

# *Augmented Segmentation and Visualization for Presentation Videos*

Alexander Haubold and John R. Kender  
Department of Computer Science  
Columbia University, New York  
{ahaubold,jrk}@cs.columbia.edu

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Motivation

- Videos of student team presentations
- 1 semester  $\approx$  160 students, 30 teams, 8 hours of video for midterm presentations
- How to best review?
- Need automatic index for videos
- Need visual browser for searching

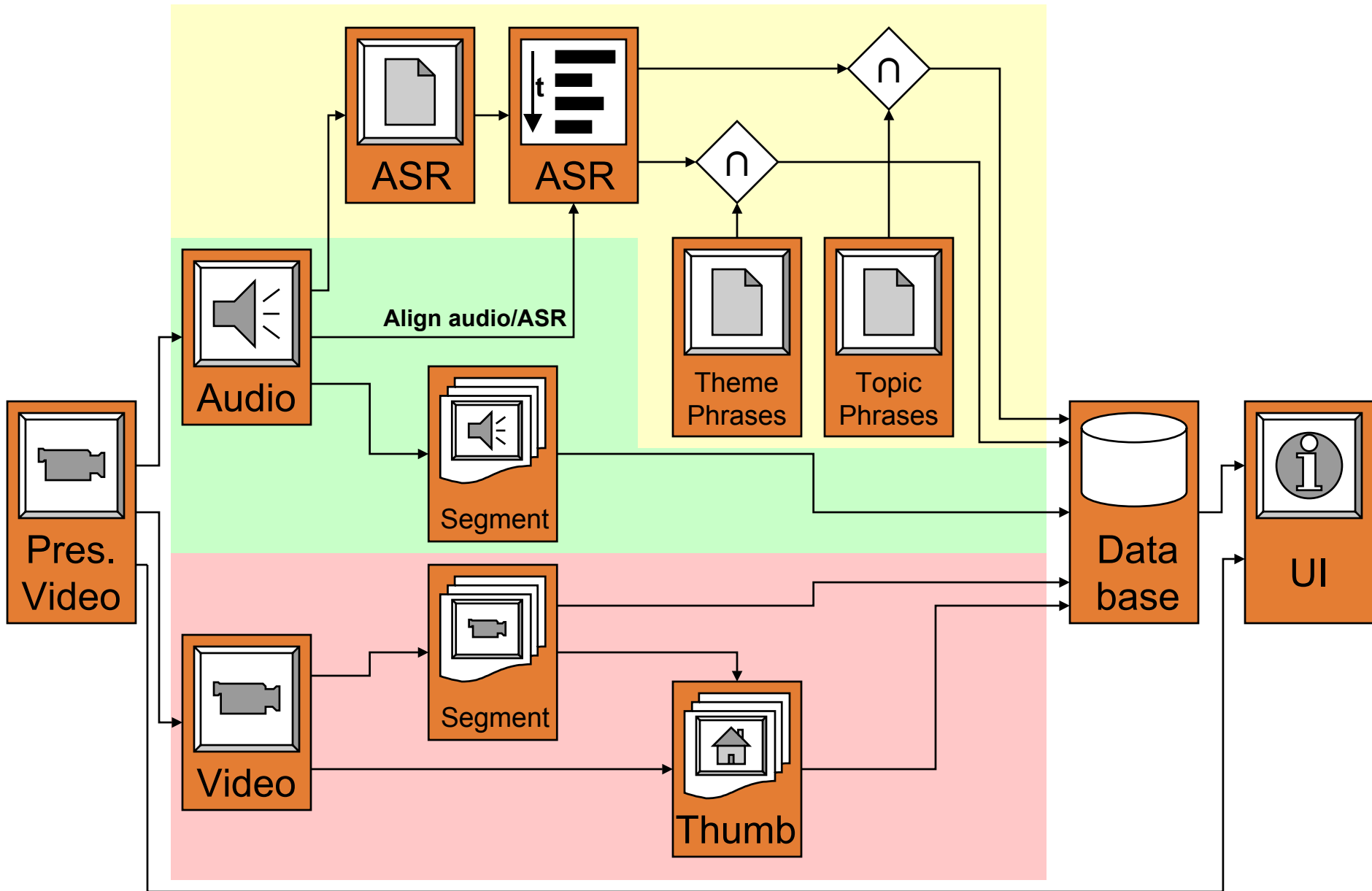
# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# *Characteristics*

- Multiple speakers:  $\approx 5$  / team,  $\approx 20$  / hour
- Not professionally recorded or edited
- Lighting conditions vary
- Long shots without distinct visual cuts
- Audio quality varies (handling of microphone)
- But: known structure of thematic sections

# Characteristics



# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations



# Segmentation (Audio)

- Identify audio segments for each student
- MFCCs for representing features of speech
- Bayesian Information Criterion detects speaker changes
- Results encouraging, even for varying audio quality

Precision	Recall	# Segments
88.5%	95.7%	395

# Segmentation (Visual)

- Boundaries from non-overlapping sources:
  - Presentation slide changes
    - Not all presentations have slides
  - Speaker gesture changes
    - Long-term change in speaker pose
    - Reconfiguration of speaker position
    - Amount of gesture

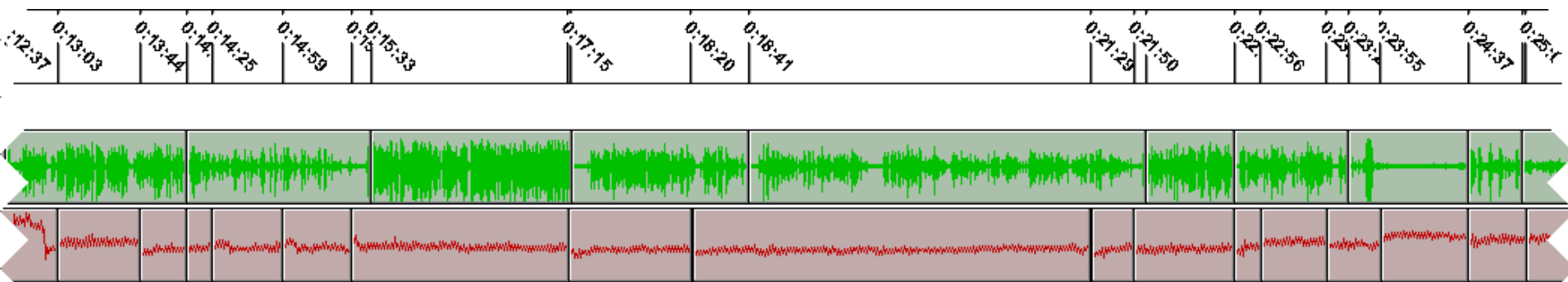
Precision	Recall	# Segments
89.4%	82.7%	594

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Segmentation (Both)

- Combination of audio and video cues results in more natural segmentation
  - Not every speaker change is accompanied by visual change, and vice versa
  - Presentation Unit: Union of A/V change



# Segmentation (Both)

Precision	Recall	# Segments
89.3%	92.7%	710

- Compare to separate segmentations w.r.t. presentation units:

	Precision	Recall
Audio	51.3%	53.2%
Video	66.6%	69.2%

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# *Text Augmentation*

- ASR transcript from IBM<sup>®</sup> ViaVoice<sup>®</sup>
  - Poor audio quality
  - No training (would require 160 / semester)
  - Word Error Rate of 75%
- Apply 2 filters
  - Manually assembled list of “theme phrases”
    - Phrases / titles of required sections
  - Automatic list of “topic phrases” from presentation slides (if available)
    - Appear in presentation AND transcript

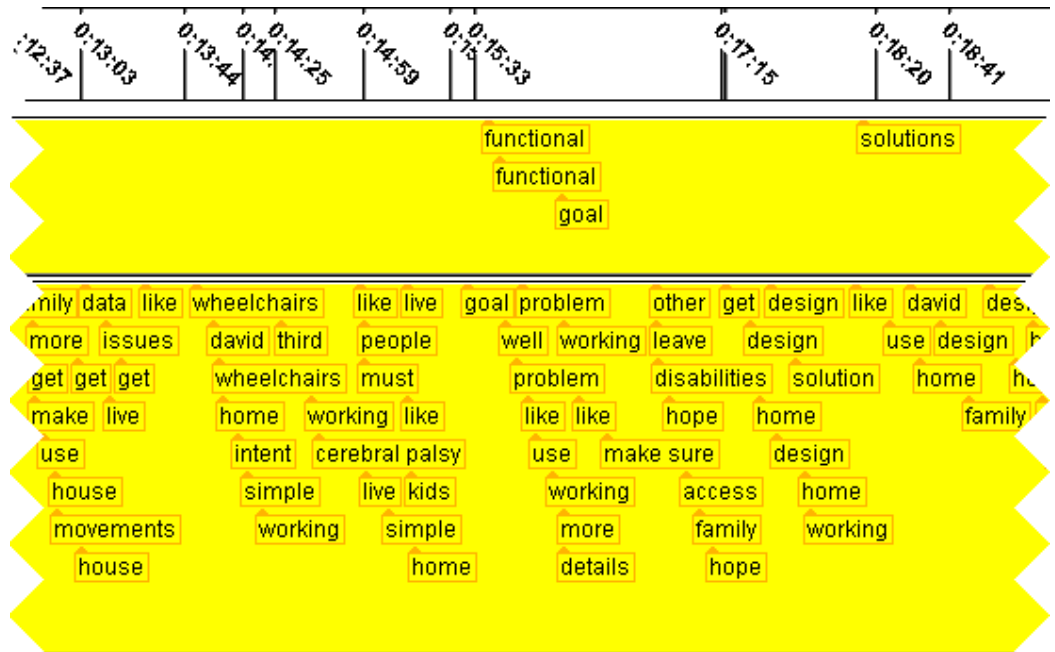
# Text Augmentation

## Theme Phrases:

Alternative solutions	Objective tree	Background	Functional	Prototype
Continuity Plan	Problem statement	Chart	Future	Requirements
Design Constraints	Project goals	Constraints	Goal	Schedule
Functional Requirements	Tasks performed	Continuity	Implementation	Solutions
Future directions	Team process	Deliverables	Limitations	Statement
Gantt chart	Team development	Demo	Objective	Tasks

Theme Phrases →

Topic Phrases →





# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Interface

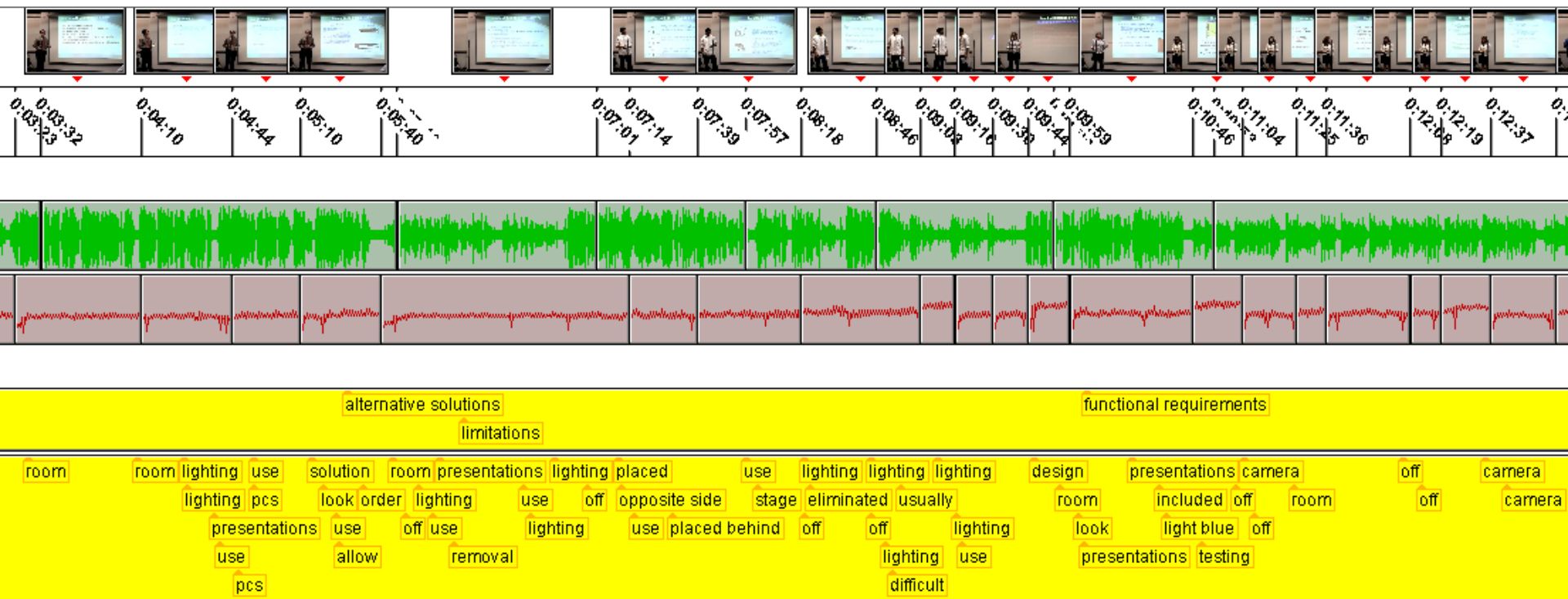
- List of Videos
  - Zoomable Summary →
  - Video Playback
- Thumbnails
  - Timeline
  - Audio, video tracks
  - Text tracks

The screenshot displays a video player interface with the following components:

- Video List:** A sidebar on the left showing a list of video presentations, including titles like "E1112 Section 4 (Part 1) Midterm Spring 2005 Presentation" and their durations.
- Selected Video:** The main area shows the selected video with a "Zoom Frames Per Pixel" control set to 21. It includes checkboxes for "Video", "Audio", "Timeline", "Thumbnails", "Subtitles", and "Text Tracks".
- Timeline:** A horizontal timeline at the top of the main area with markers for various time points.
- Audio and Video Tracks:** Below the timeline are tracks for audio (green waveform) and video (red waveform).
- Text Track:** A large yellow area at the bottom displaying a text track with words and phrases such as "functional solutions", "functional requirements", "implementation", "solutions", "tasks", "goal", "deliverable", "ideas get group develop children child able question needs build removing designing swing swing designing swing problems tasked goal question far", "last semester problems number wheelchair wheelchair question number able designing making ideas starting regular build for used", "group designing able swing number comfort operate child easy making height children build used preliminary research", "fun preliminary research able needs able operate designing hearing build child used needs", "light designing used number requirements needs dimensions simulation secured used designing", "group swing wheelchair simulation wheelchair making designing used people", "being transferred", "wheelchair making designing used people", and "determine".
- Streaming Video:** A small window at the bottom left shows a live streaming video of a presentation titled "Wheelchair Swing Team A".

# Interface: Timeline

- Portrait notebook-style not well received
- Re-modeled to horizontal continuous timeline



# Interface: Text Graph

- Zoomable interface distributes text
- 10 minutes
- Deeply nested text
- 1.5 minutes
- More precise browsing

The screenshot shows a zoomable interface with a timeline at the top and a dense network of terms below. The timeline includes segments with timestamps: 0:41:40, 0:42:40, 0:43:24, 0:44:23, 0:45:54, 0:48:08, and 0:50:08. The terms are arranged in a hierarchical, nested structure, with some terms highlighted in yellow boxes. The terms include: objective tree, future, functional requirements, functional, goal, functional, constraints, future, constraints, goal, limitations, future, limitations, continuity, staff, way, prices, getting, will, will, letter, room, looking, decide, divided, will, desk, learn, re, ss, increase, visual, number, meetings, usefulness, will, considering, looking, redesign, f, room, back, prices, put, getting, replaced, aged, some, room, know, make, fill, some, ar, possible, provides, room, needed, good, possible, computer, lab, know, staff, design, usefulness, room, condition, will, some, project, needed, visit, will, planned, s, h, usefulness, small, number, planned, will, will, getting, learn, will, planned, beginn, function, fill, fill, some, looking, prices, present, know, will, visit, see, furniture, function, will, room, room, work, learn, usefulness, possible, design, some, way, includes, function, some, small, design, usefulness, cla, usefulness, possible, walls, were given, learn, possible, beginning, ness, size, windows, design, number, incorporates, will, lab.

The screenshot shows a zoomable interface with a timeline at the top and a sparse network of terms below. The timeline includes segments with timestamps: 0:44:23, 0:45:43, and 0:45:54. The terms are arranged in a hierarchical, nested structure, with some terms highlighted in yellow boxes. The terms include: constraints, functional requirements, limitations, needed, fill, will, will, planned, lack, reviewed, room, room, room, pos, meetings, will, replaced, room, some, usefulness, ag, includes, some, furniture, design, will, walls, design, lot, good, will, looking, planned, function, letter, room.

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# User Study

- 176 students, mostly appearing in videos
- Questions answered using UI

Find your appearance during presentation
Find beginning of your team's presentation
Find you team's discussion on topic X
Find presentation Y (Y of different team & class)
Summarize segment using only text

- 1/2 students: summaries + video playback
- 1/2 students: only summaries

# *User Study: Results*

- Video + Summaries vs. Summaries only
  - Overall same accuracy
  - 20% less time spent without video
  - But: no comparison to linear search (VCR)



# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# Conclusions

- System
  - External structure of contents important
    - Apply and visualize in browser
  - Zoomable text requires ranking (structure)
- User
  - Thumbnails good: focus on task
  - Video bad: easily sidetracked

# Overview

- Motivation
- Characteristics of Presentation Video
- Segmentation (Audio, Visual)
- Segmentation (Combined Audio-Visual)
- Text Augmentation
- Interface
- Demo
- User Study
- Conclusion
- Future Investigations

# *Future Investigations*

- Active displays
  - What you see on UI must be clickable
- Topological grouping
  - Temporally group similar audio/visual sources
- Speaker gesture
  - Classification and labeling of speakers
- Annotation tool
  - Instructors / students annotate presentations

*Thank you!*

*Questions / Answers?*