An Efficient Methodology for The Identification of Multiple Music Works within a Single Query

Emanuele Di Buccio[1], Nicola Montecchio[1], Nicola Orio[2]

1 - Department of Information Engineering
2 - Department of Cultural Heritage
University of Padova

Introduction

- Increasingly larger media collections

last.fm

YouTube

Spotify

- Increasingly bigger challenges, e.g.:
  - Metadata inconsistency
  - New querying methods
  - Interaction with the content

Case Study

- An efficient identification methodology for improved access to music heritage collections
  - Real application scenario
    - Scalability issues
  - Applicable to many other contexts
    - Public sound archives
    - Theater and concert halls sound archives (mainly rehearsals)

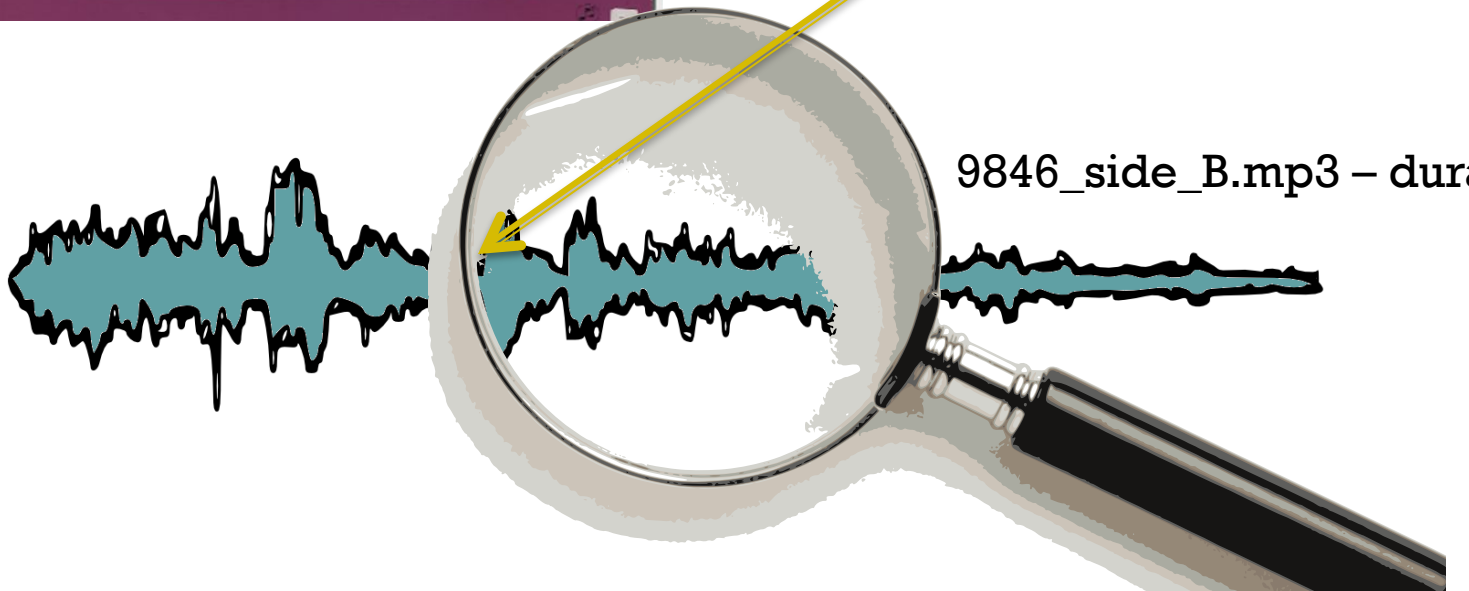
Metadata Inconsistency

Maurice Ravel
Trio para violin, violoncello y piano - op. 120
Madrid - D.L. 1980 - edita Columbia

A: Moderado, Pantoum, Passacaglia, Final
B: Allegro ma non troppo Andantino, Allegro vivo

P. Fontanarosa: violin. R. Fontanarosa: cello, F. Fontanarosa: piano

Catalogue metadata of the
Fonoteca of the University of Alicante

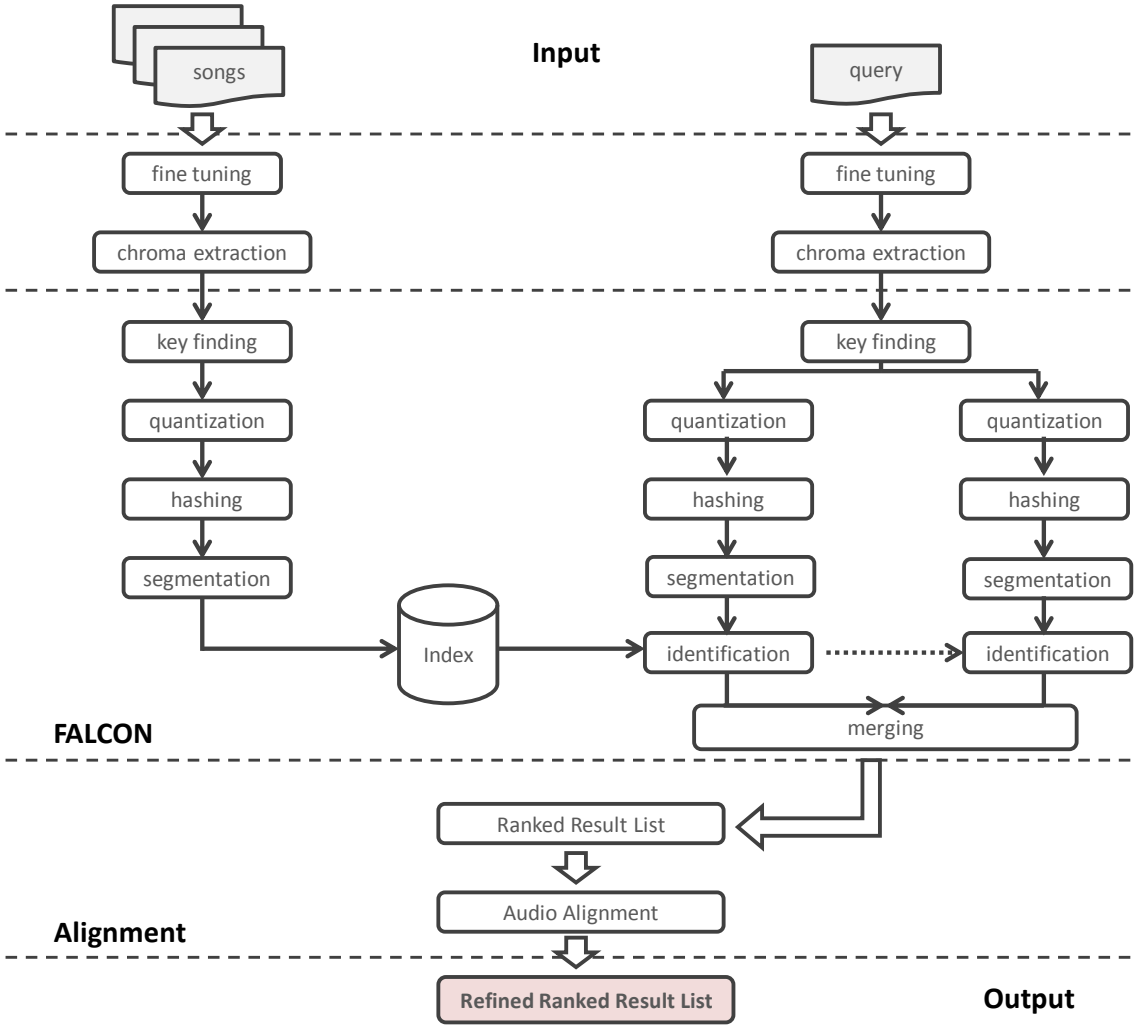


9846_side_B.mp3 – duration 22'34"

Overview

- Enrich and validate the descriptors of recordings digitized by a sound archive institution
  - **Content-based identification**
    - Allows a music digital library to automatically retrieve metadata about recordings
  - Automatic **segmentation** of digitized material
    - Allows to direct access to individual tracks

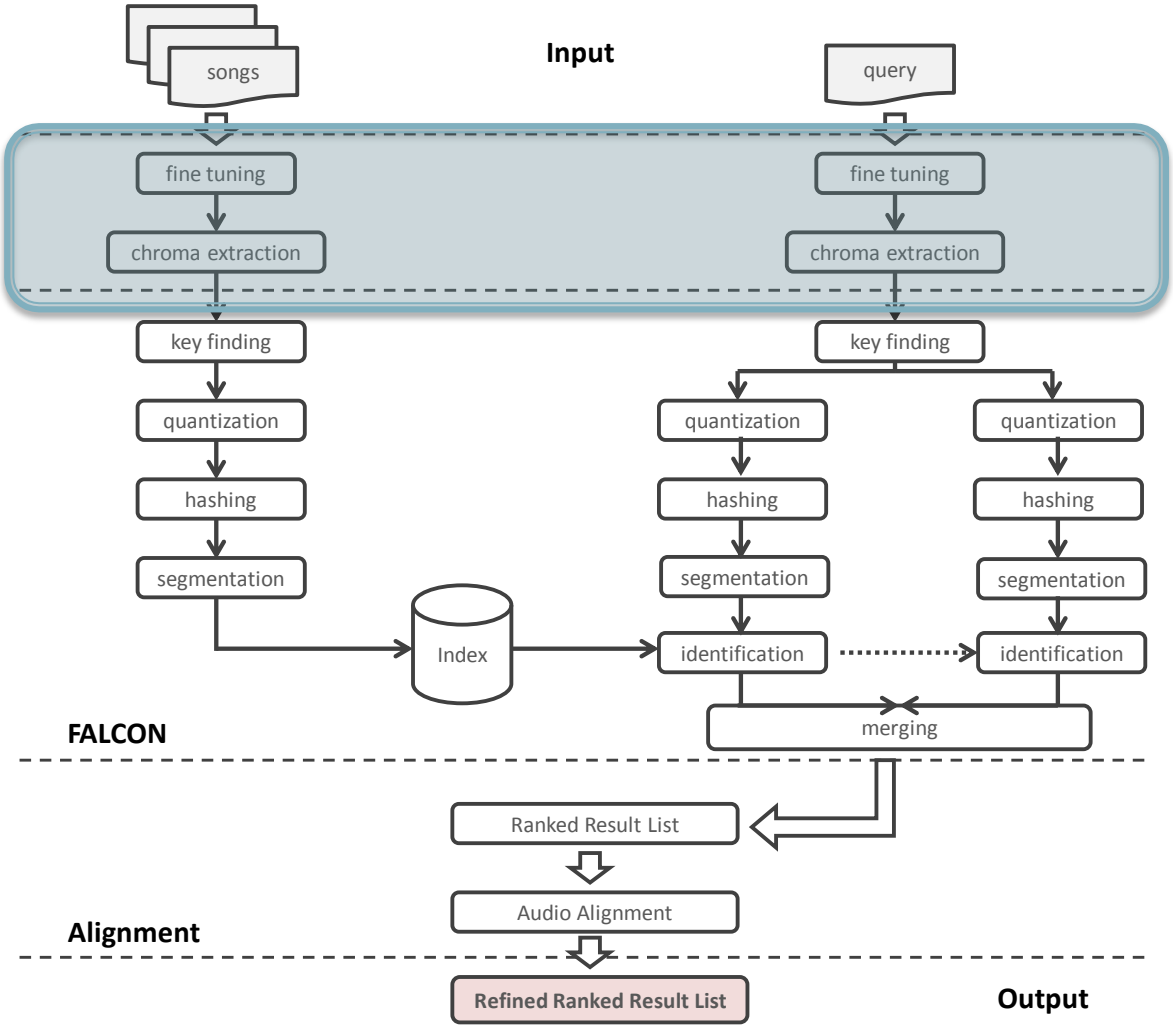
System Overview



efficient identification

accurate reranking

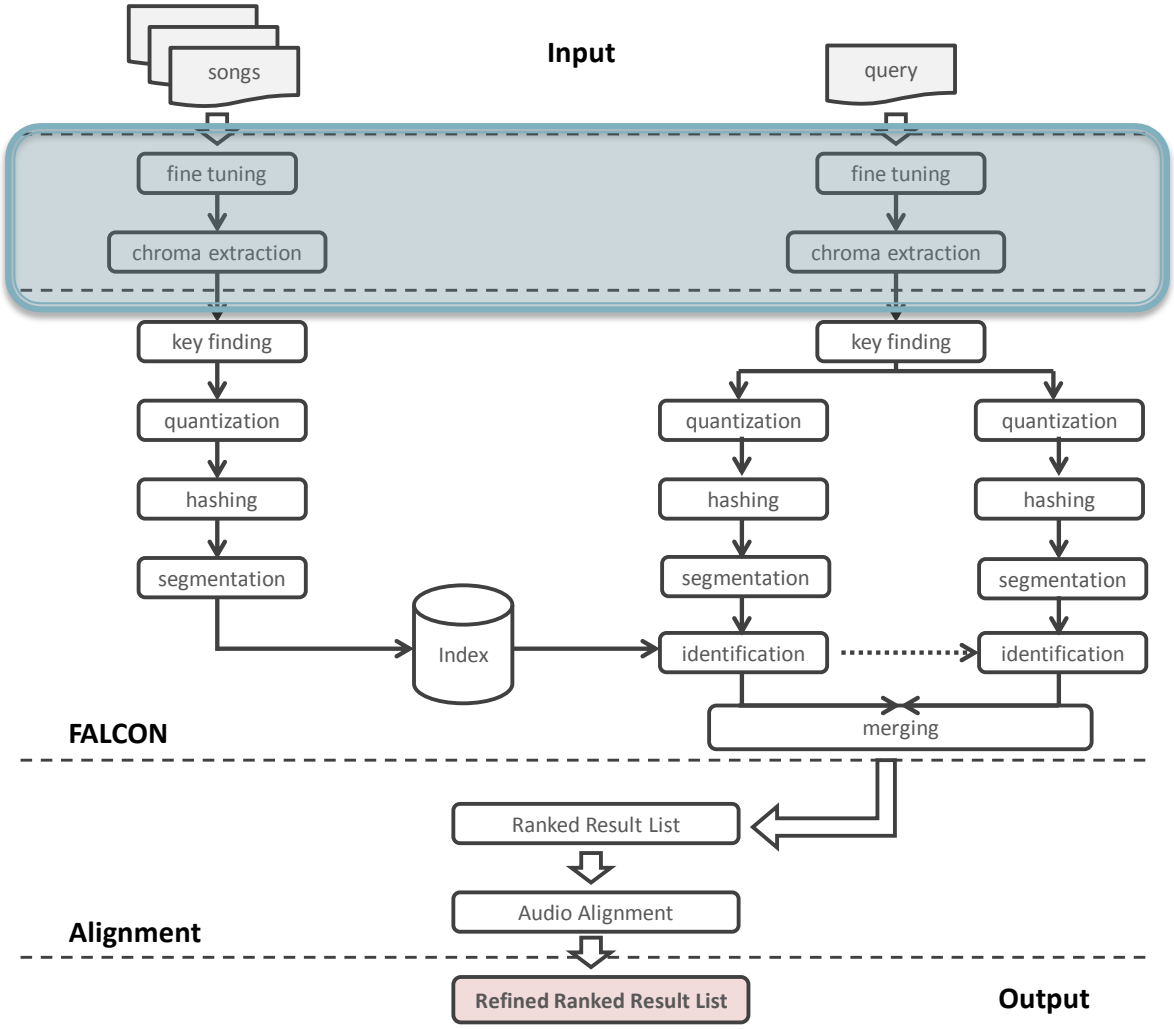
Feature Extraction



CD: Fs = 44.1 kHz

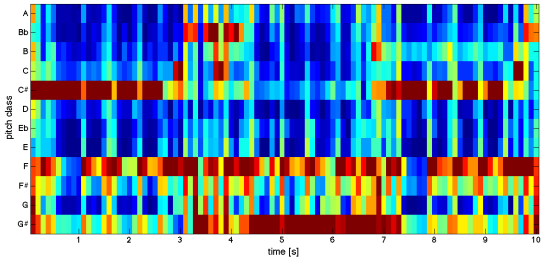
1.4 MBps

Feature Extraction



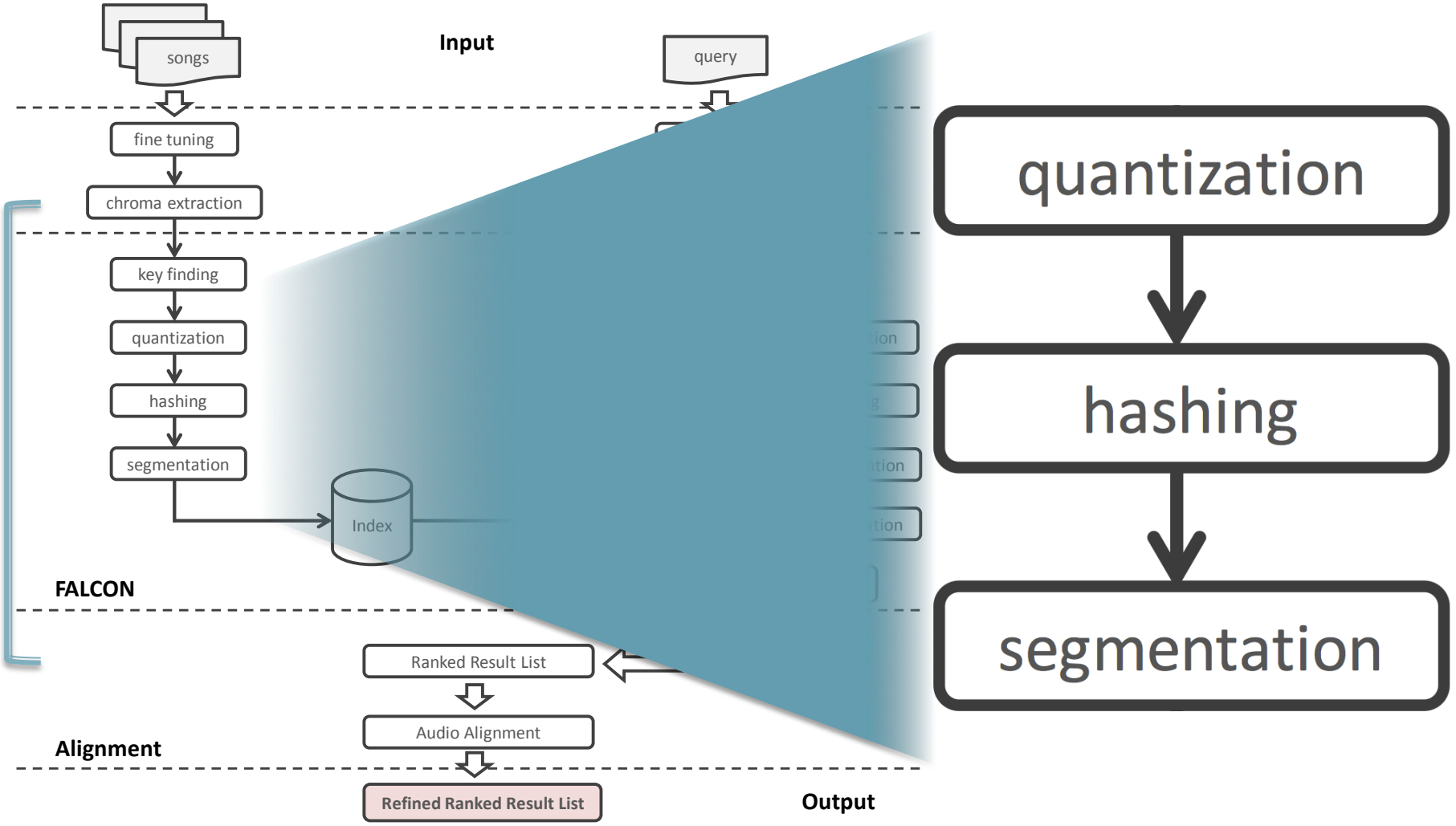
CD: Fs = 44.1 kHz
1.4 Mbps

CHROMA extraction

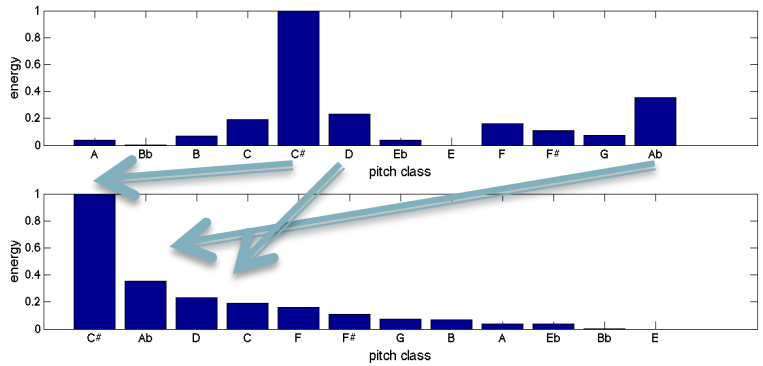


Fs = 100Hz 3.8 Kbps

Representation



Representation - hashing



3-level **quantization:**

C# Ab D

$$4 \cdot 12^2 + 11 \cdot 12^1 + 5 \cdot 12^0 = 713$$

SONG_i

(1)

1812 1880 1192 1876 1953 1876 1857 2676 1876 1147 1867 1876

(2)

Segment size: **7**

Segment overlap: **2**

segment i_1 1812 1880 1192 1876 1953 1876 1857

segment i_2 1876 1857 2676 1876 1147 1867 1876

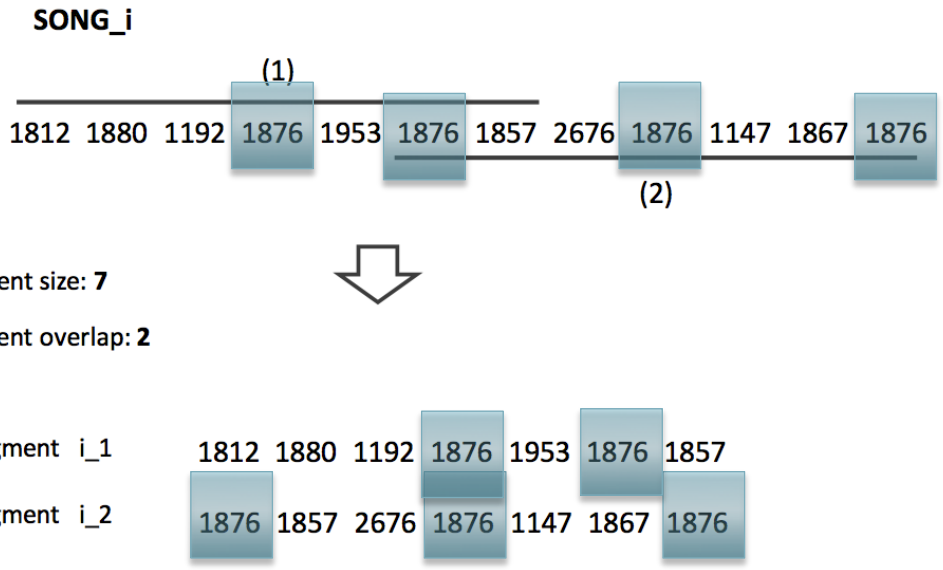
quantization

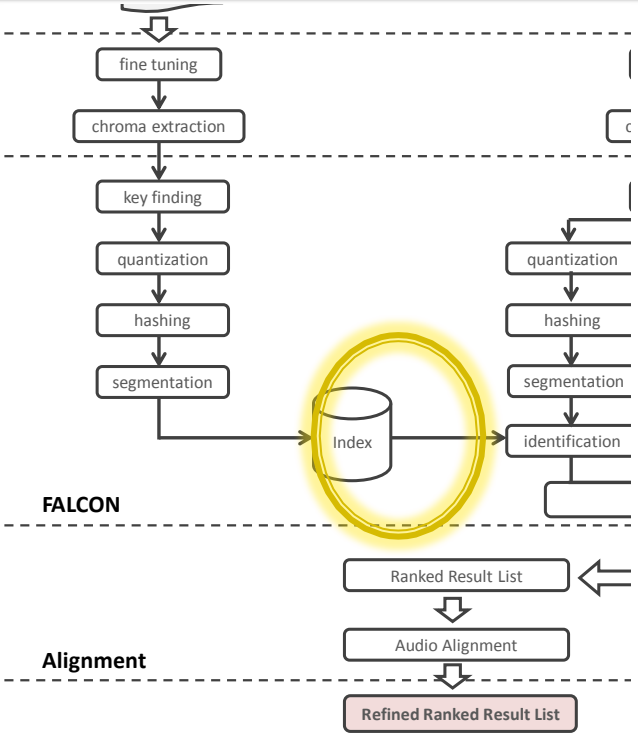
hashing

Fs=100Hz 320bps

segmentation

Indexing





1147	$[i_2, 1]$
1876	$[i_1, 2]$ $[i_2, 3]$

Retrieval

$$S(Q,D) = \underbrace{\sqrt[|Q|]{\prod_{q \in Q} \max_{d \in D} \underbrace{\left\{ \sum_{t \in q \cap d} \min\left(\frac{\mathrm{tf}(t,d)}{|d|}, \frac{\mathrm{tf}(t,q)}{|q|} \right) \right\}}_{\text{Similarity at segment level}}}}_{\text{Similarity at recording level}}$$

- Based on **counting**
- Discard information about **ordering**

- Can be computed **efficiently** using an **index**

Collection – 7k dataset

- Around 7k recordings of classical music, manually selected from the database of a TV broadcasting company
  - A 20 year effort made by music experts, yet
    - A lot of metadata inconsistencies
    - Unnecessary (and expensive) replication of the material
      - Around 0.4% of the records are duplicates

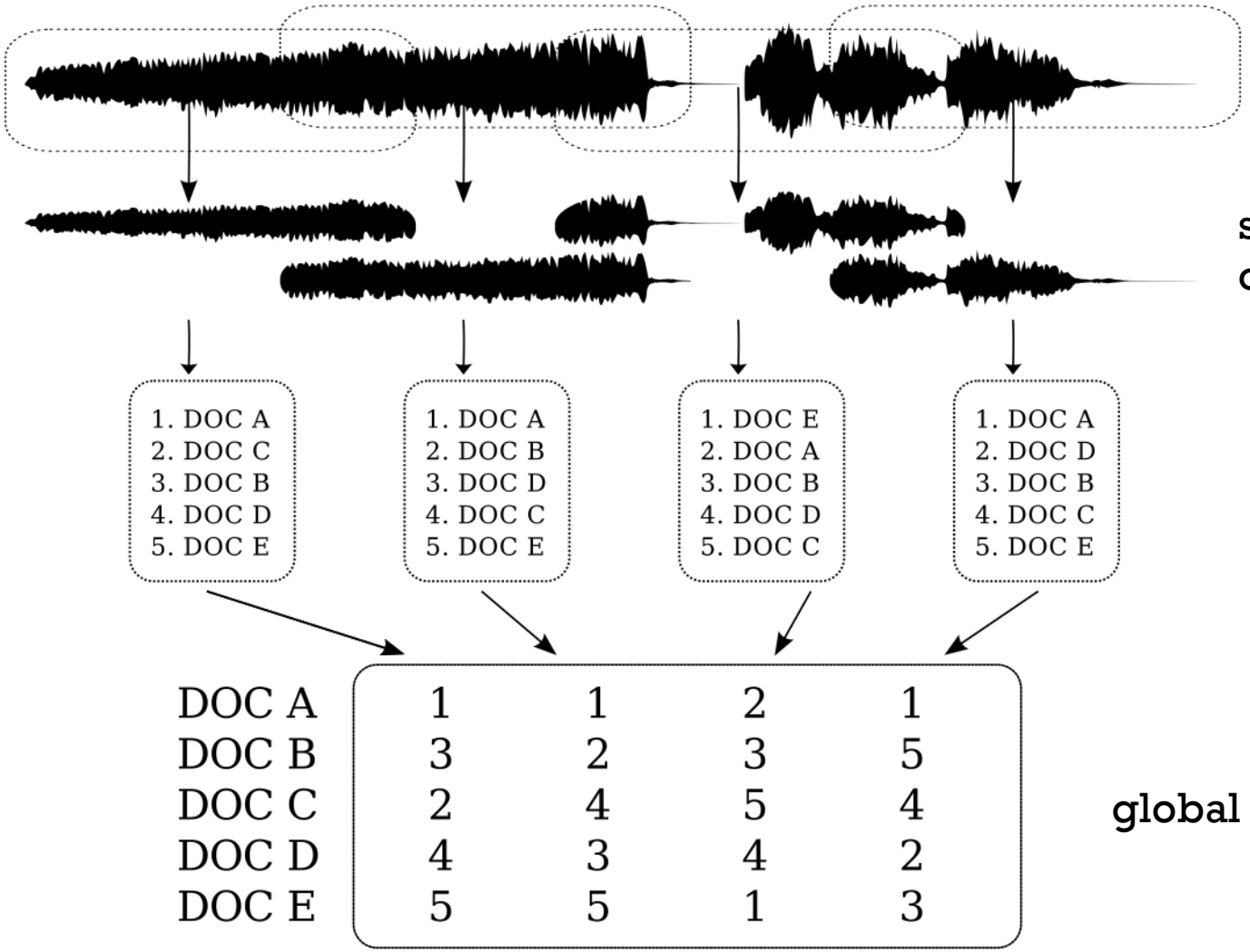
Identification – 7k dataset

- Investigated parameterizations:
  - quantization levels [2,3,4]
  - segment length (sec) [20,30,40,50,60]
  - segment overlap [0,50%]
  - chroma extraction algorithm [3 variants]
  - tuning freq. adjustment [on,off]
- Best configuration so far:
  - Long segments, no overlap, *in-house* chroma

Identification – 7k dataset

- Collection characteristics
  - 6680 tracks
  - 2671 works have at least two tracks
    - 945 cover sets
- Results
  - Mean Reciprocal Rank: **0.83**
  - Mean Average Precision: **0.77**
  - Av. query time: **2”** (av. query length: 5’25”)

Addressing compound queries



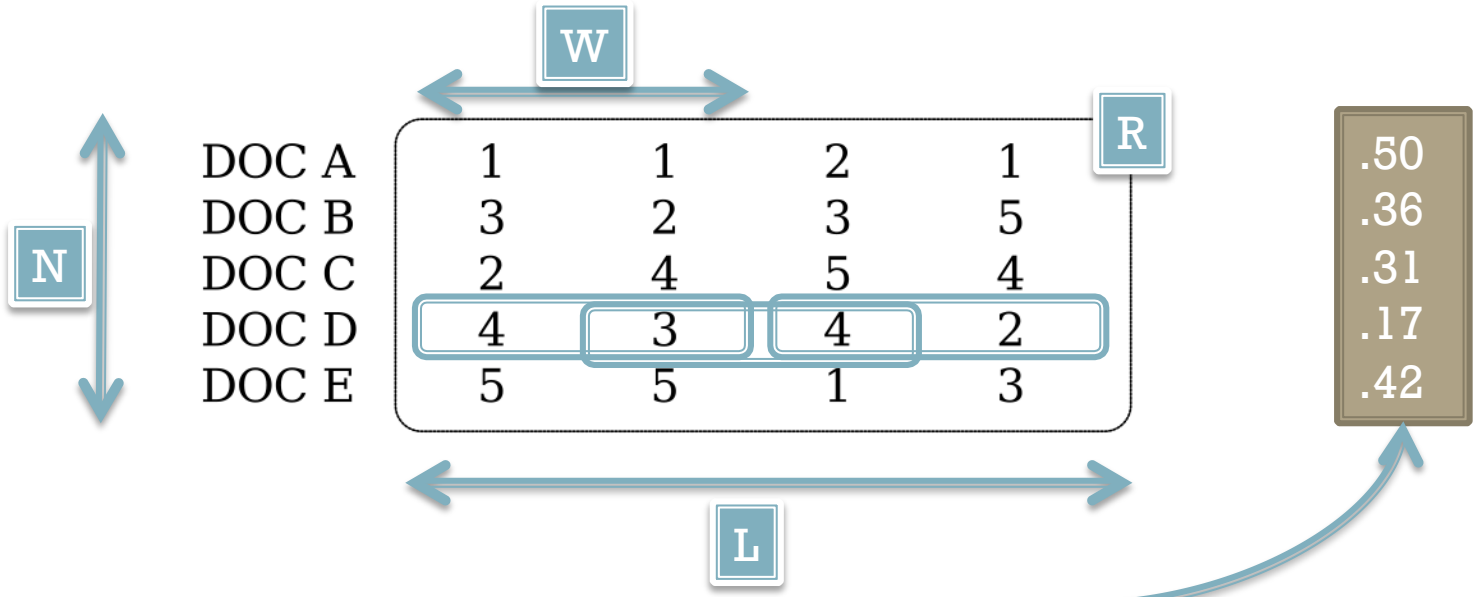
LP side A: 18'35''

segmentation into 5' overlapping windows

rank lists for the 5' windows

global RANK MATRIX

Addressing compound queries



$$\text{score}_i = \max_{w=1\ldots L-W} \sum_{j=w}^{w+W-1} \frac{1}{(\max(R_{i,j}, C))^2)}$$

Collection – LP dataset

- 100 digitized LPs selected from the Fonoteca of the University of Alicante
  - Previously a radio archive
    - The digitization effort took most of the available economic resources
  - Each LP contains at least (a different version of) a recording among the 7k set
    - The 7k set was used as the ground truth for the identification
      - Manual inspection of results was required anyway

Identification – LP dataset

- Collection characteristics
  - 100 audio files (A+B sides) to identify
  - 358 works overlap with 7k dataset
- Results
  - Mean Reciprocal Rank: **0.90**
  - Mean Average Precision: **0.74**
  - Av. query time: **46”** (av. query length: 47’13”)

Conclusion

- Increasing availability of multimedia data demands appropriate techniques for retrieval, access and interaction
- Identification is a fundamental tool for many applications
  - Efficient approaches can be developed
    - Application on a large scale
  - Real case scenarios can be faced