Music Information Geometry

Arshia Cont

Researcher, Realtime Musical Interactions Team,

Ircam - Centre Pompidou, Paris, France.

cont@ircam.fr

Collaborators: Shlomo Dubnov (UCSD) and Gérard Assayag (IRCAM)

ircam
Centre
Pompidou

ircam
Centre
Pompidou

Motivations

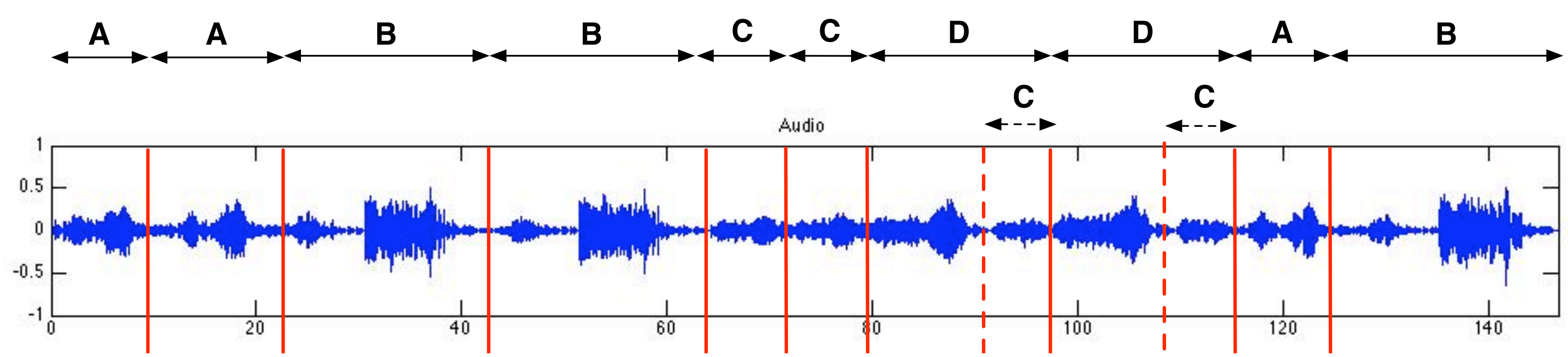
- The gap between ***symbolic*** and ***signal*** aspects of music
  - for composers
  - for musicologists
  - for researchers/engineers
- Artificial Intelligence / Complexity
  - Rodney A. Brook keynote speech during AAAI 1996 Conference: "Despite research advances in speech processing and recognition, our knowledge has much less progressed in artificial perception of audio outside the speech domain. This presents one of the most important challenges of artificial intelligence."

Motivations

- “Effortless” abilities among music listeners and/or trained musicians that still pose challenging problems for machine intelligence:
  - Real-time cognition of recurrent musical structures:
    - Example from Beethoven’s first piano sonata, 3rd movement. (performed by Friedrich Gulda in 1950s)
    - Symbolic domain... Easy analysis: Four recurring structures

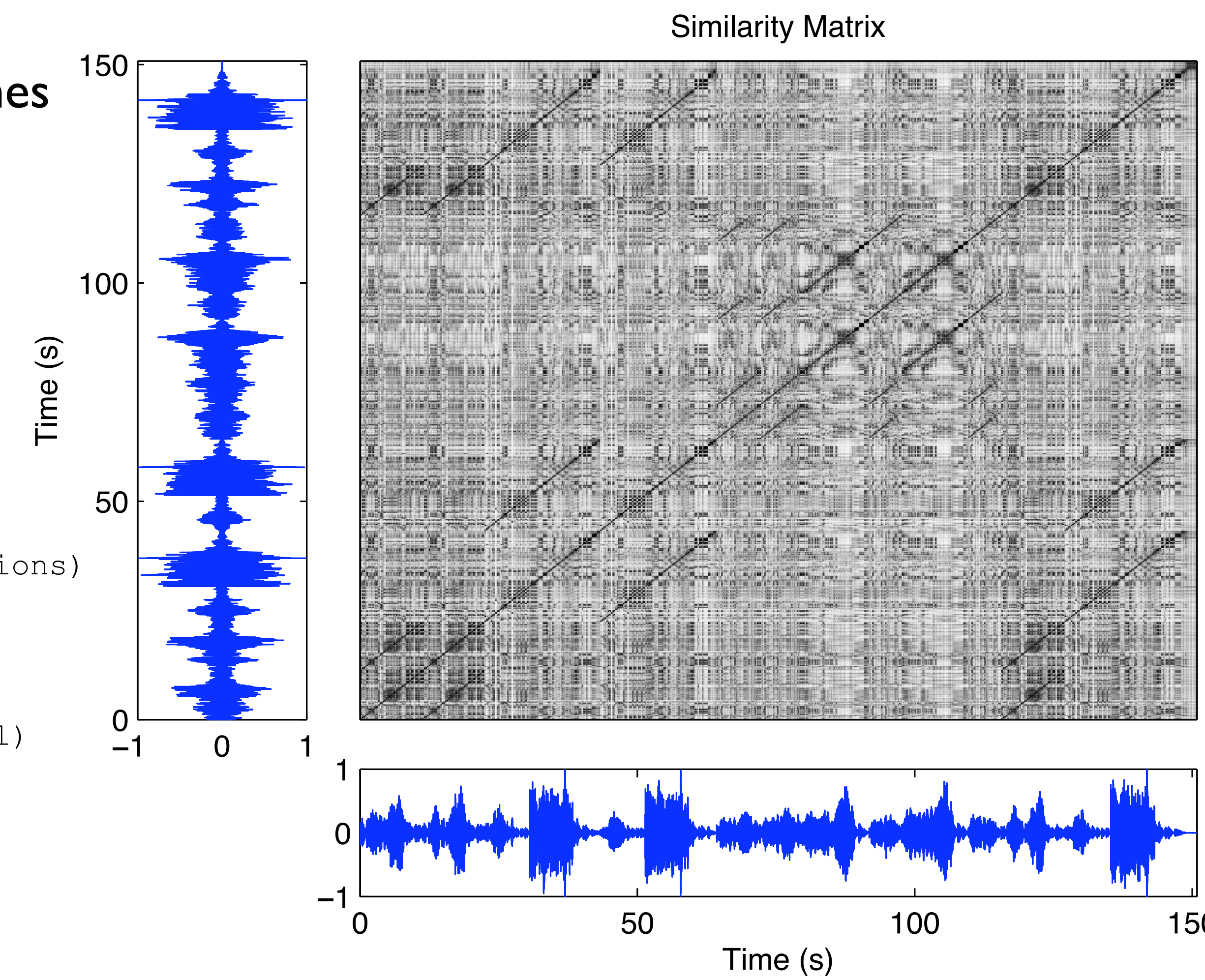
Motivations

- "Effortless" abilities among music listeners and/or trained musicians that still pose challenging problems for machine intelligence:
  - Real-time cognition of recurrent musical structures:
    - Example from Beethoven's first piano sonata, 3rd movement.
    - Analysis in the signal domain? Not evident... .
      - Specially with no a priori knowledge of music
      - Worse in real time!



Motivations

- Information Retrieval
  - Central Question: *What information is present in the signal and to what relevancy?*
  - Current trends: *Self-Similarity Computations in the field of Music Information Retrieval*
    - Compute similarity between *every* possible combination of analysis frames
    - Non-realtime
    - Far from being "effortless"
    - Requires further processing
  - "Philosophical" drawbacks:
    - Brains do not store sounds but relations
      (D. Huron, Psychology of Musical Expectations)
    - Complex problems do not require complex solutions
      (H.A. Simon, The Science of the Artificial)



Motivations

- Information Theory
  - Classical IT has few answers! Signals have entropy whether they have relevant information or not.... .
  - Worse when it comes to the complexity of musical patterns (Pressing 1999)
  - Rate distortion theories based on relative entropy. Problems: Non-stationarity of music signals and strong temporality of musical structures
  - *Information Rate* (Dubnov '08) has proven to address both issues in limited context
  - Necessity for similarity spaces to obtain perceptual equivalent classes
- Advances in Machine Learning
  - Many approaches and applications: Max. Entropy algorithms, Clustering, Probabilistic Systems and exponential distributions
  - Mostly based Self-similarity methods and *analysis* of information content using a priori information on the content itself and also on the structure
  - Necessity for *metric spaces* to obtain equivalent classes
- Goal: To make such transitions possible and more...
  - A general framework to fill in the following gap for musical applications:

    signal ↔ symbol

  - Towards a *metric similarity space* on the signal domain

ircam Centre Pompidou

Motivations

<u>Role of *expectations* in musical experience</u>

- In listening experience
  - Expectations imply mental representations in which our daily musical experience is being examined and updated.
  - Major responsibility for musical emotions
- In musical creativity
  - Meyer (1954): composition = choreography of musical expectations
  - Huron (2006): Demonstrates explicit cases of these "choreographies"
  - Grisey (1987): "A composer's reflections on musical time"
    - *The skin of time*
    - *From the time of music to the music of time...*
- No major consideration for *expectation* in music computation

"I think that the search for a *universal* answer to the questions raised by musical experience will never be completely fulfilled; but we know that a question raised is often more significant than the answer received. Only a reckless spirit, today, would try to give a total explanation of music, but anyone who would never pose the problem is even more reckless."

Remembering the future
Luciano Berio

ircam Centre Pompidou

Information Geometry

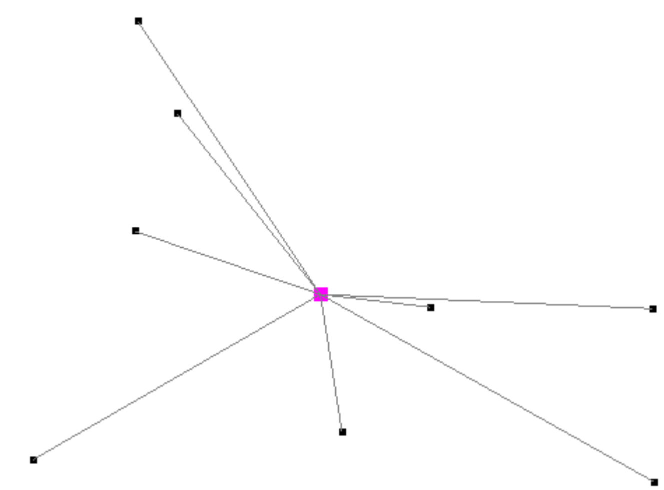
Intuition

- Consider the following geometry:
  - *Points* are probability distributions $p(x, \xi)$ (instead of *dots*)
  - *Distance* between two points is some measure of *information* between them
- Welcome to the world of *Information Geometry!*
  - Geometric manifolds with *information metrics* on *probability space*
    - Marriage of *Differential Geometry*, *Information Theory*, and *Machine Learning*
  - Considering probabilistic representations as well-behaved geometrical objects, with intuitive geometric properties
    - Spheres, lines (geodesics), rotations, volumes, lengths, angles, etc.
- Getting real...
  - Riemannian Manifolds over probability spaces with *Fisher Information* measure
  - Characterized by the type of employed *distance* (called *divergences*)
  - Our interest, canonical elements:
    - Space of *exponential distributions*
    - with *Bregman divergences*
      - Bijection between the two

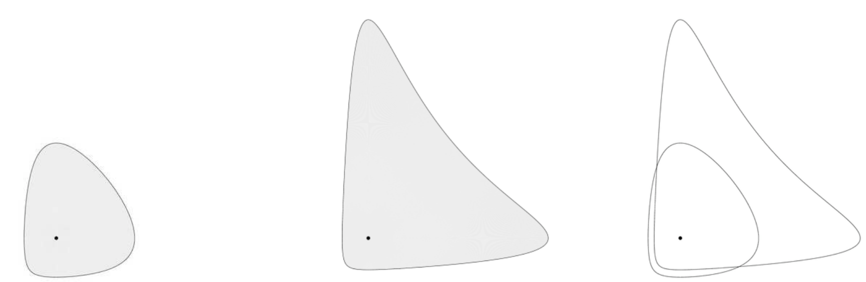
Information Geometry

<u>*Elements of Bregman Geometry*</u>

- *Bregman Centroids*
  - Significant property
    - The "right type" centroid is independent of the choice of Bregman divergence and is equal to the mean:



- *Bregman Balls*
  - In analogy to Euclidean geometry, we can define balls using Bregman divs, centered at μ_k with radius R_k



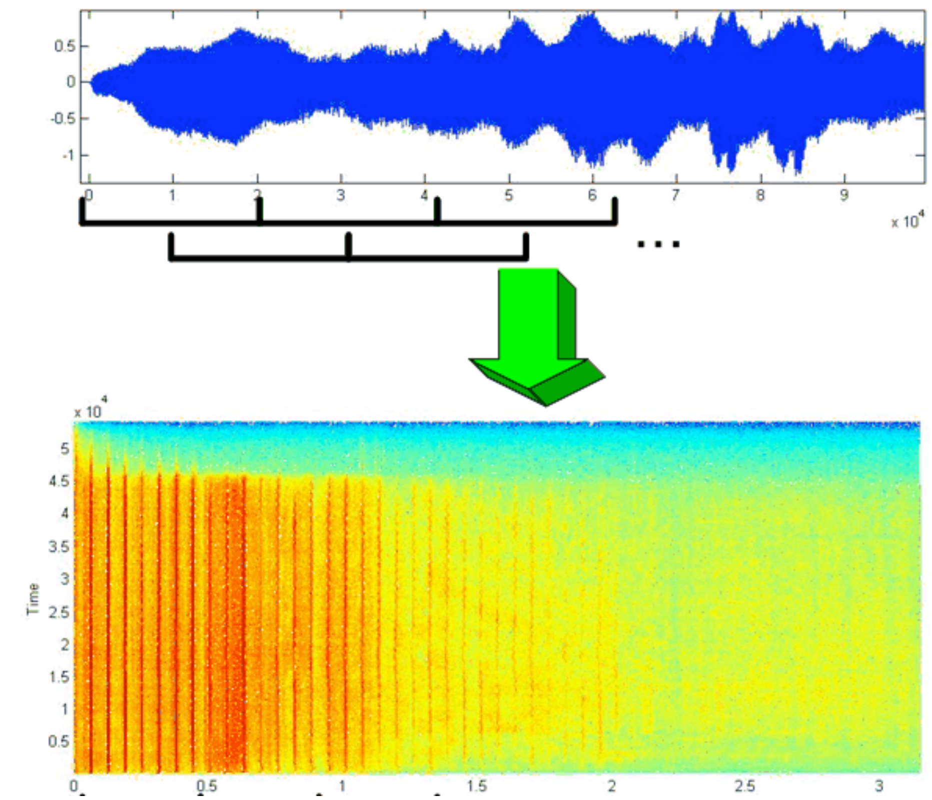
(a) (b) (c)

- *Bregman Information* of a random variable *X*
  - Defined as the expectation over divergences of all *points* from the centroid
  - Special cases: *variance, mutual information*

Music Information Geometry

<u>*One Possible Music Information Geometry*</u>

- Points = time domain windows of audio signal X_t, represented by their *frequency distributions* $S_t(\omega)$
  - Arriving incrementally / in real time
  - Corresponding to normalized log-scale Fourier transform amplitudes
  - Mapped to *Multinomial points* in the information geometry (one-to-one)
    - Corresponding Bregman divergence is *Kullback-Leibler divergence*
    - Therefore, Bregman Information is equivalent to *mutual information*
- Can be extended to other frameworks



Music Information Geometry

Quantifying and Qualifying Relevant Information

- Do *not* formalize information content!
- Control *changes* of information content instead
  - Using some *metric d*, that gives rise to the notion of similarity:

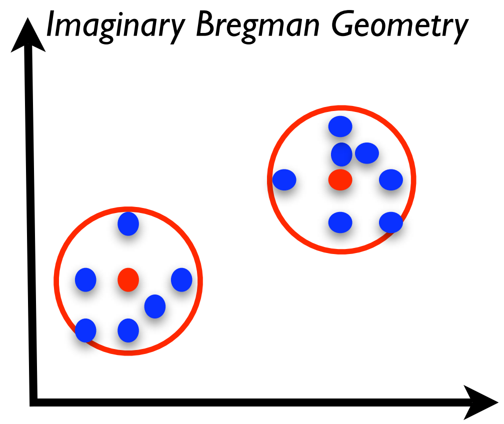
> **Definition** Two entities $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \mathcal{X}$ are assumed to be *similar* if the information gain by passing from one representation to other is zero or minimal; quantified by $d_X(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) < \epsilon$ which depends not on the signal itself, but on the probability functions $p_X(x; \boldsymbol{\theta}_0)$ and $p_X(\boldsymbol{x}; \boldsymbol{\theta}_1)$.

- How to choose *d(.,.)?*

Music Information Geometry

Appoach

- Proposal: Use the bijected Bregman divergence of the information geometry of audio data streams
- Data-IR:
  - For stationary data == Information carried between the signal's past {t=1...(n-1)} and present {t=1...n} *or* carried into the future
  - Is proven (mathematically) to be equal to *Bregman Information* on iid data
- Model-IR:
  - For non-stationary data
  - Requires *segmenting* audio stream into chunks.
  - Proposal:



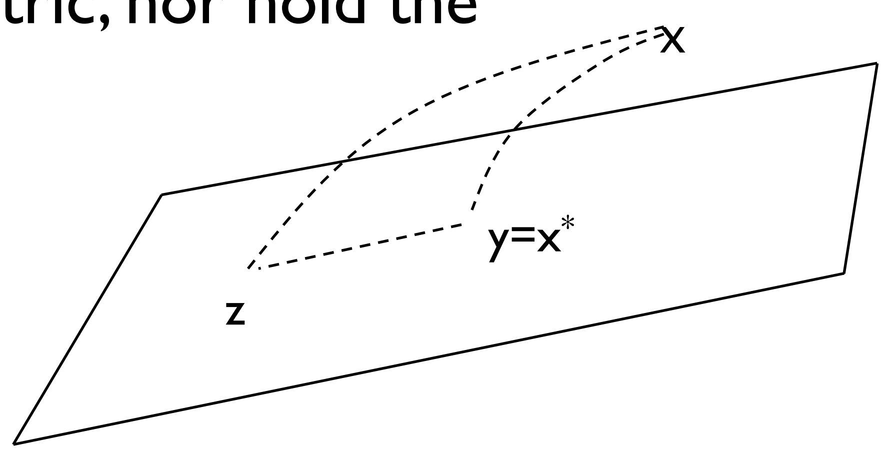
Definition Given a dual structure manifold $(\mathcal{S}, g, \Delta^D, \Delta^{D^*})$ derived on a regular exponential family formed on data-stream X_k, a *model* θ_i consist of a set $\mathcal{X}_i = \{\boldsymbol{x}_k | k \in \mathcal{N}, \mathcal{N} \subset \mathbb{N}\}$ that forms a *Bregman Ball* $B_r(\boldsymbol{\mu}_i, R_i)$ with center $\boldsymbol{\mu}_i$ and radius R_i.

Music Information Geometry

From Divergence to Similarity Metric

- Further requirements for *d:*
  - *symmetric* $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$
  - and to hold the *triangular inequality* $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y})$

  to obtain equivalent classes.
- Similarity ~ (1/Divergence)
- Problem: Bregman divergences are neither symmetric, nor hold the triangular inequality!
- Solutions: (Nielsen and Nock, 2007)
  a. Triangular inequality hold IFF *y* is the geometric projection of *x* onto the tangent plane passing through *zy*.
  b. In our geometry, the notions of max. likelihood and projection are ***equivalent***! `(thanks to Duality!)`
  c. Symmetrize Bregman divergence using a max. likelihood formulation!



We can approach both notions of symmetry and triangular inequality.

Music Information Geometry

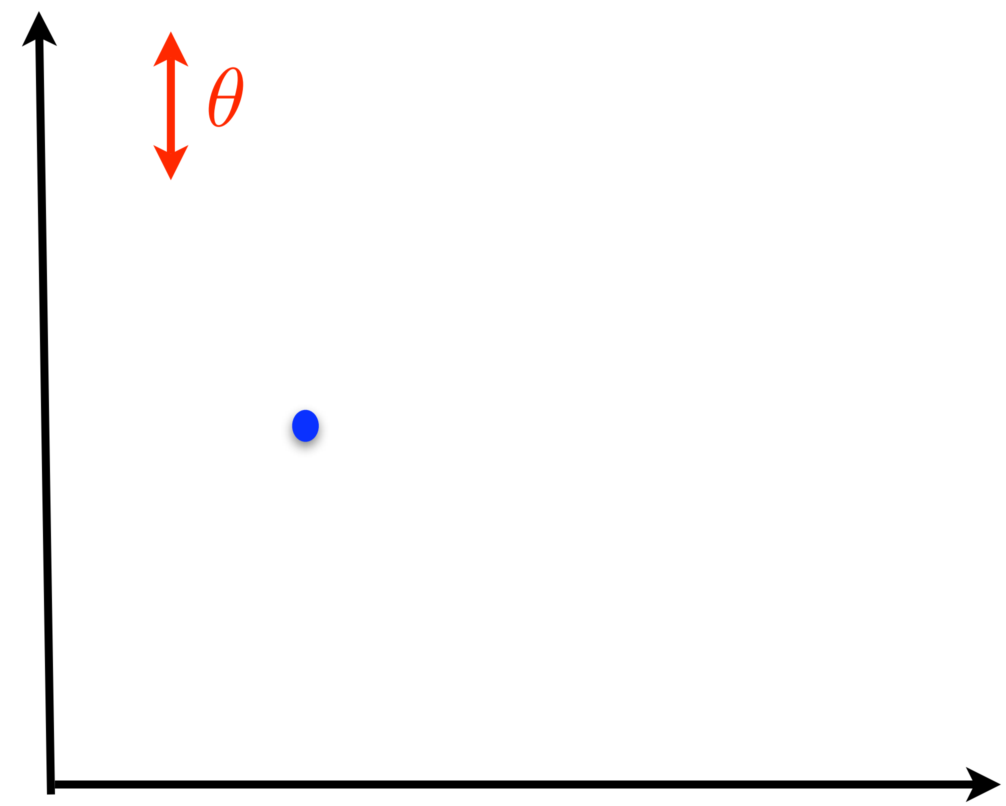
Model Formation

- Incrementally segment information regions into quasi-stationary chunks
  - A model is a Bregman ball whose information radius reveals the maximum distance in terms of *mutual information* within the ball.
  - Detect balls with *jumps* in information distance between a new point and a forming ball
  - *Computationally cheap:* Linear construction in time and space

Music Information Geometry

Incremental Segmentation

Imaginary Bregman Geometry



: Moving Centroid

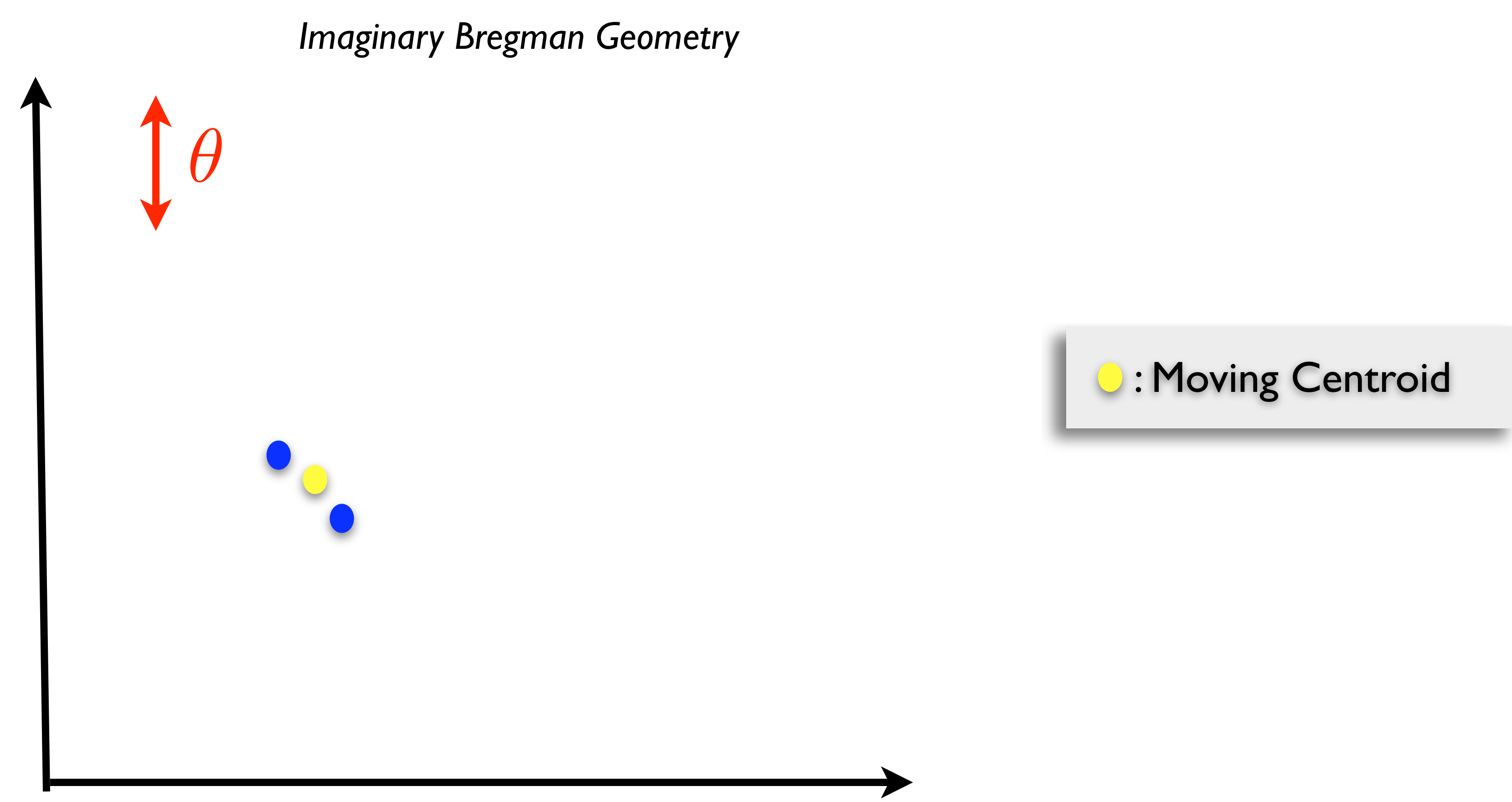
REALTIME Scheduler

Time = 1

No event!

Music Information Geometry

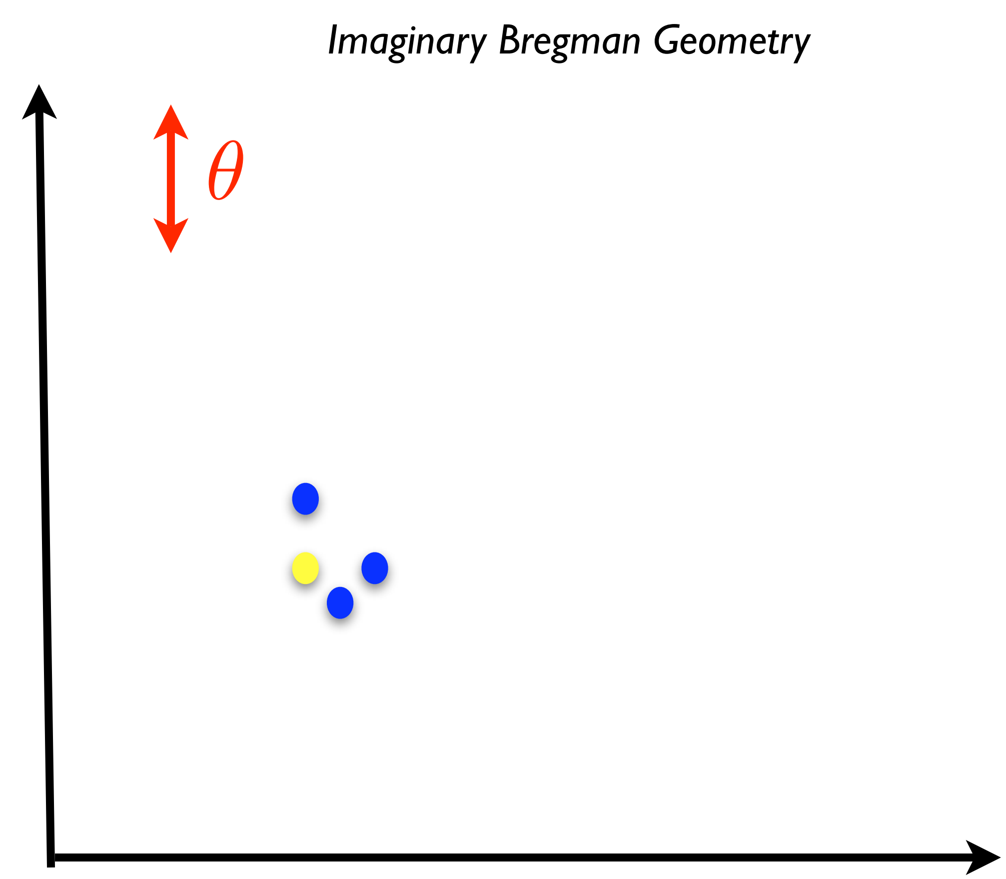
Incremental Segmentation



REALTIME Scheduler
Time = **2**

Music Information Geometry

Incremental Segmentation

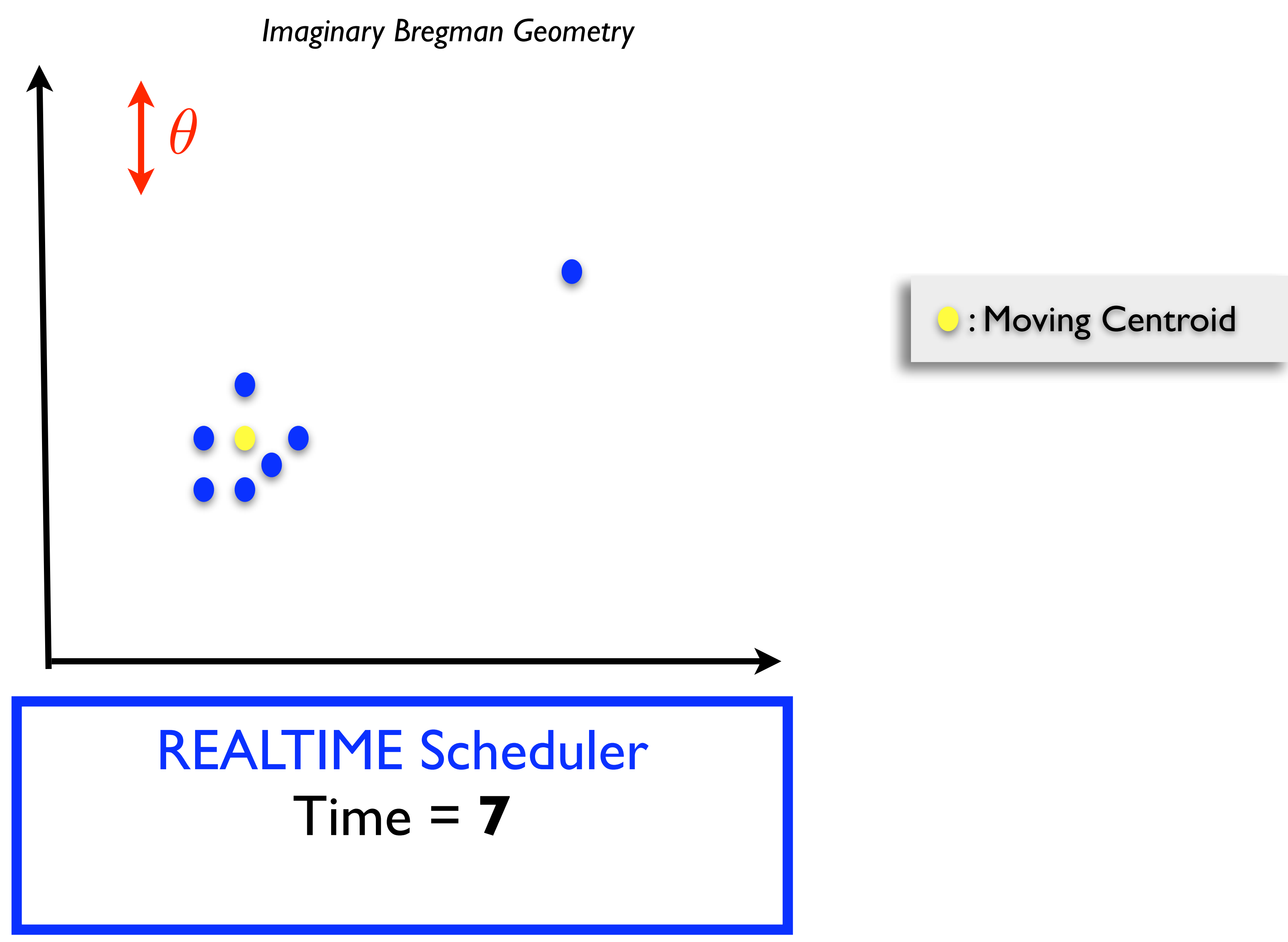


: Moving Centroid

REALTIME Scheduler
Time = **3**

Music Information Geometry

Incremental Segmentation

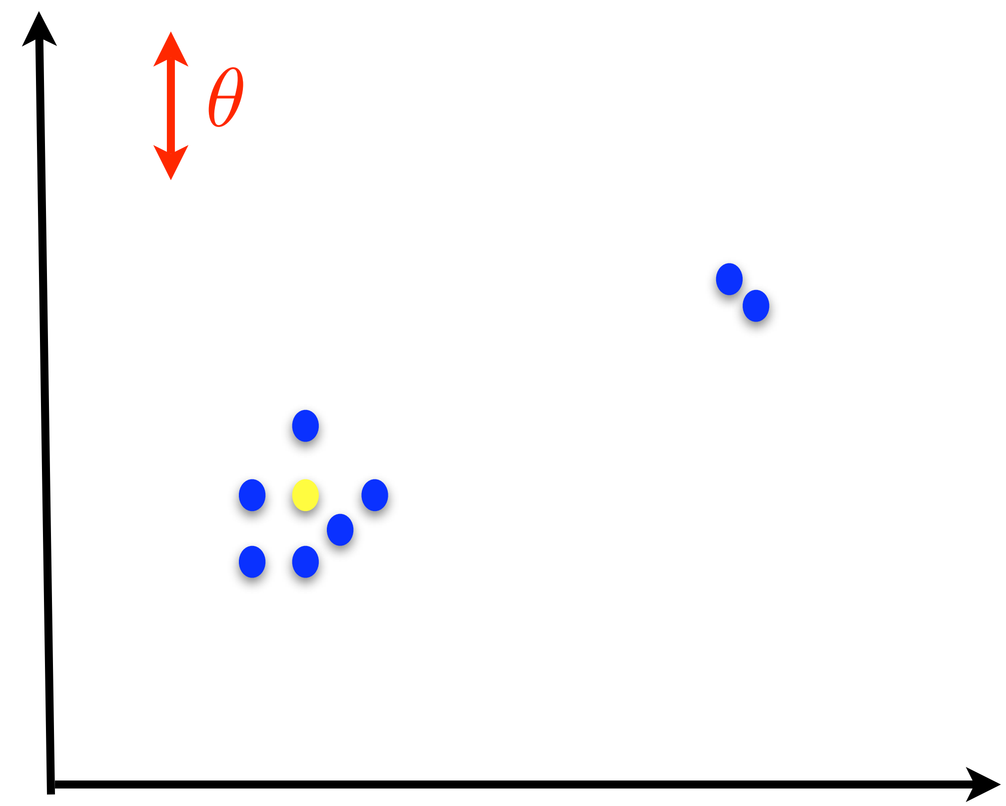


ircam
Centre
Pompidou

Music Information Geometry

Incremental Segmentation

Imaginary Bregman Geometry



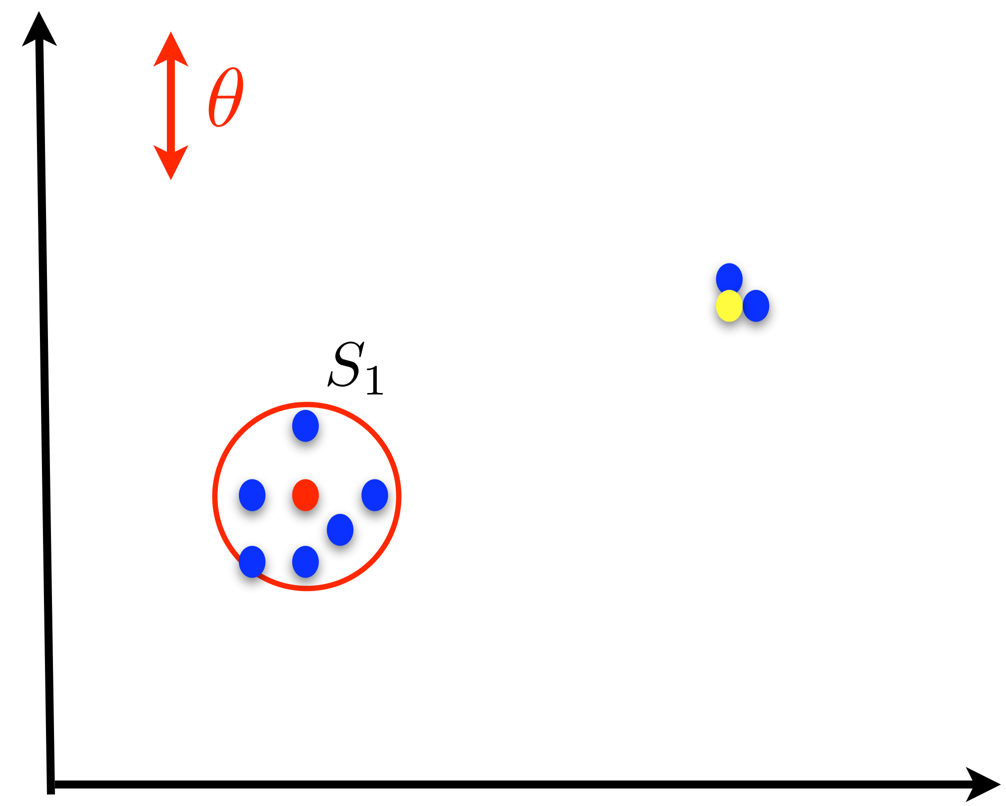
: Moving Centroid

REALTIME Scheduler
Time = **8**

Music Information Geometry

Incremental Segmentation

Imaginary Bregman Geometry



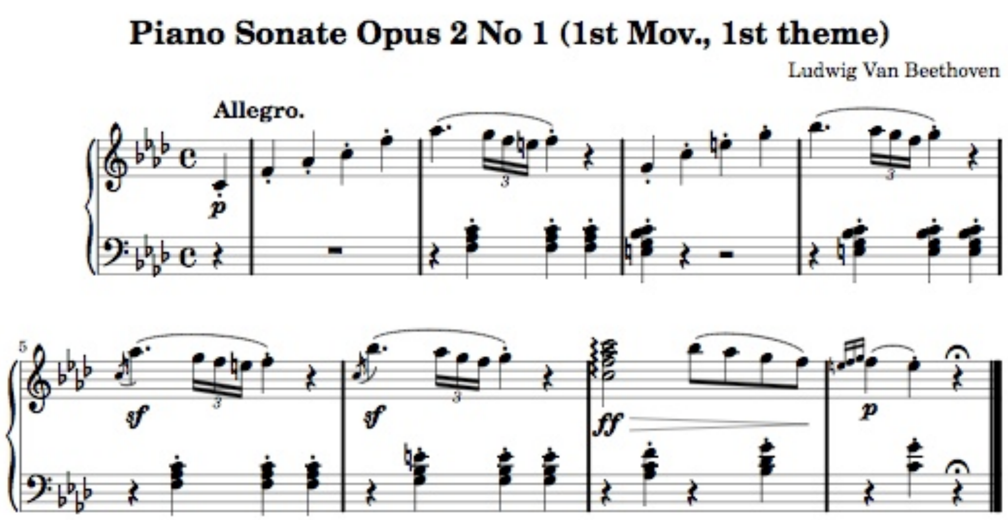
: Moving Centroid

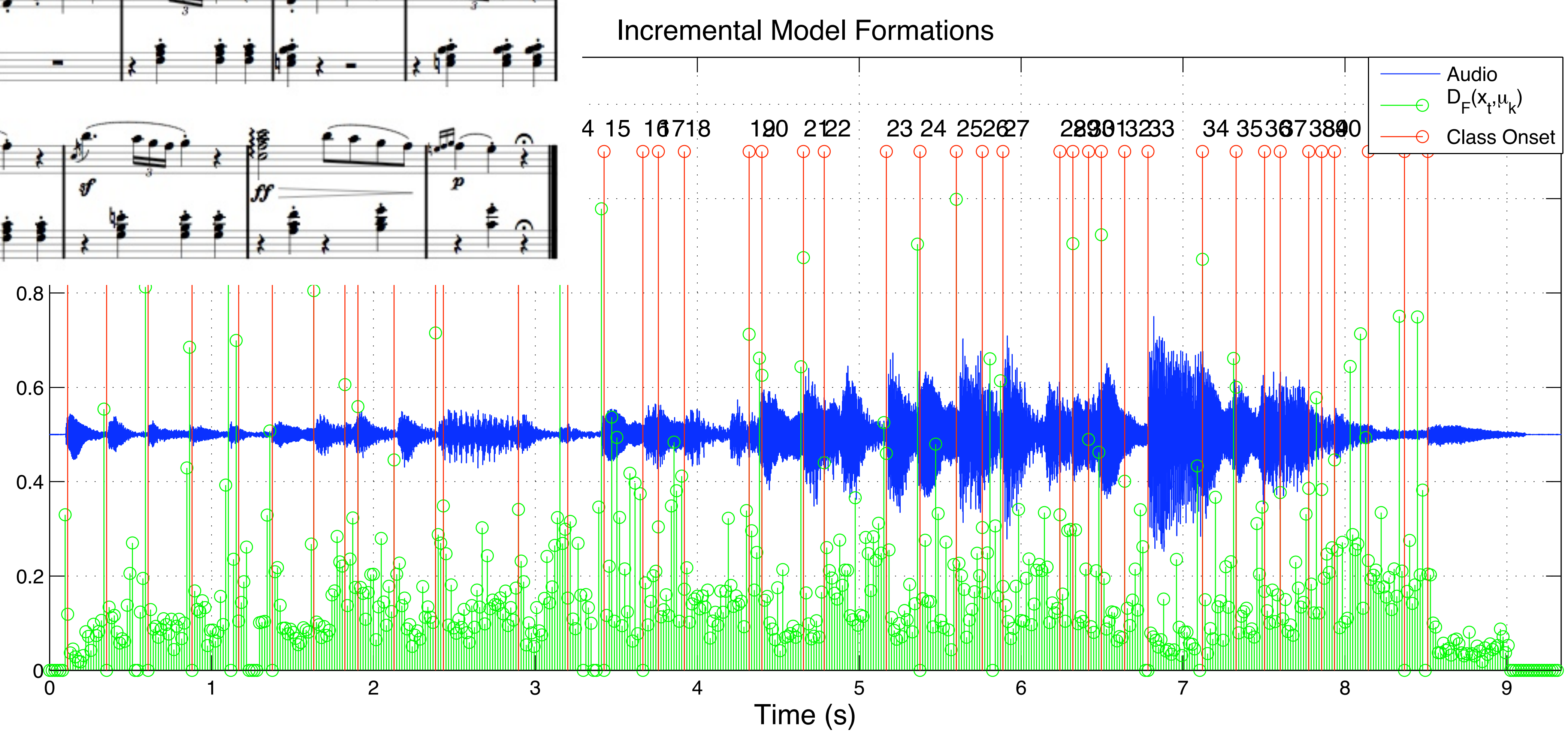
REALTIME Scheduler
Time = **8**
Model Formation

Music Information Geometry

Incremental Segmentation

- Sample Result: Beethoven's first piano sonata, first movement
  - performed by Friedrich Gulda (1958)





ircam Centre Pompidou

Methods of Information Access

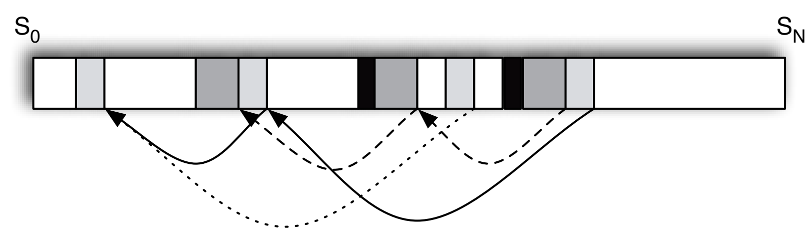
Incremental Structure Discovery

- Idea: The *models* in music information geometry provide instantaneous similarities between consequent models.
  - What about similarities between subsets of models at different time intervals?
  - What about grabbing long term regularities in the music signal?
- Literature of *Audio Structure Discovery* algorithms: Usually off-line and/or incorporate a priori beliefs over music structure
- Our goal:
  - Do it *online* and *incrementally* as audio signals arrive
  - Grab and learn regularities on-the-fly from the signal itself and without a priori knowledge
  - *Key for Anticipatory Modeling*: Grabbing stabilities and regularities of information in the environment

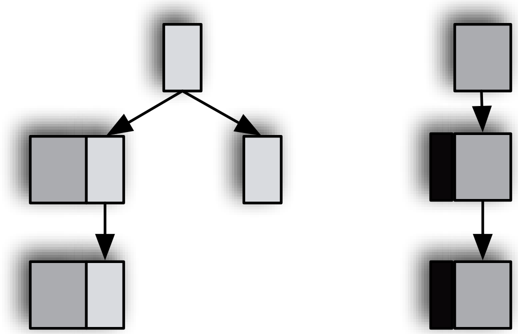
Methods of Information Access

Incremental Structure Discovery

- Proposal: Extend an existing algorithm in the symbolic domain to the continuous audio domain by passing through information geometry and *Models*.
- Point of departure: *Factor Oracles*
  - Used primarily on text and DNA data to detect repeating structures.
  - A finite-state automaton learned incrementally.
  - A state-space representation of repeating structures in a sequence



  - Provides *forest of suffix tree structures*

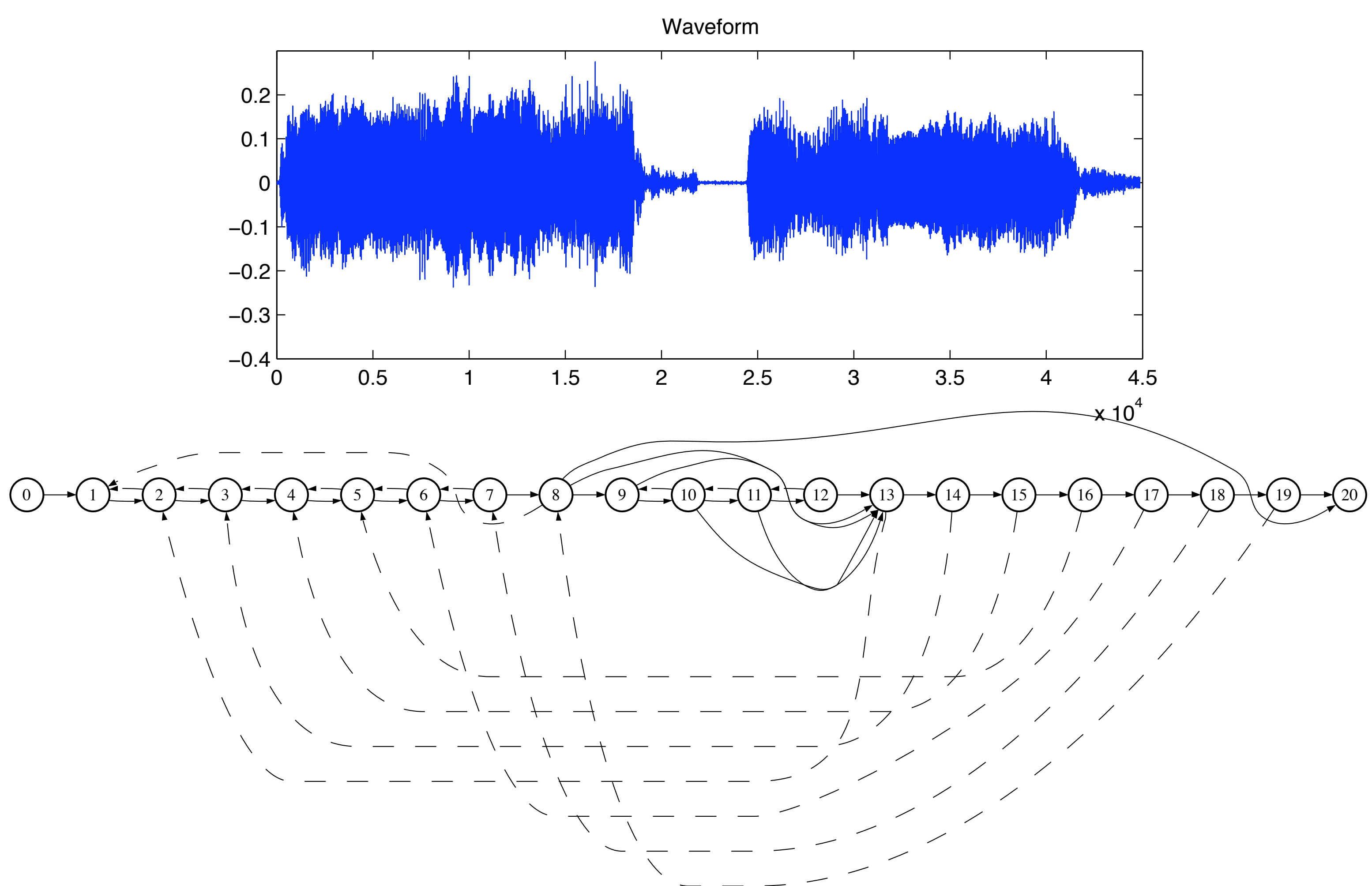


- *The beauty of MIG*
  - Keep the algorithm, replace symbols by *models* or *points* and equivalence by *similarity* in a music information geometry!

  ⇨ **Audio Oracle**

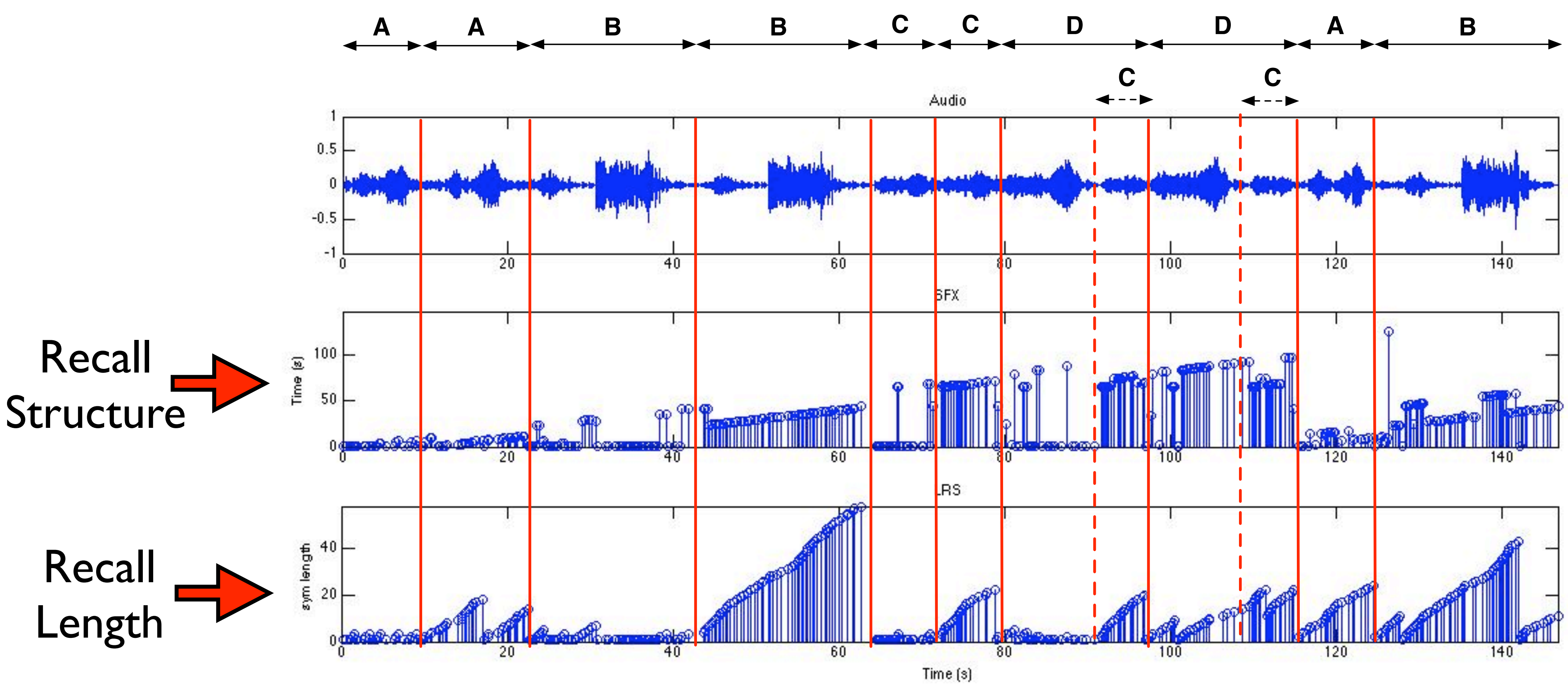
Methods of Information Access

- Audio Oracle results
  - On *points:* (each state=one analysis window)
    - *Natural bird uttering (natural repetition)*
    - *Using MFCC audio features on Multinomial music information geometry*



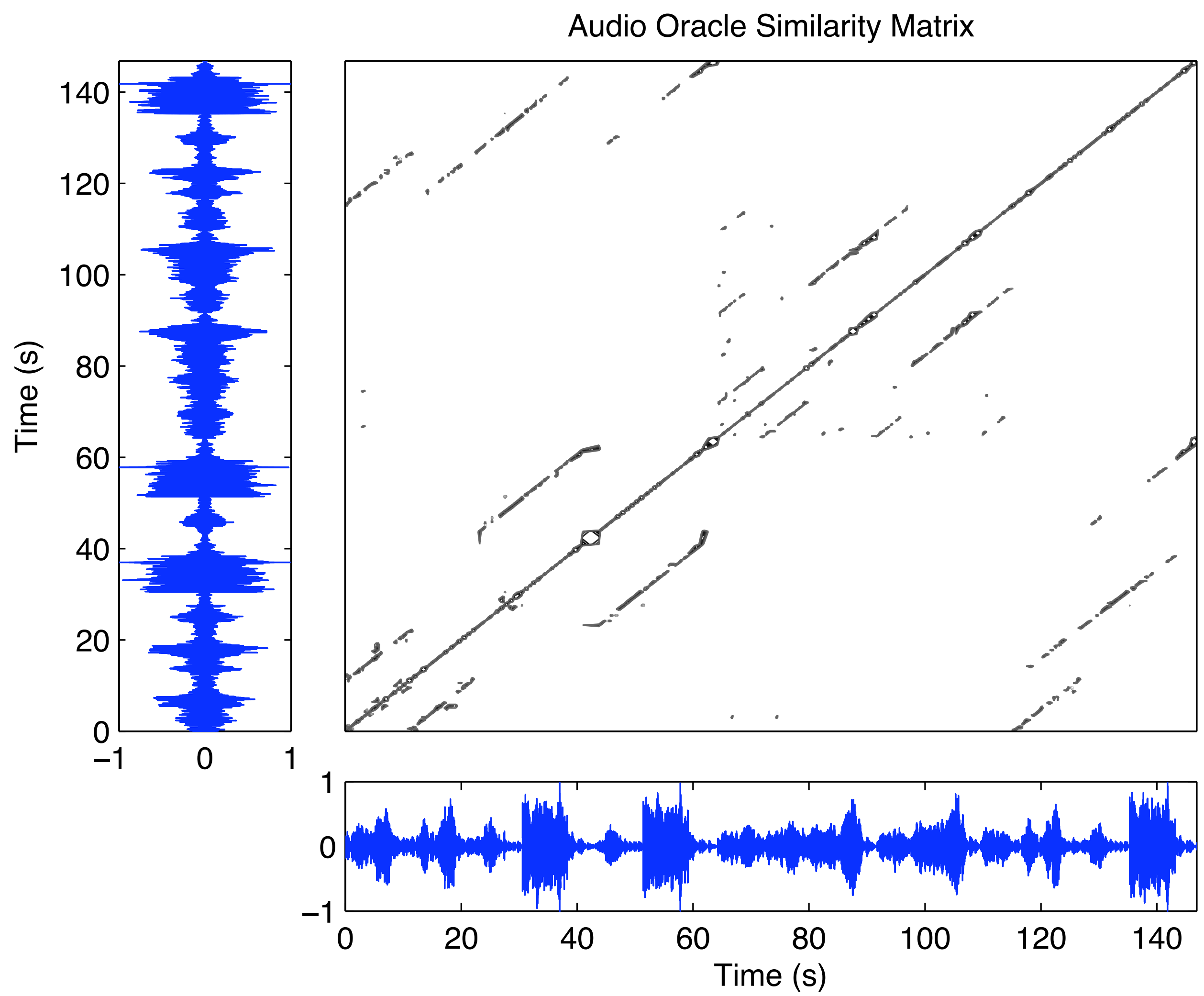
Methods of Information Access

- Audio Oracle results:
  - On *models*
    - *Beethoven's first Piano Sonata, Third Movement (Gulda, 1958)*
    - *Using Constant-Q amplitude spectrum on Multinomial music information geometry*
      - *150 seconds, > 9500 analysis frames, resulting to 440 states*

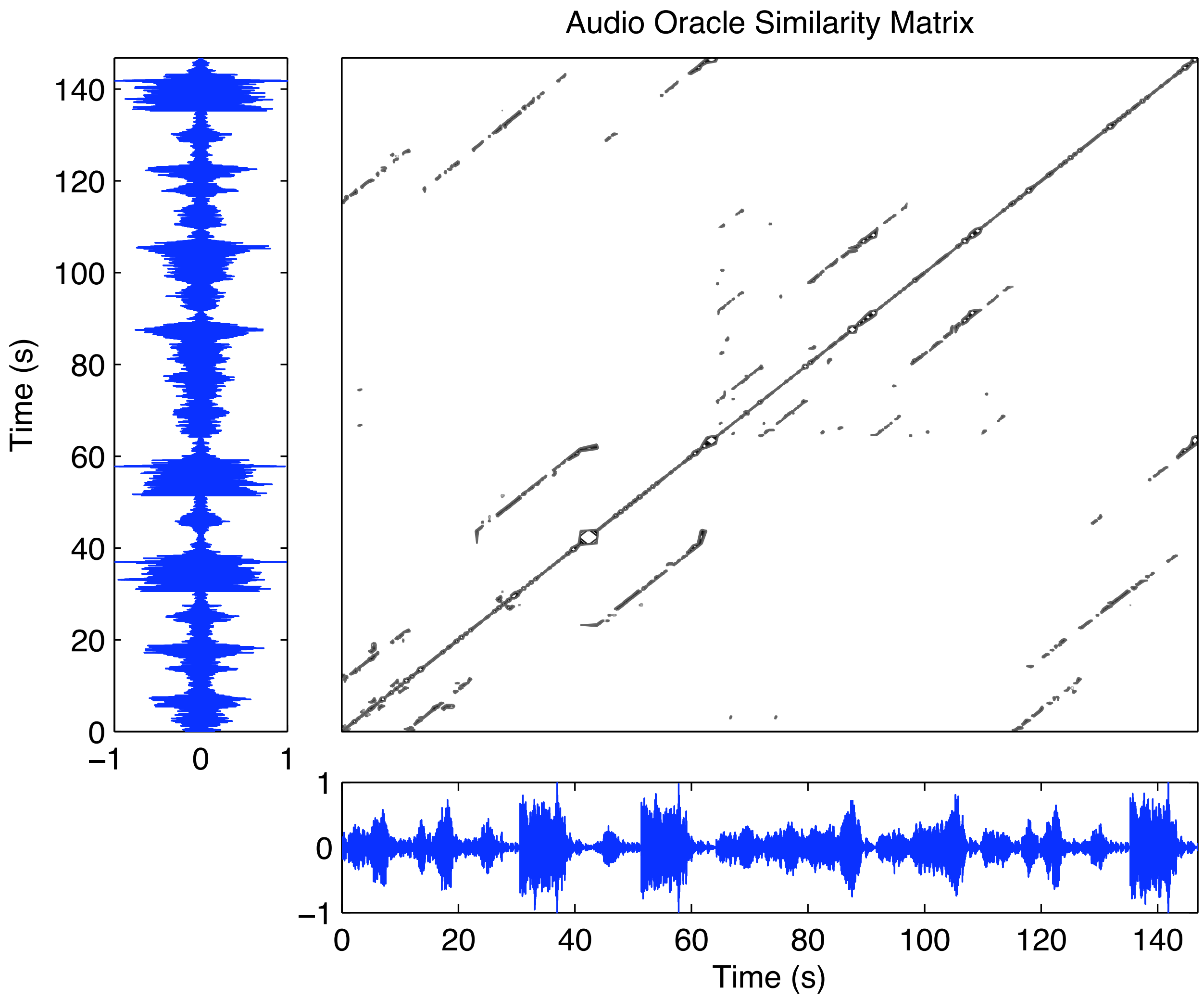


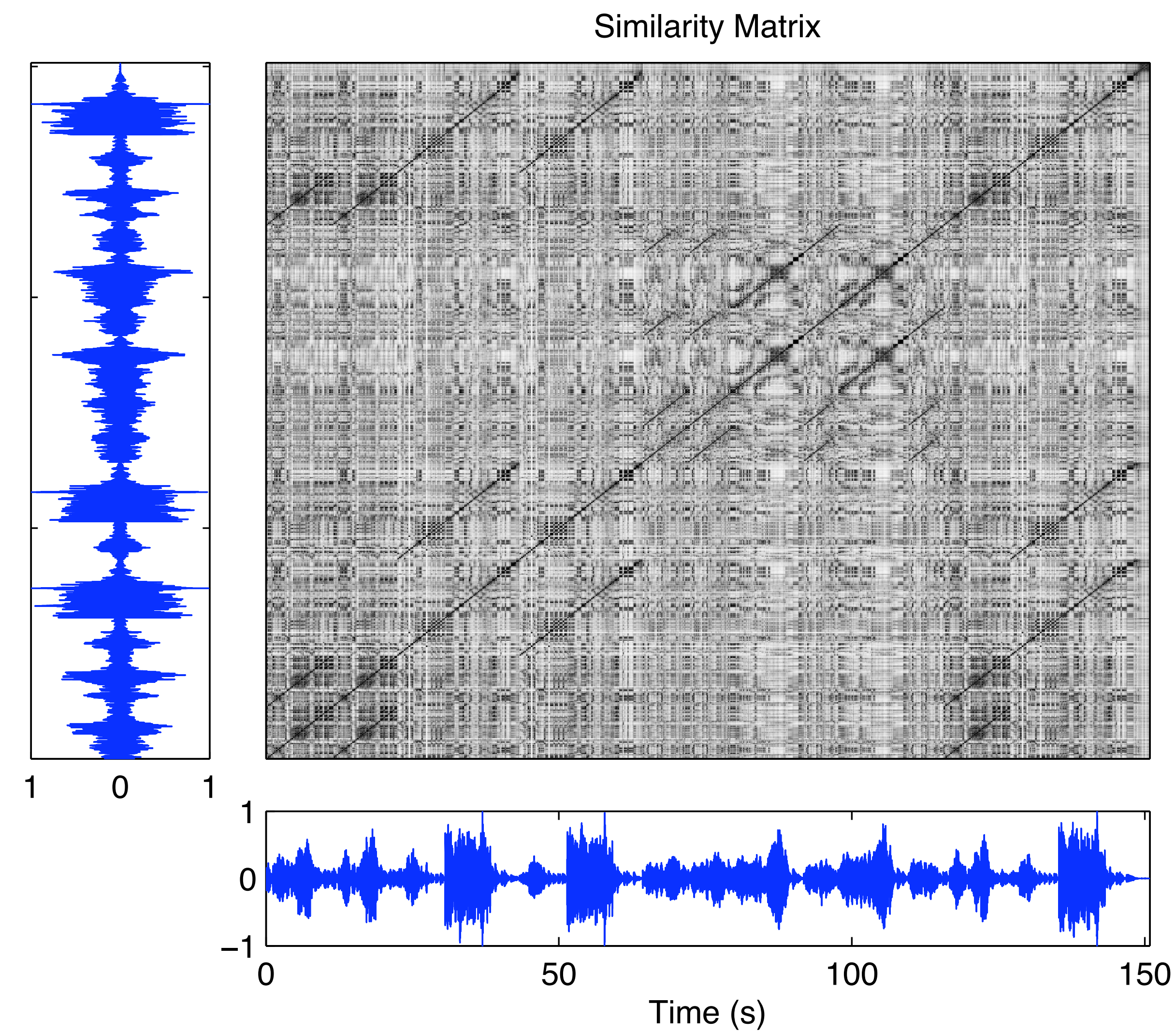
Methods of Information Access

- Audio Oracle results:
  - On *models*
    - *Beethoven's first Piano Sonata, Third Movement (Gulda, 1958)*
    - *Realtime computation, sparsity, less complexity, more robust to complex changes in the environment:*



Compare...


Audio Oracle Similarity Matrix
Time (s)
140
120
100
80
60
40
20
0
−1
0
1
1
0
−1
0
20
40
60
80
100
120
140
Time (s)


Similarity Matrix
1
0
1
1
0
−1
0
50
100
150
Time (s)

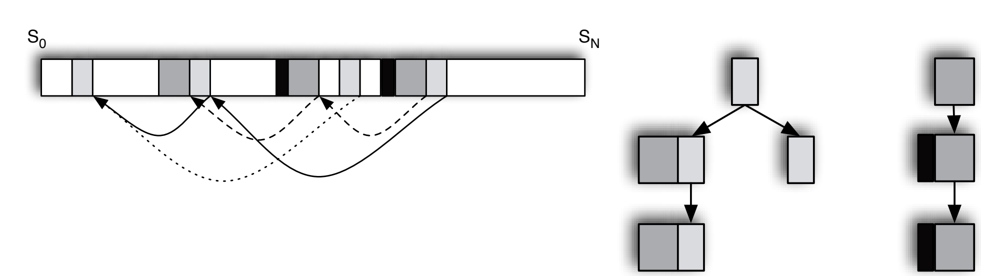
ircam Centre Pompidou

Methods of Information Access

<u>*Fast Information Retrieval*</u>

- Proposal: Compile an *search engine* over a database of audio and using an outside audio query
  - That is also capable of *recombining/reassembling* chunks of audio within a large target, to reconstruct the query.
- Idea: Do not search on the audio itself but on *audio structures*
  - Audio Oracle as Meta data
  - (ab)use the long-term structures of Audio Oracle to maintain *perceptual continuity* of the results (access to long term structures)
- Simple *Dynamic Programming* algorithm:
  - Follow the *forest of suffix tree structures* to find the longest and best possible result
  - Maintains all the results (paths) at all times!

⇨ **Guidage**

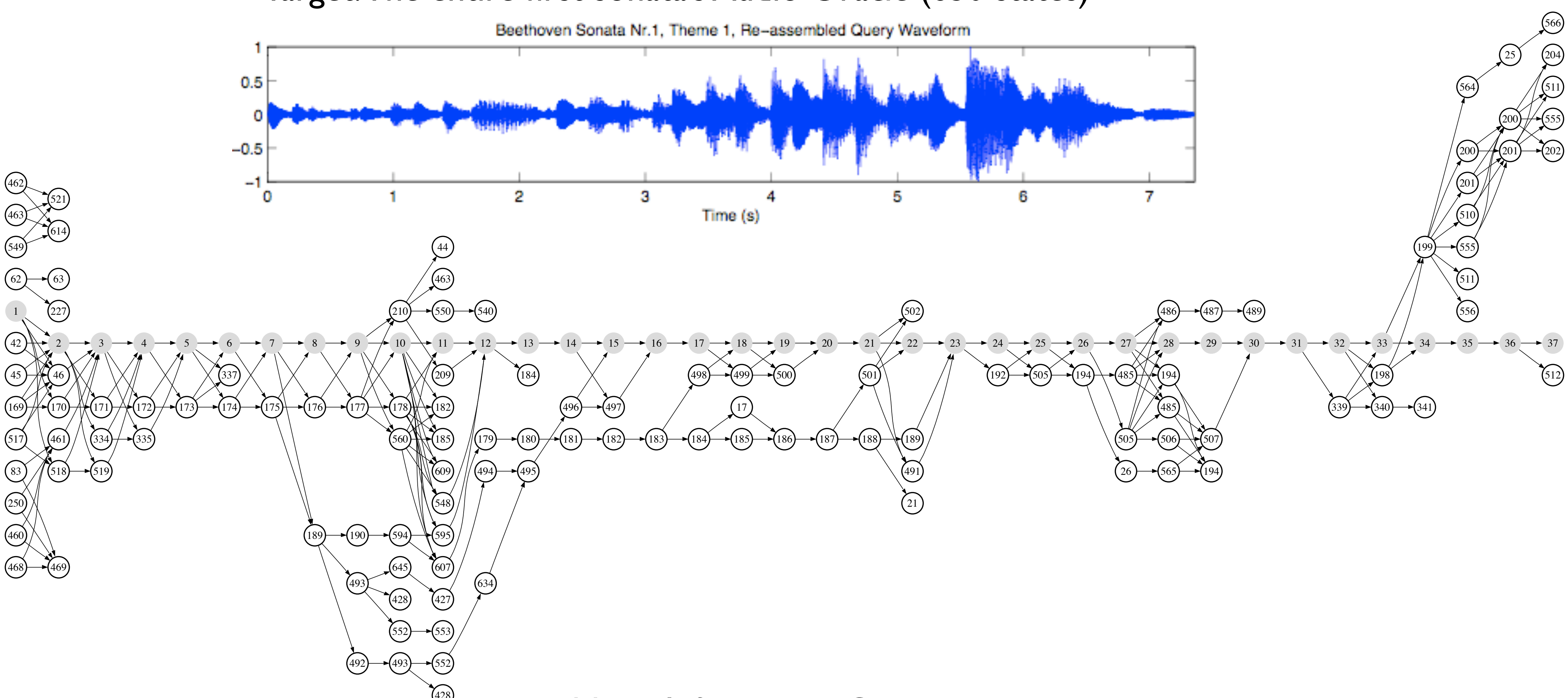


ircam Centre Pompidou

Methods of Information Access

Guidage Results

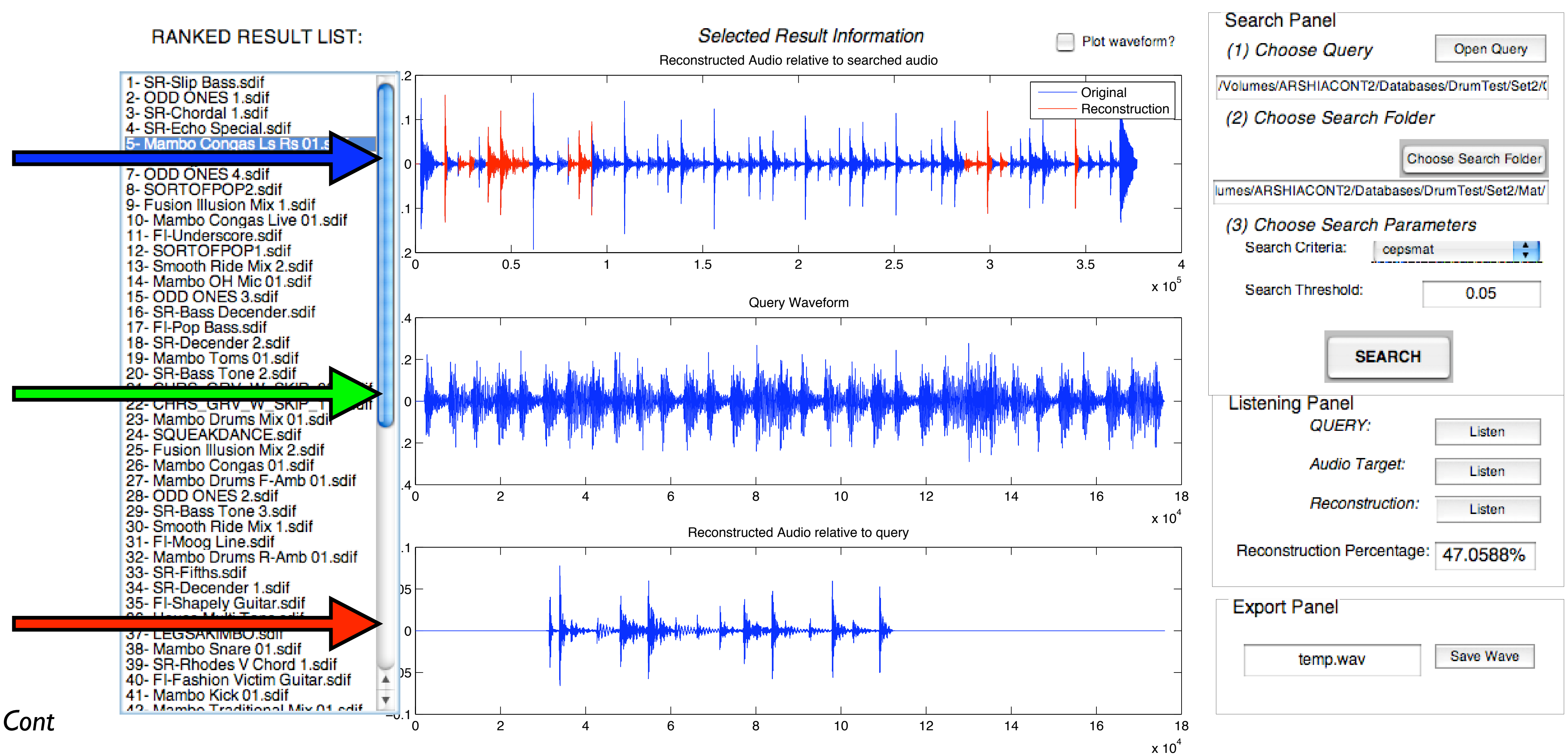
- Self-Similarity test
  - Task: Search for the *first theme* of the first Beethoven's sonata in the entire sonata.
    - Query: Audio of the first theme
    - Target: The entire first sonata's Audio Oracle (650 states)



Methods of Information Access

Guidage Results

- Database Search
  - Task: Find audio, or a recombination within a file that are similar to query
    - Query: African drum sample
    - Database: Loop database (*Kontakt*) 140 audio files, 200Mb, Mean duration 7s
      - Convergence time: 20s in *Matlab* on a 2.3Ghz unicore Intel machine



ircam Centre Pompidou

Future

- This is just a preliminary (and promising) step towards our goal...
  - Study the geometric behavior of music information entities
    - Information geometric Analysis and Resynthesis of audio streams
    - Formalization of theoretical foundations with regard to music processing
    - Validation with perceptual experience
    - PhD project of Arnaud Dessein
  - Extension to other media streams (video)
  - Applications for composers and music scholars
    - OMax audio
    - Concatenative Synthesis
    - High-level (and geometric) control of sound synthesis
  - Engineering / Industrial applications
    - Music Data Mining
    - Audio Thumbnailing
    - Music Information Retrieval...

Merci!

Arshia Cont
cont@ircam.fr

Shlomo Dubnov
sdubnov@ucsd.edu

Gérard Assayag
assayag@ircam.fr

Arnaud Dessein
arnaud.dessein@ircam.fr

ircam
Centre Pompidou

Further Readings:

Arshia Cont, Shlomo Dubnov and Gérard Assayag. *On the Information Geometry of Audio Streams with Applications to Automatic Structure Discovery*. IEEE Transactions on Audio, Speech and Language Processing, 2009. (forthcoming)

Arshia Cont. *A coupled duration-focused architecture for realtime music to score alignment.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.

Arshia Cont, *Modeling Musical Anticipation: From the time of music to the music of time*. PhD thesis, University of Paris 6 and University of California in San Diego, October 2008.