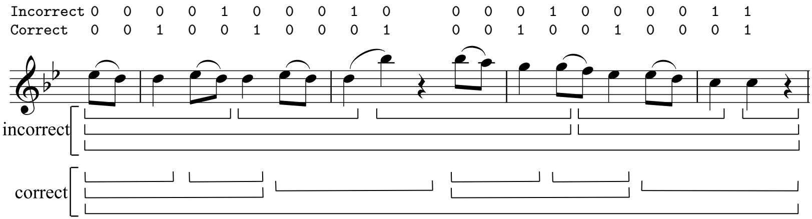
An information-dynamic model of melodic segmentation

Marcus T. Pearce, Daniel Müllensiefen and Geraint A. Wiggins

Centre for Cognition, Computation and Culture
Goldsmiths, University of London, UK

- melodic grouping (single level)
- e.g., Mozart Symphony 40 in G minor [Lerdahl and Jackendoff, 1983]

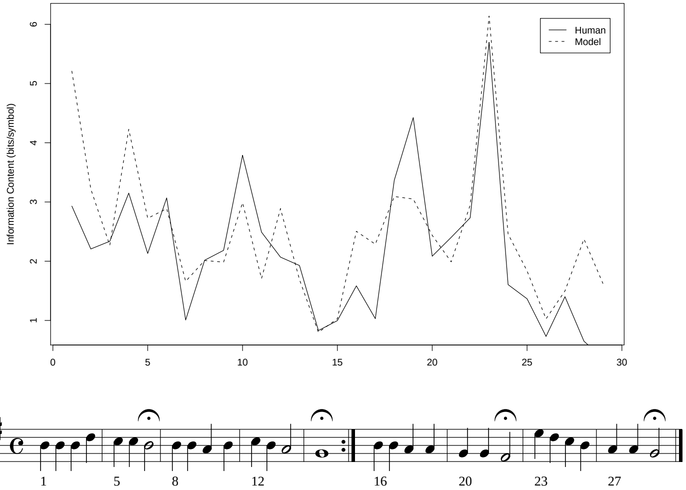


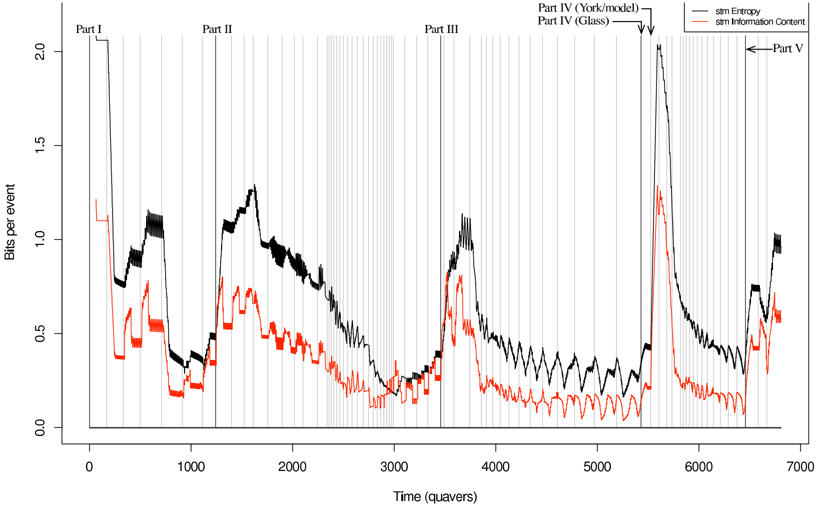
- musicology: groups associated with uncertainty and expectation violation [Narmour, 1990, Meyer, 1957]
- sequences of elements e from an alphabet $\mathcal{E}$
- model: $p(e_i|e_1^{i-1})$
- Information content (unexpectedness):

$$h(e_i|e_1^{i-1}) = \log_2 \frac{1}{p(e_i|e_1^{i-1})}.$$

Entropy (uncertainty):

$$H(e_1^{i-1}) = \sum_{e \in \mathcal{E}} p(e_i|e_1^{i-1}) h(e_i|e_1^{i-1}).$$


Human
Model
Information Content (bits/symbol)
1
2
3
4
5
6
0
5
10
15
20
25
30
1
5
8
12
16
20
23
27


Part I
Part II
Part III
Part IV (York/model)
Part IV (Glass)
Part V
stm Entropy
stm Information Content
Bits per event
0.0
0.5
1.0
1.5
2.0
0
1000
2000
3000
4000
5000
6000
7000
Time (quavers)

- *n*-gram model
- combines models of different order
- uses a long- and short-term models
- estimates the probability of an event based on its pitch and onset (IOI and preceding silence).
- 10-fold cross-validation for training/testing
- [Pearce, 2005]

- focus here on information content profile
- we interpret this as a boundary strength profile s_n
- pick peaks in the profile at locations where:
  - $S_n > S_{n-1}$
  - $S_n \geq S_{n+1}$
  - $$S_n > k\sqrt{\frac{\sum_{i=1}^{n-1}(w_i S_i - \overline{S}_{w,1\dots n-1})^2}{\sum_1^{n-1} w_i}} + \frac{\sum_{i=1}^{n-1} w_i S_i}{\sum_1^{n-1} w_i}$$

Experimental Psychology: infants/adults segment syllable/tone sequences on the basis of local statistics [Saffran et al., 1999]

Cognitive Linguistics: difficulty of word comprehension related to information content [Levy, 2008] and entropy [Hale, 2006];

Machine Learning: algorithms based on information content and entropy can identify word boundaries with some success [Brent, 1999, Cohen et al., 2007]

IDyOM: with $k = 2$

Grouper: [Temperley, 2001]

LBDM: [Cambouropoulos, 2001] with $k = 0.5$

GPR2a: [Lerdahl and Jackendoff, 1983] with $k = 0.5$

GPR2b: [Lerdahl and Jackendoff, 1983] with $k = 0.5$

GPR3a: [Lerdahl and Jackendoff, 1983] with $k = 0.5$

GPR3d: [Lerdahl and Jackendoff, 1983] with $k = 2.5$

Always: every note falls on a boundary

Never: no note falls on a boundary

k optimised from set $\{0.25, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$

- Essen Folk Song Collection: Erk
  - 1705 German folk melodies
  - 78,995 sounding events
  - average of 46 events per melody
- annotated with phrase boundaries by musicologists
  - 12% of notes fall on boundaries

Model	Precision	Recall	F1
Hybrid	0.87	0.56	0.66*
Grouper	0.71	0.62	0.66*
LBDM	0.70	0.60	0.63*
GPR2a	0.99	0.45	0.58*
IDyOM	0.76	0.50	0.58*
GPR2b	0.47	0.42	0.39
GPR3a	0.29	0.46	0.35
GPR3d	0.66	0.22	0.31
Always	0.13	1.00	0.22
Never	0.00	0.00	0.00

1. GPR2a does well - importance of rests
2. LBDM/Grouper comparable to other studies
3. Hybrid model outperforms its component models
4. information dynamic model performs surprisingly well
   - developed as a model of pitch prediction
   - not optimised for melodic grouping

- focus on boundaries not indicated by rests
- use boosting to create a hybrid model
    - optimise viewpoints for segmentation
    - other information dynamic measures
        - entropy
        - predictive information
- explicit Bayesian models of phrase segmentation [Brent, 1999]

References I

Brent, M. R. (1999).
An efficient, probabilistically sound algorithm for segmentation and word discovery.
Machine Learning, 34(1-3):71–105.

Cambouropoulos, E. (2001).
The local boundary detection model (LBDM) and its application in the study of expressive timing.
In *Proceedings of the International Computer Music Conference*, pages 17–22, San Francisco. ICMA.

Cohen, P. R., Adams, N., and Heeringa, B. (2007).
Voting experts: An unsupervised algorithm for segmenting sequences.
Intelligent Data Analysis, 11(6):607–625.

Hale, J. (2006).
Uncertainty about the rest of the sentence.
Cognitive Science, 30(4):643–672.

Lerdahl, F. and Jackendoff, R. (1983).
A Generative Theory of Tonal Music.
MIT Press, Cambridge, MA.

Levy, R. (2008).
Expectation-based syntactic comprehension.
Cognition, 16(3):1126–1177.

References II

Meyer, L. B. (1957).
Meaning in music and information theory.
Journal of Aesthetics and Art Criticism, 15(4):412–424.

Narmour, E. (1990).
The Analysis and Cognition of Basic Melodic Structures: The Implication-realisation Model.
University of Chicago Press, Chicago.

Pearce, M. T. (2005).
The Construction and Evaluation of Statistical Models of Melodic Structure in Music Perception and Composition.
PhD thesis, Department of Computing, City University, London, UK.

Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (1999).
Statistical learning of tone sequences by human infants and adults.
Cognition, 70(1):27–52.

Temperley, D. (2001).
The Cognition of Basic Musical Structures.
MIT Press, Cambridge, MA.