# MUSIC INFORMATION RETRIEVAL

## AND THE PRINCIPLES OF AUDIO PROCESSING AND ANALYSIS

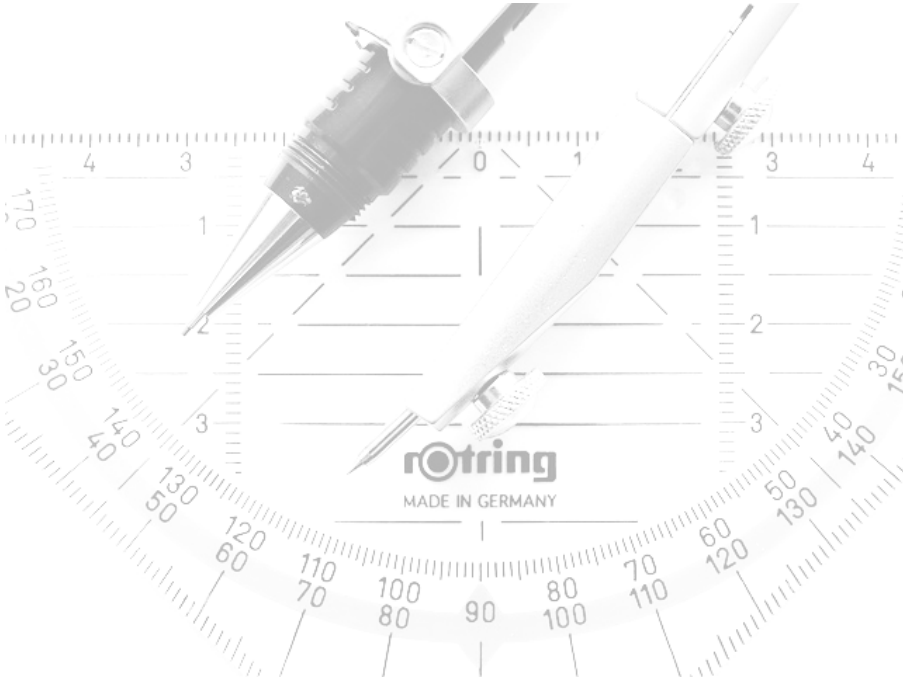
**Alexander Schindler**
Scientist
Information Management
Center For Digital Safety & Security

**AIT Austrian Institute Of Technology Gmbh**
Alexander.schindler@ait.ac.at

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

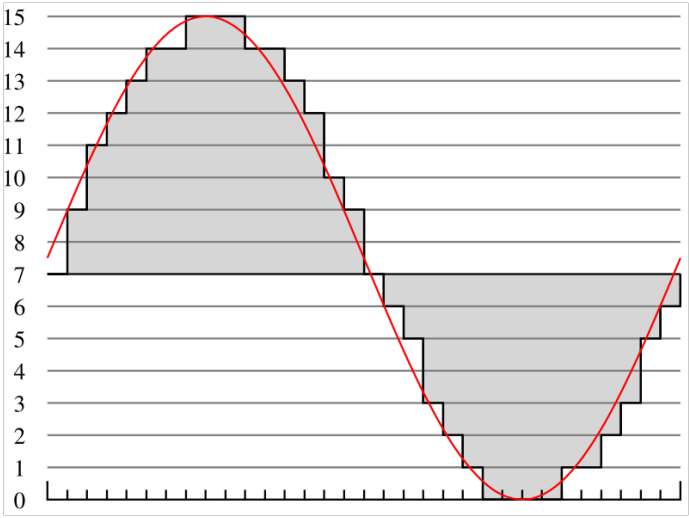
# FEATURE EXTRACTION FROM MUSIC

## Extracting Music Information



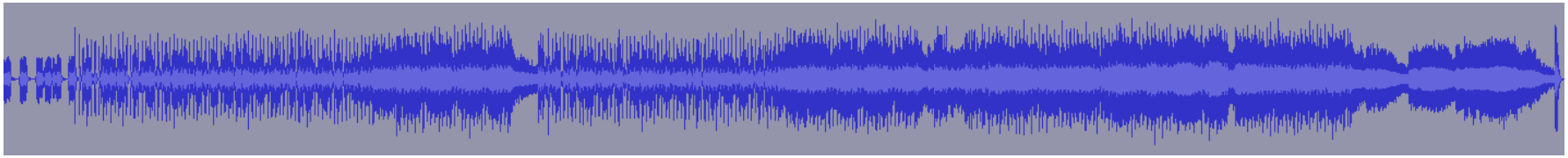
# TOO MUCH AUDIO DATA

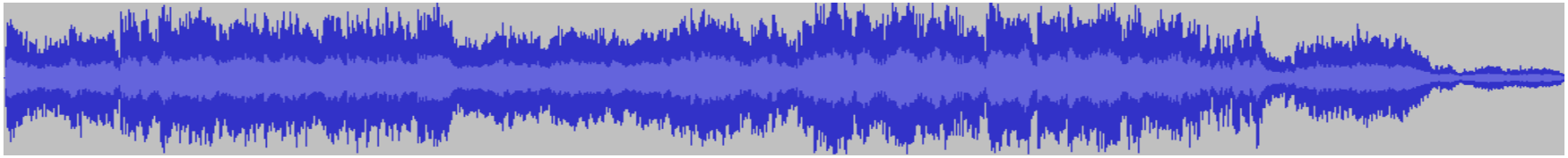
- Digital Audio
  - Sampling Rate: 44,100 Hz
  - 16-bit resolution for each channel
    - 2 channels for stereo
  - 88,200 Integers per second
  - 15,9 Millions for mainstream music (3min)
  - 60,5 MB

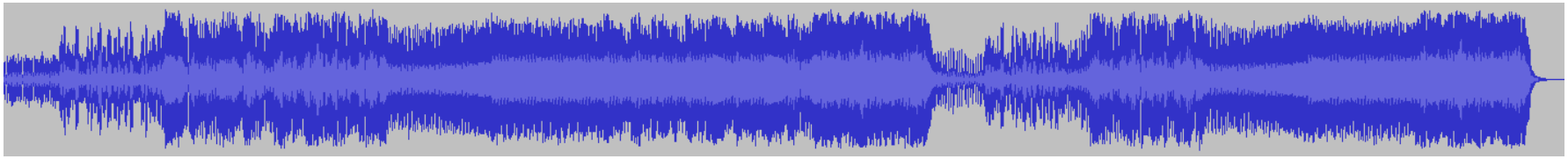


AUSTRIAN INSTITUTE OF TECHNOLOGY

# EXCERCISE: SAME GENRE?

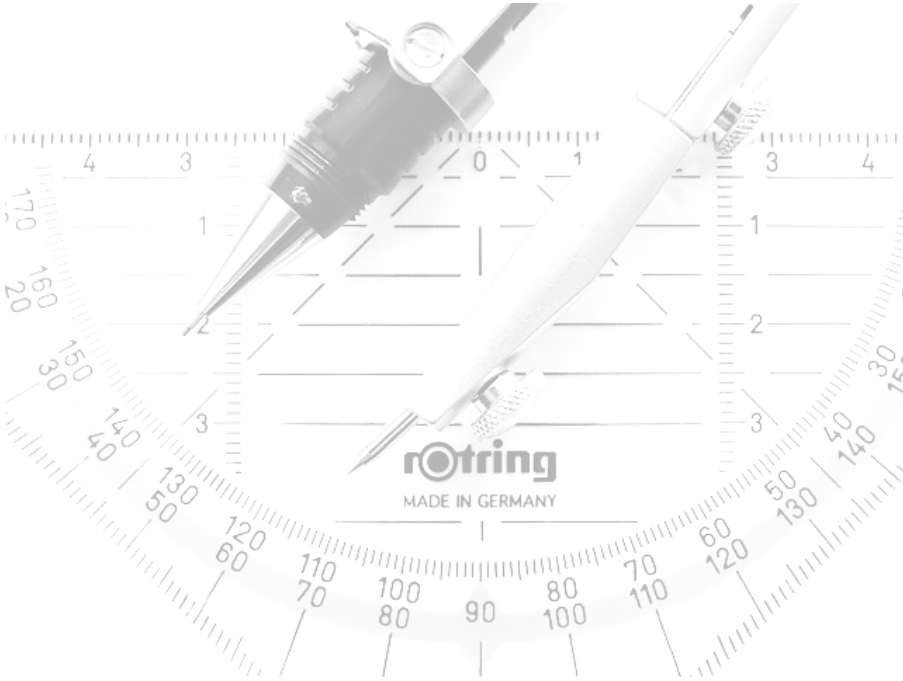






AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

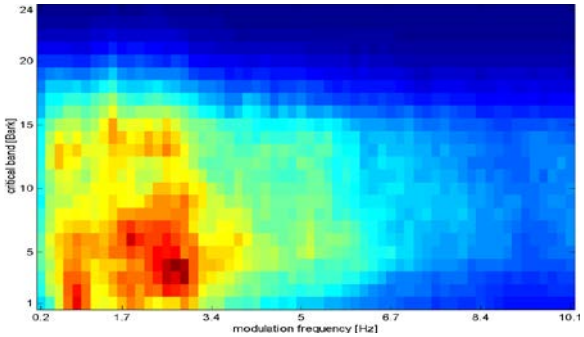
# FEATURE EXTRACTION FROM MUSIC



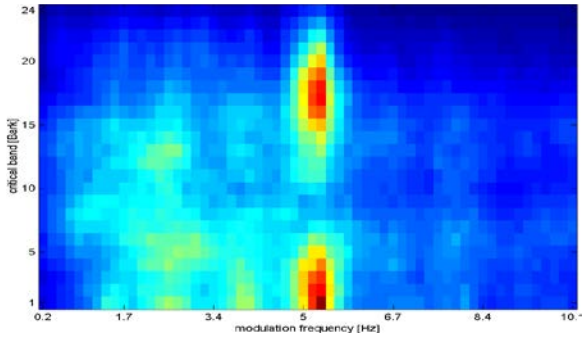

By example…

# RHYTHM PATTERN (RP)

- fluctuations on critical frequency bands (a.k.a. Fluctuation Pattern)
- covers rhythm in the broad sense

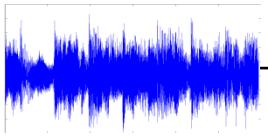



Classical




Rock

# RHYTHM PATTERN (RP)



Segmentation (6 s)

Spectrogram (FFT)

Bark scale (24 bands)

Spectral Masking

Decibel

Phon

Sone

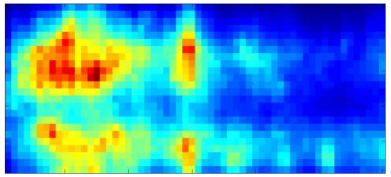
Modulation Amplitudes

Fluctuation Strength

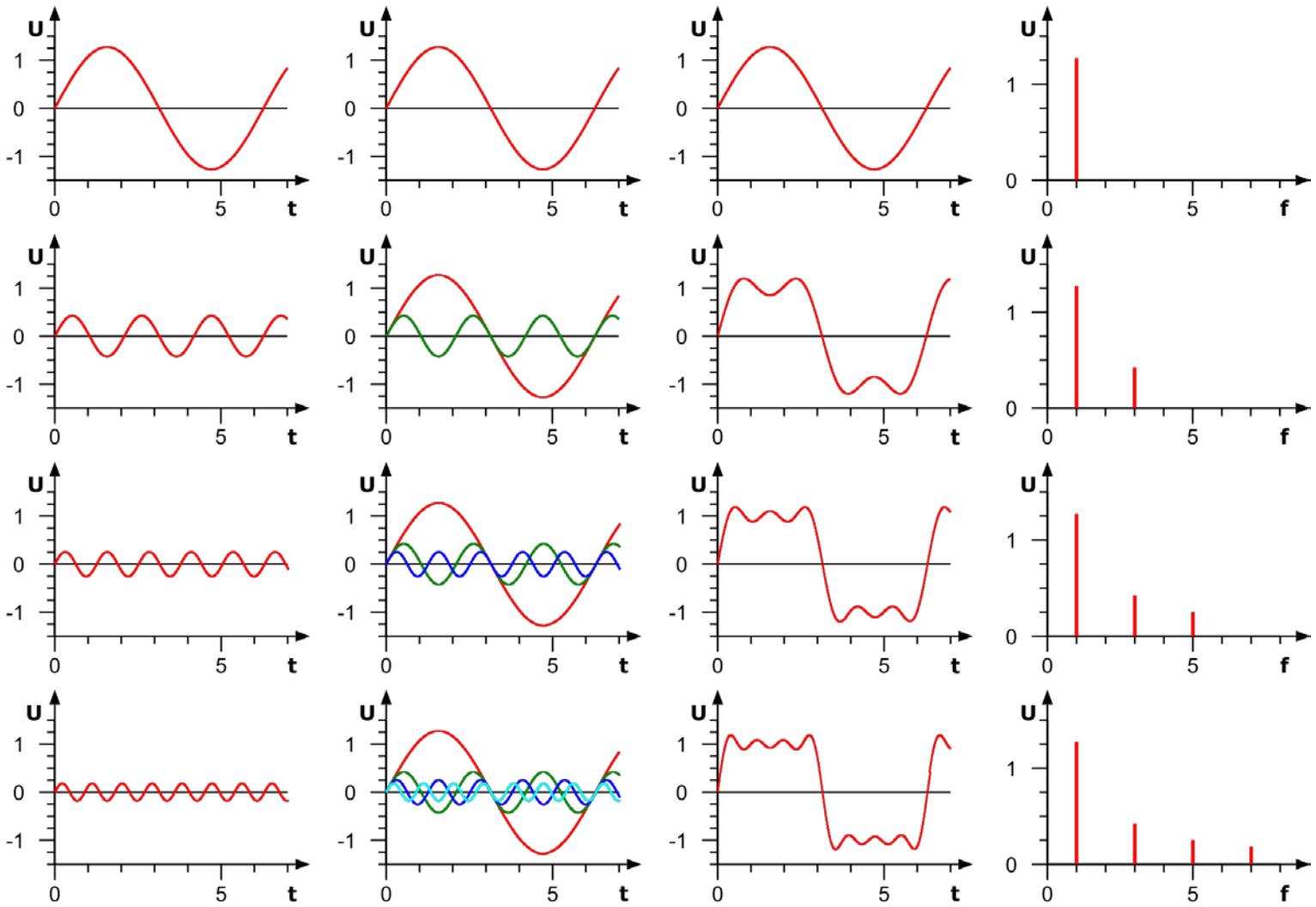
Filtering/Smoothing

Rhythm Pattern features

(1440 dimensions)

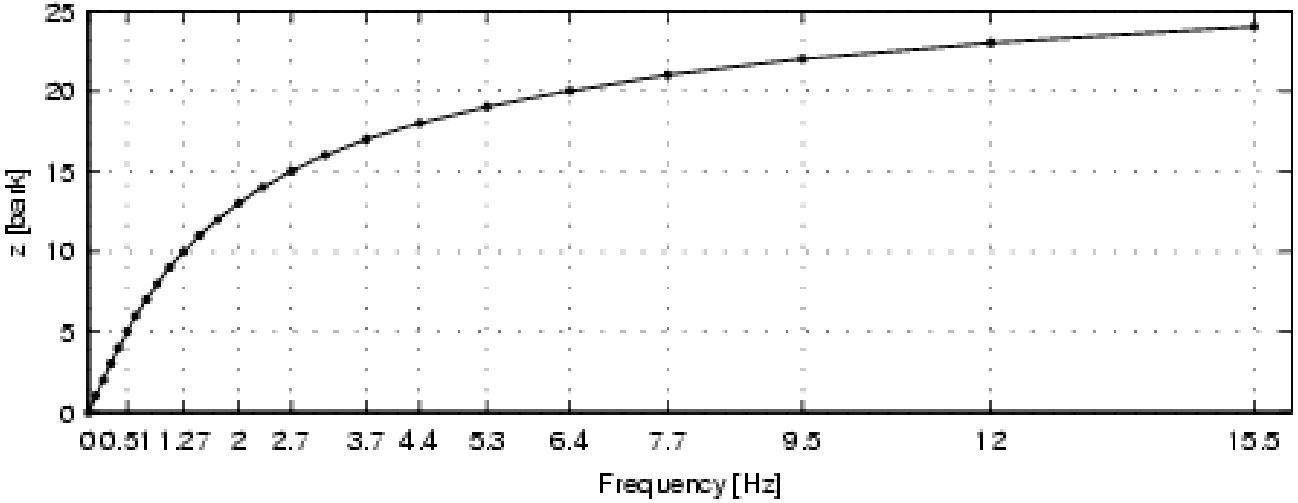


# FOURIER TRANSFORM



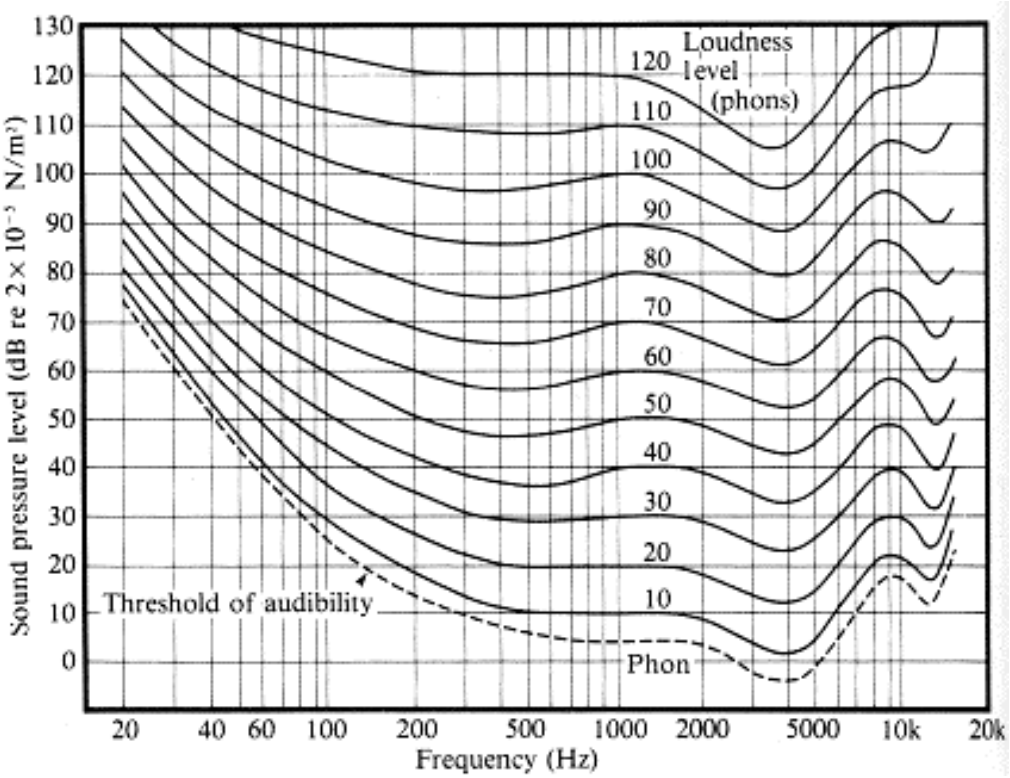
# BARK SCALE

- psychoacoustical scale (related to Mel scale)
- 24 „critical bands“ of hearing (non-linear)
- proposed by Eberhard Zwicker in 1961



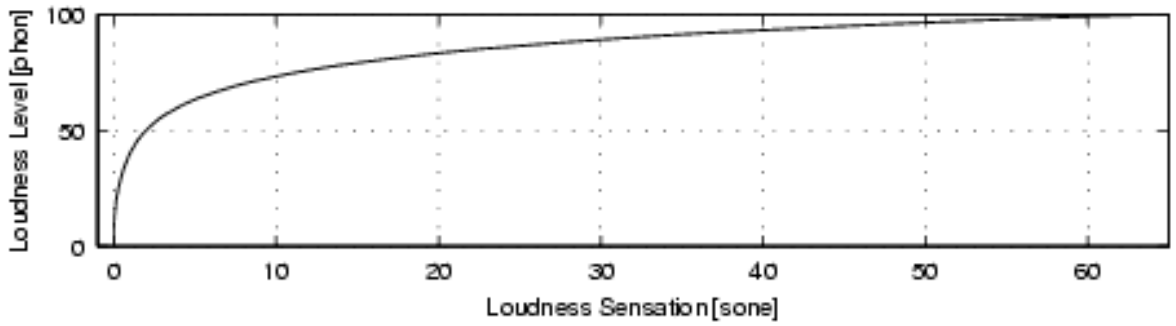
# Equal loudness curves (Phon)

- Relationship between sound pressure level in decibel and hearing sensation is not linear
- Perceived loudness depends on frequency of the tone
- equal loudness contours for 3, 20, 40, 60, 80, 100 phon




on-line test: http://www.phys.unsw.edu.au/jw/hearing.html

# Sone Transformation



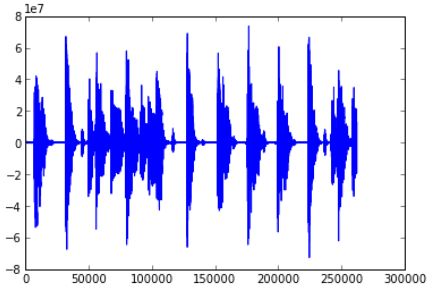

| Sone | 1 | 2 | 4 | 8 | 16 | 32 | 64 |
|---|---|---|---|---|---|---|---|
| Phon | 40 | 50 | 60 | 70 | 80 | 90 | 100 |

- Perceived loudness measured in Phon does not increase linearly
- Transformation into Sone
- Up to 40 phon slow increase in perceived loudness, then drastic increase
- Higher sensibility for certain loudness differences

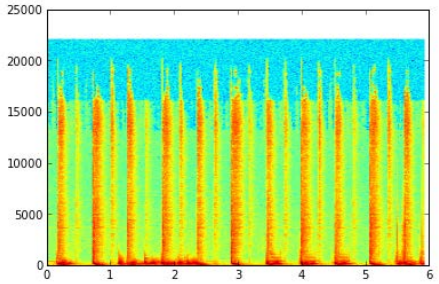
# RHYTHM PATTERN (RP): 2 EXAMPLES

**Queen – Another One Bites The Dust** (first 6 seconds)

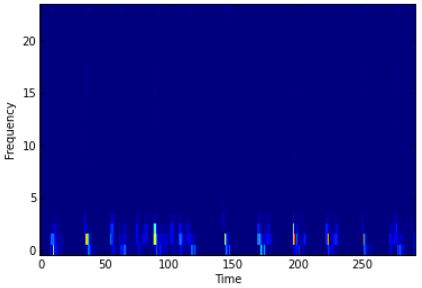
PCM Audio Signal



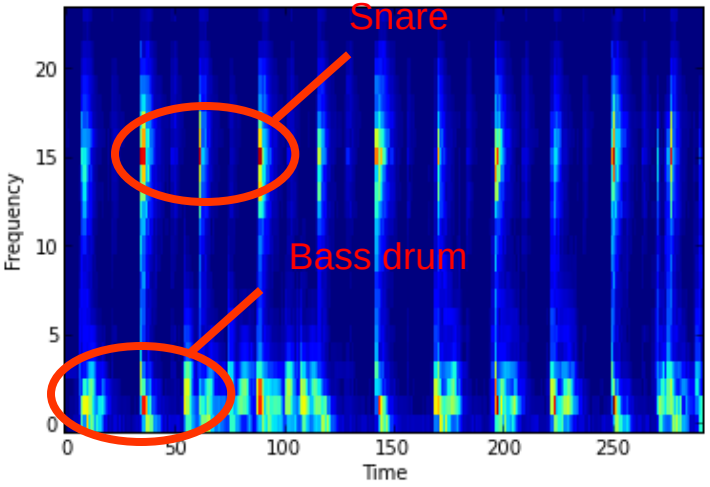

Power Spectrum



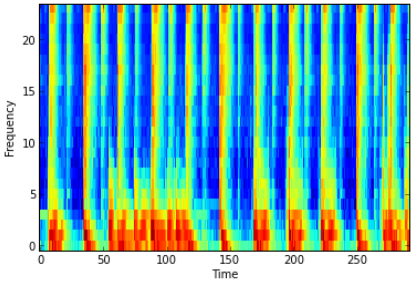

Bark Scale



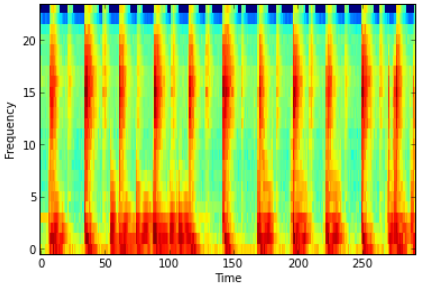

Sone




Decibel



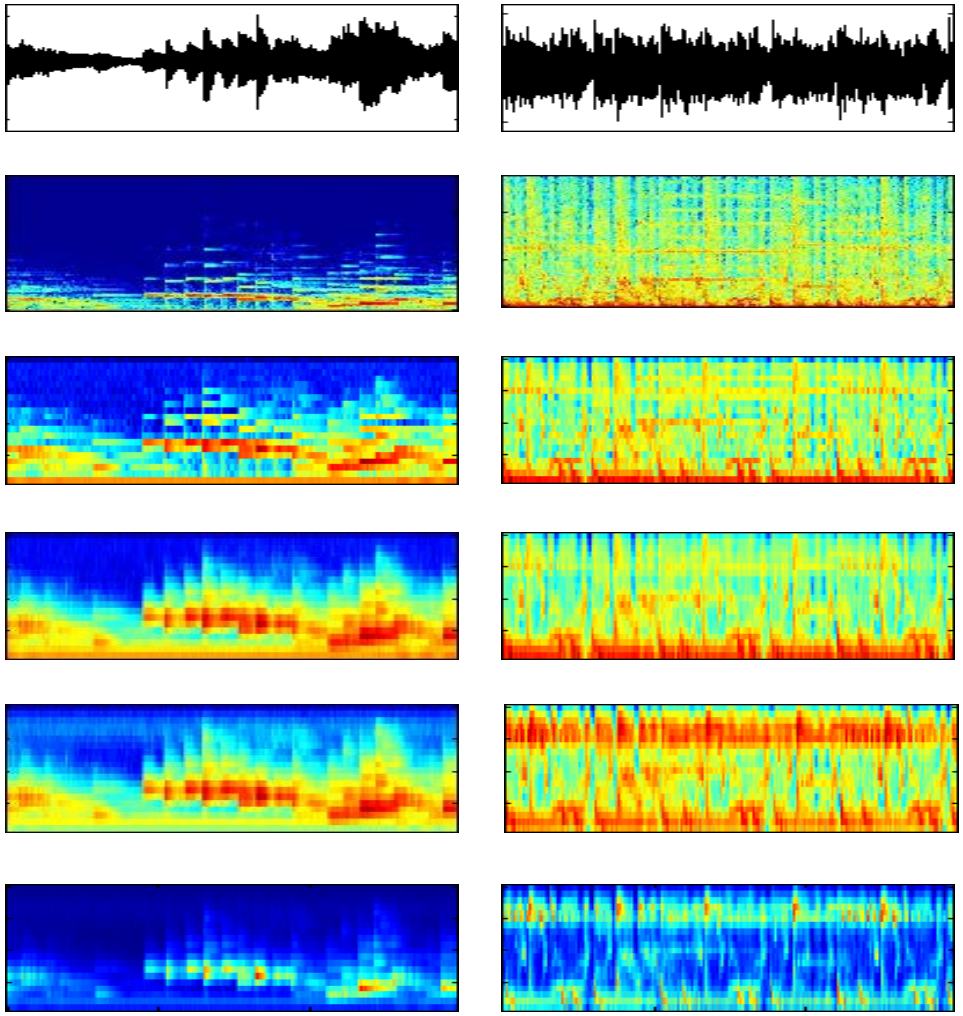

Phon



# RHYTHM PATTERN (RP): 2 EXAMPLES

Classical

Metal



**PCM Audio Signal**

Power Spectrum

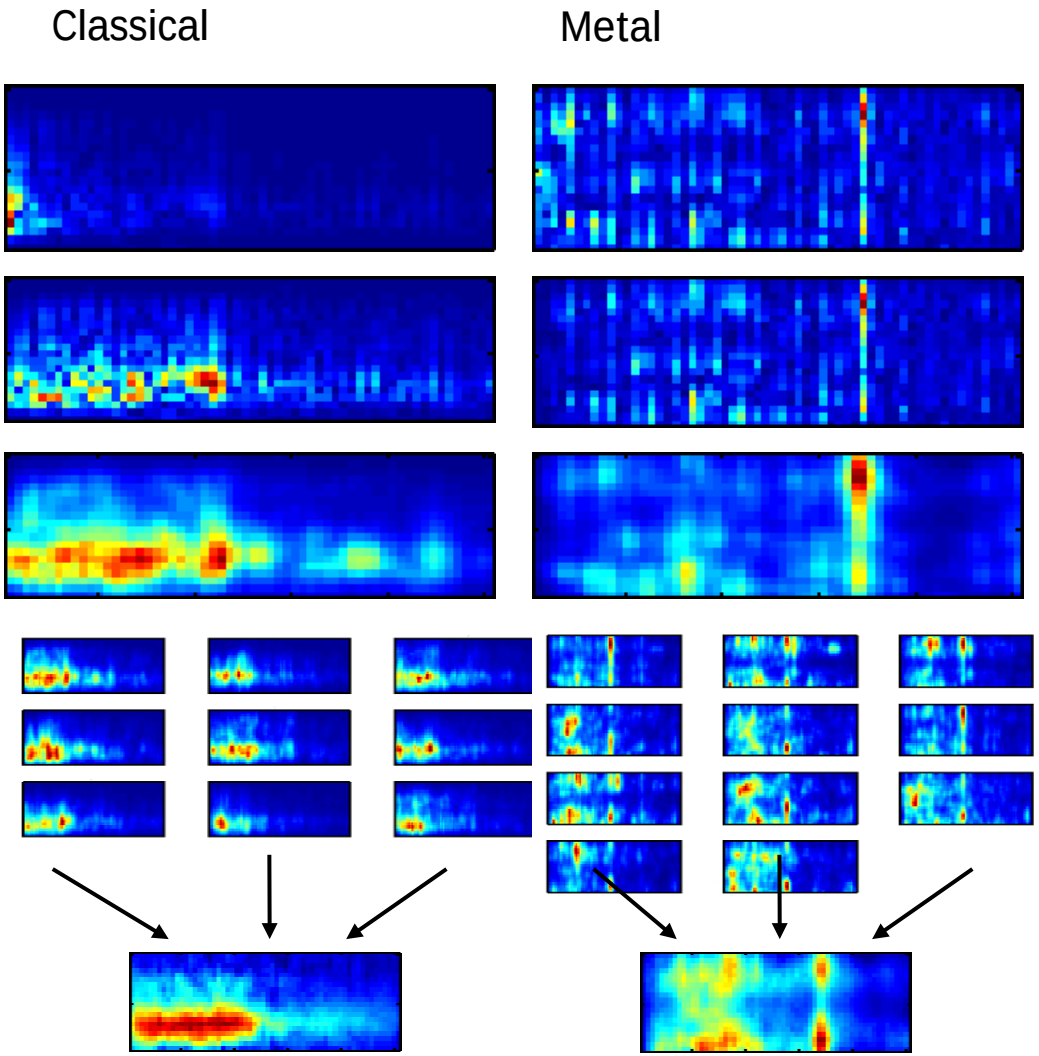
Frequency Bands

Masking Effects

Phon

Sone

# RHYTHM PATTERN (RP): 2 EXAMPLES




modulation amplitude spectrum (“cepstrum”)

Fluctuation Strength

Filter (Gradient, Gauss)

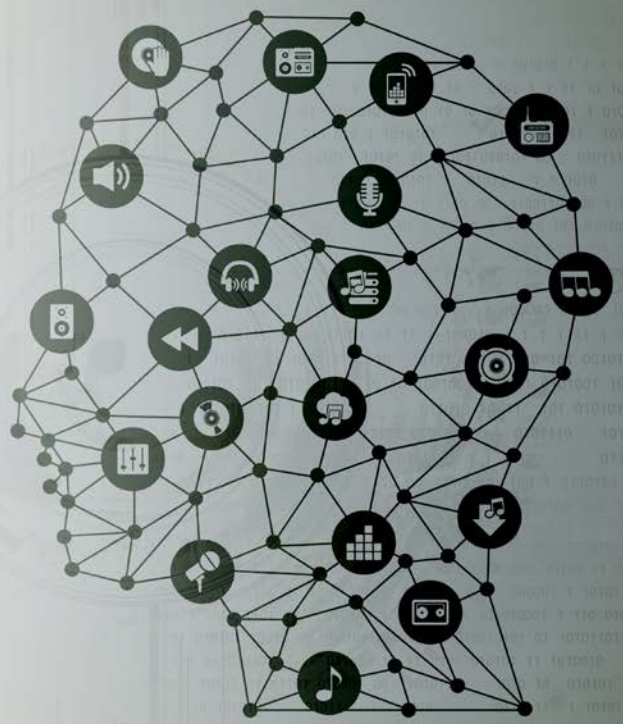
Median

24*60=
1.440-dim feature vec.

AUSTRIAN INSTITUTE OF TECHNOLOGY

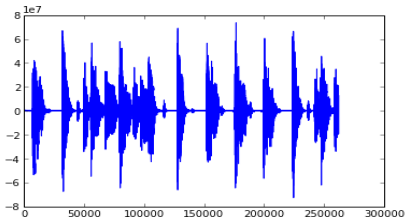
# DEEP LEARNING

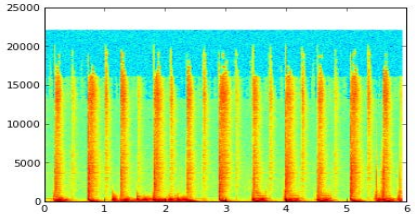
## for
## Music Information Retrieval

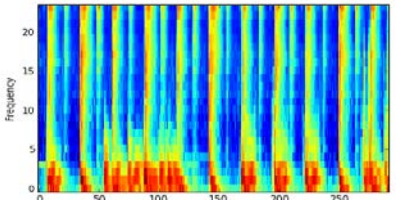


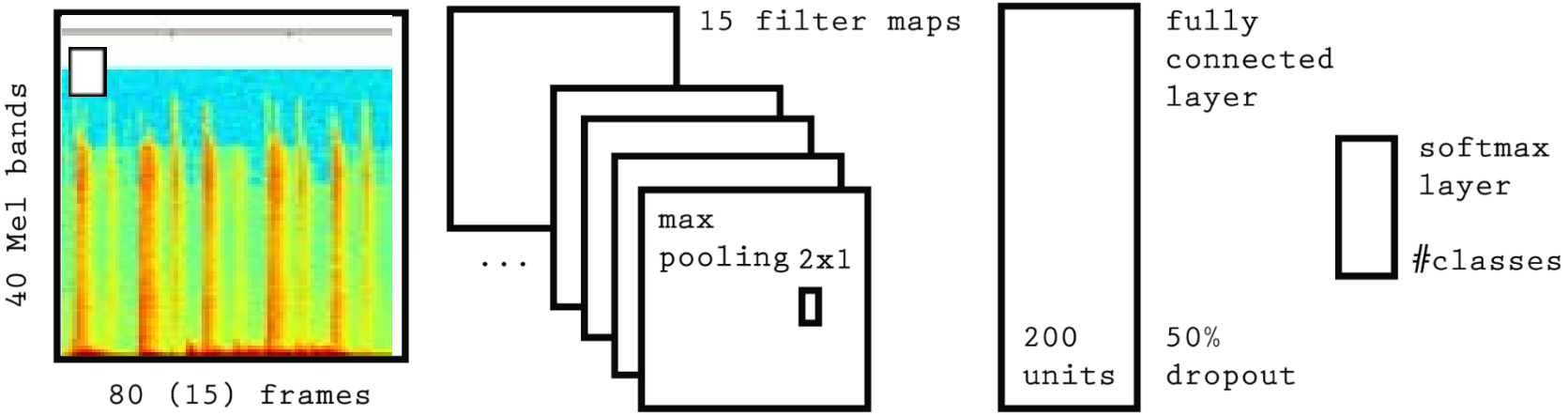
# DEEP LEARNING FOR MUSIC IR

Pre-Processing: Waveform → Spectrogram → 40 Mel bands → Log scale



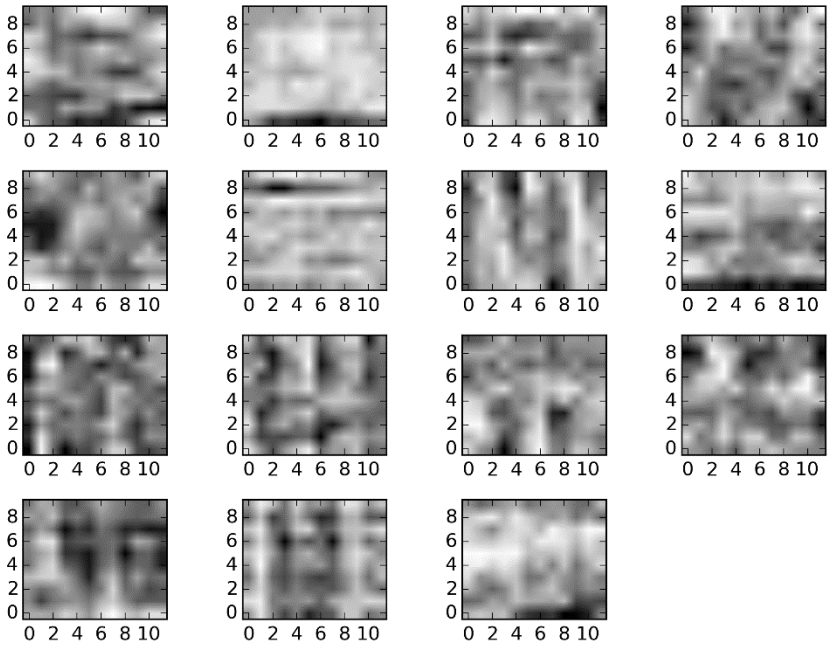




Winning algorithm MIREX 2015 music/speech classification task (99.73%) by Thomas Lidy

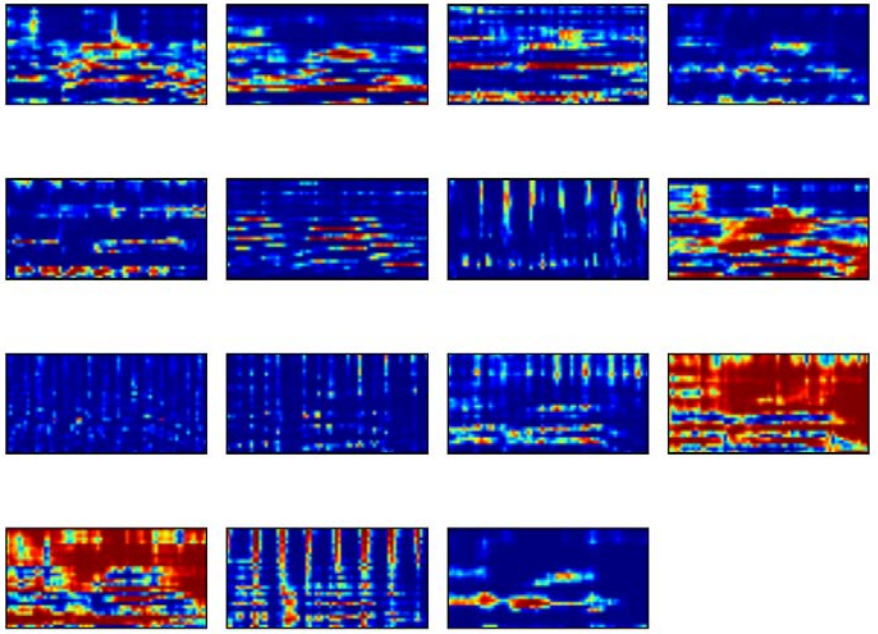
# VISUALIZING CNN FILTERS LEARNED FOR MUSIC/SPEECH CLASSIFICATION

AUSTRIAN INSTITUTE OF TECHNOLOGY

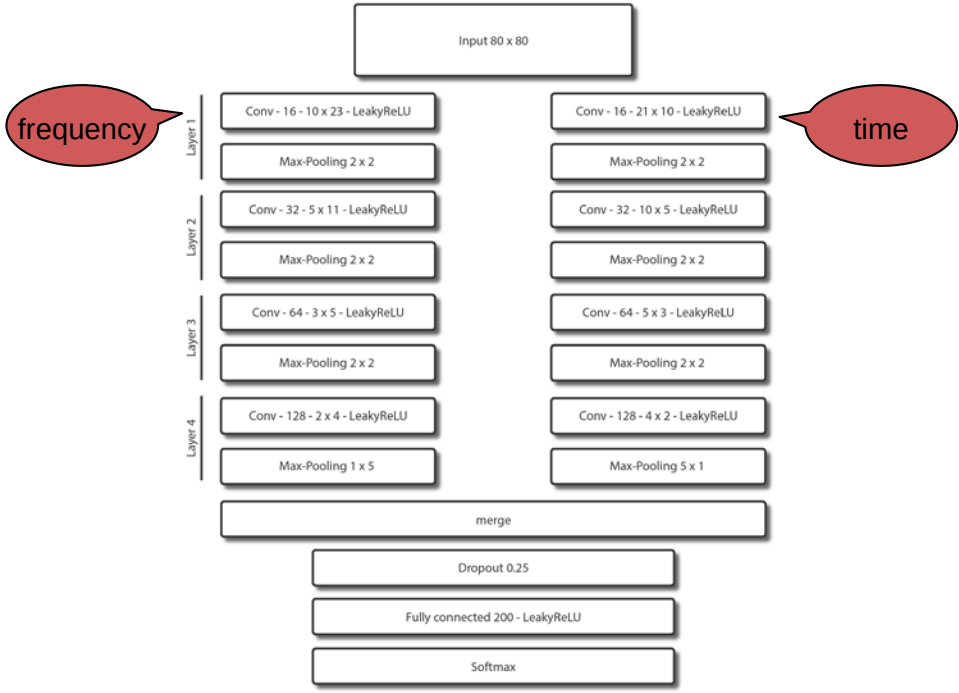
## Learned Filter Weights




## Convolved Spectrograms



# DEEP VS. SHALLOW




| | 100 epochs | | 200 epochs | |
|---|---|---|---|---|
| | **Shallow** | **Deep** | **Shallow** | **Deep** |
| GTZAN | 78.1 | **78.6** | **80.8** | 80.6 |
| ISMIRgenre | **85.5** | 84.1 | 84.9 | **85.1** |
| Latin | 92.4 | **94.4** | 93.5 | **95.1** |
| MSD | 63.9 | **67.2** | / | / |

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

# THANK YOU!