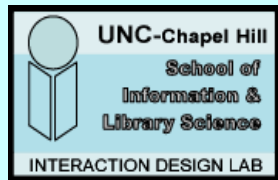
Open Video R&D Case

INLS 818



Open Video Vision/Contributions

- An open repository of video ***files*** that can be re-used in a variety of ways by the education and research communities
  - Encourages contributions
  - A testbed for interactive interfaces
- An easy to use DL based upon the *agile views interface design framework*
  - Multiple, cascading, easy to control views (pre, over, re, shared, peripheral)
  - Views based upon empirically validated surrogates
  - An environment for building theory of human information interaction
- A set of methods and metrics that reveal how people understand digital video through surrogates

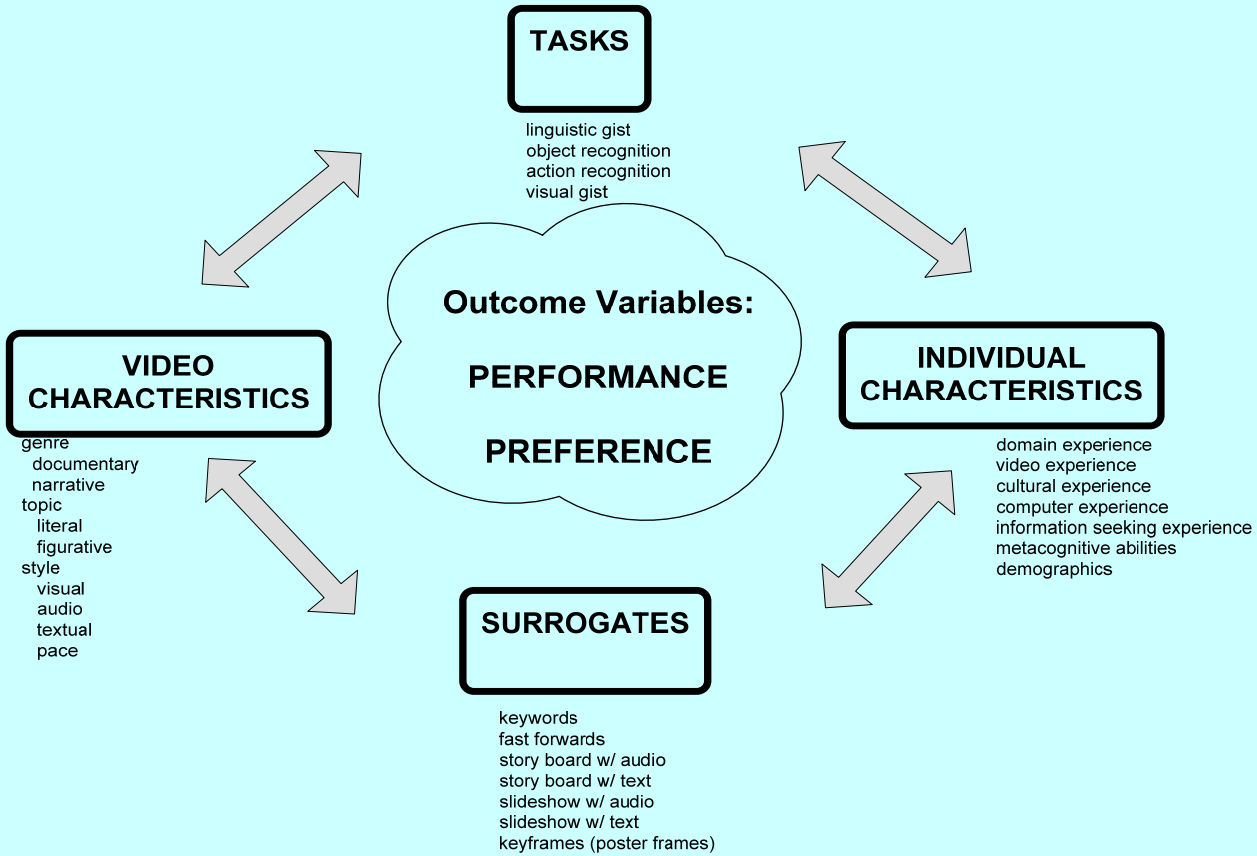
Background & Status

- Begun 1995 with colleagues at UMD & BCPS
- Funding: NSF# IIS-0099538 (2000-04); IIS-0455970
  - NASA, IBM, Google
- Collaborators/Contributors: I2-DSI, ibiblio, CMU, UMD, NIST, Internet Archive, NASA, ACM
- ~4000 video segments
- ~30000 unique visitors per month (32,000 in Oct 06)
- Ibiblio and Google file hosting yields more activity
- MPEG-1, MPEG-2, MPEG-4, QT
- OAI provider
- Ongoing user studies and spin offs

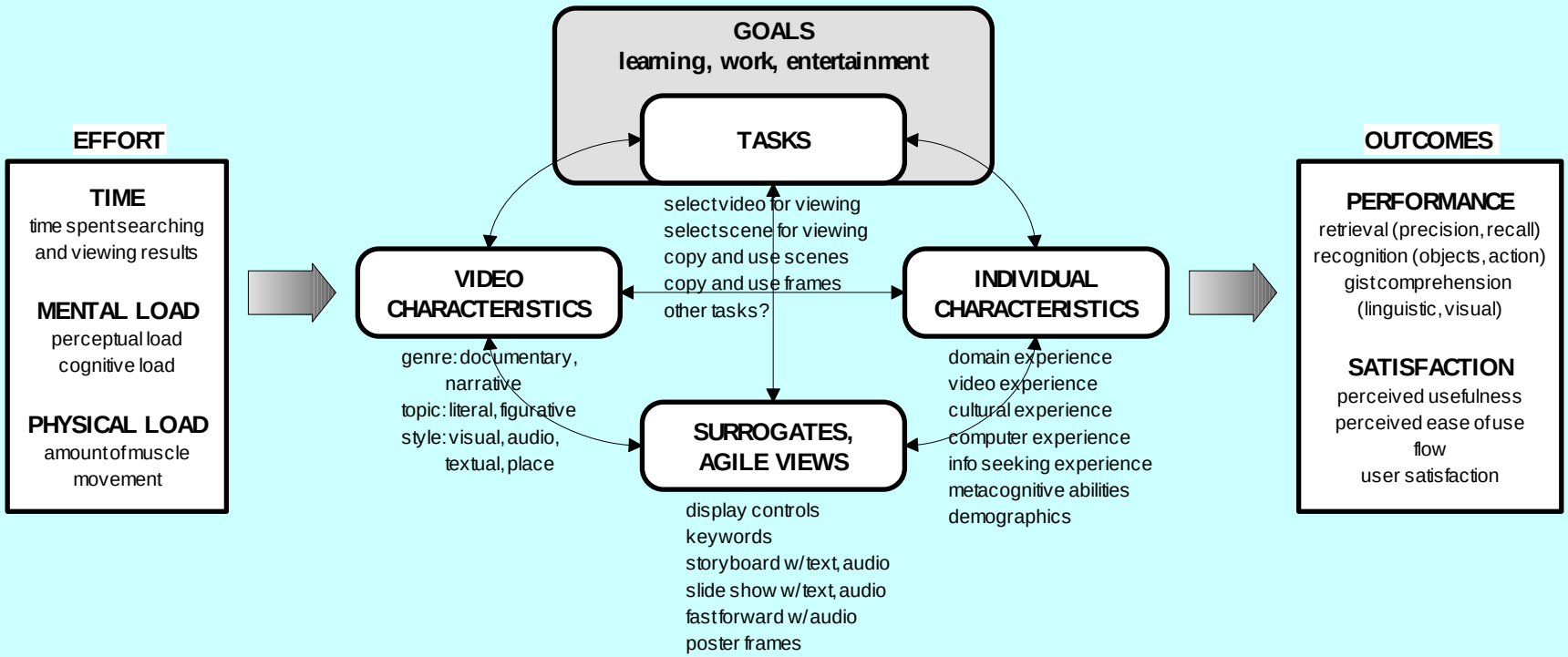
Agile Views Interface Research

- Provide a variety of access representations (e.g., indexes) and control mechanisms
- Usual search and browse capabilities
- Leverage both visual and linguistic cues
- Create and test surrogates for overview preview, shared and history views

User Study Research Framework



Revised User Study Framework



THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

The Surrogates

- Storyboard with text keywords (20-36 per board@ 500 ms)
- Storyboard with audio keywords
- Slide show with text keywords (250ms repeated once)
- Slide show with audio keywords
- Fast forwards 32X, 64X, 128X, 256X
- Poster frames
- Real time clips (7 sec. excerpts)
- Text titles
- Spoken descriptions, keywords
- Audio abstractions?

Surrogate Examples

Type of surrogate	Examples
Text surrogate	Title, keyword, description, etc.
Still image surrogate	Poster frame, storyboard/filmstrip, slide show, video stream, key-frame-based table of contents, etc.
Moving image surrogate	Skim, fast forward, etc.
Audio surrogate	Spoken keywords, environmental sounds, music, etc.
Mutlimodal surrogate	Text surrogate + still image surrogate, still image surrogate + audio surrogate, etc.

Metrics

	Text	**Still image**	**Action**
Recognition	Object recognition (text)	Object recognition (graphical)	Action recognition
Inference	Gist determination (free text) Gist determination (multiple-choice)	Visual gist determination	
Affect	Learnability, usability Enjoyment, engagement		

User Studies

- Qualitative Comparison of Surrogates (ECDL 02)
- Fast Forwards (JCDL 03)
- Narrativity (CHI 02)
- Shared views and History Views (Geisler dissertation)
- Poster frames and text (eyetracking, CIVR 03)
- TREC evaluation
- Hughes MP, Gruss MP
- Relevance judgments (Yang dissertation)
- ISEE collaboration (Mu dissertation)
- Redesign effects and integration of surrogates in AV
- Spoken description vs SB vs combined

Study 1: Compare Surrogates

- What are the strengths and weaknesses of different surrogates from the users' perspective?
- Are any of the surrogates better than the others in supporting user performance?

The Surrogates

- Storyboard with text keywords (20-36 per board@ 500 ms)
- Storyboard with audio keywords
- Slide show with text keywords (250ms repeated once)
- Slide show with audio keywords
- Fast forward (~ 4X)

Method

- 7 video segments (2-10 min), 5 surrogates created for each
- 10 subjects with high video and computer experience
- Three phases (all multi-camera videotaped)
  - View full video then use 3 surrogates, repeat
    - Participant observation and debriefing
  - Do NOT view full video, use 3 surrogates, repeat
    - Participant observation and debriefing
  - Complete 3 assigned tasks with surrogates of choice
    - Think aloud and debriefing
- http://www.open-video.org/experiments/chi-2002/methods/study1.mov

Tasks

- Gist determination—free text
- Gist determination—multiple choice
- Object recognition—textual
- Object recognition—graphical
- Action recognition (2-3 second clips)
- Visual gist (predict which frames belong)
  - http://www.open-video.org/experiments/chi-2002/surrogates/index.html

Preferences

- In debriefing after each phase, subjects asked about preferences.
- Some preferences changed over the phases
- 2 subjects preferred ff
- 4 subjects said ff if audio keywords added
- 1 storyboard with audio keywords
- 2 slide show with audio keywords
- → drop ss with text keywords, develop ff

Performance

- No SRD on gist (both free text and multiple choice)
- SRD on action recognition favoring ff
- ‘Near’ SRD on text object recognition favoring SB/w audio keywords
- 8:1 to 29:1 compaction rates suitable for tasks
- Psychometric and face validity support for the tasks (means and variances; relevant to real tasks)
- SRD in gist and visual gist for one video
  - →Homogeneity of frames diminishes surrogate value
  - →Keywords help when visual variability decreases

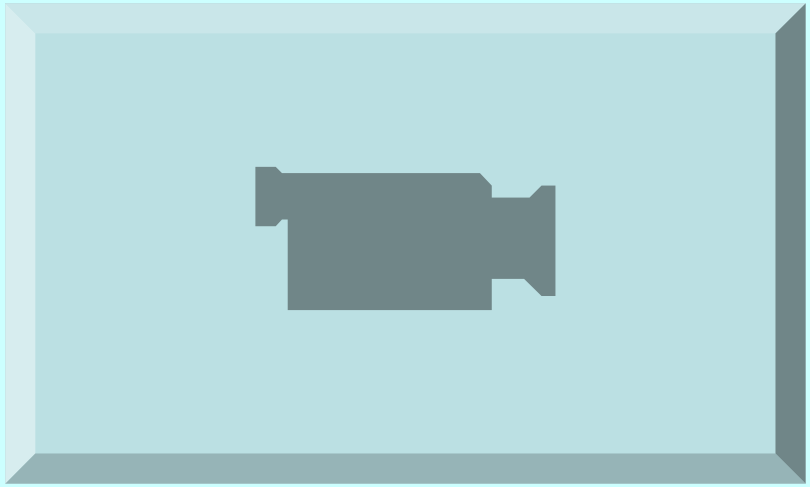
Qualitative Results

- Subjects suggested different surrogates for different tasks (e.g., ff for judging kid safe, sb for identifying images, ff for video styles)
- Three senses of gist
  - Topic (T)
  - Narrativity (N)
  - T+N+visual style
- Individual preferences and experiences influence surrogate effectiveness

Study 2: Fast Forward

- How fast can we make fast forwards?
  - 4 ff conditions (32X, 64X, 128X, 256X)
  - Four video segments for each condition
  - 45 subjects (1/2 UG, 1/2 grad, 2/3 female)
  - 6 tasks (full text gist, multiple choice gist, word object recognition, graphical object recognition, action recognition, visual gist)
  - Counterbalance speed and videos
  - Web-driven experimental condition, 3-camera video tapes, single subject at a time in usability laboratory

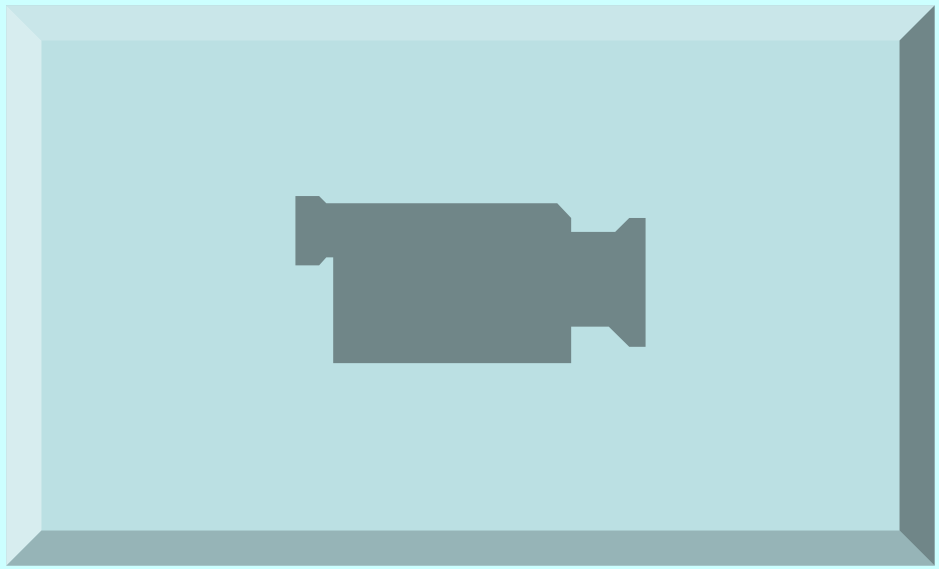
Example 1: 9:19 at 32X



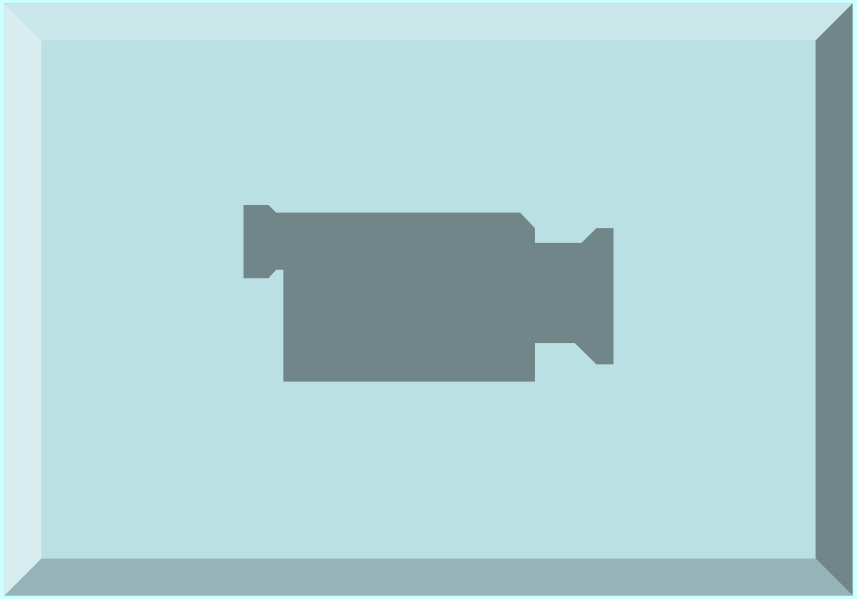
THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

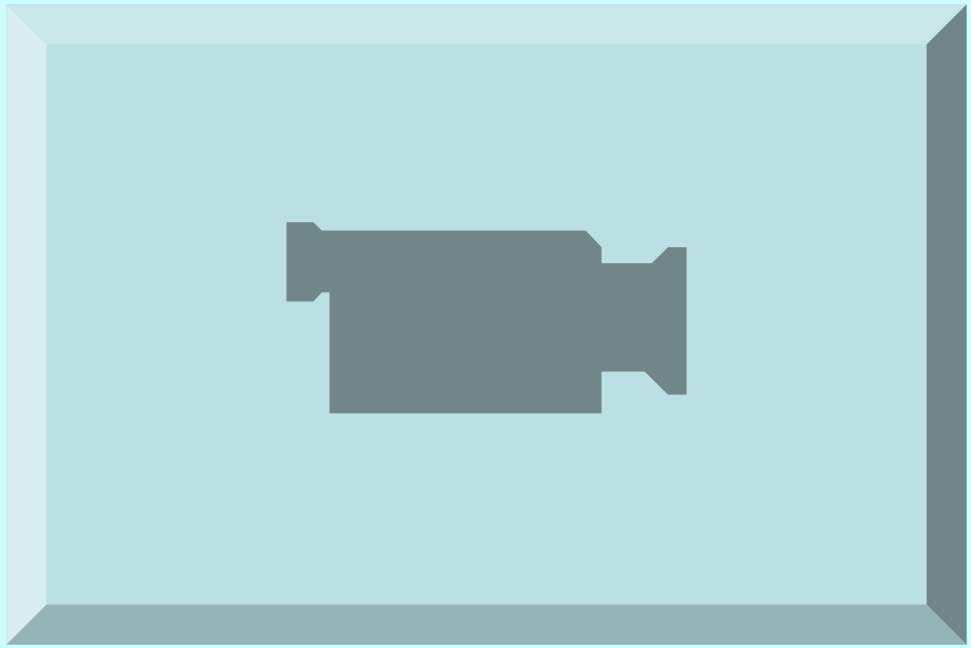
Example 2 19:48 at 64X



Example 3: 14:00 at 128X



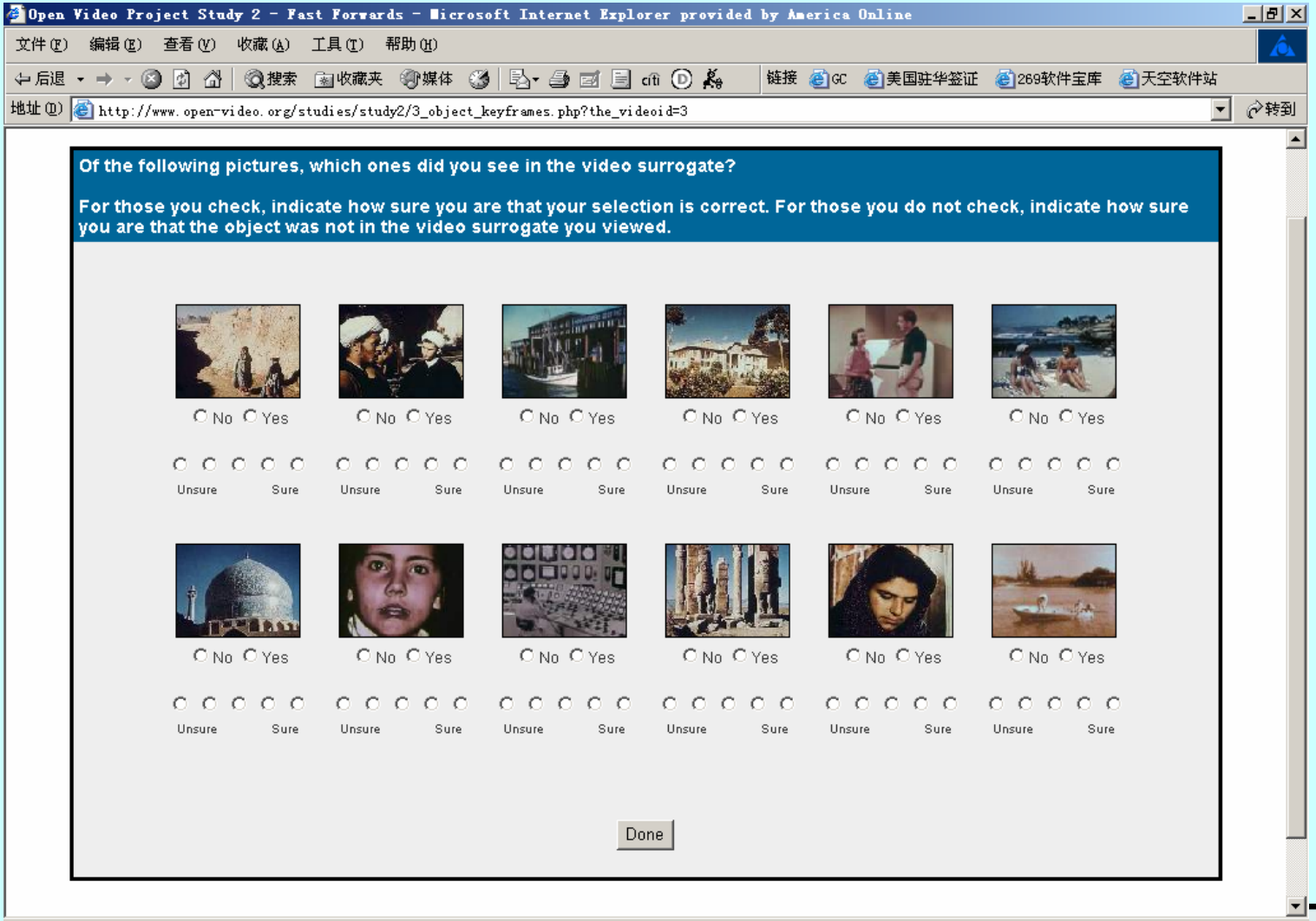
Example 4: 14:09 at 256X



THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

Example Image Recognition Stimulus



THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

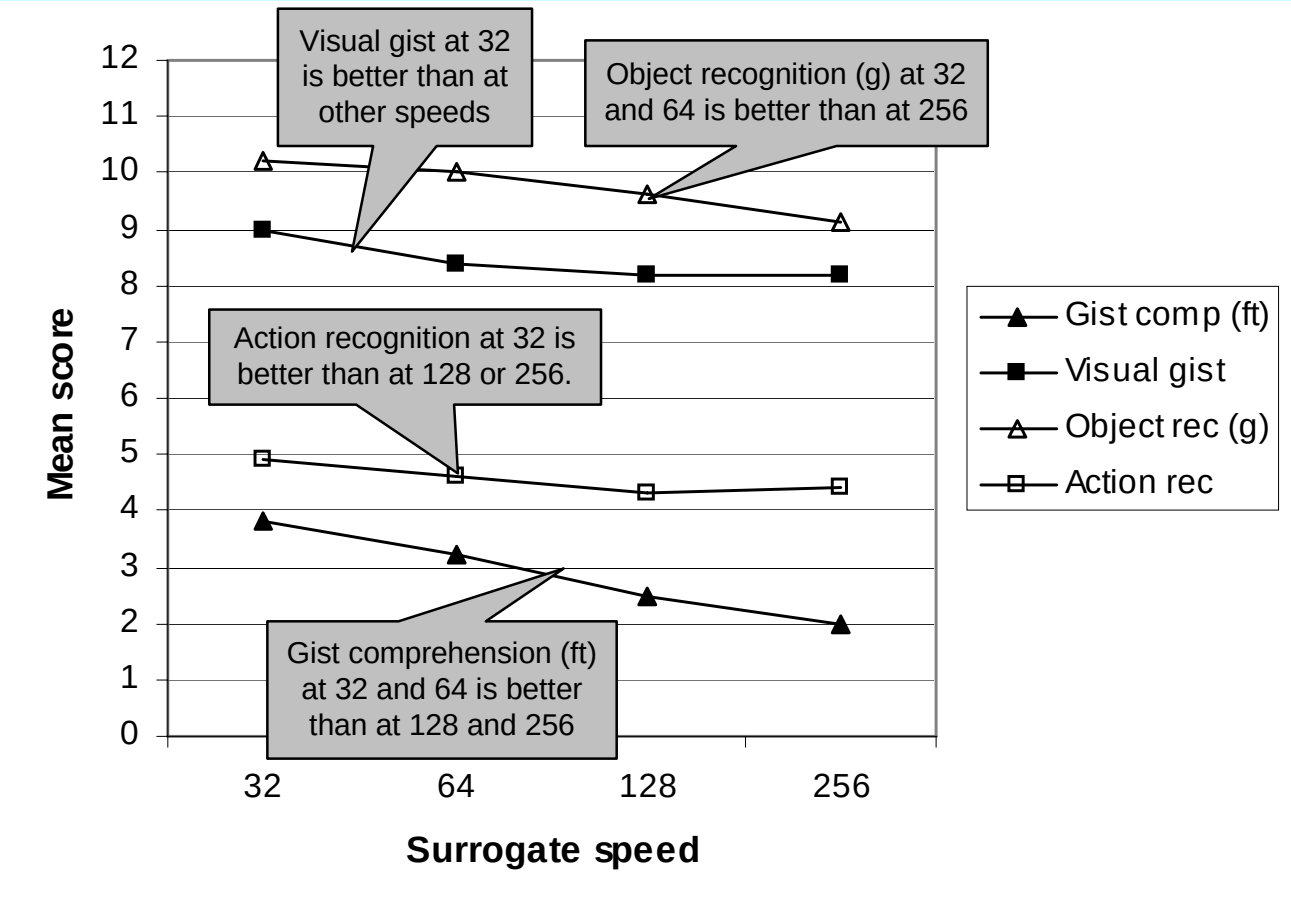
Results

- SRD on 4 of 6 tasks as speed increases, however, reasonable performance at even the highest rate
- Video content/genre interacts with performance
- Preference does not parallel performance (people can perform well under extreme conditions but do not like/enjoy)
- No user characteristic differences (age, sex)
- →Give users control but select appropriate defaults
- Caveat: controlled, independent focus on FF, likely a lower bound on performance

Performance Summary

	Maximum possible score	Mean	s.d.	Actual Min/Max
Object recognition (textual)	12	8.6	1.35	5/11
Object recognition (graphical)	12	9.7	1.65	5/12
Action recognition	6	4.5	0.93	2/6
Linguistic gist comprehension (full text)	8	2.9	1.72	0/8
Linguistic gist comprehension (multiple choice)	100%	46%		
Visual gist comprehension	12	8.4	1.41	5/12

Speed Effects on Performance



Mean Performance by Video

	Iran	Coney Island	On the Run	How Much Affection
*Gist comprehension (full text) (max=8)	3.2	2.5	2.5	3.3
*Gist comprehension (multiple choice)	89%	49%	24%	22%
*Visual gist (max=12)	8.0	8.4	9.0	8.3
*Object recognition (textual) (max=12)	7.9	9.2	9.1	8.3
*Object recognition (graphical) (max=12)	10.1	8.9	8.6	11.2
Action recognition (max=6)	4.8	4.5	4.6	4.3

Confidence-Performance Relationship

	Confidence when item was correct	Confidence when item was incorrect
Object recognition (textual)	3.9	3.4
Object recognition (graphical)	3.9	3.3
Action recognition	3.8	3.1
Visual gist determination	3.9	3.5

Confidence by Speed

	1:32	1:64	1:128	1:256
Object recognition (textual)	4.0	3.8	3.6	3.6
Object recognition (graphical)	3.9	3.9	3.8	3.6
Action recognition	3.8	3.7	3.5	3.6
Visual gist determination	3.9	3.8	3.7	3.7

Narrativity Study

- CHI walk up kiosk, 20 people used
- 20 one-minute clips (half b&w, no audio) selected on 2 criteria: contain characters, have cause/effect relations between scenes (5 in each category)
- SRD on chars, cause, and interaction

Shared Views and History Views Studies

- Evaluate AV Design Framework by instantiating and evaluating a design
- Shared (based on recommendations) and History Views (based on logs)
- Phase 1: compare OV to Views interface (28 participants). OV>accuracy; NSRD on time, but learning effect; AV>navigation/efficiency; AV>satisfaction
- Phase 2: qualitative analysis of shared and history views

Poster Frame Study

- Research Questions:
  - Given both textual and visual metadata; which surrogate will be ***utilized***, which surrogate will be ***preferred***?
  - Does the ***placement*** of the surrogates affect how they are used?
  - Does the assigned ***task*** affect how surrogates are used?
  - Does ***personal preference*** play a role in how surrogates are used?

Study Methods / Procedures

- 12 undergraduate students (paid volunteers)
- Pre-Study questionnaire
  - Demographics
  - Visual vs. Verbal learning style (VVQ)
- 10 search problems
  - Counter-balanced
- Design 1 and 2
  - 1 : text on left / visuals on right
  - 2 : visuals on left / text on right
- Eyetracking
- Post-study questionnaire
  - Follow up questions

Tasks

- Please find a video that discusses the destruction earthquakes can do to buildings. These search results are from a search on the word “Earthquake”.
- Please find a video that discusses nurses and their contributions to the United States Army. These search results are from a search on the word “Work”.
- Please choose a video from the following list that you think would be entertaining for you and your friends to watch.

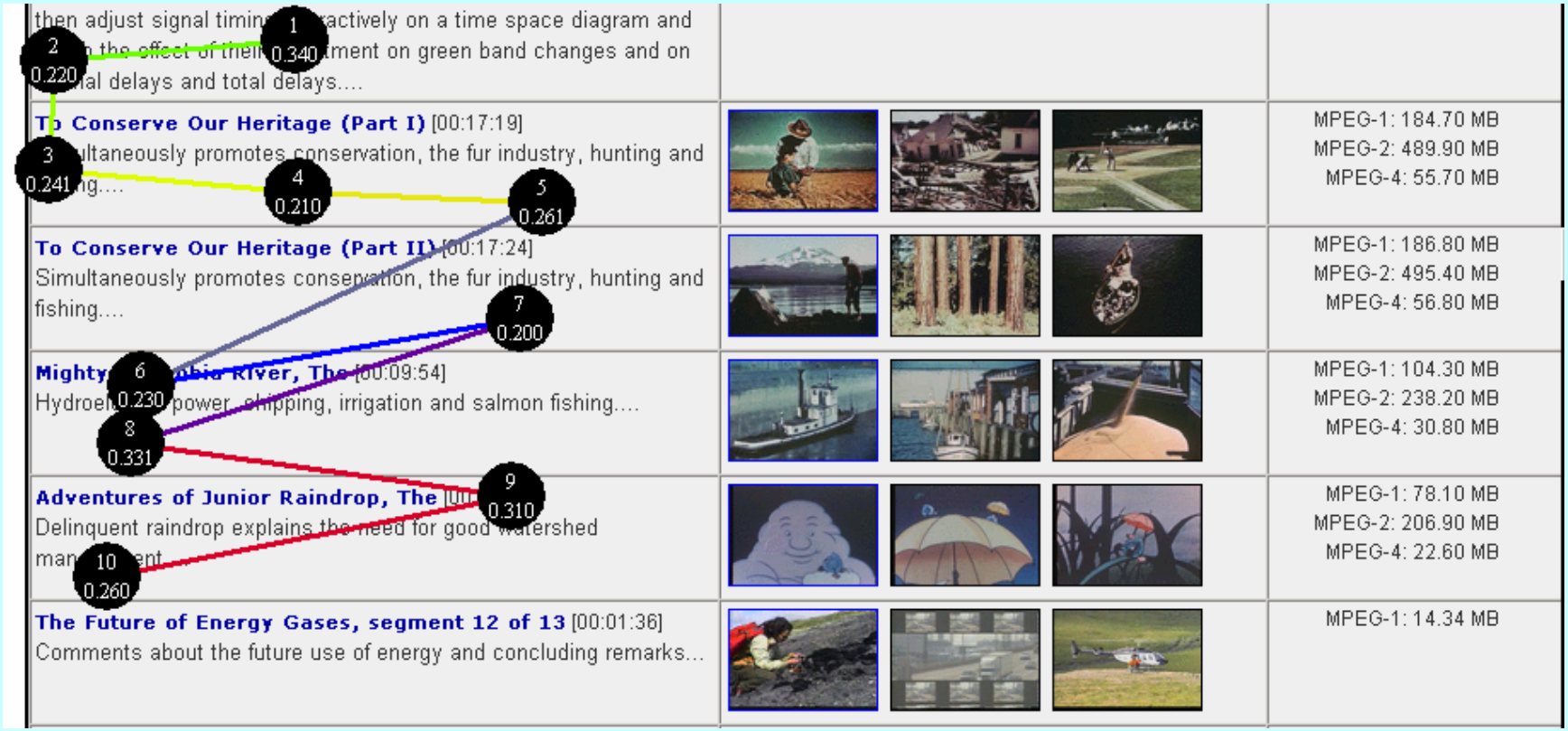
Results

- All participants over all tasks:
  - Mean time looking at text = 29.7 sec.
  - Mean time looking at pics = 6.8 sec.
  - 75% of fixations over text
  - 18% of fixations over pics
  - First fixations over text = 65
  - First fixations over pics = 54
- Text requires and gets more user attention

Results cont'd

- Design 1 vs. Design 2
  - When text was placed on the left, mean time per fixation was slightly higher
- VVQ
  - Balanced group spent more time looking at text
- Tasks
  - Varied by task:
    - Time spent looking at text
    - Time spent per fixation over text
    - Frequency of fixations over text

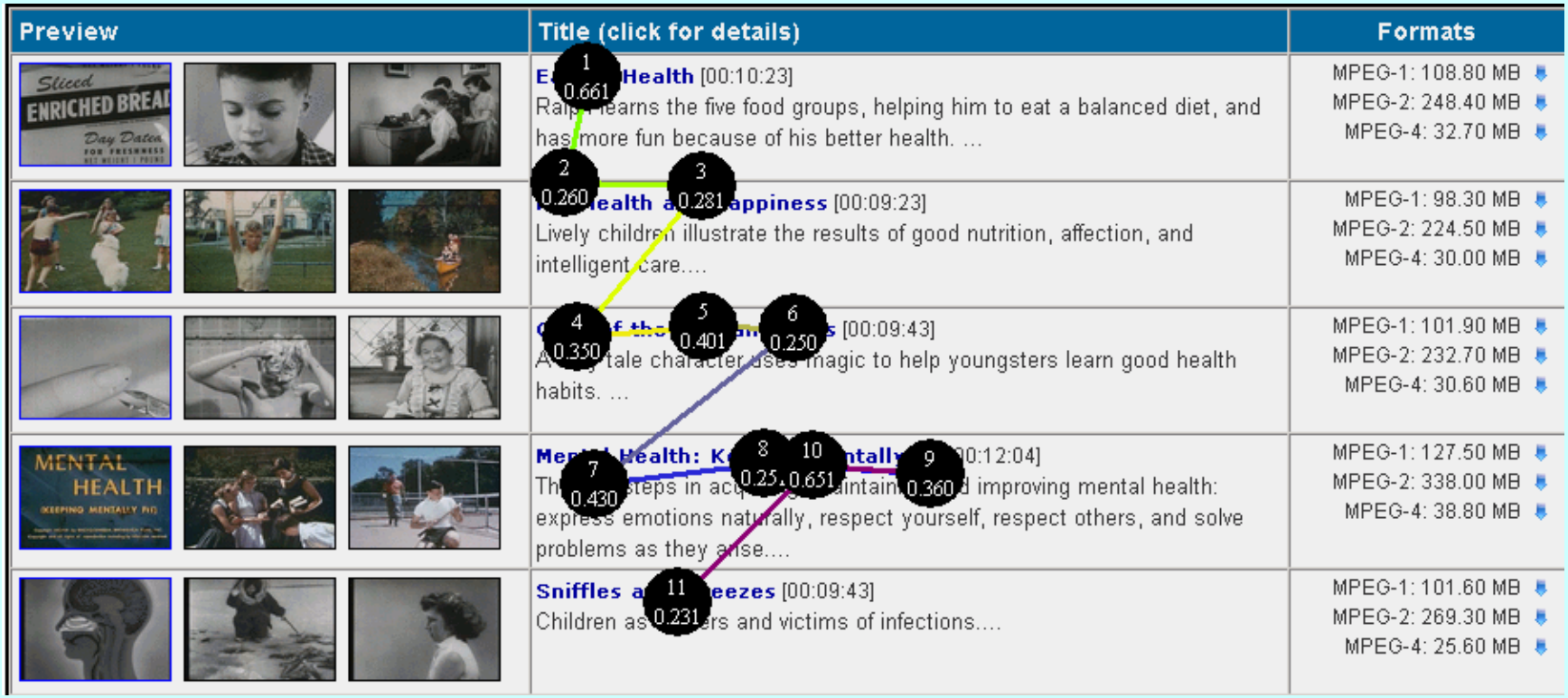
Screen Shots



THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

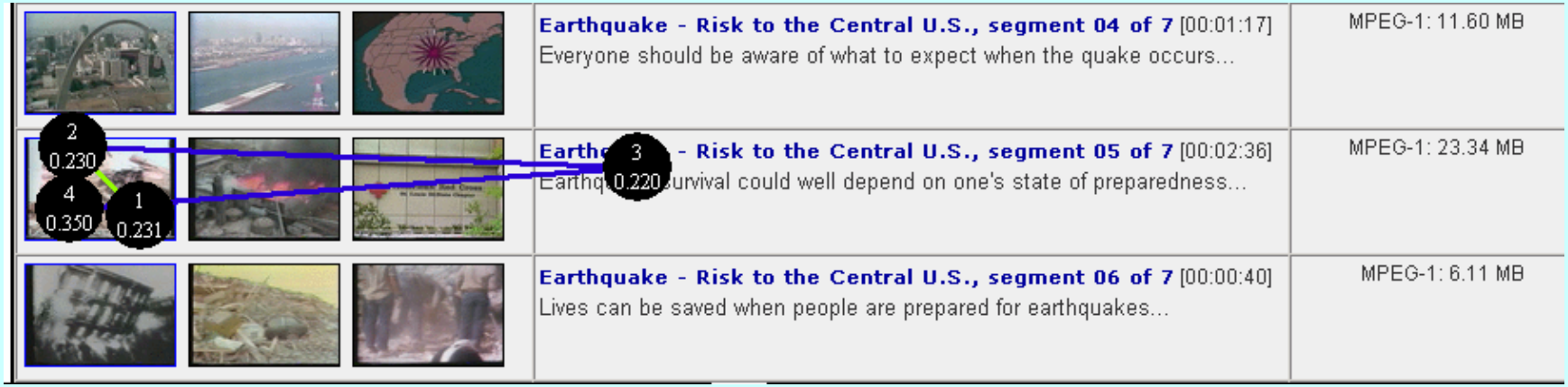
Screen Shots

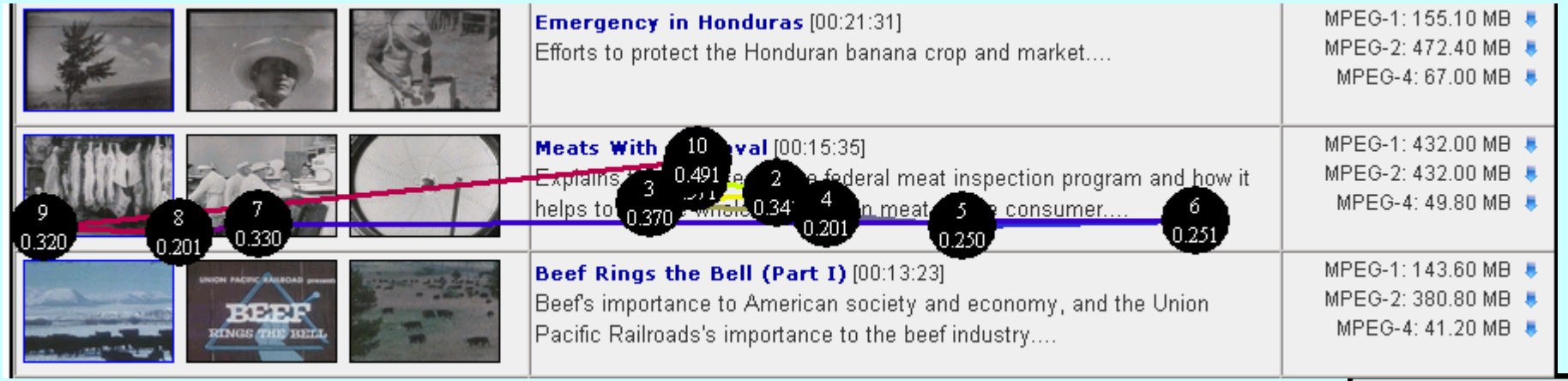


THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

Screen Shots





THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

Discussion

- In this restricted situation (i.e. pre-formulated results page) participants used text as the main anchor point
  - ? Because text is a better surrogate?
  - ? Because text contains more information?
  - ? Because text is more familiar to people
  - ? Because tasks directed users to text?

Discussion cont'd

- Layout seemed to have little effect on how surrogates were used.
  - Difference of .03 of a second
  - Participants didn't report a significant preference for layout
    - Some liked design 1 and some liked design 2
- VVQ
  - Hypothesis that visual learners would use visual surrogates and verbal learners would use verbal surrogates was not supported

Discussion cont'd

- Tasks
  - Some tasks took more time to complete
    - Regardless of:
      - Counterbalancing order
      - Participant
      - Layout design

Text or Pictures?

- Text was reported as:
  - + Being the search anchor
  - + Containing significant topical information
  - – Taking longer to read than pictures
- Visuals were reported as:
  - + Being globally liked
  - + Being used to quickly narrow down choices
  - + Taking less time to decode than text
  - + All participants said the results page would be weaker without them
  - – Often lacking in reference points

Conclusion

- Visual metadata was used to make (confirm???) relevance judgments
- Combination of visual & verbal stronger than one or the other
- Generalize with caution:
  - Small number of study participants
  - Specific set of search results pages
  - Ten specific search tasks.

VisOR study (Fall 03)

- Interface effects of automatically extracted features (TREC 02 features); 17 subjects each doing 14 search tasks
- Sliders to adjust weights of different features did not affect performance
- Keywords, indoors/outdoors and cityscape/landscape most useful
- Use of color and brightness helped with exact match searches
- General satisfaction with using different features
- (Gruss Master's paper, 2004)

Look vs Read Study (Sp 03)

- Twelve subjects think aloud while viewing results pages for five search tasks with text (titles, descriptions) or visual (3 keyframes, storyboard) surrogates
- Surrogates used differently depending on task; neither primary with considerable switching and combining (e.g., find airplane, most used visual first)
- Time a factor in deciding which to use and when
- (Hughes Master's Paper, 2004)

TREC 03 Study

- Compare transcript only, feature only, and combined surrogates with 36 subjects
- NSRD in precision across 3 surrogates, transcript only and combined yielded SR higher recall in less time and SR greater satisfaction results.

(TREC notebook; ACM MM 04)

Video Relevance Judgments

- How do people make relevance judgments for video? Qualitative study (Yang dissertation; CHI 05; ASIST 04)
  - 3 groups
    - Video editors/producers
    - Video librarians
    - Video users (professors)
- 9 visual gist attributes
- Differences across users

Other Studies

- Relative value of surrogates in context
  - Four sets of surrogates (ff, sb, excerpt, combined) compared (in analysis)
- Mu dissertation: cognitive load effects on collaborative learning with video (ISEE) Investigation of tasks
- TREC 05 study

The Integration Study

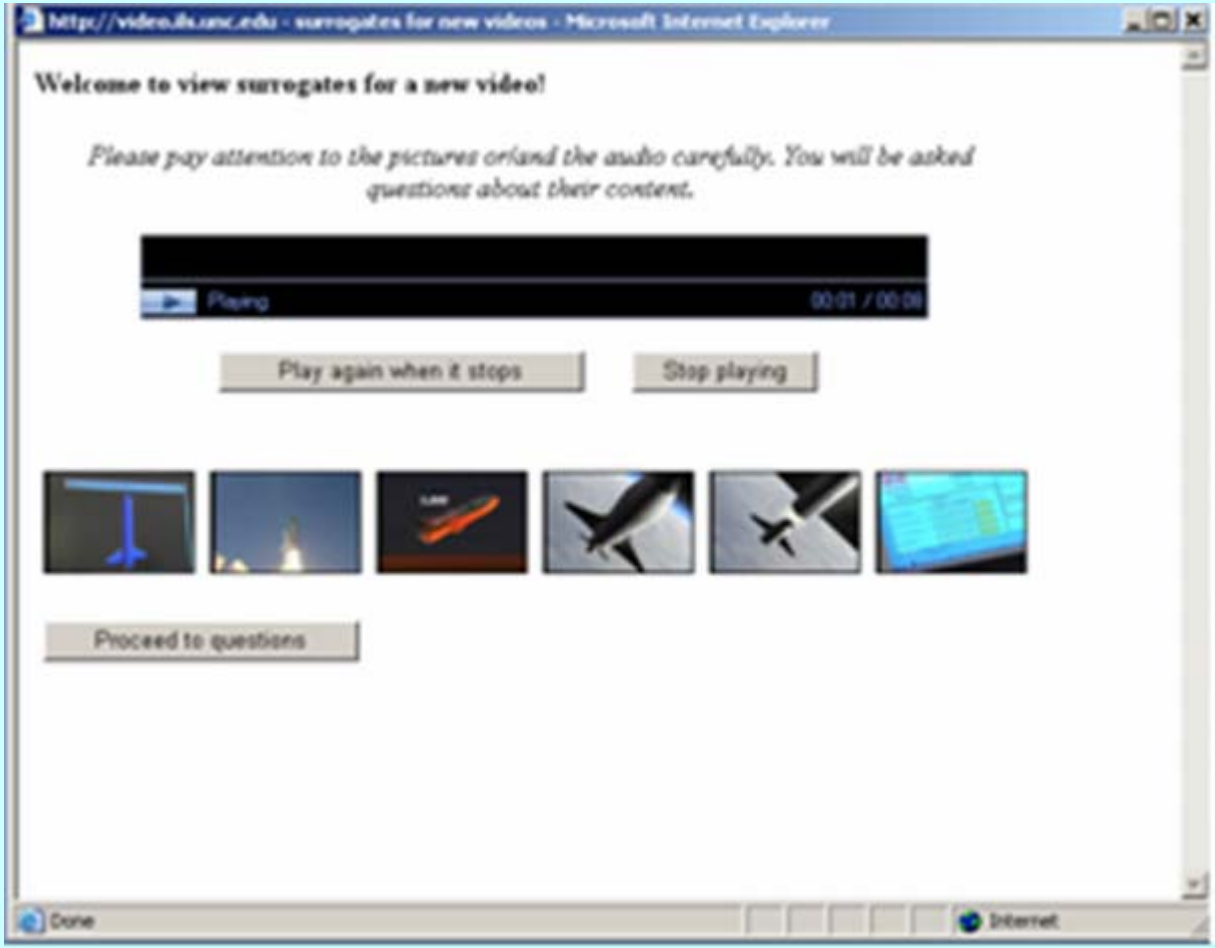
- Compare old OV to redesign? Compare to Internet archive?
- How do multiple surrogates and agile control mechanisms affect understanding of video?
- Accuracy? Time? Satisfaction? Cognitive load? Navigational overhead?

Audio and Visual Surrogates

- How does spoken descriptions compare to storyboards?
- How much value is added by combined audio and visual surrogates?
- Summer 06, submitted to CHI 07

Method

- 15 NASA video segments
- Three types of surrogate (spoken description, storyboard, combined)
- 36 participants (within subjects)
- 5 trials with each surrogate (counter balanced)
- Procedure
  - Experience (viewing/listening) surrogate
  - 5 sense making (gist) tasks
    - Written Gist Determination (open-ended question)
    - Keyword Recognition (multiple choices)
    - Title Selection (single choice from four alternatives)
    - Keyframe Recognition (multiple choices)
    - Gist Recognition (single choice from four alternatives)
  - Self-report feedback during and after the trials
- Repeated measures ANOVA
- Qualitative comments



THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

Performance on Five Tasks, by Surrogate

Surroga	N	Write Gist Mean	Keyword Recog Mean	Title Select Mean	Keyframe Select Mean	Gist Select Mean
Visual	36	**1.06**	0.89	0.95	0.83	**0.88**
Audio	36	1.88	**0.86**	0.98	0.82	0.97
Both	36	1.92	0.9	0.99	0.83	0.99
Main Effects		p<.001	p=.006	p=.174	p=.813	p<.001
Contrasts		V<A~B	A<B,V~A			V<A~B

Confidence on Five Tasks, by Surrogate

Surroga	N	Write Gist Mean	Keyword Recog Mean	Title Select Mean	Keyframe Select Mean	Gist Select Mean
Visual	36	**3.54**	**3.97**	**4.01**	3.73	**3.97**
Audio	36	4.20	4.26	4.61	3.84	4.75
Both	36	4.21	4.32	4.60	3.89	4.71
Main Effects		p<.001	p=.003	p<.001	p=.319	p<.001
Contrasts		V<A~B	V<A~B	V<A~B		V<A~B

THE OPEN VIDEO PROJECT
a shared digital video repository

UNC-Chapel Hill
School of Information & Library Science
INTERACTION DESIGN LAB

Time to Experience Surrogate and to Complete Task, by Surrogate

Surrogate	**N**	**Mean Time to Experience Surrogate (sec)**	**Mean Time to Complete Task (sec)**
Visual	36	19.47	90.75
Audio	36	27.22	87.93
Both	36	28.81	88.09
Main Effects		p=.001	p=.894
Contrasts		V<A, V<B, A~B	

Subjective Measures, by Surrogate

Surrogate	N	Usability Mean	Engagement Mean	Enjoyment Mean
Visual	36	3.09	4.16	3.61
Audio	36	3.78	4.56	3.90
Both	36	4.02	4.80	4.71
Main Effects		p<.001	p=.022	p<.001
Contrasts		A<V<B	V<B, A~B	A~V<B

New Set of Follow on Studies

- How to construct spoken surrogates
  - Descriptions (manual vs automatic—MAGIC)
  - Keywords (manual vs automatic)
- Different visual surrogates
  - Storyboards
  - Fast forwards
- How to combine (degree of synchronization)
- Audio abstractions

Take Away Summary

- User studies inform good design
- Give people multiple views and easy control mechanisms
- No silver bullets (many factors determine performance and preference); people make context-dependent tradeoff decisions
- Video offers new kinds of potentials for learning and communication; do not ignore audio channel semantic richness
- Good user interfaces get used