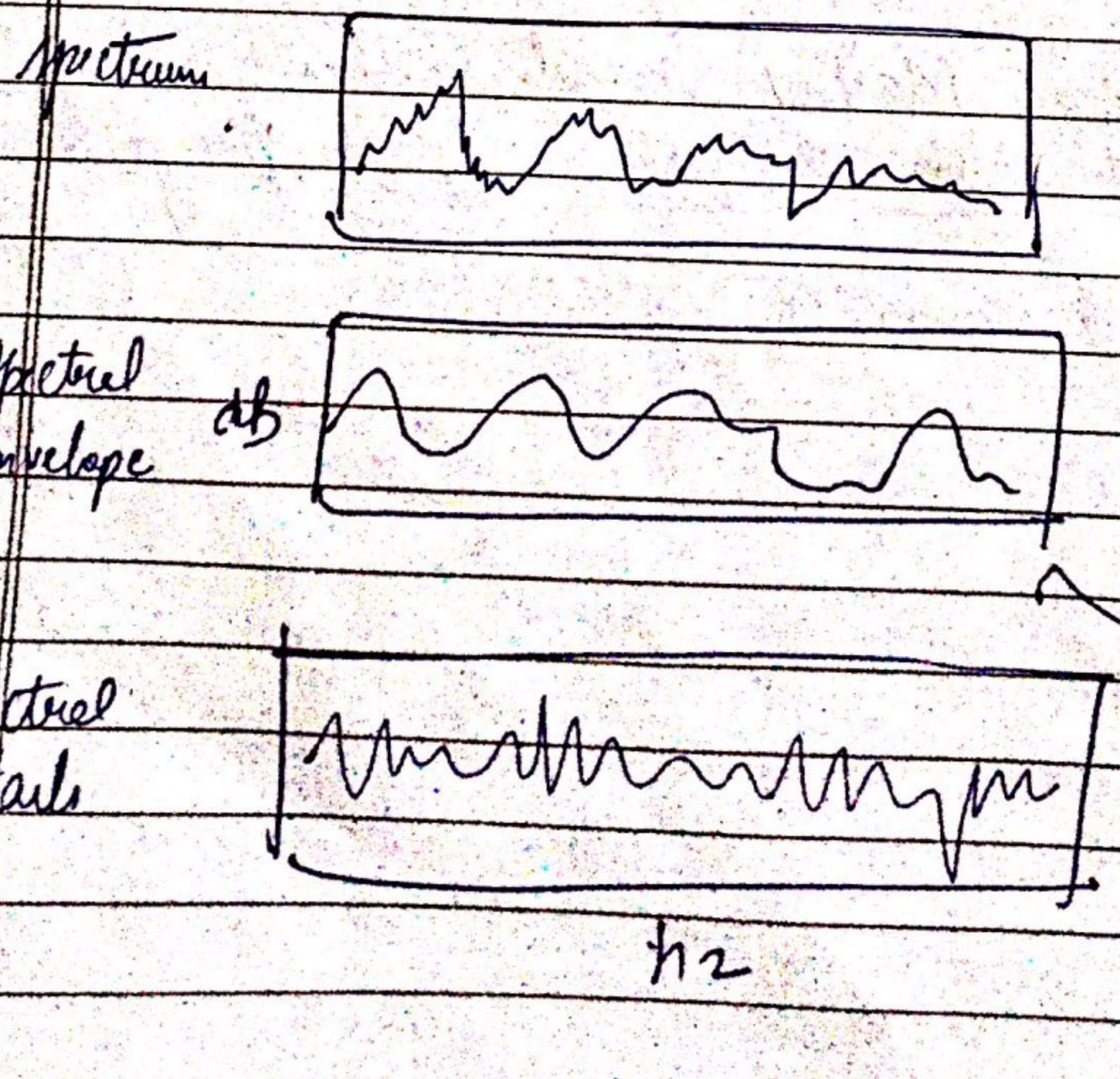
- Phones and their properties are better observed in spectrograms
- Easily detecting phone boundaries
- Sounds can be identified much better by the formants and by their transitions.
  - dark region (high amplitude)
- Formants carry the identity of sound.

Note: Hidden Markov models implicitly model these spectrograms to perform speech recognition

Date ☺ Page No.

Usefulness of spectrogram

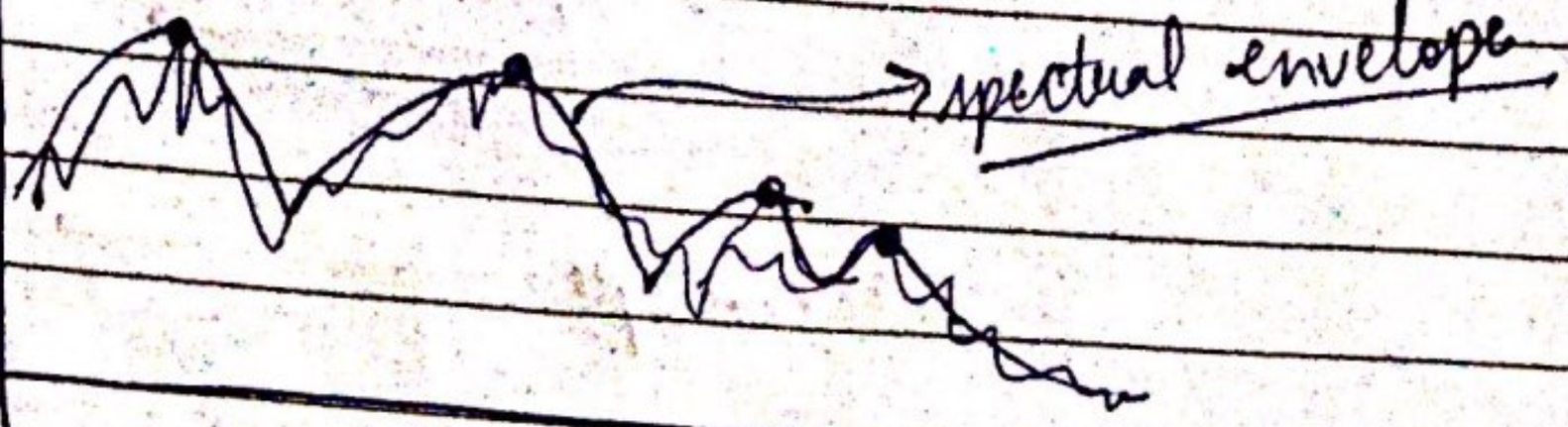
- Time - Frequency representation of the speech signal
- Spectrogram is a tool to study speech sounds (phones)
- Phones and their properties are visually studied by phoneticians
- HMM - simplicity models spectrograms for speech to text systems
- Useful for evaluation of text to speech systems
  - A high quality text to speech system should produce synthesized speech whose spectrogram should nearly match with the natural sentences.



Date 00 Page No. / /

Cepstral Analysis

A sample speech spectrum



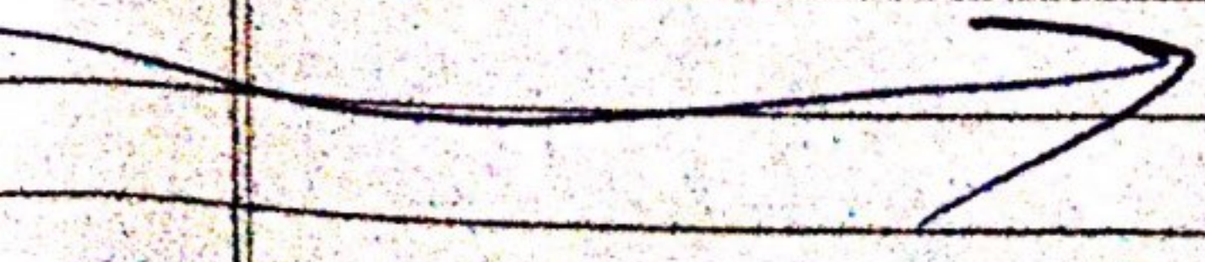
- Peaks denote dominant frequency components in speech signals
- Peaks are referred to as formants
- formants carry identity of sound

What we want to Extract

Spectral Envelope

- Formants and a smooth curves connecting them
- This smooth curve is refered to as spectral envelope

Find peaks and curves connecting them



Date / / Page No.

After Extracting spectral envelope from spectrum what we left there is spectral details

Spectrum ——— represented as ——→ log X[k]

Spectral envelope ——————→ log H[k]

Spectral details ——————→ log E[k]

log X[k] = log H[k] + log E[k]

1) Goal :- Separate spectral envelope and spectral details from the spectrum

2) Given log X[k] separate log H[k] and log E[k]

How to achieve separation

⇒ As we know spectrum is already a FFT of wave plot. We have converted amplitude into log(amplitude).

⇒ An FFT of an spectrum referred to as IFFT (Inverse FFT)

Date (☺) Page No.

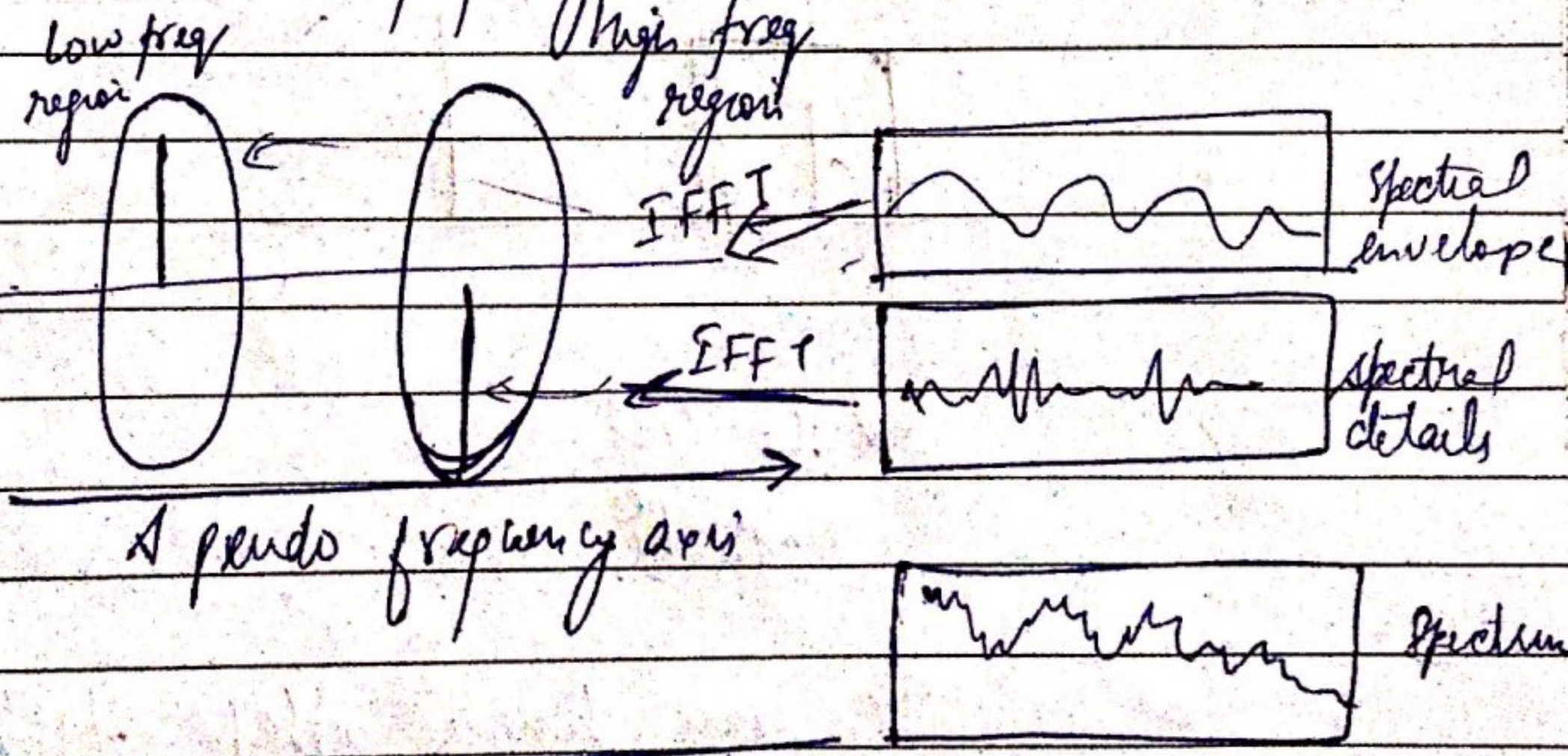
IFFT (FFT (waveplot)) = waveplot

But we after taking FFT (waveplot) we obtain y axis (amplitude) & IFFT (spectrum) we obtain more return frequency.

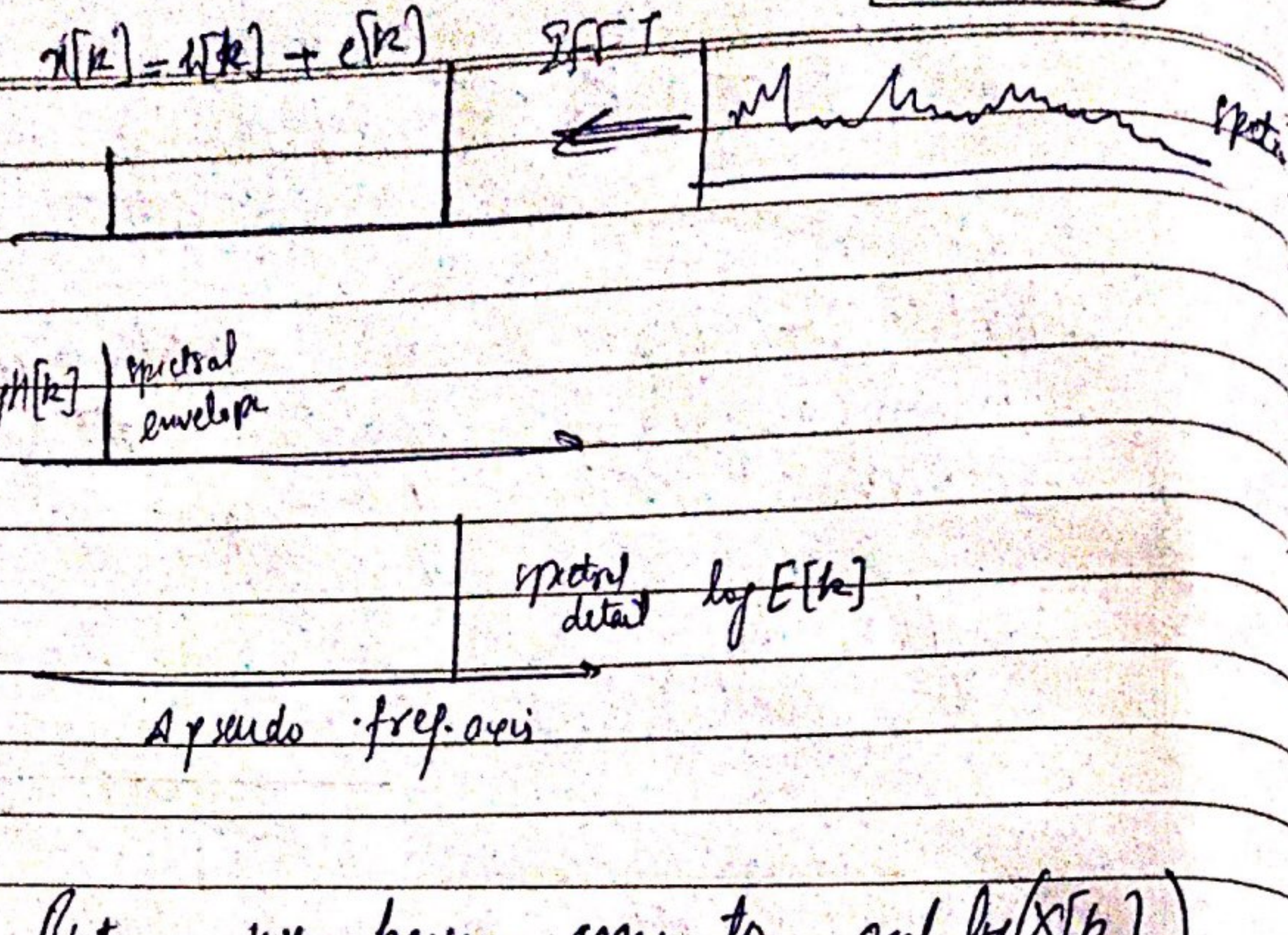
→ IFFT of log spectrum would represent the signal in pseudo - frequency axis

Bottom - up approach

Assume spectral envelope as a sine wave with four cycles per sec (assum x-axis as time) If we take its FFT then it gives a peak at 4Hz in frequency axis

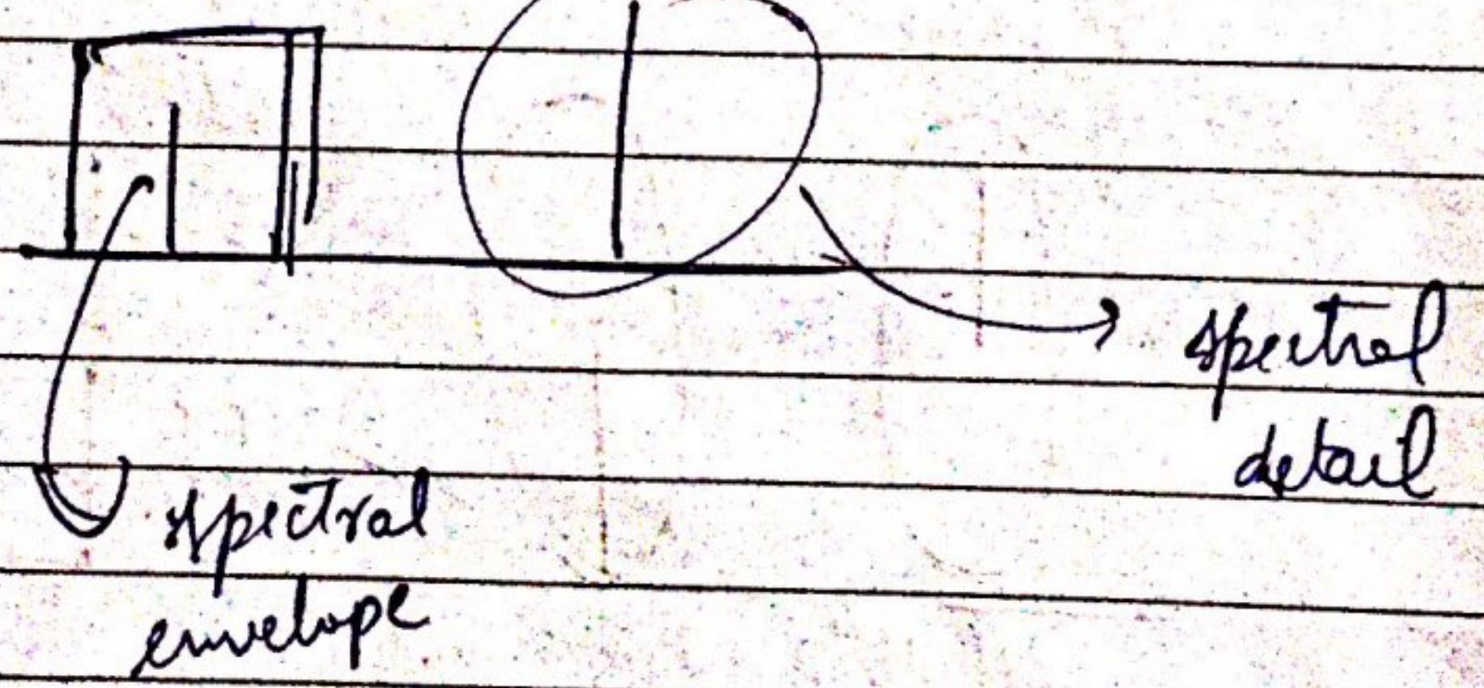


Similarly assume spectral details as sine wave with 100 cycles per second. If we take IFFT we get peaks at 100Hz



→ But we have access to only $\log(X[k])$ and hence you can obtain $x[k]$

and it have a lower freq and higher freq component that correspond to spectral envelope and spectral detail.



→ If you know $x[k]$ filter the low frequency regions to get $h[k]$.

$x[k]$ is referred to as CEPSTRUM

Date ☺ Page No.
/ /

$h[k]$ represents the spectral envelope and is widely used as feature for speech recognition.

Mel-Frequency Analysis (Speech recognition behave a human ear)

Review : What we did

- We captured spectral envelope (curve connecting all formants).
- But perceptual experiment says human ear concentrates on certain region rather than whole of the spectral envelope. So how to find those regions.

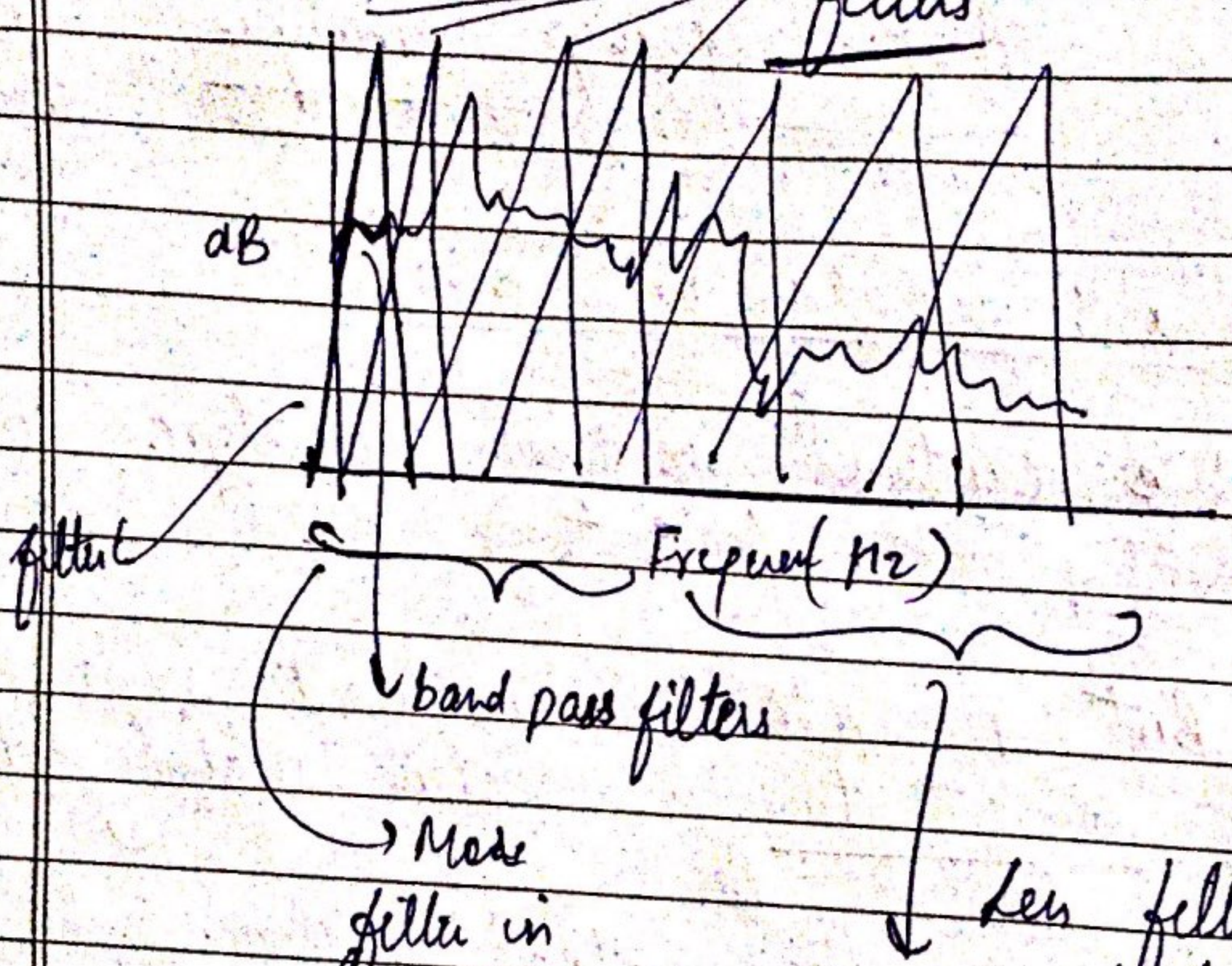
(Concentration points also known as speech perception)

⇒ Mel-Frequency analysis of speech is based on human perceptual experiments

⇒ It is observed that human ear acts as a filter and concentrates on only certain frequency components.

⇒ These filters are non uniformly spaced on frequency axis i.e. more filters on low freq

regions. and less filters in high frequency regions."



→ Filters only pass freq ~~inside~~ that fall within.

→ 41 filters are found to be effective.

There 41 filters ~~fit~~ have particular positions to be placed based on Mel-Scale

Mel- Frequency Cepstral Coefficients (MFCC)

Waveform $\xrightarrow{FFT}$ Spectrum $\xrightarrow{\text{Mel. Filter}}$ MEL Spectrum

$$\log X[k] = \log(\text{Mel-spectrum})$$

$$\log x[k] = \log H[k] + \log E[k]$$

Taking IFFT (we get low freq and high freq region)

$$x[k] = h[k] + e[k]$$

Cepstral coefficient $h[k]$ obtained for Mel-spectrum are called MFCC

2) Apply this to all spectrums we get MFCCs for all time windows called cepstral vectors.