Applications de la géométrie de l'information au traitement des flux audio temps réel

Arshia Cont

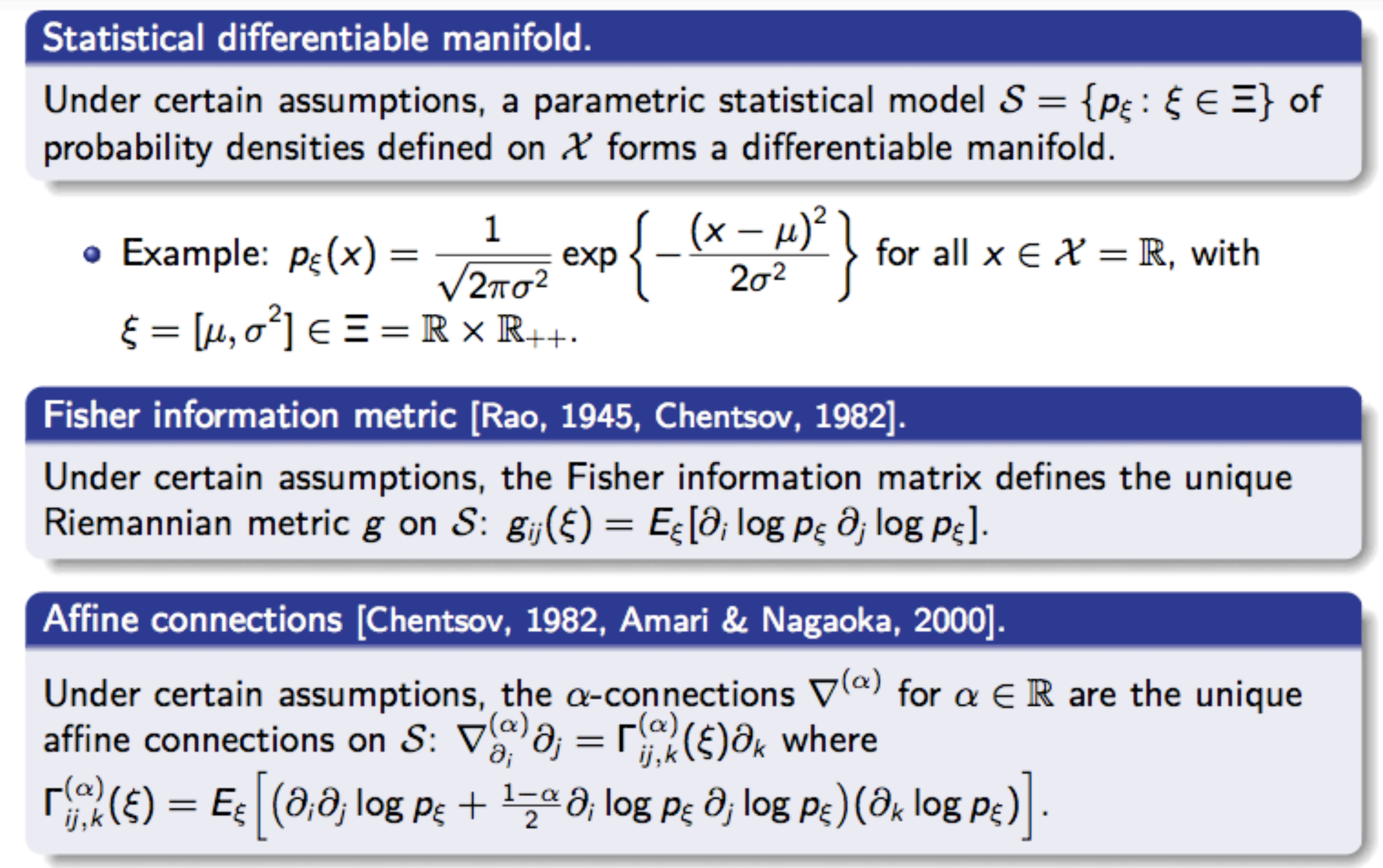
Researcher, Musical Representations Team +

Head of Musical Research

cont@ircam.fr

ircam
Centre Pompidou

Background



Statistical differentiable manifold.

Under certain assumptions, a parametric statistical model $\mathcal{S} = \{p_\xi : \xi \in \Xi\}$ of probability densities defined on $\mathcal{X}$ forms a differentiable manifold.

- Example: $p_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ for all $x \in \mathcal{X} = \mathbb{R}$, with $\xi = [\mu, \sigma^2] \in \Xi = \mathbb{R} \times \mathbb{R}_{++}$.

Fisher information metric [Rao, 1945, Chentsov, 1982].

Under certain assumptions, the Fisher information matrix defines the unique Riemannian metric g on $\mathcal{S}$: $g_{ij}(\xi) = E_\xi[\partial_i \log p_\xi \, \partial_j \log p_\xi]$.

Affine connections [Chentsov, 1982, Amari & Nagaoka, 2000].

Under certain assumptions, the α-connections $\nabla^{(\alpha)}$ for $\alpha \in \mathbb{R}$ are the unique affine connections on $\mathcal{S}$: $\nabla^{(\alpha)}_{\partial_i} \partial_j = \Gamma^{(\alpha)}_{ij,k}(\xi) \partial_k$ where $\Gamma^{(\alpha)}_{ij,k}(\xi) = E_\xi\left[\left(\partial_i \partial_j \log p_\xi + \frac{1-\alpha}{2} \partial_i \log p_\xi \, \partial_j \log p_\xi\right)\left(\partial_k \log p_\xi\right)\right]$.

Background

Exponential family.

$p_\theta(x) = \exp\left(\theta^T T(x) - F(\theta) + C(x)\right)$ for all $x \in \mathcal{X}$.

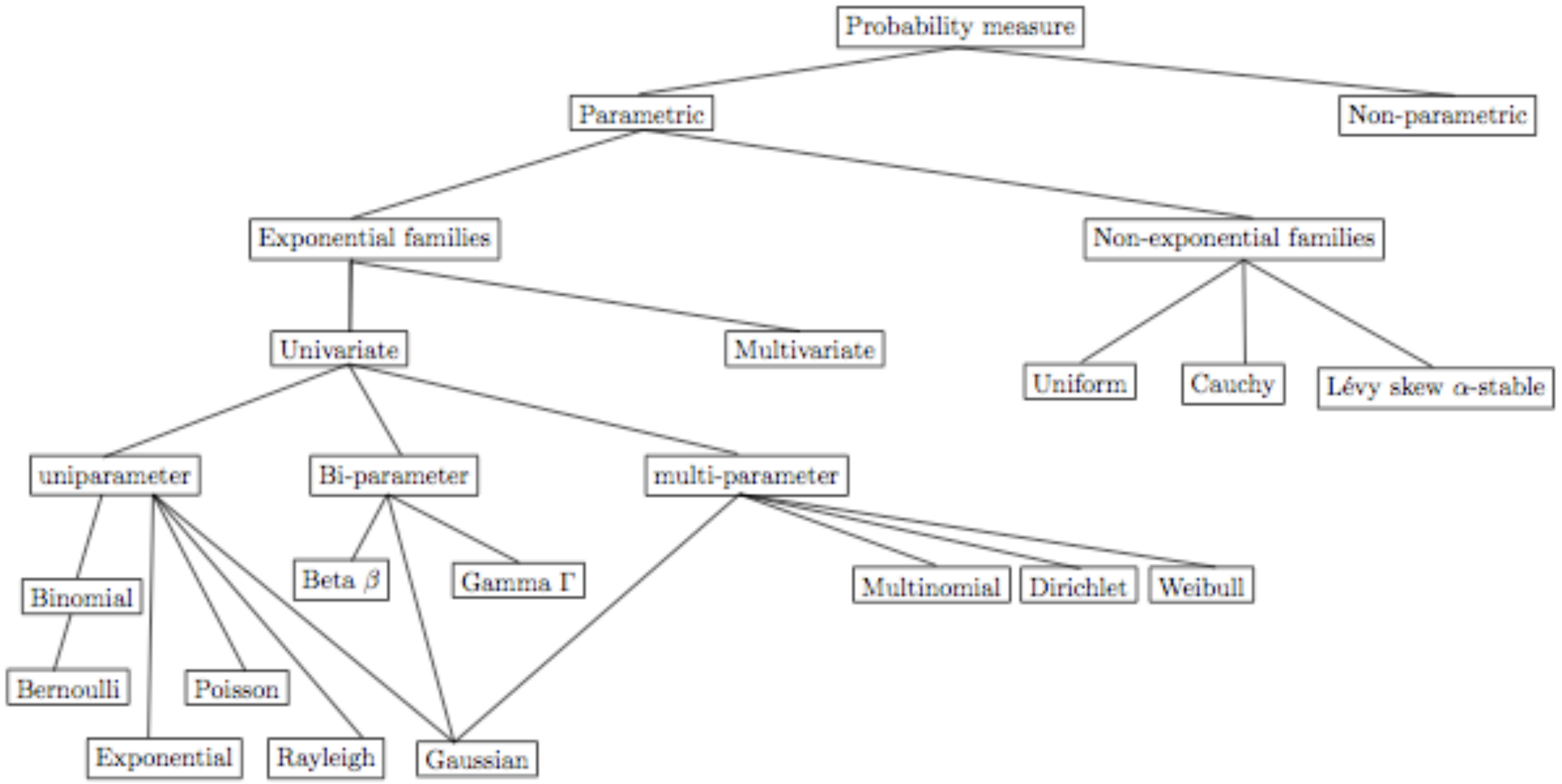


Figure: A taxonomy of probability measures [Nielsen & Garcia, 2009].

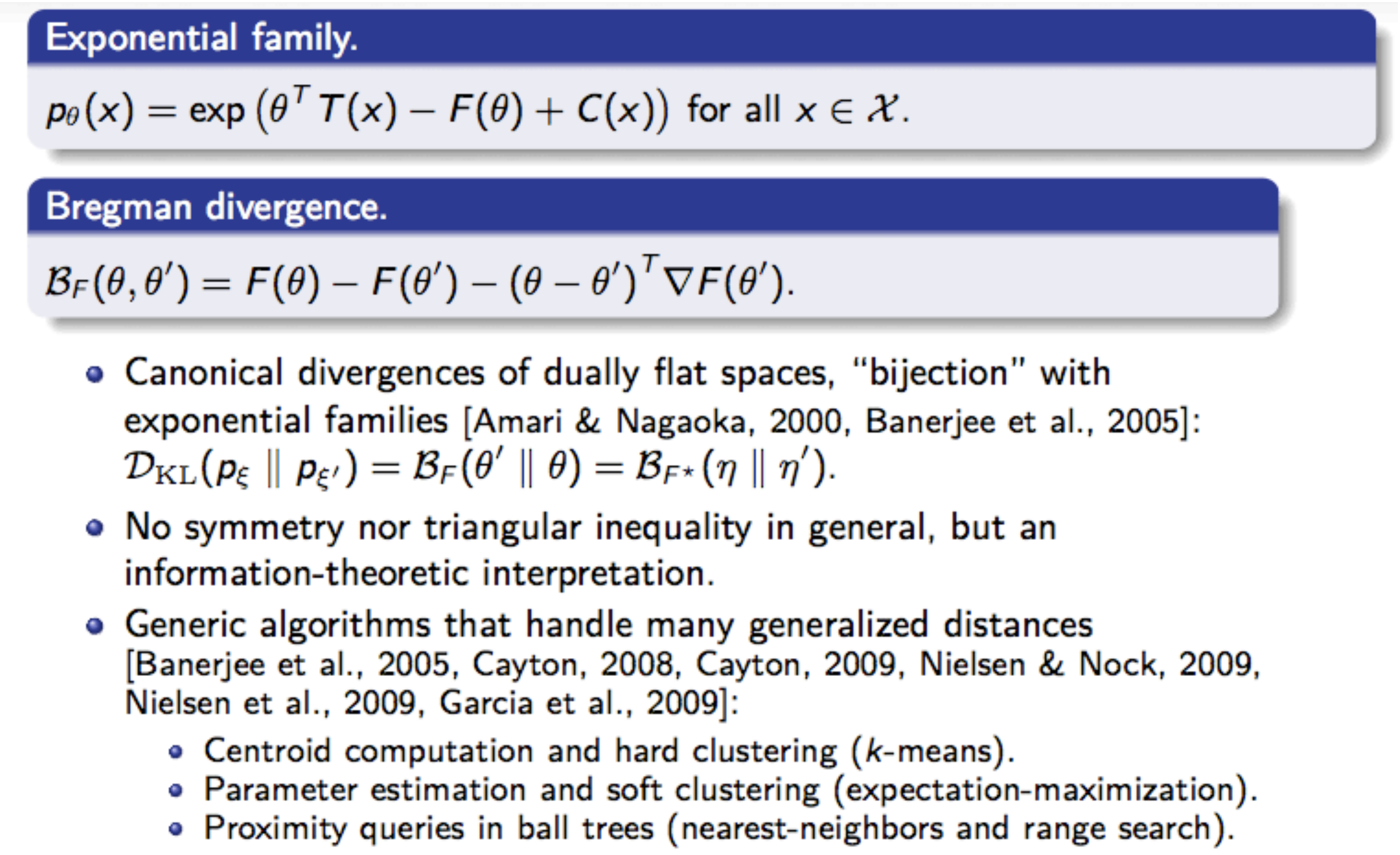
Background

Exponential family.

$p_\theta(x) = \exp\left(\theta^T T(x) - F(\theta) + C(x)\right)$ for all $x \in \mathcal{X}$.

- We consider a statistical manifold $\mathcal{S} = \{p_\theta : \theta \in \Theta\}$ equipped with g and the dual exponential and mixture connections $\nabla^{(1)}$ and $\nabla^{(-1)}$.
- $(\mathcal{S}, g, \nabla^{(1)}, \nabla^{(-1)})$ possesses two dual affine coordinate systems, natural parameters θ and expectation parameters $\eta = \nabla F(\theta)$.
- Dually flat geometry, Hessian structure, generated by the potential F together with its conjugate potential $F^\star$ defined by the Legendre-Fenchel transform: $F^\star(\eta) = \sup_{\theta \in \Theta} \theta^T \eta - F(\theta)$, which verifies $\nabla F^\star = (\nabla F)^{-1}$ so that $\theta = \nabla F^\star(\eta)$.
- Generalizes the self-dual Euclidean geometry, with notably two canonically associated Bregman divergences $\mathcal{B}_F$ and $\mathcal{B}_{F^\star}$ instead of the self-dual Euclidean distance, but also dual geodesics, a generalized Pythagorean theorem and dual projections.

Background



Exponential family.

$p_\theta(x) = \exp\left(\theta^T T(x) - F(\theta) + C(x)\right)$ for all $x \in \mathcal{X}$.

Bregman divergence.

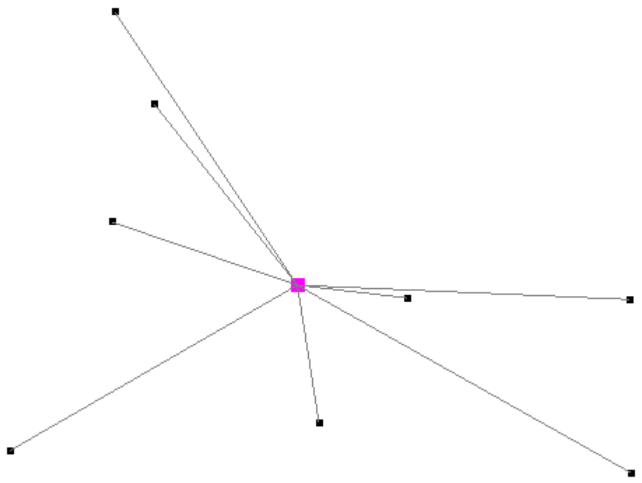
$\mathcal{B}_F(\theta, \theta') = F(\theta) - F(\theta') - (\theta - \theta')^T \nabla F(\theta')$.

- Canonical divergences of dually flat spaces, "bijection" with exponential families [Amari & Nagaoka, 2000, Banerjee et al., 2005]: $\mathcal{D}_{\mathrm{KL}}(p_\xi \parallel p_{\xi'}) = \mathcal{B}_F(\theta' \parallel \theta) = \mathcal{B}_{F^\star}(\eta \parallel \eta')$.
- No symmetry nor triangular inequality in general, but an information-theoretic interpretation.
- Generic algorithms that handle many generalized distances [Banerjee et al., 2005, Cayton, 2008, Cayton, 2009, Nielsen & Nock, 2009, Nielsen et al., 2009, Garcia et al., 2009]:
  - Centroid computation and hard clustering (k-means).
  - Parameter estimation and soft clustering (expectation-maximization).
  - Proximity queries in ball trees (nearest-neighbors and range search).

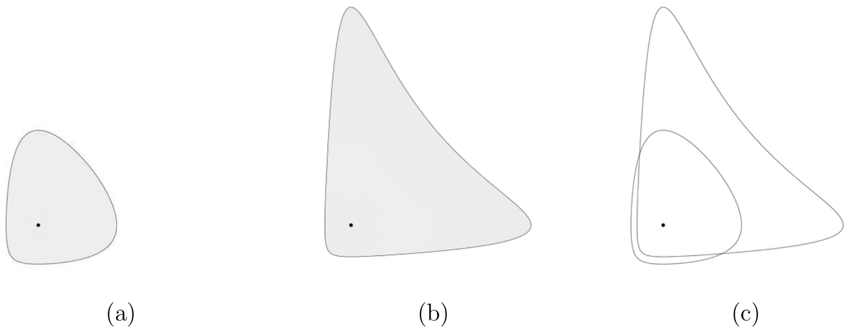
Background

Elements of Bregman Geometry

- *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



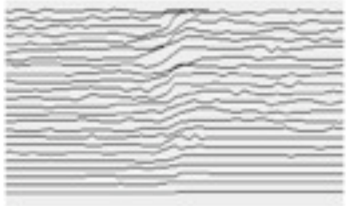
- *Bregman Information* of a random variable *X*
  - Defined as the expectation over divergences of all *points* from the centroid
  - Special cases: *variance, mutual information*

Representational levels of music

Information Theoretical

		Data rate

`<genre> Romantic </genre>`
`<form> Sonata </form>`
Grazioso
	Semantic	< 0.1 Hz

	Symbolic	0.1-25 Hz
	Control	10Hz-1kHz
	Signal	10-100 kHz
p(x,y,z,t)	Physical	10-100 kHz

Low information quantity
Implicit knowledge

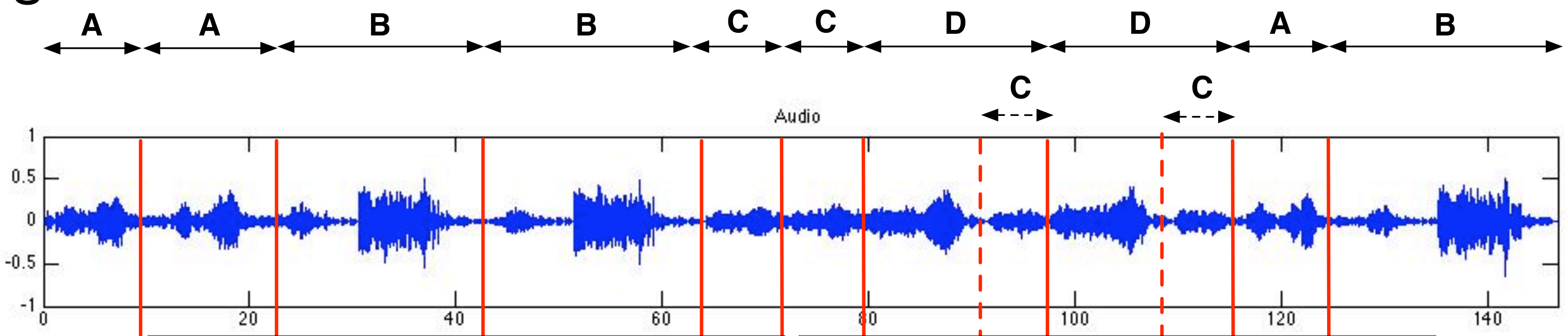
Information Generation: Synthesis ↓

↑ Information Reduction: Analysis

High information quantity
Explicit representations

Motivation

- Bridge the gap between *signal* and *symbolic* aspects of music information
- Example: “Effortless” abilities among music listeners and/or trained musicians that still pose challenging problems for machine intelligence:
  - Signal domain



  - Symbolic domain

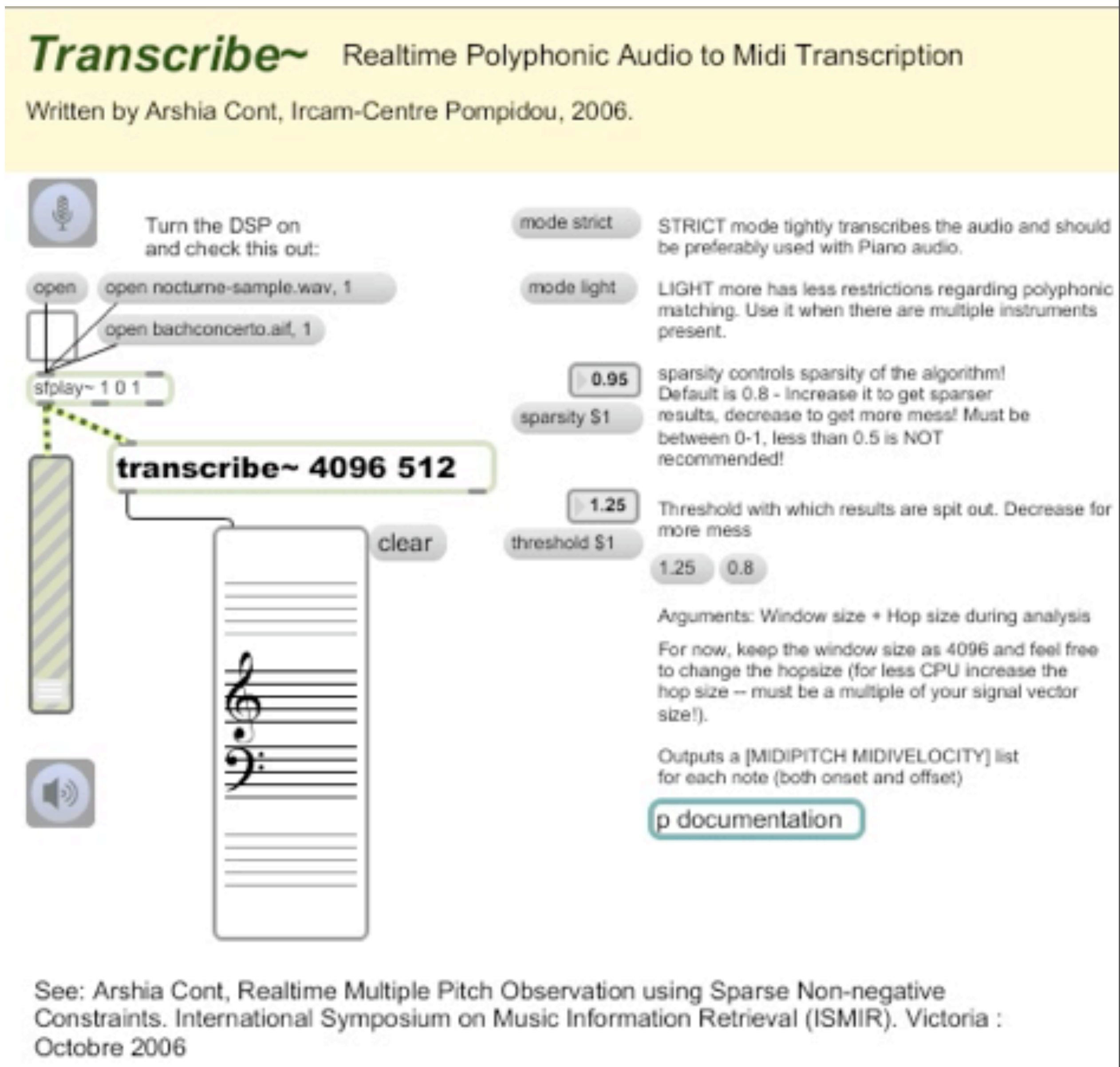
Piano Sonate Opus 2 No 1 (3rd Movement)
Ludwig Van Beethoven
MENUETTO.
Allegretto.

Motivation

- Current approaches:
  - Automatic transcription:
    - Explicit map of signal to symbol
    - But do humans really do that?
    - What about untrained listeners?
    - ill-posed approach.. .

Arshia Cont, "Realtime multiple pitch observation using sparse non-negative constraints". *Proceedings of the 7th International Symposium on Music Information Retrieval (ISMIR)*. Victoria, Canada, October 2006.

Arnaud Dessein, Arshia Cont, Guillaume Lemaitre. "Real-time polyphonic music transcription with non-negative matrix factorization and beta-divergence". *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR)*. Utrecht, Netherlands, August 2010, pp. 489-494.



Transcribe~ Realtime Polyphonic Audio to Midi Transcription

Written by Arshia Cont, Ircam-Centre Pompidou, 2006.

Turn the DSP on
and check this out:

STRICT mode tightly transcribes the audio and should be preferably used with Piano audio.

LIGHT more has less restrictions regarding polyphonic matching. Use it when there are multiple instruments present.

sparsity controls sparsity of the algorithm! Default is 0.8 - Increase it to get sparser results, decrease to get more mess! Must be between 0-1, less than 0.5 is NOT recommended!

Threshold with which results are spit out. Decrease for more mess

Arguments: Window size + Hop size during analysis

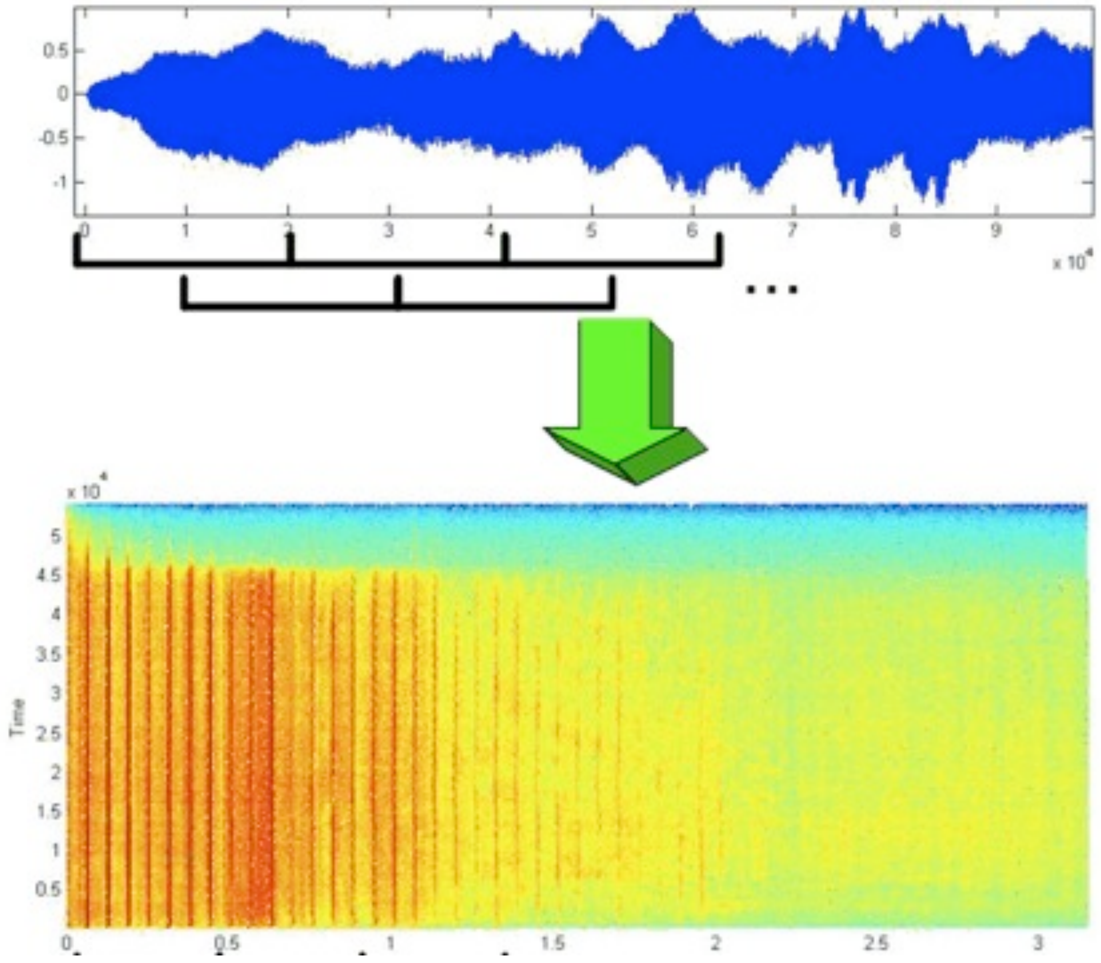
For now, keep the window size as 4096 and feel free to change the hopsize (for less CPU increase the hop size -- must be a multiple of your signal vector size!).

Outputs a [MIDIPITCH MIDIVELOCITY] list for each note (both onset and offset)

See: Arshia Cont, Realtime Multiple Pitch Observation using Sparse Non-negative Constraints. International Symposium on Music Information Retrieval (ISMIR). Victoria : Octobre 2006

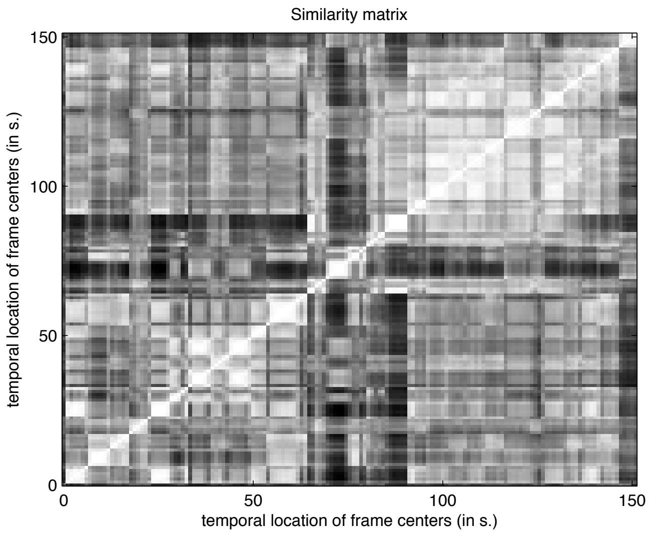
Motivation

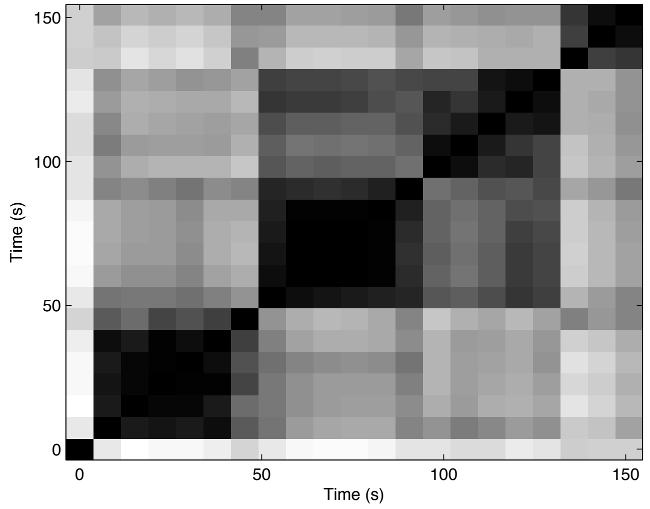
- Current approaches:
  - Signal processing front-ends
    - Time-frequency representations
    - Audio features
    - etc.

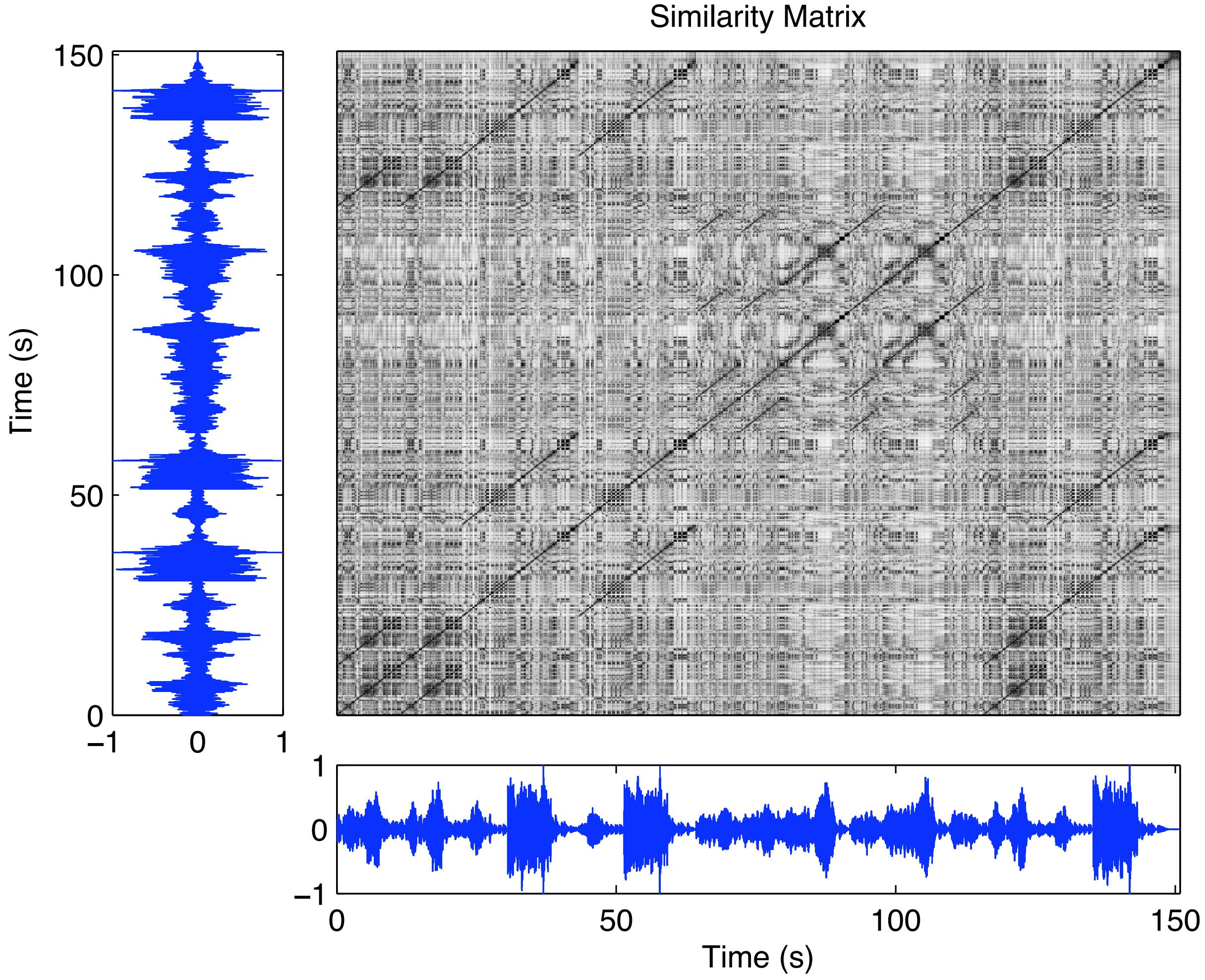


Motivation

- Current approaches:
  - Self-Similarity Analysis
    - Music Information Retrieval community
    - Computing distances between every occurrence of signal
    - Non-realtime
    - Far from being effortless
    - Requires further processing
      - Using kernels:







    - WORSE: Bag of features methods (very common in MIR)
      - Loosing all temporal ordering! -> ill-posed

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 equivalence 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

Information Geometry

Intuition

- Compute on the geometry constituted by audio streams:
  - *Points* are probability distributions $p(x, \xi)$, representing continuous audio
  - *Distance* between two points is some measure of *information* between them
  - 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

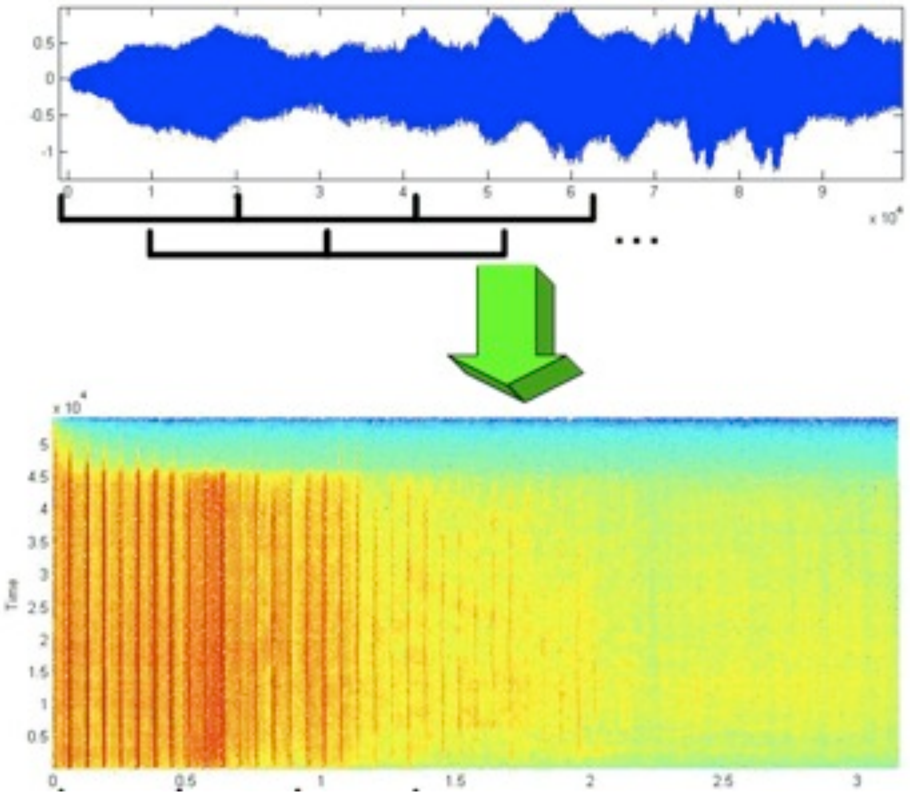
Information Geometry

- The beauty of Information Geometry
  - Unifying common engineering notions within one framework and with geometric intuition
    - Maximum Likelihood
    - Maximum Entropy
    - Projection on manifolds
    - All equivalent!
  - Computational gain: Thanks to duality (common convex optimizations will become linear in the Legendre dual space)
  - Complex Math but Simple Algorithmics
- Construction:
  - Bottom-up: Define points, connections, distances, transports, ... to reach metrics (e.g. Amari, X. Pennec)
  - Top-down: Automatically derive the geometry (with its tools) from represented data with prior engineering assumptions

Music Information Geometry

One Possible Music Information Geometry

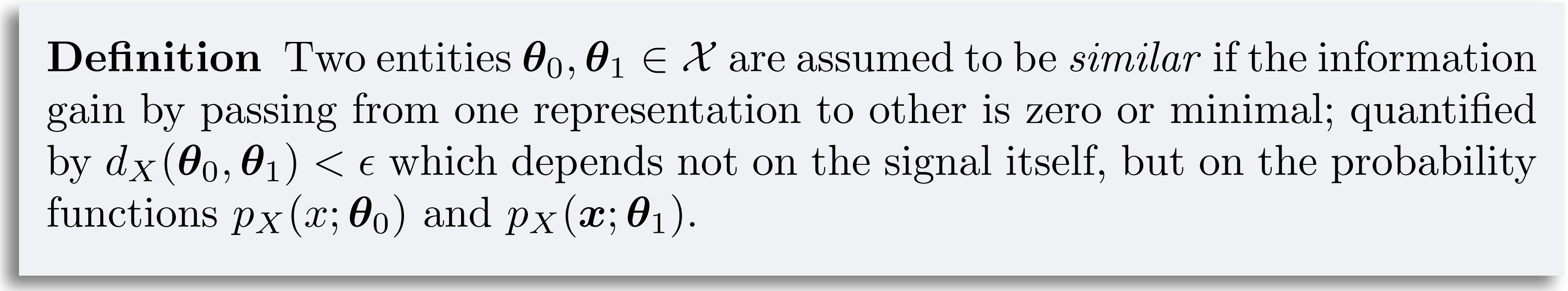
- Points = time domain windows of audio signal X_t, represented by their *frequency distributions* $S_t(\omega)$
  - Arriving incrementally / in real time
  - Corresponding to normalized log-scale Fourier transform amplitudes
  - Mapped to *Multinomial points* in the information geometry (one-to-one)
    - 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:



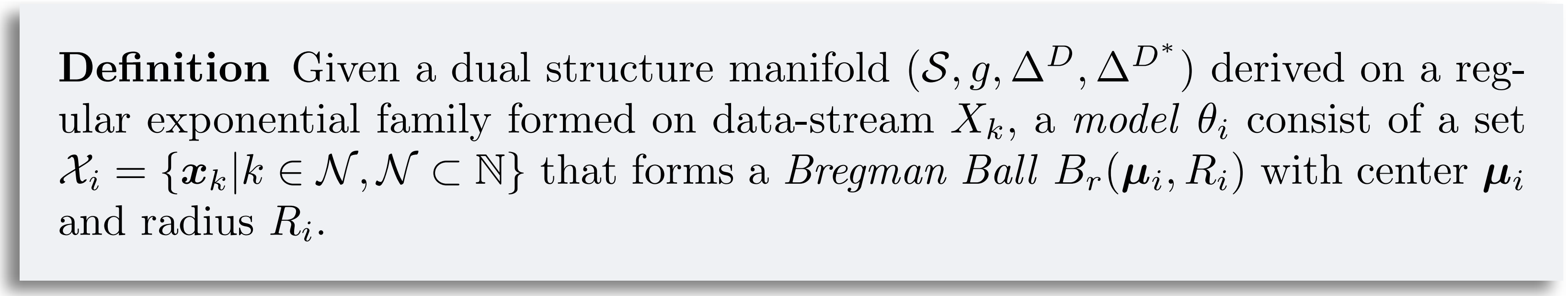
Definition Two entities $\boldsymbol{\theta}_0, \boldsymbol{\theta}_1 \in \mathcal{X}$ are assumed to be *similar* if the information gain by passing from one representation to other is zero or minimal; quantified by $d_X(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1) < \epsilon$ which depends not on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

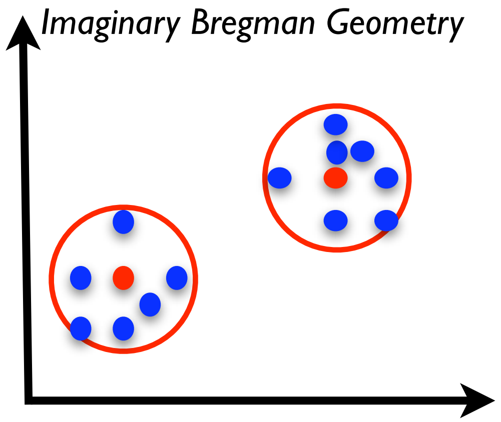
Music Information Geometry

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.



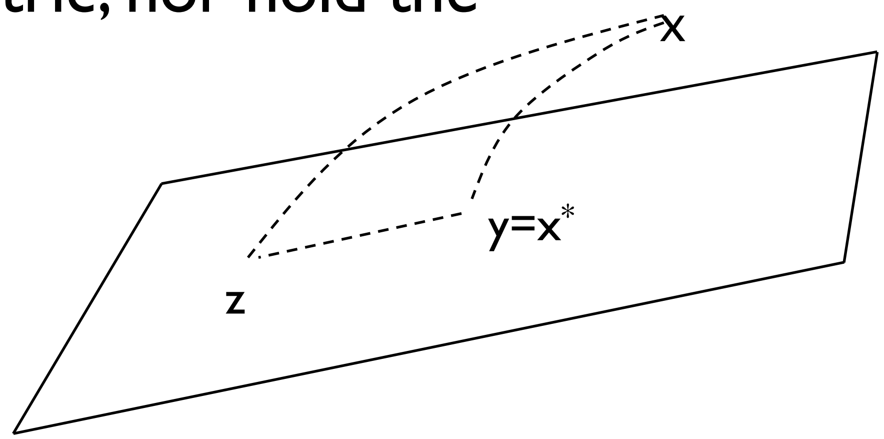
Music Information Geometry

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 equality 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!)`

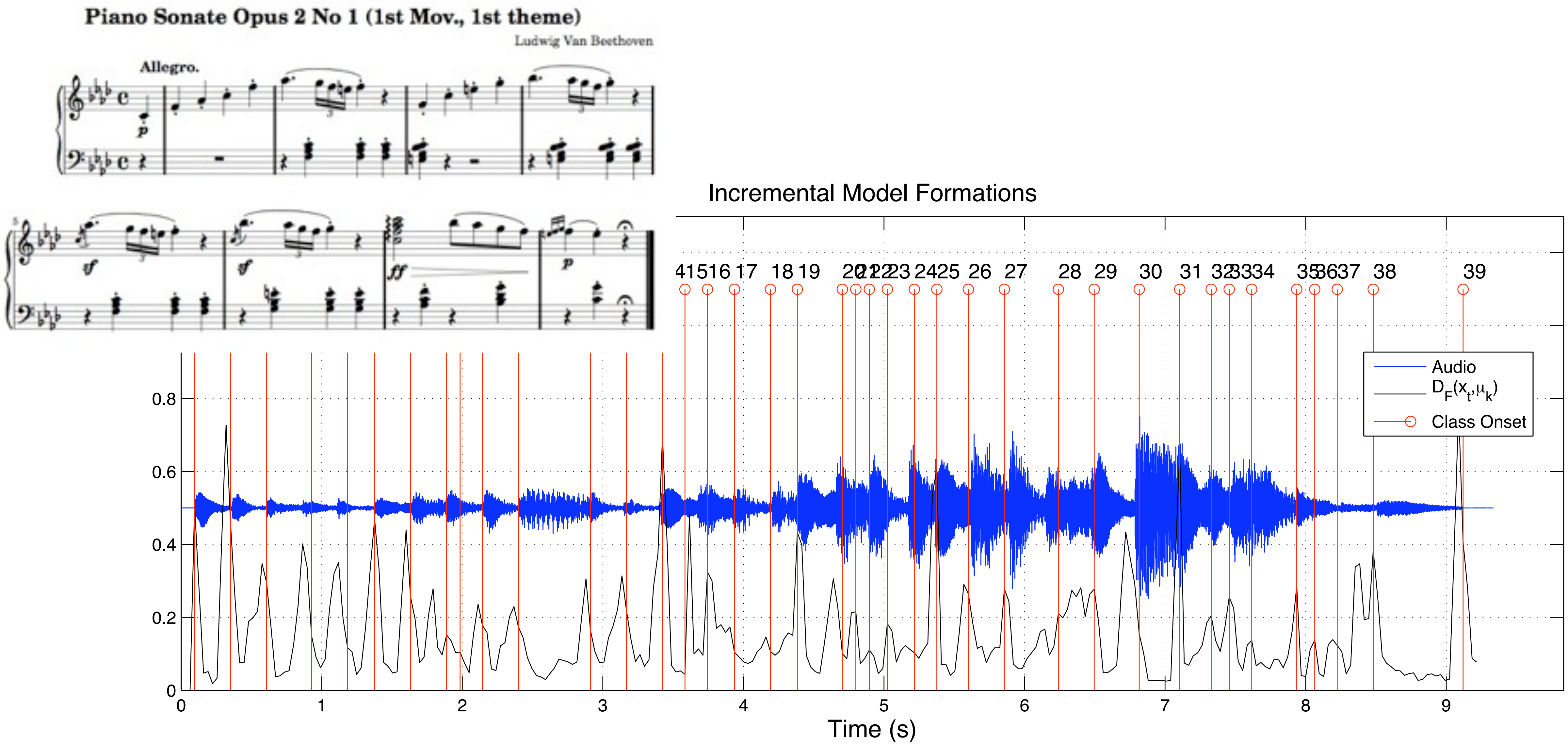
  c. Symmetrize Bregman divergence using a max. likelihood formulation!

We can approach both notions of symmetry and triangular inequality.

Music Information Geometry

Incremental Segmentation

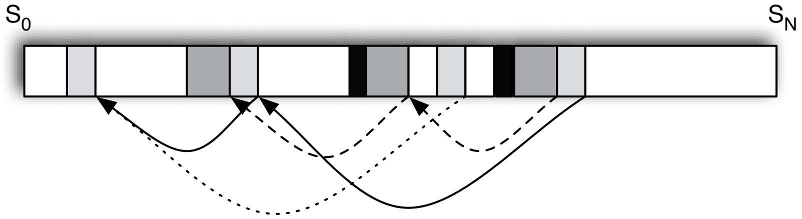
- Sample Result: Beethoven's first piano sonata, first movement
  - performed by Friedrich Gulda (1958)

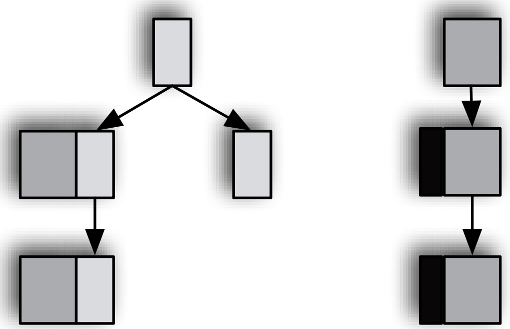


Methods of Information Access

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



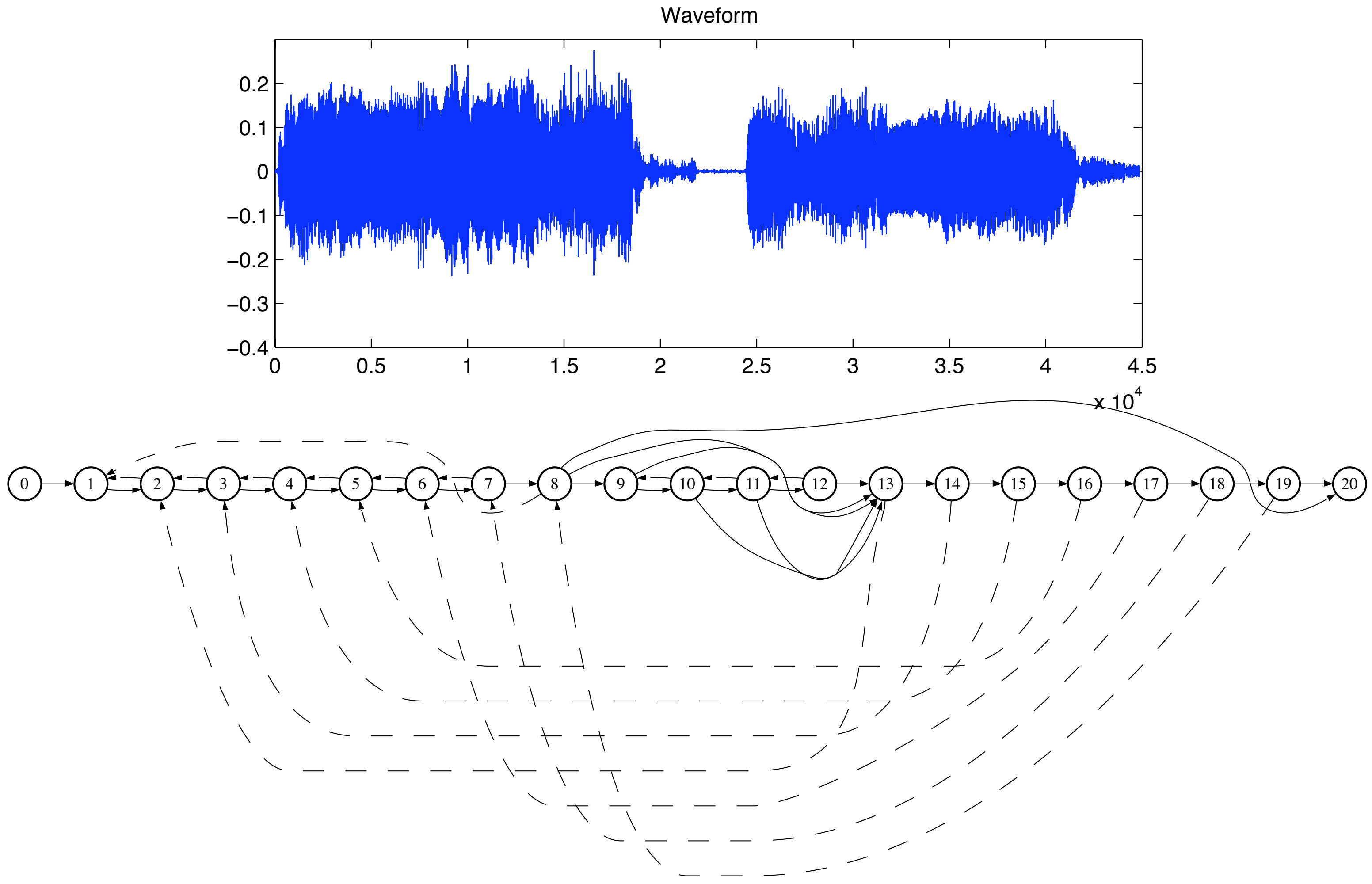


  - Provides *forest of suffix tree structures*
- *The beauty of MIG*
  - Keep the algorithm, replace symbols by *models* or *points* and equivalence by *similarity* in a music information geometry!

⇨ **Audio Oracle**

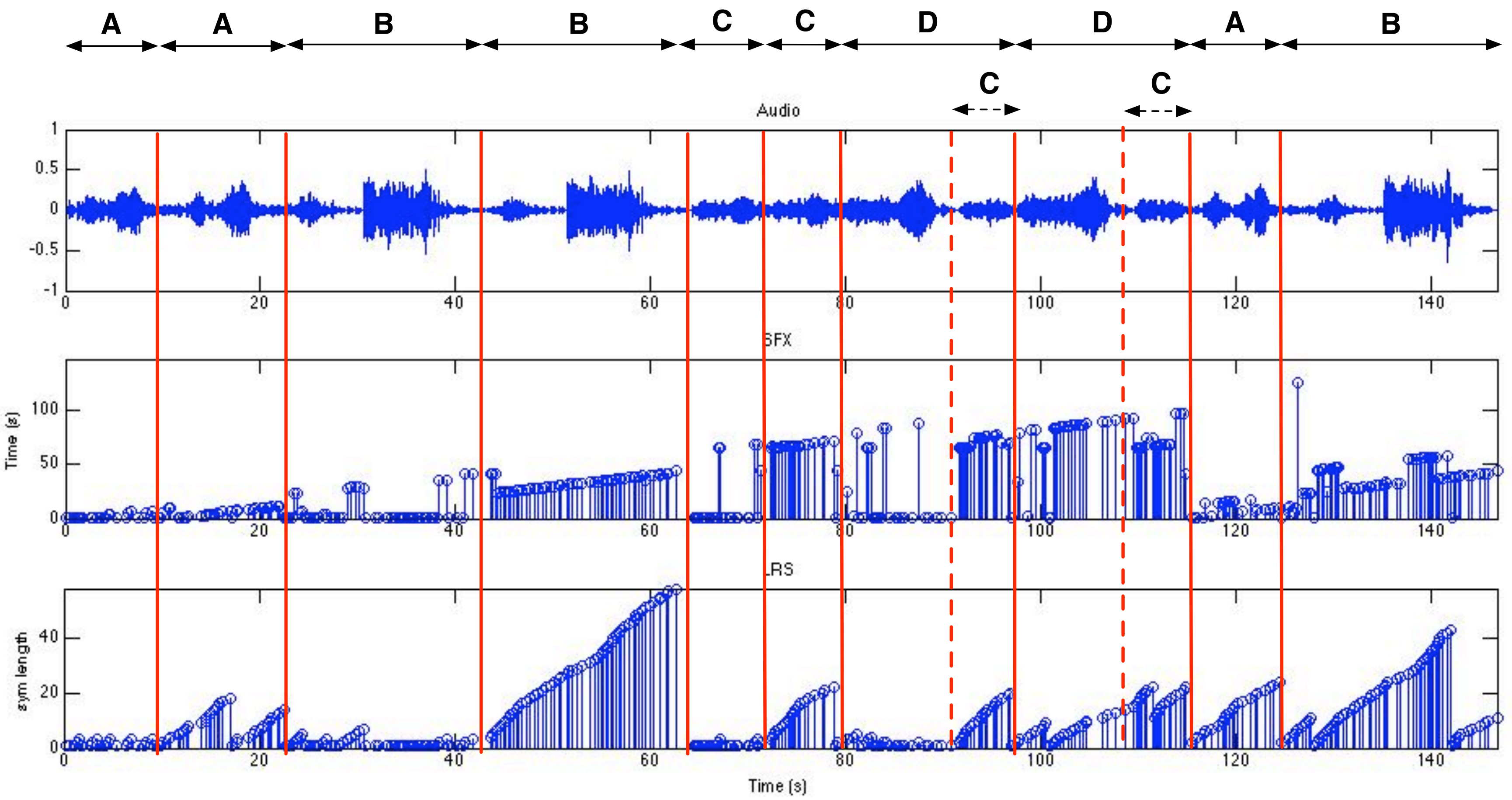
Methods of Information Access

- 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*

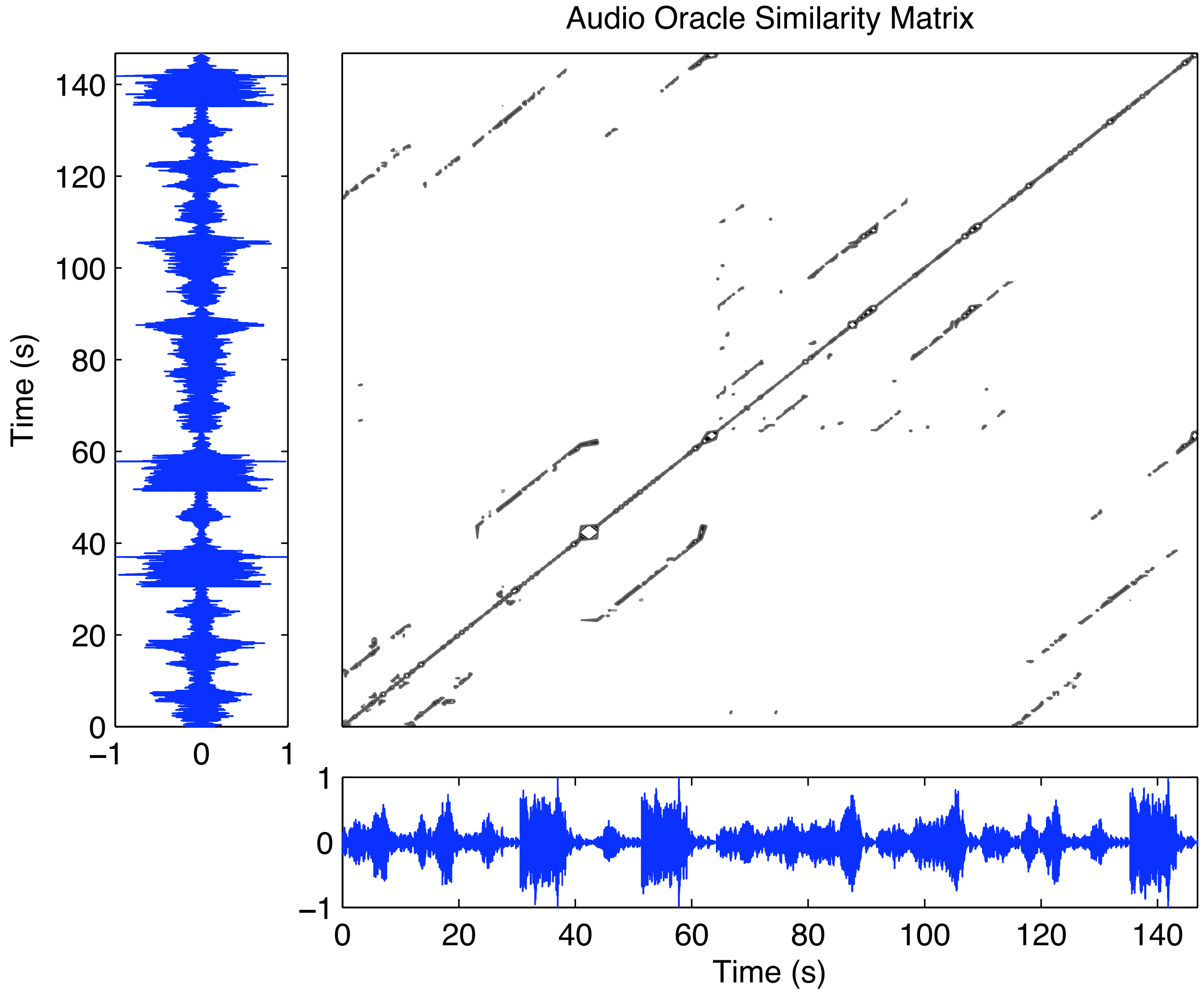


Recall Structure

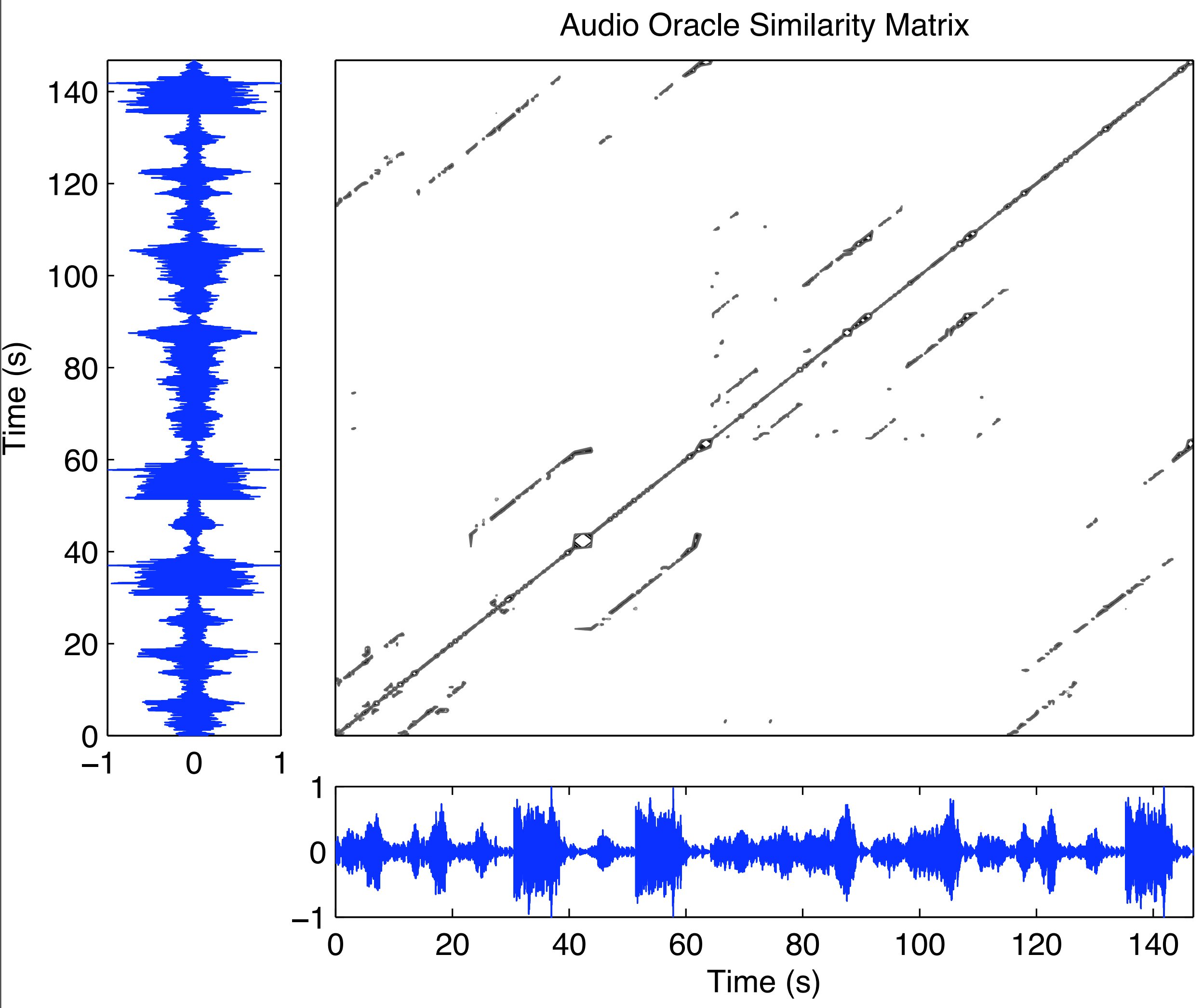
Recall Length

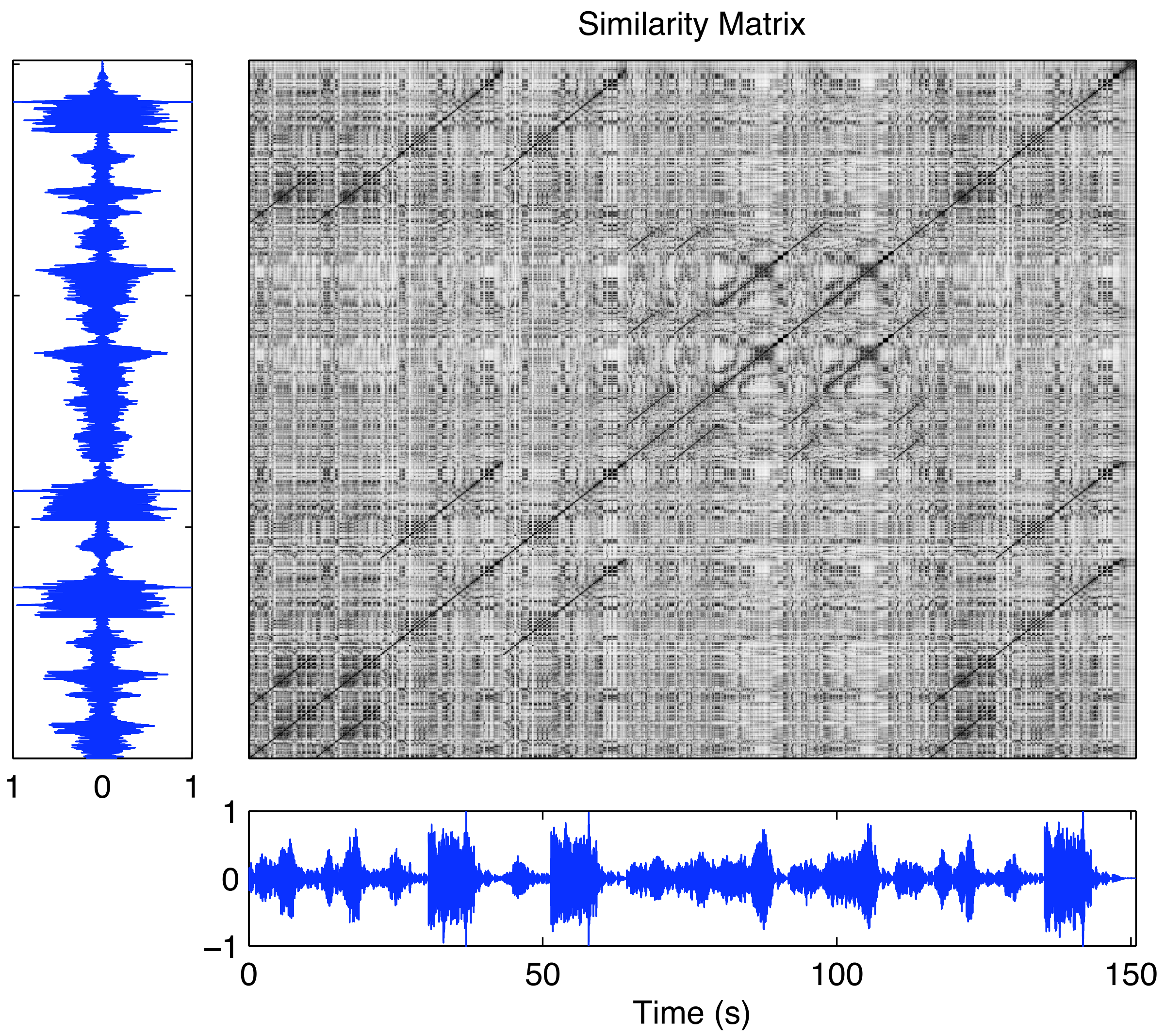
Methods of Information Access

- 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


Similarity Matrix

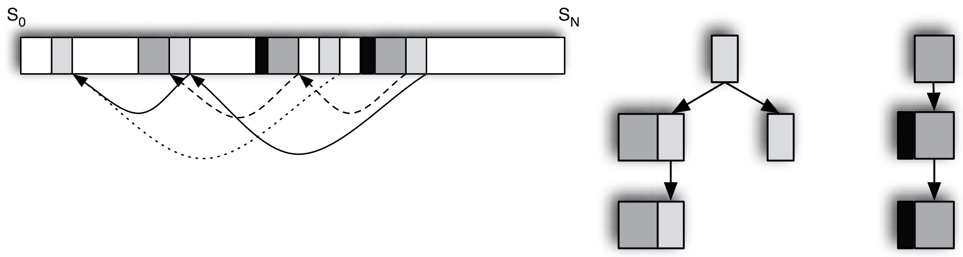
More examples on: http://imtr.ircam.fr/imtr/Audio_Oracle

Methods of Information Access

Fast Information Retrieval

- 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!

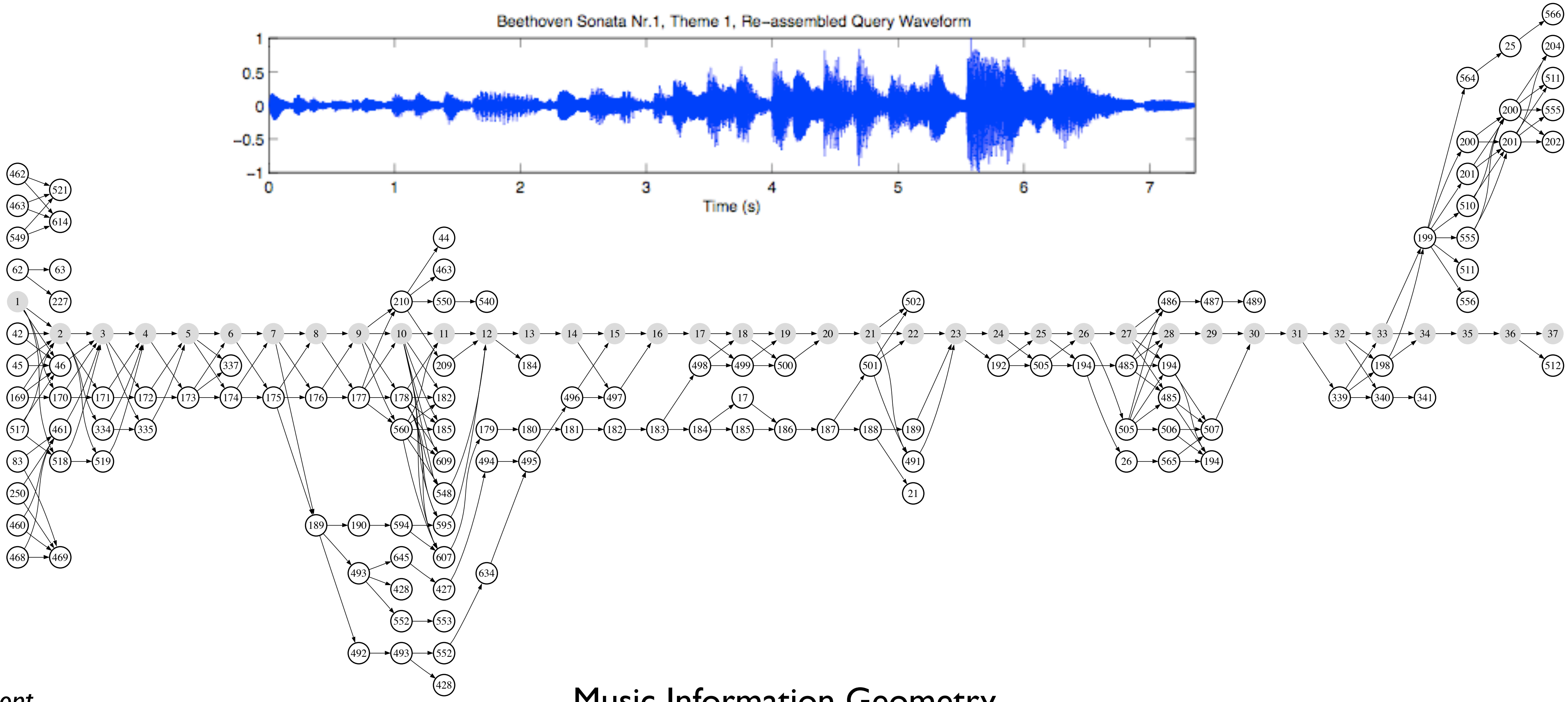
⇨ **Guidage**



Methods of Information Access

Guidage Results

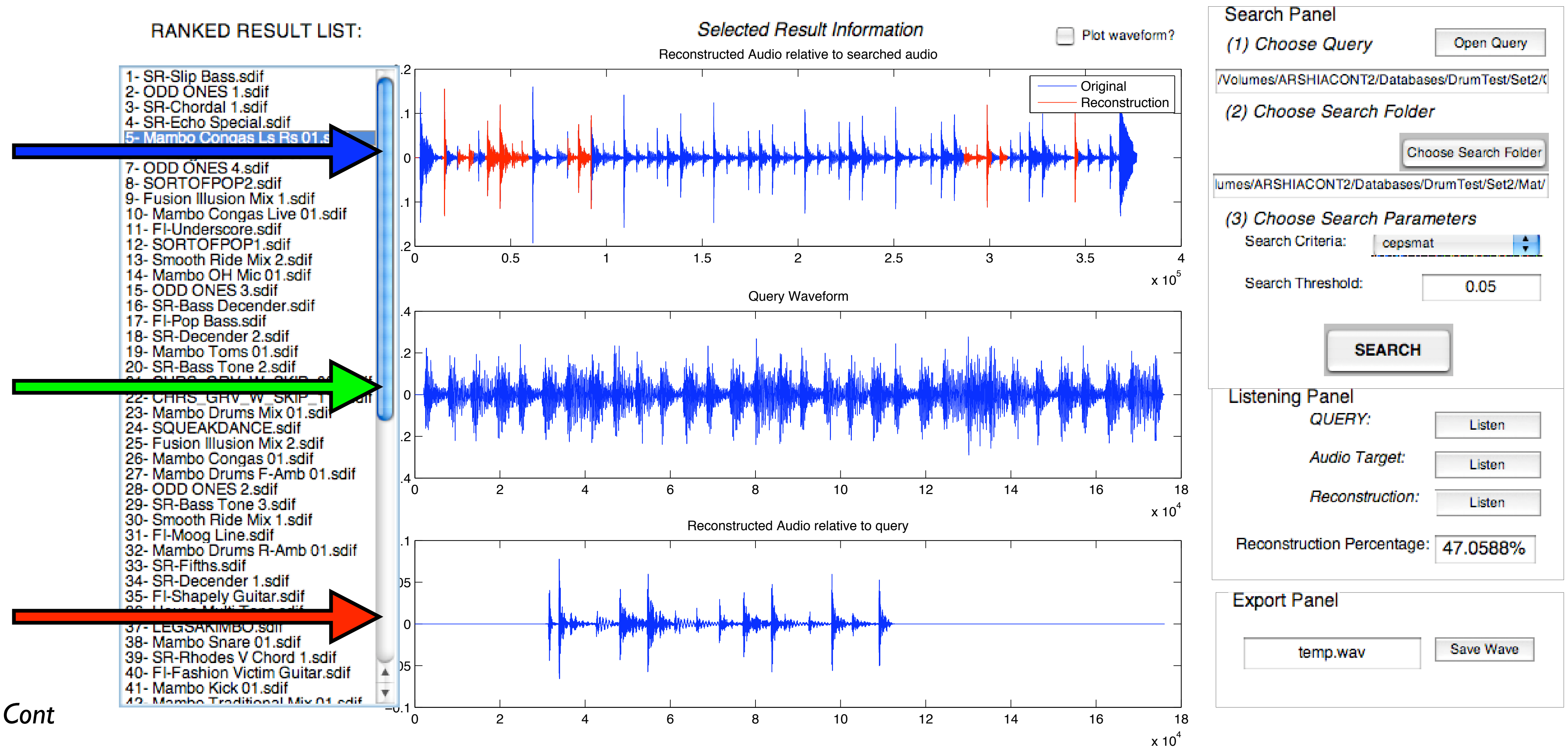
- Self-Similarity test
  - Task: Search for the *first theme* of the first Beethoven's sonata in the entire sonata.
    - Query: Audio of the first theme
    - Target: The entire first sonata's Audio Oracle (650 states)



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



RANKED RESULT LIST:

1- SR-Slip Bass.sdif
2- ODD ONES 1.sdif
3- SR-Chordal 1.sdif
4- SR-Echo Special.sdif
5- Mambo Congas Ls Rs 01.s
7- ODD ONES 4.sdif
8- SORTOFPOP2.sdif
9- Fusion Illusion Mix 1.sdif
10- Mambo Congas Live 01.sdif
11- FI-Underscore.sdif
12- SORTOFPOP1.sdif
13- Smooth Ride Mix 2.sdif
14- Mambo OH Mic 01.sdif
15- ODD ONES 3.sdif
16- SR-Bass Decender.sdif
17- FI-Pop Bass.sdif
18- SR-Decender 2.sdif
19- Mambo Toms 01.sdif
20- SR-Bass Tone 2.sdif
22- CHRS_GRV_W_SKIP_1 sdif
23- Mambo Drums Mix 01.sdif
24- SQUEAKDANCE.sdif
25- Fusion Illusion Mix 2.sdif
26- Mambo Congas 01.sdif
27- Mambo Drums F-Amb 01.sdif
28- ODD ONES 2.sdif
29- SR-Bass Tone 3.sdif
30- Smooth Ride Mix 1.sdif
31- FI-Moog Line.sdif
32- Mambo Drums R-Amb 01.sdif
33- SR-Fifths.sdif
34- SR-Decender 1.sdif
35- FI-Shapely Guitar.sdif
37- LEGSAKIMBO.sdif
38- Mambo Snare 01.sdif
39- SR-Rhodes V Chord 1.sdif
40- FI-Fashion Victim Guitar.sdif
41- Mambo Kick 01.sdif
42- Mambo Traditional Mix 01.sdif

Selected Result Information

Search Panel

(1) Choose Query

/Volumes/ARSHIACONT2/Databases/DrumTest/Set2/(

(2) Choose Search Folder

lumes/ARSHIACONT2/Databases/DrumTest/Set2/Mat/

(3) Choose Search Parameters

Search Criteria: cepsmat

Search Threshold: 0.05

Listening Panel

Reconstruction Percentage: 47.0588%

Export Panel

temp.wav

and further..

- This is just the beginning!

  Arshia Cont, Shlomo Dubnov and Gérard Assayag. On the Information Geometry of Audio Streams with Applications to Similarity Computing. IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, Nr. 4, Pp. 837-846, May 2011.

- Study the behavior of audio streams on Riemannian information manifolds...
  - Thèse de Arnaud Dessein

- “Groupe de Recherche” sur la géométrie de l’information
  - Ircam, École Polytechnique, THALES, École des mines, Univ. de Poitiers, etc.
  - Séminaire Léon Brillouin

    http://www.informationgeometry.org/Seminar/seminarBrillouin.html

  - Session spécial GRETSI 2011, Bordeaux.

ircam
Centre
Pompidou

Further Readings:

Arshia Cont, Shlomo Dubnov and Gérard Assayag. *On the Information Geometry of Audio Streams with Applications to Similarity Computing*. IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, Nr. 4, May 2011. (in press)

Arshia Cont. *A coupled duration-focused architecture for realtime music to score alignment.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32(6), Pp. 974-987, June 2010.

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.

http://repmus.ircam.fr/music-information-geometry

QUESTIONS?

Merci!
cont@ircam.fr