


Tutorial

Music Structure Analysis

Meinard Müller
International Audio Laboratories Erlangen
Universität Erlangen-Nürnberg
meinard.mueller@audiolabs-erlangen.de

Jordan B. L. Smith
Electronic Engineering and Computer Science
Queen Mary University of London
j.smith@qmul.ac.uk

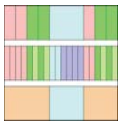
AUDIO LABS FAU Fraunhofer IIS Queen Mary University of London

Overview

Part I: Principles & Techniques
(Meinard Müller)

Coffee Break

Part II: Evaluation & Annotation
(Jordan Smith)



Overview

Music structure analysis

General goal: Divide an audio recording into temporal segments corresponding to musical parts and group these segments into musically meaningful categories.

Overview

Music structure analysis

General goal: Divide an audio recording into temporal segments corresponding to musical parts and group these segments into musically meaningful categories.

Evaluation

General goal: Determine how well an algorithm achieves the goal above

Overview

Music structure analysis

General goal: Divide an audio recording into temporal segments corresponding to musical parts and group these segments into musically meaningful categories.

Evaluation

General goal: Determine how well an algorithm achieves the goal above

Problem: What metric is appropriate?

Overview

Music structure analysis

General goal: Divide an audio recording into temporal segments corresponding to musical parts and group these segments into musically meaningful categories.

Evaluation

General goal: Determine how well an algorithm achieves the goal above

Problem: What metric is appropriate?

…More problems:

- What is the performance floor? Ceiling?
- What differences in performance are significant?
- Do the annotations mean what we think?

Overview

- Introduction
- Part 1: Evaluation techniques
  - Metrics
  - Evaluation Design
  - Meta-evaluation
- Part 2: Annotations and listeners
  - Annotation procedures
  - Disagreements

Metrics

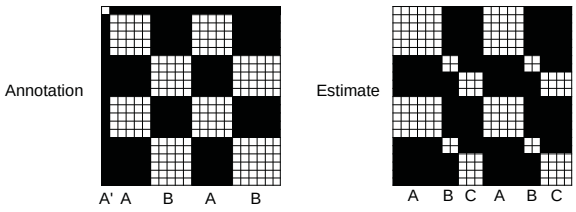
- Labelling metrics vs. boundary metrics (vs. summary metrics)
- Over-segmentation vs. under-segmentation
- Compiled in Lukashevich 2008

Metrics

- Pairwise retrieval
- Main idea: Consider M_a, the set of all *pairs of frames* **annotated** with the same label. This is a set of similarity relationships to **estimate**
  - precision: $\boldsymbol{pw_p} = | M_a \cap M_e | / | M_e |$
  - recall: $\boldsymbol{pw_r} = | M_a \cap M_e | / | M_a |$
  - f-measure: $\boldsymbol{pw_f} = 2\, pw_p\, pw_r / (pw_p + pw_r)$

Metrics

- Pairwise retrieval
- Main idea: Consider M_a, the set of all *pairs of frames* **annotated** with the same label. This is a set of similarity relationships to **estimate**

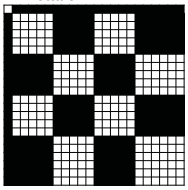


"Blue Fiddle" by Claude Lamothe

Metrics

Missed similarity = red
Spurious similarity = blue
Correct pairs = white

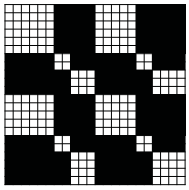
Annotation:



- 484 pixels total
- 262 black
- 222 white pixels

M_a = 222

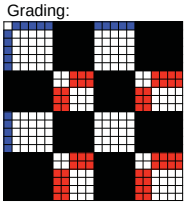
Estimate:



- 186 white pixels

M_e = 186

Grading:

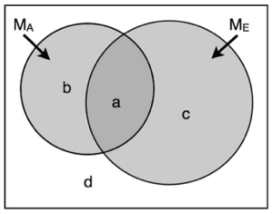


- 166 correct pairs

- Recall = 166/222
  r = 0.75
- Precision = 166/186
  p = 0.89
- f-measure
  f = 0.81

Metrics

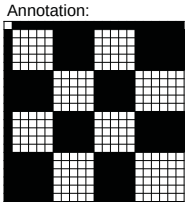
- Rand index
- Main idea: like pairwise retrieval, but consider pairwise **dissimilarities** as also necessary to estimate
  - recall = a / (a+b)
  - precision = a / (a+c)
  - **Rand** = (a+d) / (a+b+c+d)



Metrics

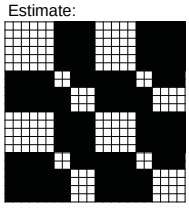
Spurious dissimilarity = red
Spurious similarity = blue
Correct pairs = white or black

Annotation:



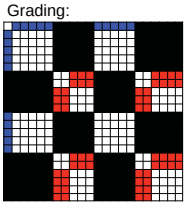
- 484 pairs
- 222 similar pairs (a+b)
- 262 dissimilar pairs (c+d)

Estimate:



- 484 off-diagonal pixels
- 186 similar pairs (a+c)
- 298 dissimilar pairs (b+d)

Grading:



- a = 166 true positive
- **b = 56** false negative
- **c = 20** false positive
- **d = 242** true negative

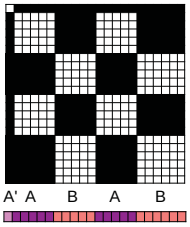
- **Rand** = (166+242) / (166+56+20+242) **= 0.84**

Metrics

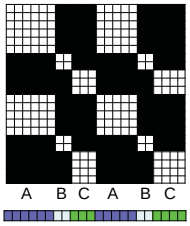
- Average speaker purity (ASP) and average cluster purity (ACP)
- Main idea: estimate the level of fragmentation of each label category
  - Consider each annotated label L_i separately
  - Given L_i, consider the parallel estimated frames and compute the sum of squares for each label
  - Normalise and tally these sums to get **ASP**

Metrics

Annotation



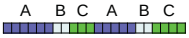
Estimate



Metrics

Annotation

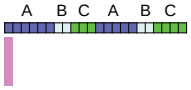
Estimate



Metrics

Annotation

Estimate

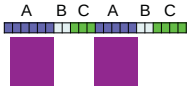


	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1

Metrics

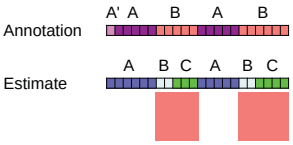
Annotation

Estimate



	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1
A	1	100	10	10

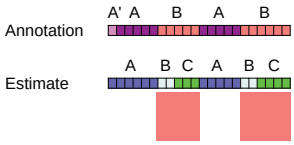
Metrics



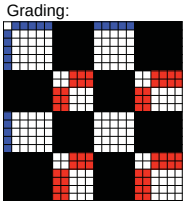
	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1
A	1	100	10	10
B	2	$4^2 + 7^2 = 65$	11	5.91

ASP = normalized sum = (1+10+5.91)/22 = 0.77

Metrics



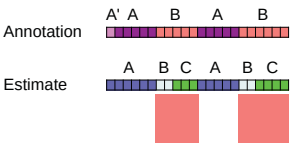
Grading:



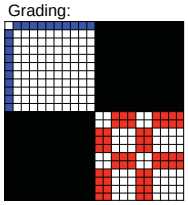
	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1
A	1	100	10	10
B	2	$4^2 + 7^2 = 65$	11	5.91

ASP = normalized sum = (1+10+5.91)/22 = 0.77

Metrics



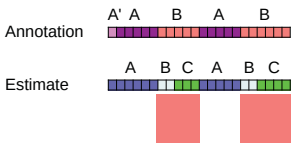
Grading:



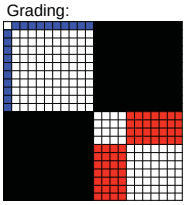
	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1
A	1	100	10	10
B	2	$4^2 + 7^2 = 65$	11	5.91

ASP = normalized sum = (1+10+5.91)/22 = 0.77

Metrics



Grading:



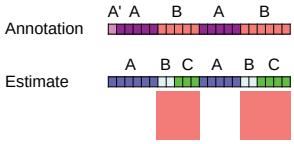
	# of labels	Sum of squares	length	SSQ/length
A'	1	1	1	1
A	1	100	10	10
B	2	$4^2 + 7^2 = 65$	11	5.91

ASP = normalized sum = (1+10+5.91)/22 = 0.77

Metrics

- Average speaker purity (ASP) and average cluster purity (ACP)
- Main idea: estimate the level of fragmentation of each label category
  - Consider each annotated label L_i separately
  - Given L_i, consider the parallel estimated frames and compute the sum of squares for each label
  - Normalise and tally these sums to get **ASP**
  - Do the reverse to get **ACP**
  - Summary metric **K = (ASP*ACP)$^{1/2}$**

Metrics



	# of labels	Sum of squares	length	SSQ/length
A	2	$1^2 + 10^2 = 101$	11	9.18
B	1	16	4	4
C	1	49	7	7

ASP = normalized sum = (1+10+5.91)/22 = **0.77**
ACP = normalized sum = (9.18 + 4 + 7)/22 = **0.92**
K = (0.77 * 0.92)$^{1/2}$ = **0.84**

- **R = 0.75**
- **P = 0.89**
- **f = 0.81**

Metrics

- Over- and under-segmentation scores

- Main idea:
  - Over-segmentation: $S_O = H(E|A)$, normalized
    - given the annotation, how much more is there to know about the estimated analysis?
  - Under-segmentation: $S_U = H(A|E)$, normalized
    - given the estimated analysis, how much more is there to know about the annotation?

Metrics

Annotation: A' A B A B

Estimate: A B C A B C

find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

Metrics

Annotation: A' A B A B

Estimate: A B C A B C

find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

a_i	A'	A	B
$P(A|a_i)$	1	1	0
$P(B|a_i)$	0	0	4/11
$P(C|a_i)$	0	0	7/11

Metrics

Annotation: A' A B A B

Estimate: A B C A B C

find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

a_i	A'	A	B
$\log P(A|a_i)$	0	0	-inf
$\log P(B|a_i)$	-inf	-inf	-1.459
$\log P(C|a_i)$	-inf	-inf	-0.6521

Metrics

Annotation: A' A B A B

Estimate: A B C A B C

find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

a_i	A'	A	B
$P(A|a_i) * \log P(A|a_i)$	0	0	0
$P(B|a_i) * \log P(B|a_i)$	0	0	-0.5307
$P(C|a_i) * \log P(C|a_i)$	0	0	-0.4150

Metrics

Annotation: A' A B A B

Estimate: A B C A B C

find all $p(a_i)$:

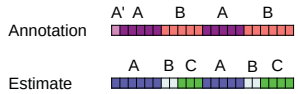
a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

a_i	A'	A	B
sum $(P(e_j|a_i) * \log P(e_j|a_i)$	0	0	-0.9457

Metrics



find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

find all $p(e_j|a_i)$:

a_i	A'	A	B
sum $(P(e_j\|a_i)$ * log $P(e_j\|a_i)$	0	0	-0.9457

$H(E|A) = 0*1/22 + 0*10/22 - 0.95*11/22$
$= 0.473$

Metrics

find all $p(a_i)$:

a_i	A'	A	B
$p(a_i)$	1/22	10/22	11/22

$$H(E|A) = -\sum_i p(a_i) \sum_j p(e_j|a_i) \cdot \log_2 p(e_j|a_i)$$

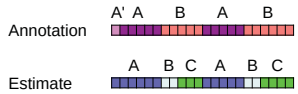
find all $p(e_j|a_i)$:

a_i	A'	A	B
sum $(P(e_j\|a_i)$ * log $P(e_j\|a_i)$	0	0	-0.9457

$H(E|A) = 0*1/22 + 0*10/22 - 0.95*11/22$
$= 0.473$

$$S_o = 1 - \frac{H(E|A)}{\log_2 N_e} = 1 - 0.473/1.585 = 0.70$$

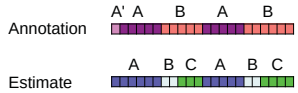
Metrics



Metrics

$$S_o = 1 - \frac{H(E|A)}{\log_2 N_e} = 1 - 0.473/1.585 = 0.70$$

Metrics



$$S_o = 1 - \frac{H(E|A)}{\log_2 N_e} = 1 - 0.473/1.585 = 0.70$$

$$S_s = 1 - \frac{H(A|E)}{\log_2 N_a} = 1 - 0.02/1.585 = 0.99$$

Metrics

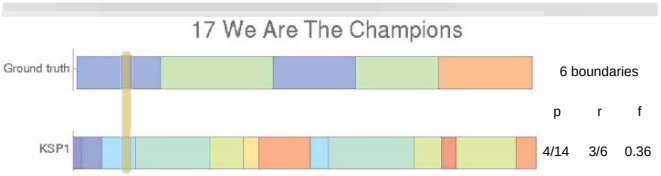
$$S_o = 1 - \frac{H(E|A)}{\log_2 N_e} = 1 - 0.473/1.585 = 0.70$$

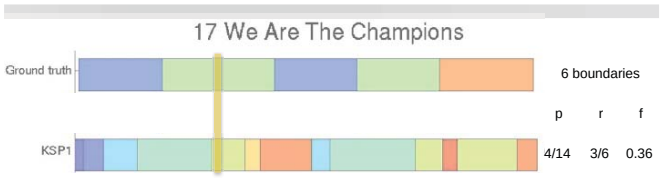
$$S_s = 1 - \frac{H(A|E)}{\log_2 N_a} = 1 - 0.02/1.585 = 0.99$$

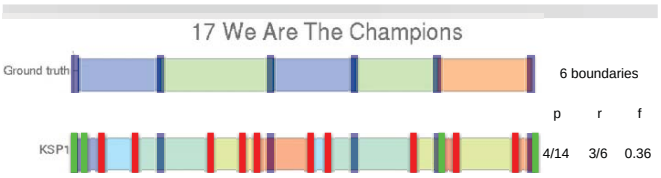
$I(A,E) = H(E) - H(E|A) = 1.473 - 0.473 = 1.00$

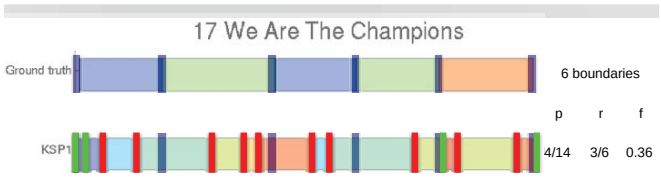
Metrics

- Boundary retrieval
- Main idea: treat all boundaries within a fixed threshold of the true boundaries as correct

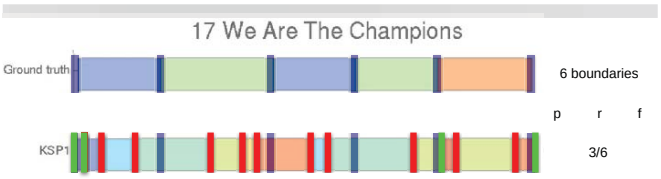


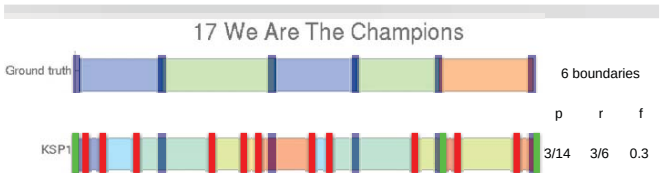


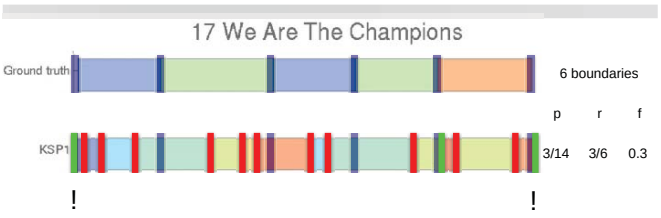


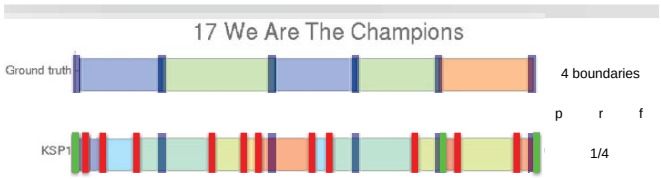


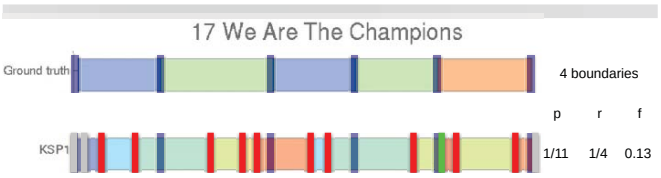
!

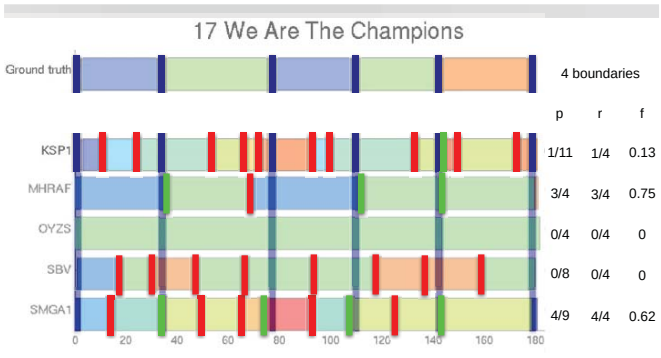






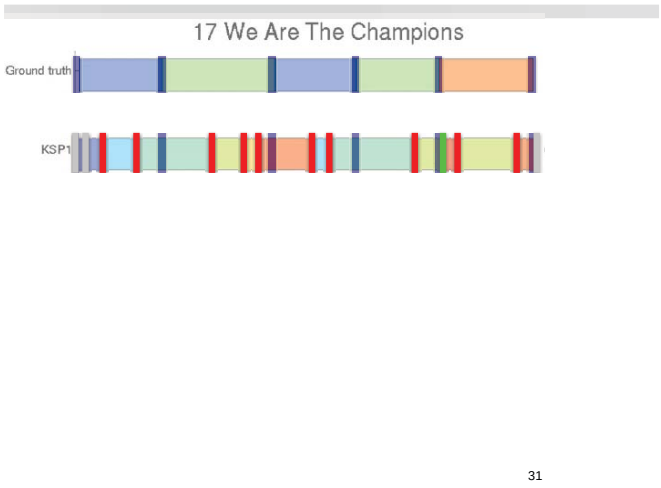




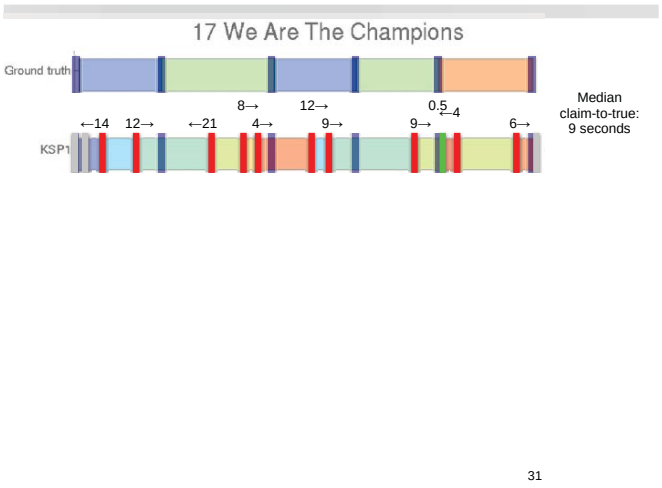


Metrics

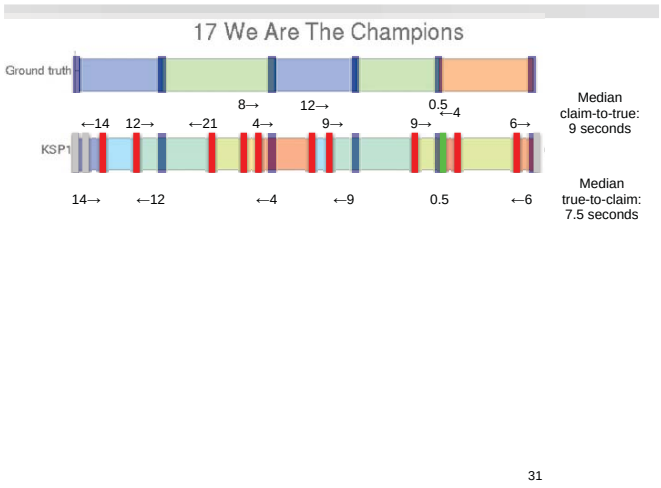
- Median claim to true
- Main idea: estimate the median proximity of the estimated boundaries to the true ones



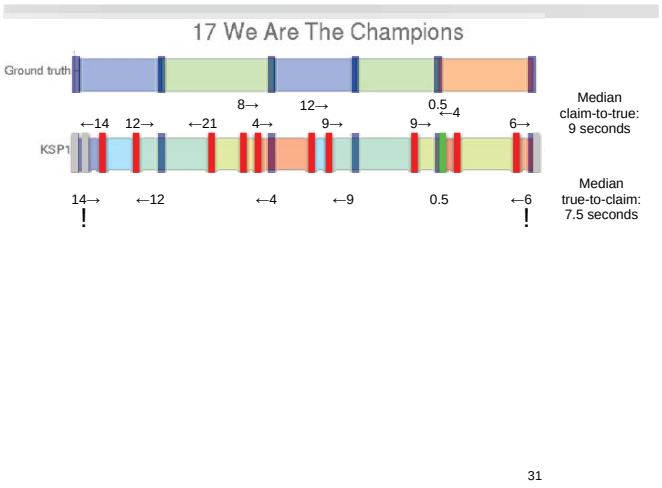
31



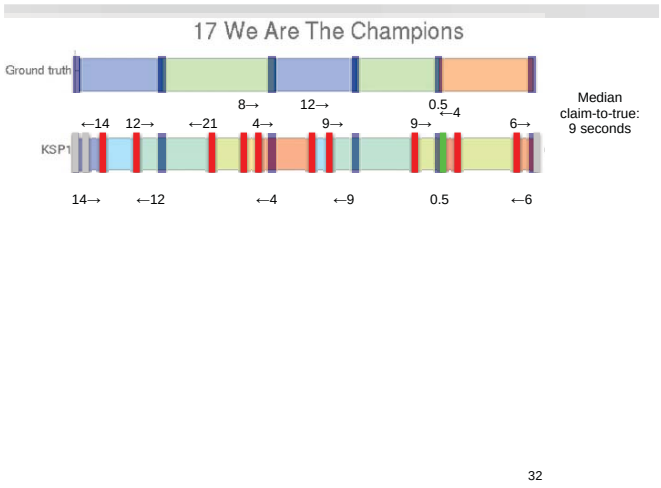
31



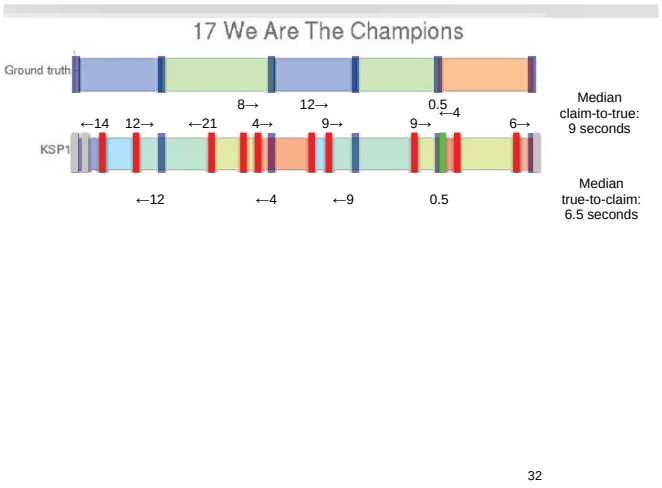
31



31



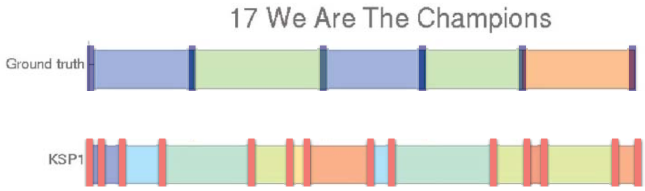
32



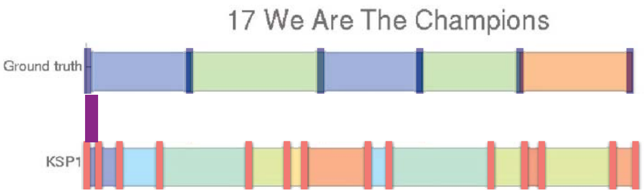
32

Metrics

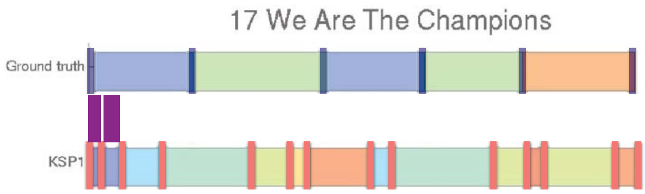
- Directional Hamming distance
- Main idea: estimate the level of fragmentation of each section



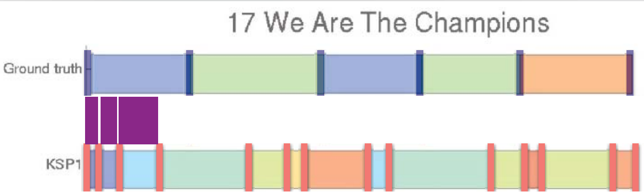
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation



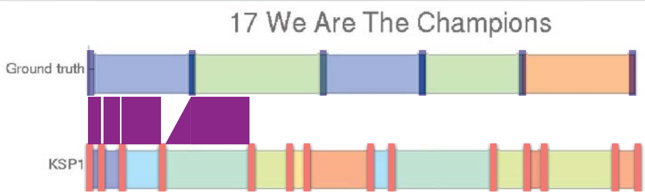
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation



Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

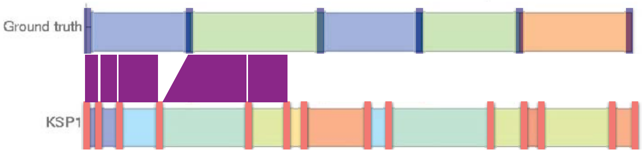


Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation



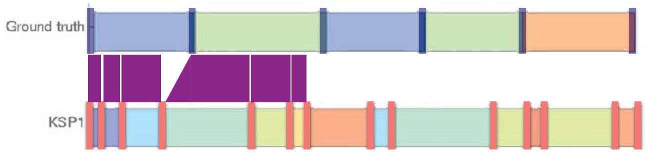
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

17 We Are The Champions



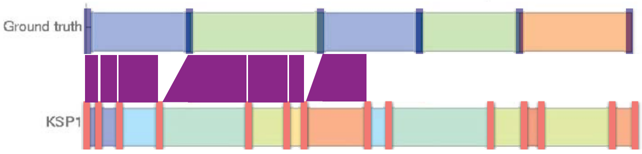
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

17 We Are The Champions



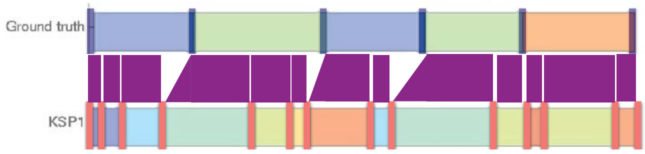
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

17 We Are The Champions



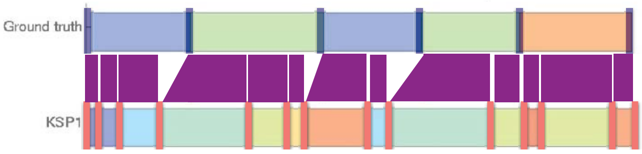
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

17 We Are The Champions



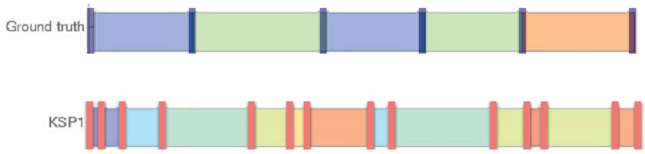
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation

17 We Are The Champions



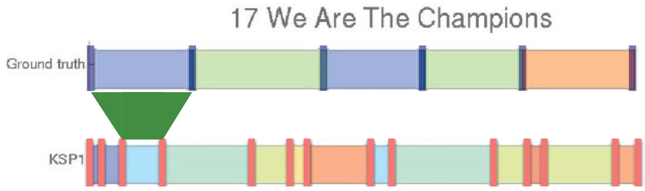
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

17 We Are The Champions



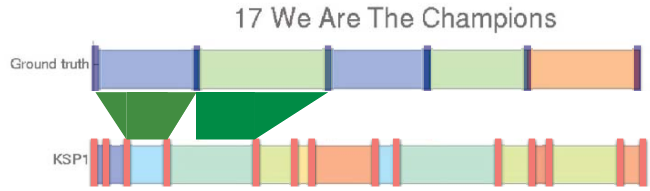
Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

Fraction of annotation covered by single estimated segments
= 1 – *m* = over-segmentation



Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

Fraction of annotation covered by single estimated segments
= 1 – *m* = over-segmentation

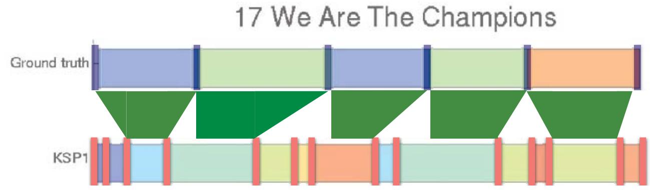


Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

Fraction of annotation covered by single estimated segments
= 1 – *m* = over-segmentation

Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

Fraction of annotation covered by single estimated segments
= 1 – *m* = over-segmentation



Fraction of estimated analysis covered by single annotated segments
= 1 – *f* = under-segmentation = 0.86

Fraction of annotation covered by single estimated segments
= 1 – *m* = over-segmentation = 0.58

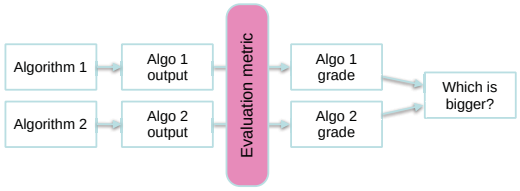
Metrics

- Can someone do this all for me?
  - Raffel et al.: (PS2-20) MIR_EVAL: A Transparent Implementation of Common MIR Metrics
  - Structural Analysis Evaluation code.soundsoftware.ac.uk/projects/structural_analysis_evaluation

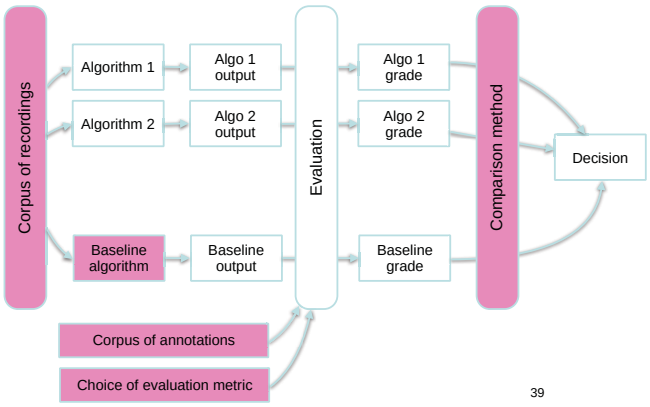
Overview

- Introduction
- Part 1: Evaluation techniques
  - Metrics
  - Evaluation Design
  - Meta-evaluation
- Part 2: Annotations and listeners
  - Annotation procedures
  - Disagreements

Evaluation design



Evaluation design



Evaluation design

- Choice of corpus
  - restricts view to subset of all music
  - choose to match needs of evaluation

Evaluation design

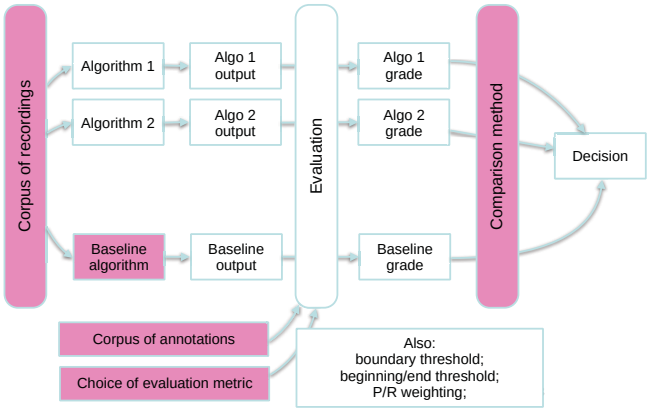
- Choice of baseline
  - Segments:
    - fixed number of random boundaries
    - boundaries at fixed interval
  - Labels:
    - all the same labels
    - all different labels
    - random labels from fixed-size vocabulary

Evaluation design

- Choice of annotations
  - merging of segments
  - simplification of labels

SALAMI		INRIA	
0.0	Silence	0.0000000	-
0.411065759	A	0.3697289	G/H--
49.840770975	A'	9.0107311	H-
65.980725623	B	17.2752319	H
80.060748299	C	25.3954022	H+
96.227619047	B	33.5430528	H+
110.325170068	C	41.6209960	H~
126.354331065	B	49.7735622	I
140.525782312	C	57.7395708	I
156.690249433	B	65.6869237	J/2
164.745873015	silence	69.8471577	(5/8)A-
166.826825396	end	80.0145458	(5/4)C
		100.1254529	(5/8)A
		109.9383816	(5/4)C'
		130.2545020	(5/8)A
		140.4218901	(3/2)C*
		164.6930311	end

Evaluation design



Evaluation design

- Choice of comparison method
  - compare mean values
  - normal statistics
    - student's t-test
    - ANOVA
  - non-normal statistics
    - Wilcoxon Signed-Rank test
    - Kruskal-Wallis test

Evaluation design

- Choice of comparison method
  - compare mean values
  - ~~normal statistics~~
    - ~~student's t-test~~
    - ~~ANOVA~~
  - non-normal statistics
    - Wilcoxon Signed-Rank test
    - Kruskal-Wallis test

Evaluation design

- Decision:
  - "Our algorithm performs better than the leading MIREX competitor!"

    vs.
  - "According to a Mann–Whitney U Test ($U=43029$, $N=298$, $p < 0.05$), our algorithm performs better than the leading MIREX competitor, when performance is evaluated with pairwise *f*-measure, on a version of the Beatles dataset with labels reduced to their main categories (*intro, verse, chorus, other, outro*). We achieved a median *f*-measure of 0.68 (IQR: 0.48, 0.75). The best-performing random baseline achieved a median *f*-measure of 0.35, and a comparison of different annotators indicates a performance ceiling with median *f*-measure 0.92.

Overview

- Introduction
- Part 1: Evaluation techniques
  - Metrics
  - Evaluation Design
  - Meta-evaluation
- Part 2: Annotations and listeners
  - Annotation procedures
  - Disagreements

Meta-evaluation

- Julián Urbano: "Information retrieval meta-evaluation: Challenges and opportunities in the music domain." ISMIR 2011
- 7 kinds of validation:
  - **construct:** does metric match goal?
  - **content:** is corpus representative?
  - **convergent:** do different results agree?
  - **criterion:** agreement with other experiments?
  - **internal:** any factors unaccounted for?
  - **external:** does sampling justify extrapolation?
  - **conclusion:** are conclusions justified?

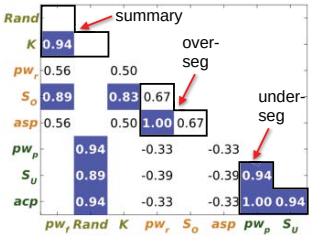
Meta-evaluation

- Julián Urbano: "Information retrieval meta-evaluation: Challenges and opportunities in the music domain." ISMIR 2011
- 7 kinds of validation:
  - **construct:** does metric match goal?
  - Nieto, Farbood, Jehan and Bello: "Perceptual analysis of the *f*-measure for evaluating section boundaries in music." ISMIR 2014, PS2-3
  - **internal:**
  - **external:**
  - **conclusion:**

Meta-evaluation

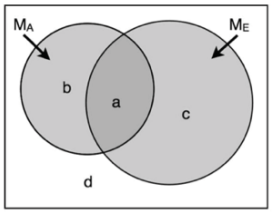
- Julián Urbano: “Information retrieval meta-evaluation: Challenges and opportunities in the music domain.” ISMIR 2011
- 7 kinds of validation:
  - **construct:**
  - **content:**
  - **convergent:** do different results agree?
  - Smith and Chew 2013a: “A meta-analysis of the MIREX structure segmentation task.” ISMIR
  - **criterion:**
  - **external:**
  - **conclusion:**

Meta-evaluation

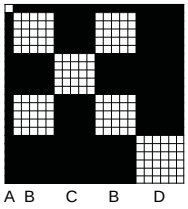


Correlation in labelling metrics in ranking algorithms

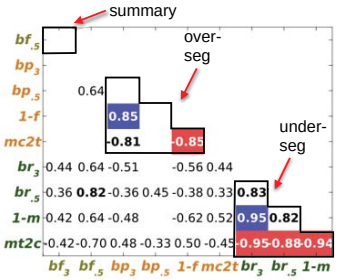
Meta-evaluation



Rand = (*a*+*d*) / (*a*+*b*+*c*+*d*)

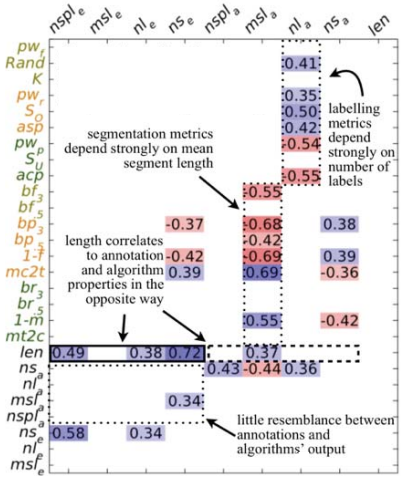


Meta-evaluation



Correlation in segmentation metrics in ranking algorithms

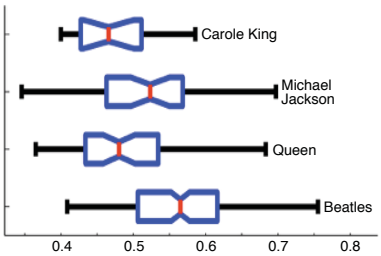
Correlation of evaluation metrics and properties of annotations and algorithm output



Meta-evaluation

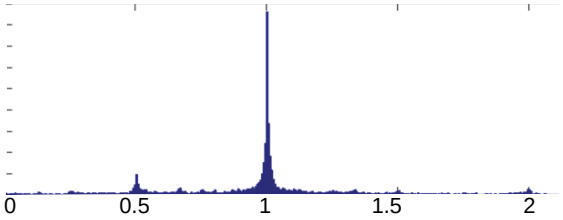
- Julián Urbano: “Information retrieval meta-evaluation: Challenges and opportunities in the music domain.” ISMIR 2011
- 7 kinds of validation:
  - **construct:**
  - **content:**
  - **convergent:**
  - **criterion:**
  - **internal:**
  - **external:** does sampling justify extrapolation?
  - **conclusion:**

Meta-evaluation



Distribution of mean pairwise *f*-measure

Meta-evaluation



Histogram of ratio between a song's median segment length and the length of all of its segments

See also Bimbot et al. 2014: "Semiotic Description of Music Structure: an Introduction to the Quaero/Metiss Structural Annotations." AES

Part 1: Summary

- Metrics
  - over- and under-segmentation metrics
  - boundary and grouping metrics
- Evaluation Design
  - corpus
  - baseline
  - annotation interpretation
  - decision method
- Meta-evaluation
  - human tests to align metrics with perceived quality
  - observe real-world performance of metrics

Summary

Music structure analysis

General goal: Divide an audio recording into temporal segments corresponding to musical parts and group these segments into musically meaningful categories.

- How much agreement is there about what the musical parts are?
- What is the significance of the disagreements?
- Who creates the ground truth?
- What procedure do they follow?

Overview

- Introduction
- Part 1: Evaluation techniques
  - Metrics
  - Evaluation Design
  - Meta-evaluation
- Part 2: Annotations and listeners
  - Annotation procedures
  - Disagreements

Annotation procedures

- Early Beatles annotations based on Alan Pollack's analyses

Notes on "Come Together"

Notes on ... Series #177 (CT)

by Alan W. Pollack

```
      Key: d minor / D Major
    Meter: 4/4
     Form: Intro/Verse | Intro/Verse | Refrain |
                       | Intro/Verse | Refrain |
                       | 1/2 Intro/Verse (Instrumental) |
                       | 1/2 Intro/Verse | Refrain |
                       | Intro/Outro (fade-out)
       CD: "Abbey Road", Track 1 (Parlophone CDP7 46446-2)
 Recorded: 21th, 22nd, 23rd July 1969, Abbey Road 3;
           25th, 29th, 30th July 1969, Abbey Road 2
UK-release: 26th September 1969 (LP "Abbey Road")
US-release: 1st October 1969 (LP "Abbey Road")
```

http://www.icce.rug.nl/~soundscapes/DATABASES/AWP/ct.shtml

Notes on "Come Together"

Notes on ... Series #177 (CT)

by Alan W. Pollack

```
      Key: d minor / D Major
    Meter: 4/4
     Form: Intro/Verse | Refrain |
           | Intro/Verse | Refrain |
           | 1/2 Intro/Verse (Instrumental) |
           | 1/2 Intro/Verse | Refrain |
           | Intro/Outro (fade-out)
       CD: "Abbey Road", Track 1 (Parlophone CDP7 46446-2)
 Recorded: 21th, 22nd, 23rd July 1969, Abbey Road 3;
           25th, 29th, 30th July 1969, Abbey Road 2
UK-release: 26th September 1969 (LP "Abbey Road")
US-release: 1st October 1969 (LP "Abbey Road")
```

General Points of Interest

Style and Form

"Come Together" opens the 'Abbey Road" album with a stylistic gesture that remains, over the long run of their career as well as from our historical view of it 30 years later, one of the Beatles key strengths and accomplishments. Call it what you will: "stylized", "neo-classical"; maybe even "rubber soul" (read: music style, not

http://www.icce.rug.nl/~soundscapes/DATABASES/AWP/ct.shtml

Annotation procedures

- Conflation of similarity, function, and instrumentation noted by Peeters and Deruty (2009)

Annotation procedures

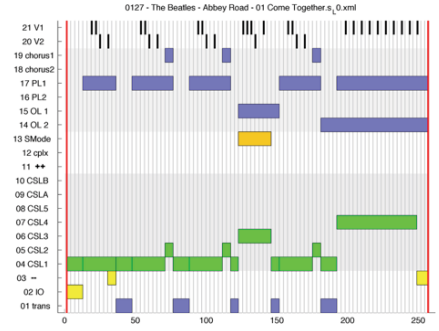
- Conflation of similarity, function, and instrumentation noted by Peeters and Deruty (2009)

Beatles annotation:

0.000	1.000	silence
1.000	35.861	intro/verse
35.861	70.617	intro/verse
70.617	76.487	refrain
76.487	111.236	intro/verse
111.236	116.995	refrain
116.995	145.717	1/2_intro/verse_(instrumental)
145.717	174.955	1/2_intro/verse
174.955	180.829	refrain
180.829	254.248	intro/outro_(fade-out)
254.248	260.627	silence

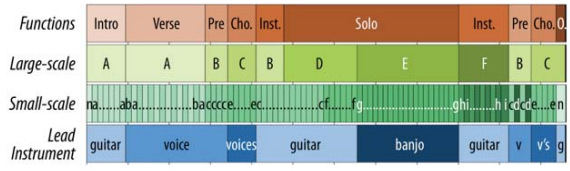
Annotation procedures

- Conflation of similarity, function, and instrumentation noted by Peeters and Deruty (2009)



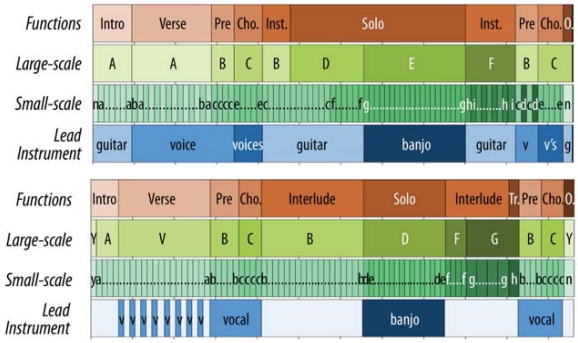
Annotation procedures

- SALAMI dataset used simplified version of their proposal



Annotation procedures

- SALAMI dataset used simplified version of their proposal

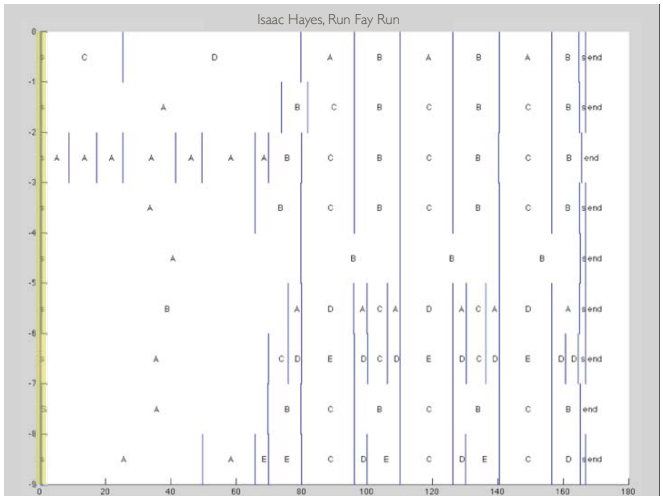


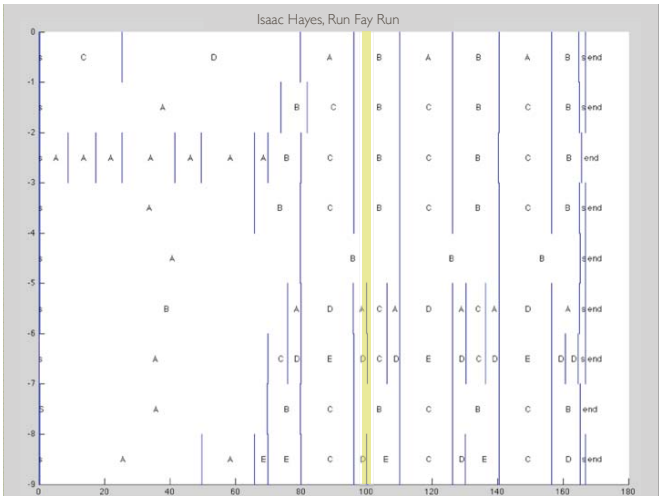
Annotation procedures

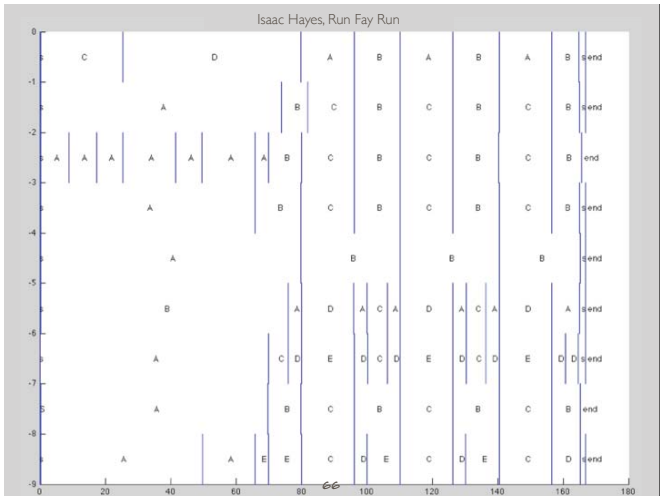
- Bimbot et al. 2010 & 2012:
  - Segmentation:
    - Set standard segment length for each song
      - Ideal segment length: 15 seconds
    - Criteria for being a segment:
      - Interchangeability
      - Similarity
      - etc.
  - Labelling:
    - System & Contrast model
      - standard segment form: a-b-c-d
      - taxonomy of transformations and exceptions

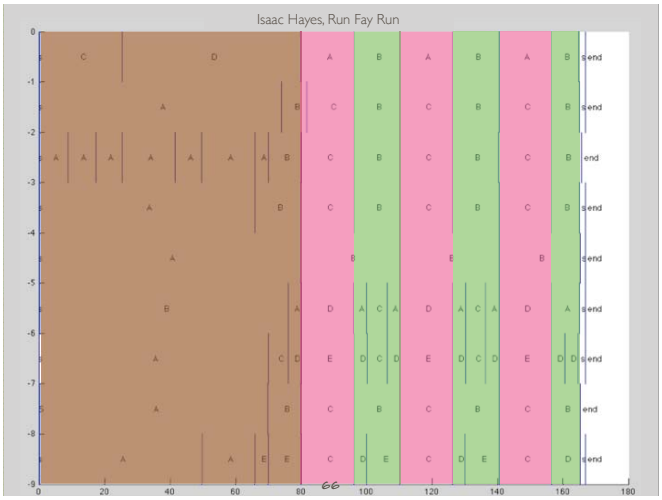
Overview

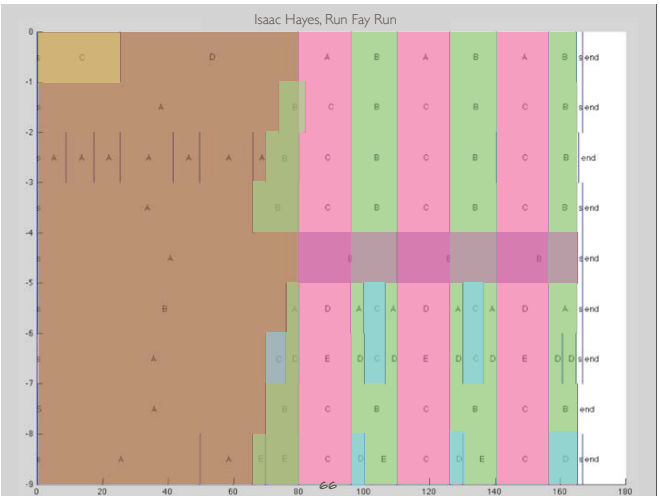
- Introduction
- Part 1: Evaluation techniques
  - Metrics
  - Evaluation Design
  - Meta-evaluation
- Part 2: Annotations and listeners
  - Annotation procedures
  - Disagreements



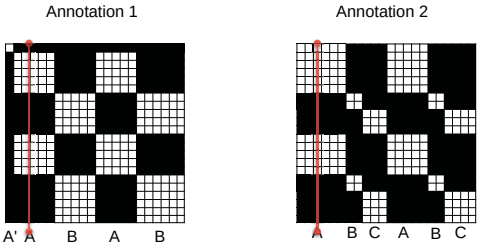






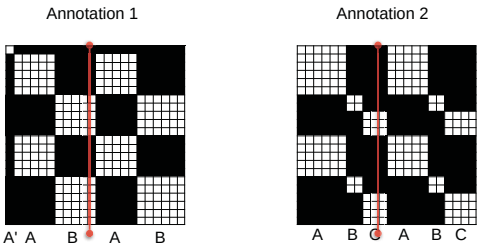


Disagreements



"Blue Fiddle" by Claude Lamothe

Disagreements



"Blue Fiddle" by Claude Lamothe

Disagreements

- How to minimise disagreements?
  - SALAMI: collect extra annotations to reflect variety of interpretations
  - INRIA: constrain annotation format to improve repeatability

Disagreements

- Perception of structure depends on:
  - Familiarity with the piece (e.g., Margulis 2012)
  - Level of musical training (e.g., Bamberger 2006)
  - Attention?

Disagreements

- Smith 2014
- Question: Do differences in attention cause listener disagreements, or merely accompany them?
- Goal: Observe the impact that attention to specific features has on the grouping preferences of listeners
- Method: Experiment presenting listeners with ambiguous stimuli and controlling the attention condition

Disagreements

Part 3 of 4: Salience of change

Every excerpt in this part has a single pattern repeated 4 times, with a change in some feature between the 2nd and 3rd instances; i.e., it has form AABB. We ask you to focus on a particular aspect of the music while listening, and tell us: how significant was the change at the half-way point?

This section should take less than 6 minutes.

Trial 4 of 12

Please pay attention to the **chords** of the following excerpt.

Question 1. How strong is the change at the midpoint of the excerpt?

- 5. Extremely strong
- 4.
- 3.
- 2.
- 1. Not strong at all

Next >>

Disagreements

Part 3 of 4: Salience of change

Every excerpt in this part has a single pattern repeated 4 times, with a change in some feature between the 2nd and 3rd instances; i.e., it has form AABB. We ask you to focus on a particular aspect of the music while listening, and tell us: how significant was the change at the half-way point?

This section should take less than 6 minutes.

Trial 4 of 12

Please pay attention to the chords of the following excerpt.

Question 1. How strong is the change at the midpoint of the excerpt?

5. Extremely strong
4.
3.
2.
1. Not strong at all

Next >>

71

Disagreements

Part 3 of 4: Salience of change

Every excerpt in this part has a single pattern repeated 4 times, with a change in some feature between the 2nd and 3rd instances; i.e., it has form AABB. We ask you to focus on a particular aspect of the music while listening, and tell us: how significant was the change at the half-way point?

This section should take less than 6 minutes.

Trial 4 of 12

Please pay attention to the chords of the following excerpt.

Question 1. How strong is the change at the midpoint of the excerpt?

5. Extremely strong
4.
3.
2.
1. Not strong at all

Next >>

71

Disagreements

Part 3 of 4: Salience of change

Every excerpt in this part has a single pattern repeated 4 times, with a change in some feature between the 2nd and 3rd instances; i.e., it has form AABB. We ask you to focus on a particular aspect of the music while listening, and tell us: how significant was the change at the half-way point?

This section should take less than 6 minutes.

Trial 4 of 12

Please pay attention to the chords of the following excerpt.

Question 1. How strong is the change at the midpoint of the excerpt?

5. Extremely strong
4.
3.
2.
1. Not strong at all

Next >>

71

Disagreements

Part 3 of 4: Salience of change

Every excerpt in this part has a single pattern repeated 4 times, with a change in some feature between the 2nd and 3rd instances; i.e., it has form AABB. We ask you to focus on a particular aspect of the music while listening, and tell us: how significant was the change at the half-way point?

This section should take less than 6 minutes.

Trial 4 of 12

Please pay attention to the chords of the following excerpt.

Question 1. How strong is the change at the midpoint of the excerpt?

5. Extremely strong
4.
3.
2.
1. Not strong at all

Next >>

71

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Please listen to the following **chord progression**

Please listen to the following excerpt of music and indicate whether that **chord progression** appears in it.

Question 1. Did the **chord progression** appear in the excerpt?

- Yes
- Yes, but only a variation
- No
- I do not know

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Please listen to the following **chord progression**

0:00

Please listen to the following excerpt of music and indicate whether that **chord progression** appears in it.

0:00

Question 1. Did the **chord progression** appear in the excerpt?

- Yes
- Yes, but only a variation
- No
- I do not know

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Please listen to the following **chord progression**

0:00

Please listen to the following excerpt of music and indicate whether that **chord progression** appears in it.

0:00

Question 1. Did the **chord progression** appear in the excerpt?

- Yes
- Yes, but only a variation
- No
- I do not know

72

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Please listen to the following **chord progression**

0:00

Please listen to the following excerpt of music and indicate whether that **chord progression** appears in it.

0:00

Question 1. Did the **chord progression** appear in the excerpt?

- Yes
- Yes, but only a variation
- No
- I do not know

72

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Please listen to the following **chord progression**

0:00

Please listen to the following excerpt of music and indicate whether that **chord progression** appears in it.

0:00

Question 1. Did the **chord progression** appear in the excerpt?

- Yes
- Yes, but only a variation
- No
- I do not know

72

Disagreements

Part 2 of 4: Does the pattern occur?

In this set of questions, a musical pattern of some kind will be shown to you. Your goal is to judge whether this pattern occurs in the longer musical excerpt that follows. We then ask you to re-listen to the excerpt, and state whether you prefer form AAB or ABB.

This section should take less than 12 minutes.

Trial 5 of 12

Now, please listen to the excerpt again. (The following clip is identical to the previous clip.)

0:00

Question 2. Which of the following analyses do you think best fits the excerpt?

- A A B
- A B B

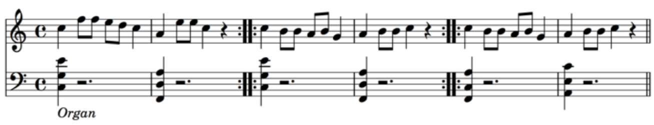
Question 3. How certain are you about your choice of analysis?

- Totally certain
- Very certain
- Both certain and uncertain
- Very uncertain
- Not at all certain

Next >>

Disagreements

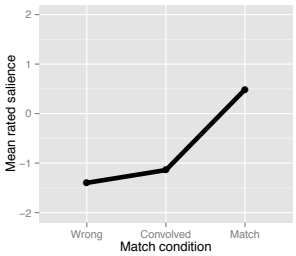
Melody: A A B

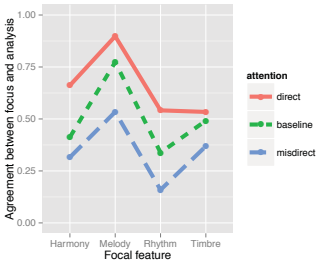


Chords: A B B

Disagreements

- Results:



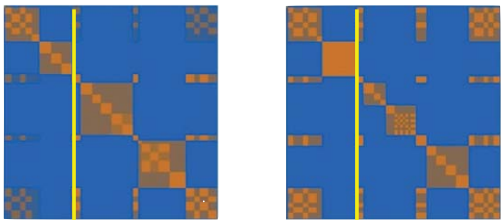


Disagreements

- Smith and Chew 2013b
- The perception of structure is influenced by attention
- Can we infer what a listener was paying attention to?
- Can this help to explain listener disagreements?

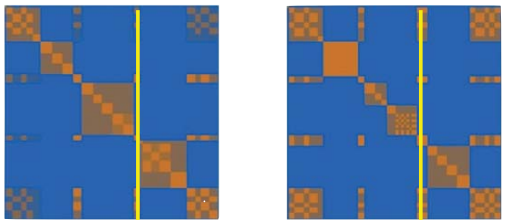
Disagreements

- Two different annotations of Chago Rodrigo's "Garrotin"



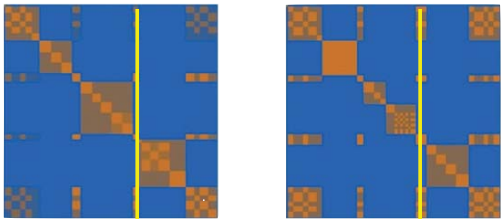
Disagreements

- Two different annotations of Chago Rodrigo's "Garrotin"

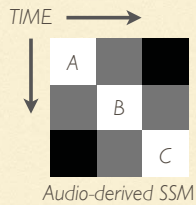


Disagreements

- Two different annotations of Chago Rodrigo's "Garrotin"

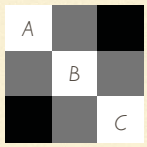


A TOY EXAMPLE



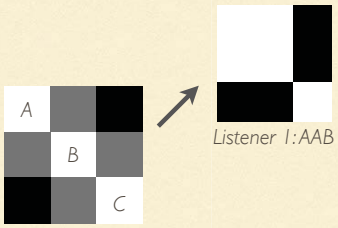
Audio-derived SSM

A TOY EXAMPLE



Audio-derived SSM

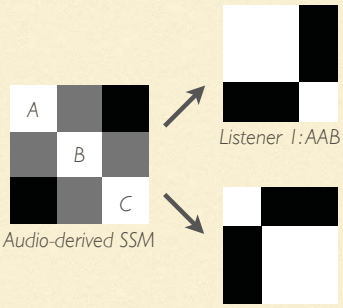
A TOY EXAMPLE



Audio-derived SSM

Listener 1: AAB

A TOY EXAMPLE

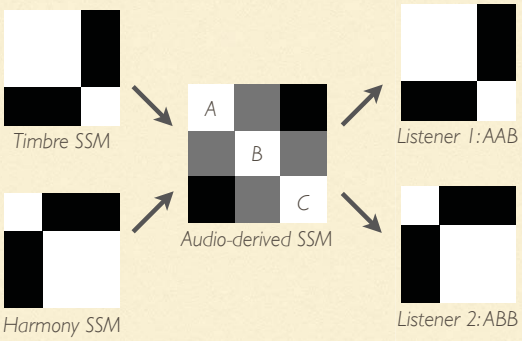


Audio-derived SSM

Listener 1: AAB

Listener 2: ABB

A TOY EXAMPLE



Timbre SSM

Audio-derived SSM

Harmony SSM

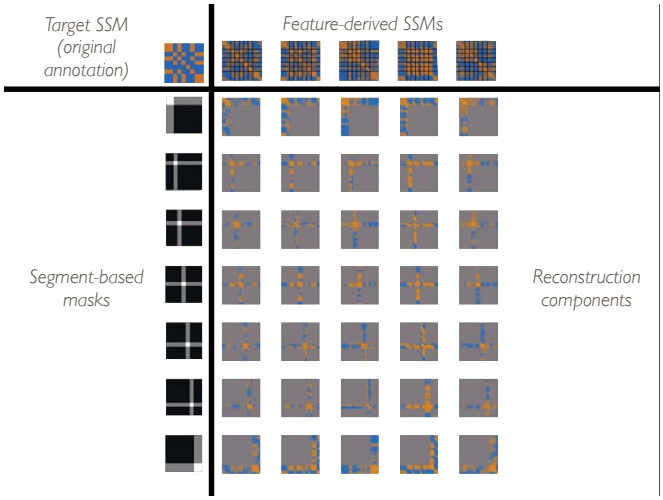
Listener 1: AAB

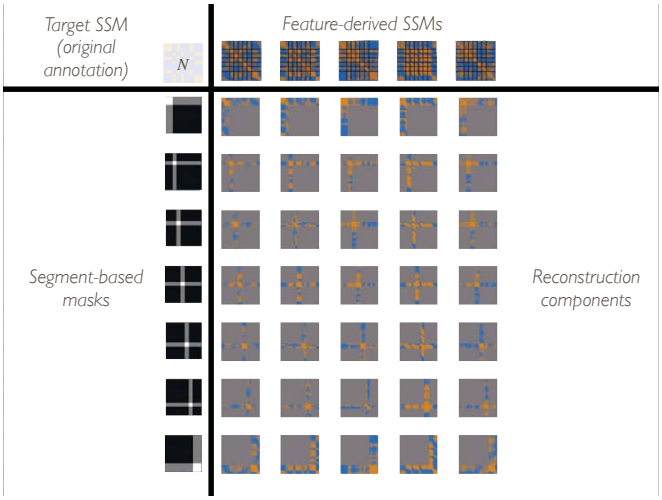
Listener 2: ABB

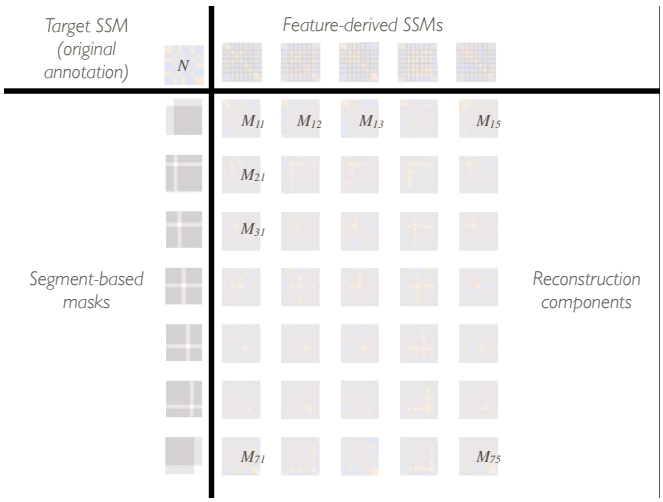
A TOY EXAMPLE

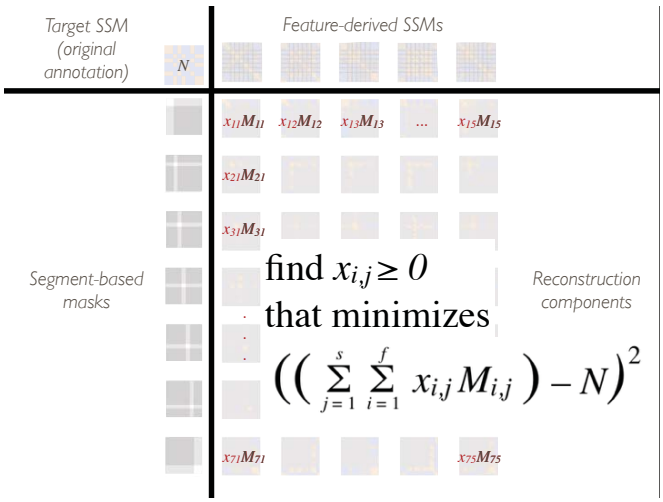
Suggestion: We might explain the differences between two annotations by matching them with the musical features.

A TOY EXAMPLE

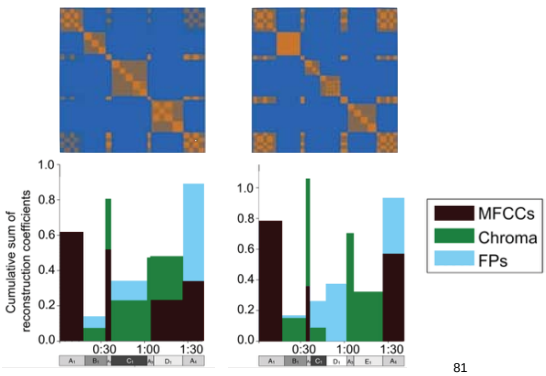








Disagreements



Part 2: Summary

- Annotations and listeners
  - Defining ground truth is a fraught task because listeners often disagree
  - Solutions:
    - poll many listeners
    - define the task more narrowly
  - Music perception research demonstrates personal factors affect analysis
    - Content-based approach may never be perfect
  - Attention may be an important factor, and we can try to estimate it

Final thoughts

- Part 1: Be aware of how you evaluate!
  - Use proper statistics
  - Need for more meta-analysis of metrics
- Part 2: Be aware of what you're using!
  - Know the limitations of annotations
  - Need for more music cognition studies

Acknowledgement (Jordan Smith)

- Elaine Chew (Queen Mary University of London)
- Marcus T. Pearce (Queen Mary University of London)
- Isaac Schankler (University of Southern California)
- Ching-Hua Chuan (University of North Florida)
- SALAMI annotators (McGill University)
- Frédéric Bimbot (IRISA)
- Corentin Guichaoua (IRISA)

This work has been supported by: the Social Sciences and Humanities Research Council; a PhD studentship from Queen Mary University of London; a Provost's Ph.D. Fellowship from the University of Southern California. Some material is also based in part on work supported by the National Science Foundation under Grant No. 0347988.

Works cited

- Jeanne Bamberger. 2006. What develops in musical development? A view of development as learning. In Gary McPherson, editor, The Child as Musician: Musical Development from Conception to Adolescence, pages 69–92. Oxford: Oxford University Press.
- Frédéric Bimbot, Gabriel Sargent, Emmanuel Deruty, Corentin Guichaoua and Vincent, Emmanuel. 2014. Semiotic description of music structure: An introduction to the Quaero/Metiss structural annotations. Proceedings of AES Conference on Semantic Audio, London, UK.
- Frédéric Bimbot, Emmanuel Deruty, Gabriel Sargent, and Emmanuel Vincent. 2012. Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions. Proceedings of ISMIR, pages 235–240, Porto, Portugal.
- Frédéric Bimbot, Olivier Le Blouch, Gabriel Sargent, and Emmanuel Vincent. 2010. Decomposition into autonomous and comparable blocks: A structural description of music pieces. In Proceedings of ISMIR, pages 189–194, Utrecht, Netherlands.
- Hanna Lukashevich. 2008. Towards quantitative measures of evaluating song segmentation. Proceedings of ISMIR, pages 375–380, Philadelphia, PA, USA.
- Elizabeth Margulis. 2012. Musical repetition detection across multiple exposures. Music Perception, 29 (4): 377–385.
- Oriol Nieto, Morwaread M. Farbood, Tristan Jehan, and Juan Pablo Bello. 2014. Perceptual analysis of the f-measure for evaluating section boundaries in music. Proceedings of ISMIR, pages 265–270, Taipei, Taiwan.
- Geoffroy Peeters and Emmanuel Deruty. 2009. Is music structure annotation multi-dimensional? A proposal for robust local music annotation. In Proceedings of the International Workshop on Learning the Semantics of Audio Signals, pages 75–90, Graz, Austria.

Works cited

- Alan Pollack. 2001. "Notes on ... Series." http:// www.icce.rug.nl/~soundscapes/DATABASES/AWP/awp-notes_on.shtml
- Colin Raffel, Brian McFee1, Eric J. Humphrey, Justin Salamon, Oriol Nieto, Dawen Liang, and Daniel P. W. Ellis. 2014. mir_eval: A transparent implementation of common MIR metrics. Proceedings of ISMIR, pages 367–372, Taipei, Taiwan.
- Jordan B. L. Smith, J. Ashley Burgoyne, Ichiro Fujinaga, David De Roure, and J. Stephen Downie. 2011. Design and creation of a large-scale database of structural annotations. Proceedings of ISMIR, pages 555–560, Miami, FL, USA.
- Jordan B. L. Smith and Elaine Chew. 2013. A meta-analysis of the MIREX Structure Segmentation task. Proceedings of ISMIR, pages 251–256, Curitiba, Brazil.
- Jordan B. L. Smith and Elaine Chew. 2013. Using quadratic programming to estimate feature relevance in structural analyses of music. In Proceedings of the ACM International Conference on Multimedia, pages 113–122, Barcelona, Spain.
- Jordan B. L. Smith. 2014. Explaining listener disagreements in the perception of musical structure. PhD thesis. Queen Mary University of London.
- Jordan B. L. Smith. Structural Analysis Evaluation. https://code.soundsoftware.ac.uk/projects/structural_analysis_evaluation
- Julián Urbano. 2011. Information retrieval meta-evaluation: Challenges and opportunities in the music domain. Proceedings of ISMIR, pages 609–614, Miami, FL, USA.

Music

- "Blue Fiddle" by Claude Lamothe (SALAMI ID 104)
- "We Are The Champions" by Queen (SALAMI ID 1606)
- "Come Together" by The Beatles
- "Run Fay Run" by Isaac Hayes
- "Garrotin" by Chado Rodrigo (SALAMI ID 842)

Thank you!