UNIVERSITY of HOUSTON
DEPARTMENT OF COMPUTER SCIENCE

Text Based Indexing to Ease Navigation in Lecture Video

Tayfun Tuna, Varun Varghese and Jaspal Subhlok

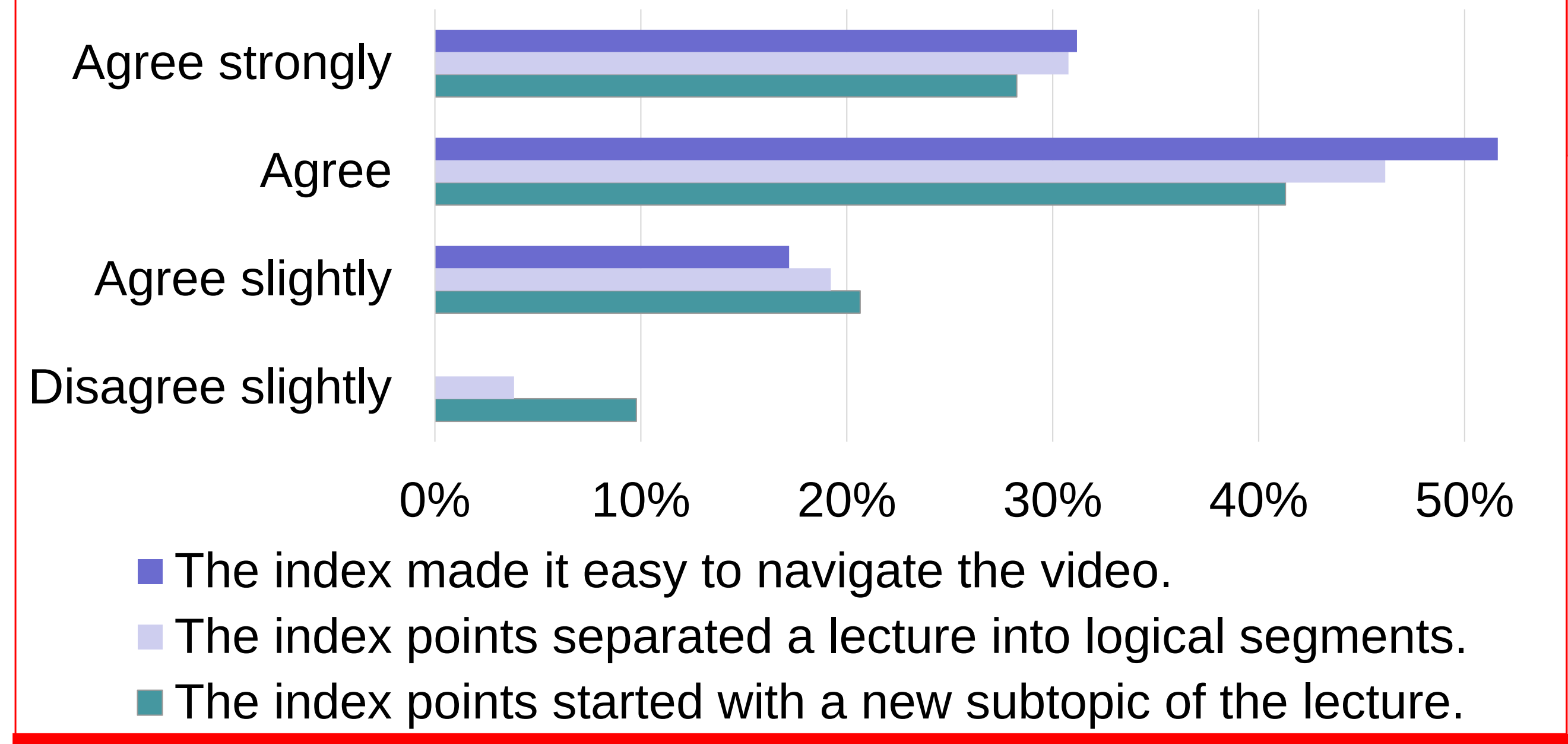
Background

- Videos of classroom lectures have proven to be a popular and versatile learning resource.
- A major weakness of recorded lecture videos is the inability to quickly access the content of interest.
- “Indexed Captioned Searchable (ICS) Videos” framework aims to provide quick access to video content of interest by ICS:
  - Indexing: Segmented videos
  - Search: Keyword search in video
  - Captioning: Scrolling text for audio

What is Video Indexing?

- Videos are automatically divided into logical segments, each represented by a visual index snapshot.
- User can access/switch to these segments without watching whole video.

Value Of Video Indexing

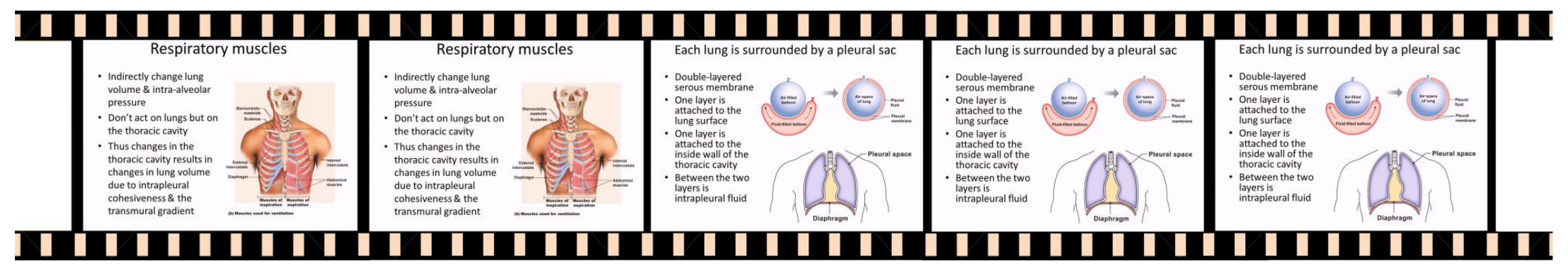


Research Question

- What is the best way to do video indexing to provide conceptual segmentation which each index points will represent subtopics?

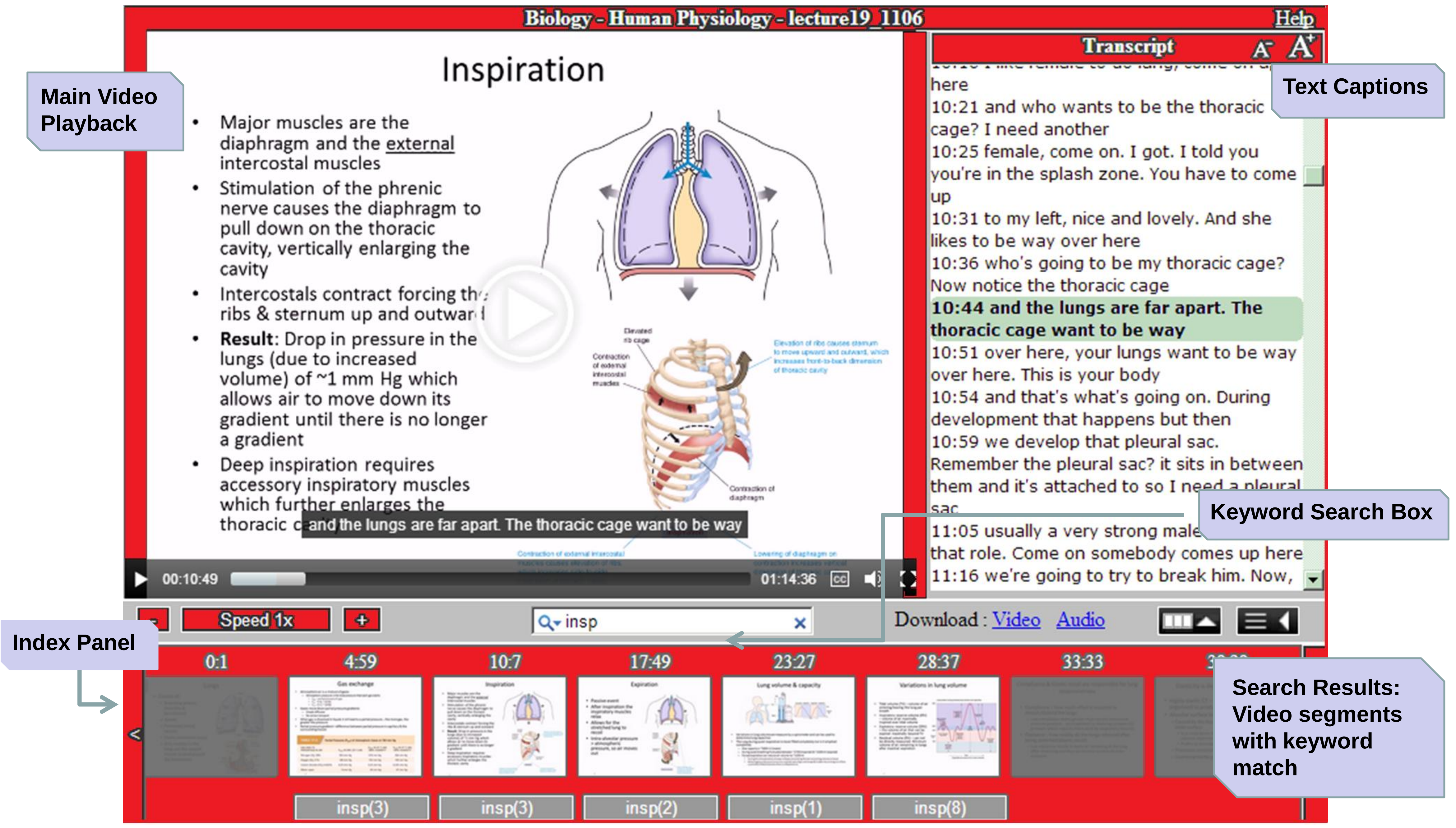
How To Do Video Indexing?

- Index points should be:
  - Meaningful: can represent subtopics
  - Not too many :scrollable
  - Not too few: broad
- Video indexing requires:
  - Identifying Transition Points (TP) where video scene changes.



  - Identifying Index Points (IP) : Select some TPs as IP based on text similarity.
- Assumption is that topics within the video are associated with different groups of terms/words.

ICS Video Player



Biology - Human Physiology - lecture19_1106

Inspiration

- Major muscles are the diaphragm and the external intercostal muscles
- Stimulation of the phrenic nerve causes the diaphragm to pull down on the thoracic cavity, vertically enlarging the cavity
- Intercostals contract forcing th ribs & sternum up and outwar
- **Result**: Drop in pressure in the lungs (due to increased volume) of ~1 mm Hg which allows air to move down its gradient until there is no longer a gradient
- Deep inspiration requires accessory inspiratory muscles which further enlarges the thoracic c

and the lungs are far apart. The thoracic cage want to be way

here
10:21 and who wants to be the thoracic cage? I need another
10:25 female, come on. I got. I told you you're in the splash zone. You have to come up
10:31 to my left, nice and lovely. And she likes to be way over here
10:36 who's going to be my thoracic cage? Now notice the thoracic cage
10:44 and the lungs are far apart. The thoracic cage want to be way
10:51 over here, your lungs want to be way over here. This is your body
10:54 and that's what's going on. During development that happens but then
10:59 we develop that pleural sac. Remember the pleural sac? it sits in between them and it's attached to so I need a pleural sac
11:05 usually a very strong male that role. Come on somebody comes up here
11:16 we're going to try to break him. Now,

Algorithm for Video Indexing

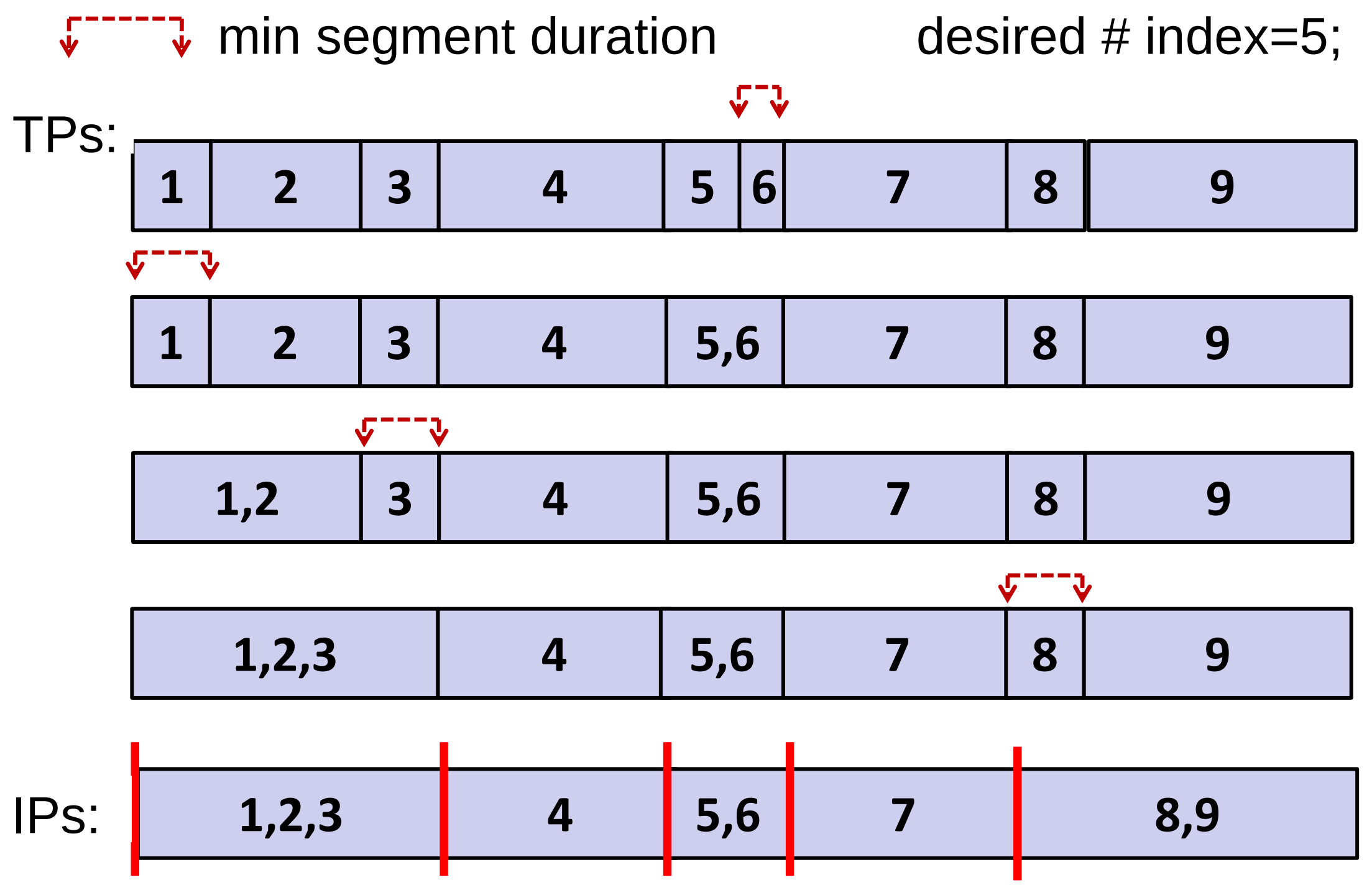
1. Set “desired number of IP”
2. Set “min segment duration”
3. Find the segment has the smallest duration
4. Compare the **color/text difference** with left and right and merge:

```
IF dif(current,left)>dif(current,right)
    THEN  merge(current,right)
    ELSE  merge(current,left)
```

5. Repeat 3-4 until:

smallest_segment duration > min segment duration
and
total number of segments == desired number of IP

Example Steps for Video Indexing



Text Based Video Indexing

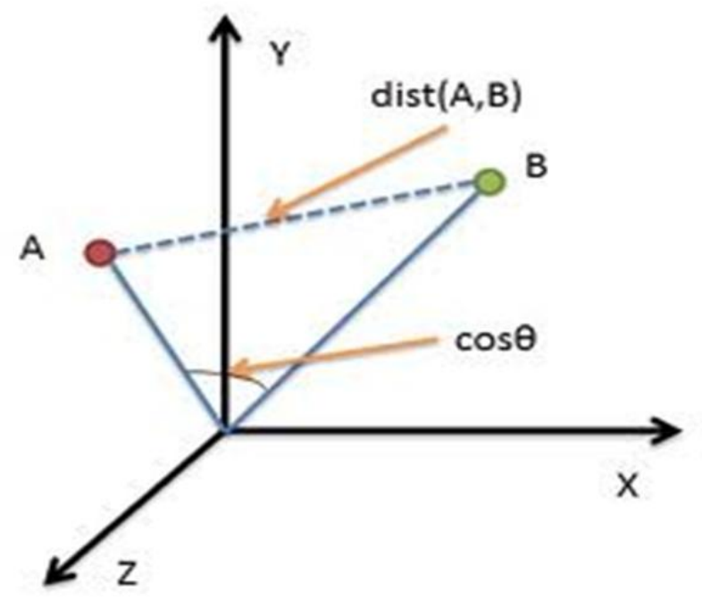
- Text on the video frames is extracted using OCR technology.
- The similarity between video sections is determined by analyzing term-frequency vector of the text sections.

Frame1	Frame2	Frame3
cavity, cavity	cavity	-
deep, deep	deep, deep	deep, deep
-	nerve, nerve	nerve, nerve, nerve

- Each frame is vector of frequency of the all words.

Word/ Frame	Frame1	Frame2	Frame3
cavity	2	1	0
deep	2	2	2
nerve	0	2	3

- Text Similarity of two frame is measured by the “**Cosine angle**” of the vectors:



$$\boldsymbol{cos(\theta)} = \frac{\boldsymbol{A.B}}{\|\boldsymbol{A}\|.\|\boldsymbol{B}\|}$$

$$\text{sim(Frame1,Frame2)} = \frac{2*1 + 2*2 + 0*2}{\sqrt{(2^2+2^2+0^2)} * \sqrt{(1^2+2^2+2^2)}} = 0.70$$

$$\text{sim(Frame2,Frame3)} = \frac{1*0 + 2*2 + 2*3}{\sqrt{(1^2+2^2+2^2)} * \sqrt{(0^2+2^2+3^2)}} = 0.92$$

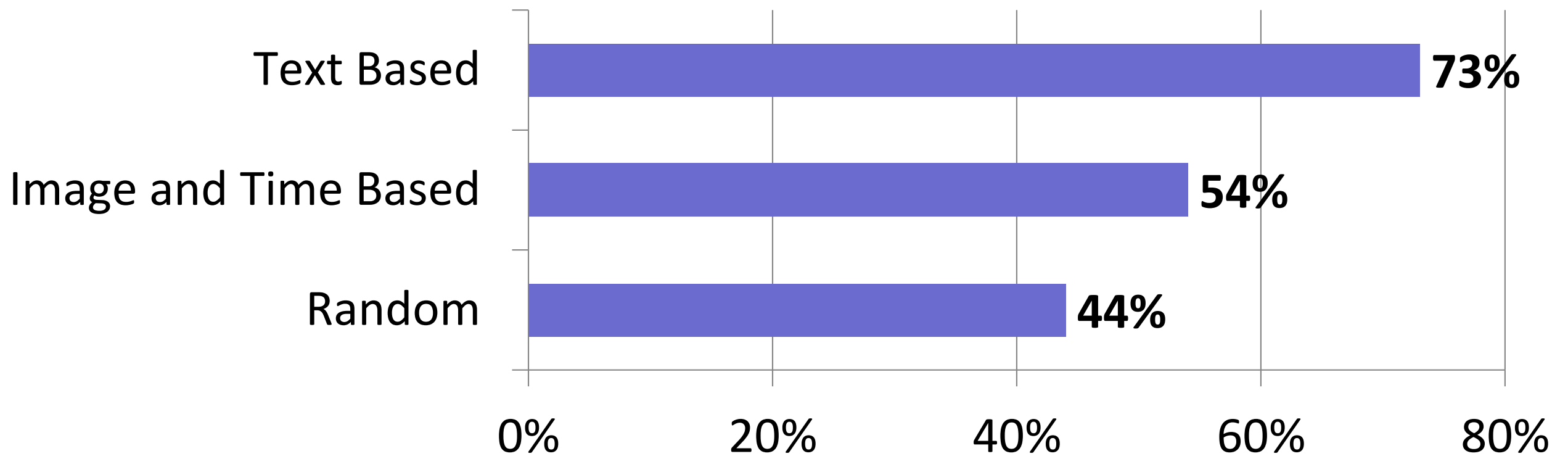
Evaluation of Text Based Indexing

- 25 diverse lecture videos were selected from computer science, biology and geology and were indexed manually to determine the ground truth.
  - Average of 75 minutes per video
  - Total 30+ hours of video
- 1700 TPs tagged as “definitely IP”, “probably IP”, “probably NOT IP’ and “definitely NOT IP”.
- 3 Different indexing methods are compared: Random(IPs randomly selected) , image and time based(IPs selected by scene/color changes) and Text Based.

Ground Truths

Algorithm Output	Definitely Not IP	Probably Not IP	Probably IP	Definitely IP
0 (Not IP)	(+2)	(+1)	(-1)	(-2)
1 (IP)	(-2)	(-1)	(+1)	(+2)

- Experiment results shows that text based indexing method provides far more accuracy than others.



Conclusion

- Text based indexing algorithm provides far more accuracy than image based and random indexing algorithms, 73% vs. 54% and 44%.
- Text based indexing was successfully used to index over hundreds of videos and got positive feedbacks from user surveys.
- Text based indexing is integrated with Indexed Captioned Searchable (ICS) Videos framework that includes indexing, search, and captioning in video playback and has been used by dozens of courses and 1000s of students.

Challenges and Future Work

- Incremental slide progress, irrelevant text appearing in a concept, image dominated slides with little texts are found as some challenges for finding the correct index points. Instead of comparing the slide with immediate left and right, comparing it to all slides in both sides in a weighting schema (so that closer frames will have more effect) is proposed to overcome these challenges.
- Each video has its own profile (# of words per slide, duration per slide etc..). A machine learning approach to define thresholds for different profiles is expected to increase text based indexing accuracy.