Music Information Retrieval

Derek Tingle (IDAGIO / Freelan)

(Traditional) Information Retrieval

- “Information Retrieval (IR) is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections” - Christopher Manning
  - https://nlp.stanford.edu/IR-book/information-retrieval-book.html
- Natural Language Processing (NLP)
- Typical problems:
  - Search
  - Ranking
  - Recommendations

Multimedia Information Retrieval

- Documents aren’t strictly text
- Computer Vision
  - Massively popular / important
  - Facial recognition
  - Object detection
- Video
  - Classification
  - Fingerprinting
- Music Information Retrieval
  - ...

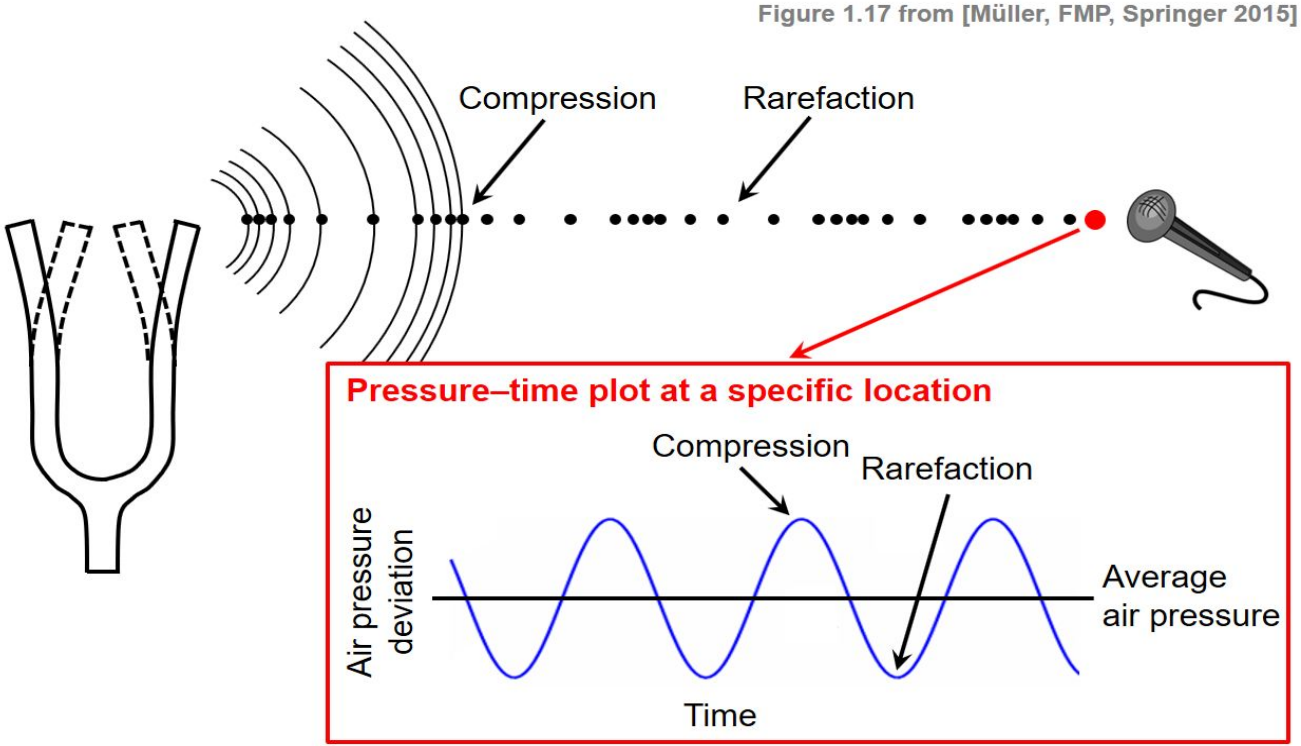
MIR Tasks

- Audio fingerprinting
- Cover song identification
- Classification (genre, mood, instrument, vocalist, etc.)
- Drum transcription
- Downbeat detection / tempo estimation
- Fundamental frequency tracking
- Chord / key estimation
- Lyrics-to-audio alignment
- Score following
- Source separation
- .... (see MIREX)

How do we hear stuff?

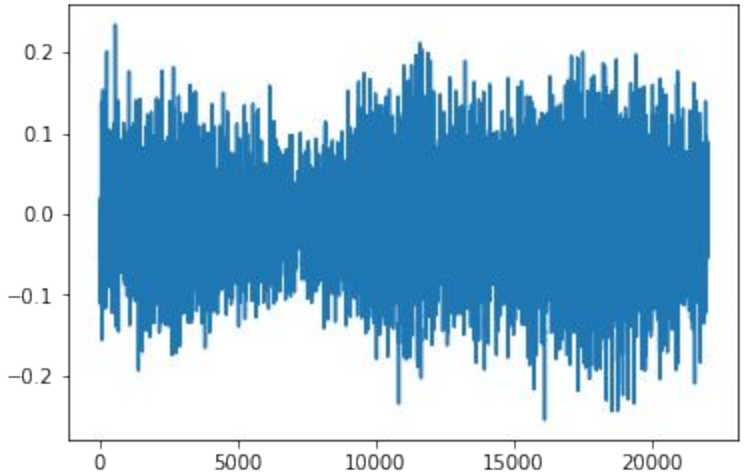
https://www.audiolabs-erlangen.de/resources/MIR/FMP/C1/C1S3Waveform.html

How do we hear stuff?



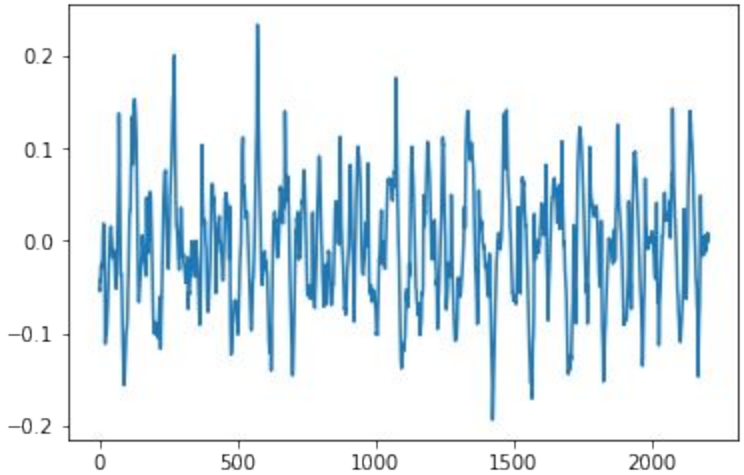
https://www.audiolabs-erlangen.de/resources/MIR/FMP/C1/C1S3_Waveform.html

Digital Audio



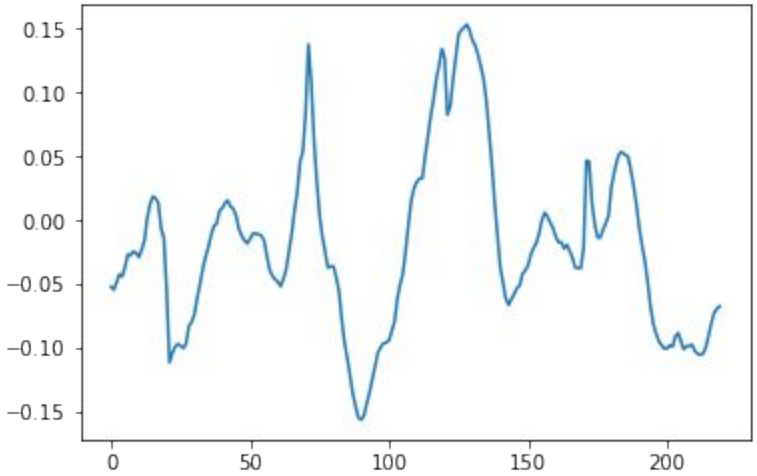
1 second at 22.05 kHz

Digital Audio



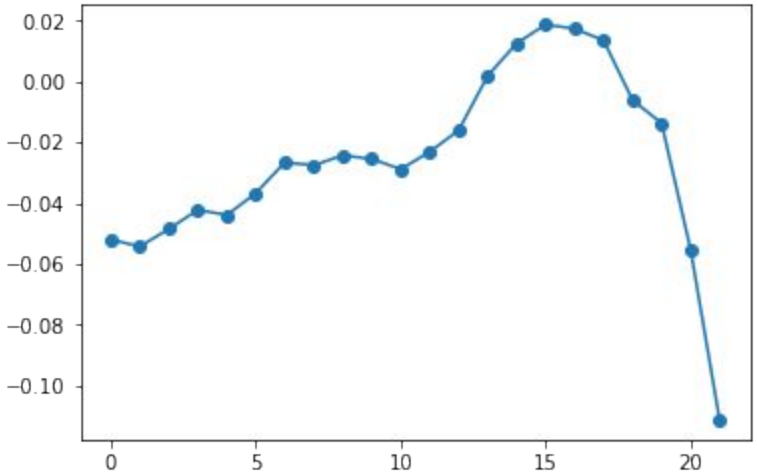
0.1 second at 22.05 kHz

Digital Audio



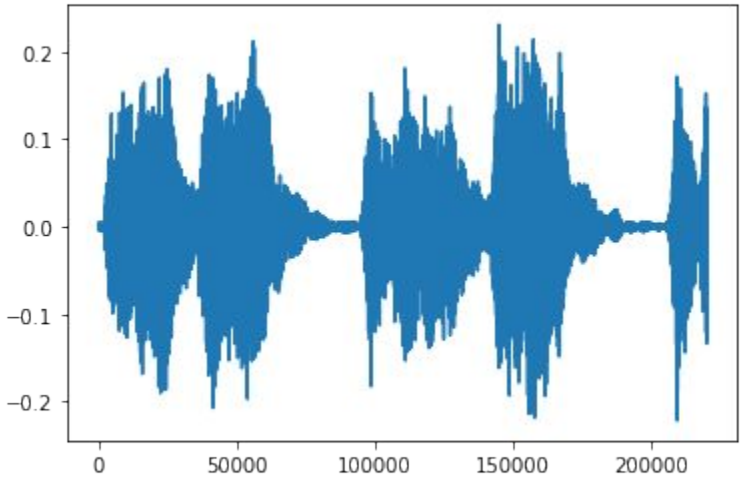
0.01 second at 22.05 kHz

Digital Audio



0.001 second at 22.05 kHz

So… what can we “retrieve” from this (time domain)?



(10 seconds)

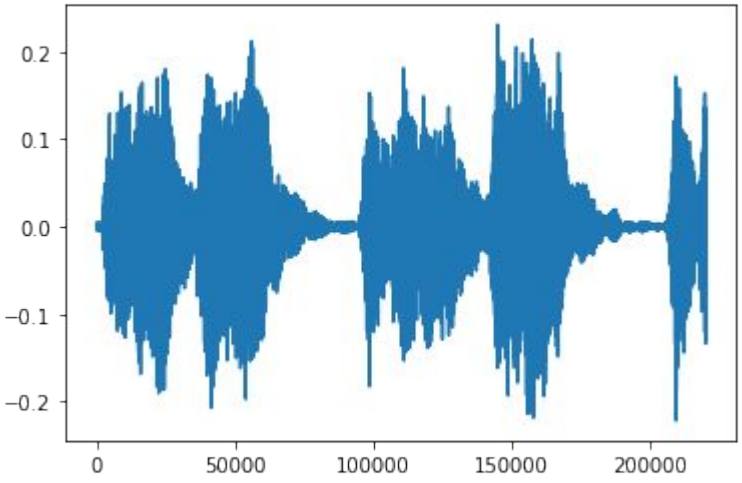
What to do…

Fourier Transform

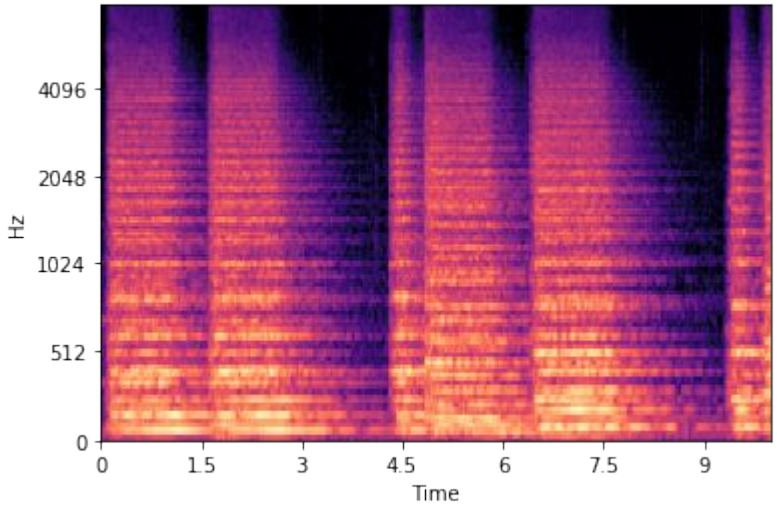
Time Duration		
Finite	Infinite	
Discrete FT (DFT)	Discrete Time FT (DTFT)	discr.
$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\omega_k n}$	$X(\omega) = \sum_{n=-\infty}^{+\infty} x(n)e^{-j\omega n}$	time
$k = 0, 1, \ldots, N-1$	$\omega \in (-\pi, +\pi)$	n
Fourier Series (FS)	Fourier Transform (FT)	cont.
$X(k) = \int_0^P x(t)e^{-j\omega_k t}dt$	$X(\omega) = \int_{-\infty}^{+\infty} x(t)e^{-j\omega t}dt$	time
$k = -\infty, \ldots, +\infty$	$\omega \in (-\infty, +\infty)$	t
discrete freq. k	continuous freq. ω	

https://ccrma.stanford.edu/~jos/sasp/Fourier_Transforms_Continuous_Discrete_Time_Frequency.html

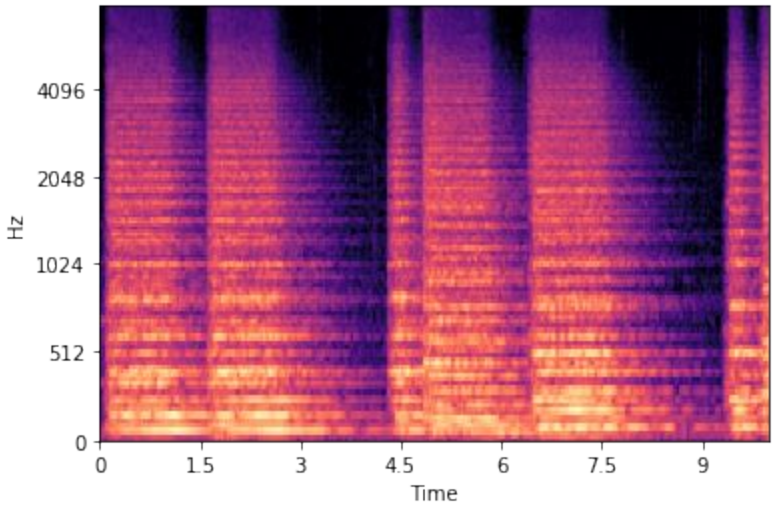
Fourier Transform



(10 seconds)

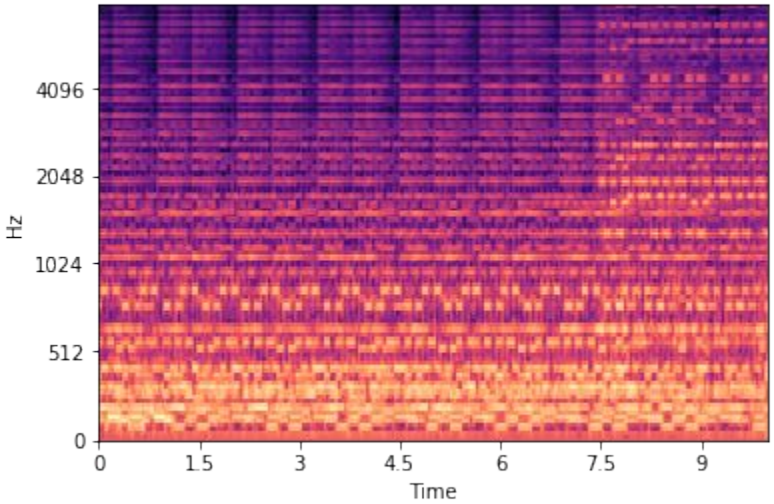


So… what can we “retrieve” from this (frequency domain)?



https://app.idagio.com/recordings/18966045

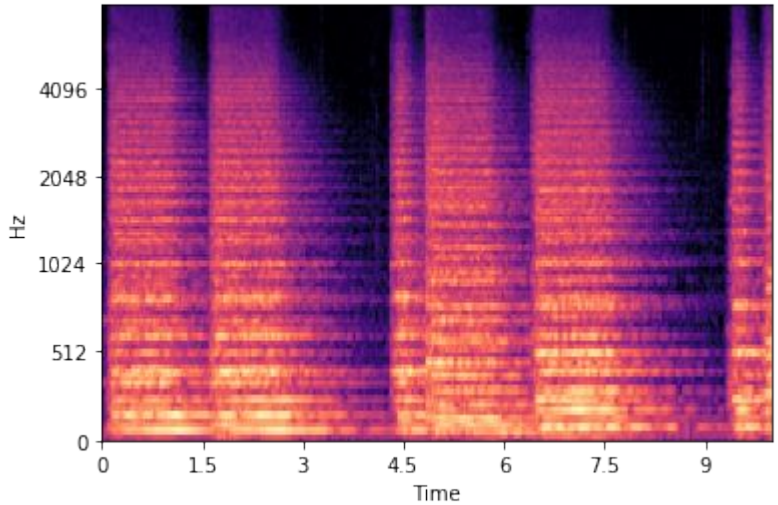
So… what can we “retrieve” from this (frequency domain)?



https://app.idagio.com/recordings/19104571

MIR Tasks - “Recapitulation”

- Audio fingerprinting
- Cover song identification
- Classification (genre, mood, instrument, vocalist, etc.)
- Drum transcription
- Downbeat detection / tempo estimation
- Fundamental frequency tracking
- Chord / key estimation
- Lyrics-to-audio alignment
- Score following
- Source separation
- .... (see MIREX)



Lab Session:
Genre Classification