Sound and Music Description

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra, Barcelona
Index

- Introduction
- Sound and music content analysis
- Towards a semantic description
- Applications
Introduction
Music Information Retrieval

A (much simplified) MIR map

User

Rights owner

Taste, mood

Annotation

DRM

Metadata

Playlists

Genre

Motives

Features

Symbolic

Digital

Physical

Abstract

Fingerhut, 2004

Musical

Acoustical

Concept

Sign

Information

Stored data

Music

Cognitive, social

Semantic

Fingerhut, 2004

Xavier Serra - 2008
# Taxonomy of musical features

<table>
<thead>
<tr>
<th>STRUCT</th>
<th>CONCEPT LEVEL</th>
<th>MUSICAL CONTENT FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CONTEXTUAL</strong></td>
<td>HIGH II</td>
<td>EXPRESSIVE</td>
</tr>
<tr>
<td>global beyond 3 sec</td>
<td></td>
<td>melody</td>
</tr>
<tr>
<td></td>
<td>HIGH I</td>
<td>FORMAL</td>
</tr>
<tr>
<td></td>
<td>key</td>
<td>tonality</td>
</tr>
<tr>
<td></td>
<td>profile</td>
<td>cadence</td>
</tr>
<tr>
<td></td>
<td>MID</td>
<td>PERCEPTUAL</td>
</tr>
<tr>
<td></td>
<td>successive intervallic pattern</td>
<td>simultane intervallic pattern</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOW II</td>
<td>SENSORIAL</td>
</tr>
<tr>
<td></td>
<td>pitch</td>
<td>time</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>LOW I</td>
<td>PHYSICAL</td>
</tr>
<tr>
<td></td>
<td>fundamental frequency</td>
<td>note duration</td>
</tr>
<tr>
<td></td>
<td>periodicity pitch</td>
<td>pitch deviations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fundamental frequency</td>
</tr>
</tbody>
</table>

Lesaffre et al., 2003

Xavier Serra - 2008
Disciplines involved in MIR

Music
- Music Theory & Analysis
- Music Cognition
- Musicology

Psychology
- Psychoacoustics
- Experimental Design & Statistics

Computer Science
- Computer Music
- Database Management
- Digital Rights Management
- Digital Libraries

Information Science
- Artificial Intelligence
- Signal Processing
- Information Retrieval

Engineering
- Human-Computer Interaction
- Electronic Devices
- Ontology Engineering
- Text Mining

Information Technology Group

Xavier Serra - 2008
Sound and Music Content Analysis
Audio content analysis

• **Content:** The implicit and explicit information that is related to a sound or a piece of music and that is embedded in the signal itself.

• **Goal:** Automatically describe and deal (search, edit, transform) with audio data in a meaningful way.
Audio content classification

Feature Extraction
- Decode
- Windowing
- FFT
- Log
- DCT
- MEL Scale
- MFCC

Training
- Labeled Examples
- Machine Learning
- Model

Classifying
- Unknown Examples
- Labeled Examples

Xavier Serra - 2008
Levels of description

- Low-level (signal-centered) descriptors: computed from the audio signal in a direct or derived (ex: spectral analysis) way: average energy, spectral centroid, MFCCs ....

- Mid-level (object-centered) descriptors: requiring an induction operation or data modeling: key, genre, instrument ...

- High-level (user-centered) descriptors: requiring a user model: mood (ex: happy, sad), ...
Facets of music content

- Timbre
- Rhythm
- Melody / Harmony
- Structure

Music Content Analysis
Structure description

- Partitioning the sound stream into *homogeneous* regions

- Detecting special roles for the segmented regions: intro, verse, chorus, bridge,

- Other segments can also be identified: instrumental / singing; solo / ensemble; chords...

(Ong, 2006)
Structure description

(Ong, 2006)
Tonal description

- Extract:
  - Melody (predominant melody or score)
  - Harmony (chords)
  - Key, modulations

- Much research is related to **automatic transcription** of music (*Klapuri PhD 2004*)
  - Fundamental frequency / Multipitch estimation (*de Cheveigné*)
  - Melody extraction (Predominant pitch, note segmentation)
  - Still unsolved, even for monophonic signals.

- Pitch class distribution of a piece

- Mid and high level features -> apply a tonal model / musical analysis (*Krumhansl, Leman, Temperley, ....*)
Tonal description

Audio Frame

Preprocessing

Spectral Analysis

Peak Detection

Spectral Peaks

HPCP Computation

HPCP vector

Normalization

Normalized HPCP vector

Normalized HPCP

Transient location

Frames: 4096 samples (93 Windowing
FFT

Local maxima
Amplitude threshold -100 dB

Compute the HPCP vector
size = 12, 24, 36,...

Normalize with respect
to its maximum value

HPCP Mean Vector

Correlation

Correlation Matrix

Max

KeyProfileMatrix

(K)

Mode

Tonality strength

Key Note
Rhythm description

Extraction of the metrical structure of a piece

(Gouyon, 2005)
Rhythm description

(Gouyon, 2005)
Towards a Semantic Description of Sound and Music
Semi-automatic annotation

Which tags should be used for the un-annotated song?
Semi-automatic annotation

Tags are suggested using content-based similarity

The user chooses the right ones from this limited set

- rock
- Fast
- Cute
- Indie
- 90s
- Weird
- twee
- guitar

- noise pop
- Fast
- Cute
- 80s
- Quirky
- Drums
- Sweet
- playful

- Fun
- rock
- pop

- thrash
- 90s
- metal
- Weird
- concert
- loud
- death
- heavy metal
- thrash rock
- gothic
- rock
- Edgy
- 90s
- Loud

Xavier Serra - 2008
**Cover detection**

* Research in:

- Descriptors sequences
- Descriptors similarity
- Sequence alignment

Applications:
audio segmentation, mood classification, perceptually-based descriptors similarity measures, song hierarchies, visualization, sequence prediction, rights management...
Cover detection

![Diagram showing the process of cover detection]

- Song A
  - HPCP Extract.
  - Global HPCP
  - OTI
- Song B
  - HPCP Extract.
- Pre-process.
- Binary Similarity Matrix
- Dynamic Program. Local Alignment
- Post-proc.
- Dist.
- Align.
Towards semantic descriptors

- Music complexity
- Genre
- Mood
- ????
Music complexity

- Acoustic complexity:
  - loudness fluctuations
- Timbre complexity
- Rhythm complexity -> “Danceability” descriptors
- Tonal complexity

Figure 1.3: The Wundt curve for the relation between music complexity and preference.
Sound description

Creation Information:
The Tape Gallery
London 1986

Media Information:
Media format
48Khz, WAV
Stereo,Duration
G:/SFX/SRC_Sclv0pt/t1andy1b/wav/golfsw00.wav

Usage Information:
Rights
The Tape Gallery
All rights reserved

CONTENT MANAGEMENT

CONTENT STRUCTURE

Whoosh:  
- Impulsive
- Unpitched

Hit:  
- Impulsive
- Unpitched

Gurgle:
- Iterative
- Complex

CONTENT SEMANTICS

Concept
Golf

Agent
Golfer

Event
drive

Swing
Golf stroke

object
Club
Iron

object
Hole

Place
Golf course
Grass

object
Ball
Hit
Towards a cognitive approach

Peretz and Coltheart, 2003
Towards a convergence approach

- Creation (production tools)
- Content (databases)
- Community (social tools)
- Music 3.0

Content processing
Collaborative creativity
Content sharing
Applications
Application: music retrieval

- Efficient management of sound archives, music retrieval, …
Application: personal annotation

Good Vibrations, a Winamp plugin for building “persononomies” and automatically annotating collections (Sandvold, Celma, Herrera, 2005)
Application: music recommendation

(Celma, 2006)
Application: music selection

music selection based on heart beat rate
Application: sound search

SearchSounds
The first content-based audio search engine

Keyword Search | Genres | Artists

traditional Irish

Search

Results 1 - 11 of 250 for traditional Irish

Play All Songs >>

Dogbite_LetTheCocaineBe(Traditional).mp3
http://www.jamesmorris.org/media/Dogbite_LetTheCocaineBe(Traditional).mp3
From: Fatal Film - Street Hair

StellarRoad2005-11-10T01_64kb.mp3
Title: Soundcheck Jam - Artist: Stellar Road - Album: 2005-11-10 - The Curragh Irish
From: Stellar Road Live at The Curragh Irish Pub on 2005-11-10

StellarRoad2005-11-10T01_vbr.mp3
Title: Soundcheck Jam - Artist: Stellar Road - Album: 2005-11-10 - The Curragh Irish
From: Stellar Road Live at The Curragh Irish Pub on 2005-11-10

StellarRoad2005-11-10T02_64kb.mp3
Title: Lucky Penny - Artist: Stellar Road - Album: 2005-11-10 - The Curragh Irish
From: Stellar Road Live at The Curragh Irish Pub on 2005-11-10

StellarRoad2005-11-10T02_vbr.mp3
Title: Lucky Penny - Artist: Stellar Road - Album: 2005-11-10 - The Curragh Irish
From: Stellar Road Live at The Curragh Irish Pub on 2005-11-10

StellarRoad2005-11-10T03_64kb.mp3
Title: Our Last Night (Summer Song) - Artist: Stellar Road - Album: 2005-11-10 - The Curragh Irish
From: Stellar Road Live at The Curragh Irish Pub on 2005-11-10

Xavier Serra - 2008
Application: music communities

http://freesound.org

Xavier Serra - 2008
Application: music interaction

BeatMash: sound synthesis based on mosaicing
Application: sound transformation

\[ \alpha > 1 \]

\[ \alpha < 1 \]

audio time scaling to add swing
Application: music interaction

**Reactable:**
musical instrument based on a tangible interface