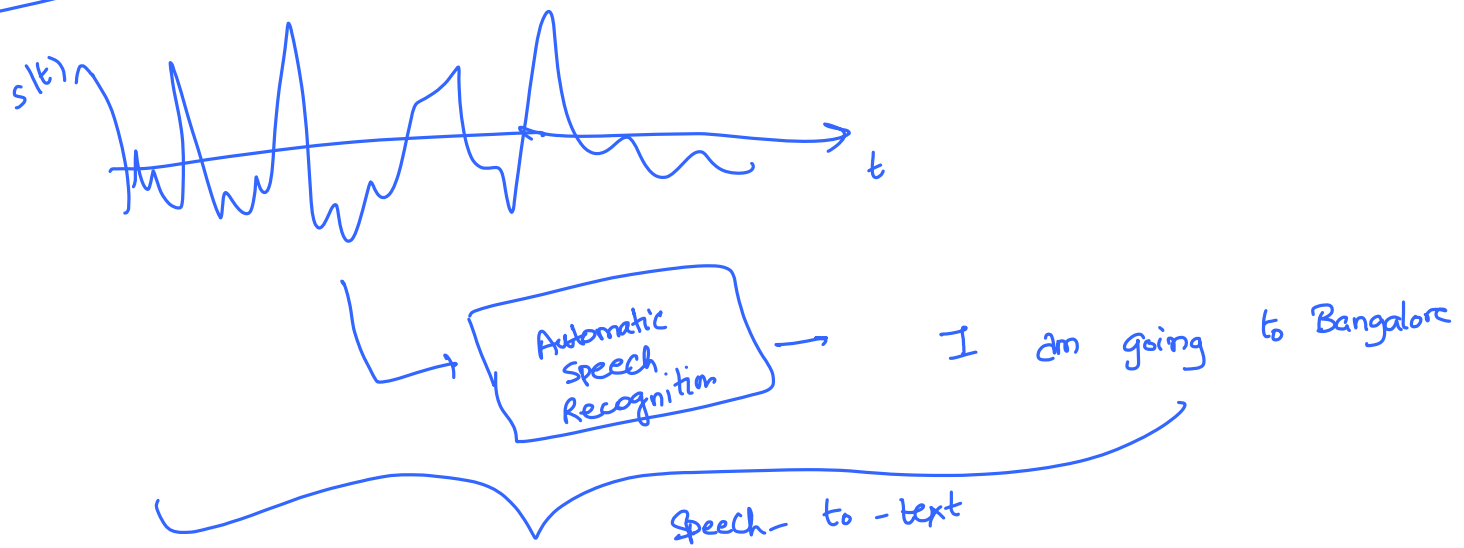
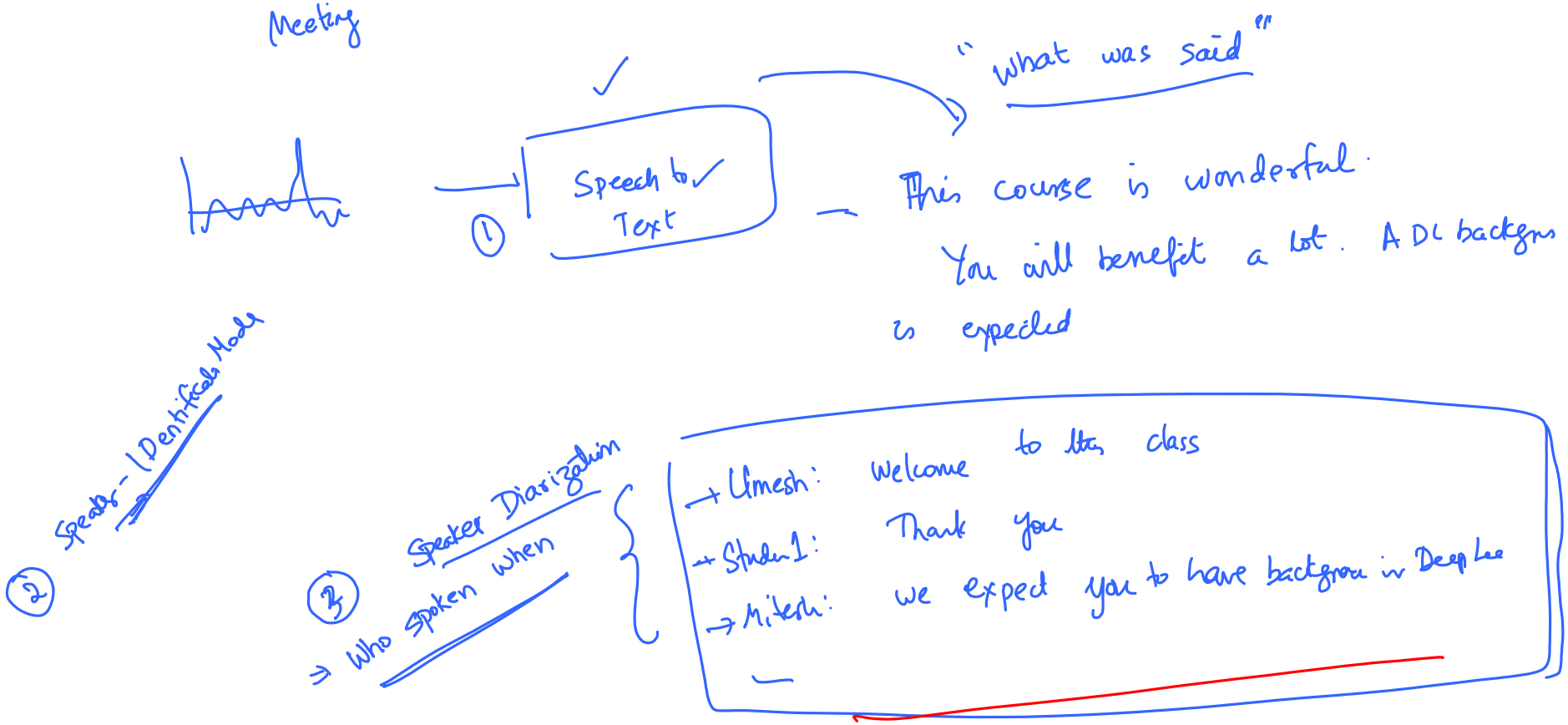
Note Title

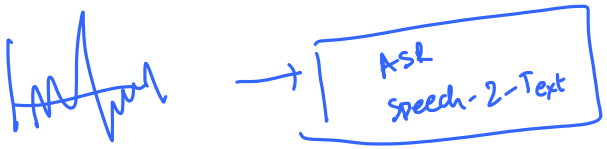
Week-2 for Speech Part → DLP- 6th

Speaker Diarisation




Meeting
Speech to Text
① "what was said"
This course is wonderful. You will benefit a lot. A DL backgrou is expected
② Speaker Identification Model
③ Speaker Diarization
→ Who spoken when
→ Umesh: Welcome to the class
→ Stud1: Thank you
→ Mitesh: we expect you to have backgrou in Deep Le

① Speech-to-Text ⇒ Whisper ASR Model ⇒ 680,000 hrs
⇒ open AI

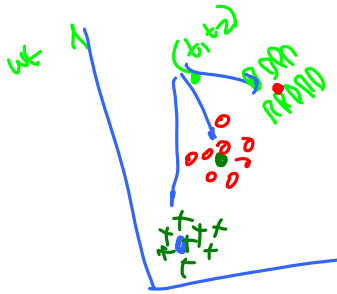


2:03 2:08 This is a Test ✓
2:08 2:17 Therefore the success depends on your inpu ✓

② Speaker-Identification Module
⇒ Speaker-Embedding Extractor
"Speaker-feature" Extractor

} Speech Brain Tool Kit
ECAPA-TDNN

Speech Tool-Kits
Kaldi → JHU
Espnet → JHV/CMU
Speech Brain → MILA



$$\begin{bmatrix} h_1 \\ w_1 \end{bmatrix} \hookleftarrow \begin{bmatrix} h_2 \\ w_2 \end{bmatrix} \quad \begin{bmatrix} h_3 \\ w_3 \end{bmatrix}$$

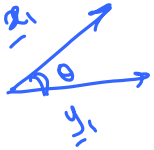
$$\begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

"closeness"

$\Rightarrow$ Euclidean Distance

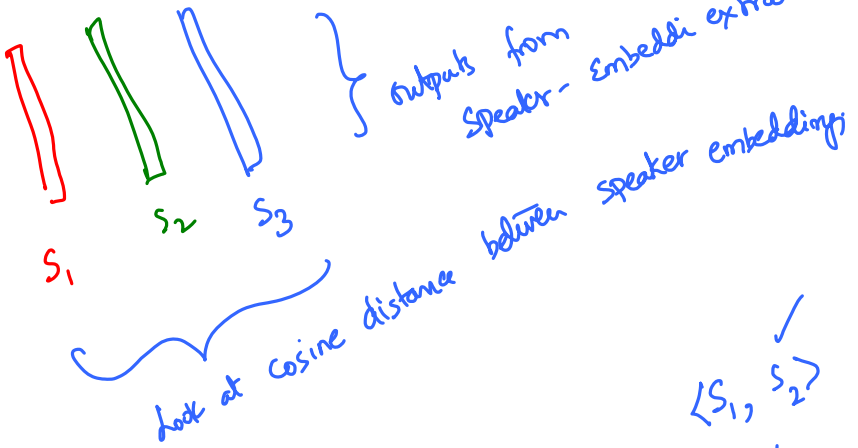
$$\underline{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} \qquad \underline{y}_1 = \begin{bmatrix} y_{11} \\ y_{21} \end{bmatrix}$$

$$\|x_1 - y_1\| = (x_{11} - y_{11})^2 + (x_{21} - y_{21})^2$$



cosine distance

$$\cos\theta = \frac{x_1^T y_1}{\|x_1\| \|y_1\|}$$



$$\langle s_1, s_2 \rangle = \frac{s_1 s_2}{\|s_1\| \|s_2\|}$$

$\langle s_1, s_3 \rangle$

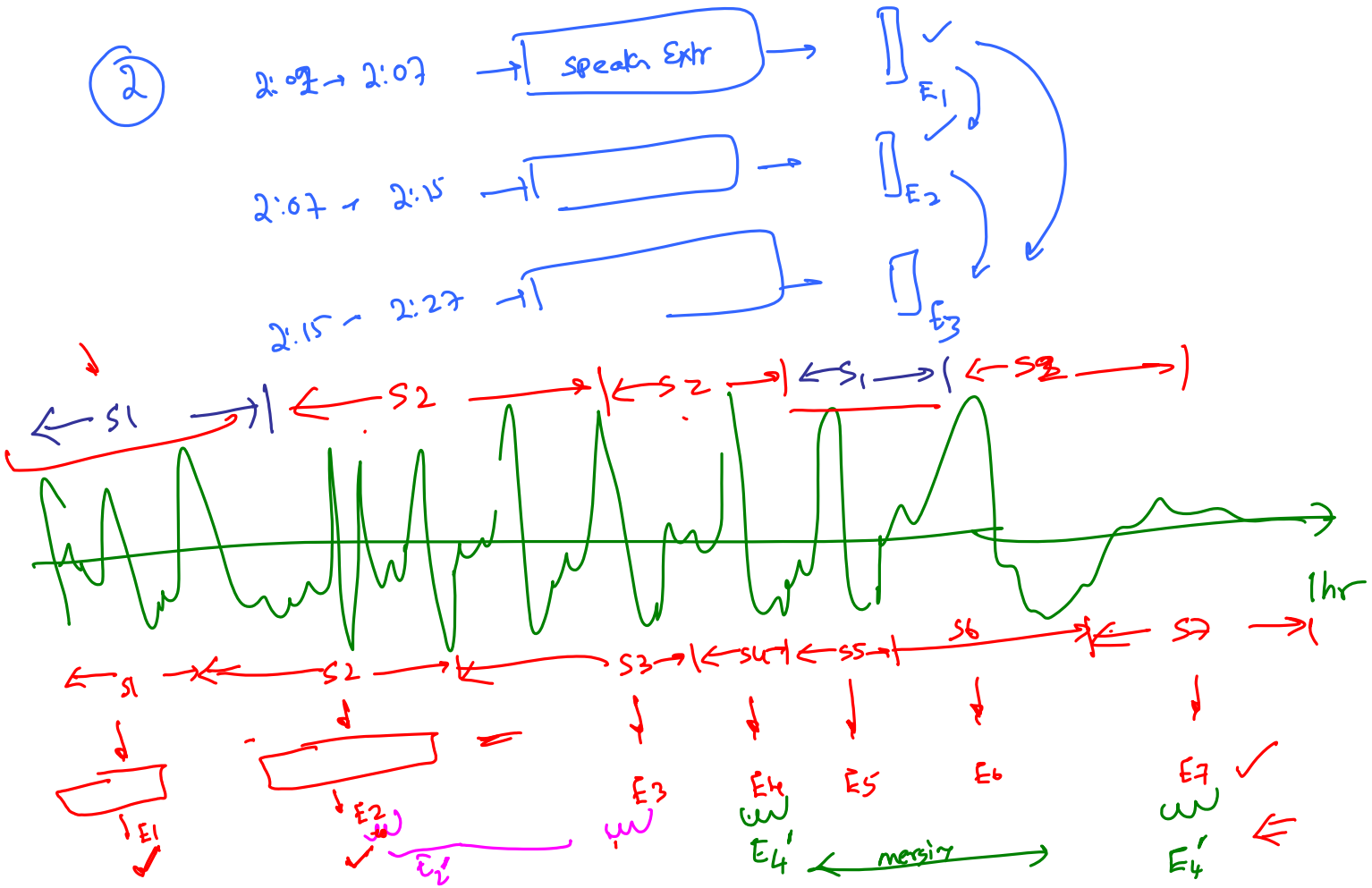
$\langle s_2, s_3 \rangle$

Recall: Speaker Diarisation: "Who spoken when"

S1 2:02 – 2:07 This is a test

S1 2:07 – 2:15 Good Test

S2 2:15 – 2:27 How are you


②
2:00 → 2:07
Speaker Extr
E1
2:07 → 2:15
E2
2:15 → 2:27
E3
S1
S2
S2
S1
S2
1hr
S1
S2
S3
S4
S5
S6
S7
E1
E2
E2'
E3
E4
E5
E6
E7
E4'
E4'
merge

I know only "2 speakers"

AGGLOMERATIVE CLUSTERING ⇒ "clusters" ⇒ and keep merging clusters

⇒ until I get "desire number of clusters"

Initial 7 segments → 7 speakers

$\langle E_1, E_2 \rangle$ $\langle E_1, E_3 \rangle$ $\langle E_1, E_4 \rangle$ .. $\langle E_1, E_7 \rangle$

$\langle E_2, E_3 \rangle$ $\langle E_2, E_4 \rangle$ .. $\langle E_2, E_7 \rangle$

.. $\langle E_6, E_7 \rangle$

} Miniminum Cosine Distance

Let us say E_4, E_7 ⇒ has min. cosine distan