What and Why?
- Segmentation is the process of partitioning a melody
- Gives structure to the melody
- Useful as a preprocessing stage for:
  - Pattern discovery
  - Music search

The LBDM
- This model is based on Gestalt principles
- The ‘strength’ of the interval is given by:

\[ s_i = x_i \times \left( r_{i-1} + r_{i+1} \right) \]

\[ r_{ij} = \begin{cases} 
\frac{1}{\theta_x} \ln \left| x_{i+1} - x_i \right| & x_{i+1} \neq 0 \land x_i, x_{i+1} \geq 0 \\
\frac{1}{\theta_{IOI}} & x_i = x_{i+1} = 0
\end{cases} \]

- A change rule assigns boundaries to intervals with strength proportional to the degree of change between consecutive interval pairs.
  - Proximity rule scales the previous boundary down.
  - For each parameter \( k \) a sequence is calculated and all sequences are normalized and combined to give the overall boundary strength profile.
  - Suggested weights are:
    - Pitch = rest = 0.25
    - IOI = 0.5
  - Local peaks indicate boundaries.
Melodic Density Segmentation Model

- In contrast with the LBDM, the melodic density model calculates melodic cohesion between pitch intervals.
- Uses a sliding window system and an attenuation function to model short-term memory.

<table>
<thead>
<tr>
<th>$P_i$</th>
<th>$F_i$</th>
<th>$F_{i+1}$</th>
<th>$F_{i+2}$</th>
<th>$F_{i+3}$</th>
<th>$P_{i+3}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

- More recent events have a stronger contribution to the melody than earlier ones.
- Size of window is fixed – the tempo is the determining factor in boundary perception.

Formulae

- Melodic Density in a sequence of N events:
  - The density at event $i$ is:
  
  $d = \sum_{k=0}^{N} \sum_{n=1}^{N-k} f(n(m,n)) \cdot a(n, n)

  - $F(r)$ returns the frequency of the interval; $r(m, n) = |p_i - p_{i-n}|$ is pitch interval in semitones.
  
  $a_i(m, n) = (1 - \frac{t_i - t_{i-m-n}}{M})^2$

  - Attenuation function where $t_i$ = onset time of $e_i$ and $M$ is duration of memory window.

Results and Comparison

- The MDSM approach led to fewer boundaries detected than the LBDM approach.
  - MDSM had higher Precision and higher Recall than the LBDM.

  Author’s Claim:
  - The MDSM is a better method than the LBDM method.
  - MDSM is a more cognitively realistic approach.
Unsupervised Learning of Melodic Segmentation: A Memory Based Approach

Miguel Ferrand, Peter Nelson, Geraint Wiggins

What is Melody?

Melody is seen to be a “temporal process where sound events unfold in time”

- From basic information like pitch, duration and inter-onset intervals we get:
  - Pitch Step: the interval distance between consecutive notes (in semitones)
  - Duration Ratio: the ratio between the duration of consecutive events.

Aims

- To provide an automatic method of performing melodic segmentation
- To do the above task with no prior musical knowledge

Example

<table>
<thead>
<tr>
<th>No.</th>
<th>Pitch</th>
<th>Onset</th>
<th>Dur</th>
<th>PS</th>
<th>DR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>82</td>
<td>3998</td>
<td>1000</td>
<td>-1</td>
<td>-5 (0.14)</td>
</tr>
<tr>
<td>2</td>
<td>81</td>
<td>4998</td>
<td>143</td>
<td>+2</td>
<td>-1 (0.80)</td>
</tr>
<tr>
<td>3</td>
<td>83</td>
<td>5141</td>
<td>115</td>
<td>-3</td>
<td>6 (8.70)</td>
</tr>
<tr>
<td>4</td>
<td>80</td>
<td>5256</td>
<td>1000</td>
<td>-1</td>
<td>-5 (0.17)</td>
</tr>
<tr>
<td>5</td>
<td>79</td>
<td>6256</td>
<td>167</td>
<td>+2</td>
<td>0 (0.99)</td>
</tr>
<tr>
<td>6</td>
<td>81</td>
<td>6413</td>
<td>165</td>
<td>-2</td>
<td>1 (1.64)</td>
</tr>
<tr>
<td>7</td>
<td>78</td>
<td>6589</td>
<td>273</td>
<td>-3</td>
<td>0 (1.18)</td>
</tr>
<tr>
<td>8</td>
<td>77</td>
<td>6862</td>
<td>322</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Markov Models and Mixed Markov Models

Markov model (nth order n-gram):
- The probability of occurrence of a depends on the prior occurrence of n-1 other symbols and the probability of a sequence \( P(s) \) is given by:

\[
P(s) = \prod_{i=1}^{n} P(w_i | w_{i-1}, \ldots, w_1)
\]

This causes problems if any of the terms has zero probability

Mixed Markov Models:

\[
P(w_i | w_{i-1}, \ldots, w_1) = \sum_{\mu} \phi(\mu) \alpha^{\mu} (w_i | w_{i-1})
\]

Entropy and Boundary Prediction

- Boundaries occur when there is a change in entropy
- Entropy in a context \( 'c' \) is given by:

\[
H_c = \sum_w P(w | c) \log_2 P(w | c)
\]

- \( w \) = all symbols that can be successors of context \( c \)
- Context \( c \) is a sequence of size \( n-1 \), where \( n \) is the order of the model
- Entropy vectors are calculated by taking successive context sequences from the feature vectors of the target melody and calculating their means and standard deviations. All values outside the standard deviation are rejected.
- Of the remaining, only those that have a contiguous high to low or low to high variation with respect to the mean are considered.

Results

- Compared results from actual listeners to results from the computational model
- Listeners were made to mark boundaries in the melodies (called L-boundaries)
- The computational model also marked boundaries
- In the case of Debussy's Syrinx, the 11 L-boundaries were predicted correctly by the software, but it also generated 5 extra ones.

Mixed Markov Models (contd.)

- \( a(w_i | w_{i-1}) \) is a \( k x k \) transition matrix containing probabilities of occurrence of a symbol at position \( i \) given that it has occurred at position \( i-1 \)
- Mixing coefficients \( (\cdot) \) are estimated using an iterative process
Conclusion

- An entropy based model was constructed to evaluate boundaries using pitch and duration features.
- The experiment seems to corroborate the idea that variations in entropy constitute boundaries.