Information geometry for real-time processing of audio signals

Arnaud Dessein, Ph.D. student at the UPMC under supervision of
Arshia Cont and Gérard Assayag

Institute for Research and Coordination of Acoustics and Music, Paris, France
Japanese-French Laboratory for Informatics, Tokyo, Japan

November 11th 2010
Outline

1. Introduction
2. Background
3. Proposed system
4. Results
5. Conclusion
Introduction

Background

Proposed system

Results

Conclusion

Motivations

Results obtained with non-negative matrix factorization

Going further with information geometry
## Motivations

<table>
<thead>
<tr>
<th>Representation</th>
<th>Type</th>
<th>Data rate</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic</td>
<td>&lt; 0.1 Hz</td>
<td>Implicit knowledge</td>
<td></td>
</tr>
<tr>
<td>Symbolic</td>
<td>0.1-25 Hz</td>
<td>Low information</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>10Hz-1kHz</td>
<td>Generation: Synthesis</td>
<td></td>
</tr>
<tr>
<td>Signal</td>
<td>10-100 kHz</td>
<td>Reduction: Analysis</td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>10-100 kHz</td>
<td>High information</td>
<td></td>
</tr>
</tbody>
</table>

### Figure
Levels of representation of audio, waveform and spectrogram representations.

- **Waveform**
- **Spectrogram**

**Motivations**
- Results obtained with non-negative matrix factorization
- Going further with information geometry

**Figure**
- Levels of representation of audio, waveform and spectrogram representations.
Motivations

- Fill in the gap between signal and symbolic representations.

Figure: Levels of representation of audio, waveform and spectrogram representations.
Motivations

Figure: Levels of representation of audio, waveform and spectrogram representations.

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.
Motivations

Figure: Levels of representation of audio, waveform and spectrogram representations.

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.
- Two approaches:
  - Non-negative matrix factorization: current trend.
  - Information geometry: new trend.
Polyphonic music transcription (demo).

International evaluation: 2nd rank at MIREX 2010 (against off-line systems) for note-level transcription of polyphonic music (not only piano).

Other applications: drum transcription, environmental scene analysis.
Non-negative matrix factorization: reductionist approach, supervised system, structural a priori.

“But do humans really do that? What about untrained listeners?”
Non-negative matrix factorization: reductionist approach, supervised system, structural a priori.

“But do humans really do that? What about untrained listeners?”

Information geometry: holistic approach, unsupervised system, no structural a priori.

“Discover the environment structure without prior knowledge, through the variation of its information content as it unfolds in time.”
Information geometry framework

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold:

\[ S = \{ p_\xi = p(x; \xi) : \xi = [\xi^1, \ldots, \xi^n] \in \Xi \} \]
Information geometry framework

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold:

\[ S = \{ p_\xi = p(x; \xi): \xi = [\xi_1, \ldots, \xi_n] \in \Xi \}. \]

- Example: 
  \[ p(x; \xi) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \text{ with } \xi = [\mu, \sigma^2]. \]

Figure: A statistical model and its differentiable manifold structure.
Information geometry framework

Statistical differentiable manifold.

Under certain assumptions, a statistical model forms a differentiable manifold:
\[ S = \{ p_\xi = p(x; \xi): \xi = [\xi^1, \ldots, \xi^n] \in \Xi \}. \]

- Example: \( p(x; \xi) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\} \) with \( \xi = [\mu, \sigma^2] \).

- Exponential families and Bregman divergences
  [Amari & Nagaoka, 2000, Banerjee et al., 2005].

- Generic algorithms that handle many generalized distances (demo)
Application to real-time audio processing

Scheme:
1. Represent the incoming audio stream with short-time sound descriptors $d_j$.
2. Model these descriptors as probability distributions $p_{\theta_j}$ from a given exponential family.
3. Use the framework of computational information geometry on these distributions.
Scheme:
1. Represent the incoming audio stream with short-time sound descriptors $d_j$.
2. Model these descriptors as probability distributions $p_{\theta j}$ from a given exponential family.
3. Use the framework of computational information geometry on these distributions.

In particular, it allows to define the notion of similarity in an information setup through divergences.
Application to real-time audio processing

Scheme:
1. Represent the incoming audio stream with short-time sound descriptors $d_j$.
2. Model these descriptors as probability distributions $p_{\theta_j}$ from a given exponential family.
3. Use the framework of computational information geometry on these distributions.

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

Potential applications:
- Audio content analysis.
- Segmentation of audio streams.
- Automatic structure discovery of audio signals.
- Sound processing and synthesis.
Figure: Schema of the general architecture of the system.
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.
Sound descriptors modeling

- 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.
  - Many other possibilities.

Figure: Sound descriptors modeling.
Temporal information modeling

- Model formation: from signal to symbol.
  - Assumption of quasi-stationary audio chunks.
  - Change detection adapted from CuSum [Basseville & Nikiforov, 1993].

*Figure:* Model formation at time $t$. 
Temporal information modeling

- Model formation: from signal to symbol.
  - Assumption of quasi-stationary audio chunks.
  - Change detection adapted from CuSum [Basseville & Nikiforov, 1993].

![Model formation at time t.](image)

**Figure:** Model formation at time $t$.

- Factor oracle: from symbol to syntax (and from genetics to music!).
  - Forward transitions: original sequence factors.
  - Backward links: suffix relations, common context.

![Factor oracle of the word abbbbaab.](image)

**Figure:** Factor oracle of the word abbbbaab.
Outline

1. Introduction
2. Background
3. Proposed system
4. Results
5. Conclusion
Audio segmentation

Introduction

Background

Proposed system

Results

Conclusion

Audio segmentation

Figure: Segmentation of the 1st Piano Sonate, 1st Movement, 1st Theme, Beethoven.
Music similarity analysis

Figure: Similarity analysis of the 1st Piano Sonate, 3rd Movement, Beethoven.
Musical structure discovery

Figure: Structure discovery of the *1st Piano Sonata, 3rd Movement*, Beethoven.
Figure: Computer-assisted improvisation, Fabrizio Cassol and Philippe Leclerc.
Introduction

Background

Proposed system

Results

Conclusion
Motivations:

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.
Motivations:
- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.

Resources on NMF: http://imtr.ircam.fr/imtr/Realtime_Transcription
Motivations:
- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.

Resources on NMF: http://imtr.ircam.fr/imtr/Realtime_Transcription

Resources on IG: http://imtr.ircam.fr/imtr/Music_Information_Geometry

National research group: IRCAM, Ecole Polytechnique, Thales, etc.
Brillouin seminar: http://www.informationgeometry.org/Seminar/seminarBrillouin.html
IGAIA 2012.
Motivations:

- Fill in the gap between signal and symbolic representations.
- Devise computational tools for complex real-time settings.

Resources on NMF: http://imtr.ircam.fr/imtr/Realtime_Transcription


Resources on IG: http://imtr.ircam.fr/imtr/Music_Information_Geometry


National research group: IRCAM, Ecole Polytechnique, Thales, etc.

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

IGAIA 2012.

Thanks for your attention! Questions?


