Statistical segmentation of audio streams in real-time within the framework of information geometry

Arnaud Dessein, Arshia Cont
Institut de Recherche et Coordination Acoustique/Musique

September 7th 2011
Outline

1. Introduction
2. Information geometry
3. Proposed system
4. Obtained results
5. Conclusion
Outline

1 Introduction
   - Context
   - Motivations
   - Contributions

2 Information geometry

3 Proposed system

4 Obtained results

5 Conclusion
What is audio segmentation?

Audio segmentation
Partitioning a sound signal into continuous and homogeneous temporal regions, called segments, which exhibit inhomogeneities with adjacent regions.

Figure: Audio segmentation.
What is audio segmentation?

Audio segmentation

Partitioning a sound signal into continuous and homogeneous temporal regions, called segments, which exhibit inhomogeneities with adjacent regions.

- Temporality:
  - Continuity along the time dimension.
  - Causality principle.
  - On-line or real-time setups.

Figure: Audio segmentation.
What is audio segmentation?

Audio segmentation

Partitioning a sound signal into continuous and homogeneous temporal regions, called segments, which exhibit inhomogeneities with adjacent regions.

- Temporality:
  - Continuity along the time dimension.
  - Causality principle.
  - On-line or real-time setups.
- Homogeneity:
  - Intrinsic homogeneity.
  - Inhomogeneity with contiguous segments.
  - Criterion for homogeneity.

Figure: Audio segmentation.
What is audio segmentation?

Audio segmentation

Partitioning a sound signal into continuous and homogeneous temporal regions, called segments, which exhibit inhomogeneities with adjacent regions.

- Temporality:
  - Continuity along the time dimension.
  - Causality principle.
  - On-line or real-time setups.

- Homogeneity:
  - Intrinsic homogeneity.
  - Inhomogeneity with contiguous segments.
  - Criterion for homogeneity.

- Examples include speech, music, radio broadcasts

Figure: Audio segmentation.
What do we need?

- **In general:**
  - High level criteria and automatic classification (e.g., speakers, instruments, speech/non speech, voiced/unvoiced, speech/music).
  - Drawbacks: relies on an automatic classification which is unfortunately not infallible, hypothesis of classes, data for training.
What do we need?

- In general:
  - High level criteria and automatic classification (e.g., speakers, instruments, speech/non speech, voiced/unvoiced, speech/music).
  - Drawbacks: relies on an automatic classification which is unfortunately not infallible, hypothesis of classes, data for training.

- Other approaches:
  - Low level criteria and descriptors on the signal (e.g., onset detection, noise detection).
  - High level criteria but no a priori assumption on the existence of classes.
  - In particular for speaker segmentation
    [Siegler et al., 1997, Tritschler & Gopinath, 1999, Delacourt & Wellekens, 2000, Kotti et al., 2008, Grasic et al., 2010].
  - Computation of a distance between successive frames, or a statistic on the hypothesis of a change point.
What do we need?

- In general:
  - High level criteria and automatic classification (e.g., speakers, instruments, speech/non speech, voiced/unvoiced, speech/music).
  - Drawbacks: relies on an automatic classification which is unfortunately not infallible, hypothesis of classes, data for training.

- Other approaches:
  - Low level criteria and descriptors on the signal (e.g., onset detection, noise detection).
  - High level criteria but no a priori assumption on the existence of classes.
  - In particular for speaker segmentation
    [Siegler et al., 1997, Tritschler & Gopinath, 1999, Delacourt & Wellekens, 2000, Kotti et al., 2008, Grasic et al., 2010].
  - Computation of a distance between successive frames, or a statistic on the hypothesis of a change point.

- Our approach:
  - Real-time constraints.
  - Modularity with various types of signals and criteria.
  - No a priori assumption on the existence of classes.
  - Control on the variation of the information content.
What do we propose?

- Real-time modular segmentation scheme.

Figure: Audio segmentation in the framework of information geometry.
What do we propose?

- Real-time modular segmentation scheme.
- Framework of information geometry for exponential families.

Figure: Audio segmentation in the framework of information geometry.
What do we propose?

- Real-time modular segmentation scheme.
- Framework of information geometry for exponential families.
- Statistical grounds through sequential generalized likelihood ratio tests.

Figure: Audio segmentation in the framework of information geometry.
What do we propose?

- Real-time modular segmentation scheme.
- Framework of information geometry for exponential families.
- Statistical grounds through sequential generalized likelihood ratio tests.
- Geometrical interpretation through dually flat Bregman geometry.

**Figure:** Audio segmentation in the framework of information geometry.
What do we propose?

- Real-time modular segmentation scheme.
- Framework of information geometry for exponential families.
- Statistical grounds through sequential generalized likelihood ratio tests.
- Geometrical interpretation through dually flat Bregman geometry.
- Link between distance and statistic-based methods in a unified framework.

**Figure:** Audio segmentation in the framework of information geometry.
What do we propose?

- Real-time modular segmentation scheme.
- Framework of information geometry for exponential families.
- Statistical grounds through sequential generalized likelihood ratio tests.
- Geometrical interpretation through dually flat Bregman geometry.
- Link between distance and statistic-based methods in a unified framework.
- Quantization of each segment with an information geometric prototype.

**Figure:** Audio segmentation in the framework of information geometry.
Outline

1. Introduction
2. Information geometry
   - Background
   - Exponential families
3. Proposed system
4. Obtained results
5. Conclusion
What is information geometry?

Statistical differentiable manifold.

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

**Statistical differentiable manifold.**

Under certain assumptions, a parametric statistical model $S = \{p_\xi : \xi \in \Xi\}$ 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 $\xi = [\mu, \sigma^2] \in \Xi = \mathbb{R} \times \mathbb{R}_{++}$. 
What is information geometry?

**Statistical differentiable manifold.**

Under certain assumptions, a parametric statistical model \( S = \{ p_\xi : \xi \in \Xi \} \) 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 \( S \): \( g_{ij}(\xi) = E_\xi[\partial_i \log p_\xi \partial_j \log p_\xi] \).
What is information geometry?

Statistical differentiable manifold.

Under certain assumptions, a parametric statistical model $S = \{p_\xi : \xi \in \Xi\}$ 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 $S$: $g_{ij}(\xi) = E_\xi[\partial_i \log p_\xi \partial_j \log p_\xi]$.


Under certain assumptions, the $\alpha$-connections $\nabla^{(\alpha)}$ for $\alpha \in \mathbb{R}$ are the unique affine connections on $S$: $\nabla^{(\alpha)} \partial_j = \Gamma^{(\alpha)}_{ij,k}(\xi) \partial_k$ where

$$\Gamma^{(\alpha)}_{ij,k}(\xi) = E_\xi\left[ (\partial_i \partial_j \log p_\xi + \frac{1-\alpha}{2} \partial_i \log p_\xi \partial_j \log p_\xi) (\partial_k \log p_\xi) \right].$$
How to use information geometry from a computational viewpoint?

**Exponential family** [Darmois, 1935, Koopman, 1936, Pitman, 1936].

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

**Figure**: A taxonomy of probability measures [Nielsen & Garcia, 2009].
How to use information geometry from a computational viewpoint?

**Exponential family** [Darmois, 1935, Koopman, 1936, Pitman, 1936].

\[ p_\theta(x) = \exp (\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

- 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 \( \theta \) and expectation parameters \( \eta = \nabla F(\theta) \).
- Dually flat geometry, Hessian structure, generated by the potential \( F \) together with its conjugate potential \( F^* \) defined by the Legendre-Fenchel transform: \( F^*(\eta) = \sup_{\theta \in \Theta} \theta^T \eta - F(\theta) \), which verifies \( \nabla F^* = (\nabla F)^{-1} \) so that \( \theta = \nabla F^*(\eta) \).
- Generalizes the self-dual Euclidean geometry, with notably two canonically associated Bregman divergences \( B_F \) and \( B_{F^*} \) instead of the self-dual Euclidean distance, but also dual geodesics, a generalized Pythagorean theorem and dual projections.
How to use information geometry from a computational viewpoint?

**Exponential family** [Darmois, 1935, Koopman, 1936, Pitman, 1936].

\[ p_\theta(x) = \exp(\theta^T T(x) - F(\theta) + C(x)) \text{ for all } x \in \mathcal{X}. \]

**Bregman divergence** [Bregman, 1967].

\[ B_G(\theta, \theta') = G(\theta) - G(\theta') - (\theta - \theta')^T \nabla G(\theta'). \]

- Canonical divergences of dually flat spaces, “bijection” with exponential families [Amari & Nagaoka, 2000, Banerjee et al., 2005]:
  \[ D_{KL}(p_\xi \parallel p_{\xi'}) = B_F(\theta' \parallel \theta) = B_F^*(\eta \parallel \eta'). \]
- No symmetry nor triangular inequality in general, but an information-theoretic interpretation.
  - Centroid computation and hard clustering (k-means).
  - Parameter estimation and soft clustering (expectation-maximization).
  - Proximity queries in ball trees (nearest-neighbors and range search).
Outline

1. Introduction
2. Information geometry
3. Proposed system
   - General architecture
   - Sound descriptors modeling
   - Statistical segmentation
4. Obtained results
5. Conclusion
How to segment audio streams in the framework of information geometry?

Scheme:

1. Represent the incoming audio stream with short-time sound descriptors $d_j$.
2. Model these descriptors with probability distributions $p_{\theta_j}$ from a given exponential family.
3. Use the computational tools from information geometry to segment these distributions.

Figure: Segmentation at time $t$.

Figure: Schema of the general architecture of the system.
How to segment audio streams in the framework of information geometry?

Scheme:

1. Represent the incoming audio stream with short-time sound descriptors $d_j$.
2. Model these descriptors with probability distributions $p_{\theta_j}$ from a given exponential family.
3. Use the computational tools from information geometry to segment these distributions.

In particular, it allows to define the notion of similarity in an information setup through divergences.

Figure: Segmentation at time $t$. 

Figure: Schema of the general architecture of the system.
How to model sounds?

- Computation of a sound descriptor $d_j$:
  - Fourier or constant-Q transforms for information on the spectral content.
  - Mel-frequency cepstral coefficients for information on the timbre.
  - Many other possibilities.

**Figure:** Sound descriptors modeling.
How to model sounds?

- **Computation of a sound descriptor** $d_j$:
  - Fourier or constant-Q transforms for information on the spectral content.
  - Mel-frequency cepstral coefficients for information on the timbre.
  - Many other possibilities.
- **Modeling with a probability distribution** $p_{\theta_j}$ from an exponential family:
  - Categorical distributions.
  - Multivariate Gaussian distributions.
  - Many other possibilities.

![Waveform and Spectrogram](image)

**Figure**: Sound descriptors modeling.
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     \[ B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \parallel \hat{\theta}) \leq r \}. \]
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     \[
     B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \parallel \hat{\theta}) \leq r \}.
     \]
  2. Update the Bregman ball centroid:
     \[
     \hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} B_F(\theta_j \parallel \theta) = \frac{1}{n} \sum_{j=1}^{n} \theta_j.
     \]

Figure: Segmentation at time $t$. 

arnaud.dessein@ircam.fr September 7th 2011 Presentation GRETSI 2011 13/21
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     \[
     B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \| \hat{\theta}) \leq r \}.
     \]
  2. Update the Bregman ball centroid:
     \[
     \hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} B_F(\theta_j \| \theta) = \frac{1}{n} \sum_{j=1}^{n} \theta_j.
     \]
  3. Update the Bregman ball radius:
     \[
     r = \max_{j \in \{1, \ldots, n\}} B_F(\theta_j \| \hat{\theta}).
     \]

Figure: Segmentation at time $t$. 
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     $$B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \parallel \hat{\theta}) \leq r \}.$$  
  2. Update the Bregman ball centroid:
     $$\hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} B_F(\theta_j \parallel \theta) = \frac{1}{n} \sum_{j=1}^{n} \theta_j.$$  
  3. Update the Bregman ball radius:
     $$r = \max_{j \in \{1, \ldots, n\}} B_F(\theta_j \parallel \hat{\theta}).$$  
  4. Segment if the radius becomes greater than a threshold $\gamma$.

Figure: Segmentation at time $t$. 

Drawbacks: robustness issues, in particular for observed points, kind of one-sample-estimation after change.

New scheme: sequential generalized likelihood ratio tests inspired by CuSum.
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     
     $$B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \parallel \hat{\theta}) \leq r \}.$$
  2. Update the Bregman ball centroid:
     
     $$\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} B_F(\theta_j \parallel \theta) = \frac{1}{n} \sum_{j=1}^{n} \theta_j.$$
  3. Update the Bregman ball radius:
     
     $$r = \max_{j \in \{1, \ldots, n\}} B_F(\theta_j \parallel \hat{\theta}).$$
  4. Segment if the radius becomes greater than a threshold $\gamma$.

- Drawbacks: robustness issues, in particular for observed points, kind of one-sample-estimation after change.
How to segment audio streams? (1)

- Previous scheme:
  1. Aggregate incoming points $\theta_j$ in a left/right Bregman ball:
     \[ B_F(\hat{\theta}, r) = \{ \theta \in \Theta : B_F(\theta \parallel \hat{\theta}) \leq r \} \].
  2. Update the Bregman ball centroid:
     \[ \hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{j=1}^{n} B_F(\theta_j \parallel \theta) = \frac{1}{n} \sum_{j=1}^{n} \theta_j. \]
  3. Update the Bregman ball radius:
     \[ r = \max_{j \in \{1, \ldots, n\}} B_F(\theta_j \parallel \hat{\theta}). \]
  4. Segment if the radius becomes greater than a threshold $\gamma$.
- Drawbacks: robustness issues, in particular for observed points, kind of one-sample-estimation after change.
- New scheme: sequential generalized likelihood ratio tests inspired by CuSum.
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  
  \( H_0: x_1, \ldots, x_n \sim p_{\theta_0}. \)
  
  \( H_1^i: x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \)
  
  (Generalized) likelihood ratio test: \( LR^i = -2 \log \frac{p(x|H_0)}{p(x|H_1^i)} > \lambda. \)
  
  - Known parameters before and after change.

\[
\frac{1}{2} LR^i = \sum_{j=i+1}^{n} \log \frac{p_{\theta_1}(x_j)}{p_{\theta_0}(x_j)}.
\]
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \[ H_0: x_1, \ldots, x_n \sim p_{\theta_0}. \]
  \[ H^i_1: x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  (Generalized) likelihood ratio test: \[ LR^i = -2 \log \frac{p(x|H_0)}{p(x|H^i_1)} > \lambda. \]
  - Known parameters before and after change.
  - Unknown parameter after change: \( \hat{\theta}_1^i \).

\[ \frac{1}{2} LR^i = \sum_{j=i+1}^{n} \log \frac{p_{\theta_1}(x_j)}{p_{\theta_0}(x_j)}. \]
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \[ H_0 : x_1, \ldots, x_n \sim p_{\theta_0}. \]
  \[ H_{1i} : x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]

  (Generalized) likelihood ratio test: \[ LR_i = -2 \log \frac{p(x|H_0)}{p(x|H_{1i})} > \lambda. \]

  - Known parameters before and after change.
  - Unknown parameter after change: \( \hat{\theta}_1^i \).
  - Unknown parameters after and before change: \( \hat{\theta}_0^i, \hat{\theta}_1^i \).

  \[ \frac{1}{2} LR_i = \sum_{j=i+1}^{n} \log \frac{p_{\theta_1}(x_j)}{p_{\theta_0}(x_j)}. \]
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \[ H_0: x_1, \ldots, x_n \sim p_{\theta_0}. \]
  \[ H_1^i: x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  (Generalized) likelihood ratio test: \( LR^i = -2 \log \frac{p(x|H_0)}{p(x|H_1^i)} > \lambda. \)

- Proposed change detection scheme.
  \[ H_0: x_1, \ldots, x_n \sim p_{\theta_0}'. \]
  \[ H_1^i: x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \( H_0 : x_1, \ldots, x_n \sim p_{\theta_0} \).
  \( H_{i1} : x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1} \).
  (Generalized) likelihood ratio test: \( LR^i = -2 \log \frac{p(x|H_0)}{p(x|H_{i1})} > \lambda \).

- Proposed change detection scheme.
  \( H_0 : x_1, \ldots, x_n \sim p_{\theta'_0} \).
  \( H_{i1} : x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1} \).

  - Unknown parameters after and before change: \( \hat{\theta}'_0, \hat{\theta}_0, \hat{\theta}_1 \).
  \[
  \frac{1}{2} LR^i = i \left( B_F (\hat{\theta}'_0 \parallel \hat{\theta}'_0 \text{mle}) - B_F (\hat{\theta}_0 \parallel \hat{\theta}_0 \text{mle}) \right) + (n - i) \left( B_F (\hat{\theta}'_0 \parallel \hat{\theta}_1 \text{mle}) - B_F (\hat{\theta}_1 \parallel \hat{\theta}_1 \text{mle}) \right).
  \]
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \[ H_0: x_1, \ldots, x_n \sim p_{\theta_0}. \]
  \[ H'_i: x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  (Generalized) likelihood ratio test: \( LR^i = -2 \log \frac{p(x|H_0)}{p(x|H'_i)} > \lambda. \)

- Proposed change detection scheme.
  \[ H_0: x_1, \ldots, x_n \sim p_{\theta'_0}. \]
  \[ H'_1: x_1, \ldots, x_i \sim p_{\theta'_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  - Unknown parameters after and before change: \( \hat{\theta}'_0, \hat{\theta}_0, \hat{\theta}'_1. \)
    \[
    \frac{1}{2} LR^i = i \left( B_F(\hat{\theta}'_0 \parallel \hat{\theta}'_{0 \text{mle}}) - B_F(\hat{\theta}_0 \parallel \hat{\theta}'_{0 \text{mle}}) \right) + (n-i) \left( B_F(\hat{\theta}'_0 \parallel \hat{\theta}'_{1 \text{mle}}) - B_F(\hat{\theta}'_1 \parallel \hat{\theta}'_{1 \text{mle}}) \right).
    \]

- On-line change detection.
  - Sequential generalized likelihood ratio tests, growing window.
  - Heuristics: minimum/maximum window size, sliding/growing factor.
How to segment audio streams? (2)

- Problems of existing CuSum change detection [Basseville & Nikiforov, 1993].
  \[ H_0: \ x_1, \ldots, x_n \sim p_{\theta_0}. \]
  \[ H_{i1}^i: \ x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  (Generalized) likelihood ratio test: \( LR^i = -2 \log \frac{p(x|H_0)}{p(x|H_{i1}^i)} > \lambda. \)

- Proposed change detection scheme.
  \[ H_0: \ x_1, \ldots, x_n \sim p_{\theta'_0}. \]
  \[ H_{i1}^i: \ x_1, \ldots, x_i \sim p_{\theta_0}, \text{ and } x_{i+1}, \ldots, x_n \sim p_{\theta_1}. \]
  - Unknown parameters after and before change: \( \hat{\theta}'_0, \hat{\theta}'_0, \hat{\theta}'_1. \)
  \[
  \frac{1}{2} LR^i = i \left( \mathcal{B}_F(\hat{\theta}'_0 \parallel \hat{\theta}'_{0, \text{mle}}) - \mathcal{B}_F(\hat{\theta}'_0 \parallel \hat{\theta}'_{0, \text{mle}}) \right) + (n-i) \left( \mathcal{B}_F(\hat{\theta}'_0 \parallel \hat{\theta}'_{1, \text{mle}}) - \mathcal{B}_F(\hat{\theta}'_1 \parallel \hat{\theta}'_{1, \text{mle}}) \right).
  \]

- On-line change detection.
  - Sequential generalized likelihood ratio tests, growing window.
  - Heuristics: minimum/maximum window size, sliding/growing factor.
  - Here no heuristic, computationally efficient updates with the maximum likelihood estimator, incremental scheme.
  \[
  \frac{1}{2} LR^i = i F^*(\hat{\eta}'_0) + (n-i) F^*(\hat{\eta}'_1) - n F^*(\hat{\eta}'_0).
  \]
Introduction

Information geometry

Proposed system

Obtained results
  - Synthetic data
  - Well-log data
  - Speaker segmentation
  - Music segmentation

Conclusion
Synthetic data

Figure: Segmentation of synthetic data.
Figure: Segmentation of well-log data.
Speaker segmentation

**Figure:** Segmentation of a speech fragment in speakers.
Music segmentation

Figure: Segmentation of a polyphonic musical excerpt.
Music segmentation

Figure: Segmentation of a polyphonic musical excerpt.
Music segmentation

**Figure:** Segmentation of a polyphonic musical excerpt.
Outline

1 Introduction
2 Information geometry
3 Proposed system
4 Obtained results
5 Conclusion
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.

- Many possibilities.
- Combinations of descriptors.
- Feature selection.
<table>
<thead>
<tr>
<th>Summary and perspectives.</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Representations.</td>
</tr>
<tr>
<td>• Descriptors modeling.</td>
</tr>
<tr>
<td>• Temporality of events.</td>
</tr>
<tr>
<td>• Applications.</td>
</tr>
</tbody>
</table>

- Exponential families and Bregman divergences, mixture models.
- Model selection.
- Other geometries, divergences, test statistics.
What we (don't) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.

- Assumption of quasi-stationarity.
- Non-stationarity modeling.
- Conditional distributions, linear/non-linear systems, time series.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.

- Evaluation on large datasets in audio and other domains.
- Onset detection, music segmentation, speaker segmentation, etc.
- First stage in real-time systems for polyphonic music transcription [Dessein et al., 2010], music similarity analysis [Cont et al., 2011], computer-assisted improvisation.
What we (don’t) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.

- Brillouin seminar.
- Tutorial on the applications of information geometry to audio signal processing at DAFx 2011.
What we (don't) have

Summary and perspectives.

- Representations.
- Descriptors modeling.
- Temporality of events.
- Applications.

- Brillouin seminar.
- Tutorial on the applications of information geometry to audio signal processing at DAFx 2011.
- Thank you for your attention! Questions?
- This work was supported by a doctoral fellowship from the UPMC (EDITE) and by a grant from the JST-CNRS ICT (Improving the VR Experience).
Bibliography I

Methods of information geometry, volume 191 of Translations of Mathematical Monographs.
American Mathematical Society.

Clustering with Bregman divergences.

Detection of abrupt changes: Theory and application.

The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming.
USSR Computational Mathematics and Mathematical Physics, 7(3), 200–217.

Fast nearest neighbor retrieval for Bregman divergences.

Efficient Bregman range search.

Chentsov, N. N. (1982).
Statistical decision rules and optimal inference, volume 53 of Translations of Mathematical Monographs.
American Mathematical Society.


On distributions admitting a sufficient statistic.

Computationally efficient and robust BIC-based speaker segmentation.

Statistical exponential families: A digest with flash cards.

Sided and symmetrized Bregman centroids.
*IEEE Transactions on Information Theory, 55*(6), 2882–2904.

Tailored Bregman ball trees for effective nearest neighbors.
In *Proceedings of the 25th European Workshop on Computational Geometry (EuroCG)* (pp. 29–32). Brussels, Belgium.

Blind change detection for audio segmentation.

Sufficient statistics and intrinsic accuracy.
Information and accuracy attainable in the estimation of statistical parameters.

Automatic segmentation, classification and clustering of broadcast news audio.
*In Proceedings of the DARPA Speech Recognition Workshop* (pp. 97–99). Chantilly, VA, USA.

Audio scene segmentation using multiple features, models and time scales.

Improved speaker segmentation and segments clustering using the Bayesian information criterion.