Goldsmiths
UNIVERSITY OF LONDON

Statistical Models of Music Cognition

Geraint A. Wiggins, Daniel Müllensiefen & Marcus T. Pearce
Intelligent Sound and Music Systems Group
Goldsmiths, University of London

Outline

Goldsmiths
UNIVERSITY OF LONDON

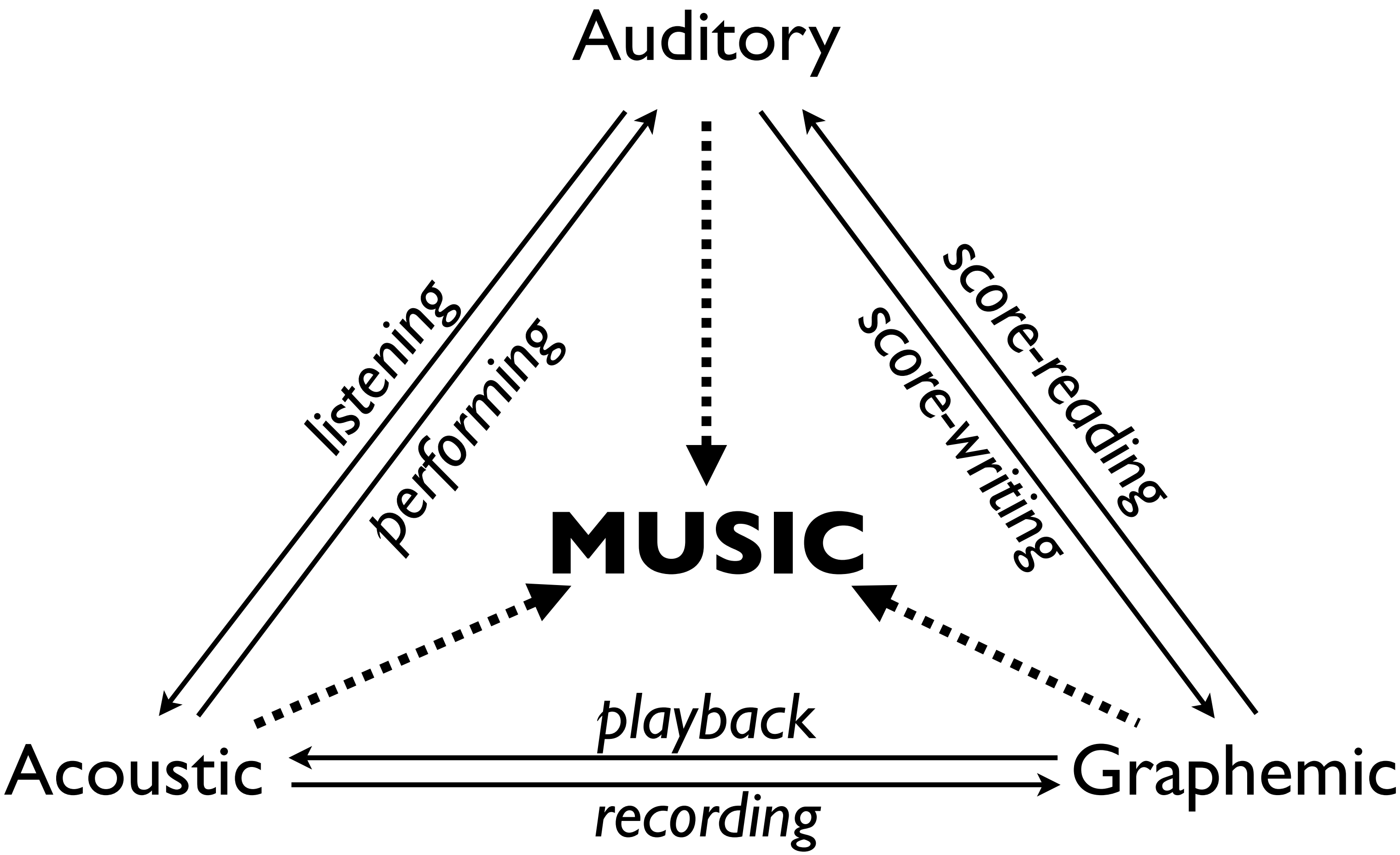
1. General methodology: statistical corpus-based musicology
2. Musical similarity
3. Melodic pitch expectation
4. Melodic segmentation
5. Specific methodology
6. Summary and future work

Part I

General metholodogy

General methodological issues

- What is music (for the purposes of this work)?



General methodological issues

Goldsmiths
UNIVERSITY OF LONDON

- Study music entirely as **cognitive/perceptual phenomenon**
  - ask whether perceptual primitives and learning together can give rise to music
- Approach based on **unsupervised learning** over a musical surface based on **perceptual primitives**
  - no hand-coded rules
  - no appeal to music theory beyond notes and intervals
- Attempt to model **low-level cognition of musical structure** - at level as close to perception as possible
  - no imposed high-level structural rules
- **Ockham's Razor:** simple models wherever possible

General methodological issues

Goldsmiths
UNIVERSITY OF LONDON

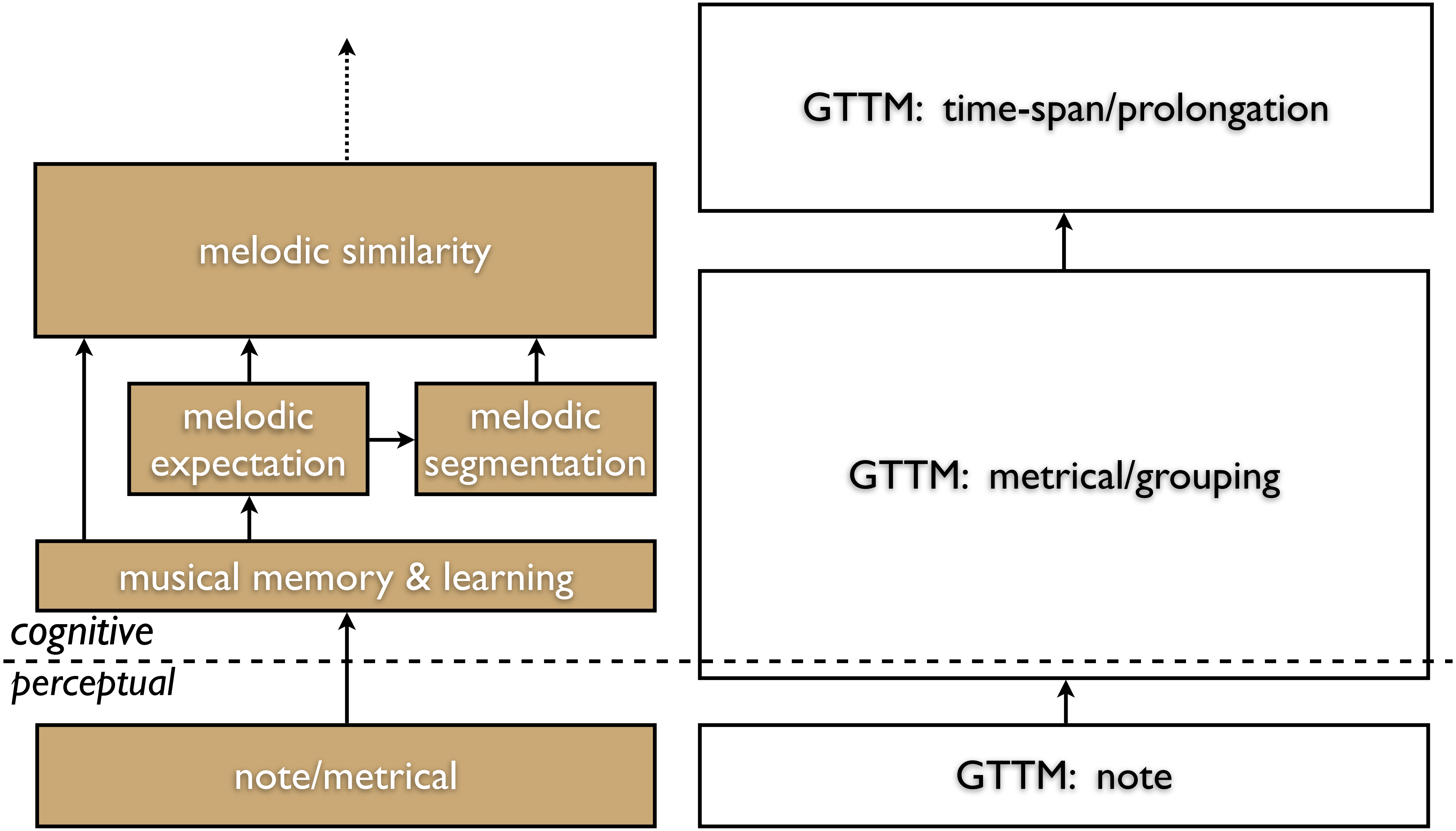
- Use **corpus-based methods** involving **large bodies of data** from which **generalisations** can be made
  - borrow **successful methods** from **computational linguistics**
  - use theories of **implicit learning**
- Use **general models** of learning applied to **specific data**
  - achieve **explanatory** model where the process is explained as well as the effect (descriptive/explanatory, explanandum/explanation, final/efficient cause)
- **Reuse existing models** to explain **further (related) effects**
  - this strengthens case for original model (Popper/Honing)
  - we call this **meta-modelling** (μετα, Gr. *beyond*)

General methodological issues

Goldsmiths
UNIVERSITY OF LONDON

- Apply **reductionist scientific methods** but in a **musically realistic** way
  - use music written by **real composers**
  - choose music **suitable for particular enquiries and experiments**
  - use **complex sound** (not sine waves)
  - **avoid artificial alterations** in music
  - maintain as **normal** a **listening environment** as possible
- In short, **maximise ecological validity**

Relation to GTTM



Part 2

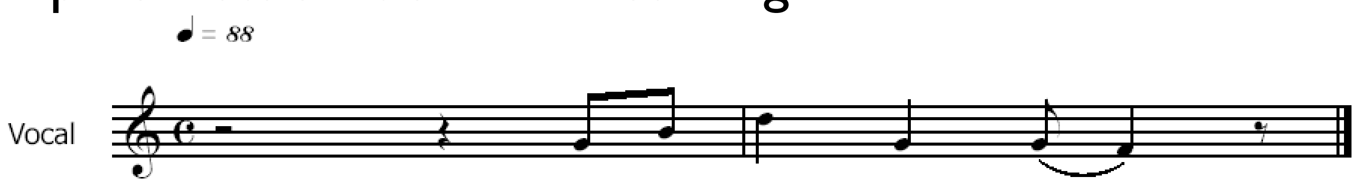
Goldsmiths
UNIVERSITY OF LONDON

Melodic similarity

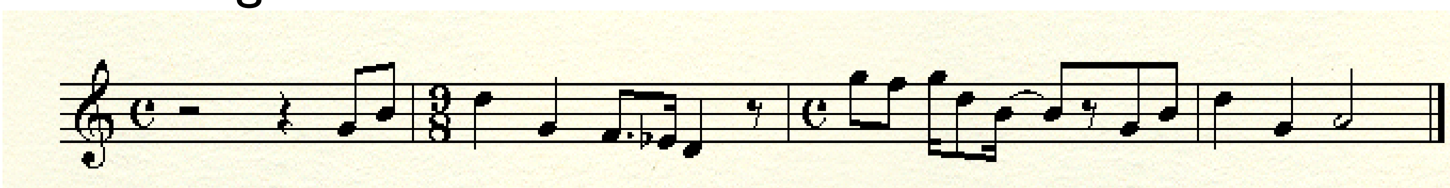
Models of musical memory

- Unfamiliar tune

- Memory representation after 2nd listening



- After 6th listening



Models of musical memory

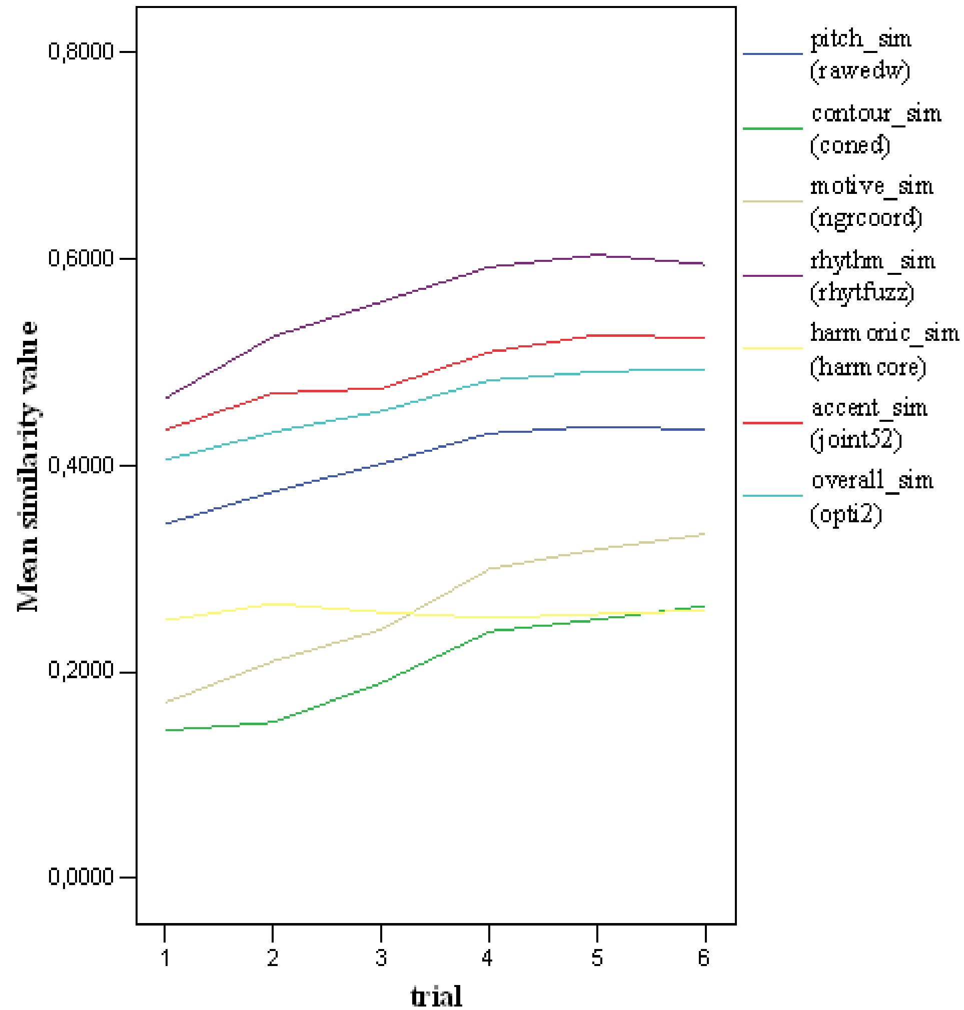
Goldsmiths
UNIVERSITY OF LONDON

- Can we predict melodic memory?
- What is remembered?
- How does memory improve over repeated listenings?
- Which musical features enhance memory?

Melodic similarity as a tool

Goldsmiths
UNIVERSITY OF LONDON

- 10 subjects (740 observations)
- Rank correlation between similarity and trial number for each dimension:
  - Micro-Motives: 0.252**
  - Rhythm: 0.219**
  - Accents: 0.209**
  - Contour: 0.196**
  - Pitch: 0.14**
  - Implied harmony: .005
- Interpretations:
  - Similar learning curves for melody structure and details
  - Representation of harmonic structure depends on musical background



Approaching melodic similarity

Goldsmiths
UNIVERSITY OF LONDON

- The SIMILE toolbox (Müllensiefen & Frieler, 2004, 2007):
    - Systematise, implement, and combine many suitable similarity algorithms and transformations for melodic data
    - Test algorithms against data from psychological experiment
    - Algorithms: Edit Distance, substring frequency comparisons, vector correlations and differences
    - Transformations: Interval and rhythm classification, contour, harmonic implications, accent weighting

Approaching melodic similarity

Goldsmiths
UNIVERSITY OF LONDON

- SIMILE's 50 Similarity Measures

Raw pitch edit distance
Raw pitch edit distance weighted
Raw pitch Pears. Brav. correlation
Raw pitch P-B. corr, weighted, 0-1
Raw pitch Pears. Brav. Corr. Weighted
Raw pitch P-B. Corr. weighted, 0-1
Raw pitch crosscorrelation
Raw pitch crosscorrelation weighted
Contour (Steinbeck) edit distance
Contour (Steinbeck), P-B. correlation
Contour (Steinbeck), P-B. corr., 0-1
Contour (Steinbeck), Crosscorrelation
Contour, Edit distance
Contour, Pearson-Bravais correlation
Contour, Pearson-Bravais corr., 0-1
Contour, Crosscorrelation
Fourier (ranks)
Fourier (ranks), weighted, 0-1
Fourier (ranks), weighted
Fourier (ranks), weighted, 0-1
Fourier (ranks, intervals)
Rhythm (gaussified onset points)
Rhythm (fuzzy, Edit distance)
Accent similarity measure

Intervals (Edit distance)
Intervals (Mean difference)
Intervals (Mean difference, exp.)
Intervals (fuzzy), Edit Distance
Intervals (fuzzy contour)
n-grams Sum Common (intervals)
n-grams Ukkonnen (intervals)
n-grans Coordinate Matching (intervals)
n-grams Sum Common (interval dir.)
n-grams Ukkonnen (interval dir.)
n-grams Coord. Match. (interval dir.)
n-grams Sum Common (fuzzy int.)
n-grams Ukkonnen (fuzzy int.)
n-grams Count distinct (fuzzy int.)
n-grams sum common (fuzzy rhythm)
n-grams Ukkonnen (fuzzy rhythm)
n-grams Coord. Match. (fuzzy rhythm)
Selfridge-Field (max.)
Selfridge-Field (modus I)
Selfridge-Field (modus II)
Selfridge-Field (signs)
Harmonic correlation (type I)
Harmonic correlation (type II)
Harmonic correlation (Edit distance)
Harmonic correlation (circle)

Approaching melodic similarity

Goldsmiths
UNIVERSITY OF LONDON

- What is missing in SIMILE?
  - A model of experience with melodies from the real world (all previous similarity measures behave as though they had never seen a melody before)
- How to model melodic knowledge in a similarity measure?
  - Term Frequency - Inverted Document Frequency (TF-IDF) measures
  - Idea: Take statistical frequency of melodic feature / formula into account when comparing melodies for feature
  - Prerequisite: Corpus of melodies that is representative for style

Approaching melodic similarity

Goldsmiths
UNIVERSITY OF LONDON

- Example: TF-IDF measures for substrings (interval transformation)

A rare motive

More diagnostic,
if present in two melodies
(f = 7.1×10^{-5})

A common motive

Less diagnostic,
if present in two melodies
(f = 4.9×10^{-3})

Approaching melodic similarity

Goldsmiths
UNIVERSITY OF LONDON

TF-IDF similarity for melodies s, t from corpus C:

$$\sigma_{C;n}(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)(TF_s(\tau) \cdot TF_t(\tau))}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}$$

TF-IDF correlation similarity:

$$\sigma_{C;n}(s,t) = \frac{\sum_{\tau=1}^{N} TFIDF_{(s,C)}(\tau) \cdot TFIDF_{(t,C)}(\tau)}{\sqrt{\sum_{\tau=1}^{N} (TFIDF_{(s,C)}(\tau))^2 \cdot \sum_{\tau=1}^{N} (TFIDF_{(t,C)}(\tau))^2}}$$

with Term-Frequency for term τ in melody m with m_n different terms:

$$TF_{(m,\tau)} = \frac{f_m(\tau)}{\sum_{m_n} f_m(m_n)}$$

with combined TF-IDF weighting:

$$TFIDF_{(m,C)}(\tau) = TF_{(m)}(\tau) \cdot IDF_C(\tau)$$

and Inverted Document Frequency for term τ in corpus C with |C| melodies:

$$IDF_C(\tau) = \begin{cases} \log \frac{|C|}{|m:\tau \in m|} & \exists m \in C : \tau \in m \\ \log \frac{|C+2|}{2} & else \end{cases}$$

Part 3

Goldsmiths
UNIVERSITY OF LONDON

Melodic expectation

Expectation in music

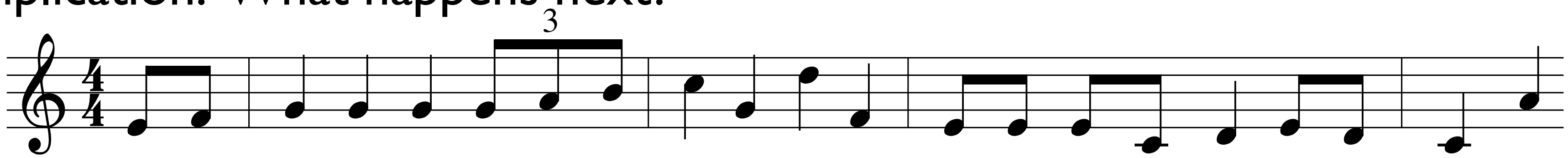
Goldsmiths
UNIVERSITY OF LONDON

- Introduction: expectation in music
- A statistical learning account
    - theory
    - model
- Results

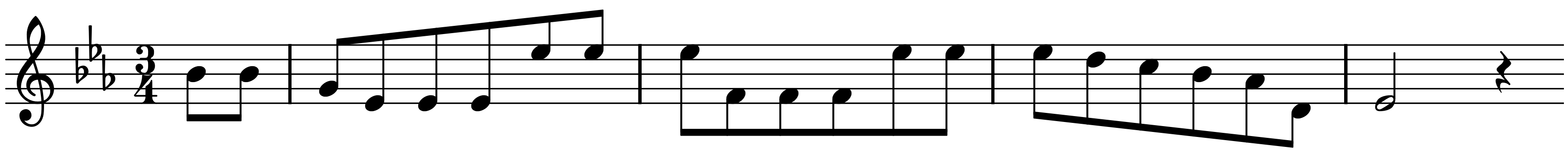
Expectation in music

Goldsmiths
UNIVERSITY OF LONDON

- Implication: What happens next?



- Closure: What about here?



- Focus on:
  - monody
  - pitch (interval)

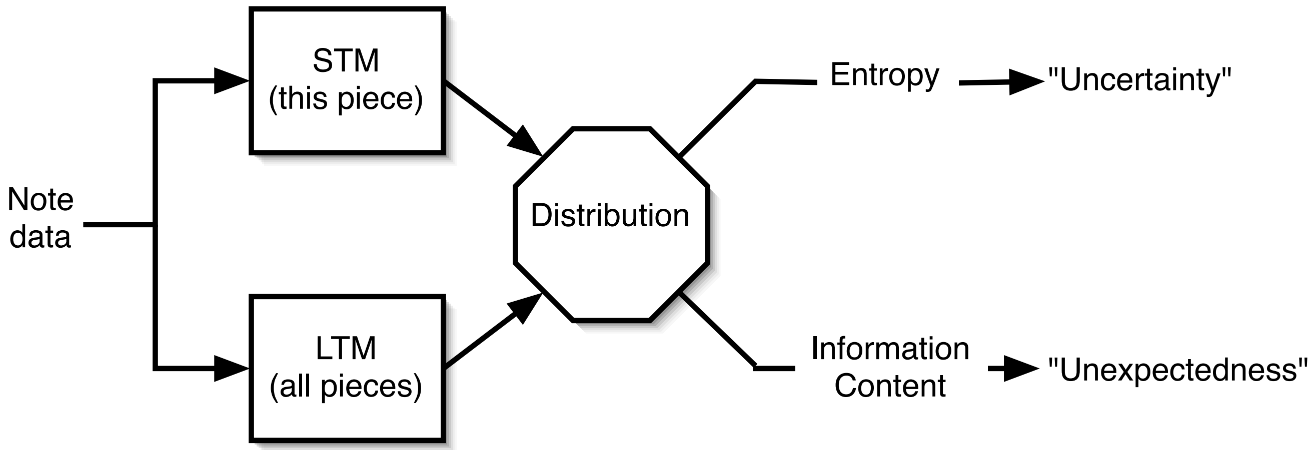
Why study expectation?

Goldsmiths UNIVERSITY OF LONDON

- Theoretical Perspective (Meyer)
    - aesthetic experience
    - communication of emotion and meaning
- Empirical Perspective:
    - recognition memory for music
    - production of music
    - perception of music
    - transcription of music
- **Can a purely unsupervised statistical model account for observed patterns of expectation as well as other models?**

The IDyOM model

Goldsmiths
UNIVERSITY OF LONDON



- STM: n-gram (arbitrary n) model
  - complex backoff/smoothing strategy
  - dynamic weighting of features used for prediction, according to information content
- LTM: same as STM
  - but trained with a database of >900 tonal melodies

The IDyOM model

Goldsmiths UNIVERSITY OF LONDON

- No domain-specific a priori rules
- STM and LTM can be used independently or together
- “optimised” for pitch expectancy prediction

Learning data

Goldsmiths
UNIVERSITY OF LONDON

- Chosen to reflect a range of Western tonal musical styles:

Description	Compostions	Events	Mean event/composition
Canadian folk songs	152	8,553	56.27
Chorale melodies	185	9,227	49.88
German folk songs	566	33,087	58.46
Total	903	50,867	56.33

Method: model comparison

Goldsmiths
UNIVERSITY OF LONDON

- Compare with two-factor+tonality model of Schellenberg (1997) against behavioural data from
  - Cuddy & Lunney (1995): single-interval context
  - Schellenberg (1996): within British folk songs
  - Manzara et al (1992): throughout chorale melodies
- Criteria (Cutting et al., 1992)
  - Scope: compare correlations with behavioural data
  - Selectivity: compare performance on random data
  - Simplicity: multiple regression analysis of nested models

Results

Goldsmiths
UNIVERSITY OF LONDON

Scope	Two-factor R	IDyOM R	t-test of difference
Cuddy & Lunney (1995)	0.83	0.85	n.s.
Schellenberg (1997)	0.87	0.91	$p < 0.05$
Manzara et al (1992)	0.41	0.80	$p < 0.01$

- IDyOM accounts for the data at least as well as the two-factor model

Part 4

Goldsmiths
UNIVERSITY OF LONDON

Melodic segmentation

Melodic segmentation

Goldsmiths
UNIVERSITY OF LONDON

- **Phrase**
  - (most?) important **unit of melodic content**
  - often relates to **practical musical parameters**, e.g. articulation, breathing, agogics, tempo change
  - **smallest melodic unit** where many features can be meaningfully computed, e.g. contour, length, complexity, event density, implied harmony
  - (feature description approach in melodic similarity work depends on phrases)
- How can we segment a melody into musically meaningful phrases?
- What musical information contributes to cognitive judgements of boundaries?

Gestalt rules & models

Goldsmiths UNIVERSITY OF LONDON

- Strong support that perceived boundaries are influenced by:
  - the presence of rests or pauses (GPR2a)
  - the presence of notes with relatively long duration or inter-onset interval (GPR2b)
- Less definite support for other dimensions
- Many symbolic, rule-based segmentation models in literature
  - Lerdahl & Jackendoff, 1983; Deliège 1987-1997; Rowe, 1993; Friberg et al., 1998; Cambouropoulos, 1998-2008; Temperley, 2001; Ahlbäck, 2004; Weyde, 2004; etc
- A few evaluation studies published, but as phrase boundaries are in many instances not consistently given by composer, they use
  - Analytical annotations by 1 subject (e.g. Bod, 2001)
  - Experimentally collected ratings by several subjects (e.g. Deliège, 1987-98; Thom et al., 2001; Melucci & Orio, 2002; Spiro & Klebanov, 2006; Bruderer et al., 2008)

Segmentation from expectation

Goldsmiths
UNIVERSITY OF LONDON

- Evidence from computational linguistics for a statistical segmentation strategy
  - Statistical learning of syllable/tone sequences (Saffran et al., 1996, 1999)
  - Predicting word boundaries in speech processing (Brent, 1999)
- Perceptual groups associated with points of closure where expectations are weak (Meyer, 1957; Narmour, 1990)
- In information theoretic terms (Shannon, 1948):
  - Uncertainty: entropy ≈ information content of a probability distribution over events predicted from context
  - Unexpectedness: information content (or low probability) of an event in context
- **Our hypothesis:**
  - Closure: increasing certainty followed by lack of certainty ≈
    decreasing entropy/IC followed by (relatively) high entropy/IC

The study

Goldsmiths
UNIVERSITY OF LONDON

- What's new here?
  - Comparing programmed Gestalt models and information theoretic segmenters
  - Explicit search for optimal hybrid model
  - Explicit distinction between items (melodies) with low/high agreement
  - Explicit search for consistent but diverging rating patterns

The models

Goldsmiths
UNIVERSITY OF LONDON

- Gestalt, programmed-rule-based:
  - GPRs (Lerdahl & Jackendoff, 1983; Frankland & Cohen, 2004)
  - Grouper (Temperley, 2001)
  - LBDM (Cambouropoulos, 2001)
  - SimpleSegmenter (Müllensiefen & Frieler, 2004)
- Information theoretic (no programmed musical knowledge)
  - Saffran et al (1999)
  - IDyOM (now)

The method

Goldsmiths UNIVERSITY OF LONDON

- Subjects: 25 adults, musicology (graduate) students, mean age: 28.4 (Std: 8), mean years playing instrument: 16.4 (Std: 9), mean number of paid gigs 36.3 (Std: 60.6), mean months paid instrumental lessons: 100.8 (Std: 72.3), mean practice hours in most active musical phase: 27.4 (Std: 17.5)
- Preliminary task: Indicate phrase boundary strength (on 3-point scale) while listening; 2 consecutive listenings for each melody
- Definition of phrase boundary: end of musical segment where a performer would make phrase indication; tested in group
- Material: 15 monophonic melodies from pop or folk songs, 50-132 notes at natural tempo, MIDI piano renditions
- Questionnaire about musical background
- Dependent variables: binary indicator of majority vote

The study

Goldsmiths
UNIVERSITY OF LONDON

- Obtain segmentations of each of 15 melodies by human experts
- Check consistency of ratings between subjects for each melody using κ measure for inter-rater agreement (Fleiss, 1971; Spiro & Klebanov, 2006)
- 7 melodies with $\kappa \geq .6$ on binary ratings ('moderate agreement', Landis & Koch, 1977) taken as reliable melodies
- 8 unreliable melodies ($\kappa < .6$) saved for later analysis
- **Main Task**: predict segmentation boundaries on notes where ≥ 50% of subjects agreed on boundary using segmentation algorithm
- Measures of accuracy: precision/recall; F1; d'; κ

Hybrid model

Goldsmiths UNIVERSITY OF LONDON

- Hybrid model combining several algorithms into logistic regression model, using stepwise model selection (Bayes' Information Criterion)
  - Predictors (binary): IDyOM, Grouper, LDBM, SimpleSeg, Saffran, GPRs
  - Criterion (binary): majority vote
  - Model: Logistic regression
  - Variable selection by model comparison using Bayes' Information Criterion
- Optimal model:

$$p(\text{boundary}) = \frac{1}{1 + e^{-(6.04\ .\ \text{Grouper} + 2.73\ .\ \text{IDyOM} - 5.17)}}$$

Comparing models

Goldsmiths
UNIVERSITY OF LONDON

	Precision (1 - specificity)	Recall (sensitivity)	F1	d’	*K*
never	0	0	-	-	-.04
saffran.p.pitch	.10	.04	.06	.14	.01
always	.08	1	.15	-	-.86
GPR2b	.13	.19	.15	.38	.07
GPR3a	.16	.37	.22	.69	.12
SimpleSeg.	.25	.35	.29	.99	.22
IDyOM	.60	.63	.61	2.15	.58
LBDM2001	.86	.57	.69	2.61	.67
GPR2a	.95	.55	.70	2.93	.68
Grouper	.67	.87	.76	2.94	.73

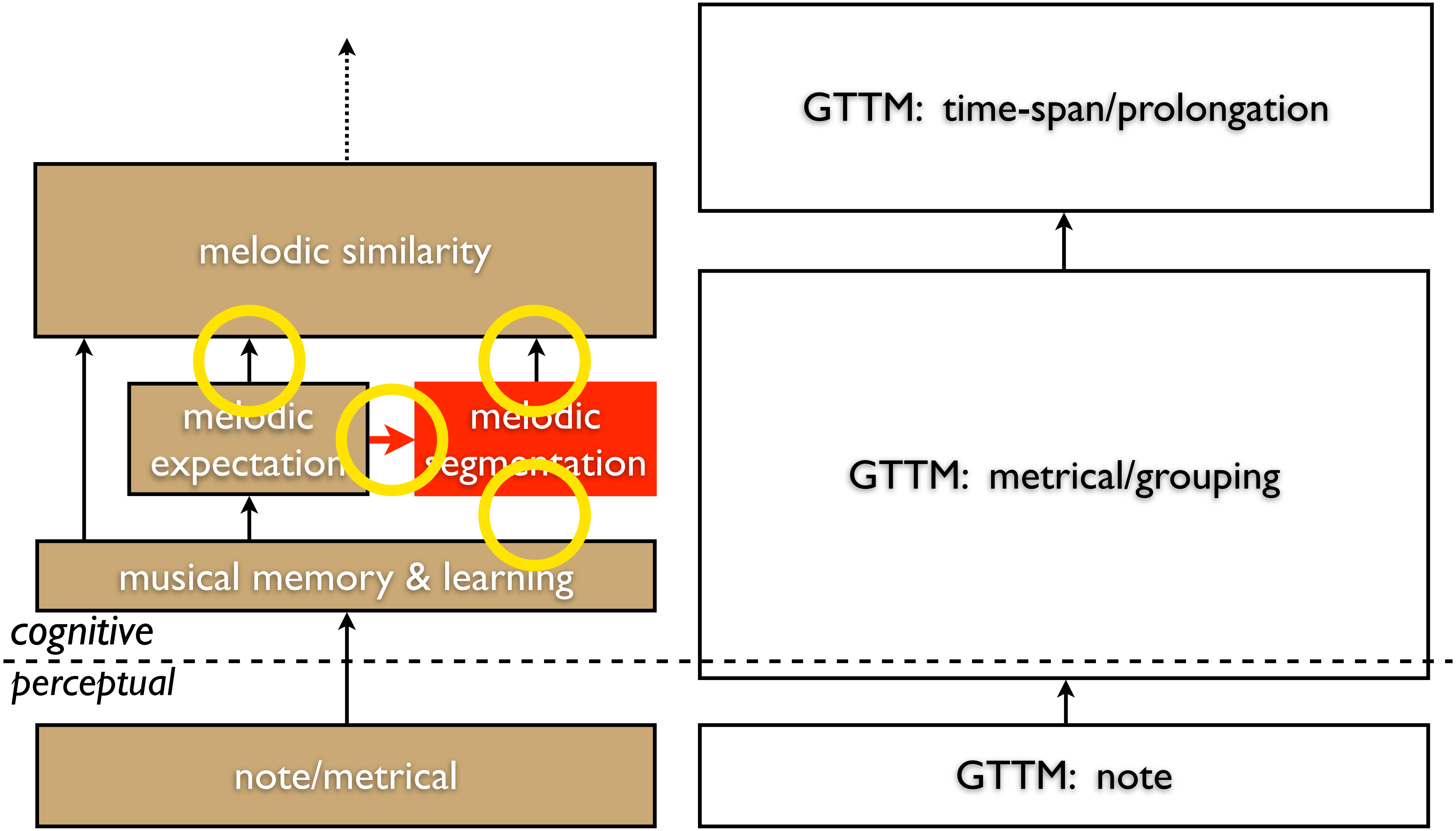
A meta-model

Goldsmiths UNIVERSITY OF LONDON

- We use the IDyOM expectation model as the basis for a **meta-model** for predicting **melodic segmentation**
  - a **meta-model** uses an existing model (without changing it) to predict a different, related phenomenon
  - IDyOM uses the **information-theoretic properties** of the **distributions** generated by the pitch expectation model
- We give this simple idea a name because the existence of meta-models adds evidence for the correctness of the models on which they are based

A meta-model

Goldsmiths UNIVERSITY OF LONDON



Part 5

Goldsmiths
UNIVERSITY OF LONDON

Specific methodology

Analysis methodology

Goldsmiths
UNIVERSITY OF LONDON

- 'Traditional' data analysis: 'Majority vote approach'
  - Add all subjects' boundary indications
  - Use threshold (e.g. 50%) to determine 'true' boundaries
  - Model 'true' boundaries only

Segmentation clustering

Goldsmiths UNIVERSITY OF LONDON

- Problems with majority vote:
  - Low subject agreement on certain melody items and exclusion of melody items from dataset
  - Incomplete segmentation solutions at fixed threshold

Segmentation clustering

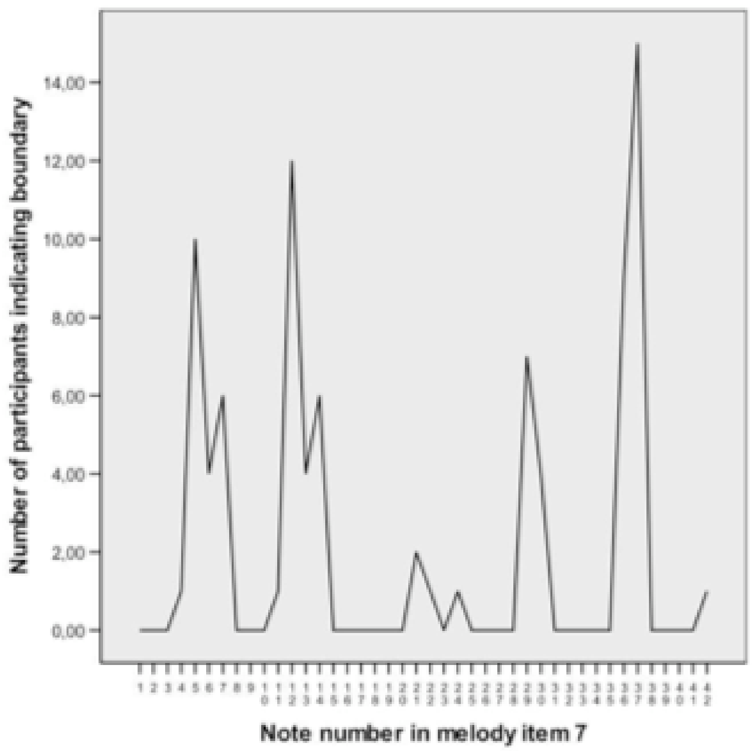
Goldsmiths
UNIVERSITY OF LONDON

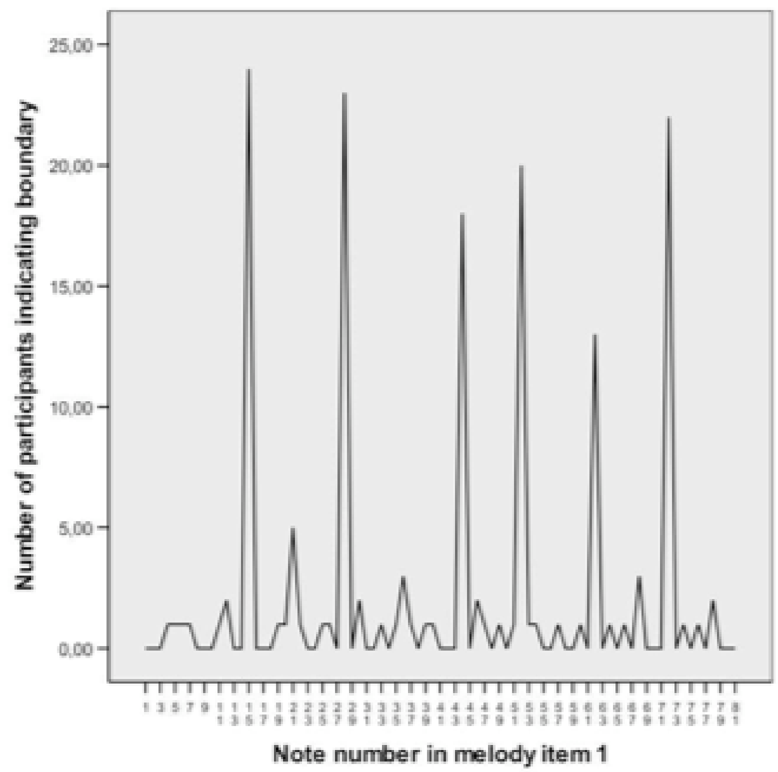
Melody no.	Mean no. boundaries	StDev. boundaries	Boundaries at 50% agreement	% of part. req. to agree f. mean no. of boundaries	K ≥ 0.6
1	6.88	4.30	6	21	-
2	9.00	3.13	7	42	+
3	4.13	1.84	3	43	-
4	8.74	3.63	4	43	-
5	9.96	2.12	8	46	+
6	9.78	3.22	9	30	+
7	3.82	1.37	1	41	-
8	10.48	3.22	10	56	+
9	9.64	2.69	10	64	+
10	4.39	1.73	3	52	-
11	9.36	3.58	7	36	-
12	7.84	1.82	8	80	+
13	3.08	1.51	1	28	-
14	9.72	3.10	8	40	+
15	11.16	4.44	9	36	-

Segmentation Clustering

Goldsmiths UNIVERSITY OF LONDON

- Problems with majority vote:
  - Incomplete segmentation solutions

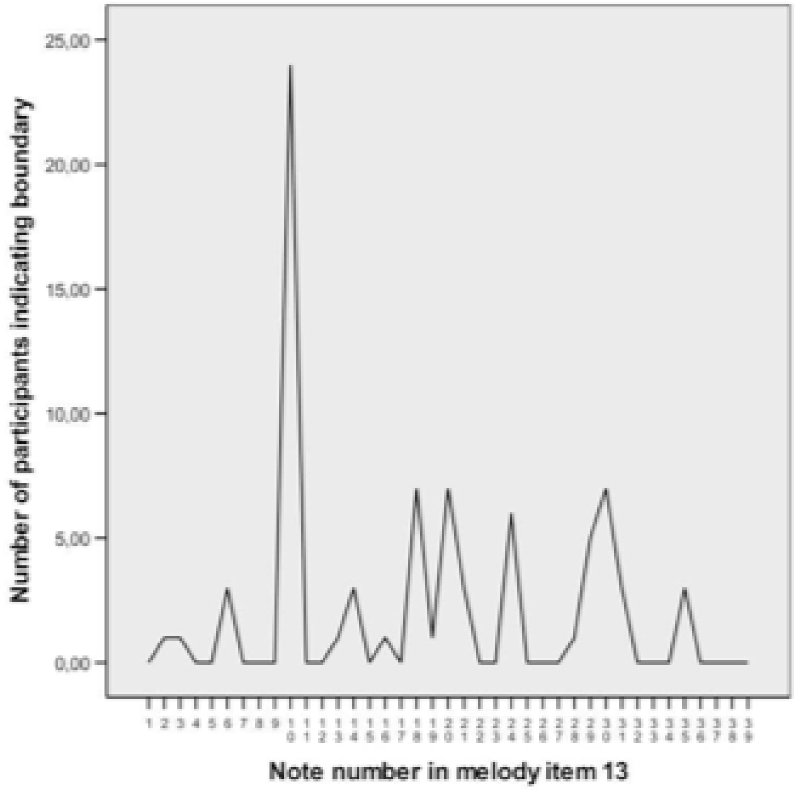




Segmentation Clustering

Goldsmiths UNIVERSITY OF LONDON

- Problems with majority vote:
  - No concept of multiple valid solutions



Segmentation Clustering

Goldsmiths UNIVERSITY OF LONDON

- Alternative approach to data analysis:
    - **Clustering of subjects with similar strategies**
- For each melody, measure similarity between ratings of all pairs of subjects using Jaccard distance
- Project distances onto lower-dimensional space (4D) using metric MDS
- Use model-based clustering (Fraley & Raftery, 1998) to
    - determine outliers among subjects (NNclean procedure from MCLUST)
    - cluster subjects
    - decide on cluster model and number by BIC
- Compute inter-subject agreement within cluster using Fleiss' kappa (1971)

$$\kappa = \frac{P(actual) - P(\mathrm{exp}ected)}{1 - P(\mathrm{exp}ected)}$$

Segmentation Clustering

Goldsmiths UNIVERSITY OF LONDON

- Results: κ values for inter-rater agreement

Melody no.	No. clusters	k cl. 1	k cl. 2	k cl. 3	k cl. 4	No. part. in noise cl.
1	3	.97	.93	.82		8
2	2	.91	.72			9
3	2	.90	1.00			15
4	3	.90	.70	.88		16
5	3	.83	.83	.90		8
6	1	.90				14
7	4	.68	.62	.80	.69	1
8	2	.91	.96			17
9	4	.83	.91	1.00	1.00	9
10	4	.86	.58	.85	.63	7
11	2	.67	.69			11
12	2	.97	.72			9
13	4	1.00	1.00	.54	.47	14
14	4	.97	.89	.96	.80	10
15	4	.60	.64	.66	.66	3

Segmentation Clustering: Results

Goldsmiths
UNIVERSITY OF LONDON

- Evaluation of segmentation algorithms according to optimal clustering
  - Compute 'true' boundaries for each cluster (on each melody) by model-based clustering (2 clusters only) from aggregated data
  - For each melody compute F1-performance for each algorithm with all clusters
  - Select only highest F1-value from each melody
  - Compute average over all test items

Mean performance over 15 melodies

	Mean F1	StDev. F1
never	0	0
always	.24	.05
saffran.p.pitch	.26	.17
GPR3a	.28	.16
SimpleSeg.	.29	.30
GPR2b	.42	.21
IDyOM	.55	.28
GPR2a	.56	.40
LBDM2001	.69	.18
Grouper	.76	.25

Performance over all notes of all 15 melodies

	Mean F1
never	0
always	.24
saffran.p.pitch	.24
GPR3a	.28
SimpleSeg.	.33
GPR2b	.42
IDyOM	.63
GPR2a	.75
LBDM2001	.71
Grouper	.79

Segmentation Clustering: Results

- Next Steps:
- Have IDyOM learn from interval and onset data
- Have IDyOM learn from different melody corpora (pop music)
- Associate subject clusters with segmentation strategies
- Combine models to build hybrid model

Part 6: Summary

Goldsmiths
UNIVERSITY OF LONDON

1. General methodology: statistical corpus-based musicology
   1.1. Music as psychological phenomenon
   1.2. Unsupervised-learning-based models over large corpora
2. Musical similarity
   2.1. Statistical techniques to model structural similarity
   2.2. Need segmentation before we start
3. Melodic pitch expectation
   3.1. Learn a model from a corpus
   3.2. Use it to predict sequences given context
4. Melodic segmentation
   4.1. Use the information-theoretic properties of 3 in a meta-model
   4.2. Predict segmentation (and other musicological properties?) from statistical signals
5. Specific methodology
   5.1. How to cope with inter-subject ambiguity in human perception

Acknowledgements

Goldsmiths
UNIVERSITY OF LONDON

- This work is funded by UK Engineering and Physical Sciences Research Council grants
  - GR/S82220: "Techniques and Algorithms for Understanding the Information Dynamics of Music"
  - EP/D038855: "Modelling Musical Memory and the Perception of Melodic Similarity"
- Thanks to Keith Potter (Goldsmiths Dept. of Music) for introducing us to "Two Pages" and helping us understand it