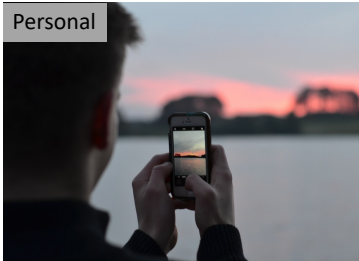


Computational Video Editing

CS448V – May 20, 2019

Personal



Indie



Feature



Making a video means different things to different people

Documentary



Action



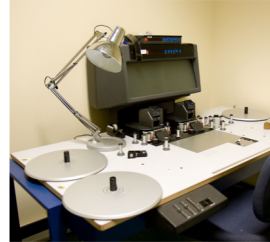
Sitcom



Video Editing



1920's
Early editing machines



1970's
Nonlinear Editing

2000's
Transition to digital media

Early 1900's
First video edits with cut and taped film



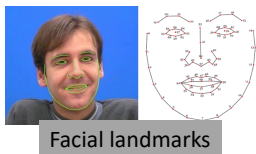
1960's
Electronic editing

1990's
Nonlinear editing software

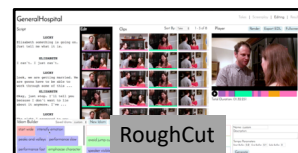
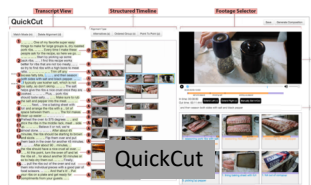
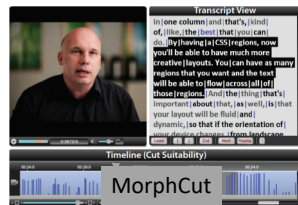


Computational Video Tools in Research

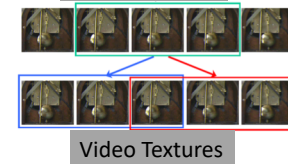
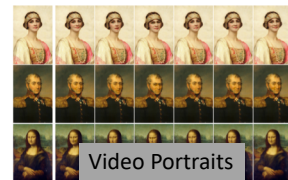
Analyzing Video

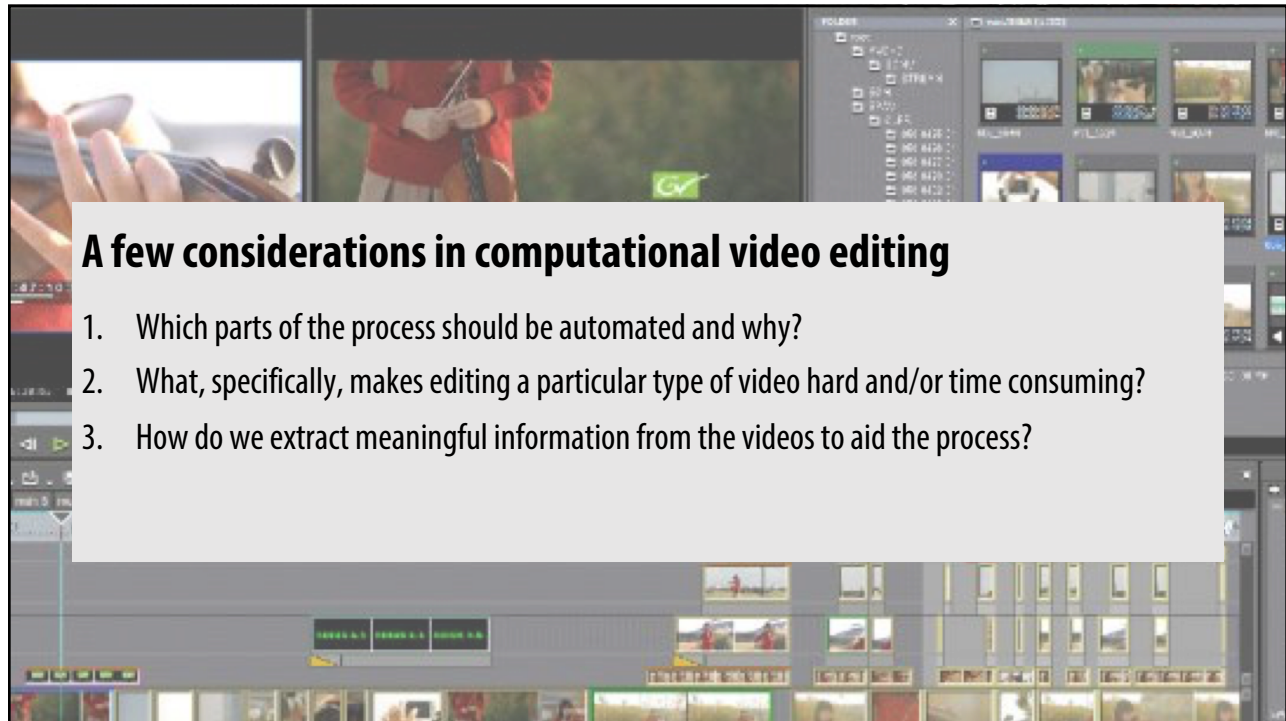


Creating "Traditional" Media



Creating "New" Media

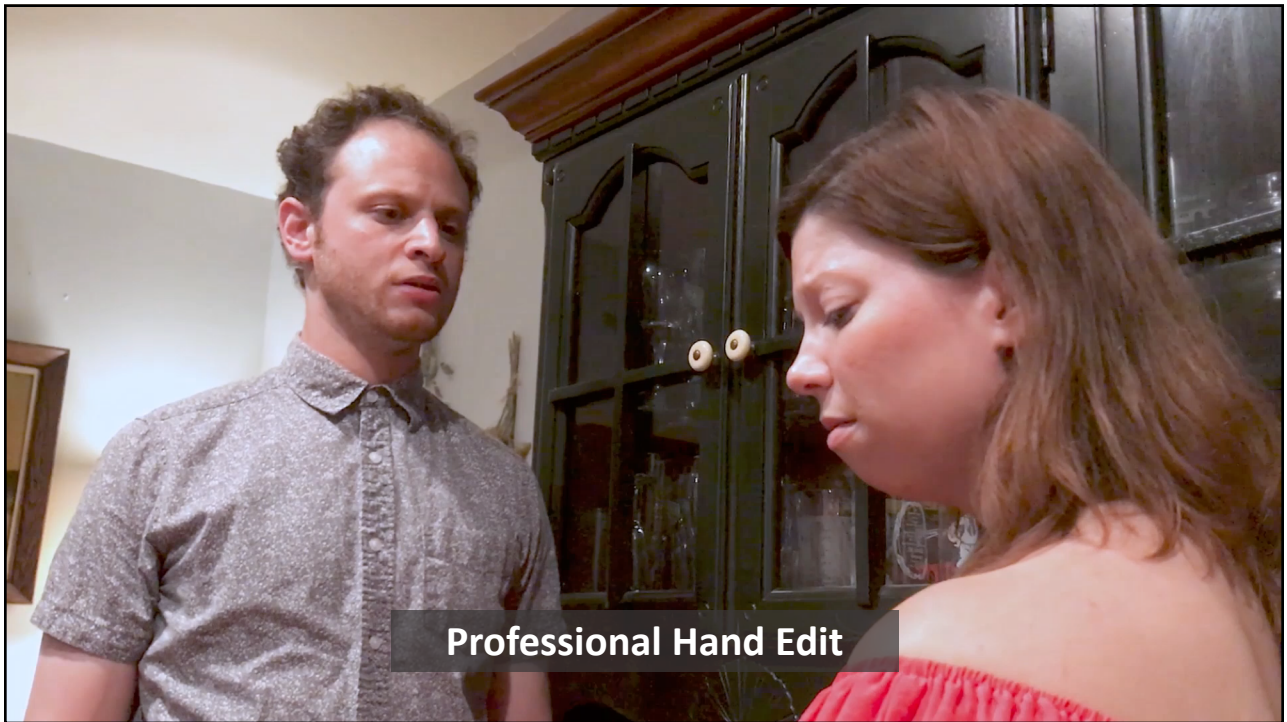




Computational Video Editing for Dialogue-Driven Scenes

Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala
SIGGRAPH 2017





Avid Media Composer (20x speed)

Name	Duration	IN-OUT	Mark IN	Mark OUT
GenHosp Sequence	56:18			
take1.mp4	1:34:04	5:22 1:00:19:01	01:00:2	
take2.mp4	1:26:09	7:20 1:00:50:01	01:00:5	
take3.mp4	1:21:21	18:13 1:01:03:08	01:01:2	
take4.mp4	1:26:07		1:00:38:17	
take5.mp4	1:34:11	14:12 1:00:59:19	01:01:0	
take6.mp4	1:23:07			
take7.mp4	1:27:21	2:04 1:00:13:11	01:00:1	
take8.mp4	1:36:17		1:00:58:13	

Raw Takes Length: 18 mins

Edited Scene Length: 89 secs

Editing Time: 3 hrs

Dialogue-Driven Scenes



"Eternal Sunshine of the Spotless Mind" [2004]



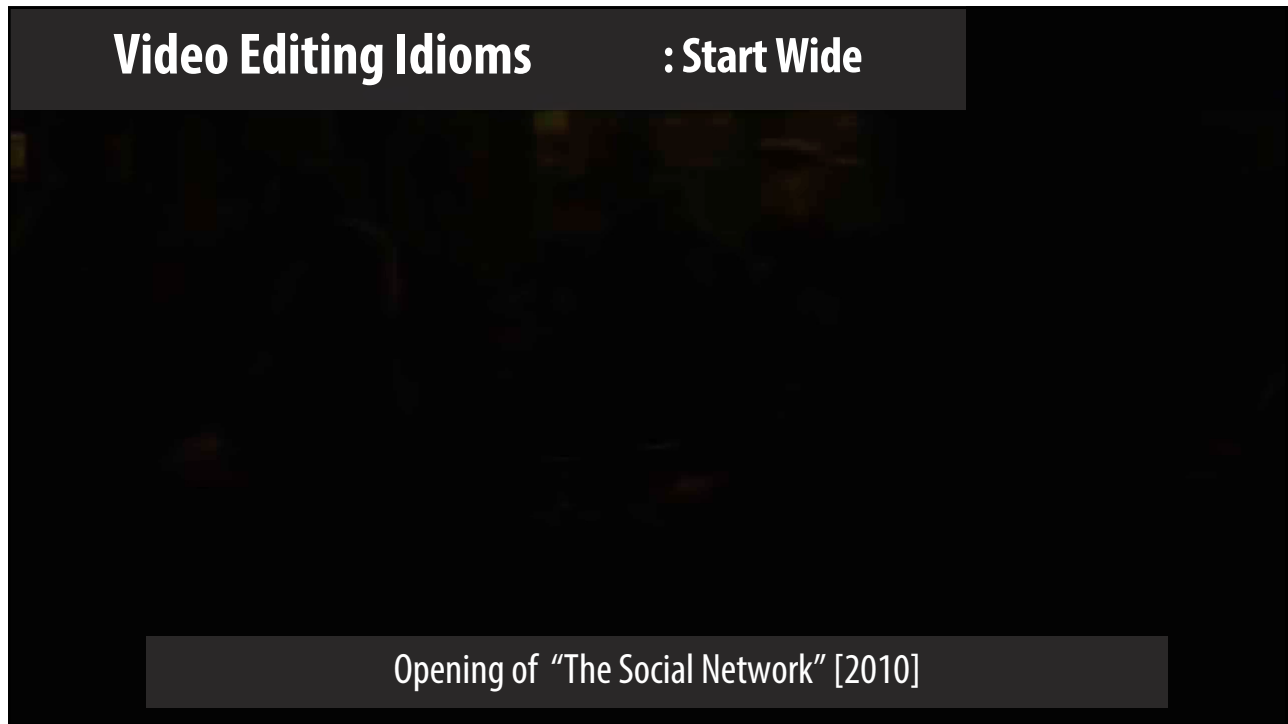
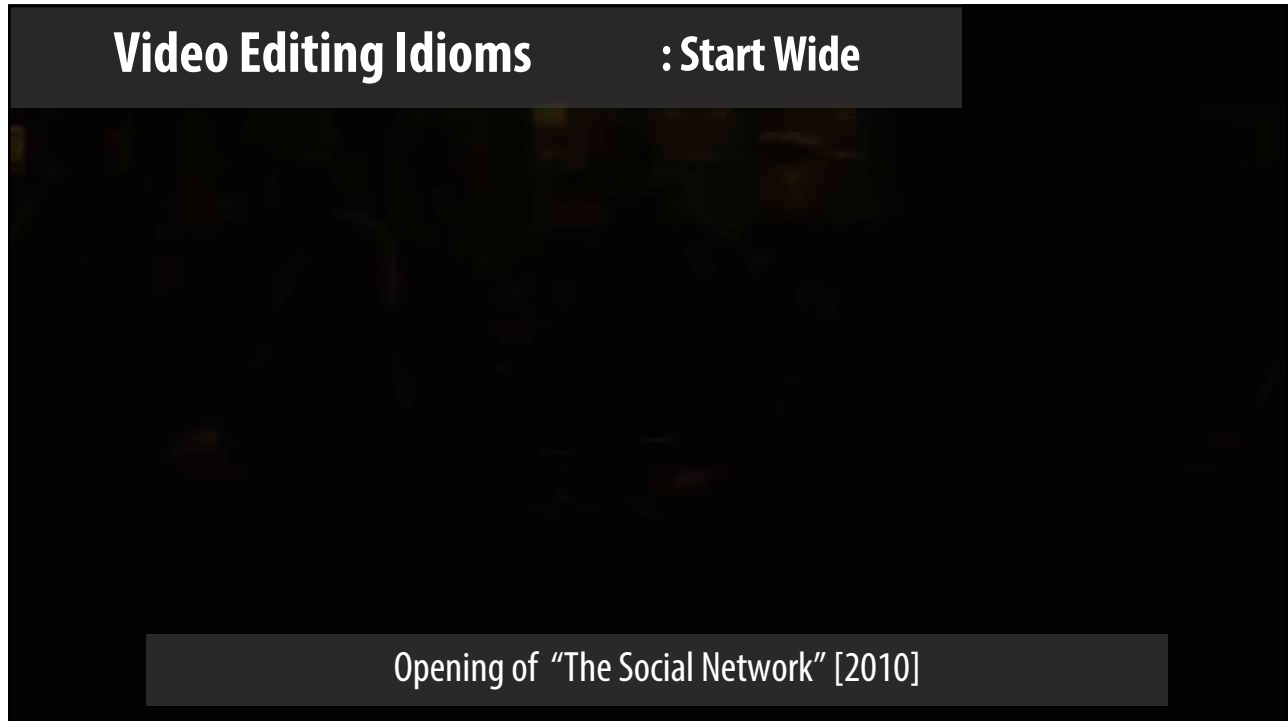
"Casablanca" [1942]

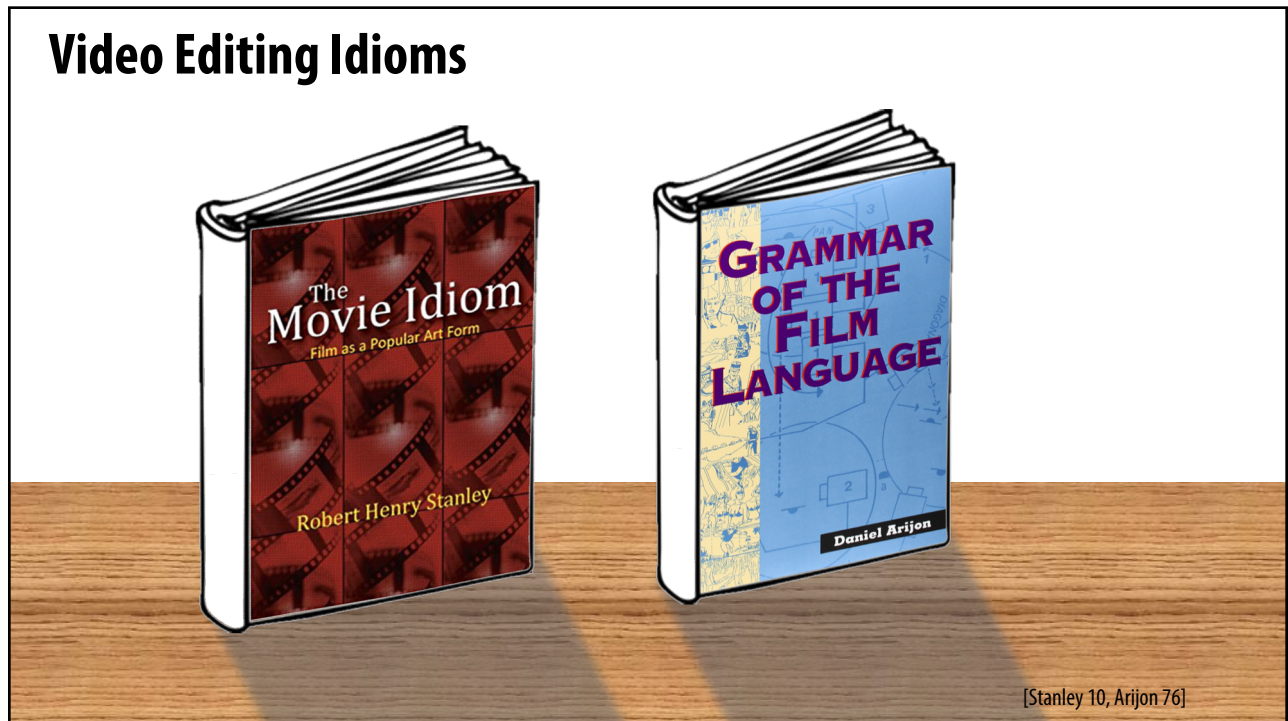


"Doubt" [2008]

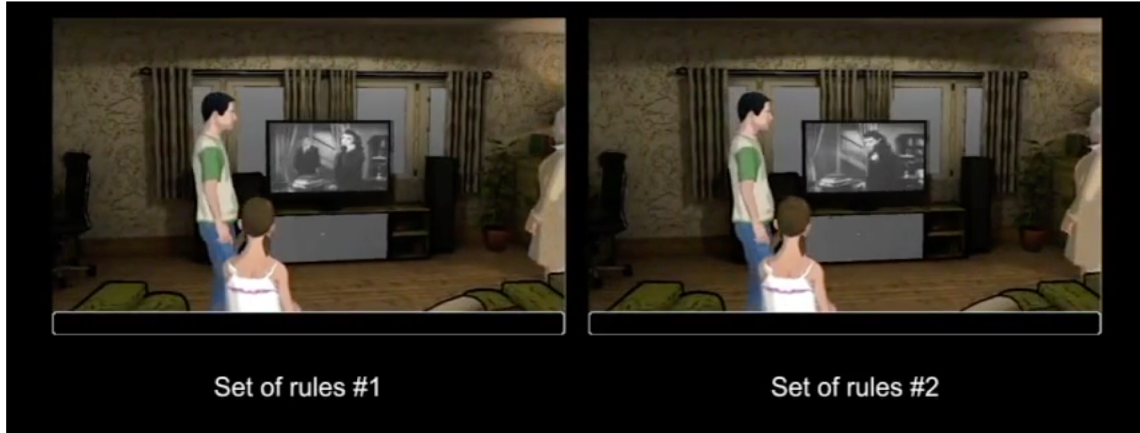


"Dumb and Dumber" [1994]





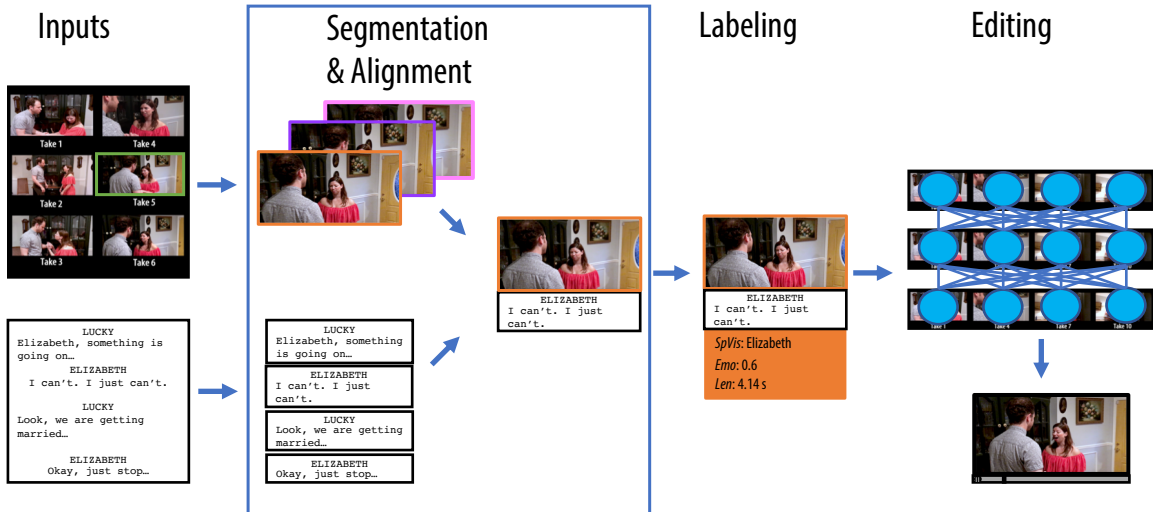
Prior Work on Virtual Cinematography



“Narrative-Driven Camera Control for Cinematic Replay of Computer Games.” [Galvane 14]

Additional work: [Karp 93, Christianson 96, He 96, Jhala 05, Elson 07, Lino 11, Galvane 15]

Our System



Segmentation & Alignment

Script

Script

LUCKY
Elizabeth something is going on. Just tell me what it is.

ELIZABETH
I can't. I just can't.

LUCKY
Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.

ELIZABETH
Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...

LUCKY
The night I proposed to you.

ELIZABETH
Yes. It wasn't planned. It just. It just happened.

LUCKY
Things like that don't just happen, Elizabeth.

Input Takes



Take 0



Take 1



Take 2



Take 3

Segmentation & Alignment

Script

Script

LUCKY
Elizabeth something is going on. Just tell me what it is.

ELIZABETH
I can't. I just can't.

LUCKY
Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.

ELIZABETH
Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...

LUCKY
The night I proposed to you.

ELIZABETH
Yes. It wasn't planned. It just. It just happened.

LUCKY
Things like that don't just happen, Elizabeth.

Input Takes



based script-video alignment [Robin 13]



Segmentation & Alignment

Script

Script

LUCKY
Elizabeth something is going on. Just tell me what it is.

ELIZABETH
I can't. I just can't.

LUCKY
Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.


ELIZABETH
Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...

LUCKY
The night I proposed to you.

ELIZABETH
Yes. It wasn't planned. It just. It just happened.

LUCKY
Things like that don't just happen, Elizabeth.

Input Takes



Segmentation & Alignment

Script

Script

LUCKY
Elizabeth something is going on. Just tell me what it is.

ELIZABETH
I can't. I just can't.

LUCKY
Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.

ELIZABETH
Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...

LUCKY
The night I proposed to you.

ELIZABETH
Yes. It wasn't planned. It just. It just happened.

LUCKY
Things like that don't just happen, Elizabeth.

Input Takes

	Line 1	Clip 1	Clip 2	Clip 3	Clip 4
	Line 2	Clip 1	Clip 2	Clip 3	Clip 4
		Clip 2	Clip 3	Clip 4	
		Clip 2	Clip 3	Clip 4	
	Line 5	Clip 1	Clip 2	Clip 3	Clip 4
	Line 6	Clip 1	Clip 2	Clip 3	Clip 4
	Line 7	Clip 1	Clip 2	Clip 3	Clip 4

Alignment provides structure for editing

Segmentation & Alignment

Script	Edit	Input Takes
<p>Script</p> <p style="text-align: center;">LUCKY</p> <p>Elizabeth something is going on. Just tell me what it is.</p> <p style="text-align: center;">ELIZABETH</p> <p>I can't. I just can't.</p> <p style="text-align: center;">LUCKY</p> <p>Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.</p> <p style="text-align: center;">ELIZABETH</p> <p>Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...</p> <p style="text-align: center;">LUCKY</p> <p>The night I proposed to you.</p> <p style="text-align: center;">ELIZABETH</p> <p>Yes. It wasn't planned. It just. It just happened.</p> <p style="text-align: center;">LUCKY</p> <p>Things like that don't just happen, Elizabeth.</p>		

Segmentation & Alignment

Script	Edit	Input Takes
<p>Script</p> <p style="text-align: center;">LUCKY</p> <p>Elizabeth something is going on. Just tell me what it is.</p> <p style="text-align: center;">ELIZABETH</p> <p>I can't. I just can't.</p> <p style="text-align: center;">LUCKY</p> <p>Look, we are getting married. We are gonna have to be able to work through some of this stuff and trust each other.</p> <p style="text-align: center;">ELIZABETH</p> <p>Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've been sleeping with Nicholas. Ever since the night of ...</p> <p style="text-align: center;">LUCKY</p> <p>The night I proposed to you.</p> <p style="text-align: center;">ELIZABETH</p> <p>Yes. It wasn't planned. It just. It just happened.</p> <p style="text-align: center;">LUCKY</p> <p>Things like that don't just happen, Elizabeth.</p>		

Challenge: Choose appropriate clip for each line of dialogue

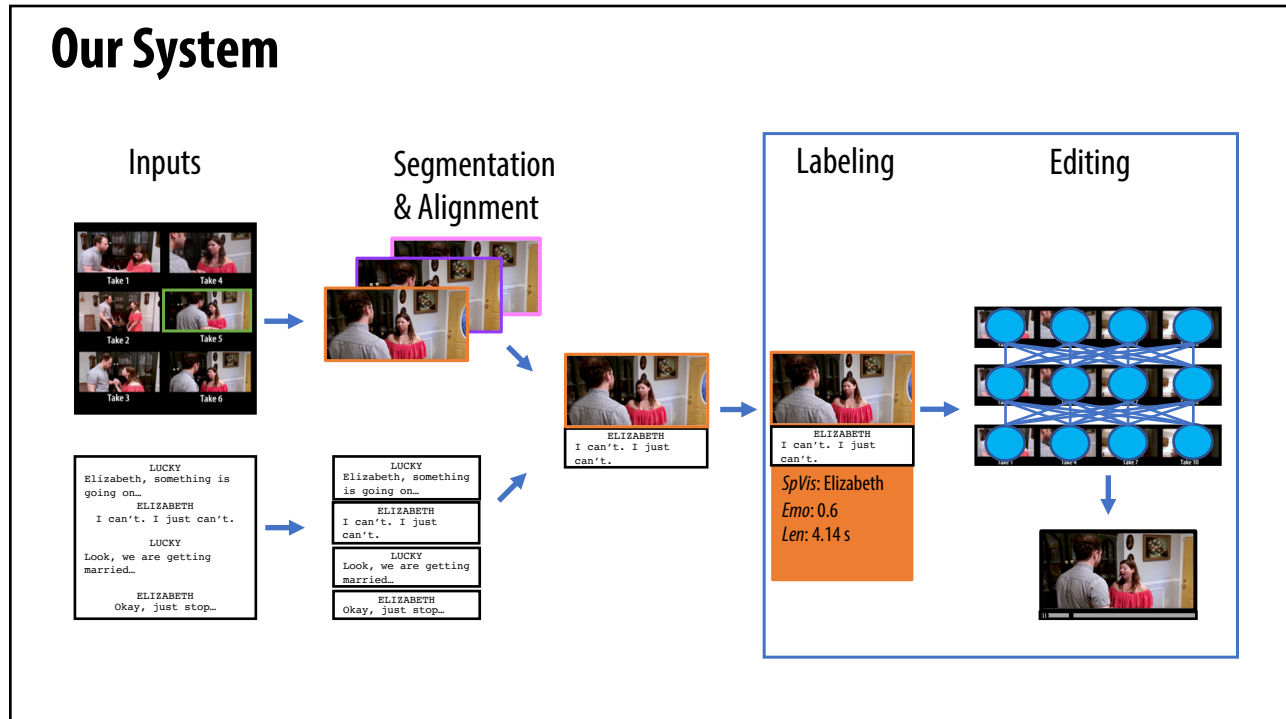
The screenshot shows a web-based video editing interface for a project named "GeneralHospital". The interface is divided into several panels:

- Script:** Contains dialogue for characters LUCKY and ELIZABETH. LUCKY says, "Elizabeth something is going on. Just tell me what it is." ELIZABETH replies, "I can't. I just can't." LUCKY continues, "Look, we are getting married. We are gonna have to be able to work through some of this ..." ELIZABETH responds, "Okay, just stop. I'll tell you because I don't want to lie about it anymore. I've ..." LUCKY says, "The night I proposed to you".
- Edit:** A vertical strip of video thumbnails corresponding to the script lines.
- Clips:** A 3x3 grid of video thumbnails, with the middle row highlighted in green.
- Player:** A large video preview window showing a scene with a man and a woman. Below it is a progress bar and the text "Total Duration: 01:32:251".
- Idiom Builder:** A bottom panel with various editing options like "start wide", "intensify emotion", "peaks and valleys", "performance slow", "avoid jump cuts", "performance fast", "emphasize character", and "speaker visible". It also includes a "Generate" button and tempo parameters.

This screenshot is identical to the one above but features colored overlays and labels to identify the main components of the interface:

- The **Script** panel is highlighted in orange.
- The **Edit** panel is highlighted in blue.
- The **Clip Grid** panel is highlighted in yellow.
- The **Viewer** panel is highlighted in green.
- The **Idiom Builder** panel is highlighted in purple.
- Large black text labels "Script", "Edit", "Clip Grid", "Viewer", and "Idiom Builder" are placed over their respective panels.

Our System



Labeling: Speaker

Input Script

```

          LUCKY
Elizabeth, something is going on. Tell me
what it is.

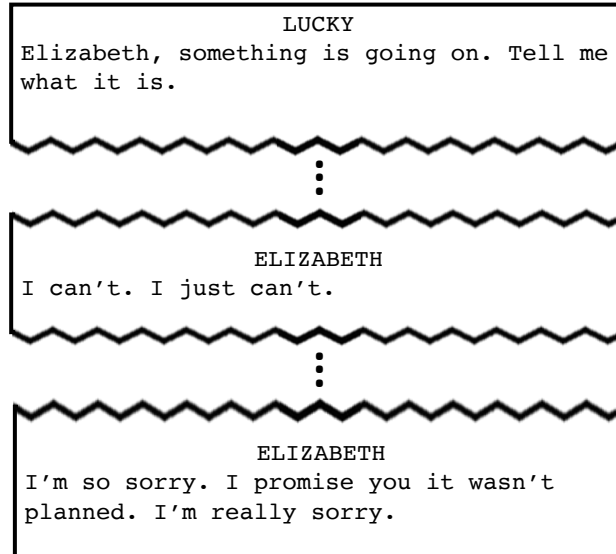
          ELIZABETH
I can't. I just can't.

          ELIZABETH
I'm so sorry. I promise you it wasn't
planned. I'm really sorry.

```

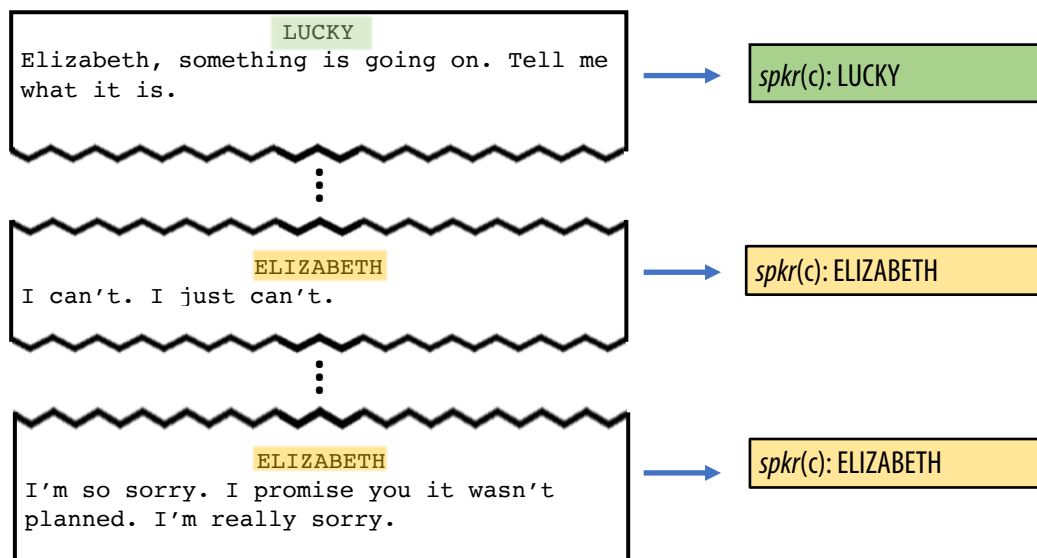
Labeling: Speaker

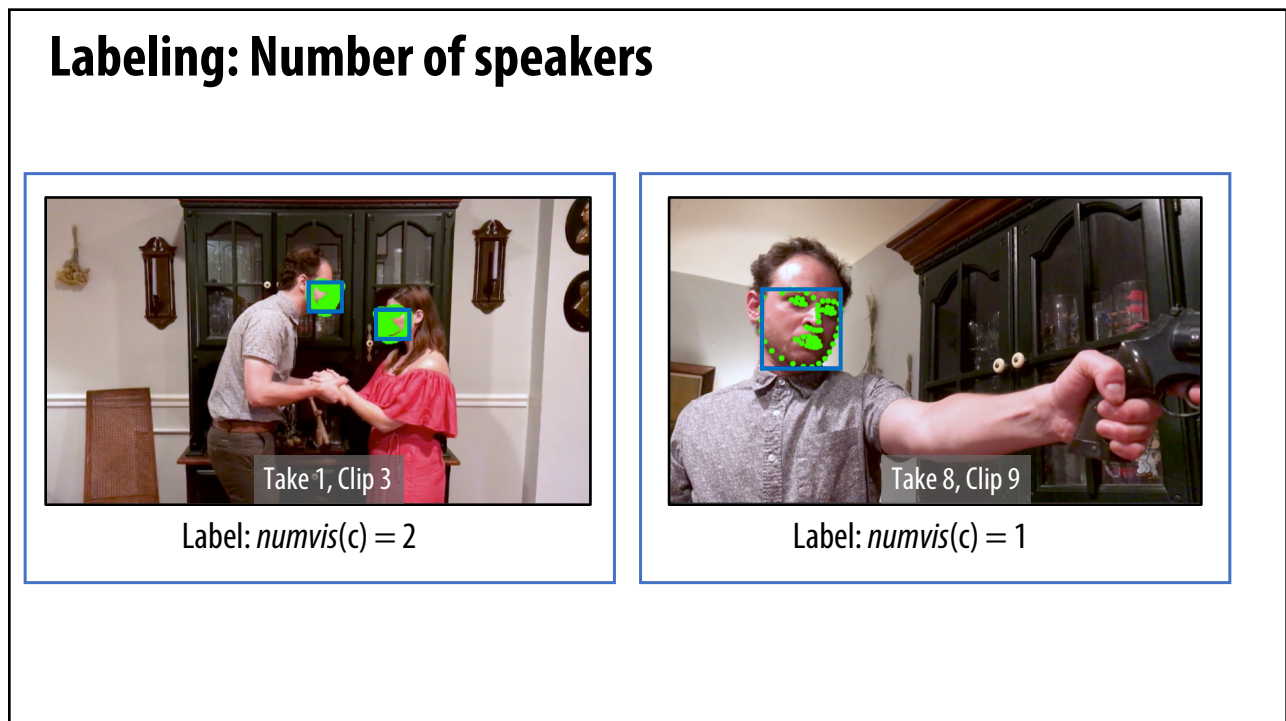
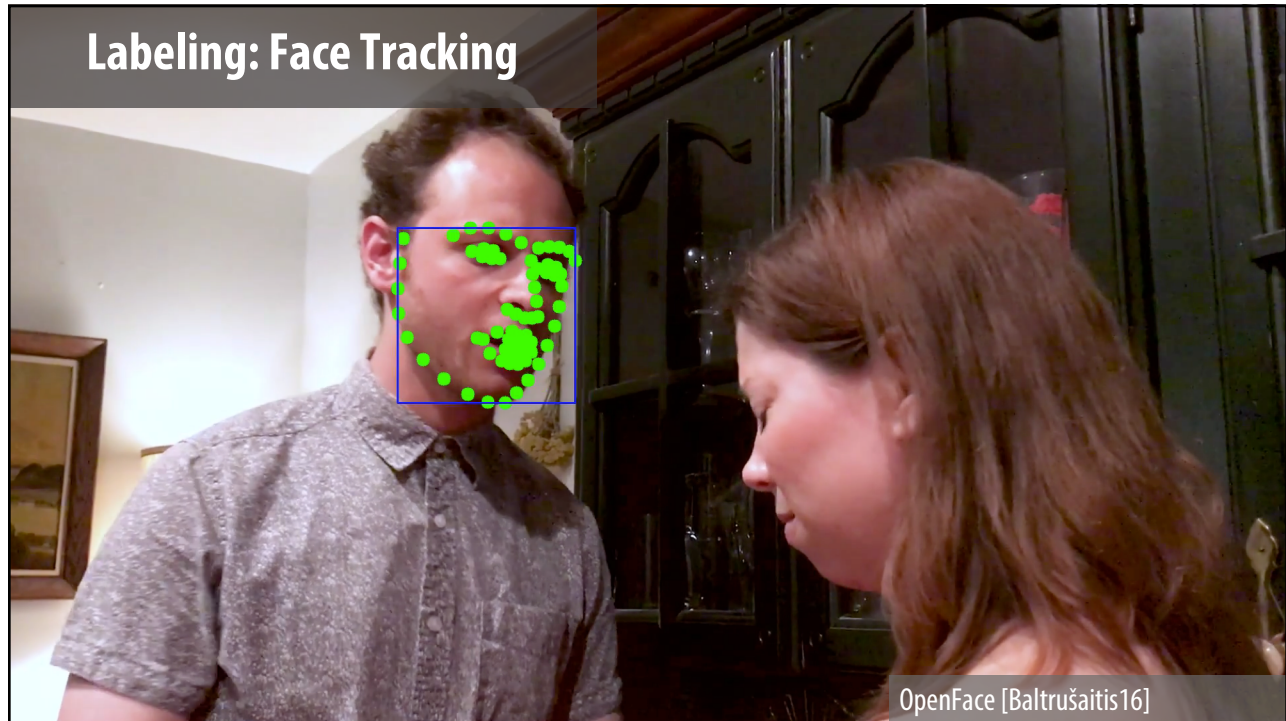
Input Script



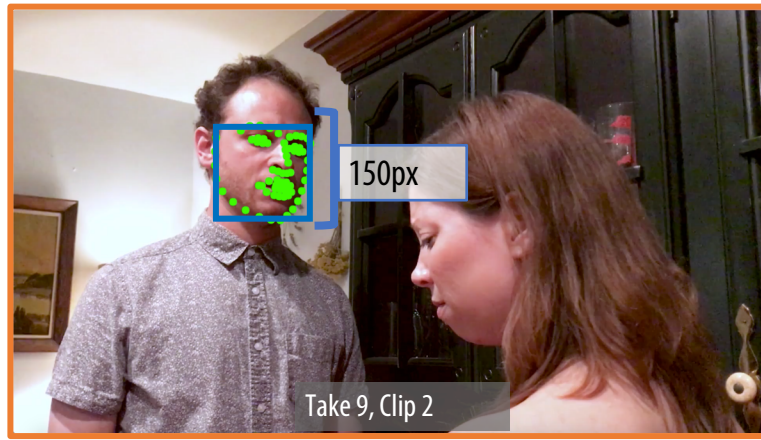
Labeling: Speaker

Input Script





Labeling: Shot Type: Zoom



600px

$$zoom(c) = \frac{150}{600} = 0.25$$



Long



Medium Long



Medium



Medium Close Up

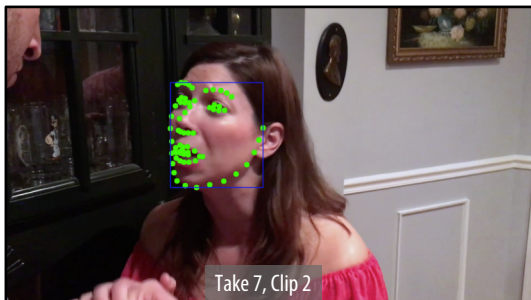


Close Up



Extreme Close Up

Labeling: Visible Speaker



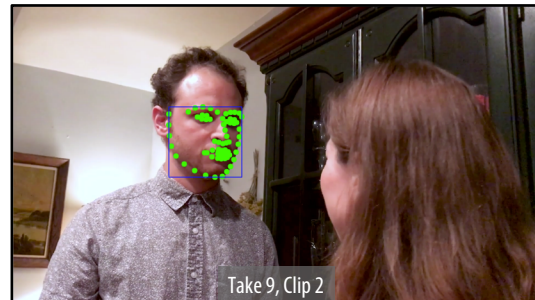
Take 7, Clip 2

ELIZABETH

I can't. I just can't.



mouth motion



Take 9, Clip 2

ELIZABETH

I can't. I just can't.

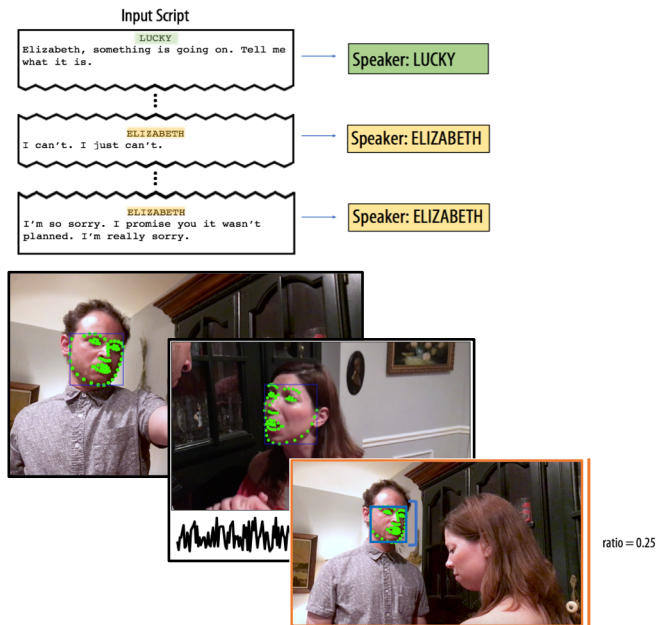


mouth motion

Label: $SpVis(c) = ELIZABETH$

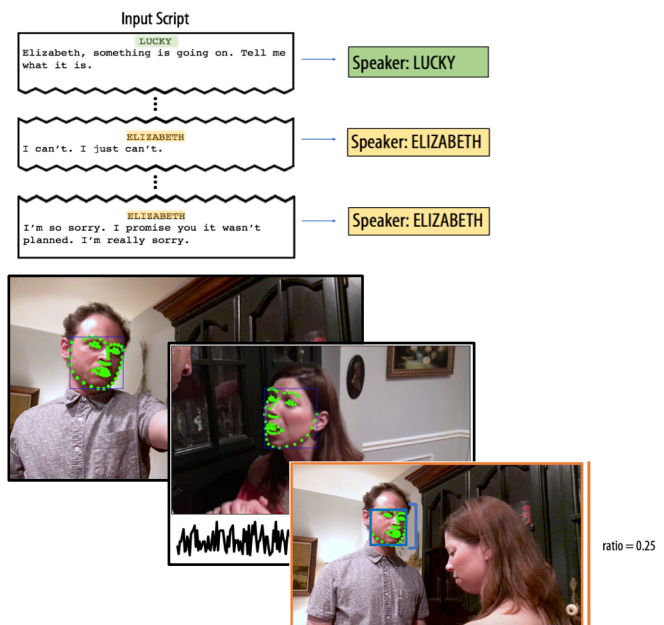
Additional Labels

- Script
 - **Speaker**
- Video
 - **Shot type: NumVis**
 - **Shot type: Zoom**
 - **Speakers visible**



Additional Labels

- Script
 - **Speaker**
 - Emotional Sentiment
- Video
 - **Shot type: NumVis**
 - **Shot type: Zoom**
 - **Speakers visible**
 - Screen position
 - Clip volume
 - Clip length



goldfish

Script

LINDA
What time do you think she'll call?

CRAIG
Not today. Remember?

LINDA
Oh, yeah.

CRAIG
She called yesterday.

LINDA
Where was she going to dinner tonight?

CRAIG

Clips

Sort By: Take 1.4 of 8

Player

Total Duration: 00:53:920

Current Clip Info

Line: The Red Chili Hut.
Take: 6
Visible Actors: CRAIG
Shot Type: CU

Idiom Builder

Saved Idioms: Goldfish

Intensify emotion, peaks and valleys, performance slow, emphasize character, zoom consistent, zoom in/out, zoom in/out gradually, short lines, start wide, avoid jump cuts, speaker visible, performance fast

Name: Goldfish
Description:
Tempo Parameters
Start Buffer: 0 End Buffer: 0 Safe Buffer: 0
Generate

goldfish

Script

LINDA
What time do you think she'll call?

CRAIG
Not today. Remember?

LINDA
Oh, yeah.

CRAIG
She called yesterday.

LINDA
Where was she going to dinner tonight?

Edit

Clips

goldfish

Script	Edit	Clips
<p>LINDA What time do you think she'll call?</p>		
<p>CRAIG Not today. Remember?</p>		
<p>LINDA Oh, yeah.</p>		
<p>CRAIG She called yesterday.</p>		
<p>LINDA Where was she going to dinner tonight?</p>		

goldfish

Script	Edit	Clips
<p>LINDA What time do you think she'll call?</p>		
<p>CRAIG Not today. Remember?</p>		
<p>LINDA Oh, yeah.</p>		
<p>CRAIG She called yesterday.</p>		
<p>LINDA Where was she going to dinner tonight?</p>		

goldfish

Script

LINDA
What time do you think she'll call?

CRAIG
Not today. Remember?

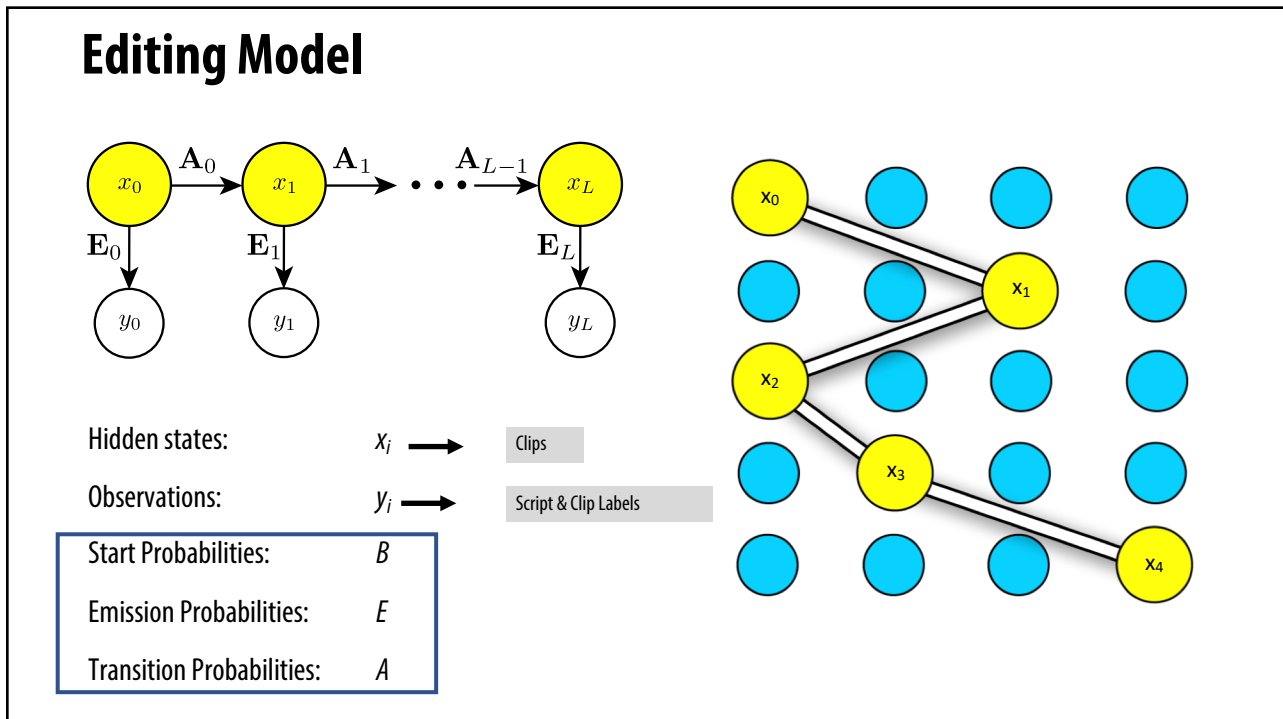
LINDA
Oh, yeah.

CRAIG
She called yesterday.

LINDA
Where was she going to dinner tonight?

Edit

Clips

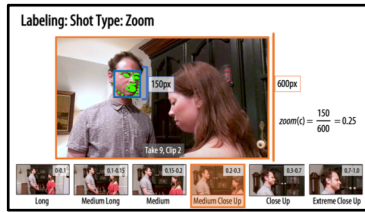


Start Probabilities (B)

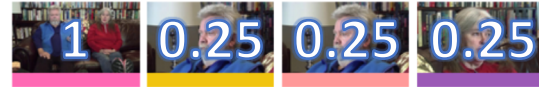
Idiom: Start Wide

Function:
$$B(c) \propto \frac{1}{\text{zoom}(c)}$$

Label: **Shot Type: Zoom**



Clips



Emission Probabilities (E)

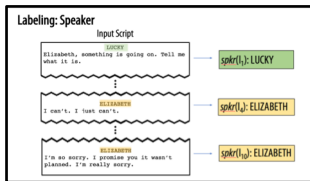
Idiom: Speaker Visible

Function:
$$E(c) \propto \begin{cases} 1, & SpVis(c) \\ \epsilon, & \text{not } SpVis(c) \end{cases}$$

Labels:

Visible Speakers

Speaker



Clips

Linda



Emission Probabilities (E)

Idiom: Speaker Visible

Function:

$$E(c) \propto \begin{cases} 1, & SpVis(c) \\ \epsilon, & \text{not } SpVis(c) \end{cases}$$

Labels:

Visible Speakers

Speaker

Labeling: Visible Speaker

Labeling: Speaker

Input Script

ELIZABETH: Elizabeth, something is going on. Tell me what it is.

LUCKY: I don't know.

ELIZABETH: I don't know.

LUCKY: I don't know.

ELIZABETH: I'm so sorry. I promise you it wasn't planned. I'm really sorry.

Labels: speaker(): LUCKY, speaker(): ELIZABETH, speaker(): ELIZABETH

Clips



LINDA
What time do you think she'll call?

Emission Probabilities (E)

Idiom: Speaker Visible

Function:

$$E(c) \propto \begin{cases} 1, & SpVis(c) \\ \epsilon, & \text{not } SpVis(c) \end{cases}$$

Labels:

Visible Speakers

Speaker

Labeling: Visible Speaker

Labeling: Speaker

Input Script

ELIZABETH: Elizabeth, something is going on. Tell me what it is.

LUCKY: I don't know.

ELIZABETH: I don't know.

LUCKY: I don't know.

ELIZABETH: I'm so sorry. I promise you it wasn't planned. I'm really sorry.

Labels: speaker(): LUCKY, speaker(): ELIZABETH, speaker(): ELIZABETH

Clips

Linda				
Craig				
Linda				
Craig				
Linda				

Labels: 1, ε

Transition Probabilities (A)

Idiom: Avoid jump cuts

Function:

$$A(c_a, c_b) \propto \begin{cases} 1, & \text{not jump}(c_a, c_b) \\ \epsilon, & \text{jump}(c_a, c_b) \end{cases}$$

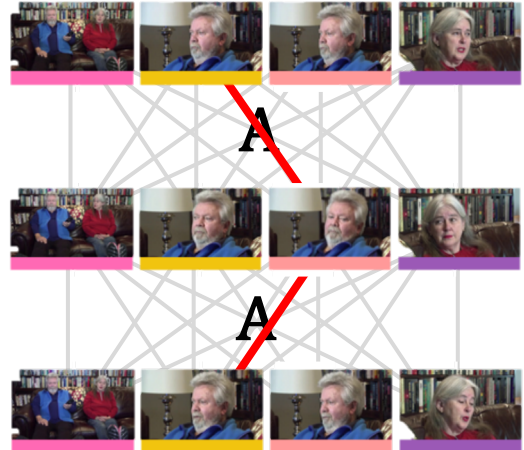
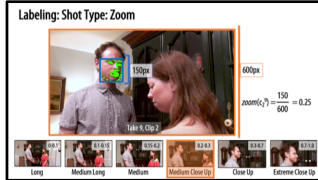
$\text{jump}(c_a, c_b)$ **True** if c_a and c_b have same zoom and visible performers
False, otherwise

Labels:

Visible Speakers



Shot Type: Zoom



Transition Probabilities (A)

Idiom: Avoid jump cuts

Function:

Labels:

Visible



Transition Probabilities (A)

Idiom: Avoid jump cuts

Function:

$$A(c_a, c_b) \propto \begin{cases} 1, & \text{not } jump(c_a, c_b) \\ \epsilon, & jump(c_a, c_b) \end{cases}$$

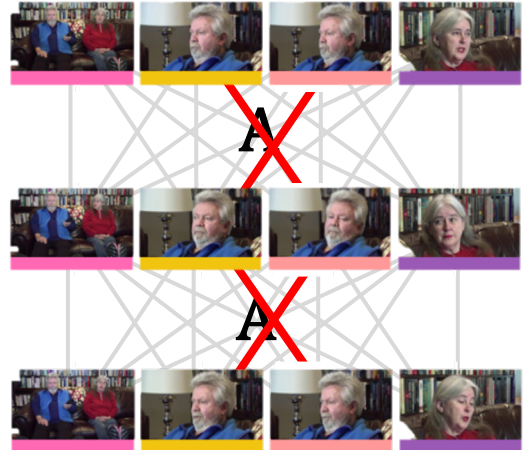
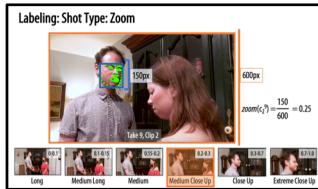
$jump(c_a, c_b)$ **True** if c_a and c_b have same zoom and visible performers
False, otherwise

Labels:

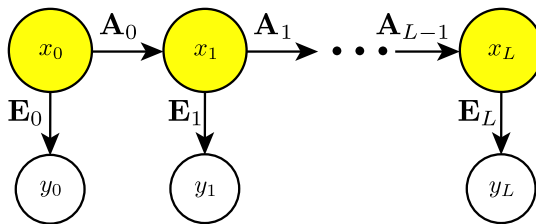
Visible Speakers



Shot Type: Zoom



Combining Idioms



Vectorized Idiom $J = \begin{bmatrix} B \\ E \\ A \end{bmatrix}$

Hidden states: x_i

Observations: y_i

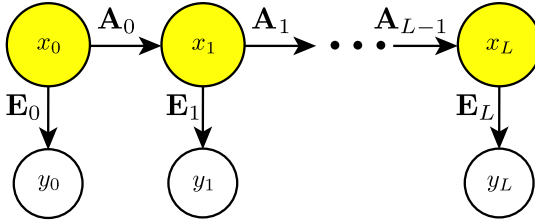
Start Probabilities: B

Emission Probabilities: E

Transition Probabilities: A

Idiom

Combining Idioms



Hidden states: x_i

Observations: y_i

Start Probabilities: B

Emission Probabilities: E

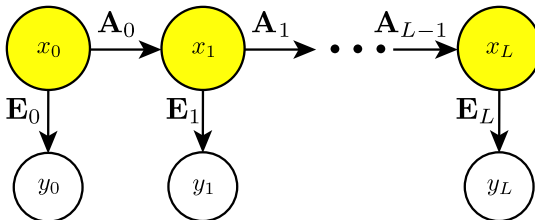
Transition Probabilities: A

Idiom

Vectorized Idiom $\mathcal{J} = \begin{bmatrix} B \\ E \\ A \end{bmatrix}$

Idiom Combination $\mathcal{J}_1 \mathcal{J}_2 = \begin{bmatrix} B_1 \circ B_2 \\ E_1 \circ E_2 \\ A_1 \circ A_2 \end{bmatrix}$

Combining Idioms



Hidden states: x_i

Observations: y_i

Start Probabilities: B

Emission Probabilities: E

Transition Probabilities: A

Idiom

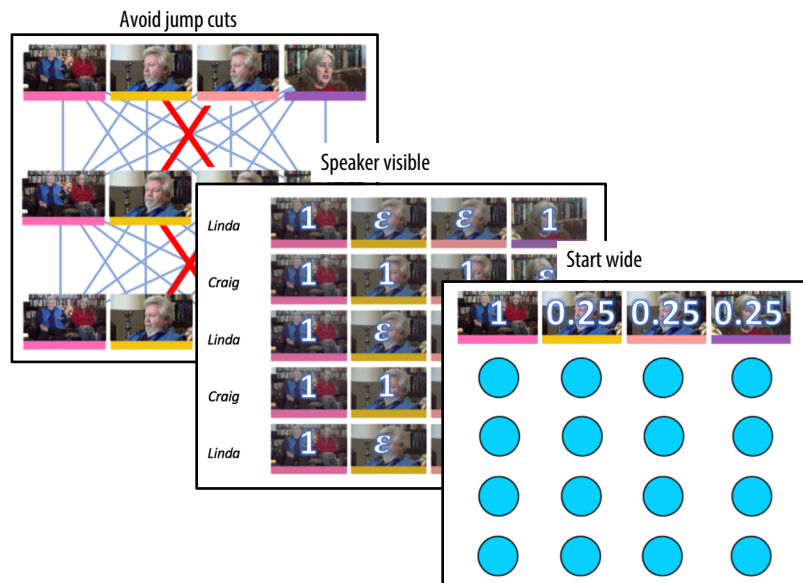
Vectorized Idiom $\mathcal{J} = \begin{bmatrix} B \\ E \\ A \end{bmatrix}$

Idiom Combination $\mathcal{J}_1 \mathcal{J}_2 = \begin{bmatrix} B_1 \circ B_2 \\ E_1 \circ E_2 \\ A_1 \circ A_2 \end{bmatrix}$

Weighted Combination $\mathcal{J}_1^\alpha \mathcal{J}_2^\beta = \begin{bmatrix} B_1^\alpha \circ B_2^\beta \\ E_1^\alpha \circ E_2^\beta \\ A_1^\alpha \circ A_2^\beta \end{bmatrix}$

Additional Idioms

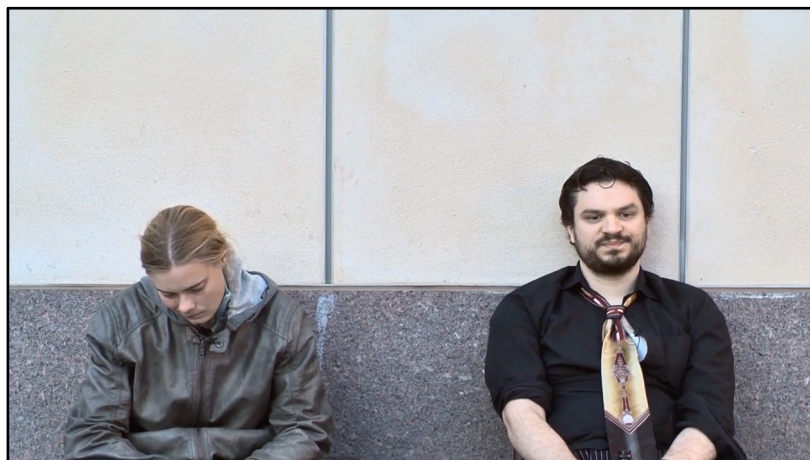
1. *Avoid jump cuts*
2. *Speaker visible*
3. *Start wide*
4. Change zoom gradually
5. Emphasize character
6. Intensify emotion
7. Mirror position
8. Peaks and valleys
9. Performance fast/slow
10. Performance loud/quiet
11. Short lines



Fired: Cinematic Style

Style 1: Cinematic Style

- *Start wide*
- *Speaker visible*
- *Avoid jump cuts*
- *Performance fast*



Takes: 9

Total takes duration: 16.8m

Fluffles: Cinematic Style

Style 1: Cinematic Style

- *Start wide*
- *Speaker visible*
- *Avoid jump cuts*
- *Performance fast*
- ***Emphasize Stacy***



Takes: 15

Total takes duration: 18.3m

Style 2: YouTube Style

Fluffles: YouTube Style

Style 2: YouTube Style

- *Speaker visible*
- *Performance fast*
- *Zoom consistent*
- *Avoid jump cuts*
 - *Weight -1*



Takes: 15

Total takes duration: 18.3m

Comparing Edits

Style 1: Cinematic Style

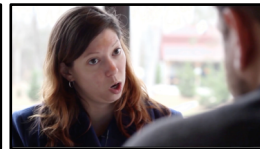
- *Start wide*
- *Speaker visible*
- *Avoid jump cuts*
- *Performance fast*
- *Emphasize Stacy*



Line 1



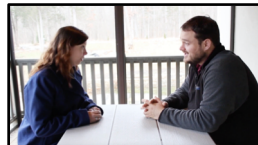
Line 2



Line 3

Style 2: YouTube Style

- *Speaker visible*
- *Performance fast*
- *Zoom consistent*
- *Avoid jump cuts*
 - *Weight -1*



Line 1



Line 2



Line 3

Evaluation

Scene
Baby Steps
Fired
Fluffles
Friend
Goldfish
Krispies
Princess Bride
Social Network

Evaluation

Scene	Inputs	
	# Takes	Duration
Baby Steps	8	9.1m
Fired	9	16.8m
Fluffles	15	18.3m
Friend	8	14.4m
Goldfish	8	9.6m
Krispies	15	14.7m
Princess Bride	15	13.3m
Social Network	13	7.6m
Summary	8-15	7.6-18.3m

Evaluation

Scene	Inputs		Batch Pre-processing
	# Takes	Duration	
Baby Steps	8	9.1m	155m
Fired	9	16.8m	165m
Fluffles	15	18.3m	209m
Friend	8	14.4m	201m
Goldfish	8	9.6m	110m
Krispies	15	14.7m	171m
Princess Bride	15	13.3m	217m
Social Network	13	7.6m	149m
Summary	8-15	7.6-18.3m	110-217m

Evaluation

Scene	Inputs		Batch Pre-processing	Editing
	# Takes	Duration		Auto
Baby Steps	8	9.1m	155m	2s
Fired	9	16.8m	165m	2s
Fluffles	15	18.3m	209m	3s
Friend	8	14.4m	201m	2s
Goldfish	8	9.6m	110m	2s
Krispies	15	14.7m	171m	2s
Princess Bride	15	13.3m	217m	2s
Social Network	13	7.6m	149m	2s
Summary	8-15	7.6-18.3m	110-217m	2-3s

Evaluation

Scene	Inputs		Batch Pre-processing	Editing	
	# Takes	Duration		Auto	Hand
Baby Steps	8	9.1m	155m	2s	105m
Fired	9	16.8m	165m	2s	105m
Fluffles	15	18.3m	209m	3s	180m
Friend	8	14.4m	201m	2s	135m
Goldfish	8	9.6m	110m	2s	105m
Krispies	15	14.7m	171m	2s	90m
Princess Bride	15	13.3m	217m	2s	235m
Social Network	13	7.6m	149m	2s	90m
Summary	8-15	7.6-18.3m	110-217m	2-3s	90-235m

Evaluation

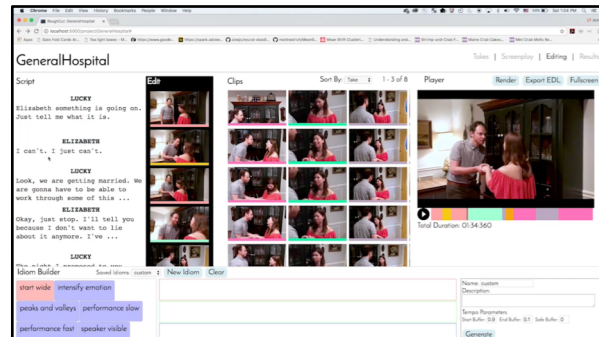
Scene	Inputs		Batch Pre-processing	Editing	
	# Takes	Duration		Auto	Hand
Baby Steps	8	9.1m	155m	2s	105m
Fired	9	16.8m	165m	2s	105m
Fluffles	15	18.3m	209m	3s	180m
Friend	8	14.4m	201m	2s	135m
Goldfish	8	9.6m	110m	2s	105m
Krispies	15	14.7m	171m	2s	90m
Princess Bride	15	13.3m	217m	2s	235m
Social Network	13	7.6m	149m	2s	90m
Summary	8-15	7.6-18.3m	110-217m	2-3s	90-235m

Takeaways:

- Focus user time on creative decisions
- Quickly iterate on different styles

Conclusion & Future Work

- **Structure** enables more powerful editing tools
- Script alignment can provide this structure
- What other types of structure can we use?



Further thoughts

1. We took a heuristics-based approach as opposed to a learning approach – why?
2. What other parts could we have automated? Why did we choose to automate what we did?
3. How else could we have evaluated our system?
4. We did not discuss our system being for “novices” or “experts.” Who do you think are the intended users?

Additional Results & Scenes

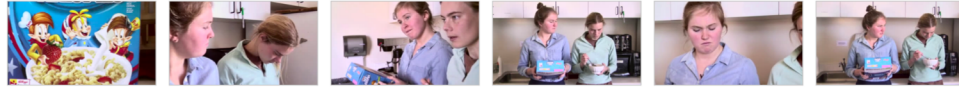
"Baby Steps"



Idiom: Start Wide, Speaker Visible, Avoid Jump Cuts

Objective: Establish the scene, and then show the actors who are speaking. Avoid jump cuts, but allow any kinds of zoom jumps otherwise.

"Krispies"



Idiom: Start Wide, Speaker Visible, Avoid Jump Cuts, Emphasize Speaker-B

Objective: Establish the scene, and then show the actors who are speaking. Avoid jump cuts. Choose a dominant speaker -- Betty.

"Social Network"



Idiom: Start Wide, Speaker Visible, Avoid Jump Cuts, Emphasize Character-Mark

Objective: Establish the scene, and then show the actors who are speaking. Avoid jump cuts. Choose a dominant speaker -- Mark.

"Princess Bride"



Idiom: Start Uniform, Speaker Visible, Performance Slow, Avoid Jump Cuts

Objective: Show the speaker and choose the longest edit while introducing no jump cuts.

Additional results and raw materials at:

<https://graphics.stanford.edu/papers/roughcut/>

RIP Fluffles: <http://www.fluffles.org>