

Indexed Captioned Searchable Videos

Jaspal Subhlok

**University of Houston
Department of Computer Science**



Outline

- Technology and Learning (& MOOCs)
- Motivation for Employing & Enhancing videos
- How does it work: Demo
- Building It: The Technology
- Using It: Deployment & Lessons from Surveys
- Wrap Up

[The opinions in this talk are those of the speaker and do not represent the official view of Department of Computer Science or any funding agency]



Technologies Influence Education

1. (1480s) **The Printing Press -> Textbook**
 - + Fears of Professors were misplaced (settled in ~20 years)
 - + for education quality (not sure about memory capacity)
2. **Computer, Viewgraphs/Powerpoint Lectures**
 - + Professor & Student convenience
 - Influence on education less clear
3. **Clickers, Video, Social Media, Online xxx, Tablets, Mobile Devices, Wikipedia,....**
 - Not clear where this is heading
 - **Model of lecturing+homeworks+exams intact so far!**



What About MOOCs

Presented as an alternative model, but

- Motivated by cost reduction, but no convincing business model or integration with classrooms
- Evolution of distance learning and Evolution of textbooks
- Very valuable in many scenarios but overhyped
- Distraction from other technological developments with potential for impact



Drivers of Change ?

- Learning literature says “students not learning much” (critical thinking, reasoning, writing)
- Financial squeeze as governments reduce/stop funding education. More like a business.

Education enterprise will change with maturing of new technologies and approaches (e.g. “Inverted/Flipped” classrooms)

- Events like shutting/scaling down of (smaller) universities more likely than ever!
- Combination of technology and external pressures



ICS Videos

Research project to enhance the value of video as a learning resource employing automated (where possible) Indexing, Search and Captioning.

Motivation for ICS Videos Project

- University of Houston has been a leader in supporting video for coursework (NSM IT)
 - Tablet PCs to teach and record lectures and make them available online as study material

Surveys showed videos are a powerful, versatile learning resource



Survey Results - 1

- **N=2,349** taken from 43 sections in biology, computer science, geology, chemistry, and mathematics/physics between 2009 to 2011.
- Course sizes varied from 8 students to >300.
- Each professor posted ~25 videos per semester

These surveys are recent but the basic results have been similar even before this project started

Survey instrument details are often skipped in this talk to save time – available in papers.



Students Watch Videos

84%

used lecture
videos at
least once

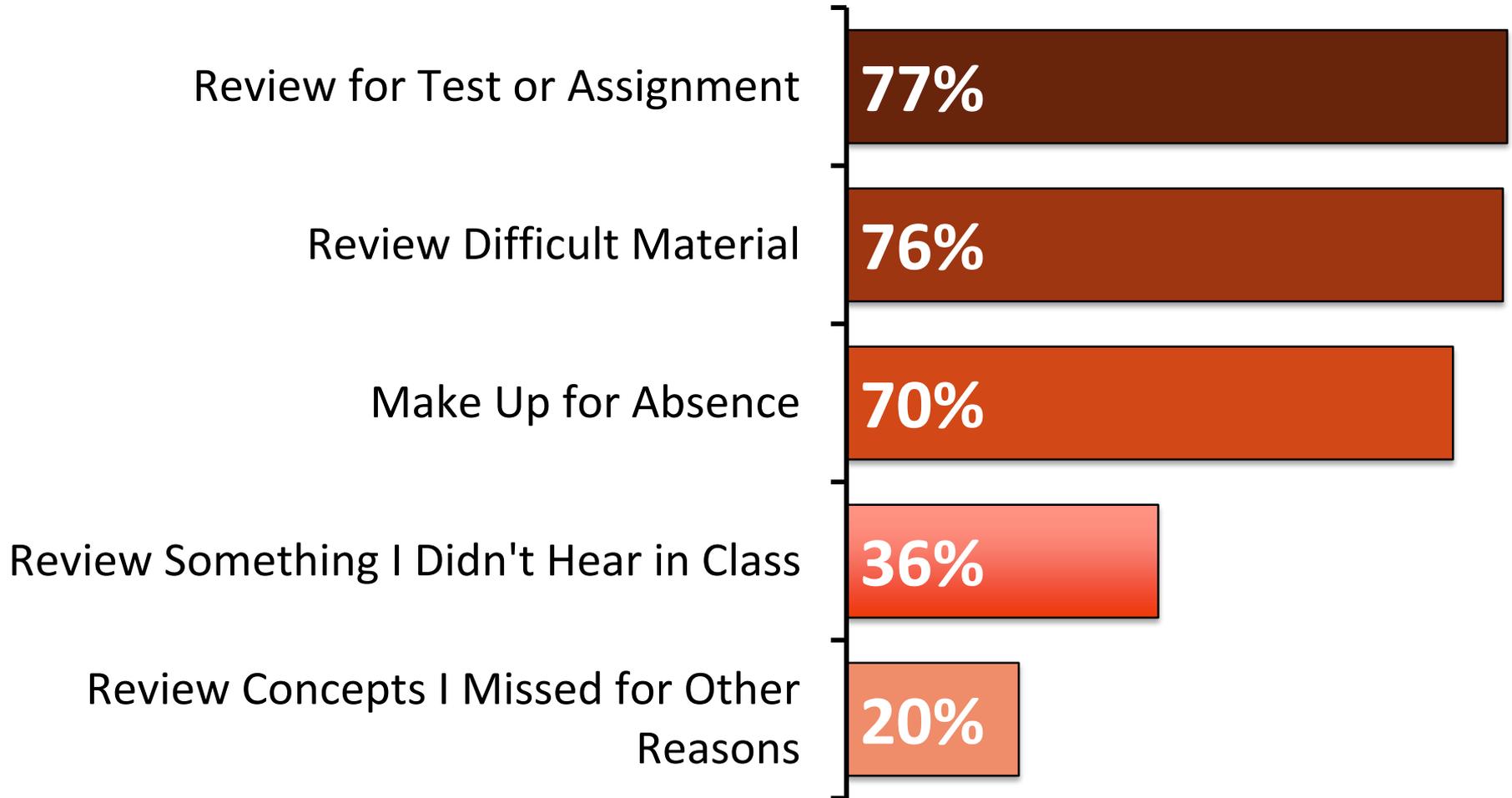
81%

watched the
entire video,
not only a
part they
needed

47%

watched a
single video
more than
once

Reasons Students Used Videos

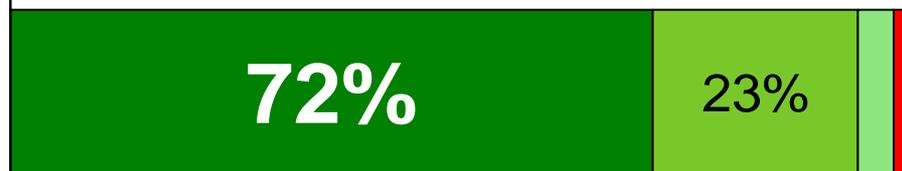


Students Strongly Value Videos

"Having access to lecture videos for this class is important to me."



"Lecture videos are useful for reviewing."



"The lecture videos helped me to study for quizzes or tests."

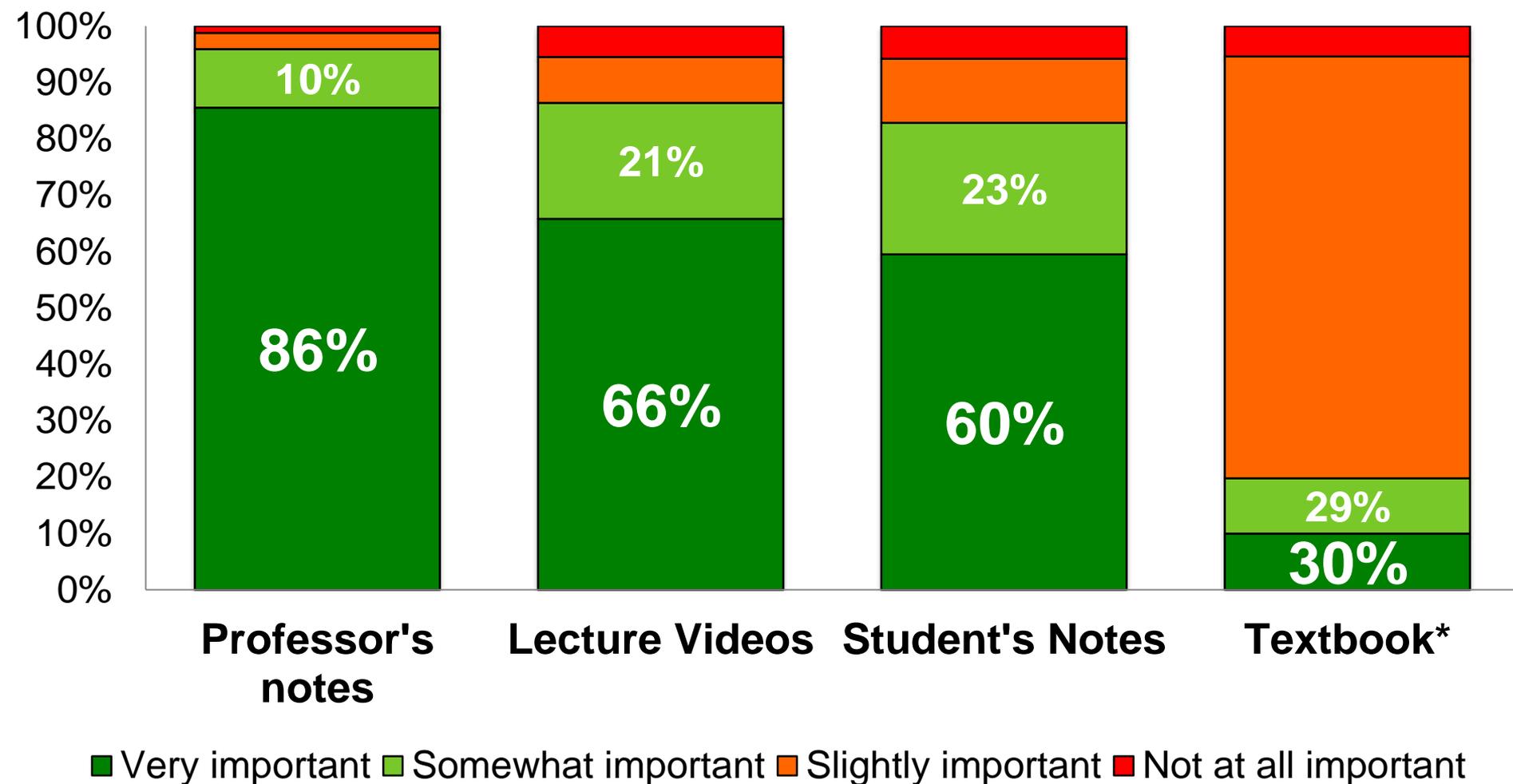


"Lecture videos help me to clarify material that was not clear in class."

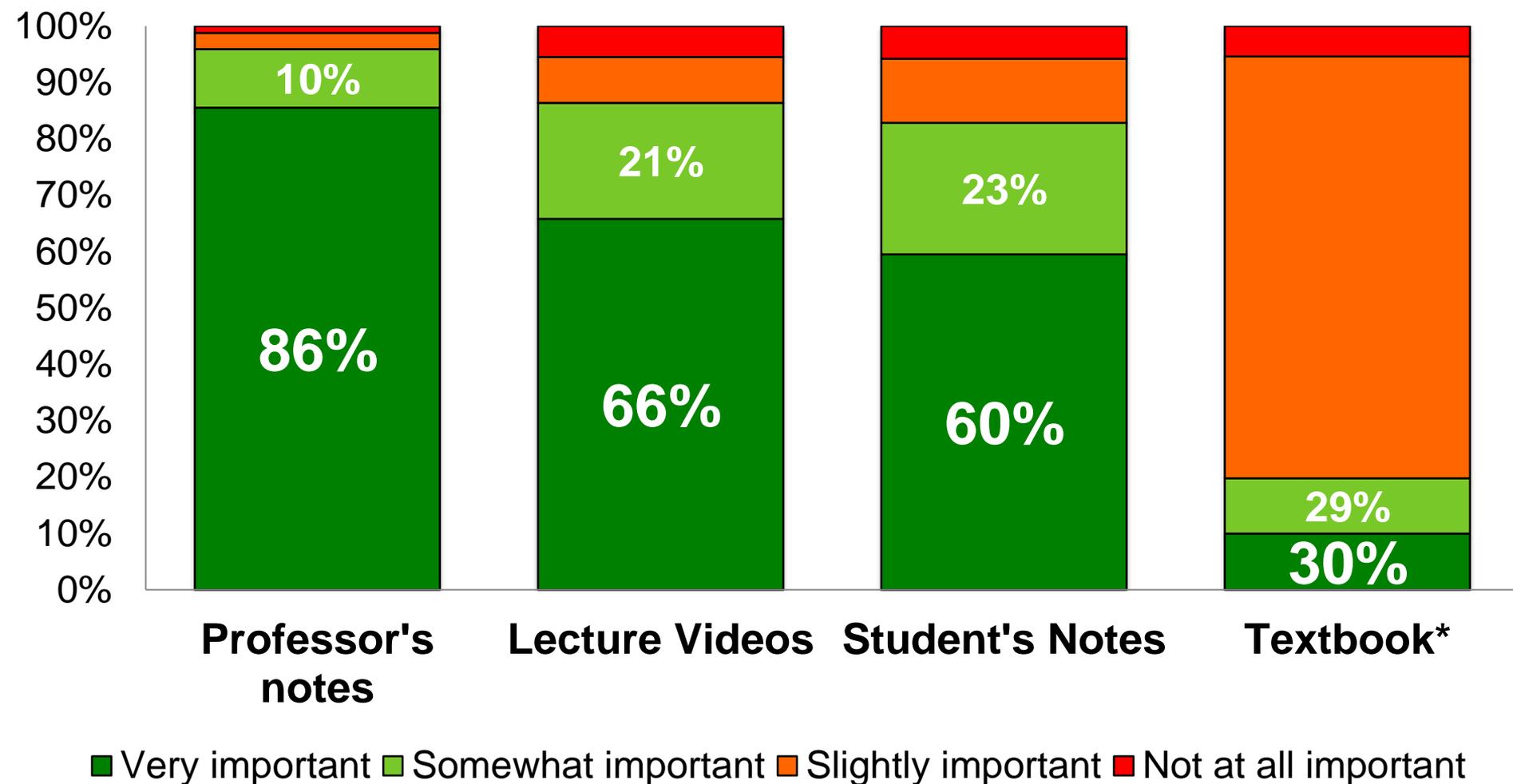


■ Agree strongly ■ Agree ■ Agree slightly ■ Disagree

Comparison with other Resources



Comparison with other Resources



Motivation (continued)

Videos are valuable but:

- Key shortcoming of video format is the inability to quickly access content of interest
 - Loud and clear in surveys and interviews (not shown)
 - Students want answers to questions for review, not watch an hour long video!

Goal: ICS video player with advanced Indexing, Captioning, and Search

A Project to bring clear and present benefits to students with cool computing research on the way!



ICS Videos

Computer Science - Introduction to Computing - COSC 1300 Lecture 21 Help

On **Transcript** **A⁻** **A⁺**

0:05 Okay so I need 3 more volunteers for today to give your powerpoint.

0:10 presentation. Let's see you are Shannon, are you Shannon?

0:38 Jasmine, okay so you're in the third. Jasmine Scott.

0:57 Whats your name now? Your last name is?

1:35 okay I got 2, I need one more? We'll go with these two okay. First one is going to be Jasmine Scott here.

2:38 So Jasminen is going to talk to us about TranSwitch: Engines for Global Connectivity. Hello everybody I'm Jasmine. The company I chose was TranSwitch.

2:56 Basically its a company, they make integrated circuits and intellectual property solutions.

Speed 1x

00:01 04:05 06:33 15:55 23:05 25:51 31:01 35:31

quiz(1) quiz(1) quiz(9)

Building It – Challenges

Search: Keyword search Inside video

- OCR on video frames
- Semantic search

Indexing: Divide a lecture into topic segments

- Identifying topic changes
- Images, text, or audio?

Captioning:

- Speech recognition ineffective for classroom videos

Keyword Search

- Keyword Search requires text detection in video frames
- Can be accomplished by OCR tools



- Accuracy on lecture video images?

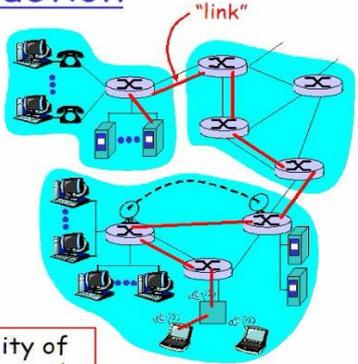
Images from Lecture Videos

- OCR: good for scanned images
- Lecture video images with colorful different layouts

Link Layer: Introduction

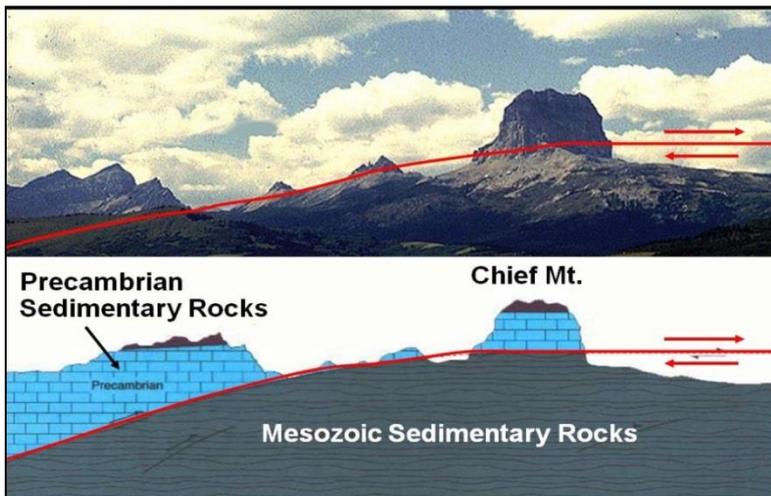
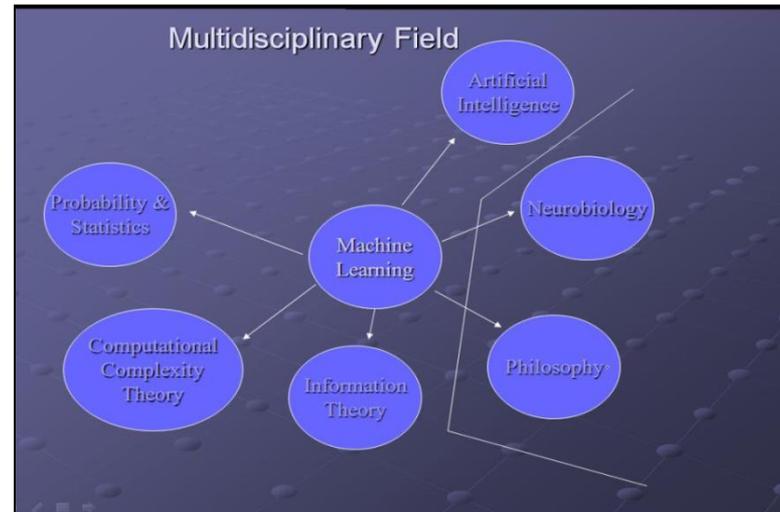
Some terminology:

- hosts and routers are **nodes**
- communication channels that connect adjacent nodes along communication path are **links**
 - wired links
 - wireless links
 - LANs
- layer-2 packet is a **frame**, encapsulates datagram



data-link layer has responsibility of transferring datagram from one node to adjacent node over a link

5: DataLink Layer 5-4



Testcross

- How can you determine the genotype of a dominant phenotype?
 - Purple-flowered pea plant is either PP or Pp : cross with a homozygous recessive (pp) white-flowered plant:

	If PP		or	If Pp	
	Sperm			Sperm	
	P	P		P	p
Eggs	P	Pp	Pp	Pp	Pp
	p	Pp	Pp	pp	pp
	All purple			$\frac{1}{2}$ purple; $\frac{1}{2}$ white	

- By definition, the **testcross** is used to determine the genotype of an organism expressing a dominant phenotype by breeding with a recessive homozygote.

Output of OCR Tools

Question 3

Where did the story say that there was a statue raised in Mrs. Bethune's honor?

Miami, Florida

Mayesville, South Carolina

TEXT

GOOCR

MODI

Question 3
Where did the story say that there was a statue raised in Mrs. bethune's honor?

Where did the story say that there was a statue raised in Mrs. Bethune's honor?

Washington, D.C.
Miami, Florida
Mayesville, South Carolina

Where did the story say that there was a statue raised in Mrs Bethune's honor?
Washington, D.C.
Miami, Florida
Mayesville, South Carolina

B-Nik is

Segmentation of Text Regions and Enlargement

Color Inversion

Segmentation of Text Regions

Image Example



Noisy, $\sigma=10$ (MSE=100)

denoised (T=3, MSE=56)

39

a) Original image

Image Example

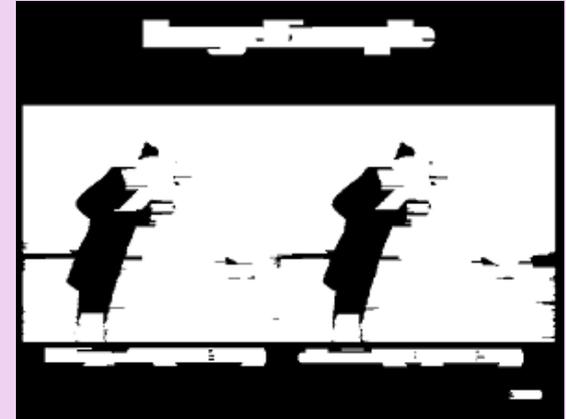


Noisy, $\sigma=10$ (MSE=100)

denoised (T=3, MSE=56)

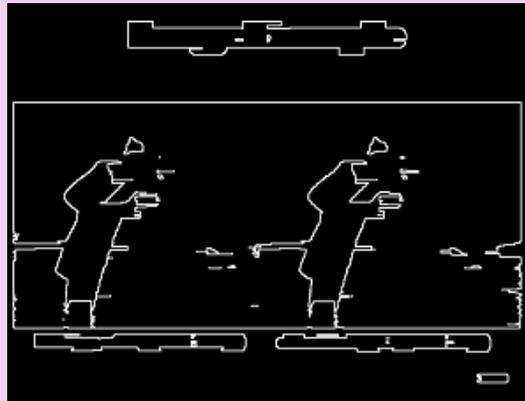
39

b) Binarization



c) Dilation effect

Image Example



d) Edge detection

Noisy, $\sigma=10$ (MSE=100)

denoised (T=3, MSE=56)

39

e) Blob extraction

Image Example

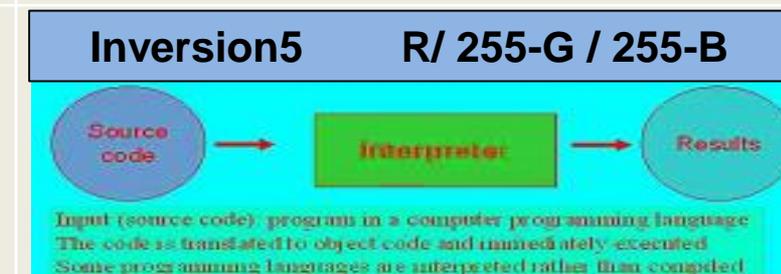
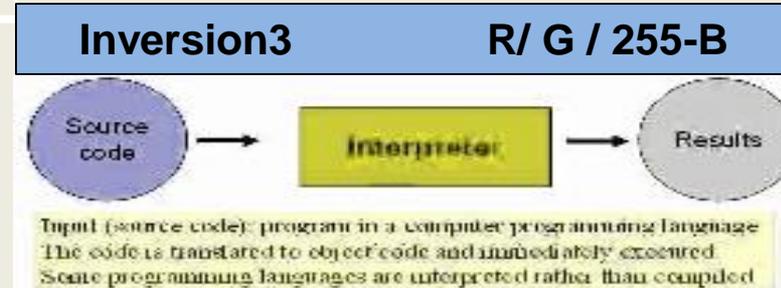
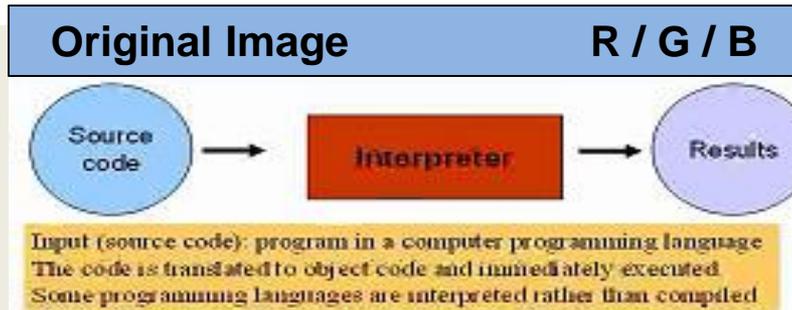
Noisy, $\sigma=10$ (MSE=100)

denoised (T=3, MSE=56)

39

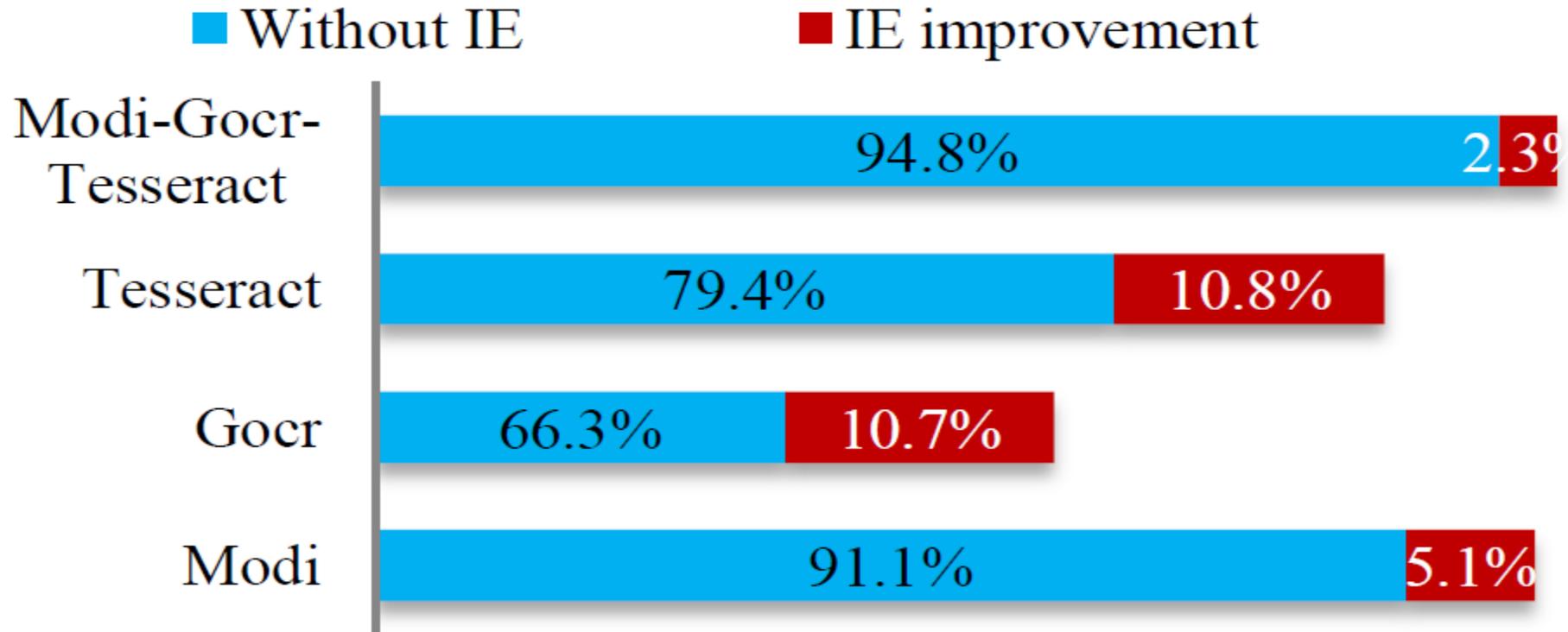
f) Enlargement

Color Inversion



Impact of Image Enhancement (IE)

Accuracy = %age of words detected correctly:



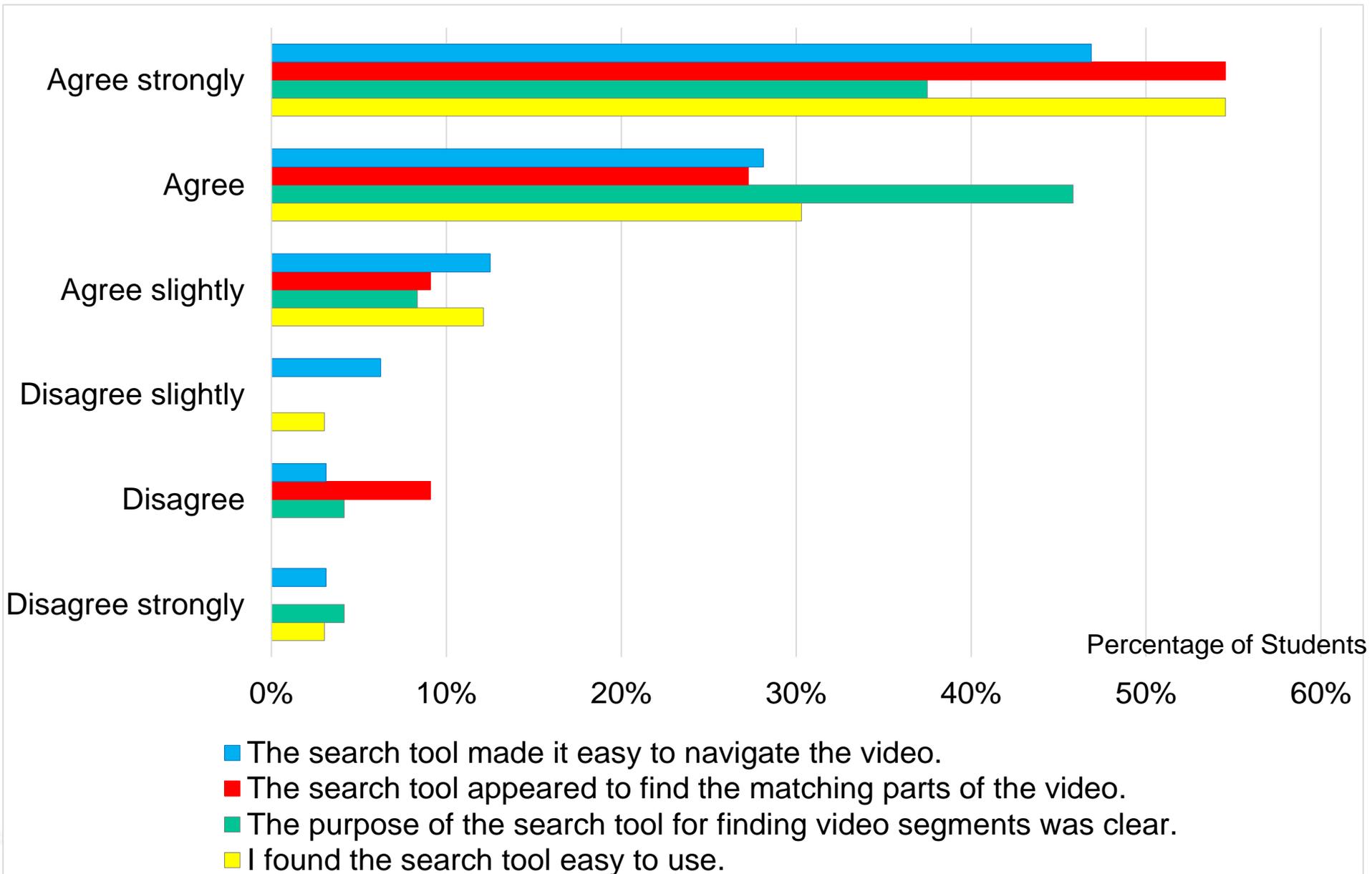
Many false positives – *not a major issue for Search*

Value of Search

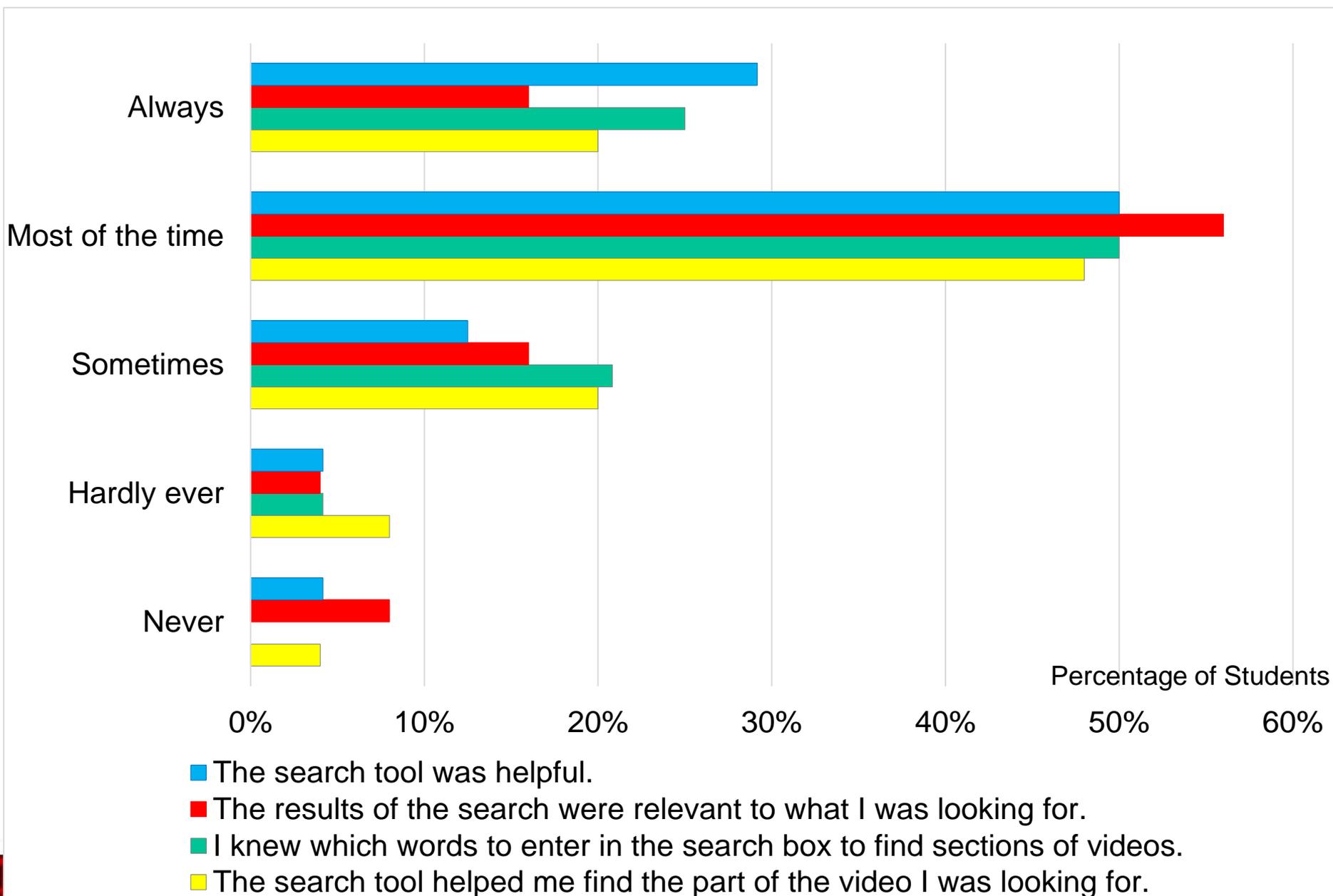
Results from Survey of Student Users



Survey-Search (1)



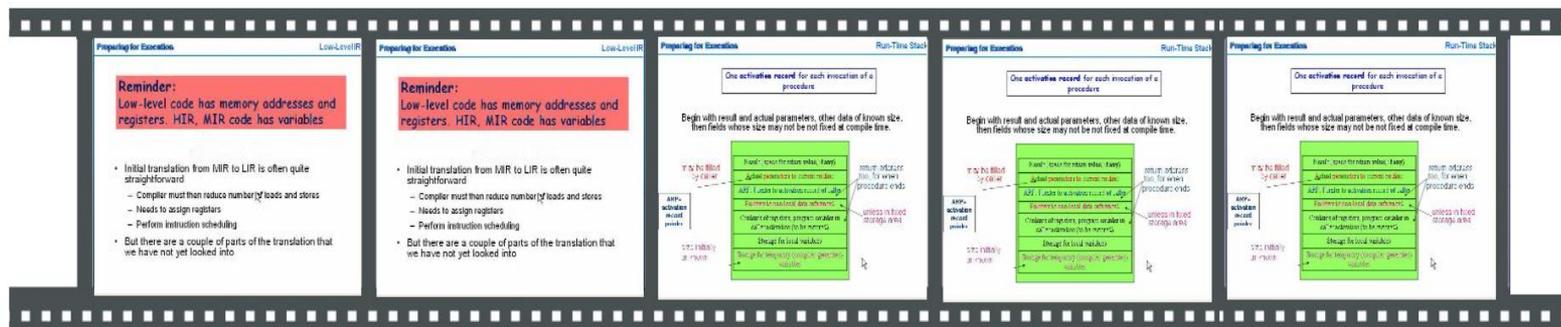
Survey-Search (2)



Video Indexing

Objective : Split video into meaningful, topical segments

1. Identify *Transition Points (TP)* where video scene changes (image- difference)



2. Identify subset of Transition Points that represent Topic changes!

Methods for Video Indexing

Split video in equal intervals of time

- Lazy reference method

Split video based on image difference

- Does not work well (J. Li MS thesis)

Split video based on text difference

Indexing by Machine Learning

Text Based Indexing Algorithm

Input:

A list of transition points

Required number of index points

Output:

List of index points

Repeat:

1. Select transition point/segment with smallest duration
2. Merge it left or right neighbor **based on text similarity**

Until

Reach required number of index points



Uniform Indexing Algorithm

desired # index=5

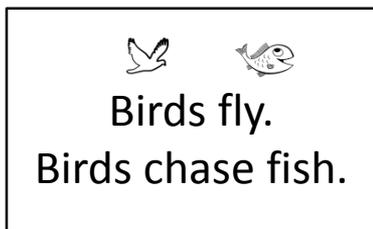
TPs:



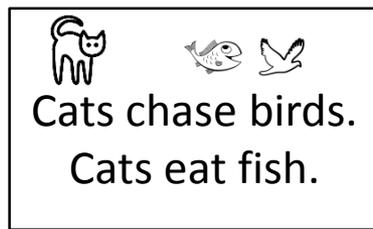
IPs: 1,2,4,5,7

Text Similarity Metric: Cosine

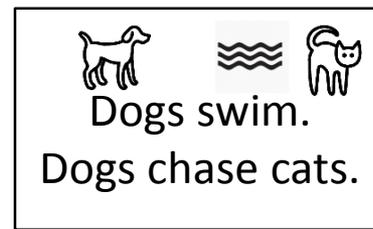
Frame 1



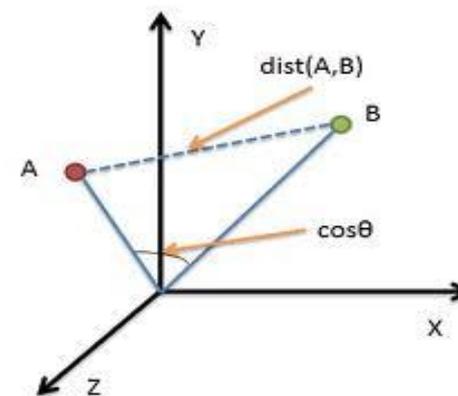
Frame 2



Frame 3



Word/Frame	Frame1	Frame2	Frame3
birds	2	1	0
cats	1	2	1
chase	1	1	1
dogs	0	0	1
fish	1	1	0
fly	1	0	0
swim	0	0	1



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

Cos(frame2,frame1)=0.80

Cos(frame2,frame3)=0.57



Video Indexing Experiment

- 25 different videos
 - Computer Science 19, Biology 3, Geology 3
 - 10 different course
 - Average of 75 minutes per video,
 - Total 30+ hours of video
- 3 Manual Ground Truths
 - 1700 transition points

Interface to Identify Ground Truth

icsvideos.uh.edu/ICS/player/indexManual.php?id=526

Index Rating Completion: 100% DI(3):12 PI(2):8 PN(1):1 DN(0):97 ICS Videos - icstesting - GEO-JEO-EUS

Basin Fill

00:00:00 01:12:35

Speed 1x

1__0:1 2__0:7 3__0:13

Basin Fill

Basin Fill

Basin Fill

❖ In this segment we will examine the

❖ In this segment we will examine the

❖ In this segment we will examine the

3: Definitely Index Points,
2: Probably Index Point,
1: Probably Not an Index Point,
0: Definitely Not an Index Point

3 2 1 0 3 2 1 0 3 2 1 0

basin fill

basin fill this segment we will examine the factors that come into play with respect to the architecture of basin fill these include the interactions between

basin fill this segment we will examine the factors that come into play with respect to the architecture of basin fill these include the interactions between

Indexing Accuracy Metric

Manual Ground Truths

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

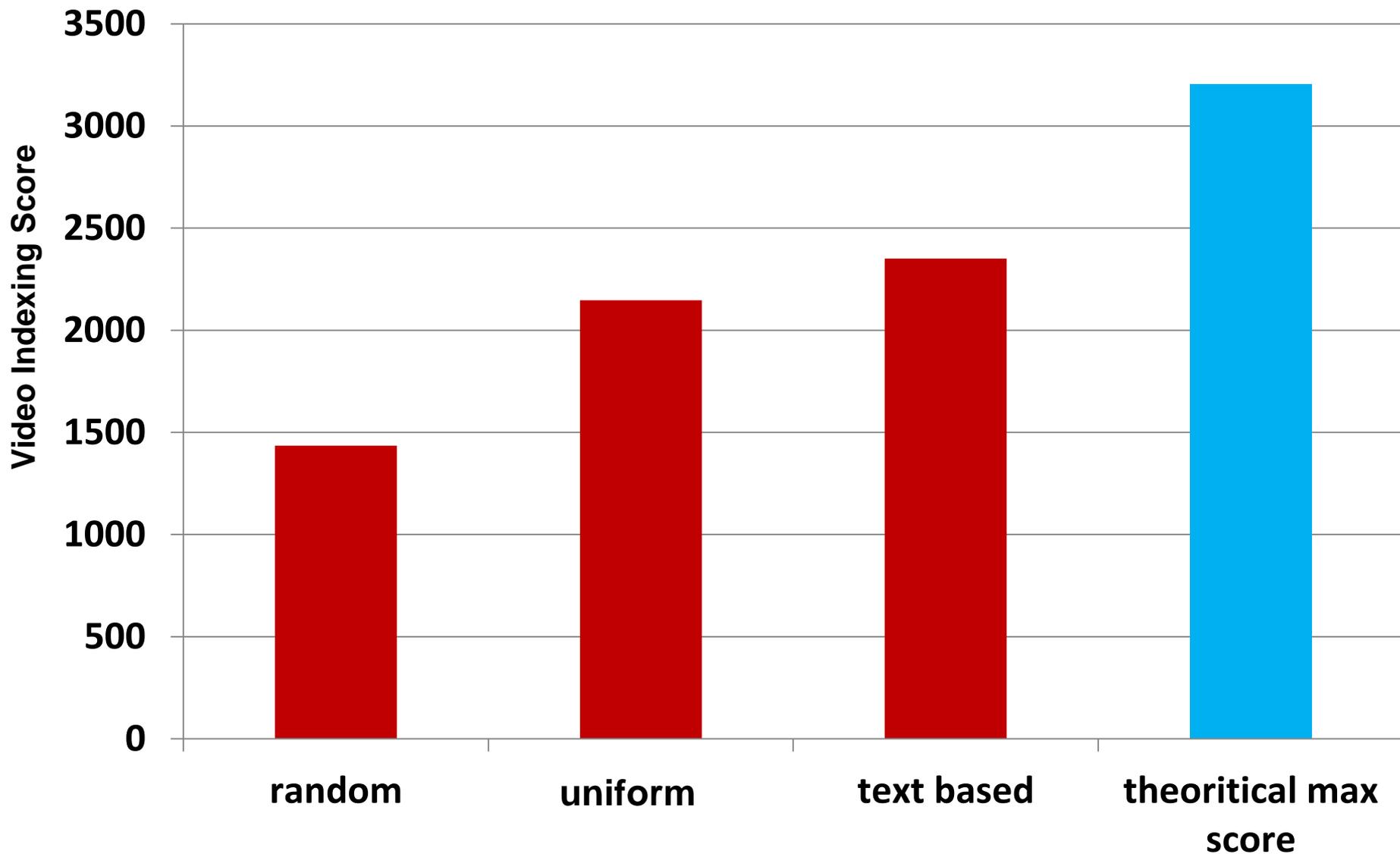
TIS: Transition Indexing Score

n: number of TP in a video

$$\text{Video Indexing Score} = \sum_{i=1}^n (TIS_i)$$

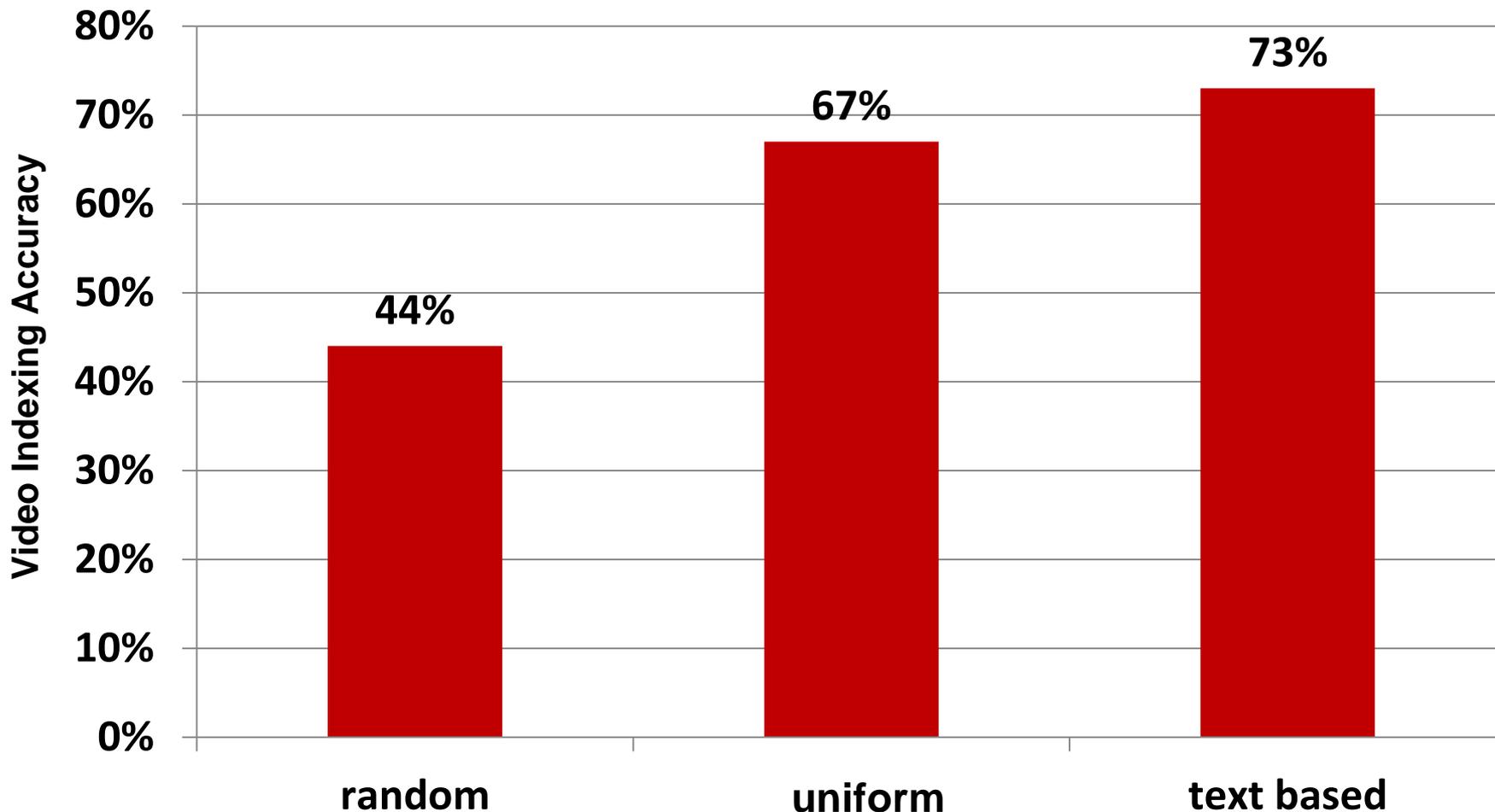
Indexing Algorithms Accuracy

Total Video Indexing Scores for 25 Videos



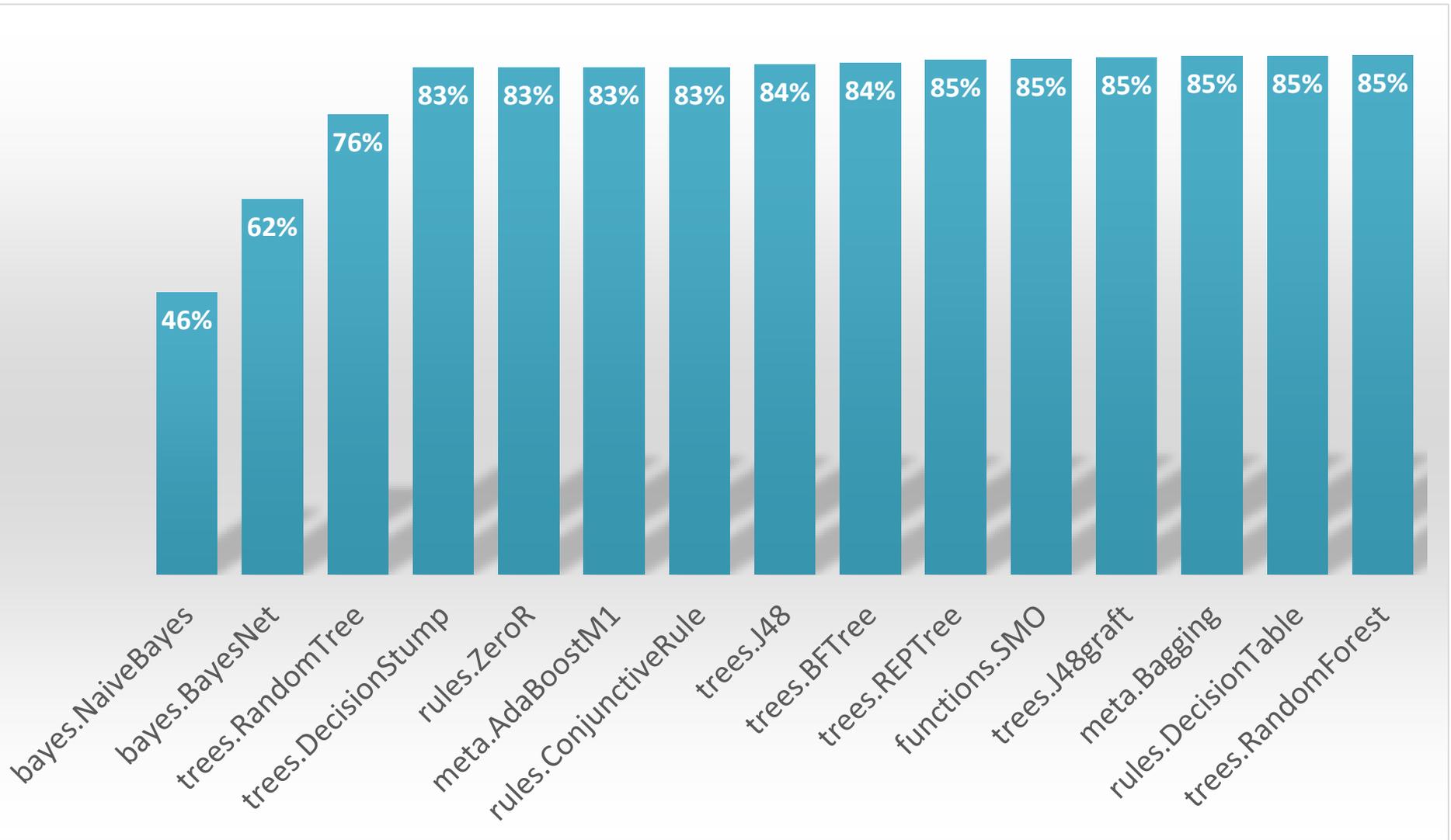
Indexing Algorithms Accuracy

Indexing Accuracy for Total of 25 videos



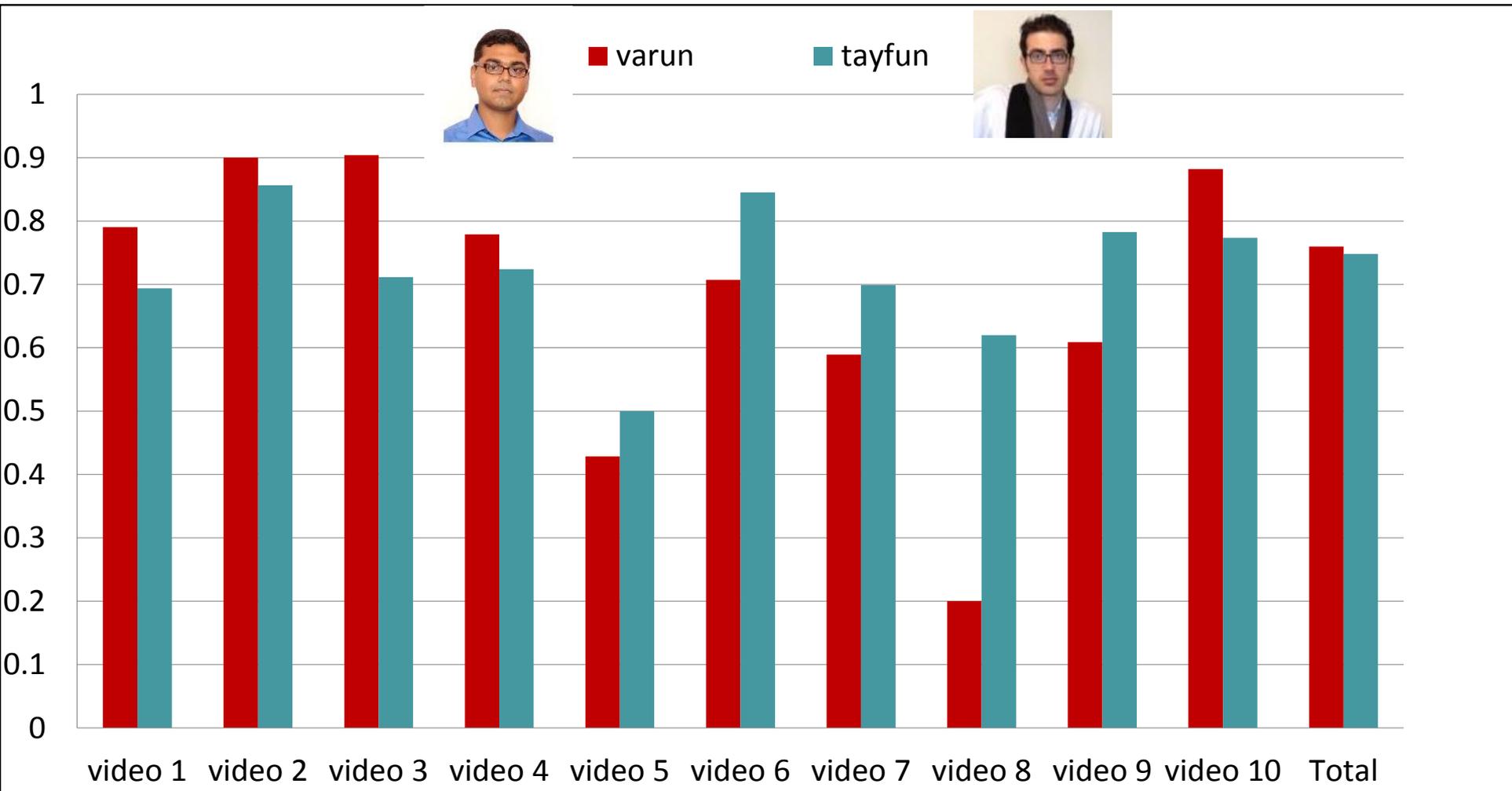
$$\text{Video Indexing Accuracy} = \frac{\text{Video Indexing Score}}{\text{Theoretical Max Score}}$$

Indexing by Machine Learning



- Promising but not fully validated! (Tuna's Ph.D. thesis)

Limits of Indexing: Humans

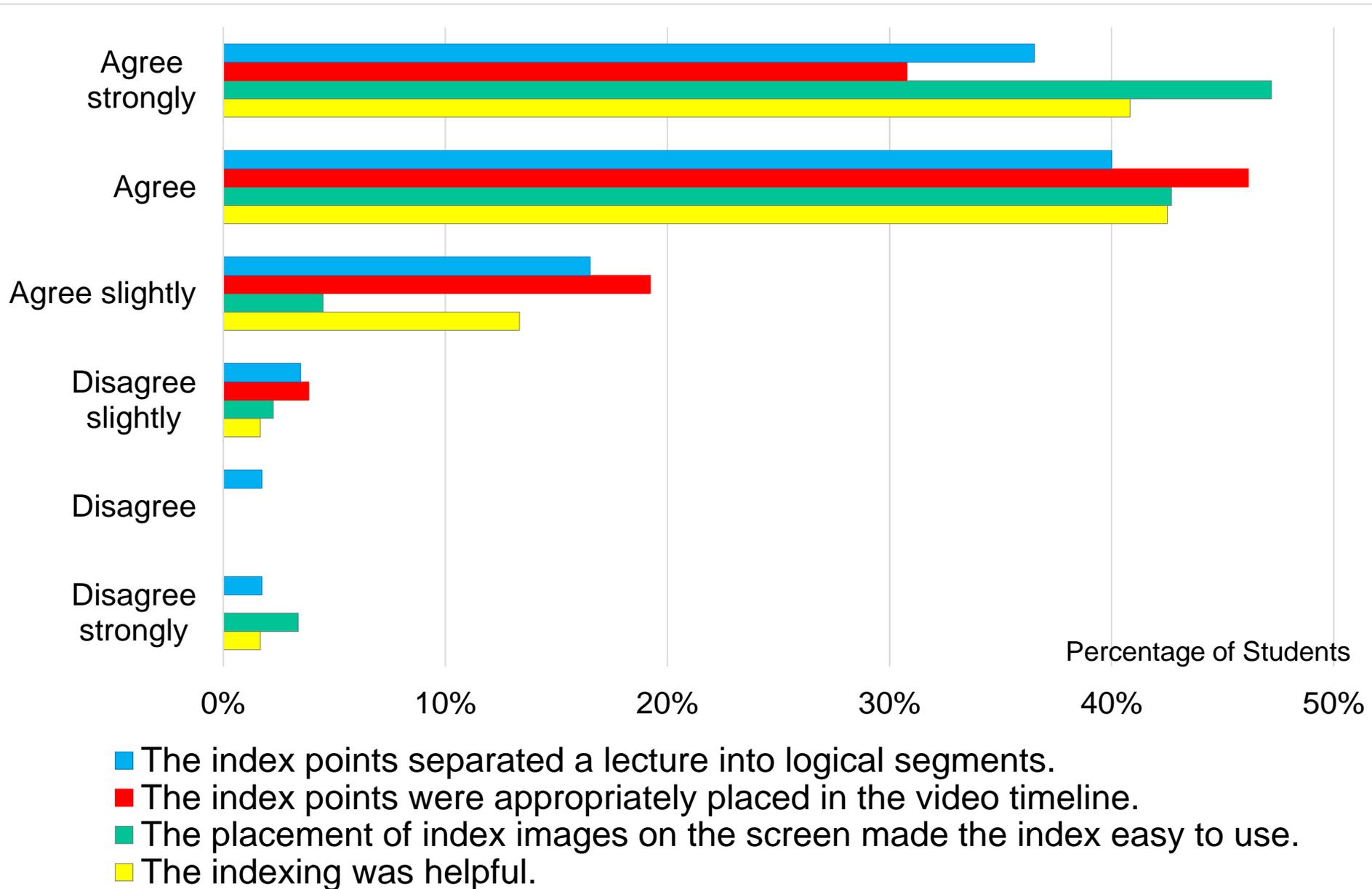


More on Indexing

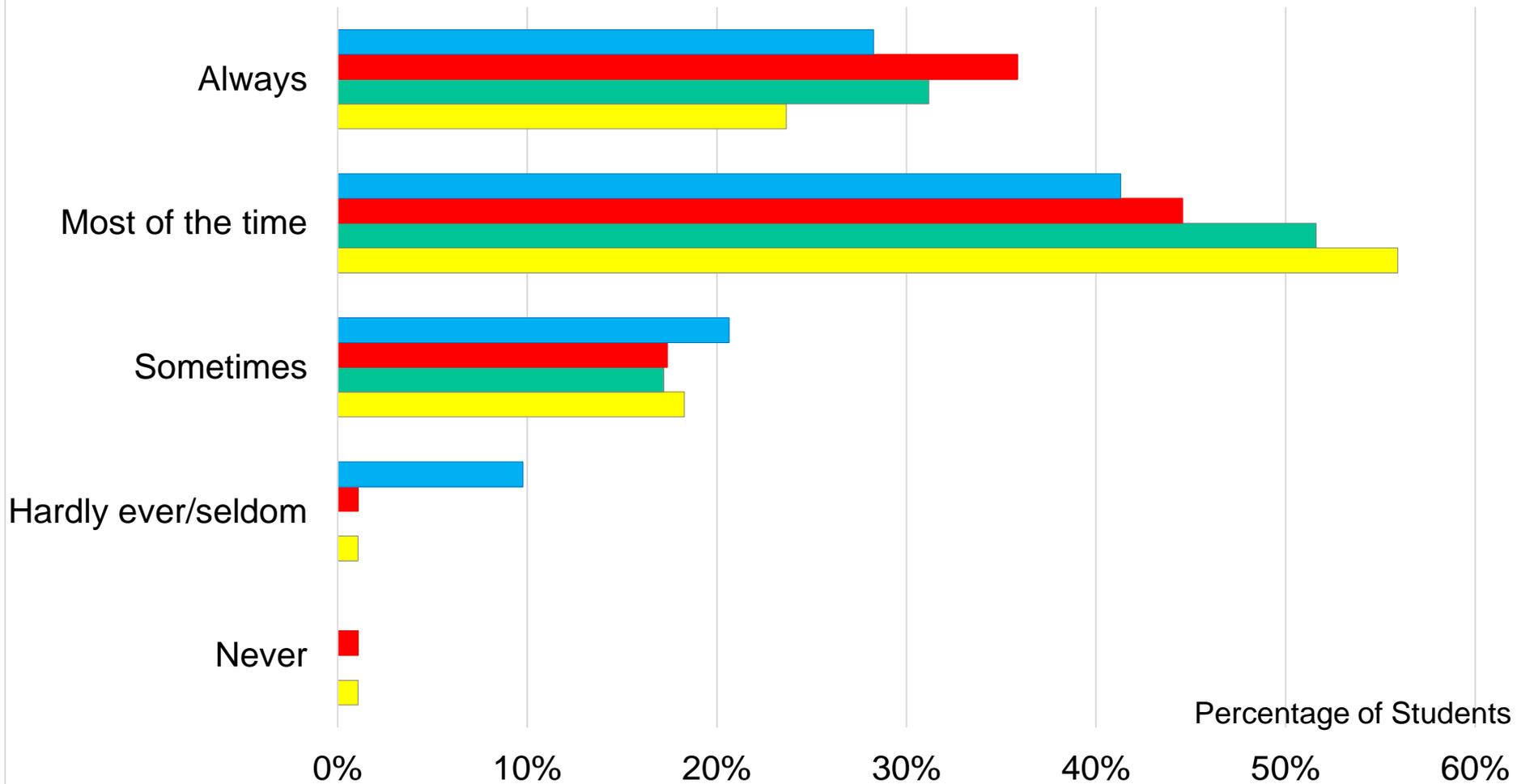
- Employment of Speech-text with slide-text being studied (M. Joshi M.S. Thesis)
- Semi-automatic user driven indexing may be a useful option

Student Surveys of Value of Indexing

Survey – Indexing (1)



Survey – Indexing (2)



- An index point started a new subtopic of the lecture.
- The index functioned well.
- The index made it easy to navigate the video.
- The index provided enough information to allow me to identify the video segment I needed.



ICS Videos: Captions

The screenshot shows a video player interface. The main video area displays a presentation slide titled "Company Products" with a list of items: "Super Computer" and "Telecommunication equipment". Below the list are images of a telephone and a computer system. A red box highlights the caption text at the bottom of the slide: "Their products are supercomputers, telecommunication products, servers, consumer electronics". To the right of the video is a transcript panel with a red border, containing a list of timestamps and corresponding text. The transcript text is: "6:05 So our first presenter today is going to be Steven Nguyen. Steven come on up please.", "6:51 So Steven's going to talk to us about NEC. I did my company", "7:22 on NEC. I did my company on NEC. It stands for Nippon Electric Company.", "7:42 I did my company on NEC. Is that better? It stands for Nippon Electric Company.", "7:53 Their products are supercomputers, telecommunication products, servers, consumer electronics", "8:02 domestic appliance they make everything electronic", "8:12 for their country...for their company. In 1899". The video player controls at the bottom include a speed control set to "1x", a search bar "Search Inside the Video", and a timeline with chapter markers at 00:01, 03:31, 07:57, 11:49, 16:11, 20:37, 29:05, and 38:21.

Computer Science - Introduction to Computing - COSC 1300 Lecture 20

Company Products

- Super Computer
- Telecommunication equipment

rs

Their products are supercomputers, telecommunication products, servers, consumer electronics

00:07:58 01:11:45

Speed 1x

Search Inside the Video

00:01 03:31 07:57 11:49 16:11 20:37 29:05 38:21

Transcript

6:05 So our first presenter today is going to be Steven Nguyen. Steven come on up please.

6:51 So Steven's going to talk to us about NEC. I did my company

7:22 on NEC. I did my company on NEC. It stands for Nippon Electric Company.

7:42 I did my company on NEC. Is that better? It stands for Nippon Electric Company.

7:53 **Their products are supercomputers, telecommunication products, servers, consumer electronics**

8:02 domestic appliance they make everything electronic

8:12 for their country...for their company. In 1899

On screen Caption and Transcript panel

Captioning

- Captions (and transcript) are valuable for students
- Automatic speech recognition is inadequate to generate meaningful captions today
- Goal is to leverage ASR to generate captions efficiently

Automatic Speech Recognition (ASR)

Participant	Lecture Speech Recognition	Dictation	Parroting
Professor A	71.40	89.93	94.76
Professor B	62.14	83.13	96.63
Professor C	70.80	83.03	96.1
Average	68.11	85.36	95.83

Percentage Accuracy

Results shown for the best of 3 ASR tools: YouTube, Windows, Dragon

Dictation: Participants read a prescribed text that is recorded

Parroting: A trained speaker repeats the text

Normal Lectures cannot be transcribed with ASR today



Caption Editor

Motivation: A mechanism to provide captions for STEM classroom lectures.

Design Objectives/Challenges:

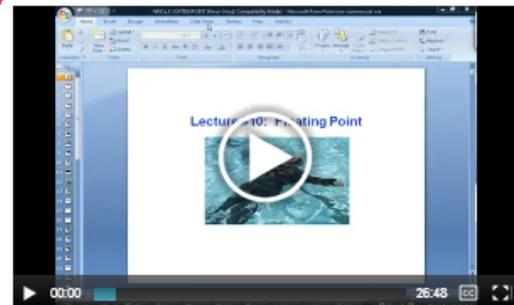
- *Teamwork:* Allow co-operative/crowdsourced captioning
- *Technical content+Poor audio quality:* Captions/transcript may not be obvious
- *Easy navigation:* for non-professionals.

Caption Editor

Welcome rucha! [Logout](#)

ICS Caption Editor

[Help](#) | [Watch Video](#)



Legend

- Section currently being edited
- Caption is Complete
- Caption Needs a Review

PlaySpeed:

0 Completed!, 0 Requested for Review, 14 remaining.

Section No.	Start Time (mm:ss)	Caption Text	Save my Changes	Status	Review Count
Edit Section 1	0:05	Lets talk about floating point today.We have seen the ways of representing numbers in binary so far	<input type="button" value="Save"/>	Needs a Review <input type="button" value="Mark as Completed"/>	
	0:13	the largest we've gotten with the unsigned integers we preferred edited by Rucha	<input type="button" value="Save"/>	Complete <input type="button" value="Request a Review"/>	1
	0:18	two to the n minus one ending number in bits you have represent and with sign integer the two complement we have negative two to the n minus one	<input type="button" value="Save"/>	<input type="button" value="Mark as Completed"/>	
	0:26	two to the n minus one minus one because one less positive number than the negative number. So the question you might	<input type="button" value="Save"/>	<input type="button" value="Mark as Completed"/>	
	0:37	represents a very large numbers or even decimal number, rational number like pi or the natural number e. The way to do this	<input type="button" value="Save"/>	<input type="button" value="Mark as Completed"/>	
Edit Section 2	0:50	to go back and look at scientific notation base tenth	<input type="button" value="Save"/>	<input type="button" value="Mark as Complete"/>	
	0:56	things to note the number before the decimal point were call the mantissa	<input type="button" value="Save"/>	<input type="button" value="Mark as Complete"/>	
	1:04	the radix or the base that you're in so we're are talking about base 2 that would be 2 and an exponent number	<input type="button" value="Save"/>	<input type="button" value="Mark as Complete"/>	
	1:15	multiple ways to represent this number, of course, like we have point 1 and ten to the negative ten. However, to make sense you want to have stuffs in normalise form	<input type="button" value="Save"/>	<input type="button" value="Mark as Complete"/>	
	1:27	based ten that just means you have one number before the decimal point and that is also true with binary	<input type="button" value="Save"/>	<input type="button" value="Mark as Complete"/>	



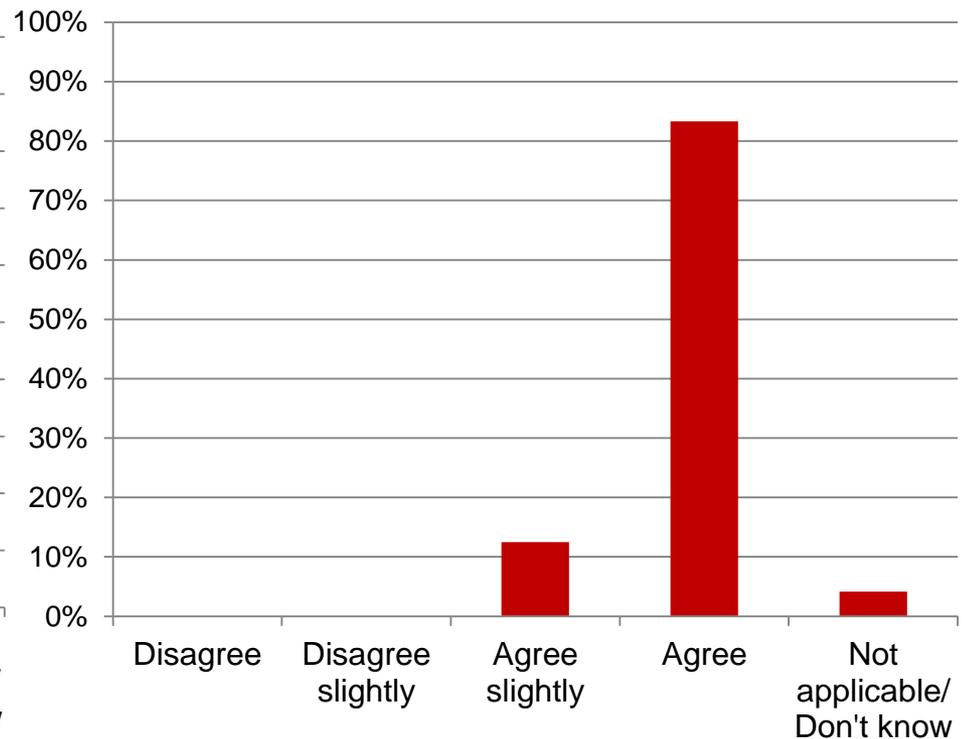
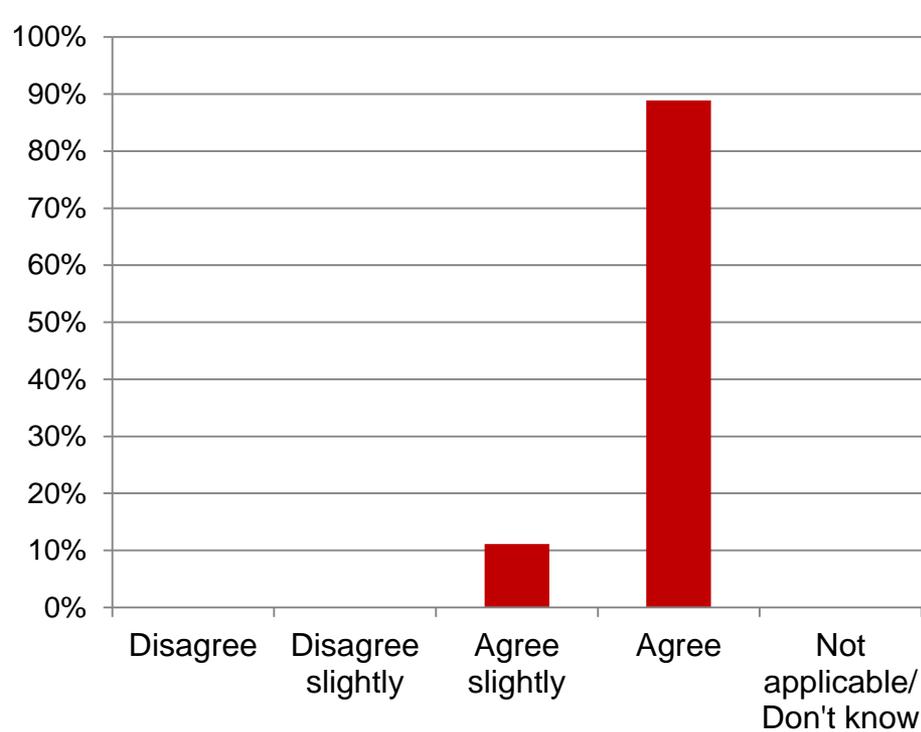
Caption Editor: Evaluation

- Deployed across 2 Computer Science courses to successfully caption selected lectures
- Work distributed across participating students.
Typical example: one lecture with 11 participants
 - Each student worked on between 9 and 52 segments
 - Spent between 10 and 76 minutes over 4 days

(hour long lecture takes around 8-10 hours to caption by an untrained person)



Captions Helpful & Valuable



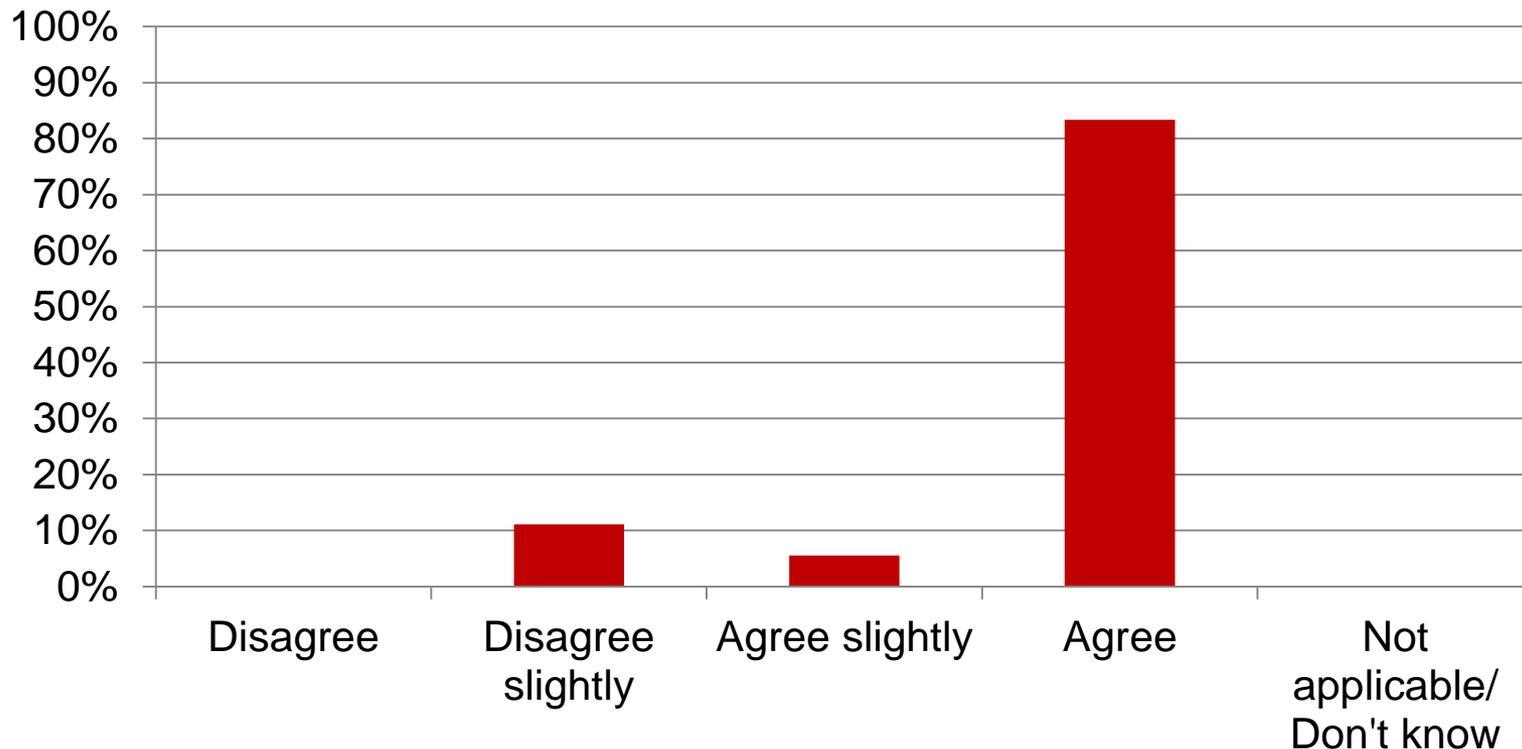
Q:The captions and transcript helped me understand what the professor was saying

Q:The videos with caption (& transcript) are preferable than videos without them

Most helpful for: **Learning, Attention, Efficiency, Note-taking**
Somewhat helpful for: **Motivation, Quiz Performance**

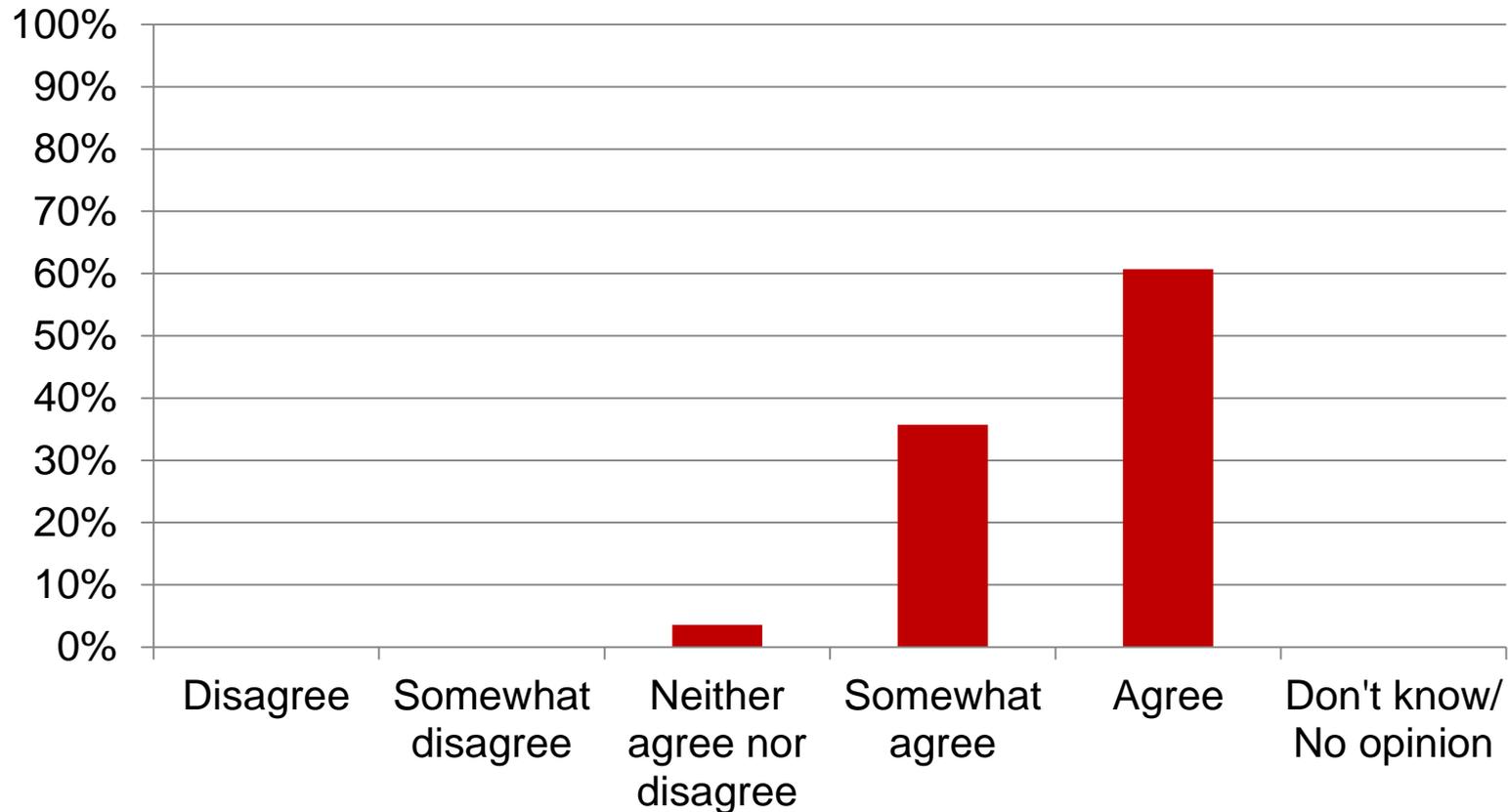


Caption Editor: Evaluation



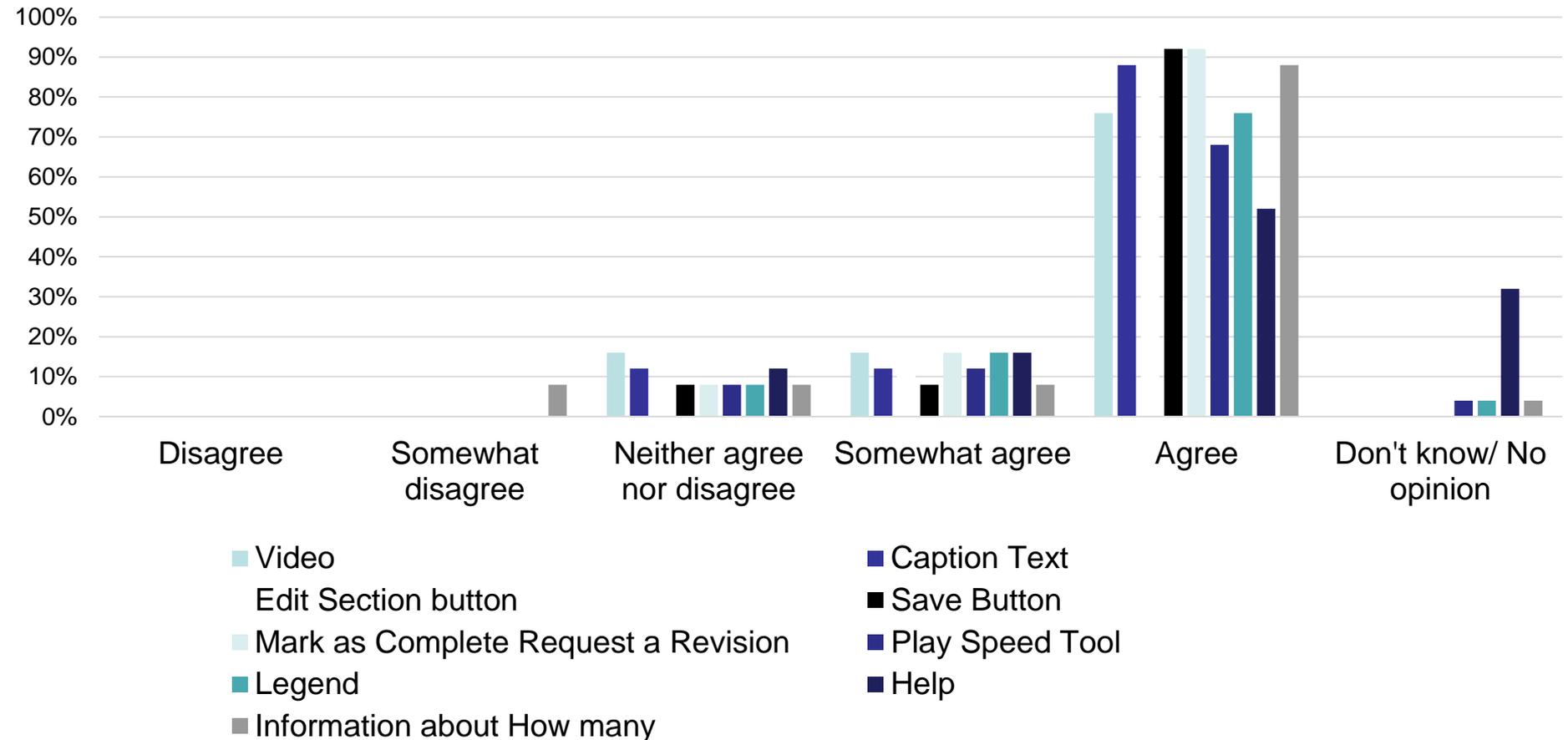
Question: The captions and the transcript represented accurately what the professor said

Caption Editor: Evaluation



Question: The ICS Caption Editor is easy to use.

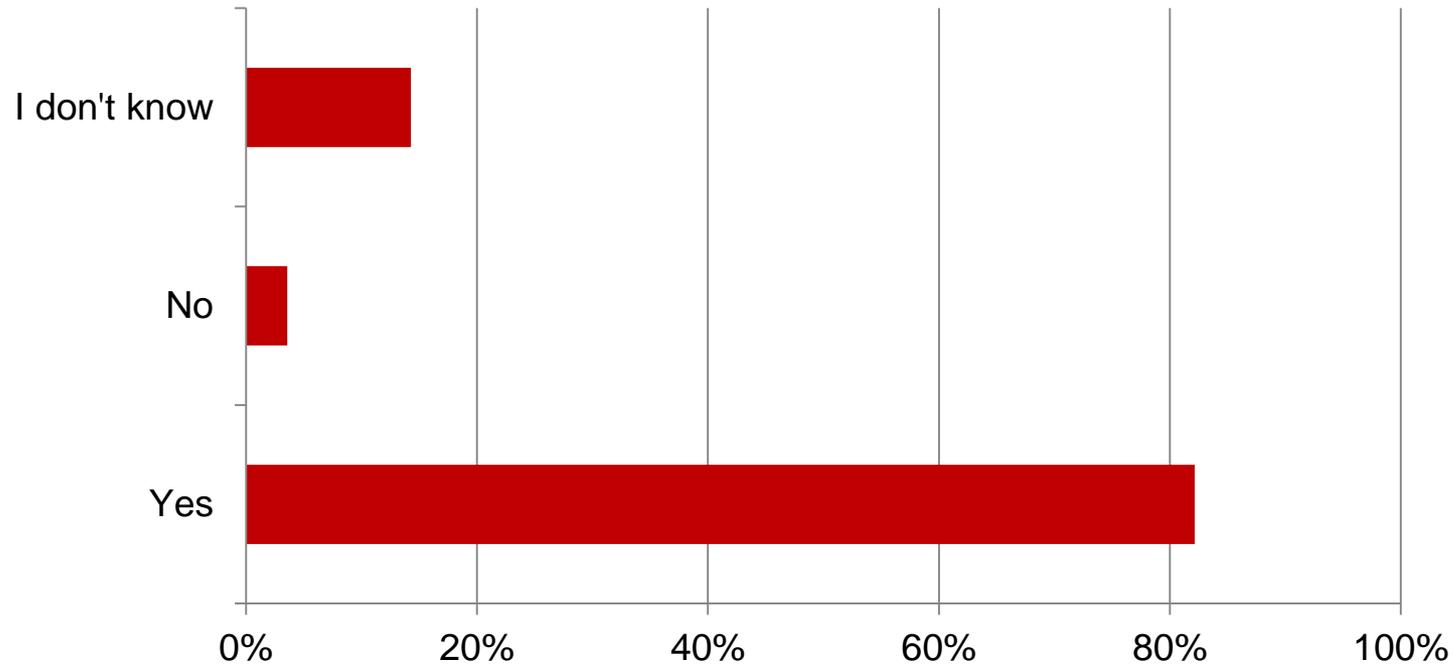
Caption Editor: Evaluation



Question: The placement (position) of the following elements and controls on the Caption Editor interface was appropriate?



Caption Editor: Evaluation



Would you be interested in working with other students to correct captions for your class lectures using this caption editor if you receive some incentive (for example academic credit)?



ICS Videos: Usage and Experience

Employed in over 100 class sections at UH over the past 5 with many 1000s of student users total!

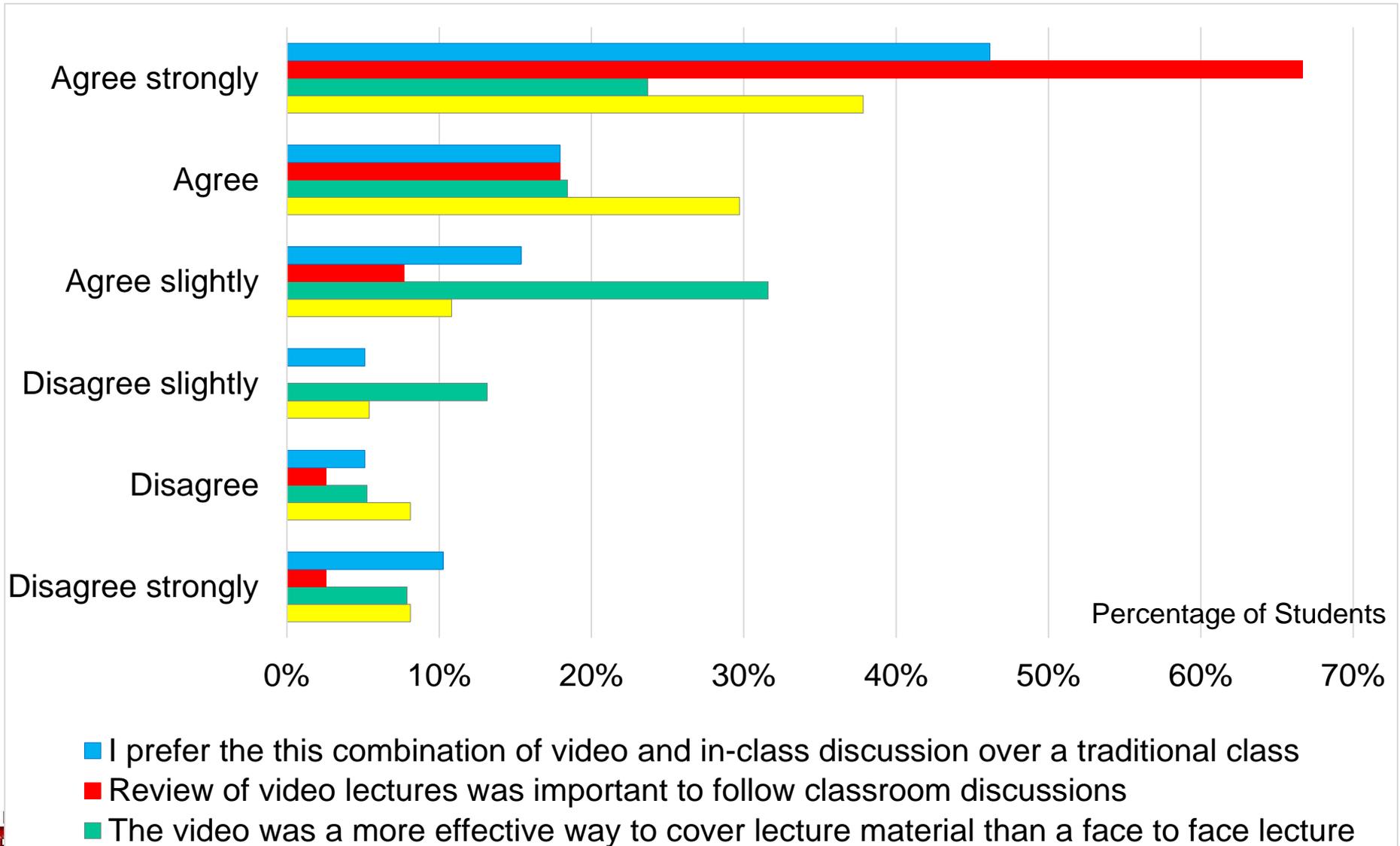
Freely available

Recently being employed as the technology to enable flipped classrooms (UH TIP Award 2014)



Inverted Classroom

Dr. Leigh Leasure – Physiological Psychology



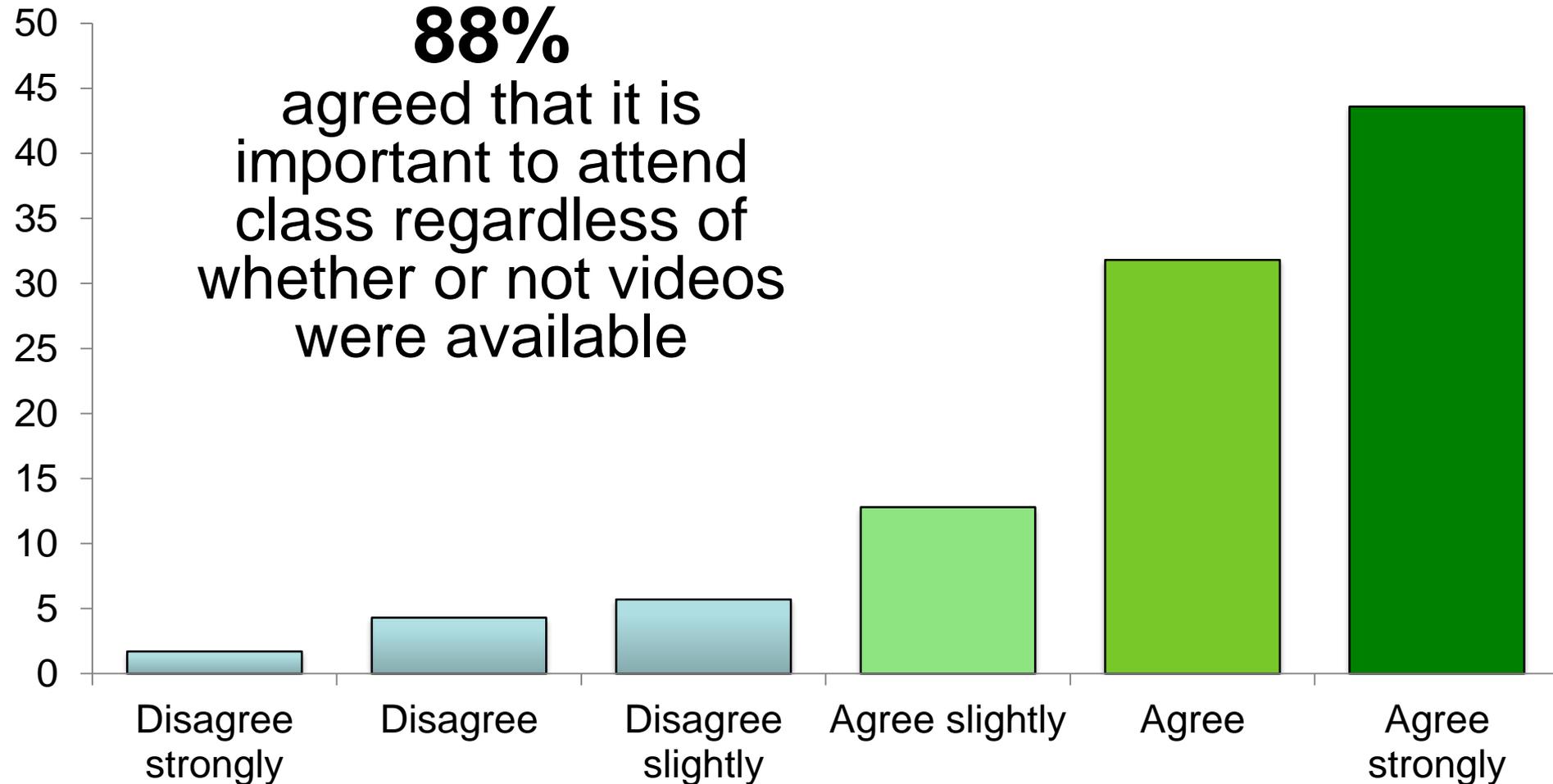
Videos & Attendance

Students who watched at least one video had higher attendance rates than those who did not watch any videos.

Video use and **attendance** were slightly **positively** correlated

Videos & Attendance

88%
agreed that it is
important to attend
class regardless of
whether or not videos
were available



Survey Comments

Comments in surveys and focus groups were positive beyond expectations (& very satisfying)

“I can tell that I learned 80% of the class by videos”

“This is a great tool for us as a student to go over everything. I would hope that this would be available in all classes”

“Indexing enabled me to jump directly to my trouble spots”.

“I’m used to listening and then looking at the caption because it helped me tremendously to learn the new words and vocabulary”

“The search feature function have instant results, like a Google search box”

Conclusions (sort of)

Classrom Videos are an important learning companion (akin to a textbook)

- Automatic methods are effective at making video content more accesible
- CS research challenges remain
 - Smarter/Semantic Indexing and Search
 - Merging of audio, text and image information
 - HCI for a better learning experience
- Ongoing efforts to make the technology widely available

Co-conspirators

CS Faculty: Zhigang Deng, Olin Johnson, Shishir Shah, Rakesh Verma, Christoph Eick

Students: Tayfun Tuna, Varun Varghese, Mahima Joshi, Tuhin Dey (Education)

X-Students: J. Li, C. Yun, G. Bhatt, T. Tuna. A. Verma, R. Kushalnagar, Rucha Borgaonkar,

NSMIT Staff: S. Baez-Franseschi, Pradeep Krishnan, Andrea Arias

Assessment: Lecia Barker (UT Austin), Yumei Liu, Chris Hovey (NorthEastern)

Deployment, usage and assessment

UH Computer Science, UH Geosciences, UH Biology and Biochemistry

UH Downtown (Richard Alo), Texas School for the Deaf (David Coco)



Last Slide

- ICS Videos System

www.icsvideos.uh.edu

- Username: student
- Password : icsstudent

- Contact email:

icsvideoscontact@gmail.com

or jaspal@uh.edu

