KEYNOTE DAFX-2014

AUDIO INDEXING FOR MUSIC ANALYSIS AND MUSIC CREATIVITY

Geoffroy Peeters
peeters@ircam.fr

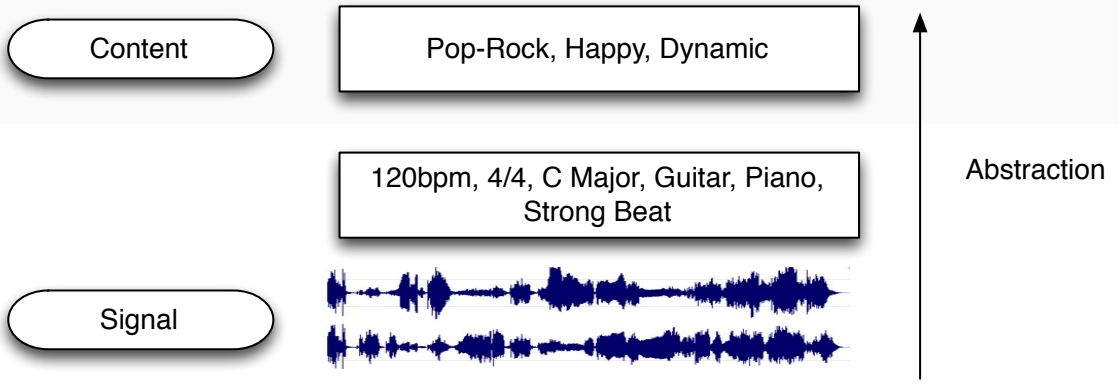
STMS-IRCAM-CNRS-UPMC

Geoffroy Peeters is partly founded by the French government Programme Investissements d'Avenir (PIA) through the Bee Music Project

Introduction

What is Audio Indexing for Music ?

- Audio indexing
  - = describing audio (speech, sound-FX, environmental sounds, music instruments, polyphonic music)
- Audio indexing for music
  - = part of Music Information Retrieval (MIR)
- M.I.R.
  - = not only audio, but also symbolic representation (scores), web-mining, text-mining, user behavior
- M.I.R. from audio (Audio Content Analysis)
  - = extract any user-relevant information from an audio signal
    (= information understandable by a user >< sinusoidal phase)
  - Mostly application driven
  - Examples:
    - genre/mood recognition, tempo/beat/chord/notes estimation, query-by-humming, search by similarity, cover-version detection, …



peeters@ircam.fr

Historically

<u>When did it started ?</u>

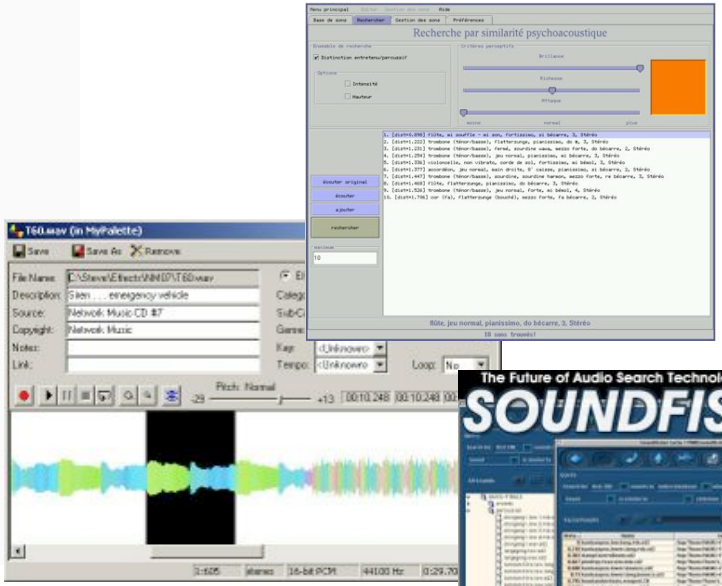
It was existing from a long time under other terms (audio indexing)

- Score following (Vercoe, Dannenberg for music performances)
- Speech/music segmentation
- Musical instrument identification using CASA (MIT/MediaLab)
- Beat estimation
- Object representation of audio sources (MPEG-4 SAOL)



<u>Before 2000</u>

- ISMIR does not exist
- Evaluations
  - On few files
- Applications
  - Mainly sound sample description (Studio-OnLine, Find Sound, MuscleFish)
- Music
  - Napster



peeters@ircam.fr

1994 Goto Real-Time Beat-Tracking

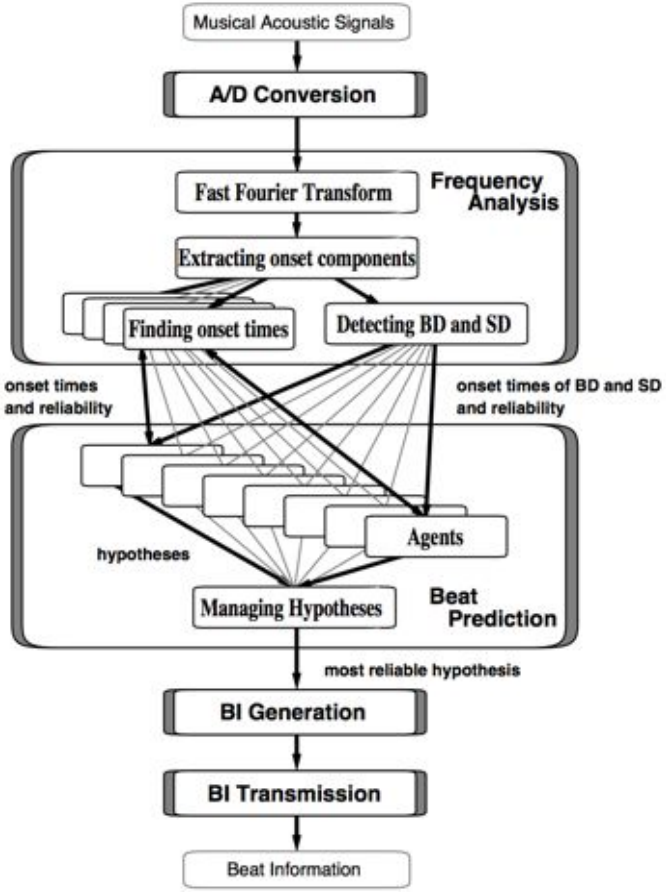
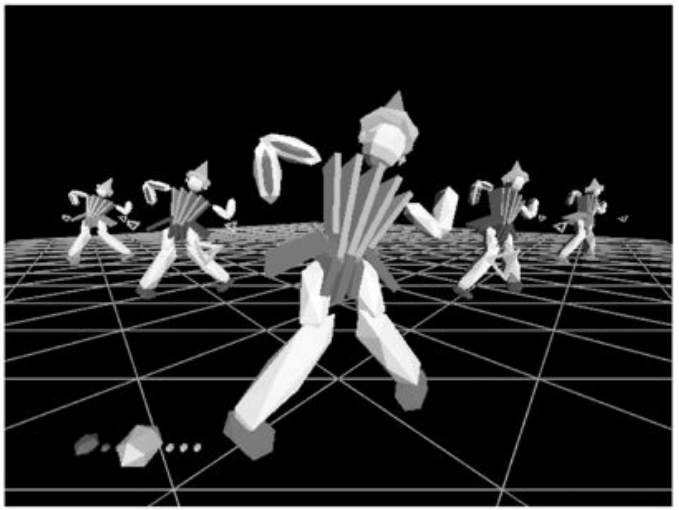
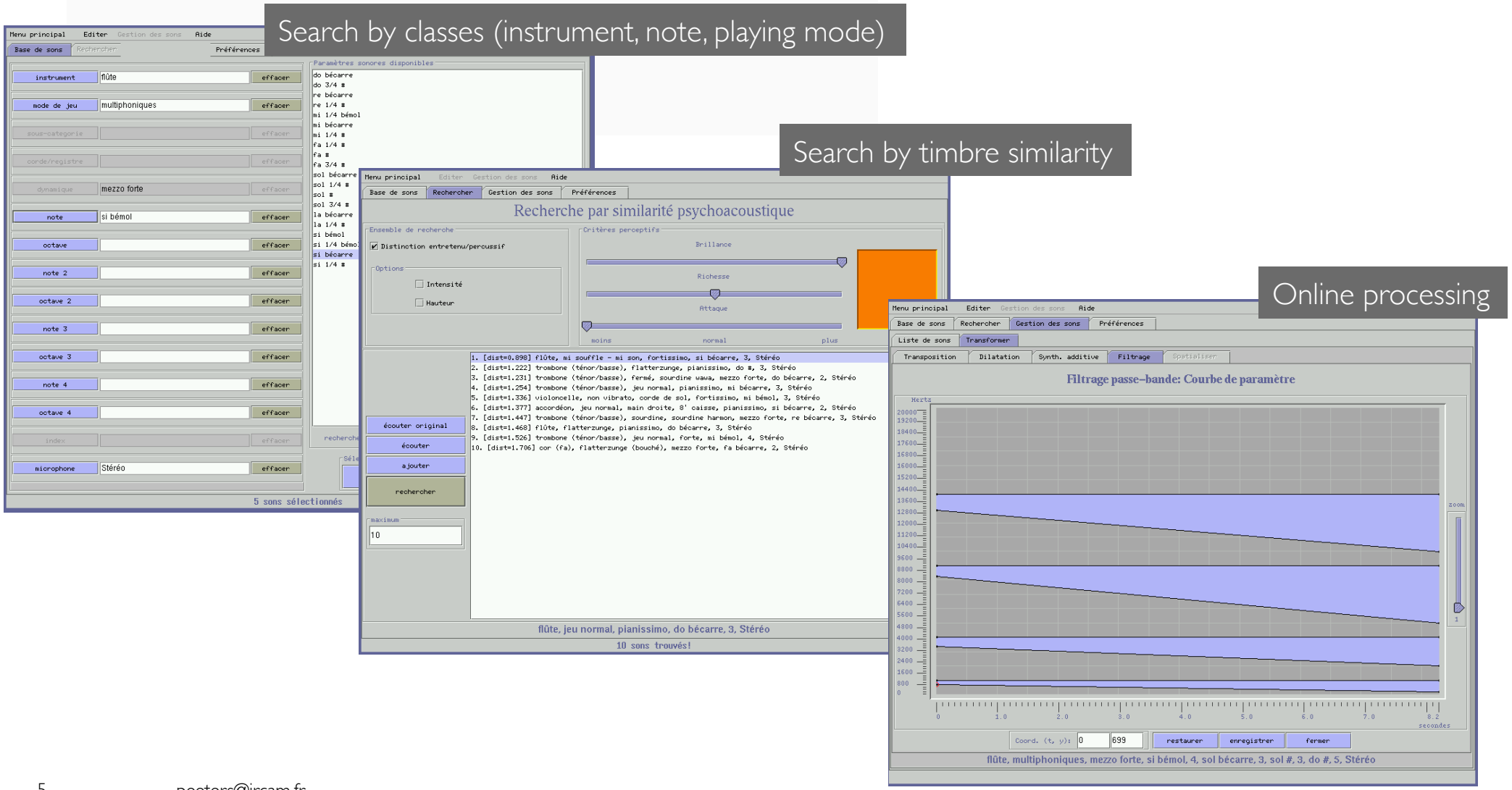


Figure 1: Overview of BTS



1996 IRCAM Studio OnLine

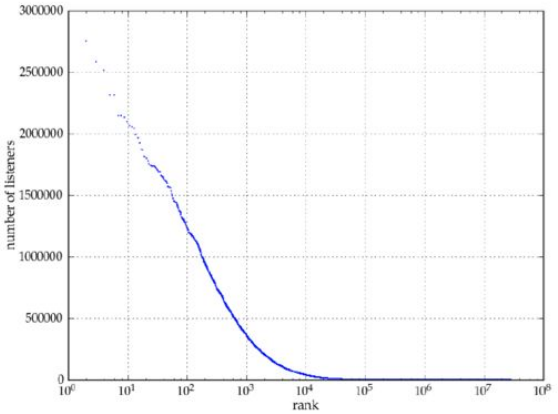


peeters@ircam.fr

Introduction

Motivation ?

- Digital music => many data accessible, how to access ?
- Meta-data: manual, web-crowd-based, content-based
- How to speed up annotation time ?
- Long-tail, cold start



peeters@ircam.fr

Introduction

Motivation ?

- Digital music => many data accessible, how to access ?
- Meta-data: manual, web-crowd-based, content-based
- How to speed up annotation time ?
- Long-tail, cold start

In 2000

- ISO MPEG-7 Audio (1999) [Herre]
- Creation of ISMIR community in 2000
  - ISMIR: fusion between communities: audio processing, machine learning, IR, librarian, …

peeters@ircam.fr

Introduction

Motivation ?

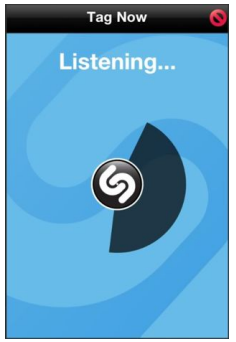
- Digital music => many data accessible, how to access ?
- Meta-data: manual, web-crowd-based, content-based
- How to speed up annotation time ?
- Long-tail, cold start

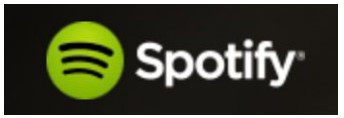
In 2000

- ISO MPEG-7 Audio (1999) [Herre]
- Creation of ISMIR community in 2000
  - ISMIR: fusion between communities: audio processing, machine learning, IR, librarian, …

Today ?

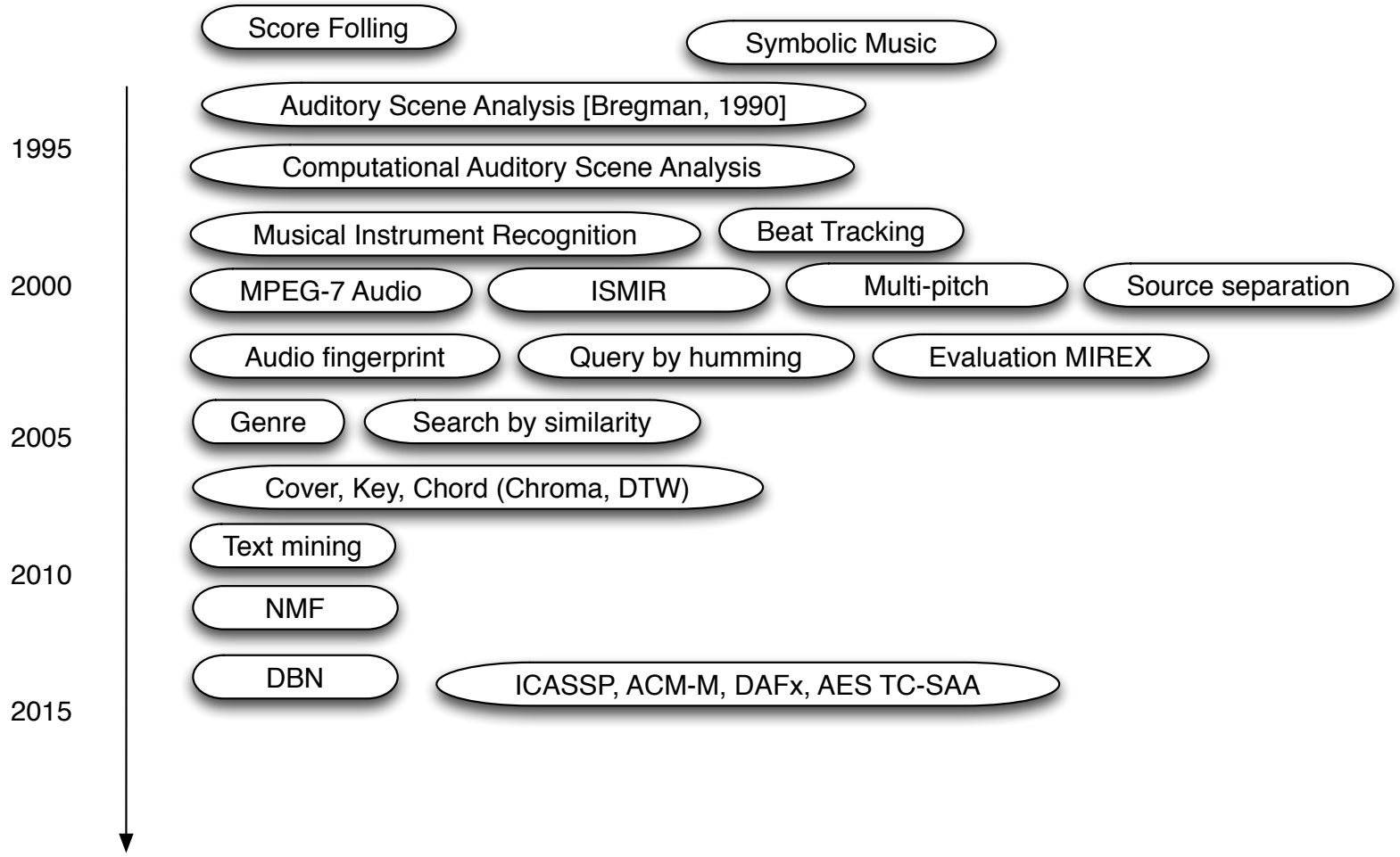
- Conferences
  - ISMIR, special sessions at ICASSP, ACM-M, AES TCAA
- Evaluation
  - On one million titles
- Applications
  - Shazam/MIDOMI
- Music
  - Listening through streaming (YouTube, Last-FM, Spotify, Deezer),
  - Meta-data provided by web-services (Echo-Nest, BMAT)





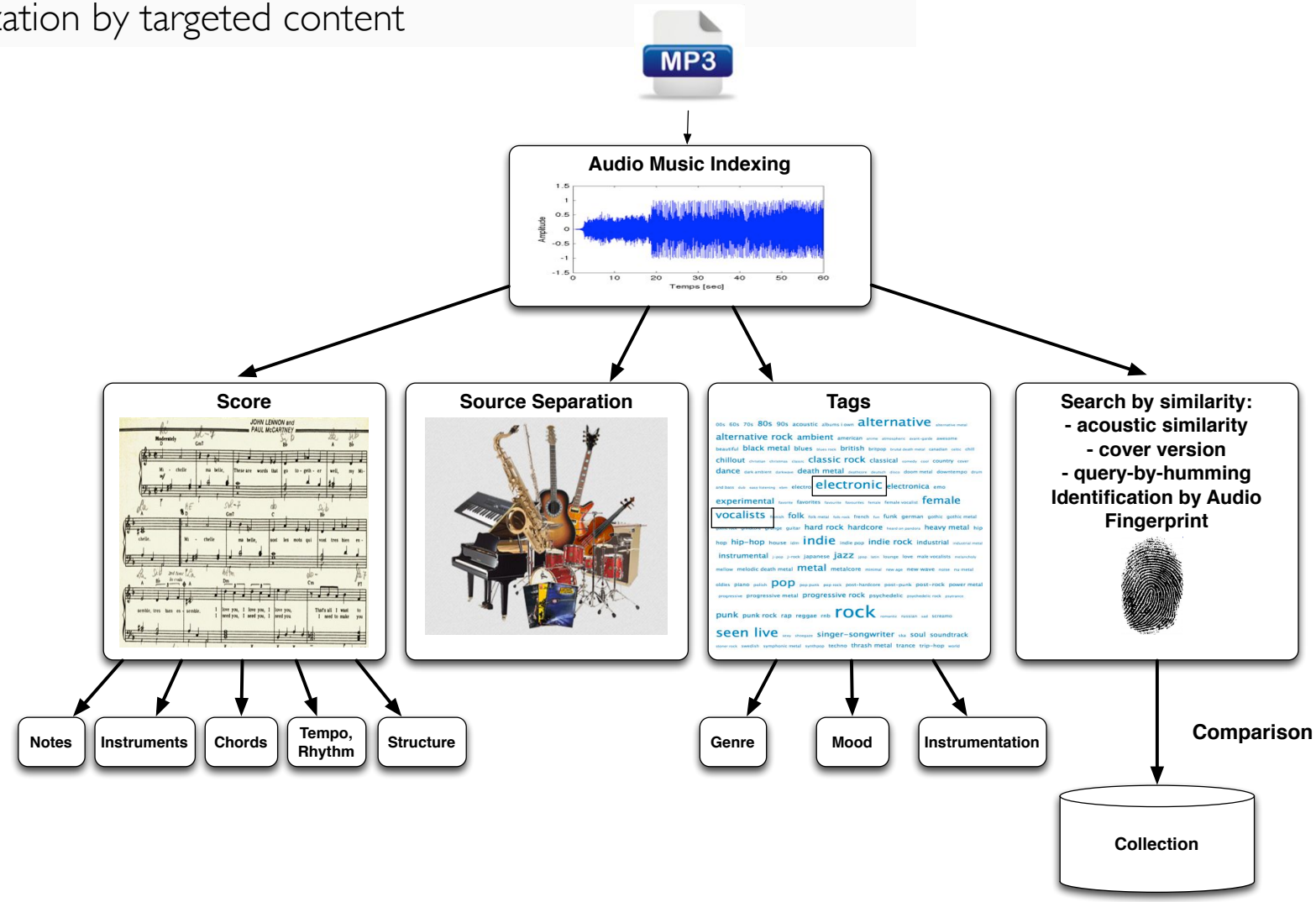
peeters@ircam.fr

MIR evolution over time



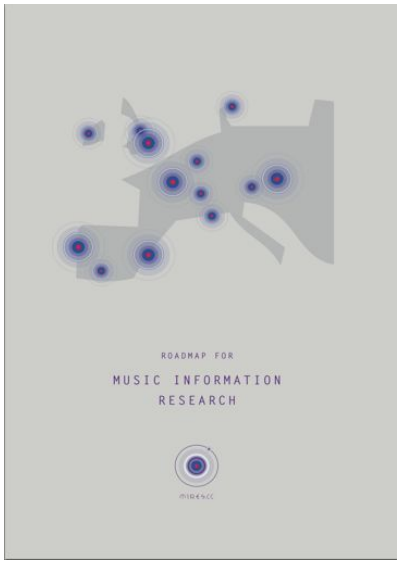
peeters@ircam.fr

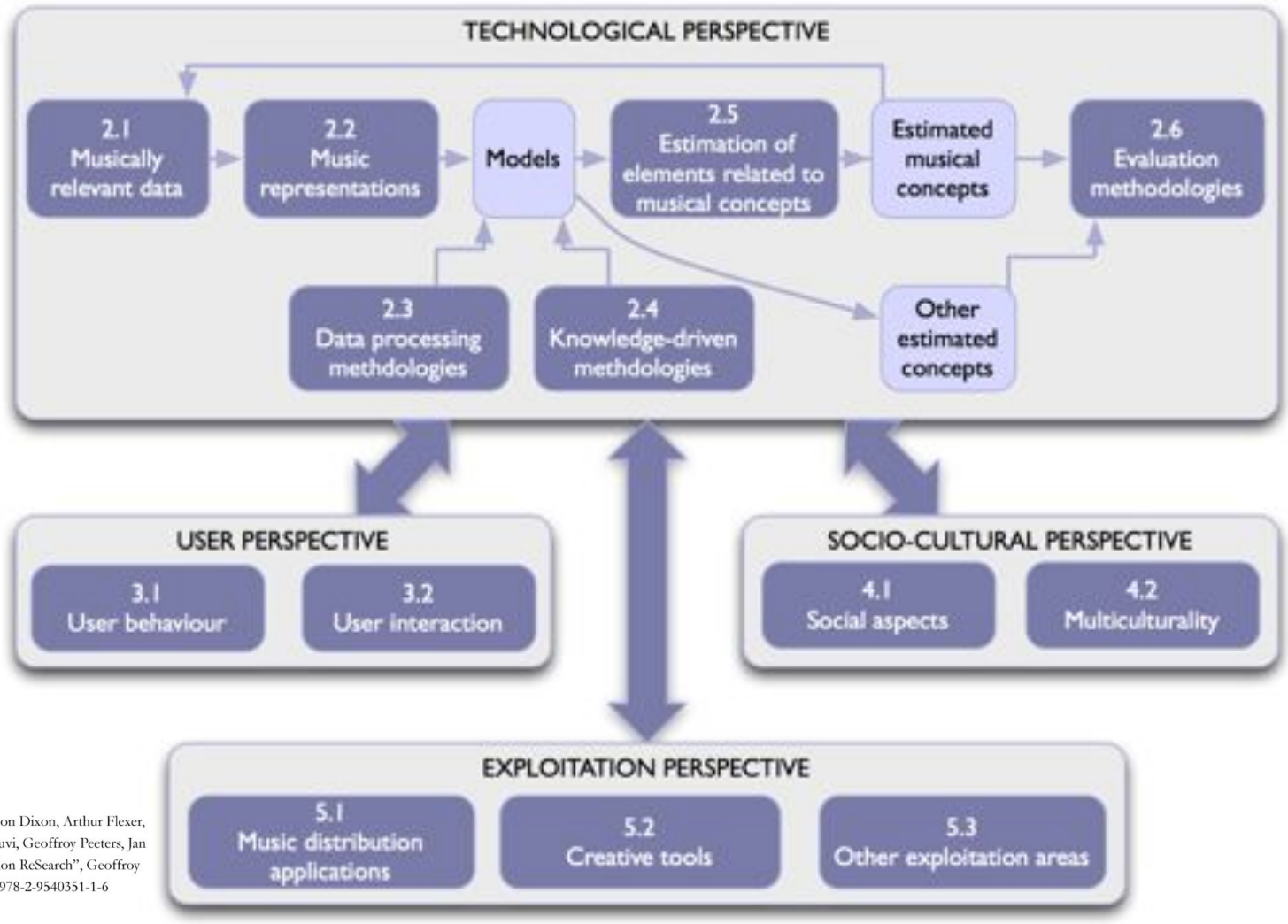
Organization by targeted content



peeters@ircam.fr

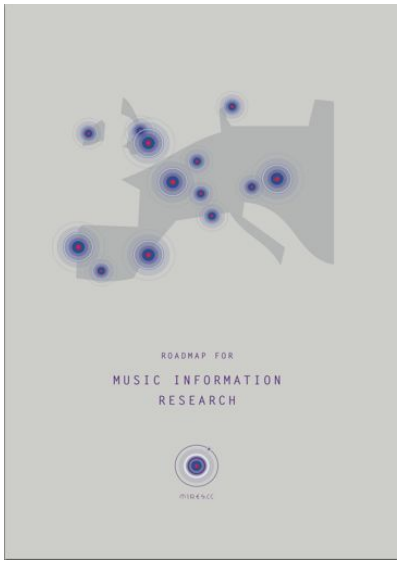
Introduction: « Roadmap for Music Information Research »

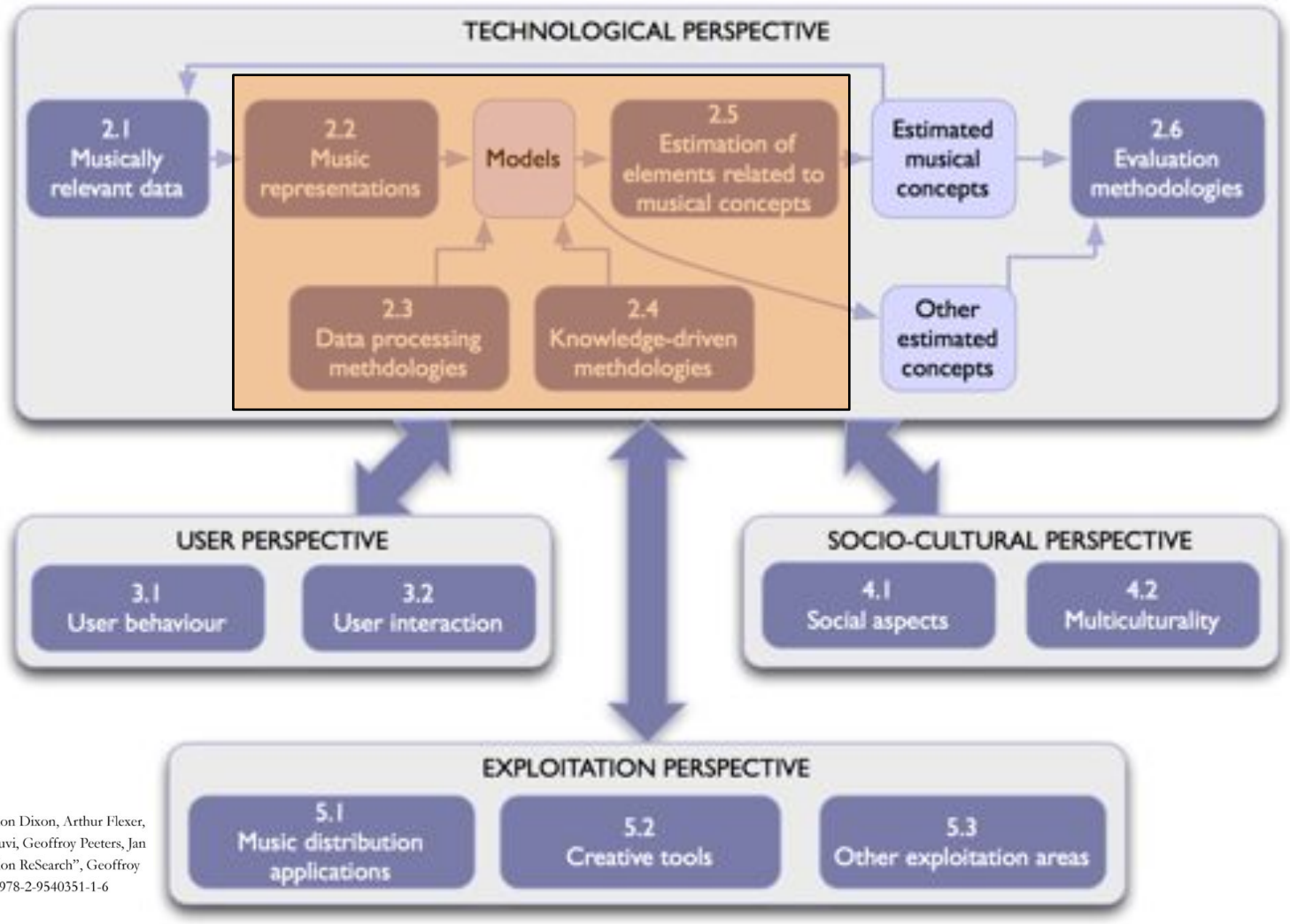




Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

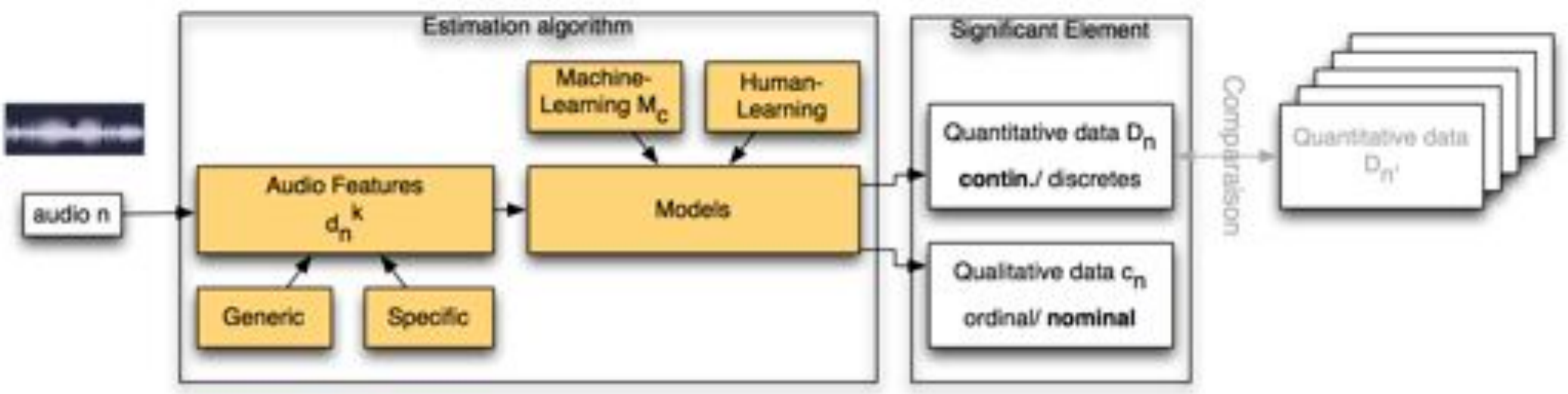
Introduction: « Roadmap for Music Information Research »





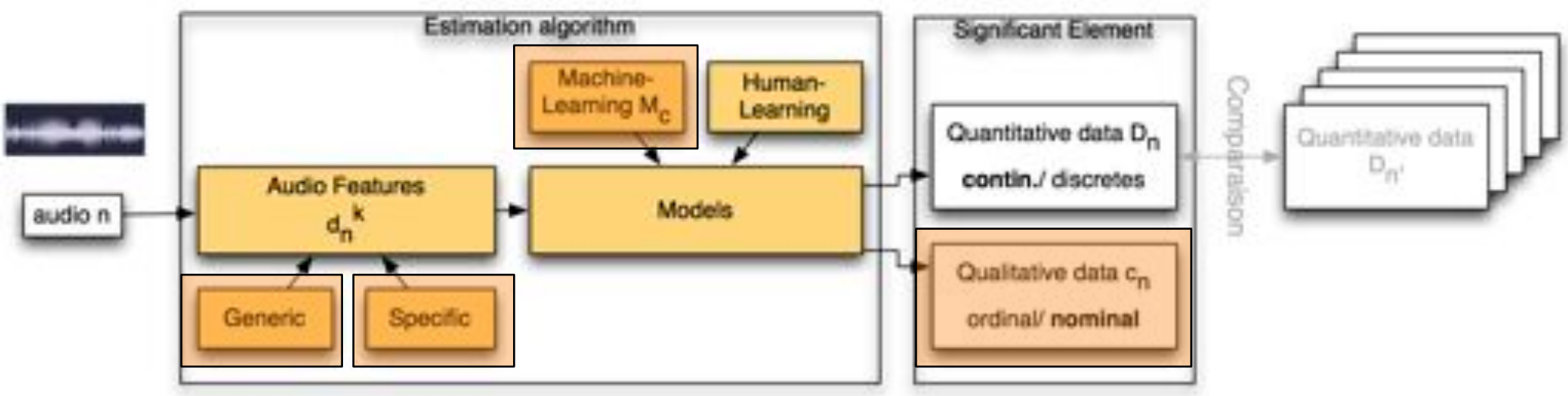
Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

M.I.R. Techniques

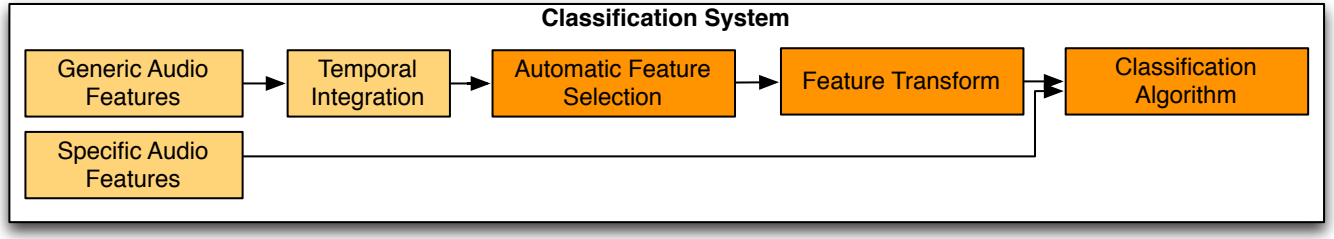


peeters@ircam.fr

M.I.R. Techniques



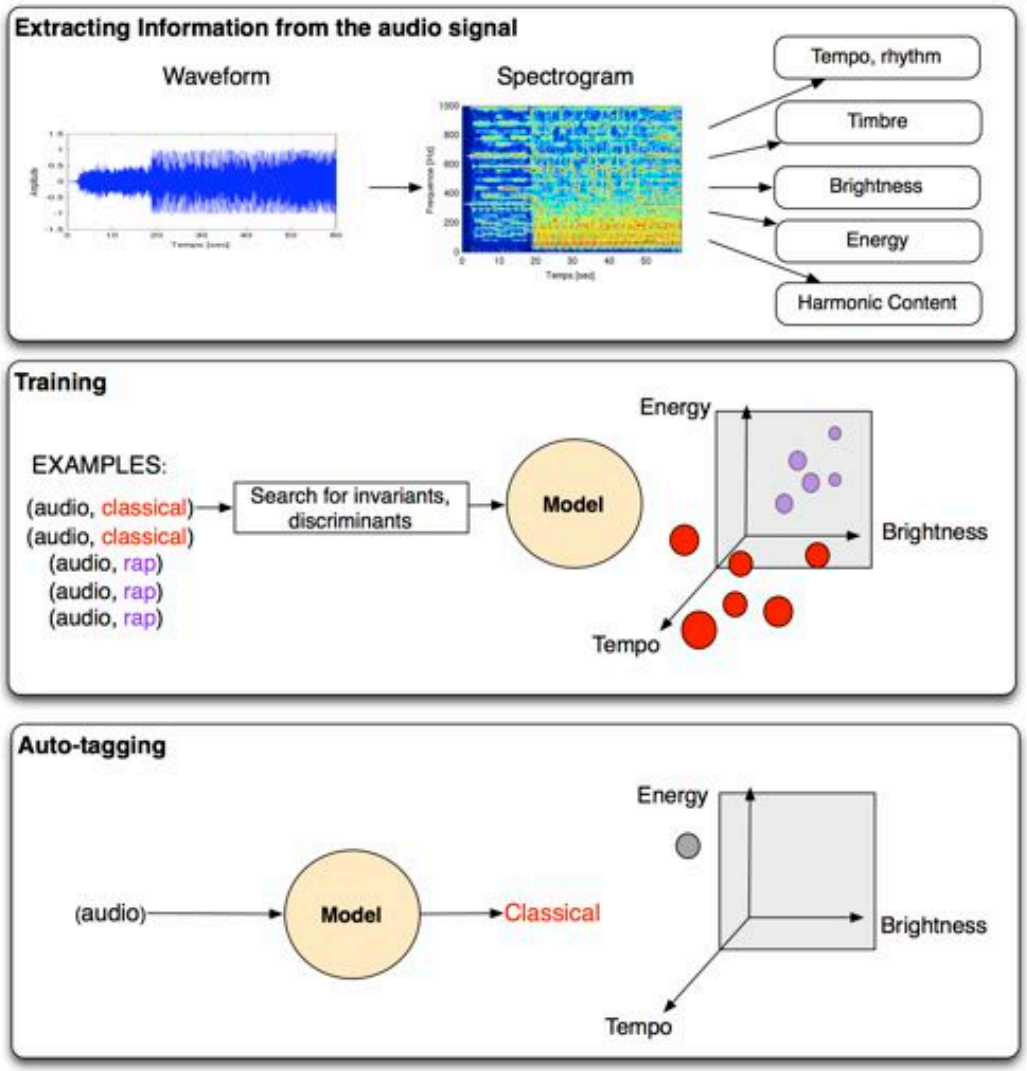
Machine-Learning for class problems



peeters@ircam.fr

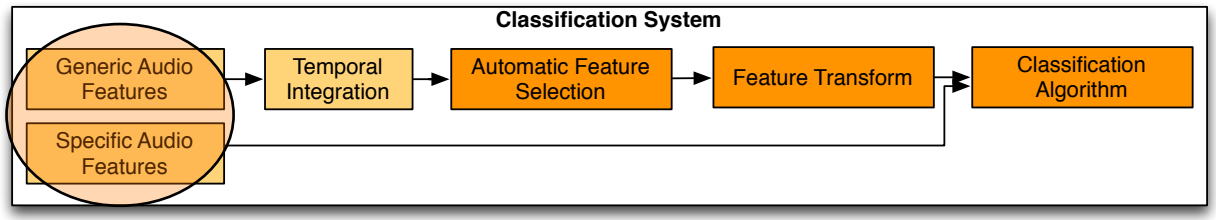
Machine Learning for Class Problems

- Simplified working flow
  - Extraction
  - Training
  - Auto-tagging

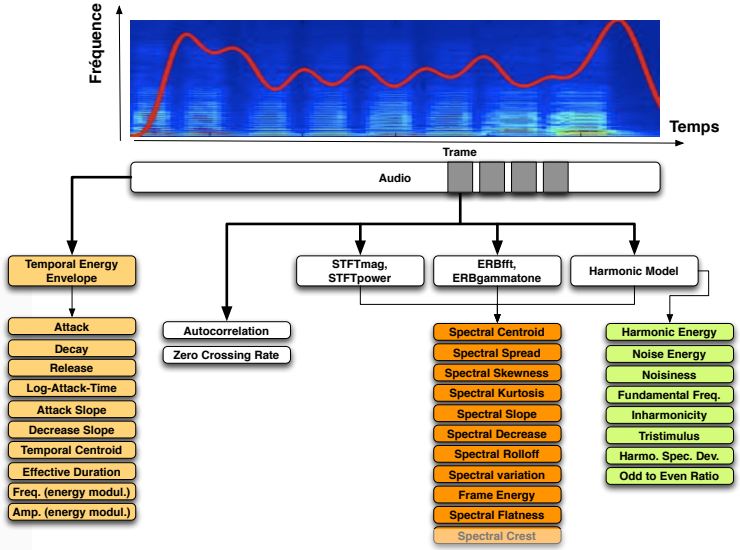


peeters@ircam.fr

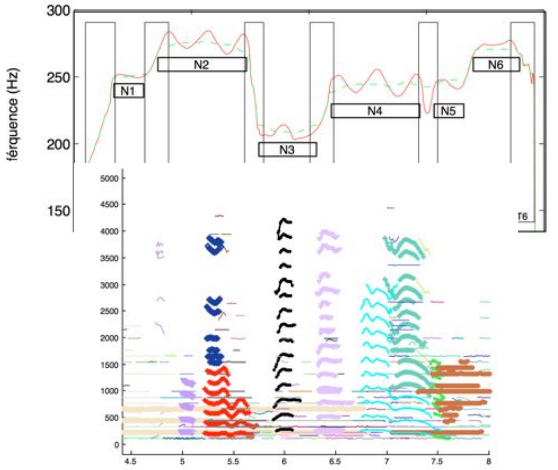
M.I.R. algorithms are « systems »



- Systems:
  - Many stage:, many possible choices at each stage
- Choice of the audio feature
  - Generi cAudio Features
    - MFCC [Rabiner]
    - Chroma/PCP [Bratsch,Wakefield], CENS, CRP [Mueller]
    - Block Features [Seyerlehner]
    - "A Large set of Audio Desciptors for …" [Peeters]
  - Specific Audio Features
    - Odd to even harmonic ratio
    - Intonative features [Regnier]
  - Automatic Feature Design
    - EDS [Pachet]
    - Deep Believe Network [Hamel,Humphrey]



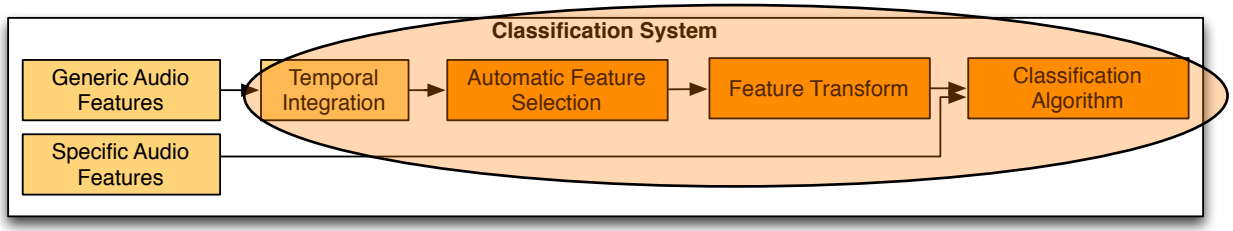
Ircamdescriptor flowchart



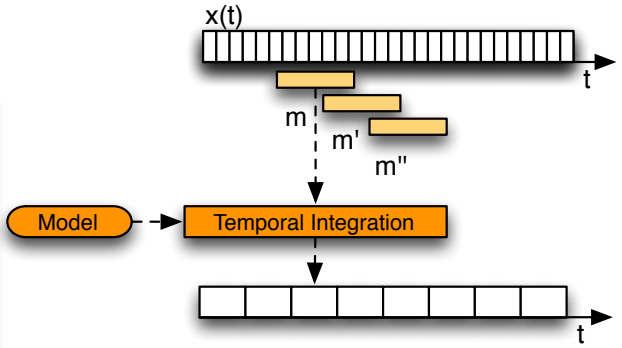
Intonative features [Regnier]

peeters@ircam.fr

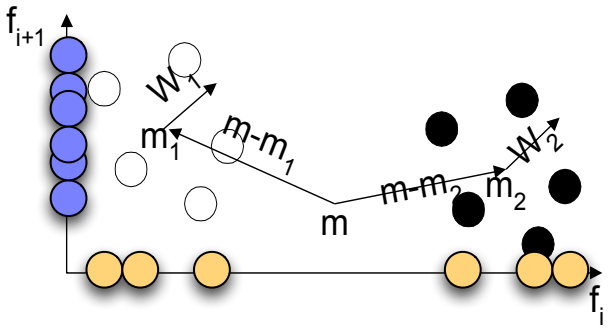
M.I.R. algorithms are « systems »



- Choice of the feature processing (temporal integration)
  - Texture window (short-term statistical moments)
  - Super-vector, i-Vector, Multi-dimensional Auto-Regressif Model
- Choice of feature Selection
  - Filter, Embedded, Wrapped
- Choice of the classification paradigm:
  - Single-label / multi-label (binary relevance, power-set)
- Choice of the machine learning algorithm
  - GMM, SVM, K-NN, Random Forest
- Segmentation over time:
  - Segmentation (BIC/AKAIKE) then classification of the segments
  - or Segmentation by class variation over time



Temporal integration of features

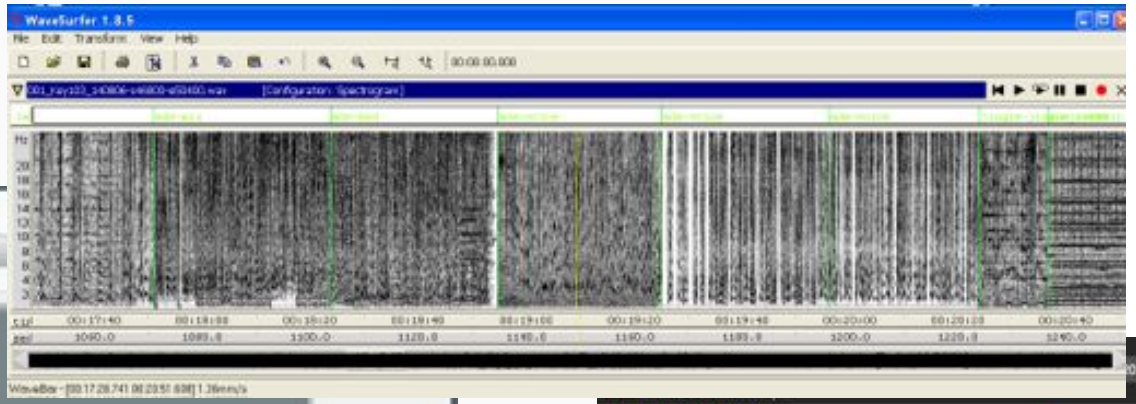


Automatic Feature Selection:
IRMFSP [Peeters]

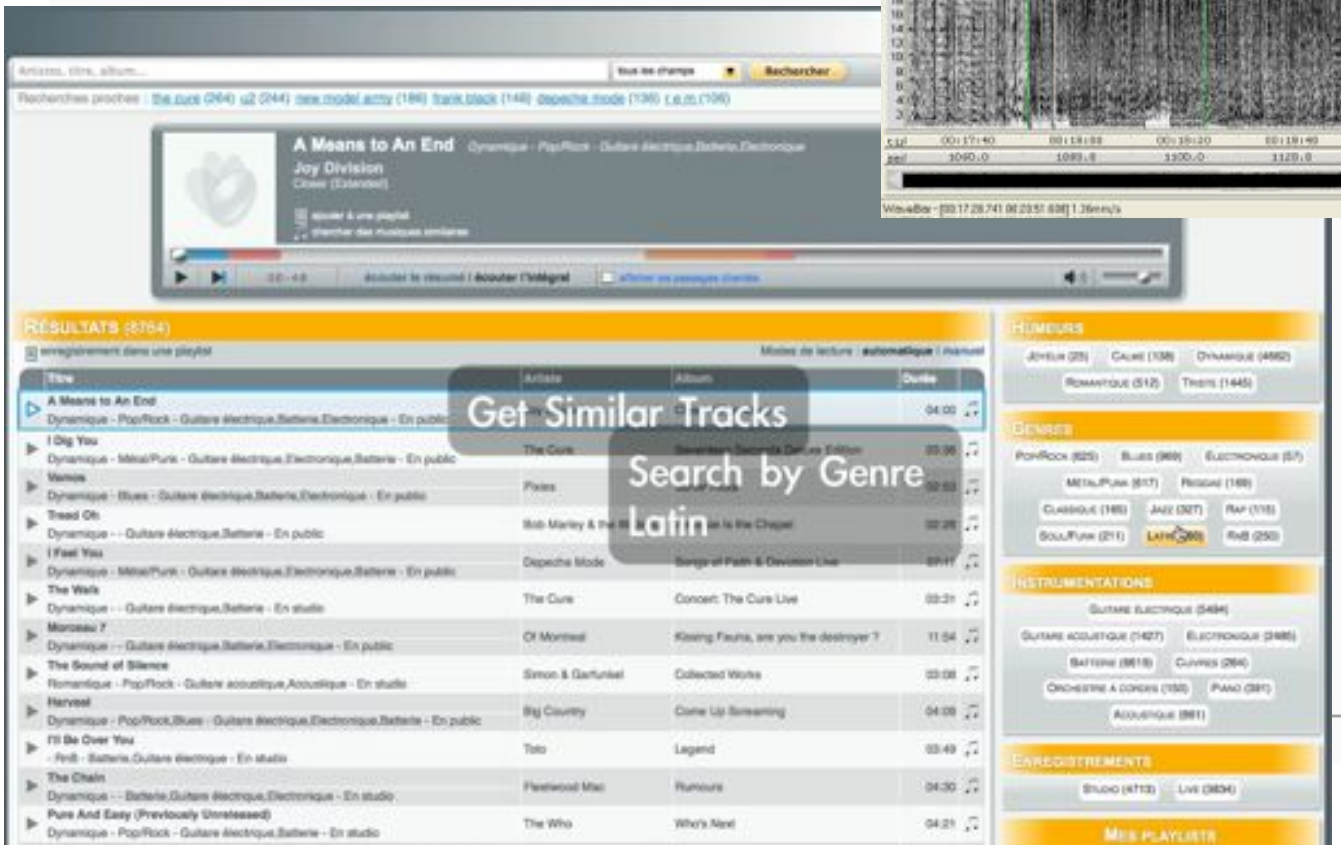
peeters@ircam.fr

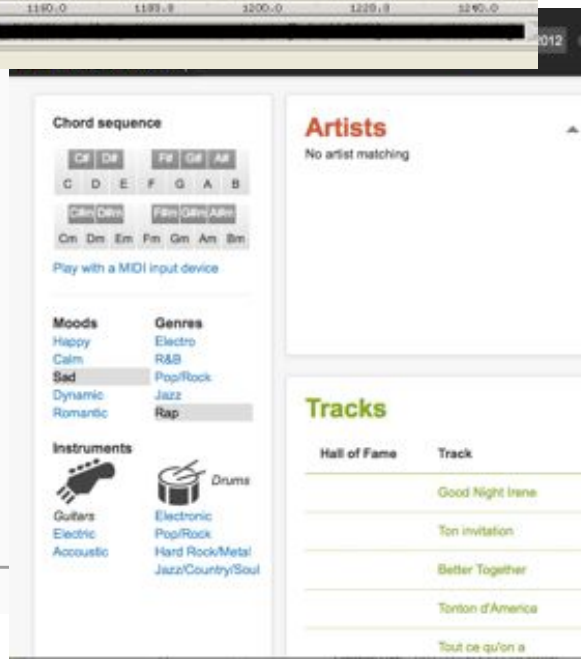
M.I.R. applications: Auto-tagging, Temporal-segmentation

VIDEO Time Segmentation



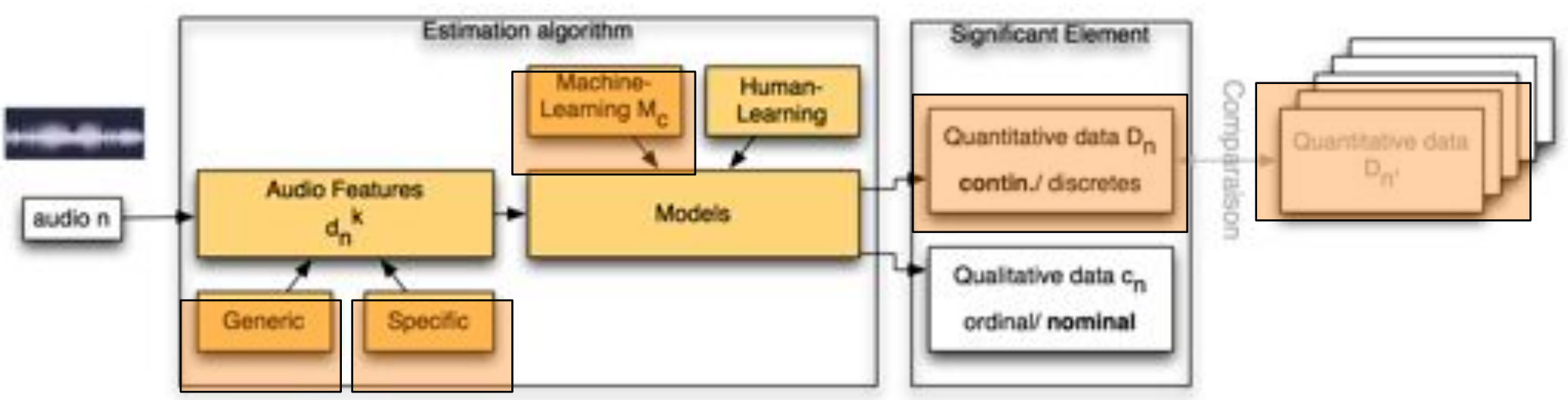
VIDEO Quaero Orange MSSE





VIDEO Quaero Exalead MUMA

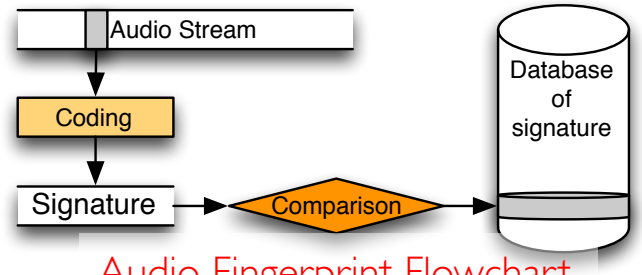
Content comparison problems



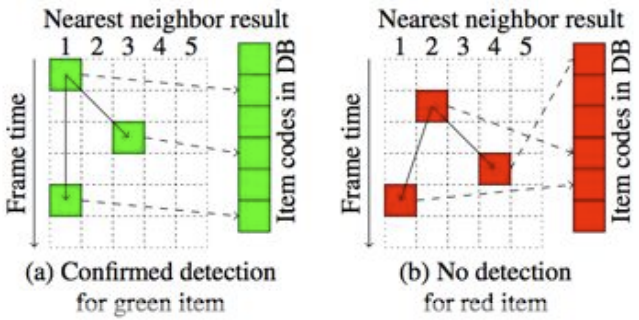
peeters@ircam.fr

M.I.R. Applications: Identification by Audio Fingerprint

- Audio identification
  - Watermarking
  - Fingerprint
- Audio fingerprint
  - Coding: from each existing ISRC store the corresponding fingerprint in a table
  - Decoding: compare the fingerprint of the target track to the fingerprints of the table
- Many existing systems
  - Shazam, Philips, FHG, B-Mat, …
- Challenges
  - Fingerprint must be robust (to audio degradation), discriminant (not to confuse two different ISRCs), scalable
  - Search must be highly scalable (>1e9 fingerprints in the table)
- Applications
  - Live-audio identification (user or copyright management)
  - Cleaning collection
  - Synchronization



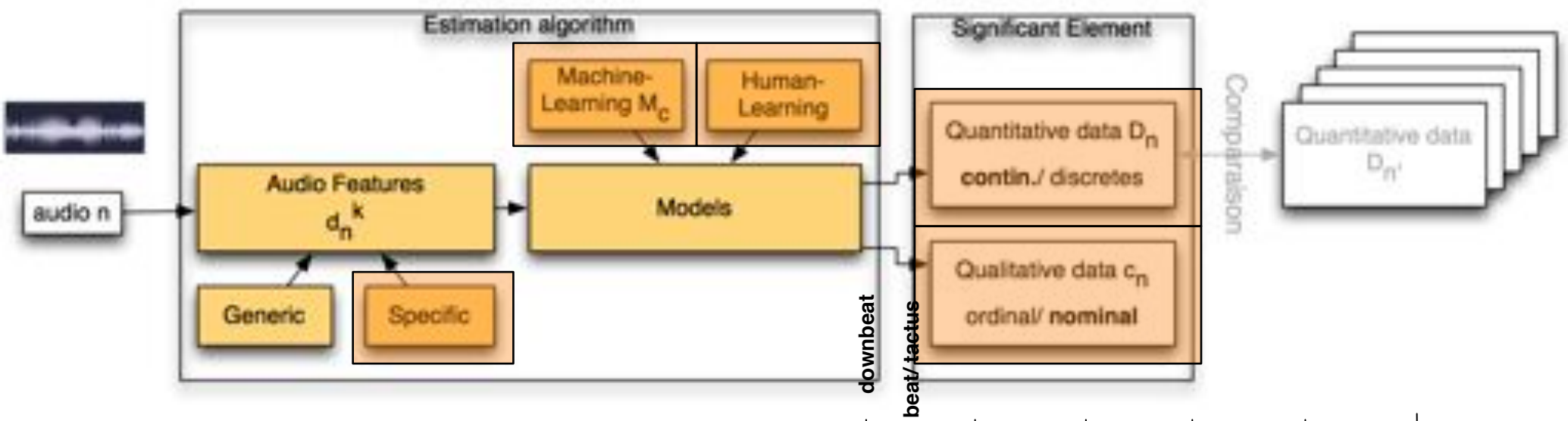
Audio Fingerprint Flowchart



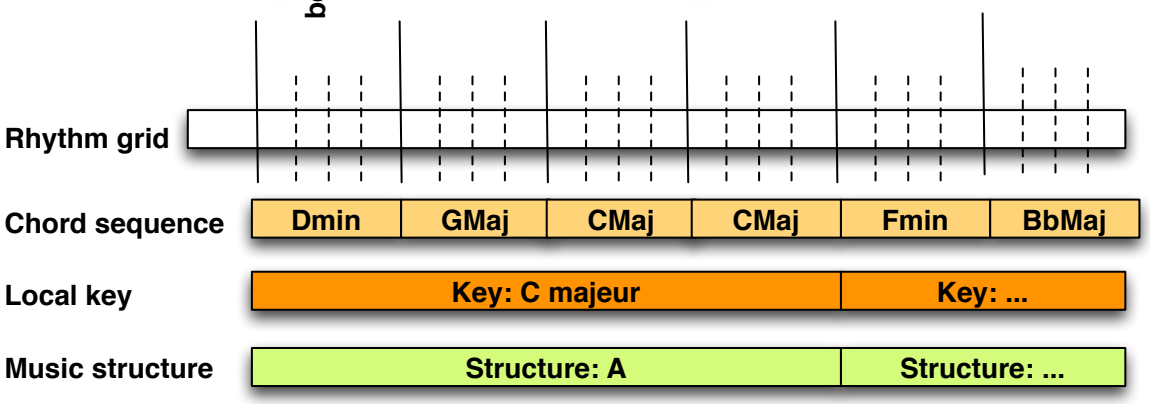
AudioPrint post-processing [Ramona]

VIDEO AudioPrint for Lyrics Alignment

peeters@ircam.fr



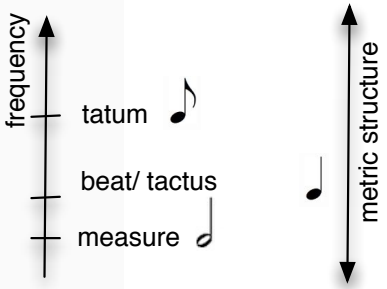
Musical and local concepts
Rhythm content



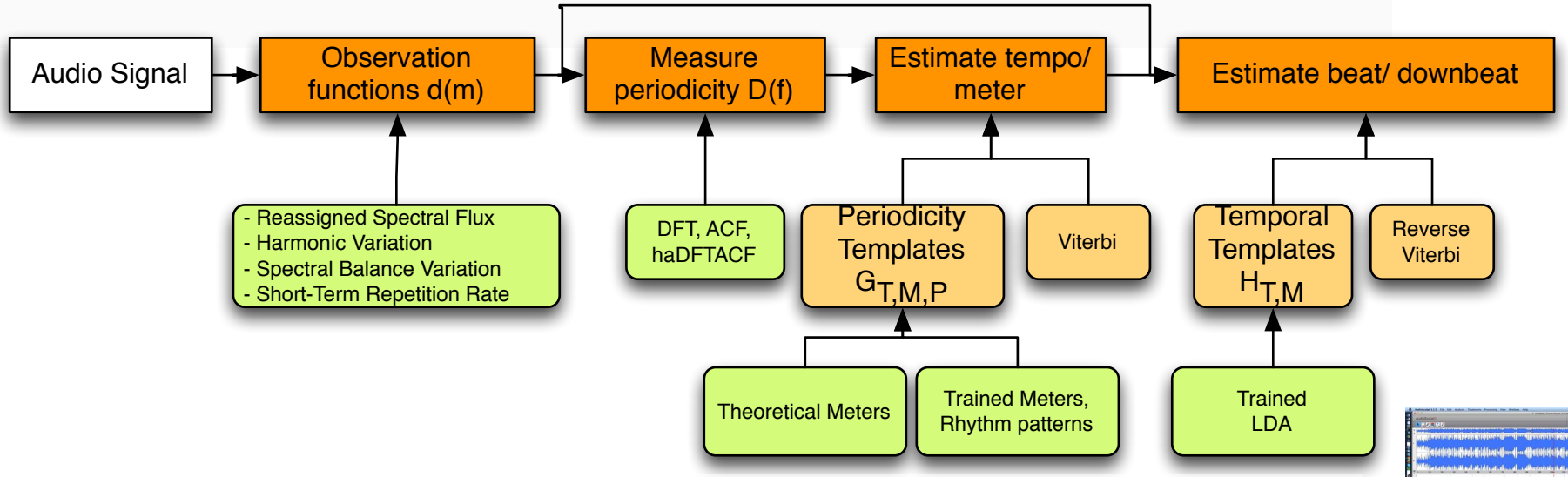
peeters@ircam.fr

M.I.R. Applications: Tempo/ Meter/ Beat/ Downbeat Tracking

- Various approaches for beat-tracking:
  - **Cognitive motivation**: Scheirer [works well for simple music, but …]
  - **Knowledge-based**: Klapuri, Peeters [need to introduce the rules for each music style]
  - **Purely Machine Learning:** Bock [recurrent Neural Network]



- Challenges of beat-tracking
  - Non-pre-eminence of events
  - Ambiguity of metrical-level to estimate
  - Rhythmic complexity
  - Temporal variability



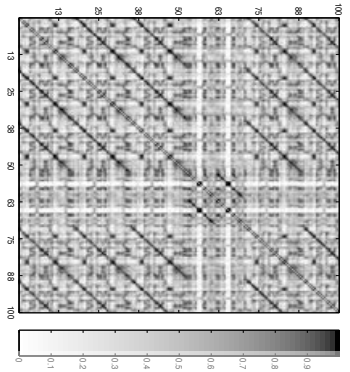
Global flowchart of ircambeat for tempo/meter/beat and downbeat estimation

VIDEO ircambeat in Audiosculpt 3.0

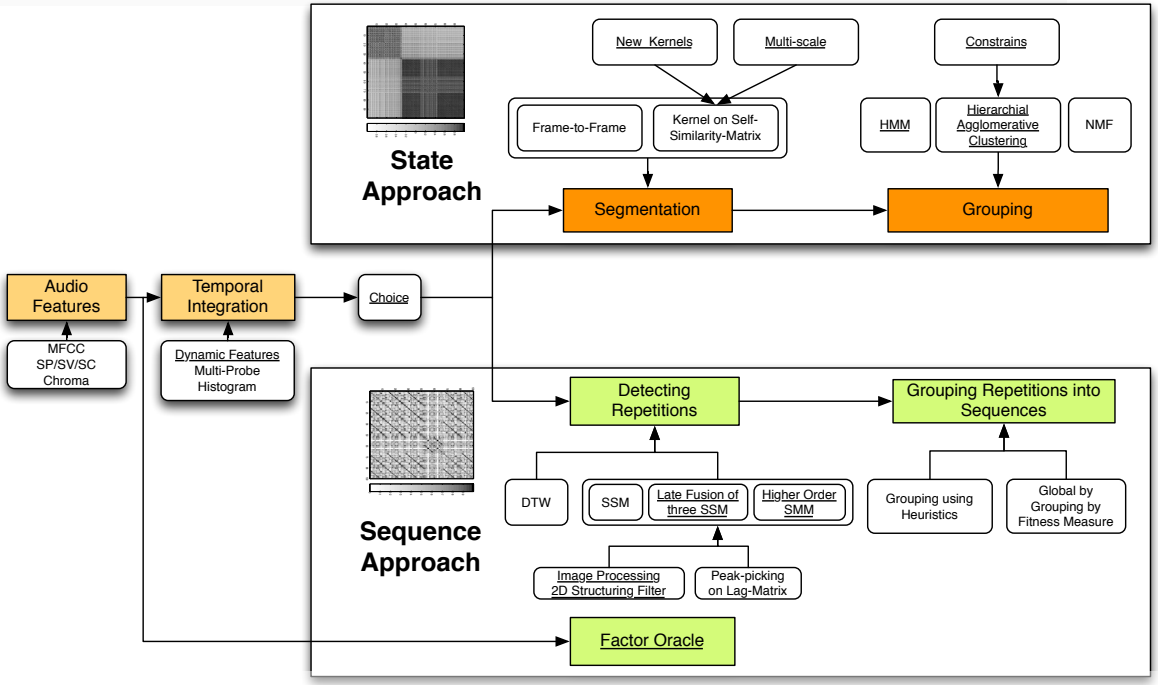
peeters@ircam.fr

M.I.R. Applications: Music Structure Discovery / Audio Summary

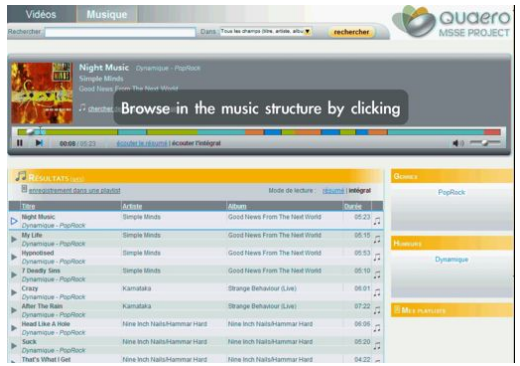
- Objective:
  - Find the underlying structure of the structure (verse, chorus)
  - Ill-defined problem !
  - Structure specific to each track -> non supervised learning
  - See Meinard Muller tutorial
- Use:
  - Understanding music
  - Interactive music players
  - Audio summary



Self-Similarity Matrix



Global flowchart of Ircamsummary for tmusic structure estimation and audio summary generation

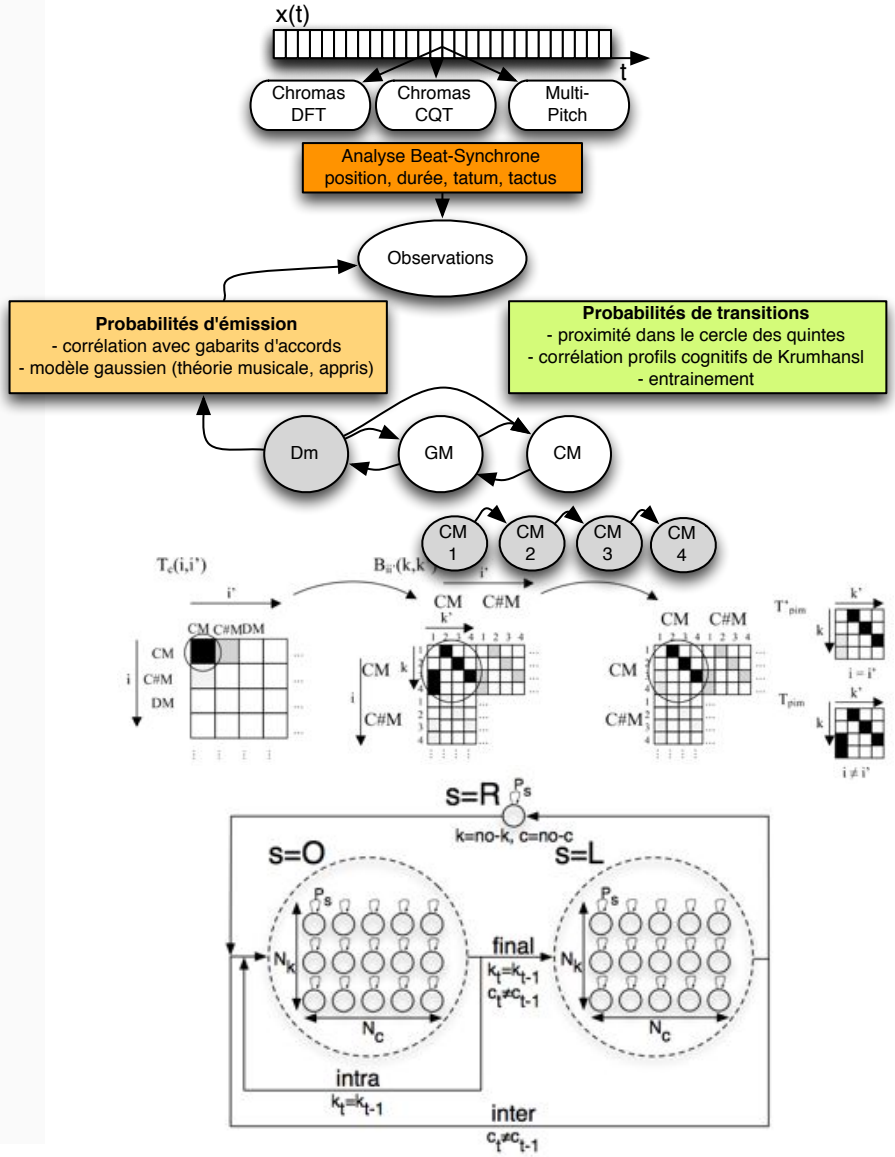


VIDEO: Browsing by Music Structure in Orange MMSE

AUDIO: Examples of audio summarie

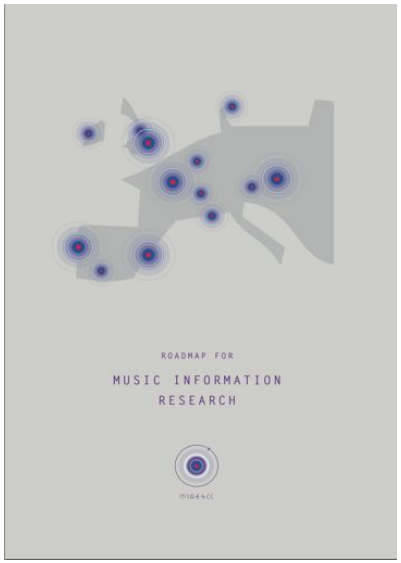
M.I.R. Applications: Key/ Chord Detection

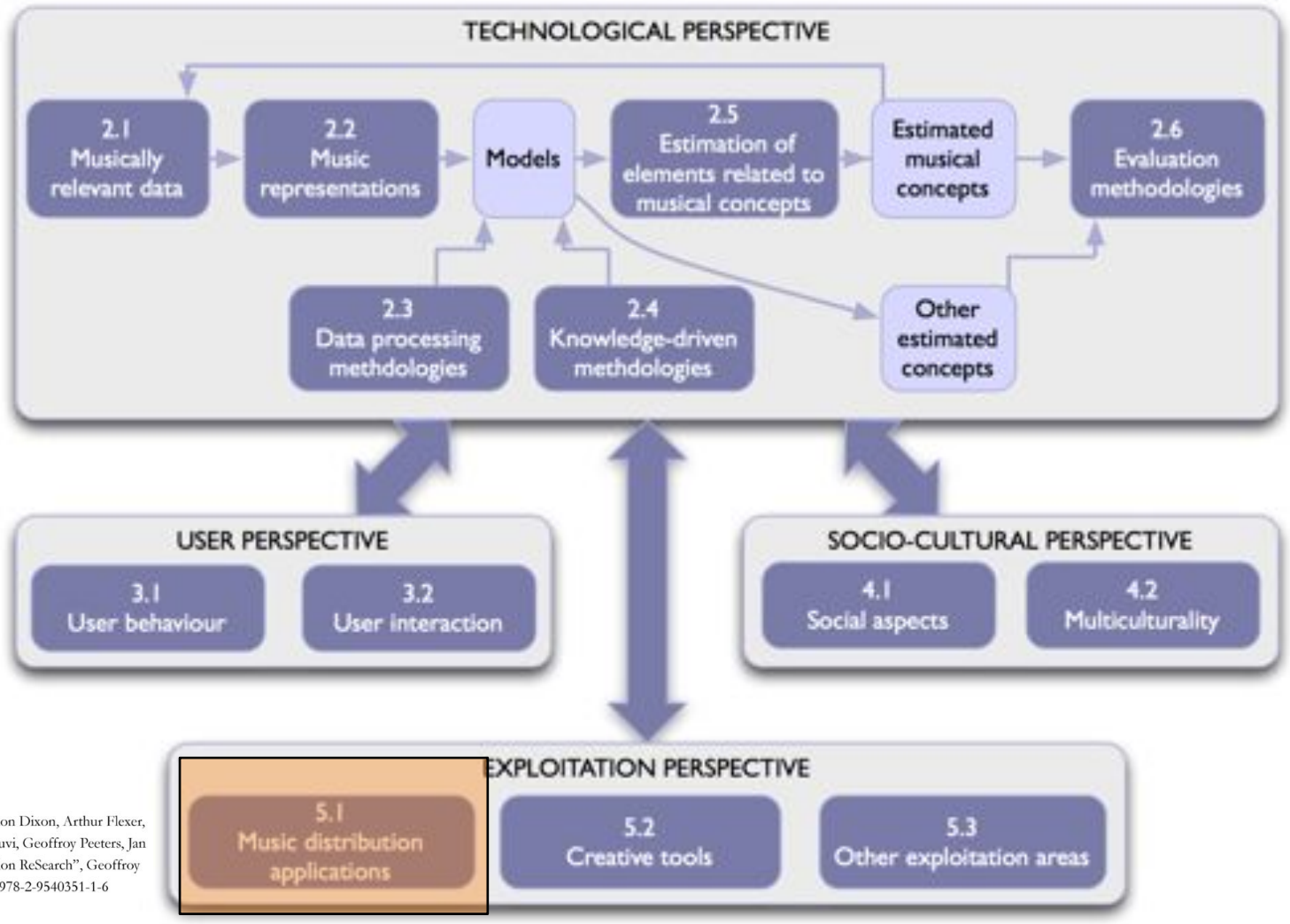
- How to encode musical knowledge ?
  - Key, chord: observation probability, template
  - Chords over time: hidden Markov model
- How to encode interdependence ?
  - Joint estimation of chords and downbeat [Papadopoulos]
    - Hypothesis: chord changes at 4/1 or 2/3
    - Chord → CPIM « Chord Position inside Measure »
    - Transition matrix
  - Joint estimation of chords, key and downbeat [Papadopoulos]
    - 24 HMM, one for each key
  - Joint estimation of structure, chords and key [Pauwels]
    - Chord bi-gram perplexity: final < inter < intra
- How to take into account errors ?
  - Possible errors of beat estimation are represented in the transition matrix [Papadopoulos]



peeters@ircam.fr

Introduction: « Roadmap for Music Information Research »



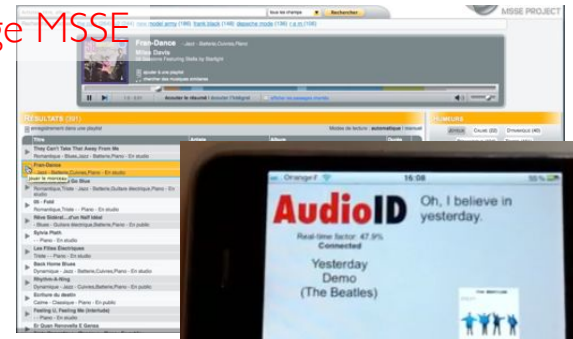


Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

M.I.R. Applications: Access

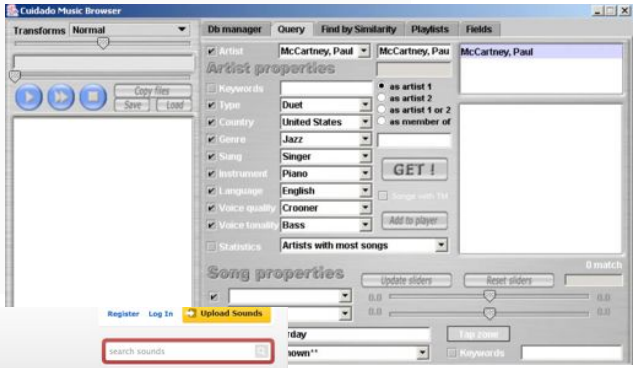
- Examples of applications
  - Quaero Orange MSSE Prototype
  - Quaero AudioPrint Lyrics Alignment
  - Sony CSL Music Browser
  - Semantic HIFI
  - MTG Freesound
  - BMat
  - EchoNest

Quaero Orange MSSE

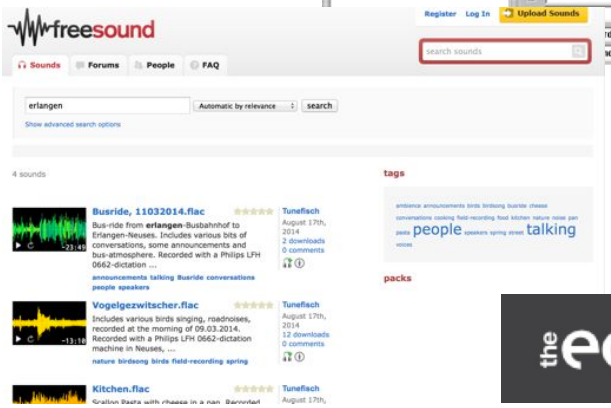


AudioPrint

Sony CSL Music Browser

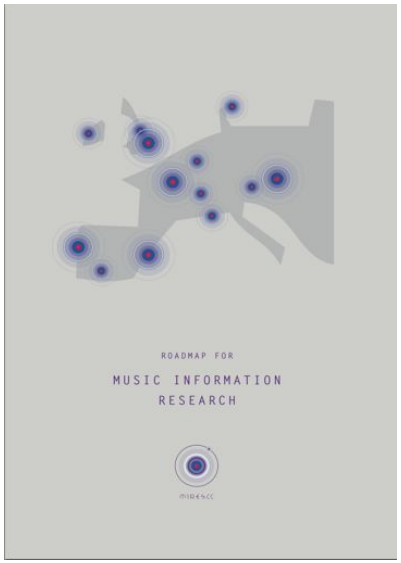


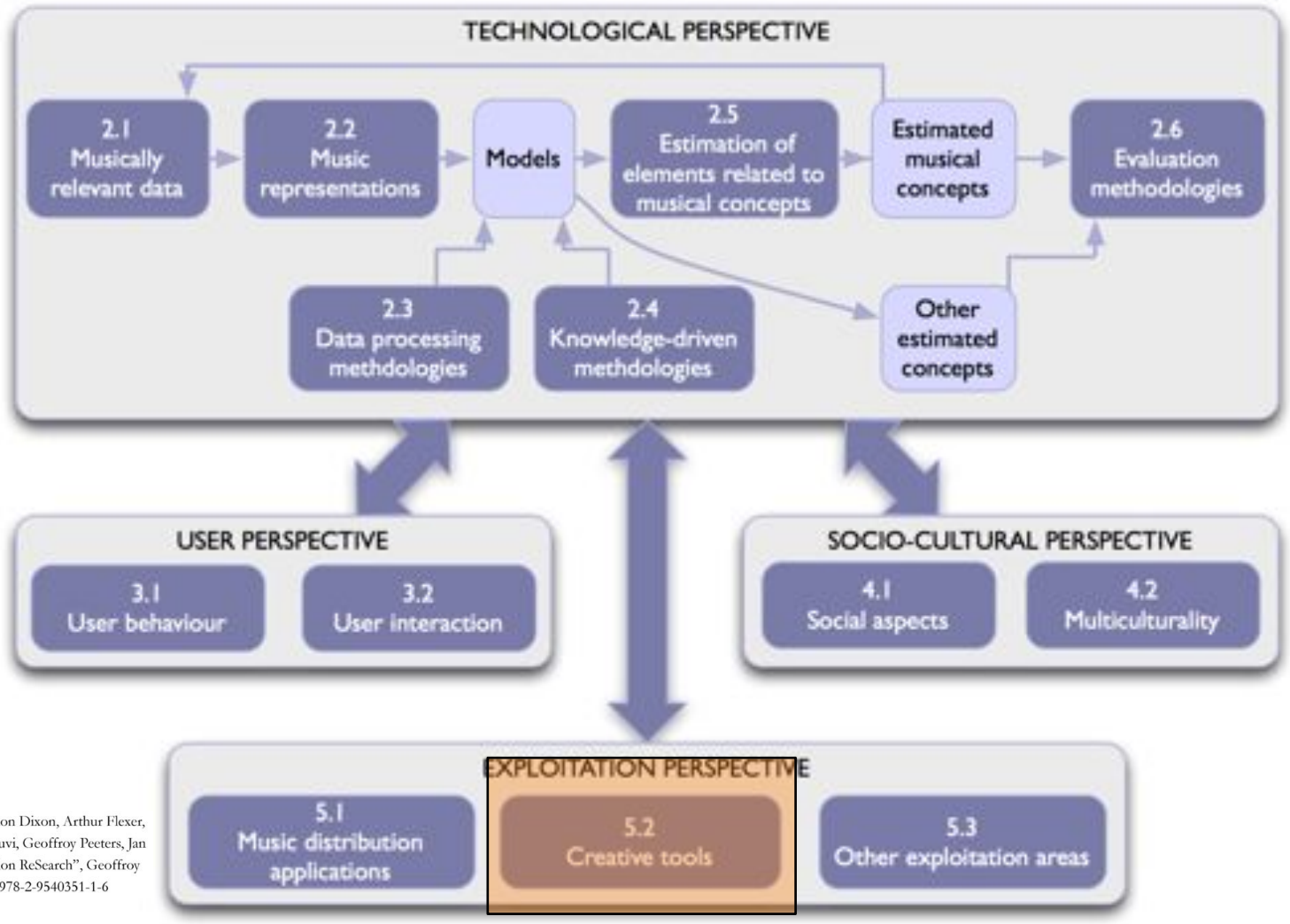
Semantic HIFI



peeters@ircam.fr

Introduction: « Roadmap for Music Information Research »





Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

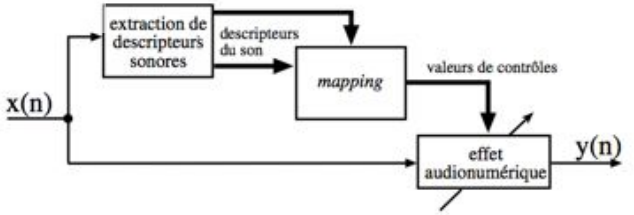
M.I.R. Applications: Creativity

- Initially MIR was developed for
  - Access (browsing, search, identification)
- Currently MIR developments include
  - Creativity

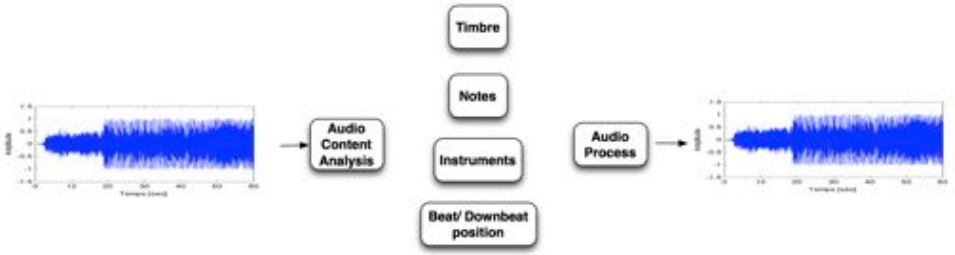
peeters@ircam.fr

Main current paradigms

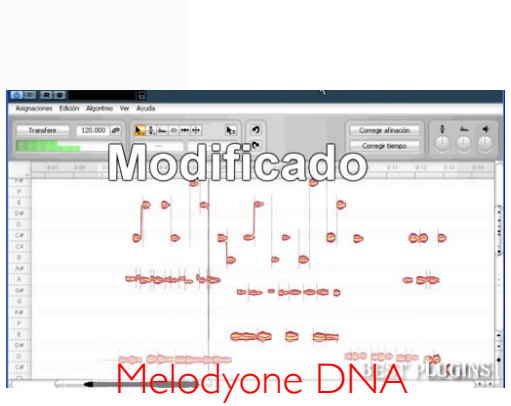
- In Audio
  - Modified the sounds using sound description
    - [Verfaille, Arfib] control soundFX using audio content (audio descriptors)
    - Vibrato control, attack removal



- In Music
- I. Music processing driven by a musical description
  - Changing tonal content (pitch)
    - Melodyne DNA
    - IRCAM Audio2Note (Ableton Live 9) [AUDIO]
  - Adding/ removing swing
    - Audiosculpt swing transform [AUDIO]
    - EchoNest : swinger, waltzify



Ableton Live 9 Audio2Note



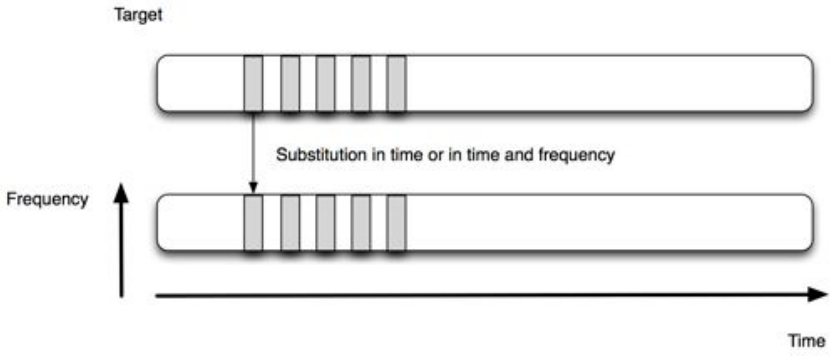
Melodyone DNA

Audiosculpt 3.0

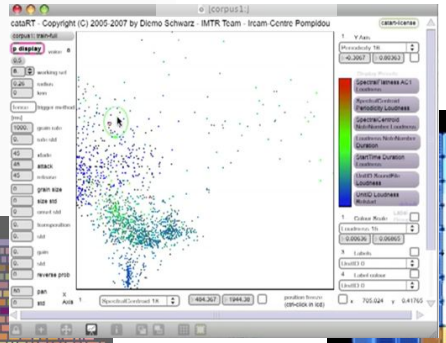
peeters@ircam.fr

Main current paradigms

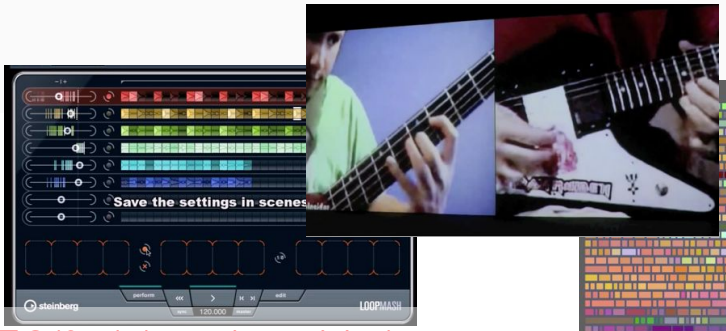
- II. Music slicing in time (or in time and frequency) using a music collection
  - Replacing elements in time:
    - MTG/Steinberg LoopMash [VIDEO]
  - Create in time: Musaicing, re-madonnisation [Zils, Pachet]
    - Corry Arcengel [VIDEO],
    - EchoNest Global Sampler, Bohemian Rhapsichord [VIDEO]
    - IRCAM CATaRT
  - Create in time and frequency [VIDEO IRCAM X-Micks]
  - Mix in time
    - Goto « MashUpper » [VIDEO]



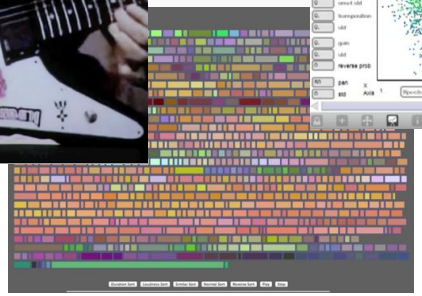
CATaRT



Corry Arcengel

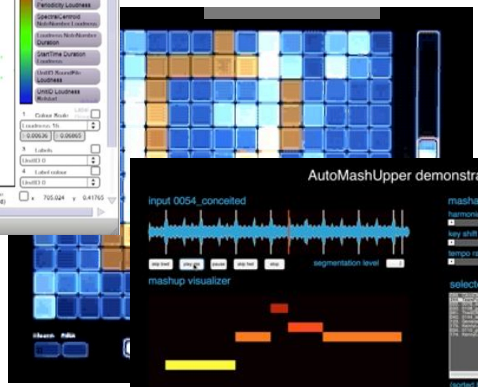


MTG/Steinberg LoopMash

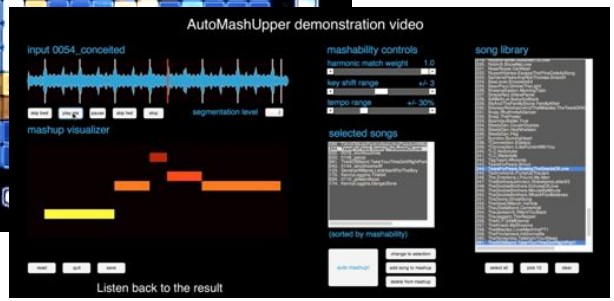


Bohemian Rhapsochord

X-Micks



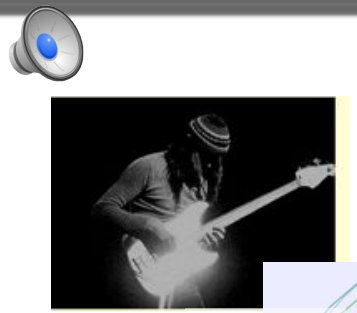
MashUpper

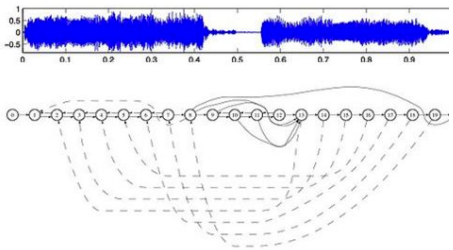


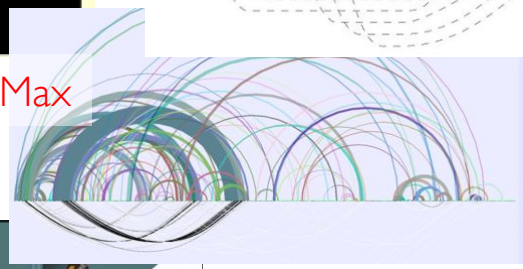
peeters@ircam.fr

Main current paradigms

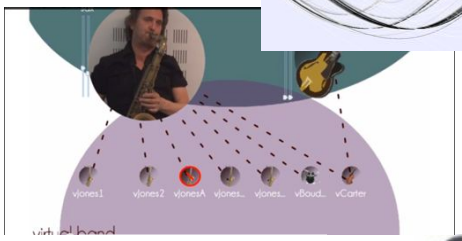
- III a. Learn the music model
  - Audio Factor Oracle [AUDIO, Pastorius], [VIDEO Ircam O-Max]
  - Constrained Markov Chains: Pachet, Vitual Band [VIDEO]







O-Max



Virtual Band

LaDiDa

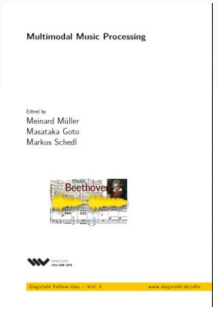
SongSmith

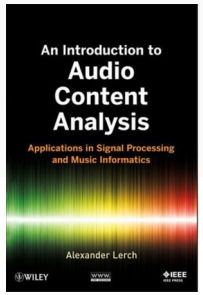
peeters@ircam.fr

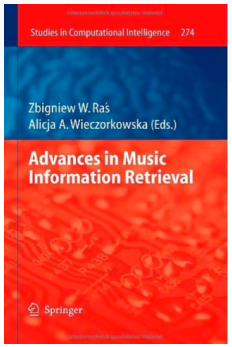
Community

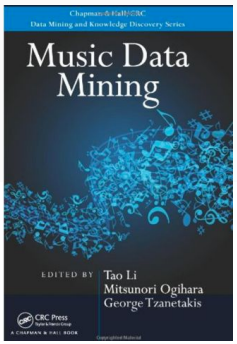
- Main conference: ISMIR
  - But also: ACM-M, ICASSP, AES TC-SAA, DAFx, …
- Mailing-list: music-ir
- Music Hack Days
  - With support by EchoNest, SoundCloud, …

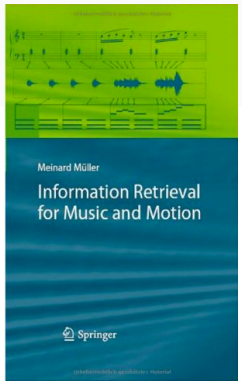
- Books
  - Mueller, Goto, Schedl, « Mutimodal Signal Processing », 2012
  - Alexander Lerch « Audio Content Analysis », 2012
  - Ras, Wierczorkowska, « Advances in Music Information Retrieval », 2012
  - Li, Ogihara, Tzanetakis, 'Music Data Mining », 2011
  - Mueller « Information Retrieval for Music and Emotion », 2007
  - Klapuri, Davy « Signal Processing Methods for Music Transcription », 2006

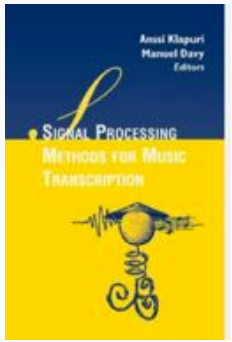








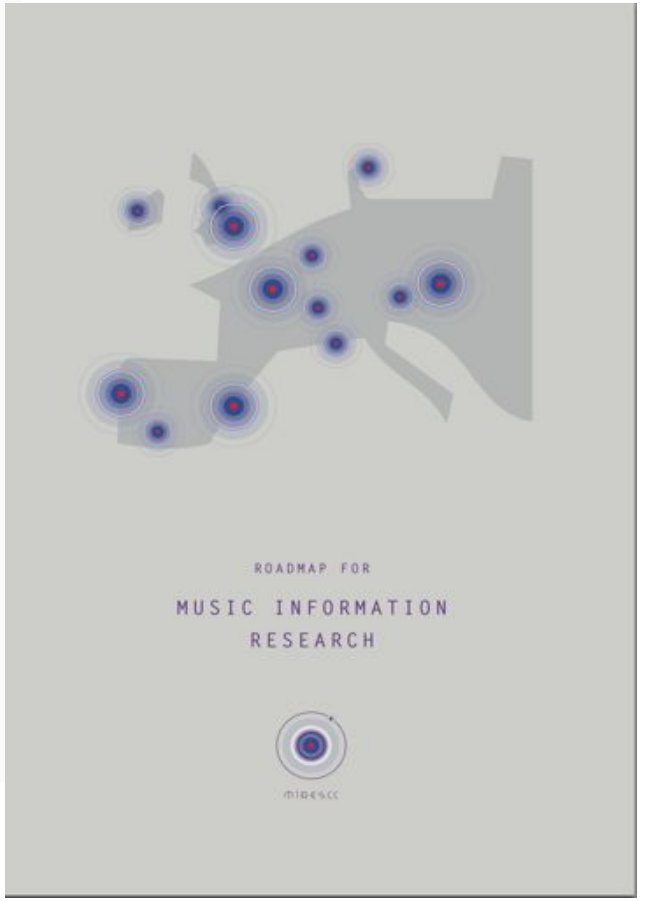




Community

- Book: Roadmap for Music Information Research
- MIReS project
- http://www.mires.cc/

  - MIReS project aims to create a research roadmap of **MIR field**, by **expanding its context** and **addressing challenges** such as multimodal information, multiculturalism and multidisciplinarity. MIR has the potential for a major impact on the future economy, the arts and education, not merely through applications of technical components, but also by evolving to address questions of fundamental human understanding, with a view to building a digital economy founded on"uncopiable intangibles": personalisation, interpretation, embodiment, findability and community. Within this wider context we propose to refer to the field of MIR as Music Information ReSearch (MIReS) and thus widen its scope, ensuring its focus is centered on quality of experience with greater relevance to human networks and communities.



Xavier Serra, Michela Magas, Emmanouil Benetos, Magdalena Chudy, Simon Dixon, Arthur Flexer, Emilia Gómez, Fabien Gouyon, Perfecto Herrera, Sergi Jorda, Oscar Paytuvi, Geoffroy Peeters, Jan Schlüter, Hugues Vinet, Gerhard Widmer, "Roadmap for Music Information ReSearch", Geoffroy Peeters (editor), 2013, Creative Commons BY-NC-ND 3.0 license, ISBN: 978-2-9540351-1-6

peeters@ircam.fr

Thank you. Questions ?