Boundaries and novelty: the correspondence between points of change and perceived boundaries

Jordan B. L. Smith, Ching-Hua Chuan and Elaine Chew
DMRN+7 18 December 2012

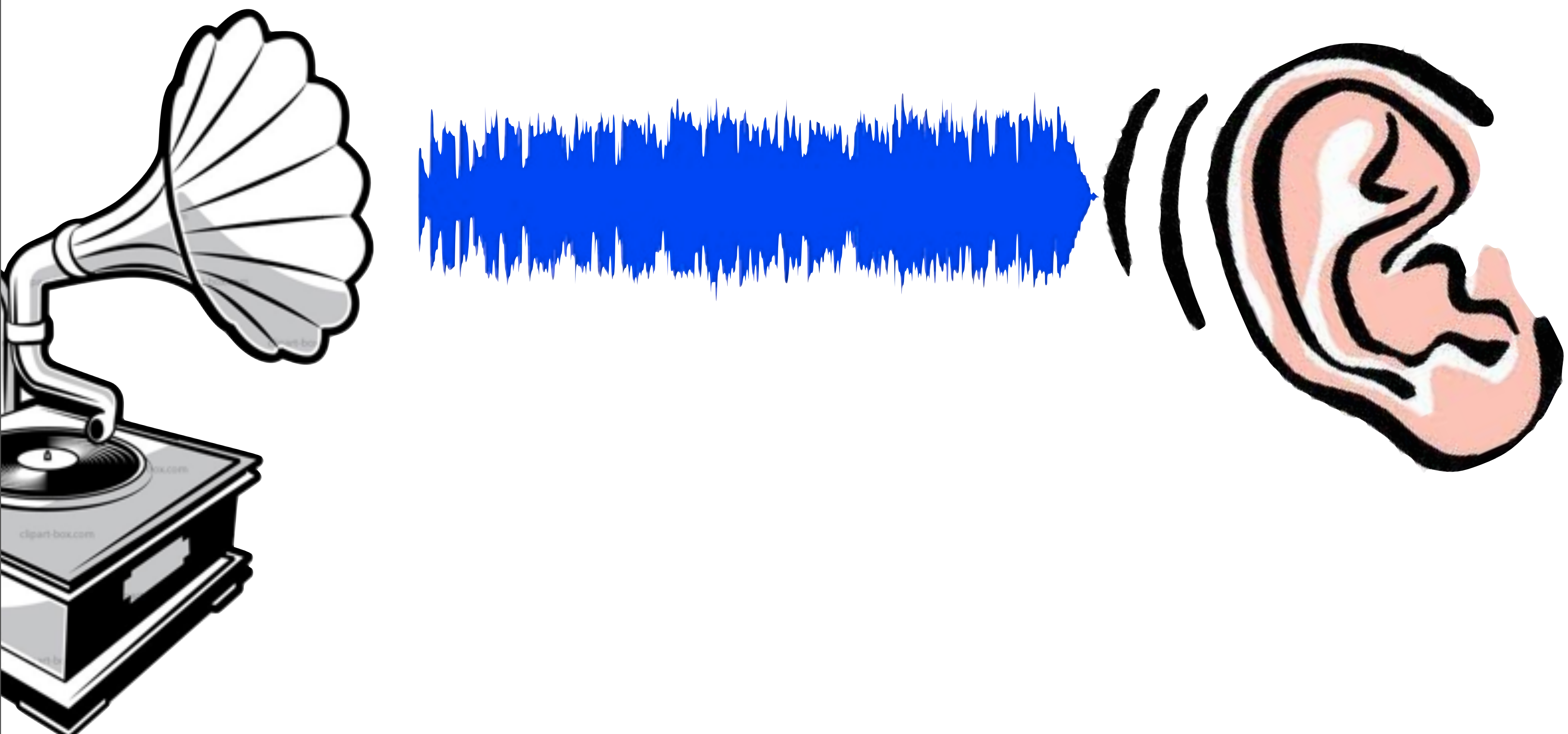
Outline

I. What the research is about and why it is very interesting

II. How the data were assembled and analyzed

III. What the results of the analysis are

Music is continuous,
but we hear it in chunks



Music is continuous, but we hear it in chunks

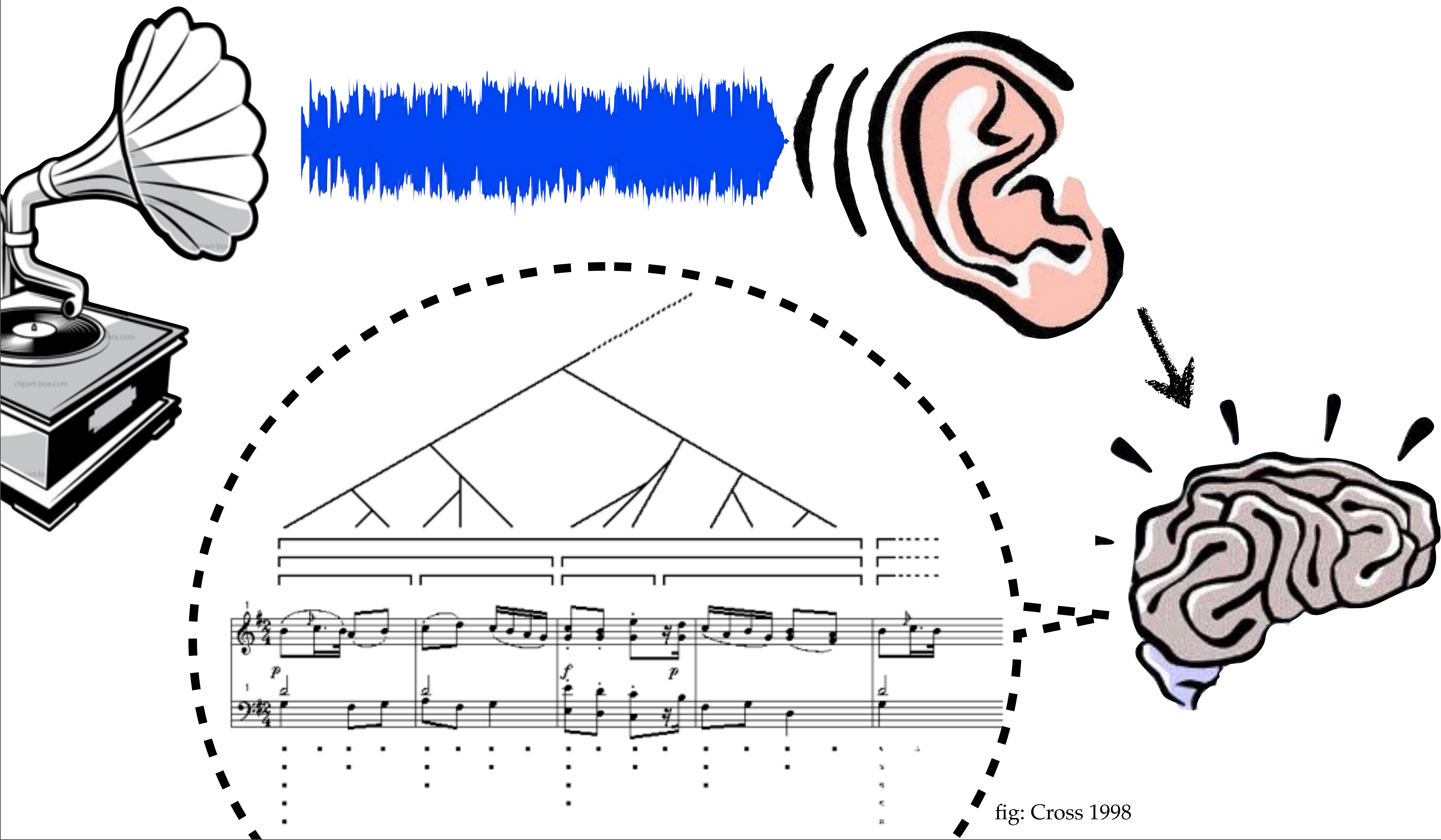
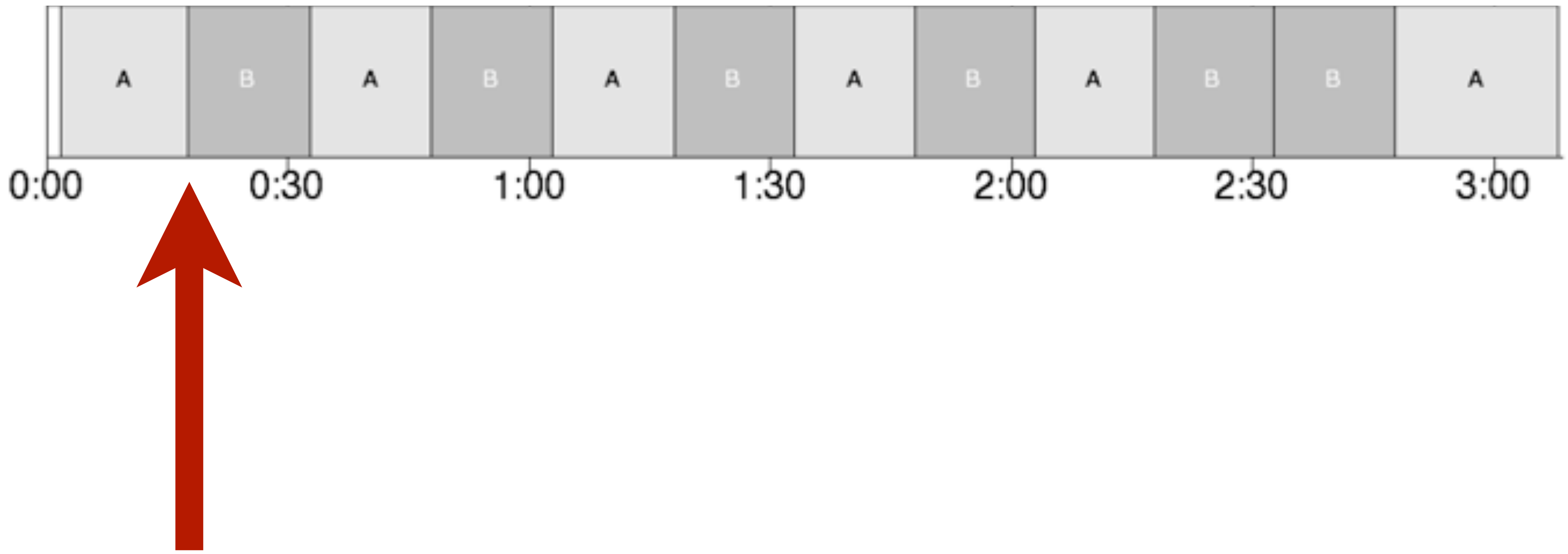


fig: Cross 1998

I’m going to talk about large-scale structure

I’m going to talk about large-scale structure



What causes a listener to believe there is a boundary here?

What causes a listener to hear a boundary?

Pause (silence) (4)
Return of material (chordal) (2)
Change of dynamic (2)

New material (chords changing to melody) (5)
Pause (silence) (2)
Change of rhythm (2)
Change of pitch content (2)
Change of articulation (1)

Return of first material (chordal) (5)

New material (change of pitch content) (4)
Start of Development (3)
Change of rhythm (2)
Change of articulation (2)
Change of register (expansion) (5)
Change of dynamic contour (3)
Change of texture (2)
Pause (1)

Return of material (chromatic run) (5)
Relaxation of tension (1)
Change of register (2)
Change of dynamic (1)

Return of material (chordal) (4)
Change of dynamic (1)

Return of material (chordal with new pitches) (4)
Introduction of trill (2)
Change of dynamic (1)
Pause (1)

New material (isolated block chords) (4)
Change of tempo (1)
Change of register (1)
Change of pitch content (1)

Clarke and Krumhansl 1990

change in harmonic progression
change in melody
change in tempo
change in rhythm
change in timbre
change in loudness / dynamics
breaks
global structure
repetitions

Bruderer 2008

Aviezer, Trope and
Todorov 2012

Aviezer, Trope and
Todorov 2012

We can use large-scale MIR studies to learn about perception of structure

novelty-based algorithm $\xrightarrow{X}$ ground truth boundaries

We can use large-scale MIR studies to learn about perception of structure

$$\begin{tikzcd}
\text{novelty-based algorithm} \arrow[r, "X"] & \text{ground truth boundaries} \\
\text{naive baseline algorithm} \arrow[ur, "Y"'] &
\end{tikzcd}$$

X – Y = the extent to which a novelty-based algorithm explains the ground truth better than a naive algorithm

We can use large-scale MIR studies to learn about perception of structure

$$\begin{tikzcd} \text{novelty-based algorithm} \arrow[r, "X"] \arrow[dr, "Y"'] & \text{ground truth boundaries} \\ & \text{random set of non-boundaries} \end{tikzcd}$$

X – Y = the extent to which novelty explains the boundaries better than it explains the non-boundaries

II. How the data were assembled and analyzed

SALAMI database: Structural Analysis of Large Amounts of Music Information

Social Sciences and Humanities Research Council of Canada

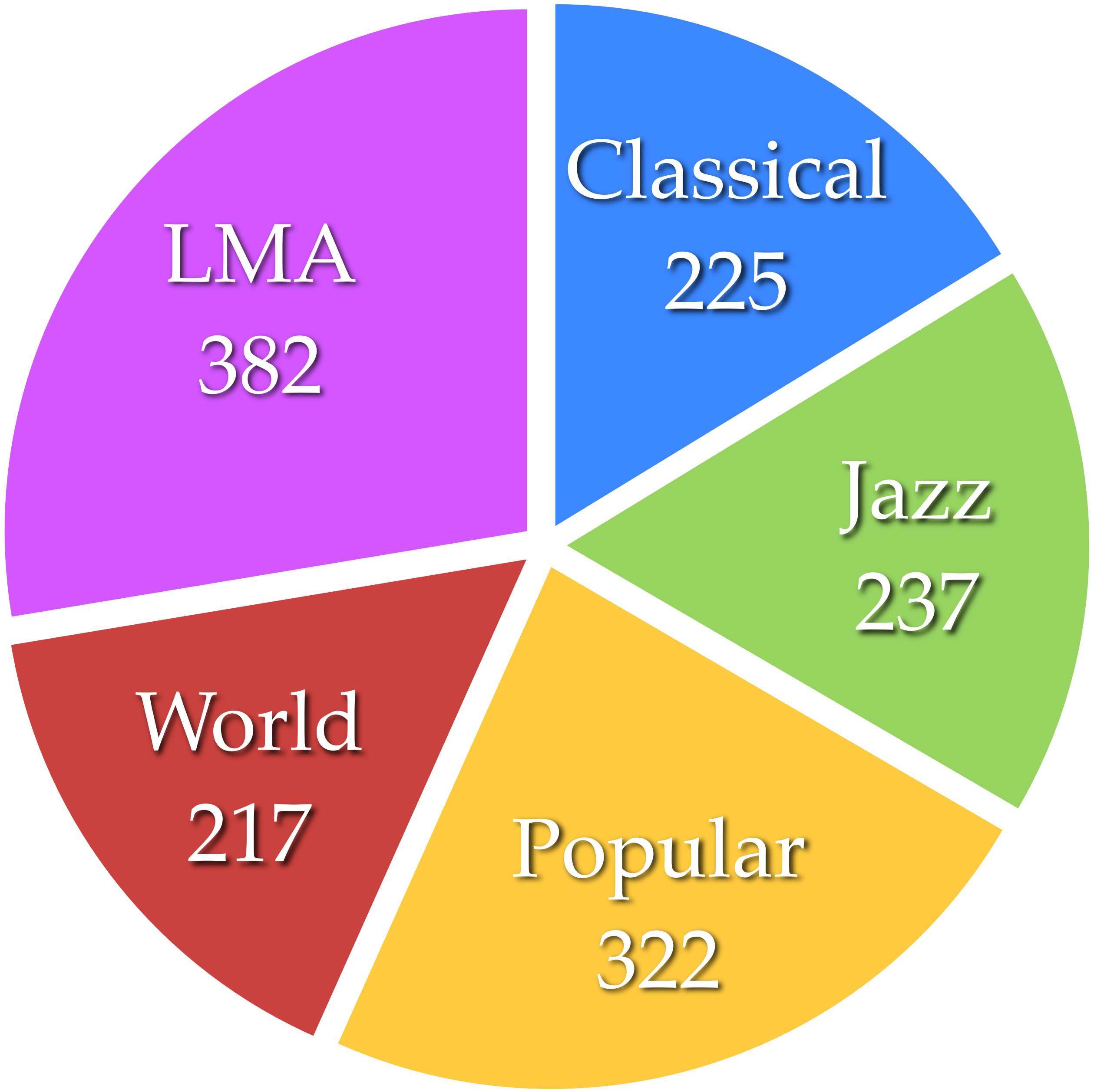
Conseil de recherches en sciences humaines du Canada

Canada

NSF

JISC

SALAMI by genre




?
LMA
382
Renaissance / Medieval
Baroque
Classical
Romantic
20th Century
Classical
225
Jazz
237
Acid Jazz
Avant-Garde
Bebop
Cool Jazz
Contemporary
Blues
Country Blues
Dixieland
Hard Bop
Latin Jazz
Post-Bop
Soul Jazz
Swing
Urban Blues
World
217
African
Americas
Arabic
Asian
Balkan
Calypso
Celtic
Chanson
Cuban
European
Flamenco
Fusion
Gypsy
Indian
Klezmer
Latin American
Mixed
Traditional
Tango
U.S. Traditional
Popular
322
Alternative Pop / Rock
Alternative Metal / Punk
Alternative Folk
Classic Rock
Country
Dance Pop
Electronica
Hip Hop & Rap
Humour
Instrumental Pop
Metal
Reggae
Roots Rock
Singer/Songwriter Folk

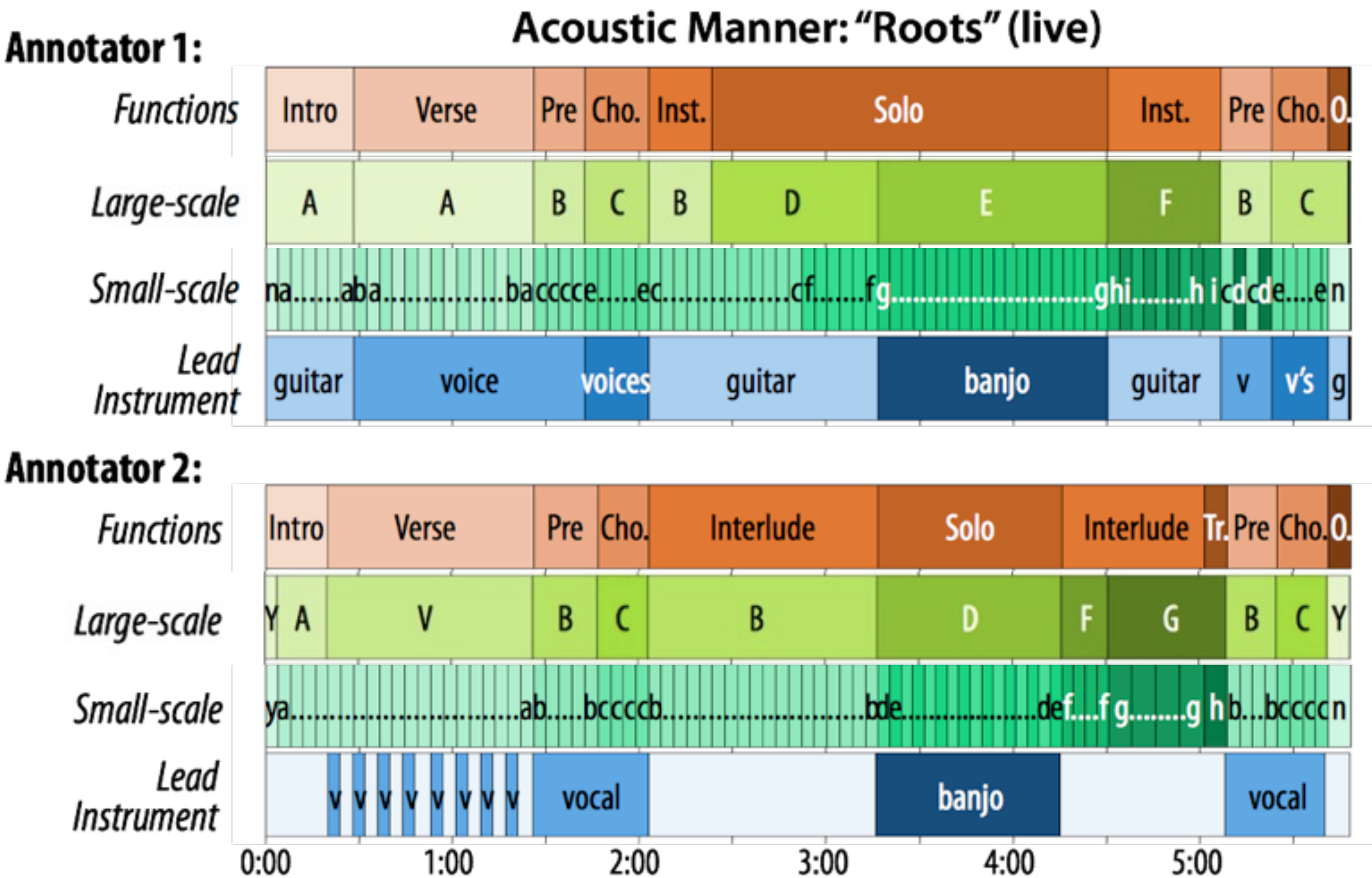
SALAMI

Nutrition Facts

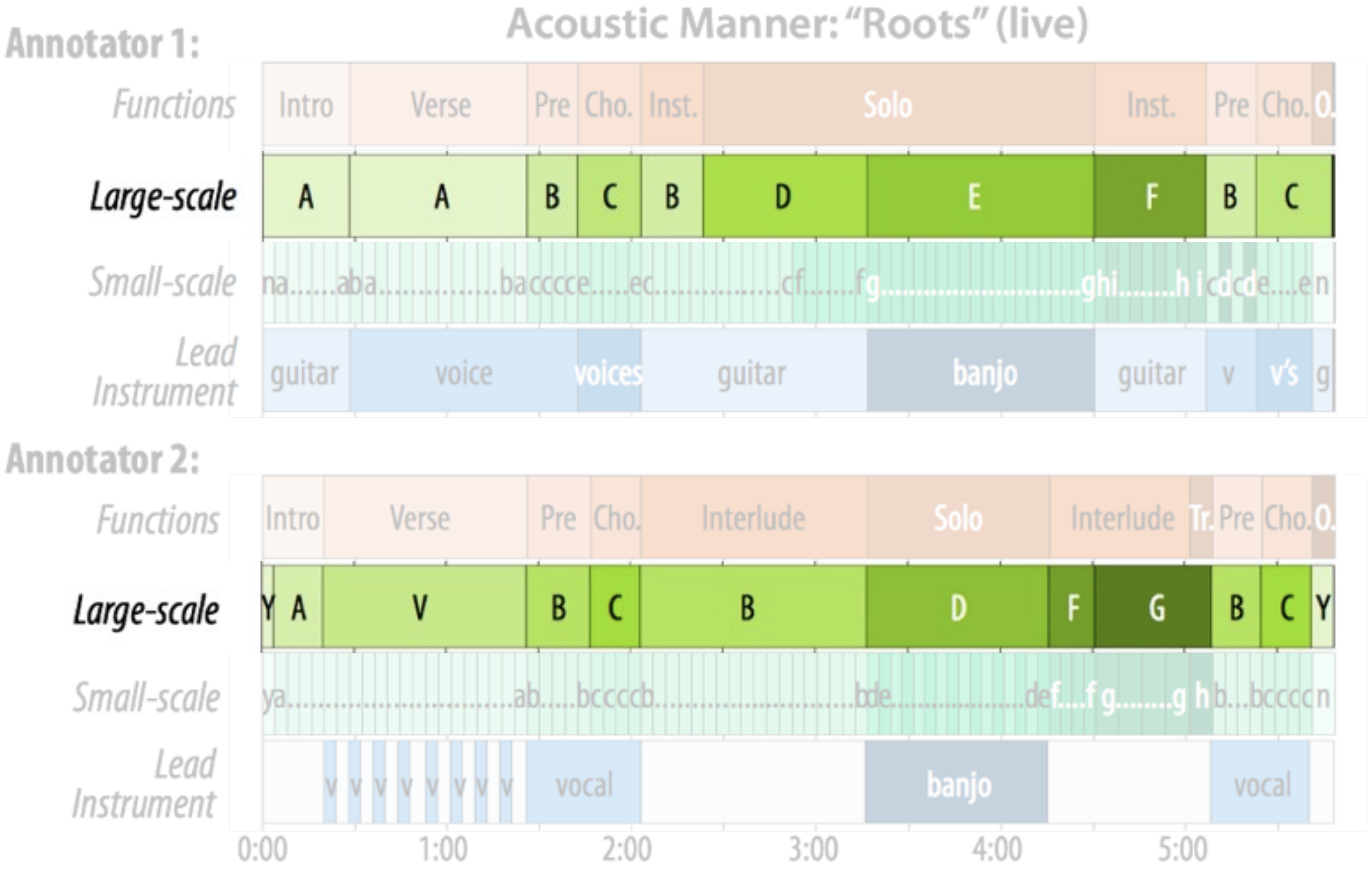
Genre	Number of recordings annotated once	Number of recordings annotated twice
Popular	51	101
Jazz	10	112
Classical	44	65
World	30	78
Live Music Archive (LMA)	113	142
Total:	146	498

Total number of annotations:	1142

Example SALAMI annotations



Example SALAMI annotations



Carte de audio features

timbre:
Mel-frequency cepstral coefficients (MFCCs)

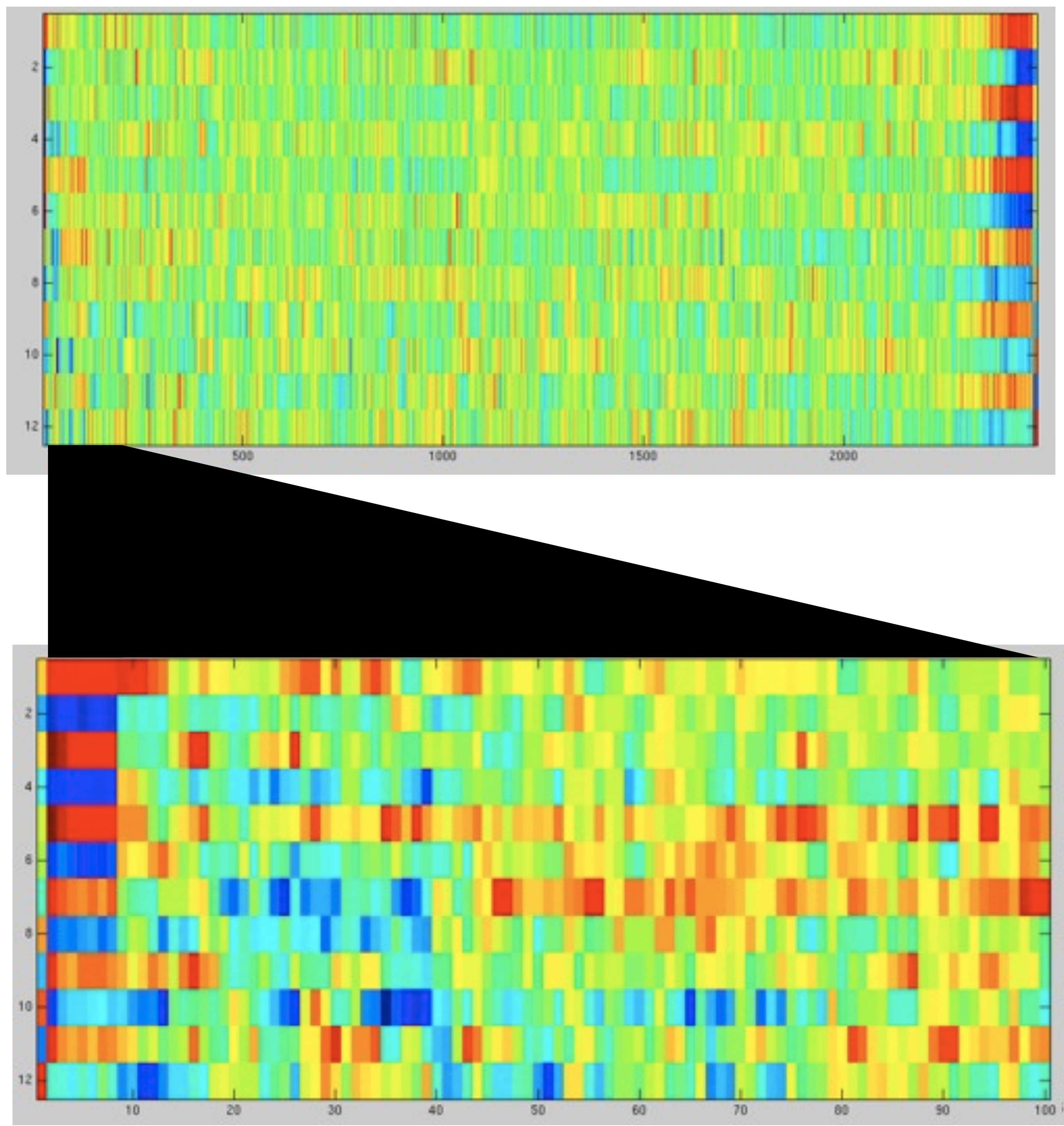
pitch:
chromagram

key:
center of effect (CE)

rhythm:
rhythmogram / fluctuation patterns (FPs)

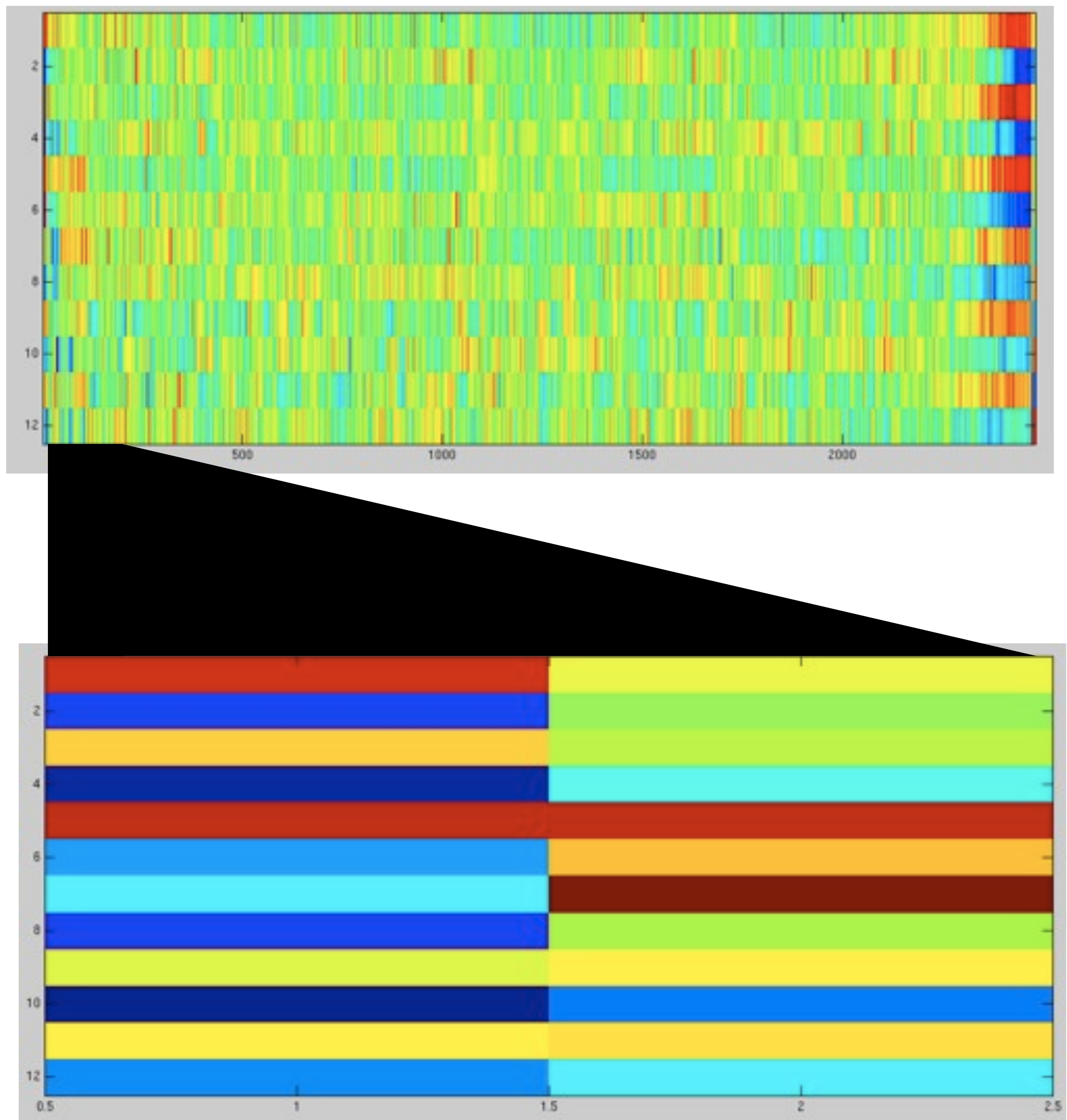
tempo:
periodicity histogram (PH)

From features to novelty functions

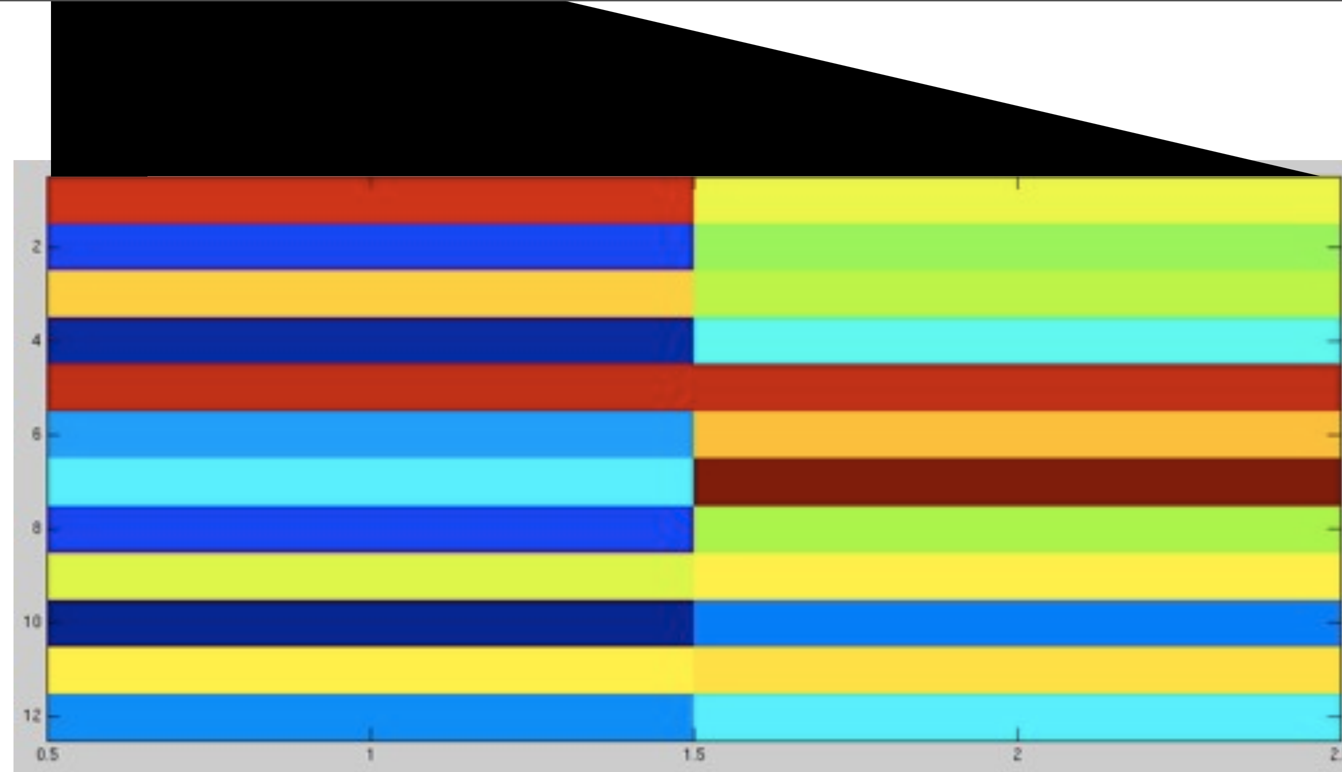


“Across the Universe” by The Beatles

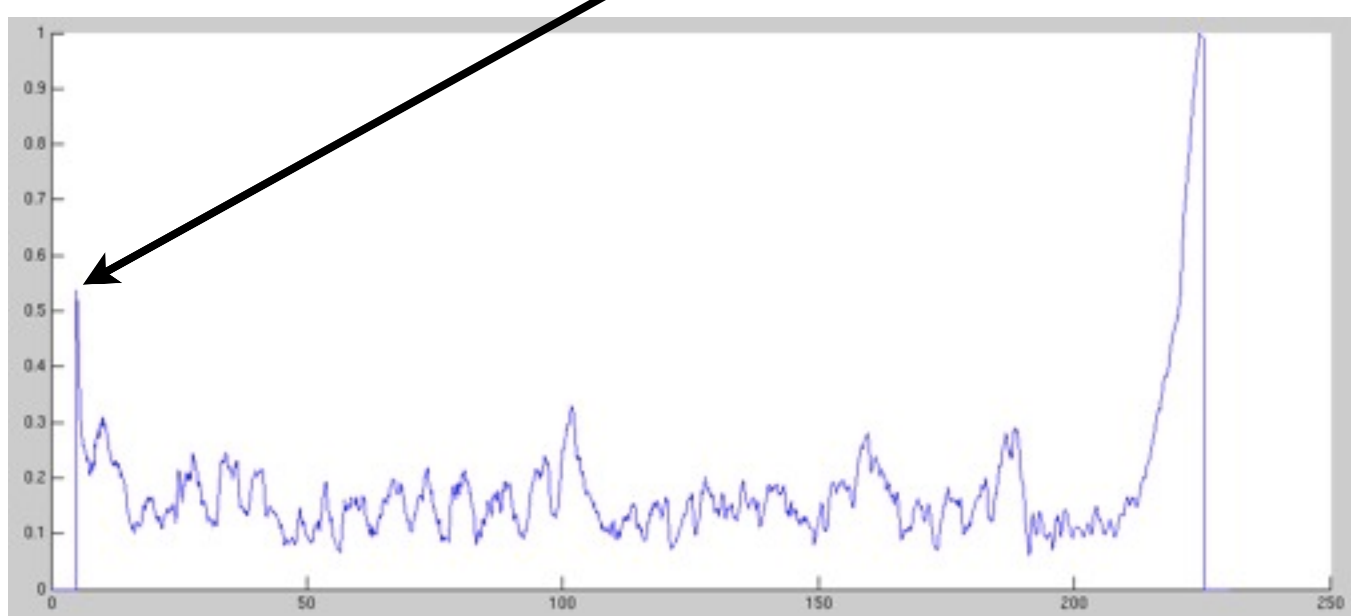
From features to novelty functions



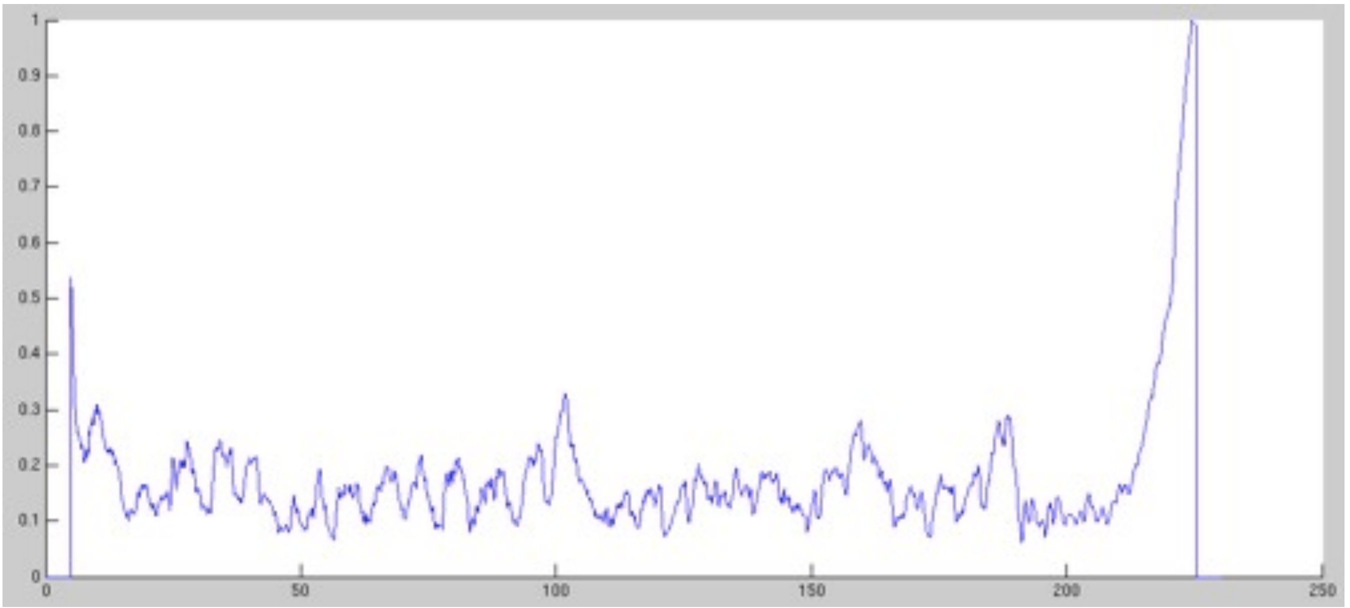
“Across the Universe” by The Beatles



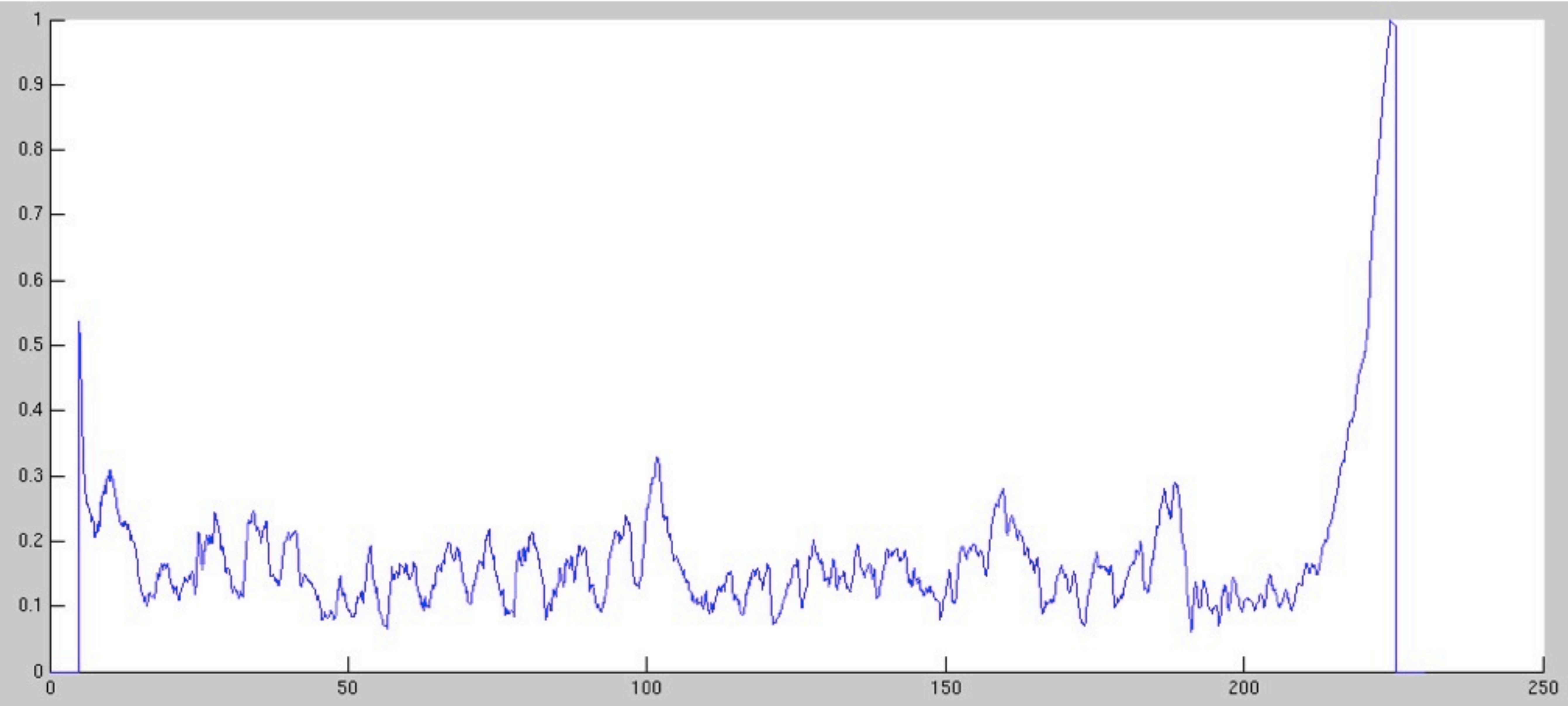
Euclidean distance



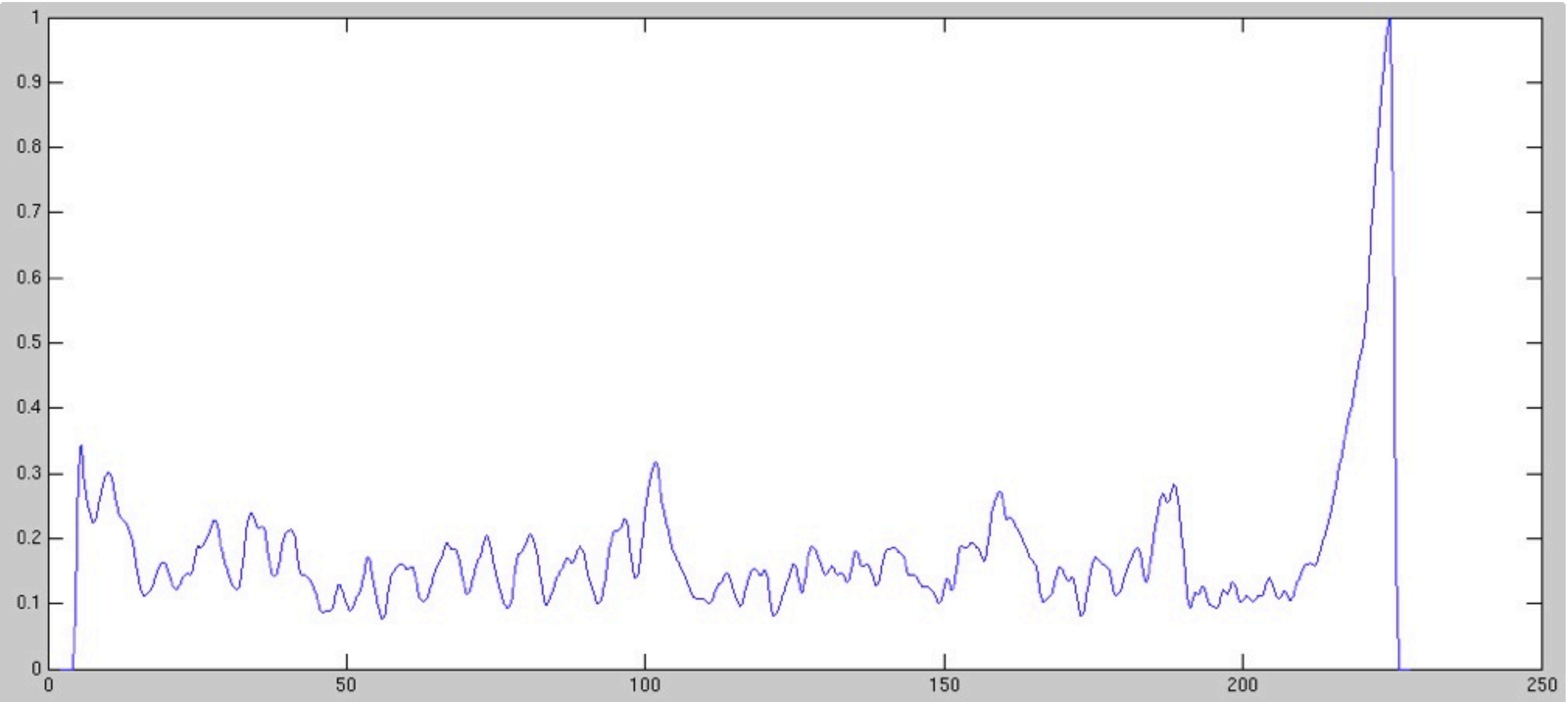
“Across the Universe” by The Beatles



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250

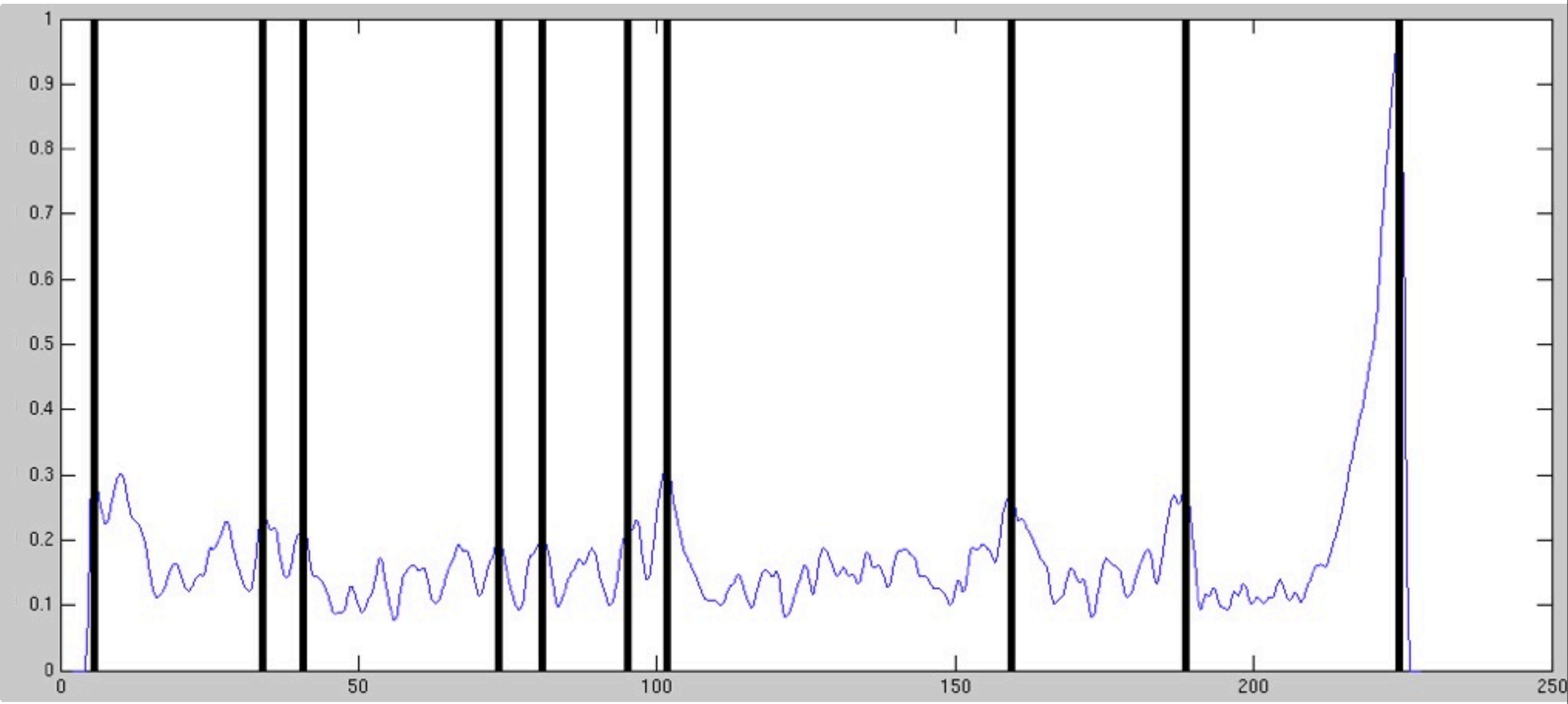


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
50
100
150
200
250

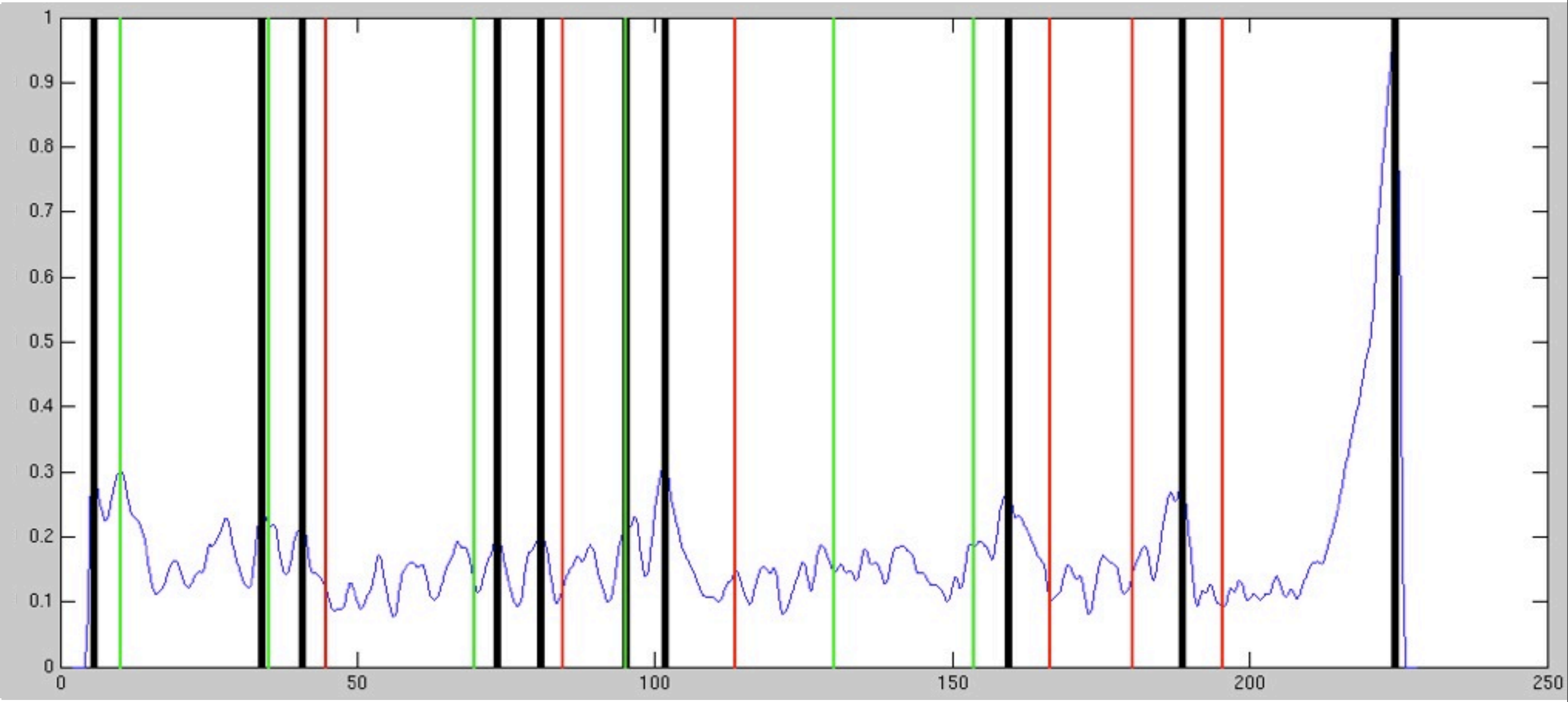
black = point of greatest change



black = point of greatest change

green = perceived as a boundary

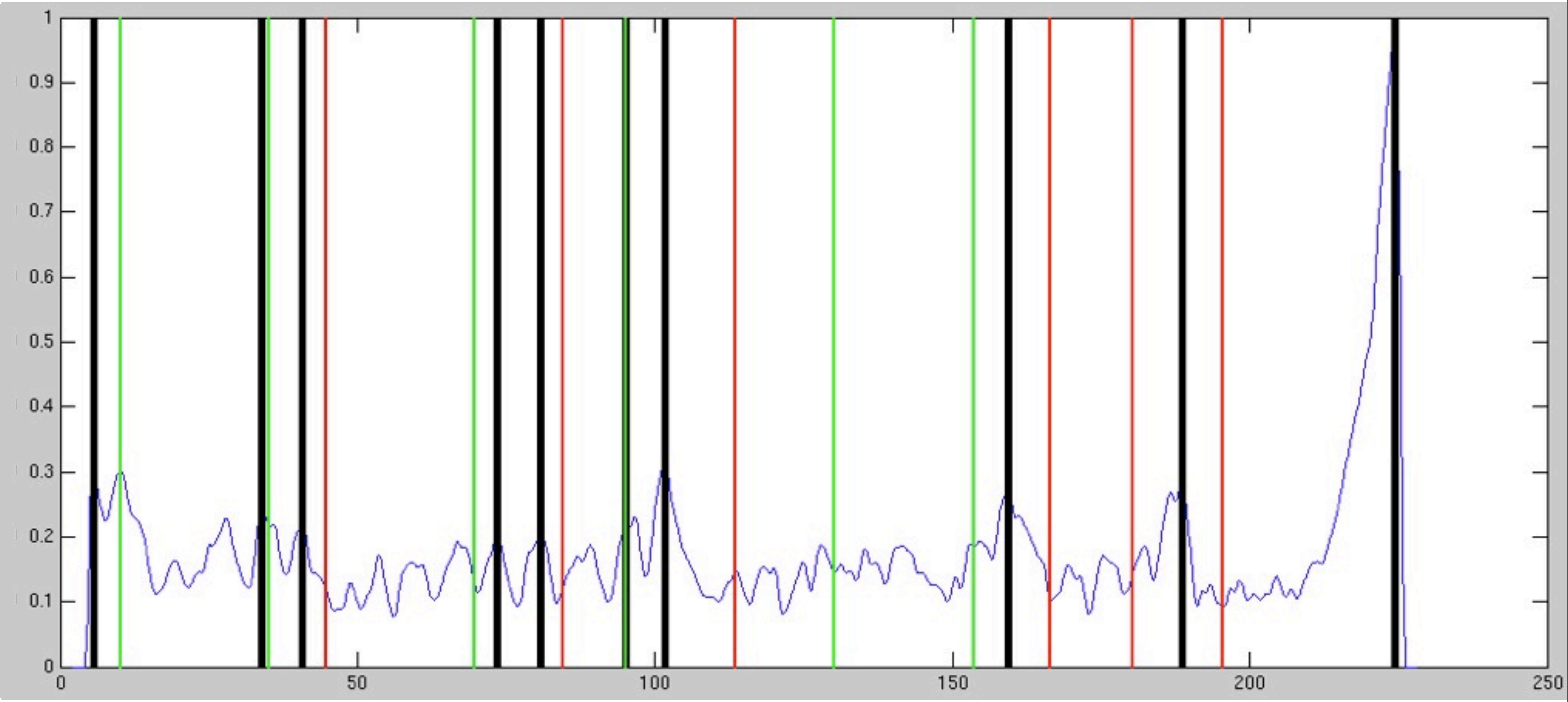
red = random point



black = point of greatest change

green = perceived as a boundary

red = random point

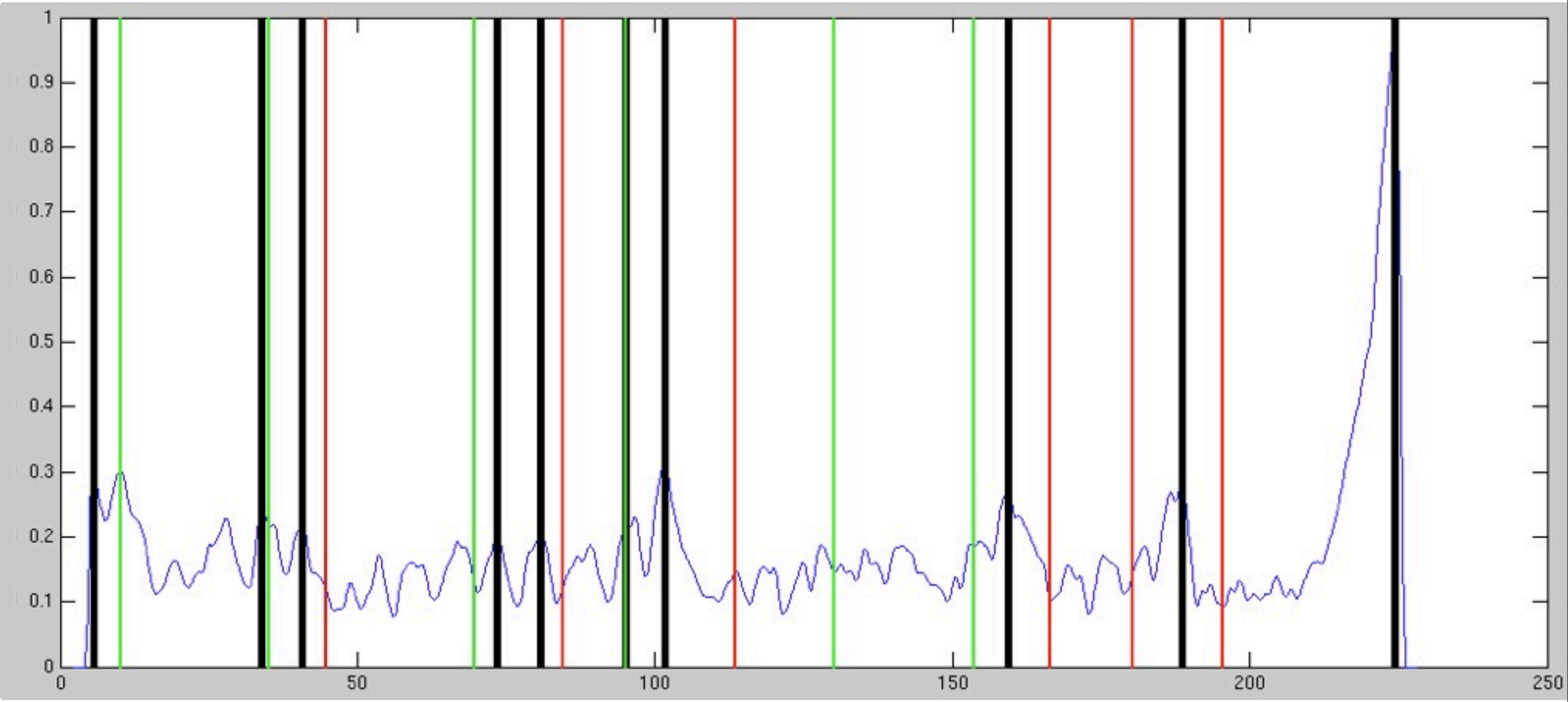


2 / 10 guesses were true boundaries: precision = 0.2
2 / 6 true boundaries were found: recall = 0.33
f-measure = 0.25

black = point of greatest change

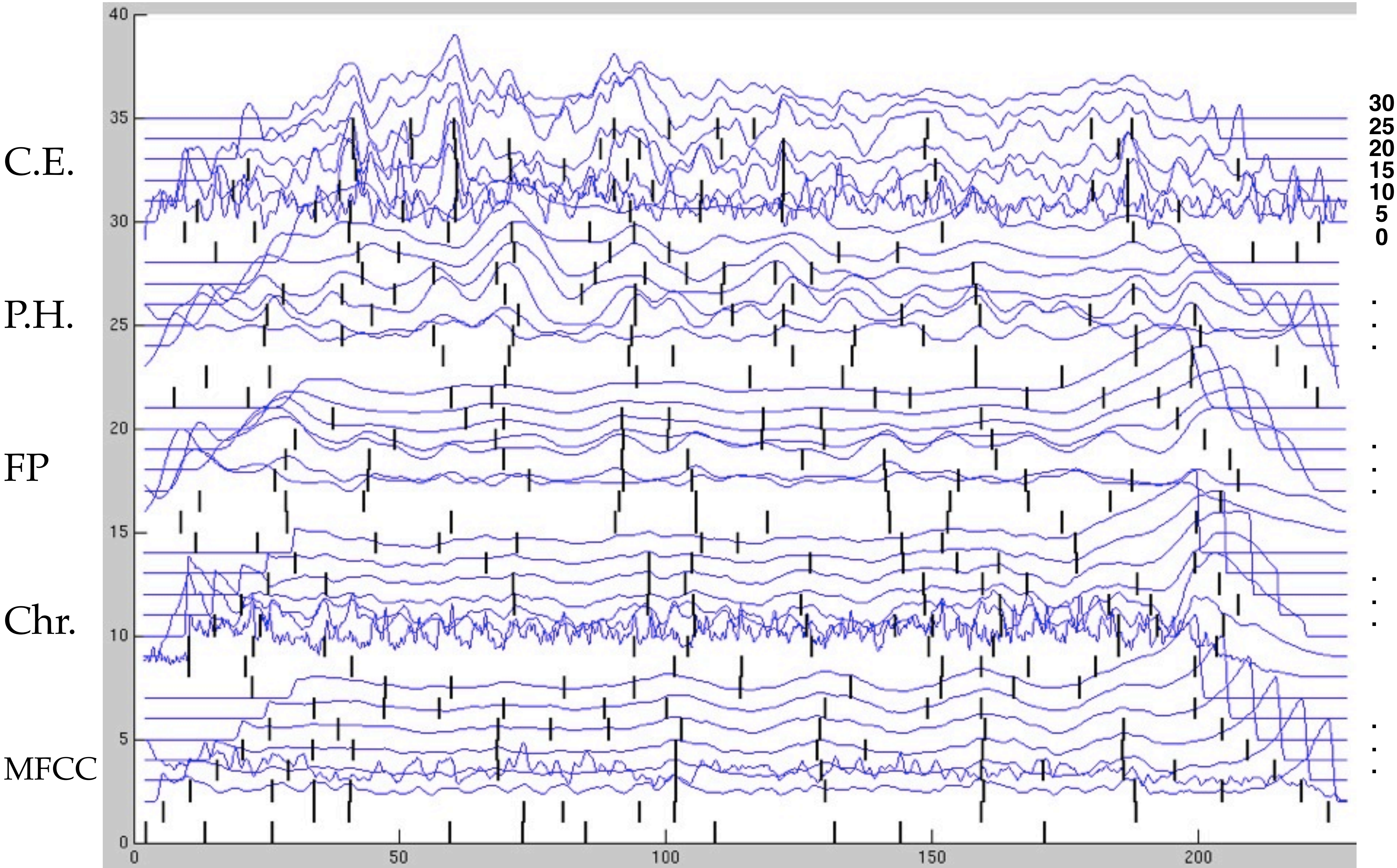
green = perceived as a boundary

red = random point



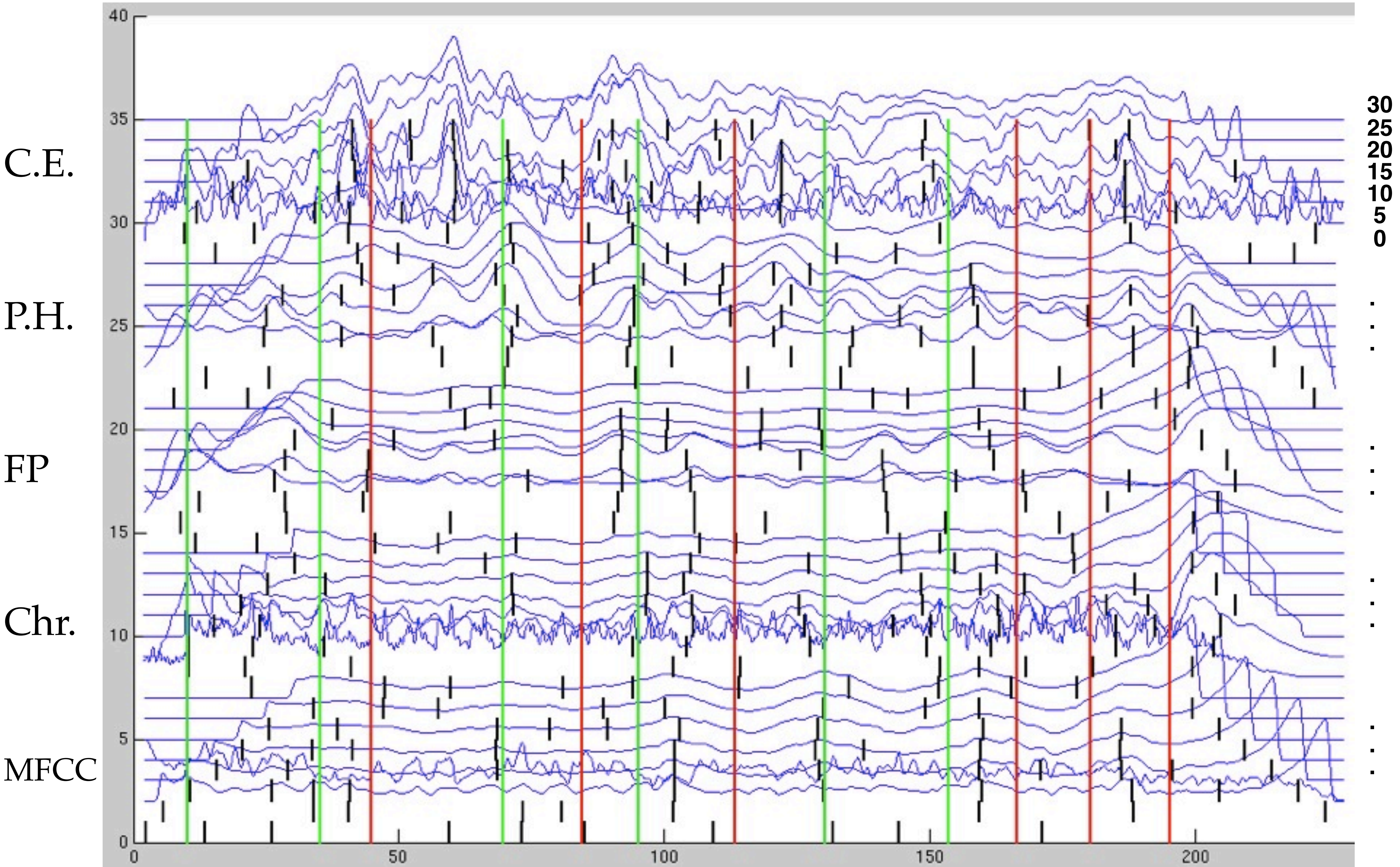
2 / 10 guesses were true boundaries: precision = 0.2
2 / 6 true boundaries were found: recall = 0.33
f-measure = 0.25

0 / 10 guesses matched red
f-measure = 0
***f*-measure contrast = 0.25**



5 different features

7 different timescales



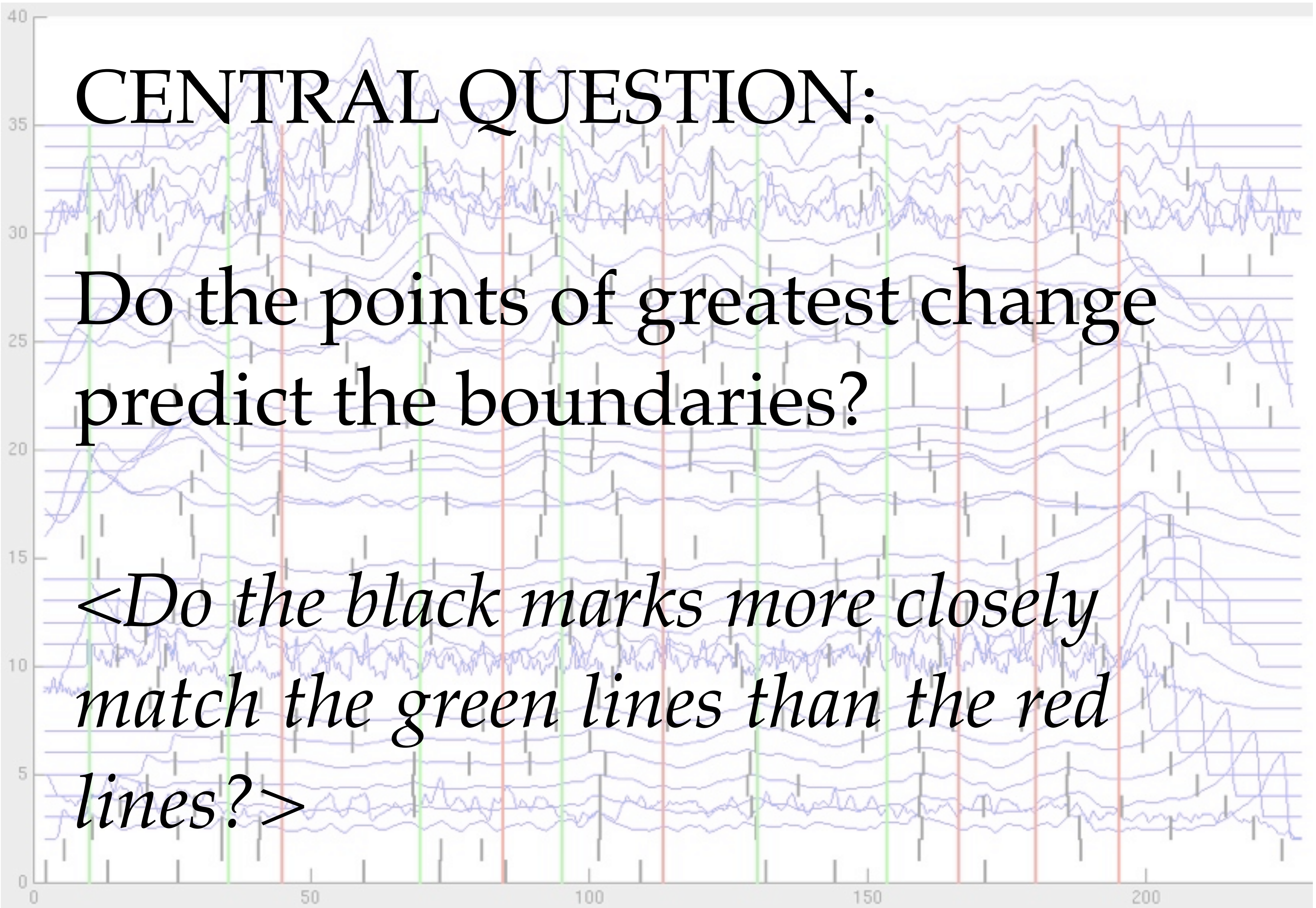
5 different features

7 different timescales

CENTRAL QUESTION:

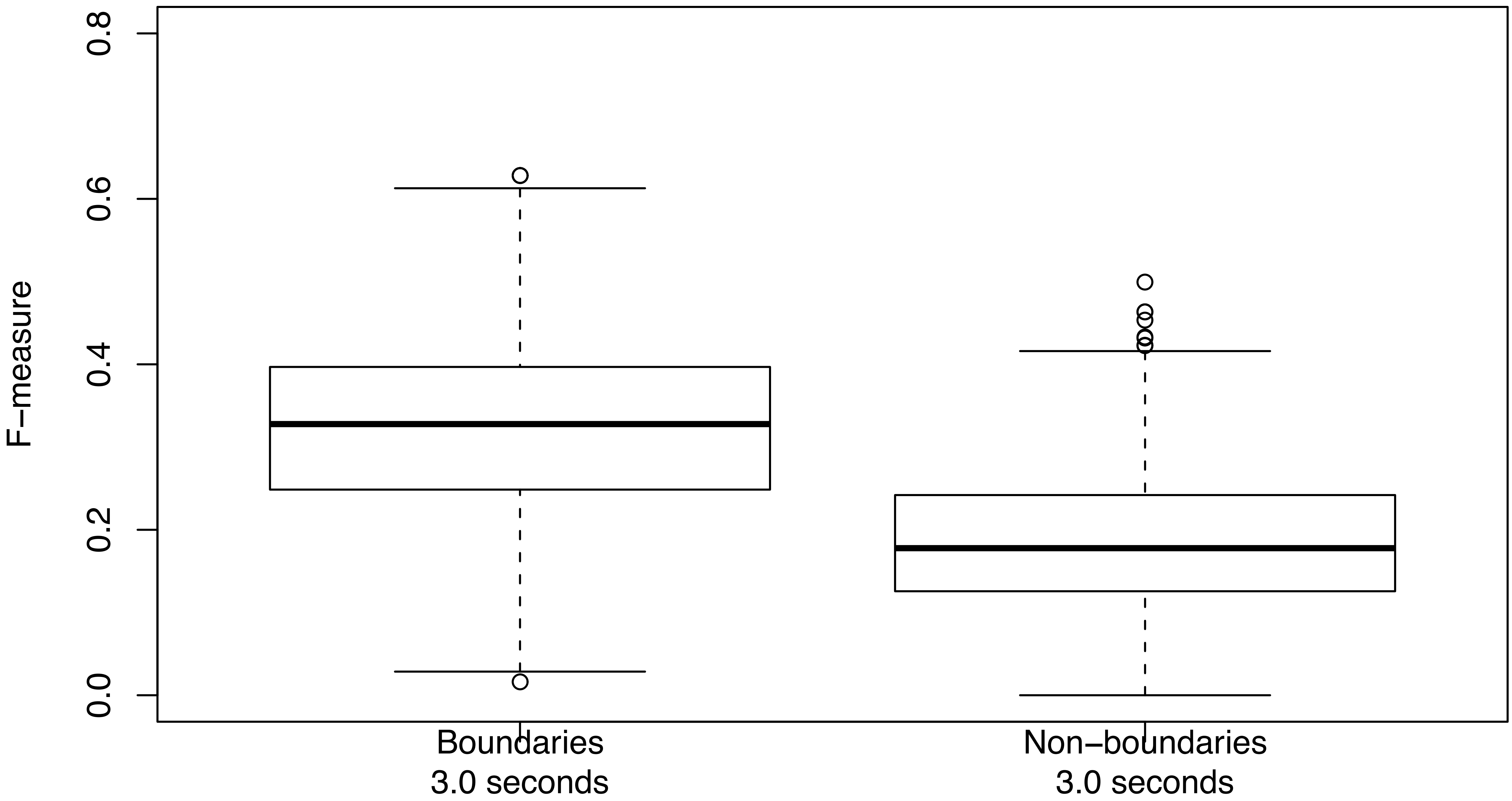
Do the points of greatest change predict the boundaries?

<Do the black marks more closely match the green lines than the red lines?>

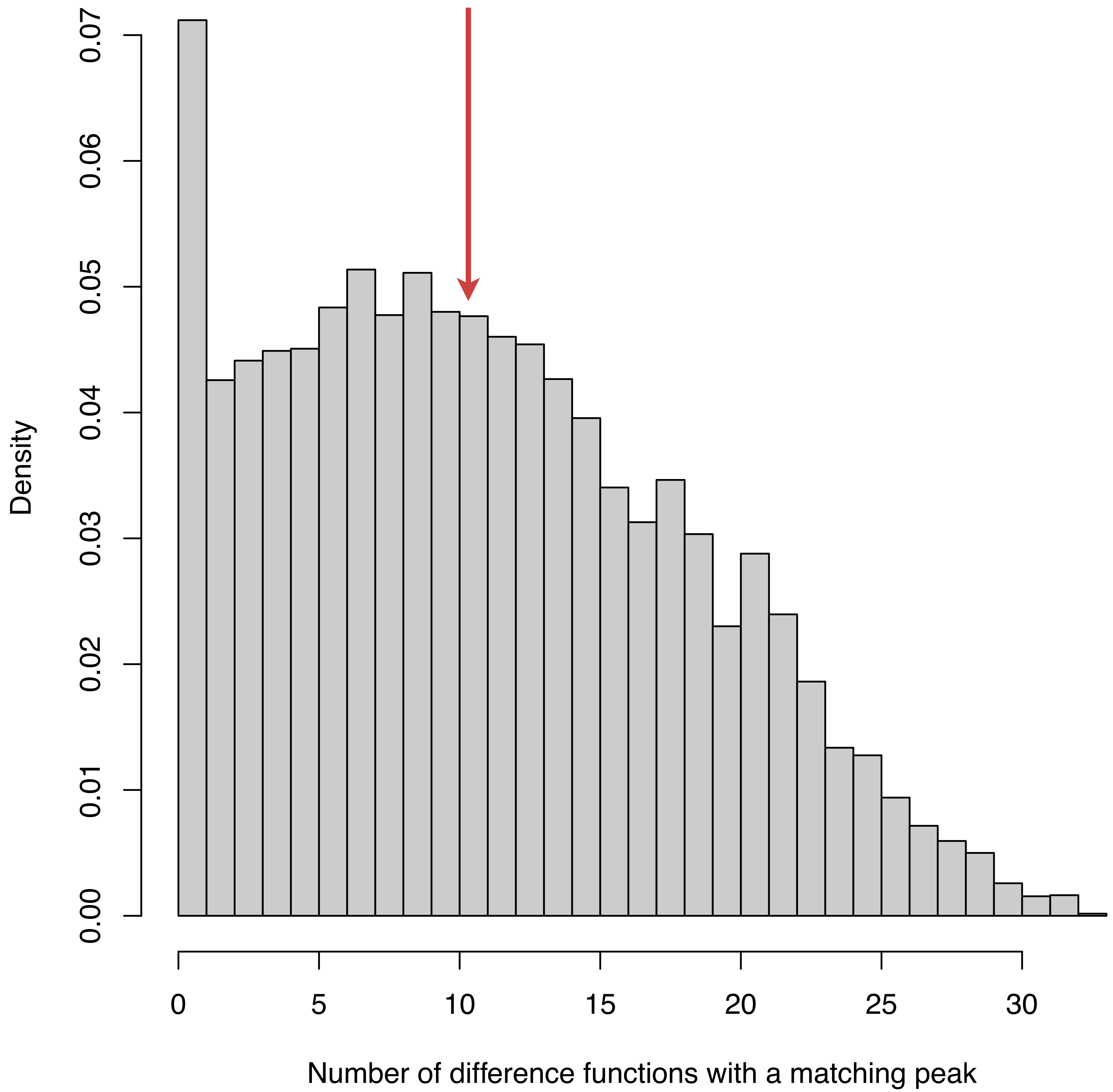


III. What the results of the analysis were.

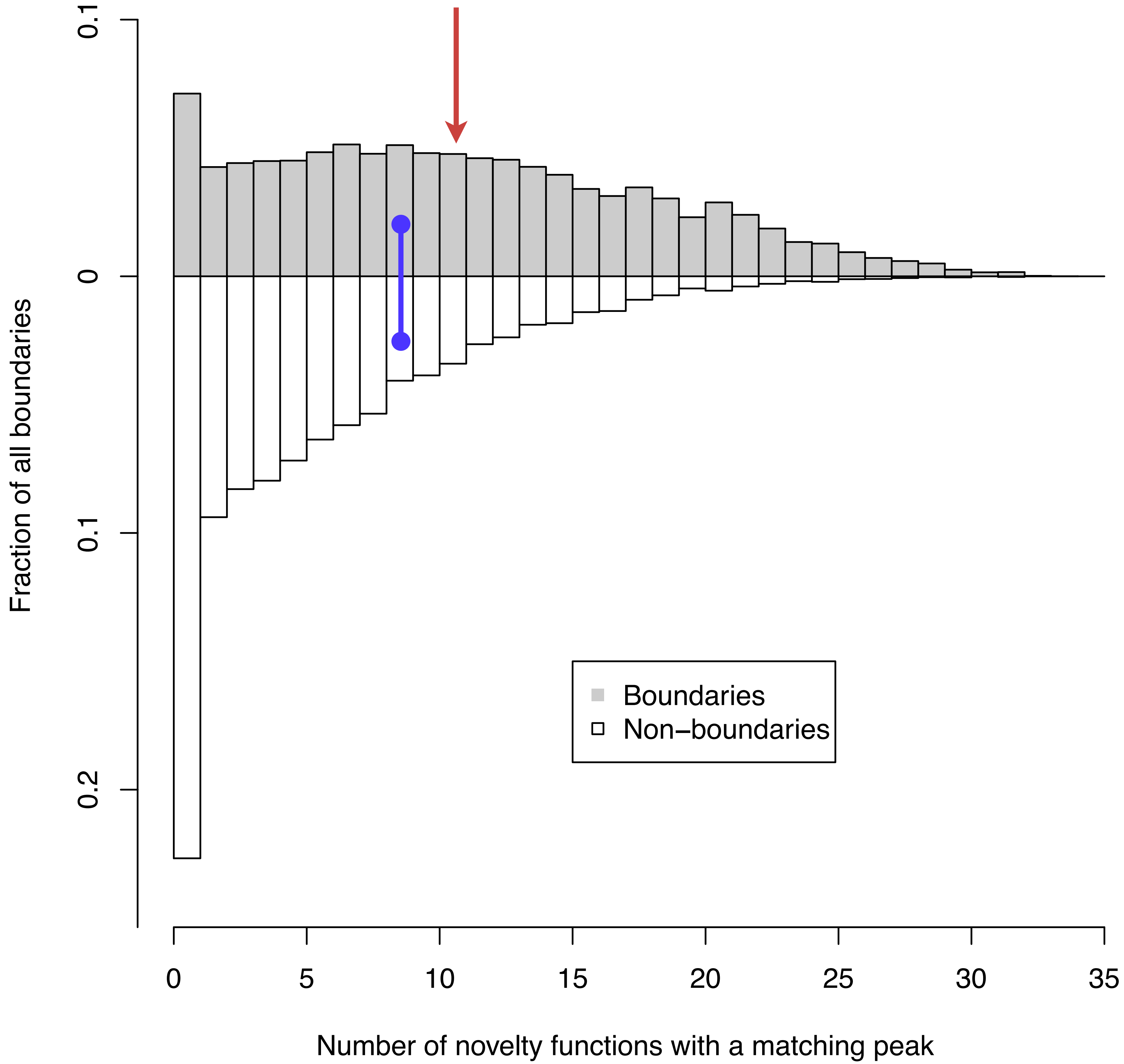
f-measure for boundaries and non-boundaries



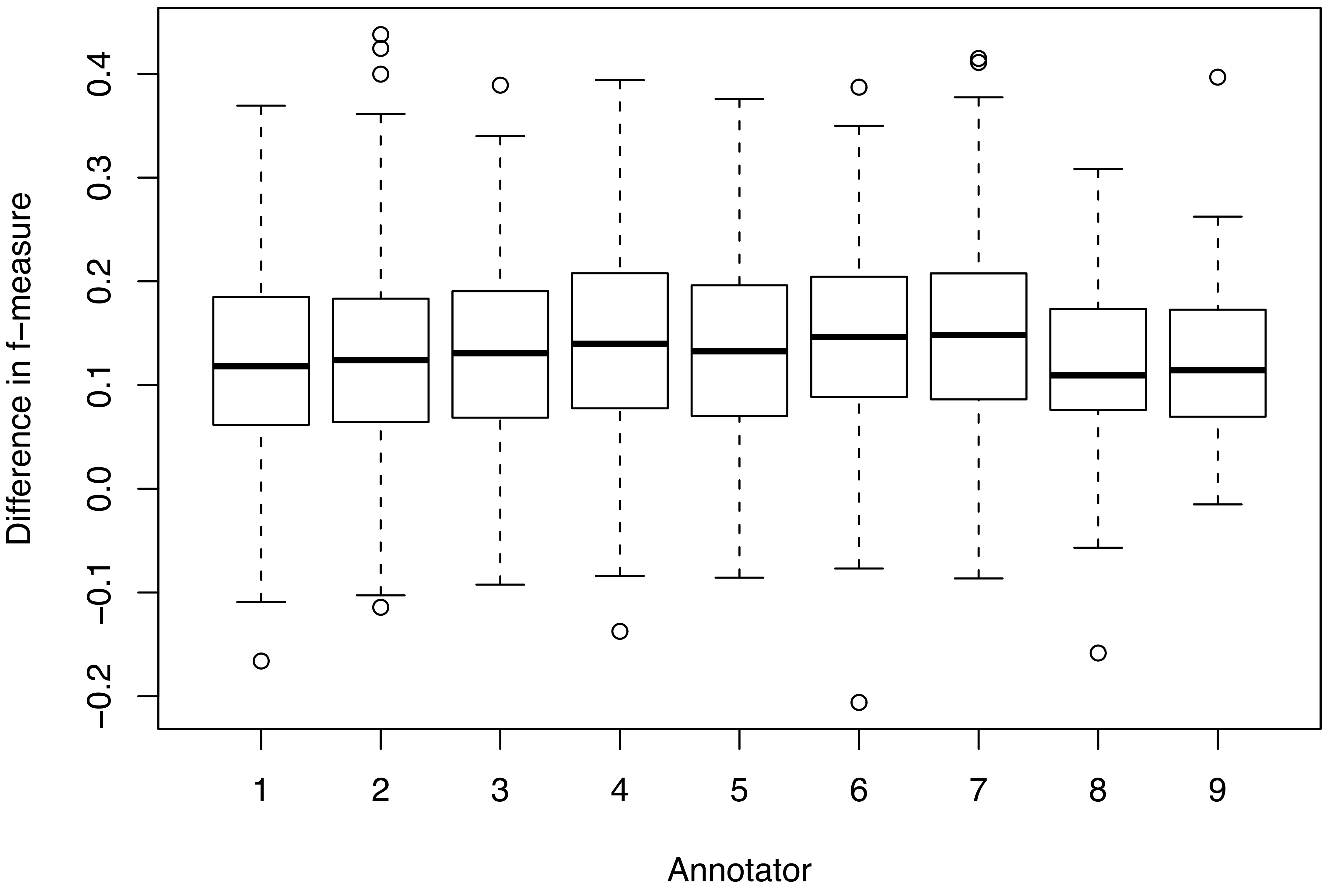
How many changes does each boundary match?



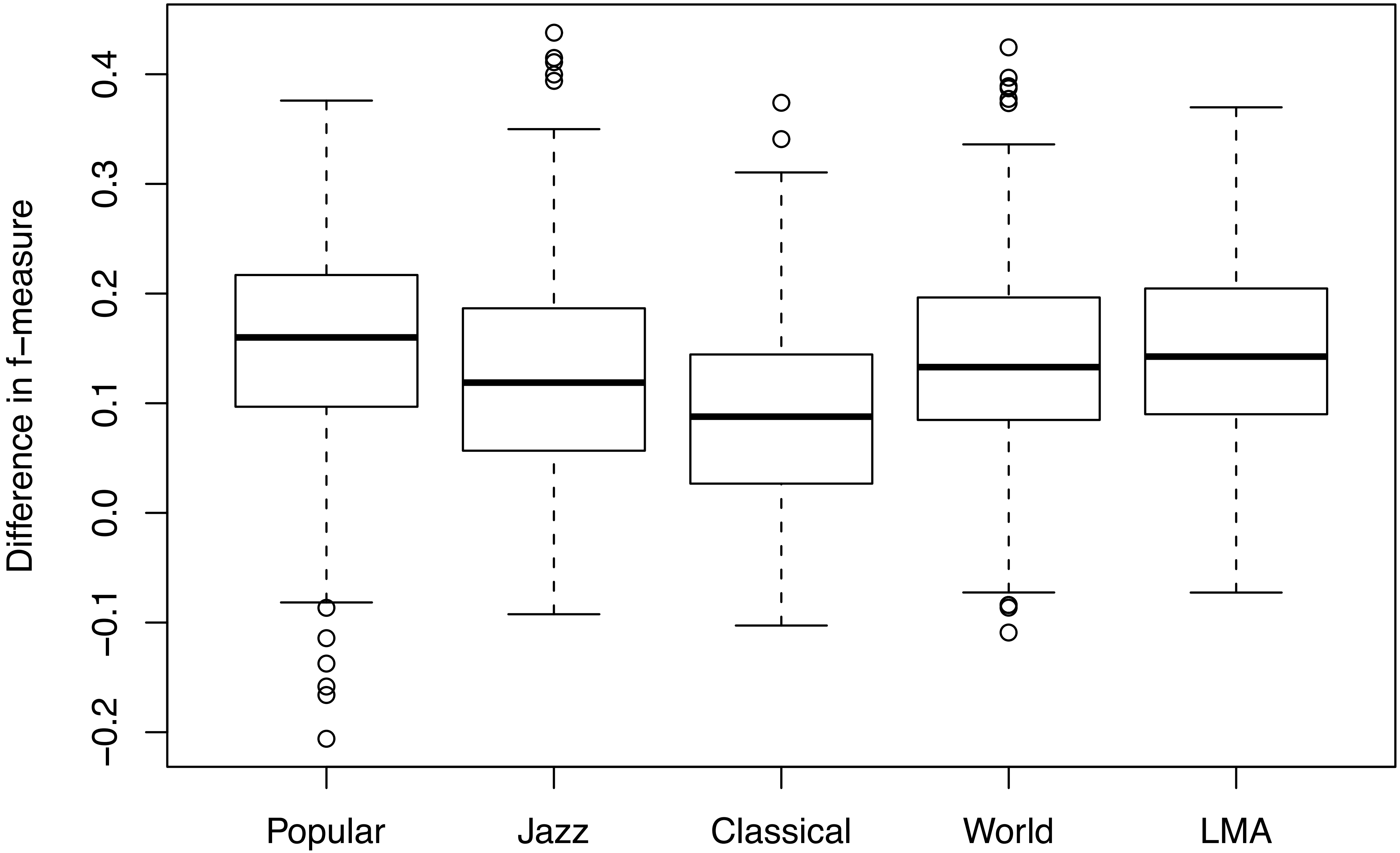
How many changes does each non-boundary match?



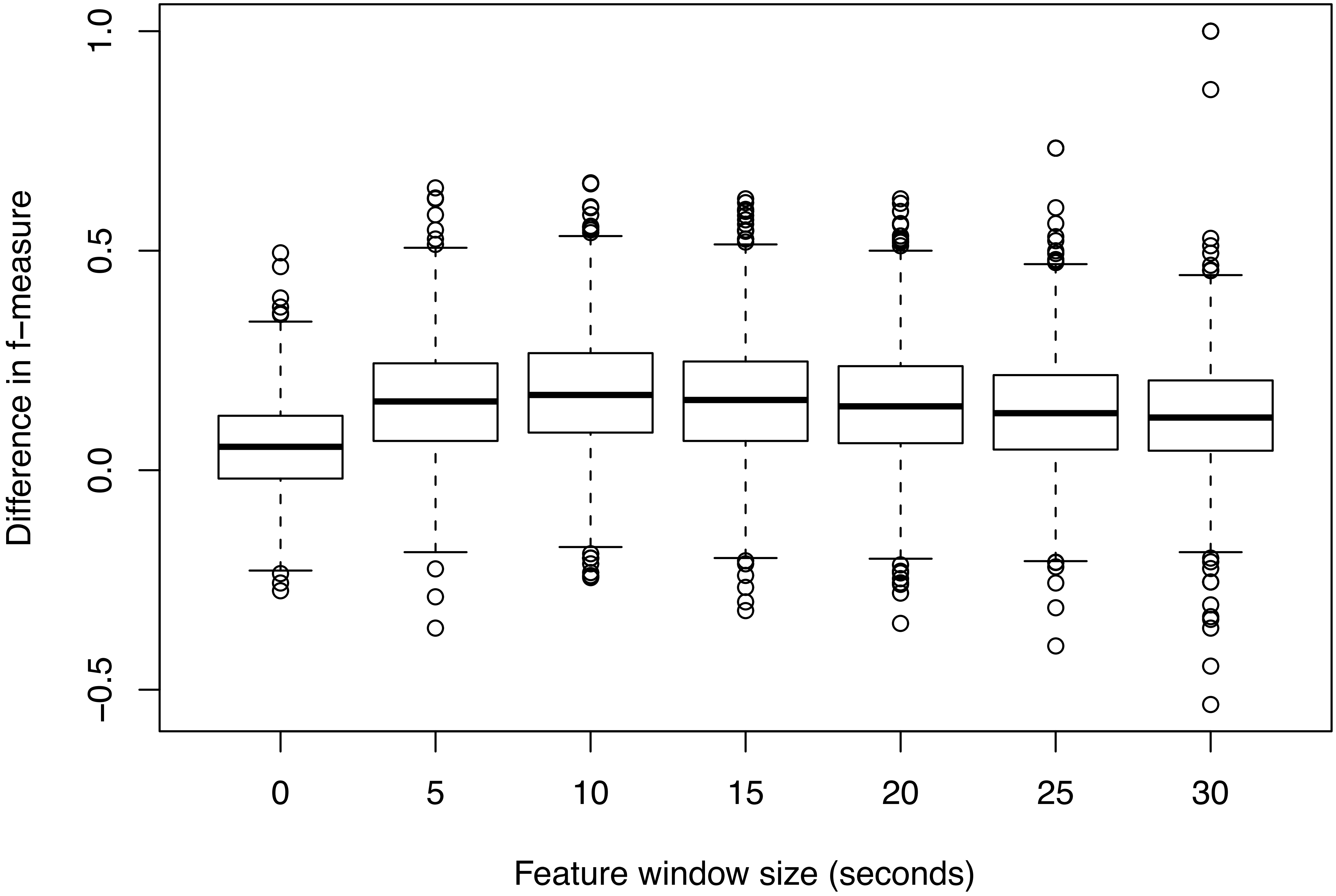
f-measure contrast for different <u>`annotators`</u>



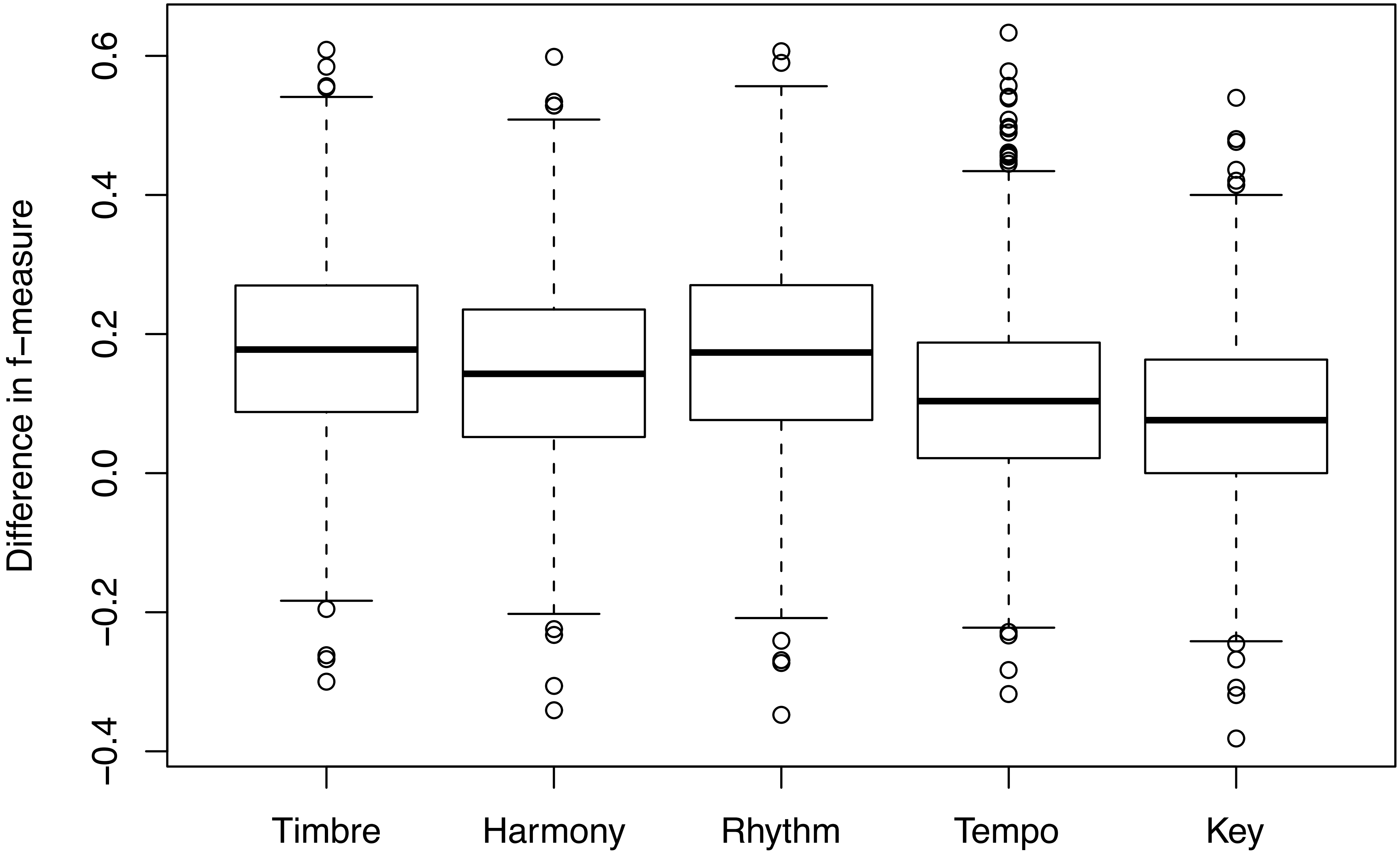
f-measure contrast for different genres



f-measure contrast for different `timescales`



f-measure contrast for different `features`



Conclusions

Large changes in acoustic features ***are*** an indicator of boundaries.

Changes indicate boundaries about ***twice*** as strongly as non-boundaries—but ***only*** twice.

The ***more*** types of change occurring, the ***greater*** the odds of being a boundary.

Being a moment of change seems to be a ***necessary*** but not ***sufficient*** condition for being a boundary.

Wrap-up

We explicitly studied the ground truth by comparing it to a randomized version of itself.

Similar studies examining the role of repetitions and breaks in boundary placement are planned.

Thanks!

This research was supported by the Social Sciences and Humanities Research Council, and by Queen Mary University of London.

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada

Queen Mary
University of London

References

H. Aviezer, Y. Trope, and A. Todorov. "Body cues, not facial expressions, discrimintate between intensive positive and negative emotions." Science*, 30, 2012, pp. 1225–1229.*

M. Bruderer. Perception and modeling of segment boundaries in popular music. Ph.D. dissertation, Technische Universiteit Eindhoven. 2008.

E. F. Clarke, and C. L. Krumhansl, "Perceiving musical time," Music Perception*, 7 (3), 1990, pp. 213–251.*

I. Cross, "Music analysis and music perception," Music Analysis*, 17 (10), 1998. [image credit]*

J. B. L. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and S. J. Downie, "Design and creation of a large-scale database of structural annotations," in Proc. ISMIR*, Miami, FL, 2011, pp. 555–560.*

More references for this research not explicitly involved in this presentation can be found in J. B. L. Smith, C.-H. Chuan, E. Chew. "Audio properties of perceived boundaries in music," submitted to IEEE Trans. Multimedia, which you can get a copy of if you email me or something.