Goldsmiths
UNIVERSITY OF LONDON

Sloboda & Parker's recall paradigm for melodic memory: A new, computational perspective

Daniel Müllensiefen and Geraint Wiggins
Centre for Cognition, Computation and Culture
Goldsmiths, University of London)

Outline

Goldsmiths
UNIVERSITY OF LONDON

- Anatomy of a paradigm
- The Why and How of computational models
- Example: Melodic Similarity
- Revisiting the Sloboda & Parker (1985) results with a computer
- Beyond the original study: A new experiment

Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

- Classic experimental paradigms for verbal memory
  - Serial Recall (Ebbinghaus, 1885; Young, 1962)
  - Paired-Associate Learning (Calkins, 1894; Battig, 1972)
  - Free recall (Murdock, 1962; Rundus, 1974)
  - Recognition (Tulving, 1968)

Anatomy of a paradigm

- Employing different paradigms allows us to
  - Study different memory effects (e.g. list length and position vs. stimulus discrimination vs. semantic organisation)
  - Generate different memory models (e.g. memory trace models vs. storage models vs. conceptual hierarchies)
  - Provide complementary evidence for same effect or model (e.g. for serial position curve)

Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

- In musical (melody) memory research
  - ~~Free recall~~
  - ~~Paired-associate learning~~
  - (Serial recall)
  - Recognition

Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

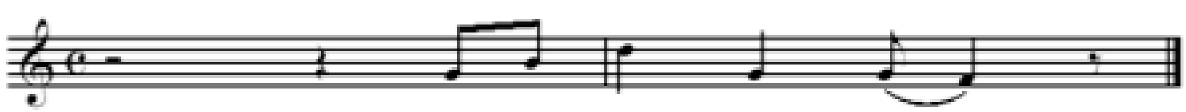
- Studies using melodic recall paradigm are much rarer and differ in:
  - Response mode
    - Notation (Deutsch, 1980)
    - Pseudo-notation (Davies & Yelland, 1977)
    - Playback on instrument (Halpern, 1989)
    - Verbal labels (Williamson, 2009)
    - Singback (Sloboda & Parker, 1985)
  - Subjects
    - Musically trained
    - Musically untrained
  - Melodic stimuli
    - Well-known
    - Unknown

Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

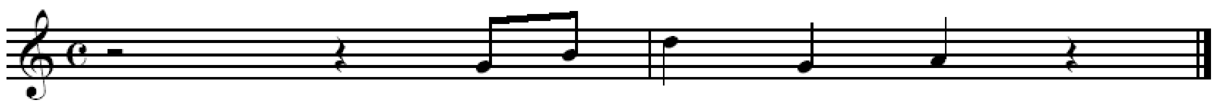
- How the Sloboda and Parker (1985) approach works:
  1. Subject listens to a melody:

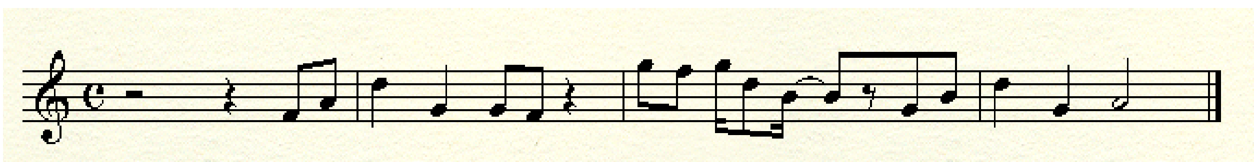
  2. Subject recalls (sings back) what s/he can remember:



Anatomy of a paradigm

3. Listening and recall is repeated over several trials for the same melody:





Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

4. Transcribe melodic data from recalls and score different types errors in comparison with original melody

5. Compare error rates and types over different trials, subject populations, melodies, and presentation modes (e.g. with vs. without lyrics)

Anatomy of a paradigm

Goldsmiths
UNIVERSITY OF LONDON

- Reasons for infrequent use of recall paradigm:
  - Generation of "dirty data" as experimental responses
  - Difficulty of comparing recalled melody fragment with original ("no theory of melodic identity", Sloboda & Parker, 1985, p. 159)
  - Limit for hand-made transcription and analysis of recalled melody fragments:
    - Sloboda & Parker (1985): 48 recalled fragments
    - Oura & Hatano (1988): 320
    - Zielinska & Miklaszewski (1992): 310
    - Ogawa et al. (1995): 80
    - Ginsborg & Sloboda (2007): 60
    - Müllensiefen & Wiggins (in prep.): 1900
- No standard scoring method for complex response data (i.e. music)

The Why and How of computational models

Goldsmiths
UNIVERSITY OF LONDON

- Idea
  - Use computational methods to process, analyse, and compare recall data
- Why?
  - Hand-made analysis is tedious and error-prone
  - Analytical methods need to be defined precisely and explicitly (in program code)
    - creates knowledge about music perception that can be studied/tested
    - lead to new, testable, operationalised cognitive models
  - Computers are fast and can process large amounts of data
    - allows easy and rigorous comparison between different conditions and different data sets
  - Powerful research method/tool

The Why and How of computational models

Goldsmiths
UNIVERSITY OF LONDON

- How?
  - Adapt existing methods as tools for music cognition from e.g.:
    - Computational music analysis/theory (e.g. Temperley, 2001; Huron, 1995, 2006)
    - Mathematical music theory (e.g. RUBATO; Chew et al., 2005)
    - Music Information Retrieval (e.g. Crawford et al., 1997; Müllensiefen & Frieler, 2004)
    - Computational linguistics (e.g. Downie, 1999; Pearce & Wiggins, 2006)
    - Music cognition (Thomassen, 1982; Krumhansl, 1990; Eerola et al., 2001)

The Why and How of computational models

Goldsmiths
UNIVERSITY OF LONDON

- Aspects of melodies that can be analysed computationally
  - Melodic identity and similarity (Müllensiefen & Frieler, 2004; Müllensiefen & Pendzich, 2009)
  - Metrical structure, metre induction (Eck, 2000; Volk, 2008)
  - Phrase structure (Bod, 2002; Temperley, 2001; Pearce et al., 2008)
  - Rhythmic structure (Weyde, 2004)
  - Harmonic structure: tonality induction (Longuet-Higgins & Steedman, 1970; Krumhansl, 1990)
- Also
  - Accent strength, complexity, expectedness, high-level structure identification
- And for polyphonic music
  - Voice separation, main voice identification, motive matching, chord labelling, genre classification, *etc*.

Example: Melodic Similarity

- Two basic steps in measuring melodic similarity
  1. Transform raw melodies into abstract representation of interest (intervals, contour, tonality etc.)
  2. Compute similarity between between abstract representations of the two melodies as numerical value between 0 and 1

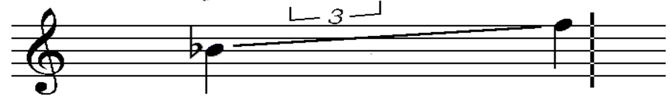
Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

1. Transformations of melodic data

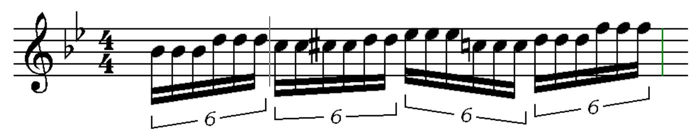
Raw data / melody

Contour (Steinbeck)



Countour (Müllensiefen & Frieler)

Rhythmically weighted



Intervals:	+4 –2 +1 +1 +1 –3 +2 +3
Interval categories:	+J –S +S +S +S –J +S +J
Intervals directions:	U D U U U D U U
Ranks:	6. 3. 5. 4. 3. 2. 5. 3. 1.
Implicit tonality:	Bb Major
Duration classes:	N N S S S N N N N

Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

2. Take sequences of number or symbols and compute similarity value between 0 and 1

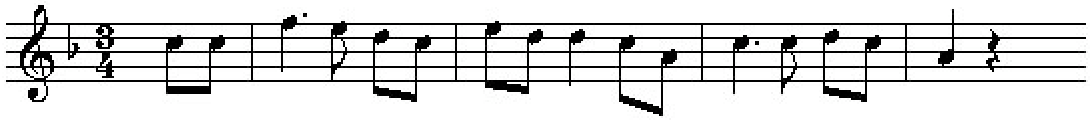
- Range of possible similarity algorithms:
  - Geometric (distance, correlation) algorithms (Ó Maidín, 1998; Aloupis et al., 2003)
  - Transportation distances (Typke et al., 2007)
  - String matching algorithms (Mongeau & Sankoff, 1990; Crawford et al., 1998)
  - n-gram algorithms measures (Downie, 1999; Uitdenbogered, 2002)
  - Probabilistic (hidden Markov) models (Meek & Birmingham, 2002)
  - Feature-matching models (Müllensiefen & Pendzich, 2009)
  - Hybrid algorithms (Müllensiefen & Frieler, 2004)

Example: Melodic Similarity

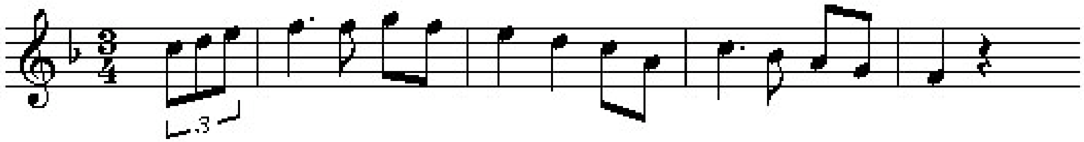
Goldsmiths
UNIVERSITY OF LONDON

2. *(contd.)* Take sequences of number or symbols and compute similarity value between 0 and 1

- Example:
- 2nd phrase of Luxembourgeois folksong „*Ist denn Liebe ein Verbrechen*“



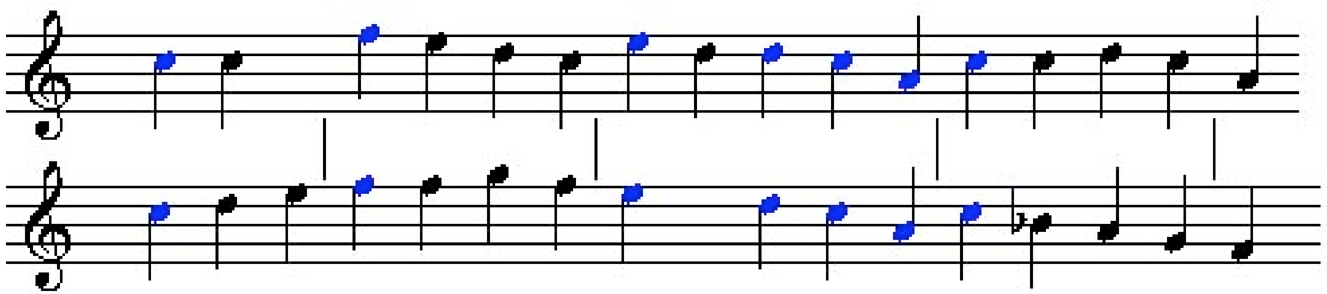
- 2nd phrase of Luxembourgeois folksong „*Ehestandslehren*“



Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

- String matching algorithm: Optimal alignment of phrases using Edit Distance



  - We need 10 operations to transform *„Ehestandlehren“* into *„Ist denn Liebe ein Verbrechen“*: 8 substitutions and 2 insertions (deletions)
  - Similarity(Edit distance) = 1- 10/16 = 6/16 = 0.375

Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

- A statistically informed alternative: Tversky's feature-based similarity model (1977)

$$\sigma(s,t) = \frac{f(s_n \cap t_n)}{f(s_n \cap t_n) + \alpha f(s_n \setminus t_n) + \beta f(t_n \setminus s_n)}, \alpha, \beta \geq 0$$

- … adapted for measuring melodic similarity by
  - taking interval chains or terms τ as features of melodies and
  - defining salience function f as inverse of frequency of term τ in melody corpus C

$$IDF_C(\tau) = \log\left(\frac{|C|}{|m : \tau \in m|}\right)$$

- … gives:

$$\sigma(s,t) = \frac{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau)}{\sum_{\tau \in s_n \cap t_n} IDF_C(\tau) + \alpha \sum_{\tau \in s_n \setminus t_n} IDF_C(\tau) + \beta \sum_{\tau \in t_n \setminus s_n} IDF_C(\tau)}$$

Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

- Example: Terms τ and corresponding IDF weights for „Ist denn Liebe“

Melodic term τ (pitch interval 2-gram)	Frequency of melodic term τ in M^4S corpus C	$IDF_C(\tau)$
0, +5	4750	1.09
+5, -1	2529	1.72
-1, -2	10474	0.29
-2, -2	11676	0.19
-2, +4	6195	0.82
+4, -2	5620	0.92
-2, 0	11291	0.22
0, -2	12009	0.16
-2, -3	8343	0.52
-3, +3	7413	0.64
+3, 0	8048	0.56
0, +2	11167	0.23
+2, -2	11857	0.17

Example: Melodic Similarity

- **Problem**: Combinatorial explosion for constructing similarity measures from transformations x comparison algorithms

RAWED	Raw pitch edit distance		
RAWEDW	Raw pitch edit distance weighted		
RAWPC	Raw pitch Pears. Brav. correlation		
RAWPCST	Raw pitch P-B. corr, weighted, 0-1	FOURR	Fourier (ranks)
RAWPCW	Raw pitch Pears. Brav. Corr. weighted	FOURRST	Fourier (ranks), weighted, 0-1
RAWPCWST	Raw pitch P-B. Corr. weighted, 0-1	FOURRW	Fourier (ranks), weighted
RAWCC	Raw pitch cross-correlation	FOURRWST	Fourier (ranks), weighted, 0-1
RAWCCW	Raw pitch cross-correlation weighted	FOURRI	Fourier (ranks, intervals)
CONSED	Contour (Steinbeck) edit distance	DIFFED	Intervals (Edit distance)
CONSPC	Contour (Steinbeck), P-B. correlation	DIFF	Intervals (Mean difference)
CONSPCST	Contour (Steinbeck), P-B. corr., 0-1	DIFFEXP	Intervals (Mean diff., exp.)
CONSCC	Contour (Steinbeck), Cross-corr.	DIFFFUZ	Intervals (fuzzy), Edit Distance
CONED	Contour, Edit distance	DIFFFUZC	Intervals (fuzzy contour)
CONPC	Contour, Pearson-Bravais correlation		
CONPCST	Contour, Pearson-Bravais corr., 0-1		
CONCC	Contour, Crosscorrelation		

Example: Melodic Similarity

Goldsmiths
UNIVERSITY OF LONDON

- **Solution**: Evaluate similarity algorithms for specific modelling purpose:
  - Experts' similarity judgements (Eerola & Bregman, 2007; Ziv & Eitan, 2007)
  - Experts' similarity rankings (Typke et al., 2007)
  - Identification of folk song variants (Müllensiefen & Frieler, 2007)
  - Classification of plagiarism cases in pop music (Müllensiefen & Pendzich, 2009)

⇒ Potential Results:

- Valid and reliable similarity algorithm as tools for specific application
- Cognitive model of similarity judgements
- Predictions and hypotheses to be tested

Revisiting Sloboda & Parker

Goldsmiths
UNIVERSITY OF LONDON

- Main finding 1: *No recall is perfect*
  - Max. similarity values: 0.93 (Pitch, Edit Distance)
    0.98 (Pitch, Tversky similarity)

- Main finding 2: *Recalls are highly related to original in many respects*
  - Mean similarity values: 0.37 (Pitch, Edit Distance)
    0.21 (Pitch Contour, Edit Distance)
    0.2 (Implied Tonalities, Edit Distance)

Revisiting Sloboda & Parker

Goldsmiths
UNIVERSITY OF LONDON

- Main finding 3: *Metrical structure is preserved in almost all recalls*
  - recalls with ind. 4/4: 18.75% (Beatometer: Frieler, 2004)
  - recalls with ind. 2/4: 81.25% (Beatometer)
  - ind. metre for original: 2/4 (Beatometer)

- Main finding 6: *Subjects vary in accuracy re. melody and harmony*
  - Diff. mean sim (nov.-exp.): 0.15 (rawEd)

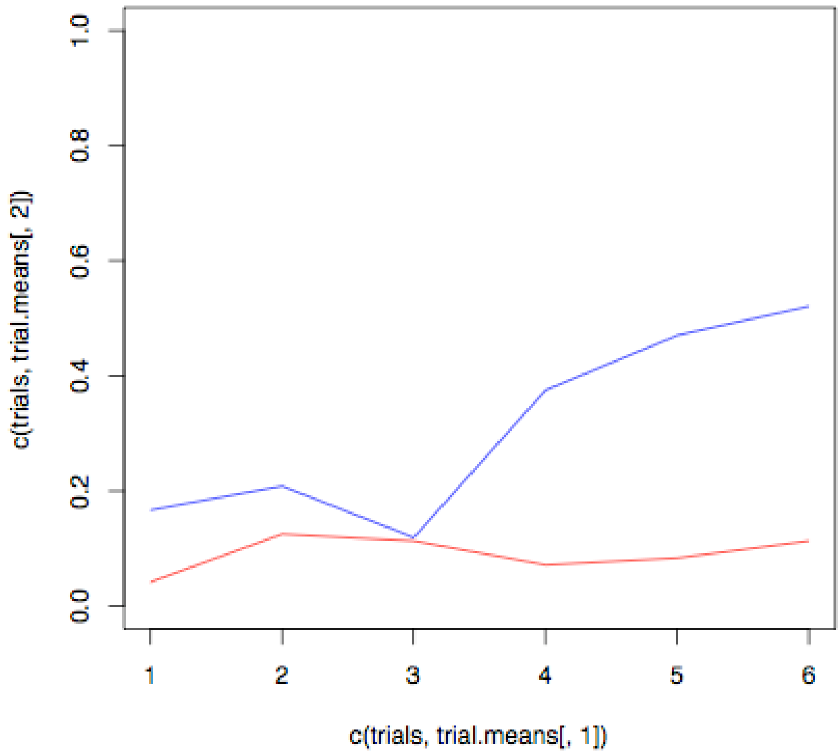
    t-test: $t(30.4)=2.9, p<0.004$
  - Diff. mean sim (nov.-exp.): 0.22 (harmCorE)

    t-test: $t(28.5)=3.2, p<0.002$

Revisiting Sloboda & Parker

Goldsmiths UNIVERSITY OF LONDON

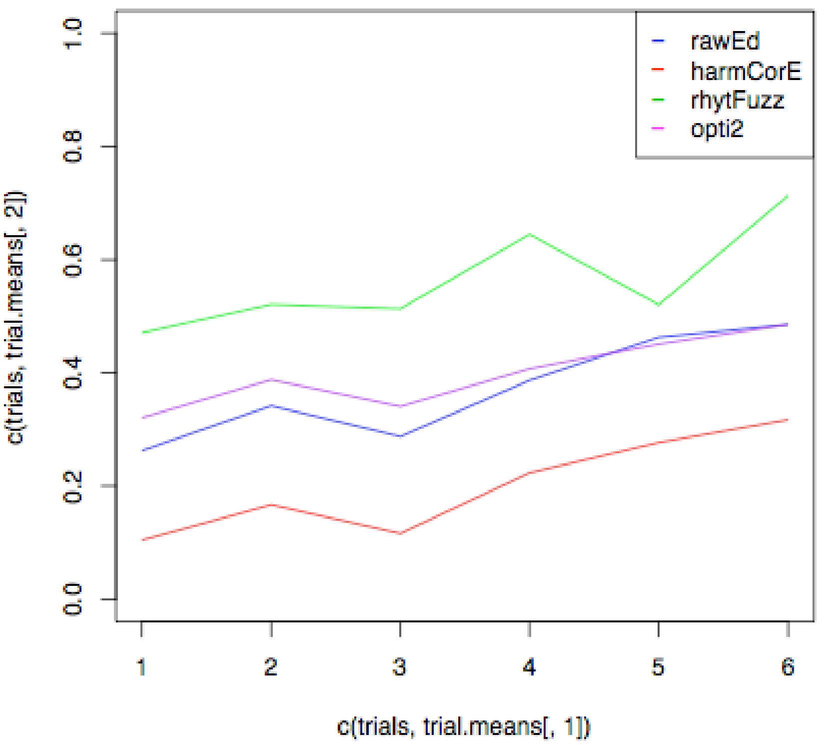
- Main finding 8: *Musicians and non-musicians differ in retention of harmony*
    - dependent sample t-test: $t(30.4)=3.1$, $p<0.02$ (harmcorE)



Revisiting Soloboda's & Parker's results

- Main finding 9: *Subjects do not show improvement on any measure*

  - cor(trial, rawEd): r= .93, p< .008
  - cor(trial, harmCorE): r= .93, p< .008
  - cor(trial, rhytFuzz): r= .77, p< .08
  - cor(trial, opti2): r= .91, p< .02



Beyond Sloboda & Parker

Goldsmiths UNIVERSITY OF LONDON

- Questions addressed by Sloboda & Parker (1985):
  - How do memory representations of melodies build up over repeated exposure?
  - What melodic parameters are easier to grasp which ones are more difficult?
  - How do musicians and non-musicians differ?
  - What is the effect of adding lyrics? (Ginsborg & Sloboda, 2007)
- Some interesting additional questions:
  - How do musical features affect recall?
  - What makes a melody easy / difficult to recall?
  - Which parts of a melody are represented first and most accurately?
  - Does commonness or rarity of melodic features play a role?
  - Are melodies in context (e.g., audio excerpt) recalled better than isolated melodies?

Beyond Sloboda & Parker

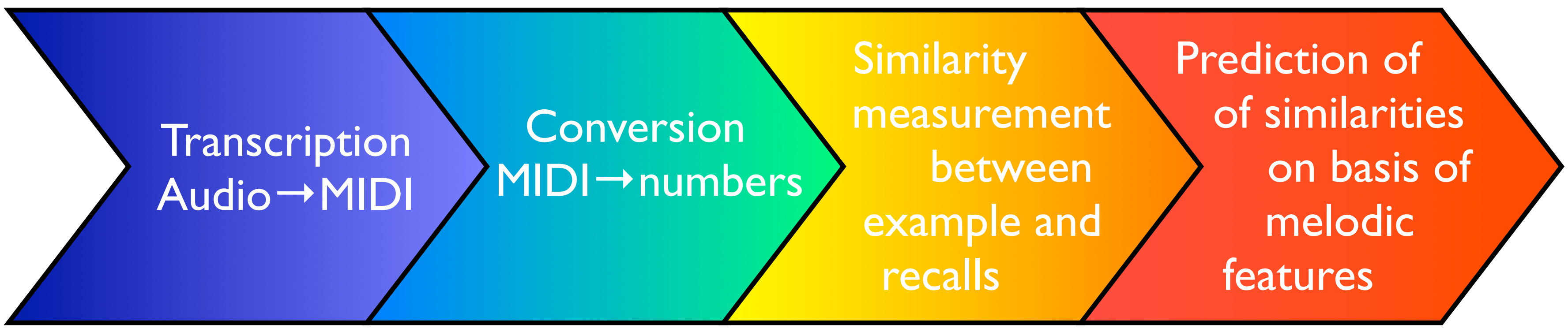
Goldsmiths UNIVERSITY OF LONDON

- A new experiment (Müllensiefen & Wiggins, in preparation):
  - Subjects: 30 adults, half with high musical background
  - Material: 14 short pop melodies as monophonic midi and real song excerpt
  - Task: Immediate recall (singing back) of melody after repeated listening
  - Data: 14 melodies x 6 trials x 23 (usable) subjects ≈ 1900 recalls (.wav files)

Beyond Sloboda & Parker

Goldsmiths
UNIVERSITY OF LONDON

- Summary of analytical steps



Beyond Sloboda & Parker

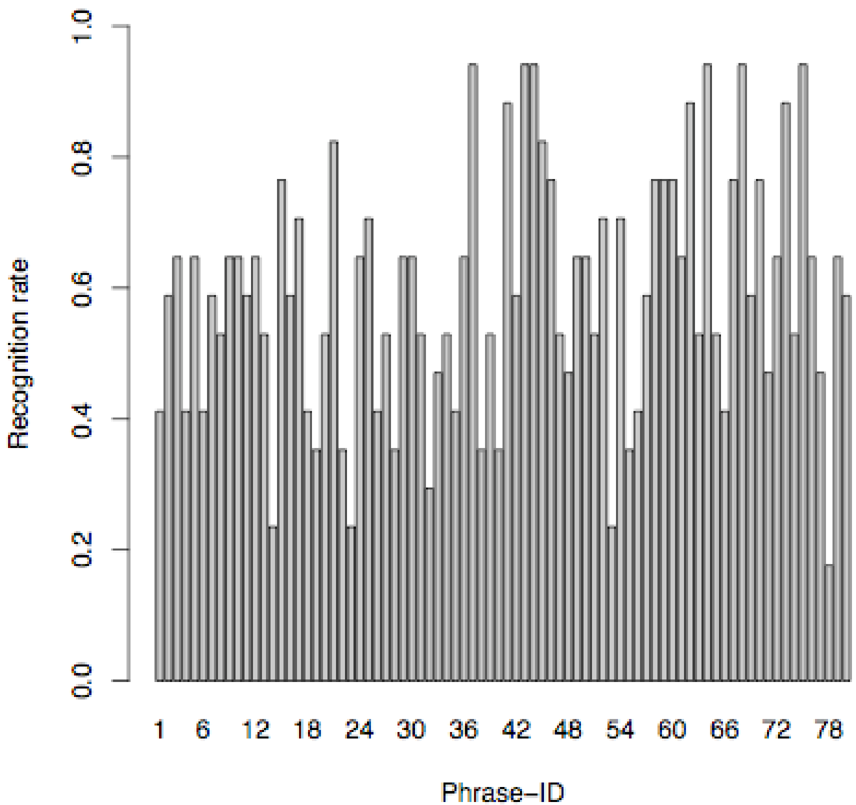
Goldsmiths UNIVERSITY OF LONDON

- A feature- and corpus-based approach to melodic analysis:
  - Assumption: Certain computable melodic characteristics (features) influence recall from memory
  - Hypothesis: Frequency / prevalence of feature in corpus of familiar music influences recall
  - Goal: Predict similarity between recall and original from melodic features
- Question: Are the same melodic features relevant for recall and recognition memory? (Halpern, Müllensiefen & Wiggins, 2008)

Beyond Sloboda & Parker

Goldsmiths UNIVERSITY OF LONDON

- Why does memory recognition performance vary so strongly between melodies? (Halpern et al., 2008)
- We have models which explain a fair proportion of the variance



Beyond Sloboda & Parker

Goldsmiths
UNIVERSITY OF LONDON

- Answer by modelling subjects' recognition and recall accuracy using sets of features and feature frequencies derived from a large corpus
- The corpus: 14,063 fully transcribed pop songs from 1950s-2006
- The feature set(s):
  1. Melodic content summary features (descriptive statistics, contour and tonality descriptors)
  2. Characteristics of text constants (distributions of "melodic terms" from computational linguistics)
- Analysis
  - 1C. Corpus-derived frequencies of summary features
  - 2C. Corpus-derived values for characteristic text constants
  - 3C. Corpus-based weights of melodic terms (cf. Latent Semantic Analysis)

Summary & Conclusion

Goldsmiths
UNIVERSITY OF LONDON

- Sloboda and Parker were modest about the utility of their paradigm
  - Gives useful results, but data very difficult to analyse
- 24 years later, computational technology is reaching the stage when it can
  - Reliably analyse participants' responses
  - Reliably compare responses with stimuli at appropriate level of abstraction
- Doing so allows the production of new computational cognitive models
  - Can be automatically run over very large data sets
  - Can therefore easily/quickly/objectively generate strong (surprising) hypotheses
  - Can therefore add strongly to empirical music cognition studies
- This paradigm has great promise for the future (Thanks, John!!)