Symbolic Music Feature Extraction: Segmentation

Marcelo E. Rodríguez-López

Department of Information and Computing Sciences
Utrecht University, Interaction Technology Group
M.E.RodriguezLopez@uu.nl

May 13, 2014
Main Course Modules

A. Sound and music for games
B. Analysis, classification, and retrieval of sound and music
C. Generation and manipulation of sound and music for games
Main Course Modules

A. Sound and music for games
B. Analysis, classification, and retrieval of sound and music
C. Generation and manipulation of sound and music for games
Today

Symbolic music feature extraction: segmentation

- Introduction to music segmentation in MIR
  - What are musical segments
  - Areas of research in automatic music segmentation
Today

Symbolic music feature extraction: segmentation

► **Introduction** to **music segmentation** in **MIR**
  - What are musical segments
  - Areas of research in automatic music segmentation

► **Computer Models of Melody Segmentation**
  - Perceptual Cues
  - Music representation (symbolic)
  - Task definition
  - 2 Cognition-inspired models of melody segmentation
Music Segmentation: MIR Task Definition

In **Music Information Retrieval (MIR)**, segmentation is the task of **dividing** a musical piece/melody/section into smaller **structural units**.
Music Segmentation: Example
Music Segmentation: Example

CORRECT
Music Segmentation: Example

INCORRECT
Music Segmentation: MIR Task Definition

Important distinction:

- **Segmentation** focuses on segregating **sequential** structural elements, such as notes, motifs, phrases, sections, etc.
Music Segmentation: MIR Task Definition

Important distinction:

- **Segmentation** focuses on segregating **sequential** structural elements, such as notes, motifs, phrases, sections, etc.

- **Streaming** focuses on segregating **simultaneous** structural elements, such as different voices of a polyphony, etc.
Music Segmentation: Global Perspective
Music Segmentation: Global Perspective
Music Segmentation: Global Perspective
Music Segmentation: Global Perspective

For more information on structuring musical processes, refer to (8)
Music Segmentation: Why

- Music Cognition
  - fundamental perceptual process
Music Segmentation: Why

- Music Cognition
  - fundamental perceptual process
- MIR
  - music music visualisation and summarisation
  - indexing for search and browsing of large music collections
  . . . and many more
Music Segmentation: Why

- Music Cognition
  - fundamental perceptual process
- MIR
  - music music visualisation and summarisation
  - indexing for search and browsing of large music collections
  \[\ldots\] and many more
- Games
  - material for automatic composition/improvisation
  - markers music-to-video/text synchronization
Music Segmentation: Subtasks

- **Subtasks**
  - *identify* boundary locations
  - *pair* boundaries (begin, end)
  - *label* segments
Music Segmentation: Subtasks

- **identify** boundary locations
- **pair** boundaries (begin, end)

**label** segments

[Diagram showing a timeline with labeled segments and boundaries]
Music Segmentation: Subtasks

- **Subtasks**
  - **identify** boundary locations
  - **pair** boundaries (begin, end)
  - **label** segments
Music Segmentation: Granularity

- Perception & action
- Waveform sample
- Threshold of pitch perception
- Fastest repetitive human gestures (~12 Hz)
- Recognition of instrumental timbre
- Common duration of a composition (2-15 m)
- Human lifetime

- ~22 µ secs
- 13 msecs
- 45 msecs
- 100 msecs
- secs mins hours days years
Music Segmentation: Granularity

Musical Time-scales

perception & action

basic musical event

micro

meso

macro

supra

~22 µ secs

13 msecs

45 msecs

100 msecs

secs

mins

hours

days

years

common duration of a composition (2-15 m)

human lifetime

fastest repetitive human gestures (~12 Hz)

recognition of instrumental timbre

threshold of pitch perception

waveform sample

~∞

∞
Music Segmentation: Granularity

Musical Time-scales

Micro
- waveform sample
- threshold of pitch perception

Meso
- recognition of instrumental timbre

Macro
- common duration of a composition (2-15 m)

Supra
- human lifetime

Music-theory analytical constructs

Perception & action

Waveform samples (~22 μsecs)
Threshold of pitch perception (13 msecs)
Recognition of instrumental timbre (45 msecs)
Common duration of a composition (100 msecs)
Human lifetime (∞)
Music Segmentation: Areas in MIR

- AUTOMATIC TRANSCRIPTION
  - identify boundary locations
  - pair boundaries (begin, end)
  - label segments
  - music-theoretic segment parallel: note
  - input: audio, monophonic/polyphonic music
  - method: onset/offset detection
Music Segmentation: Areas in MIR

- **MOTIV DISCOVERY**
  - **identify** boundary locations
  - **pair** boundaries (begin, end)
  - **music-theoretic segment parallel**: motifs
  - **input**: symbolic, monophonic/polyphonic music
  - **method**: pattern extraction
Music Segmentation: Areas in MIR

- **MELODIC SEGMENTATION**
  - identify boundary locations
  - music-theoretic segment parallel: phrases
  - input: symbolic, monophonic music
  - method: various
Music Segmentation: Areas in MIR

- **Form Segmentation**
  - *identify* boundary locations
  - *label* segments
  - *music-theoretic segment parallel*: sections, form
  - *input*: audio, monophonic/polyphonic music
  - *method*: various
Focus of this lecture

(phrase-level) melody segmentation
Melody Segmentation: What is a Phrase?

- **DEFs Western Art Music**
  Aggregation of consecutive notes “expressing a complete musical thought”\(^{(3)}\), or “containing significant tonal motion”\(^{(4)}\). Roughly 4-8 measures in length\(^{(5)}\).

- **DEFs Popular Music**
  ‘[a melodic segment] end[s] normally coinciding with the taking of breath (whether actual or nominal)’\(^{(1)}\)
  ‘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’\(^{(2)}\)
# Melody Segmentation: Perceptual Cues

<table>
<thead>
<tr>
<th>Stockhausen’s klaviersück IX</th>
<th>CUES</th>
<th>Mozart’s Fantasie</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 strongest boundaries</td>
<td></td>
<td>6 strongest boundaries</td>
</tr>
</tbody>
</table>

(chords-to-melody, pitch content, block chords) new material
(chordal, chromatic run, coda) return of material
(expansion, jump) change of register
change of rhythm
change of dynamic
(silence) pause
start of development
change of articulation
change of texture
change of pitch content
relaxation of tension
introduction of trill
change of tempo
(piano tone) change of timbre

new material (lyrical, dramatic, end of cadenza)
change of texture (thicker, thinner)
change of tempo
change of register
change of dynamic
change of key
change of harmony
change of meter
change of rhythm
change of melody

---

2 for details on this perceptual experiment please refer to (16)
Melody Segmentation: Perceptual Cues

Start

Musical piece being listened to

End
Melody Segmentation: Perceptual Cues

- a given extension of time
- start
- musical piece being listened to
- end
Melody Segmentation: Perceptual Cues

A given extension of time

Musical Entity

Cue

Relationship

Musical Entity

start

end

musical piece being listened to
Melody Segmentation: Perceptual Cues

A musical piece being listened to over a given extension of time influences the cues and musical entities, relationships, and contextual musical factors, leading to the perception of boundaries.
Melody Segmentation: Perceptual Cues

Segment Boundary Cue

<table>
<thead>
<tr>
<th>Relationship-type</th>
<th>Musical Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td>closure</td>
<td>sameness</td>
</tr>
<tr>
<td>continuation</td>
<td>difference</td>
</tr>
<tr>
<td></td>
<td>gap</td>
</tr>
<tr>
<td></td>
<td>contrast</td>
</tr>
</tbody>
</table>
Melody Segmentation: Perceptual Cues

Segment Boundary Cue

<table>
<thead>
<tr>
<th>Relationship-type</th>
<th>Musical Entities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Attribute Class</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- sameness
- difference
- gap
- contrast
- closure
- continuation
- due-to
- rhythmic
- pitch
- harmonic
- timbral
- dynamic
- tonal
- idiomatic
- templates
- feature-in
- between

Universiteit Utrecht
Melody Segmentation: Perceptual Cues

Source

stimulus (piece being listened to)

composition related

metrical & others **

performance related

schema (previously heard music)

Segment Boundary Cue

Relationship-type

sameness

difference

gap

contrast

closure

continuation

due-to

Musical Entities

Attribute Class

rhythmic

pitch

harmonic

timbral

rhythmic

pitch

tonal

timbral

dynamic

dynamic

Time-Span

feature-in

entities >= a figure

feature-in

entities <= 2-3 notes

feature-in

entities >= a figure

feature-in

entities >= a phrase
Melody Segmentation: Cue Examples

Segment boundaries in these melodies should ‘mainly’ by cued by:

- **differences**: melodic ‘gaps’
- **differences**: melodic contrast
- **sameness**: motivic repetition
- **closure**: tonal closure
Computer Melody Segmentation Models: Task & Representation

- (Revisiting) The Task
  - **identify** boundary locations
  - **pair** boundaries (begin, end)
  - **label** segments

- **Music Representation**
  - **symbolic**: mapping sound $\rightarrow$ event $\approx$ musical note (e.g. MIDI, **kern**)
Computer Melody Segmentation Models:
Historical Overview

- **development**: +30 years
- **# of models**: >30
- **comparative studies**: 4, 3-6 models evaluated
- **most successful**: Gestalt-based, $F1 = 0.60 - 0.66$
Computer Melody Segmentation Models:
General I/O Diagram

- **Segmentation Model**
- **input**
- **output**

- **Segments**:
  - Onset: 48, 72, 120, 144, 192, 216, 264, 288
  - Offset: 71, 121, 143, 191, 215, 263, 287, 335
  - Binary: 0, 1, 0, 1, 0, 0, 1, 0
  - Continuous: 0.1, 0.7, 0.3, 0.8, 0.2, 0.4, 0.6, 0.3
Computer Melody Segmentation Models: Approaches

Most Researched

- **Type:** Gestalt Based (1980-today)
  - **Principle:** difference as (Local) discontinuity detection

- **Type:** Repetition Based (∼1998-today)
  - **Principle:** sameness 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: Overview

- 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

\[ \text{\textasciitilde\textasciitilde\textasciitilde\textasciitilde\textasciitilde\textasciitilde\textasciitilde\textasciitilde} \]
Local Discontinuity Detection: Input

\[ e_1, \ldots, e_i, \ldots, e_N \]

*onset:* 24 48 72 84 96 120 \(,\ldots,\)

*pitch:* 69 69 68 66 66 69 \(,\ldots,\)

\[ 2322 \]

\[ 64 \]
Local Discontinuity Detection: Input

\[ \begin{align*}
\text{ioi} : & \quad .5 \quad .5 \quad .25 \quad .25 \quad .5 \quad .5 \quad .5 \quad .5 \quad 2 \quad .75 \quad .25 \quad .5 \quad .5 \quad .5 \quad 1 \quad .5 \quad .5 \quad .5 \quad .5 \quad .5 \\
p-inv : & \quad 0 \quad 1 \quad 2 \quad 0 \quad 3 \quad 5 \quad 0 \quad 5 \quad 1 \quad 2 \quad 3 \quad 2 \quad 5 \quad 2 \quad 2 \quad 3 \quad 3 \quad 5 \quad 2 \quad 3 \quad 7 \quad 1 \quad 2 \quad 3 \quad 5
\end{align*} \]
Local Discontinuity Detection: Profiles

\[ \text{ioi} : \]

\[ \text{p-inv} : \]
Local Discontinuity Detection: Profiles
Local Discontinuity Detection: Output

Output: 0 0 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)\(^{(10)}\)
Local Discontinuity Detection: LBDM

- Algorithm
  1. given: pitch $p[i]$, onset $on[i]$, offset $off[i]$, for $i = 1, \ldots, n$
  2. compute pitch $cp$, inter-onset-intervals $ioi$, and rest (offset-to-onset-interval) $ooi$ profiles
  3. for each profile $x$, compute profile strength $s$
     \[ s[i] = x[i] \cdot (r[i-1] + r[i]) \]
     where:
     \[ r[i] = \frac{|x[i] - x[i+1]|}{x[i] + x[i+1]} \]
  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
- 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\(^{(12;11)}\)
- Deep-Belief Networks\(^{(14)}\)
- Recurrent Neural Networks\(^{(13)}\)
- Self-Organizing Maps\(^{(15)}\)
Expectation Based: Melody Prediction

Using Markov Models:

Corpus

\[ e_{n-m} \ldots e_{n-1} e_n \]

\[ P_{\text{tm}}(e_n|C) \]

\[ P(e_n|C) \]

Input

\[ e_1 \ldots e_{n-m} \ldots e_{n-1} e_n \]

\[ C \]

\[ P_{\text{stm}}(e_n|C) \]
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’
Expectation Based Boundary Detection

**Front-end**
- Feature extraction
- Statistical models

**Back-end**
- Info-theoretic analysis
- Boundary detection

**Input**
- Read symbolic data

**Output**
- Segment boundaries

Diagram showing the process flow with musical notation and time axis.
Expectation Based: Step-by-Step

**Front-end**
- feature extraction
- statistical models

**Back-end**
- info-theoretic analysis
- boundary detection

**output**
- segment boundaries

**input**
- read symbolic data

---

<table>
<thead>
<tr>
<th></th>
<th>onset</th>
<th>deltast</th>
<th>dur</th>
<th>barlength</th>
<th>pulses</th>
<th>cpitch</th>
<th>keysig</th>
<th>mode</th>
<th>phrase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>48</td>
<td>72</td>
<td>120</td>
<td>144</td>
<td>192</td>
<td>216</td>
<td>264</td>
<td>288</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>48</td>
<td>24</td>
<td>48</td>
<td>24</td>
<td>48</td>
<td>24</td>
<td>48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>66</td>
<td>67</td>
<td>69</td>
<td>67</td>
<td>66</td>
<td>67</td>
<td>69</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td></td>
</tr>
</tbody>
</table>

---

- **onset**
- **deltast**
- **dur**
- **barlength**
- **pulses**
- **cpitch**
- **keysig**
- **mode**
- **phrase**
Expectation Based: Step-by-Step

**Front-end**
- Input: read symbolic data
- Feature extraction
- Statistical models
- Distributions

**Back-end**
- Info-theoretic analysis
- Boundary detection
- Segment boundaries

**Diagram**
- Probability distribution
  - Interval: -P8, -M6, -d5, -m3, P1, +m3, +d5, +M6, +P8
  - Pitch-class 1
  - Pitch-class 2
Expectation Based: Step-by-Step

input
read symbolic data

Front-end
feature extraction
statistical models

distributions

Back-end
info-theoretic analysis
boundary detection

output
segment boundaries

Info-theoretic measure vs. note event
Expectation Based: Step-by-Step

Front-end:
- Feature extraction
- Statistical models

Back-end:
- Info-theoretic analysis
- Boundary detection

Output:
- Segment boundaries

Input:
- Read symbolic data

Graph:
- Info-theoretic measure vs. note event
  - Note event
  - Info-theoretic measure
Expectation Based: Step-by-Step

**Front-end**
- Feature extraction
- Statistical models

**Back-end**
- Info-theoretic analysis
- Boundary detection

**Input**
- Read symbolic data

**Output**
- Segment boundaries

HH
0 0 1 0 0 1 0 0 0 1 0 0 1
Expectation Based: Assumptions

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


Bibliography (cont.)


Bibliography (cont.)


Bibliography (cont.)


Bibliography (cont.)

