Invariant Audio Prints for Music Indexing and Alignment

Rémi Mignot[1], Geoffroy Peeters[2]

[1] STMS Lab – IRCAM, Sorbonne Université, CNRS (UMR-9912), Paris, France
[2] LTCI - Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France

Companion webpage

Two tasks

- Audio ***Indexing***
  Find the "reference song" from a music catalog based on the signal content of a given audio excerpt
- Audio-to-audio ***Time Alignment***
  Search the time mapping between two occurrences of the same music

→ *use of the same method for both tasks*

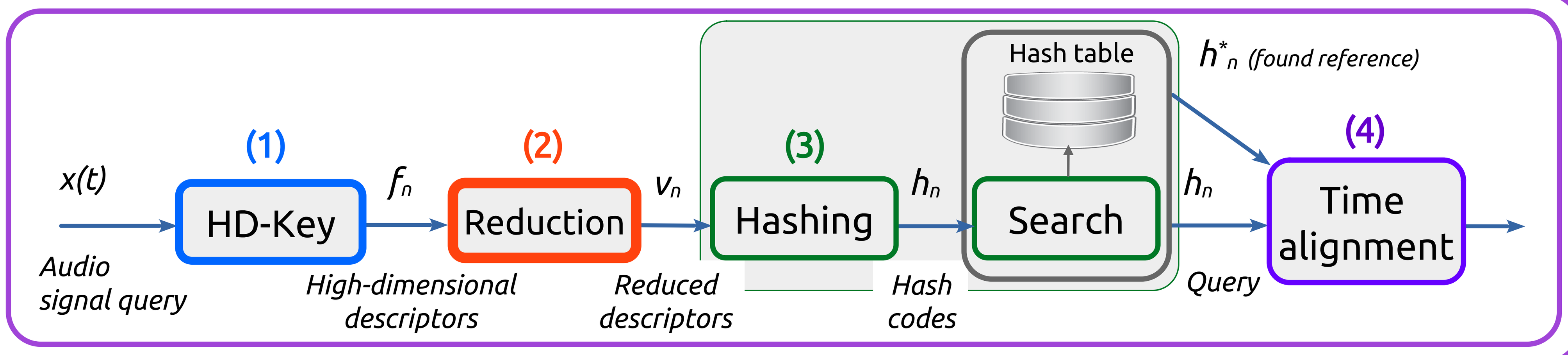
Goals

- ***Robust*** to audio ***transformations/degradations***
  time stretching, pitch shifting, noise addition, distortion, audio effects, and different instruments (for alignment)
- ***Relevant*** *to the* ***music content***
  melodies, chords, rhythms and possibly the instrument timbres

→ *computation of* ***music distances*** *based on* ***audio codes***

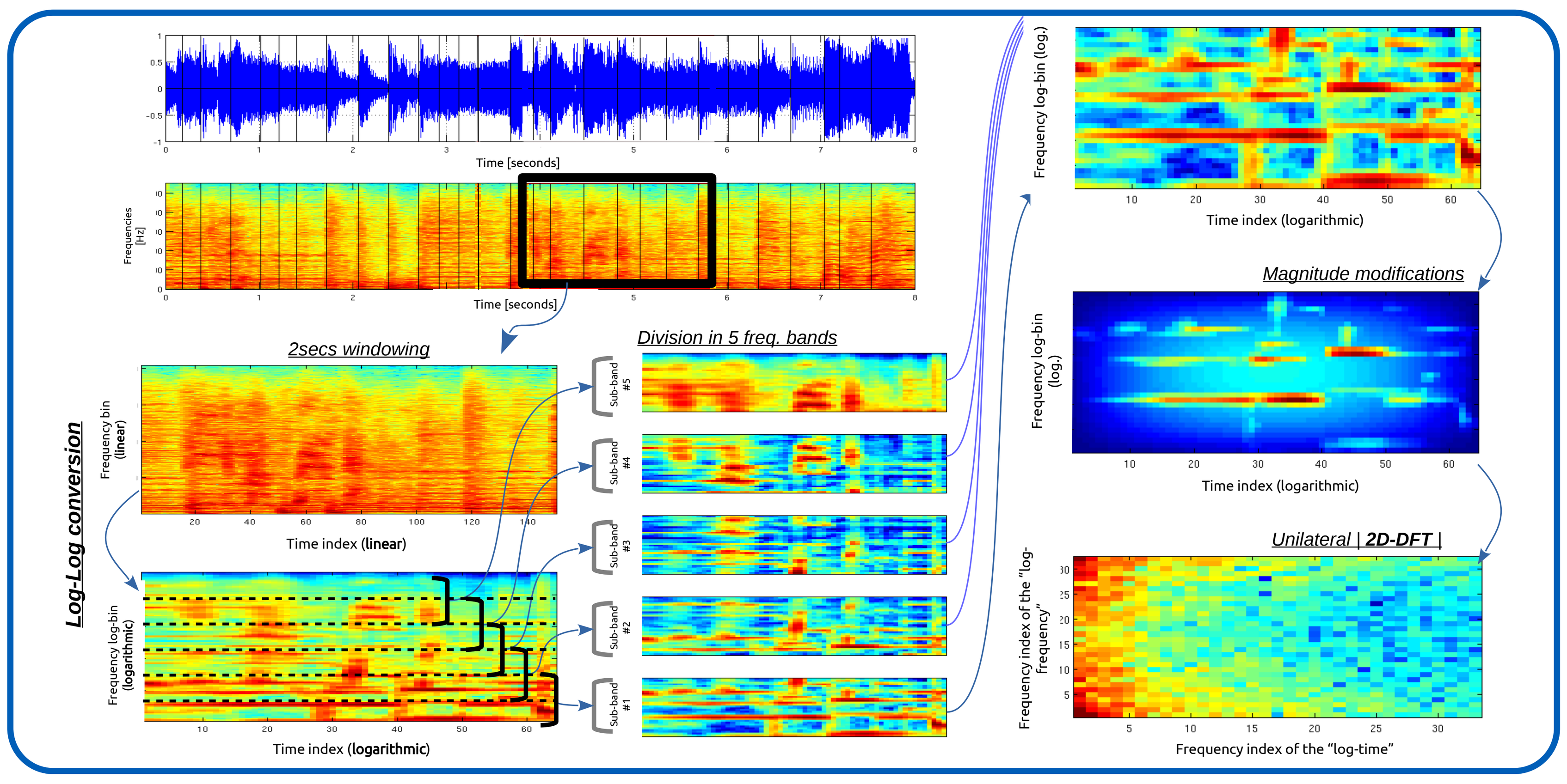
Method overview

(1) High-dimensional audio keys
(2) Robust dimension reduction
(3) *Approximate* Hashing *tolerant to bit corruption (LSH-based),*
(4) *DTW-based* Time alignment *to estimate the time mapping.*



(1) High-dimensional Audio Keys

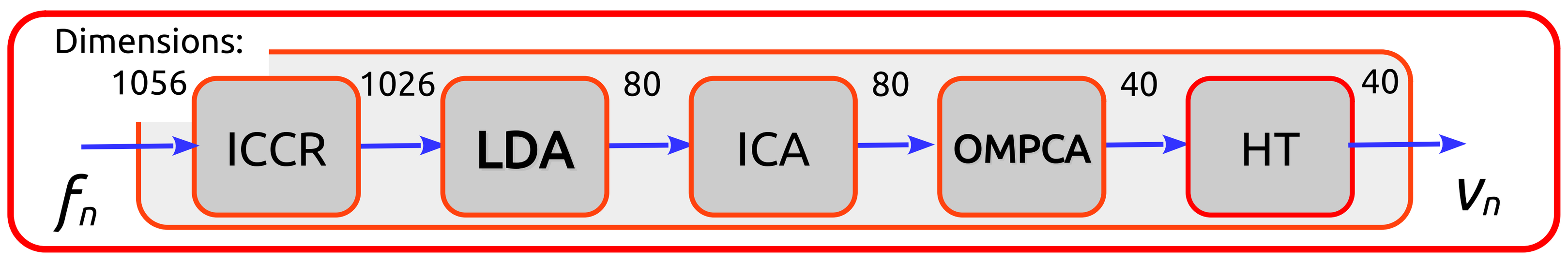
- Audio descriptors *(inspired by audio classification)*
  → **relevant** to the *music content*, and
  → **robust** *by design* to audio *transformations / degradations*
- Manipulations of sub-spectrograms:
  → *Log. scale of frequencies and time, frequency band splitting, Amplitude transformation, Magnitude of 2D-DFT.*



- Based on properties of:
  → *Log. function, Shift invariance of |DFT|, Amplitude change,*
- The descriptors are robust *by design* to:
  → *Pitch and time changes, and noise, filtering.*

(2) Robust dimension reduction

Learning of a ***linear transformation*** chain ***invariant*** to degradations

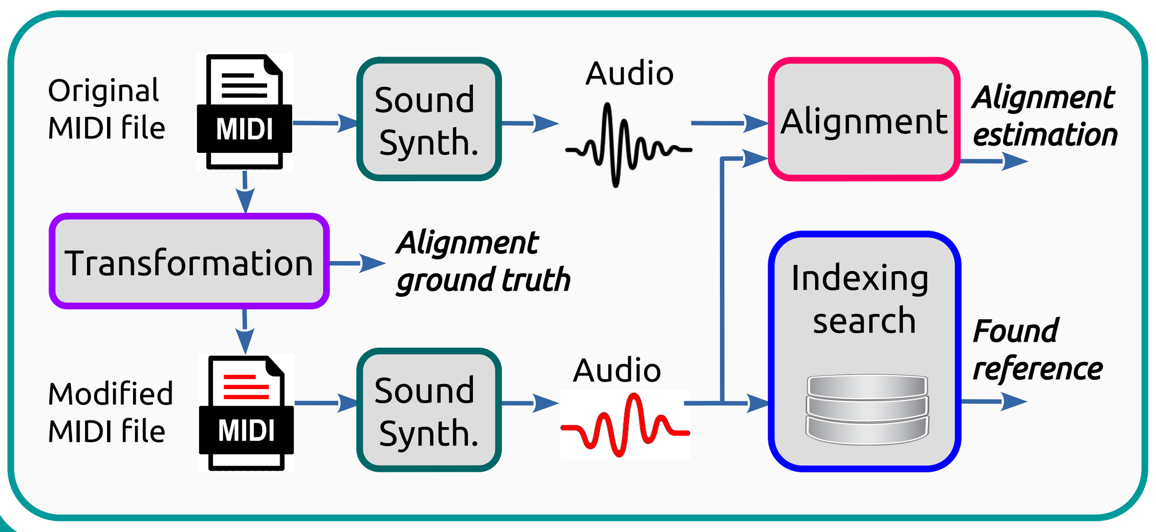


1) ICCR *(Ill-Conditioned Component Rejection)*:
   → Remove redundancies
2) LDA *(Linear Discriminant Analysis):*
   → Select robust dimensions
3) ICA *(Independent Component Analysis)*:
   → For a uniform filling of hash table because of independency
4) OMPCA *(Orthogonal Mahalanobis PCA)*:
   → Recover robustness, & preserves decorrelation
5) HT *(Hadamard Transform)*:
   → uniform robustness, prepare for hashing, & preserves decorrelation.

- Output variables v_n with properties:
  centered, normalized, *mutually uncorrelated*, *robust* to transformations, and *discriminant* to the original signal.
- Use of a ***Data Augmentation*** approach for training (LDA & OMPCA)
  → *maximize distances for different original signals, and*
  → *minimize distances for transformations of the same signal.*

Experiment: *Indexing and alignment of "MIDI covers"*

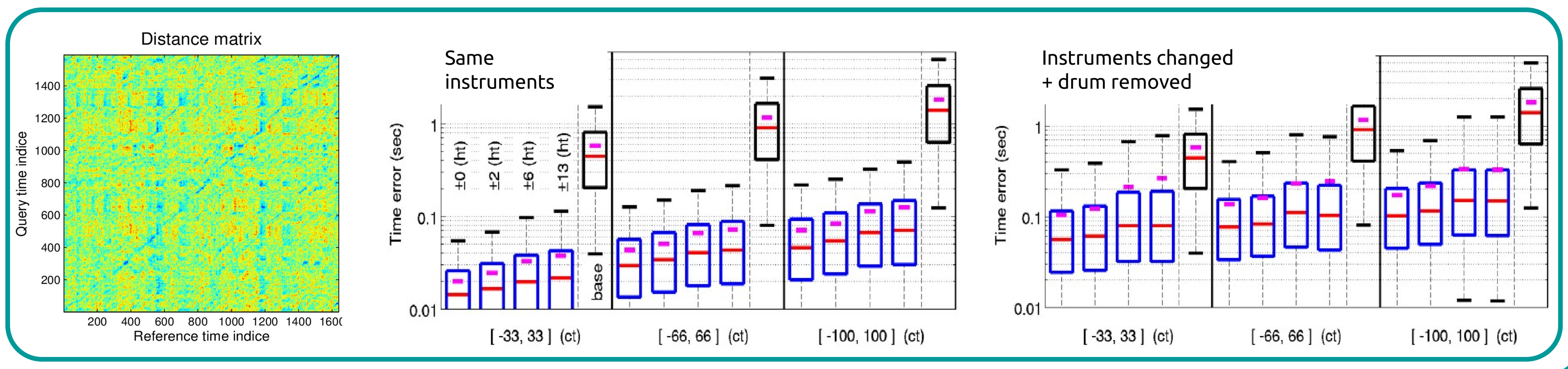
MIDI Transformations:
- ***Time variant tempo***:
  [-33, 33], [-66, 66] and [-100, 100] cents.
- ***Pitch Shifting***:
  0, ±2, ±6, ±13 half-tones.
- ***Instrument change + drum removed***
- *Remark:* 33ct → x1.25, 66ct → x1.58, 100ct→ x2.



Indexing results (full catalog: ~40 000 songs, ranks averaged over 238 tests)

	Time Stretch (cents)	[− 33, 33]				[− 66, 66]				[− 100, 100]			
	Pitch Shift ($\frac{1}{2}$ tones)	0	±2	±6	±13	0	±2	±6	±13	0	±2	±6	±13
Same instruments	**STEP1:** rank (full catalog):	1.0	1.2	4.3	17.9	1.0	1.3	5.7	22.7	1.0	1.4	5.9	29.
	STEP2: rank (over the 200 best):	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.1
Changed instruments	**STEP1:** rank (full catalog):	132.3	174.5	310.1	319.9	128.7	185.6	243.0	301.5	129.5	199.5	274.6	302.
	STEP2: rank (over the 200 best):	2.2	5.0	15.3	17.9	2.6	6.3	16.8	20.5	4.3	10.2	24.9	30.3

Alignment results (evaluation averaged over 238 tests, baseline = diagonal)



Acknowledgment

This work has been initiated with the BeeMusic Project (FSN-O14703-370483), and continued with the AQUA-Rius project (ANR-22-CE23-0022-01).

CBMI 2024
21st Int. Conf. on Content-based Multimedia Indexing
Sept. 18-20, Reykjavik, Iceland