Instrument-specific atoms for mid-level representation of music: application to music instrument recognition

P. Leveau¹,²  E. Vincent³  G. Richard¹  L. Daudet²

¹ GET-ENST (Télécom Paris)  ² Laboratoire d’Acoustique Musicale
Paris  Université Pierre & Marie Curie
Paris ³ Center for Digital Music
Queen Mary
University of London
London
Introduction

- **Music content analysis from audio**: music transcription, genre classification, music instrument recognition
- Use of **features** computed on low-level signal representations.
- Features describe **signal** as a **whole**: no source separation, limits for polyphony

=> **Mid-level representation**: towards note-like objects. used for CASA, recognition of multiple instruments, harmonic similarity …
Use of **sparse representations** to build a new mid-level representation for harmonic instruments

Signal represented as a linear combination of waveforms (*atoms*):

where $w_n$ are in a **dictionary** $D$.

- Representation is **sparse** when $N \ll \text{dim}(x)$
- to get a **sparse representation**, its elements must exhibit strong similarities with the signal
Ideally, atoms = notes … (~MIDI!)

… but it would make huge dictionaries

Solution: lower granularity (50 ms)

Goal: get an approximation of a music signal using short atoms, whose characteristics are learnt on instruments that may be playing.
Summary

I. Introduction

II. Dictionary design
   I. Atoms
   II. Gabor/Harmonic Atoms
   III. Instrument-specific atoms

III. Algorithm
   I. Matching Pursuit algorithm
   II. Sampling the dictionary

IV. Learning

V. Application
   I. Output of the decompositions
   II. Music instrument recognition
   III. Results

VI. Conclusion
Summary

I. Introduction

II. Dictionary design
   I. Atoms
   II. Gabor/Harmonic Atoms
   III. Instrument-specific atoms

III. Algorithm
   I. Matching Pursuit algorithm
   II. Sampling the dictionary

IV. Learning

V. Application
   I. Output of the decompositions
   II. Music instrument recognition
   III. Results

VI. Conclusion
Dictionary design

- Numerous types of waveforms have been used for sparse approximations of audio signals:
  - Gabor atoms (complex sinusoids)
  - Chirps
  - Local cosines
  - Haar wavelets
  - Data-driven atoms …
Dictionary design

- Numerous types of waveforms have been used for sparse approximations of audio signals:
  - Gabor atoms (complex sinusoids)
  - Chirps
  - Local cosines
  - Haar wavelets
  - Data-driven atoms …
Gabor atoms

[Mallat TSP 1993]
Harmonic atoms

\[ f_0, 2f_0, 3f_0, 4f_0, 5f_0 \]

with

\[ u, s, t \]

[Gribonval TSP 2003]
Instrument-specific harmonic atoms

- A vectors are learned from isolated notes.
- For each pitch that can be played by an instrument, several A vectors.

Example:
Cello  🎻  Clarinet  🎹
Summary

I. Introduction

II. Signal model
   1. Dictionary design
   2. Atoms
   3. Instrument-specific atoms

III. Algorithm
   1. Matching Pursuit algorithm
   2. Sampling the dictionary

IV. Learning

V. Application
   1. Output of the decompositions
   2. Music instrument recognition
   3. Results

VI. Conclusion
Decomposition algorithm

- Once the dictionary is built, how to decompose the signal with it?
- **Matching Pursuit** algorithm:
  - Compute all inner products signal \( \mid \) atoms from the dictionary
  - Subtract the most energetic atom with its weight
  - Update of the inner products and of the signal

...until a stop condition is reached (SNR or number of atoms)
Sampling the dictionary

- $s$: one scale (typically 50 ms)
- $u$: linearly sampled (with a fraction of $s$ as period)
- $f_0$: logarithmically sampled
- $A$: already discrete set
- $\Phi$: not sampled: estimated at each iteration:
Summary

I. Introduction

II. Signal model
   i. Dictionary design
   ii. Atoms
   iii. Instrument-specific atoms

III. Algorithm
   i. Matching Pursuit algorithm
   ii. Sampling the dictionary

IV. Learning

V. Application
   i. Output of the decompositions
   ii. Music instrument recognition
   iii. Results

VI. Conclusion
Learning A on isolated notes

- **Database**: RWC for five classes: Cello (Co), Clarinet (Cl), Flute (Fl), Oboe (Ob), Violin (Vi).
- Taking one instrument per class, only one atom is kept for each pitch and for each 3 velocities.
- **Method**: 
  - $f_0$ is sampled at fine fundamental frequencies around the annotated pitch
Summary

I. Introduction

II. Signal model
   I. Dictionary design
   II. Atoms
   III. Instrument-specific atoms

III. Algorithm
   I. Matching Pursuit algorithm
   II. Sampling the dictionary

IV. Learning

V. Application
   I. Output of the decompositions
   II. Music instrument recognition
   III. Results

VI. Conclusion
Output of the decomposition (1)

Flute

Cello
Clarinet
Flute
Oboe
Violin
Output of the decomposition (2)

Clarinet

Cello
Clarinet
Flute
Oboe
Violin
Music Instrument Recognition

- Decomposition ~ template-based approach
- **Score for instrument i**: Sum of the modulus of the selected atoms weights.
- Test database: Solo phrases, 2 sec excerpts
Music Instrument Recognition: Results

- **Reference**: SVM (40 features selected out of 543) [Essid TSALP 2006], trained on solos.

  - reference overall 83.9%
  - ISH atoms overall 68.5%
Results: Comments

- Less efficient than a standard feature-based approach but...
- Algorithm not optimised for classification (yet)
- Reduced training set (3 observations per pitch and instrument!)
- Only the harmonic part is taken into account
- Learnt on isolated notes, tested on solos
- Decomposition can be done on duos with the same instrument models
Summary

I. Introduction

II. Signal model
   I. Dictionary design
   II. Atoms
   III. Instrument-specific atoms

III. Algorithm
   I. Matching Pursuit algorithm
   II. Sampling the dictionary

IV. Learning

V. Application
   I. Output of the decompositions
   II. Music instrument recognition
   III. Results

VI. Conclusion
Conclusion

- Instrument-specific harmonic atoms for the decomposition of audio signals.
- Encouraging results for Music Instrument Recognition. Recent results show an improvement of 10 points.
- Perspectives:
  - **Dictionary**: chirped harmonic atoms, stéréo, inharmonic atoms
  - **Algorithms**:
    - Molecular approach to consider time dependancies
    - Selection of several atoms per time frame to better handle polyphony
  - **Applications**: optimisation for Music Instrument Recognition, evaluation of transcription, low-rate audio coding, use of symbolic algorithms …