
Thomas W. Hedges
Queen Mary University of London
t.w.hedges@qmul.ac.uk

Geraint A. Wiggins
Queen Mary University of London
geraint.wiggins@qmul.ac.uk
Introduction

• How to model higher order / hierarchical structure with bottom-up, statistically driven models?

• Forth & Wiggins (2015) present IDyOT (Information Dynamics of Thinking), a cognitive architecture which expatiates IDyOM (Pearce 2005) to account for many aspects of human behaviour across multiple domains (language and music).

• The current research focusses on tonal harmony, in particular jazz.

• Segmentation is the first stage in these bottom-up models.
Jazz Leadsheets

• The entry point of the model is at the chord symbol level.
• 348 leadsheets (15,197 events) compiled by Pachet et al. (2013), taken from the Real Book vol. 1.
• Typical sequence learnt:
  Am7, D7, DM, CM, F#halfdim7, B7, Em
IDyOM: Statistical learning and modelling of the musical surface

- Information Dynamics Of Music (Pearce 2005)
- An unsupervised probabilistic model using variable order Markov models (PPM* - Cleary & Teahan 1997), interpolated smoothing (Cleary & Witten 1984, Moffat 1990) and multiple viewpoints weighted by entropy (Conklin & Witten 1995) to model expectation.

<table>
<thead>
<tr>
<th></th>
<th>BM(^7)</th>
<th>D(^7)</th>
<th>GM(^7)</th>
<th>B(_b)(^7)</th>
<th>E(_b)M(^7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Root</td>
<td>11</td>
<td>2</td>
<td>7</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>ChordType</td>
<td>M</td>
<td>7</td>
<td>M</td>
<td>7</td>
<td>M</td>
</tr>
<tr>
<td>PosInBar</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>RootInt</td>
<td>⊥</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>i ci</td>
<td>⊥</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>RootIntFiP</td>
<td>0</td>
<td>3</td>
<td>8</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>RootInt⊙FiB</td>
<td>⊥</td>
<td>⊥</td>
<td>8</td>
<td>⊥</td>
<td>8</td>
</tr>
</tbody>
</table>
Information theoretic segmentation

- Perceived segment boundaries before difficult to predict events (Pearce et al. 2010, Wiggins 2012, Griffiths et al. submitted).

- Difficulty to predict modelled by unexpectedness, defined by information content:

\[
h (e_i|e_1^{i-1}) = - \log_2 p (e_i|e_1^{i-1})
\]

- Place segment when before large rise in information content, when ratio between \( h \) of two adjacent events exceeds a threshold, \( d \).

\[
\frac{h (e_i|e_1^{i-1})}{h (e_{i-1}|e_1^{i-2})} > d
\]
Information theoretic segmentation

- Information content profiles calculated with 10-fold cross validation.
- Viewpoint selected with forward stepwise selection algorithm.
- Viewpoints: $\text{ROOTINT} \otimes \text{ICI}$, $\text{CHORDTYPE} \otimes \text{ROOTINTTHRBAR}$, $\text{CHORDTYPE} \otimes \text{ROOTINTFIP}$, $\text{POSINBAR} \otimes \text{ROOTINTFIP}$, $\text{CHORDTYPE} \otimes \text{POSINBAR}$, $\text{CHORDTYPE} \otimes \text{ROOTINT}$, $\text{ROOT} \otimes \text{CHORDTYPE}$
Information theoretic segmentation

- Information content profiles calculated with 10-fold cross validation.
- Viewpoint selected with forward stepwise selection algorithm.
- Viewpoints: $\text{ROOTINT} \otimes \text{ICI}$, $\text{CHORDTYPE} \otimes \text{ROOTINTTHRBAR}$, $\text{CHORDTYPE} \otimes \text{ROOTINTFIP}$, $\text{POSINBAR} \otimes \text{ROOTINTFIP}$, $\text{CHORDTYPE} \otimes \text{POSINBAR}$, $\text{CHORDTYPE} \otimes \text{ROOTINT}$, $\text{ROOT} \otimes \text{CHORDTYPE}$
Phrase Structure

- No ground truth.
- Harmonic segmentation approximately reflect phrase structure.
- 4-bar phrases can be found segmenting with \( d = 2.6 \)
  - \( \kappa: .24 \), accuracy: .76
- Random segmenter segments with \( p = .2 \)
  - \( \kappa: .09 \), accuracy: .70
Giant Steps - John Coltrane

Threshold $d = 2.6$

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>D7</th>
<th>G</th>
<th>Bb7</th>
<th>Eb</th>
<th>Am7</th>
<th>D7</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>Bb7</td>
<td>Eb</td>
<td>F#7</td>
<td>B</td>
<td>Fm7</td>
<td>Bb7</td>
<td></td>
</tr>
<tr>
<td>Eb</td>
<td>Am7</td>
<td>D7</td>
<td>G</td>
<td>C#m7</td>
<td>F#7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Fm7</td>
<td>Bb7</td>
<td>Eb</td>
<td>C#m7</td>
<td>F#7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Giant Steps - John Coltrane

Threshold

\[ d = 2.6 \]

\[
\begin{array}{cccccc}
B & D7 & G & Bb7 & Eb & Am7 & D7 \\
G & Bb7 & Eb & F#7 & B & Fm7 & Bb7 \\
Eb & Am7 & D7 & G & C#m7 & F#7 \\
B & Fm7 & Bb7 & Eb & C#m7 & F#7 \\
B & & & & & & \\
\end{array}
\]
Giant Steps - John Coltrane

Threshold $d = 2.6$
Giant Steps - John Coltrane

Threshold $d = 2.6$
Giant Steps - John Coltrane

Threshold

\[ d = 2.6 \]
Segment Types

- Segmenting the whole corpus (15,197 chords) at $d = 2.6$
- 3,007 segment tokens
- 1,531 segment types (unique)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>297</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
</tr>
<tr>
<td>3</td>
<td>79</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
</tr>
<tr>
<td>6</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>22</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

• Difficult to evaluate without a ground truth.
• Useful segmentations of jazz chord sequences.
• Rough phrase structure found.
• Key tonal-harmonic units identified (e.g. ii\(^7\)-V\(^7\)-I).
• No in-built knowledge of music theory or tonal harmony.
• Future work will compare with human segmentations of harmony.
References


cclab bonus: Imperfect Cadence Problem
cclab bonus: ’Round Midnight