Machine Listening for Music and Sound Analysis

Lecture 3 – Music Information Retrieval

Dr.-Ing. Jakob Abeßer

Fraunhofer IDMT

Jakob.abesser@idmt.fraunhofer.de

https://machinelistening.github.io

Fraunhofer IDMT

Overview

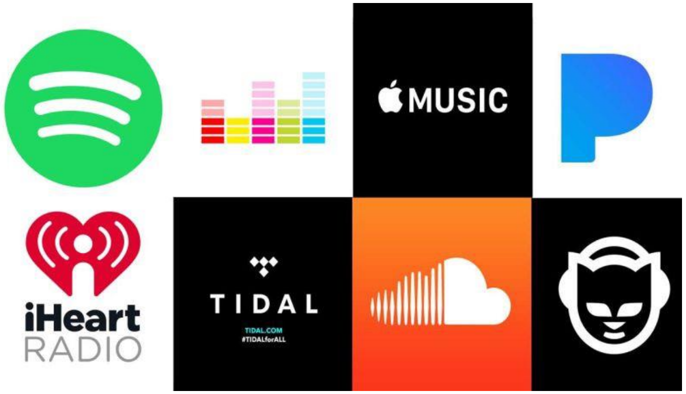
- Music Information Retrieval
- Case Studies
    - Music Tagging
    - Pitch Detection
    - Tempo Estimation
    - Instrument Recognition

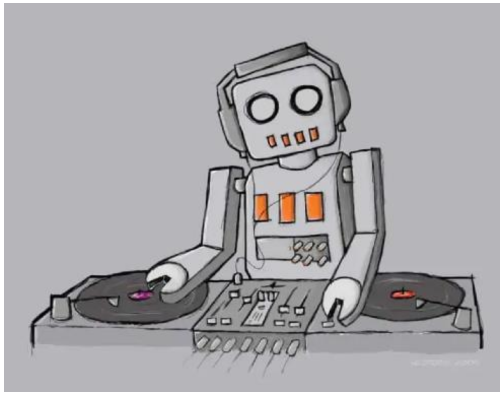
Introduction
Examples

- Examples:
    - Musical Instrument
    - Musical Genre / Tempo

Introduction
Motivation

- Large music collections
- Mobile device apps / instruments
- Music industry shifts almost completely to online products & services
- Growing market of music streaming services





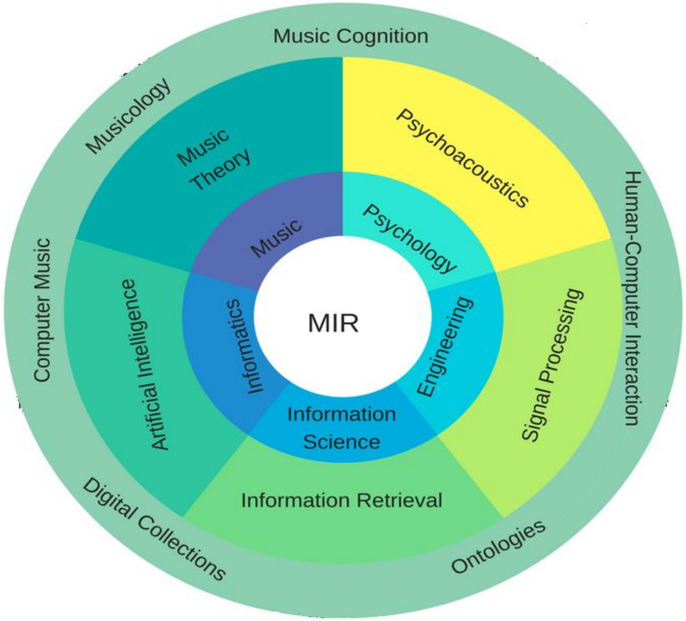
https://www.synchtank.com/wp-content/uploads/2018/06/1476277072027.jpg
https://miro.medium.com/max/800/1*cC1KOdyzzt1nazak42cBdg.jpeg

Introduction
Typical Tasks

- What's that song again? Who's singing that?
    - Audio identification
- I want to learn that song on my instrument!
    - Automatic music transcription
- What songs are similar? How to generate a playlist?
    - Audio similarity search
- How to organize my music? Which genre / style?
    - Audio classification

Introduction
MIR – Today

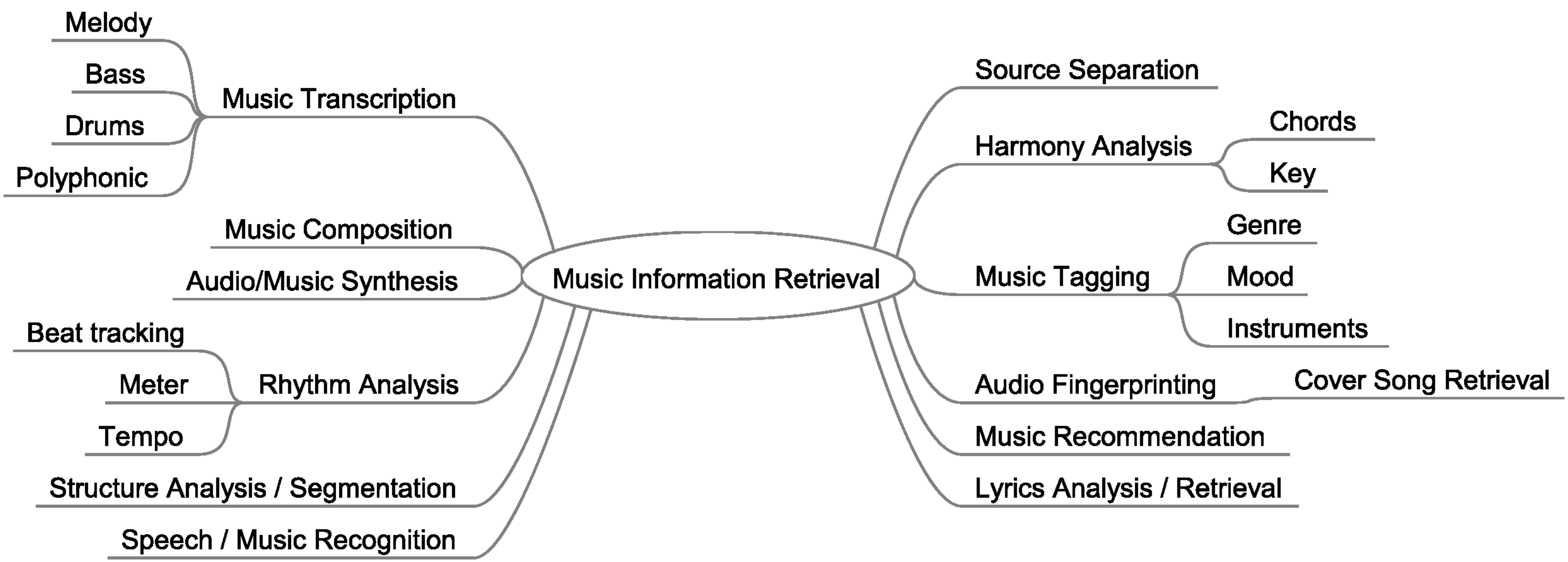
- Interdisciplinary research community since 2000
- ISMIR conference (International Society for Music Information Retrieval)
- Other conferences: ICASSP, DAFx, AES, ICMC, SMC, …
- MIREX competition (Music Information Retrieval Evaluation eXchange)



Introduction
Research Landscape

- MIR @ Fraunhofer IDMT
    - Semantic music technologies (SMT) group
        - Staff + PhD / master / bachelor students + interns
- National / international research groups
    - International Audio Laboratories Erlangen, Germany
    - Centre for Digital Music, Queen Mary University, London, UK
    - Universitat Pompeu Fabra, Barcelona, Spain
    - Institute for music/acoustic research and coordination (IRCAM), Paris, France
    - USA, China, Taiwan etc.

Introduction
MIR Tasks



Case Studies

- Music tagging → general classification tasks
- Pitch detection → melody
- Tempo estimation → rhythm
- Instrument recognition → timbre

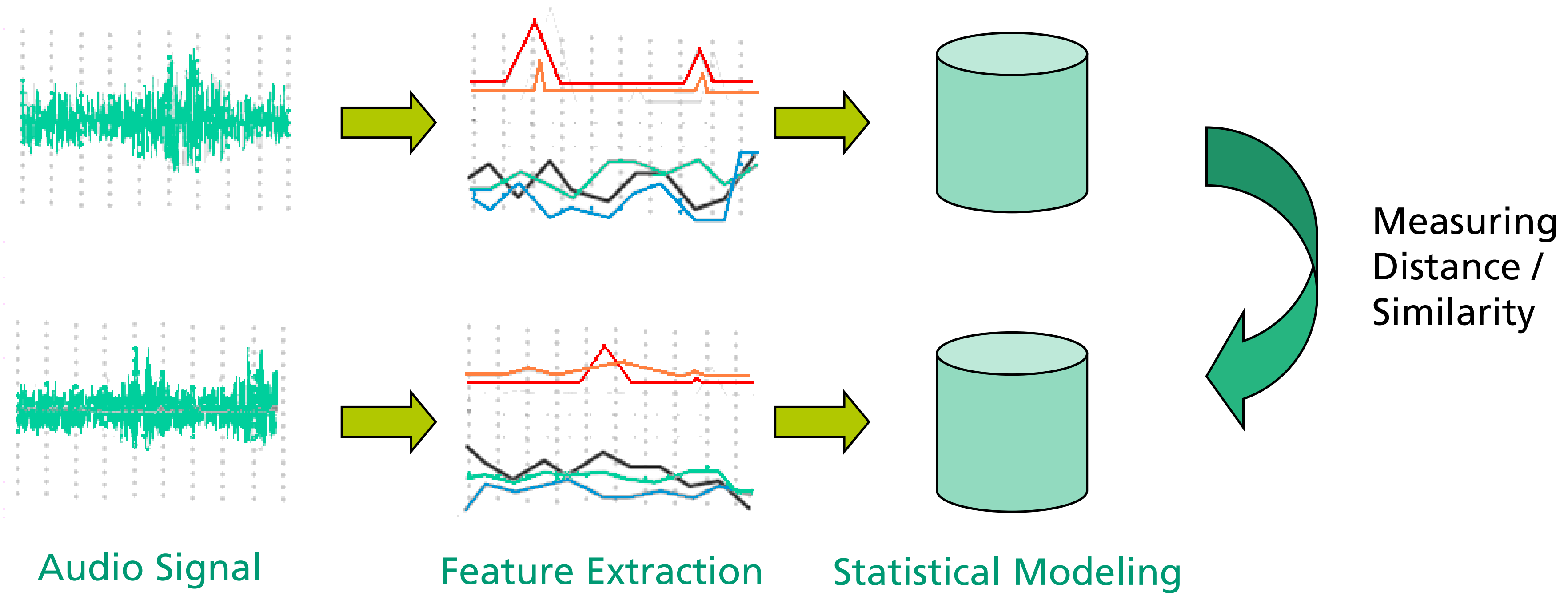
Task Introduction → Traditional Approach → Novel Approaches

Music Tagging
Introduction

- Tags
  - Textual (objective / subjective) annotations of songs
  - Examples
    - Instruments (drums, bass, guitar, vocals …)
    - Genre (classical, electro, hip hop)
    - Mood (mellow, romantic, angry, happy)
    - Miscellaneous (noise, loud, ambient)
- Challenge
  - Music pieces change their characteristics over time
    - E.g.: trumpet plays only in the chorus

Music Tagging
Traditional Approach

- Typical processing pipeline

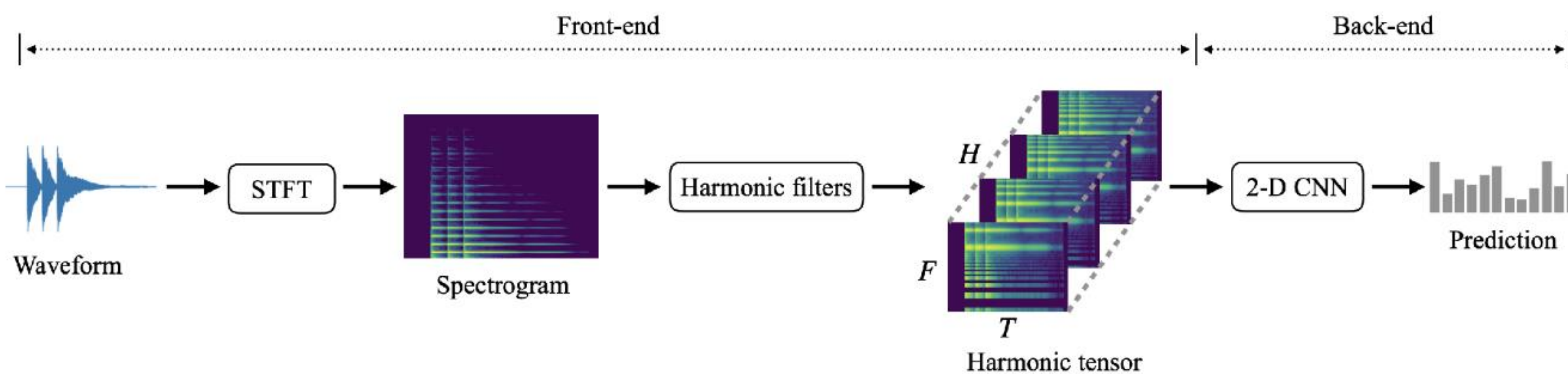


Fraunhofer IDMT

Music Tagging
Novel Approaches

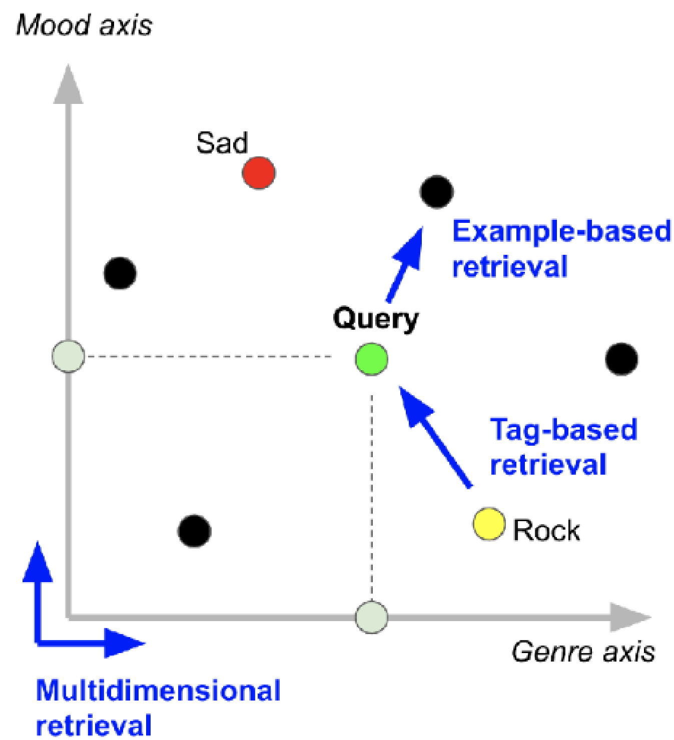
- Joint representation learning & classification using CNNs
  - Input: spectrograms (2D) or audio samples (1D end-to-end)

- Integrate musical knowledge in network design (e.g., filter shapes)



Music Tagging
Novel Approaches

- Disentanglement learning
  - Multiple semantic concepts (e.g. genre, instrument, mood)
    - are learnt jointly
    - remain separable in the embedding space
  - Improves tagging (classification) and recommendation (similarity)

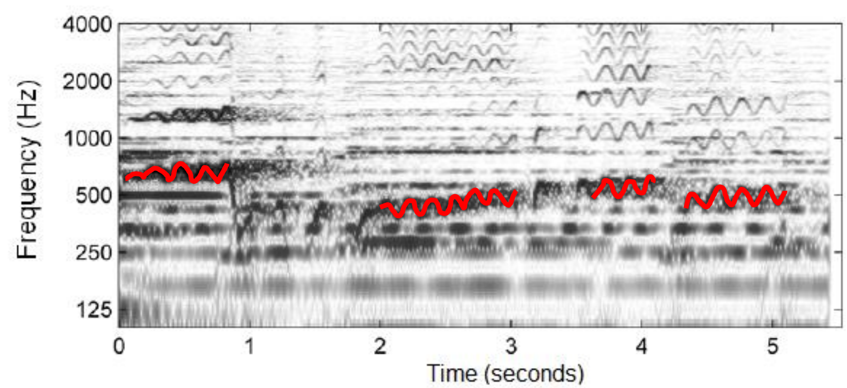


Lee et al.: Metric Learning vs. Classification for Disentangled Music Representation Learning, ISMIR 2020

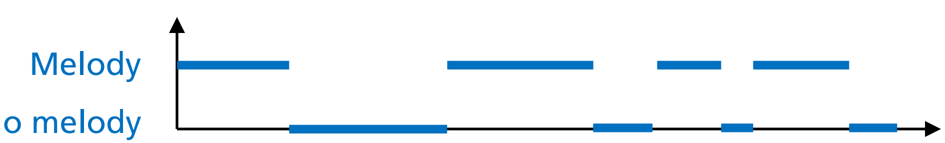
Pitch Detection
Introduction

- Pitch
    - (Subjective) psychoacoustic attribute of sound
    - Allows ordering from low to high in a frequency-related scale
        - Pitch ≠ frequency !
- Two subtasks

1) Pitch detection



2) Voicing detection

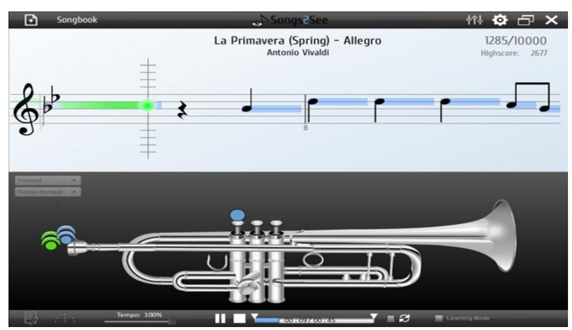


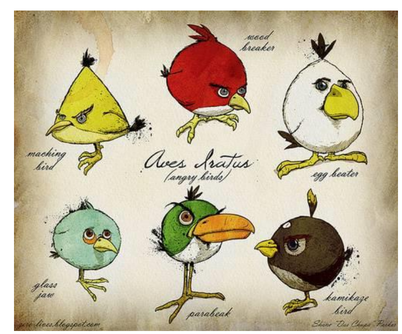
Müller, Fundamentals of Music Processing, Springer 2015, Figure 8.16(a)

Fraunhofer IDMT

Pitch Detection
Application Scenarios

- Music Instrument Tuning
- Music Education
- Music Transcription
- Bird Recognition





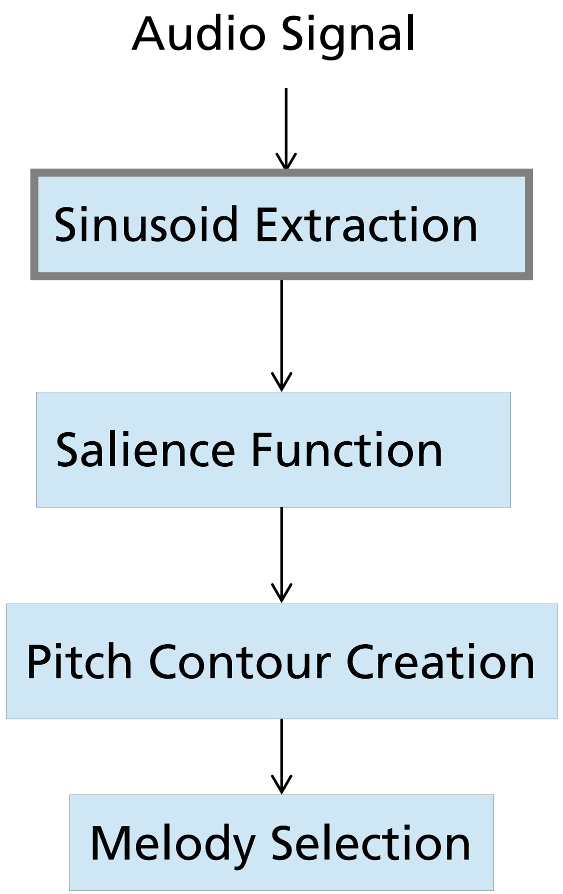
http://www.guitaradventures.com/wp-content/uploads/Tuning-your-guitar.jpg
https://cdn2.whatoplay.com/screenshots/2631slide-4.jpg
https://cdn.androidcommunity.com/wp-content/uploads/2010/11/500x_angrybirdsdarwin.jpg

Pitch Detection
Tasks

- Sorted by increasing complexity/difficulty
    - Pitch detection of isolated monophonic instruments (ex: trumpet)
    - Pitch detection of isolated polyphonic instruments (ex: guitar)
    - Predominant melody extraction in polyphonic music
    - Polyphonic melody extraction

Pitch Detection
Traditional Approach

- Sinusoid Extraction
  - Equal loudness filter
  - STFT
  - Detection of predominant peaks
  - Frequency refinement via instantaneous frequency (IF)

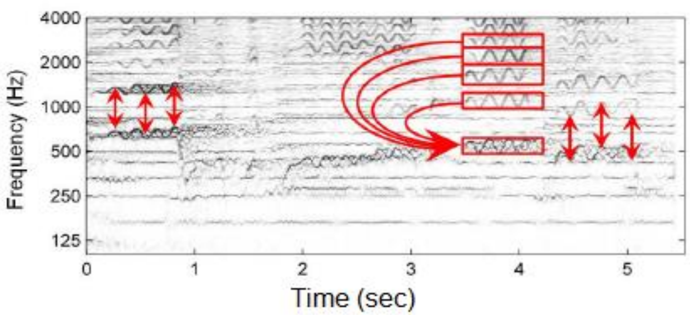


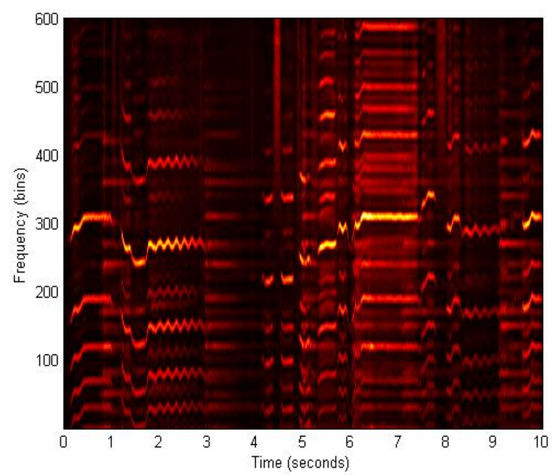
Salamon & Gómez: Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics, IEEE Transactions on Audio, Speech and Language Processing, 20(6):1759-1770, Aug. 2012.

Fraunhofer IDMT

Pitch Detection
Traditional Approach

- Salience Function
  - Harmonic summation
    - Sum over possibile harmonic frequencies
    - Frequencies → pitch candidates





Audio Signal

↓

Sinusoid Extraction

↓

Salience Function

↓

Pitch Contour Creation

↓

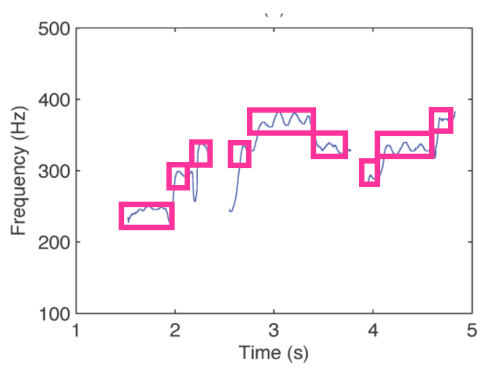
Melody Selection

Müller, Fundamentals of Music Processing, Springer 2015, Figure 8.15(a)

Fraunhofer IDMT

Pitch Detection
Traditional Approach

- Pitch contour creation & melody selection
  - Auditory streaming cues → group peaks to continuous paths (pitch contours)
  - Select melody contours using features (e.g. average pitch / salience, vibrato)
  - Note formation (one pitch value)



Pitch contour(s)

↓

Note events

Audio Signal

↓

Sinusoid Extraction

↓

Salience Function

↓

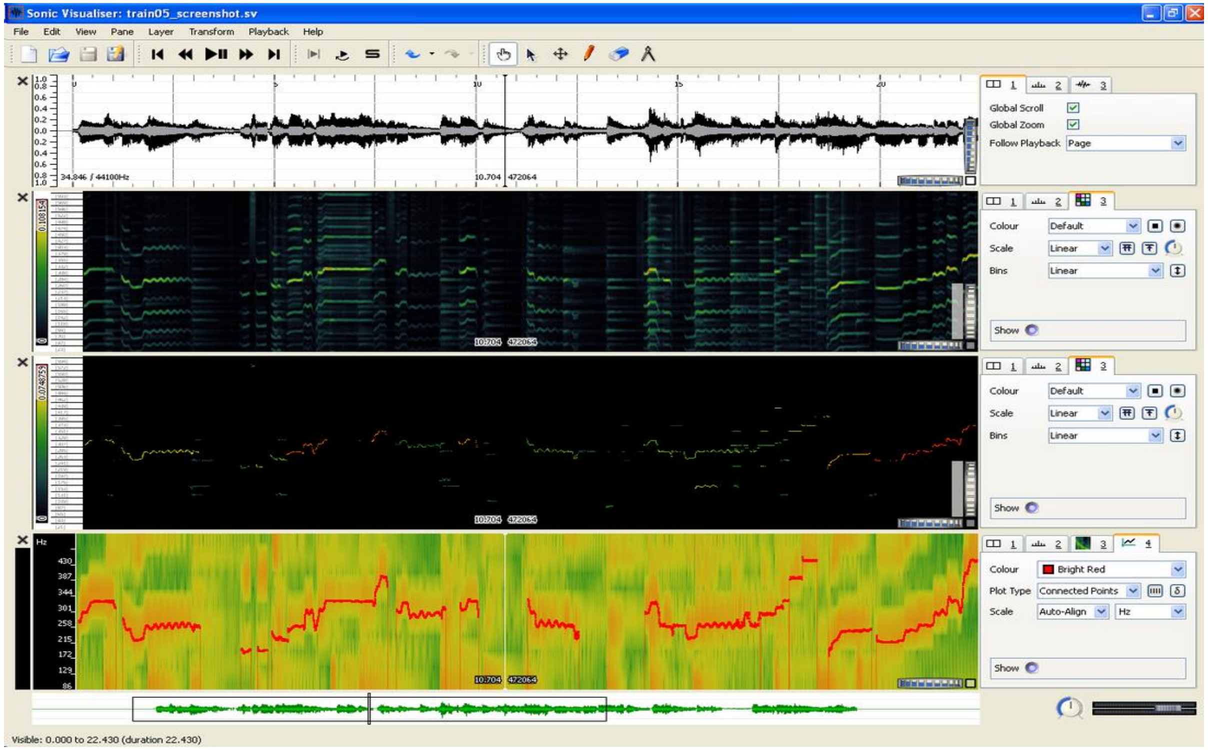
Pitch Contour Creation

↓

Melody Selection

Pitch Detection
Traditional Approach (Melodia)

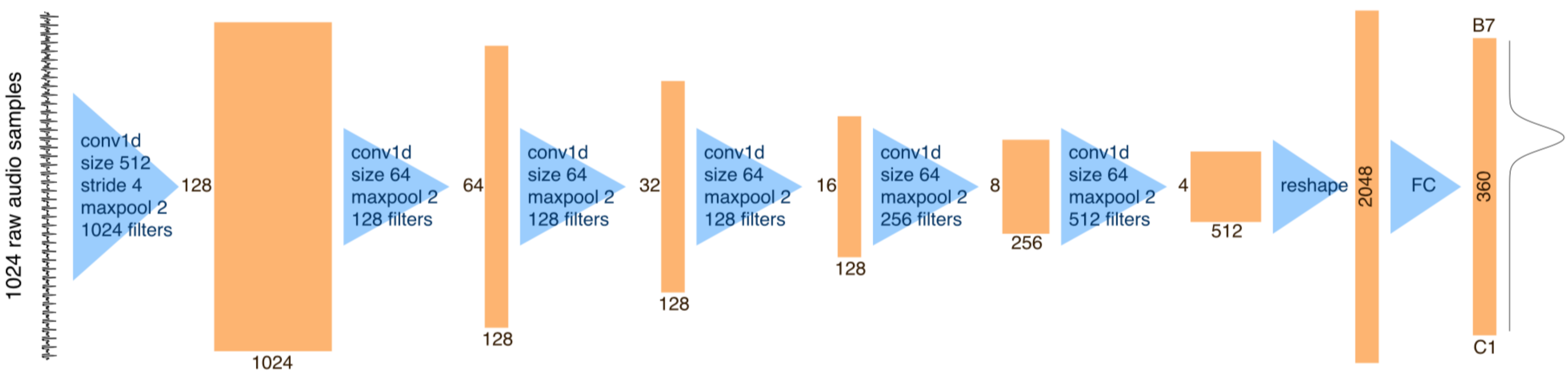
- Demo
  - Melodia plugin for Sonic Visualiser



Sonic Visualiser: http://www.sonicvisualiser.org/
Melodia plugin: http://mtg.upf.edu/technologies/melodia

Pitch Detection
Novel Approaches

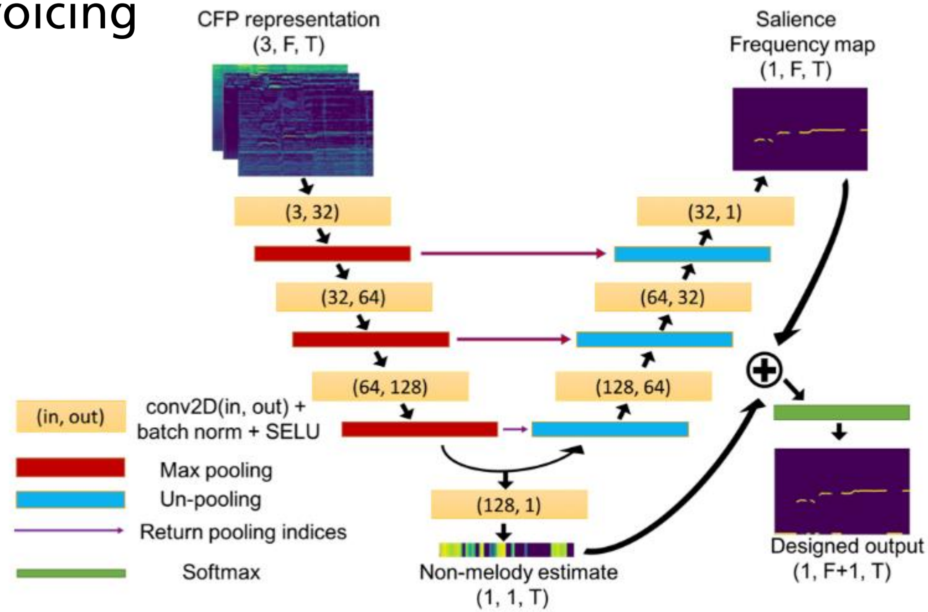
- CREPE (Convolutional Representation for Pitch Estimation)
  - Monophonic pitch tracker
  - End-to-end modeling
    - (Raw) audio samples (16 kHz) → pitch likelihoods
    - 20 cent resolution (5 pitch bins per semitones)



Kim et al.: Crepe: A Convolutional Representation for Pitch Estimation, ICASSP, pp. 161-165, 2018

Pitch Detection
Novel Approaches

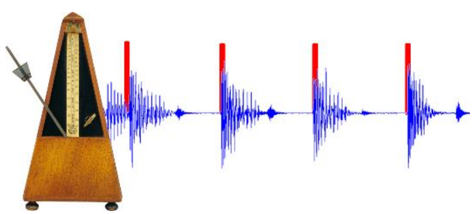
- Auto-encoder structure (U-Net)
  - Mapping from multiple time-frequency representations (2D) to pitch saliency map (2D)
  - Embedding encodes pitch voicing (melody activity)



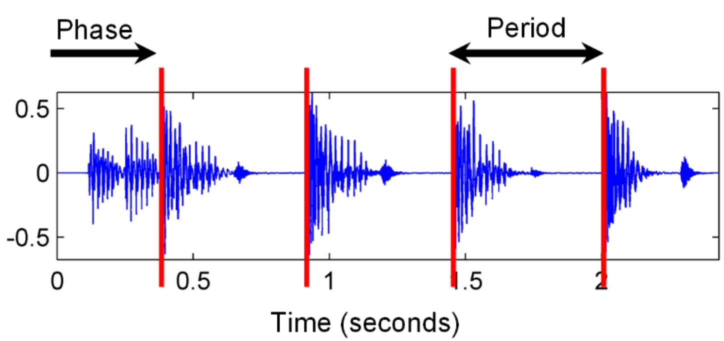
Hsieh et al.: A Streamlined Encoder/Decoder Architecture for Melody Extraction, ICASSP, pp. 156-160, 2019

Tempo Detection
Introduction

- Tempo [beats / minute]
  - Frequency with which humans tap along the beat



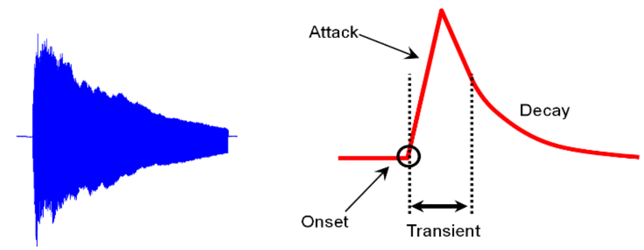
- Beat tracking
  - Estimating precise beat positions



Müller, Fundamentals of Music Processing, Springer 2015, Chapter 6 (cover image)
Müller, Fundamentals of Music Processing, Springer 2015, Figure 6.1(b)

Tempo Detection
Introduction

- Note onsets → note beginning times
  - Clearly defined for plucked string and percussion instruments
  - Ambiguous for wind & brass instruments
- Onset detection
  - Onset detection function
  - Peak picking

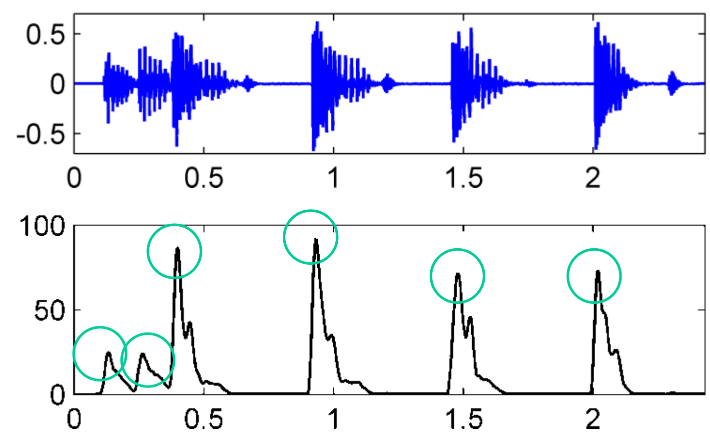


Audio samples → Note envelope

Audio samples

↓

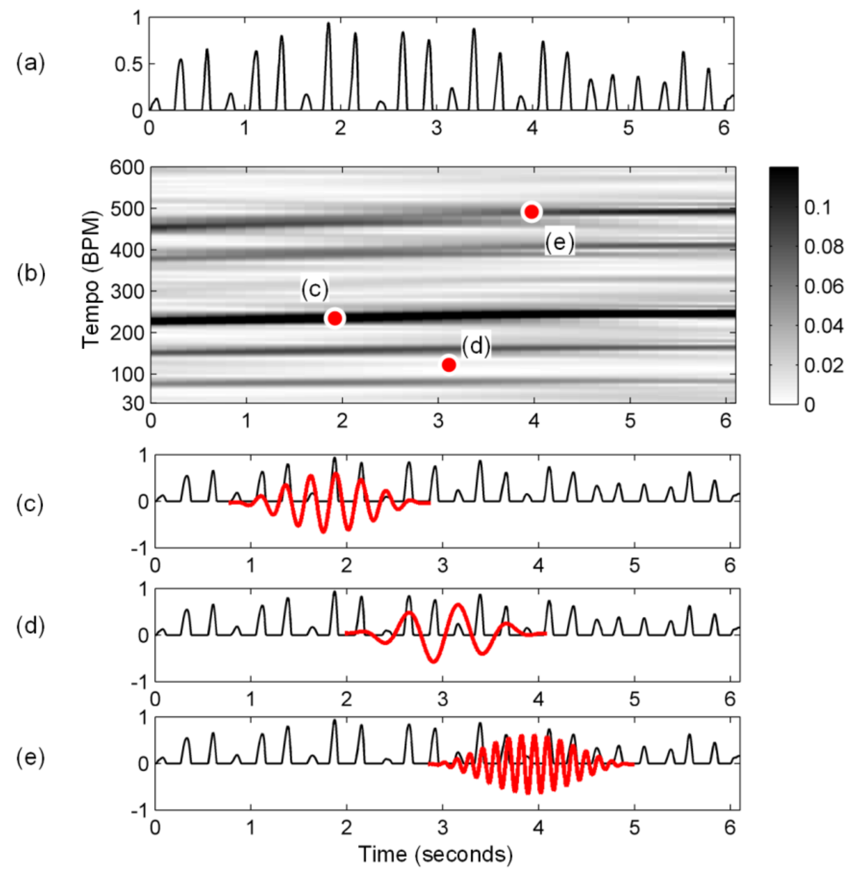
Onset detection function & peaks



Müller, Fundamentals of Music Processing, Springer 2015, Figure 6.2
Müller, Fundamentals of Music Processing, Springer 2015, Figure 6.3(a)&(b)

Tempo Detection
Traditional Approach

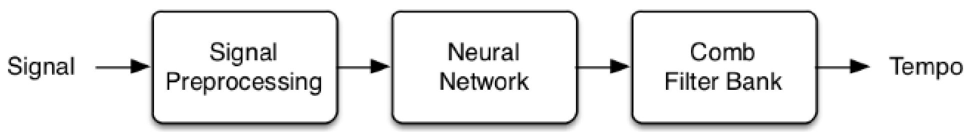
- Predominant local pulse (PLP)
  - Correlation with local (windowed) periodic patterns
- Tempogram
  - Local likelihood of different tempo candidates
  - Allows to follow tempo changes (classical music)

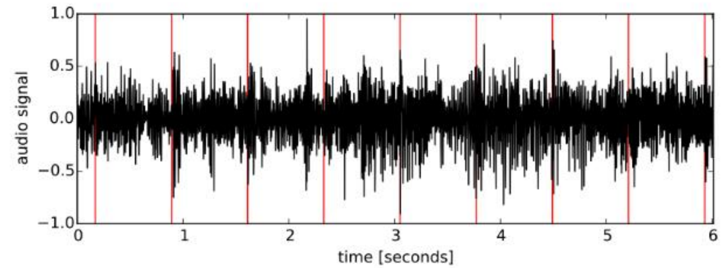


Grosche & Müller: A Mid-Level Representation for Capturing Dominant Tempo and Pulse Information in Music Recordings, ISMIR, 2009

Tempo Detection
Novel Approaches

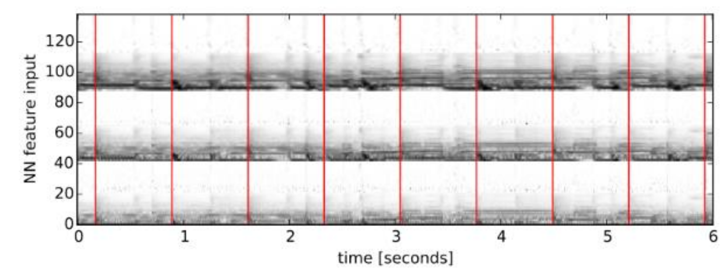
- Approach





(a) Input audio signal

- Signal representation
  - Stacking of 3 STFT magnitude spectrograms (N=1024, 2048, 4096)
  - Log-amplitude & log-frequency

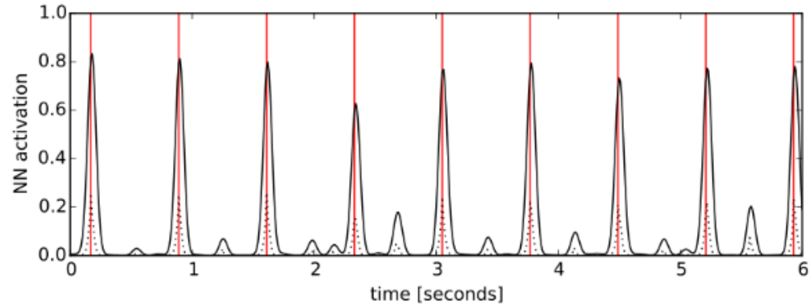


(b) Input to the neural network

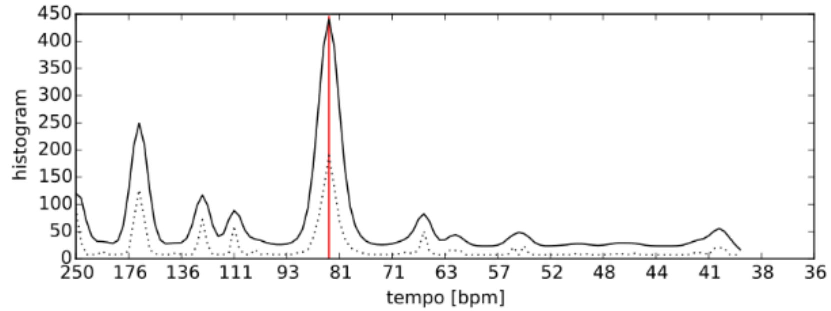
Böck et al.: Accurate Tempo Estimation based on Recurrent Neural Networks and Resonating Comb Filters, ISMIR, 2015

Tempo Detection
Novel Approaches

- Neural Network
  - Recurrent (bi-directional LSTM) layer
  - Outputs beat activation function
- Comb filter bank
  - Multiple comb filters → detect periodicities
- Estimate tempo from histogram maximum



(c) Neural network output (beat activation function)

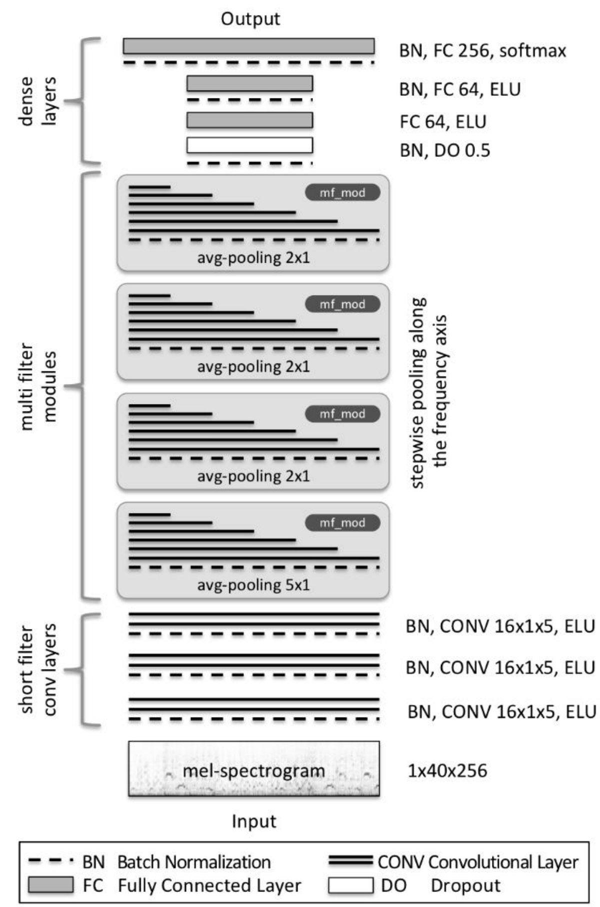


(f) Weighted histogram with summed maxima

Böck et al.: Accurate Tempo Estimation based on Recurrent Neural Networks and Resonating Comb Filters, ISMIR, 2015

Tempo Detection
Novel Approaches

- Signal representation
  - Sample rate ~ 11 kHz, 40-band mel spectrogram
- Tempo estimation → classification (256 classes: 30 – 285 bpm)
- Neural Network
  - 3 layers (short filters) → onsets
  - 4 multi-filter modules (parallel conv layers) → compress along frequency & find periodicities
  - Dense layers → tempo classification



Schreiber & Müller: A Single-Step Approach to Music Tempo Estimation using a Convolutional Neural Network, ISMIR, 2018

Instrument Recognition
Introduction

- Music ensembles include multiple instruments
  - Sound production (string / wind / brass / drum instruments)
  - Unique timbre
- Overlapping sound sources (solo recording vs. orchestra)
  - Unison (same pitch)
  - Harmonic intervals (overtone overlap)
  - Rhythmic interconnection (note attacks overlap)
- Classification on different taxonomy levels
  - Woodwind instruments → saxophone → tenor saxophone

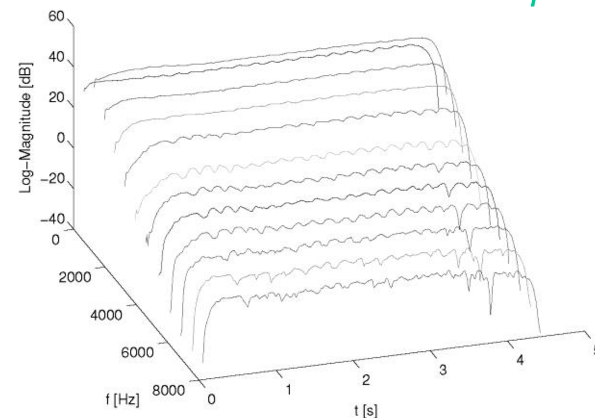
Instrument Recognition
Tasks

- Sorted by increasing complexity/difficulty
  - Instrument recognition of isolated note recordings
  - Instrument recognition on isolated instrument tracks
  - Predominant instrument recognition in ensemble recordings
  - Instrument tagging (classify all instruments)

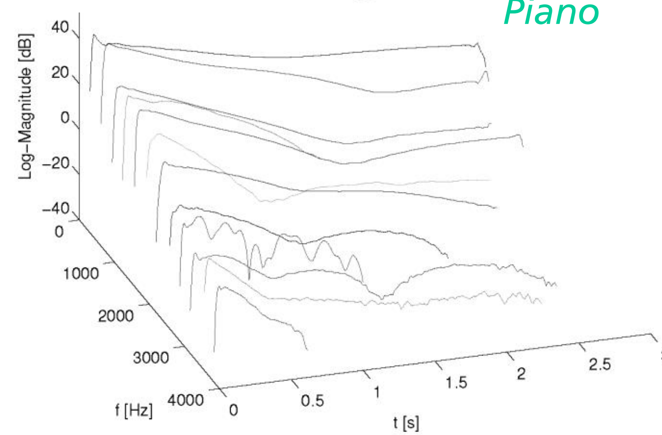
Instrument Recognition
Traditional Approach

- Multiple categories of audio features
  - Per spectral frame
    - E.g. spectral flux & flatness
  - Per overtone / partial
    - E.g. modulation rate & frequency
  - Note-event level
    - Magnitude ratios of overtones
- Aggregation
  - Features (overtones)
  - Classification results (notes)

Trumpet



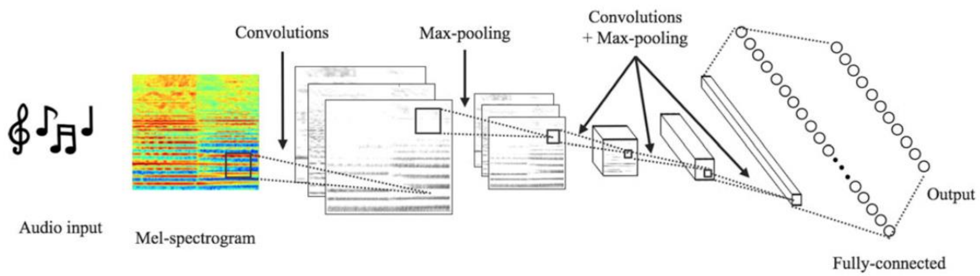
Piano



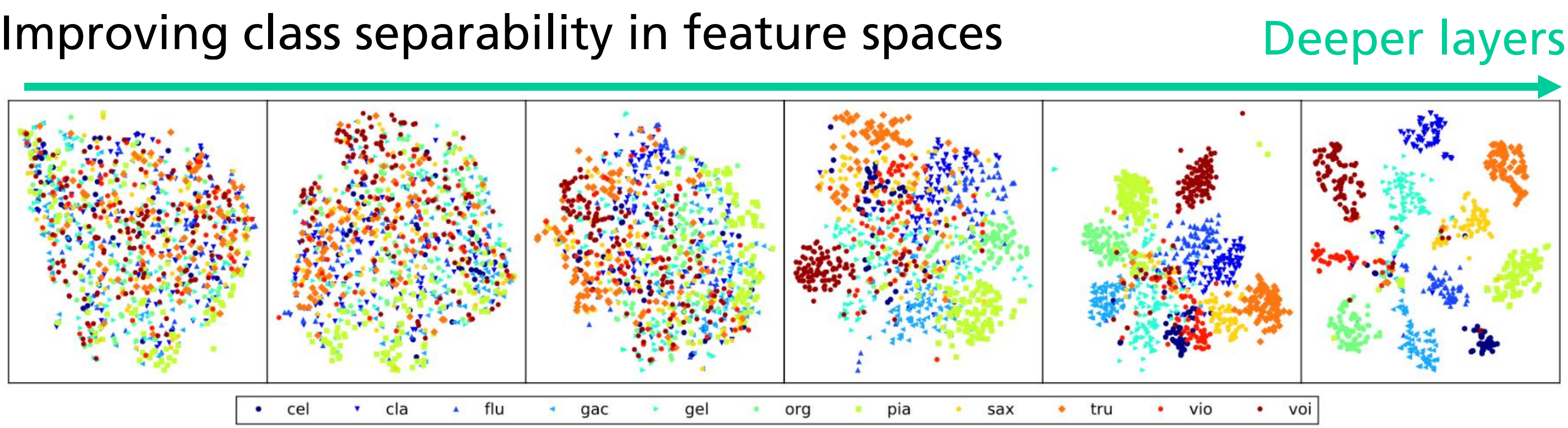
Grasis et al.: A Multiple-Expert Framework for Instrument Recognition, CMMR, 2013

Instrument Recognition
Novel Approaches

- Feature learning on mel spectrograms
- Convolutional layers & pooling & dense layers (classification)



- Improving class separability in feature spaces



Han et al.: Deep Convolutional Neural Networks for Predominant Instrument Recognition in Polyphonic Music, IEEE/ACM TASLP, 25(1), 2016

Fraunhofer IDMT

Summary

- Music Information Retrieval
- Case Studies
  - Music Tagging
  - Pitch Detection
  - Tempo Estimation
  - Instrument Recognition
- Main trends
  - Adapt (data-driven) deep learning methods to audio domain
  - Incorporate music domain knowledge