

Perceptual Segmentation of Melodies: Ambiguity, Rules and Statistical Learning

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

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

- 1 examine perceptual mechanisms in melodic grouping
- 2 particularly in pop music
 - unlike folk music, relatively poorly studied
 - need to segment 14,000 pop songs
- 3 ambiguity of the grouping percept
- 4 investigate unsupervised learning models

The Task

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

Incorrect	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	1	1	
Correct	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	1

The image shows a single staff of music in G minor (one flat). The melody consists of 20 notes. Below the staff, two sets of brackets illustrate different melodic groupings. The 'incorrect' grouping shows brackets that do not align with the natural phrasing of the melody. The 'correct' grouping shows brackets that group notes into phrases that correspond to the musical structure, such as grouping the first four notes together, the next four notes together, and so on.

Model Overview

- Gestalt rule-based:
 - GTTM GPRs [Lerdahl and Jackendoff, 1983, Frankland and Cohen, 2004]
 - GPR2a
 - GPR2b (revised)
 - GPR3a
 - GPR3d
 - LBDM [Cambouropoulos, 2001]
 - Grouper [Temperley, 2001]
- statistical models
 - TP [Saffran et al., 1999, Brent, 1999, Narmour, 1990]
 - IDyOM [Pearce et al., 2008]
- default
 - always
 - never

Model Overview

	Rule-based	Unsupervised Learning
Local	GPRs, LBDM	TP
Global	Grouper	IDyOM

Performance accuracy:

- Grouper > LBDM [Thom et al., 2002]
- LBDM > GPRs [Bruderer, 2008]

- model output is interpreted as a boundary strength profile S
- pick peaks in the profile at locations where:
 - $S_n > S_{n-1}$
 - $S_n \geq S_{n+1}$
 -

$$S_n > k \sqrt{\frac{\sum_{i=1}^{n-1} (w_i S_i - \bar{S}_{w,1\dots n-1})^2}{\sum_{i=1}^{n-1} w_i}} + \frac{\sum_{i=1}^{n-1} w_i S_i}{\sum_{i=1}^{n-1} w_i}$$

The Models

Grouper: [Temperley, 2001]

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

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

GPR2br: [Lerdahl and Jackendoff, 1983] with $k = 0.25$

GPR3a: [Lerdahl and Jackendoff, 1983] with $k = 0.25$

GPR3d: [Lerdahl and Jackendoff, 1983] with $k = 0.25$

TP: [Saffran et al., 1999] with $k = 0.25$

IDyOM: with $k = 1$

Always: every note falls on a boundary

Never: no note falls on a boundary

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

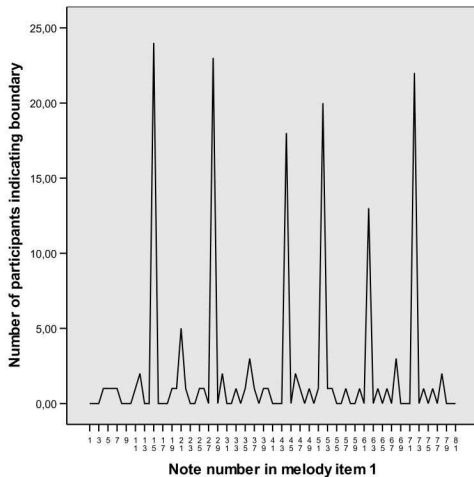
Participants and Materials

- Participants: 25 musically trained adults,
- Task: indicate phrase boundary strength (on 3-point scale) while listening; 2 consecutive listenings for each melody
- Material: 15 monophonic melodies from pop or folk songs, 50-132 notes at natural tempo, MIDI piano renditions

- How to aggregate the participants' responses?
- Majority vote?
 - low inter-rater agreement (8 melodies with kappa < 0.6)
 - Assumes a single underlying segmentation solution.

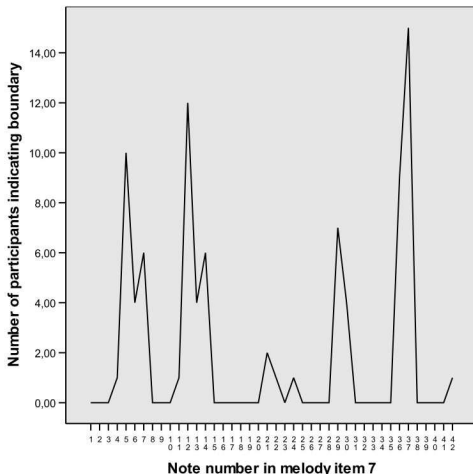
Ambiguity

Majority-vote gives incomplete segmentation solutions:



Ambiguity

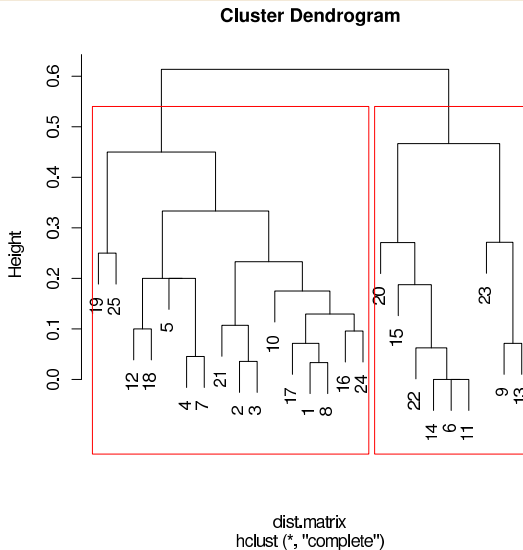
Majority-vote doesn't distinguish between equally valid, competing segmentations:



A Solution:

- cluster participants into groups for each melody
 - Kulczynski distance
 - maximum number of clusters is 5
 - clusters must have more than 3 participants
 - clusters should be compact
- generate representative segmentation for each cluster
 - sum boundary indications across participants for each note
 - k -means clustering ($k = 2$);
- test each model on the cluster it performs best on

Ambiguity



Ambiguity

Melody no.	No. clusters	K_c cl.1	K_c cl.2	K_c cl.3	K_c cl.4	K_c cl.5	No. part. excluded
1	1	.81		2			5
2	3	.87	.80	.61			0
3	3	.79	.81	.90	.76		2
4	3	.74	.64	.45	.60		9
5	3	.81	.68	.81			3
6	3	.87	.66	.81			5
7	4	.76	.76	.80	.64		4
8	2	.77	.78				0
9	1	.75	.55				4
10	2	.73	.64	.55			8
11	4	.72	.74	.71	.70	.70	1
12	1	.83					3
13	3	.54	.88	.74	.79		11
14	3	.89	.73	.84			0
15	3	.57	.63	.68	.60	.59	9

Model Performance

Model	Precision	Recall	F1
Grouper	0.86	0.82	0.83
LBDM	0.79	0.81	0.78
IDyOM	0.57	0.73	0.64
GPR2a	0.70	0.54	0.58
GPR2br	0.47	0.45	0.43
TP	0.25	0.45	0.31
GPR3a	0.26	0.43	0.30
Always	0.13	1.00	0.23
GPR3d	0.17	0.11	0.11
Never	0.00	0.00	0.00

significant: Grouper/GPR2a, Grouper/IDyOM ($F1, p < .05$)

Hybrid Model

- two sets of clusters: high- and low-level
- select models with $F1 > 0.5$
- logistic regression model
- stepwise forward selection
- raw boundary strength profiles

low-level segmentation

Model	Precision	Recall	F1
Hybrid	0.88	0.66	0.73
Grouper	0.77	0.62	0.66

high-level segmentation

Model	Precision	Recall	F1
Hybrid	0.78	0.70	0.74
Grouper	0.60	0.77	0.66

Summary

- 1 GPR2a does well - importance of rests
- 2 Grouper > LBDM > GPRs: comparable to other studies
- 3 Hybrid model able to perform better than Grouper
- 4 IDyOM model performs surprisingly well (better than TP)
 - developed as a model of pitch prediction [Pearce, 2005]
 - not optimised for melodic grouping

- focus on boundaries not indicated by rests
- use boosting to create better hybrid models
- improve IDyOM model
 - optimise viewpoints for segmentation
 - other information dynamic measures
 - entropy
 - predictive information
- use hybrid models to segment 14,000 pop songs

Thanks ...

... for listening! Any questions?

References I



Brent, M. R. (1999).

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



Bruderer, M. J. (2008).

Perception and Modeling of Segment Boundaries in Popular Music.
PhD thesis, J.F. Schouten School for User-System Interaction Research, Technische Universiteit Eindhoven, Netherlands.



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.



Frankland, B. W. and Cohen, A. J. (2004).

Parsing of melody: Quantification and testing of the local grouping rules of Lerdahl and Jackendoff's *A Generative Theory of Tonal Music*.
Music Perception, 21(4):499–543.



Lerdahl, F. and Jackendoff, R. (1983).

A Generative Theory of Tonal Music.
MIT Press, Cambridge, MA.



Narmour, E. (1990).

The Analysis and Cognition of Basic Melodic Structures: The Implication-realisation Model.
University of Chicago Press, Chicago.

References II



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.



Pearce, M. T., Müllensiefen, D., and Wiggins, G. A. (2008).

A comparison of statistical and rule-based models of melodic segmentation.

In *Proceedings of the Ninth International Conference on Music Information Retrieval*, Philadelphia, USA.



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.



Thom, B., Spevak, C., and Höthker, K. (2002).

Melodic segmentation: Evaluating the performance of algorithms and musical experts.

In *Proceedings of the 2002 International Computer Music Conference*, San Francisco. ICMA.