# Large Scale Music Information Retrieval

M. Moussallam

Peyresq 2018

# Outline

1. Introduction: Digital Representations of Music
   - Recording, Sampling and Compression
   - Music Information Retrieval Tasks
   - Music Representation and Transforms
2. Classical MIR
   - Speech/Music Classification
   - Chords Detection
   - Transcription
   - Automatic Tagging
3. Deep MIR
   - Convolutional and recurrent Networks
   - Multi-modal learning
   - Music Embedding Spaces
4. Frontiers and Open Challenges

DEEZER

# Outline

1 Introduction: Digital Representations of Music
- Recording, Sampling and Compression
- Music Information Retrieval Tasks
- Music Representation and Transforms

2 Classical MIR

3 Deep MIR

4 Frontiers and Open Challenges

DEEZER

# Recorded Music
## Definitions

- In the physical world, music is conveyed by sound waves $\phi(t, x, y, z)$

# Recorded Music
## Definitions

- In the physical world, music is conveyed by sound waves $\phi(t, x, y, z)$
- **Analog Recorded music**: time-varying signal at a fixed space location $x(t)$

# Recorded Music
## Definitions

- In the physical world, music is conveyed by sound waves $\phi(t, x, y, z)$
- **Analog Recorded music**: time-varying signal at a fixed space location $x(t)$
- **Digital music**: discrete-time varying signal $x[n]$, with *finite* size and **quantized** values

DEEZER

- Discrete-time signal $x[n]$ of length $N$
- Multichannel $x[n, c]$ (e.g. stereo $C = 2$)
- Sampling Frequency $Fe$ (e.g. 44100 Hz)
- Quantization width (e.g 16bits)

A typical musical track lasting 4min in PCM format weights approx. 40MB.

# Digital Recorded Music
## Compression

**Lossless**

Exploits time-frequency redundancies, reversible. (e.g. Flac)

# Digital Recorded Music
## Compression

**Lossless**

Exploits time-frequency redundancies, reversible. (e.g. Flac)

**Lossy**

Additionnal (clever) quantization. destructive. Minimize *perceived* distortion. (e.g. MP3, OGG, AAC, etc)

A typical musical track lasting 4min in MP3@128Kbps weights approx. 4MB.

# Digital Recorded Music

## Compression

**Lossless**

Exploits time-frequency redundancies, reversible. (e.g. Flac)

**Lossy**

Additionnal (clever) quantization. destructive. Minimize *perceived* distortion. (e.g. MP3, OGG, AAC, etc)

A typical musical track lasting 4min in MP3@128Kbps weights approx. 4MB.
Deezer catalog is currently approx. 55 millions of audio tracks.

DEEZER

# MIR Tasks
## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)

DEEZER

# MIR Tasks
## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)

# MIR Tasks
## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)
- Multi-label **Classification** (Genre, Mood, Instrument, etc.)

DEEZER

# MIR Tasks

## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)
- Multi-label **Classification** (Genre, Mood, Instrument, etc.)
- Source **Separation**

# MIR Tasks
## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)
- Multi-label **Classification** (Genre, Mood, Instrument, etc.)
- Source **Separation**
- **Similarity** and **Retrieval** (Fingerprinting, Cover Detection, etc.)

# MIR Tasks

## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)
- Multi-label **Classification** (Genre, Mood, Instrument, etc.)
- Source **Separation**
- **Similarity** and **Retrieval** (Fingerprinting, Cover Detection, etc.)
- **Structuration** (Segmentation)

DEEZER

# MIR Tasks
## In the academic community

- **Detection** (Onsets, Beats, Tempo, Key, Chords, etc.)
- **Transcription** (Notes, Lyrics, Score, Rythms, etc.)
- Multi-label **Classification** (Genre, Mood, Instrument, etc.)
- Source **Separation**
- **Similarity** and **Retrieval** (Fingerprinting, Cover Detection, etc.)
- **Structuration** (Segmentation)

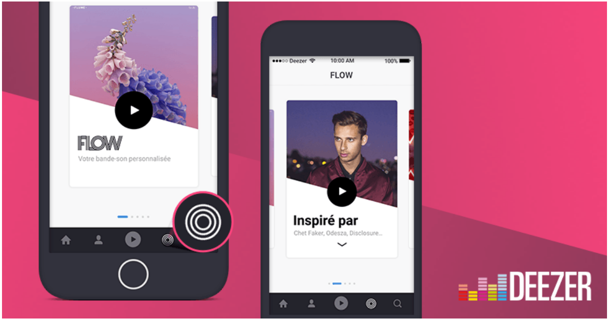
**ISMIR** Conference - **MIREX** Challenge

S. Downie *The Music Information Retrieval Evaluation Exchange*. 2008

DEEZER

# MIR Tasks
at Deezer

- Catalog Cleaning and Tagging
- Recommendation
- Exploration
- Making the world a better place



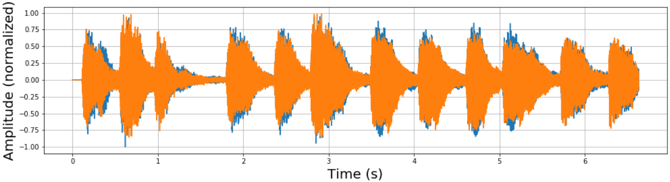

DEEZER

# High Dimensional Vector / Matrix

## Waveform

Signal $x$ is a $CxN$ matrix with real entries

$$x[c,n] \in [-1,1]$$

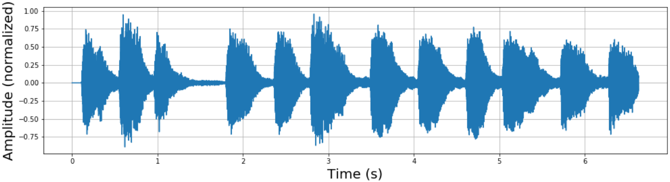



One minute long, stereo PCM at 44100Hz is a 2*x*2.646.000 matrix

DEEZER

# Symbolic Representation

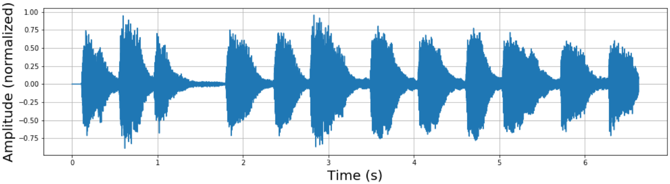
## Signal as a rendering of a Musical score





DEEZER

# Symbolic Representation
## Signal as a rendering of a Musical score





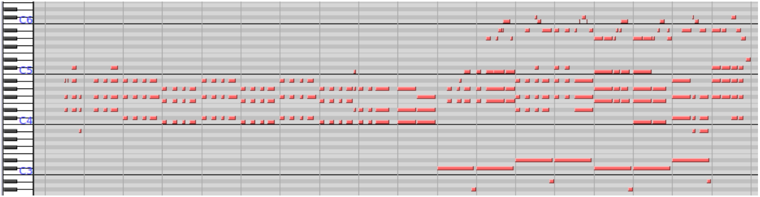
- Signal to Score: **Music Transcription***, **Score Alignment**
- Score to Signal : **Music Synthesis**

DEEZER

# Synthesis-oriented Representation
## Midi format

Synthesis control oriented

- Time-Frequency Events with Synthesis Parameters
- Finite set of frequencies mapped to Western music scale
- Signal to Midi: **Music Transcription** (much less ambiguous)




A. Klapuri, M. Davy. *Signal processing methods for music transcription*. 2007

## Short Time Fourier Transform

- Sliding Window $w$ of $Nfft$ samples, Hop Size of $\Delta_n$
- Discrete-Time Fourier Transform of $x[n]$
- Stack results in a 2D Matrix

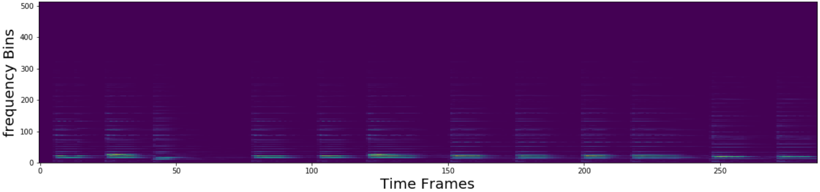
$$X[k,p] = \sum_{n=0}^{Nfft-1} w[n].x[n - p\Delta_n].\exp\left(-2i\pi n \frac{k}{Nfft}\right)$$

- Complex Matrix
- Invertible under mild conditions

# Time-Frequency Representation
## Short Time Fourier Transform

What it looks like:

- Magnitude Spectrogram: $|X[k,p]|$

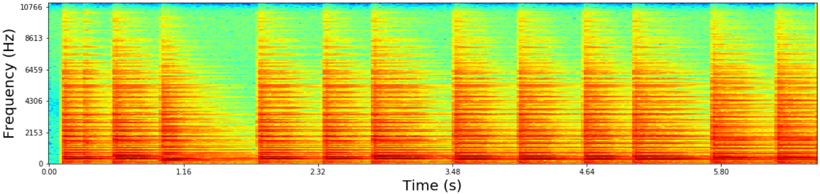



DEEZER

# Time-Frequency Representation
## Short Time Fourier Transform

What it looks like:

- Magnitude Spectrogram: $|X[k,p]|$
- Log-Spectrogram $\rightarrow \log|X[k,p]|$



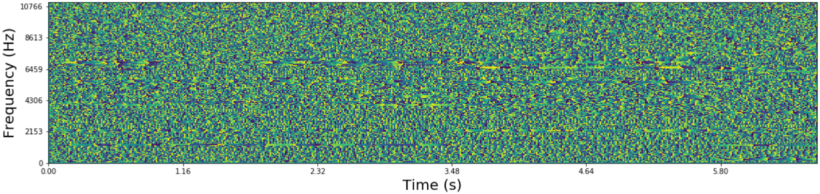

DEEZER

# Time-Frequency Representation

## Short Time Fourier Transform

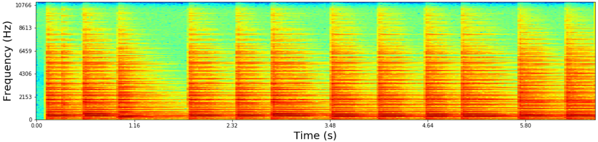
What about the Phase ?

- Phase Spectrogram looks random but it's not
- Absolutely crucial for intelligibility and reconstruction
- Very rarely used/considered in MIR systems using Machine Learning




DEEZER

# Time-Frequency Representation
## Interpretations of the STFT




- Discrete Filter Bank with $Nfft/2$ bandpass filters
- Gabor Transform (if $w$ gaussian)
- Tight-Frame decomposition into a gabor dictionary

S. Mallat *A wavelet tour of signal processing*. 2008

DEEZER

# Time-Frequency Representation

## Variants of the STFT: Log-Frequency Scale

Mammal ear perceives sound frequencies in a *logarithmic* scale

An **octave** is not an absolute frequency interval, it's a ratio of 2 between the frequencies. (e.g. if A4 is 440Hz then A5 is 880Hz and A3 is 220Hz)

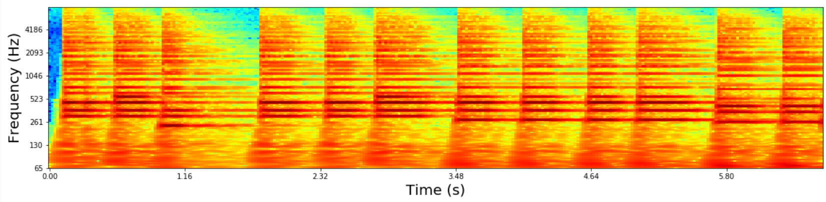
- STFT has linear frequency scaling
- Time/Freq resolution is same everywhere

DEEZER

# Time-Frequency Representation

## Variants of the STFT: Log-Frequency Scale

### Constant-Q Transform




$$X_q[k,p] = \sum_{n=0}^{Nfft_k - 1} w_k[n].x[n - p\Delta_n]. \exp\left(-2i\pi n \frac{k}{Nfft_k}\right)$$

log scale for $k$, constant number of bins per octave.

# Outline

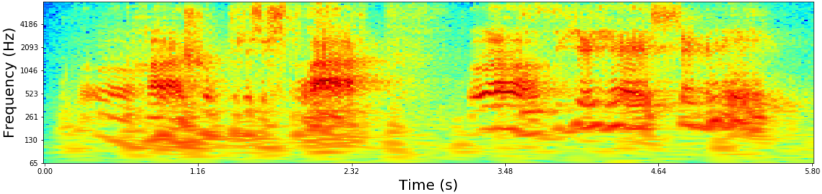
1 Introduction: Digital Representations of Music

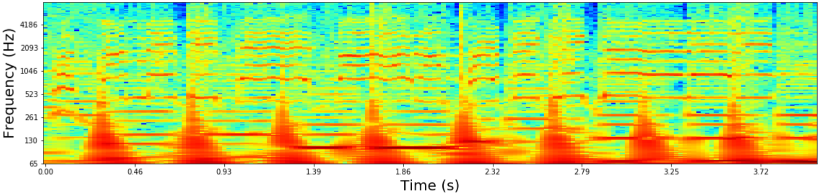
2 Classical MIR
- Speech/Music Classification
- Chords Detection
- Transcription
- Automatic Tagging

3 Deep MIR

4 Frontiers and Open Challenges

Spectral patterns are very discriminative



# *Timbral* Features

- Spectral Shape Statistics: centroid, spread, skewness and kurtosis
- Spectral Flux: Short-Term Dynamics of the Spectrum

DEEZER

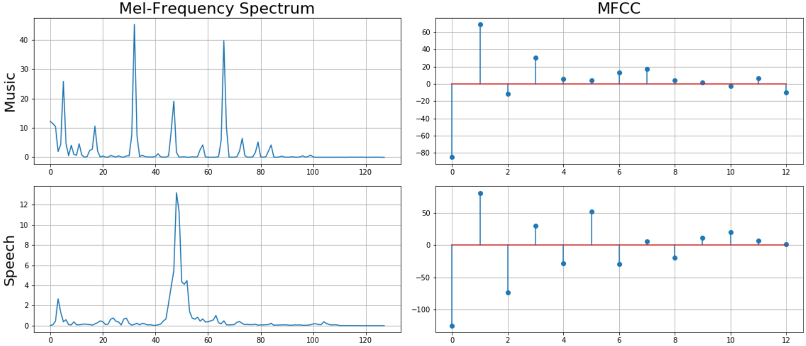
# *Timbral* Features

- Spectral Shape Statistics: centroid, spread, skewness and kurtosis
- Spectral Flux: Short-Term Dynamics of the Spectrum

## Mel Frequency **Cepstral** Coefficients (MFCC)

- Log-Magnitude in Mel-Scale Frequency
- Discrete Cosine Transform

# *Timbral* Features




## Uneasy Interpretation

- Time-Derivatives are often used
- In practice quite effective with simple linear classifiers
- Widely used on speech processing

DEEZER

# Chord Detection

*Harmonic* Features

- Basic: Local FFT Peaks
- Complex: Tonnetz transform

# Chord Detection

*Harmonic* Features

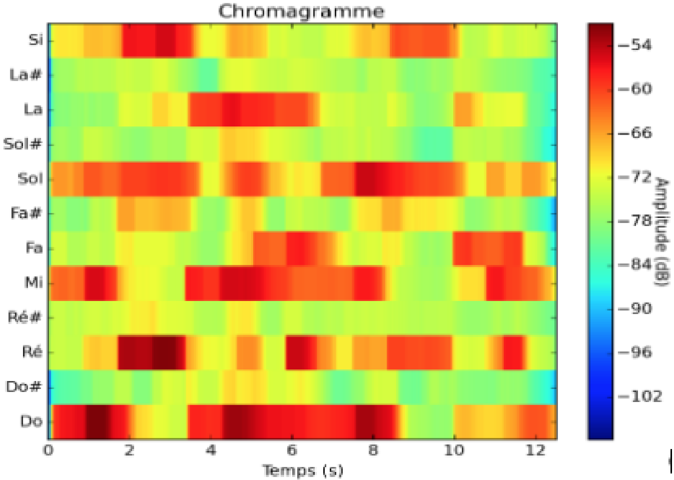
- Basic: Local FFT Peaks
- Complex: Tonnetz transform

## Pitch-Class Profile (Chromagram)

- Discrete 12 tonalities scale
- Easy to obtain from a CQT: just sum up bins belonging to same "note" on different octaves

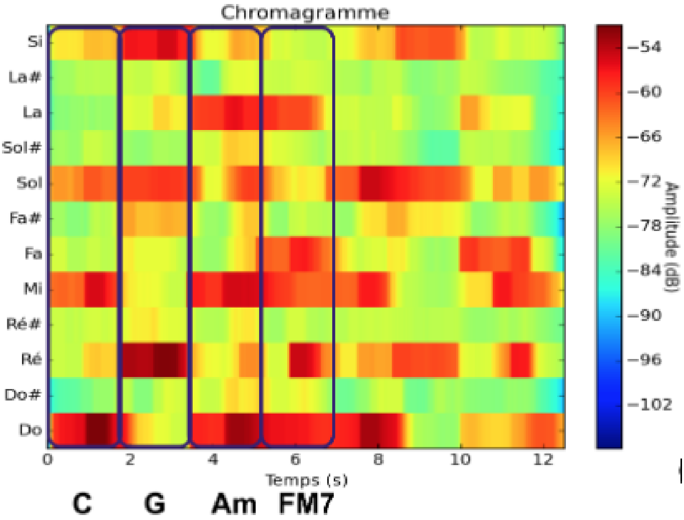
DEEZER

# Chord Detection
## Intuition




DEEZER

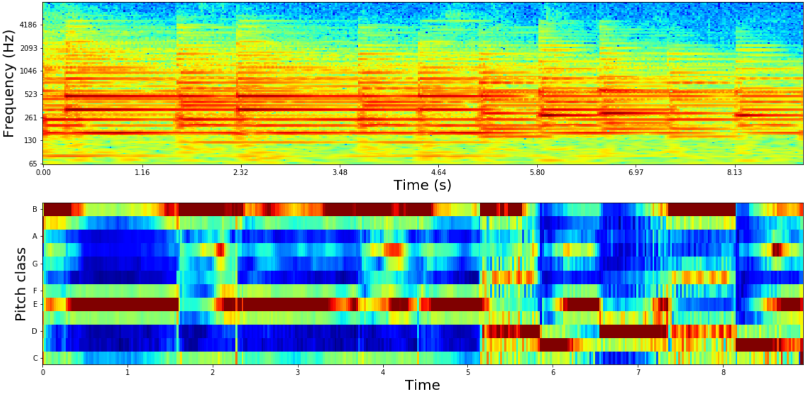
# Chord Detection
## Intuition




DEEZER

# Chord Detection

## Chromagram + HMM + Viterbi



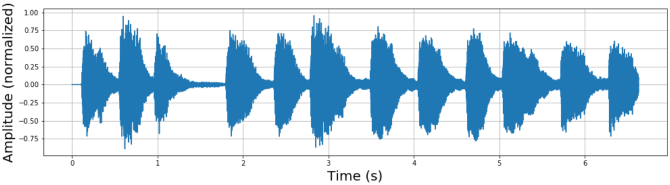

- K. Lee. *Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile*. ICMC 2006
- J.P Bello, J. Pickens *A Robust Mid-level Representation for Harmonic Content in Music Signals*. ISMIR 2005

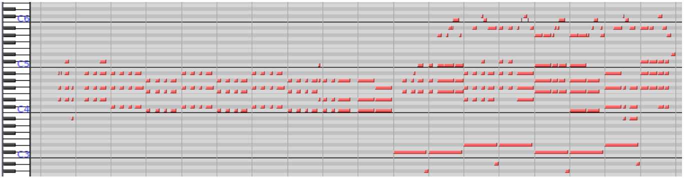
# Chord Detection
## Limits

- Becomes very hard in multi-instrumental setups
- Very sensitive to distorsions and noises
- Chord ambiguities
- Tailored for western dodecaphonism

DEEZER

# Transcription



## Non Negative Matrix Factorization

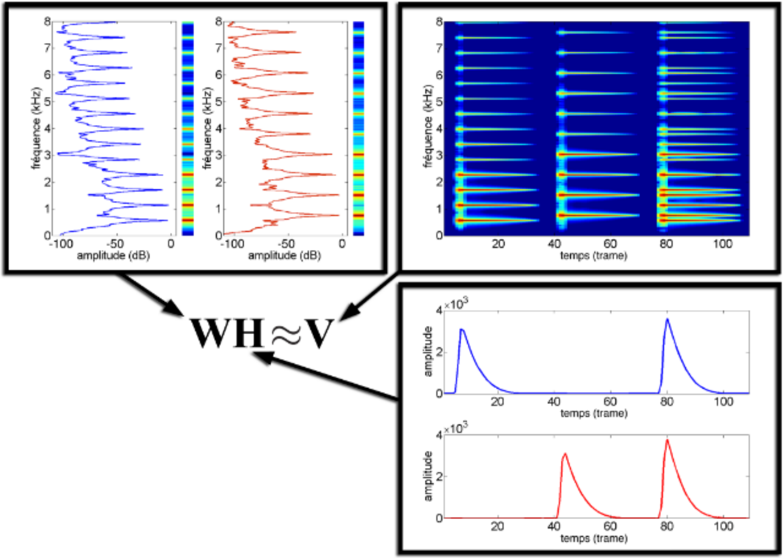
Magnitude Spectrogram as a product of Low-Rank Matrices

$$X \approx V = W \cdot H$$

$W$ is called the **Template** Matrix (or Dictionary) and $H$ is called the **Activation** Matrix

- Supervised/Unsupervised fix or learn one/both
- Multiplicative update rules
- A large variety of sparsity, structural and model constraints

# Spectrogram Factorization




DEEZER

## Additional Constraints

- Smoothness prior on $H$
- Harmonic prior on $W$
- Sparsity on $H$, $W$, or both
- Source and Noise models

## Remaining issues

- Computationaly Intensive
- Strong hypothesis: Additivity of Magnitude Spectrogram
- Component number ?

# Music transcription with NMF

Components = notes, Activations = Onsets


- R. Hennequin *Décomposition de spectrogrammes musicaux informée par des modèles de synthèse spectrale.* PhD Dissertation
- A. Dessein, A. Cont, G. Lemaitre *Real-time Polyphonic Music Transcription with Non-negative Matrix Factorization and Beta-divergence* ISMIR 2011
- B. Fuentes, R. Badeau, and G. Richard, *Harmonic Adaptive Latent Component Analysis of Audio and Application to Music Transcription* IEEE TSALP 2013

# Scaling

Can you scale a feature engineered MIR system ?

## Classical Datasets

- Genre Classification: GTZAN dataset 1000 tracks
- Chord Recognition: The Beatles dataset 225 tracks
- Chord Recognition: Billboard dataset 740 tracks
- Piano Transcription: MAPS (Synthetic) a few thousands pieces

# Scaling

Not just a size issue

## Generalization

- Most datasets provide very homogeneous sounds (e.g. Mazurka Dataset: 2700 pieces of 49 Chopin mazurkas)
- Labeled content is expensive to get
- Most music is copyrighted
- Many MIR concepts are ambiguous

# Scaling

Not just a size issue

## Generalization

- Most datasets provide very homogeneous sounds (e.g. Mazurka Dataset: 2700 pieces of 49 Chopin mazurkas)
- Labeled content is expensive to get
- Most music is copyrighted
- Many MIR concepts are ambiguous

## Realistic Datasets

- Million Song Dataset (tags, usage and features but no audio)
- Free Music Archive - 100K songs representativity issues
- AudioSet : labels from youtube video titles ...

DEEZER

# Automatic Tagging At Scale in a classical MIR Setup

- Stack as many features as you can (MFCC, Chromas) and their time derivatives
- Train a classifier on as many labeled data as you can gather
- Very poor results on truly large scale

**Real Scale Data**

Deezer receives between 2 and 20k audio tracks everyday. Both volume and variety of content is not matched by publicly available datasets.

DEEZER

# Outline

1 Introduction: Digital Representations of Music

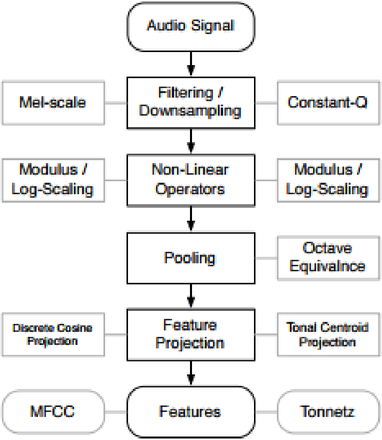
2 Classical MIR

3 Deep MIR
- Convolutional and recurrent Networks
- Multi-modal learning
- Music Embedding Spaces

4 Frontiers and Open Challenges

# The Shift

## Feature Engineering as a deep architecture



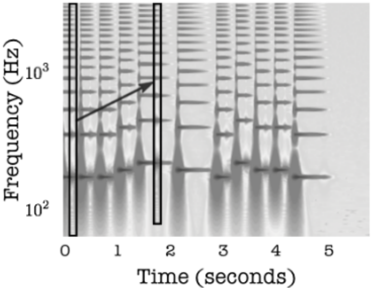

E. Humphrey, J.P. Bello and Y. Lecun. *Moving beyond feature design: deep architectures and automatic feature learning in music informatics*. ISMIR 2012

# Convolutional Layers
## Spectrograms as Oriented Images

### Interesting Invariances

- Time-Translation: Same pattern at different time localization
- For CQT: (Small) Frequency Translations: same **spectral** pattern

Intuition: use rectangular filters in first layers

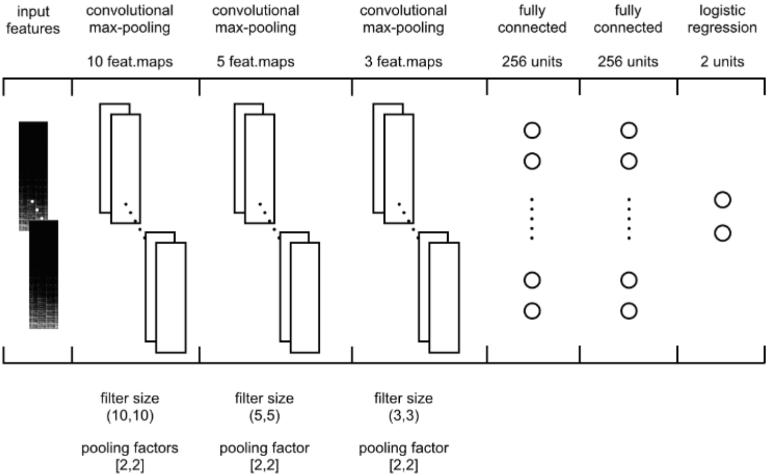



DEEZER

# Speech/Music Classification with CNNs

Work by J.Royo-Letelier in 2015

Convolutional Layers followed by Dense ones.




DEEZER

# Speech/Music Classification with CNNs

Work by J.Royo-Letelier in 2015

## Standard Datasets for the task

- GTZAN Speech/Music: **120 tracks**
- MIREX 2015 on approx 50h of audio
- MUSAN (end of 2015), 108h of audio

## Deezer Dataset

- 41000 annotated audio tracks, 58.3 hours of audio
- Semi-Annotated (third party labels)
- Huge variety of sources (language, music genres, recording conditions and audio quality)

# Speech/Music Classification with CNNs

## Feeling the Gap

Compare with feature-engineering approaches

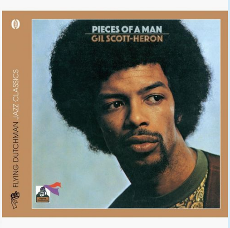
| | Input Shape | Training [s] | Recall | Precision | F-measure |
|---|---|---|---|---|---|
| SVM | (1, 3888) | 40 | 82.17 | 89.83 | 85.83 |
| RF | (1, 3888) | 16 | 76.27 | 89.11 | 82.19 |
| CNN | (3, 108, 12) | **504** | 93.10 | 90.00 | **91.53** |

Our system ranked among the first 3 on MIREX 2015

# How come it works so well?

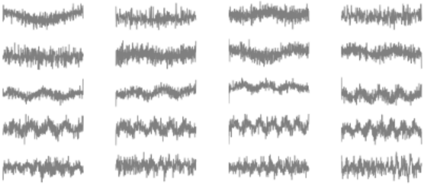
## Looking at the errors

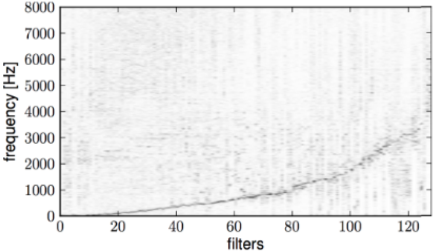
What is the kind of content that classical approaches do not manage to classify adequatly ?


# Towards End-to-End learning?

## Not necessarily






First layers seem to learn ... logarithmically spaced frequency filters.

- S. Dieleman and B. Schrauwen.*End-to- end learning for music audio* ICASSP 2014
- J. Pons, O. Nieto, M. Prockup, E. Schmidt, A. Ehmann, X. Serra *End-to-end learning for music audio tagging at scale* ICASSP 2017

DEEZER

# MIR with Deep Architecture at Deezer

## Increasing task complexities

- Instrumental/Vocal detection
- Language Identification
- Instrument detection
- Low-Quality Encoding detection
- Mood estimation

## Decisive advances

- Recurrent Networks to learn temporal dependencies
- Attention Mechanism
- Data augmentation

# MIR with Multi-modal data
## How much can be inferred from audio alone?

- Instrumentation: almost certainly

# MIR with Multi-modal data

## How much can be inferred from audio alone?

- Instrumentation: almost certainly
- Genre: probably

# MIR with Multi-modal data

## How much can be inferred from audio alone?

- Instrumentation: almost certainly
- Genre: probably
- Structure ?
- Semantics ?
- Mood ?

DEEZER

# Music info in a lot of other sources

- Images

# Music info in a lot of other sources

- Images
- Lyrics

"Pas d'direction, j'connais qu'la flèche de mon oint-j
Motivation, sang de fils de pute sur mon linge (fils de pute)
J'connais la chanson "sales négros, rentrez chez vous"
Billets de cinq cent "sales négros, bienvenue chez nous"
(bienvenue chez nous)
J'suis plus dans l'tier-quar, Castelo de São Jorge
J'tire sur un pétard, c'est la violence, maux de gorge
J'ai quitté l'terrain "dealer, parfois tu nous manques"
A dit la putain que j'baise pour un G d'Hollande (pris dans la schnek) "

DEEZER

# Mood Estimation

## Stating the problem

**Mood ?**

Emotion felt by a listener when exposed to a music

- Discrete set of moods: Multilabel classification / Clustering

| | Adjectives |
|---|---|
| Cluster 1 | passionate, rousing, confident, boisterous, rowdy |
| Cluster 2 | rollicking, cheerful, fun, sweet, amiable/good natured |
| Cluster 3 | literate, poignant, wistful, bittersweet, autumnal, brooding |
| Cluster 4 | humorous, silly, campy, quirky, whimsical, witty, wry |
| Cluster 5 | aggressive, fiery, tense/anxious, intense, volatile, visceral |

Table 1. MIREX mood tags

DEEZER

# Mood Estimation
## Stating the problem

**Mood ?**

Emotion felt by a listener when exposed to a music

- Discrete set of moods: Multilabel classification / Clustering
- Continous Space: Regression

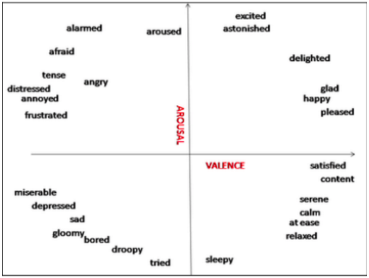



**Figure 1**. Russell's model with two dimensions

# Mood Estimation
## Problem

Can we infer the mood with audio alone ?

### sub-tasks

- Happy/Sad classification
- Arousal/Valence regression

### dataset

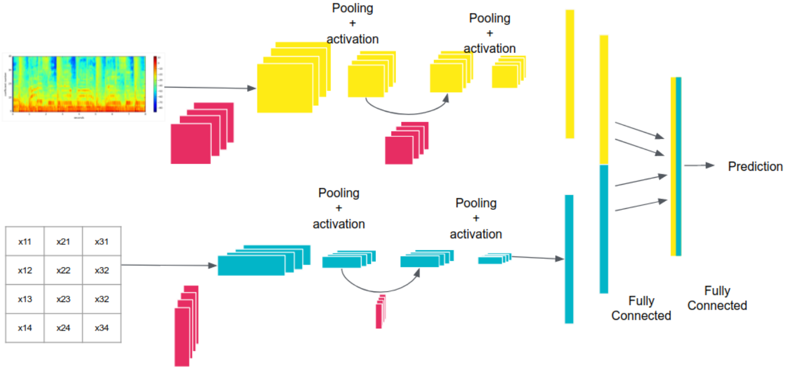
Million Song Dataset for tags + Deezer Audio + Lyrics

| Set | Happy/sad classification | Valence/arousal regression |
|---|---|---|
| Train | 3838 | 12174 |
| Test | 1220 | 3831 |
| Validation | 1302 | 4240 |

DEEZER

# Mood Estimation

## Multimodal Networks




R. Delbouys et al. *Multimodal music mood detection based on audio and lyrics*. ISMIR 2018

# Mood Estimation
## Results

| | happy/sad accuracy in % | valence MSE | arousal MSE |
|---|---|---|---|
| Audio | 80.26 | 0.9141 | 0.6721 |
| Lyrics | 72.52 | 0.9700 | 0.8870 |
| Late fusion | 82.12 | 0.9009 | 0.6721 |
| Mid-level fusion | 82.15 | 0.8701 | 0.6675 |

### Interesting findings

- Multimodal is better for Happy/Sad and Valence prediction
- Lyrics does not add much for arousal prediction

DEEZER

# Multi-modal learning

## Other Modalities

- Images (album covers, artist profile pictures)
- Usage (Collab filtering)
- Other texts (album reviews)
- Context
- ...

## Limits of the One-task one network approach

- Tedious to learn
- Long time lost in data preparation
- ad-hoc development

## Limits of the One-task one network approach

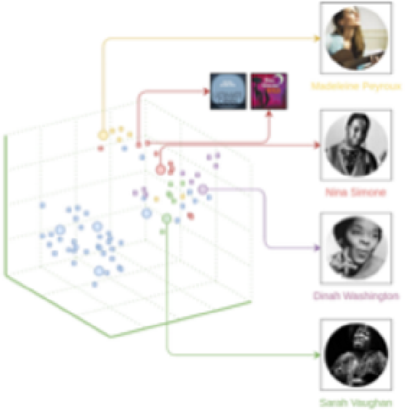
- Tedious to learn
- Long time lost in data preparation
- ad-hoc development

## Embeddings

- Assign to each item $x$ a continuous $f(x)$ in $\mathbb{R}^d$
- Core idea: $\|f(x_1) - f(x_2)\|$ encode the similarity between $x_1$ and $x_2$
- Use it to bootstrap classification/regression tasks
- Enables ranking and clustering tasks

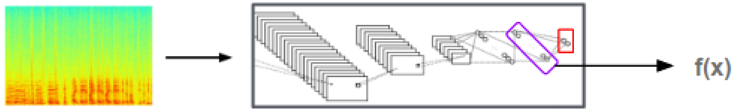
# Representation Learning

## Continuous space description of musical items



# Representation Learning
two approaches

## Intermediate layers of a trained neural net




## Metric Learning

Explicitely learn the mapping $f$, usually by optimizing a triplet loss function

$$\mathcal{L}(x^*, x^+, x^-) = \Big| \|f(x^*) - f(x^+)\|_2^2 - \|f(x^*) - f(x^-)\|_2^2 + \alpha \Big|_+$$

# Music Representations: two examples

- Genre Representation (R. Hennequin *et al*, ISMIR 2018)
- Artist Disambiguations (J. Royo-Letelier *et al*, ISMIR 2018)

# Genre Ambiguities


$\not\leftrightarrow$

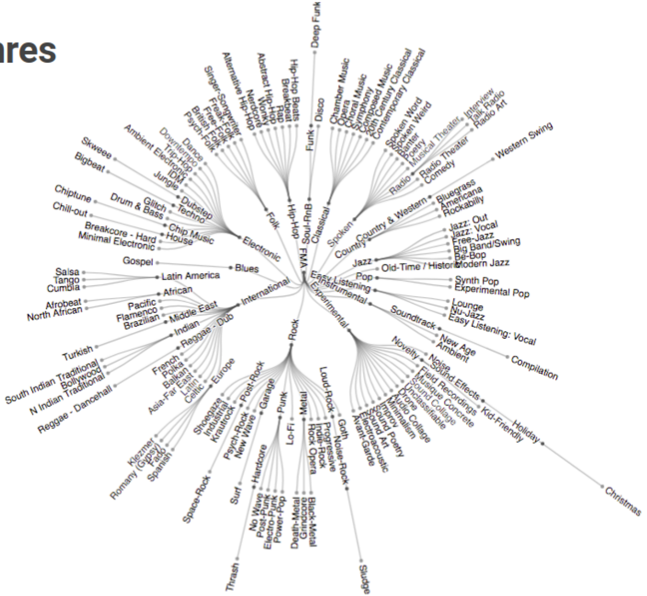


$\not\leftrightarrow$

# Genre Representations

Many Different Genre taxonomy/ontology exists

## FMA genres



# Genre Representations

Many Different Genre taxonomy/ontology exists

## Google Audio Set genres

# Genre representation
## Issues

### unsatisfactory genre representations

- Definition of tags: explicit ? variations of meaning/distribution between dataset.
- Duplication issues: Bossa Nova / Bossanova.
- Polysemy: hardcore may refer to hardcore punk or hardcore electronic.

### Tasks

- Taxonomy Inference
- Tag System Translation

# Building a Genre representation

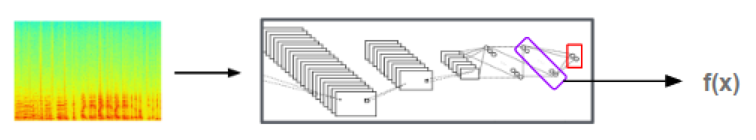
- Top-down approach: Map tags to an expert ontology (e.g. dbpedia) using string matching.
- Use the tags distribution to infer relations between tags (based on the distributional hypothesis)

## Limitations

- Meaning of tags may not be explicit in a tag set.
- The ontology has to identify every possible name for a concept.
- Polysemy is difficult to deal with.
- Needs overlap between dataset for inferring relations between tags of different tag systems.

# Building a Genre Representation from audio

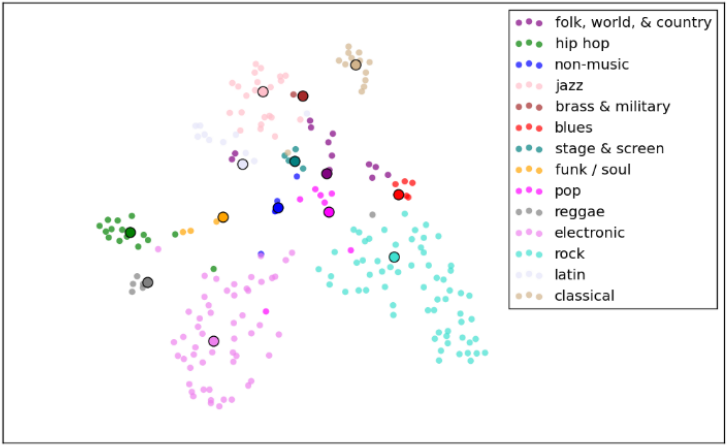
## Idea




- Use a classification system (based on a CNN).
- Build a vector representation such that the distance in the representation space is linked to the confusion of the classifier.

# Genre representation
## Taxonomy inference

Task: Given a sub-Genre (e.g. *indie rock*), retrieve its associated main genre (e.g. *rock*)




t-SNE of audio-based genre representation

DEEZER

# Genre representation
## Tag System Translation

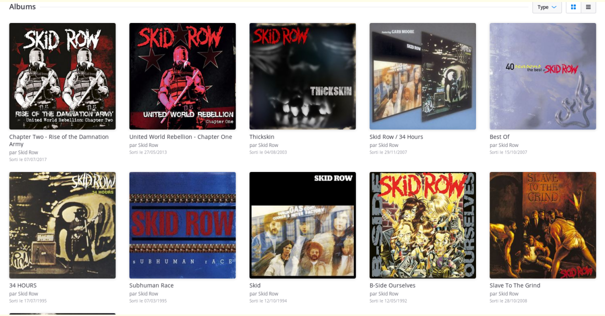
| Audio-based translation $\mathbf{f}_c$ | | Cooccurrence-based translation $\mathbf{f}_{\text{dist}}$ | |
|---|---|---|---|
| Mumu tag | Discogs tag | Mumu tag | Discogs tag |
| bebop | bop | irish folk | celtic |
| movie scores | score | contemporary big band | big band |
| indie & lo-fi | lo-fi | latin music | genre:latin |
| electric blues | modern elec. blues | rap & hip-hop | genre:hip hop |
| electronica | leftfield | vocal blues | ragtime |
| punk-pop | pop punk | dance & electronic | genre:electronic |
| modern postbebop | genre:jazz | today's country | country |
| special interest | avantgarde | electric blues | genre:blues |
| singer-songwriters | folk rock | children's music | genre:children's |
| r&b | rnb/swing | comedy & spoken word | comedy |

Embedding space seem to capture a notion of "genre" similarities that is detached from the labels.

DEEZER

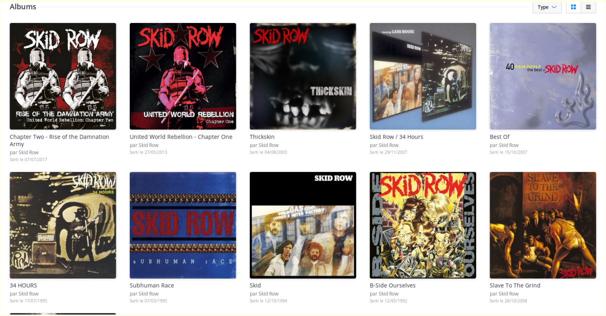
# Artist Disambiguation with Metric Learning

## No universal Identifier for Artists



# Artist Disambiguation with Metric Learning

## No universal Identifier for Artists




## Disambiguate using the audio

Sample excerpts from discographies and try to cluster them in an embedding space.

DEEZER

# Artist Disambiguation from audio

## Push-Pull Loss function

$$\mathcal{L}(x^*, x^+, x^-) = \left| \|f(x^*) - f(x^+)\|_2^2 - \|f(x^*) - f(x^-)\|_2^2 + \alpha \right|_+$$

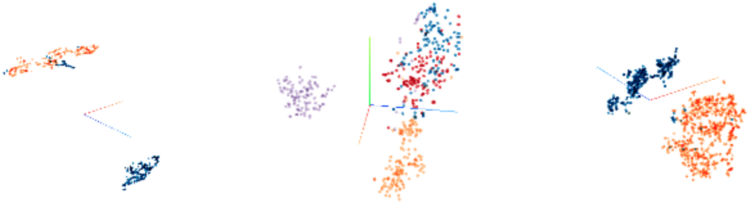
Avoid collapsing:

$$\|f(x)\| = 1$$

## Sampling triplets


DEEZER

# Artist Disambiguation
## Clusters



(a) "Ace" (b) "Scarecrow" (c) "Do or Die"

**In a nutshell**

- Generalizes to artists not seen during the training
- Best used in conjunction with metadata

# Outline


1 Introduction: Digital Representations of Music

2 Classical MIR

3 Deep MIR

4 Frontiers and Open Challenges


DEEZER

# MIR at the age of Deep Learning

## Tasks that have greatly improved

- Instrument and Chords Detection
- Genre classification
- Mood estimation
- Source separation

## Tasks not so much impacted

- Lyrics Transcription
- Structure Analysis
- Cover and version Identification

# Next Frontiers in MIR

- Understanding impact of cultural bias on musical concepts
- Understanding the impact of listening context on perception of musical information
- Captioning of music

Music notions are culturally and context dependent. Rock is not the same for a 14 year old brazilian girl and a 50 y.o. male from tenessee.

# The Andre Rieu's effect

Music notions are heavily culturally and context dependent

Classical music ?

# Open challenges

Study links between user behavior, external context and musical information concepts

## Research directions

- Audio Signal for cold start recommendation
- Embedding space alignment and projections

DEEZER

# Open challenges

Study links between user behavior, external context and musical information concepts

## Research directions

- Audio Signal for cold start recommendation
- Embedding space alignment and projections
- ANR DICTAPHONE: measuring gap between discourse and practice of music consumption
- ANR SATIE: measuring "musical satisfaction"
- EU Project MIP-Frontiers: PhD on "Behavioural music data analytics"

# The team

**Romain Hennequin**
Lead Research Scientist

**Jimena Royo-Letelier**
Research Scientist

**Viet-Anh Tran**
Research Scientist

**Anis Khlif**
Software Engineer

**Mickaël Arcos**
Software Engineer

**Soon**
Research Scientist NLP

**Andrea Vaglio**
PhD

# Our latest papers


- R. Hennequin, J. Royo-Letelier, M. Moussallam *Codec Independent Lossy Audio Compression Detection*. ICASSP 2017
- J. Royo-Letelier, R. Hennequin, M. Moussallam *Metric learning for music artist disambiguation from audio*. ISMIR 2018
- R. Delbouys, R. Hennequin, J. Royo-Letelier, F.Piccoli, M. Moussallam *Towards end-to-end multimodal music mood detection based on audio and lyrics*. ISMIR 2018
- R. Hennequin, J. Royo-Letelier, M. Moussallam *Audio Based Disambiguation Of Music Genre Tags*. ISMIR 2018


DEEZER

# Contact and question

## Jobs

deezerjobs.com

## Contact

research@deezer.com, mmoussallam@deezer.com
Twitter: @MMoussallam

## Stay tuned

deezer.io

DEEZER