Discovering Structure in Music: Automatic Approaches and Perceptual Evaluations

Oriol Nieto

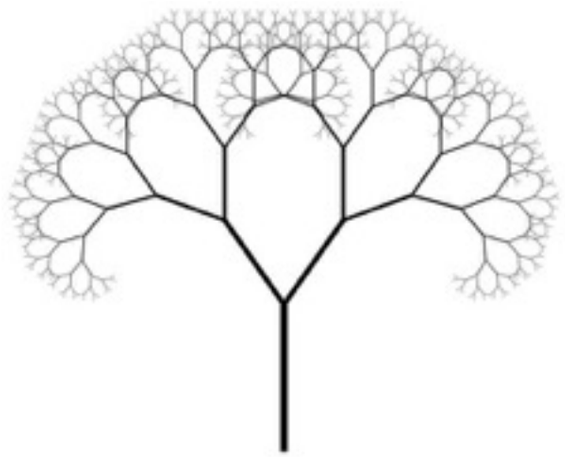
Doctoral Dissertation Defense

New York, NY
February 5th 2015

MARL
NYU Music and Audio Research Laboratory

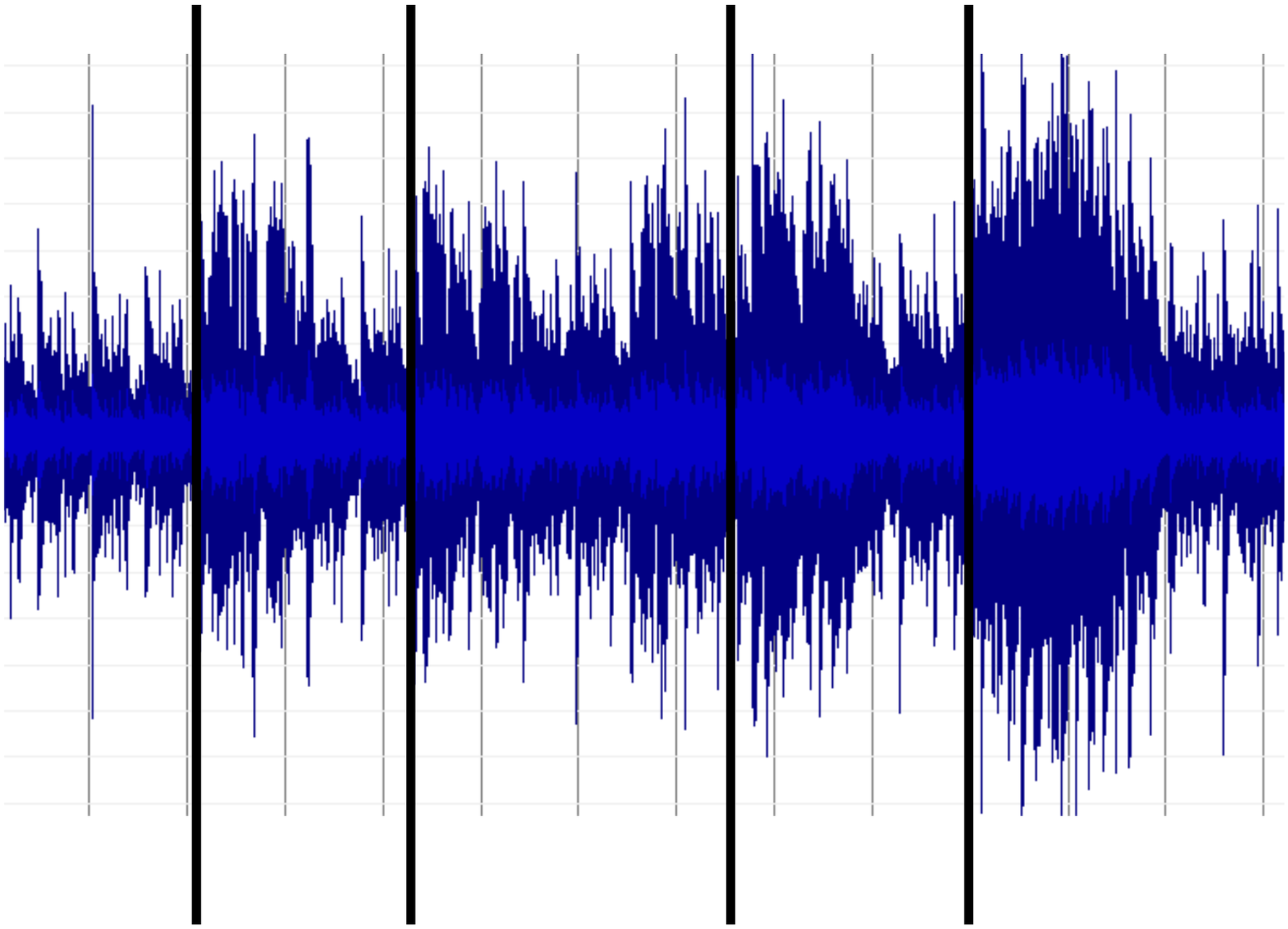
Background

The Structure of Music
(or why I am here)



MARL

Background



Results for this MIR task are far from being perfect

MARL

Why the improvement of this task matters?

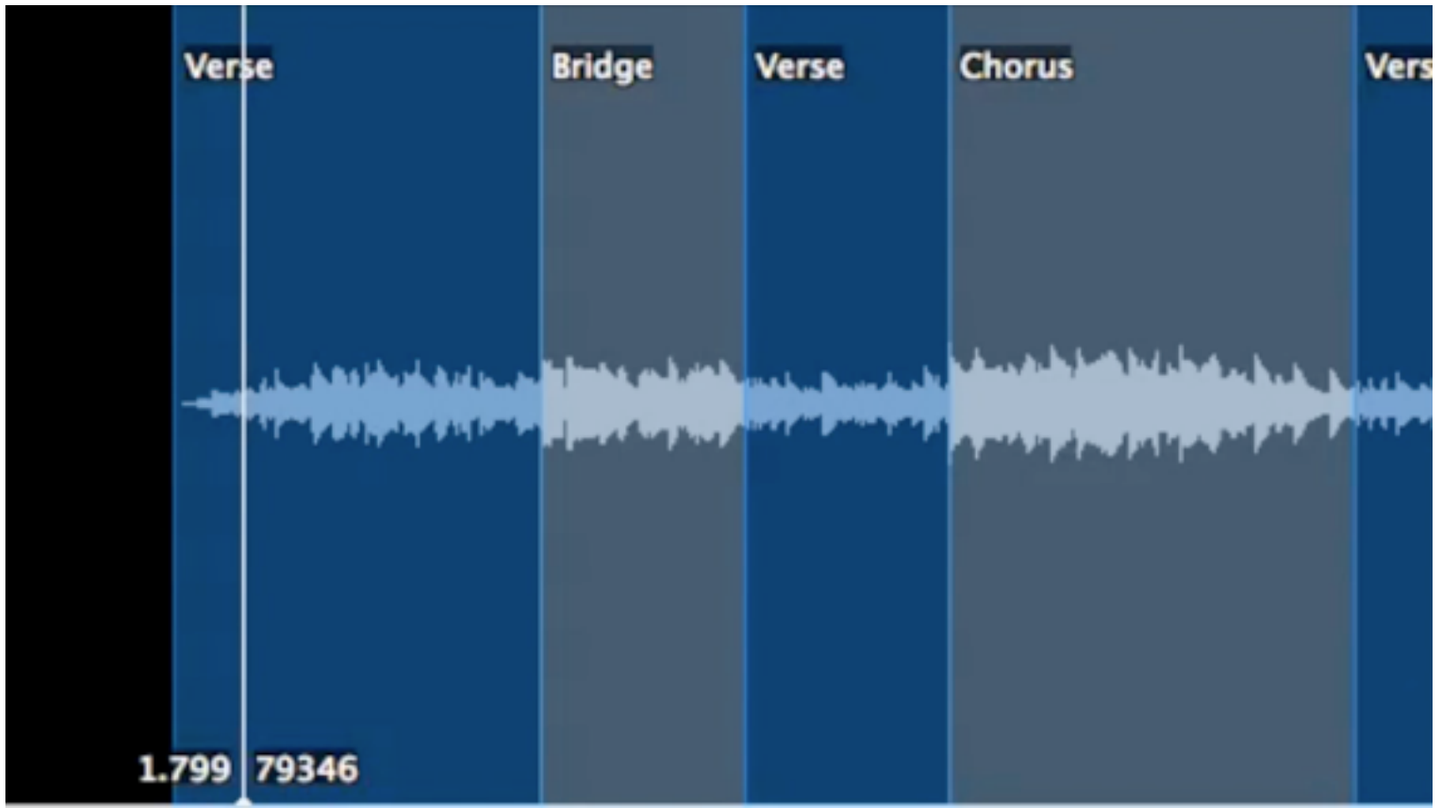
- The automatic discovery of the structure of music could:
  - Assist musicians when composing new pieces
  - Help audio engineers when editing tracks
  - Improve music recommendation systems
  - Make music players *smarter*
  - Generate music summaries to preview tracks
  - Yield better automatic dj/remix applications
  - Produce interactive visualization of musical pieces

Goal #1

Present novel automatic approaches to discover structure in music

MARL

Segment Annotation



(Trains by Porcupine Tree)

MARL

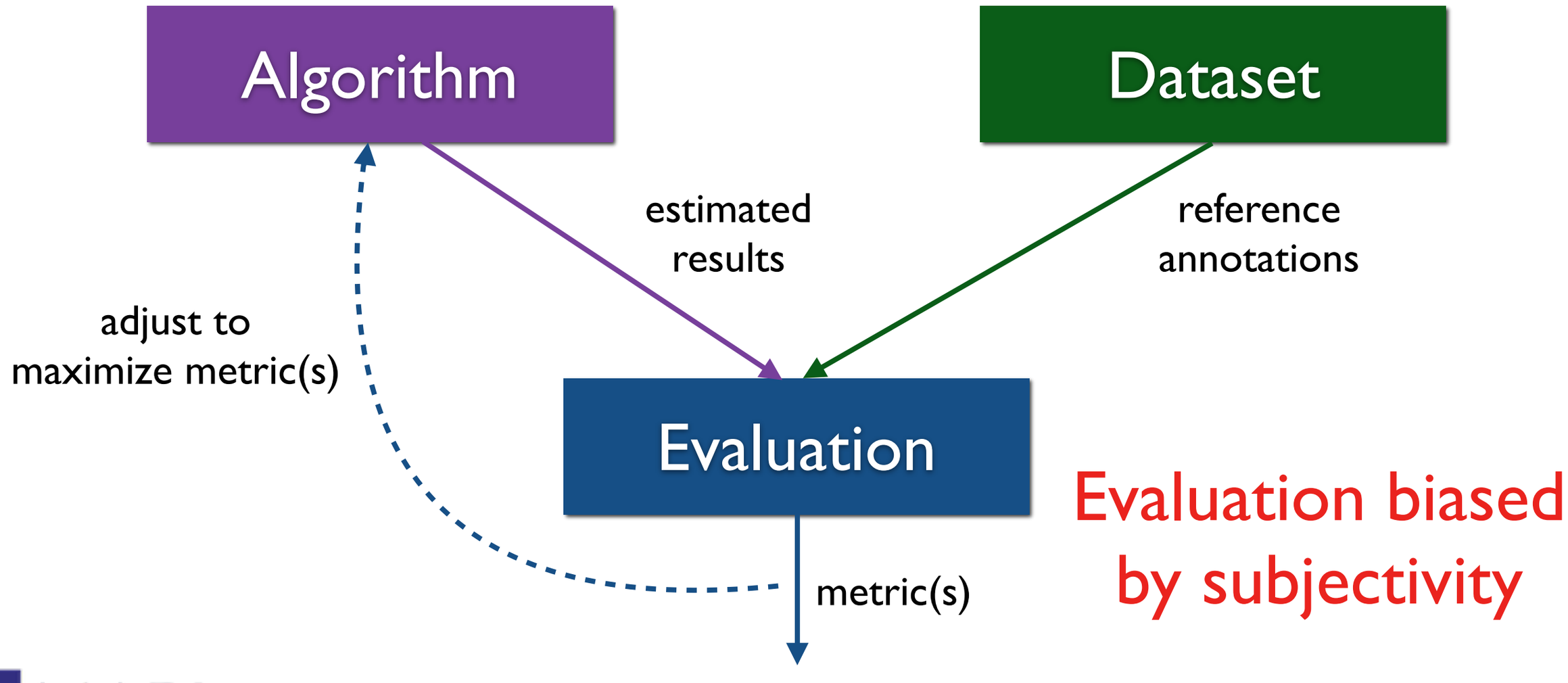
The Perception of Music Structure is Highly Subjective

MARL

Music Information Retrieval

- Standard methodology:

One annotation per track



Evaluation biased by subjectivity

Goal #2

Address the methodological issue of subjectivity inherent in the music segmentation task of MIR by proposing perceptual evaluations.

MARL

Automatic Approaches

- Four novel algorithms to discover structure in music:
  - Music Summaries
  - Pattern Discovery
  - Music Segmentation with Convex-NMF
  - Music Segmentation with 2D-FMC

MARL

Automatic Approaches

- Four novel algorithms to discover structure in music:
  - **Music Summaries**
  - Pattern Discovery
  - Music Segmentation with Convex-NMF
  - Music Segmentation with 2D-FMC

MARL

Music Summaries

- **Goal** of this MIR task:
  - Obtain a brief audio signal that *summarizes* a music piece in just a few seconds.
- **Example**:

MARL

Music Summaries

Main idea

Identify the *most* repeated parts (i.e., most relevant)

with the *least* amount of overlap (Nieto et al. 2012)

- Music Summary Criterion
  - Combine two values (harmonic mean):
    - Degree of Compression
    - Amount of Disjoint Information

MARL

Music Summaries - Results

- No standard evaluation for Music Summarization.
- Chopin's Mazurka Op. 30 No. 2.
- 3 repeated parts (AABBCC).
- Summary is composed of short parts of A, B, and C.

MARL

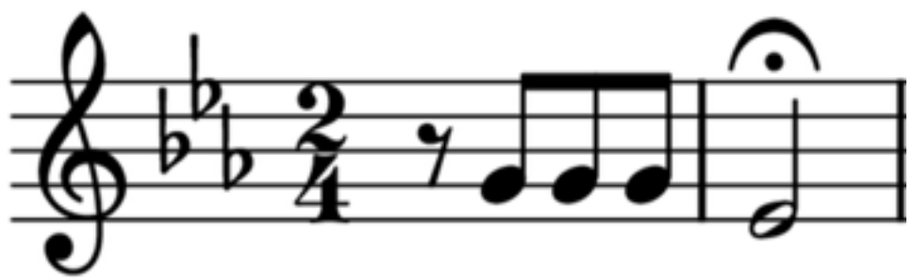
Automatic Approaches

- Four novel algorithms to discover structure in music:
  - Music Summaries
  - **Pattern Discovery**
  - Music Segmentation with Convex-NMF
  - Music Segmentation with 2D-FMC

MARL

Pattern Discovery Task

- **Goal** of this MIR task:
  - Identify the repeated parts of a given music piece.
    - Establish all the patterns contained in a piece
      - Identify all the occurrences across the piece for each pattern found
  - The shortest parts are typically *motives.*
  - The longest parts are typically *large-scale sections.*



MARL

Proposed Approach

- **Idea:** Make use of music segmentation techniques to obtain the most repeated parts of a given audio track using a greedy algorithm (Nieto and Farbood, 2014a).

Pattern Discovery - Results

- Evaluated on the JKU Development Dataset.
- Using the same metrics as in the MIR Evaluation eXchange (MIREX).
- State-of-the-art results when identifying occurrences in audio (when compared to audio-based algorithms that do not apply music transcription techniques).
  - Symbolic approaches yield superior results.
- State-of-the art when establishing patterns in audio.
  - Competitive (and sometimes better) than other symbolic approaches.

Automatic Approaches

- Four novel algorithms to discover structure in music:
  - Music Summaries
  - Pattern Discovery
  - **Music Segmentation with Convex-NMF**
  - Music Segmentation with 2D-FMC

MARL

Music Segmentation

- **Goal** of this MIR task:
  - Identify the different segments (or sections) of a music piece:
    - Determine the segment boundaries.
    - Label the different segments based on their similarity.
  - Segments tend to represent large-scale musical sections (e.g., verse, chorus, bridge).

Music Segmentation - C-NMF

- **Idea:** Factorize harmonic representations into different “segment prototypes” (centroids) using a machine learning tool (Nieto and Jehan, 2013).
  - Convex Non-negative Matrix Factorization (C-NMF)
  - Music segments can have homogenous harmonic distributions.

Results

- Evaluated on the ISO-Beatles and SALAMI datasets.
- Using the same metrics as in MIREX.
- State-of-the art (compared to other approaches that extract homogeneous segments) in terms of:
  - boundary retrieval
  - label grouping

Automatic Approaches

- Four novel algorithms to discover structure in music:
  - Music Summaries
  - Pattern Discovery
  - Music Segmentation with Convex-NMF
  - **Music Segmentation with 2D-FMC**

MARL

Proposed Approach

- **Idea**: Capture similarity between segments using a representation that is:
  - key-invariant
  - shift-invariant
  - tempo-agnostic
- Ideal candidate: 2D-Fourier Magnitude Coefficients (Nieto and Bello, 2014)

MARL

Results

- Evaluated on the ISO-Beatles and SALAMI datasets using MIREX metrics.
- Competitive results when using ground-truth boundaries.
- Strong impact on results when using estimated boundaries.
- Highly efficient in terms of computation time.

Summary of Goal# 1

- Four novel approaches to discover certain aspects of music structure:
  - Music Summaries
  - Pattern Discovery
    - https://github.com/urinieto/MotivesExtractor
  - Music Segmentation:
    - C-NMF
    - 2D-FMC
    - https://github.com/urinieto/msaf

MARL

Main Goals

- Present novel automatic approaches to discover structure in music.

- **Address the methodological issue of subjectivity inherent in the music segmentation task of MIR by proposing perceptual evaluations.**

MARL

Perceptual Evaluations

- Two types of novel evaluations:
    - Metrics for multiple annotations per track.
    - Modifying existing metrics to align better with perception.
- Tools from Music Perception and Cognition.

MARL

Perceptual Evaluations

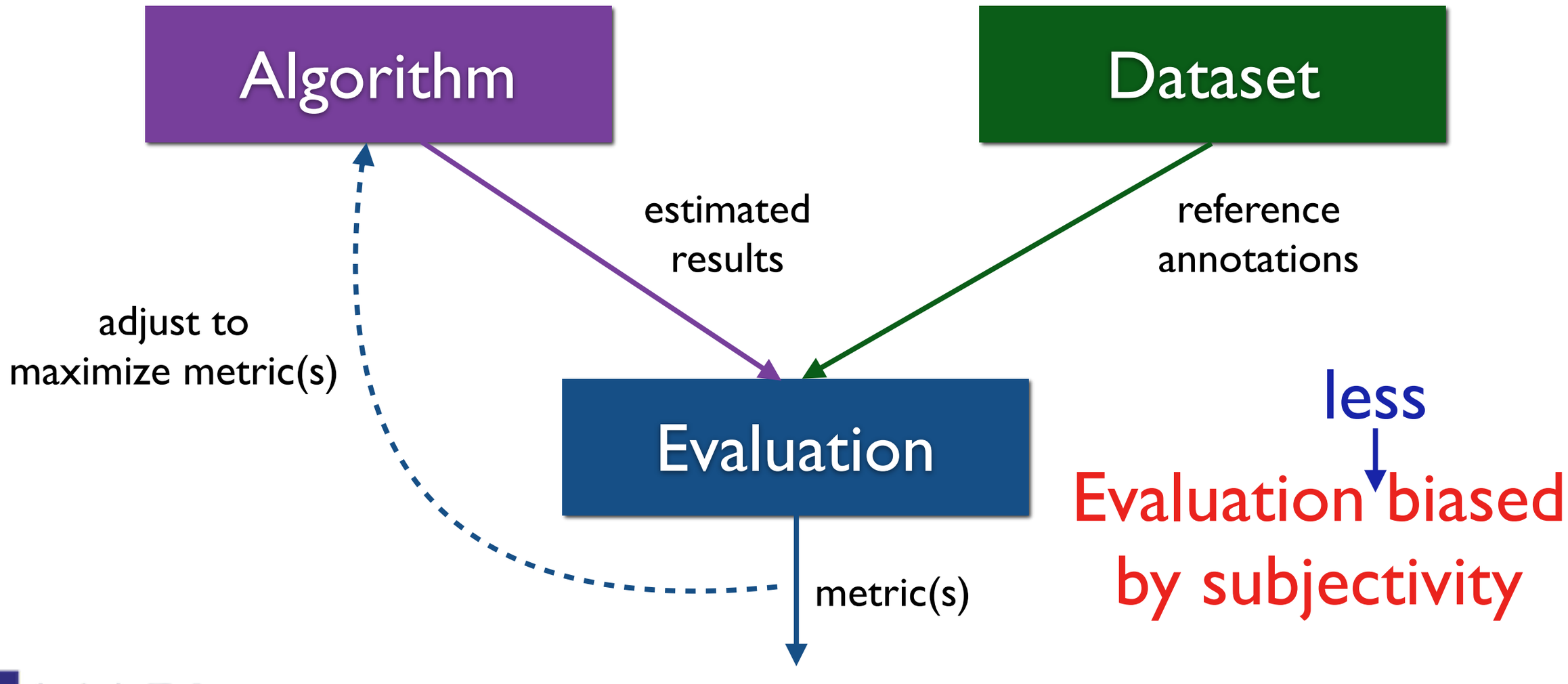
- Two types of novel evaluations:
  - **Metrics for multiple annotations per track.**
    - Modifying existing metrics to align better with perception.
- Tools from Music Perception and Cognition.

MARL

Music Information Retrieval

- Standard methodology:

Multiple
~~One~~ annotations
per track



less
↓
Evaluation biased
by subjectivity

MARL

Selecting Tracks

- From a large collection of >2,000 human annotated tracks:
    - Run multiple boundary retrieval algorithms.
    - Rank them based on a standard evaluation metric (F-measure with a 3 seconds window).
    - Choose the 45 worst performing tracks (i.e. challenging from a machine point of view).
    - Choose the 5 best performing tracks (i.e. trivial from a machine point of view).

- 5 music experts annotated the 50 selected tracks.
    - Two levels of segmentation: large and small.

- Each track will now contain five additional two-layer segmentation annotations.

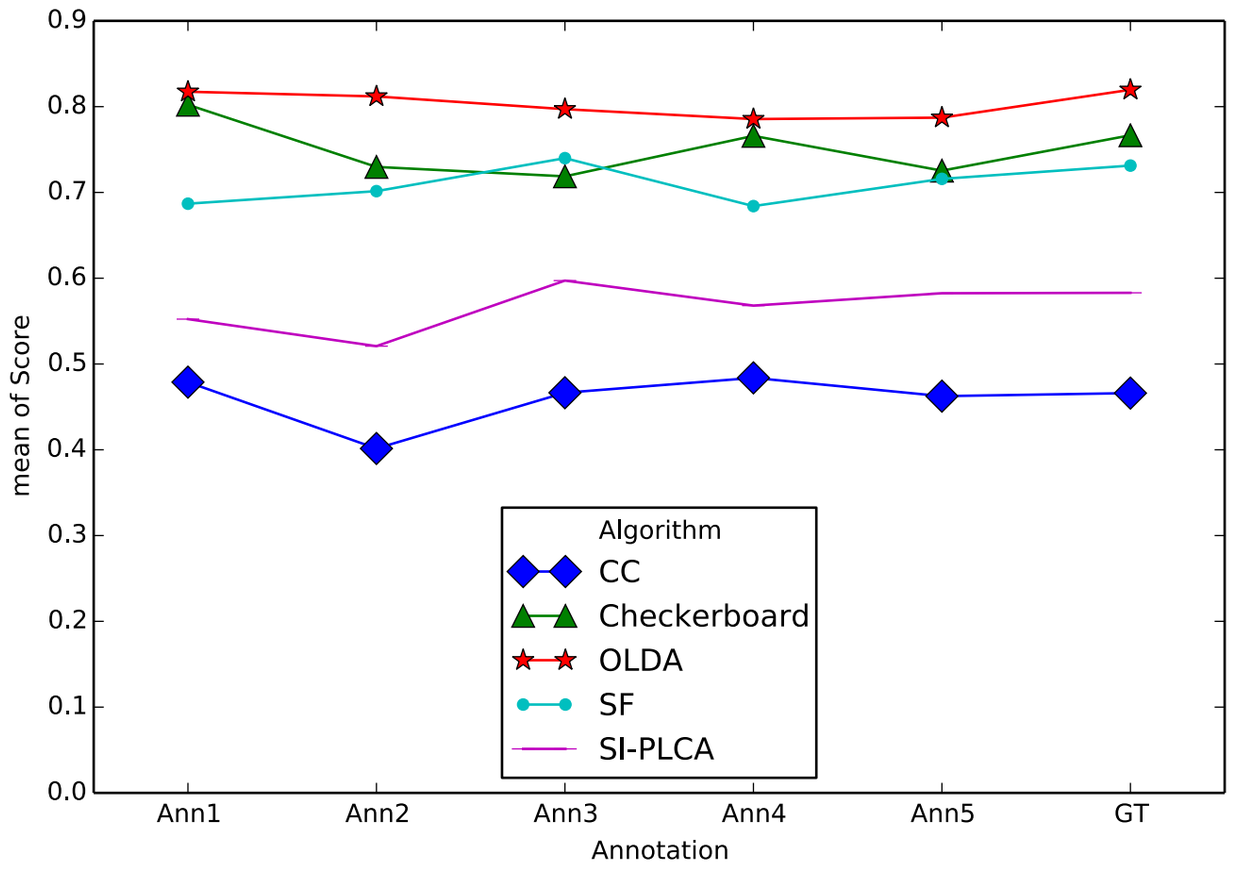
MARL

Analysis of Subjectivity

- Analyze the variation of the scores when evaluating the estimated boundaries with the new annotations.

- Use a 2-way ANOVA of the average F-measure with algorithms and annotations as factors.

- Start with the control group:



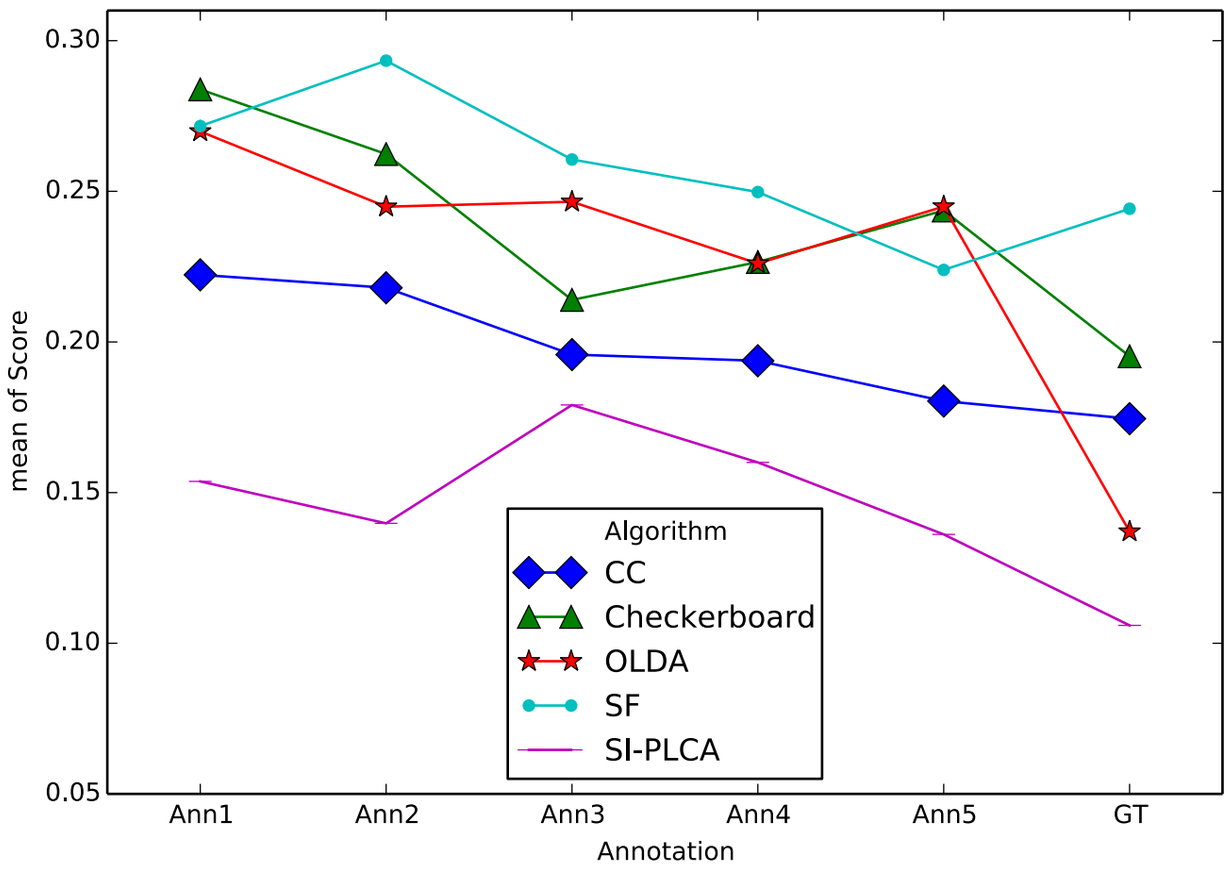
Annotations Effect:
F (5, 120) = .22, p = .95

Interaction:
F(20,120) = .13,p = .99

MARL

Analysis of Subjectivity

- No significant variation for the control group when using different annotations.
- What about the challenging group?



Annotations Effect:
$F(5, 1320) = 6.93, p < .01$

Interaction:
$F(20, 1320) = 1.13, p = .3$

MARL

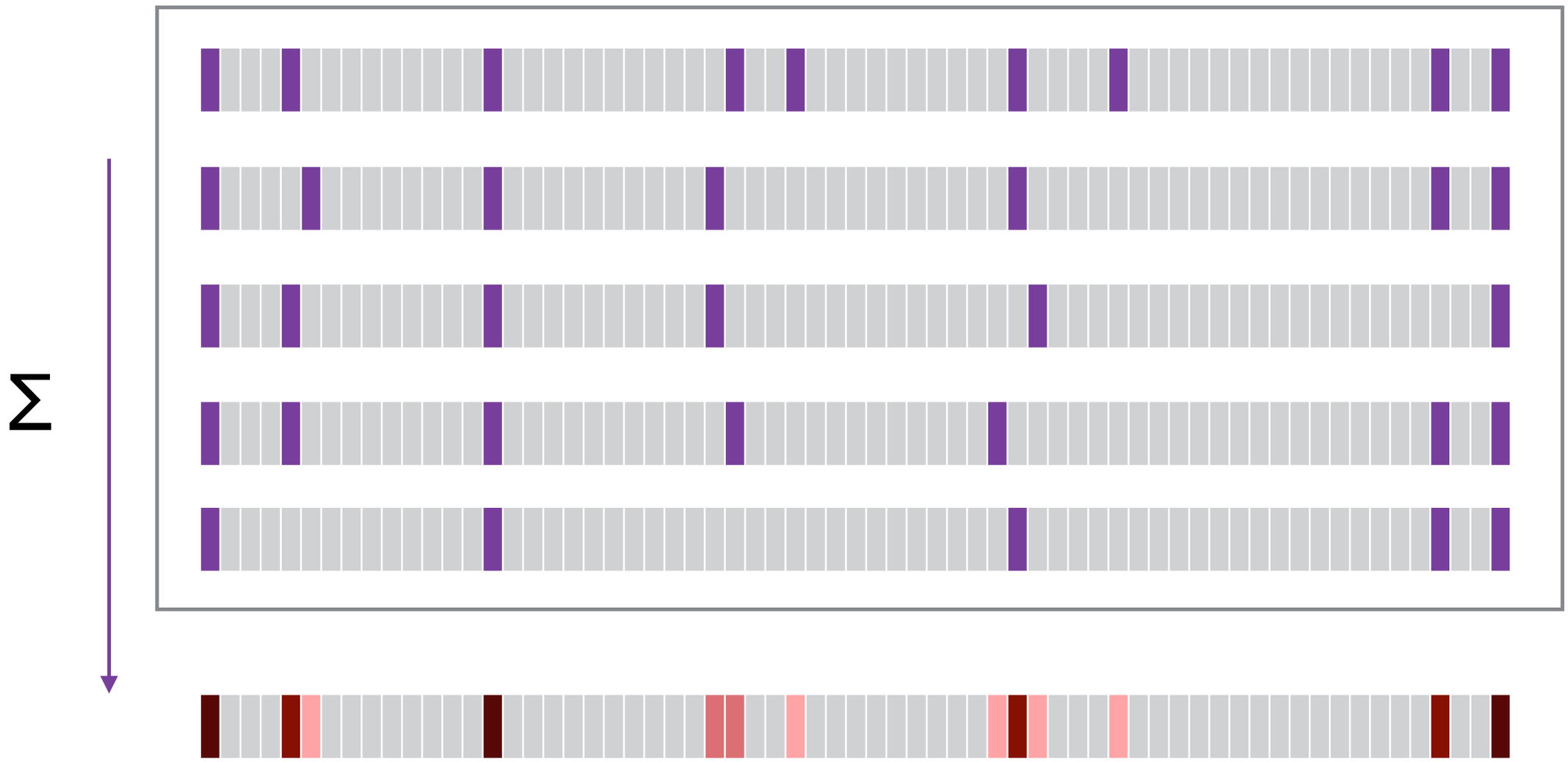
Analysis of Subjectivity

- **Significant variation** when using different annotations for the challenging tracks.

- Therefore:
  - Subjectivity is a relevant problem when evaluating music boundaries.
  - At least on the challenging tracks.

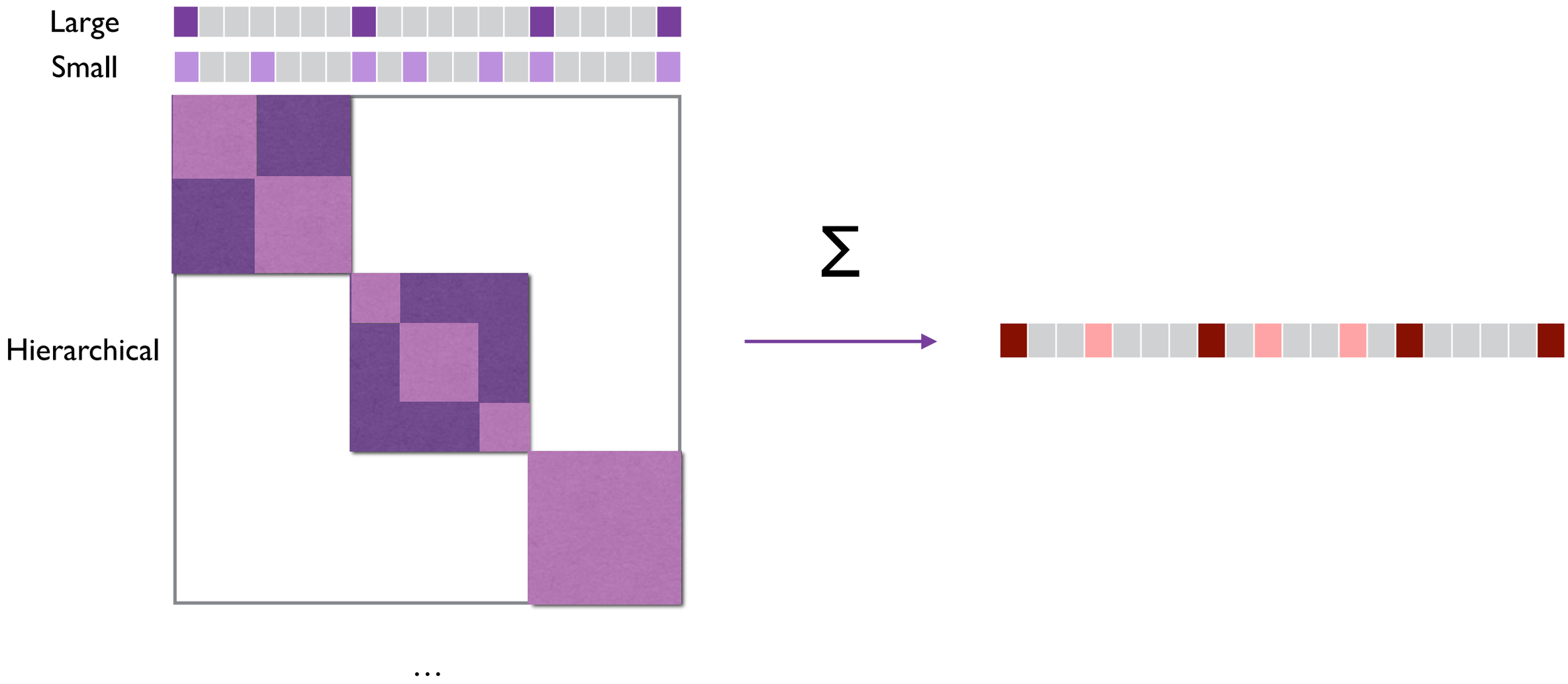
- Can we minimize the subjectivity effect for this task?
  - Yes, merging the annotated boundaries.
  - 4 types of merging

Merging Type I:
Flat to Weighted Flat



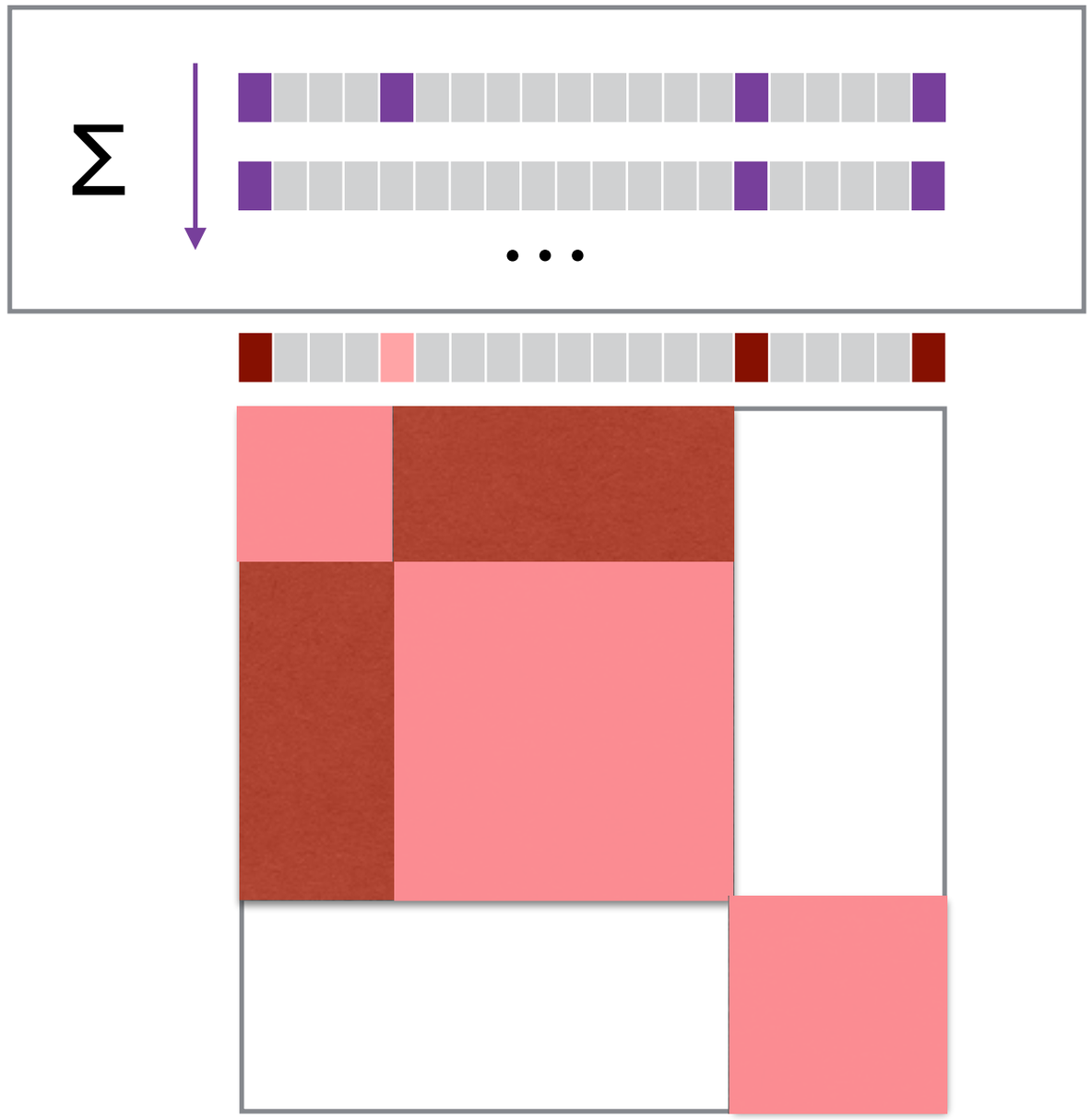
MARL

Merging Type II: Hierarchical to Weighted Flat



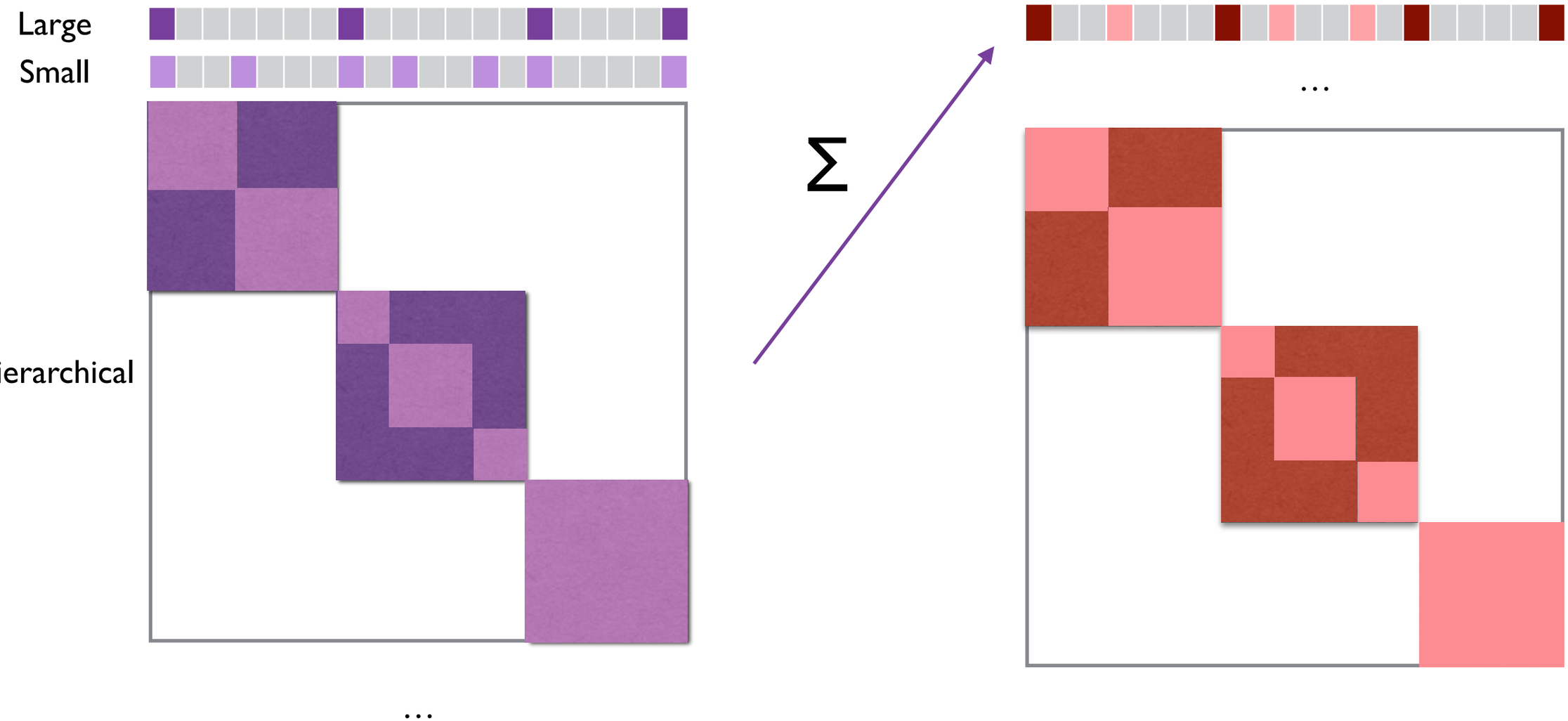
MARL

Merging Type III: Flat to Hierarchical



MARL

Merging Type VI: Hierarchical to Hierarchical



MARL

Robustness of Merged Boundaries

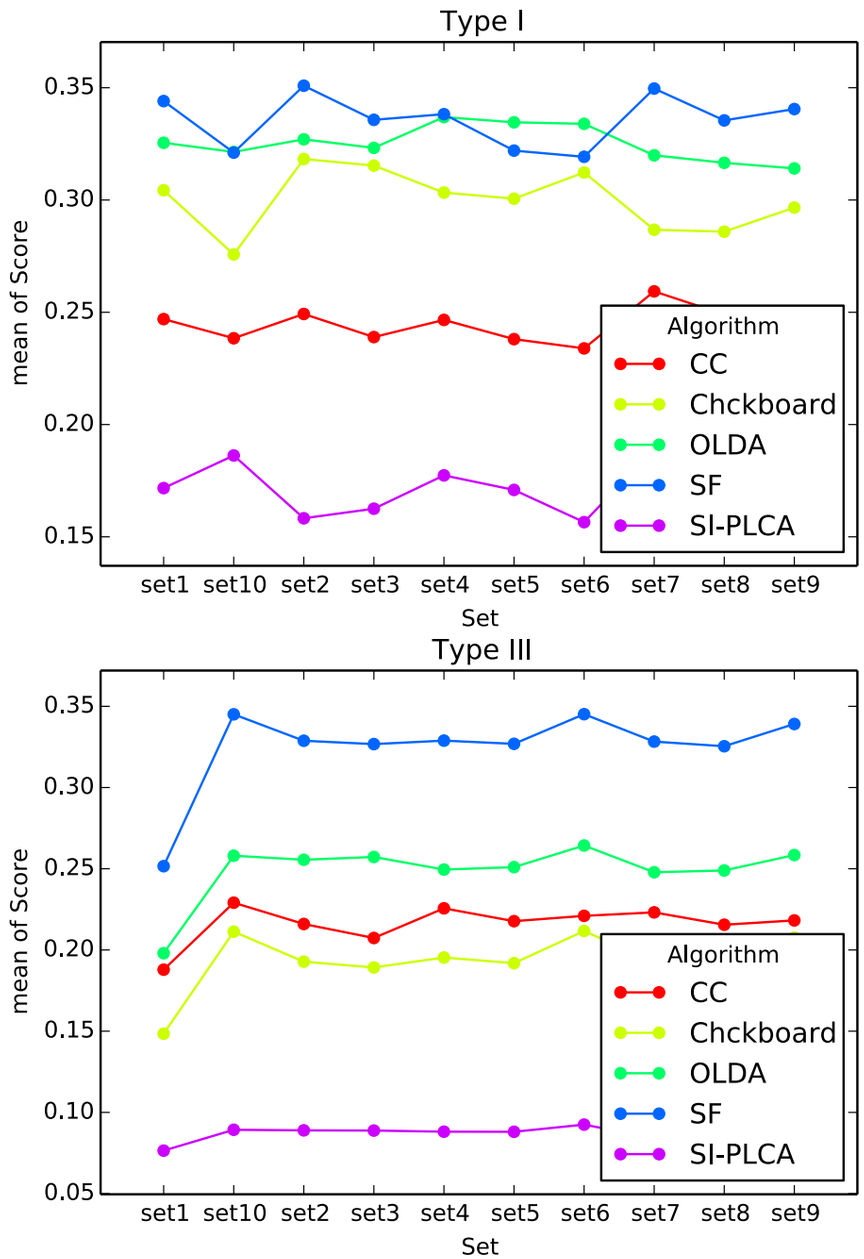
- In order to test the robustness of this merging, I divide annotations into sets of 3:
  - 5 annotators, dividing them into sets of 3. ($\binom{5}{3} = 10$)
  - Similar to cross-validation.

- For each of 10 sets, I merge their annotations using the four different types (types I, II, III and IV).
- For each type, compute two-way ANOVA with algorithm and sets as factors.

MARL

Robustness of Merged Boundaries

Merge Type	$F(9, 2200)$	p-value
I	.23	.99
II	.42	.92
III	3.35	$< .01$
IV	1.56	.12



- Except type III none of the scores significantly vary depending on the set chosen.
- No conflicts in marginal means in types III and IV.

MARL

Perceptual Evaluations

- Two type of novel evaluations:
  - Metrics for multiple annotations per track.
  - **Modifying existing metrics to align better with perception.**

MARL

Music Segmentation Evaluation

- Standard metric: the F-measure (or F1-score):
  - Quantizes the similarity between the annotations and the estimated results.
    - Is it appropriate in the framework of music segmentation? Does it align with humans' perception of the structure in music?

- I aim to perceptually redefine the F-measure for evaluating **music boundaries**.

MARL

F-measure

- Find intersection between reference annotations and estimated results:
  - Estimated boundaries are correct (hits) if they are within 3 seconds from the reference one.
- **Precision**: Ratio between hits and the total number of estimated elements.
- **Recall**: Ratio between hits and the total number of reference elements.

$$P = \frac{|\text{hits}|}{|\text{bounds}_e|} \qquad R = \frac{|\text{hits}|}{|\text{bounds}_a|}$$

- **F-measure**: Harmonic mean between P and R.
  - Weights both values equally.
  - Penalizes outliers.
  - Mitigates impact of large values.

$$F = 2\frac{P \cdot R}{P + R}$$

MARL

F-measure for Boundary Evaluation

- Higher Precision represents less false positives.
- Higher Recall represents less false negatives.
- When listening to estimated results of music segmentation, it becomes apparent that these two values are perceptually very different.
- Assess the relative effect that these differences have on human evaluations in order to redefine the F-measure.
  - Two Experiments

MARL

Experiments

- Designed to explore the preference between precision and recall.
- Conducted online, with 48 and 23 participants, respectively.
- Results suggest that **Precision** tends to be more perceptually salient than **Recall**:
  - Humans prefer to listen to “less but correct” than “more but not necessarily precise” boundaries.

MARL

Perceptually Redefining the F-measure

- The generic form of the F-measure is:

$$F_\alpha = (1 + \alpha^2) \frac{P \cdot R}{\alpha^2 P + R}$$

- If alpha = 1: R and P have the same weight (F1-score)
- If alpha > 1: more importance to R
- If alpha < 1: more importance to P

$$\alpha < 1$$

MARL

Summary of Goal#2

- Merging annotations:
  - Datasets with a single human annotation per track are prone to error.
  - Merging multiple annotations can significantly alleviate the subjectivity effect.
- Redefining existing metrics:
  - The F-measure could be redefined to better line up with perception.
  - Precision is perceptually more relevant than Recall
- Including these perceptual evaluations in the MIR methodology would result in applications that better align with human preference.

Conclusions and Future Work

- Presented 4 novel methods to automatically discover structure in music.
- Presented 2 novel evaluations for music segmentation that better align with human perception.
- Narrowed the gap between Music Information Retrieval and Music Perception and Cognition.
- Structure is regarded as hierarchical, and it is likely that future approaches to discover structure might output hierarchical results.
- Given the ambiguity of the task, in the future algorithms may produce more than one “valid” answer.
- Similar aggregation of annotations could also be employed in other subjective MIR tasks such as chords, tags, or mood.

MARL

Acknowledgments

MARL

References

- Bruderer, M. J., Mckinney, M. F., & Kohlrausch, A. (2009). The Perception of Structural Boundaries in Melody Lines of Western Popular Music. Musicæ Scientiæ, 13(2), 273–313.
- Collins, T., Arzt, A., Flossmann, S., & Widmer, G. (2014). SIARCT-CFP: Improving Precision and the Discovery of Inexact Musical Patterns in Point-set Representations. In Proc. of the 14th International Society for Music Information Retrieval Conference (pp. 549–554). Curitiba, Brazil.
- Collins, T. (2013). Discovery of Repeated Themes & Sections. Music Information Retrieval Evaluation eXchange. Retrieved January 08, 2013, from http://www.music-ir.org/mirex/wiki/2013:Discovery_of_Repeated_Themes_&_Sections
- Ellis, D. P. W., and Poliner, G. E. (2007). Identifying "Cover Songs" with Chroma Features and Dynamic Programming Beat Tracking. In Proc. of the 32nd IEEE International Conference on Acoustics Speech and Signal Processing (pp. 1429–1432). Honolulu, HI, USA.
- Müller, M., and Clausen, M. (2007). Transposition-Invariant Self-Similarity Matrices. In Proc. of the 8th International Conference on Music Information Retrieval (pp. 47–50). Vienna, Austria.
- Nieto, O., Humphrey, E. J., Bello, J. P. (2012), Compressing Music Recordings into Audio Summaries. Proc. of the 13th International Society for Music Information Retrieval Conference (ISMIR), pages 313-318. Porto, Portugal.
- Nieto, O. and Jehan, T. (2013). Convex Non-Negative Matrix Factorization For Automatic Music Structure Identification. In Proc. of the 38th IEEE International Conference on Acoustics Speech and Signal Processing, pages 236–240, Vancouver, Canada.
- Nieto, O., and Farbood, M. (2013). MIREX 2013: Discovering Musical Patterns Using Audio Structural Segmentation Techniques. In Music Information Retrieval Evaluation eXchange. Curitiba, Brazil.
- Nieto, O. and Bello, J. P. (2014). Music Segment Similarity Using 2D-Fourier Magnitude Coefficients. In Proc. of the 39th IEEE International Conference on Acoustics Speech and Signal Processing, pages 664–668, Florence, Italy.
- Nieto, O. and Farbood, M. M. (2014a). Identifying Polyphonic Patterns From Audio Recordings Using Music Segmentation Techniques. In Proc. of the 15th International Society for Music Information Retrieval Conference, pages 411–416, Taipei, Taiwan.

MARL

Thanks!

MARL