Music Information Retrieval (MIR)

Approaches for content-based music retrieval

Alberto De Bortoli

December 1, 2009

What is MIR?

- Music information retrieval or MIR is the interdisciplinary science of retrieving information from music
- This includes:
  - Computational methods for classification, clustering, and modelling
  - Formal methods and databases
  - Software for music information retrieval
  - Human-computer interaction and interfaces
  - Music perception, cognition, affect, and emotions
  - Music analysis and knowledge representation
  - Music archives, libraries, and digital collections
  - Intellectual property and rights
  - Sociology and Economy of music

Dealing with Music

Music is different from text, we have to deal with its characteristics:

Pitch which is related to the perception of the fundamental frequency of a sound; pitch is said to range from low or deep to high or acute sounds.

Intensity which is related to the amplitude, and thus to the energy, of the vibration; textual labels for intensity range from soft to loud; the intensity is also defined loudness.

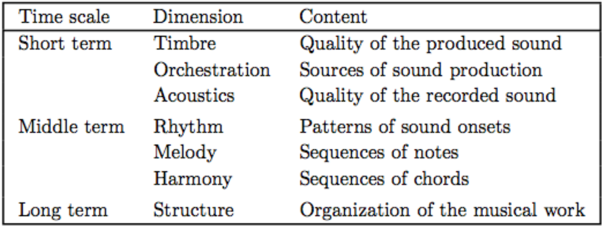
Timbre which is defined as the sound characteristics that allow listeners to perceive as different two sounds with same pitch and same intensity.

Dimensions of the Music Language (1)

- **Timbre** depends on the perception of the quality of sounds.
- **Orchestration** is due to the composers and performers choices in selecting which musical instruments are to be employed.
- **Acoustics** can be considered as the contribution of room acoustics, background noise, audio post-processing, filtering, and equalization.
- **Rhythm** is related to the periodic repetition, with possible small variants.
- **Melody** is made of a sequence of tones with a similar timbre that have a recognizable pitch within a small frequency range.

Dimensions of the Music Language (2)

- **Harmony** is the organization, along the time axis, of simultaneous sounds with a recognizable pitch.
- **Structure** is a horizontal dimension whose time scale is different from the previous ones, being related to macro-level features such as repetitions, interleaving of themes and choruses, presence of breaks, changes of time signatures, and so on.

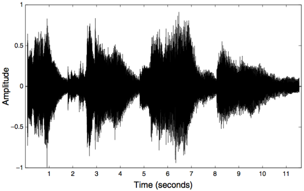


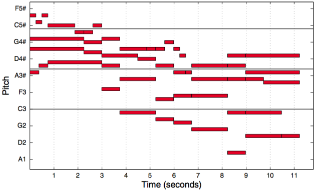
Time scale	Dimension	Content
Short term	Timbre	Quality of the produced sound
	Orchestration	Sources of sound production
	Acoustics	Quality of the recorded sound
Middle term	Rhythm	Patterns of sound onsets
	Melody	Sequences of notes
	Harmony	Sequences of chords
Long term	Structure	Organization of the musical work

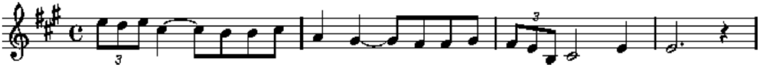
Formats of Musical Documents

Apart from the peculiar characteristics of the forms, different formats are able to capture only a reduced number of dimensions. The formats MIR deals with are:

- Audio formats (MP3, WAV...)
- Symbolic format (score, MIDI files, MPEG7)







The Role of the User (1)

Its all about Information Need. There are three main reasons why users may want to access digital music:

1. listening to a particular performance or musical work;
2. building a collection of music (playlist);
3. verifying or identifying works.

The Role of the User (2)

Potential users of MIR systems are divided in three categories:

1. casual users want to enjoy music, listening and collecting the music they like and discovering new good music;
2. professional users need music suitable for particular usages related to their activities, which may be in media production or for advertisements;
3. music scholars, music theorists, musicologists, and musicians are interested in studying music.

The Casual User

Maybe the most important kind of user, queries as follow:

1. Find me a song that sounds like this
2. Given that I like these songs, find me more songs that I may enjoy
3. I need to organize my personal collection of digital music (stored in my hard drive, portable device, MP3 player, cell phone, etc...)

The Professional User

Users may need to access music collections because of their professional activity.

1. I am looking for a suitable soundtrack for...
2. Retrieve musical works that have a rhythm (melody, harmony, orchestration) similar to this one

Music Theorists, Musicians

Differently from casual and professional users, these users are interested mostly in obtaining information from the musical works, which is the subject of their study, rather than obtaining the musical work itself.

1. access musical scores in order to develop or refine theories on the music language;
2. interested in the work of composers, performers, the role of tradition, the cultural interchanges;
3. retrieve scores to be performed and to listen to how renowned musicians interpreted particular passages or complete works

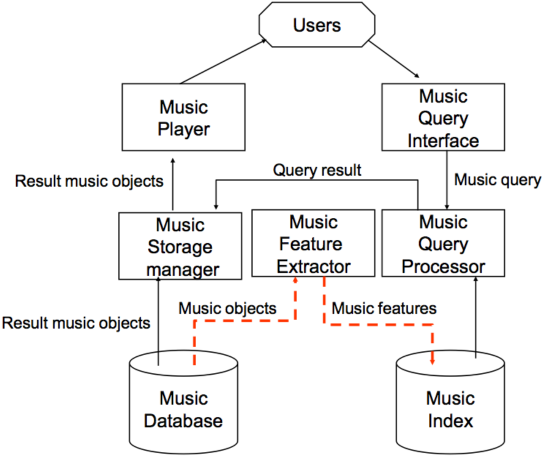
Metadata approach

- Simple I.R. queries
- Often absent (ID3 TAG), or not reliable (MPEG7)
- Not pure Music Retrieval, but something like Text Retrieval on Audio Metadata.

Content-based approach

- Dealing with intrinsic characteristics of the sounds.
- The main form is usually known by the MIR community as **query-by-humming** (QbH), the term has been introduced in 1995 (A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith. Query by humming: musical information retrieval in an audio database).
- The interaction through the audio channel is an error-prone process. The query may be dramatically different from the original song that it is intended to represent, and also different from the users intentions.

A MIR Architecture

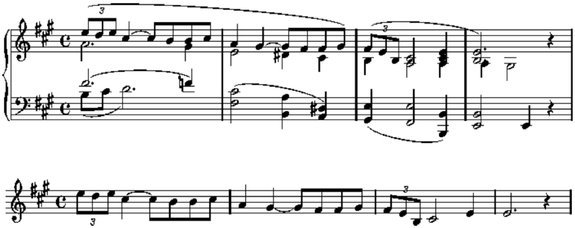


Lexical Semantical Units

- Can we treat musical notes as “words”?
- No! Notes are relative, notes are not context-free.
- What about the musical pitch intervals? Is there an analogy to the “words”?
- No! The abstraction level is too low, pitch intervals are something like characters.
- And what about pitch intervals sequences?
- In a certain way they are “word”!
- Pauses or rests do not play the same role of blanks in text documents.
- Vague concept of “stop words”.
- In general, music do not depend on intonation or on single notes!

Extraction of the main melody

- The main melody of a piece is a good representation of it.
- The computation of melodic information from a monophonic score is straightforward.
- A slightly more difficult task is given by a polyphonic score made of several monophonic voices.
- It is assumed that there is only a relevant melodic line.



Melody Segmentation: N-grams

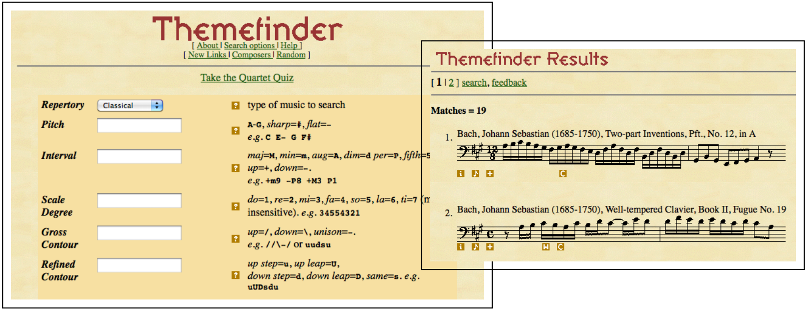
- A simple segmentation approach consists on the extraction from a melody of all the subsequences of exactly N notes, called N-grams.
- The idea underlying this approach is that the effect of musically irrelevant N-grams will be compensated by the presence of all the relevant ones.
- Each sequence, of any length, that is repeated at least K times can be used as a content descriptor of the melodic information.

The Pitch Contour (1)

- A simple but effective approach for searching melodies.
- The **Parsons Code** is a simple notation used to identify a piece of music through *melodic motion*.
  - u = "up" if the note is higher than the previous note
  - d = "down" if the note is lower than the previous note
  - r = "repeat" if the note is the same as the previous note
  - * = first tone as reference
- Examples:
  - "Love Me Tender": *uduududdduu
  - First verse in Madonna's "Like a Virgin": *rrurddrdrrurdud
  - First verse in "We Are the World": *rduduururdrddrududuu
- Boyer & Moore, KMP algorithms or Suffix Tree structures can be used.

The Pitch Contour (2)

http://themefinder.com



Themefinder

[About | Search options | Help]
[New Links | Composers | Random]

Take the Quartet Quiz

Repertory	Classical	type of music to search
Pitch		A-G, *sharp*=#, *flat*=- *e.g.* C E- G F#
Interval		*maj*=M, *min*=m, *aug*=A, *dim*=d *per*=P, *fifth*=5 *up*=+, *down*=-. *e.g.* +m9 -P8 +M3 P1
Scale Degree		*do*=1, *re*=2, *mi*=3, *fa*=4, *so*=5, *la*=6, *ti*=7 (n insensitive). *e.g.* 34554321
Gross Contour		*up*=/, *down*=\, *unison*=-. *e.g.* //\-/ or uudsu
Refined Contour		*up step*=u, *up leap*=U, *down step*=d, *down leap*=D, *same*=s. *e.g.* uUDsdu

Themefinder Results

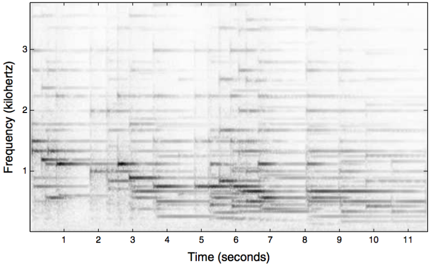
[1 | 2] search, feedback

Matches = 19

1. Bach, Johann Sebastian (1685-1750), Two-part Inventions, Pft., No. 12, in A
2. Bach, Johann Sebastian (1685-1750), Well-tempered Clavier, Book II, Fugue No. 19

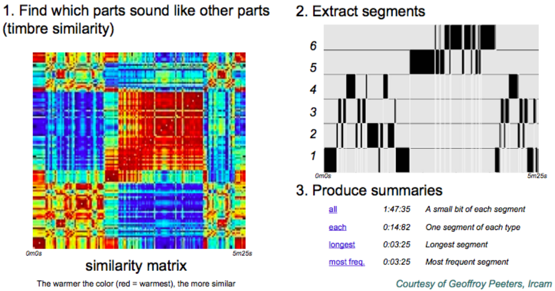
Audio form

- Fourier transform is one of the most frequently used tools for the analysis of audio.
- A common way to represent an audio excerpt is the spectrogram.
- Spectrograms are not suitable content descriptors, because of their high dimensionality.



Self Similarity Matrix

- v_i is the vector of the musical characteristics at time i.
- The self-similarity matrix $M = (m_{ij})$ is defined as follow: $m_{ij} = s(v_i, v_j)$.
- s is some similarity metric choosen.
- M is symmetric along the diagonal.



1. Find which parts sound like other parts (timbre similarity)

similarity matrix

The warmer the color (red = warmest), the more similar

2. Extract segments

3. Produce summaries

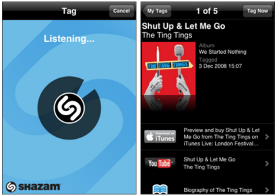
all	1:47:35	A small bit of each segment
each	0:14:82	One segment of each type
longest	0:03:25	Longest segment
most freq.	0:03:25	Most frequent segment

Courtesy of Geoffroy Peeters, Ircam

Shazam: magic?

Desiderata

- providing a service that could connect people to music by recognizing music by using their mobile phones;
- The algorithm had to be able to recognize a short audio sample of music (15 sec) that had been broadcast, mixed with heavy ambient noise, subject to reverb;
- The algorithm also had to perform the recognition quickly over a large database of music.



Shut Up & Let Me Go
The Ting Tings

Album
We Started Nothing
Tagged
3 Dec 2008 15:07

Preview and buy Shut Up & Let Me Go from The Ting Tings on iTunes Live: London Festival...

Shut Up & Let Me Go
The Ting Tings

Biography of The Ting Tings

The process (1)

- Each audio file is fingerprinted, a process in which reproducible hash tokens are extracted;
- Both database and sample audio files are subjected to the same analysis;
- The fingerprints from the unknown sample are matched against a large set of fingerprints derived from the music database;
- The candidate matches are subsequently evaluated for correctness of match;
- Each fingerprint hash is calculated using audio samples near a corresponding point in time, so that distant events do not affect the hash;

The process (2)

- Fingerprint hashes derived from corresponding matching content are reproducible independent of position within an audio file;
- Robustness means that hashes generated from the original clean database track should be reproducible from a degraded copy of the audio;
- Consider spectrogram peaks, due to their robustness in the presence of noise;
- A time-frequency point is a candidate peak if it has a higher energy content than all its neighbors in a region centered around the point;
- Generate a costellation map of peaks;

The process (3)

- Fingerprint hashes are formed from the constellation map, in which pairs of time-frequency points are combinatorially associated;
- Anchor points are chosen, each anchor point having a target zone associated with it;
- A set of hash (anchor frequency, target point frequency, time offset) records is generated;
- Create a database index: the above operation is carried out on each track in a database to generate a corresponding list of hashes and their associated offset times.

The process (4)

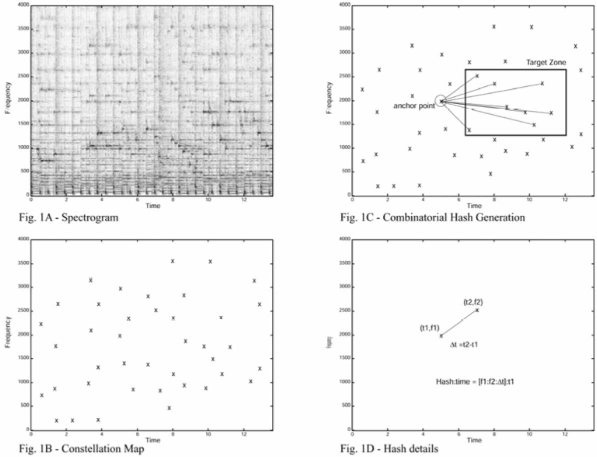


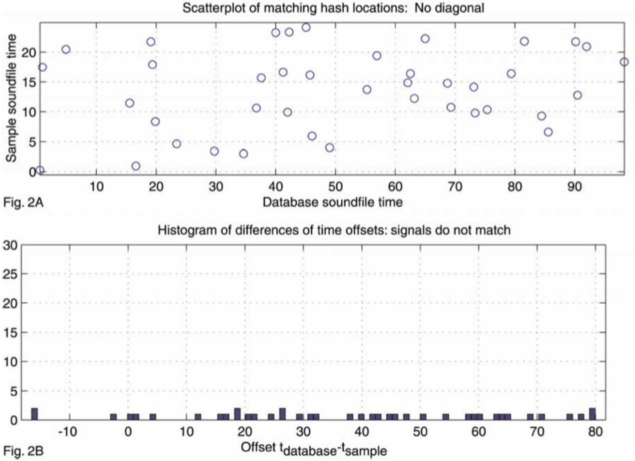
Fig. 1A - Spectrogram

Fig. 1C - Combinatorial Hash Generation

Fig. 1B - Constellation Map

Fig. 1D - Hash details

The process (5)



The process (6)

The corresponding times of matching features between matching files have the relationship

$$tk' = tk + offset$$

where tk' is the time coordinate of the feature in the matching (clean) database soundfile and tk is the time coordinate of the corresponding feature in the sample soundfile to be identified. For each (tk', tk) coordinate in the scatterplot, we calculate

$$\delta tk = tk' - tk$$

Then we calculate a histogram of these δtk values and scan for a peak. This may be done by sorting the set of δtk values.

The process (7)

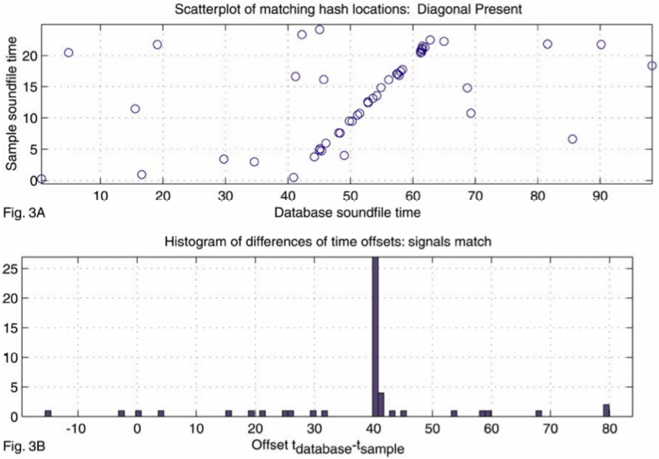


Fig. 3A

Fig. 3B

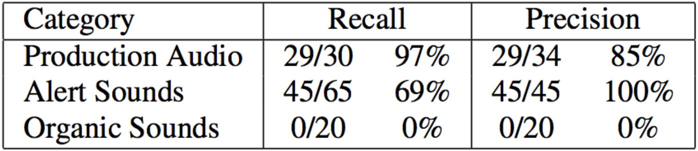
Error probability

- If each constellation point is taken to be an anchor point, and if the target zone has a fan-out of size $F = 10$, then the number of hashes is equal to F times the number of constellation points extracted from the file.
- Note that the combinatorial hashing squares the probability of point survival.
- The surviving probability of at least one hash surviving for a given anchor point would be the joint probability of the anchor point and at least one target point in its target zone surviving.
- $p =$ probability of survival for all points involved
- $p * (1 - (1 - p)^F) =$ probability of at least one hash surviving per anchor point
- $p \approx p * (1 - (1 - p)^F)$ for large values of F, e.g. $F > 10$, and reasonable values of p, e.g. $p > 0.1$.

An example

- 64 frequency bins and delta time quantized to 6 bits
- Hash (f1, f2, delta) is 18 bits
- Fan-out = 10
- Peaks/sec = 3
- 450 hashes/sample (15 sec)
- 450 raw hashes are about 1 kbyte

Considerations

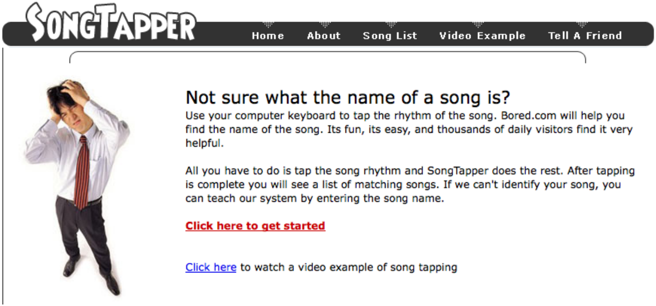


Category	Recall		Precision	
Production Audio	29/30	97%	29/34	85%
Alert Sounds	45/65	69%	45/45	100%
Organic Sounds	0/20	0%	0/20	0%

- Good for music where frequencies are clearly defined.
- Totally unacceptable for organic sounds and rumors (not the supposed

Other systems (1)

- http://www.bored.com/songtapper/



Not sure what the name of a song is?

Use your computer keyboard to tap the rhythm of the song. Bored.com will help you find the name of the song. Its fun, its easy, and thousands of daily visitors find it very helpful.

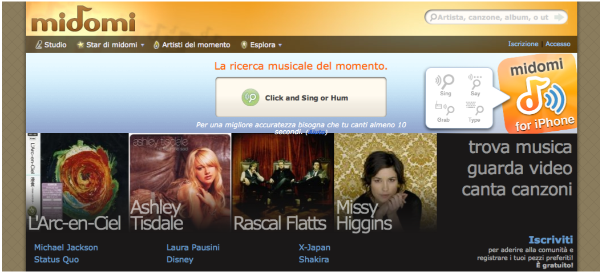
All you have to do is tap the song rhythm and SongTapper does the rest. After tapping is complete you will see a list of matching songs. If we can't identify your song, you can teach our system by entering the song name.

Click here to get started

Click here to watch a video example of song tapping

Other systems (2)

- www.midomi.com
- Sing, humming, state-of-art of *QbH*.
- Possible errors due to wrong-key, poor vocal performance, missing notes.

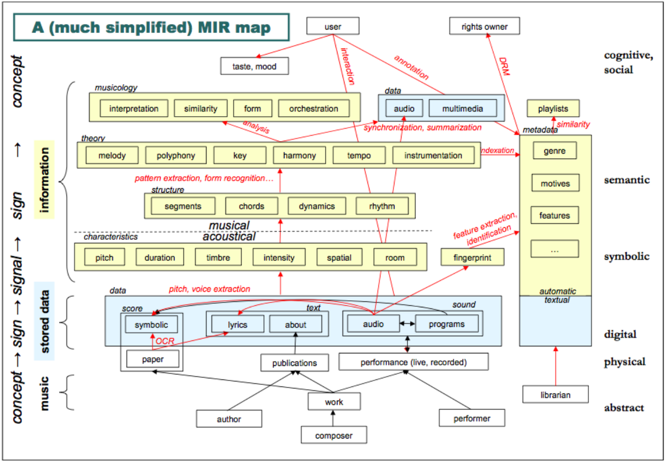


La ricerca musicale del momento.

Per una migliore accuratezza bisogna che tu canti almeno 10 secondi. (Aiuto)

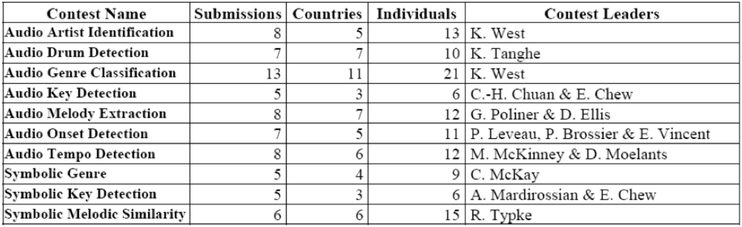
Iscriviti
per aderire alla comunità e
registrare i tuoi pezzi preferiti!
È gratuito!

MIR Map


A (much simplified) MIR map
user
rights owner
taste, mood
interaction
annotation
DRM
cognitive, social
concept
musicology
interpretation
similarity
form
orchestration
data
audio
multimedia
playlists
analysis
synchronization, summarization
similarity
metadata
theory
melody
polyphony
key
harmony
tempo
instrumentation
indexation
genre
semantic
information
pattern extraction, form recognition…
structure
segments
chords
dynamics
rhythm
motives
features
sign
musical
acoustical
characteristics
feature extraction, identification
pitch
duration
timbre
intensity
spatial
room
fingerprint
…
symbolic
signal
data
pitch, voice extraction
automatic
textual
score
text
sound
symbolic
lyrics
about
audio
programs
digital
stored data
OCR
paper
publications
performance (live, recorded)
physical
concept → sign → signal → sign → concept
music
work
librarian
abstract
author
performer
composer

MIREX 2005

- A project for the evaluation of MIR tasks.
- The campaign has been called Music Information Retrieval Evaluation eXchange (MIREX).
- There have been nine different tasks during the MIREX 2005 campaign, six on the audio form and three on thesymbolic form.



Contest Name	Submissions	Countries	Individuals	Contest Leaders
Audio Artist Identification	8	5	13	K. West
Audio Drum Detection	7	7	10	K. Tanghe
Audio Genre Classification	13	11	21	K. West
Audio Key Detection	5	3	6	C.-H. Chuan & E. Chew
Audio Melody Extraction	8	7	12	G. Poliner & D. Ellis
Audio Onset Detection	7	5	11	P. Leveau, P. Brossier & E. Vincent
Audio Tempo Detection	8	6	12	M. McKinney & D. Moelants
Symbolic Genre	5	4	9	C. McKay
Symbolic Key Detection	5	3	6	A. Mardirossian & E. Chew
Symbolic Melodic Similarity	6	6	15	R. Typke

Bibliography

- Nicola Orio. Music Retrieval: A Tutorial and Review.
- Avery Li-Chun Wang. An Industrial-Strength Audio Search Algorithm.
- James P. Ogle and Daniel P.W. Ellis. Fingerprinting to identify repeated sound events in long-duration personal audio recordings.
- Roberto Basili. Introduzione al Music Information Retrieval.
- Michael Fingerhut. Music Information Retrieval, or how to search for (and maybe find) music and do away with incipits.
- Progetto M.I.R.E.X.: http://www.music-ir.org.