Memory and Melodic Density : A Model for Melody Segmentation

By Miguel Ferrand, Peter Nelson, Geraint Wiggins

Presentation by Amit Singh

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:

$$s_i = x_i \times (r_{i-1,i} + r_{i,i+1})$$

$$r_{i,i+1} = \begin{cases} \frac{|x_i - x_{i+1}|}{x_i + x_{i+1}} & x_i + x_{i+1} \neq 0 \wedge x_i, x_{i+1} \geq 0 \\ 0 & x_i = x_{i+1} = 0 \end{cases}$$

LBDM (contd.)

- 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

e_{i-3}	e_{i-2}	e_{i-1}	e_i	*event*
53	50	50	48	*pitch*
				order(n)
	3	0	2	1
		3	2	2
			5	3
...	2	1	0	*recency(m)*

- 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:

$$d_i = \sum_{m=0}^{t_i - t_{i-m} < M} \sum_{n=1}^{t_i - t_{i-m-n} < M} f(r_i(m,n)) \cdot a_i(m,n)$$

- F(r) returns the frequency of the interval; $r_i(m,n) = |p_{i-m} - p_{i-m-n}|$ is pitch interval in semitones

$$a_i(m,n) = \left(1 - \frac{t_i - t_{i-m-n}}{M}\right)^2$$

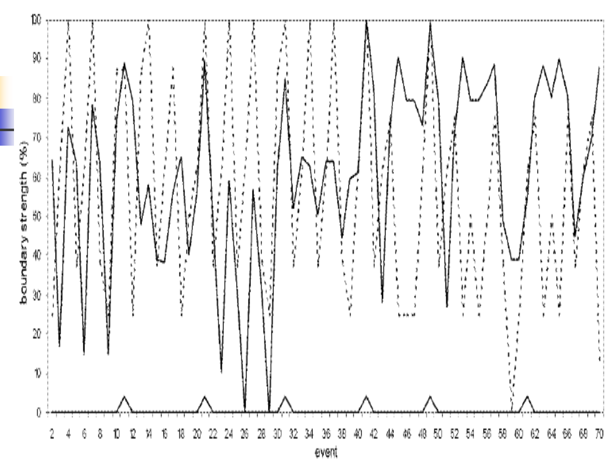
- Attenuation function where t_i = onset time of e_i and M is duration of memory window

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

Aims

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

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.

Example

No.	Pitch	Onset	Dur	PS	DR
1	82	3998	1000	-1	-5 (0.14)
2	81	4998	143	+2	-1 (0.80)
3	83	5141	115	-3	6 (8.70)
4	80	5256	1000	-1	-5 (0.17)
5	79	6256	167	+2	0 (0.99)
6	81	6423	166	-2	1 (1.64)
7	78	6589	273	-3	0 (1.18)
8	77	6862	322	…	…

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:

$$P(s) = \prod_{i=1}^{l} P(w_i \mid w_{i-1}, \ldots, w_{i-n+1})$$

This causes problems if any of the terms has zero probability

Mixed Markov Models:

$$P(w_i \mid w_{i-1}, \ldots, w_{i-n}) = \sum_{\mu=1}^{n} \phi(\mu) a^{\mu}(w_i \mid w_{i-\mu})$$

Mixed Markov Models (contd.)

- a_(w_i|w_i-_) 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-_*.
- Mixing coefficients _(_) are estimated using an iterative process

Entropy and Boundary Prediction

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

$$H_c = \sum_{\forall w} P(w \mid c) \log_2 P(w \mid c)$$

- 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.

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.