Automatic Summarization (and other stuff)

Berkeley NLP

Taylor Berg-Kirkpatrick
CS 288 UC Berkeley

Summarization

Summarization

Multidocument Summarization

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.

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

Maximum Marginal Relevance

S_9 S_2 S_5 S_1 S_4 S_7 S_3 S_8 S_6

Maximum Marginal Relevance

S_9 S_2 S_5 S_1 S_4 S_7 S_3 S_8 S_6

Maximum Marginal Relevance

S_9 S_2 S_5 S_1 S_4 S_7 S_3 S_8 S_6

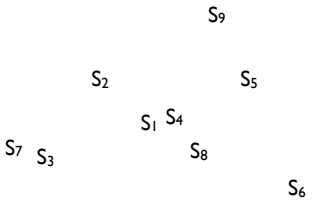
S_1: She traveled to France on Friday.

Maximum Marginal Relevance

S_9 S_2 S_5 S_1 S_4 S_7 S_3 S_8 S_6

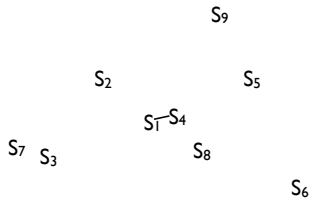
S_1: She traveled to France on Friday.

Maximum Marginal Relevance



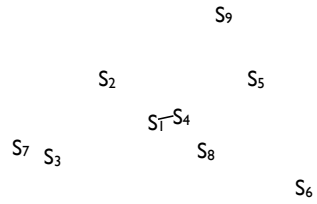
S_1: She traveled to France on Friday.
S_4: On Friday, She took a trip to France.

Maximum Marginal Relevance



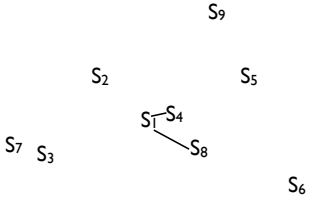
S_1: She traveled to France on Friday.
S_4: On Friday, She took a trip to France.

Maximum Marginal Relevance



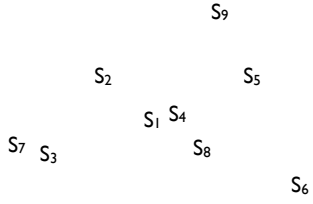
S_1: She traveled to France on Friday.
S_4: On Friday, She took a trip to France.
S_8: She plans to stay for two weeks.

Maximum Marginal Relevance

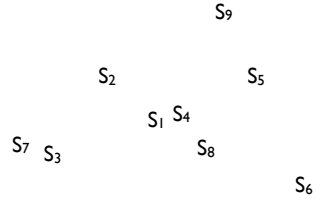


S_1: She traveled to France on Friday.
S_4: On Friday, She took a trip to France.
S_8: She plans to stay for two weeks.

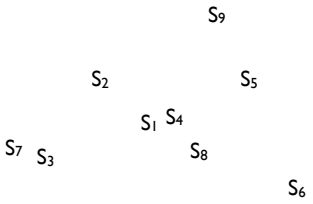
Maximum Marginal Relevance



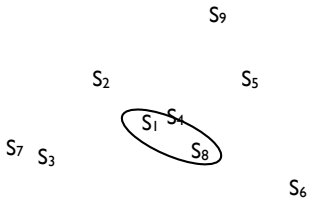
Maximum Marginal Relevance



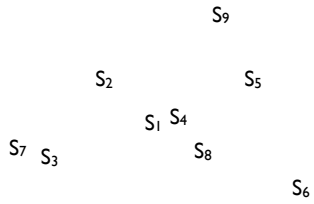
Maximum Marginal Relevance



Maximum Marginal Relevance



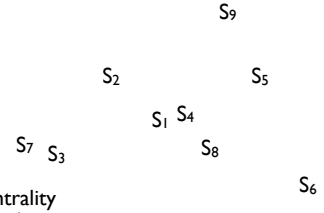
Maximum Marginal Relevance



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

[Carbonell and Goldstein, 1998]

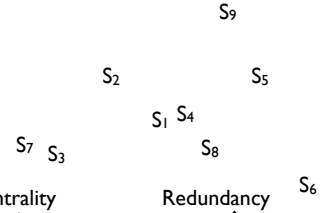
Maximum Marginal Relevance



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

[Carbonell and Goldstein, 1998]

Maximum Marginal Relevance



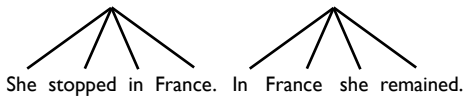
$$\underset{i \in D \setminus S}{\operatorname{argmax}} \left[\lambda \cdot \operatorname{sim}(S_i, D) - (1 - \lambda) \cdot \max_j \left(\operatorname{sim}(S_i, S_j) \right) \right]$$

[Carbonell and Goldstein, 1998]

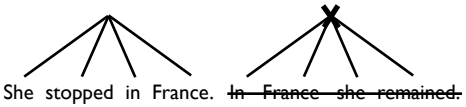
Max Coverage

She stopped in France. In France she remained.

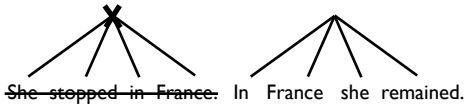
Max Coverage



Max Coverage

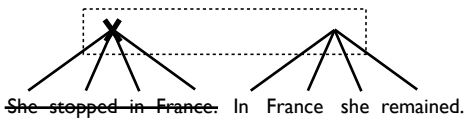


Max Coverage

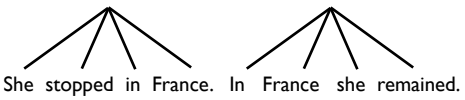


Max Coverage

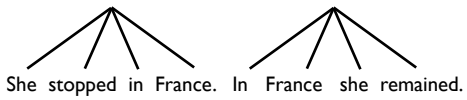
A summary is a set of cuts: S



Max Coverage

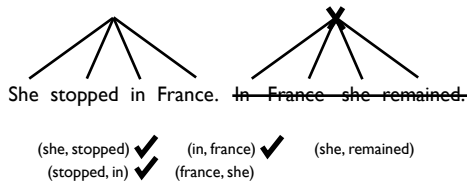


Max Coverage

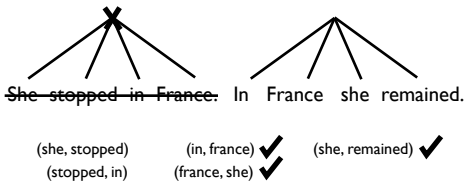


(she, stopped) ✓
(stopped, in) ✓
(in, france) ✓
(france, she) ✓
(she, remained) ✓

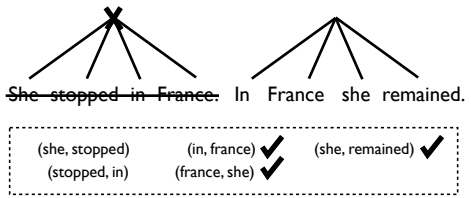
Max Coverage



Max Coverage



Max Coverage



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

Max Coverage

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

Max Coverage

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

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

Max Coverage

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

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

Max Coverage

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

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

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

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]

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

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

[Gillick and Favre 2008]

ILP for Decoding

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

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

ILP for Decoding

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

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

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

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

ILP for Decoding

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

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

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

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

Linear Model for Extraction

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

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

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

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:

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

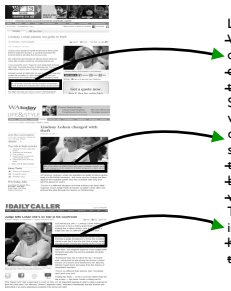
Sentence Extraction

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



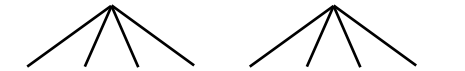
Lindsay Lohan pleaded not guilty ~~Wednesday~~ to felony grand theft of a 2,500 necklace, ~~a case that could return the troubled starlet to jail rather than the big screen.~~ Saying it appeared that Lohan had violated her probation in a 2007 drunken driving case the judge set bail at $40,000 ~~and warned 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.

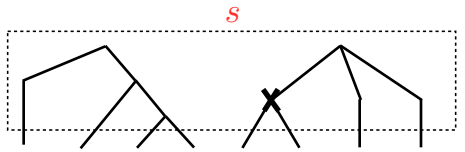
Joint Extractive / Compressive Model



Joint Extractive / Compressive Model

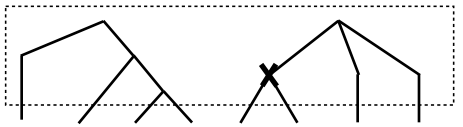
Joint Extractive / Compressive Model

Joint Extractive / Compressive Model



Joint Extractive / Compressive Model

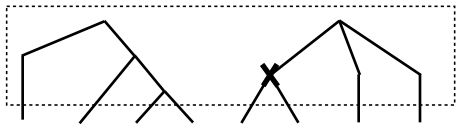
s



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

Joint Extractive / Compressive Model

s



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

$B(s)$

Joint Extractive / Compressive Model

s

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

$B(s)$

Joint Extractive / Compressive Model

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

Joint Extractive / Compressive Model

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

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:

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

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

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 $f(c)$

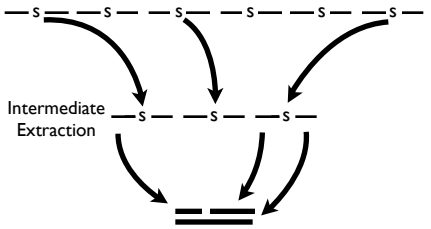
COORD:	Coordinated phrase, four versions: NP, VP, S, SBAR
S-ADJUNCT:	Adjunct to matrix verb, four versions: CC, PP, ADVP, SBAR
REL-C:	Relative clause indicator
ATTR-C:	Attribution clause indicator
ATTR-PP:	PP attribution indicator
TEMP-PP:	Temporal PP indicator
TEMP-NP	Temporal NP indicator
BIAS:	Always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

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.~~

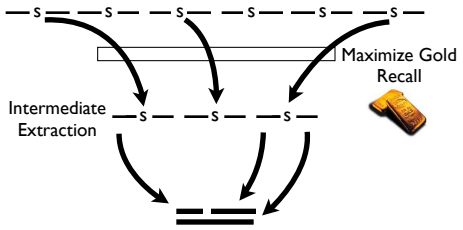
Obtaining Training Data

TAC 2009: 44 Document Collections



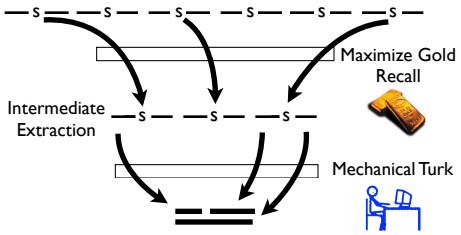
Obtaining Training Data

TAC 2009: 44 Document Collections



Obtaining Training Data

TAC 2009: 44 Document Collections



Learning

Learning

Structured SVM Training:

Learning

Structured SVM Training:

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

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

Learning

Structured SVM Training:

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

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

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

Learning

Structured SVM Training:

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

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

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

Learning

Structured SVM Training:

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

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

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

Learning

Structured SVM Training:

$$\min_w \left[\|w\|_2^2 + \xi \right]$$

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

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

End-to-end summarization quality

Learning

Structured SVM Training:

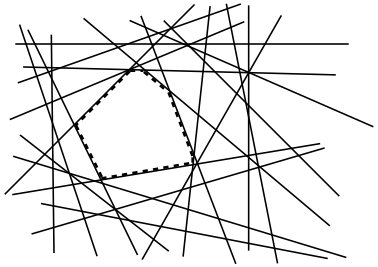
$$\min_w \left[\|w\|_2^2 + \xi \right]$$

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

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

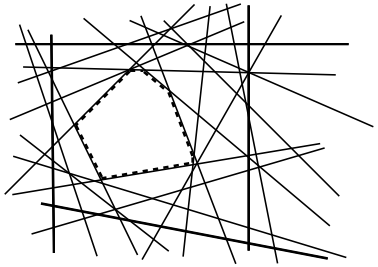
Exponentially many constraints!

Learning



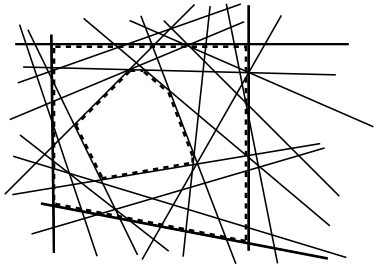
[Tsochantaridis et al. 2004]

Learning



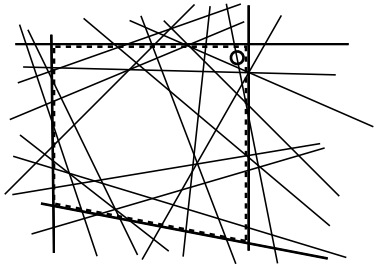
[Tsochantaridis et al. 2004]

Learning



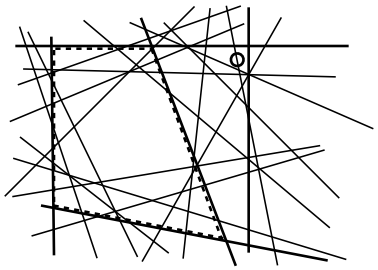
[Tsochantaridis et al. 2004]

Learning



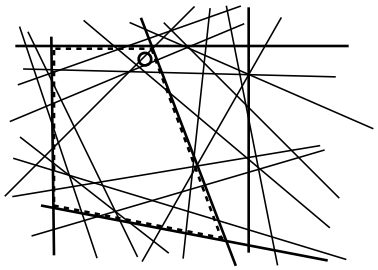
[Tsochantaridis et al. 2004]

Learning



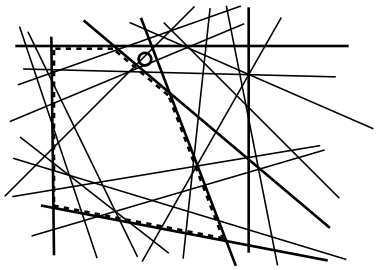
[Tsochantaridis et al. 2004]

Learning



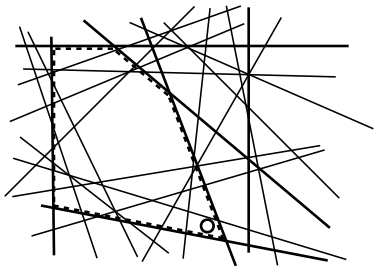
[Tsochantaridis et al. 2004]

Learning



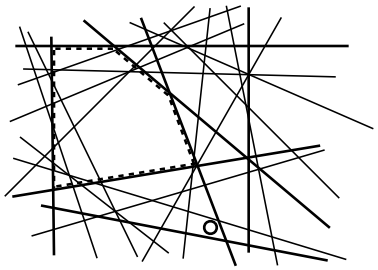
[Tsochantaridis et al. 2004]

Learning



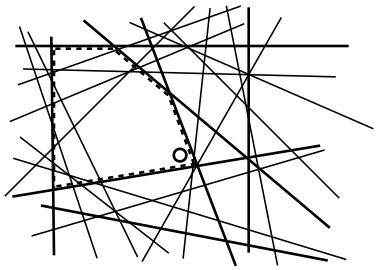
[Tsochantaridis et al. 2004]

Learning



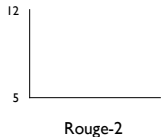
[Tsochantaridis et al. 2004]

Learning



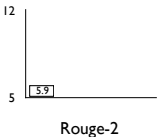
[Tsochantaridis et al. 2004]

Results



Rouge-2

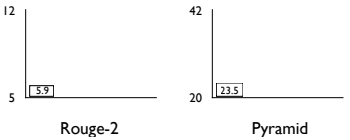
Results



Rouge-2

☐ Last Doc Baseline

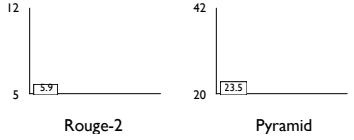
Results



Rouge-2 Pyramid

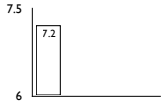
☐ Last Doc Baseline

Results



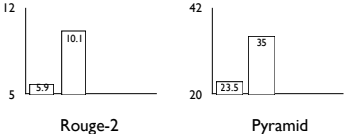
Rouge-2 Pyramid

☐ Last Doc Baseline



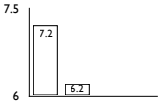
Linguistic Quality

Results



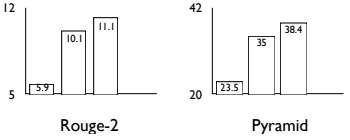
Rouge-2 Pyramid

☐ Last Doc Baseline
☐ Extractive Baseline



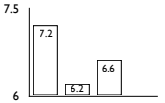
Linguistic Quality

Results



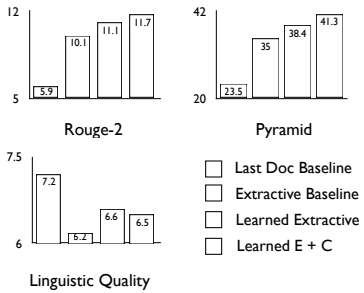
Rouge-2 Pyramid

☐ Last Doc Baseline
☐ Extractive Baseline
☐ Learned Extractive

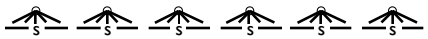


Linguistic Quality

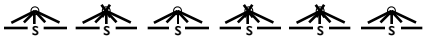
Results



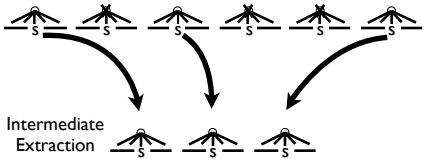
Coarse-to-fine Decoding



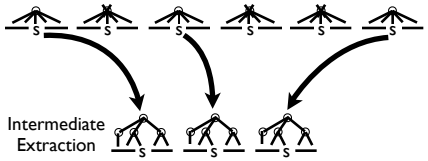
Coarse-to-fine Decoding



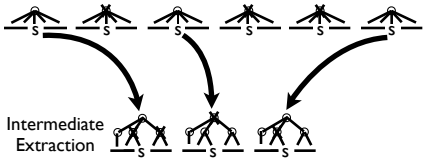
Coarse-to-fine Decoding



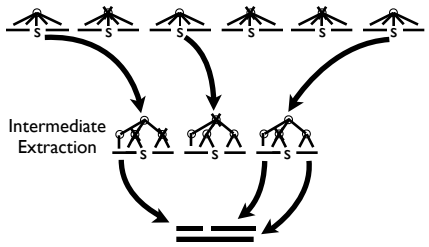
Coarse-to-fine Decoding



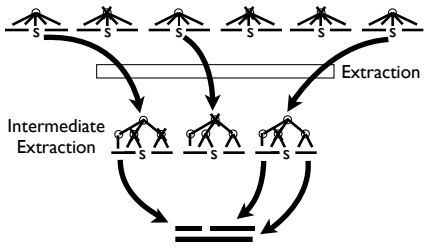
Coarse-to-fine Decoding



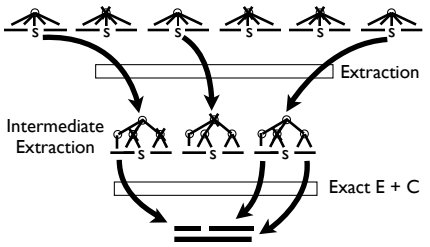
Coarse-to-fine Decoding



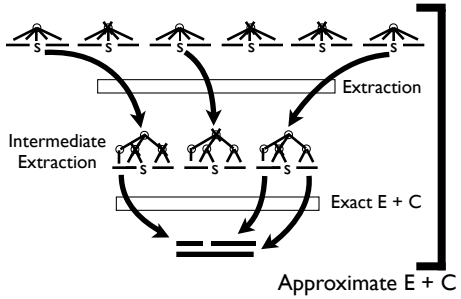
Coarse-to-fine Decoding



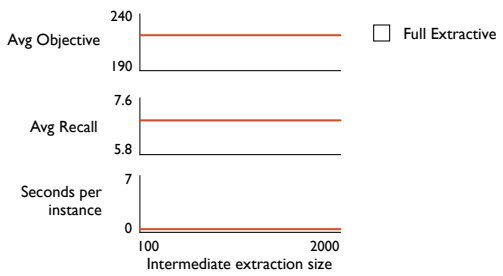
Coarse-to-fine Decoding



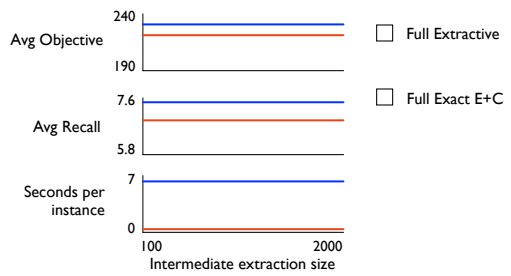
Coarse-to-fine Decoding



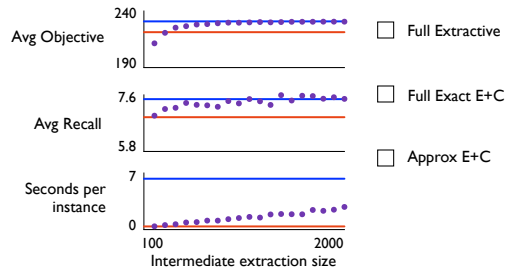
Coarse-to-fine Decoding



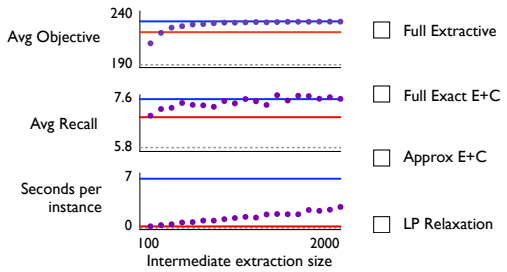
Coarse-to-fine Decoding



Coarse-to-fine Decoding



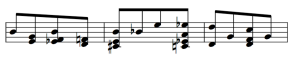
Coarse-to-fine Decoding



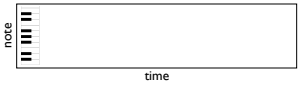
Piano Music Transcription

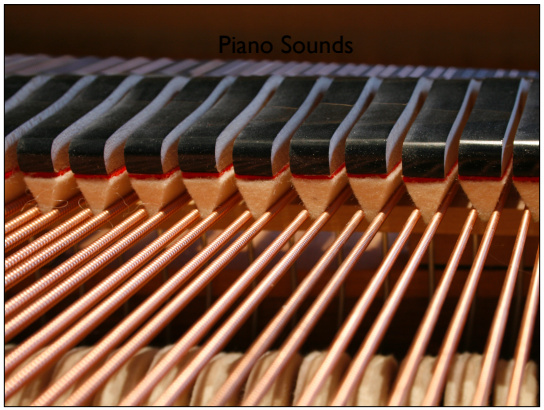
Piano Music Transcription

Piano Music Transcription

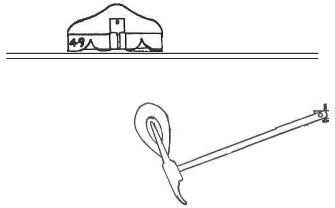


Piano Music Transcription

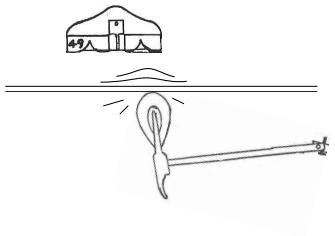




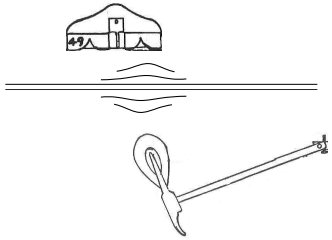
Piano Sounds



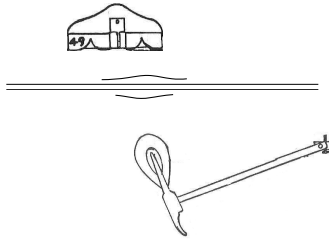
Piano Sounds



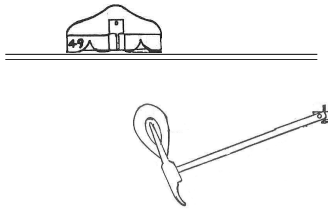
Piano Sounds



Piano Sounds



Piano Sounds



Berkeley N L P

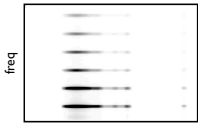
Piano Sounds

Berkeley N L P

Piano Sounds

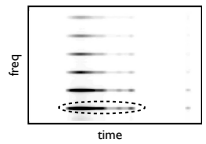
Berkeley N L P

Piano Sounds



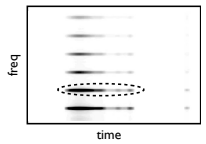
Berkeley N L P

Piano Sounds



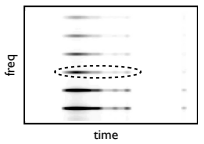
Berkeley N L P

Piano Sounds



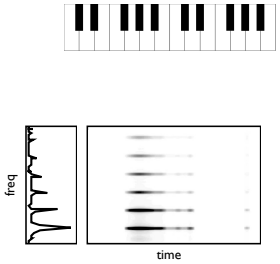
Berkeley N L P

Piano Sounds



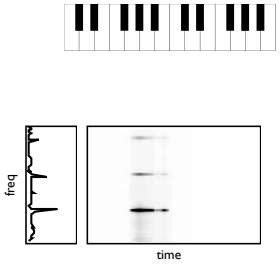
Berkeley N L P

Spectral Shape



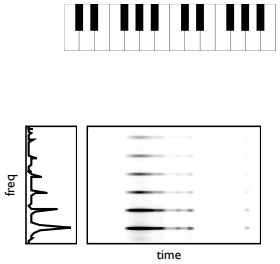
Berkeley N L P

Spectral Shape



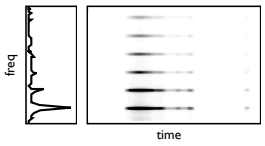
Berkeley N L P

Spectral Shape



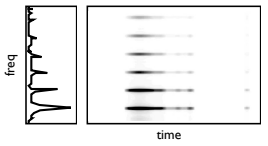
Berkeley N L P

Spectral Shape



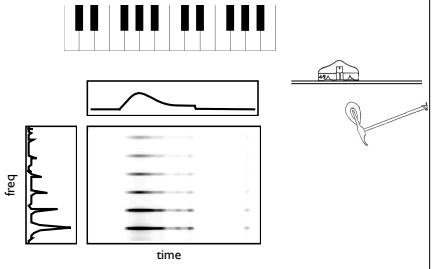
Berkeley N L P

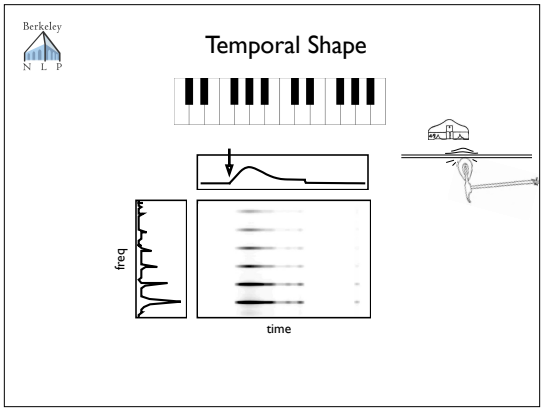
Temporal Shape

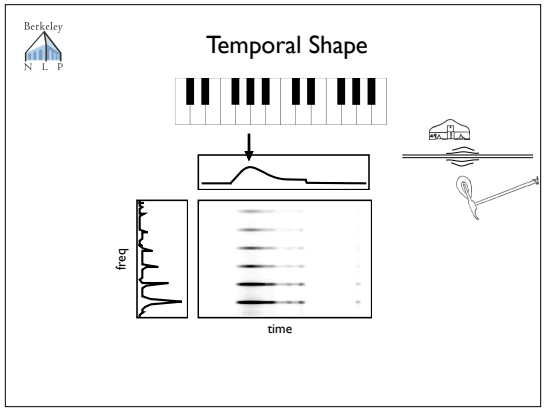


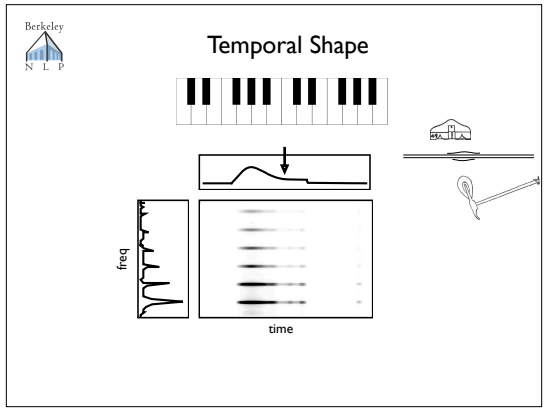
Berkeley N L P

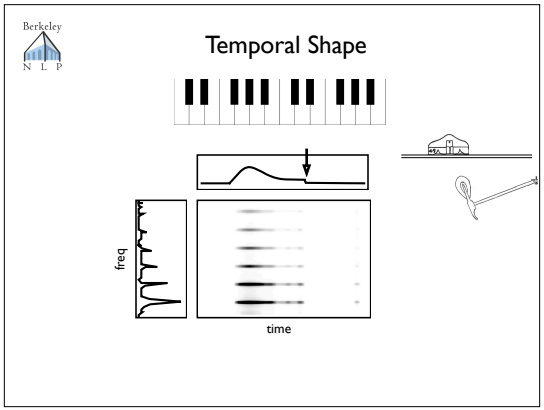
Temporal Shape

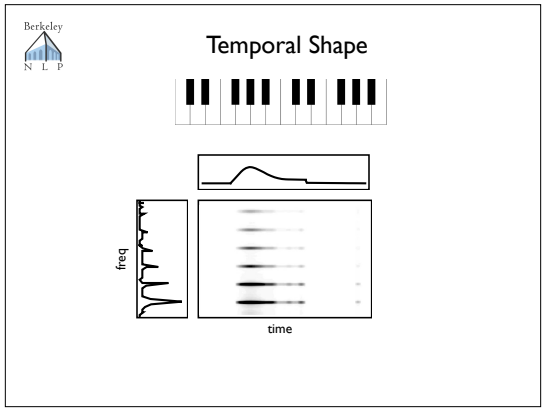


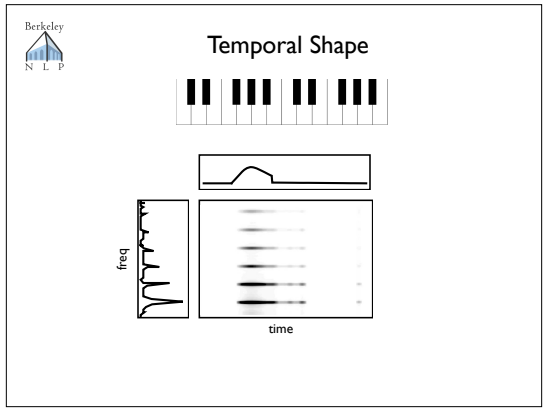

Berkeley
N L P
Temporal Shape
freq
time


Berkeley
N L P
Temporal Shape
freq
time


Berkeley
N L P
Temporal Shape
freq
time

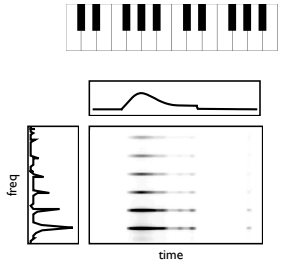

Berkeley
N L P
Temporal Shape
freq
time


Berkeley
N L P
Temporal Shape
freq
time


Berkeley
N L P
Temporal Shape
freq
time

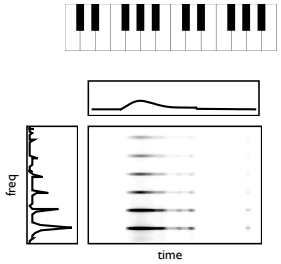
Berkeley N L P

Temporal Shape



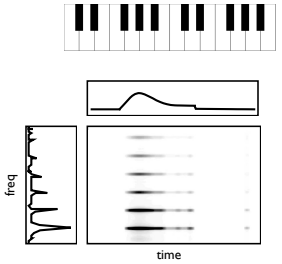
Berkeley N L P

Temporal Shape



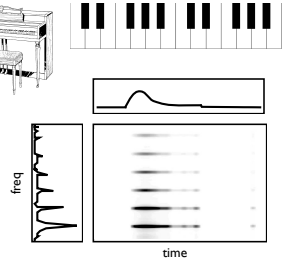
Berkeley N L P

Temporal Shape



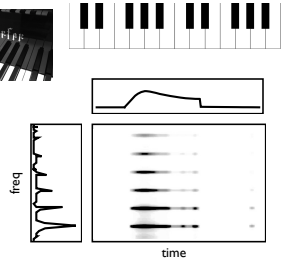
Berkeley N L P

Temporal Shape



Berkeley N L P

Temporal Shape

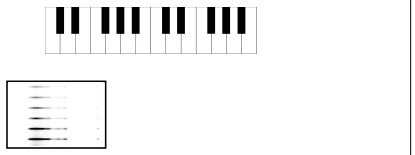


Berkeley N L P

Polyphony

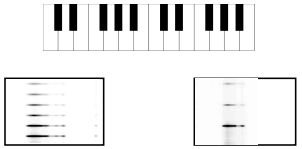
Berkeley N L P

Polyphony



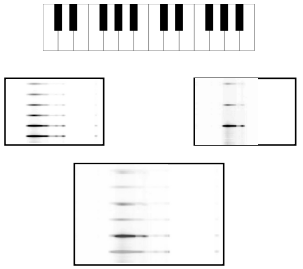
Berkeley N L P

Polyphony



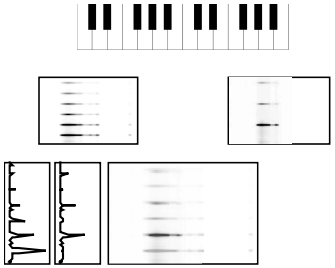
Berkeley N L P

Polyphony



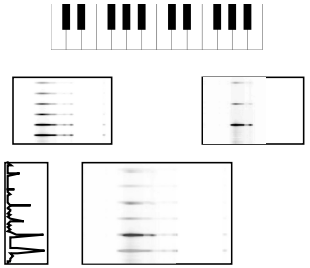
Berkeley N L P

Polyphony



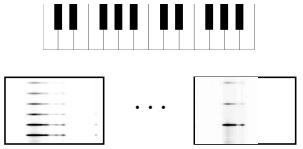
Berkeley N L P

Polyphony

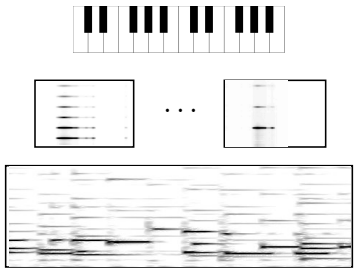


Berkeley N L P

Polyphony

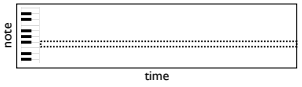


Polyphony

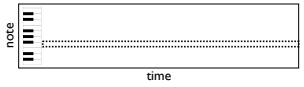


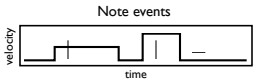
Generative Model

Generative Model

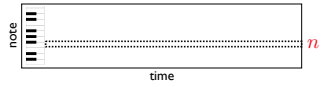


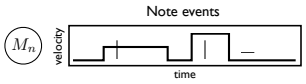
Generative Model



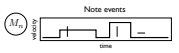


Generative Model

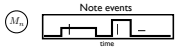




Generative Model

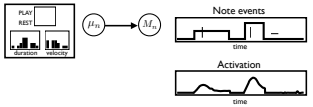


Generative Model

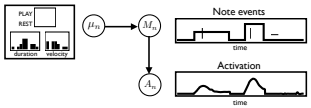


Generative Model

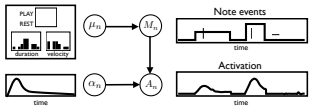
Generative Model



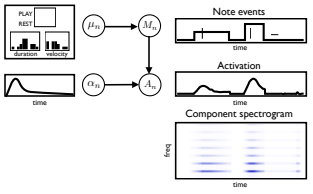
Generative Model



Generative Model

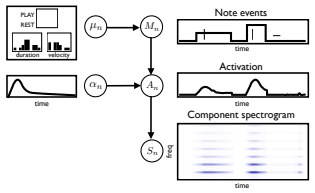


Generative Model



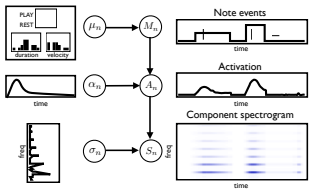
Berkeley N L P

Generative Model



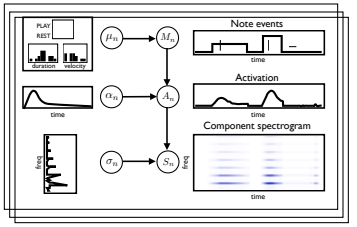
Berkeley N L P

Generative Model



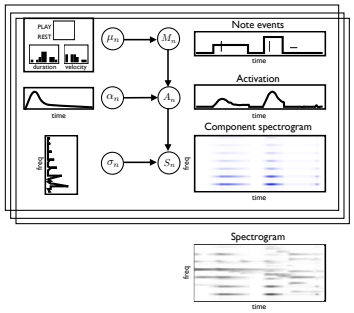
Berkeley N L P

Generative Model



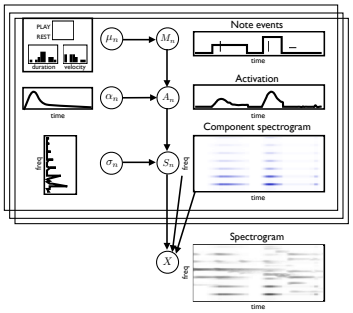
Berkeley N L P

Generative Model



Berkeley N L P

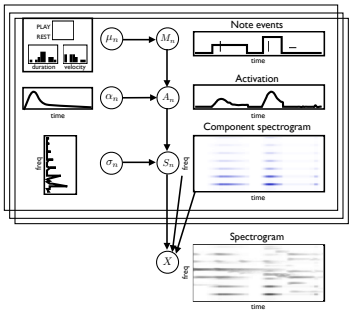
Generative Model



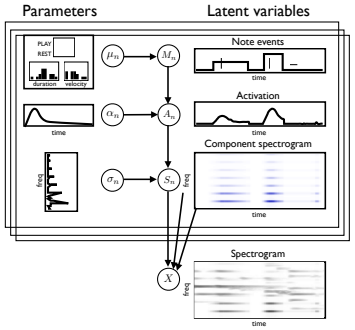
Berkeley N L P

Generative Model

Latent variables

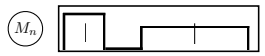


Generative Model



Note Event Model

Note Event Model

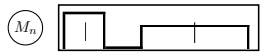


Note Event Model

Note Event Model

Event type

Note Event Model

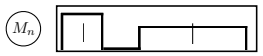


Event type

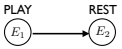
PLAY

E_1

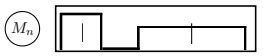
Note Event Model



Event type

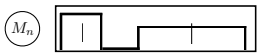


Note Event Model

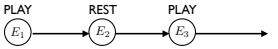


Event type

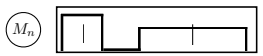
Note Event Model



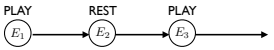
Event type



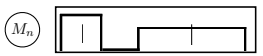
Note Event Model



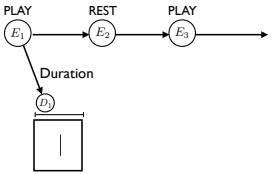
Event type



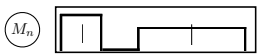
Note Event Model



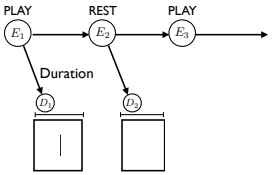
Event type

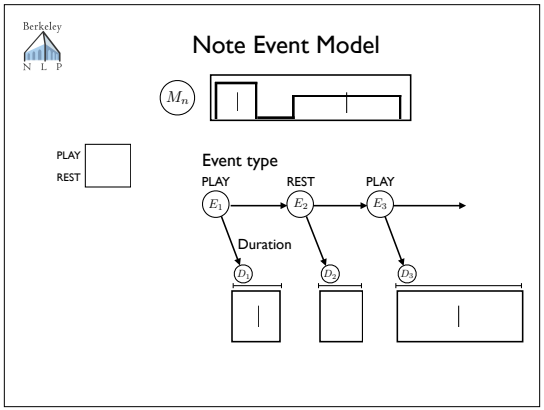


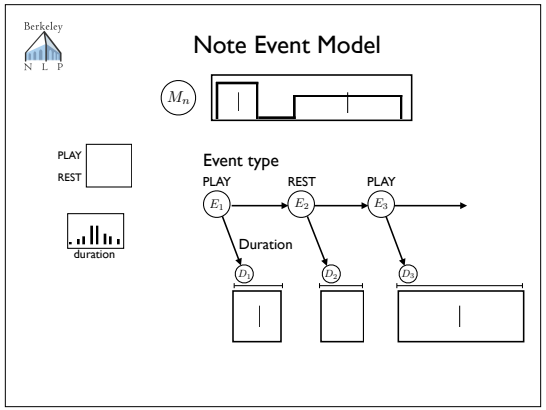
Note Event Model

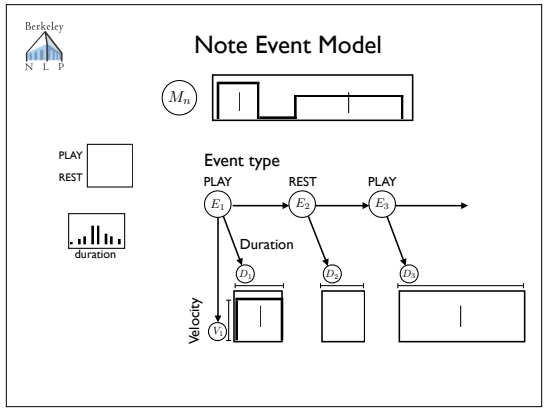


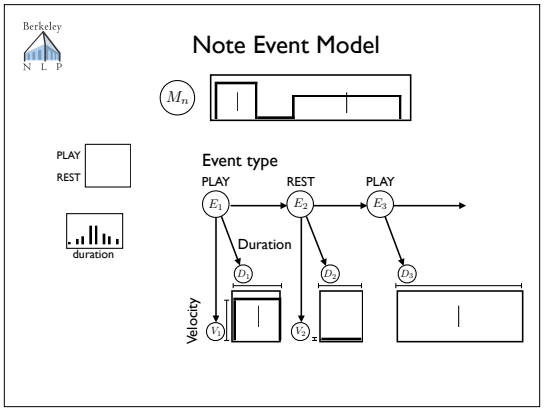
Event type

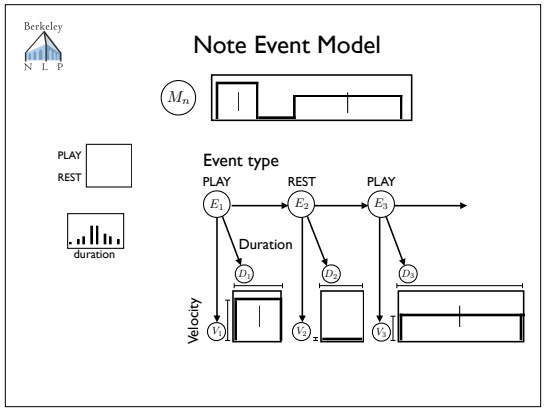


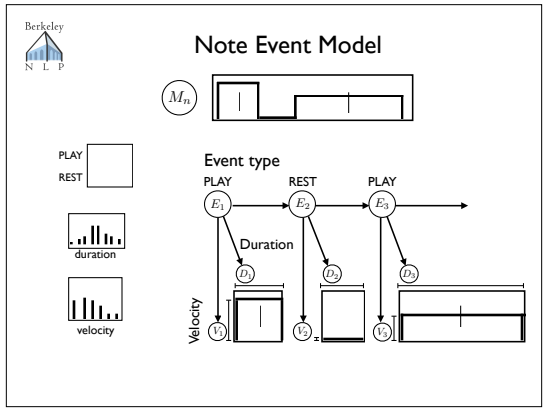

Berkeley
N L P
Note Event Model
M_n
PLAY
REST
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3


Berkeley
N L P
Note Event Model
M_n
PLAY
REST
duration
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3

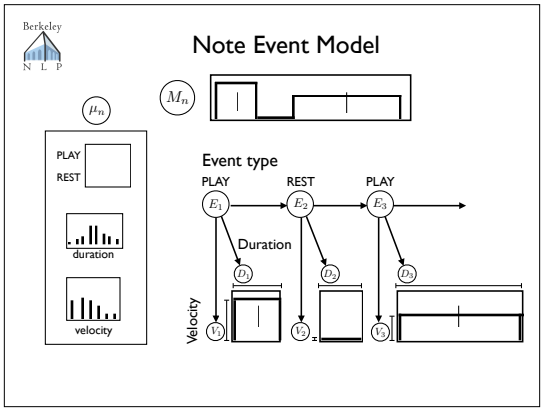

Berkeley
N L P
Note Event Model
M_n
PLAY
REST
duration
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3
Velocity
V_1


Berkeley
N L P
Note Event Model
M_n
PLAY
REST
duration
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3
Velocity
V_1
V_2

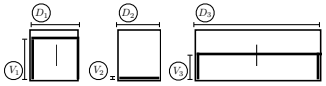

Berkeley
N L P
Note Event Model
M_n
PLAY
REST
duration
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3
Velocity
V_1
V_2
V_3


Berkeley
N L P
Note Event Model
M_n
PLAY
REST
duration
velocity
Event type
PLAY
REST
PLAY
E_1
E_2
E_3
Duration
D_1
D_2
D_3
Velocity
V_1
V_2
V_3

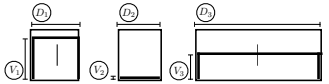
Note Event Model



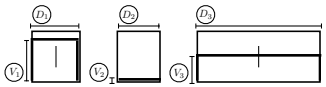
Activation Model



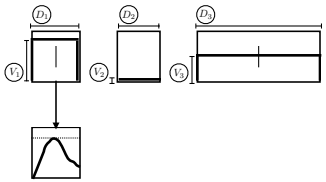
Activation Model



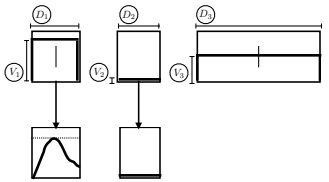
Activation Model

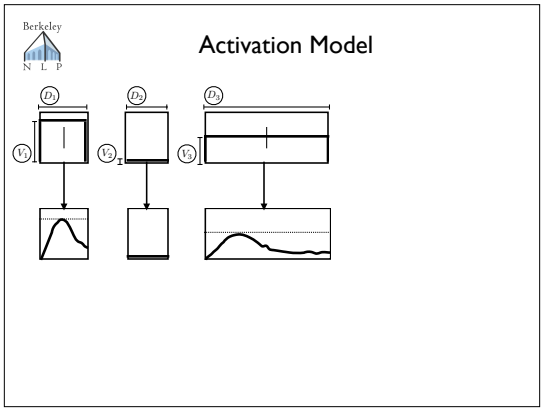


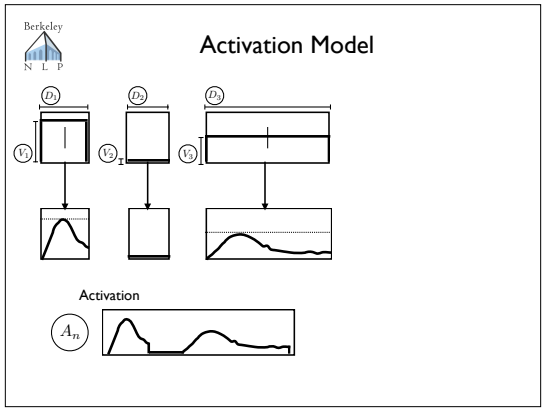
Activation Model

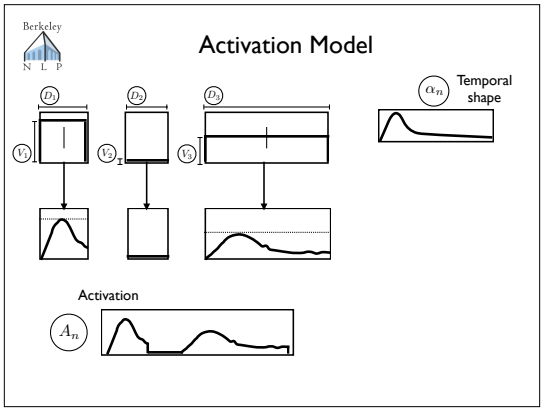


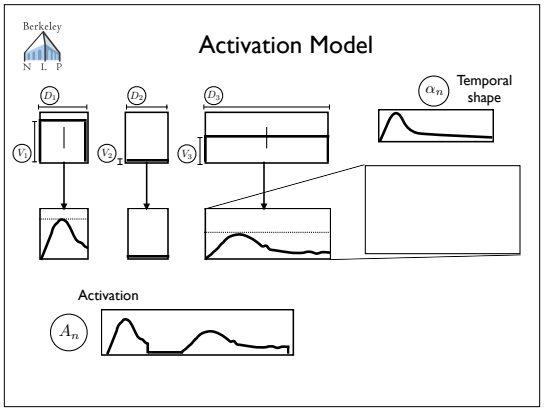
Activation Model

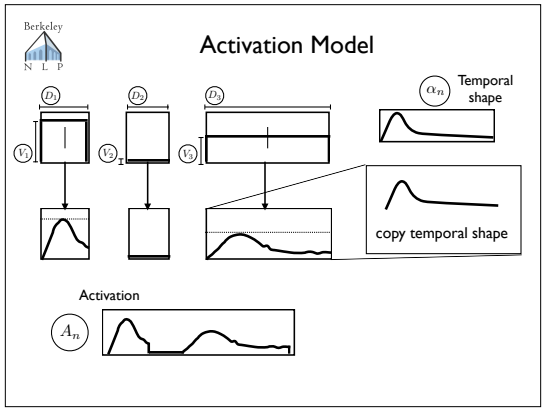


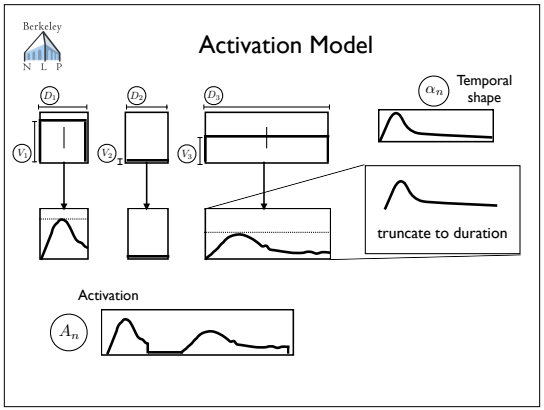

Berkeley
N L P
Activation Model

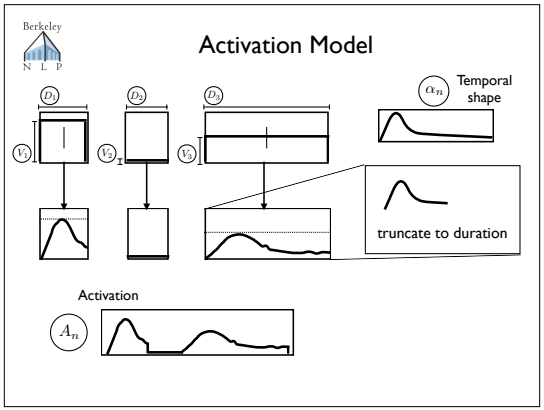

Berkeley
N L P
Activation Model
Activation

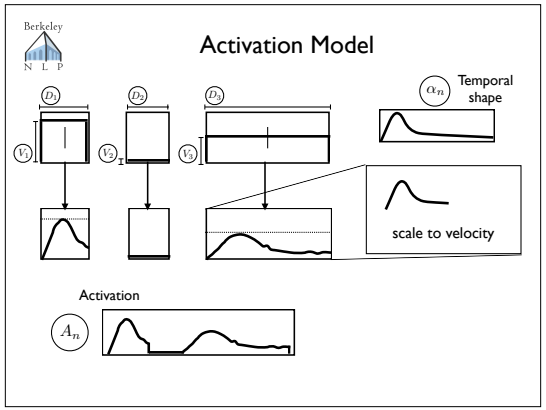

Berkeley
N L P
Activation Model
Temporal shape
Activation

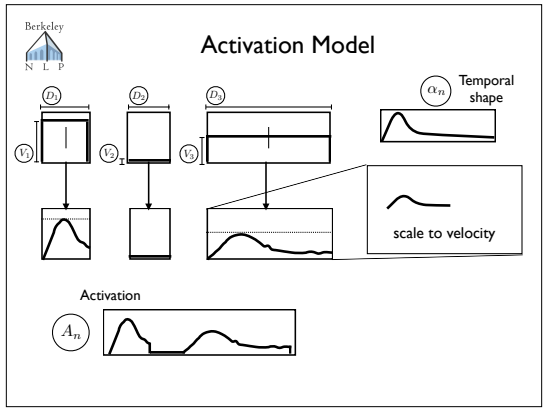

Berkeley
N L P
Activation Model
Temporal shape
Activation

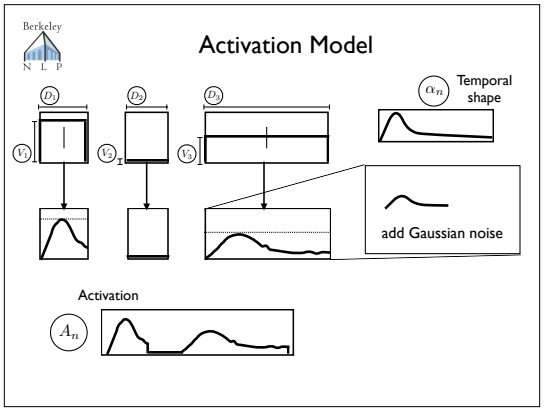

Berkeley
N L P
Activation Model
Temporal shape
copy temporal shape
Activation

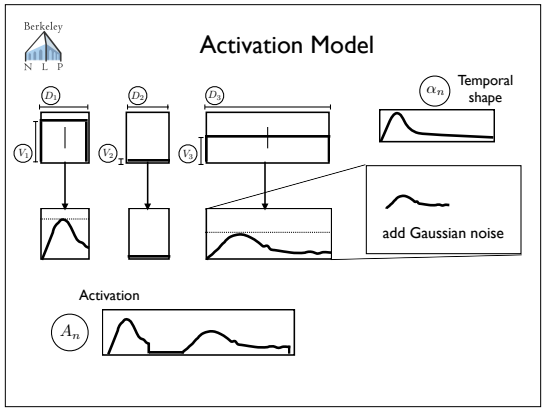

Berkeley
N L P
Activation Model
Temporal shape
truncate to duration
Activation

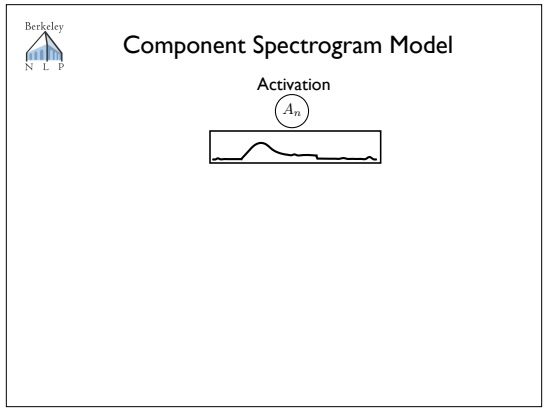

Berkeley
N L P
Activation Model
D_1
D_2
D_3
α_n
Temporal shape
V_1
V_2
V_3
truncate to duration
Activation
A_n


Berkeley
N L P
Activation Model
D_1
D_2
D_3
α_n
Temporal shape
V_1
V_2
V_3
scale to velocity
Activation
A_n

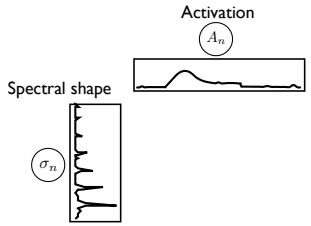

Berkeley
N L P
Activation Model
D_1
D_2
D_3
α_n
Temporal shape
V_1
V_2
V_3
scale to velocity
Activation
A_n


Berkeley
N L P
Activation Model
D_1
D_2
D_3
α_n
Temporal shape
V_1
V_2
V_3
add Gaussian noise
Activation
A_n

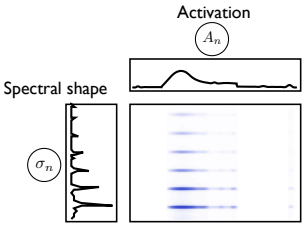

Berkeley
N L P
Activation Model
D_1
D_2
D_3
α_n
Temporal shape
V_1
V_2
V_3
add Gaussian noise
Activation
A_n


Berkeley
N L P
Component Spectrogram Model
Activation
A_n

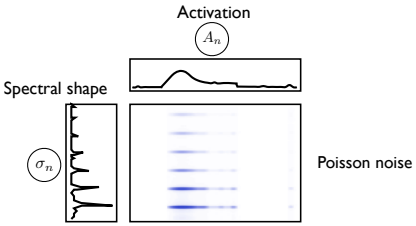
Component Spectrogram Model



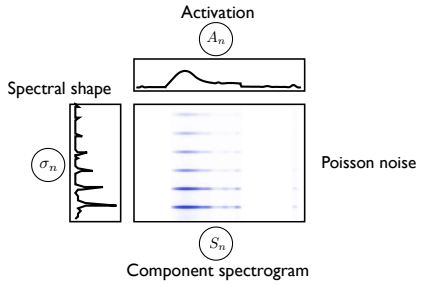
Component Spectrogram Model



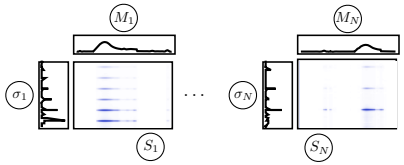
Component Spectrogram Model



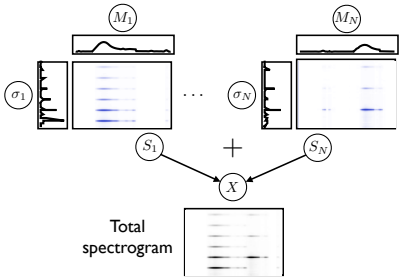
Component Spectrogram Model



Total Spectrogram Model

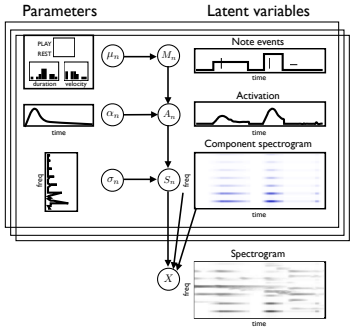


Total Spectrogram Model



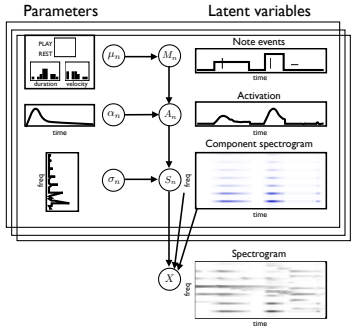
Berkeley N L P

Learning and Inference



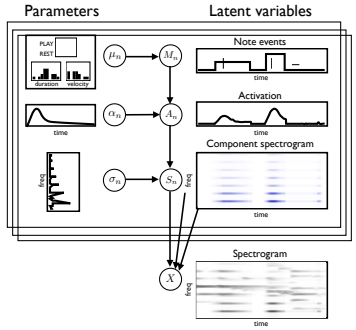
Berkeley N L P

Learning and Inference



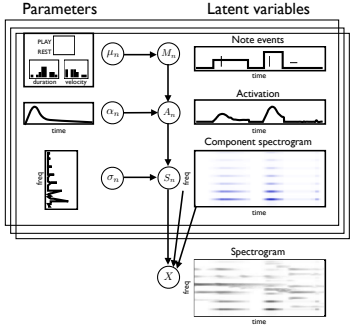
Berkeley N L P

Learning and Inference



Berkeley N L P

Learning and Inference



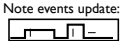
Berkeley N L P

Learning and Inference

Berkeley N L P

Learning and Inference

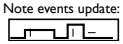
Note events update:



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

Learning and Inference

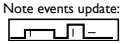
Note events update:



$M|A, \alpha, \mu$ Semi-Markov dynamic program

Learning and Inference

Note events update:



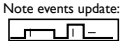
$M|A, \alpha, \mu$ Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$

Learning and Inference

Note events update:



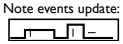
$M|A, \alpha, \mu$ Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$ Closed form update

Learning and Inference

Note events update:

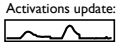


$M|A, \alpha, \mu$ Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$ Closed form update

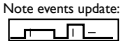
Activations update:



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

Learning and Inference

Note events update:

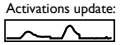


$M|A, \alpha, \mu$ Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$ Closed form update

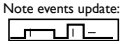
Activations update:



$A|M, X, \alpha, \sigma$ Exponentiated gradient ascent

Learning and Inference

Note events update:

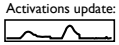


$M|A, \alpha, \mu$ Semi-Markov dynamic program

Temporal shapes update:

$\alpha|A, M$ Closed form update

Activations update:



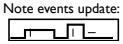
$A|M, X, \alpha, \sigma$ Exponentiated gradient ascent

Spectral shapes update:

$\sigma|A, X$

Learning and Inference

Note events update:

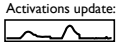


$M|A, \alpha, \mu$ — Semi-Markov dynamic program

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

Evaluation

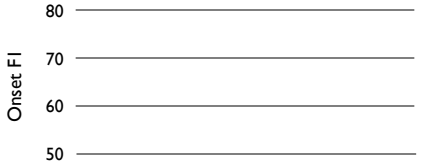
Onset F1

Evaluation

Onset F1

Results

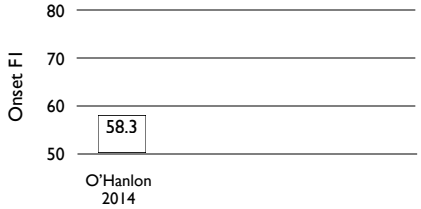
MAPS Corpus



[Valentin et al. 2010]

Results

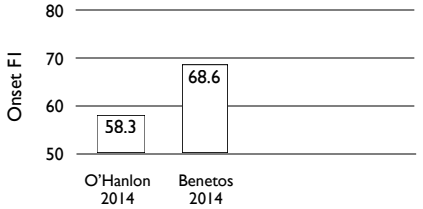
MAPS Corpus



[Valentin et al. 2010]

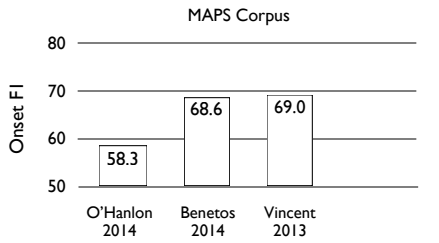
Results

MAPS Corpus



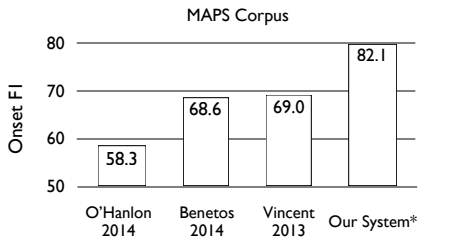
[Valentin et al. 2010]

Results



[Valentin et al. 2010]

Results

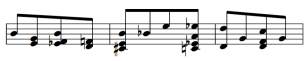


[Valentin et al. 2010]
[Berg-Kirkpatrick et al. 2014]

Transcription

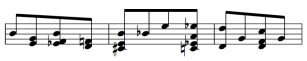
Transcription

Reference

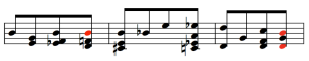


Transcription

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