

# Hierarchical Evaluation of Segment Boundary Detection

Brian McFee<sup>1,2</sup>, Oriol Nieto<sup>2,3</sup>, Juan P. Bello<sup>2</sup>

<sup>1</sup>Center for Data Science, New York University

<sup>2</sup>Music and Audio Research Laboratory, New York University

<sup>3</sup>Pandora, Inc.



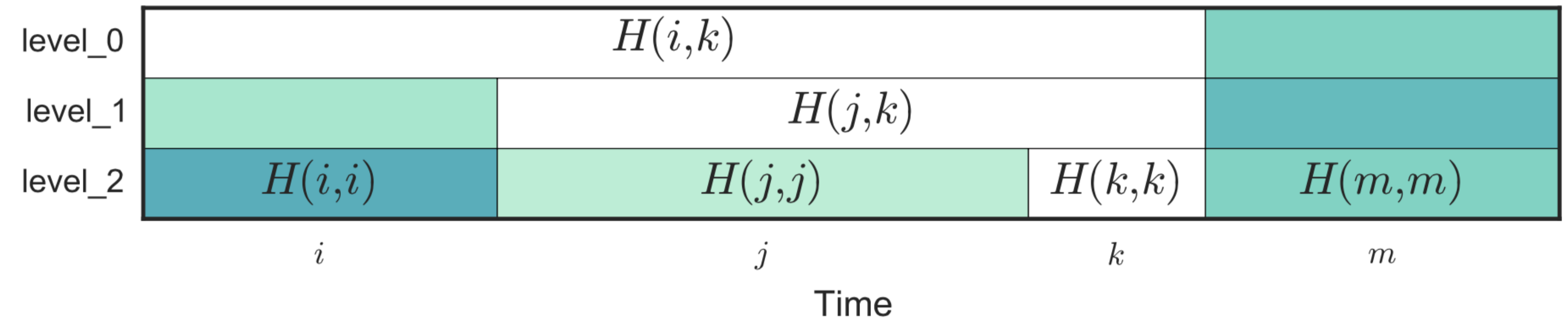
{brian.mcfee, oriol, jpbello}@nyu.edu

## Abstract

**T-measures:** new metric to evaluate hierarchical music segment boundaries:

- ▶ First measures published to assess multiple levels of music segmentation.
- ▶ Allowing flat and hierarchical references and estimations.

## Representing Hierarchical Boundaries



**n:** number of frames in a track.

**H:** hierarchical segmentation composed of multiple flat segmentations.

**H(i, j):** most refined segment containing frames i and j.

$H(j, j) \prec H(j, k)$ : segment j is deeper in hierarchy than segment k.

In the example:

$$H(j, j) \prec H(j, k) \prec H(i, j) = H(i, k)$$

## T-measures

For a given frame q, count all i, j frames in the track such that the pair (q, i) is always more refined (i.e., deeper in the hierarchy) than (q, j):

$$g(q; H_E, H_R) := \sum_{\substack{(i,j), \\ i \neq q, \\ H_R(q,i) \prec H_R(q,j)}} \frac{\llbracket H_E(q,i) \prec H_E(q,j) \rrbracket}{Z_q}$$

**H<sub>E</sub>:** Estimated hierarchy; **H<sub>R</sub>:** Reference hierarchy; **Z<sub>q</sub>:** Normalization.

Tree-recall T-measure:

$$\mathcal{T}_R(H_E, H_R) := \frac{1}{n} \sum_q g(q; H_E, H_R)$$

Tree-precision T-measure:

$$\mathcal{T}_P(H_E, H_R) := \mathcal{T}_R(H_R, H_E)$$

Can be seen as a partial ranking problem: i is considered more relevant than j based on the hierarchy.

Harmonic mean of **T<sub>R</sub>** and **T<sub>P</sub>** produces an F-measure **T<sub>F</sub>**.

## Extending the T-measures

### Windowing in Time

$$g(q; H_E, H_R, w) := \sum_{\substack{i,j \in \{x: |q-x| \leq w/2\} \\ i \neq q, \\ H_R(q,i) \prec H_R(q,j)}} \frac{\llbracket H_E(q,i) \prec H_E(q,j) \rrbracket}{Z_q(w)}$$

**w:** time window; **Z<sub>q</sub>(w):** Normalization.

### Goals:

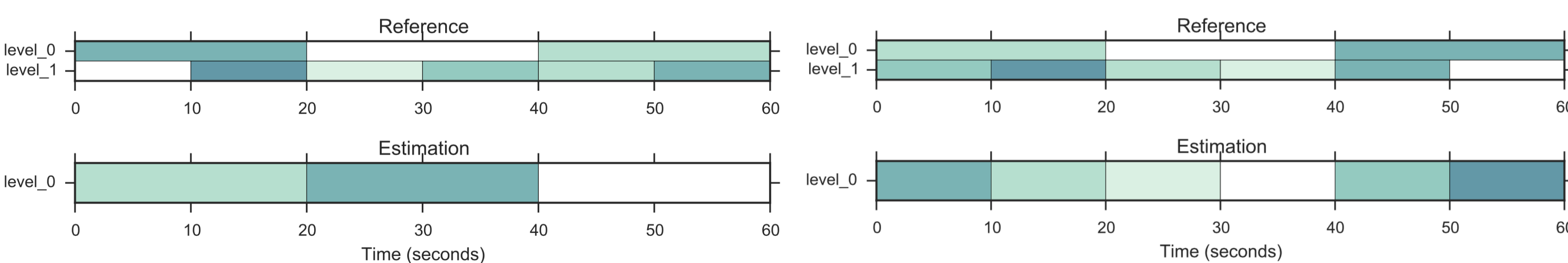
- Avoid domination of trivial comparisons in long tracks.
- Reduce computation complexity from  $O(n^3)$  to  $O(n w^2)$ .

### Transitive Reduction

Restrict summation to include only adjacent layers in the hierarchy.

**Goal:** Avoid over-counting of high levels of segmentation.

## Examples

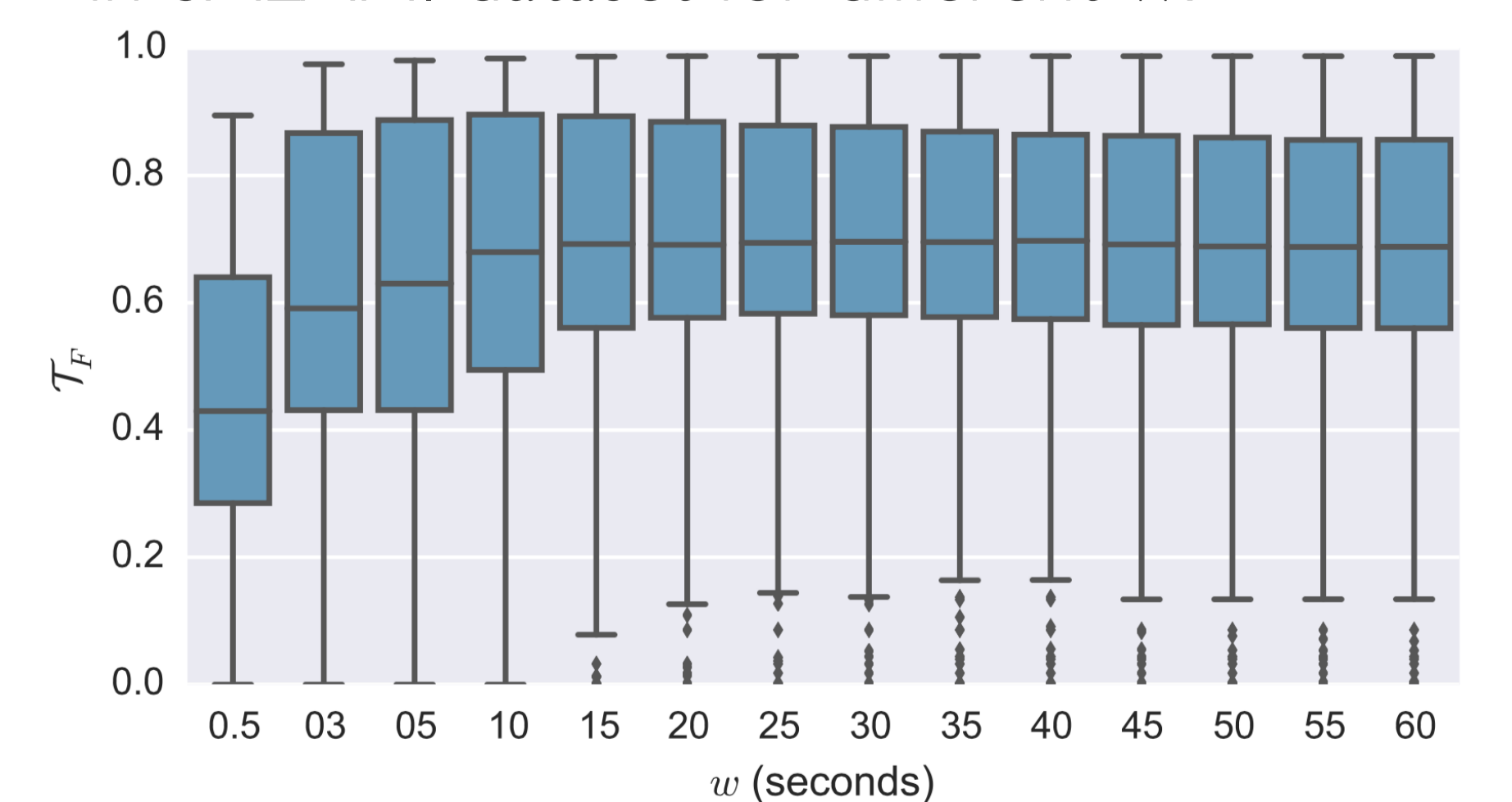


w	Reduced		Full	
	T <sub>R</sub>	T <sub>P</sub>	T <sub>R</sub>	T <sub>P</sub>
0.5	0.00	1.00	0.40	1.00
3	0.00	1.00	0.40	1.00
15	0.37	1.00	0.51	1.00
30	0.70	1.00	0.82	1.00
∞	0.80	1.00	0.89	1.00

w	Reduced		Full	
	T <sub>R</sub>	T <sub>P</sub>	T <sub>R</sub>	T <sub>P</sub>
0.5	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00
15	0.63	1.00	0.76	1.00
30	0.30	1.00	0.59	1.00
∞	0.20	1.00	0.55	1.00

## Exploring w

Compute T-measures between the 410 tracks with two hierarchical annotations in SALAMI dataset for different w:



w=15 estimated as a good value for SALAMI.

