Music Segmentation

Anja Volk

Sound and Music Technology

Jan 7, 2020

Today

- Introduction to music segmentation in MIR
  - What are musical segments
  - Perceptual cues to detect musical segments
- Computer Models of Melody Segmentation

Recap

- Transcription
  - What is transcription?
  - Why? What are user scenarios?
    - Query by humming
    - Score following (automatic accompaniment)
      - Check out Chris Raphael's "Music Plus One" project, Indiana University
      - https://www.youtube.com/watch?v=cRQ6jmZzXJE
      - Enhancing concert audience's experiences
      - Expressive Performance research: compare different performances
      - Computational Ethnomusicology: investigate oral traditions
  - What subtasks belong to transcription?
    - (multiple) F0 estimation
    - melody transcription
    - onset detection
    - chord detection
    - Instrument recognition
  - State of the art
    - no general solutions, but for controlled situations

Today

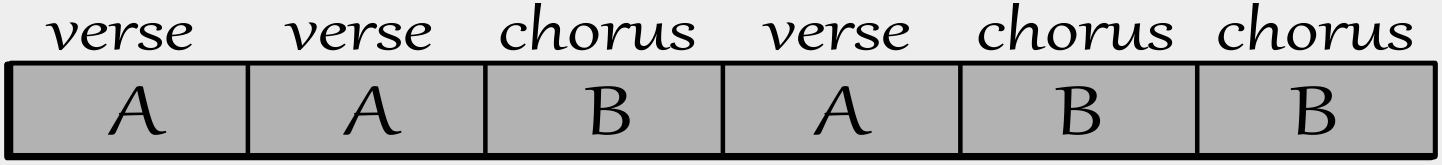
- Segmentation
  - What are musical segments
  - Perceptual cues to detect musical segments
  - Modeling segmentation of melodies (symbolic encodings)
  - Segmentation in audio domain
    - Student presentation: melodic phrases in folk singing recordings

Music Segmentation: MIR Task Definition

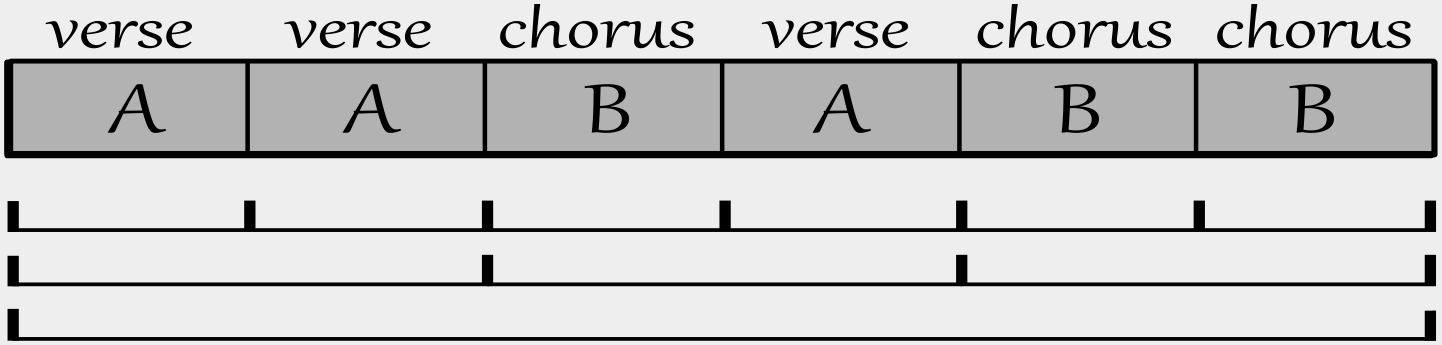
Music Segmentation: MIR Task Definition

Music Segmentation: MIR Task Definition

Music Segmentation: MIR Task Definition



Music Segmentation: MIR Task Definition



Music Segmentation: Why?

- Music Cognition
  - fundamental perceptual process
- MIR
  - music visualisation and summarisation
  - indexing for search and browsing of large music collections
  - finding the most salient part
- Games
  - material for automatic composition/improvisation
  - markers music-to-video/text synchronization
  - horizontal re-mixing

Music Segmentation: Formalization

Given a sequence of 'musical events' $e = \langle e_1 \ldots e_n \rangle$

A segmentation s of e is defined by $k + 1$ boundary locations

$1=b_1 < b_2 < \cdots < b_k < b_{k+1}=n+1$, yielding

$s = \langle s_1 \ldots s_k \rangle$, where $s_j = [e_{bj} \ldots e_{bj+1}), \forall j = 1,\ldots,k$.

Where $e=\cup^k_{j=1} s_j$ and $s_i \cap s_j = \emptyset, \forall i,j \in [1,k] \wedge i \neq j$

Music Segmentation: Musical Events?

- Audio waveform (mp3, wav, etc.)
  - $e_i \in R^d, \forall i = 1,...,n$ are **audio windows** often lasting a few tens or hundreds of milliseconds represented using *d*-dimensional feature vectors (commonly MFCCs or Chroma).
- Symbolic music encoding(Midi, MusicXML, etc.)
  - $e_i \in \xi$, where ξ is a **finite and discrete attribute space** approximating the attribute space of music theoretic notes, i.e. the space defined by ξ is at least onset $\otimes$ offset $\otimes$ pitch, with $\otimes$ denoting the Cartesian product.

How do listeners segment?

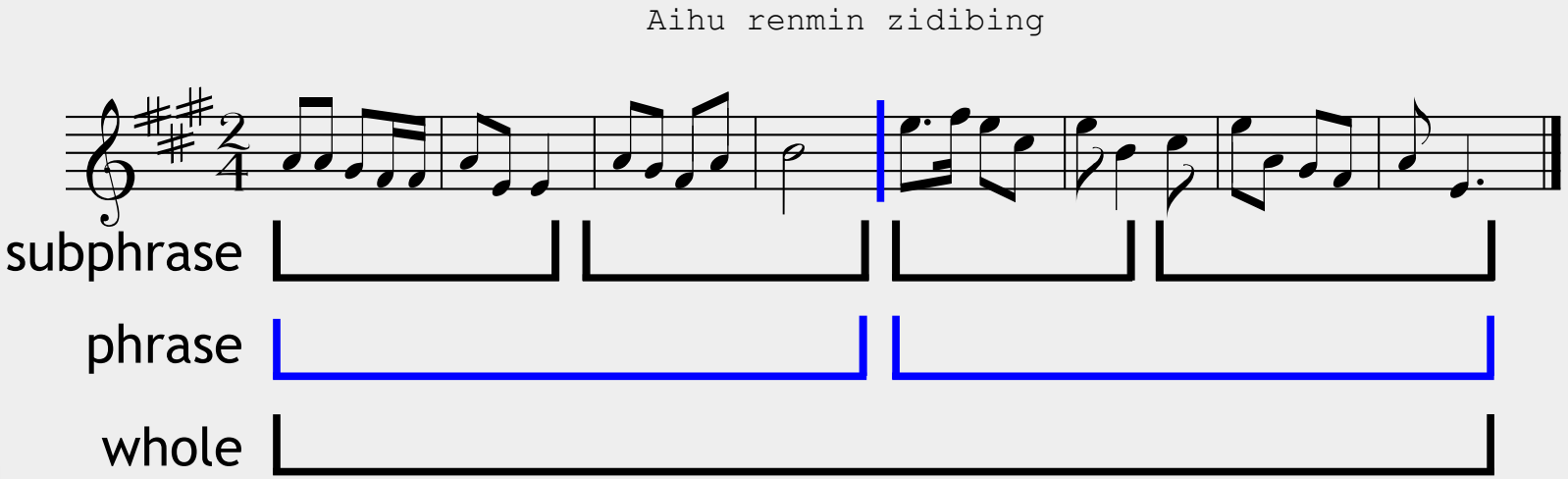
Segmentation of symbolically-encoded melodies

Why Melodies?

- Interesting musical 'object'
  - cross-cultural
  - memorable
- Offers natural constraints to study segmentation
  - Monophonic
  - manageable size

Melody Segmentation

- Melodic Phrases and Subphrases
  - units of information? . . . units of meaning?



Melody Segmentation: What is a Phrase?

- Western art music
  - consecutive notes "expressing a complete musical thought"(3)
  - "containing significant tonal motion"(4).
  - roughly 4-8 measures in length (5).
- Popular music
  - '[a melodic segment] end[s] normally coinciding with the taking of breath'(1)
  - 'A vocal melody where phrase length is based on the singer's breaths, and an instrumental accompaniment where units [...] based on the repetition of chord patterns' (2)

[1] A. Moore: The So-Called 'Flattened Seventh' in Rock Music, *Popular Music*, Vol. 14, No. 1, pp. 185–201, 1995.
[2] K. Stephenson. *What to Listen for in Rock*, Yale University Press, New Haven, 2002.
[3] L. Stein. *Structure and Style: The Study and Analysis of Musical Forms*. Summy-Birchard Company, 1979.
[4] W. Rothstein. *Phrase rhythm in tonal music*. Schirmer Books New York, 1989.
[5] David Temperley. End-accented phrases: An analytical exploration. *Journal of Music Theory*, 47(1):125–154, 2003.

Melody Segmentation: Perceptual Cues

Stockhausen´s klaviersück IX 10 strongest boundaries	CUES	Mozart´s Fantasie 6 strongest boundaries
https://www.youtube.com/watch?v=ny8ZXXRBIns		https://www.youtube.com/watch?v=yOgrqZs2L2k
(chords-to-melody, pitch content, block chords) new material		new material (lyrical, dramatic, end of cadenza)
(chordal, chromatic run, coda) return of material		change of texture (thicker, thinner)
(expansion, jump) change of register		change of tempo
change of rhythm		change of register
change of dynamic		change of dynamic
(silence) pause		change of key
start of development		change of harmony
change of articulation		change of meter
change of texture		change of rhythm
change of pitch content		change of melody
relaxation of tension		
introduction of trill		
change of tempo		
(piano tone)change of timbre		

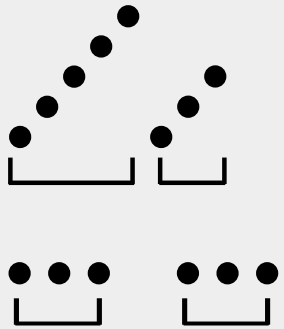
Eric F Clarke and Carol L Krumhansl. Perceiving musical time. *Music Perception*, pages 213–251, 1990.

Melody Segmentation: Cue Examples

gap detection
repetition detection
contrast detection
closure detection

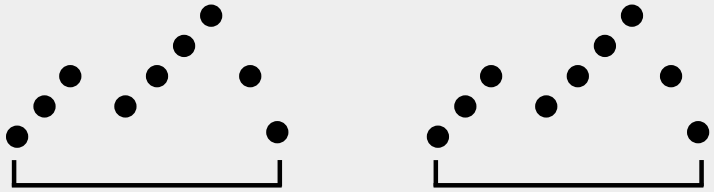
Melody Segmentation: Cue Examples

gap detection
repetition detection
contrast detection
closure detection



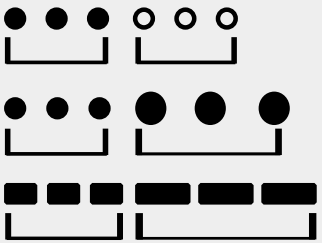
Melody Segmentation: Cue Examples

gap detection
repetition detection
contrast detection
closure detection



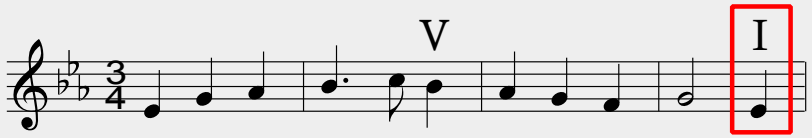
Melody Segmentation: Cue Examples

gap detection
repetition detection
contrast detection
closure detection



Melody Segmentation: Cue Examples

gap detection
repetition detection
contrast detection
closure detection



Computational models

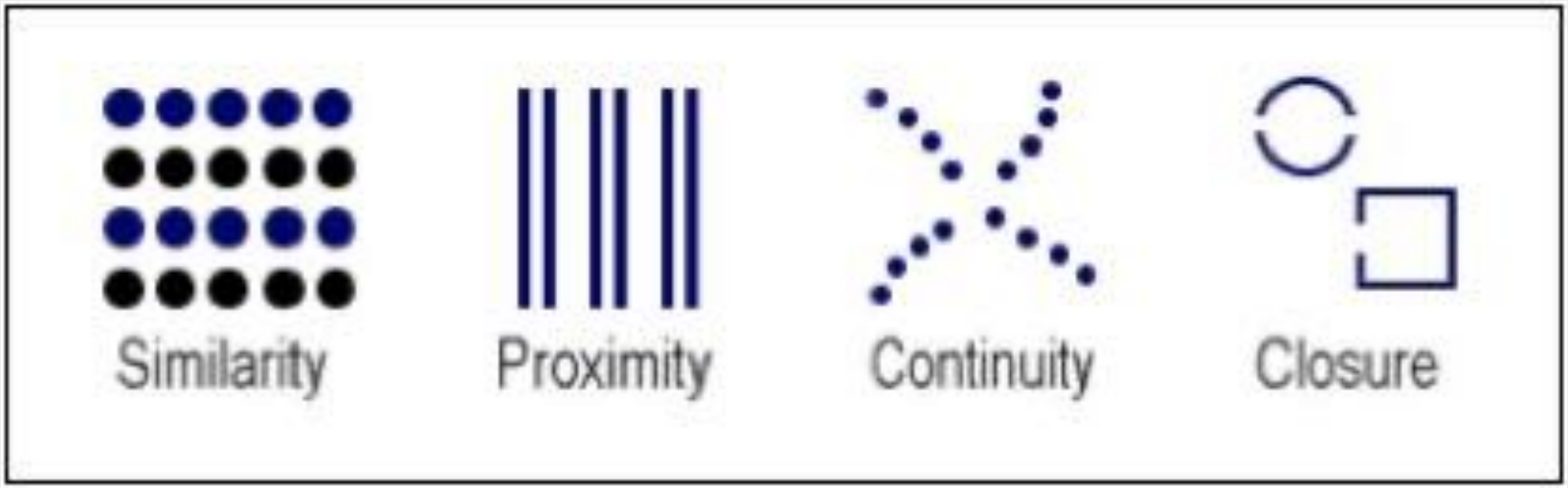
- **development**: +30 years
  number of models: >30
  most successful: Gestalt-based, $F\,1 = 0.60 - 0.66$

Approaches

Approaches

- Most Researched
  - **Type:** Gestalt Based (1980-today)
    - **Principle:** *proximity* as (Local) *discontinuity* detection
  - **Type:** Repetition Based (~1998-today)
    - **Principle:** *repetition* as *string matching*
  - **Type:** Expectation Based (~2002s-today)
    - **Principle:** *closure* as *information-theoretic surprise*
- Others
  - Using probabilistic grammars, expert systems (artificial intelligence), connectivist approaches

Gestalt-Based Models



… goes back to Gestalt psychologists in 1920s

Gestalt-Based Models

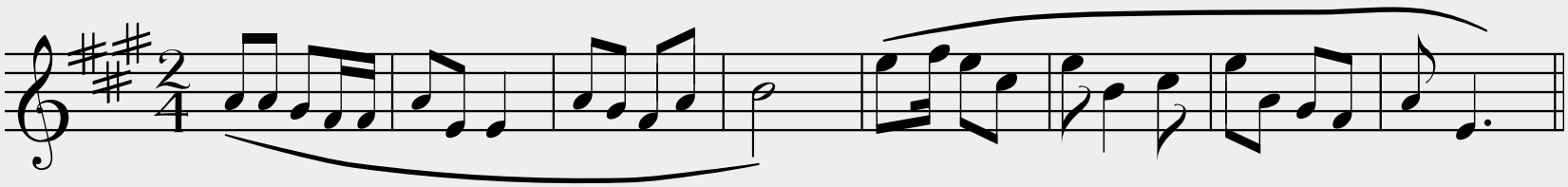
- Knowledge (rule)-based perspective
  - opposed to data based
- quantification of Gestalt principles
- use system of preference rules
- Gestalt proximity is modeled as discontinuity detection

Gestalt-Based Models

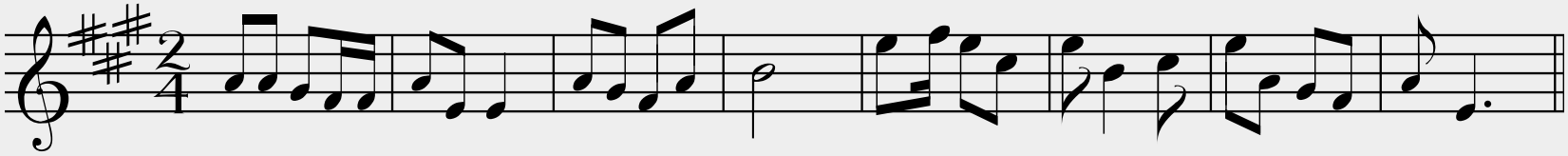
- knowledge-based perspective
- quantification of Gestalt principles
- use system of preference rules
- **Gestalt proximity is modeled as discontinuity detection**

Local Discontinuity Detection: an Example

Aihu renmin zidibing



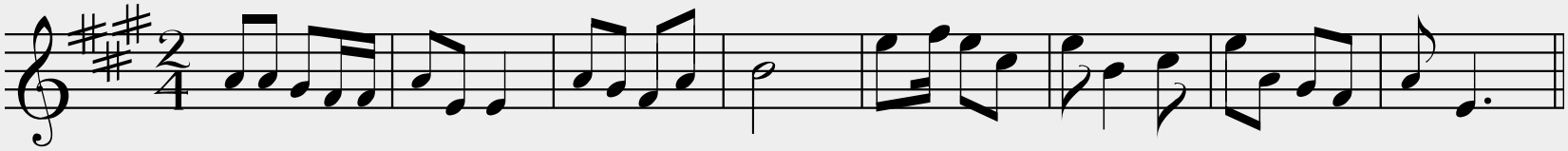
Local Discontinuity Detection: Input



e_1 , , , e_i , , , e_N

onset:	24	48	72	84	96	120	, , ,	, , ,	2322
pitch:	69	69	68	66	66	69	, , ,	, , ,	64

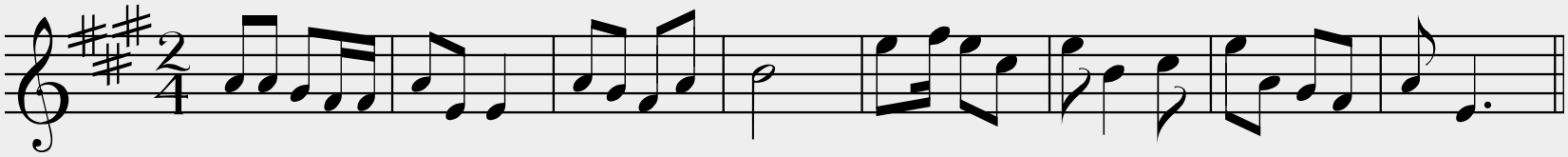
Local Discontinuity Detection: Input

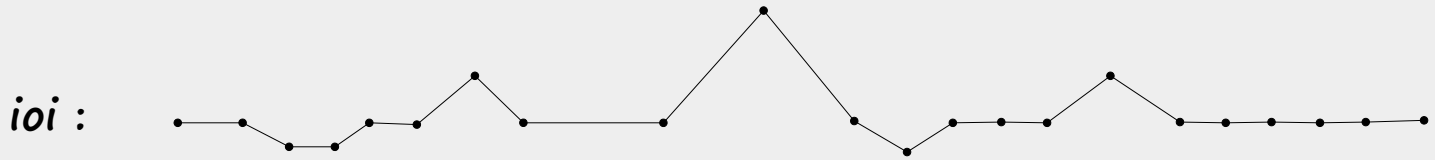


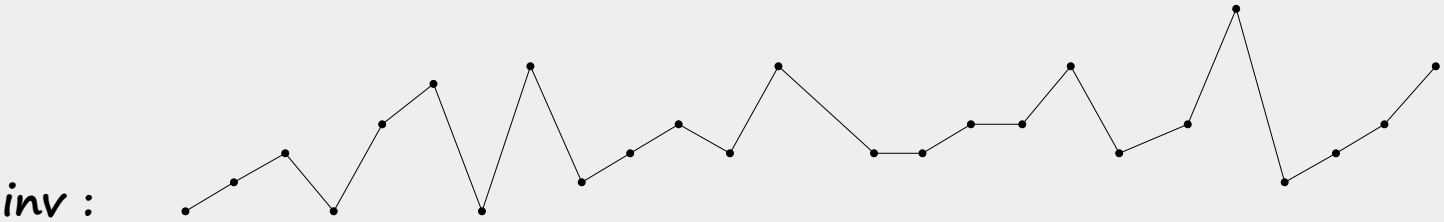
ioi :	.5	.5	.5	.25	.25	.5	.5	1	.5	.5	.5	.5	2	.75	.25	.5	.5	.5	1	.5	.5	.5	.5	.5	.5
p-inv :	0	1	2	0	3	5	0	5	1	2	3	2	5	2	2	3	3	5	2	3	7	1	2	3	5

ioi – inter onset interval
p-inv – pitch interval

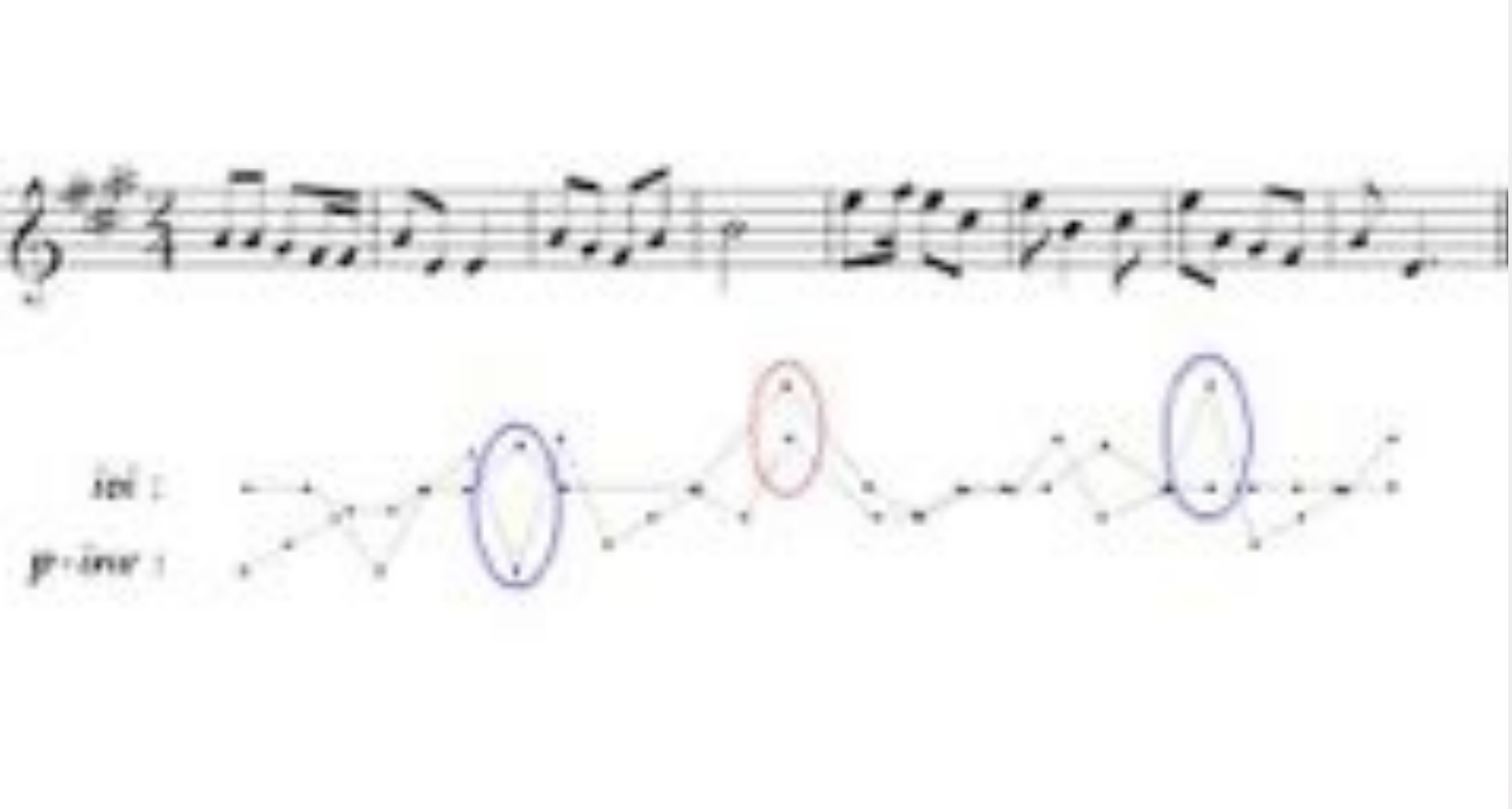
Local Discontinuity Detection: Profiles



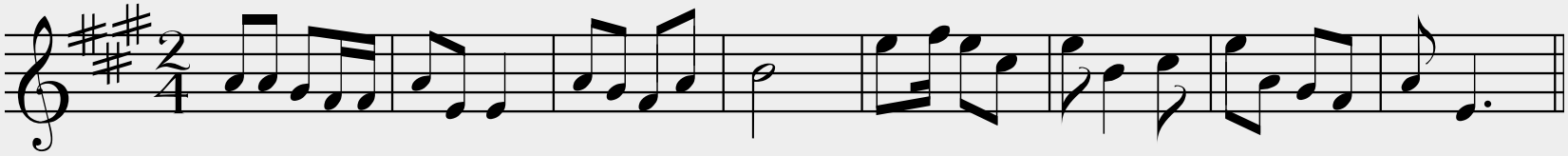




Local Discontinuity Detection: Profiles



Local Discontinuity Detection: Output



Output : 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0

Local Discontinuity Detection: Models

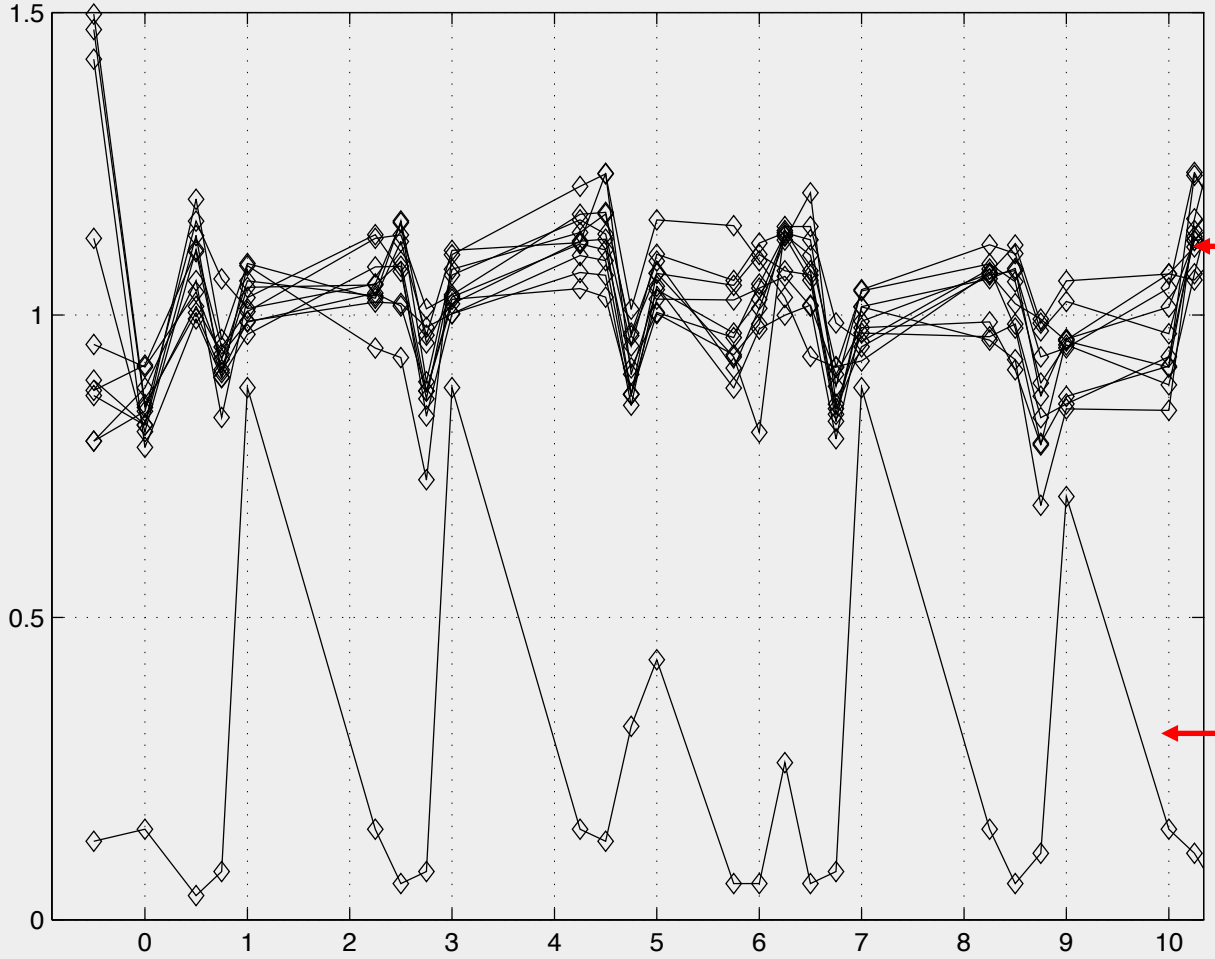
- *Some* computer models of melodic discontinuity ('gap') detection:
  - Local Boundary Detection Model (LBDM)(6)
  - Temporal Gestalt Units (TGU)(9)
  - Piece-Sensitive Segmentation (PSS)(10)

[6] E. Cambouropoulos. The local boundary detection model (LBDM) and its application in the study of expressive timing. In *Proceedings of the International Computer Music Conference (ICMC01)*, pages 232–235, 2001.

[9] J. Tenney and L. Polansky. Temporal gestalt perception in music. Journal of Music Theory, 24(2):205–241, 1980.

[10] D. Lefkowitz and K. Taavola. Segmentation in music: generalizing a piece-sensitive approach. *Journal of Music Theory*, 44(1):171–229, 2000.

Local Discontinuity Detection: LBDM



IOI deviation curves from 10 performances

LBDM

https://www.youtube.com/watch?v=mpiJbQvBP8A

Local Discontinuity Detection: LBDM

Algorithm

1. given: pitch $p[i]$, onset $on[i]$, offset $off[i]$, for $i = 1,...,n$

2. compute pitch *cp*, inter-onset-intervals *ioi*, and rest (offset-to-onset-interval) *ooi* profiles

3. for each interval x, compute boundary strength s

   $s[i] = x[i] \cdot (r[i-1] + r[i])$ with r: degree of change

   $r[i] = |x[i]-x[i+1]| \,/\, x[i]+x[i+1]$

4. computed combined boundary strength profile *bsp* as:

   $bsp = w_{cp}s_{cp}[i] + w_{ioi}s_{ioi}[i] + w_{ooi}s_{ooi}[i]$

5. local peaks in *bsp* indicate boundaries

Gestalt Models: Assumptions

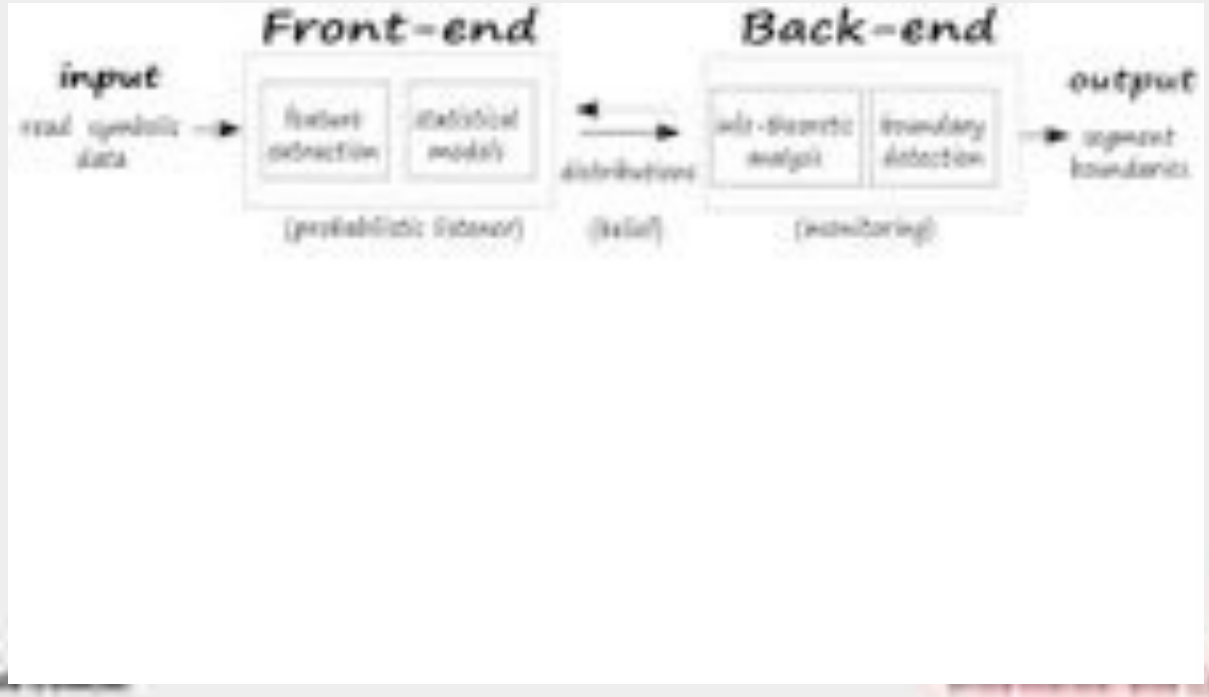
- discontinuity is **relevant** for boundary perception
- discontinuity can be treated as a **local phenomenon**
- discontinuity is **universal/idiom-independent**

Expectation Based Models: Overview

- mostly data-driven perspective
  - though there exist a rule-model on expectation by Narmour
- information-theoretic account of surprise
- use a probabilistic model of melody continuation + information theory analysis
- closure is commonly modeled as surprise detection

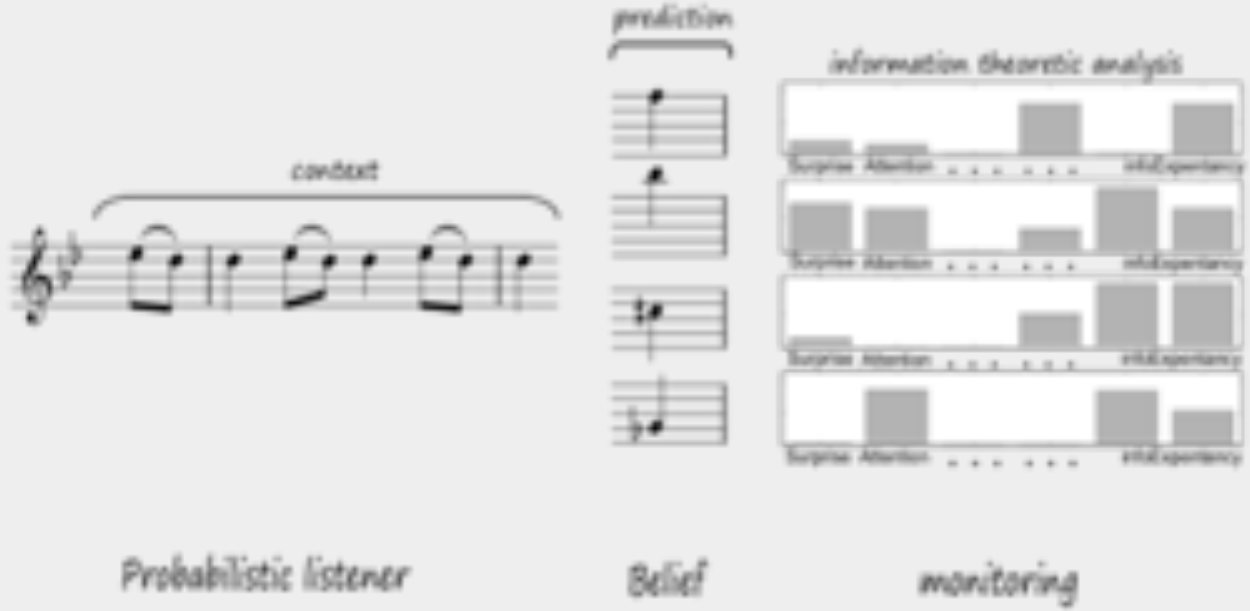
Expectation Based: Overview

- Common Framework:
  - **melody prediction** front-end
  - **information-theoretic** back-end



Expectation Based: Melody Prediction

- … perhaps more intuitively



Expectation Based: Melody Prediction

- Many different models for melody prediction
  - Markov Models (11;12)
  - Deep-Belief Networks (14)
  - Recurrent Neural Networks (13)
  - Self-Organizing Maps (15)

[11] M. Pearce and G. Wiggins. Expectation in melody: The influence of context and learning. Music Perception, 23(5):377–405, 2006.

[12] M. Pearce. *The construction and evaluation of statistical models of melodic structure in music perception and composition*. PhD thesis, Department of Computing, City University, 2005.

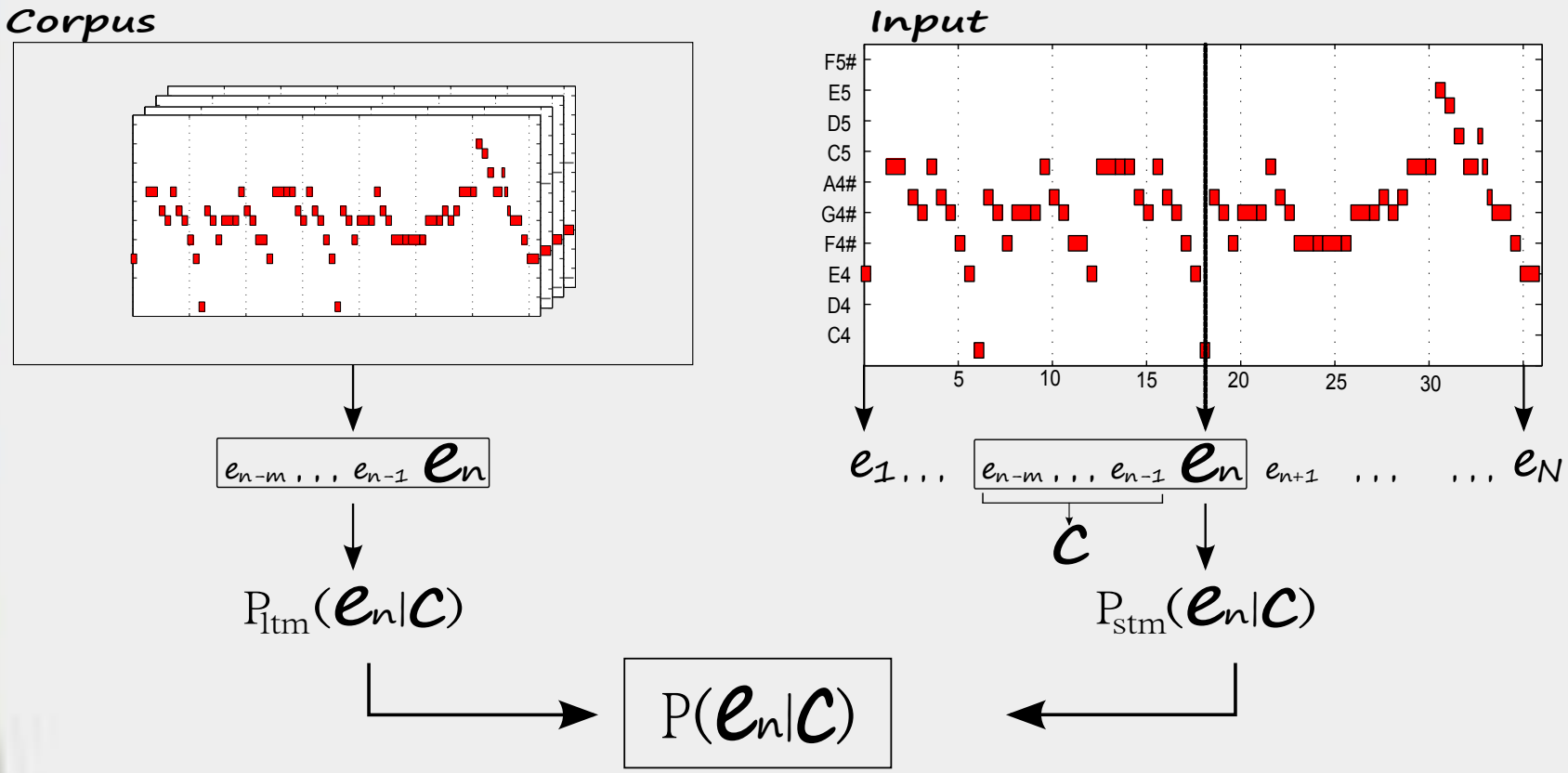
[13] G. Cox. On the relationship between entropy and meaning in music: An exploration with recurrent neural networks. In *Proceedings of the 32nd Annual Cognitive Science Society. Austin TX: CSS*, 2010.

[14] S. Cherla, et al. A Distributed Model for Multiple-Viewpoint Melodic Prediction. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, 2013.

[15] S. Harford. Automatic segmentation, learning and retrieval of melodies using a self-organizing neural network. In *Proceedings of International Conference on Music Information Retrieval, MD, Baltimore*, 2003.

Expectation Based: Melody Prediction

Using Markov Models:



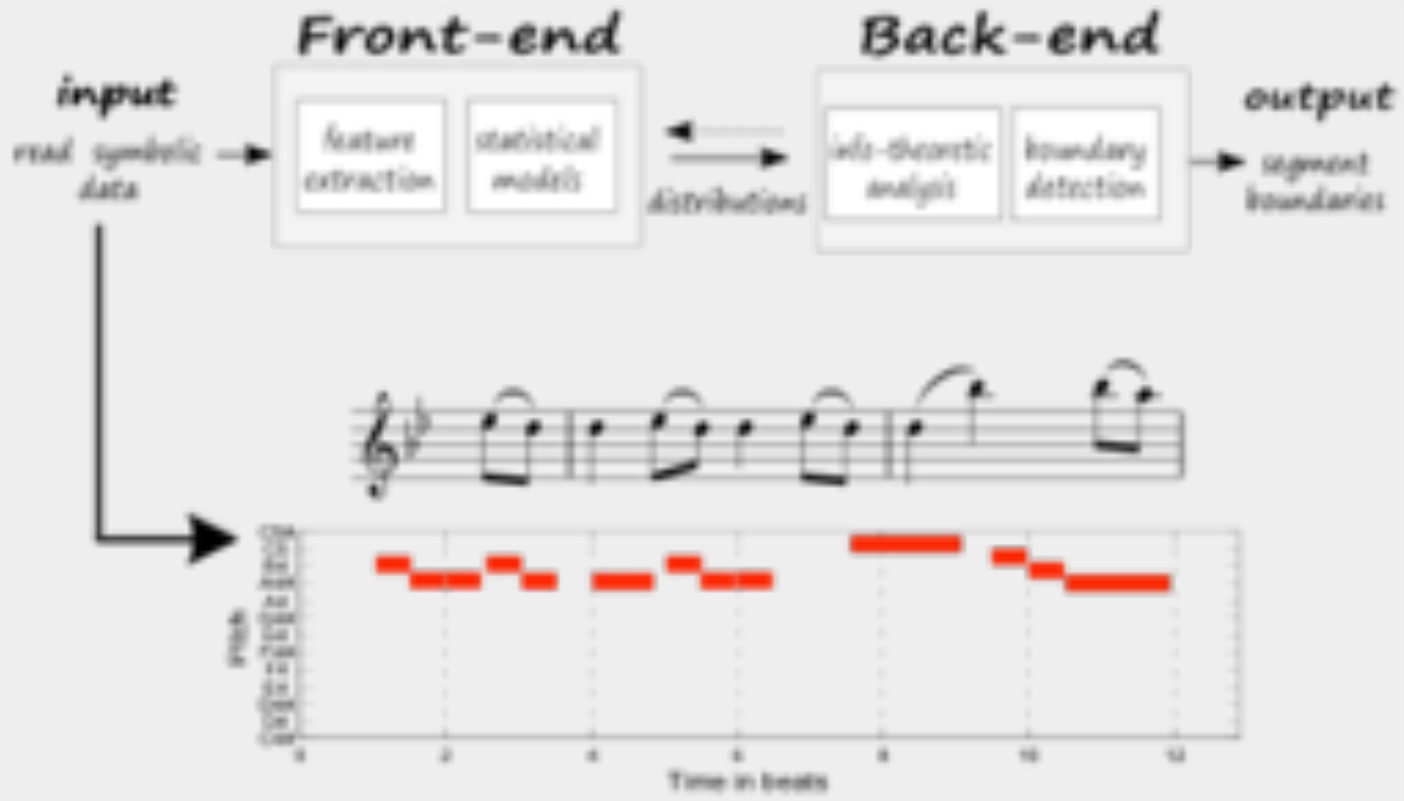
Expectation Based Boundary Detection

- A computer model of expectation-based boundary detection:
  - Information Dynamics of Music (IDyOM)(7)

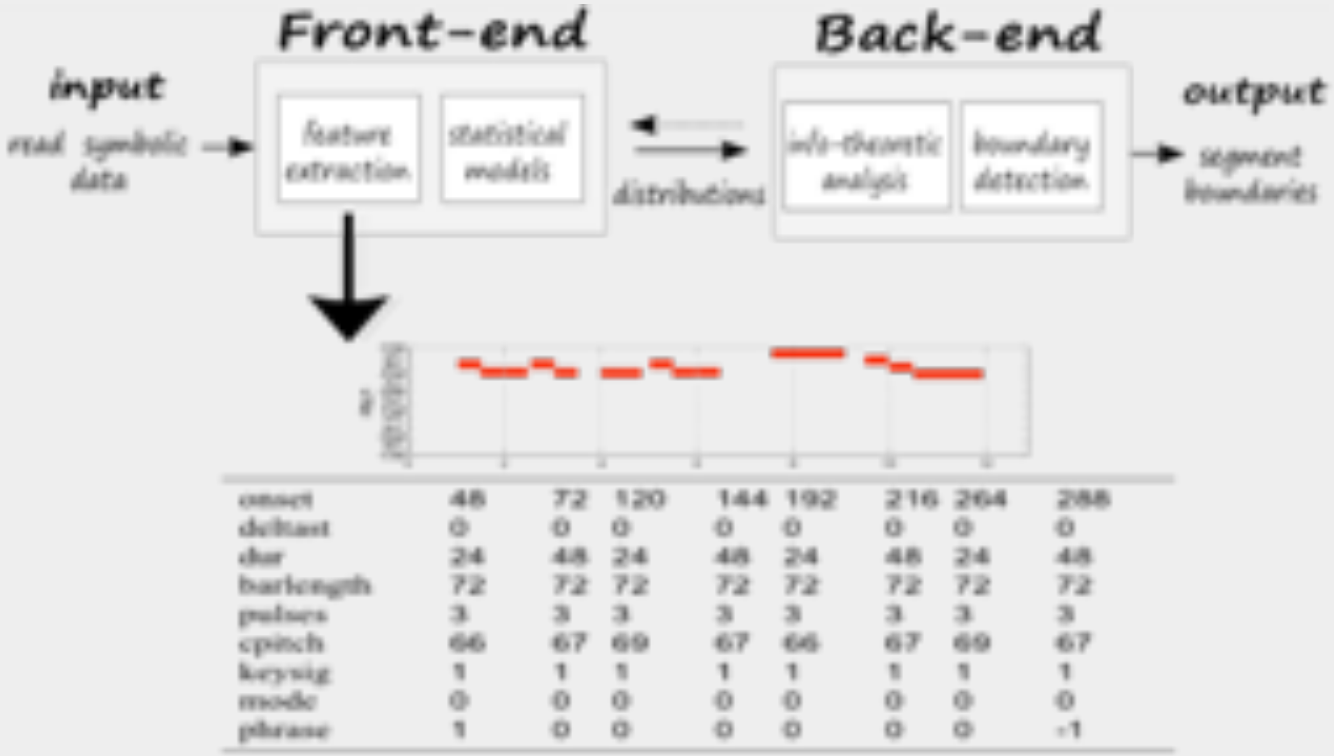
Most other existing computer models using this approach at present constitute a 'proof of principle'

[7] M. Pearce, D. Müllensiefen, and G. Wiggins. A comparison of statistical and rule-based models of melodic segmentation. In Proceedings of the Ninth International Conference on Music Information Retrieval, pages 89–94, 2008.

Expectation Based Boundary Detection

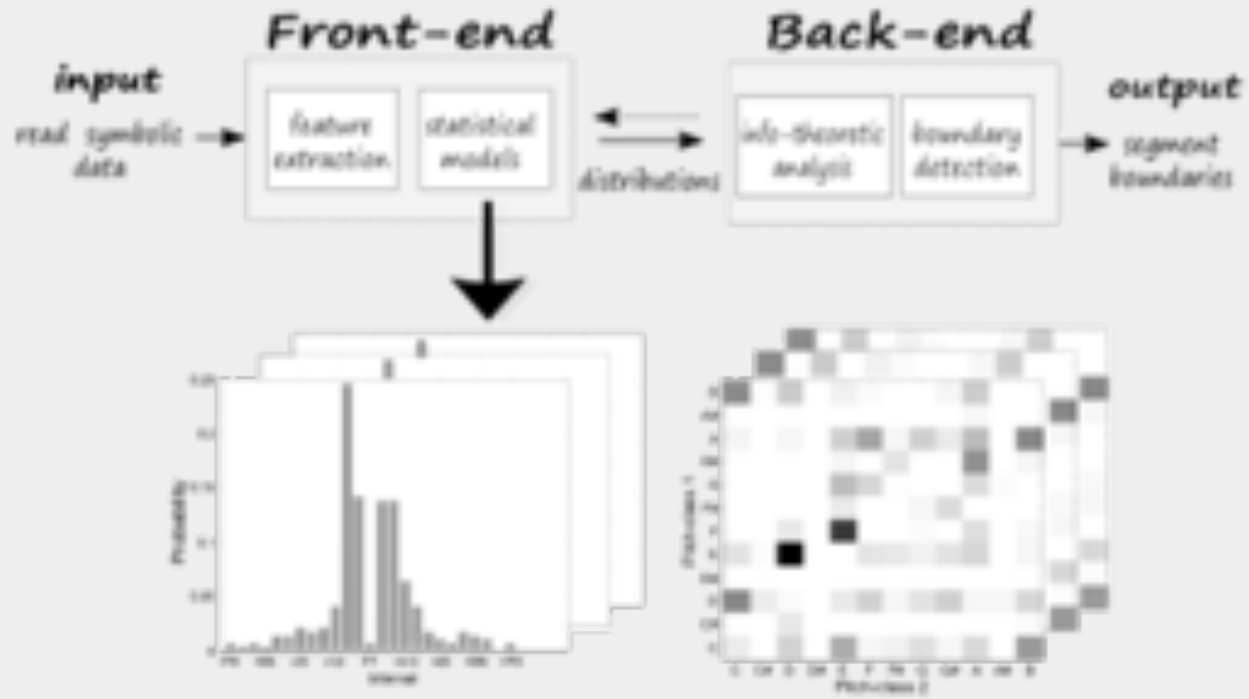


Expectation Based: Step-by-Step

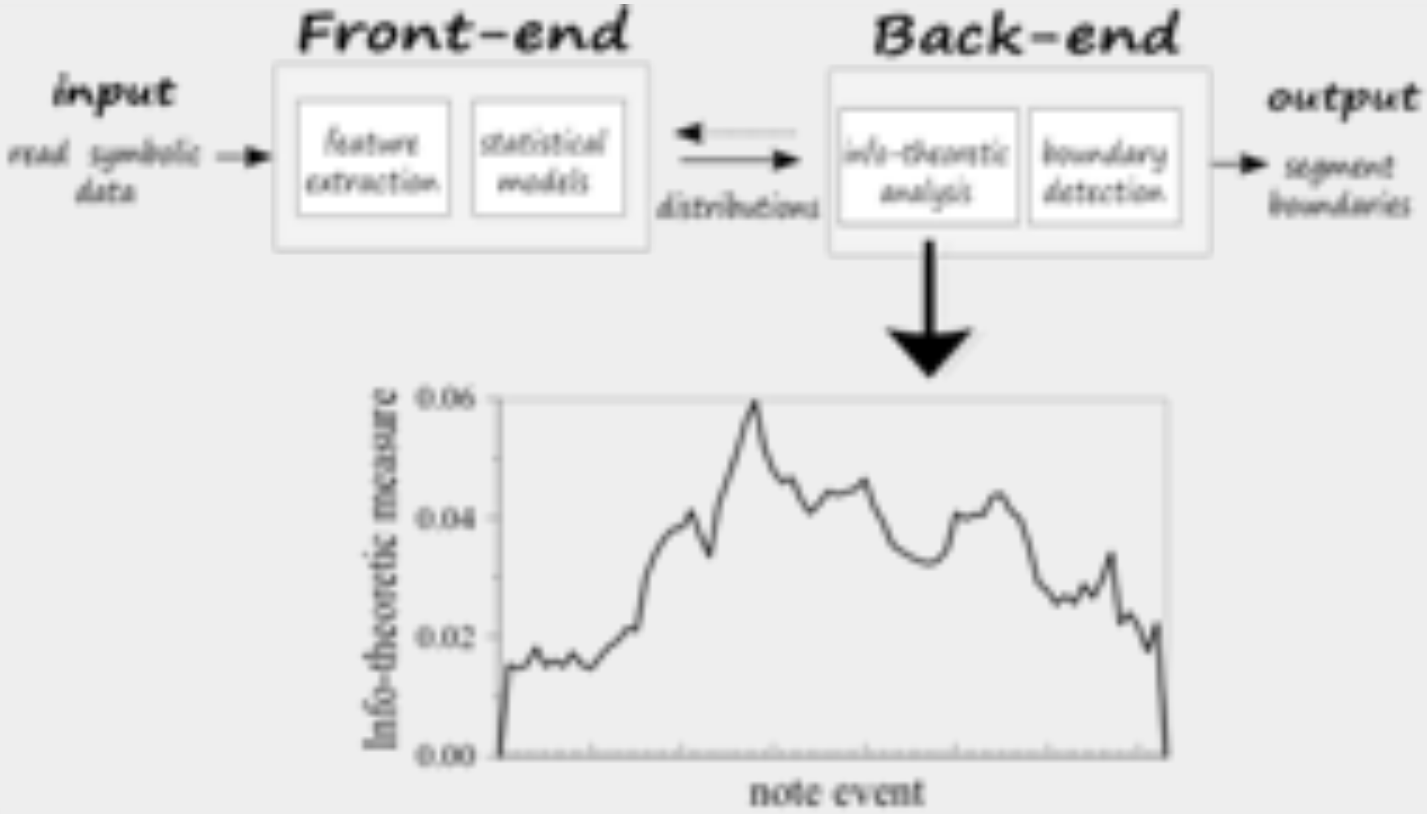


onset	48	72	120	144	192	216	264	288
deltast	0	0	0	0	0	0	0	0
dur	24	48	24	48	24	48	24	48
barlength	72	72	72	72	72	72	72	72
pulses	3	3	3	3	3	3	3	3
cpitch	66	67	69	67	66	67	69	67
keysig	1	1	1	1	1	1	1	1
mode	0	0	0	0	0	0	0	0
phrase	1	0	0	0	0	0	0	-1

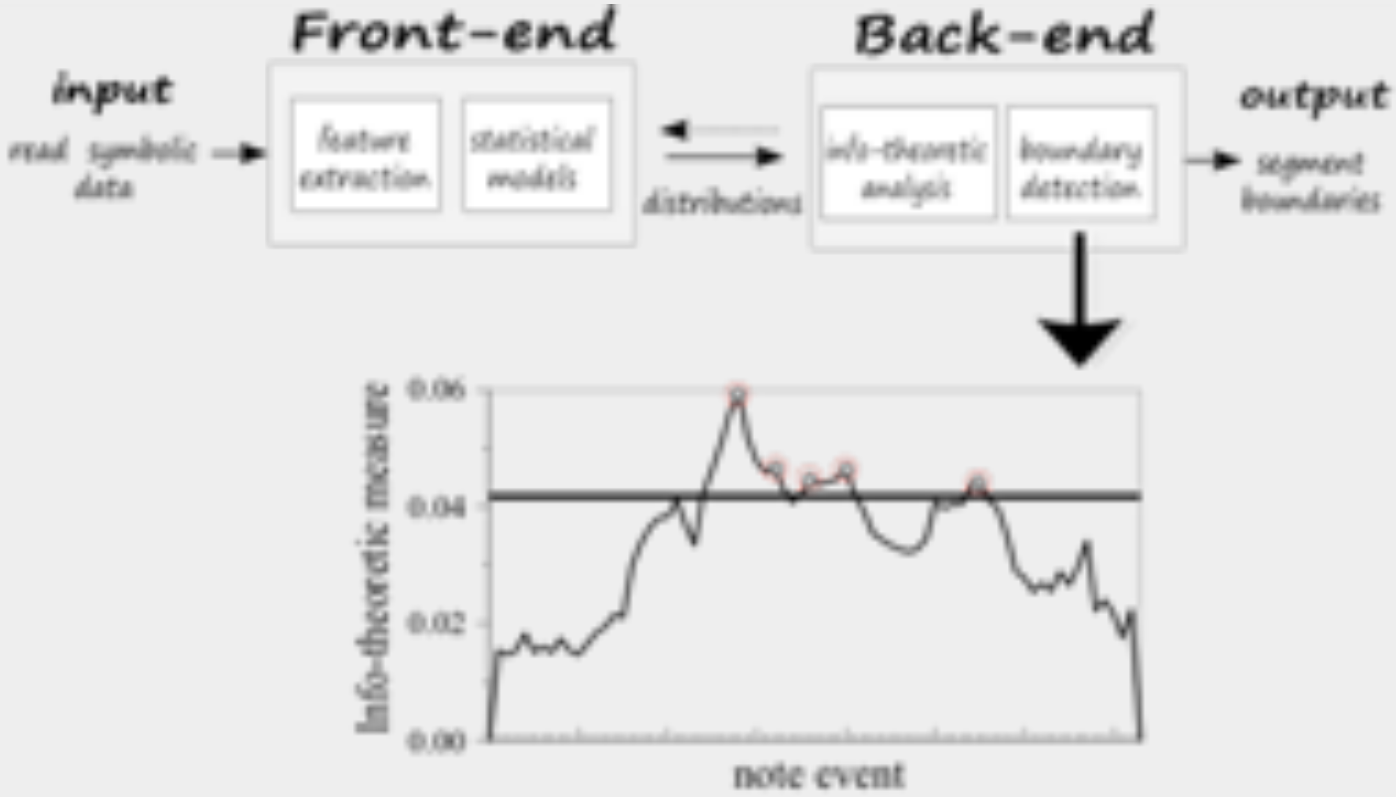
Expectation Based: Step-by-Step



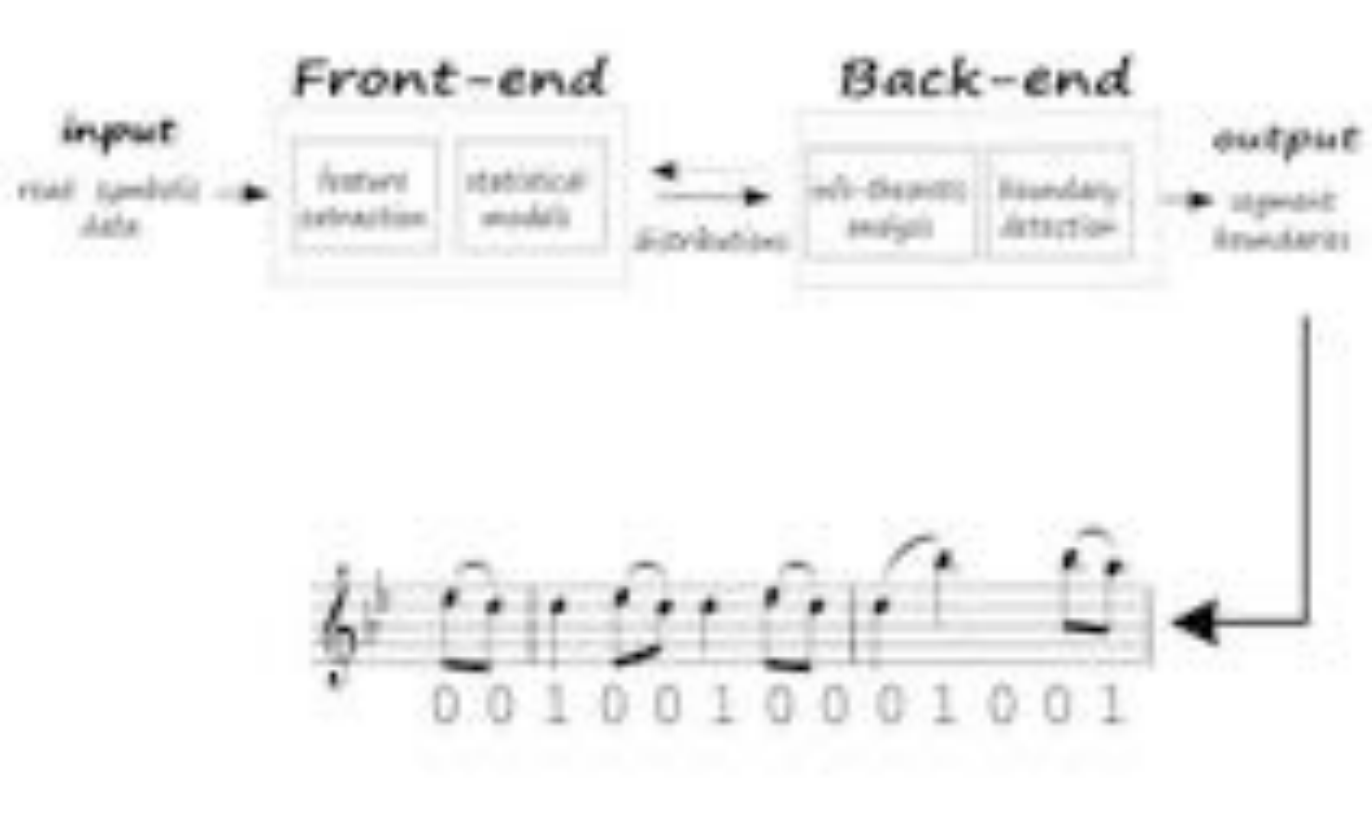
Expectation Based: Step-by-Step



Expectation Based: Step-by-Step



Expectation Based: Step-by-Step



Expectation Based: Assumptions

- closure is **relevant** for boundary perception
- closure is reflected in information-theoretic **surprise**
- closure is not **universal/idiom-independent** (LTM)

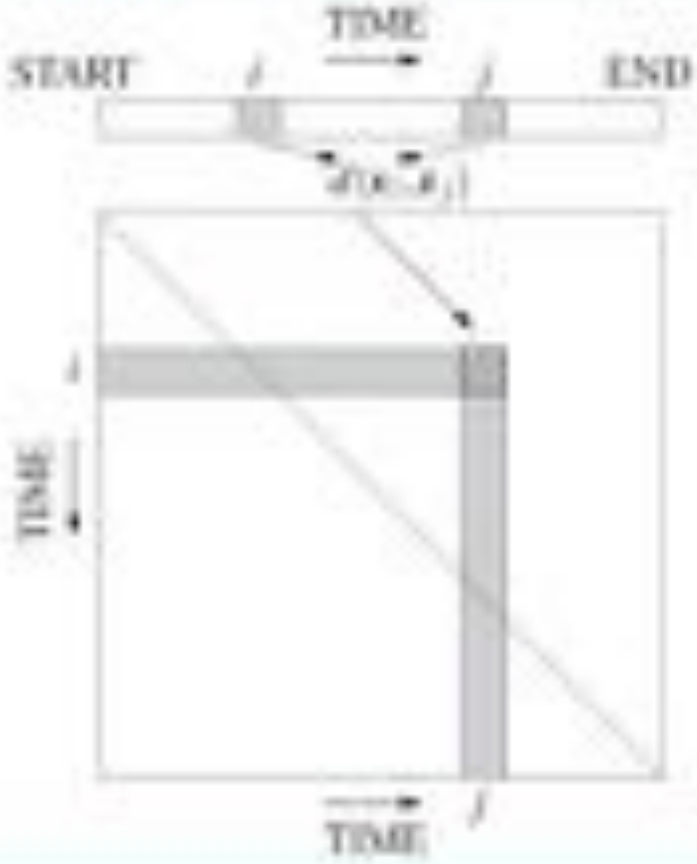
Segmentation in audio domain

- Main principles:
  - novelty-based: detect transitions between contrasting parts
  - repetition-based: identify recurring patterns
  - homogeneity-based: passages that are consistent for a certain feature

Mid-level representation

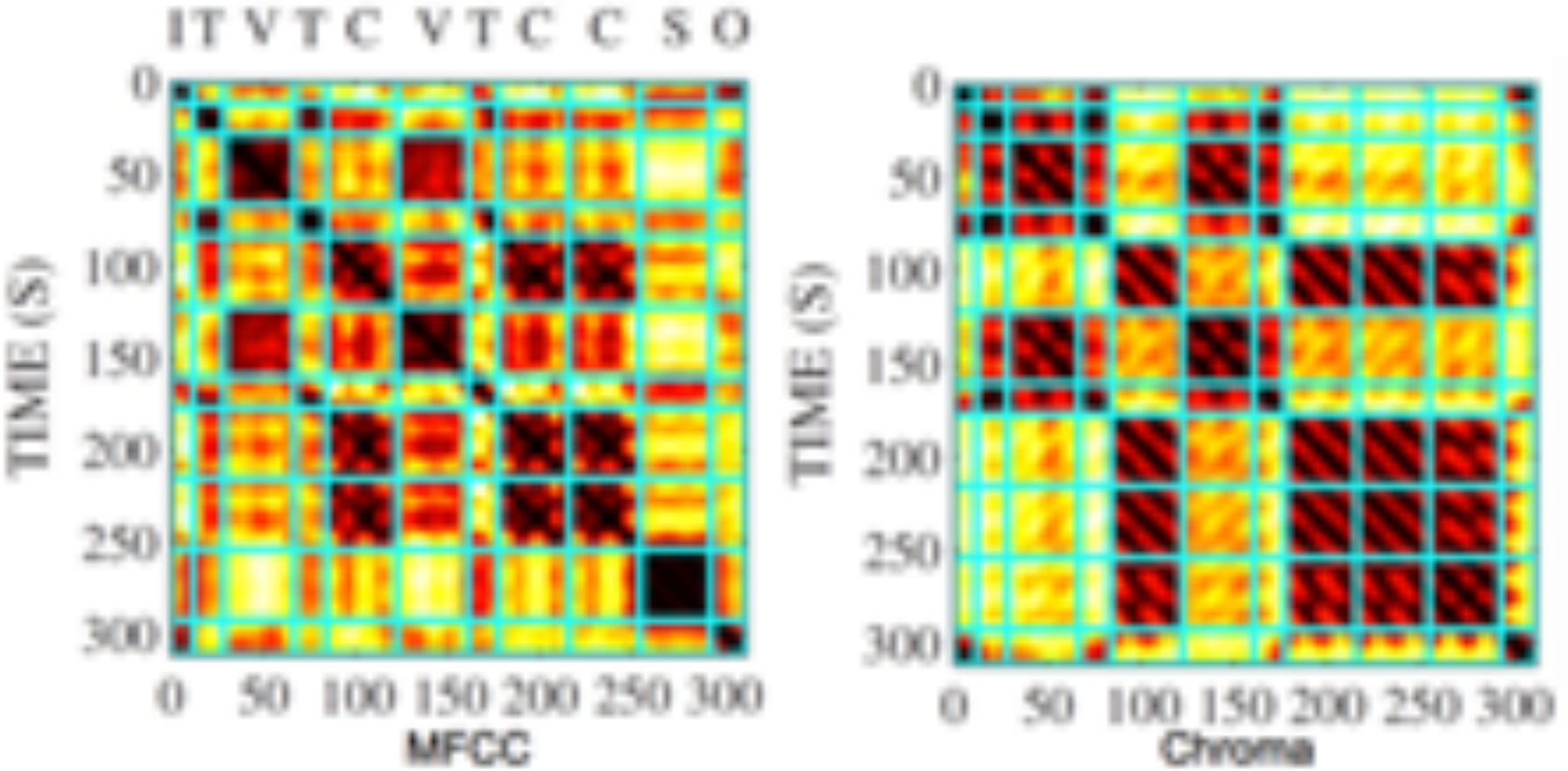
Self-distance matrix (self-similarity matrix)

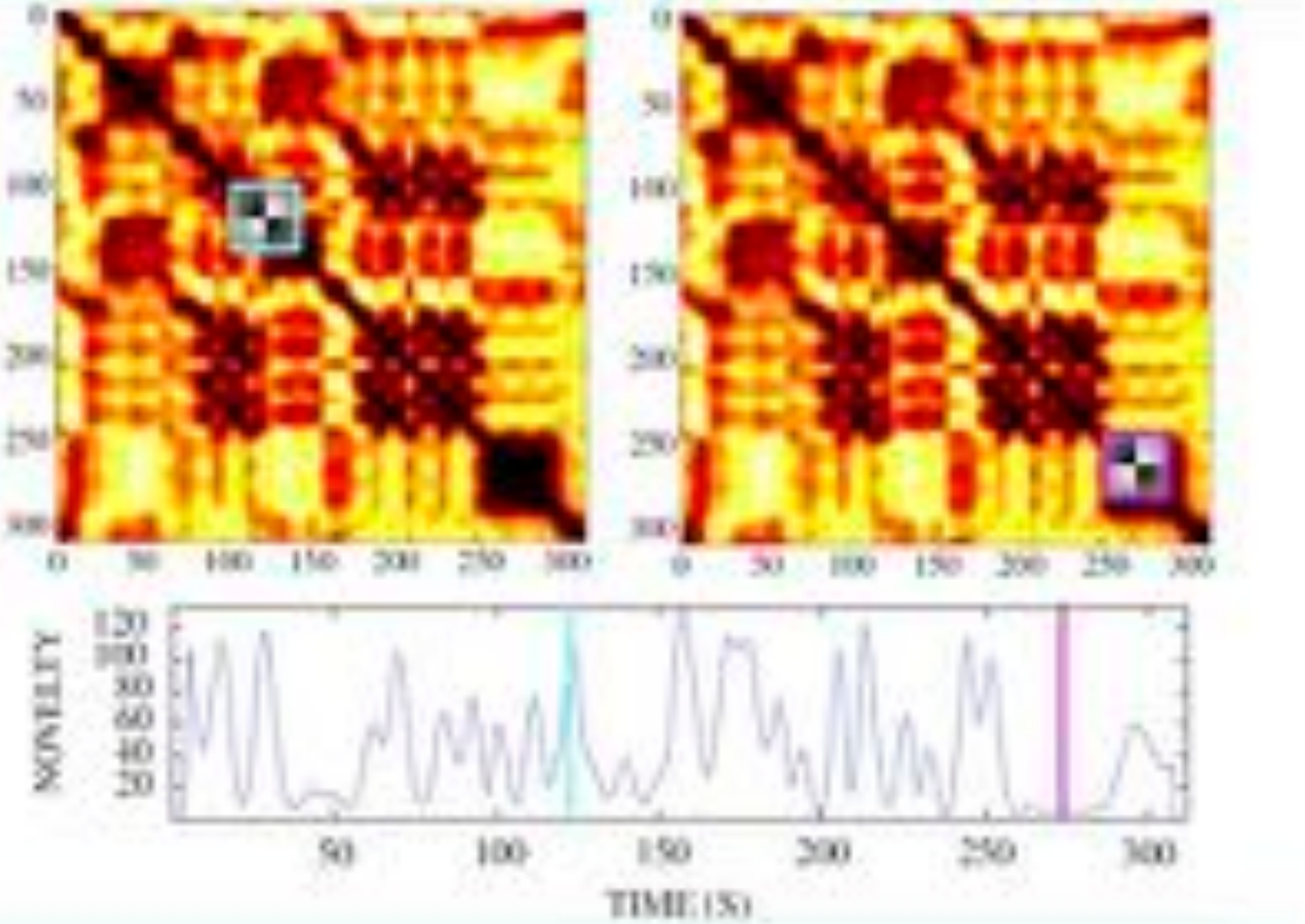
- Common mid-level representation
- Comparing each frame with all other frames
  - Each element describing the dissimilarity of two frames (or a sequence of frames)
- Informative patterns
  - Stripes for repeated sequences
  - Blocks for homogenous segments



Mid-level representation

Self-distance matrix examples




0
50
100
150
200
250
300
0
50
100
150
200
250
300
MOTILITY
120
100
80
60
40
20
50
100
150
200
250
300
TIME (S)

Summary

- Why segmentation?
- Typical Segmentation cues
- Typical approaches to computational segmentation
  - Gestalt based (rule-based)
  - Expectation based (data driven)
  - Student presentation: use of segmentation within folk music research

Final reports on projects

- Deadline for submission: Sunday February 2nd
  - Use ISMIR-style template, around 6 pages
  - Submission via Blackboard
- Presentations in week 4
  - 10 minutes per group
  - Ask questions after each presentation
  - Students NOT presenting will have to give feedback via Google document
  - Hence: participation is mandatory on both days