# Automatic Summarization (and other stuff)


Taylor Berg-Kirkpatrick
CS 288 UC Berkeley


# Summarization

Berkeley NLP

# Summarization

Berkeley NLP

# Multidocument Summarization



Berkeley NLP

# Sentence Extraction




Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.

# Sentence Extraction




Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.

Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail.

Berkeley NLP

# Sentence Extraction




Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.

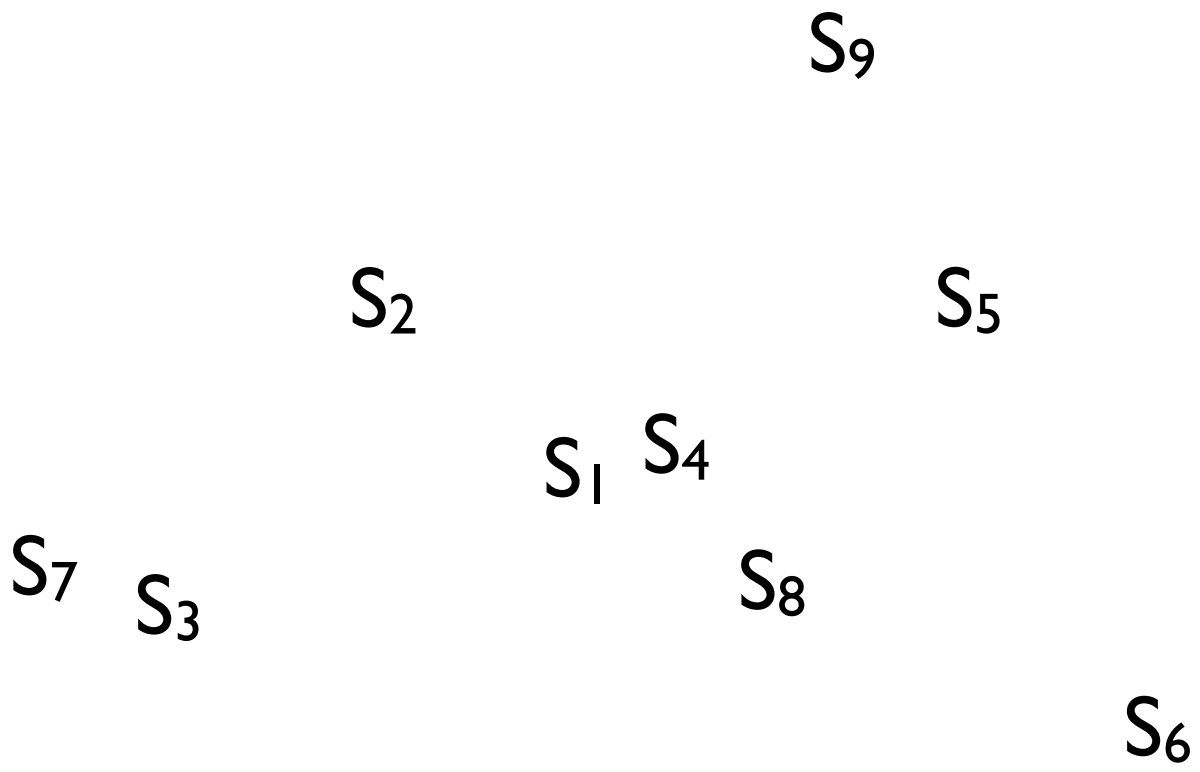
Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail.

The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.


# Maximum Marginal Relevance

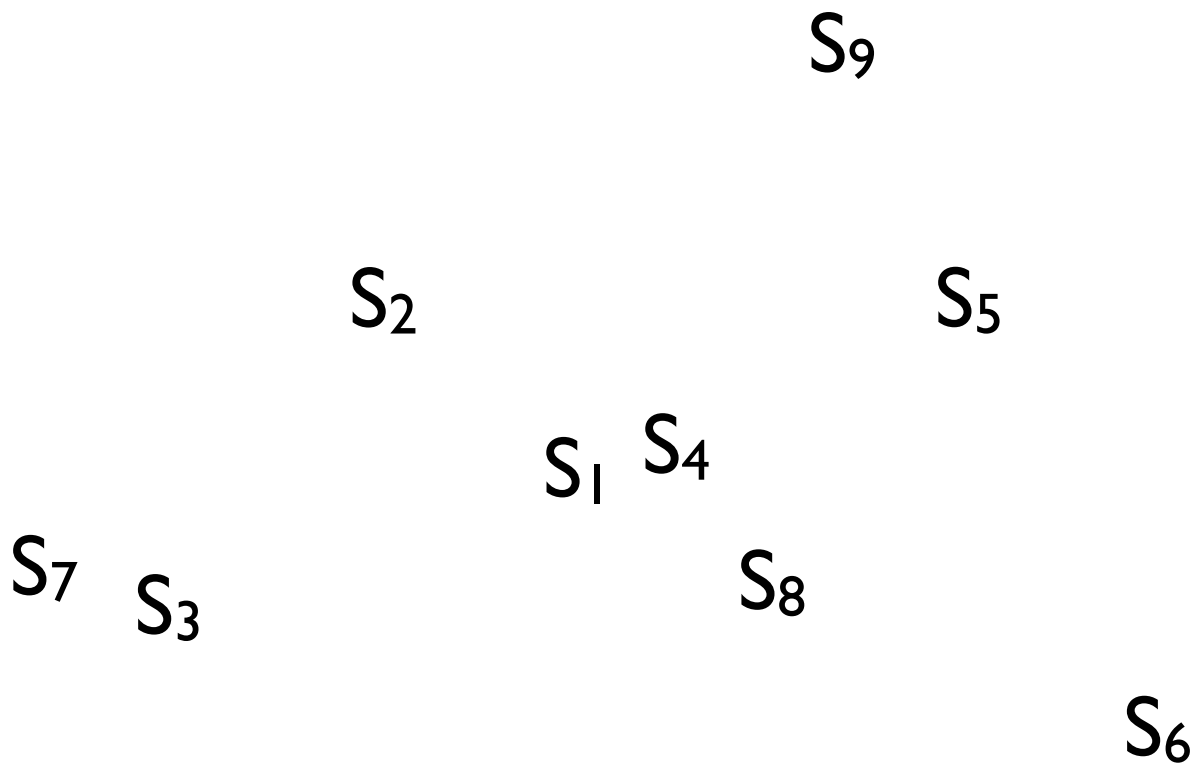
Berkeley
N L P

# Maximum Marginal Relevance



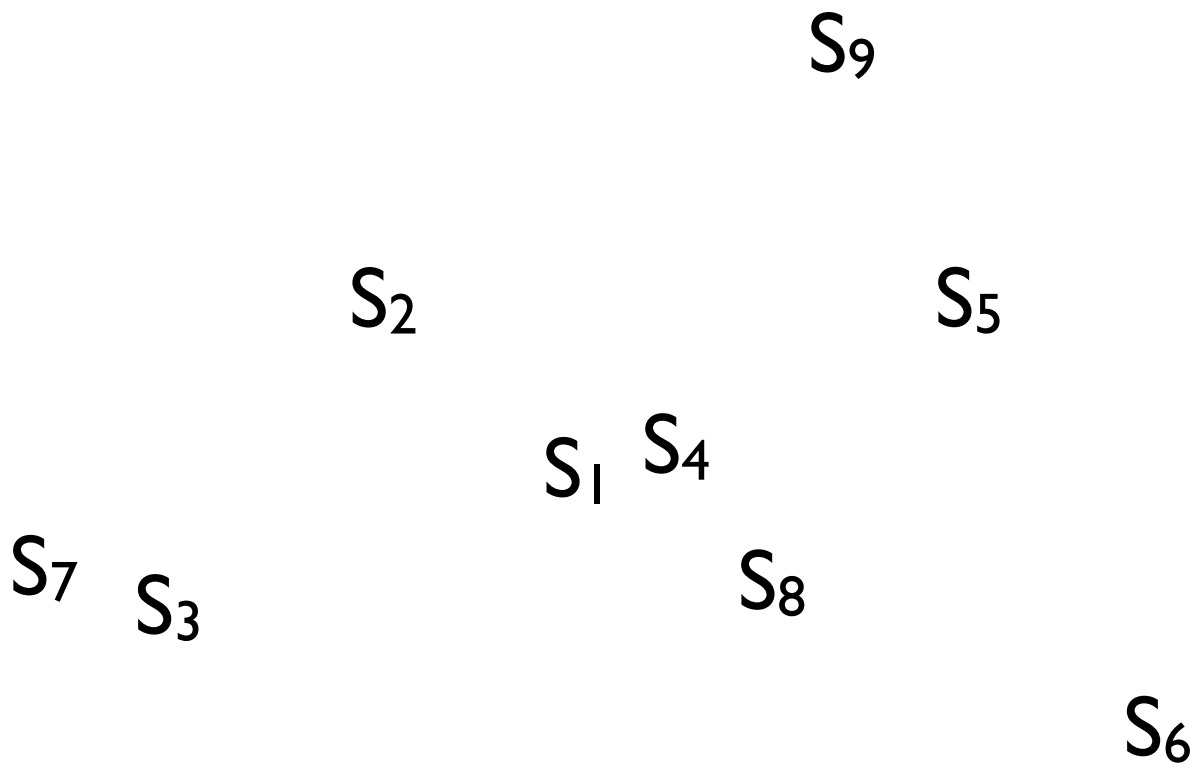
Berkeley
N L P

# Maximum Marginal Relevance



Berkeley NLP

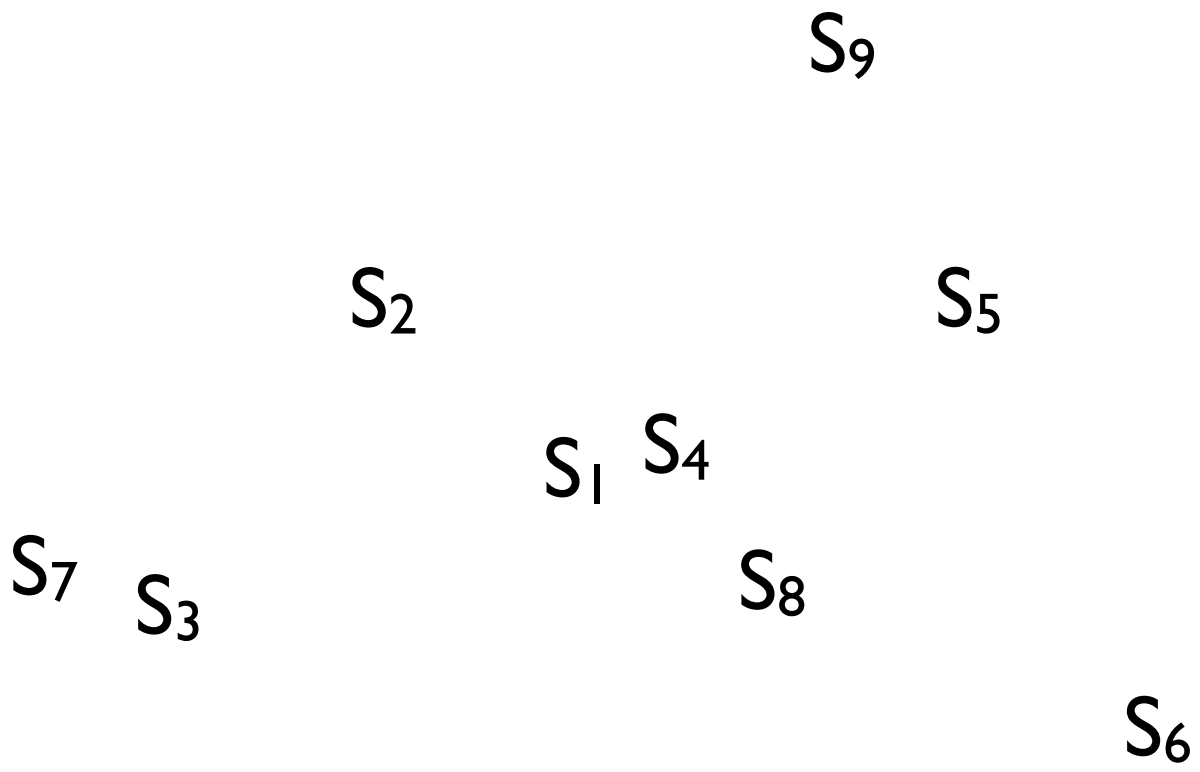
# Maximum Marginal Relevance




$S_1$: She traveled to France on Friday.

Berkeley NLP

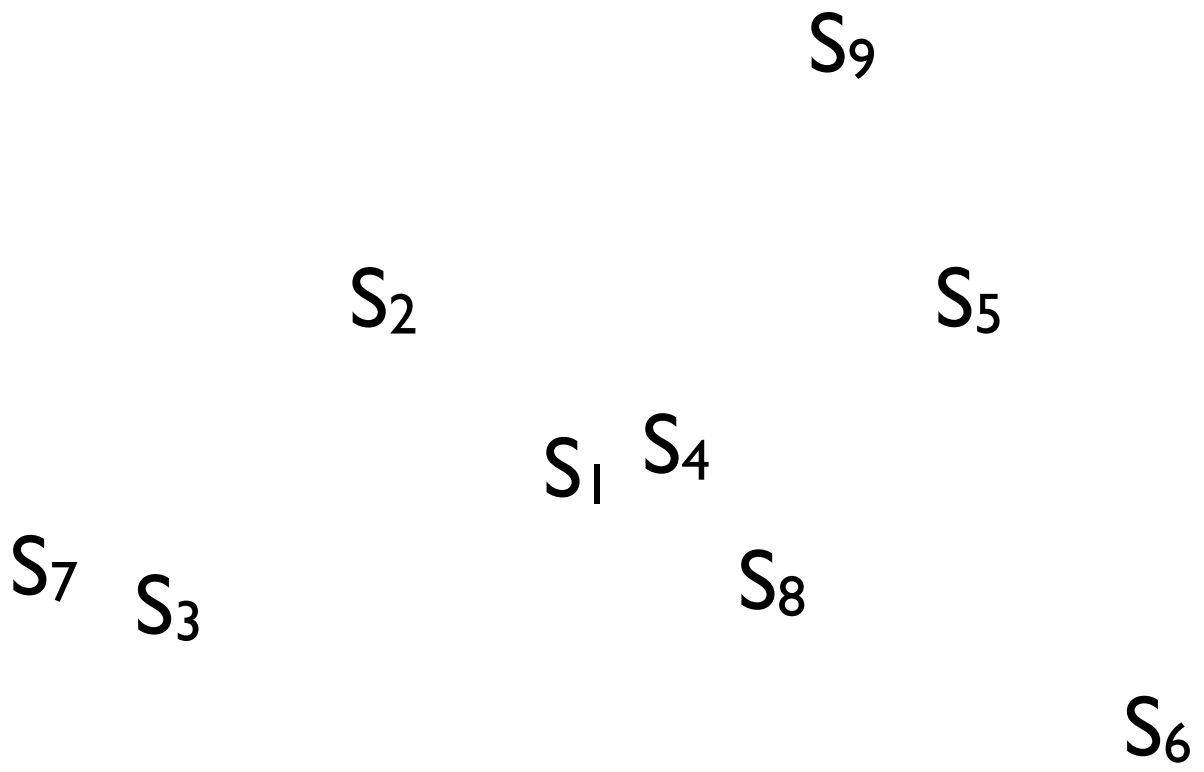
# Maximum Marginal Relevance




$S_1$: She traveled to France on Friday.

Berkeley NLP

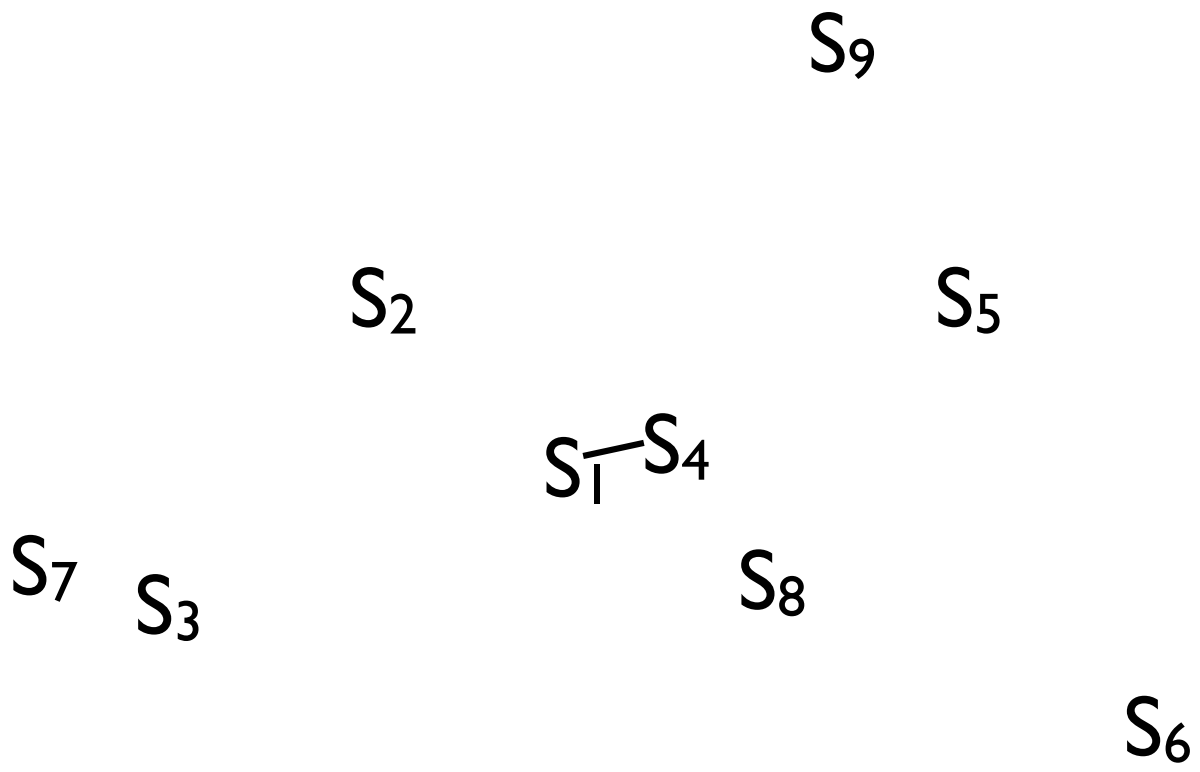
# Maximum Marginal Relevance




S1: She traveled to France on Friday.

S4: On Friday, She took a trip to France.

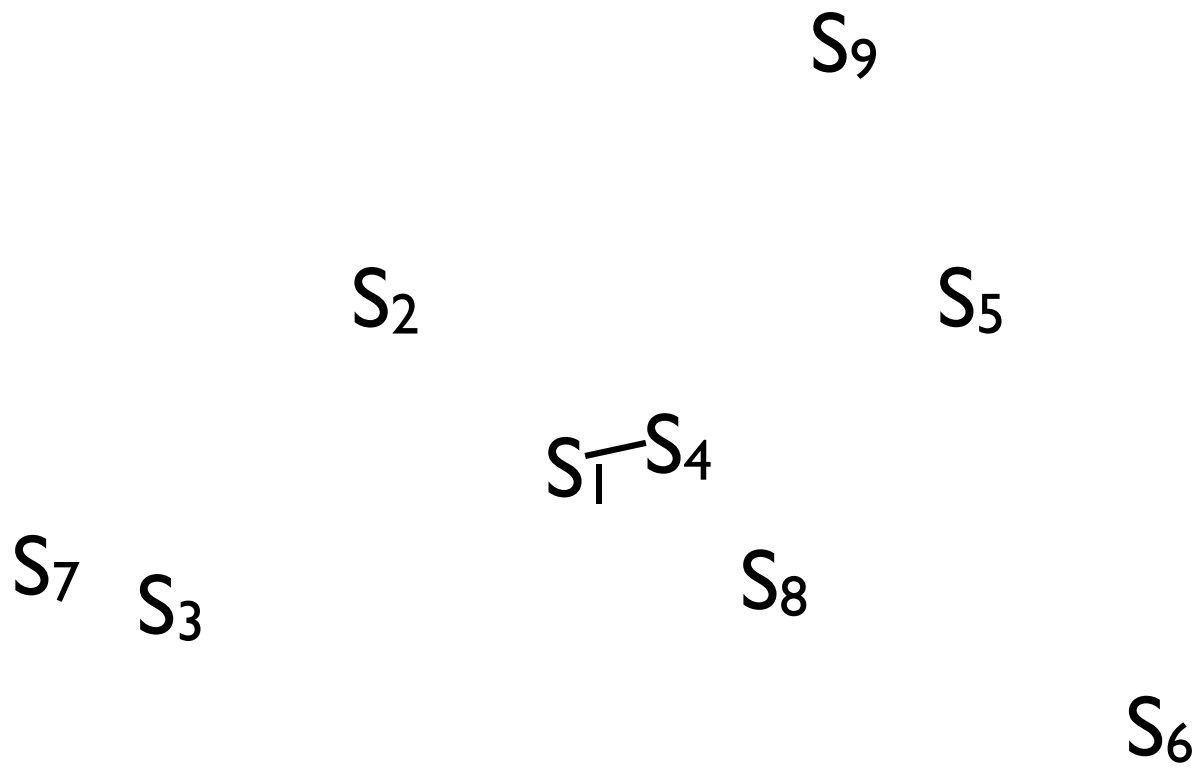
# Maximum Marginal Relevance




$S_1$: She traveled to France on Friday.

$S_4$: On Friday, She took a trip to France.

# Maximum Marginal Relevance

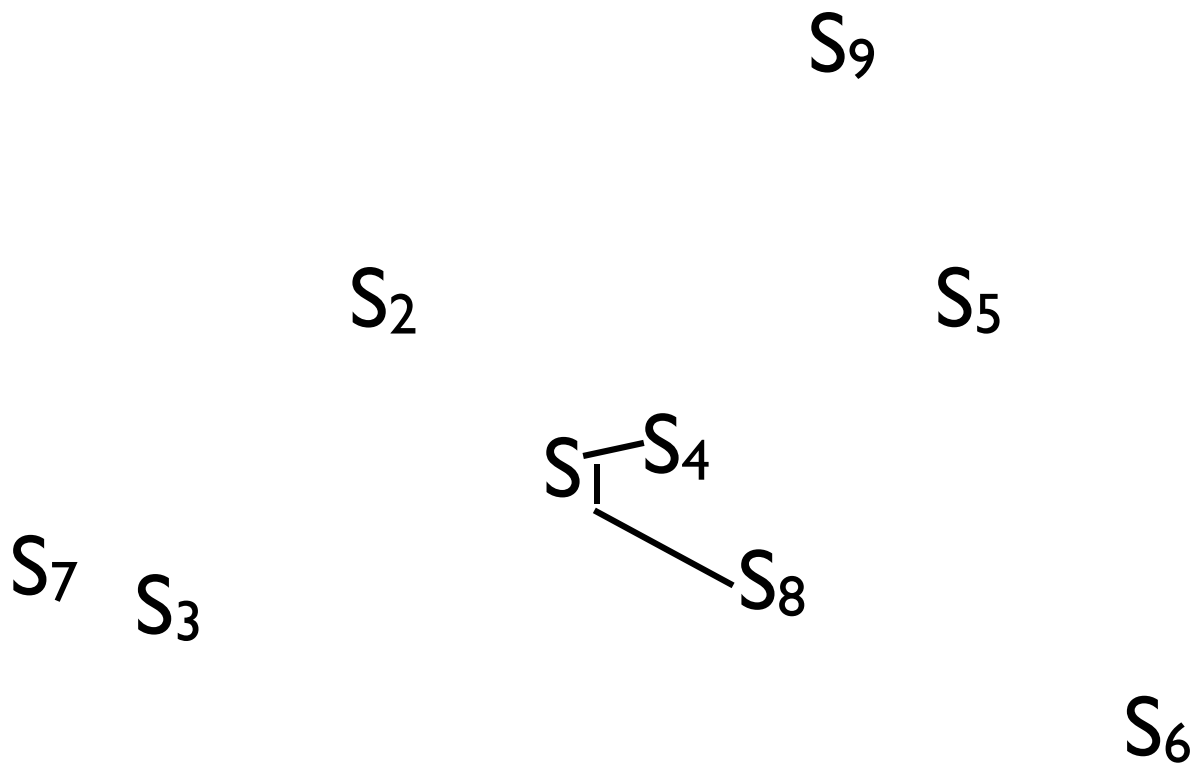



$S_1$: She traveled to France on Friday.

$S_4$: On Friday, She took a trip to France.

$S_8$: She plans to stay for two weeks.

# Maximum Marginal Relevance



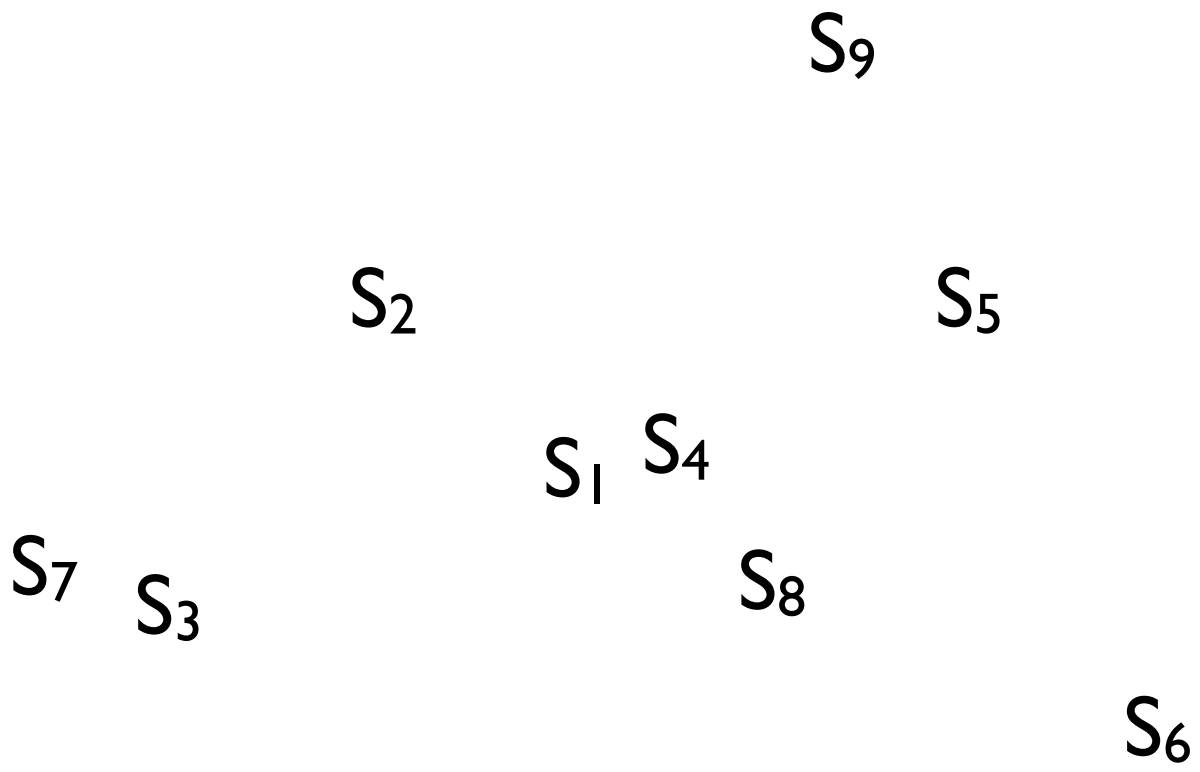

$S_1$: She traveled to France on Friday.

$S_4$: On Friday, She took a trip to France.

$S_8$: She plans to stay for two weeks.

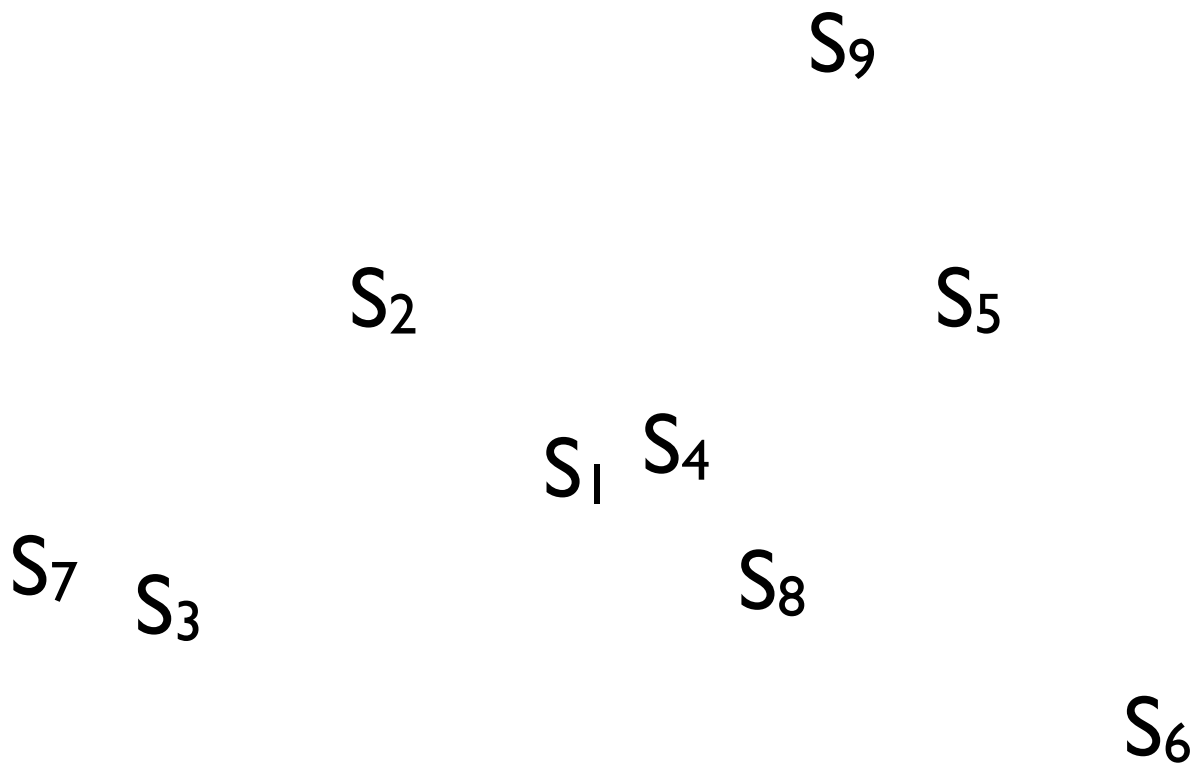
Berkeley
N L P

# Maximum Marginal Relevance



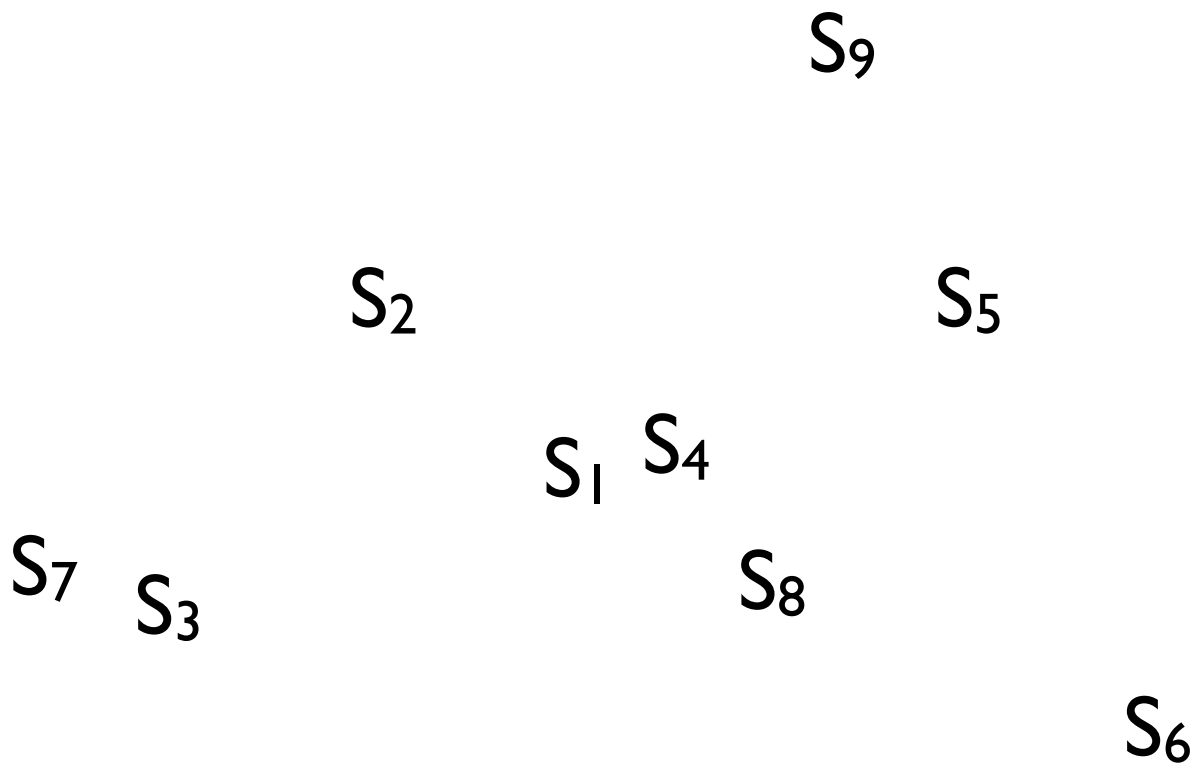
Berkeley
N L P

# Maximum Marginal Relevance

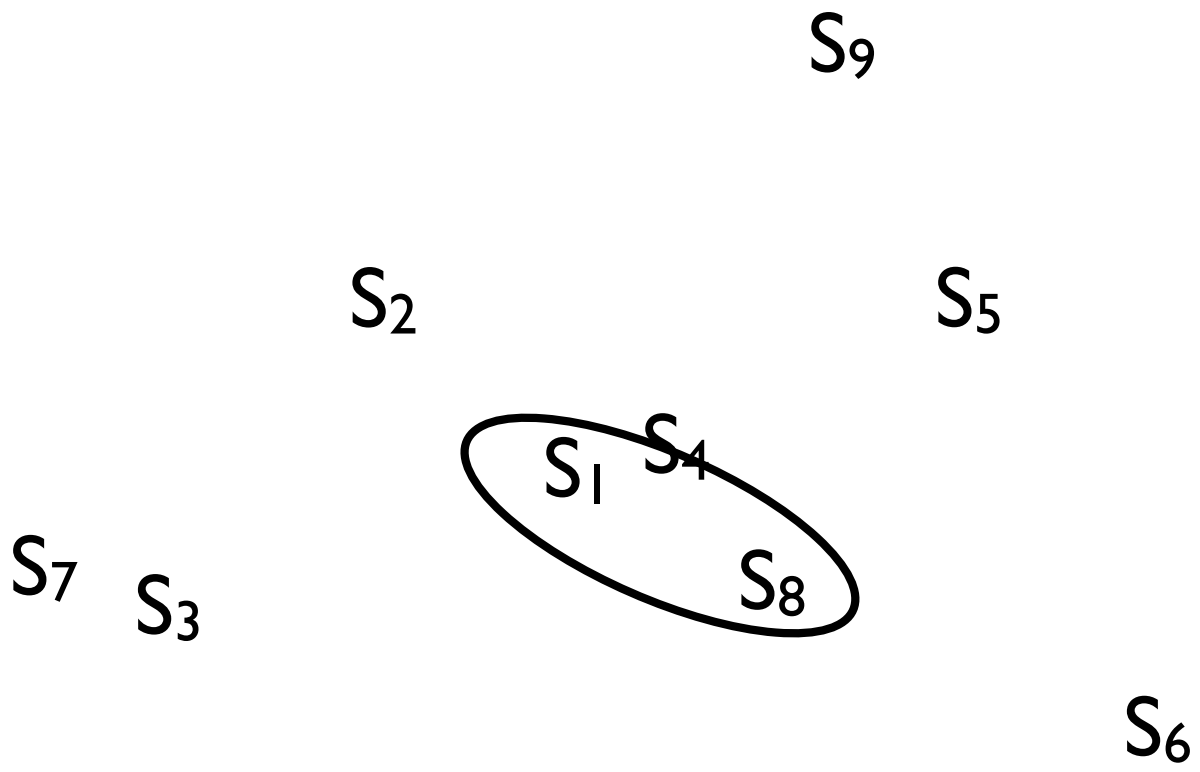


Berkeley
N L P

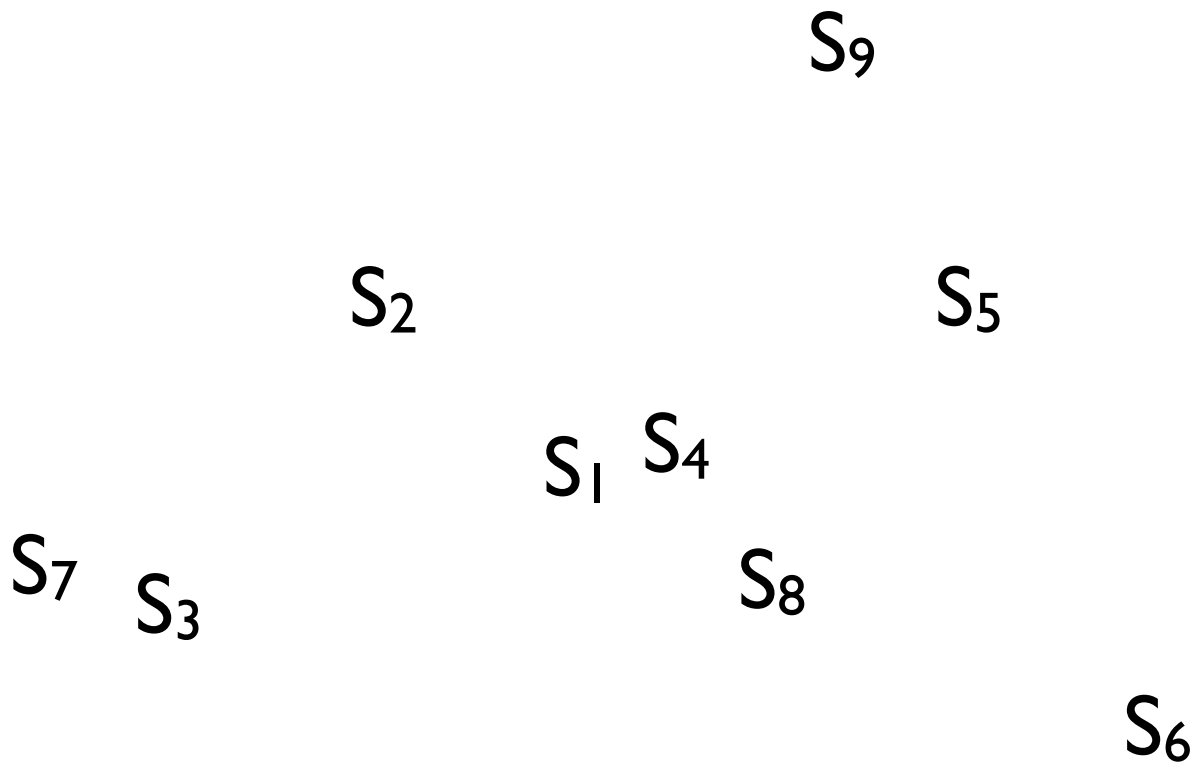
# Maximum Marginal Relevance



# Maximum Marginal Relevance



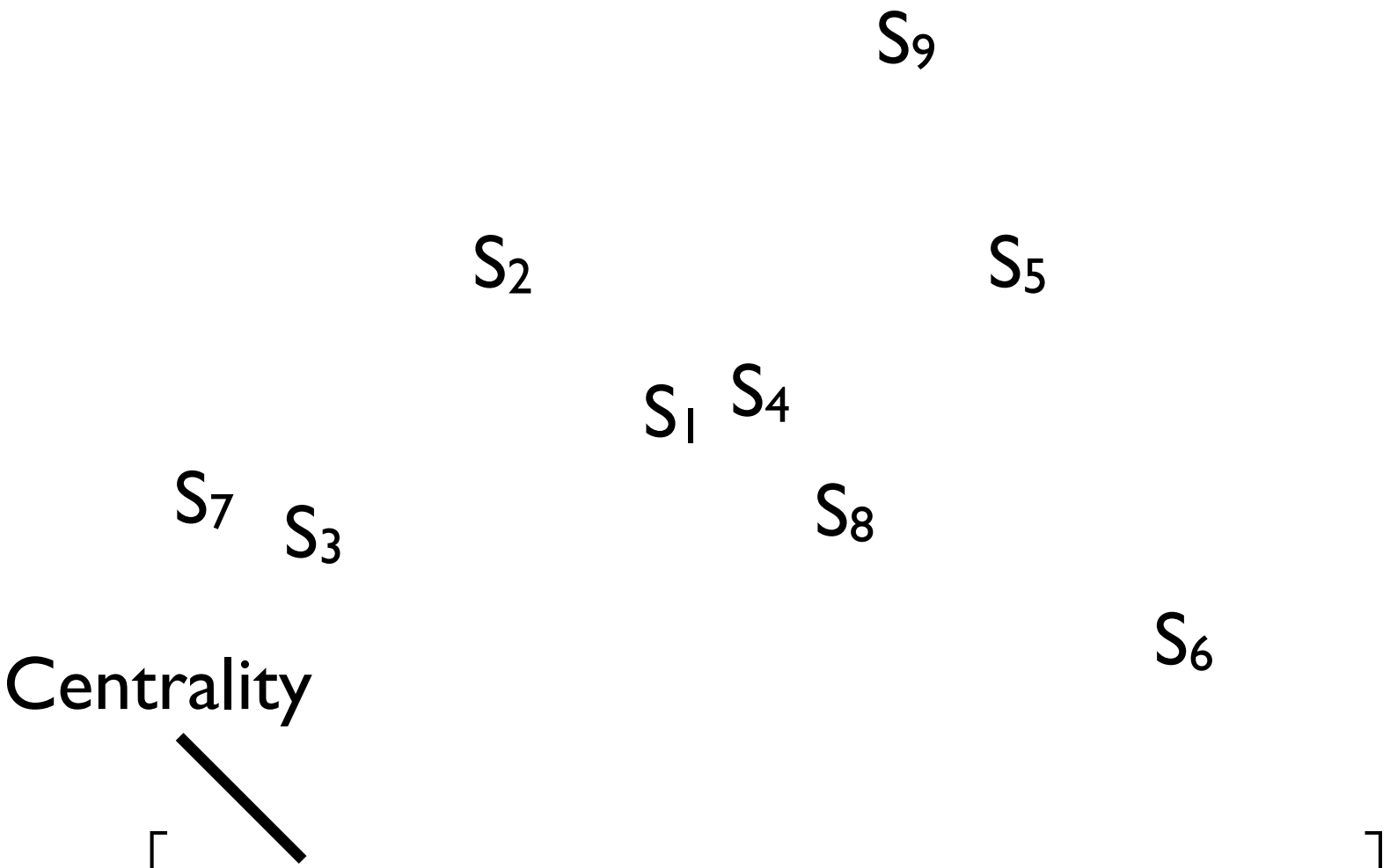
# Maximum Marginal Relevance




$$\underset{i \in D \setminus S}{\operatorname{argmax}} \left[ \lambda \cdot \operatorname{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \left( \operatorname{sim}(S_i, S_j) \right) \right]$$

[Carbonell and Goldstein, 1998]

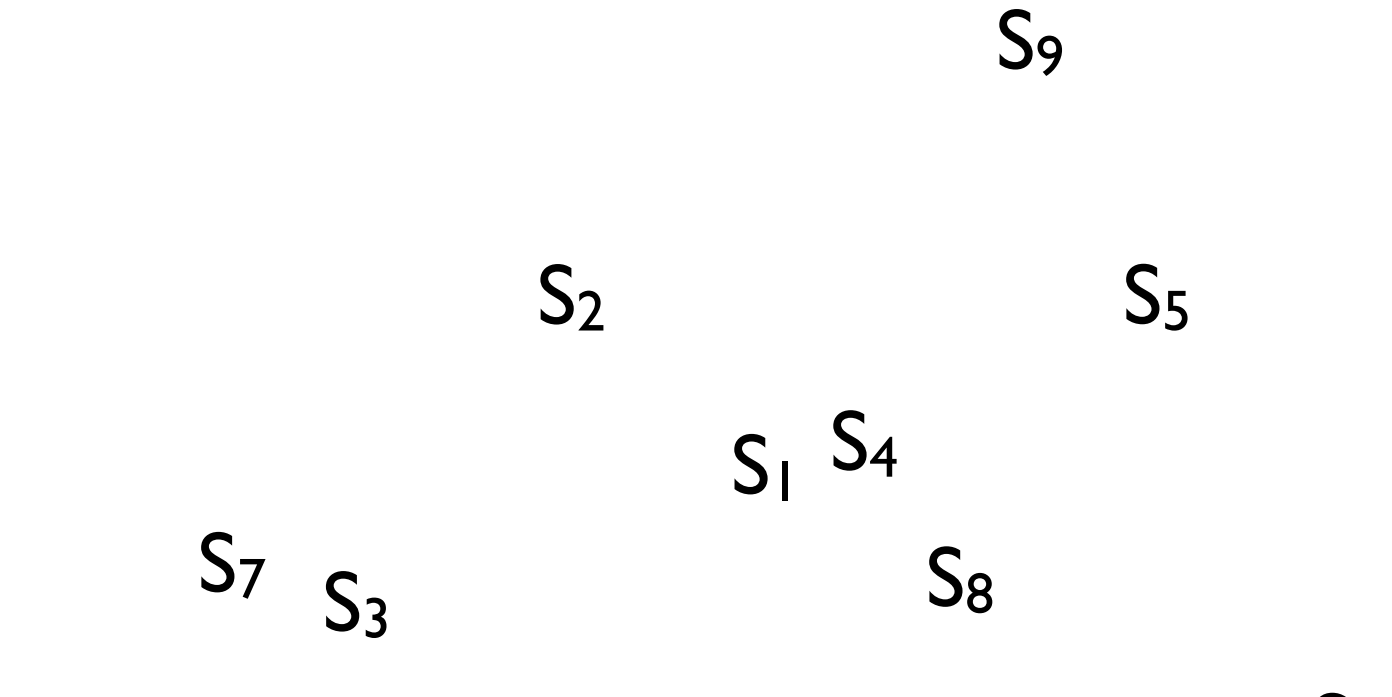
# Maximum Marginal Relevance




$$\underset{i \in D \setminus S}{\operatorname{argmax}} \left[ \lambda \cdot \operatorname{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \left( \operatorname{sim}(S_i, S_j) \right) \right]$$

[Carbonell and Goldstein, 1998]

# Maximum Marginal Relevance




$$\underset{i \in D \setminus S}{\operatorname{argmax}} \left[ \lambda \cdot \operatorname{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \left( \operatorname{sim}(S_i, S_j) \right) \right]$$

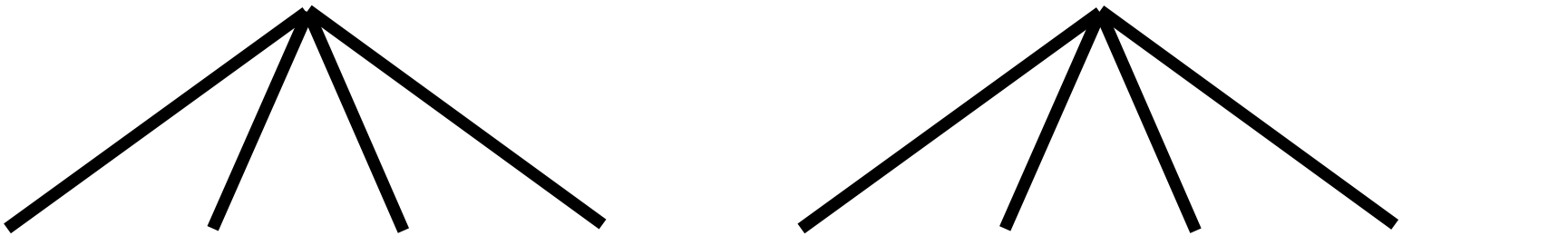
[Carbonell and Goldstein, 1998]

Berkeley NLP

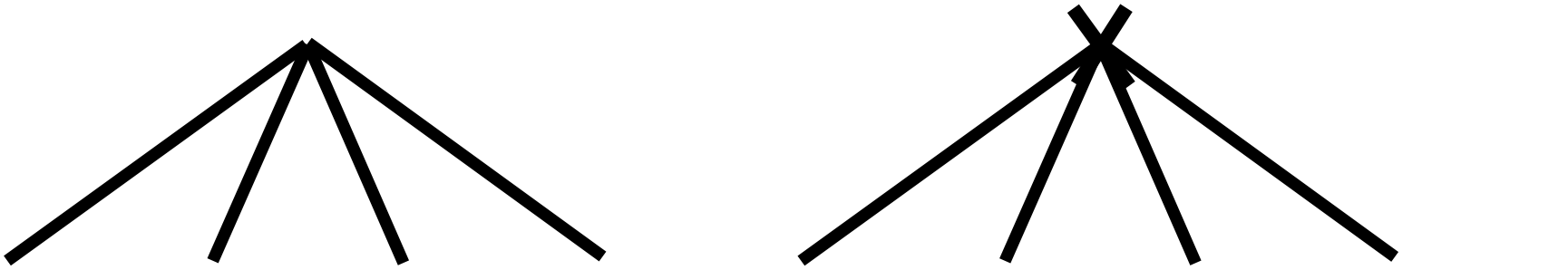
# Max Coverage

She stopped in France. In France she remained.

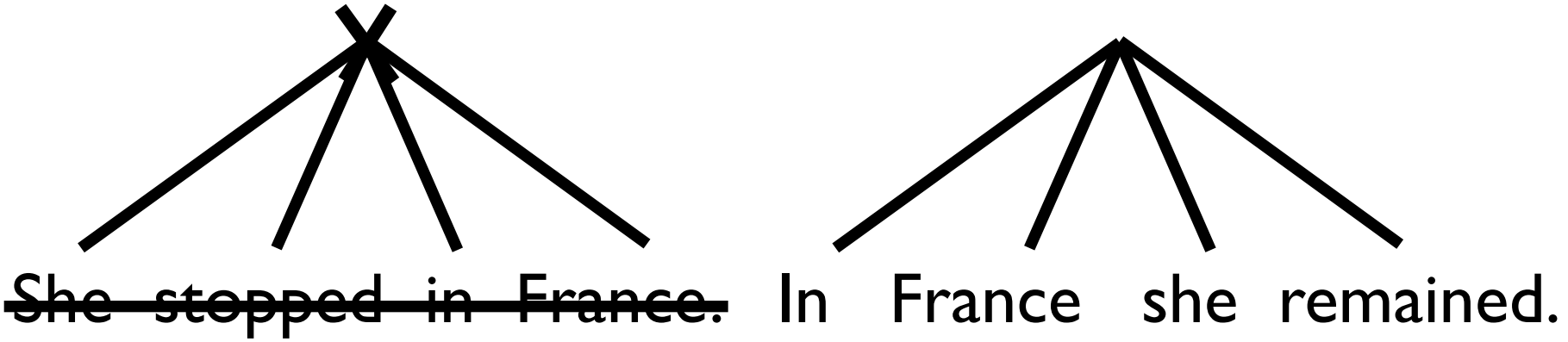
# Max Coverage



# Max Coverage

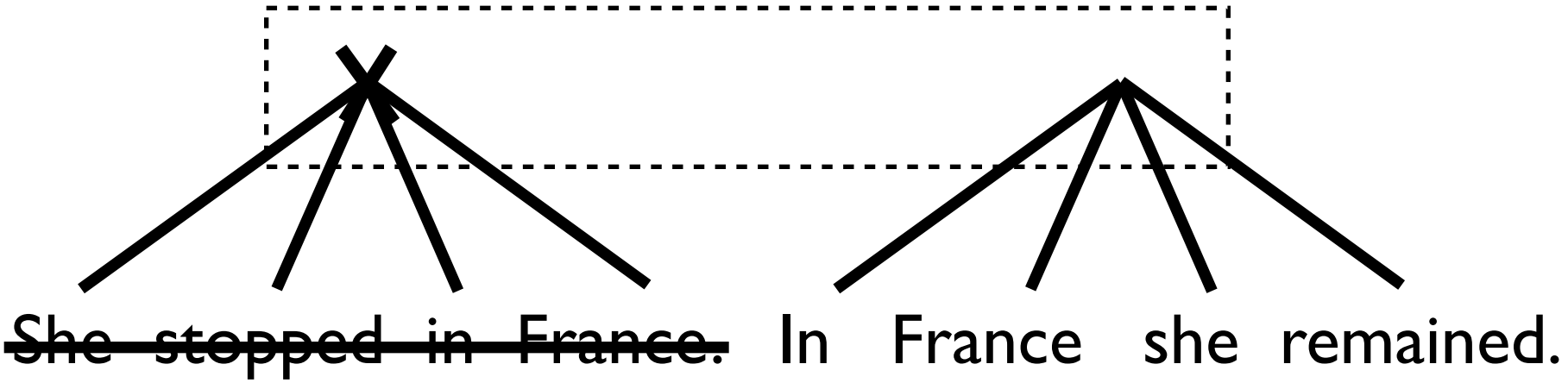


# Max Coverage

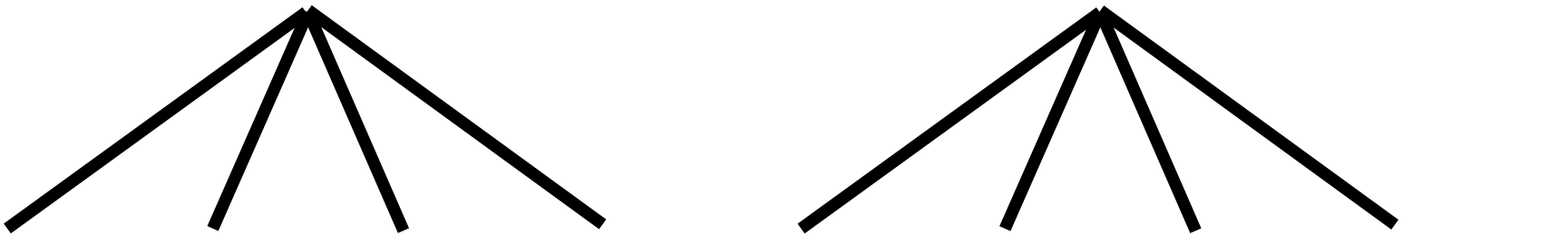


# Max Coverage

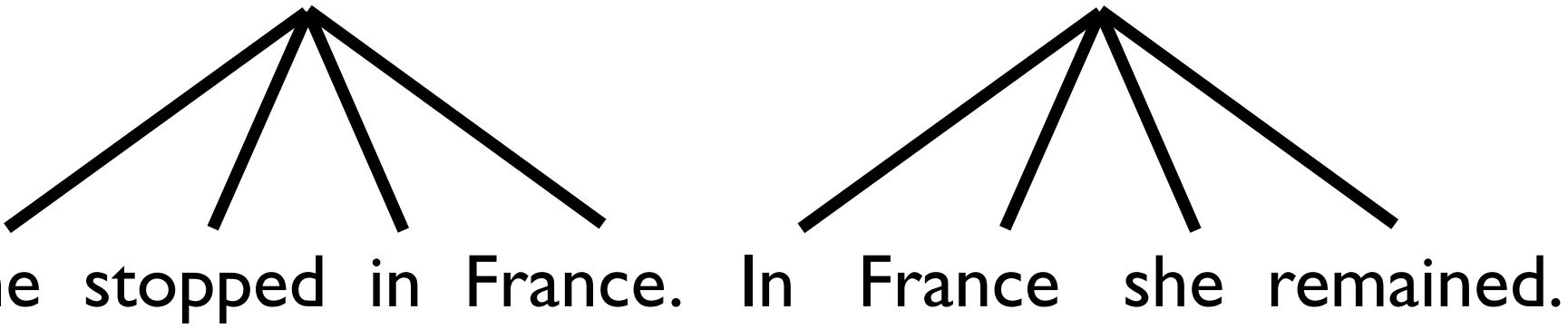
A summary is a set of cuts: $S$



# Max Coverage

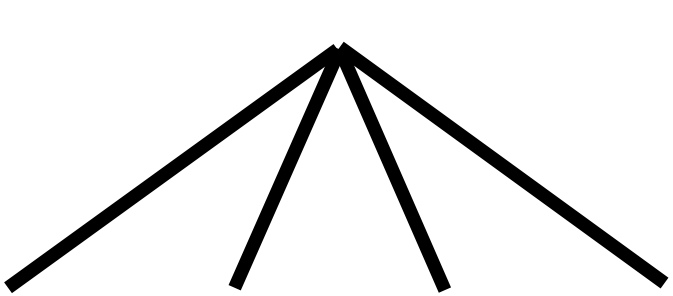


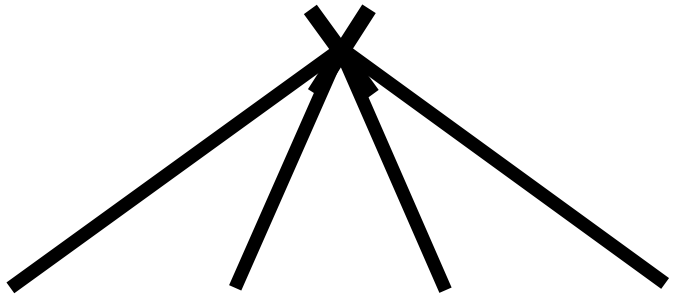
# Max Coverage




(she, stopped) ✓
(stopped, in) ✓
(in, france) ✓
(france, she) ✓
(she, remained) ✓

# Max Coverage



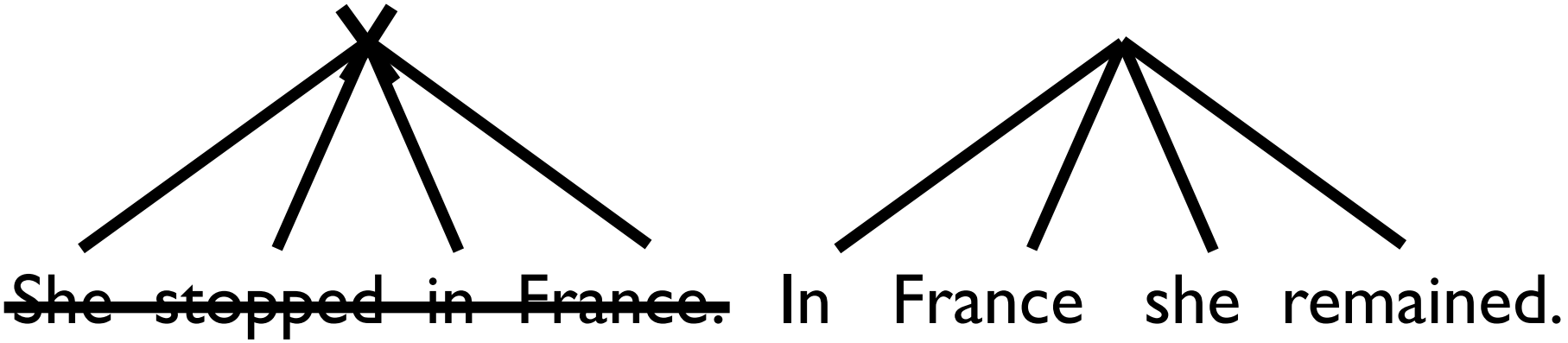


She stopped in France. ~~In France she remained.~~

(she, stopped) ✓ (in, france) ✓ (she, remained)

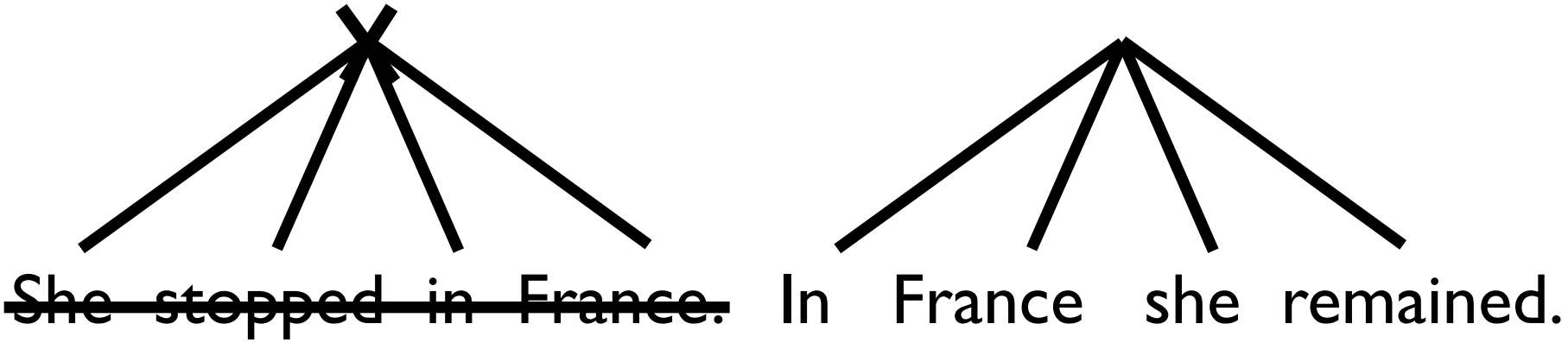
(stopped, in) ✓ (france, she)

# Max Coverage




(she, stopped)
(stopped, in)
(in, france) ✓
(france, she) ✓
(she, remained) ✓

# Max Coverage




(she, stopped) (stopped, in) (in, france) ✓ (france, she) ✓ (she, remained) ✓

Set of bigrams covered by summary: $B(s)$


# Max Coverage

$$\max_{s} \ \text{coverage}(s)$$


# Max Coverage

$$\max_{s} \ \text{coverage}(s)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

Berkeley NLP

# Max Coverage

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

Berkeley NLP

# Max Coverage

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

$$\text{value}(b) = \text{freq}(b)$$

Berkeley NLP

# Max Coverage

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

$$\text{value}(b) = \text{freq}(b)$$

[Gillick and Favre 2008]

Berkeley

N L P

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

[Gillick and Favre 2008]

Berkeley
N L P

# ILP for Decoding

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

Berkeley NLP

# ILP for Decoding

$$\max_{s,z} \sum_{b} z_b \cdot \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

Berkeley NLP

# ILP for Decoding

$$\max_{s,z} \sum_{b} z_b \cdot \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

bigrams in $s$ are covered

Berkeley NLP

# ILP for Decoding

$$\max_{s,z} \sum_{b} z_b \cdot \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

bigrams in $s$ are covered

only bigrams in $s$ are covered

Berkeley NLP

# ILP for Decoding

$$\max_{s,z} \sum_b z_b \cdot \text{value}(b)$$

$$s.t. \quad \sum_n l_n s_n \leq L_{\max}$$

$$\forall_{n,b} \quad s_n Q_{nb} \leq z_b$$

$$\forall_b \quad \sum_n s_n Q_{nb} \geq z_b$$

Berkeley NLP

# Linear Model for Extraction

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

$$\text{value}(b) = \text{freq}(b)$$

Berkeley
N L P

# Linear Model for Extraction

$$\max_{s} \sum_{b \in B(s)} \text{value}(b)$$

$$s.t. \quad \text{length}(s) \leq L_{max}$$

Parameterize using features:

$$\text{value}(b) = w^{\top} f(b)$$

Berkeley NLP

# Sentence Extraction




Lindsay Lohan pleaded not guilty Wednesday to felony grand theft of a 2,500 necklace, a case that could return the troubled starlet to jail rather than the big screen.

Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 and warned that if Lohan was accused of breaking the law while free he would have her held without bail.

The Mean Girls star is due back in court on Feb. 23 an important hearing in which Lohan could opt to end the case early.

# Extraction and Compression




Lindsay Lohan pleaded not guilty ~~Wednesday~~ to felony grand theft of a 2,500 necklace, ~~a case that could return the troubled starlet to jail rather than the big screen.~~ Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 ~~and warned that if Lohan was accused of breaking the law while free he would have her held without bail~~. The Mean Girls star is due back in court on Feb. 23 ~~an important hearing in which Lohan could opt to end the case early.~~

# Extraction and Compression



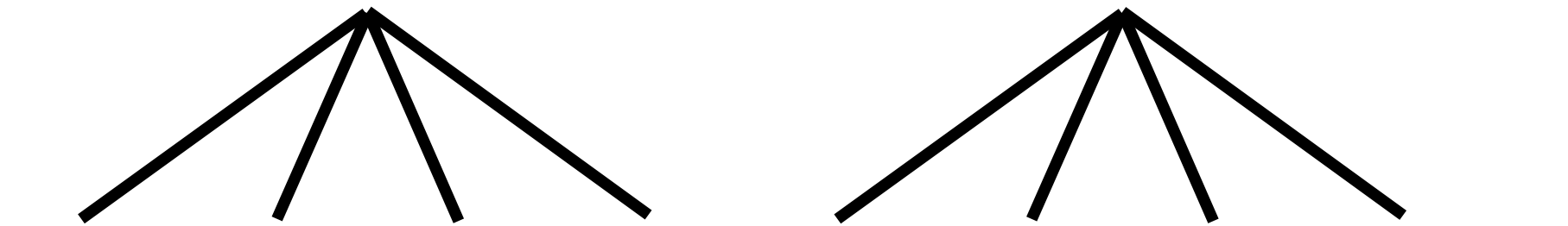

Lindsay Lohan pleaded not guilty ~~Wednesday~~ to felony grand theft of a 2,500 necklace, ~~a case that could return the troubled starlet to jail rather than the big screen.~~ Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 ~~and warned that if Lohan was accused of breaking the law while free he would have her held without bail~~. The Mean Girls star is due back in court on Feb. 23 ~~an important hearing in which Lohan could opt to end the case early.~~

[Martins and Smith 2009]
[Woodsend and Lapata 2010]

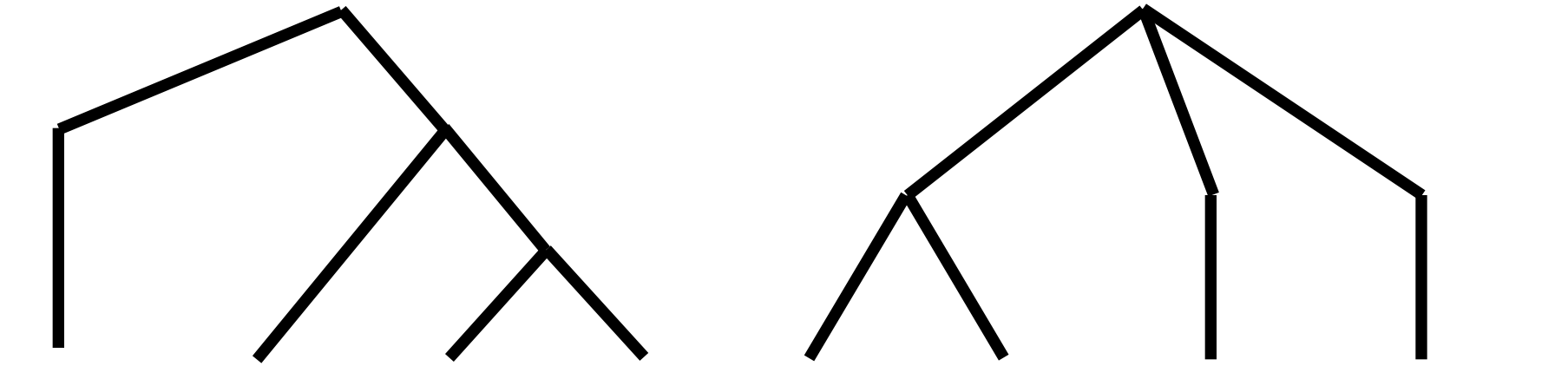

# Extraction and Compression

She stopped in France. ~~In France~~ she remained.

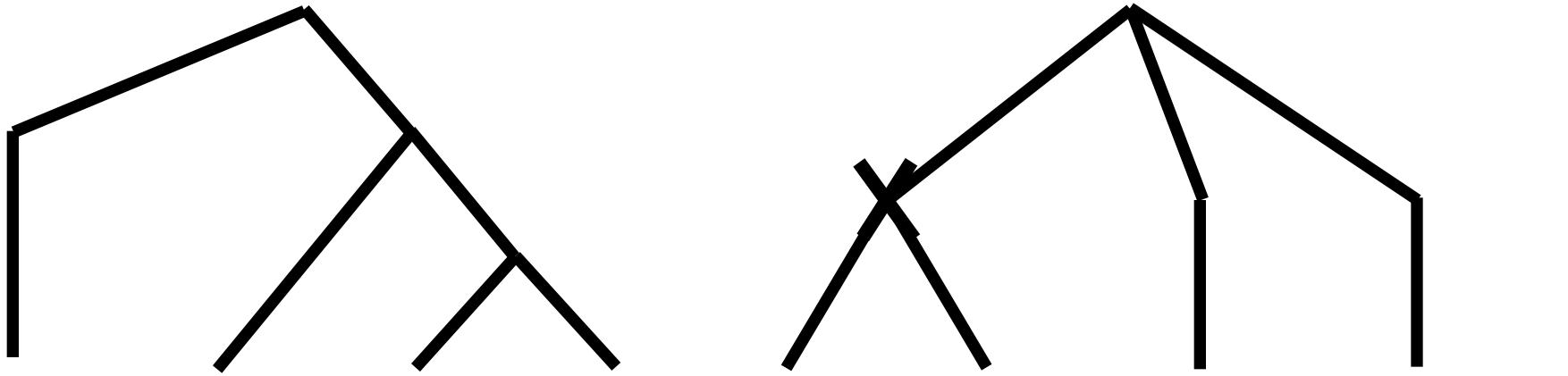
# Joint Extractive / Compressive Model



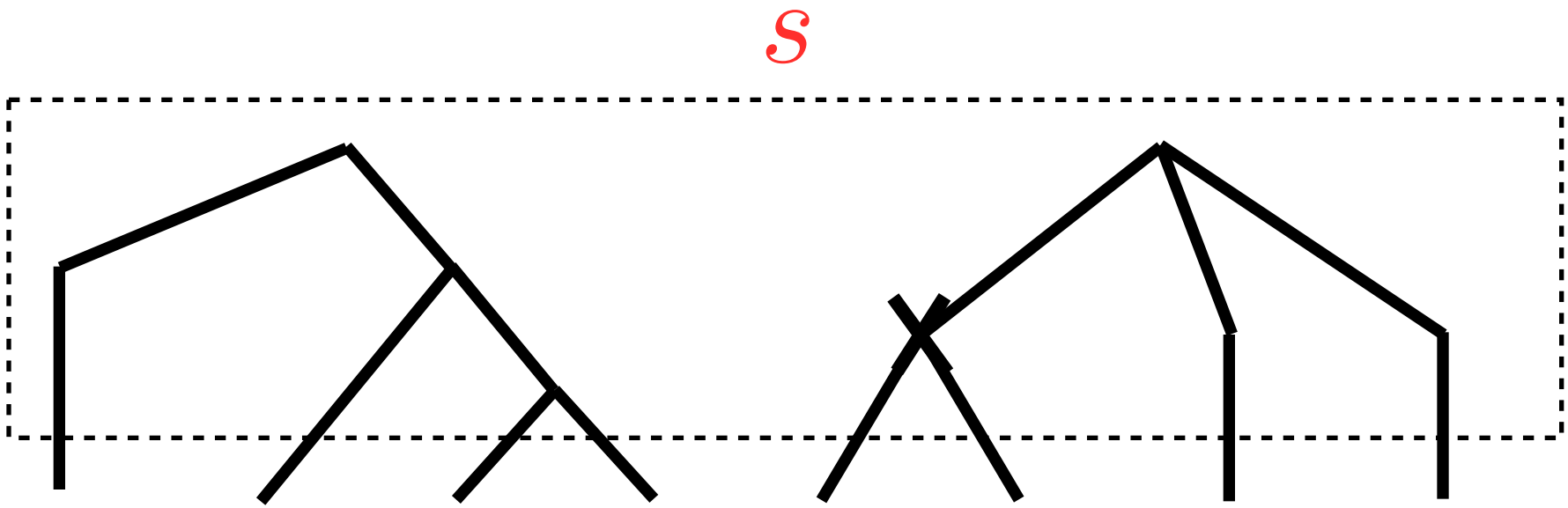
# Joint Extractive / Compressive Model



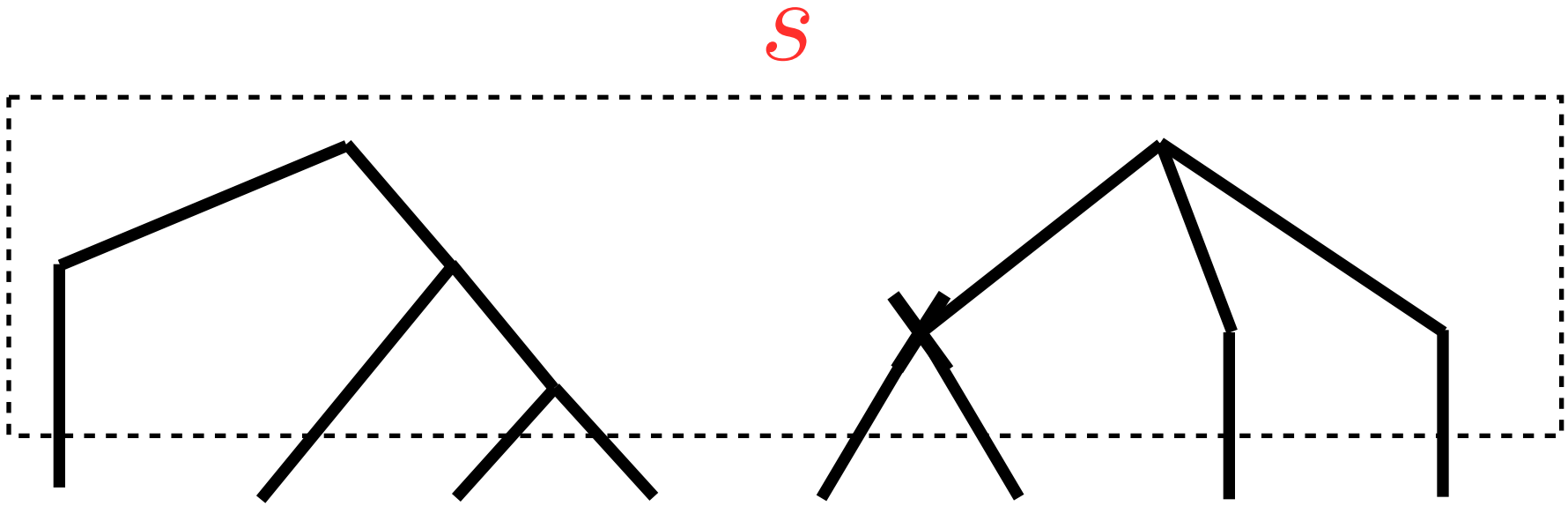
# Joint Extractive / Compressive Model



# Joint Extractive / Compressive Model



# Joint Extractive / Compressive Model

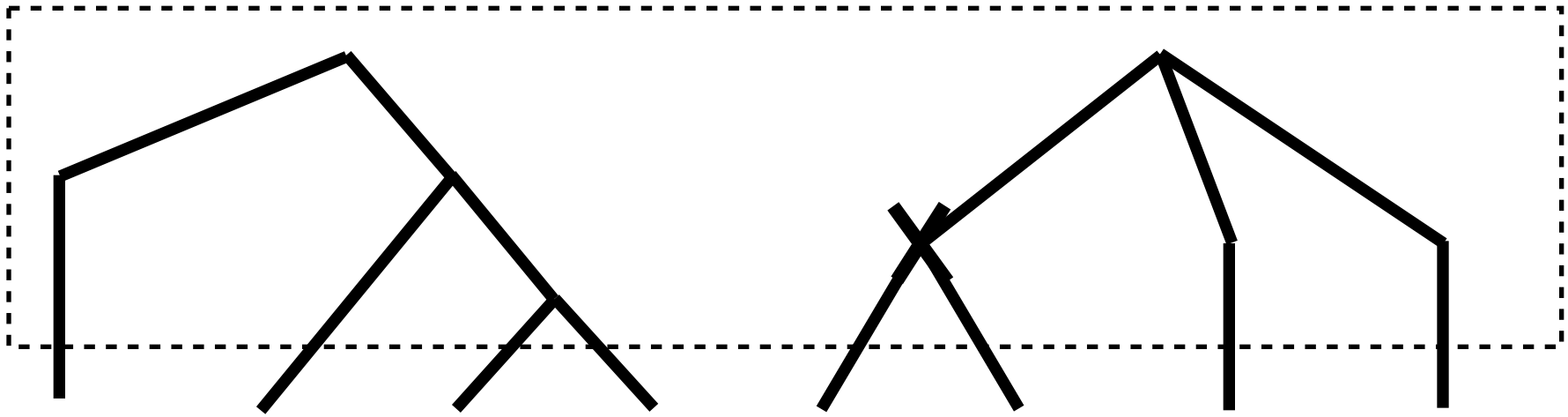



(she, stopped) ✓ (in, france) ✓ (she, remained) ✓

(stopped, in) ✓ (france, she) ✓

# Joint Extractive / Compressive Model

$s$



She stopped in France. ~~In France~~ she remained.

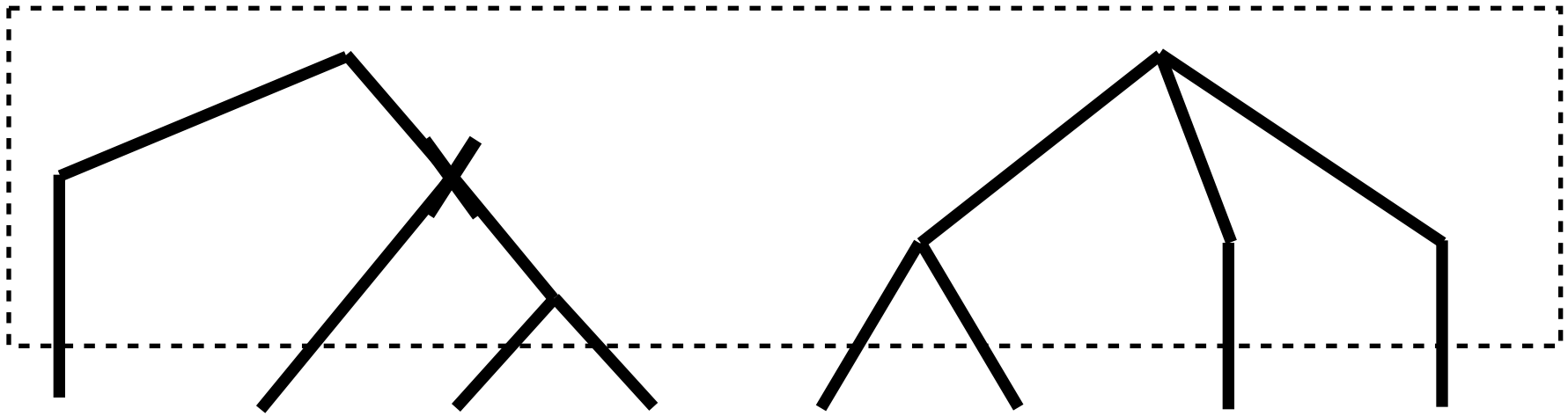
(she, stopped) ✓ (in, france) ✓ (she, remained) ✓

(stopped, in) ✓ (france, she) ✓

$B(s)$

# Joint Extractive / Compressive Model

$s$




(she, stopped) (in, france) ✓ (she, remained) ✓

(stopped, in) (france, she) ✓

$B(s)$


# Joint Extractive / Compressive Model

$$\max_{s} \left[ \sum_{b \in B(s)} \text{value}(b) \right]$$


# Joint Extractive / Compressive Model

$$\max_{s} \left[ \sum_{b \in B(s)} \text{value}(b) + \sum_{c \in s} \text{value}(c) \right]$$

Berkeley NLP

# Joint Extractive / Compressive Model

$$\max_{s} \left[ \sum_{b \in B(s)} \text{value}(b) + \sum_{c \in s} \text{value}(c) \right]$$

Parameterize using features:

$$\text{value}(b) = w^\top f(b)$$

$$\text{value}(c) = w^\top f(c)$$


# Learning

Linear prediction:

$$\text{score}(s) = w^\top f(s)$$

Berkeley NLP

# Learning

Linear prediction:

$$\text{score}(s) = w^\top f(s)$$

Feature function factors:

$$f(s) = \sum_{b \in B(s)} f(b) + \sum_{c \in s} f(c)$$


# Features


# Features

Bigram Features $f(b)$


# Features

Bigram Features $f(b)$

Cut Features $f(c)$


# Features

## Bigram Features $f(b)$

## Cut Features $f(c)$

| | |
|---|---|
| COUNT: | Bucketed document counts |
| STOP: | Stop word indicators |
| POSITION: | First document position indicators |
| CONJ: | All two- and three-way conjunctions of above |
| BIAS: | Always one |

Berkeley NLP

# Features

## Bigram Features $f(b)$

| | |
|---|---|
| COUNT: | Bucketed document counts |
| STOP: | Stop word indicators |
| POSITION: | First document position indicators |
| CONJ: | All two- and three-way conjunctions of above |
| BIAS: | Always one |

## Cut Features $f(c)$

| | |
|---|---|
| COORD: | Coordinated phrase, four versions: NP, VP, S, SBAR |
| S-ADJUNCT: | Adjunct to matrix verb, four versions: CC, PP, ADVP, SBAR |
| REL-C: | Relative clause indicator |
| ATTR-C: | Attribution clause indicator |
| ATTR-PP: | PP attribution indicator |
| TEMP-PP: | Temporal PP indicator |
| TEMP-NP | Temporal NP indicator |
| BIAS: | Always one |

Berkeley NLP

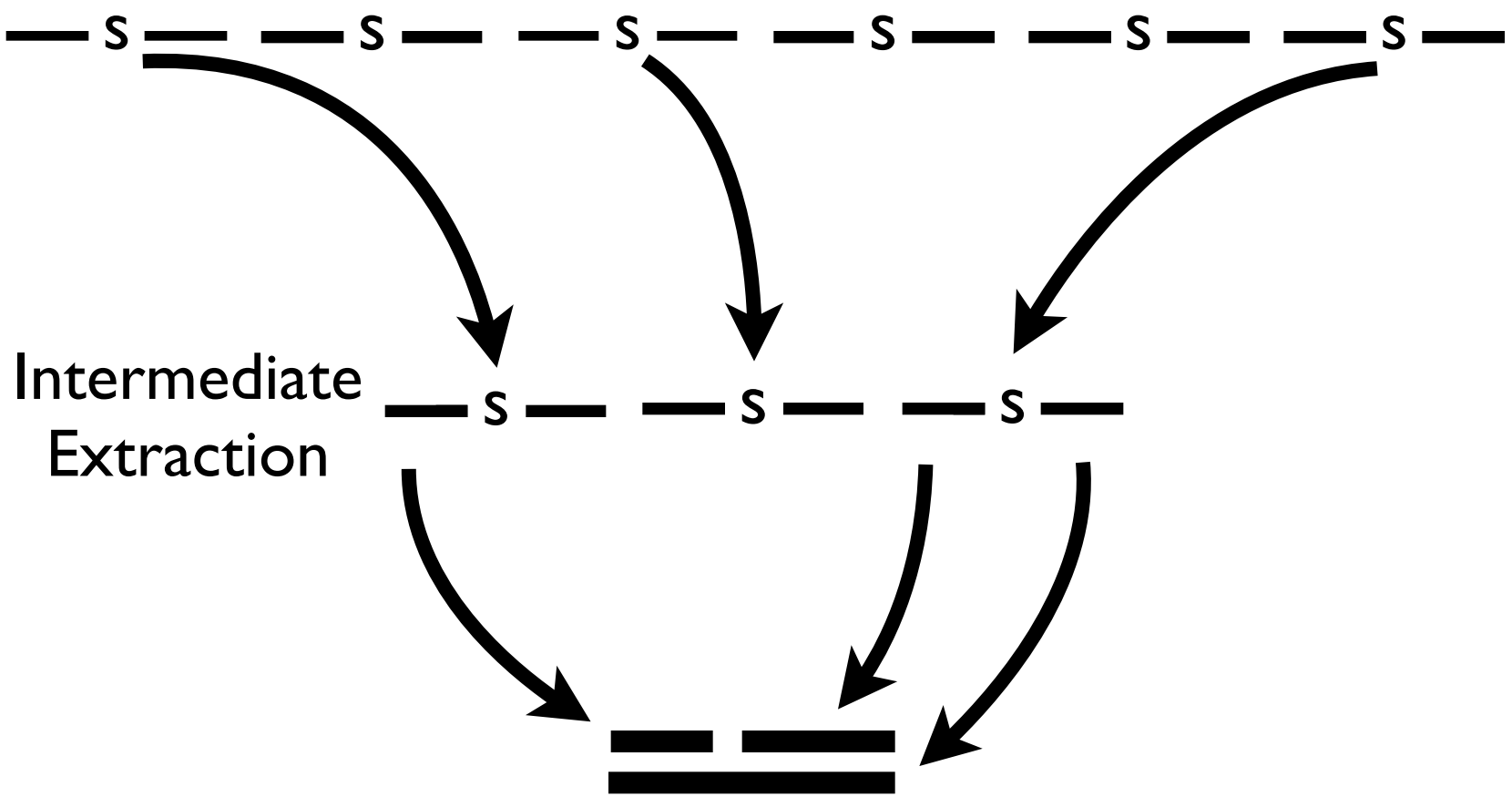
# Obtaining Training Data




Lindsay Lohan pleaded not guilty ~~Wednesday~~ to felony grand theft of a 2,500 necklace, ~~a case that could return the troubled starlet to jail rather than the big screen.~~ Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 ~~and warned that if Lohan was accused of breaking the law while free he would have her held without bail~~. The Mean Girls star is due back in court on Feb. 23 ~~an important hearing in which Lohan could opt to end the case early.~~

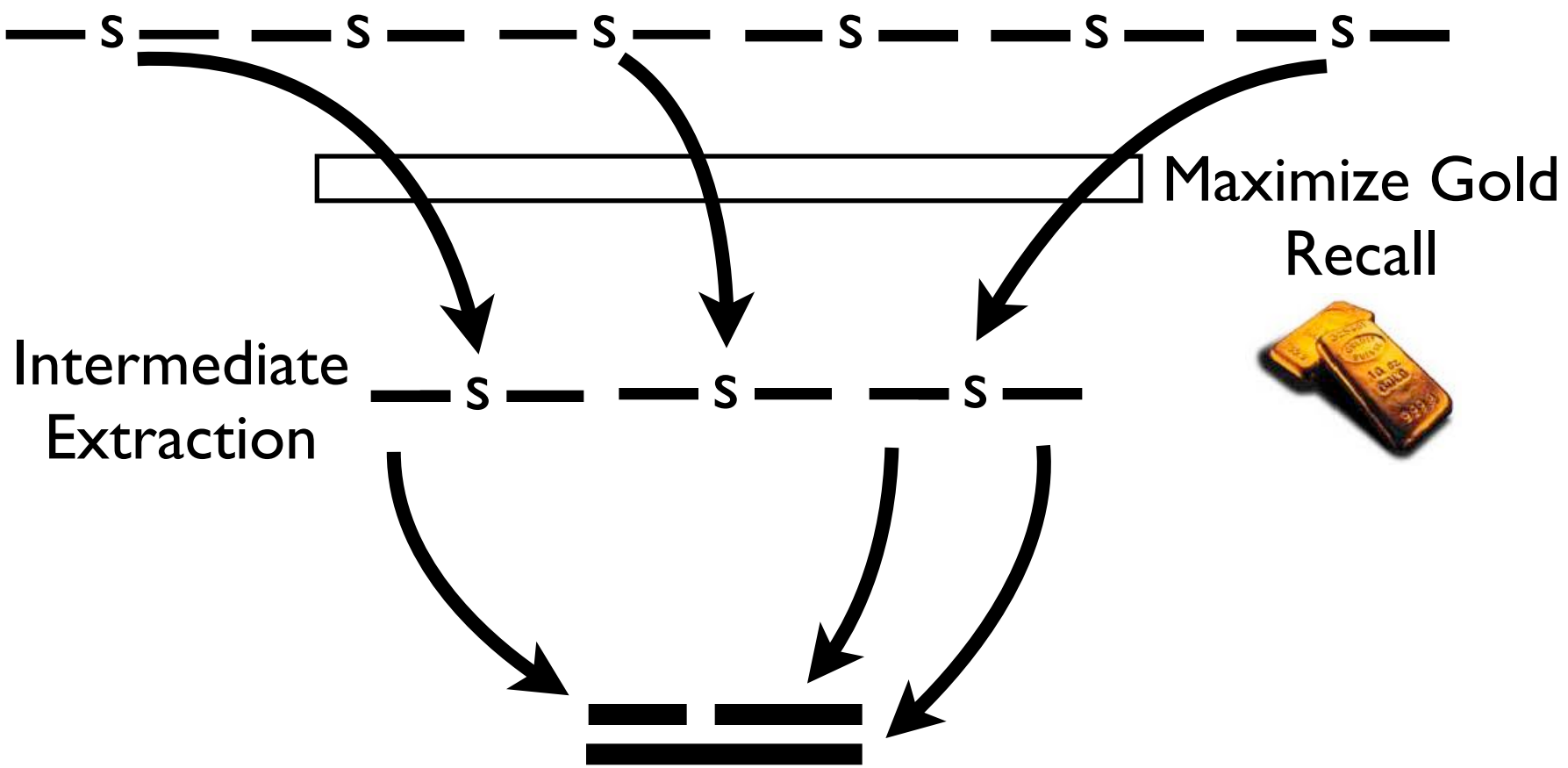
# Obtaining Training Data

TAC 2009: 44 Document Collections



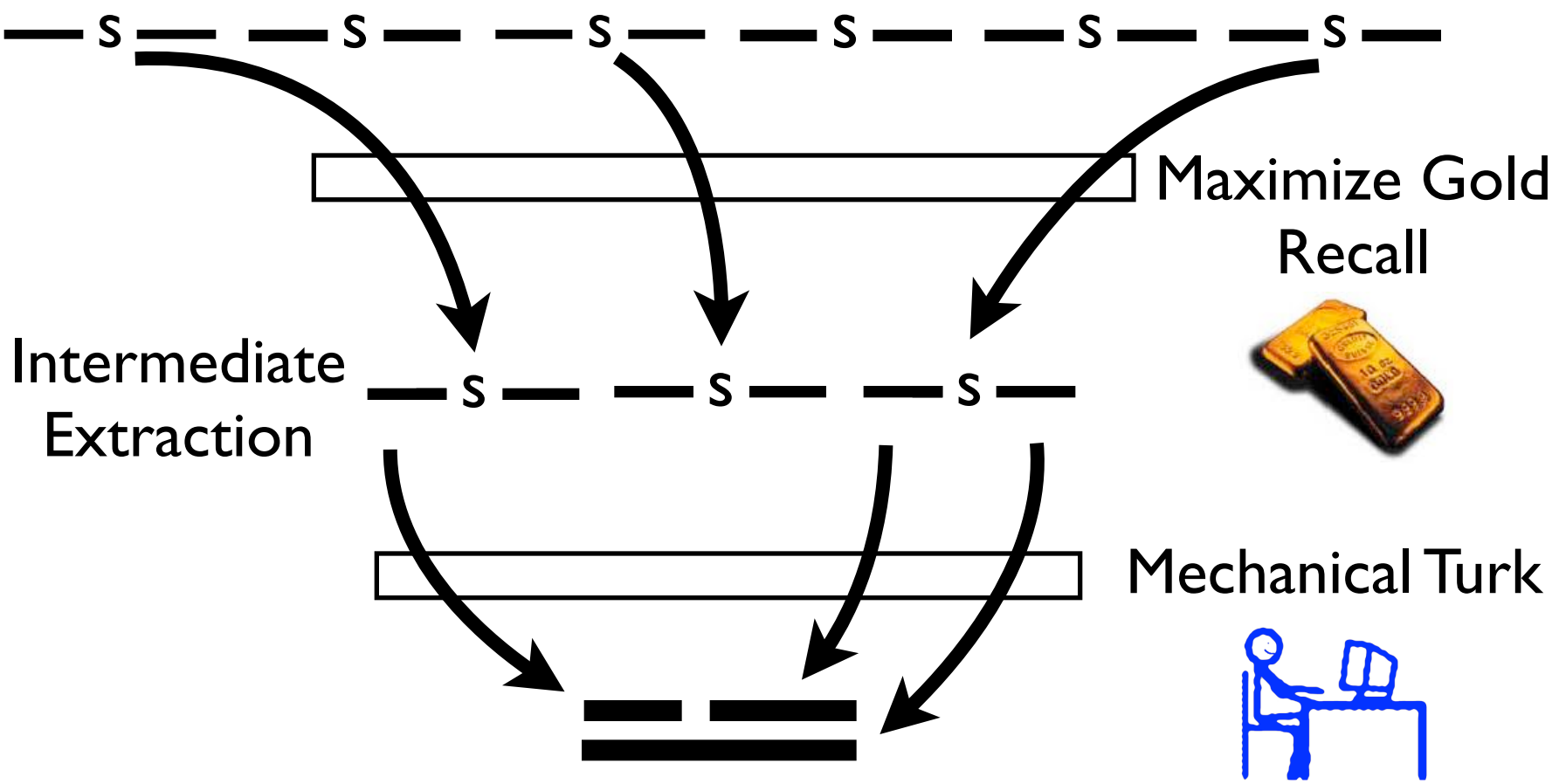
# Obtaining Training Data

TAC 2009: 44 Document Collections



# Obtaining Training Data

TAC 2009: 44 Document Collections




# Learning


# Learning

Structured SVM Training:

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \quad \|w\|_2^2$$

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \quad \|w\|_2^2$$

$s.t.$ for all possible guess summaries:

score of gold exceeds score of guess by loss

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \quad \|w\|_2^2$$

$s.t.$ for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S)$$

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \quad \|w\|_2^2$$

$s.t.$ for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \left[ \|w\|_2^2 + \xi \right]$$

$s.t.$ for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \left[ \|w\|_2^2 + \xi \right]$$

$s.t.$ for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \boxed{\text{loss}(S^*, S)} - \xi$$

End-to-end summarization quality

Berkeley NLP

# Learning

Structured SVM Training:

$$\min_{w} \left[ \|w\|_2^2 + \xi \right]$$

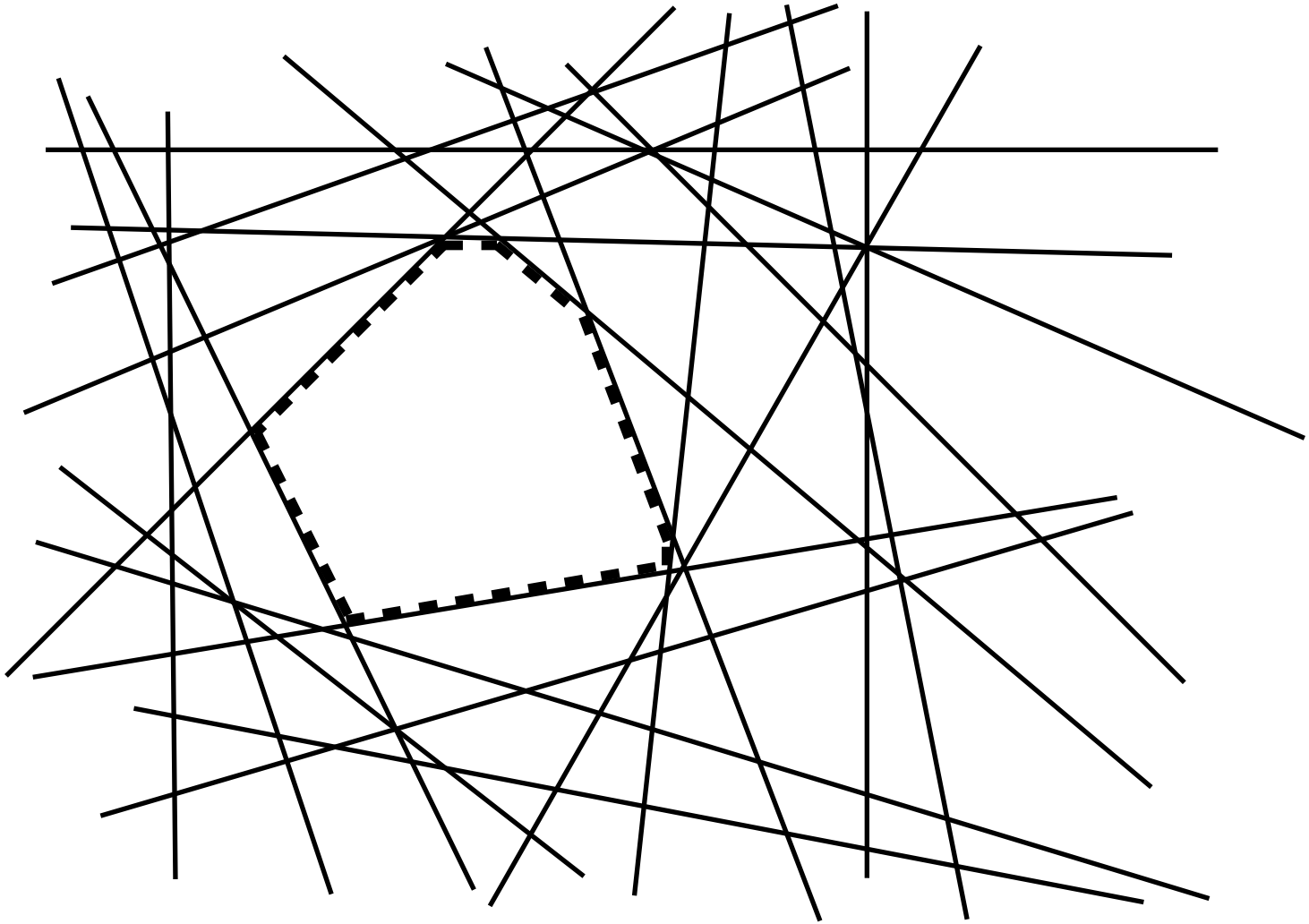
$s.t.$ for all possible guess summaries:

$$w^\top f(S^*) - w^\top f(S) \geq \text{loss}(S^*, S) - \xi$$

Exponentially many constraints!

Berkeley
N L P

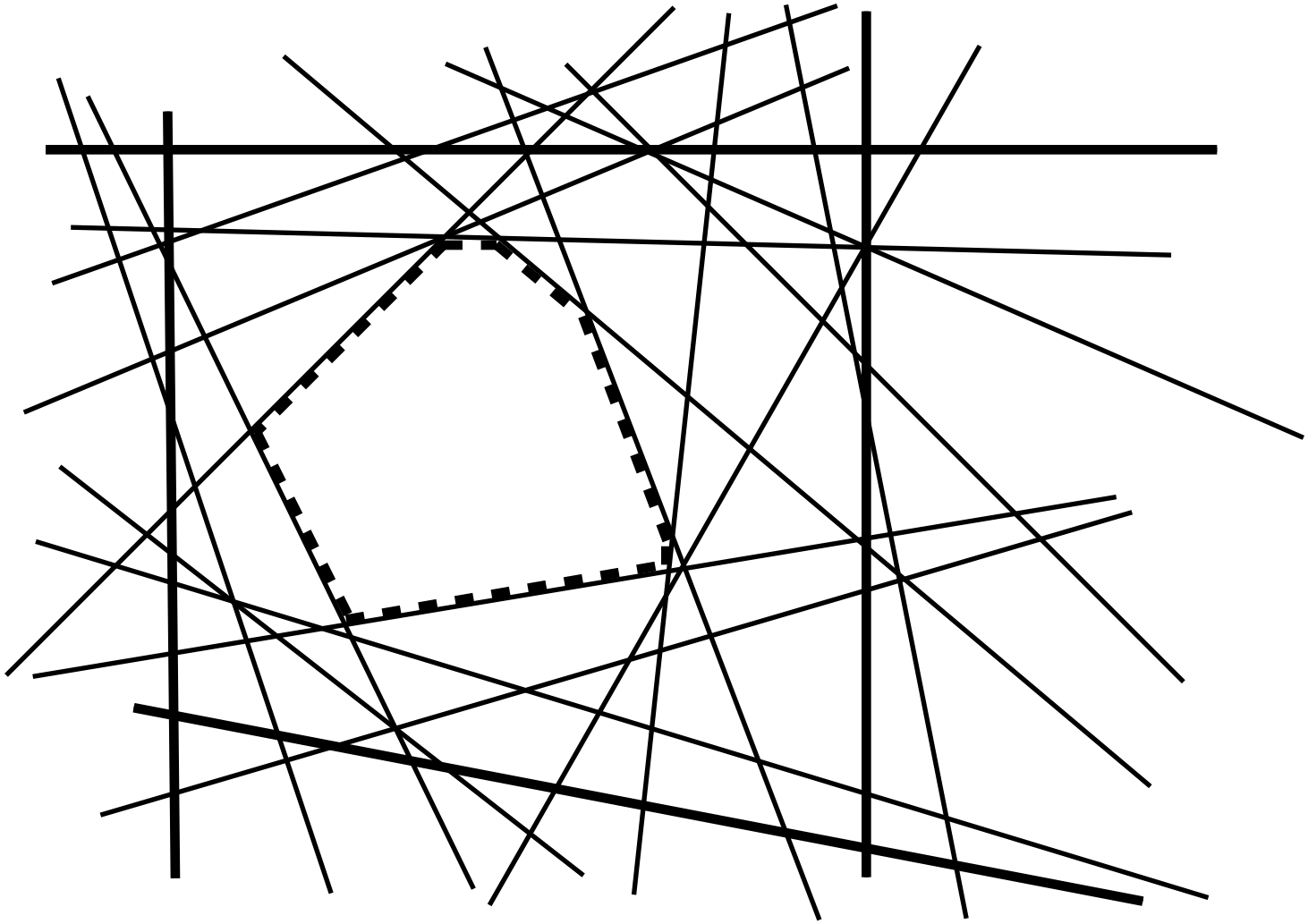
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

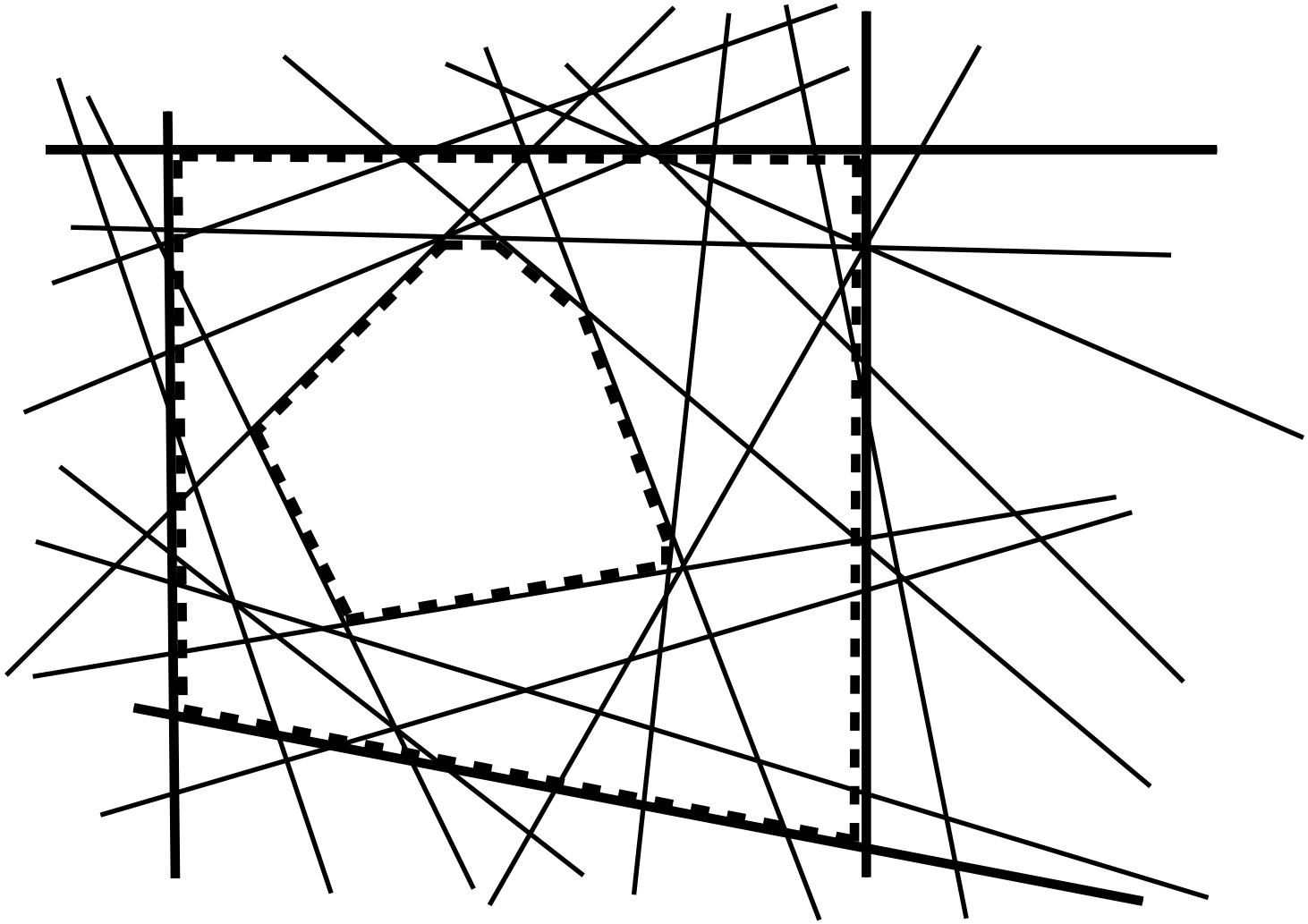
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

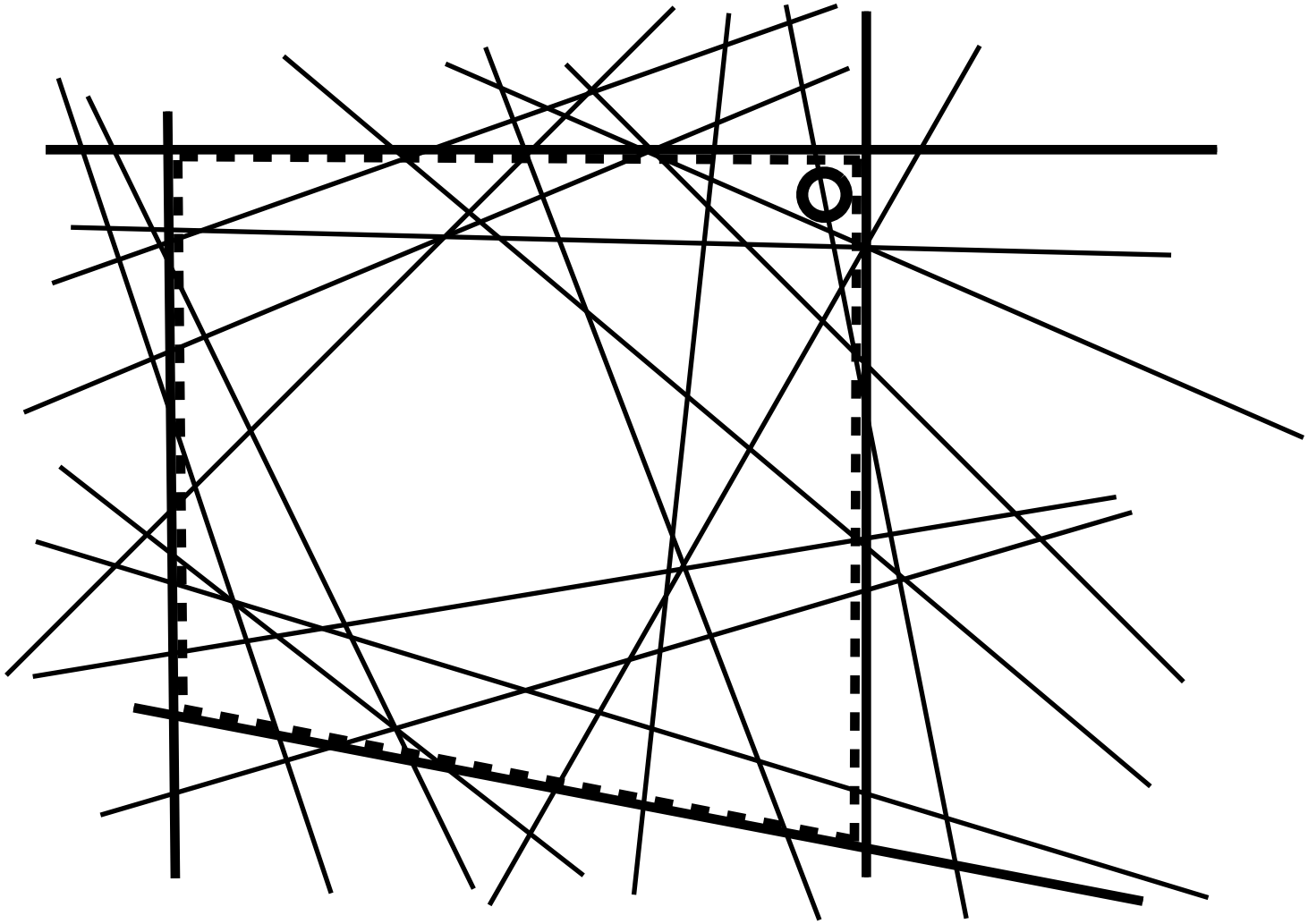
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

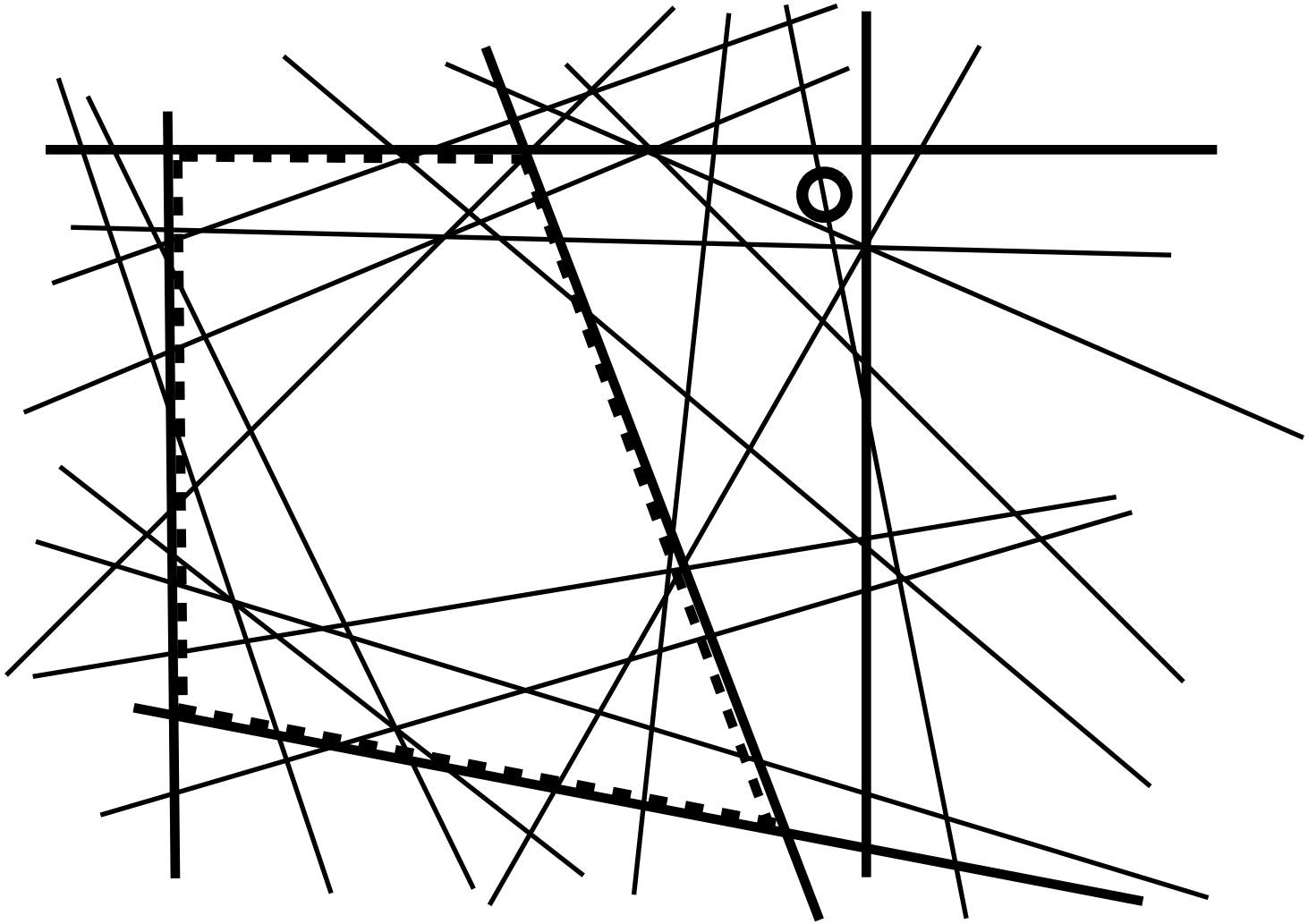
# Learning




[Tsochantaridis et al. 2004]

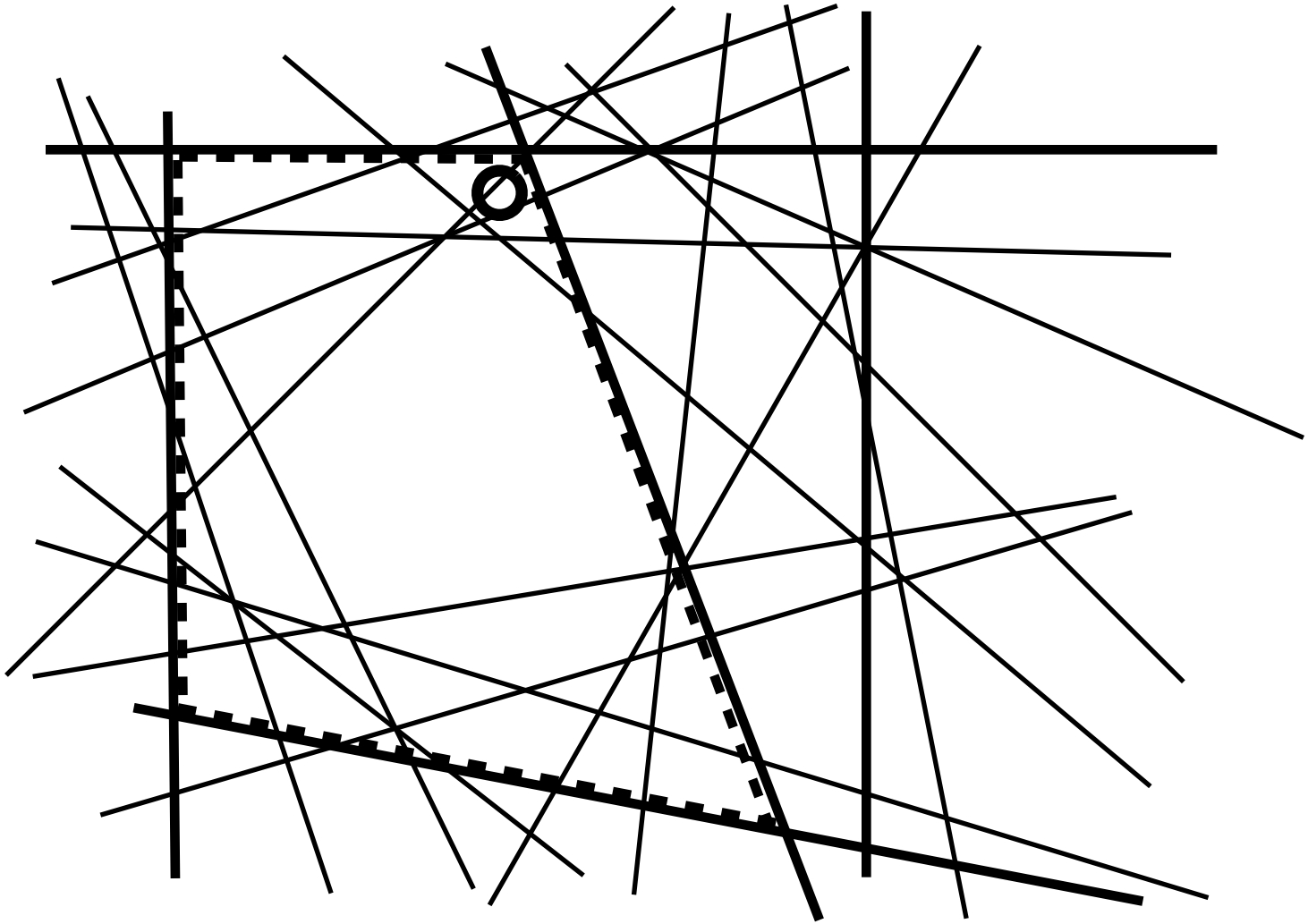
Berkeley NLP

# Learning



[Tsochantaridis et al. 2004]

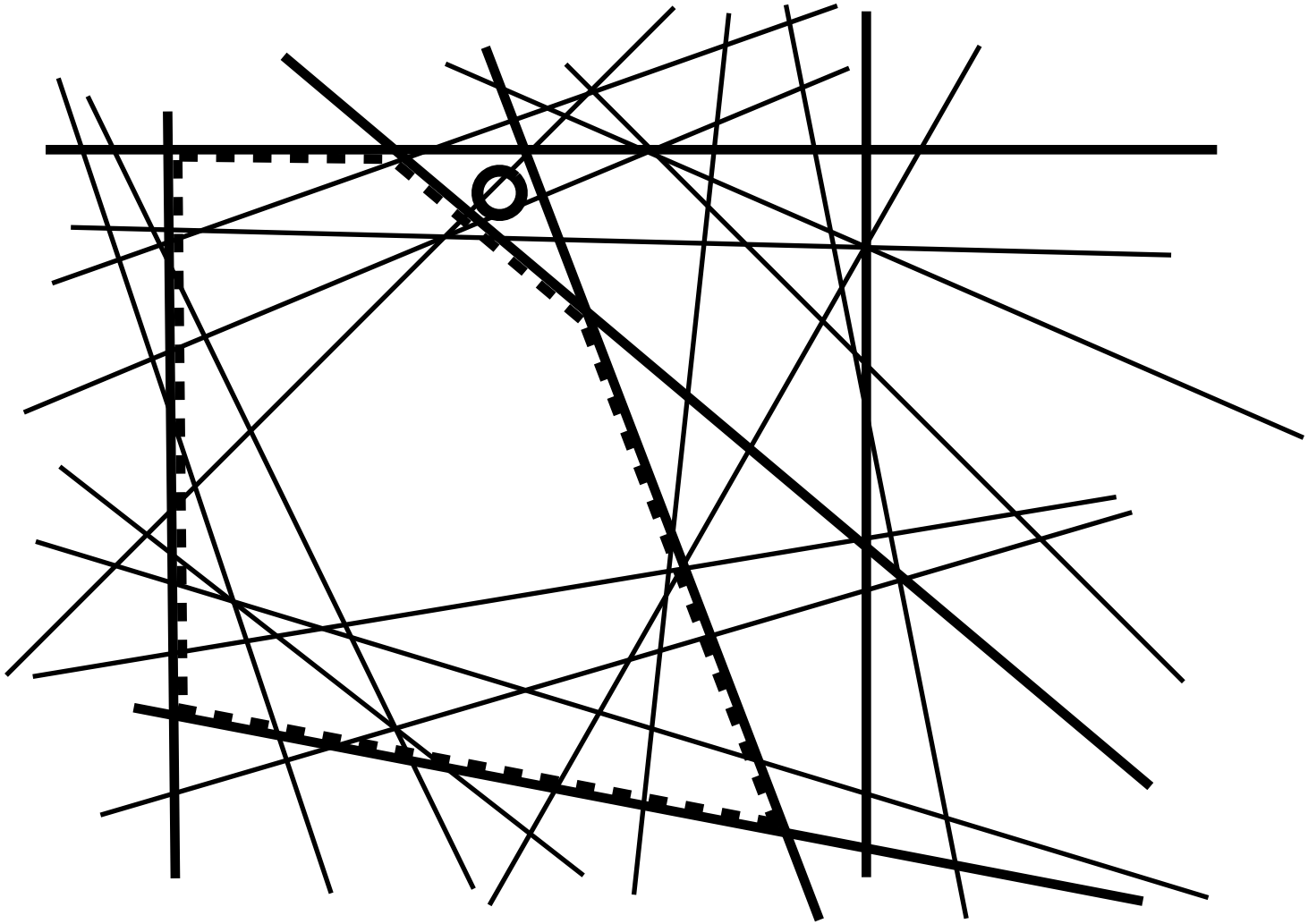
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

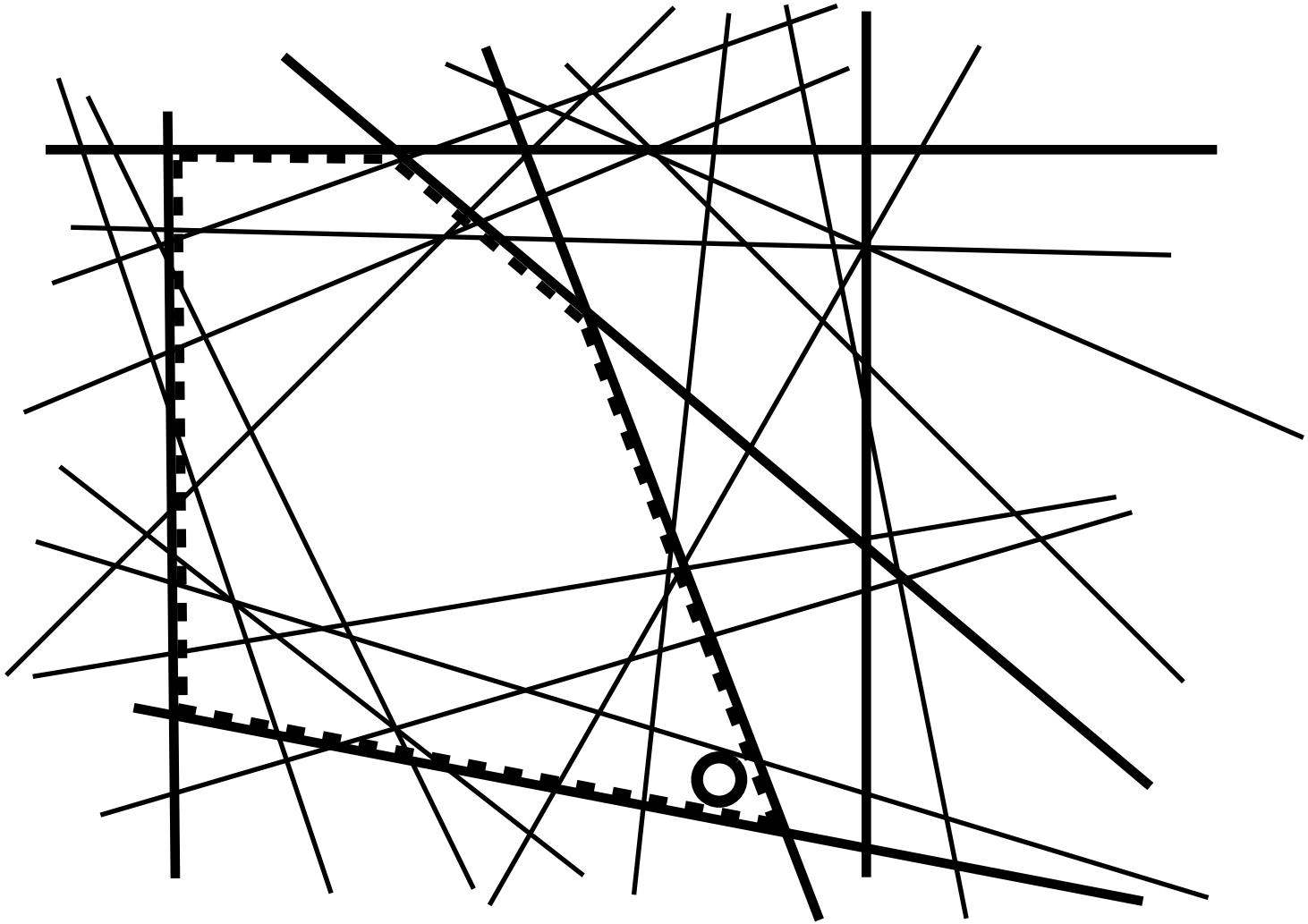
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

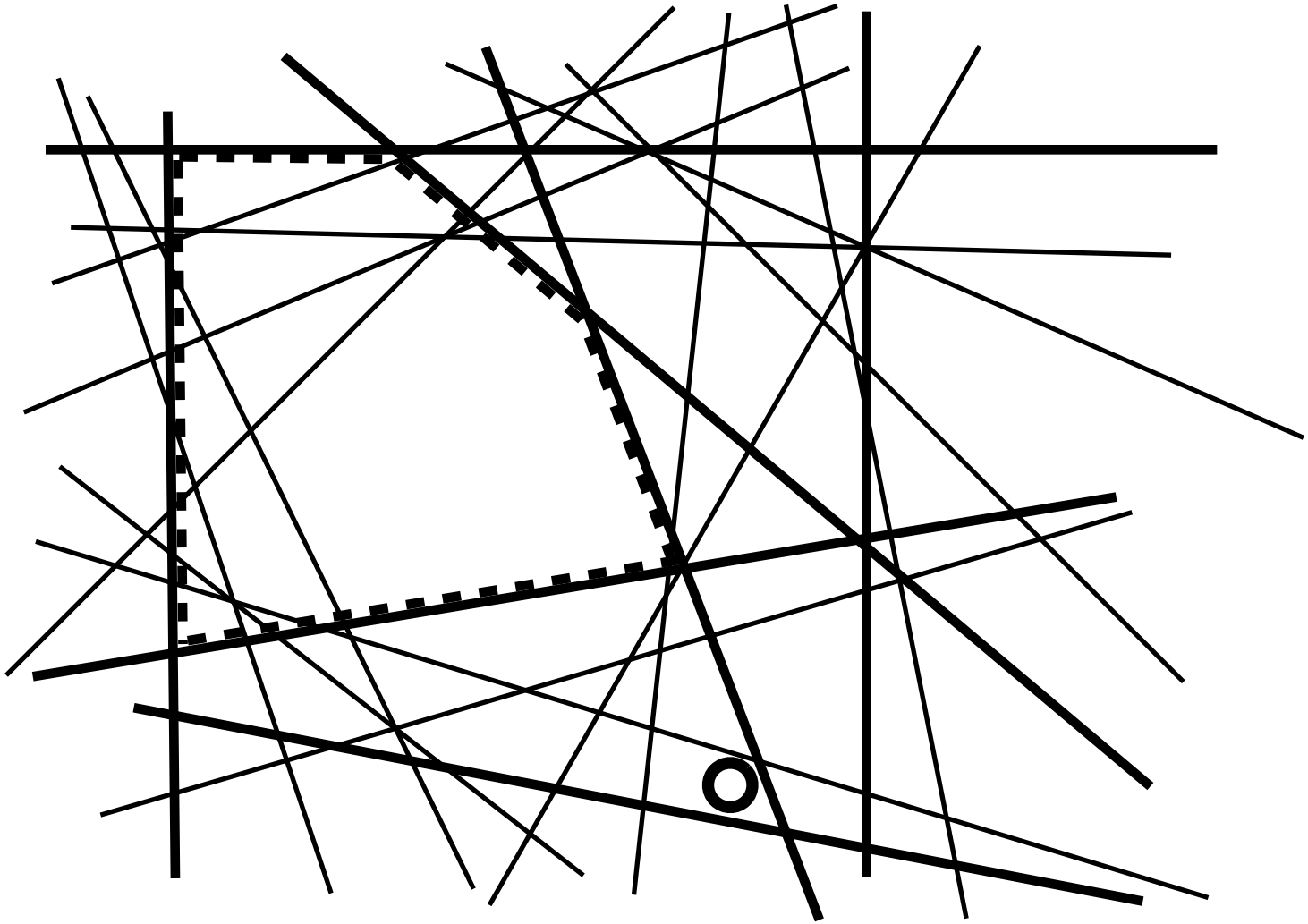
# Learning




[Tsochantaridis et al. 2004]

Berkeley NLP

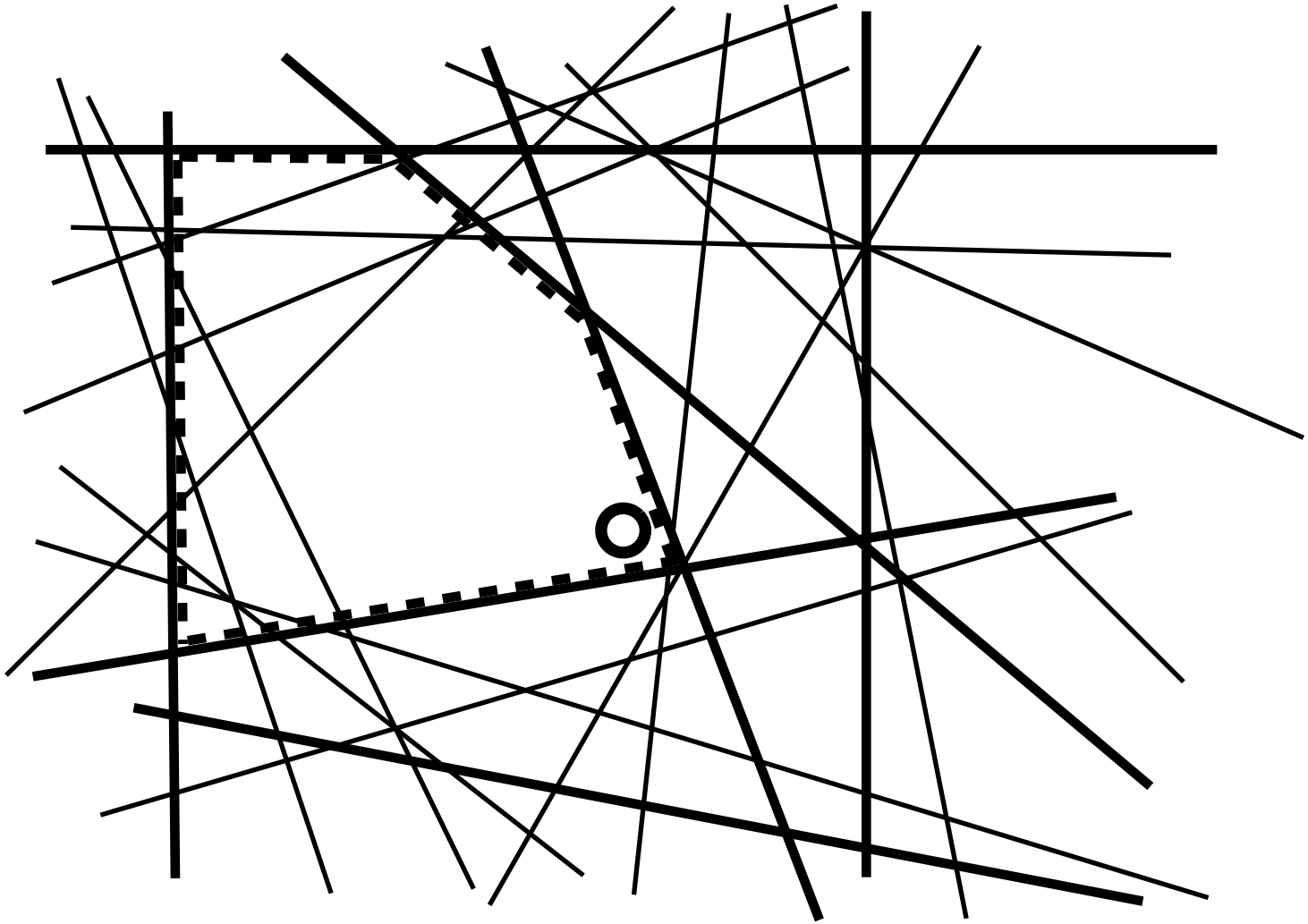
# Learning



[Tsochantaridis et al. 2004]

Berkeley NLP

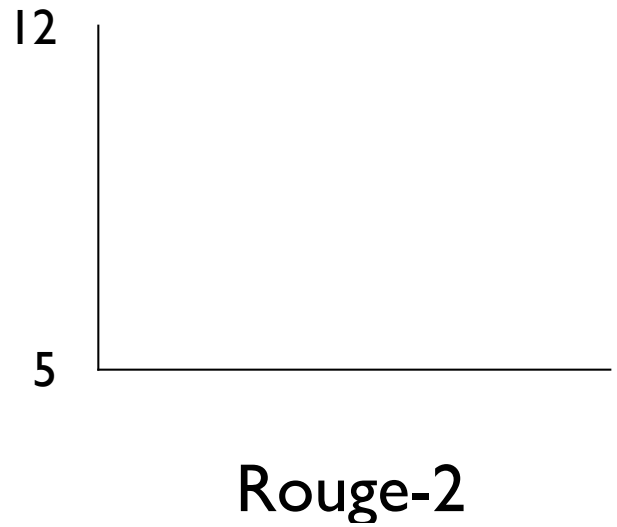
# Learning



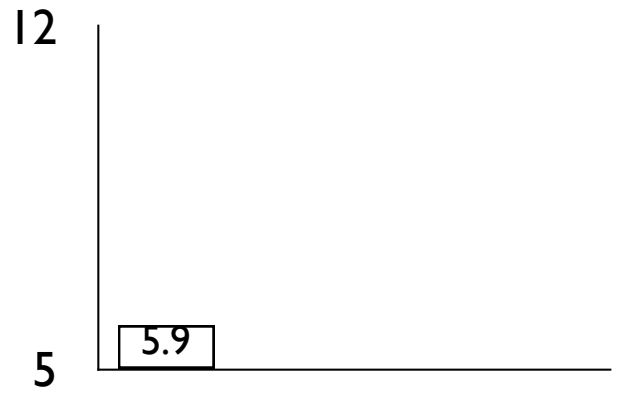
[Tsochantaridis et al. 2004]

Berkeley NLP

# Results



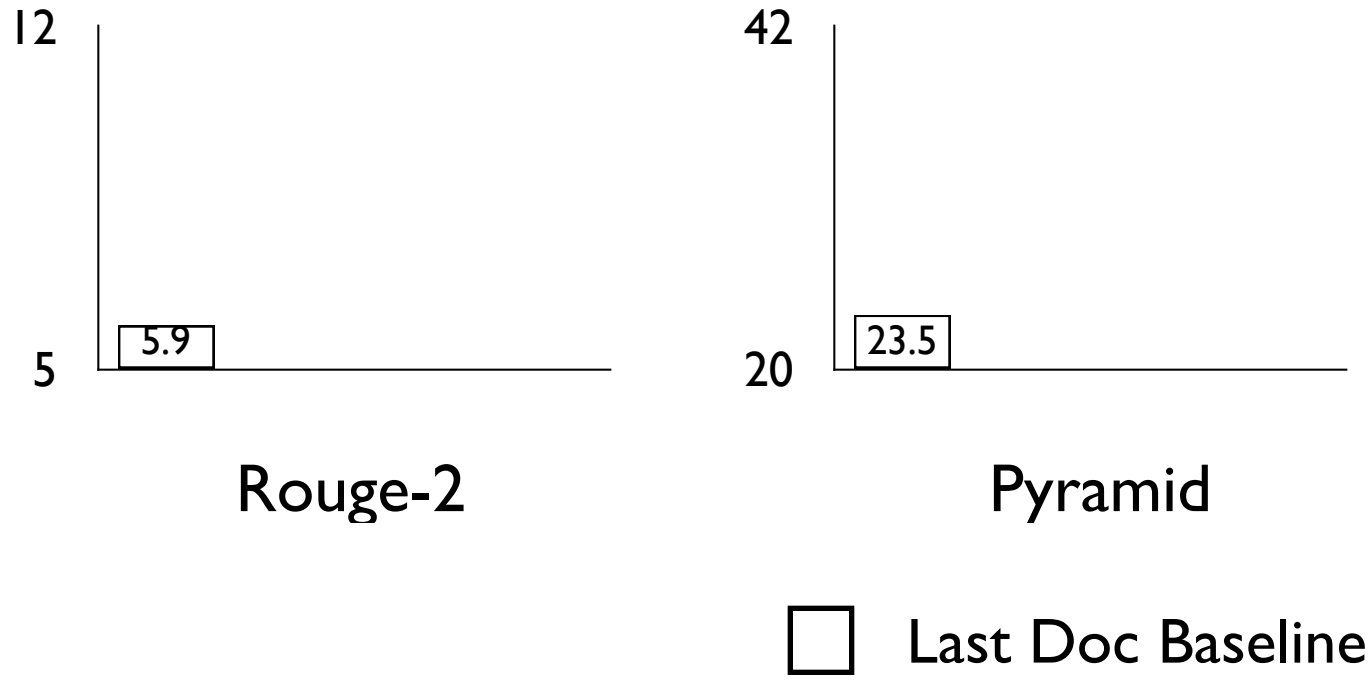
# Results



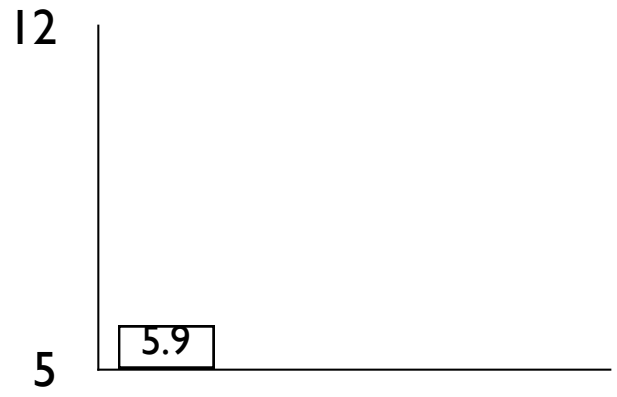

Rouge-2

Last Doc Baseline

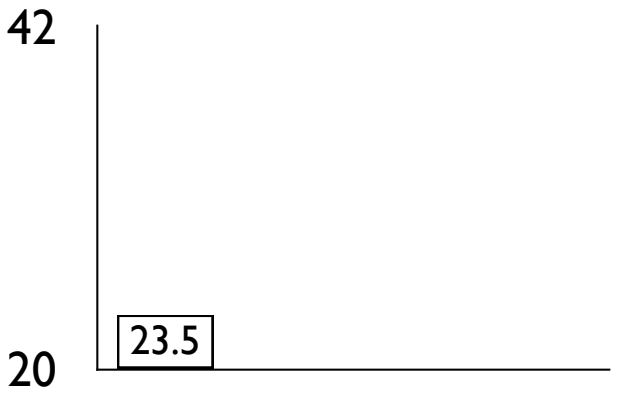
# Results



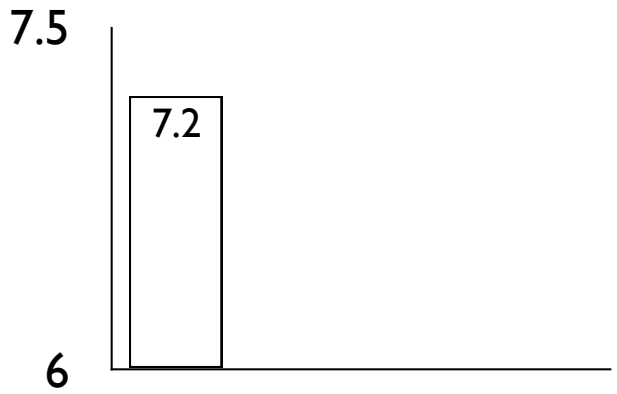
# Results




Rouge-2



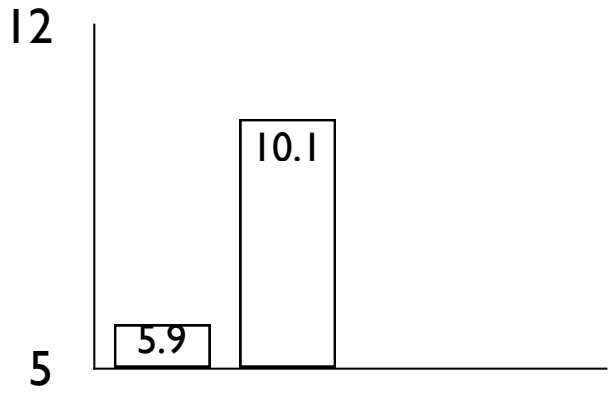

Pyramid



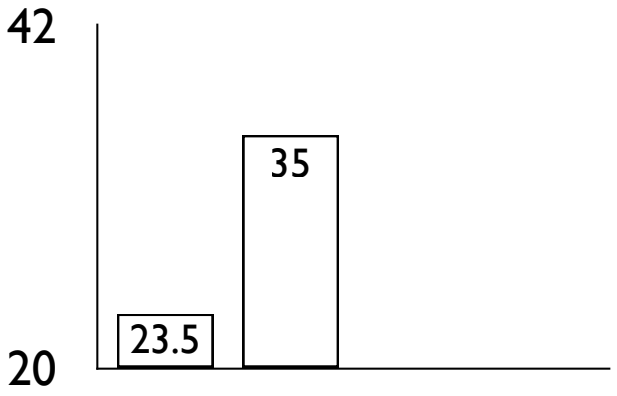

Linguistic Quality

Last Doc Baseline

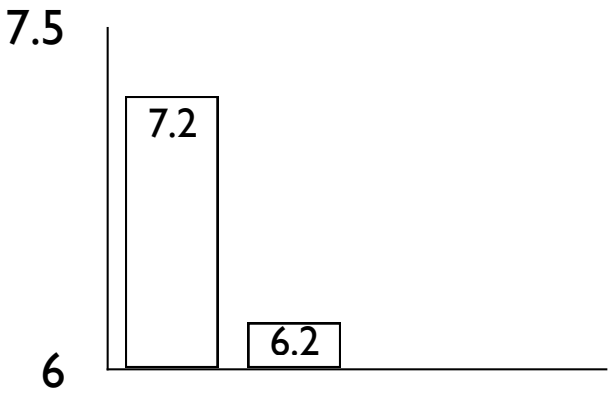
# Results




Rouge-2




Pyramid

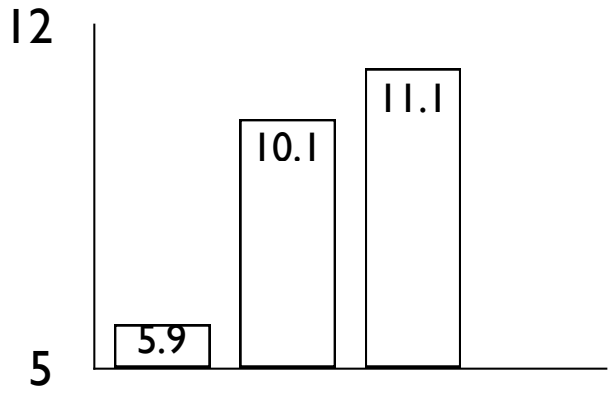



Linguistic Quality

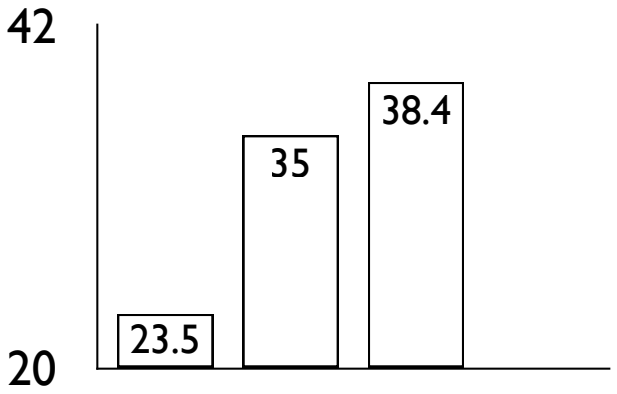
Last Doc Baseline

Extractive Baseline

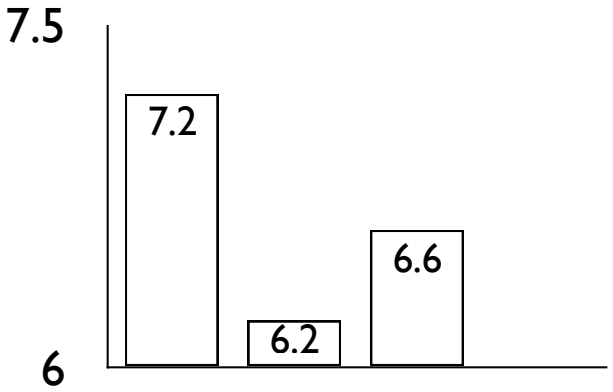
# Results




Rouge-2




Pyramid



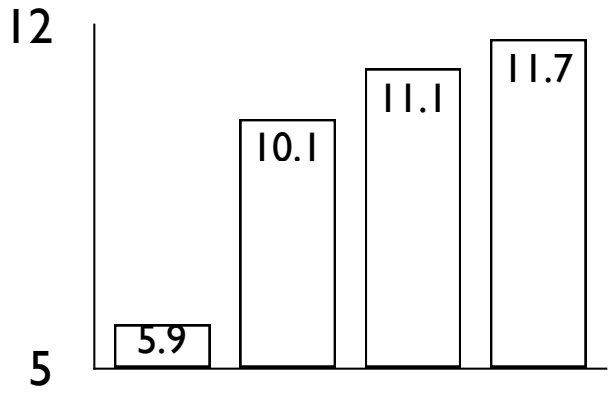

Linguistic Quality

Last Doc Baseline

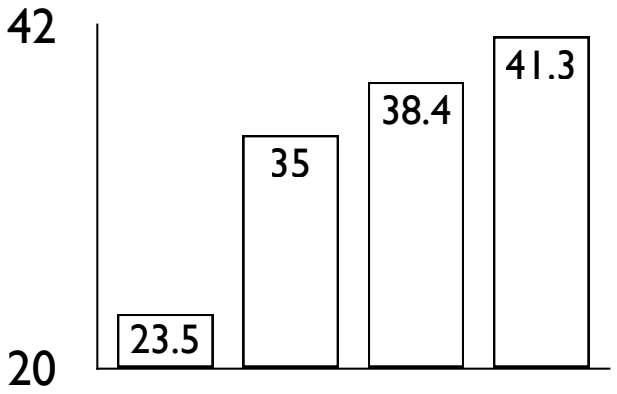
Extractive Baseline

Learned Extractive

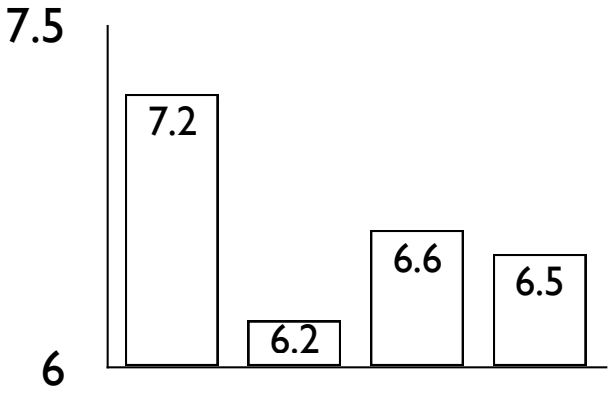
# Results




Rouge-2




Pyramid



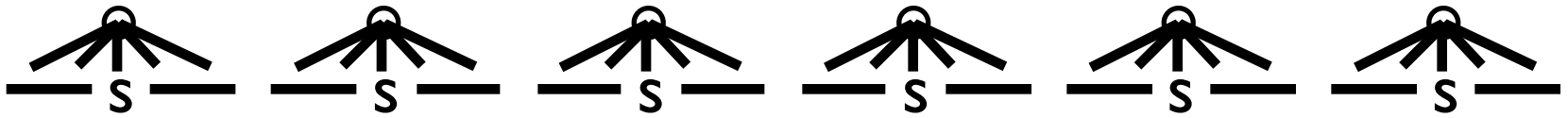

Linguistic Quality

- Last Doc Baseline
- Extractive Baseline
- Learned Extractive
- Learned E + C

Berkeley

N L P

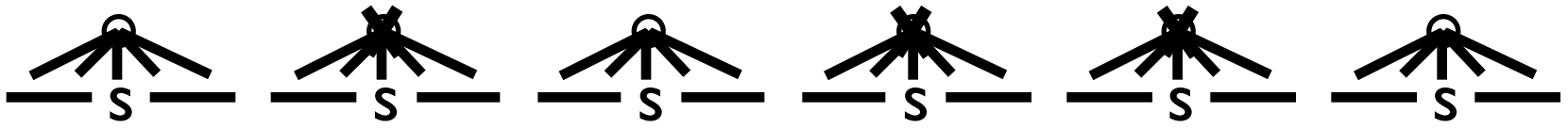
# Coarse-to-fine Decoding



Berkeley

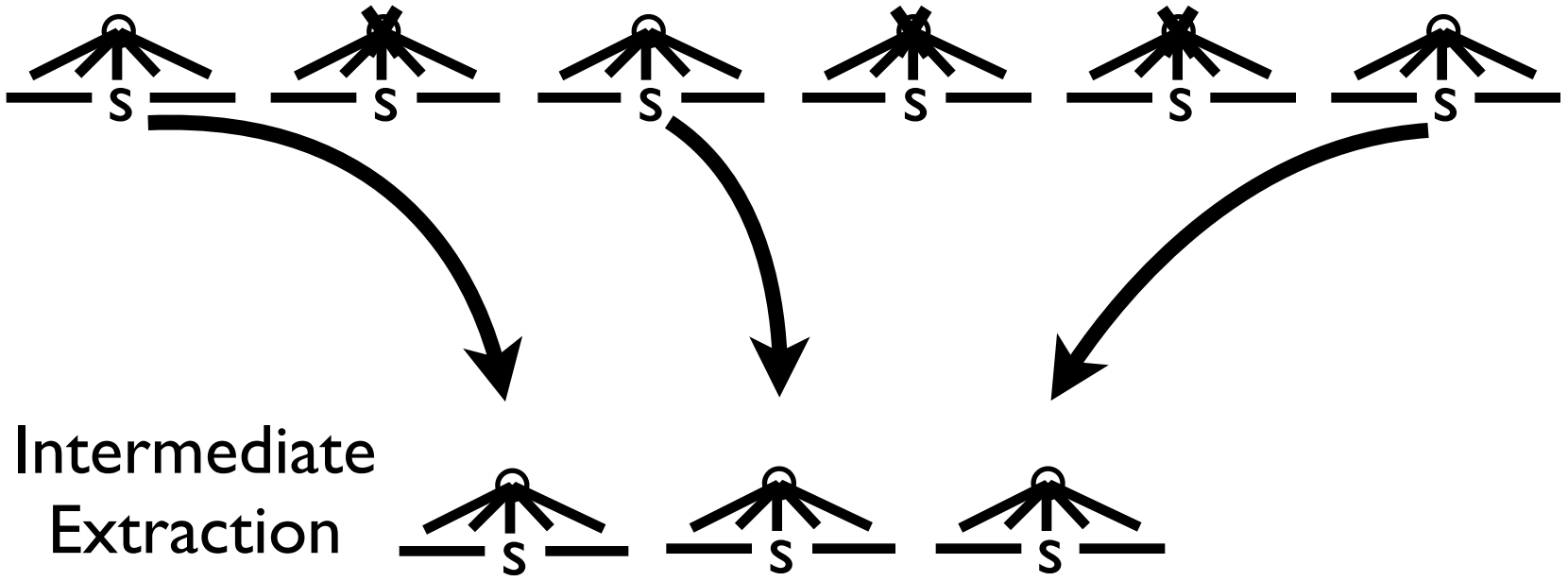
N L P

# Coarse-to-fine Decoding



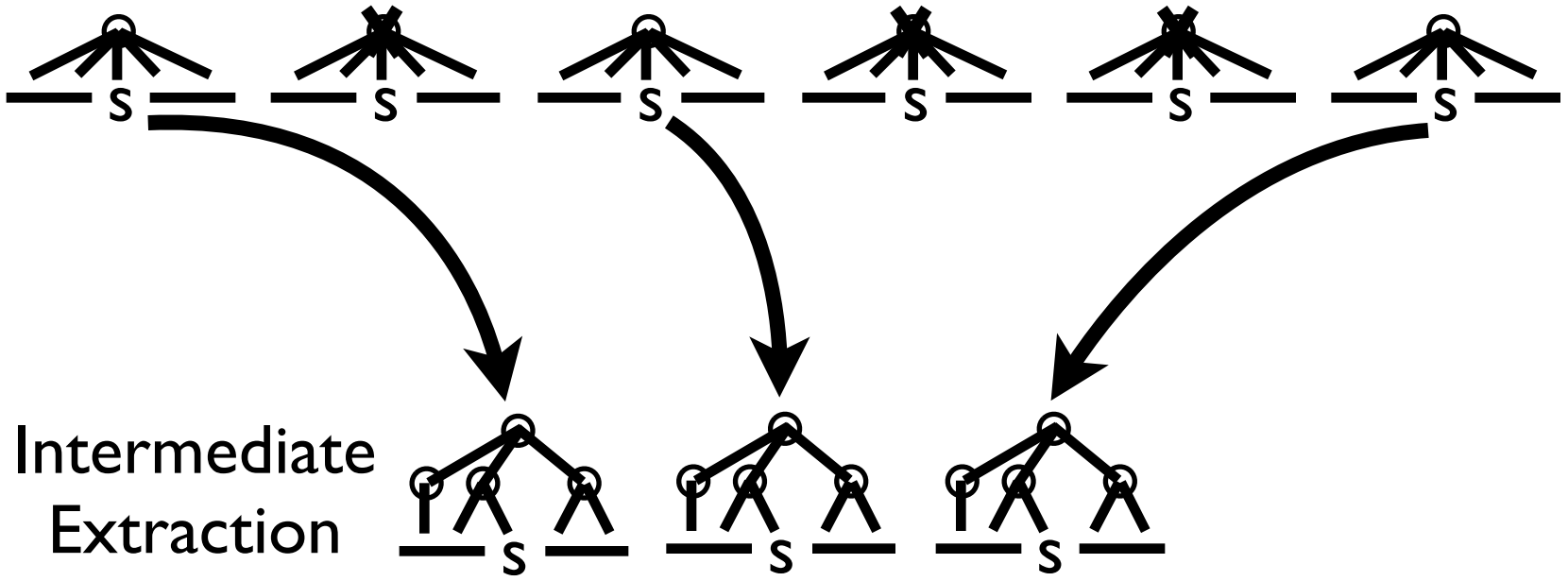
Berkeley NLP

# Coarse-to-fine Decoding



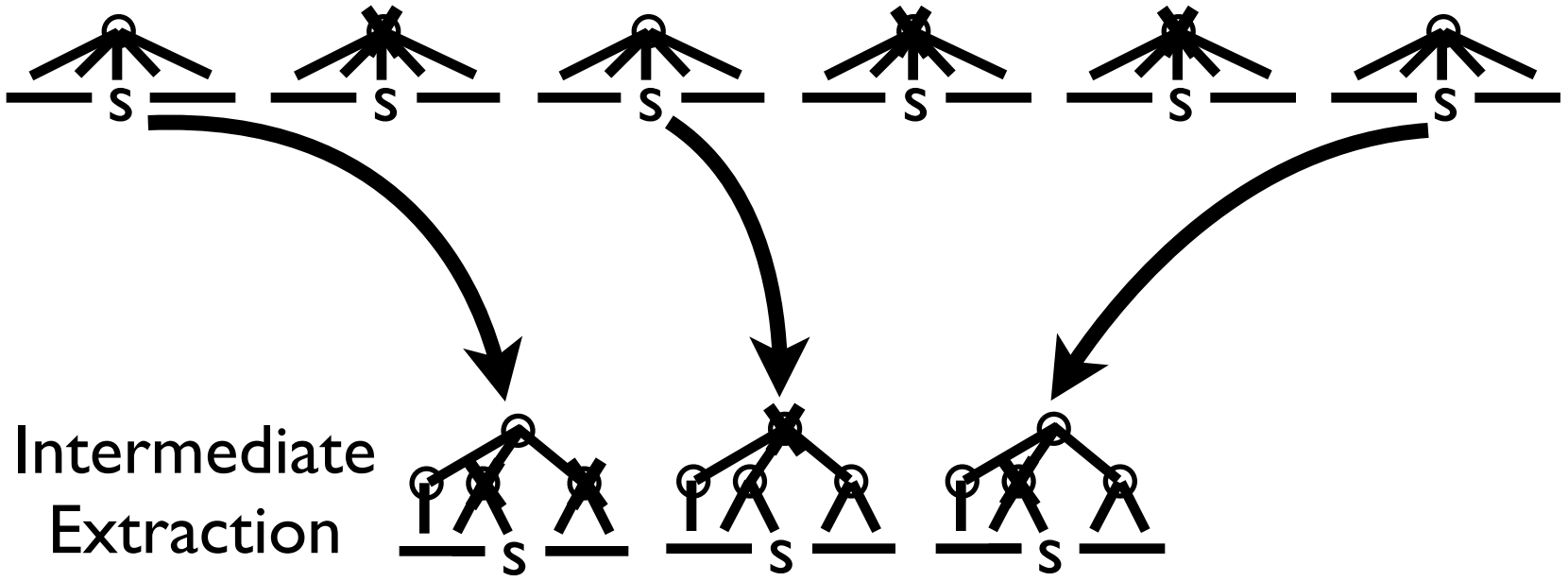
Berkeley NLP

# Coarse-to-fine Decoding



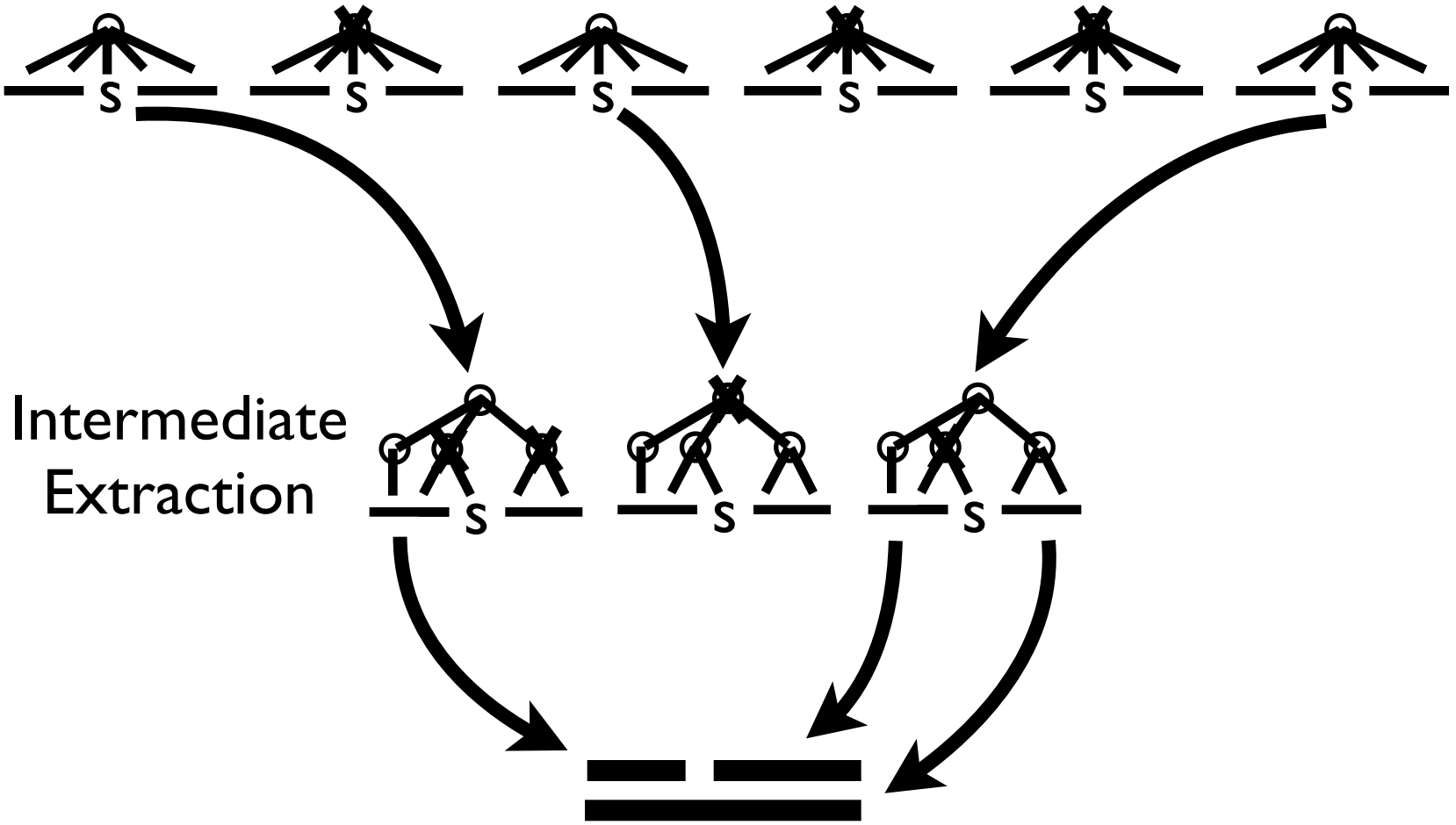
Berkeley NLP

# Coarse-to-fine Decoding



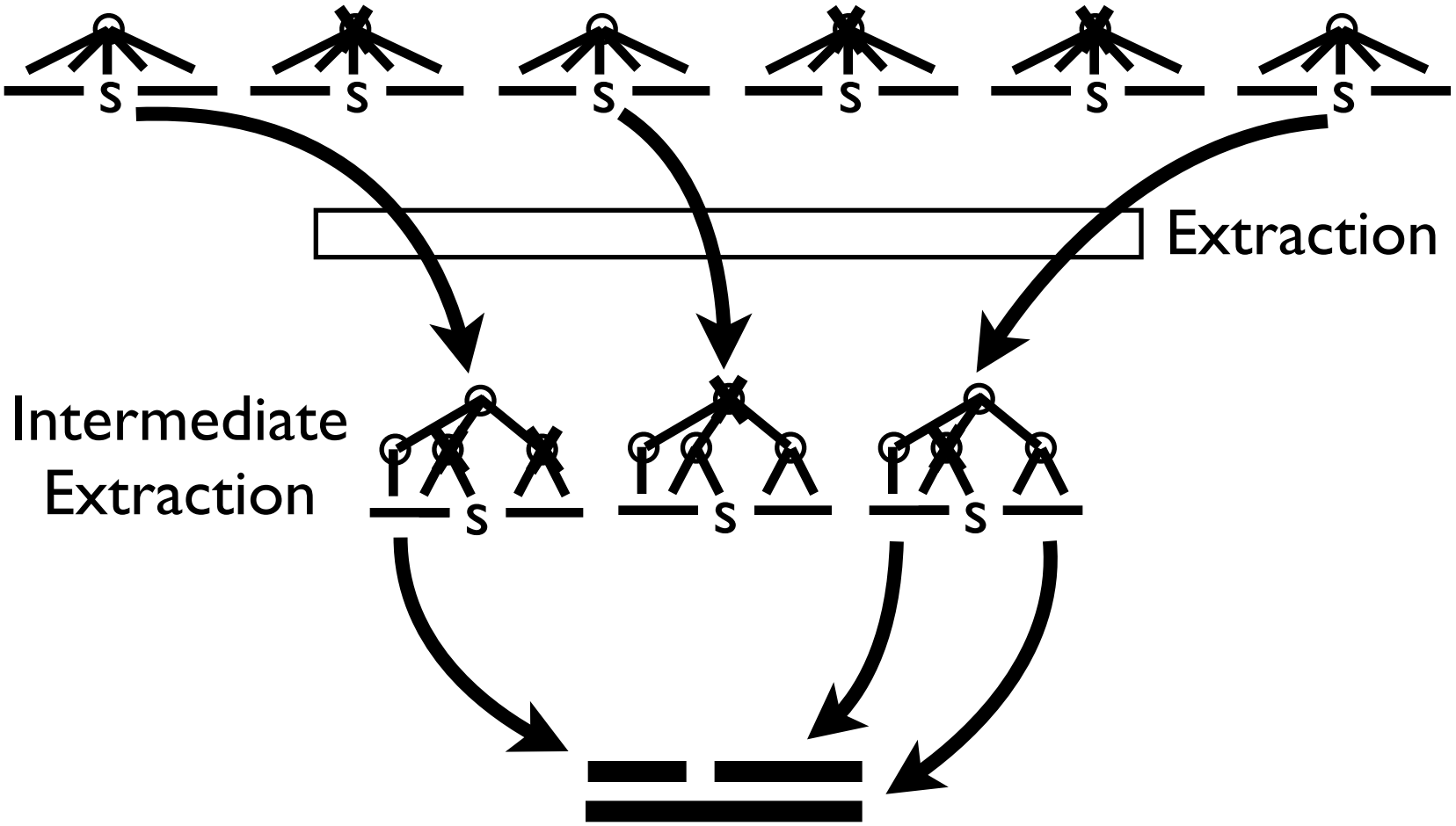
Berkeley NLP

# Coarse-to-fine Decoding



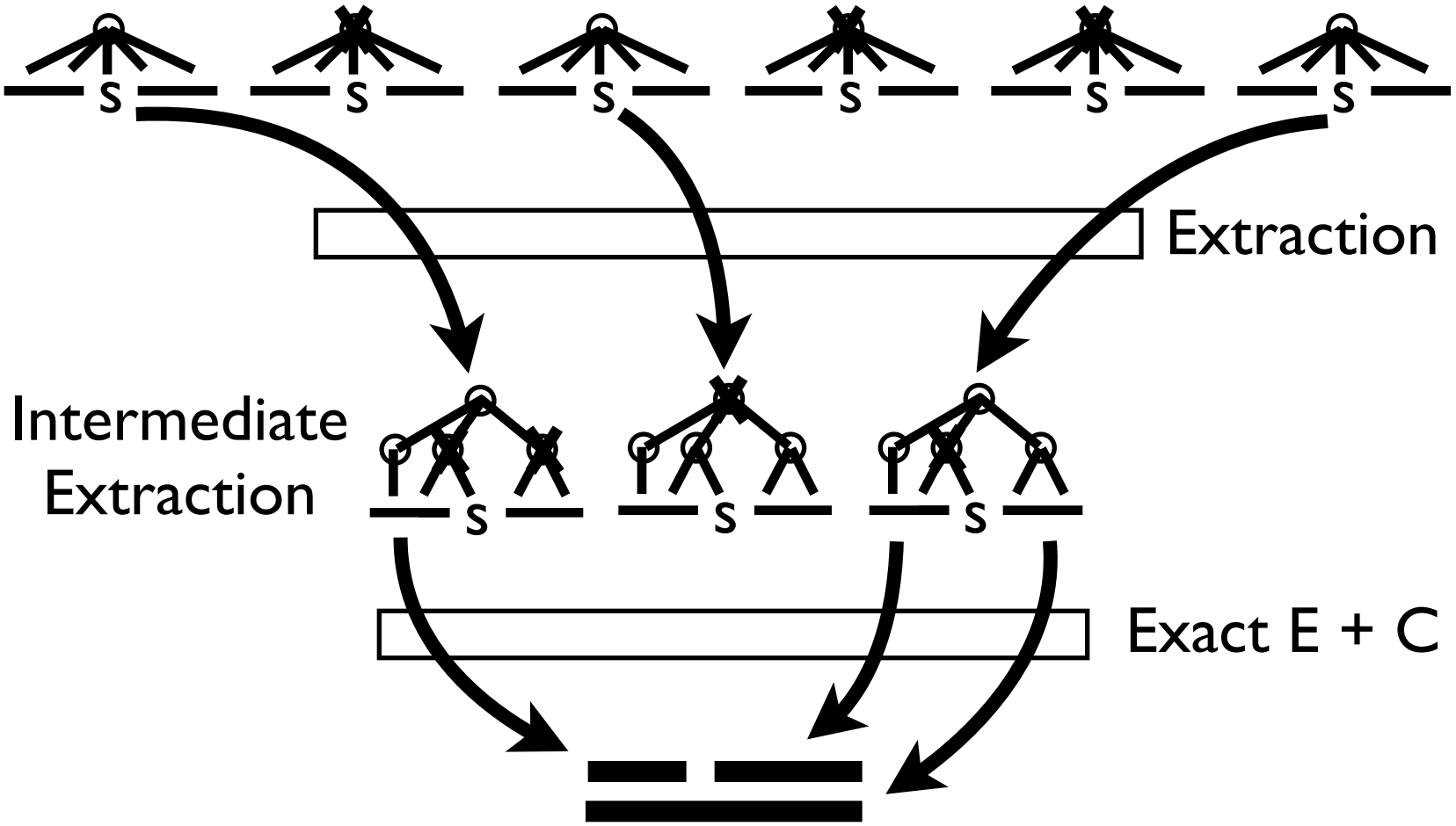
Berkeley
N L P

# Coarse-to-fine Decoding



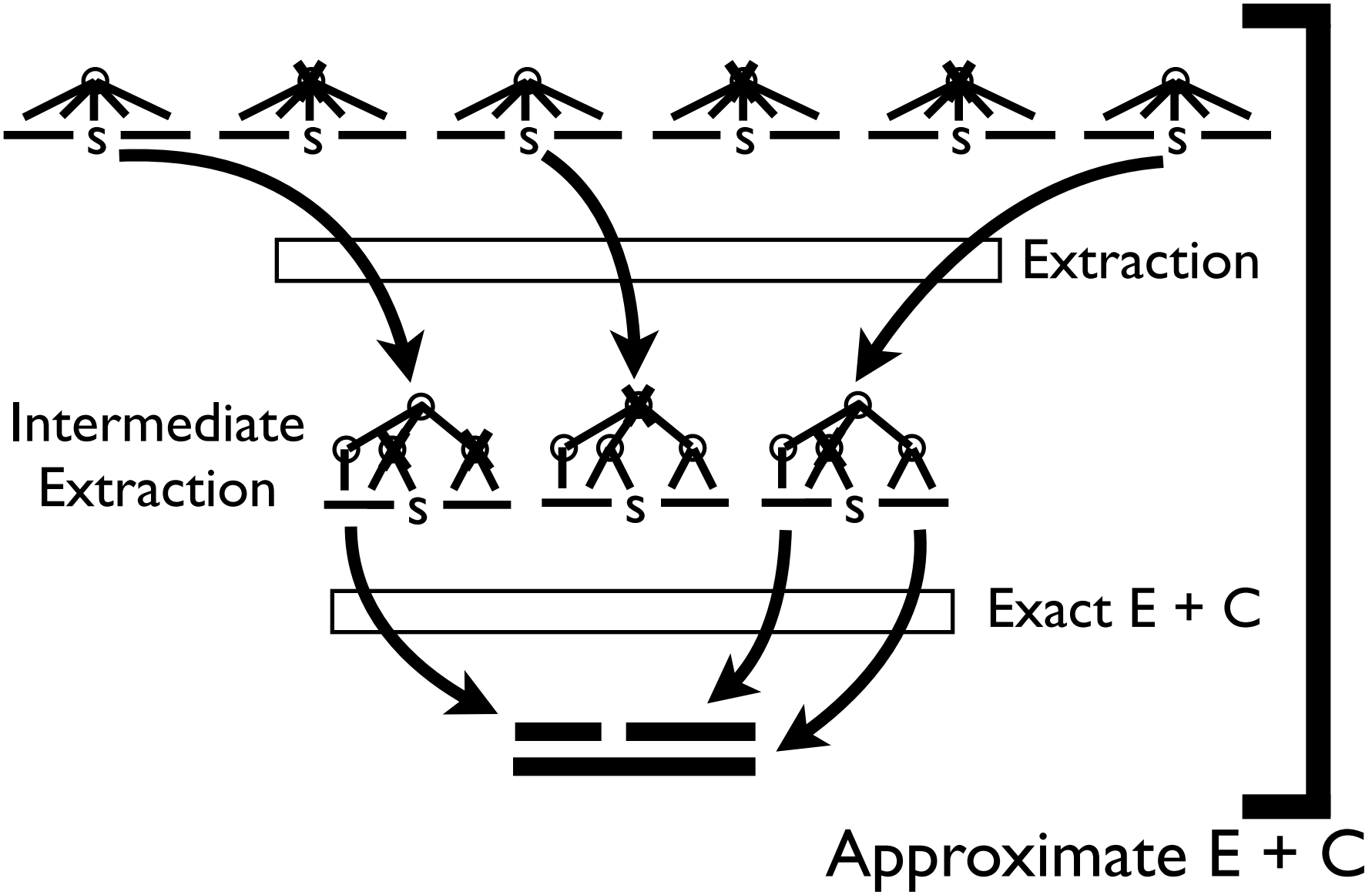
Berkeley
N L P

# Coarse-to-fine Decoding

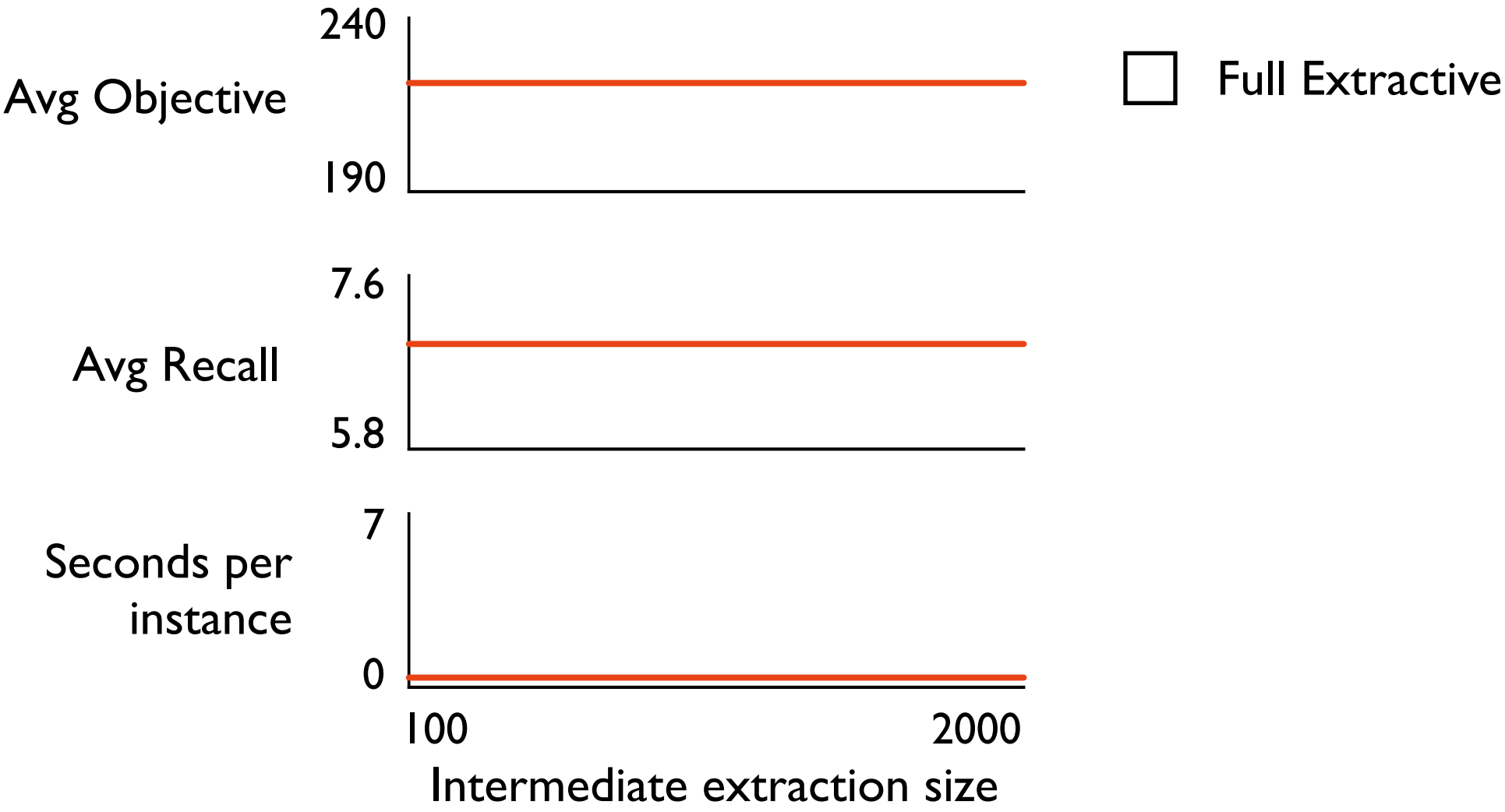


Berkeley NLP

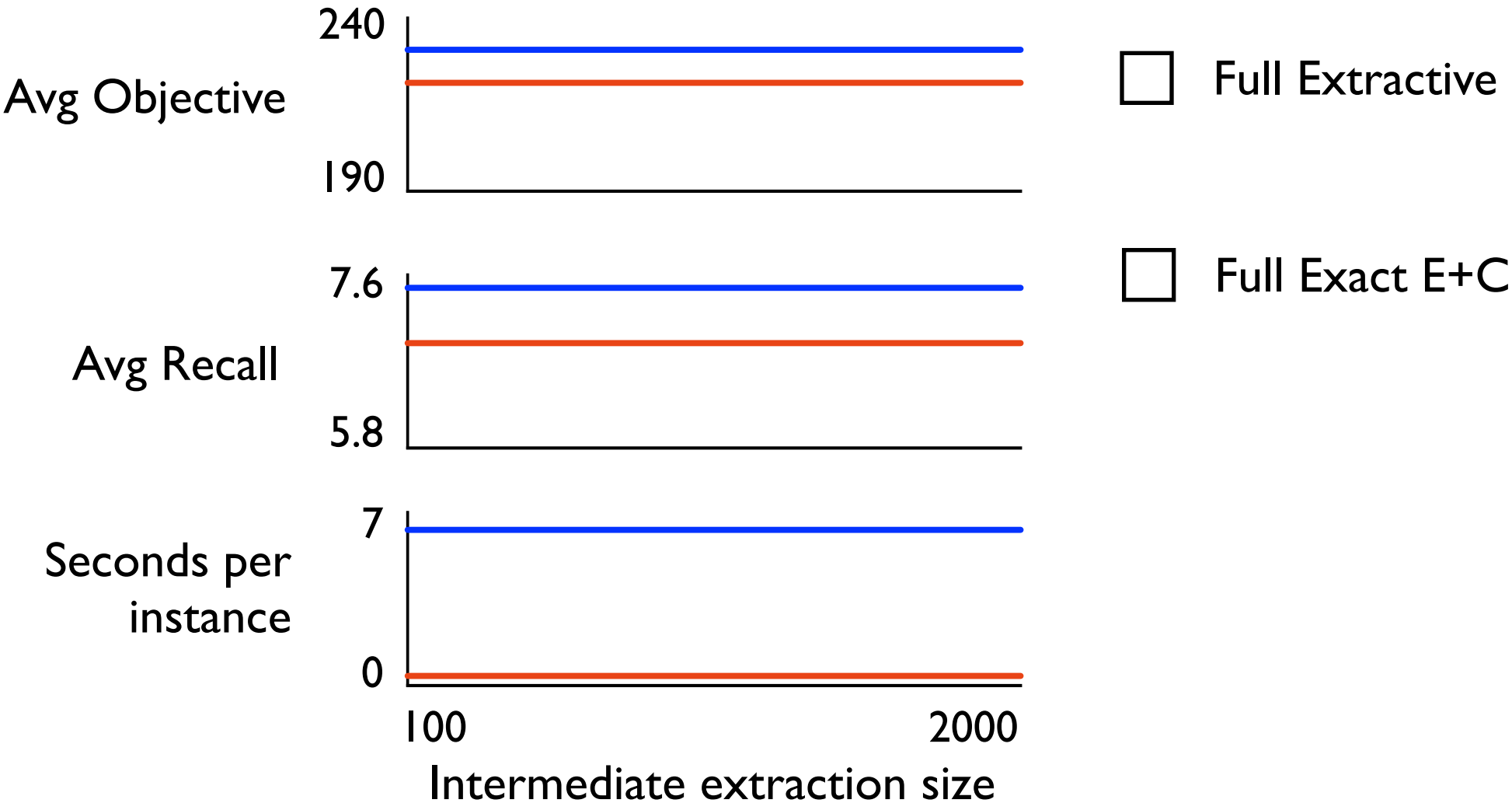
# Coarse-to-fine Decoding



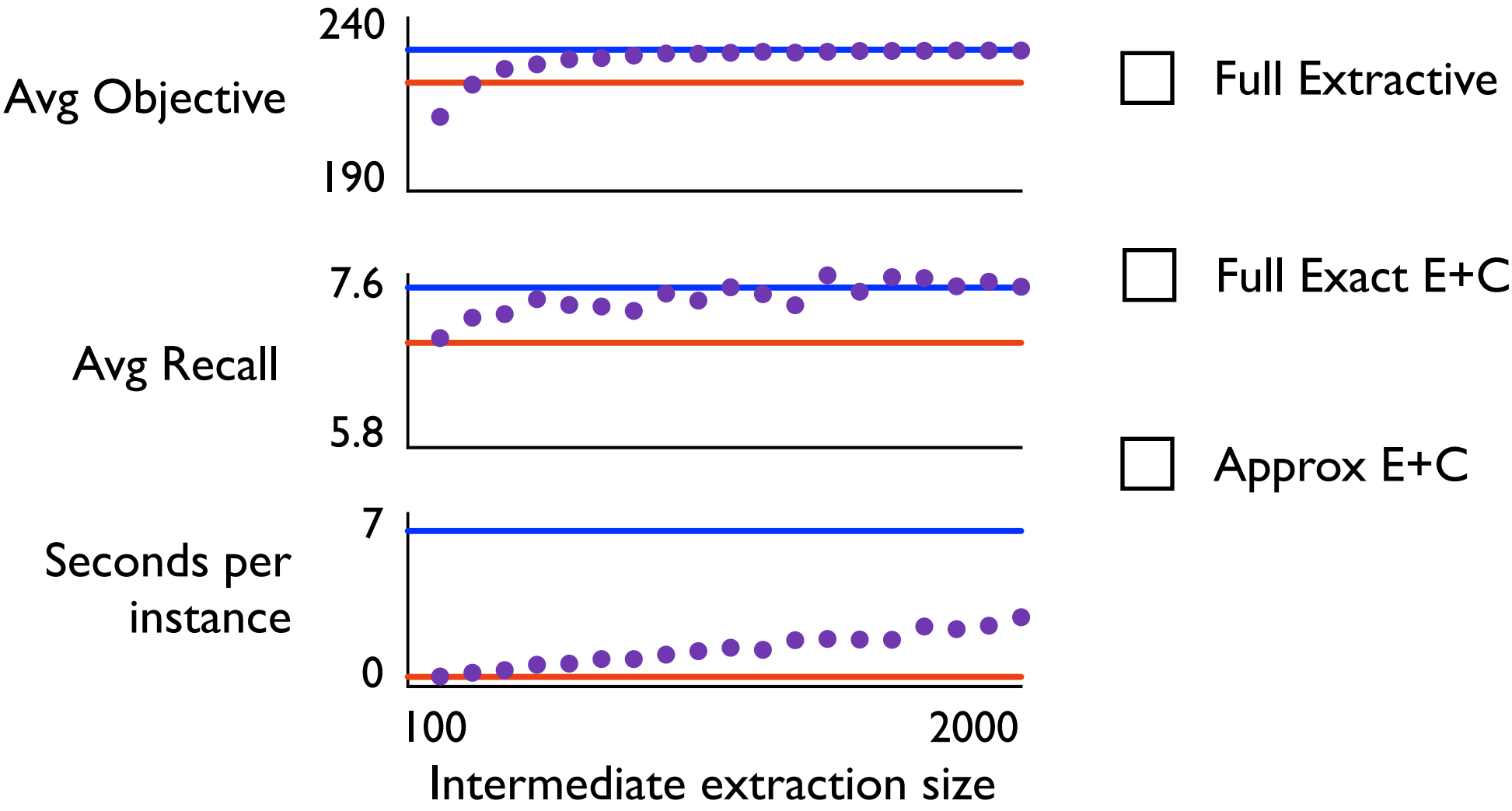
# Coarse-to-fine Decoding



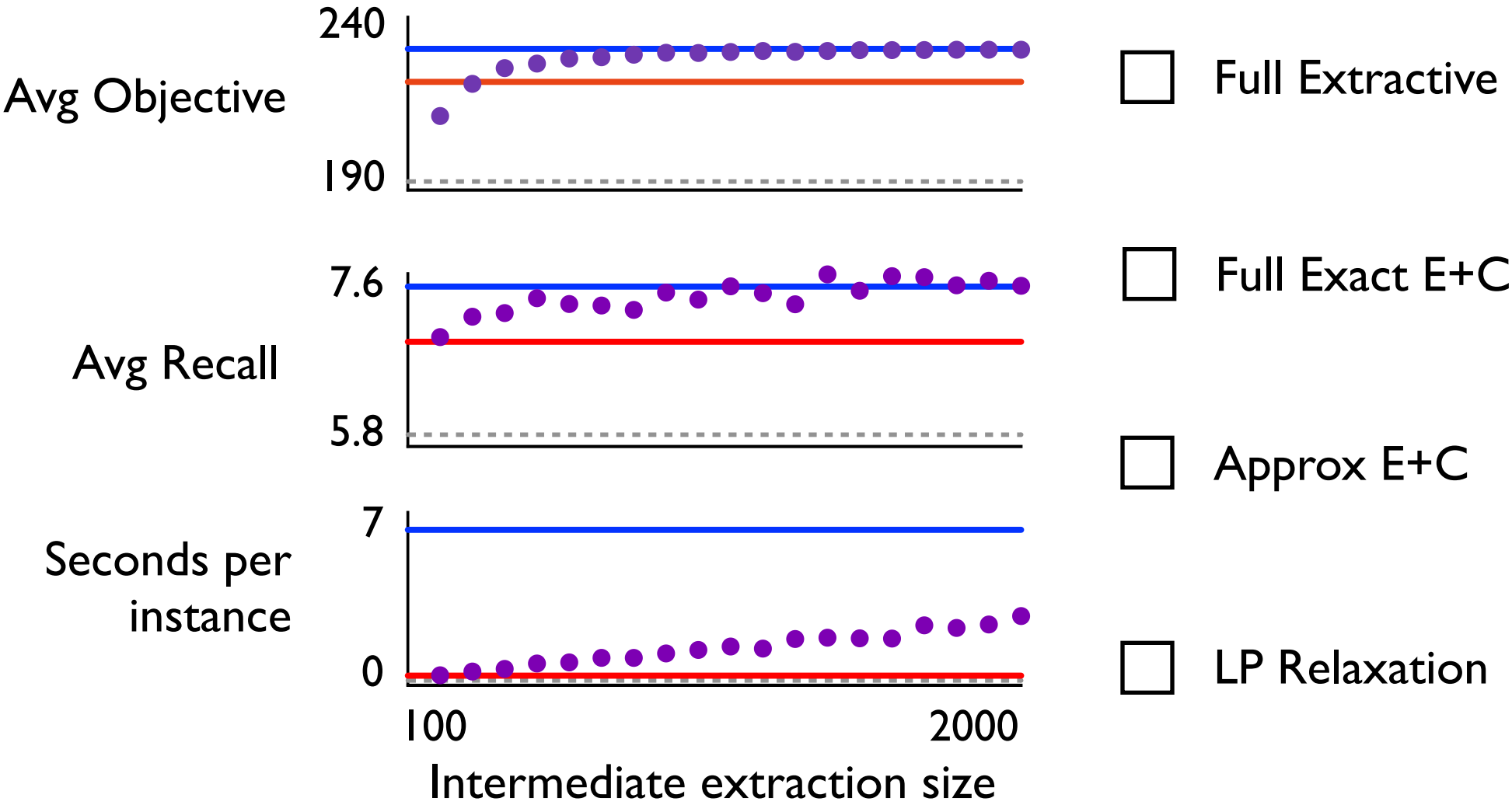
# Coarse-to-fine Decoding



# Coarse-to-fine Decoding



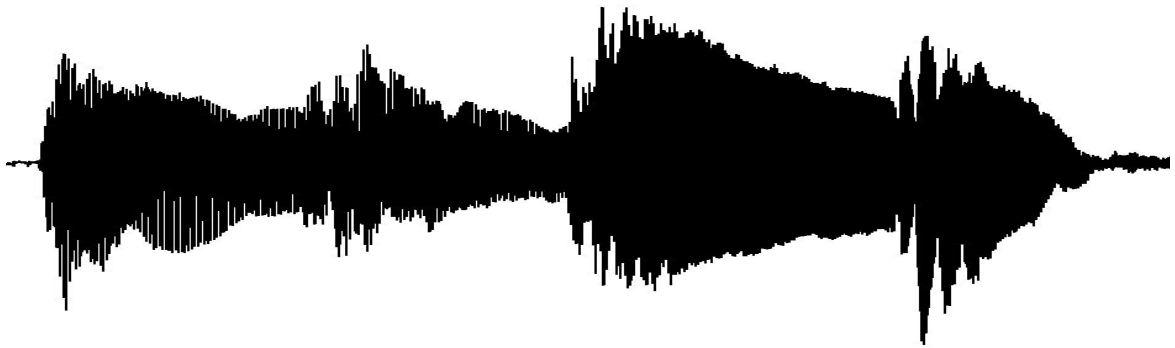
# Coarse-to-fine Decoding




# Piano Music Transcription

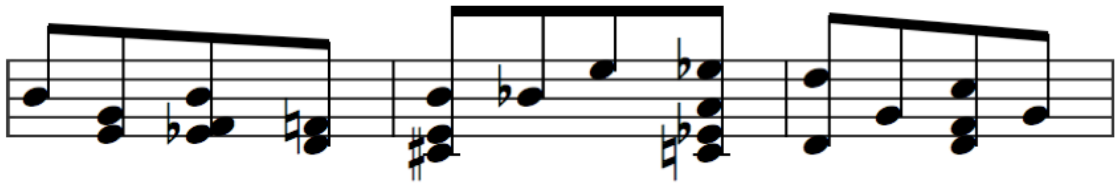
Berkeley NLP

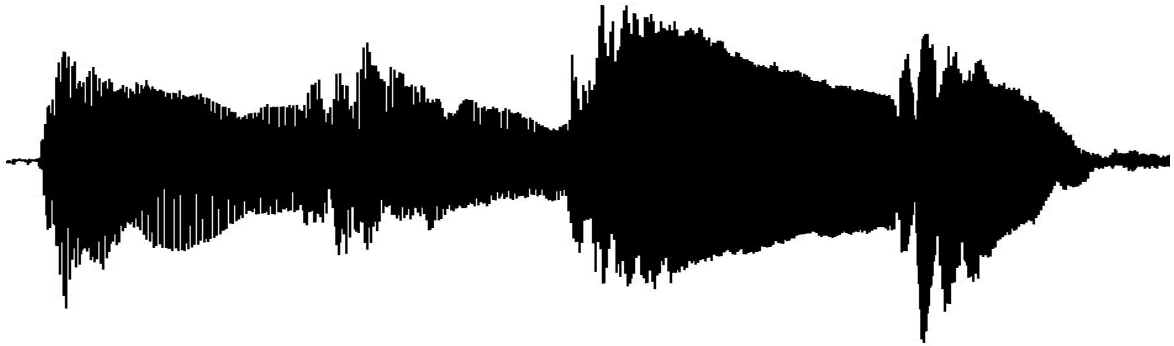
# Piano Music Transcription



Berkeley
N L P

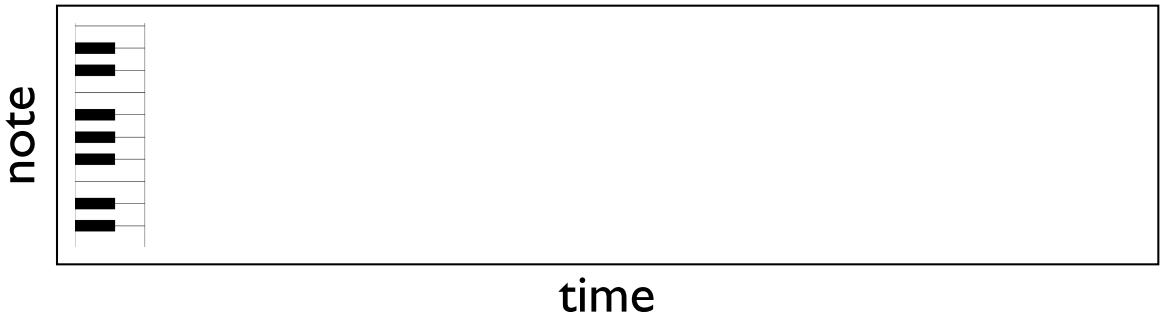
# Piano Music Transcription

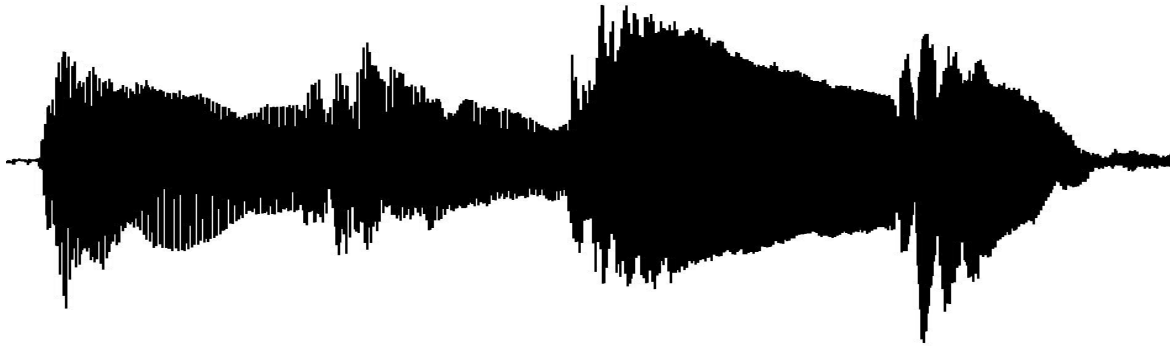




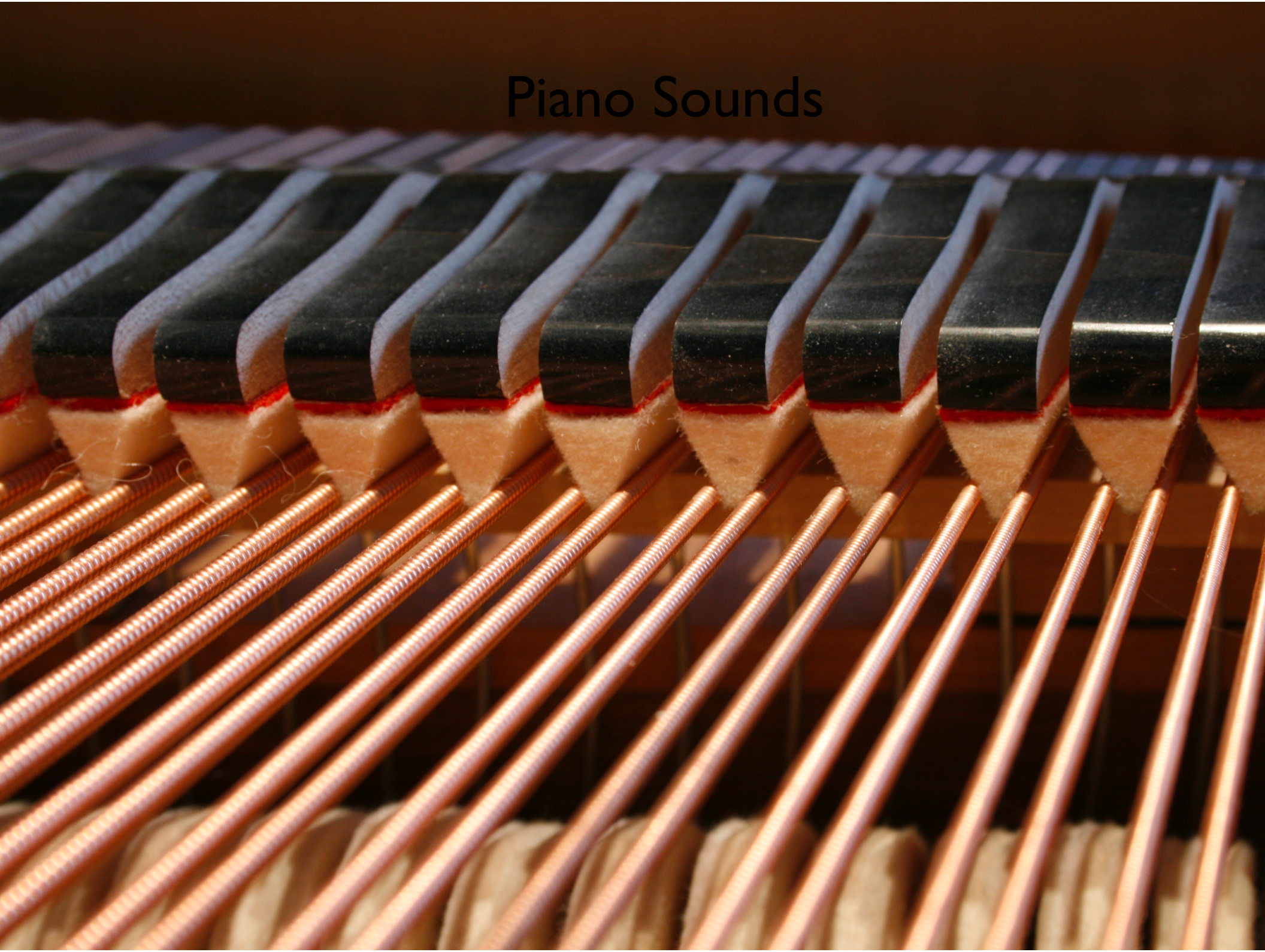
Berkeley NLP

# Piano Music Transcription



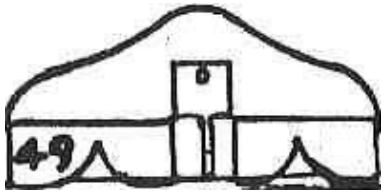



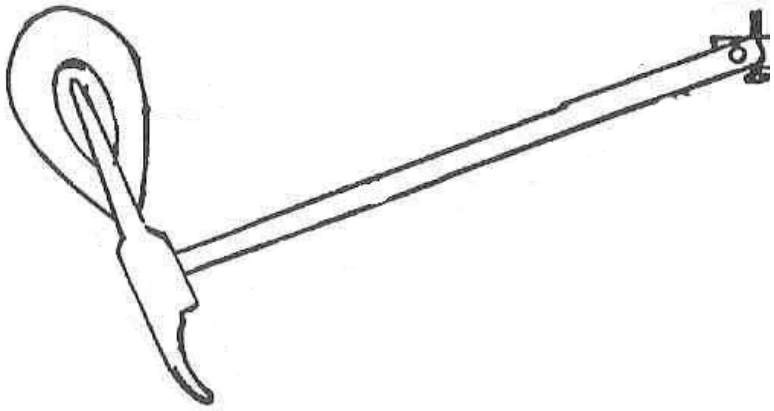
# Piano Sounds



Berkeley
N L P

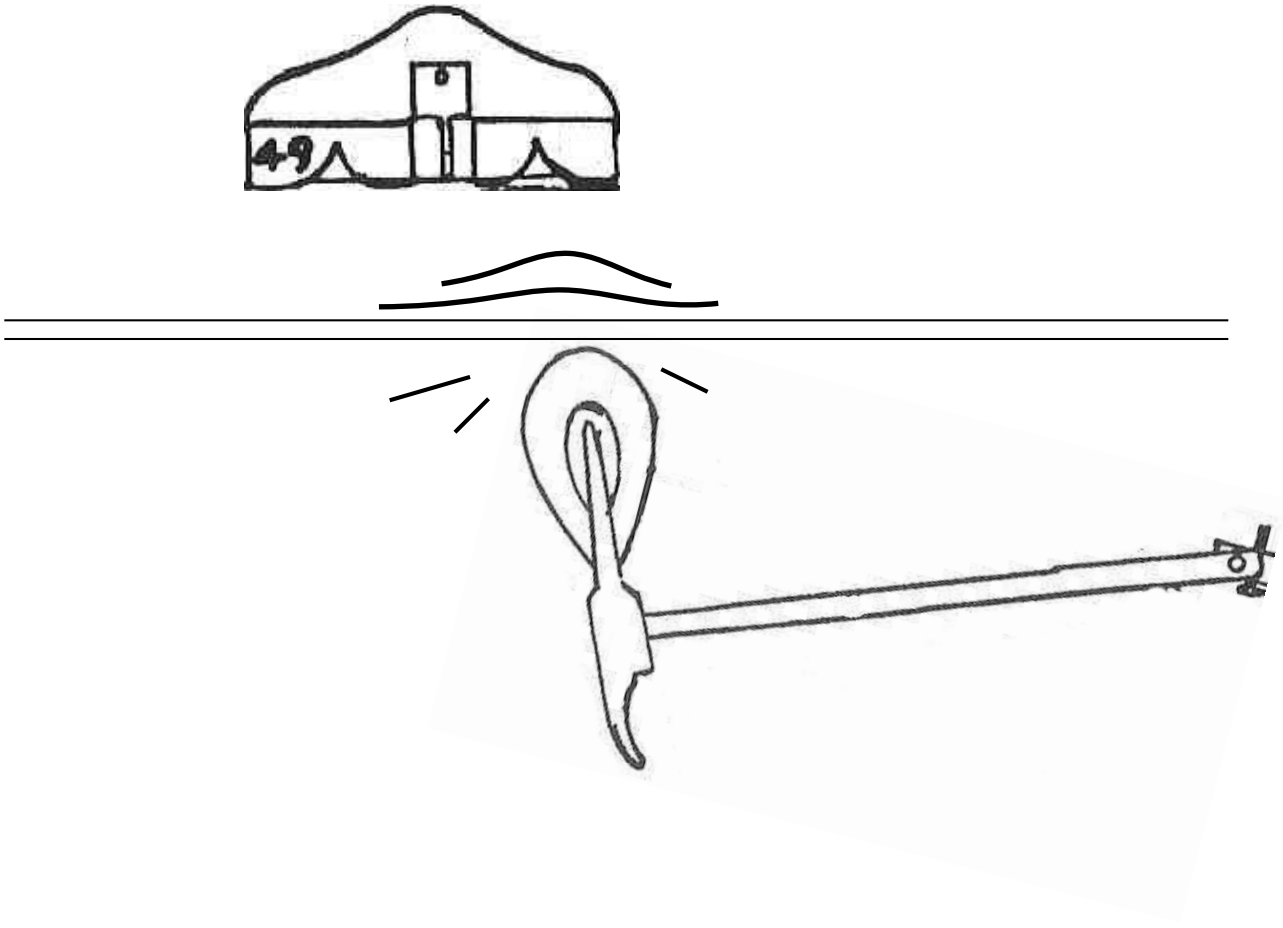
# Piano Sounds



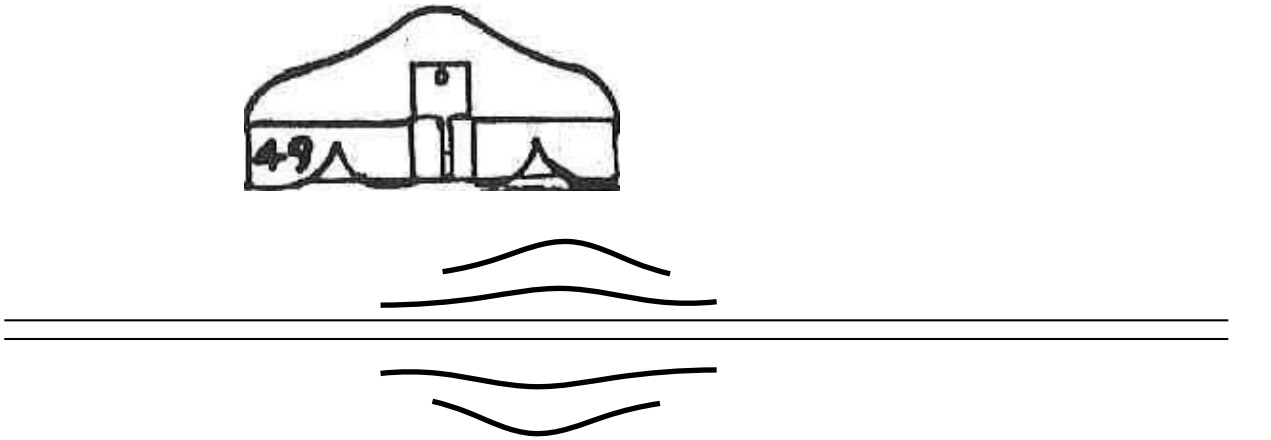
Berkeley
N L P

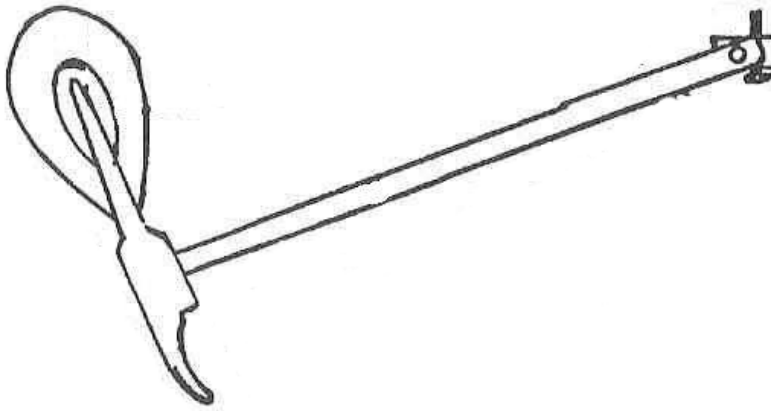
# Piano Sounds



Berkeley
N L P

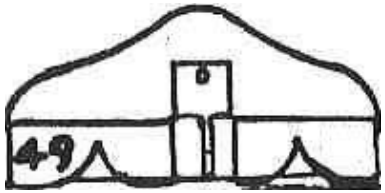
# Piano Sounds

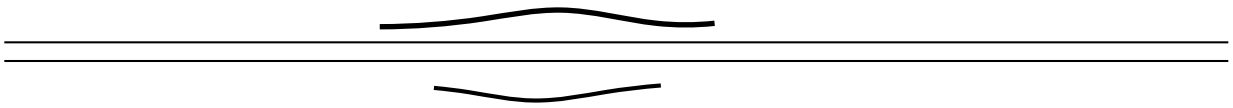


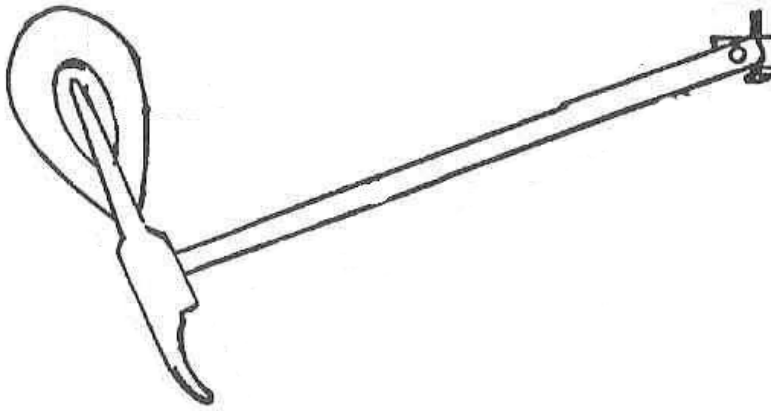



Berkeley
N L P

# Piano Sounds

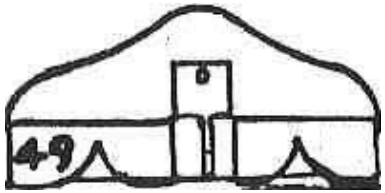


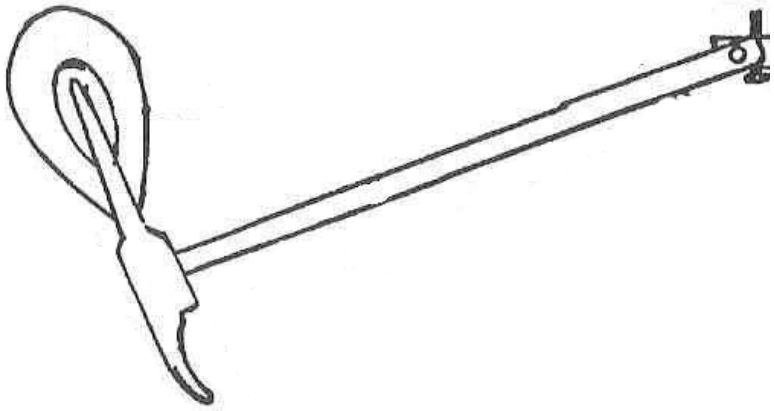





Berkeley
N L P

# Piano Sounds

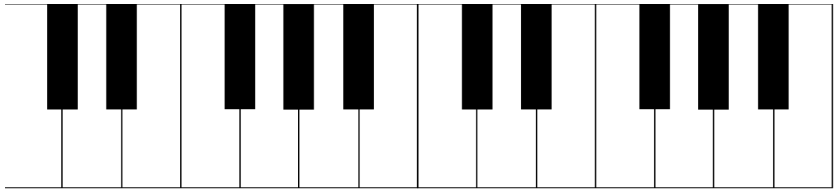


Berkeley

N L P

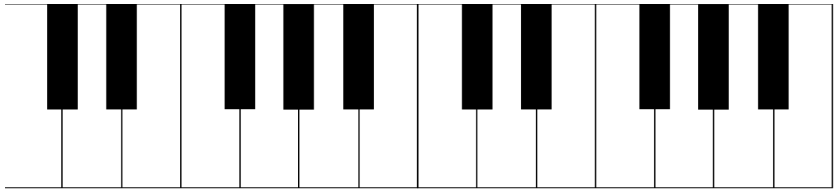
# Piano Sounds



Berkeley

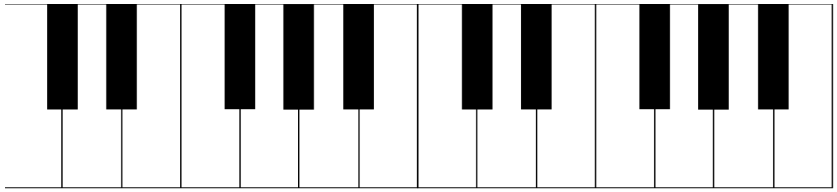
N L P

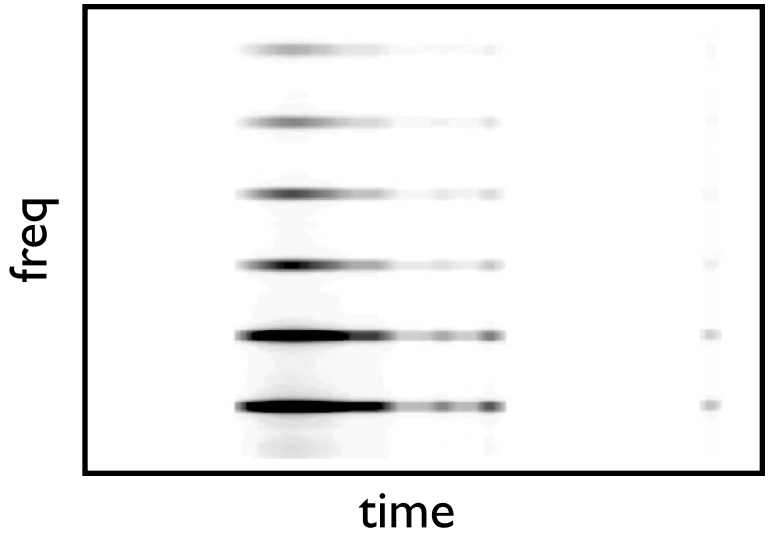
# Piano Sounds



Berkeley
N L P

# Piano Sounds

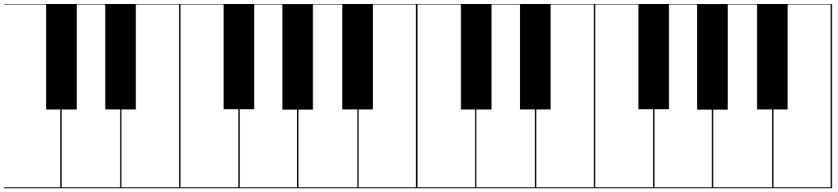


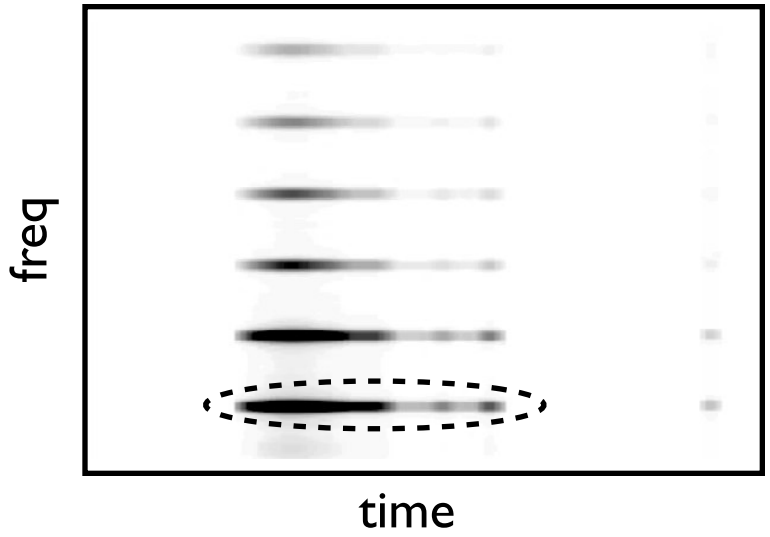


Berkeley

N L P

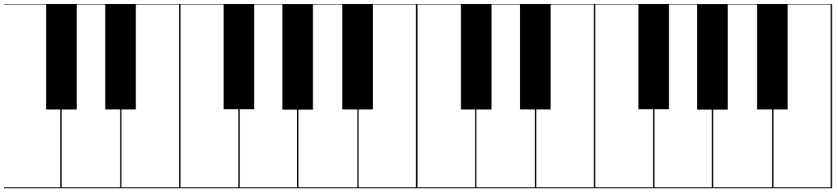
# Piano Sounds

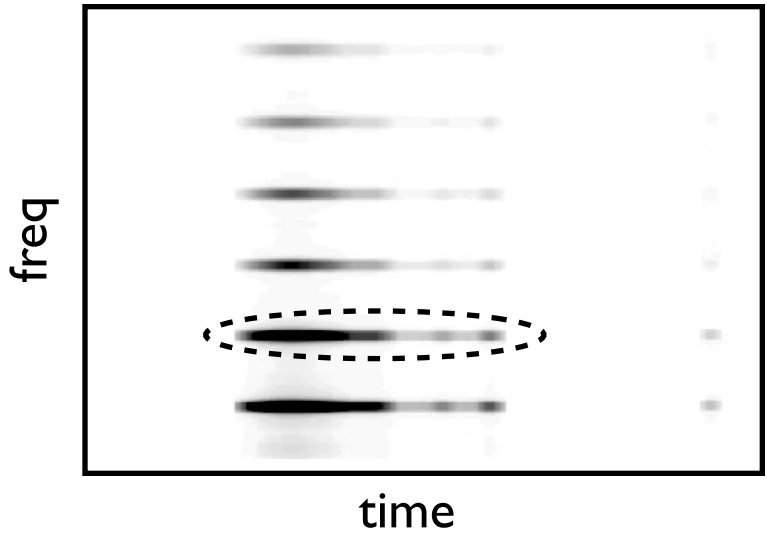




Berkeley
N L P

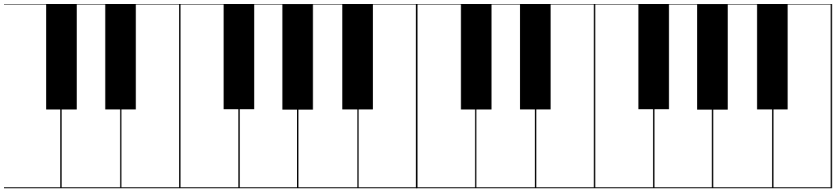
# Piano Sounds

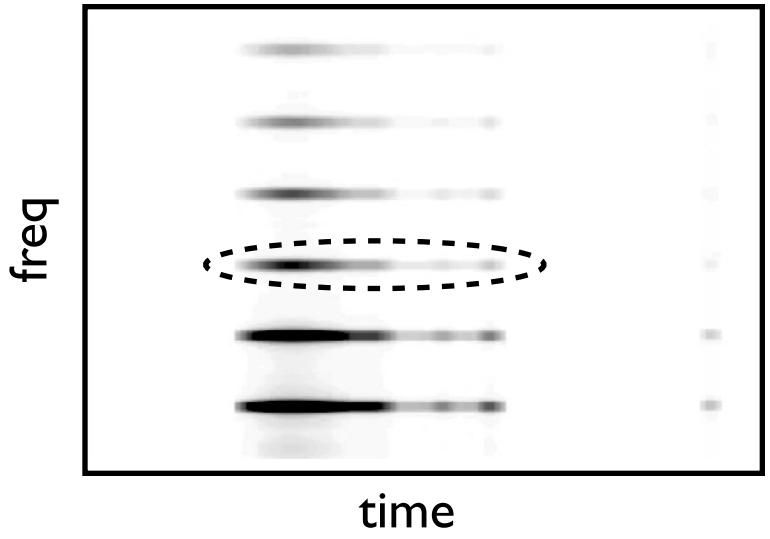




Berkeley
N L P

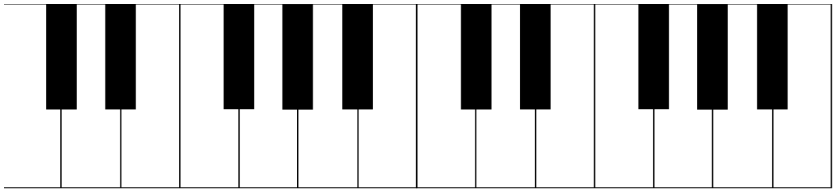
# Piano Sounds

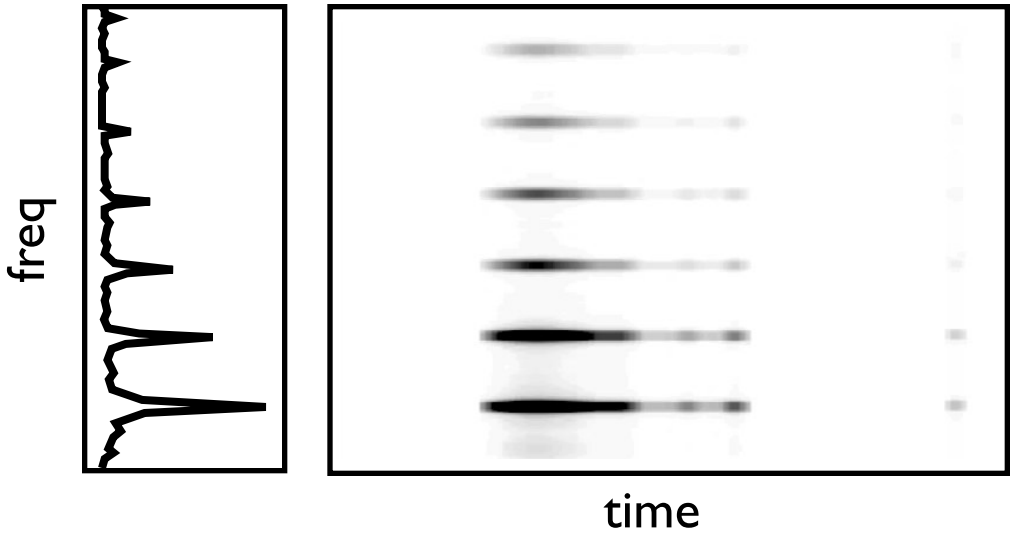




Berkeley
N L P

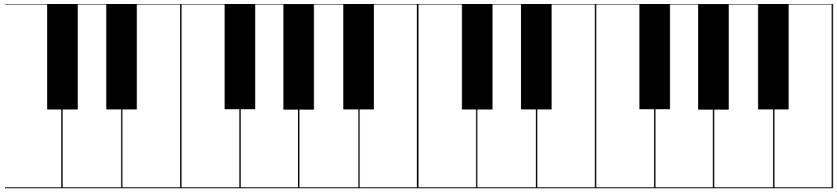
# Spectral Shape

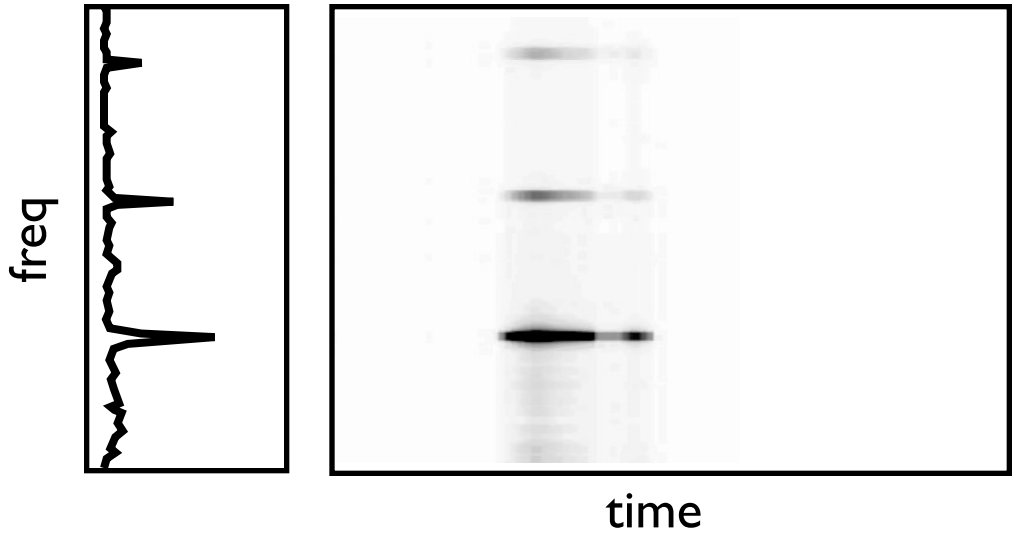




Berkeley
N L P

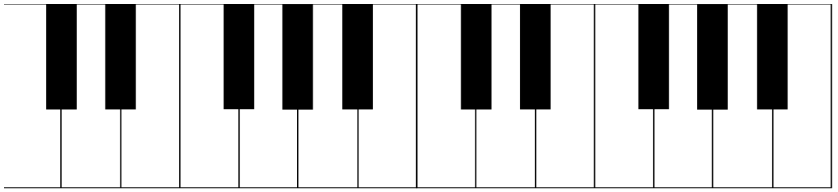
# Spectral Shape

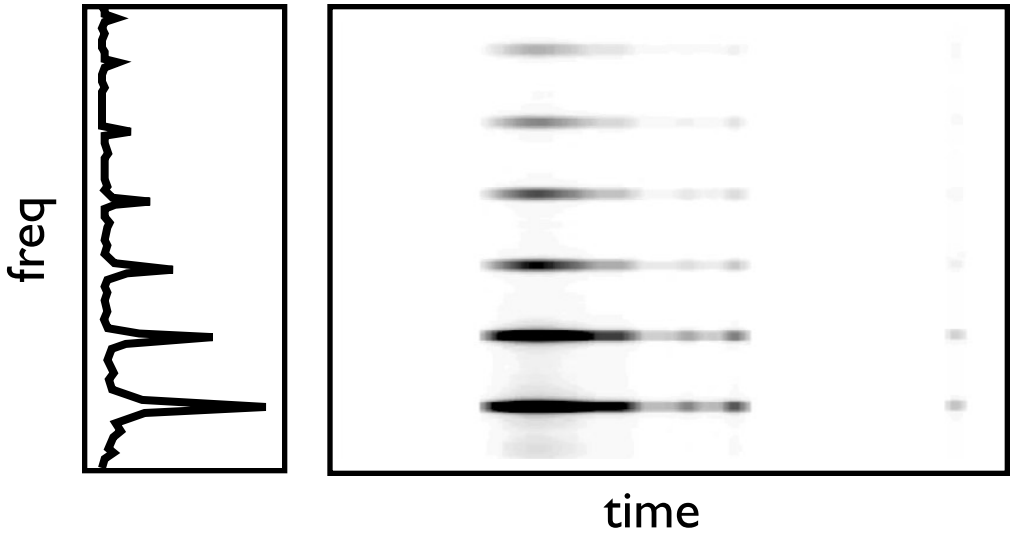




Berkeley
N L P

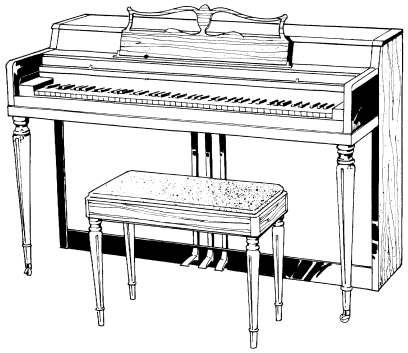
# Spectral Shape

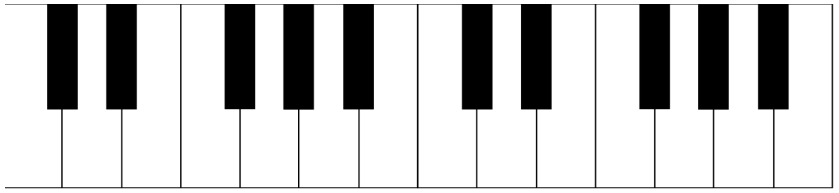


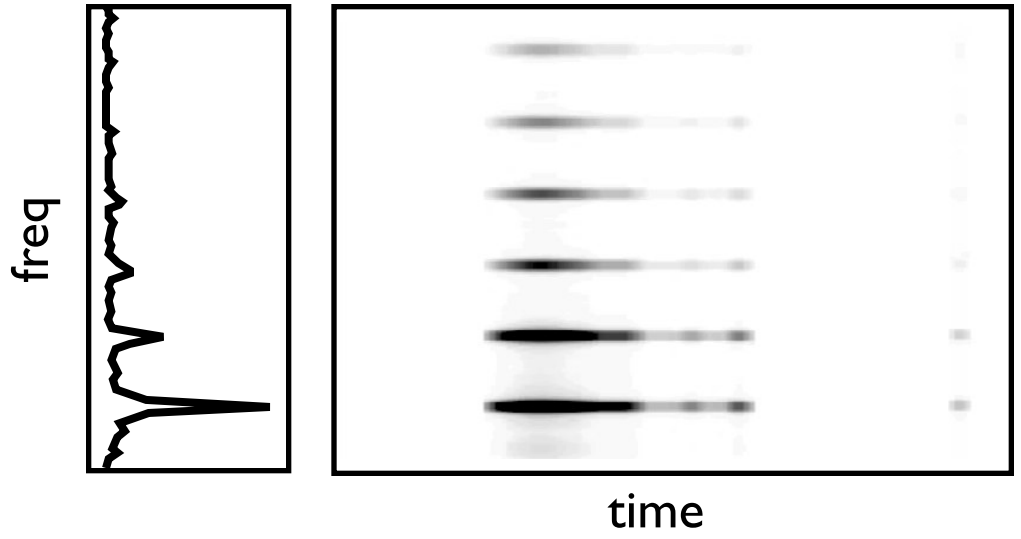


Berkeley NLP

# Spectral Shape

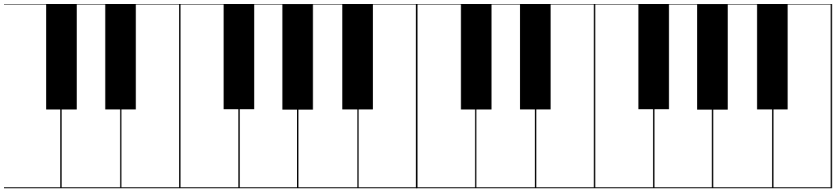


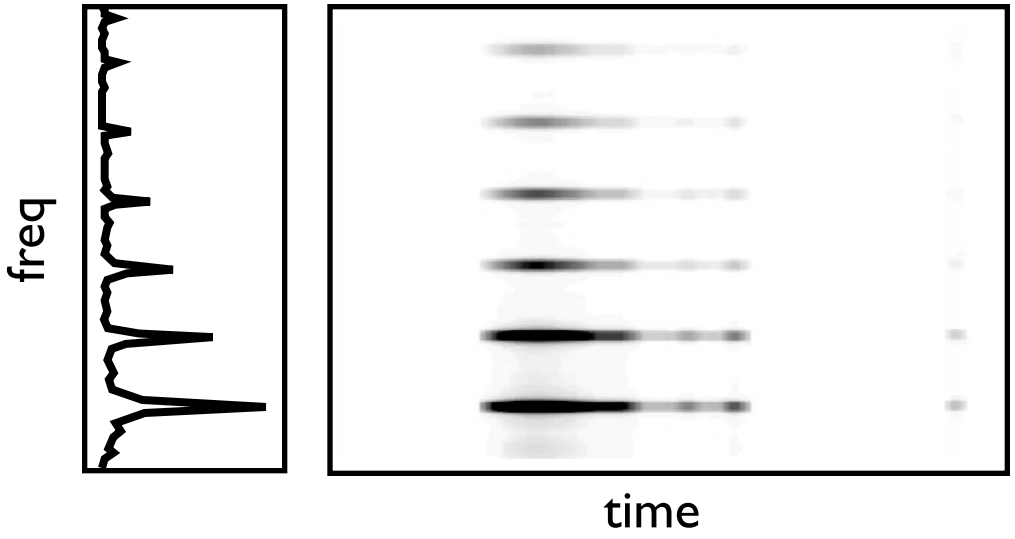




Berkeley
N L P

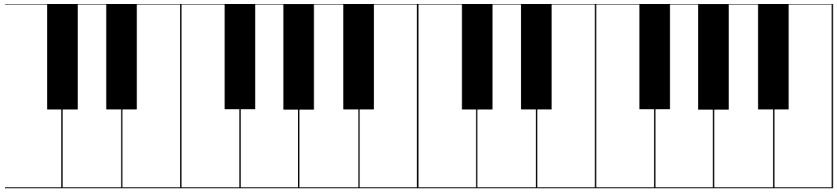
# Temporal Shape

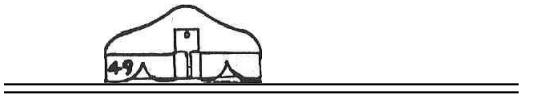


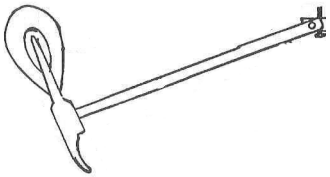


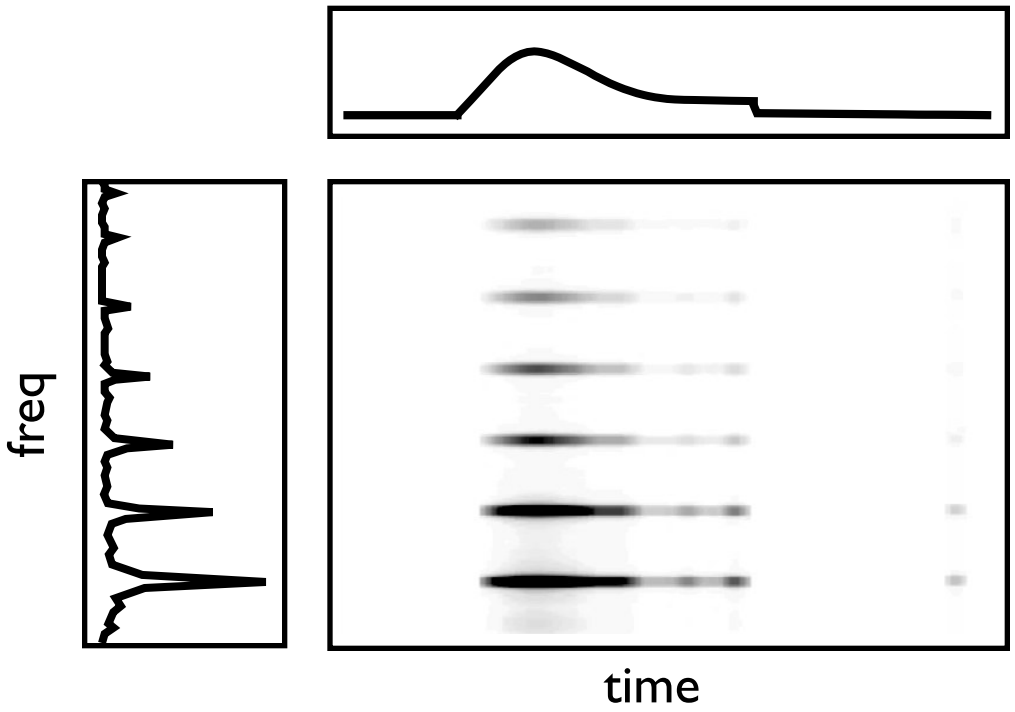
Berkeley
N L P

# Temporal Shape

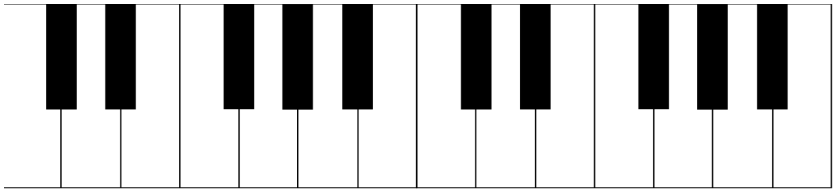


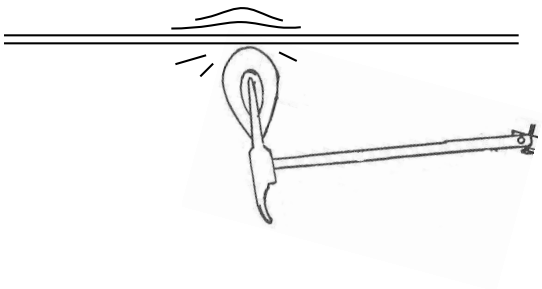


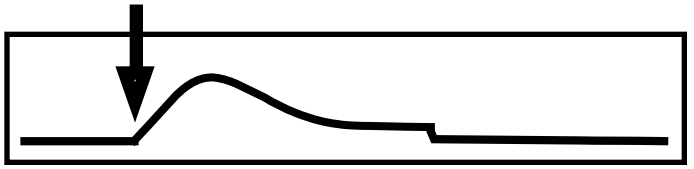


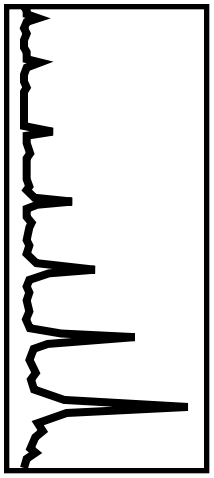


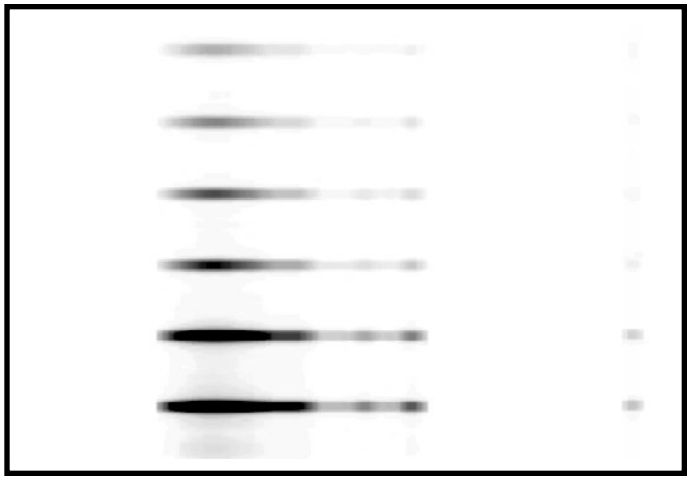
# Temporal Shape









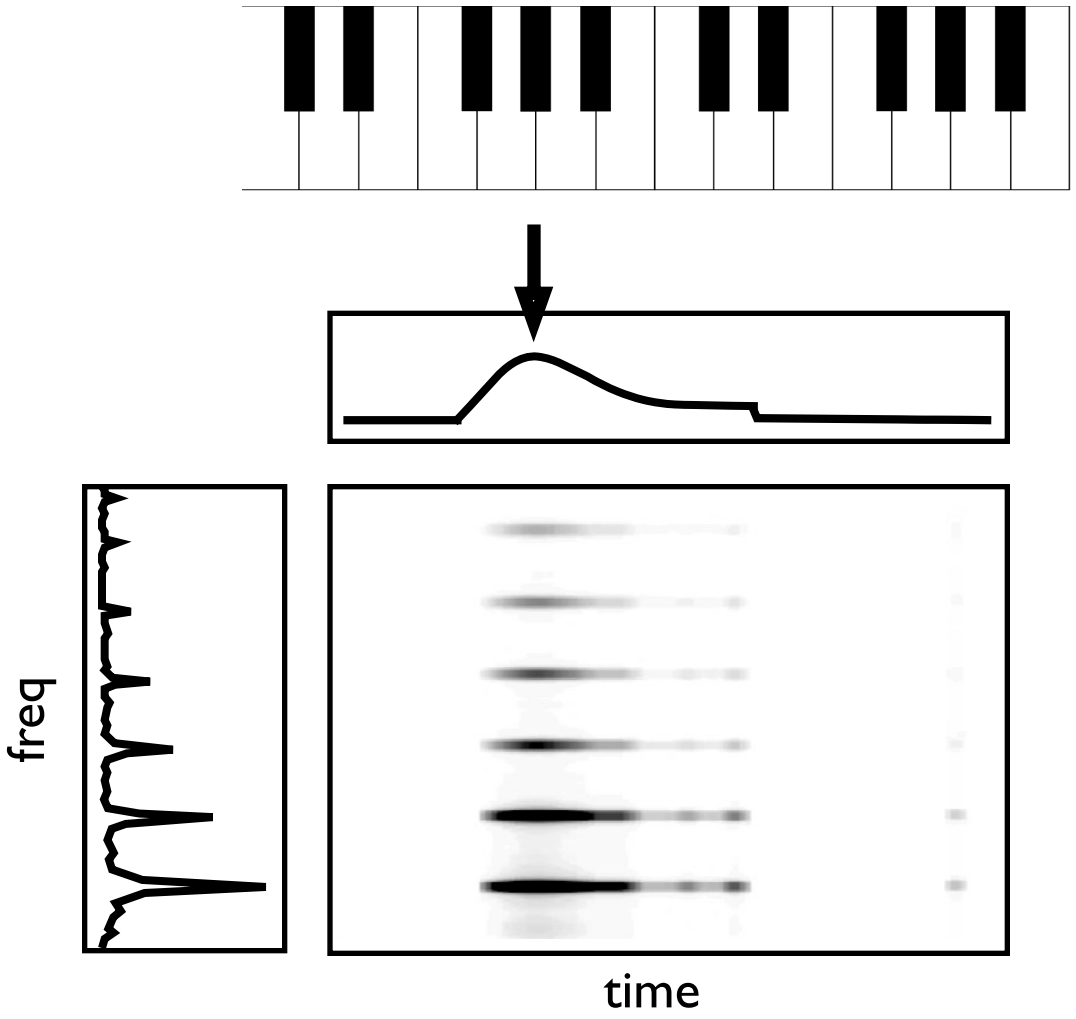


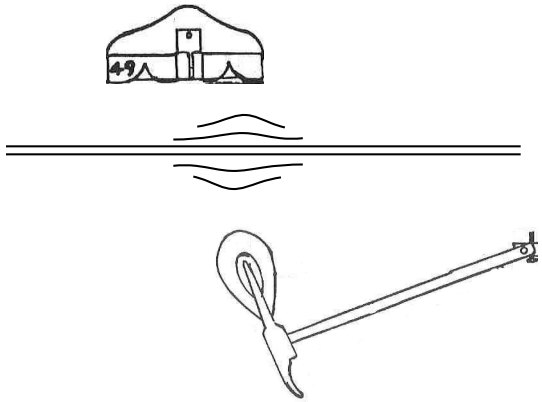
freq

time

Berkeley
N L P

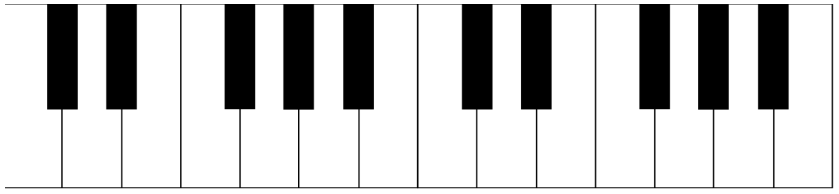
# Temporal Shape

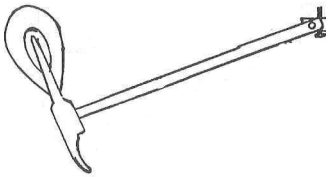


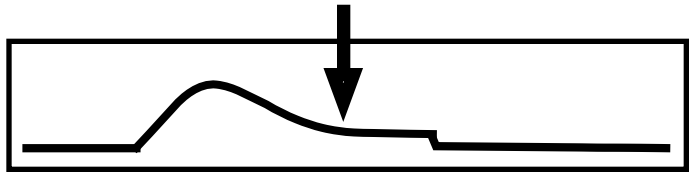



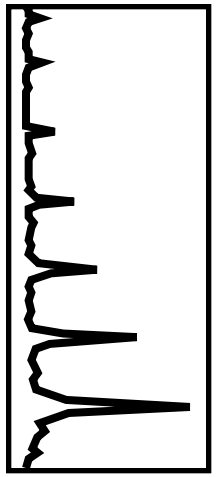
Berkeley
N L P

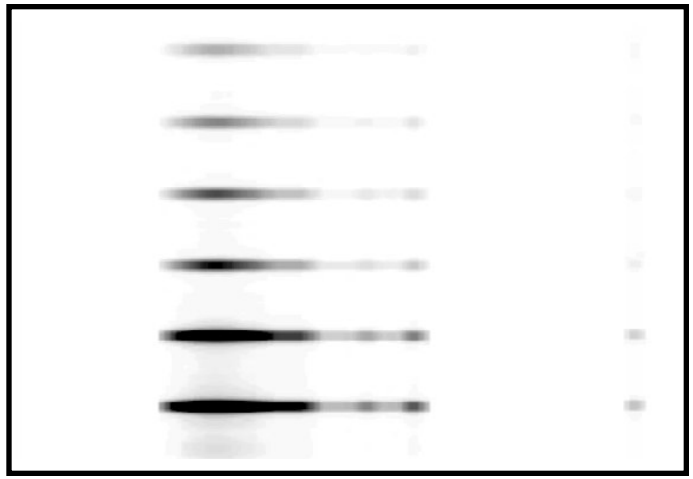
# Temporal Shape









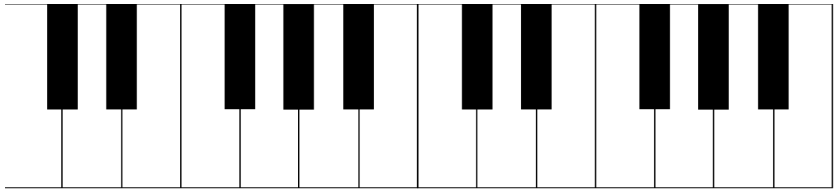


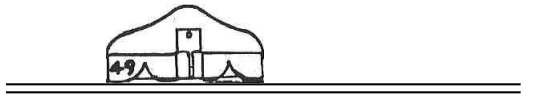
freq

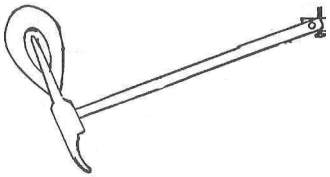
time

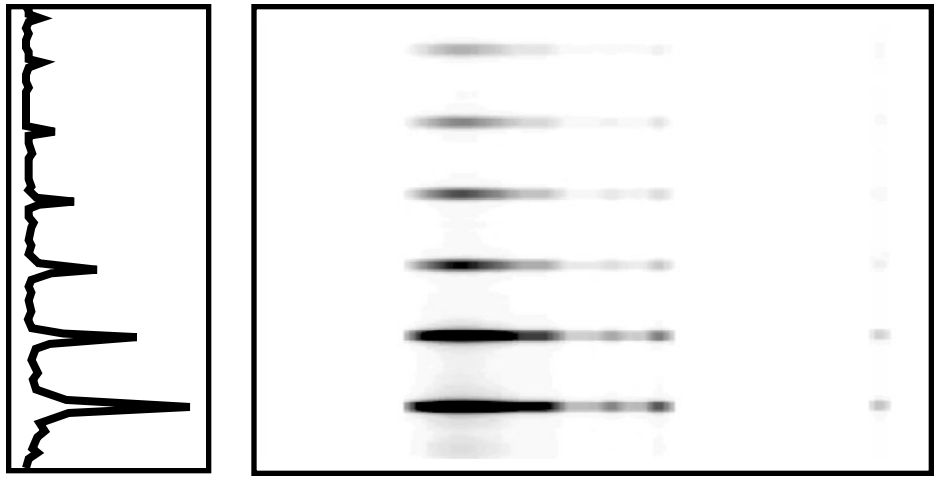
Berkeley
N L P

# Temporal Shape



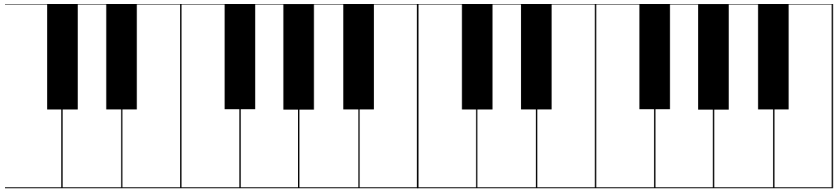


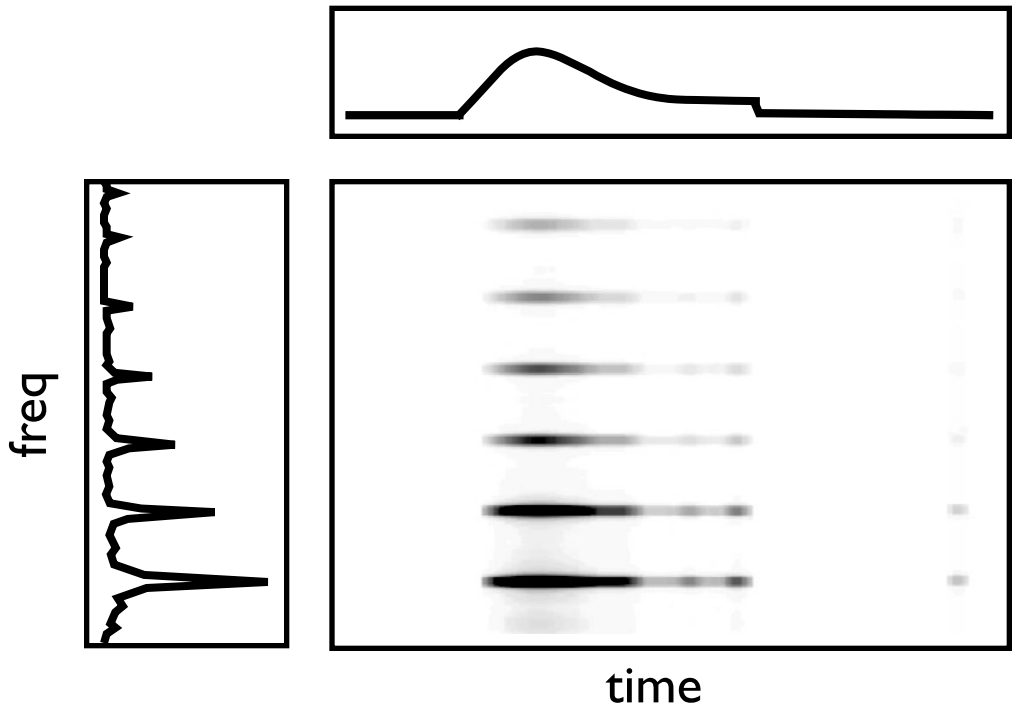




Berkeley
N L P

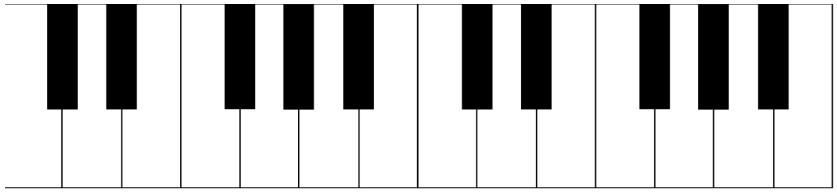
# Temporal Shape

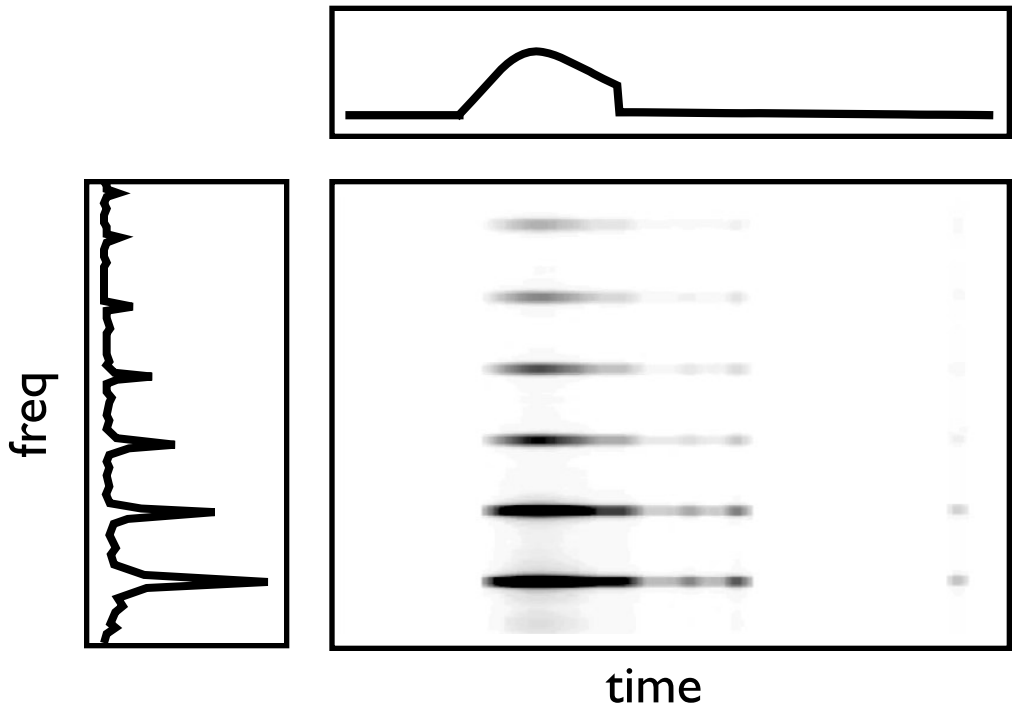




Berkeley
N L P

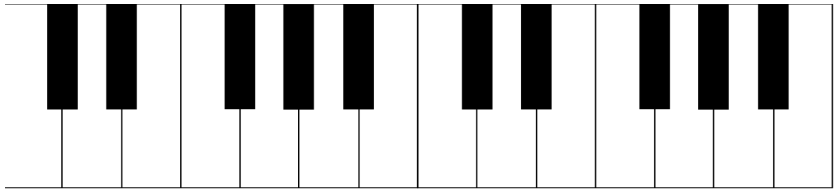
# Temporal Shape

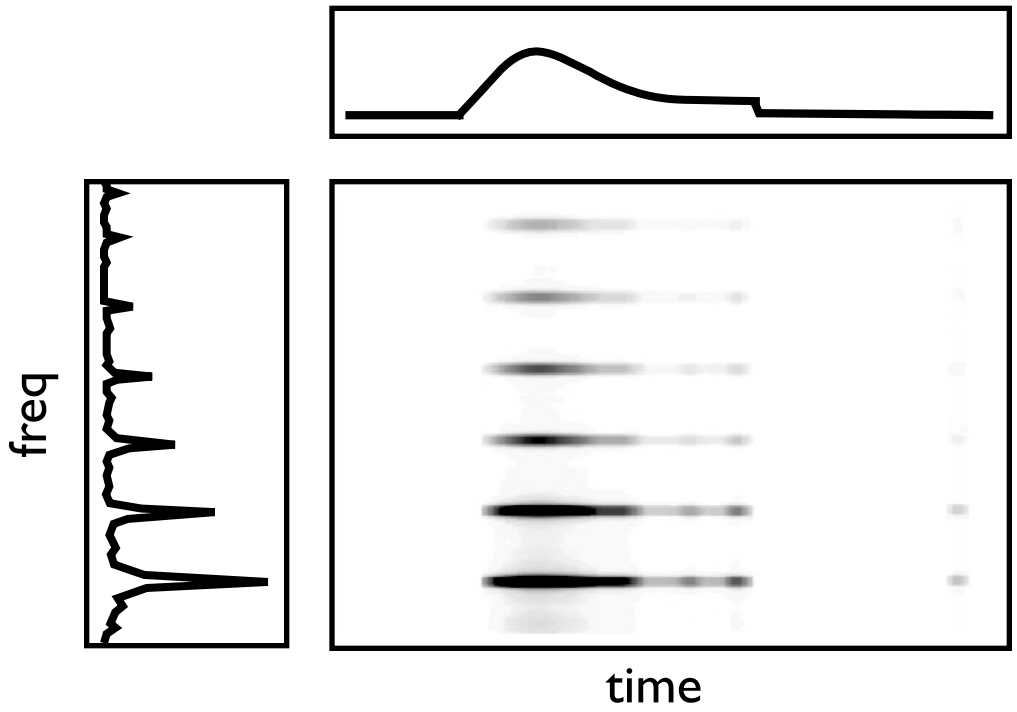




Berkeley
N L P

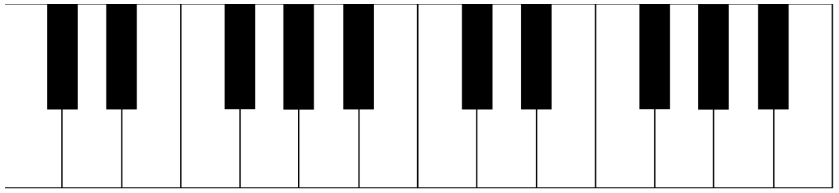
# Temporal Shape

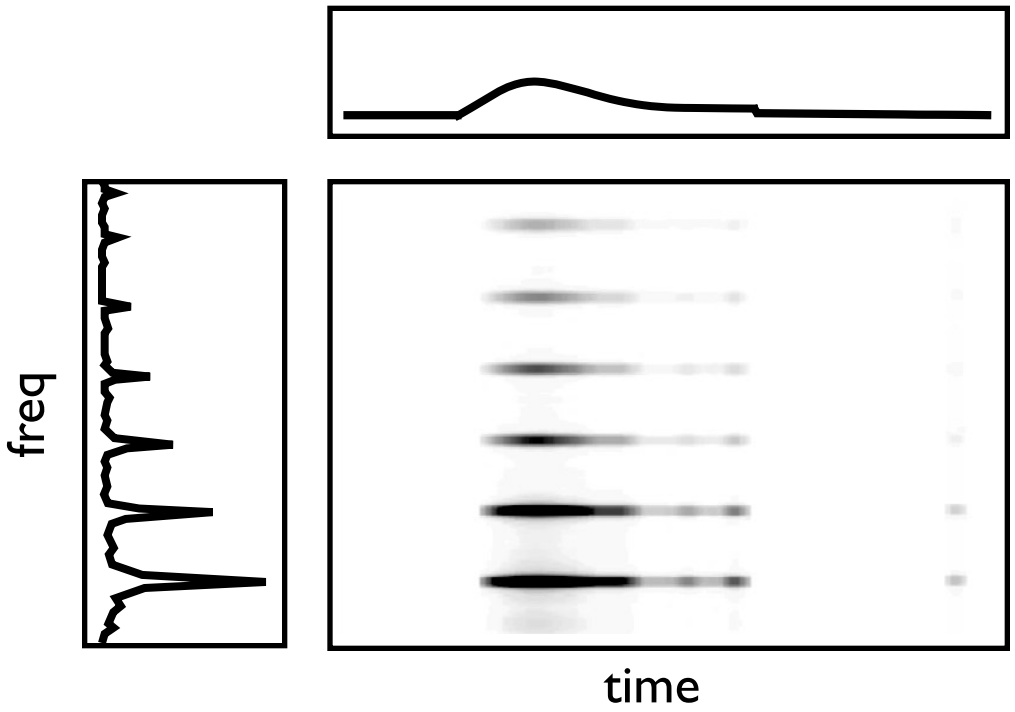




Berkeley
N L P

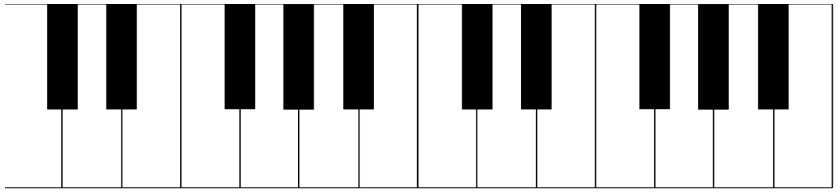
# Temporal Shape

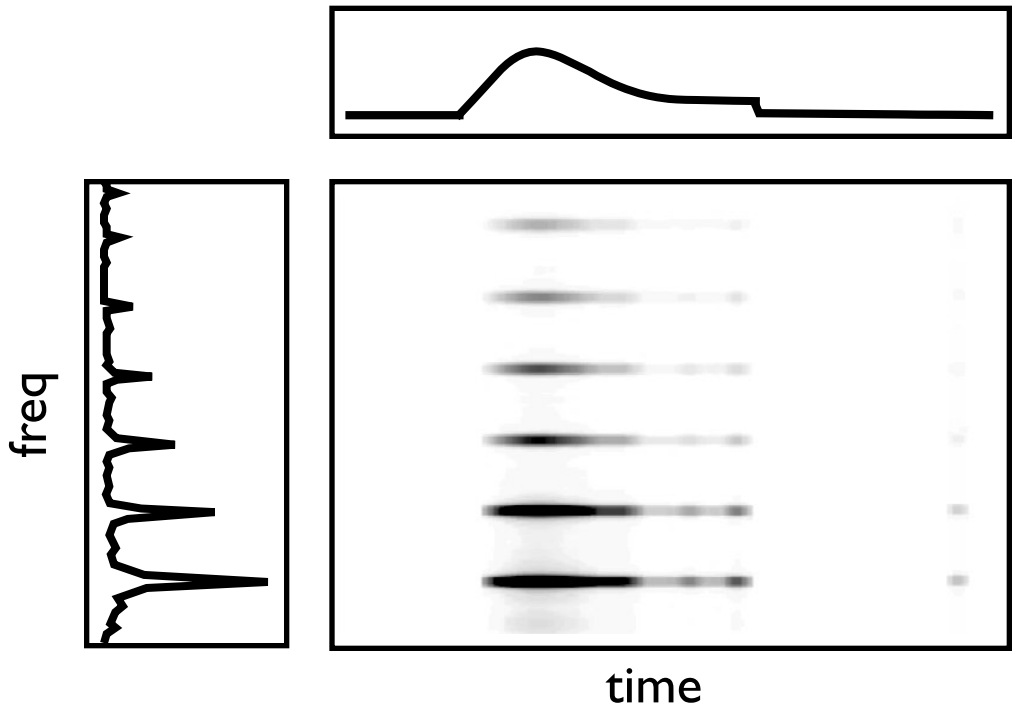




Berkeley
N L P

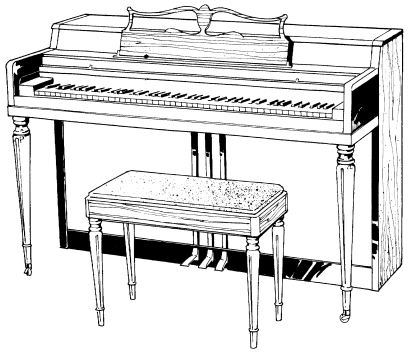
# Temporal Shape

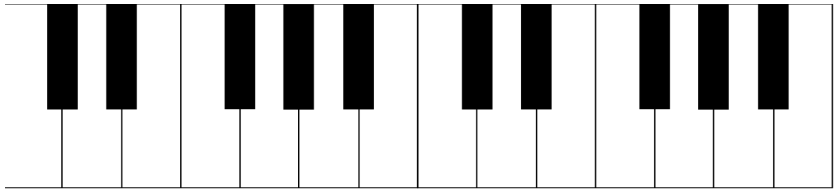


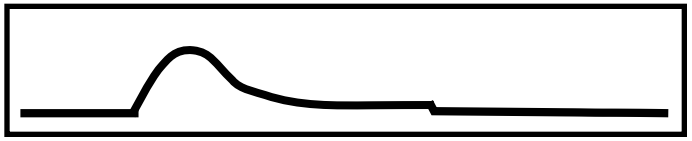


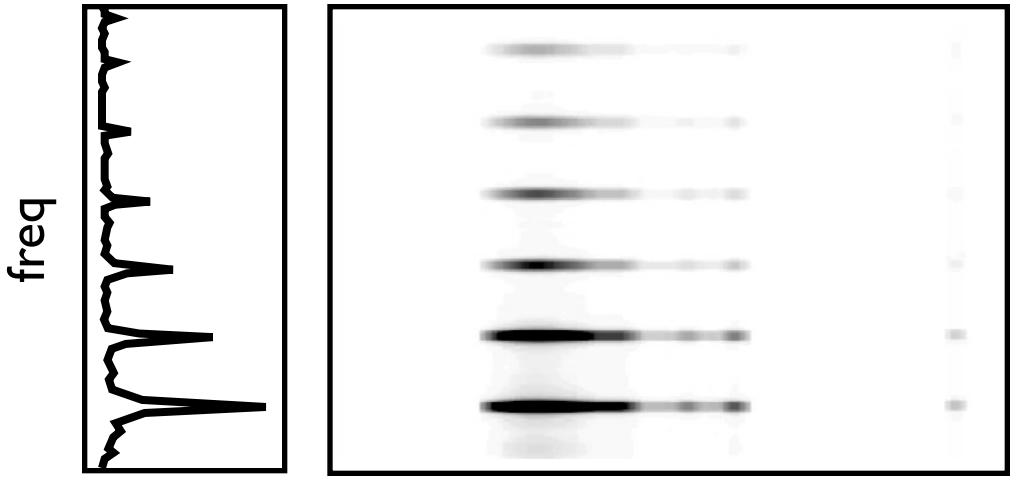
Berkeley NLP

# Temporal Shape



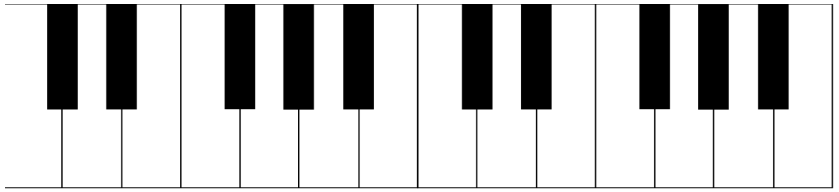


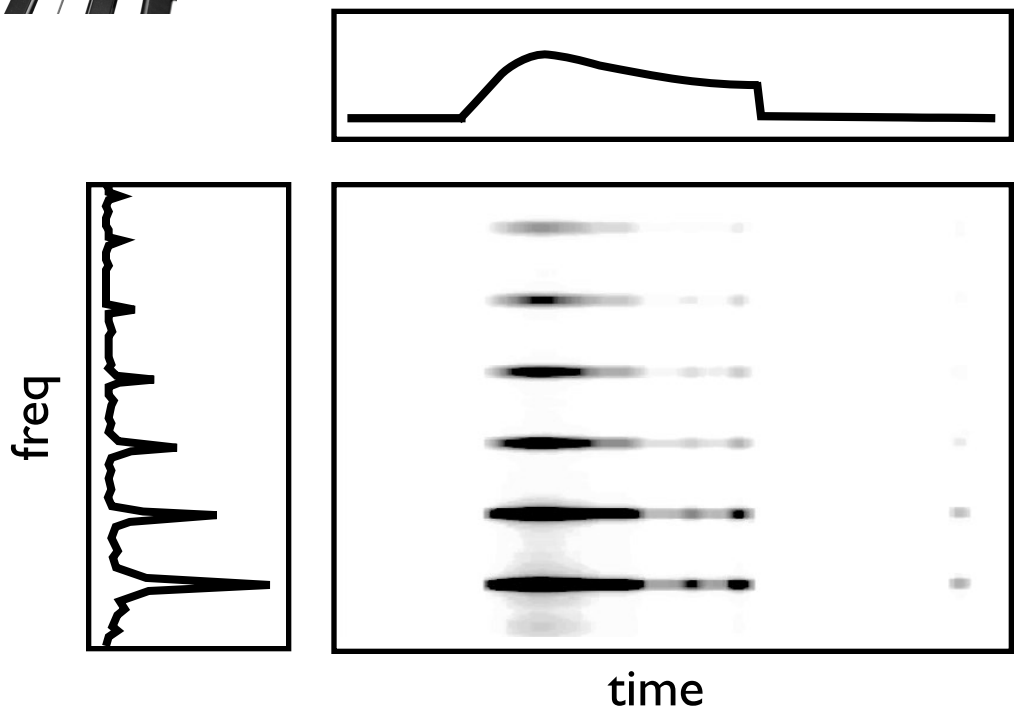




Berkeley
N L P

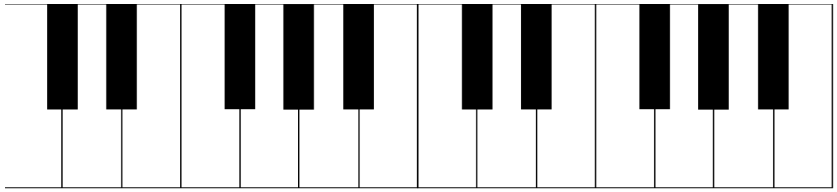
# Temporal Shape



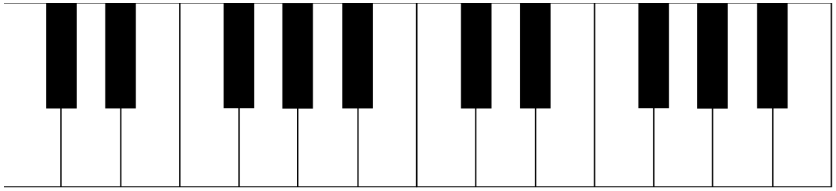
Berkeley
N L P

# Polyphony



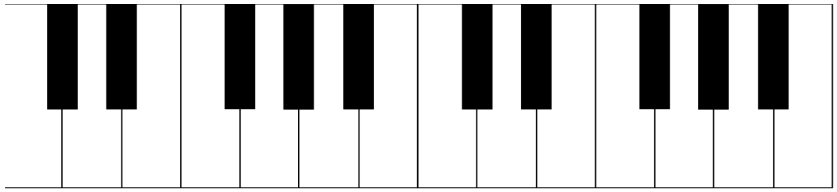
Berkeley NLP

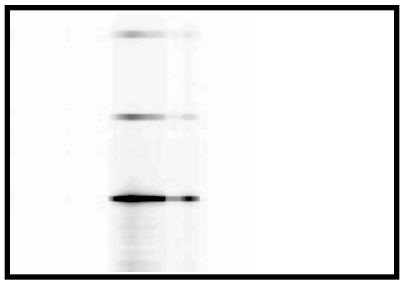
# Polyphony



Berkeley
N L P

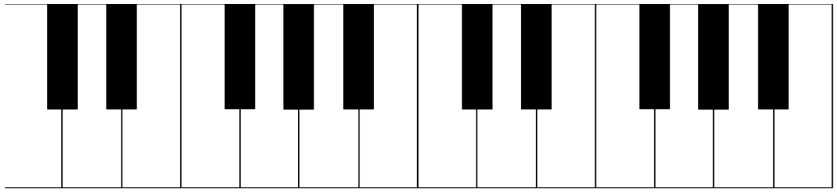
# Polyphony

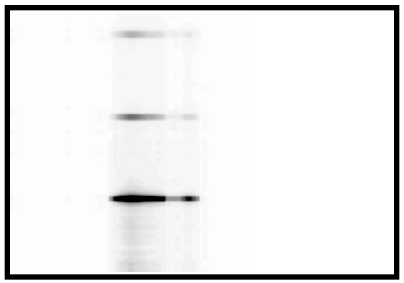


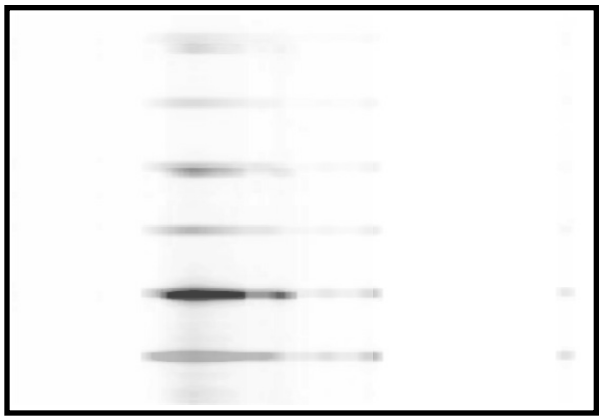


Berkeley
N L P

# Polyphony

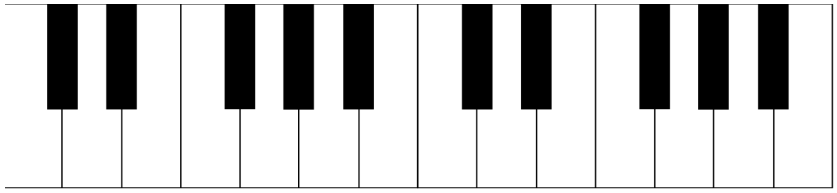


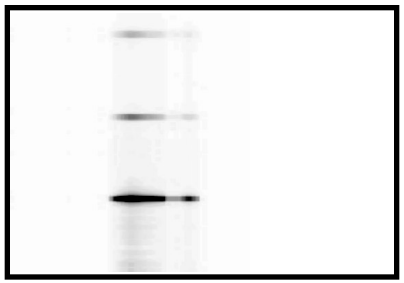


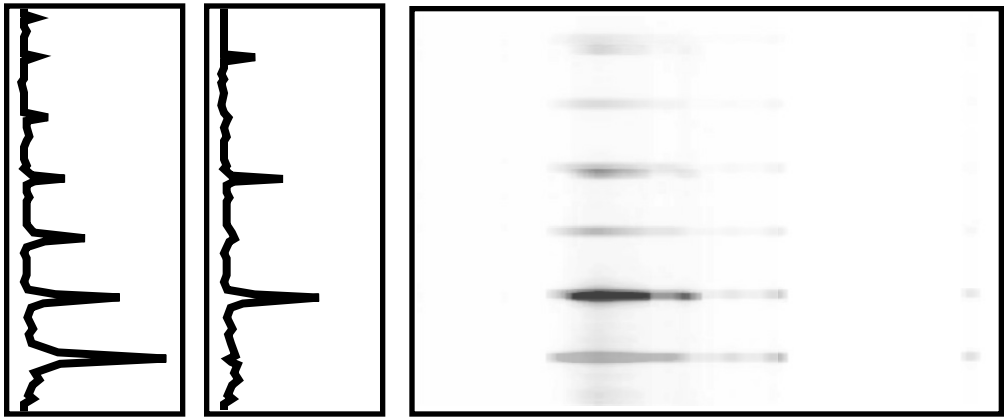


Berkeley
N L P

# Polyphony

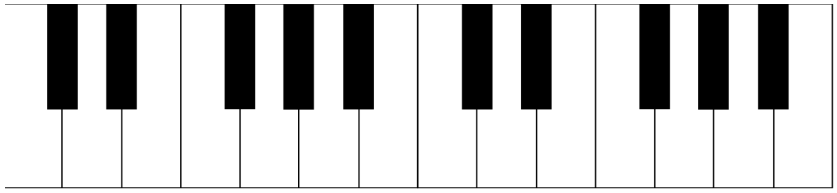


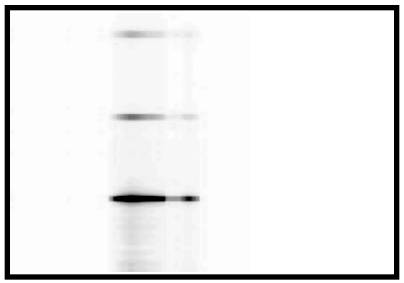


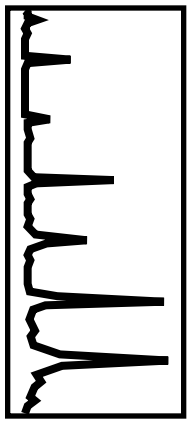


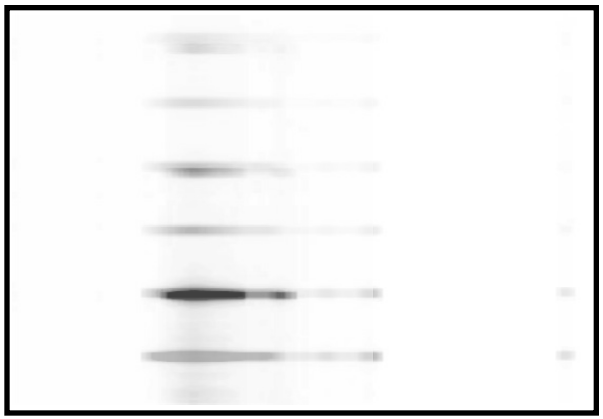
Berkeley NLP

# Polyphony



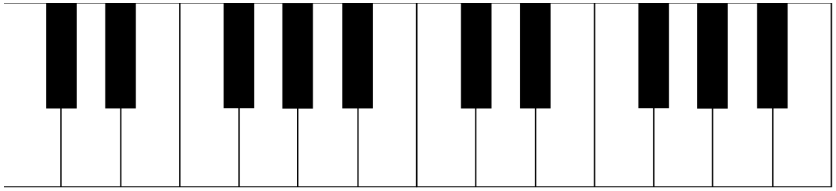


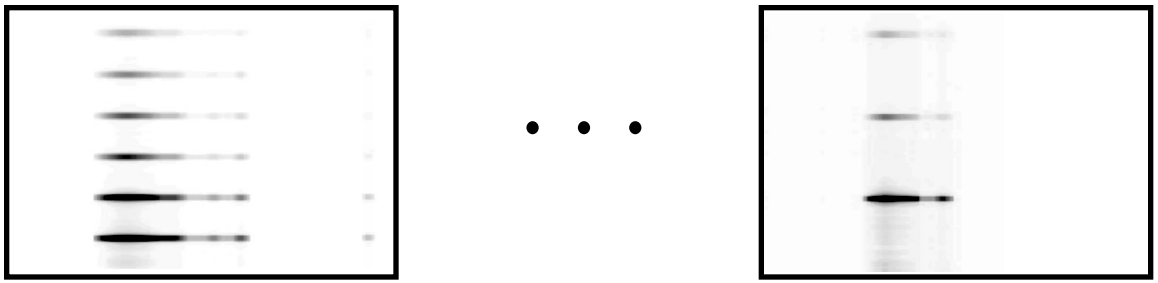




Berkeley
N L P

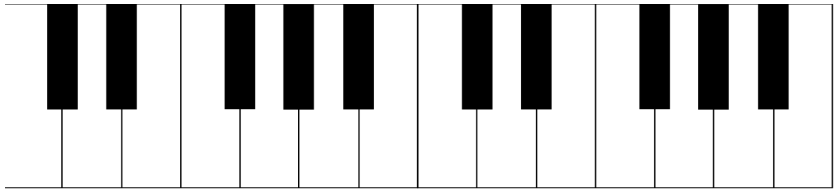
# Polyphony

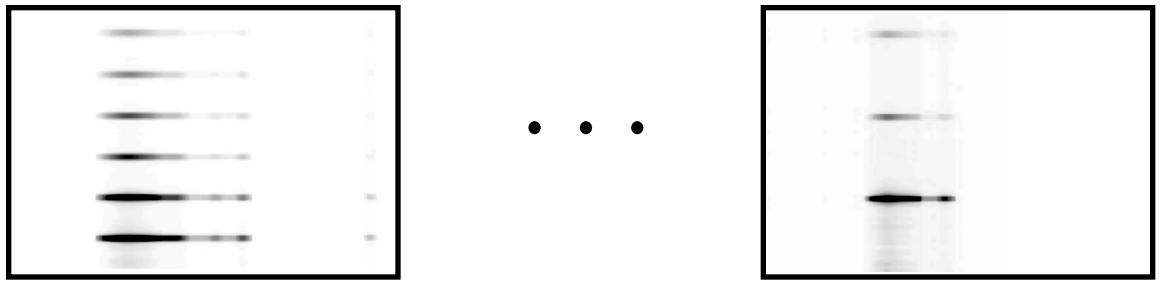




Berkeley
N L P

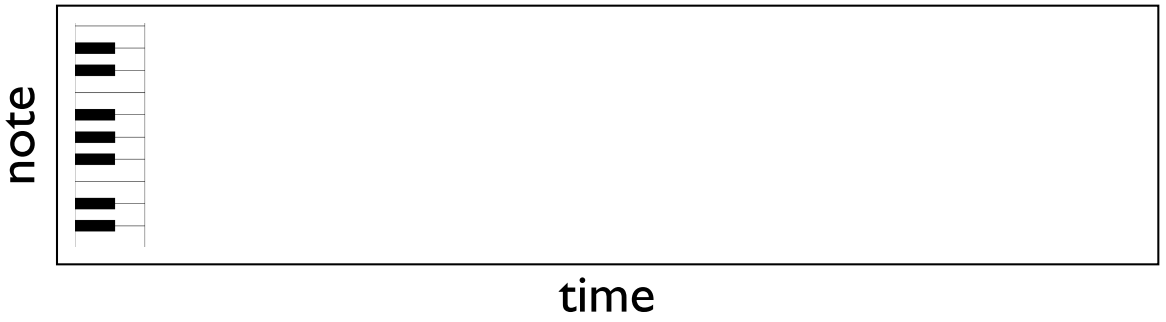
# Polyphony





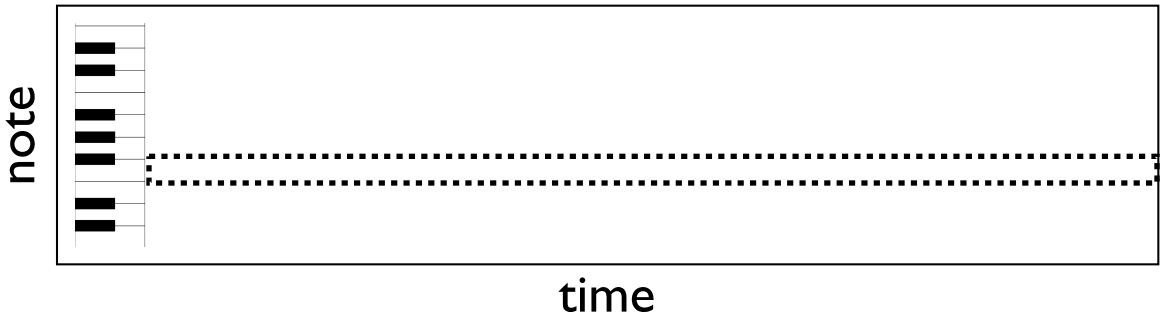
Berkeley
N L P

# Generative Model



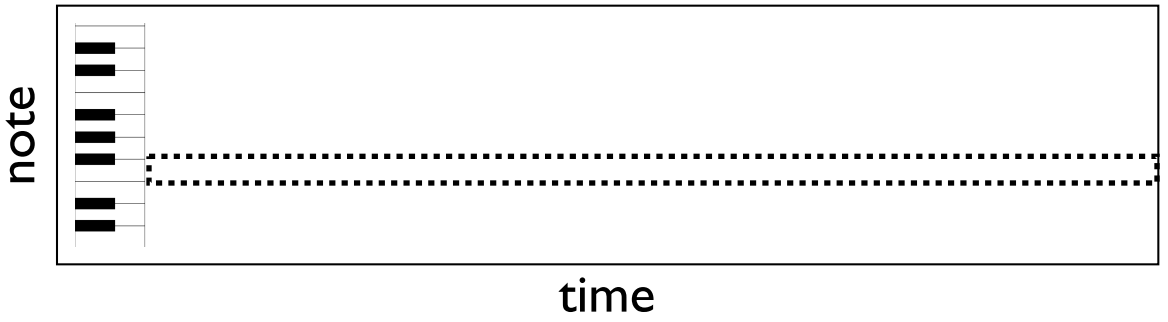
Berkeley
N L P

# Generative Model

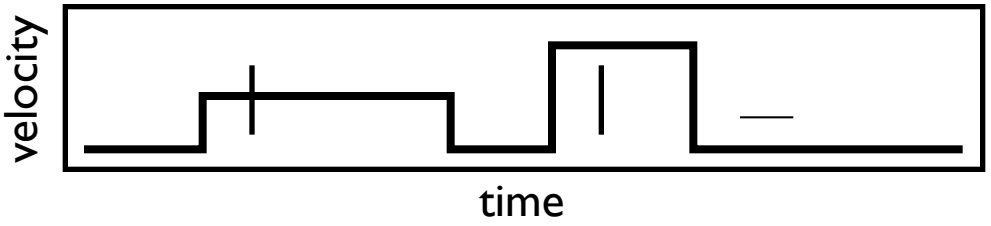


Berkeley
N L P

# Generative Model

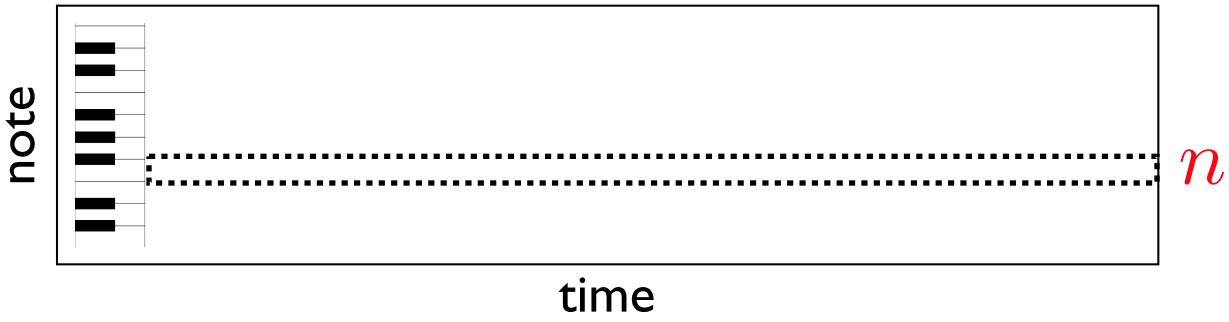


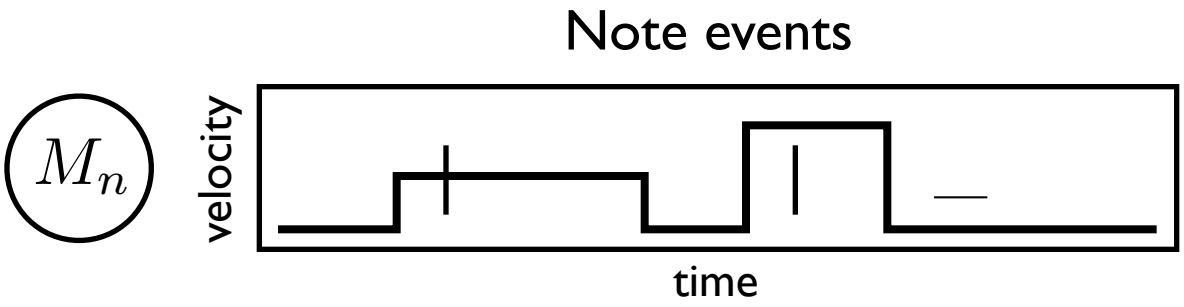

Note events



Berkeley
N L P

# Generative Model

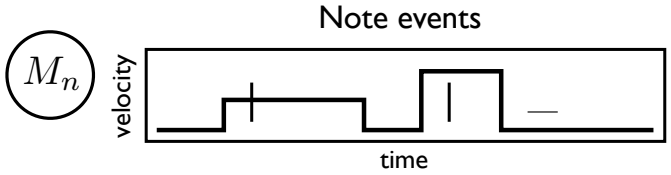


Berkeley

N L P

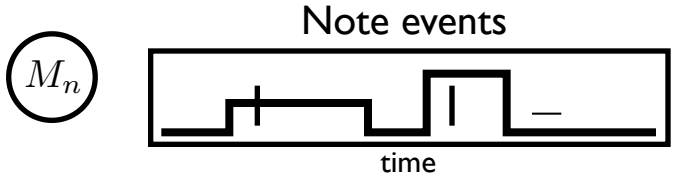
# Generative Model



Berkeley

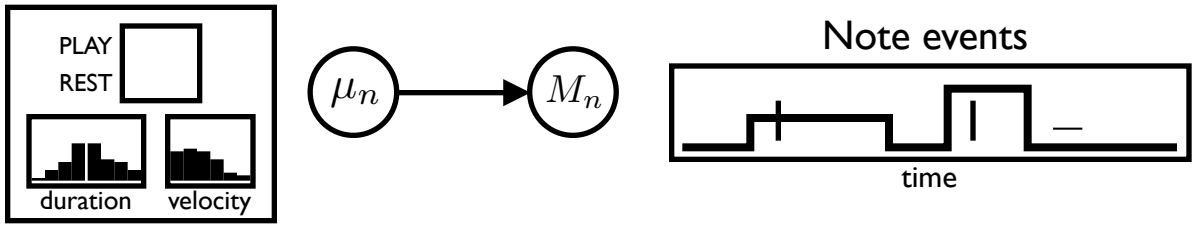
N L P

# Generative Model



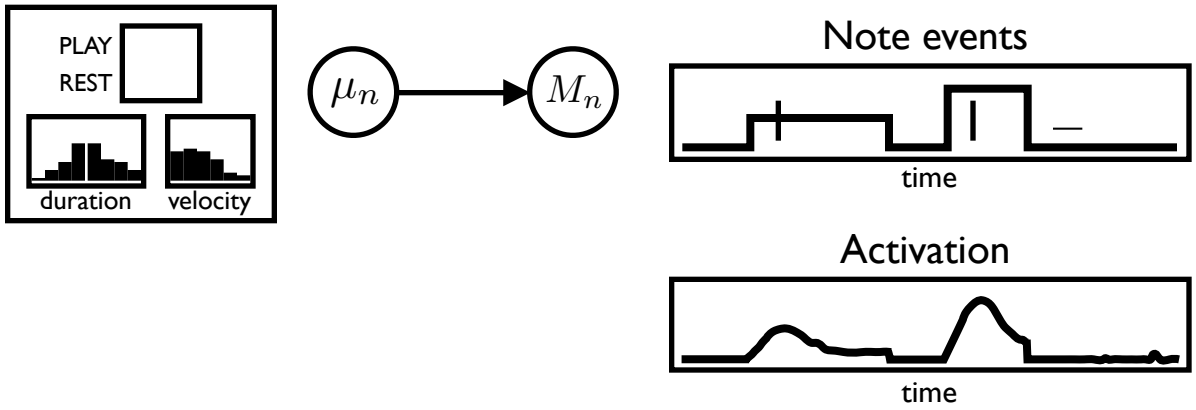
Berkeley
N L P

# Generative Model



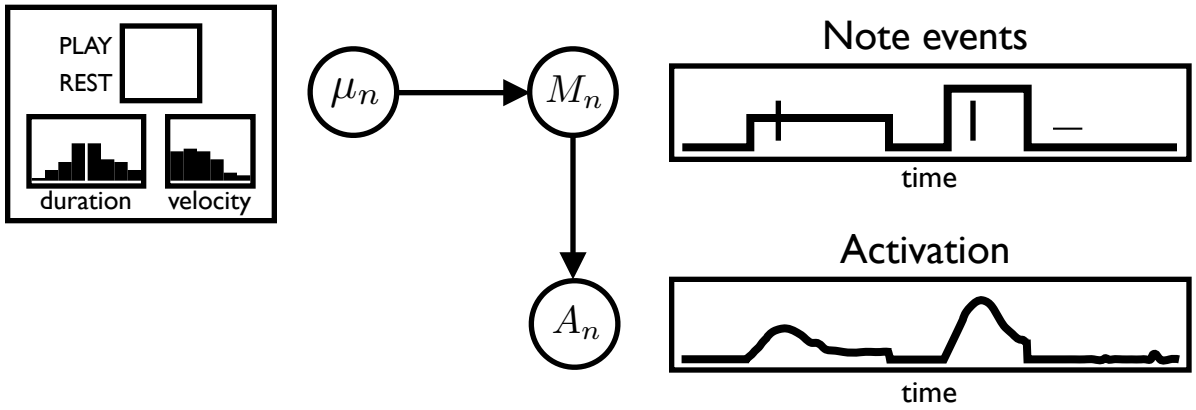
Berkeley
N L P

# Generative Model



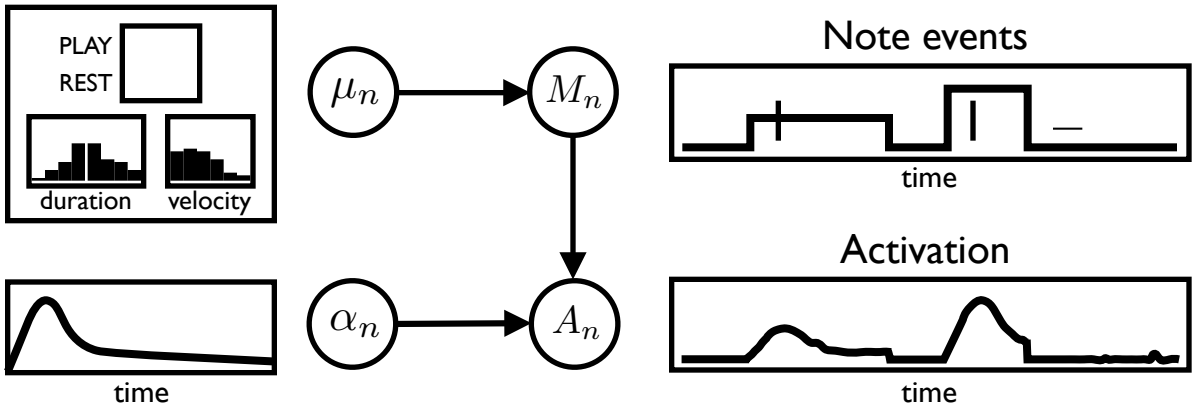
Berkeley
N L P

# Generative Model

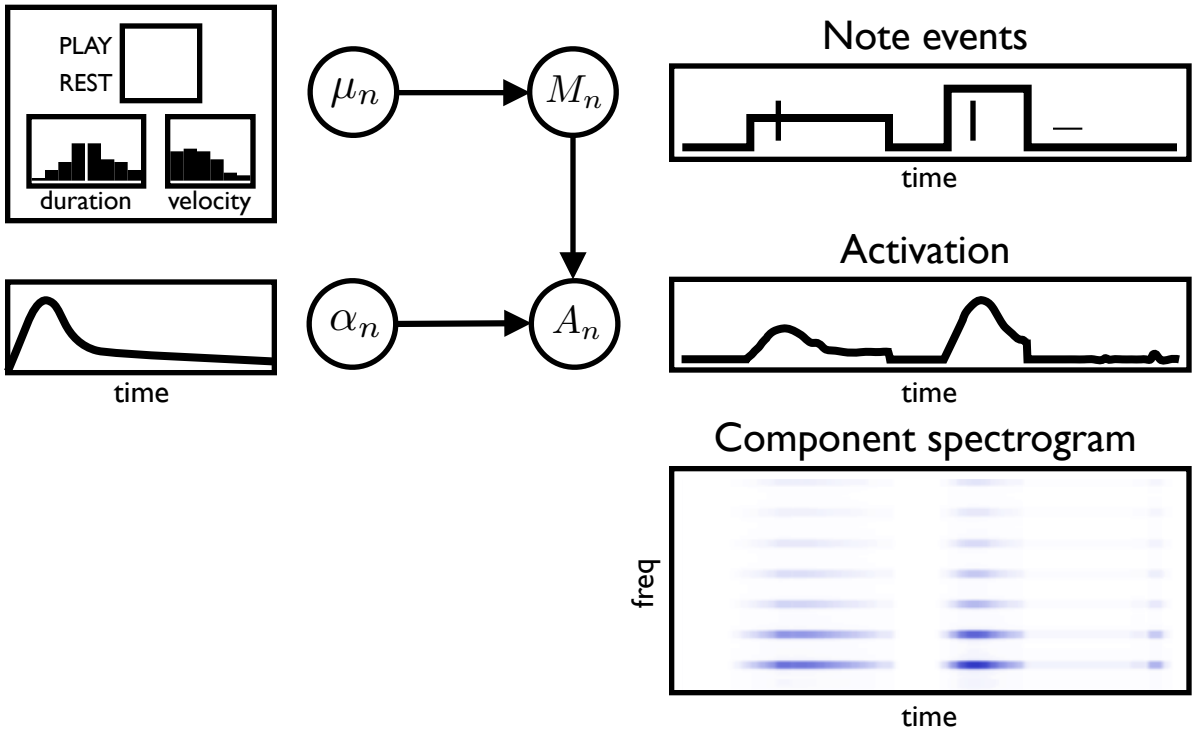


Berkeley
N L P

# Generative Model

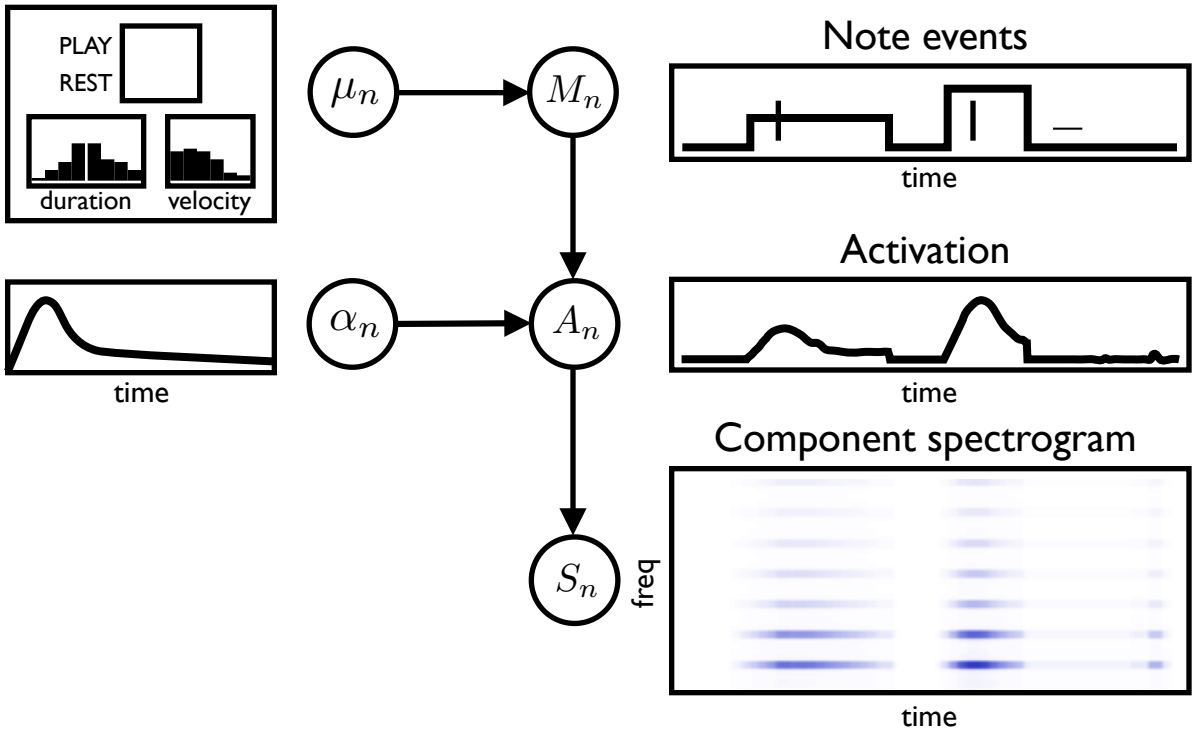


# Generative Model

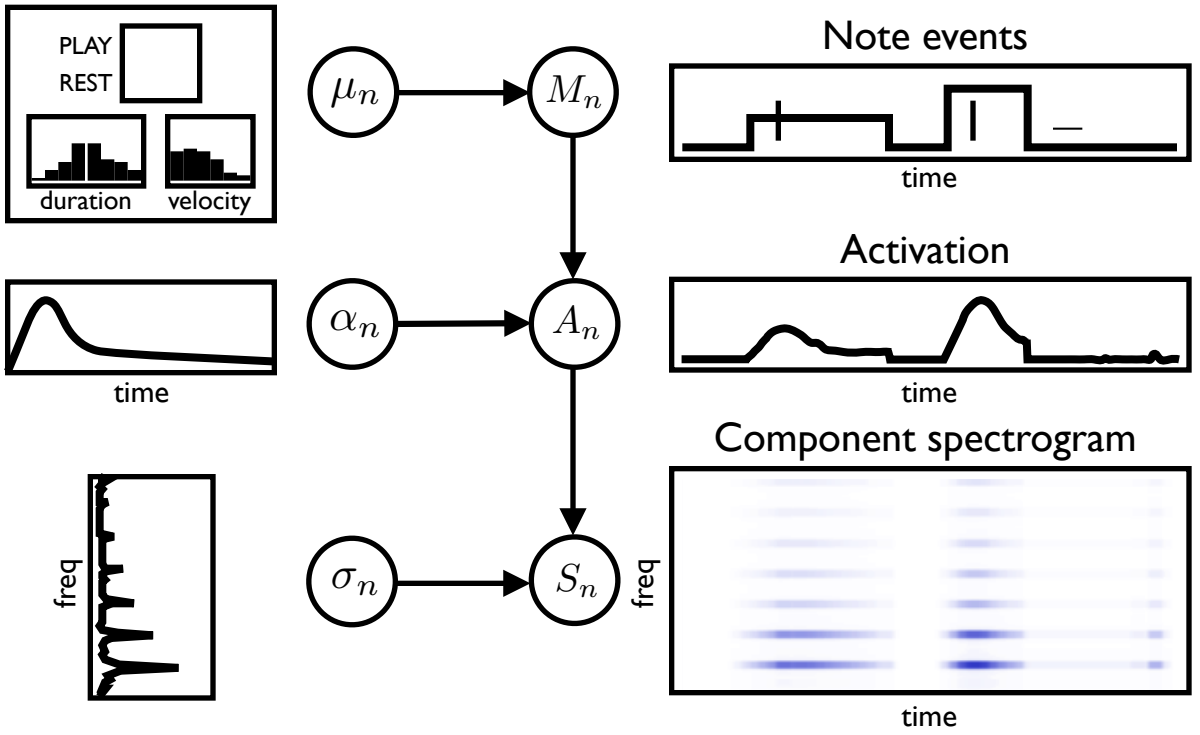


Berkeley
N L P

# Generative Model

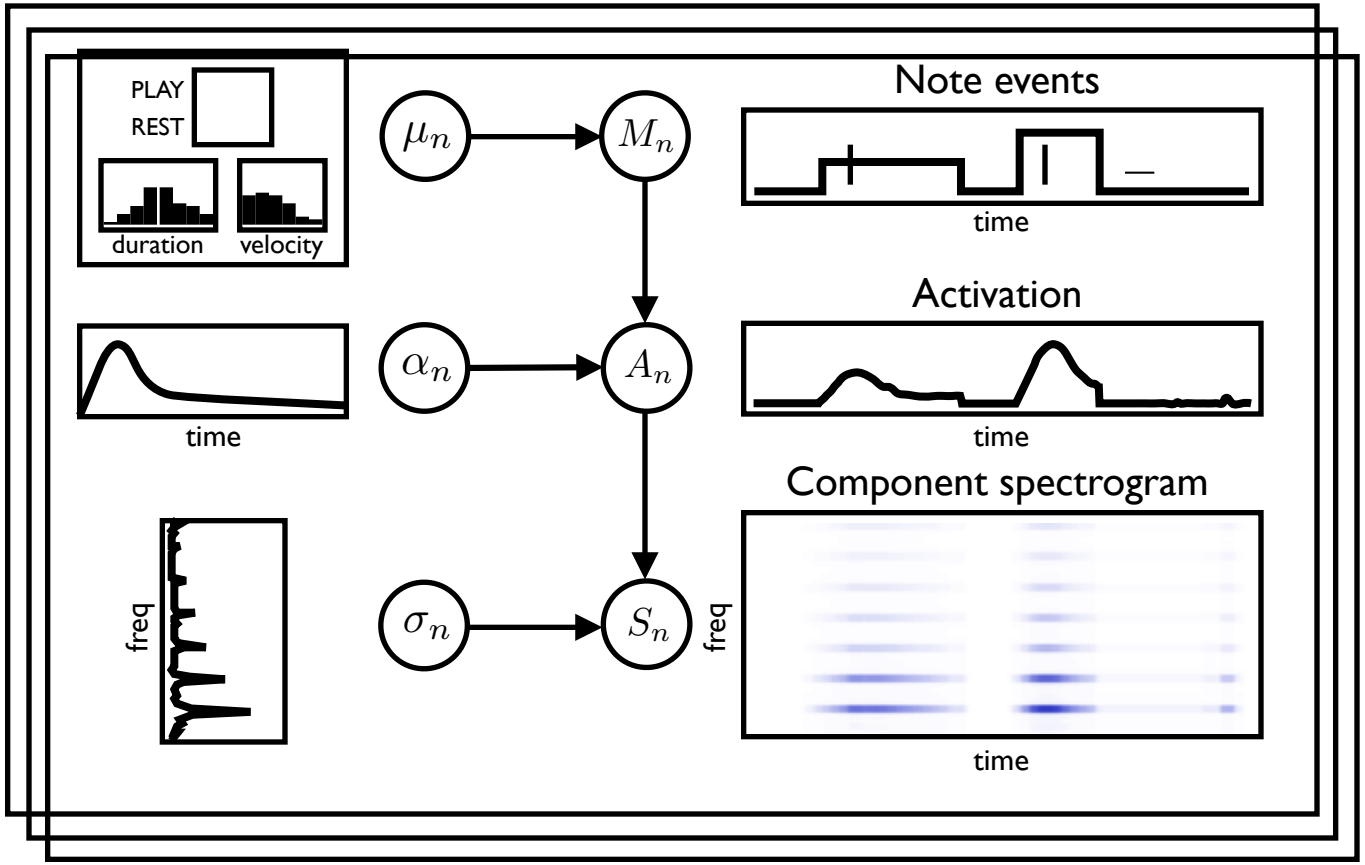


# Generative Model



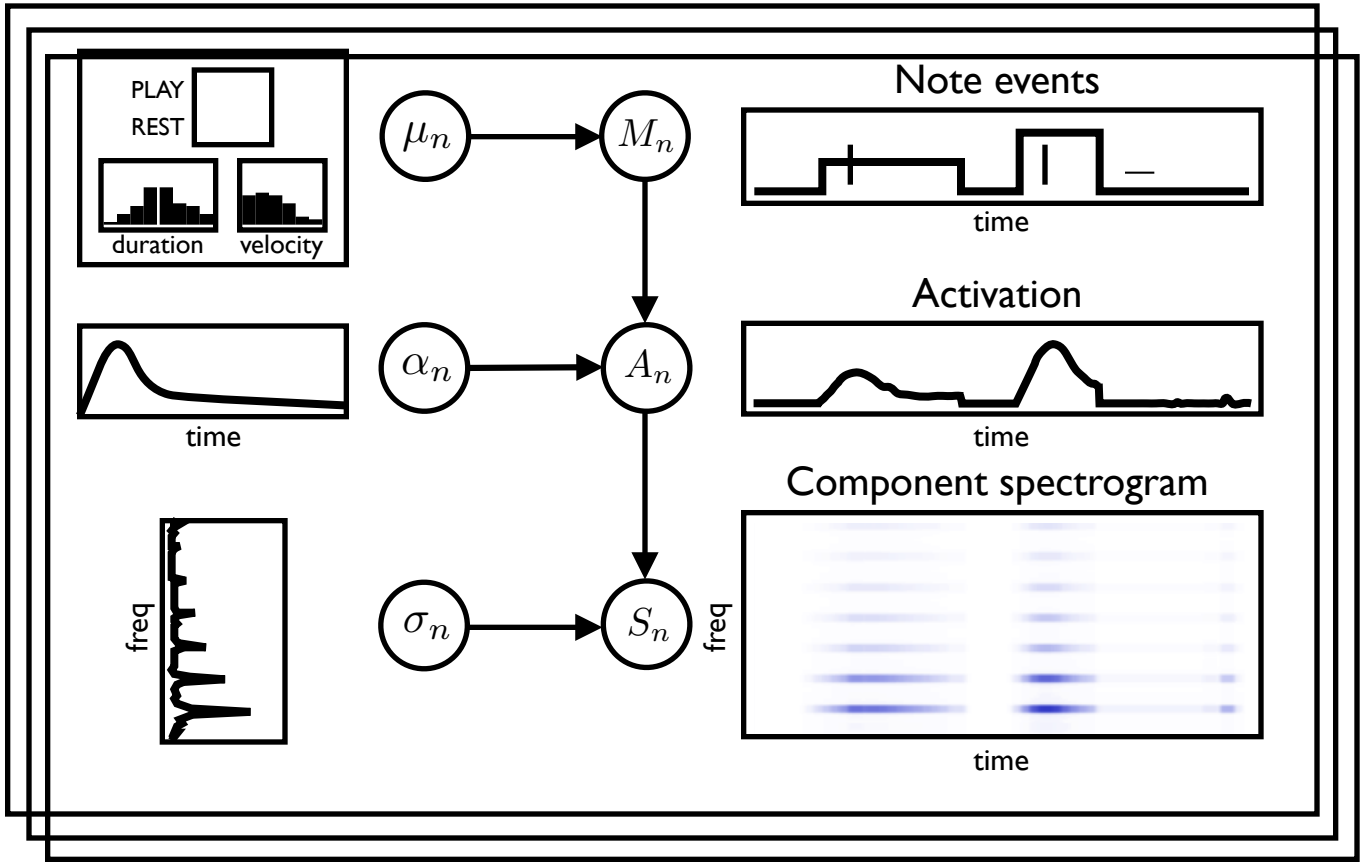
Berkeley
N L P

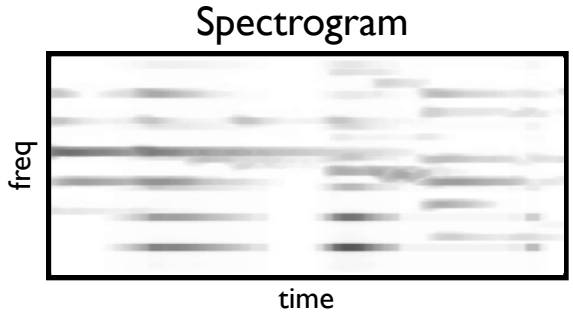
# Generative Model



Berkeley
N L P

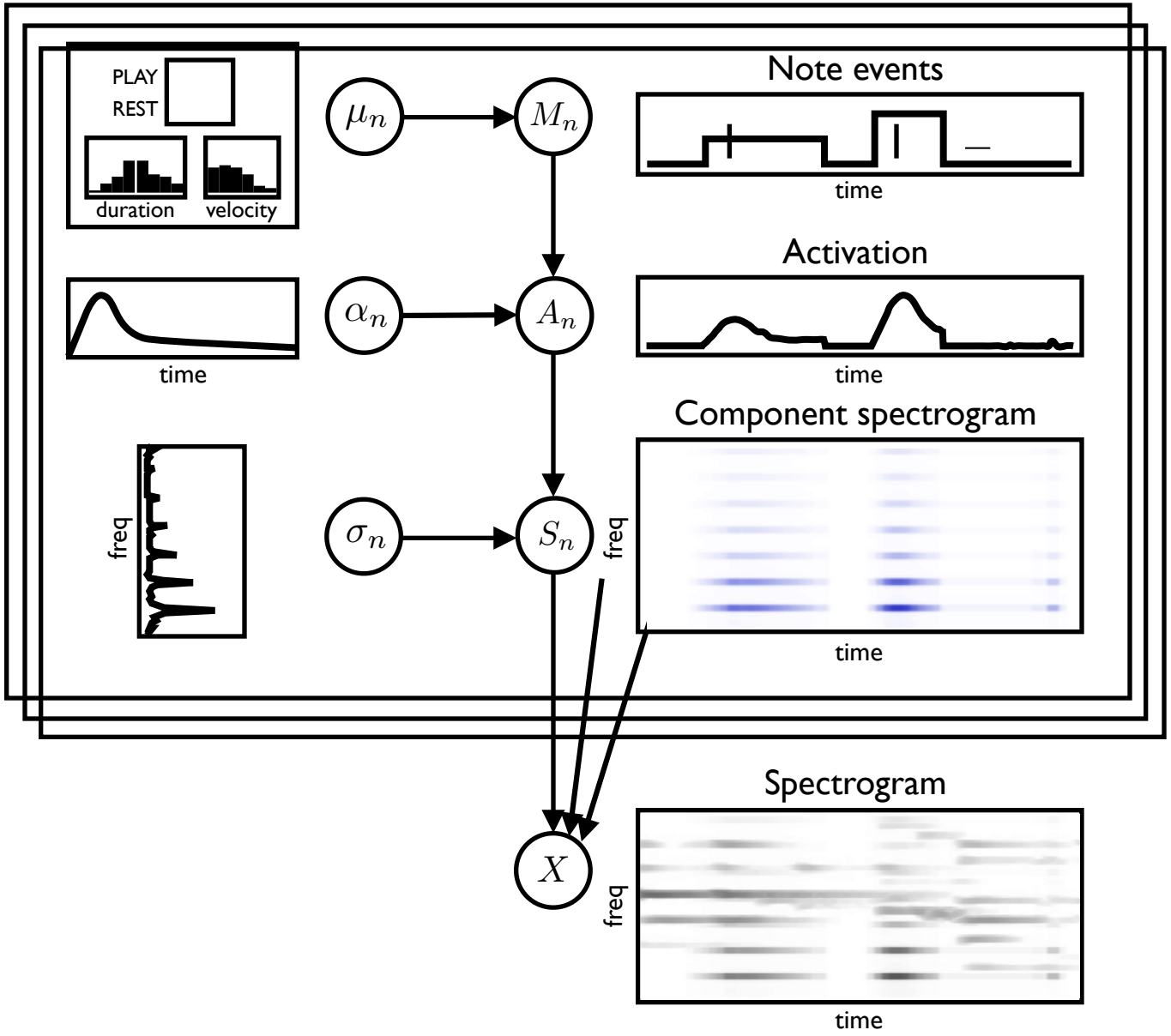
# Generative Model



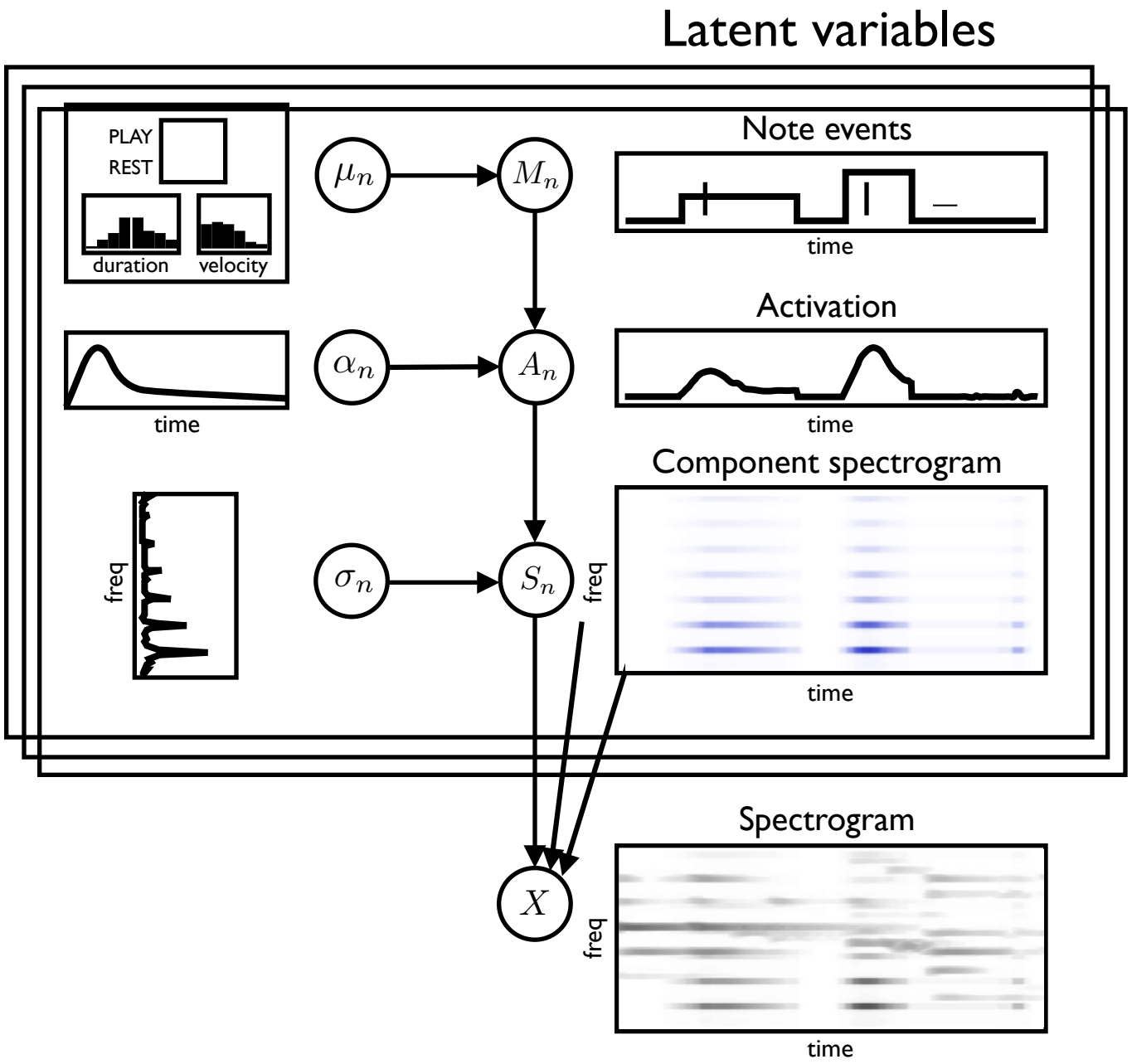
Berkeley
N L P

# Generative Model



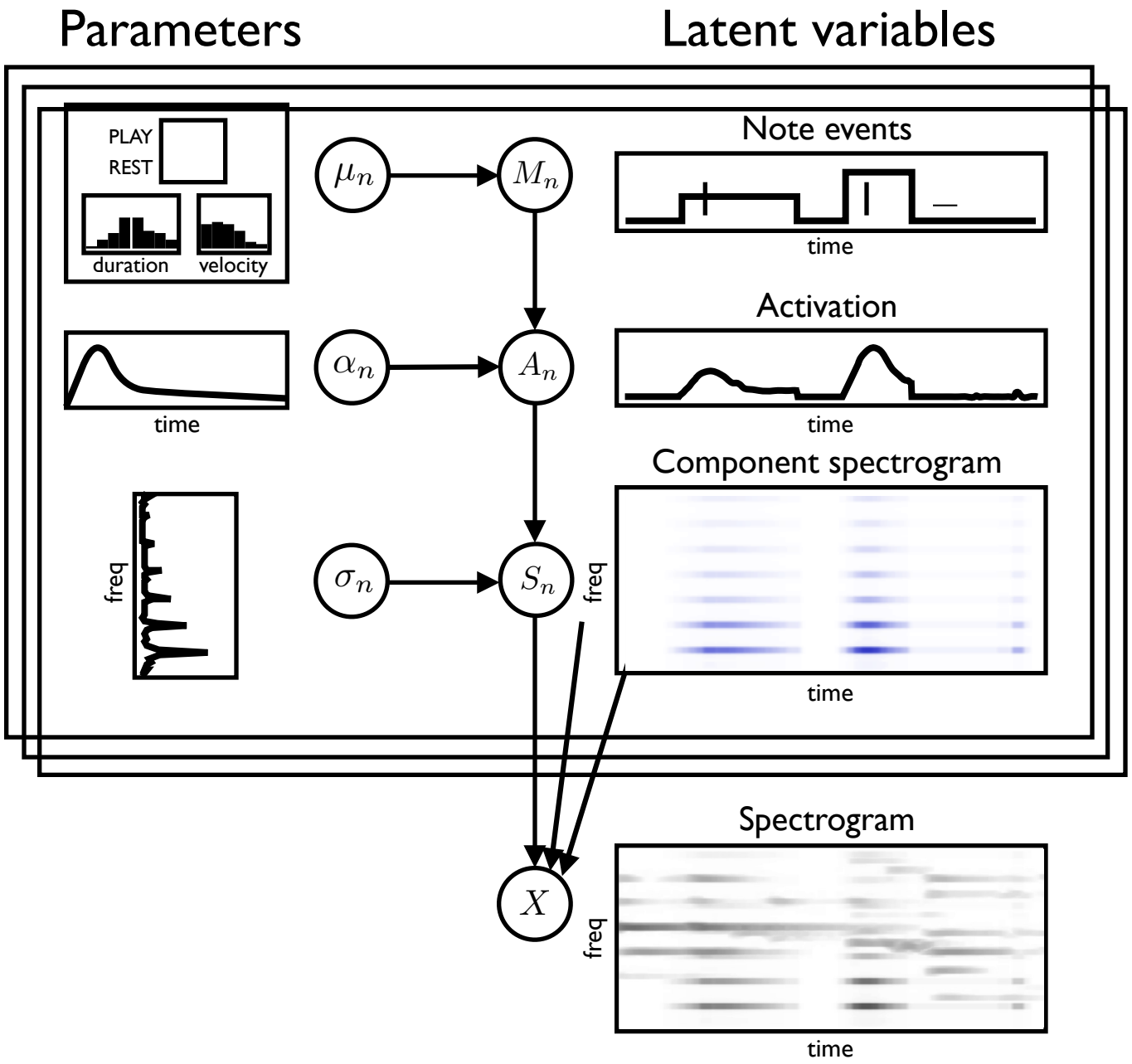
Berkeley NLP

# Generative Model



Berkeley
N L P

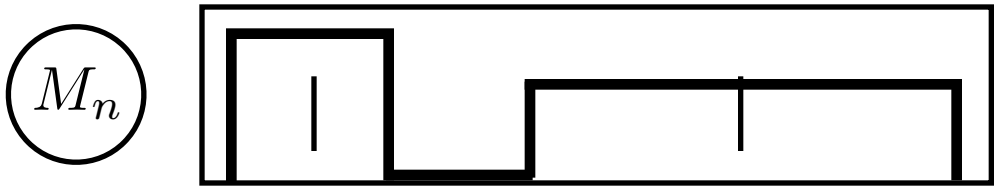
# Generative Model




# Note Event Model

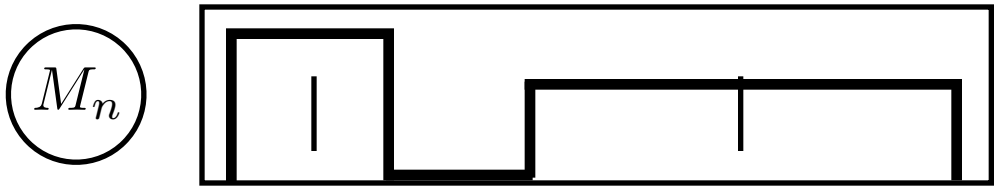
Berkeley
N L P

# Note Event Model



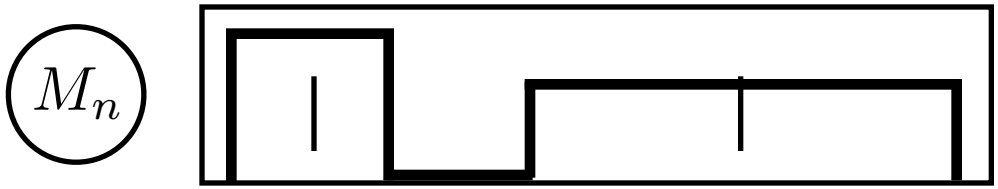
Berkeley
N L P

# Note Event Model



Berkeley
N L P

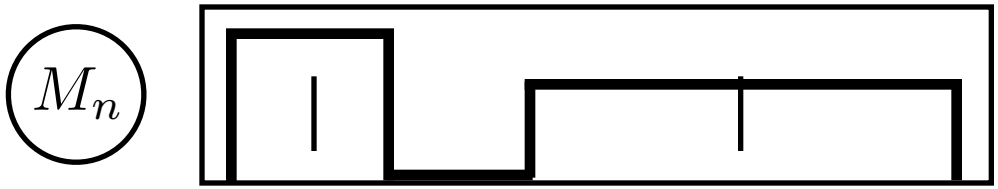
# Note Event Model




Event type

Berkeley NLP

# Note Event Model

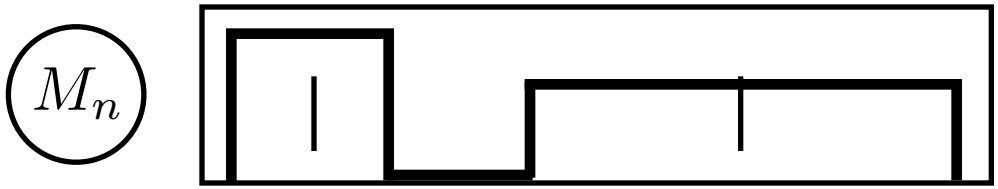



Event type

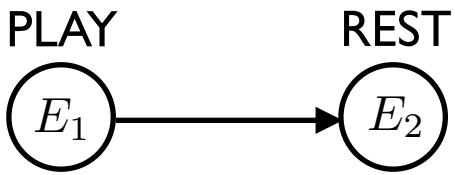
PLAY

Berkeley
N L P

# Note Event Model

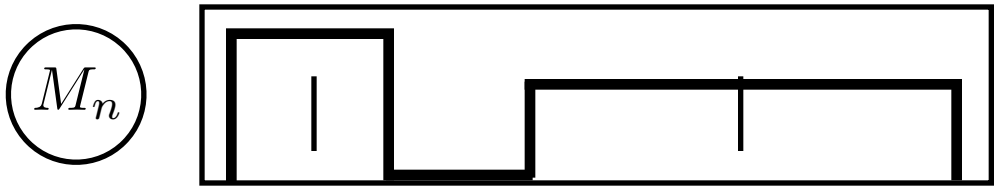



Event type

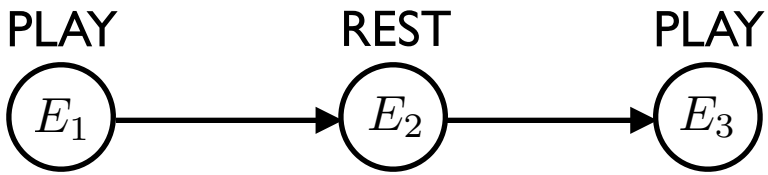


Berkeley NLP

# Note Event Model

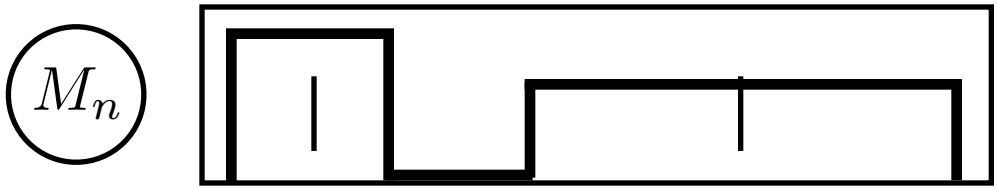



Event type

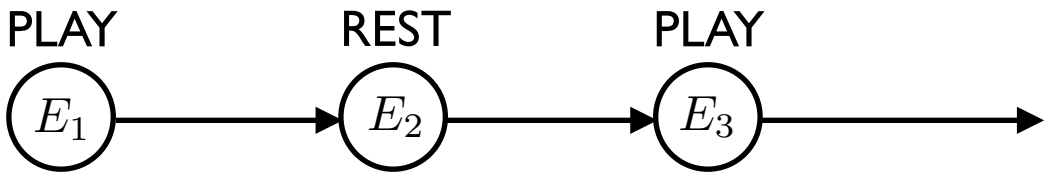


Berkeley
N L P

# Note Event Model

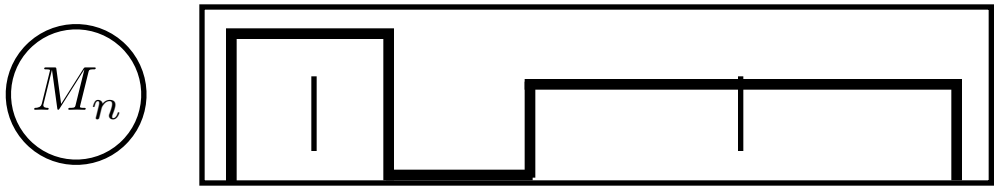


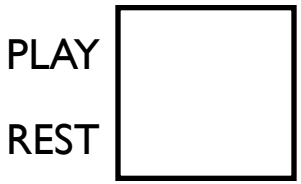

Event type



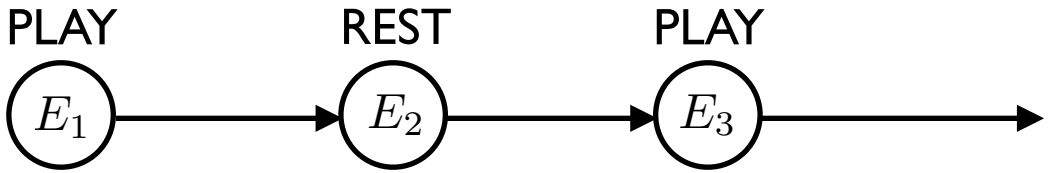
Berkeley
N L P

# Note Event Model

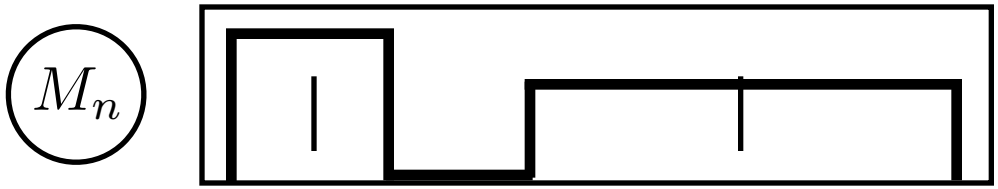


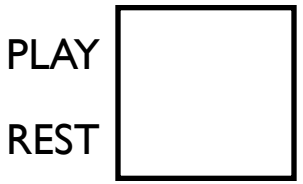



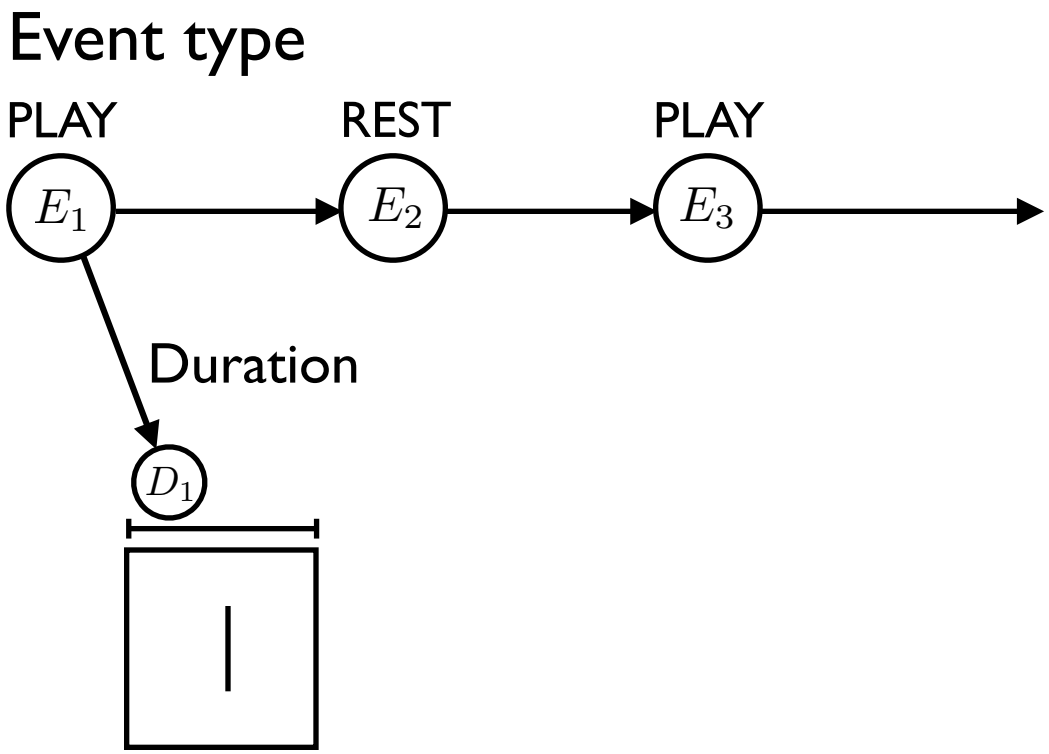

Event type



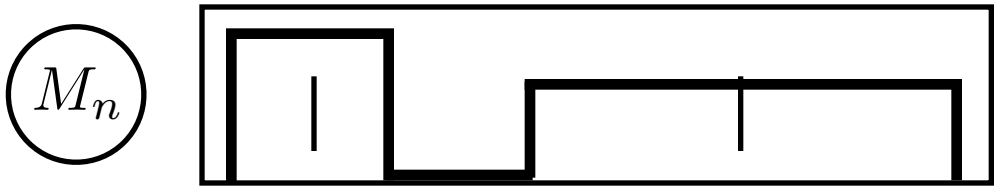
# Note Event Model

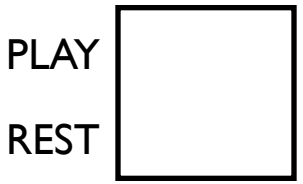


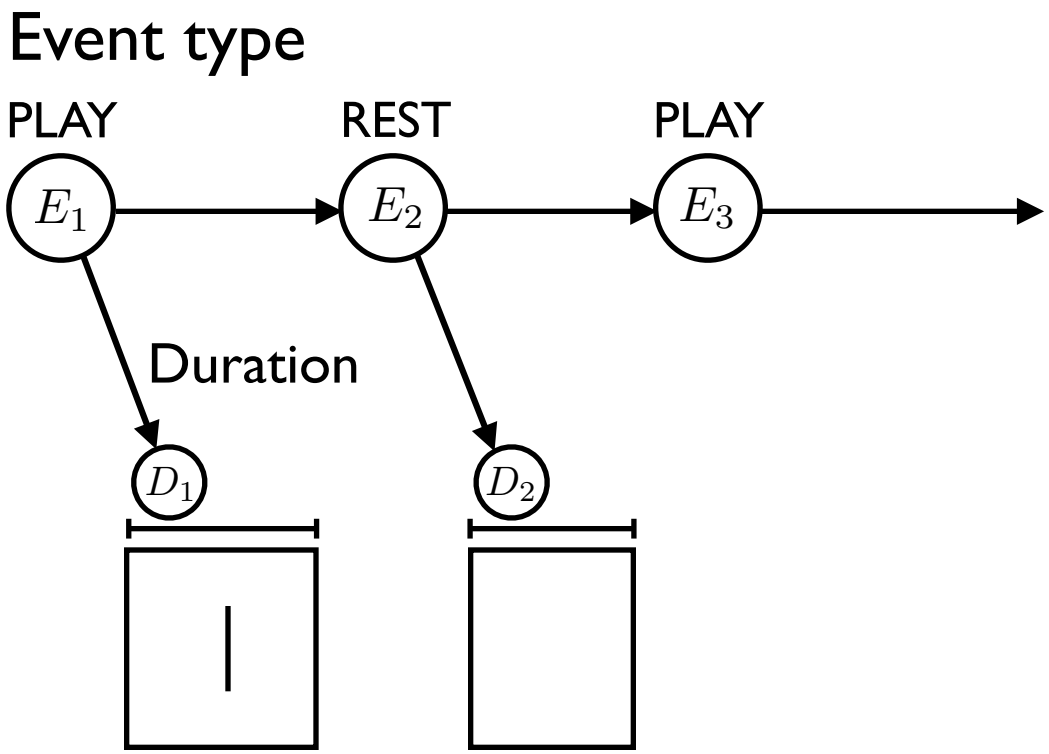



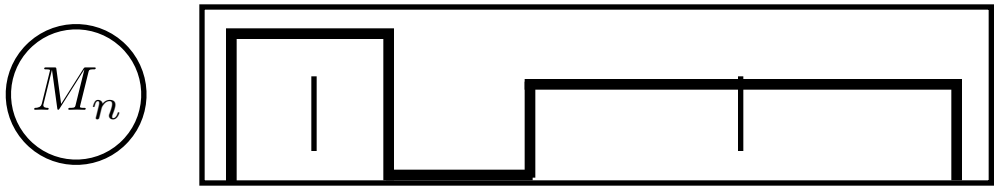
# Note Event Model

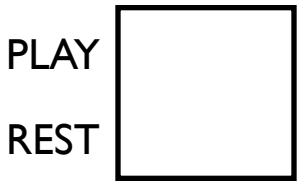


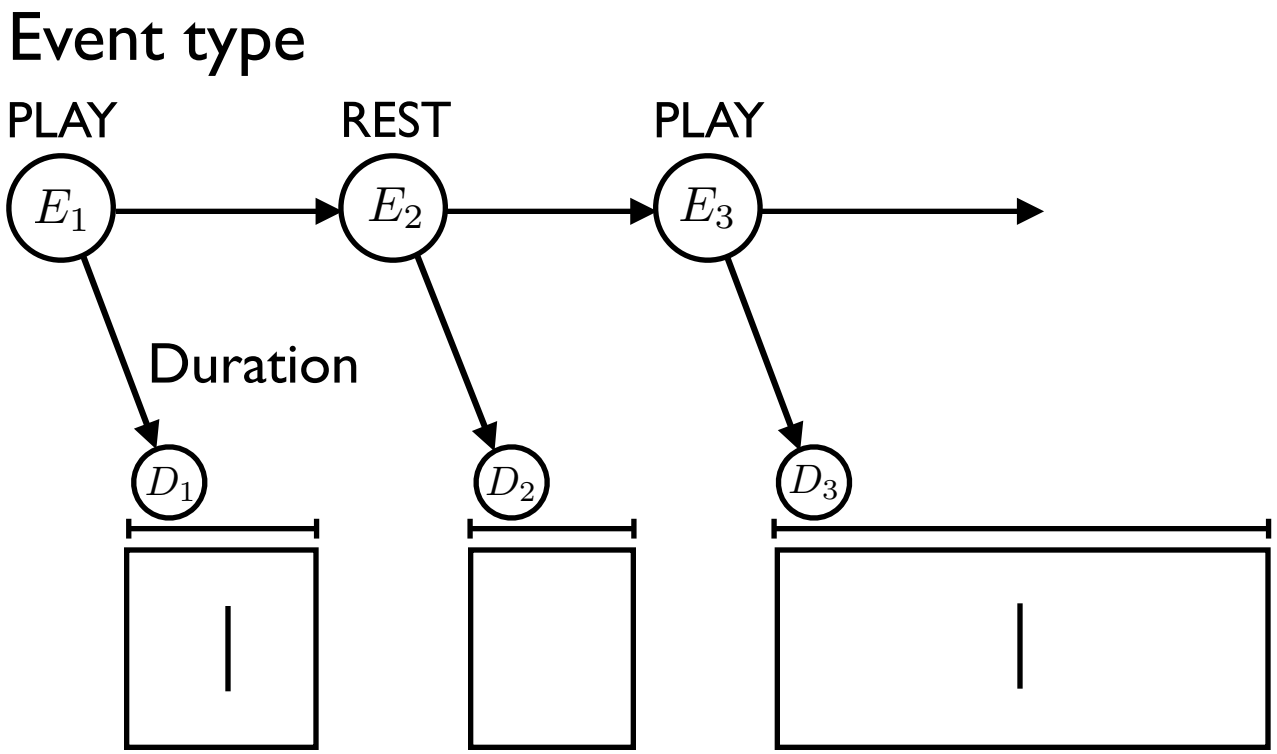



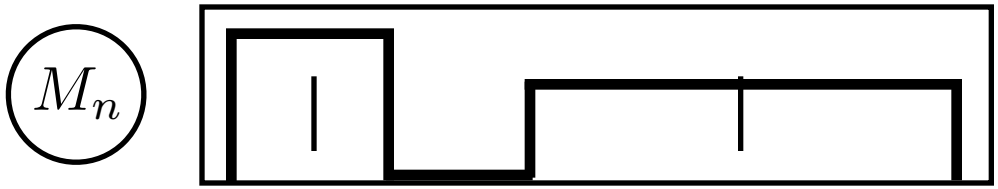
# Note Event Model

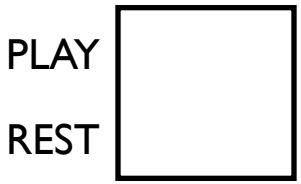


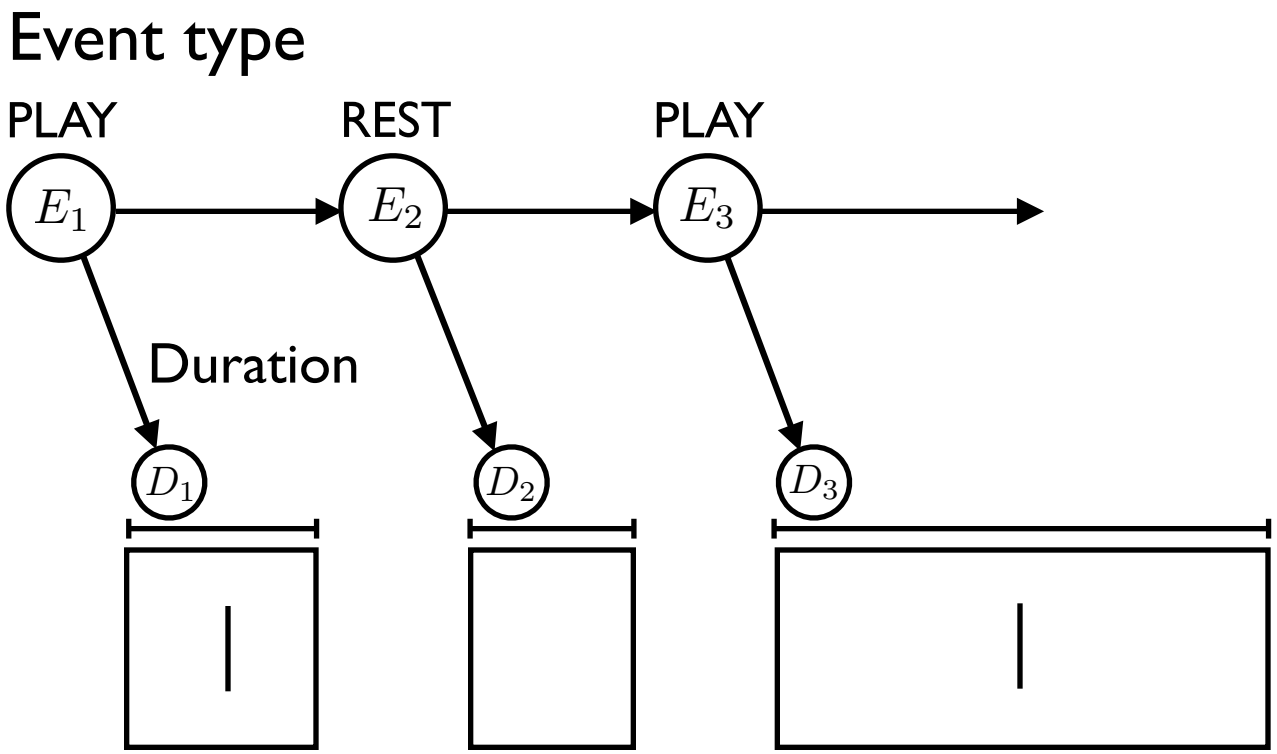



# Note Event Model

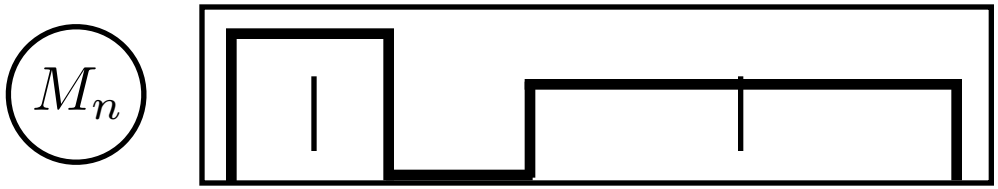


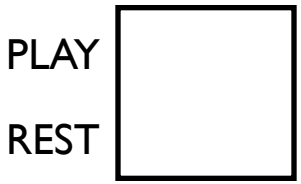





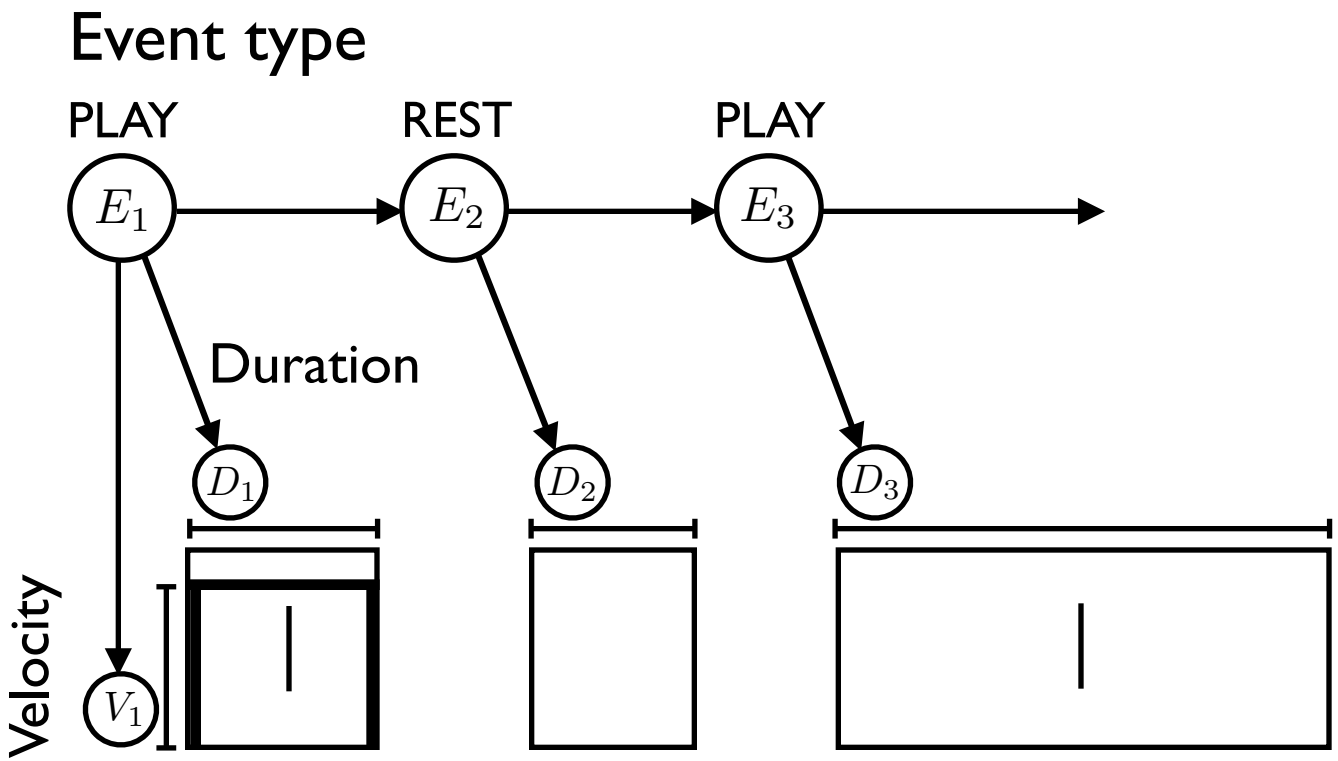


# Note Event Model

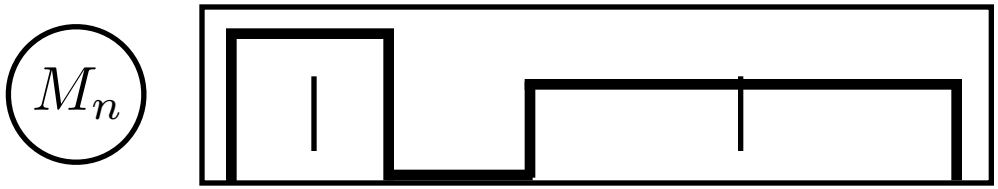


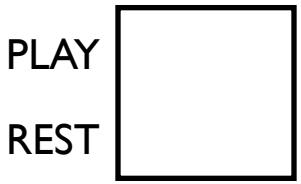





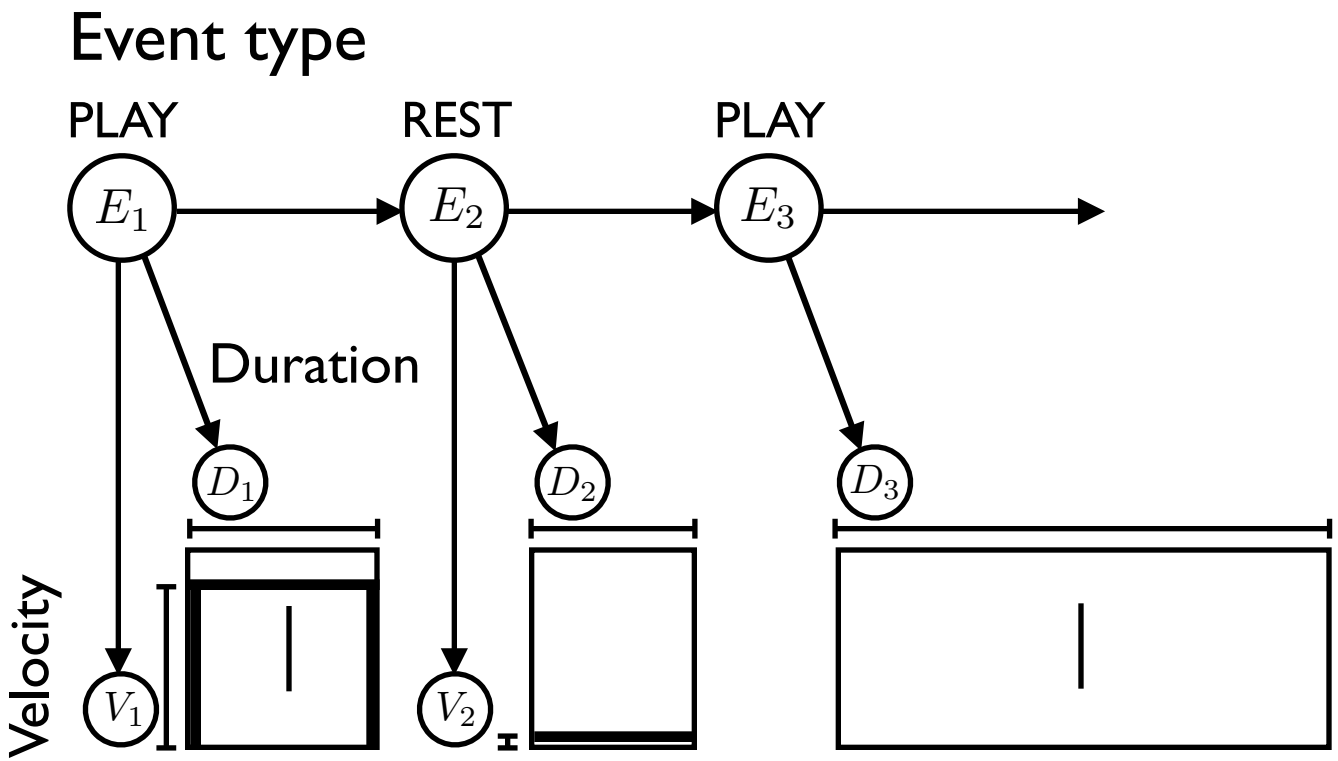


# Note Event Model

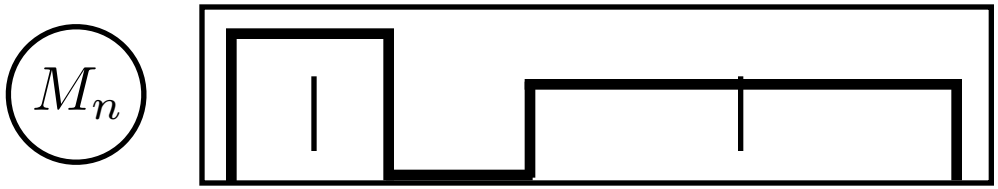


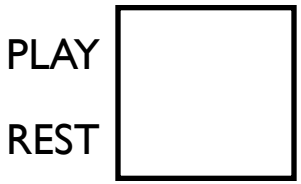





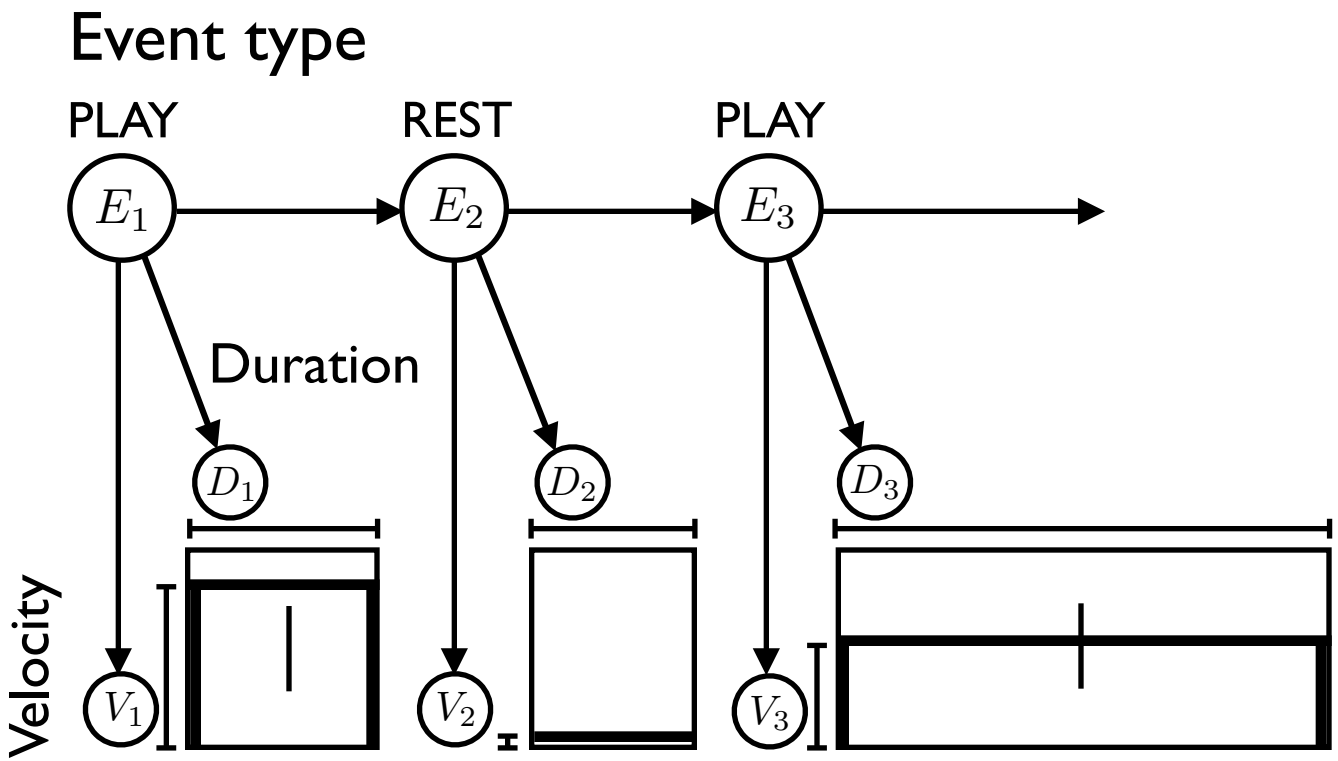


# Note Event Model

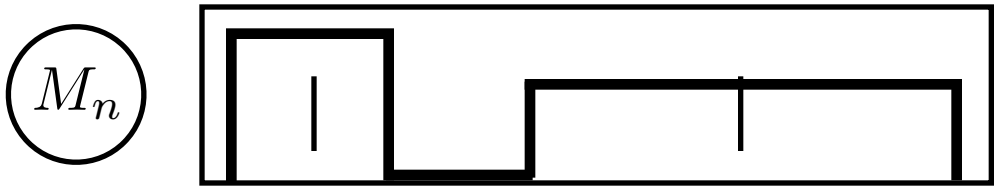


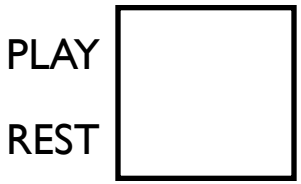



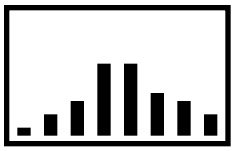


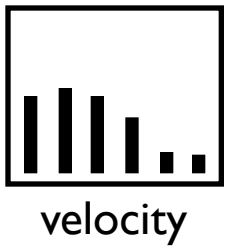


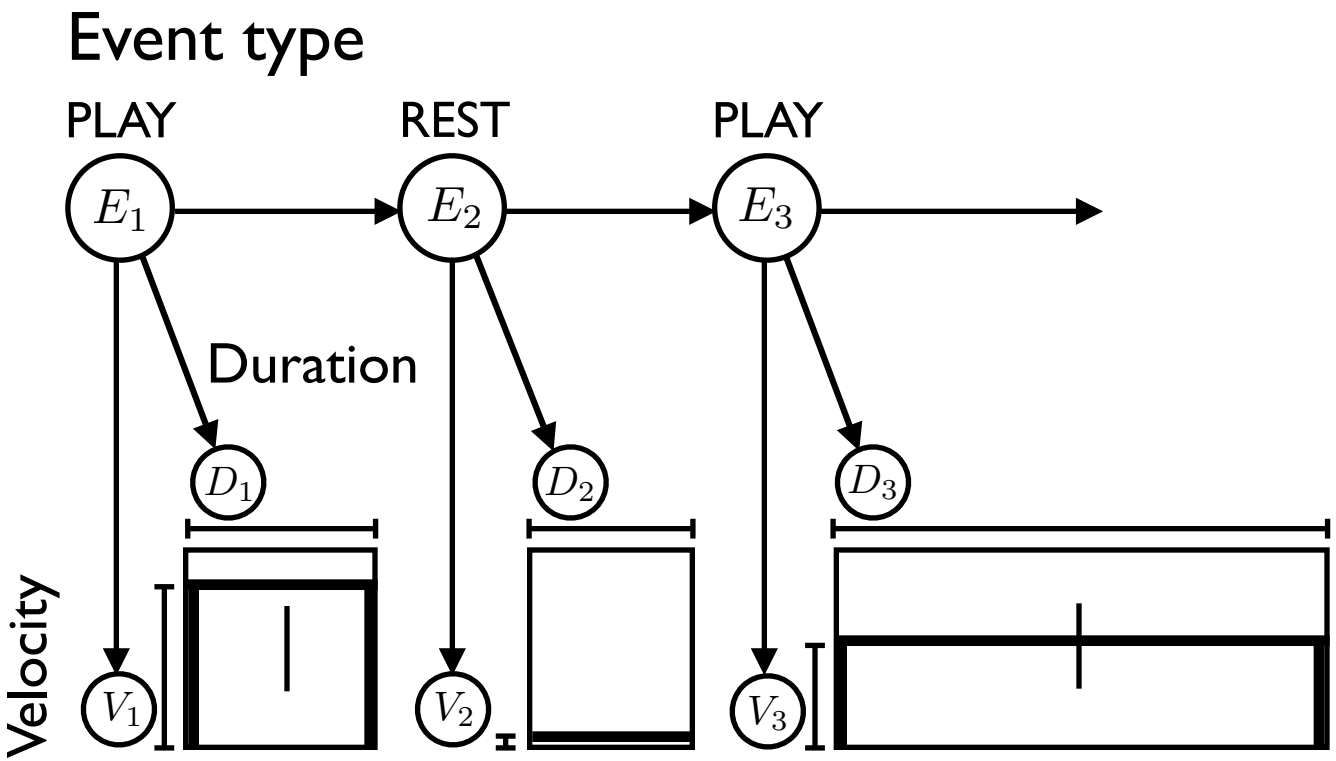
# Note Event Model



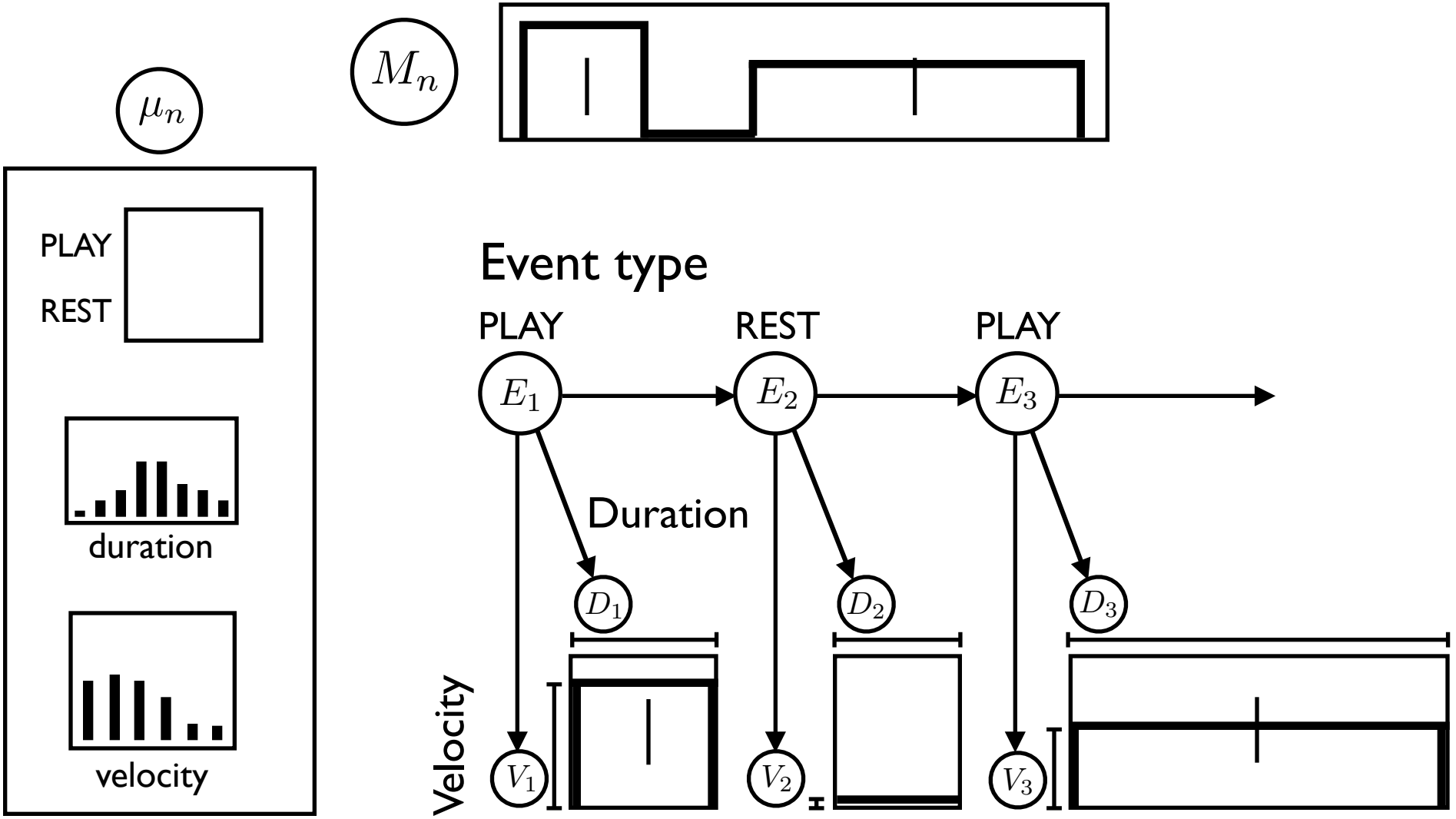



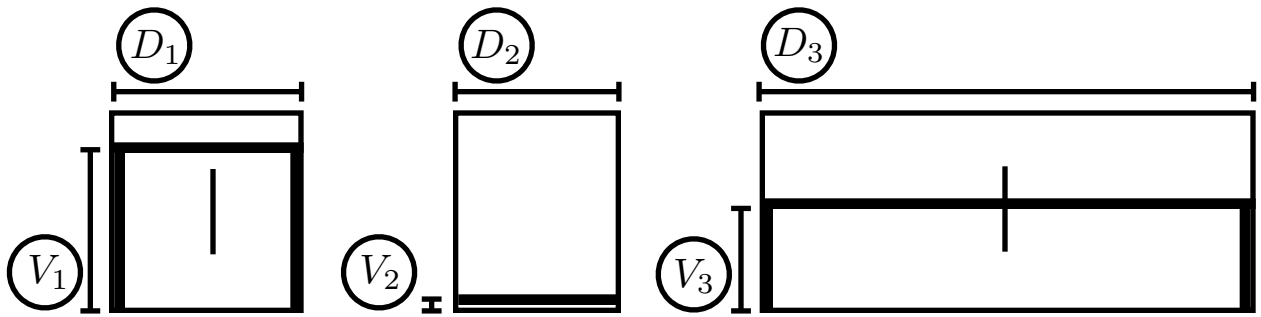
# Note Event Model



Berkeley

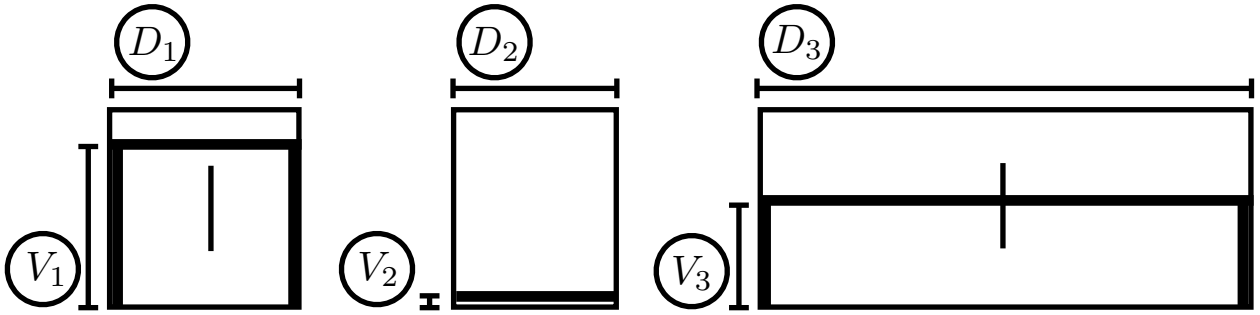
N L P

# Activation Model



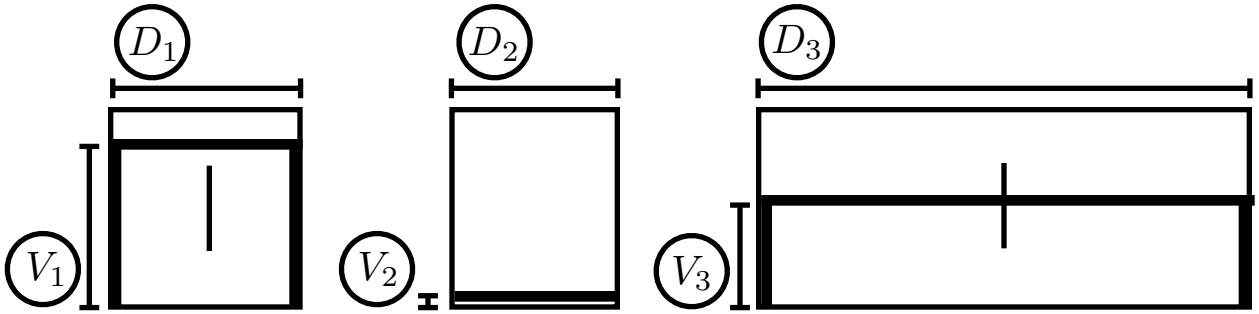
Berkeley
N L P

# Activation Model



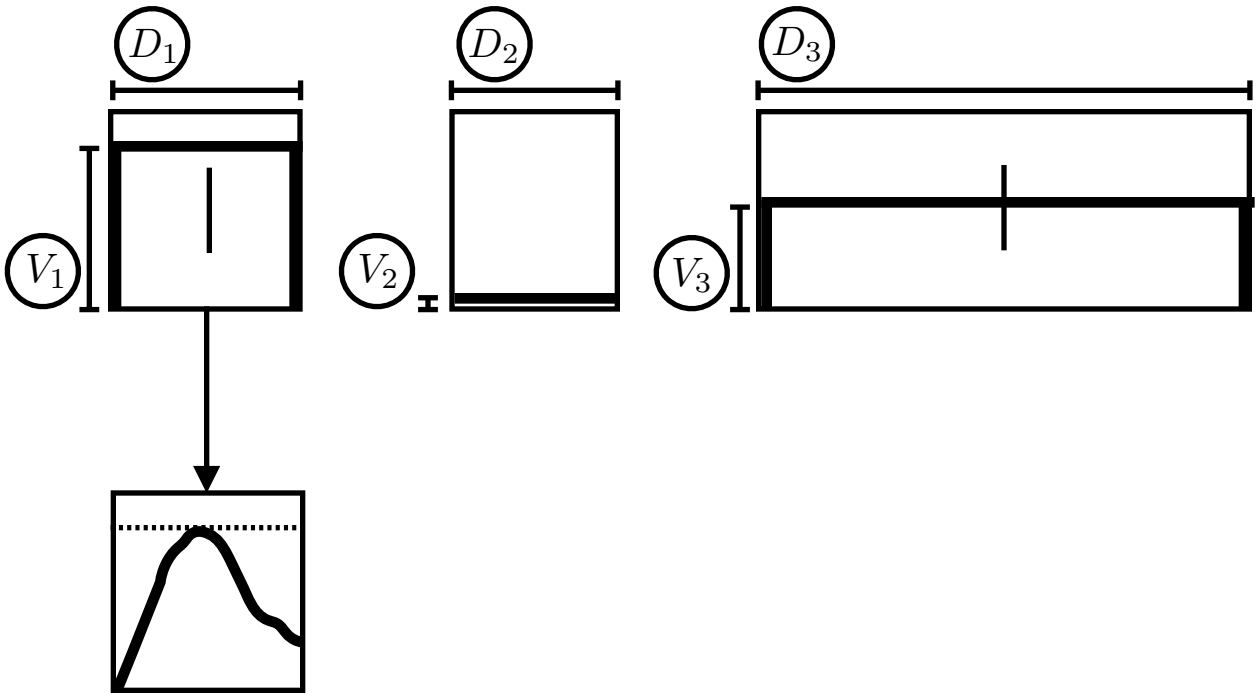
Berkeley
N L P

# Activation Model



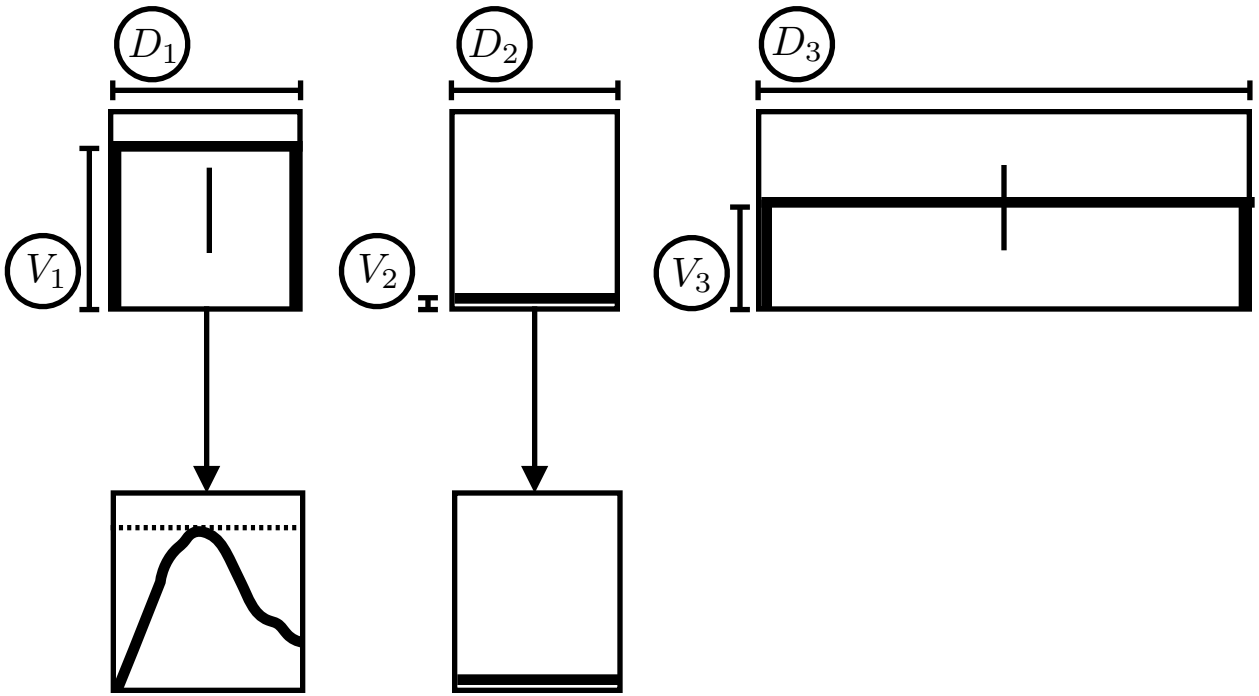
Berkeley
N L P

# Activation Model



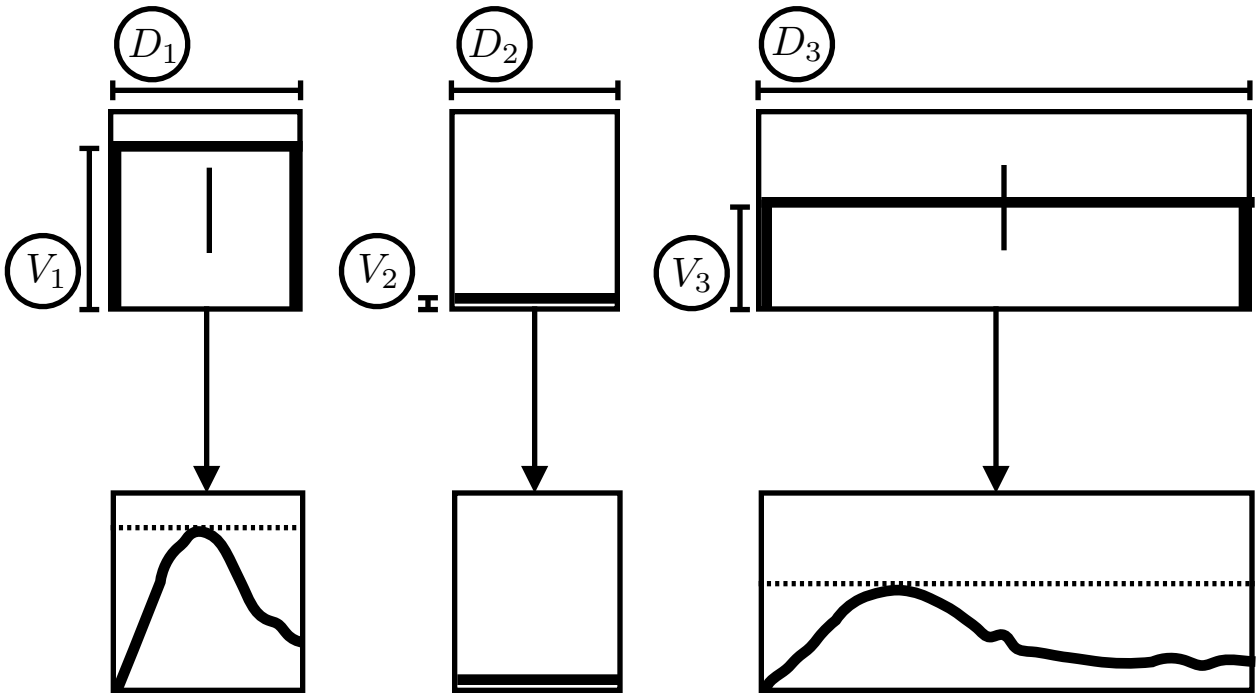
Berkeley N L P

# Activation Model

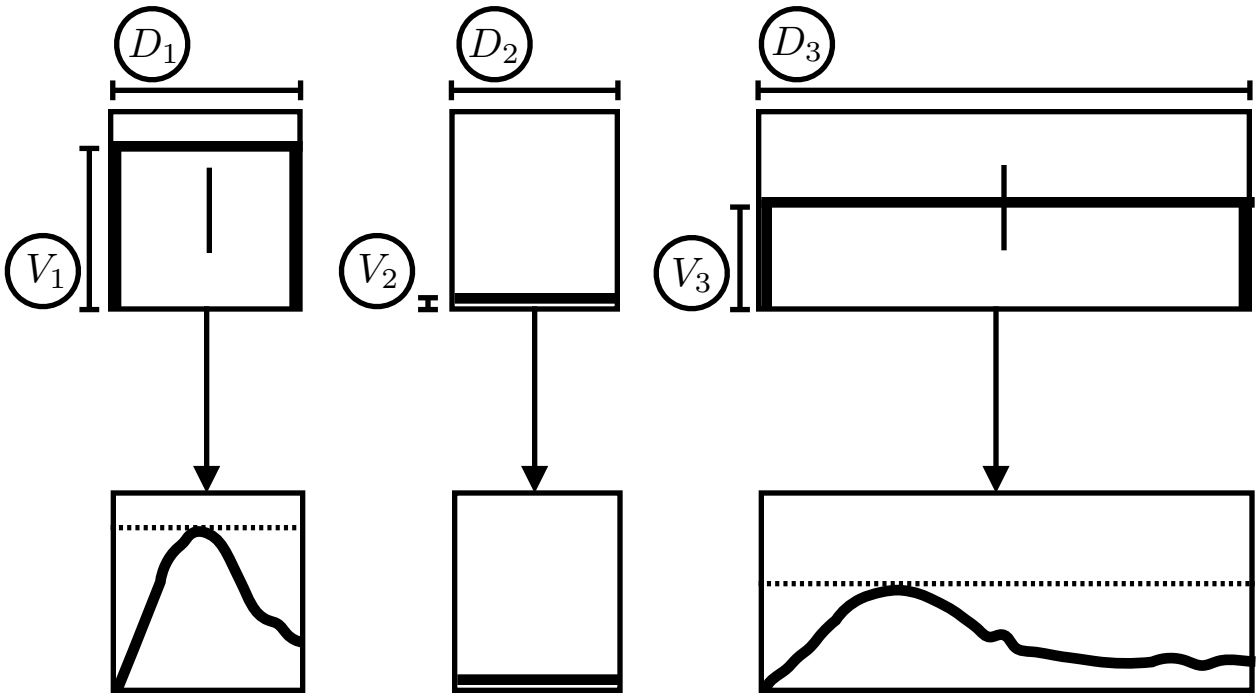


Berkeley NLP

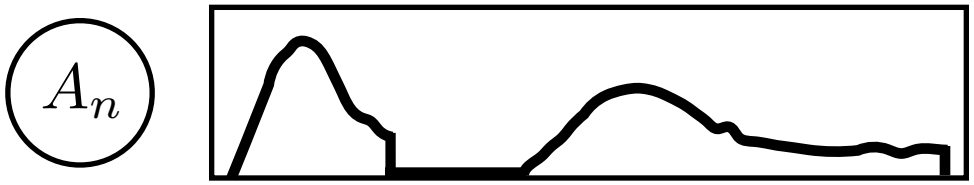
# Activation Model



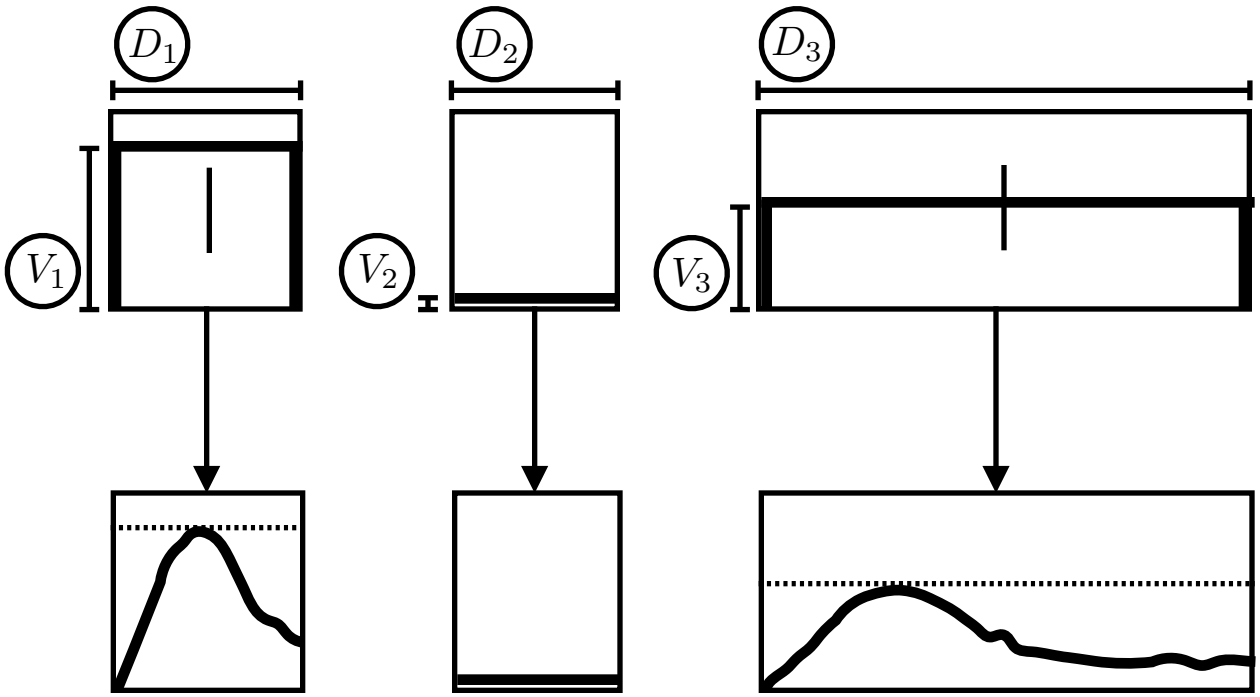
# Activation Model

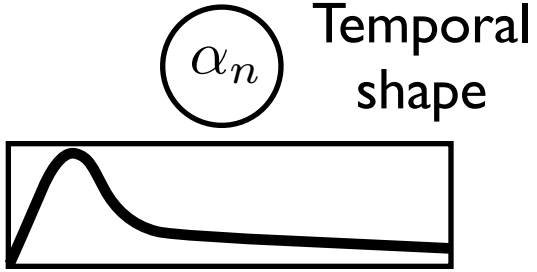



Activation

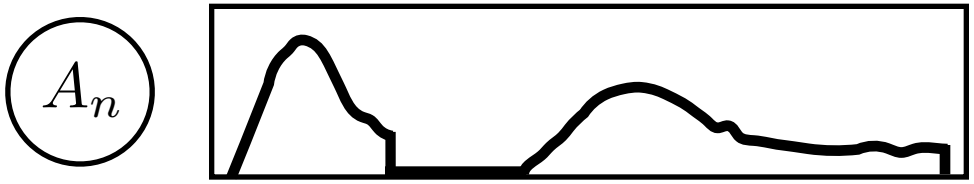


# Activation Model



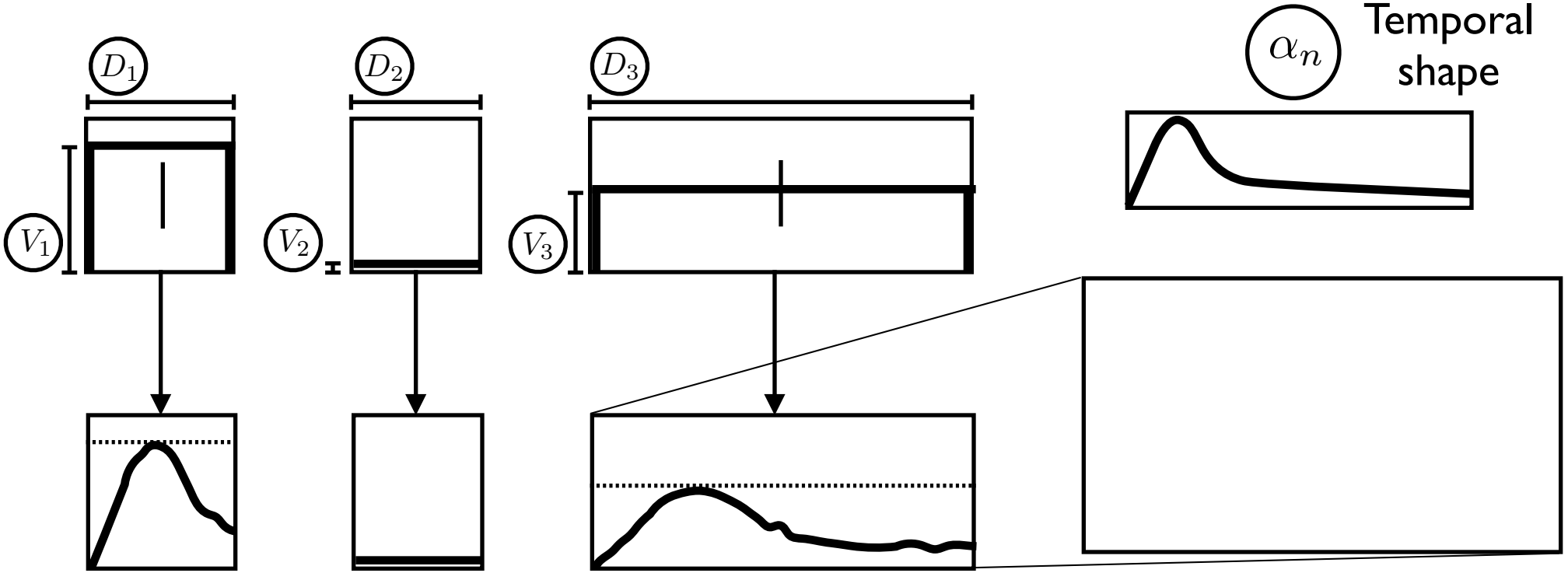




Activation

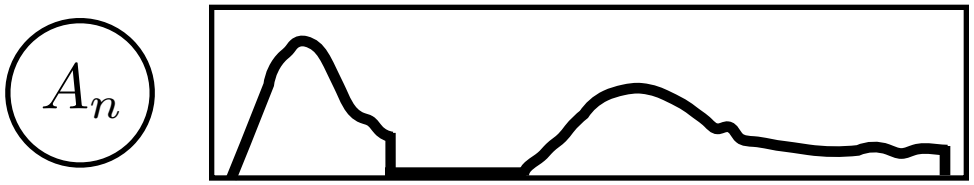


Berkeley N L P

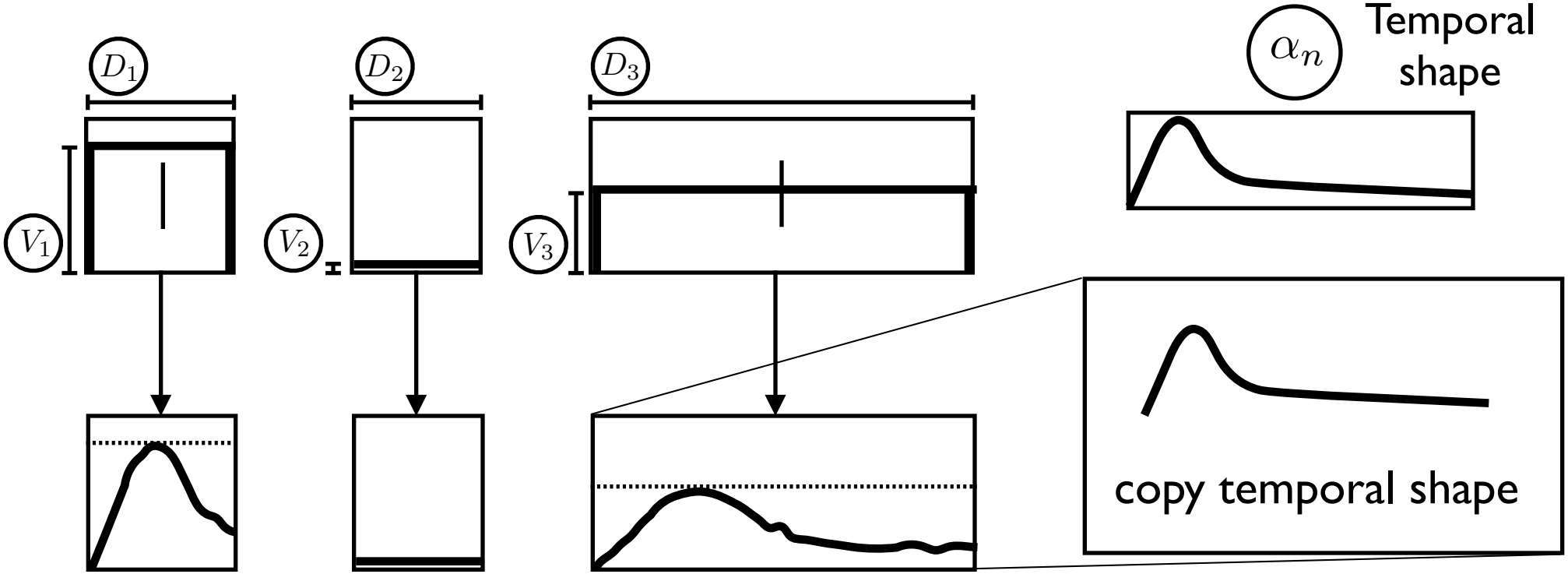
# Activation Model



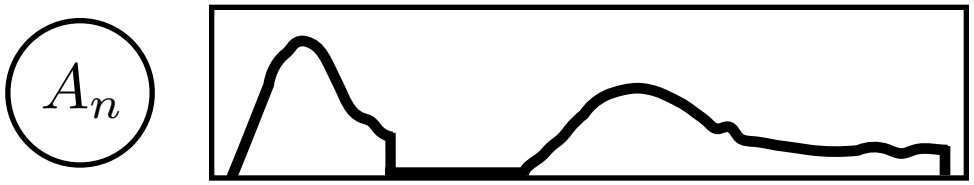

Activation



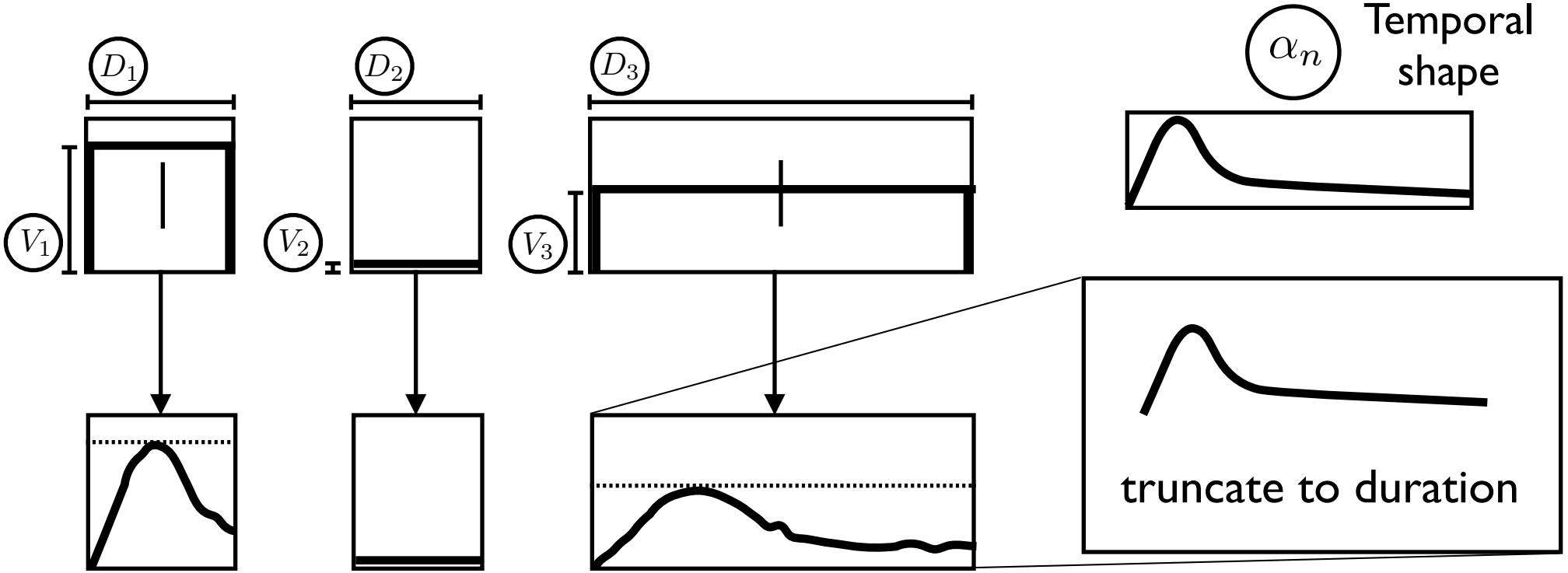
# Activation Model



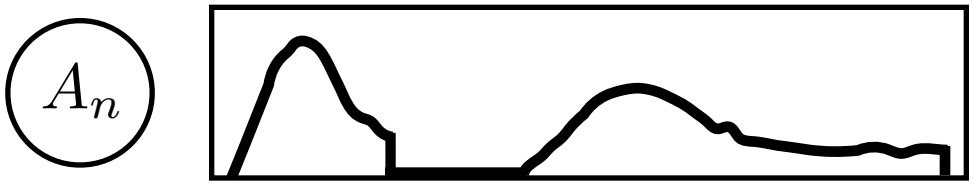

Activation



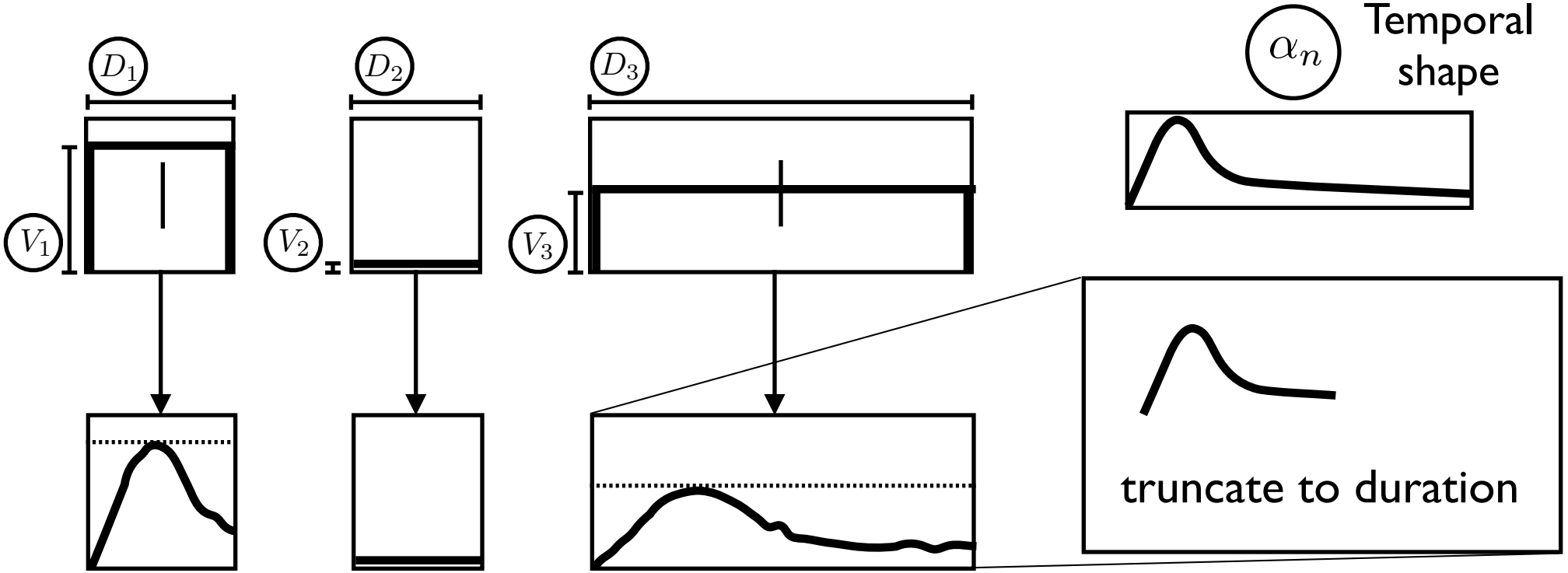
# Activation Model



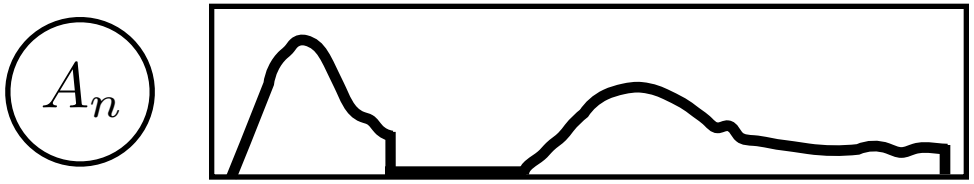

Activation



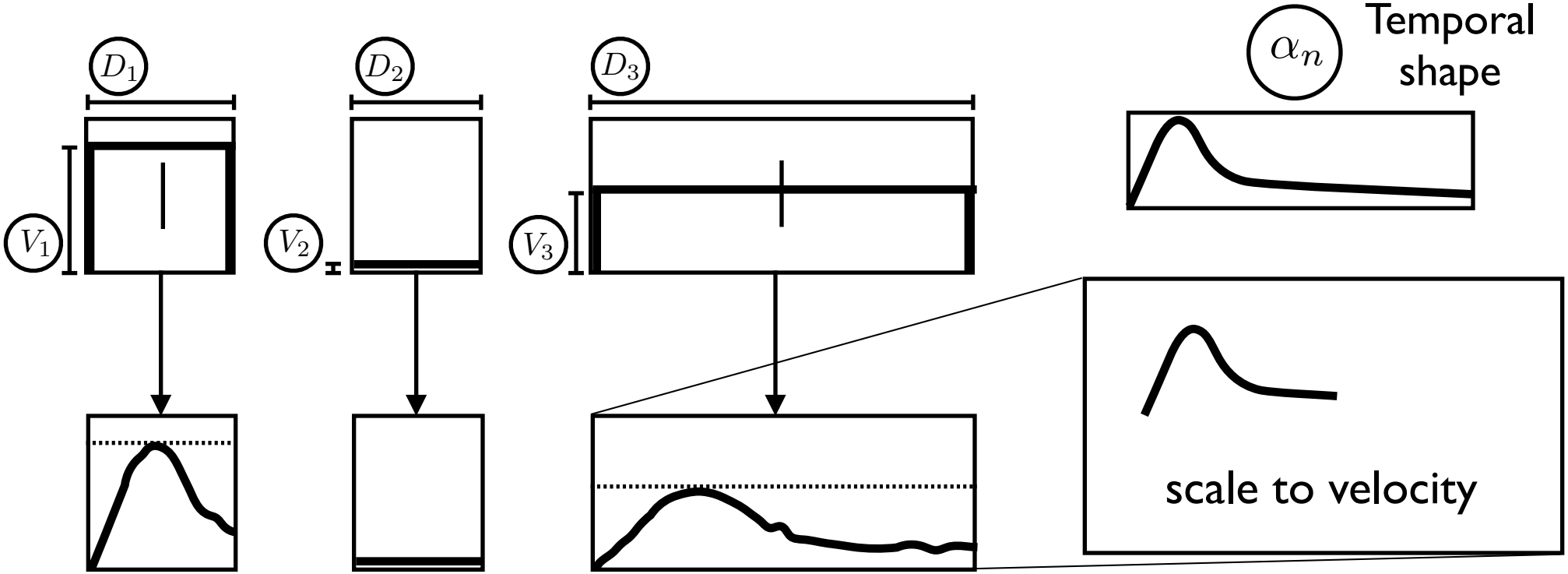
# Activation Model



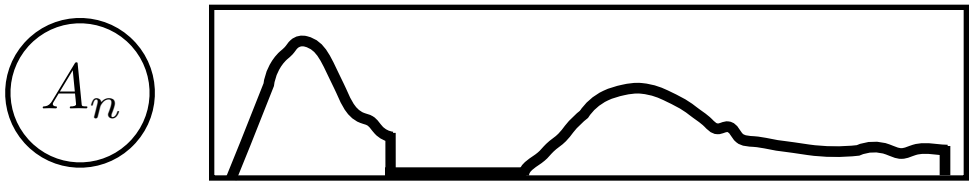

Activation



# Activation Model

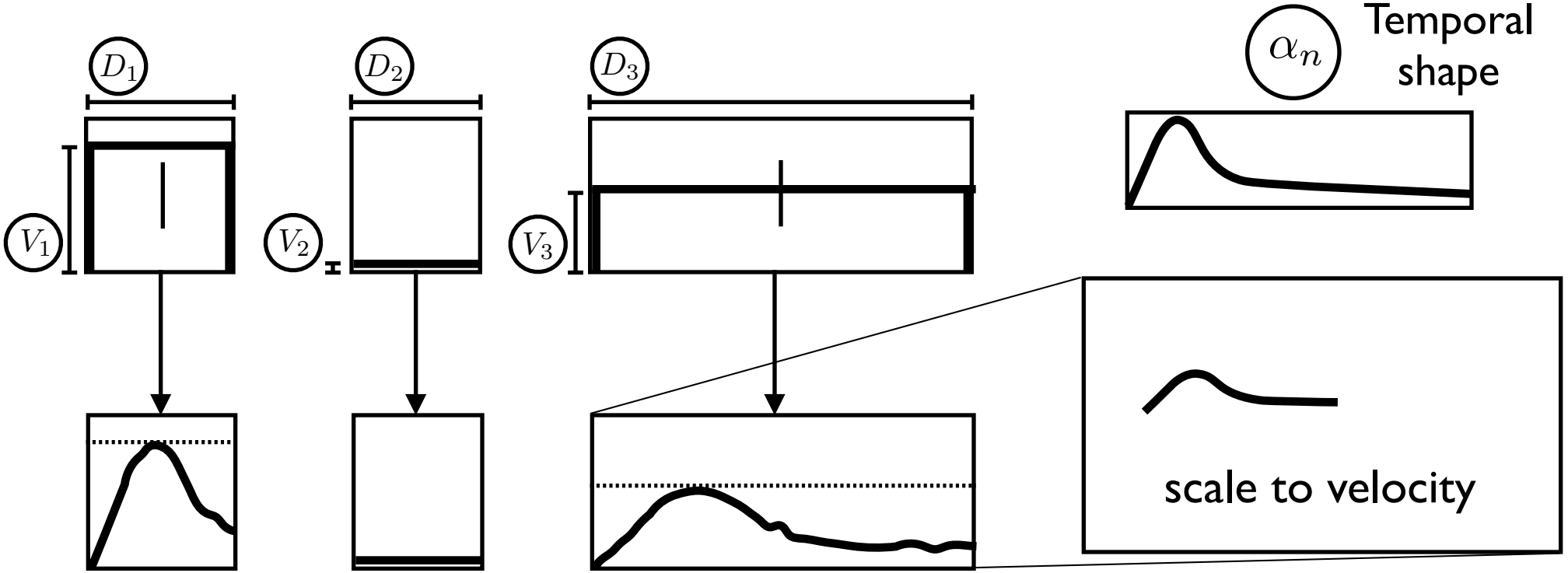



Activation

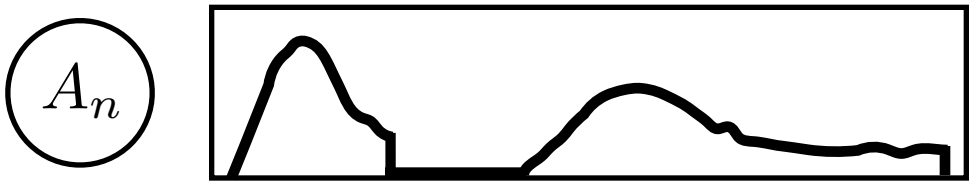


Berkeley N L P

# Activation Model

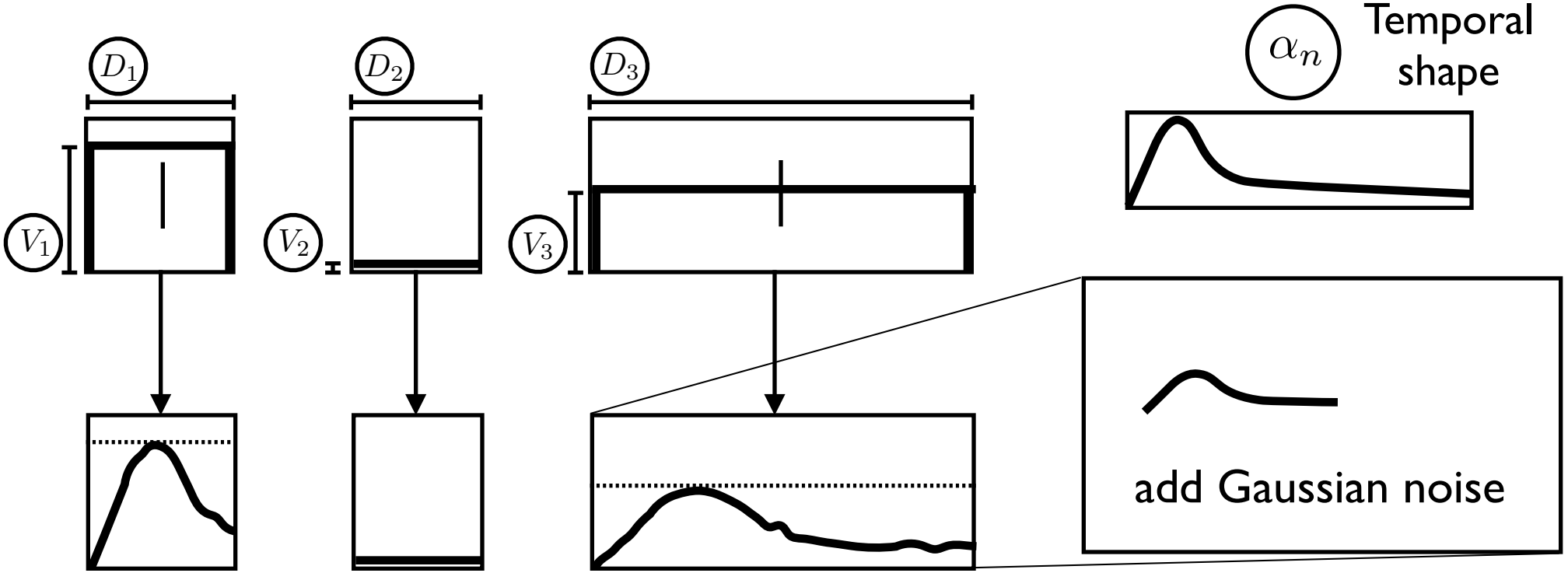



Activation

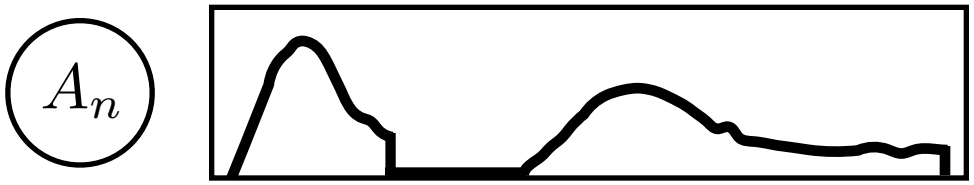


Berkeley
N L P

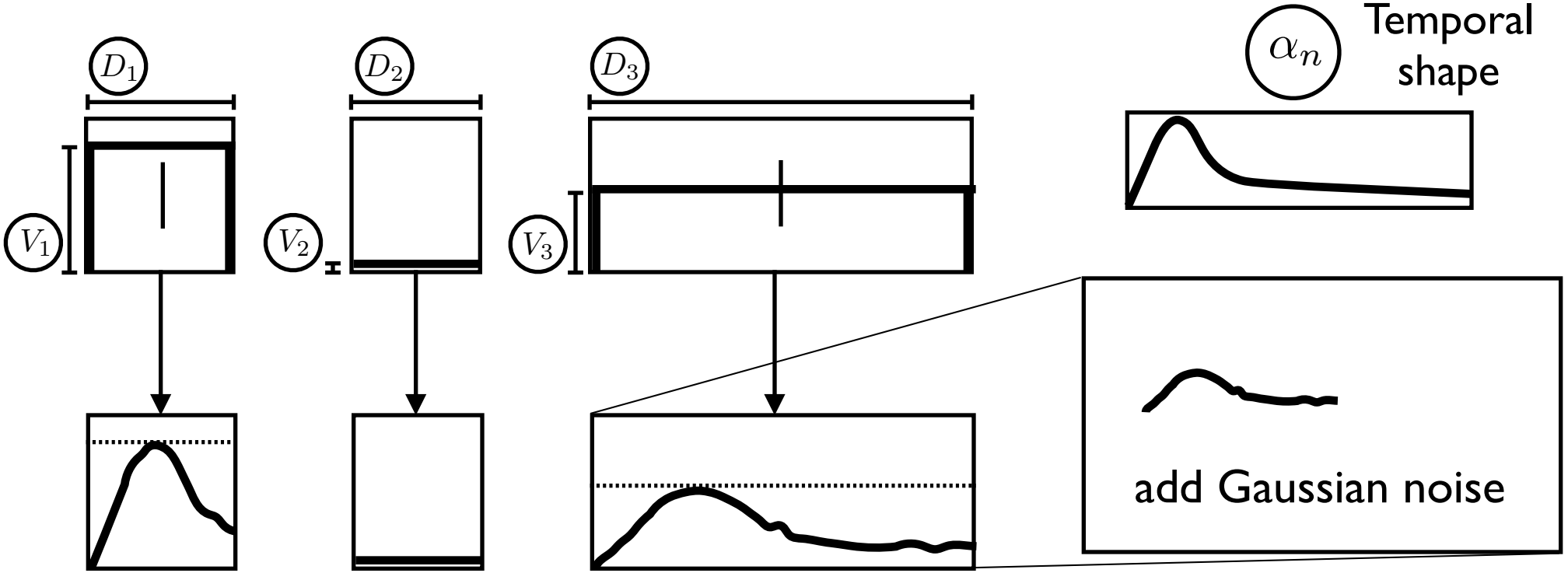
# Activation Model



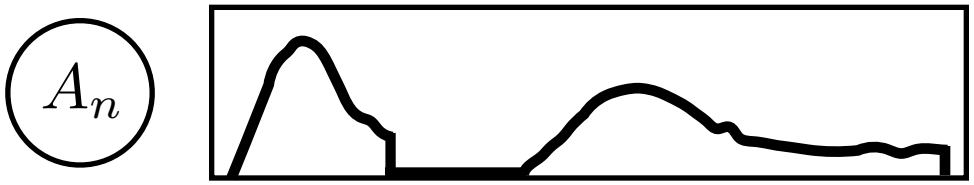

Activation



# Activation Model



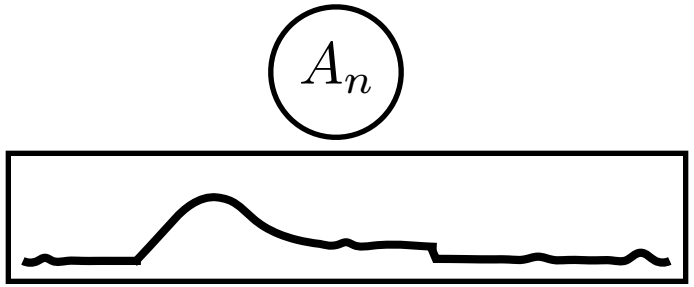

Activation



Berkeley
N L P

# Component Spectrogram Model

Activation



Berkeley
N L P

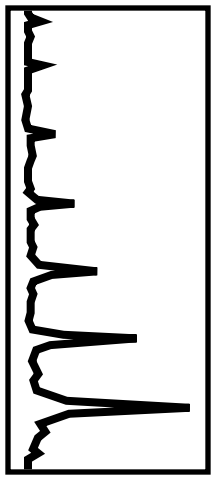
# Component Spectrogram Model

Activation

$A_n$

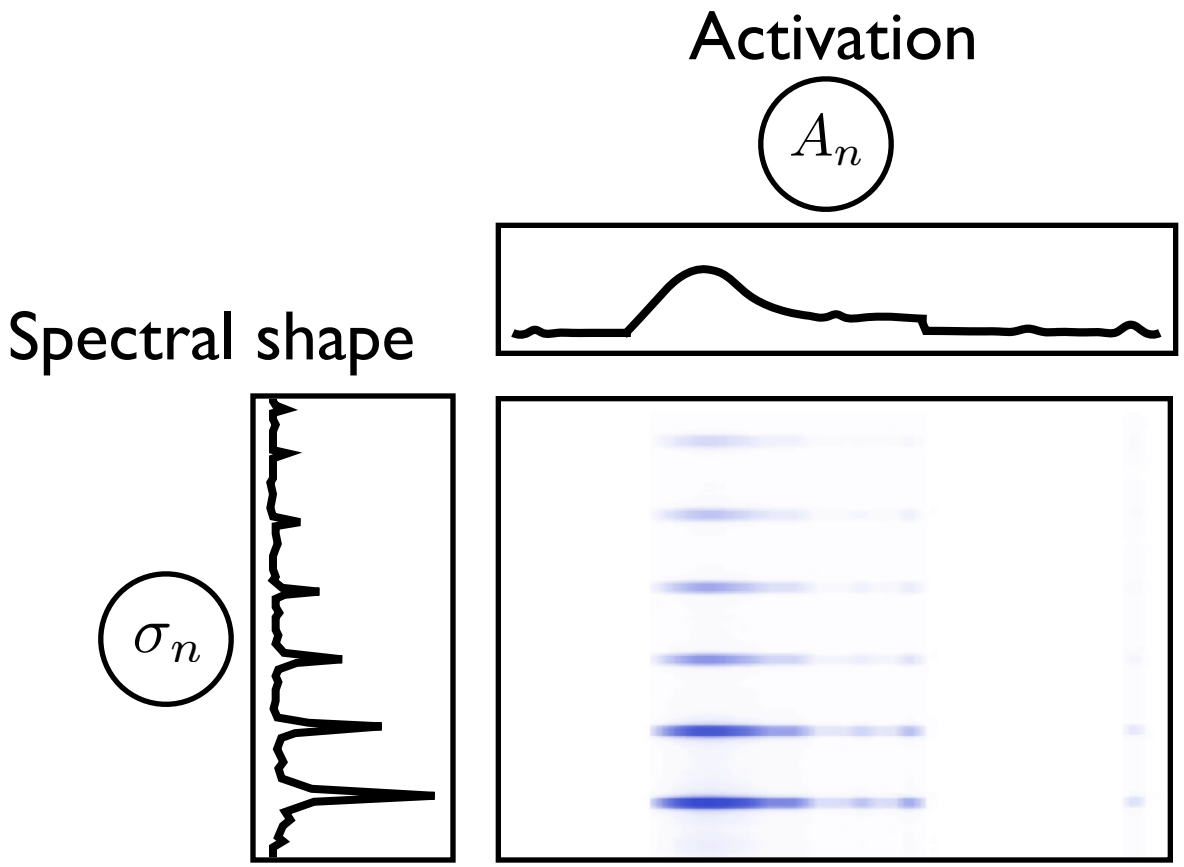
Spectral shape

$\sigma_n$

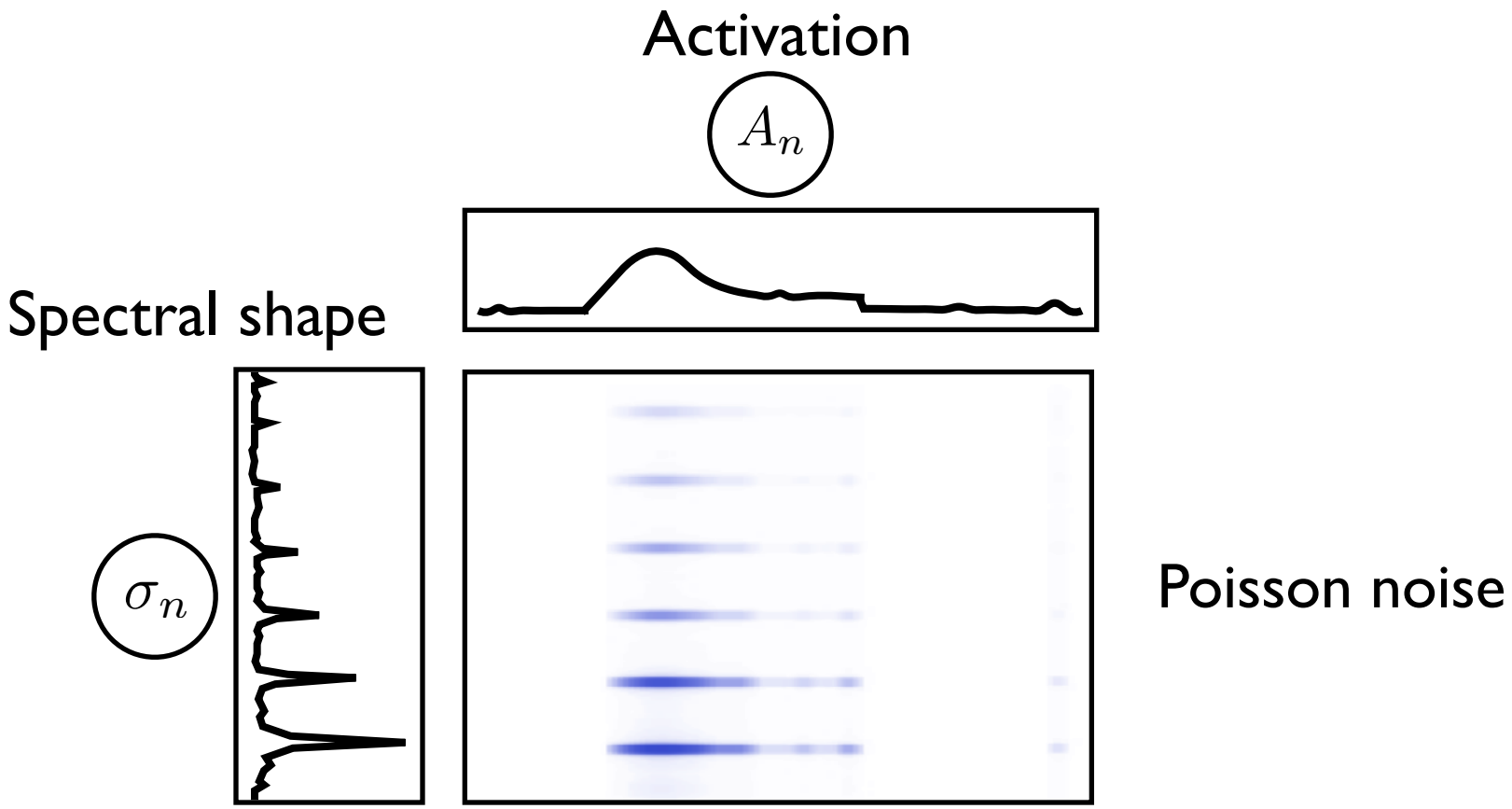


Berkeley N L P

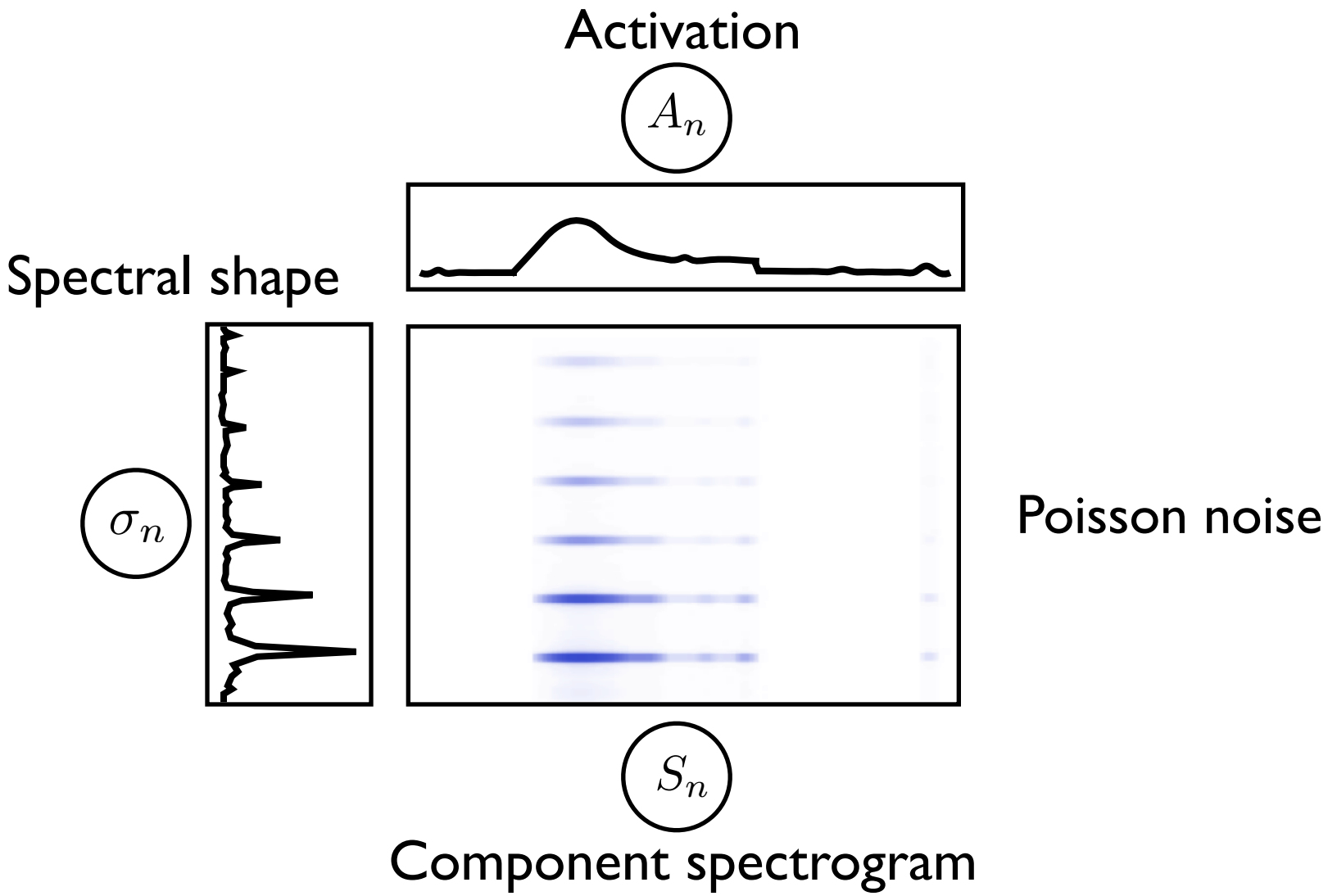
# Component Spectrogram Model



# Component Spectrogram Model

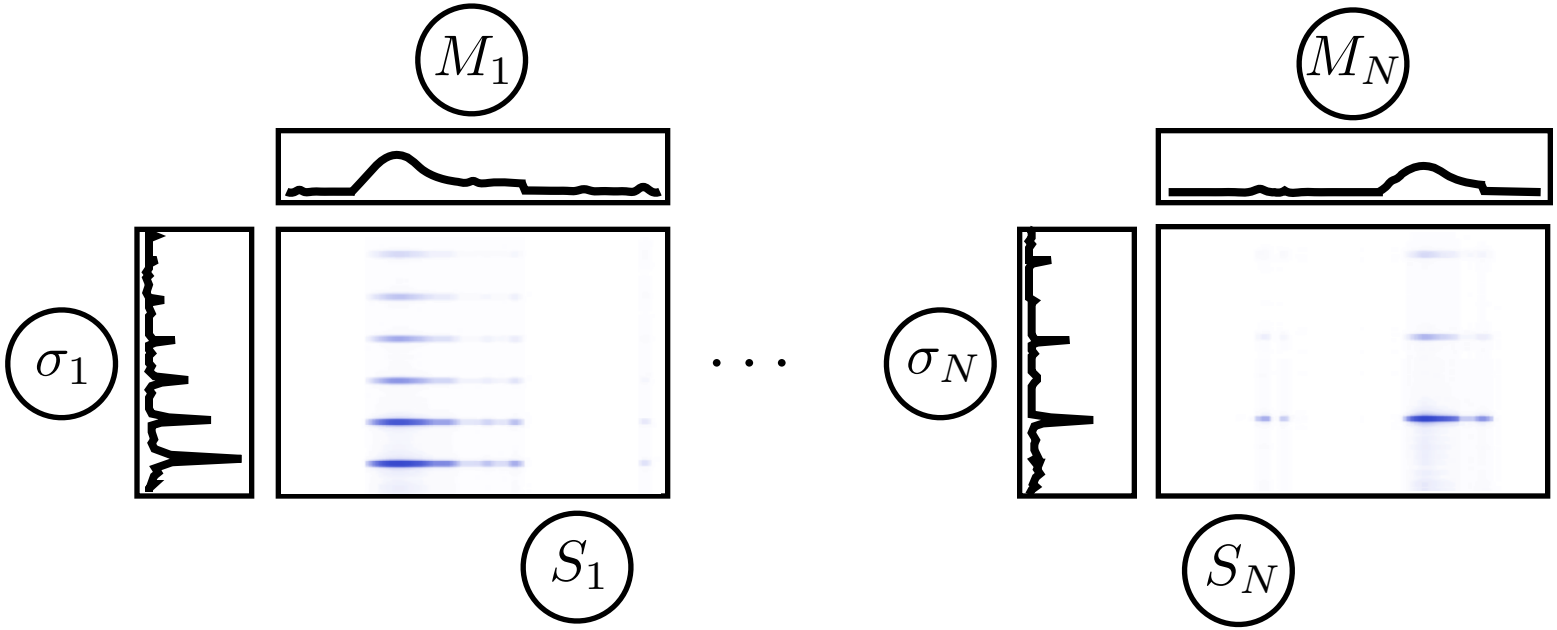


# Component Spectrogram Model



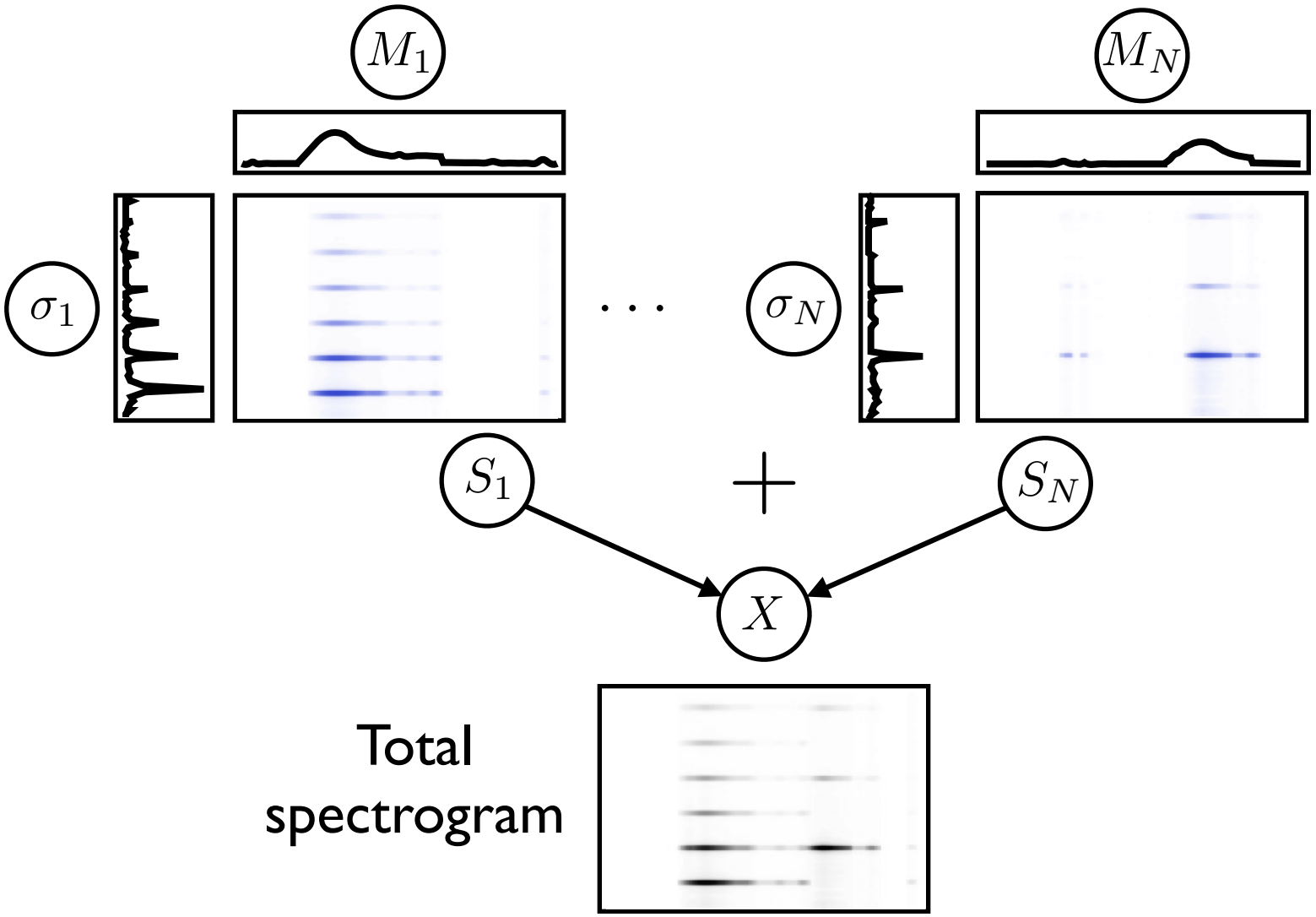
Berkeley
N L P

# Total Spectrogram Model



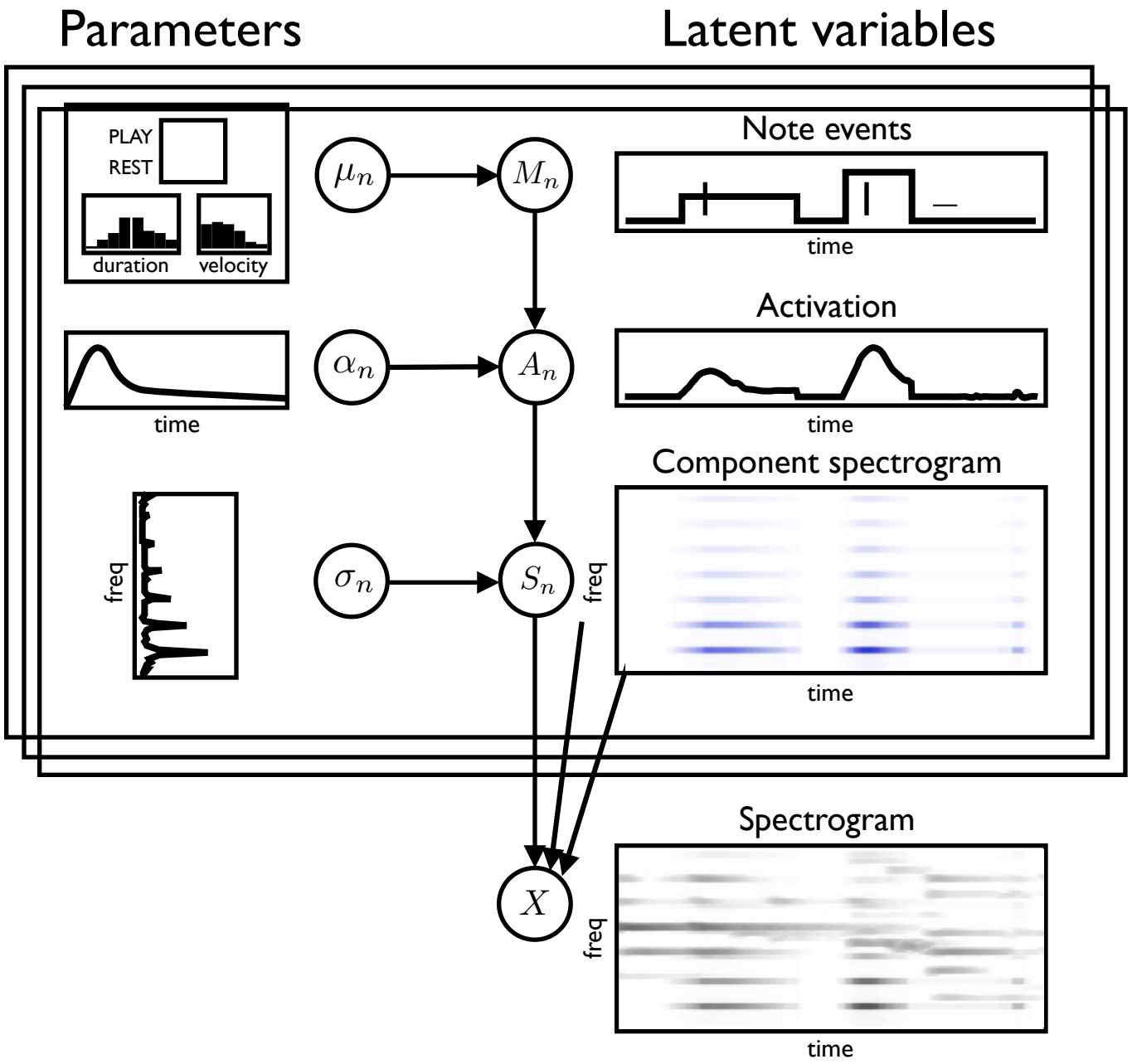
Berkeley N L P

# Total Spectrogram Model



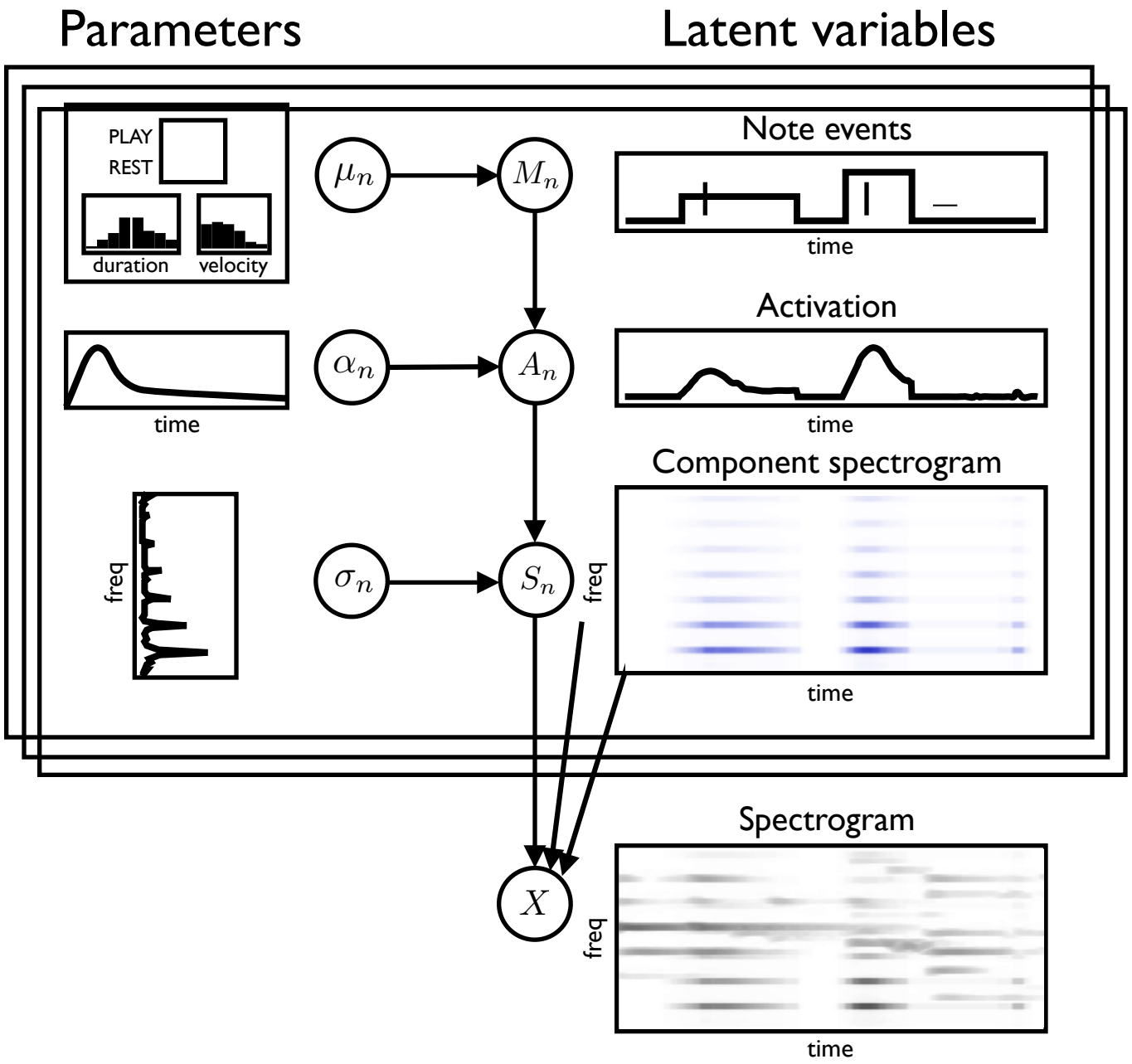
Berkeley NLP

# Learning and Inference



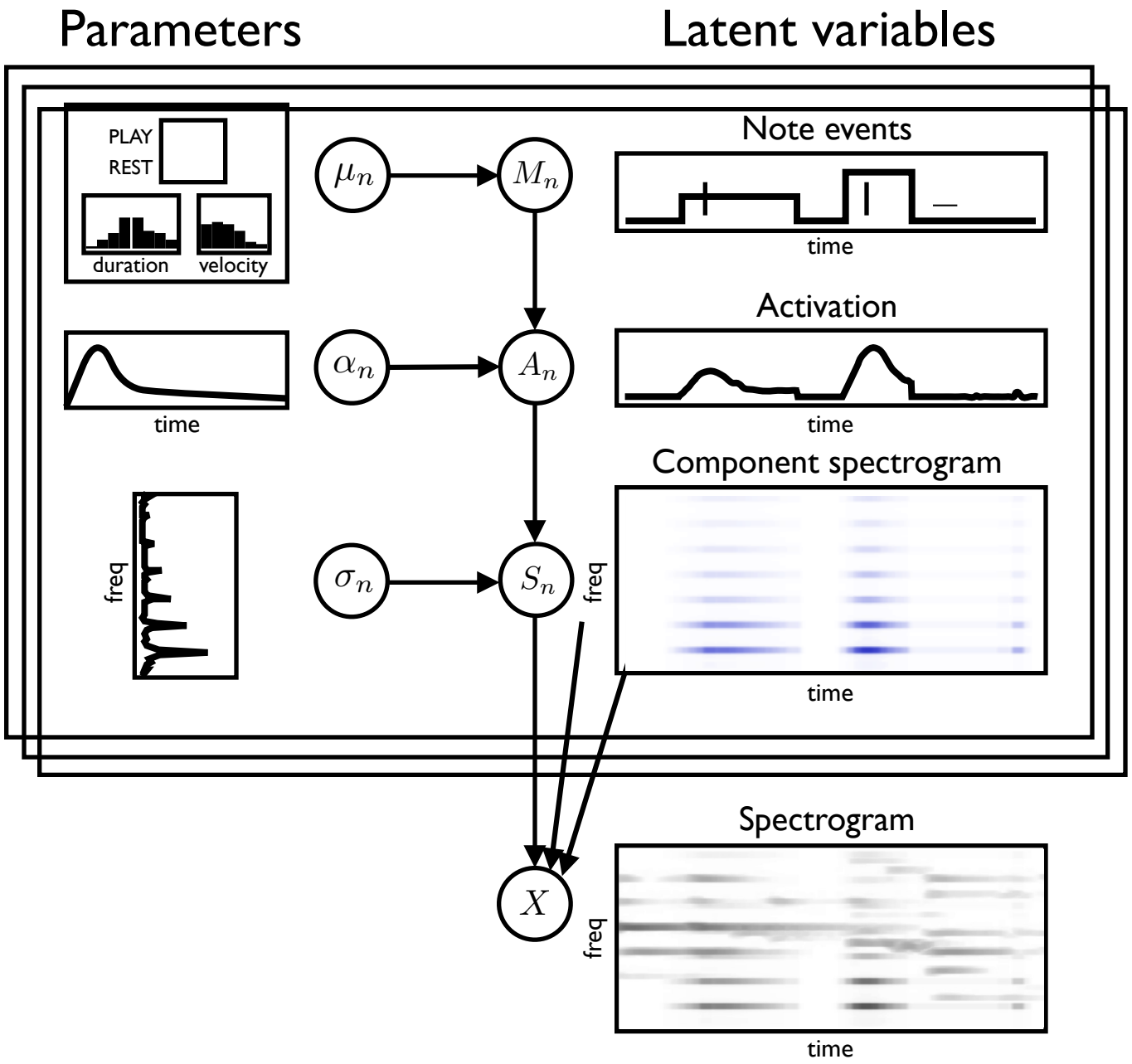
Berkeley NLP

# Learning and Inference



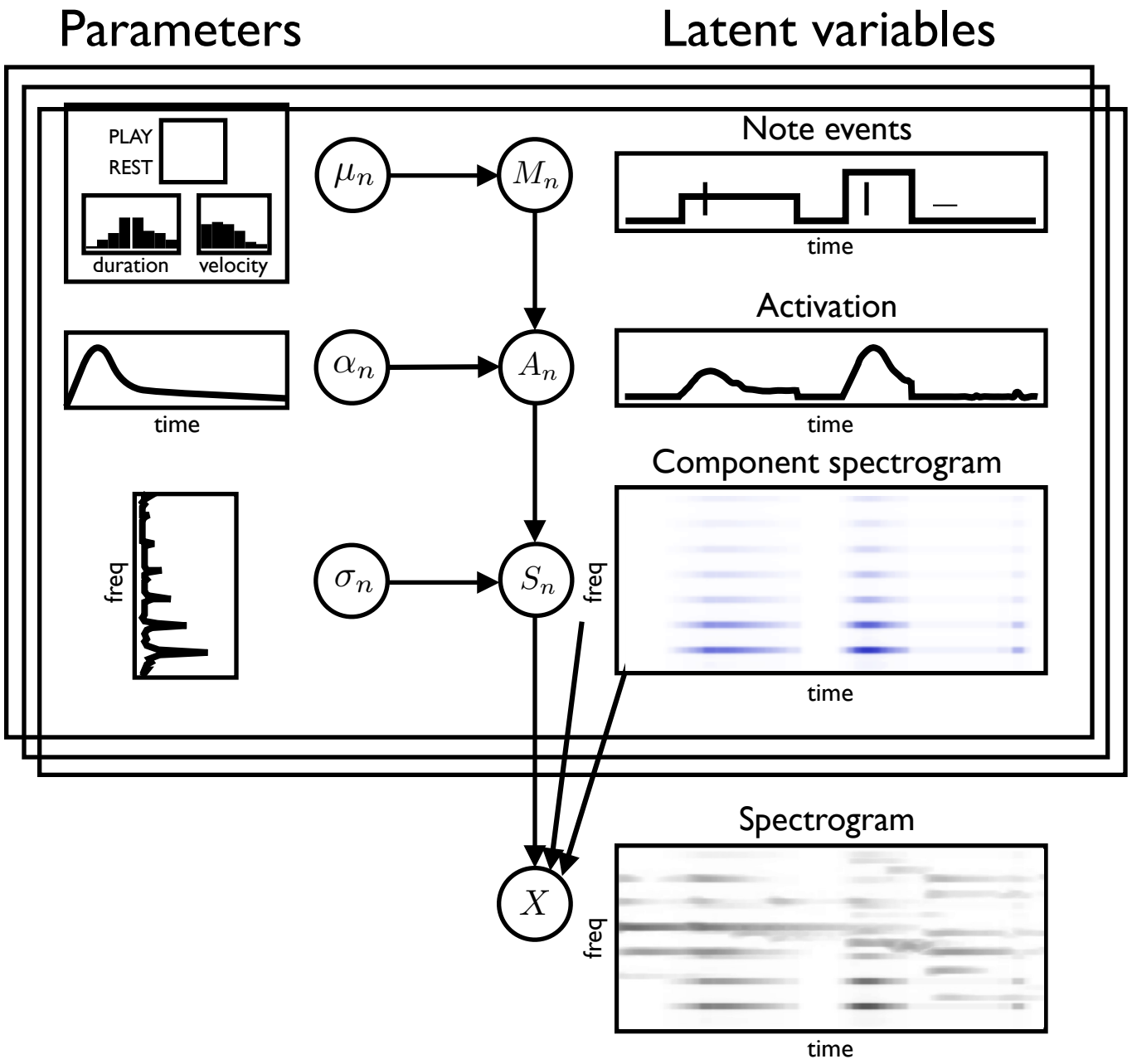
Berkeley NLP

# Learning and Inference



Berkeley NLP

# Learning and Inference




# Learning and Inference

Berkeley

N L P

# Learning and Inference

Note events update:

$$M|A, \alpha, \mu$$

Berkeley

N L P

# Learning and Inference

Note events update:

$$M|A, \alpha, \mu$$

Semi-Markov dynamic program

Berkeley
N L P

# Learning and Inference

Note events update:

$$M|A, \alpha, \mu$$

Semi-Markov dynamic program

Temporal shapes update:

$$\alpha|A, M$$

Berkeley
N L P

# Learning and Inference

Note events update:

$M|A, \alpha, \mu$

Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$

Closed form update

Berkeley
N L P

# Learning and Inference

Note events update:

$$M|A, \alpha, \mu$$

Semi-Markov dynamic program

Temporal shapes update:

$$\alpha|A, M$$

Closed form update

Activations update:

$$A|M, X, \alpha, \sigma$$

Berkeley
N L P

# Learning and Inference

Note events update:

$M|A, \alpha, \mu$

Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$

Closed form update

Activations update:

$A|M, X, \alpha, \sigma$

Exponentiated gradient ascent

Berkeley
N L P

# Learning and Inference

Note events update:

$$M|A, \alpha, \mu$$

Semi-Markov dynamic program

Temporal shapes update:

$$\alpha|A, M$$

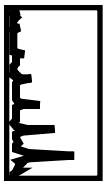
Closed form update

Activations update:

$$A|M, X, \alpha, \sigma$$

Exponentiated gradient ascent

Spectral shapes update:



$$\sigma|A, X$$

Berkeley
N L P

# Learning and Inference

Note events update:

$M|A, \alpha, \mu$

Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$

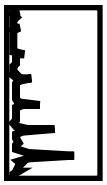
Closed form update

Activations update:

$A|M, X, \alpha, \sigma$

Exponentiated gradient ascent

Spectral shapes update:



$\sigma|A, X$

Exponentiated gradient ascent

Berkeley
N L P

# Evaluation

## Onset F1

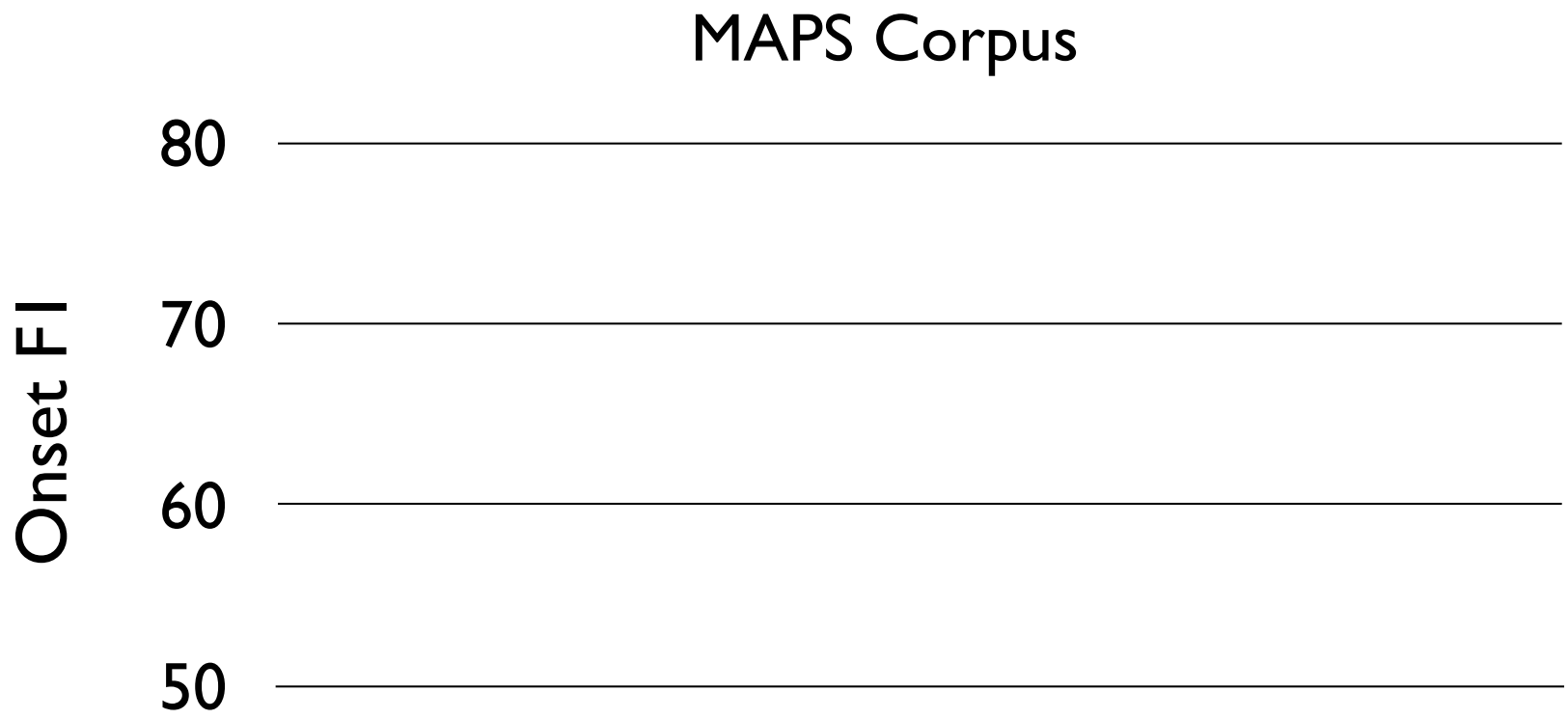
Berkeley
N L P

# Evaluation

## Onset F1

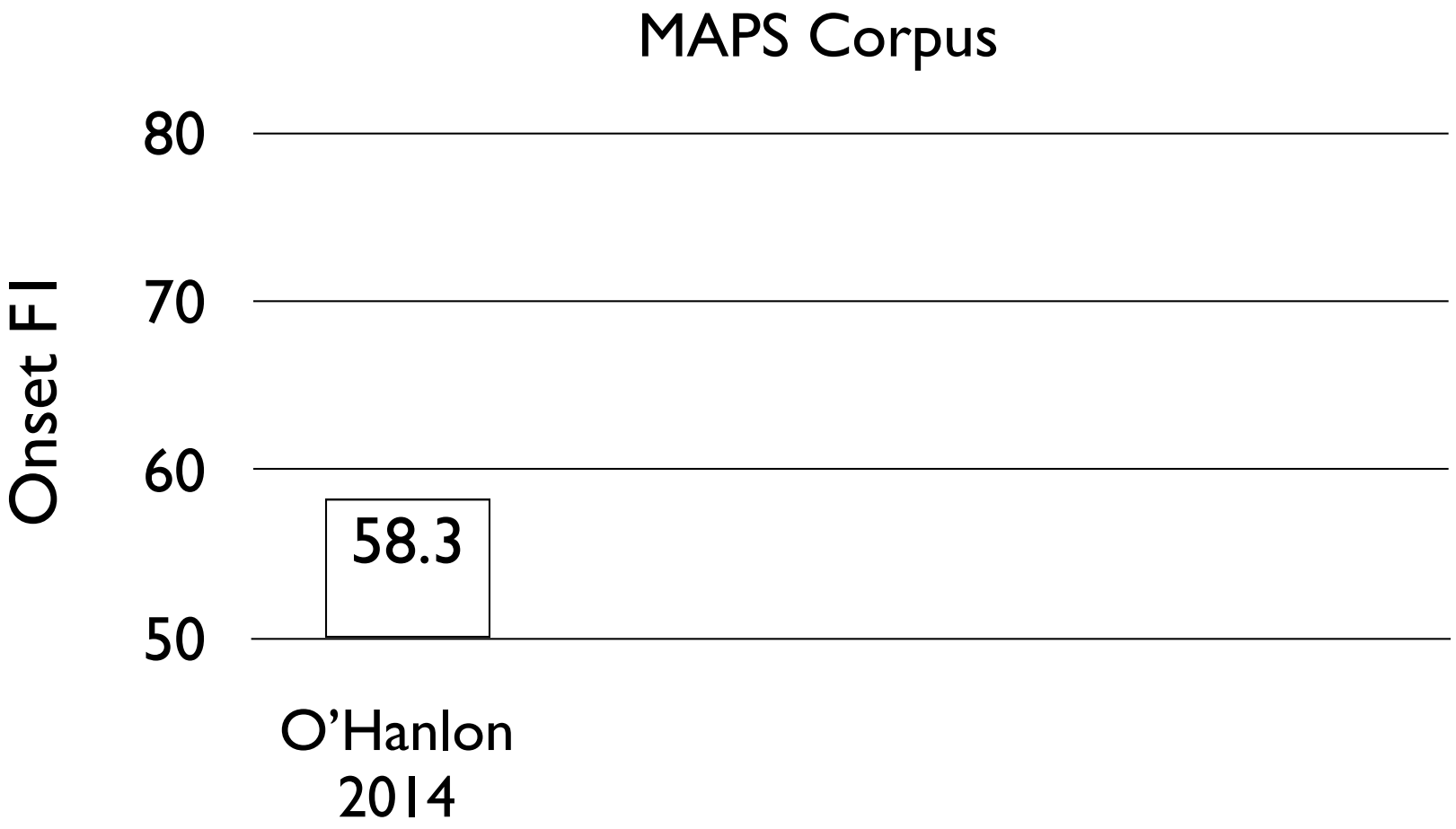
Berkeley NLP

# Results



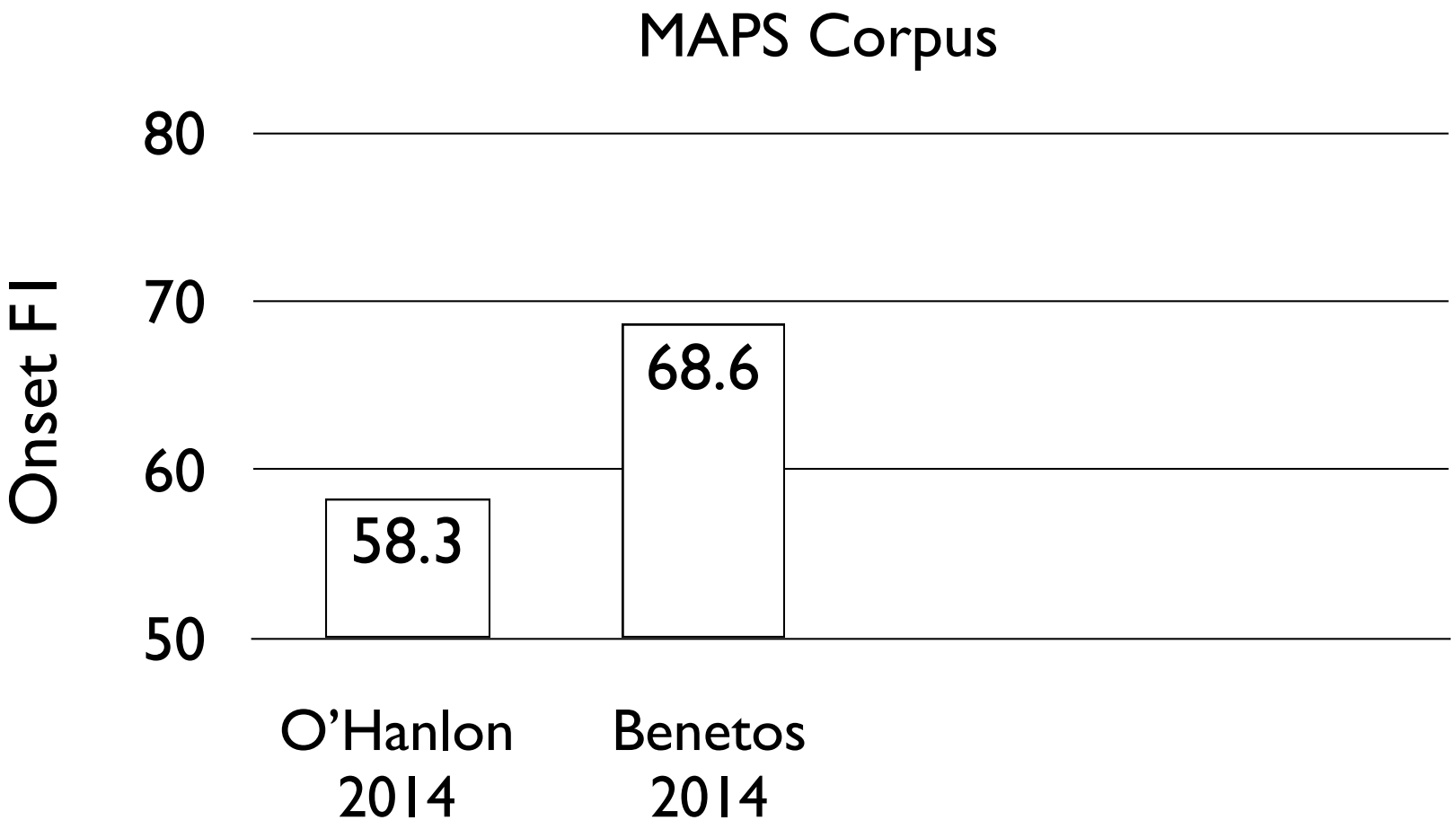

[Valentin et al. 2010]

# Results



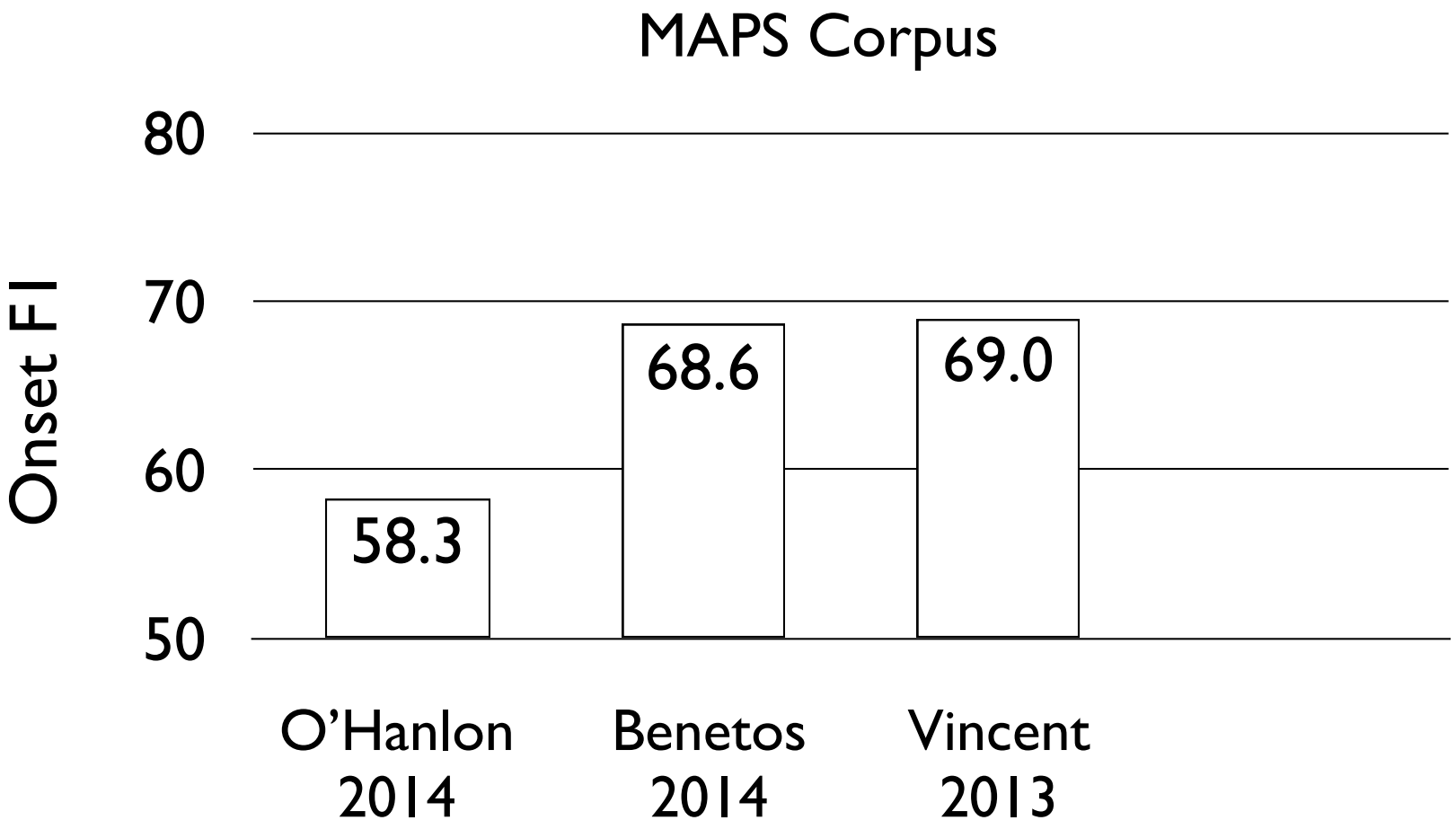

[Valentin et al. 2010]

# Results



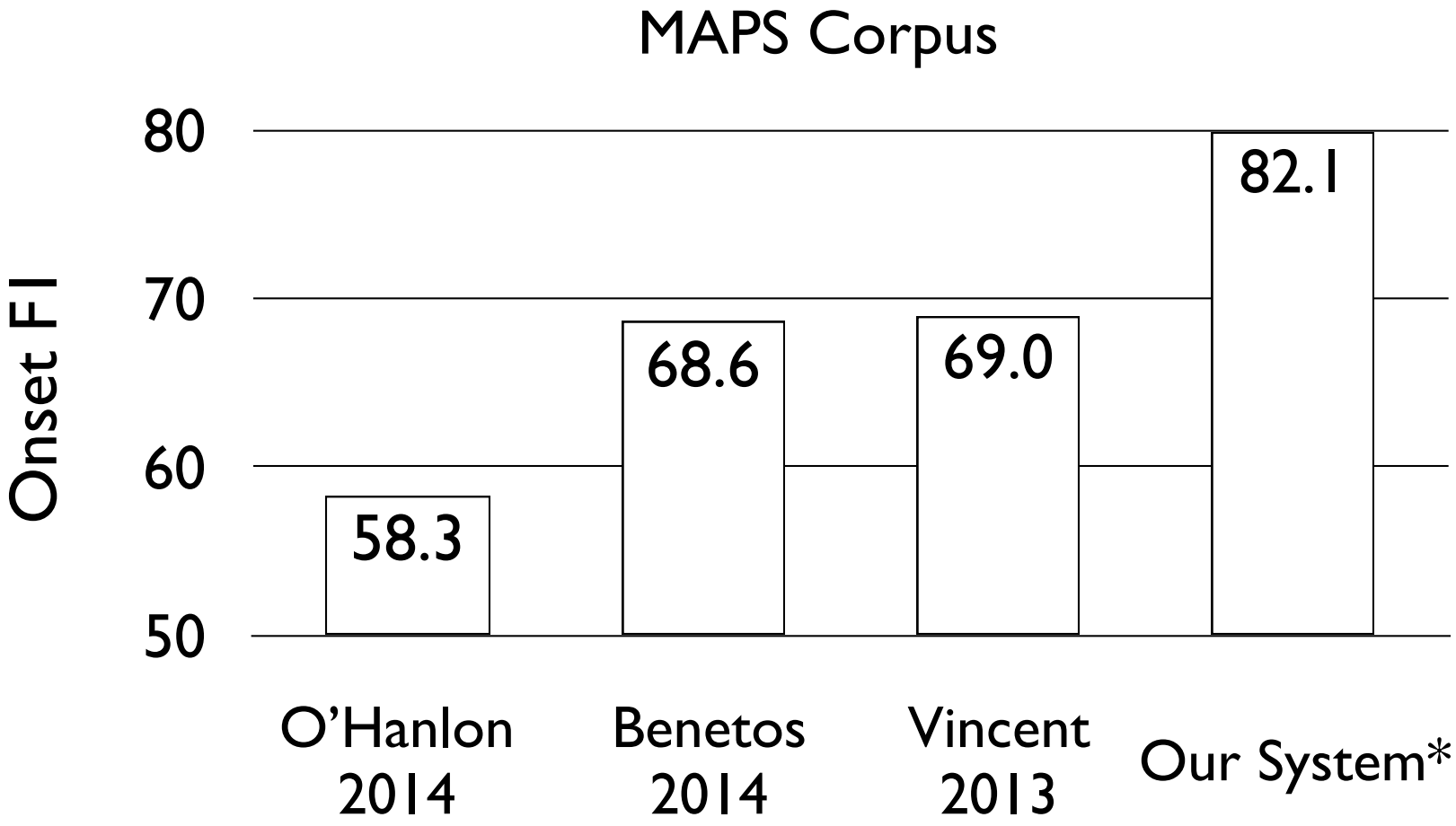

[Valentin et al. 2010]

# Results




[Valentin et al. 2010]

# Results




[Valentin et al. 2010]

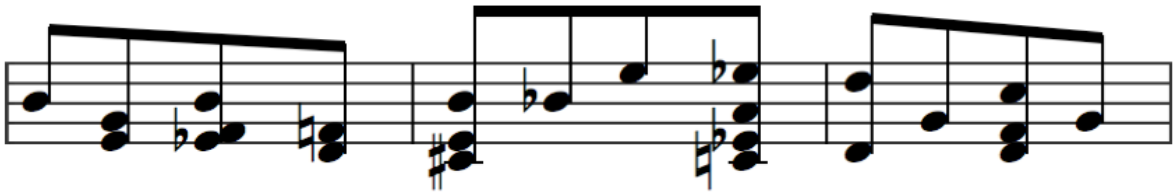
[Berg-Kirkpatrick et al. 2014]


# Transcription

Berkeley
N L P

# Transcription

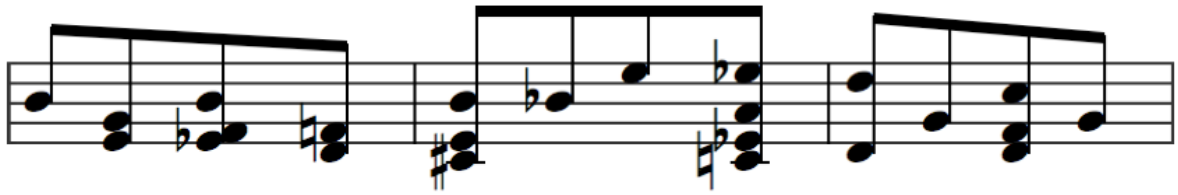
## Reference



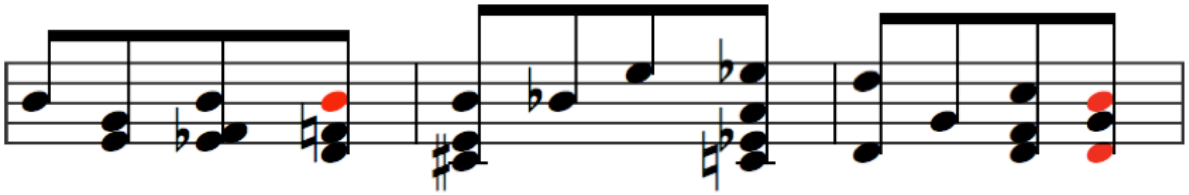
Berkeley
N L P

# Transcription

## Reference



## Predicted




# Resynthesized Examples


# Resynthesized Examples

Grieg input


# Resynthesized Examples

Grieg input

Grieg resynth piano


# Resynthesized Examples

Grieg input

Grieg resynth piano

Grieg resynth guitar