6.300

October 29, 2024

Music Information Retrieval

- Representations of Music
- Representations of Music Signals
- Features of Music Signals
- Applications
- Signal Processing at MIT

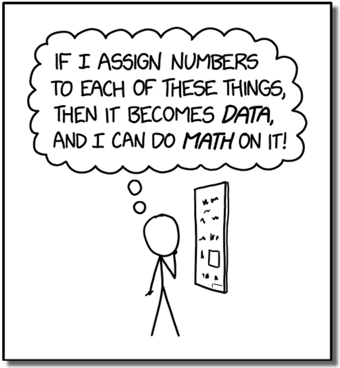
Signal processing techniques enable us to **extract features** from audio signals to understand **higher-level musical meaning**.

Signal Processing

$$x[n] \equiv x(nT)$$

$$X(e^{j\omega}) = \sum_n x[n]e^{-j\omega n}$$

$$x[n] \rightarrow$$



$$\rightarrow y[n]$$

$$X[k] = \sum_{n=0}^{N-1} x[n]W_N^{kn}$$

$$W_N^{kn} = e^{-j2\pi kn/N}$$

$$H(s) = \int_{-\infty}^{\infty} h(t)e^{-st}dt$$

$$y(t) = \sum_n y[n]p(t - nT)$$

$$y[n] = \sum_k h[k]x[n-k]$$

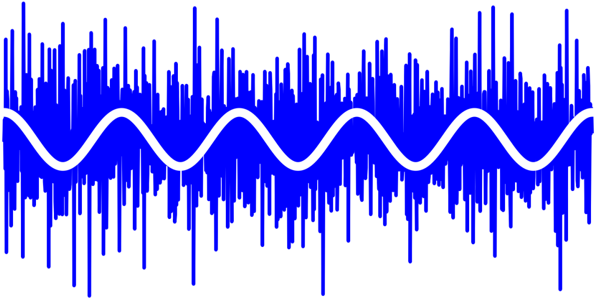
$$\frac{Y(z)}{X(z)} = \frac{b_0 + b_1z^{-1} + b_2z^{-2} + \cdots}{1 + a_1z^{-1} + a_2z^{-2} + \cdots}$$

modified from `xkcd.com`

Signal Processing

Signals convey **information**.

That information may not be in the most readily analyzed or manipulated form — hence the need for further **processing**.



Signal Processing

Many **applications** of signal processing:

- audio (speech, music) processing
- imaging and image processing
- radar signal processing
- communication systems
- data compression
- (and many more applications)

Signal processing is the art and science of
extracting information from signals
(or encoding information within signals)

Music Representations

More-or-less familiar representations of music:

- **audio recording** WAV, MP3,
- **musical score** sheet music
- **symbolic representation** piano roll, MIDI

Music and Signal Processing

Audio recordings of music common in the digital age

- samples of time-pressure waveform
- higher-level musical meaning not explicit

Two signal processing objectives:

- **analysis** extract information from signal
- **synthesis** generate signal with information

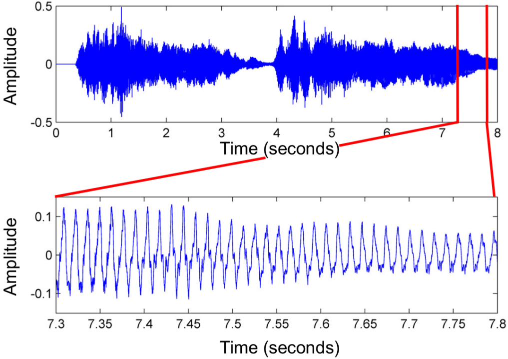
Today, focus on analysis of audio music signals
— **music information retrieval (MIR)**

MIR: Given an **audio recording** of music, extract higher-level **musical meaning**

Signal Representations

Representations of music signals we'll focus on:

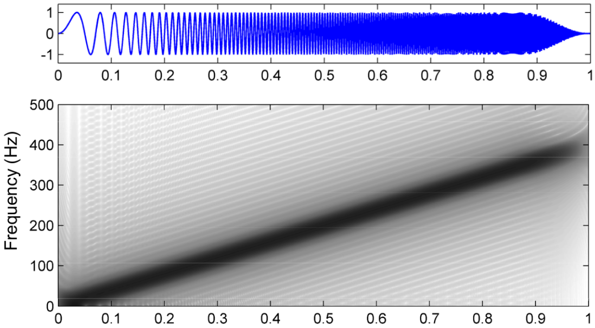
• **waveform**	1D	time-pressure signal
• spectrogram	2D	time-frequency plot
• chromagram	2D	time-pitch plot



Signal Representations

Representations of music signals we'll focus on:

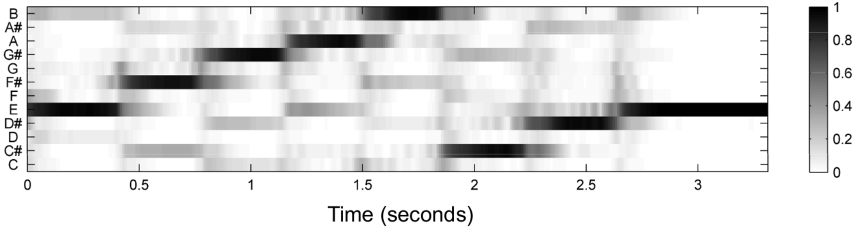
• waveform	1D	time-pressure signal
• **spectrogram**	2D	time-frequency plot
• chromagram	2D	time-pitch plot



Signal Representations

Representations of music signals we'll focus on:

• waveform	1D	time-pressure signal
• spectrogram	2D	time-frequency plot
• **chromagram**	2D	time-pitch plot



Sort of like a spectrogram / sheet music hybrid . . .

Signal Representations

Representations of music signals we'll focus on:

• **waveform**	1D	time-pressure signal
• **spectrogram**	2D	time-frequency plot
• **chromagram**	2D	time-pitch plot

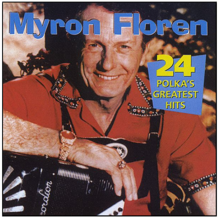
Each signal representation may emphasize some **musical features** and downplay others.

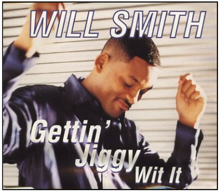
- Common thread in signal processing: Use the signal representation(s) best suited for the job!

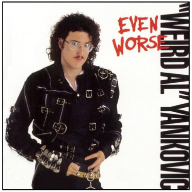
> Analyze music from several views: **waveform, spectrogram,** or **chromagram**

The Sound of Music

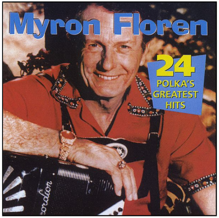
To start, we better understand music just a bit.
What is music, really?

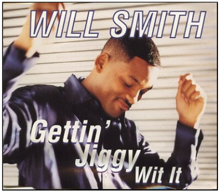



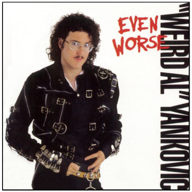




The Sound of Music







What constitutes music is **subjective**, but lots of music shares **common properties**.

Features and Applications

Some common features of music:

- pitch and harmony — middle C
- tempo, beat, and rhythm — tap along
- timbre and instrumentation — tone color
- polyphony — many voices

A few applications these features enable:

- beat tracking
- audio recognition
- music synchronization

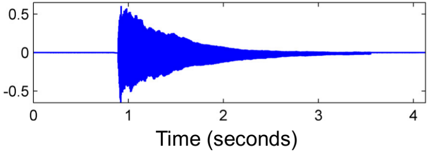
Extract musical features from music signals. Features enable further music analysis.

Beats

Start with analyzing waveforms for musical features.

Beats correspond to sudden, large changes in air pressure — which waveforms represent.

- represent "unit time" for music signals
- may enhance downstream processing tasks

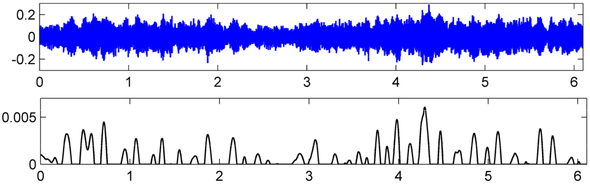


Beats: Abrupt **changes in air pressure**.

Beats

Generate an **energy novelty curve**.

- half-wave rectification
- finite differences
- thresholding and peak-finding

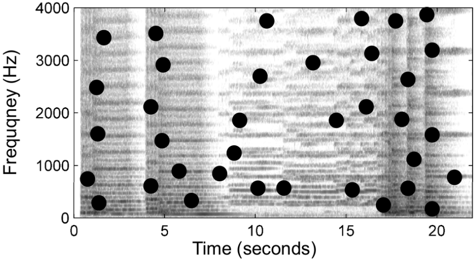


Beats correspond to peaks of an **energy novelty curve**.

Audio Recognition

Need a robust method to **identify music** given (potentially very noisy) audio recordings . . .

- compute spectrogram and determine **peaks**
- create **constellation map**
- use efficient **hashing** scheme
- **compare** with constellation maps in database



Audio Recognition

Shazam existed before smartphones did!

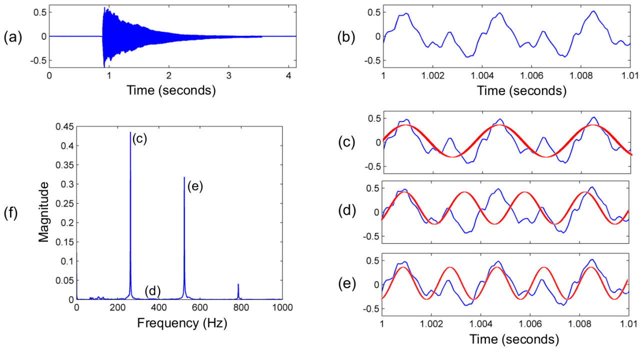
Like today, you held your phone up to the music:

(Well, you had to dial in . . .)

Periodicity and Pitch

Wind and string musical instruments produce **quasi-periodic waveforms**.

- sum of harmonically-related sinusoids
- fundamental frequency determines **pitch**



Periodicity and Pitch

Octave equivalence: Frequencies that are half or double the fundamental are perceived as the same pitch as the fundamental!

C	16.35 Hz	2×8.18 Hz
C	32.70 Hz	2×16.35 Hz
C	65.41 Hz	2×32.70 Hz
C	130.81 Hz	2×65.41 Hz
C	261.63 Hz	2×130.81 Hz
C	523.25 Hz	2×261.63 Hz

Periodicity and Pitch

Equal-tempered scale: Divide each octave into 12 equally-spaced pitch classes.

- next pitch is $2^{\frac{1}{12}}$ times higher in frequency
- refer to octave degree as pitch's **chroma**
- must also specify octave

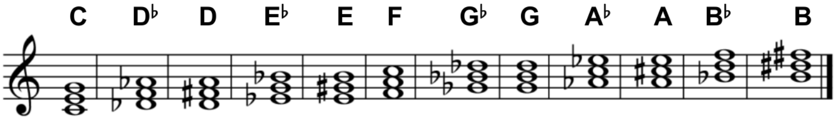
C4 (Middle C)	261.63 Hz
E4	329.63 Hz
G4	392.00 Hz
A4 (Concert A)	440.00 Hz
C5	523.25 Hz

Pitch and Harmony

Play several pitches at once to produce **harmony**.

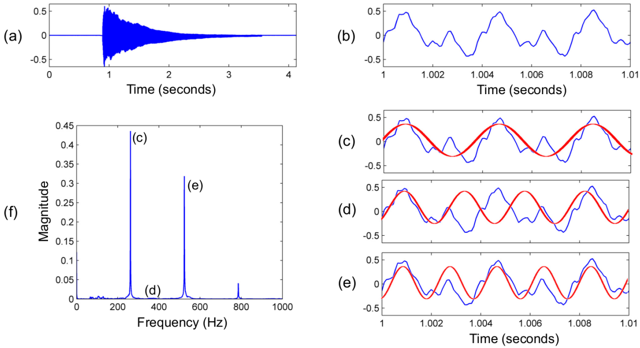
A harmony with three or more pitches is a **chord**.

- Chords that "sound pleasant" (consonant) involve pitches with simple frequency ratios, indicating **many shared harmonics**.



Combining pitches with **many shared harmonics** produces **consonant harmony**.

Frequency and Pitch



The **fundamental frequency** of a waveform determines the **pitch** one perceives. A **note** is specified by **chroma** and **octave**.

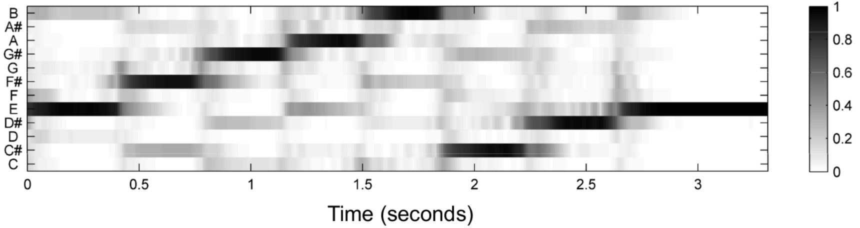
Frequency/Pitch vs. Time

Spectrograms display the magnitude-squablue (power) frequency spectrum of a signal over time.

- pitch frequencies increase exponentially

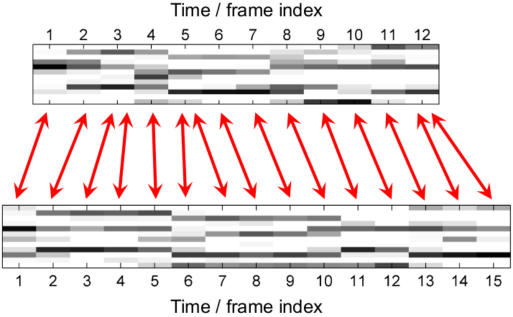
Chromagrams: Similar to spectrograms, but display the pitch (chroma) of a signal over time.

- condense spectrum into a single octave



Temporal Alignment

Temporal alignment: Synchronize two variations of the same underlying song.



Match up sections in "cover songs" to the original.

Temporal Alignment

Temporal alignment: Synchronize two variations of the same underlying song.

- Compare similarity between time-frames of each song (columns of each chromagram) to produce a **cost matrix**.

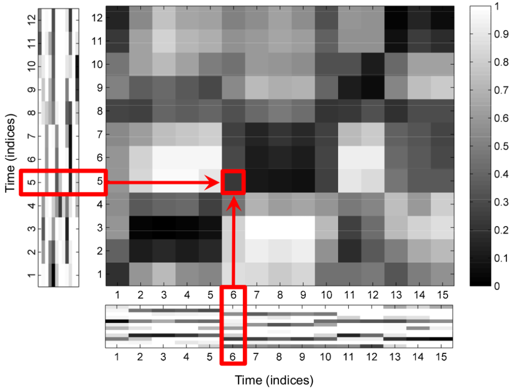
- Determine **minimum-cost path** through cost matrix, from start to end of song.

- Use minimum-cost path to determine **correspondence between songs**.

Sifting out features via **measuring similarity** . . .

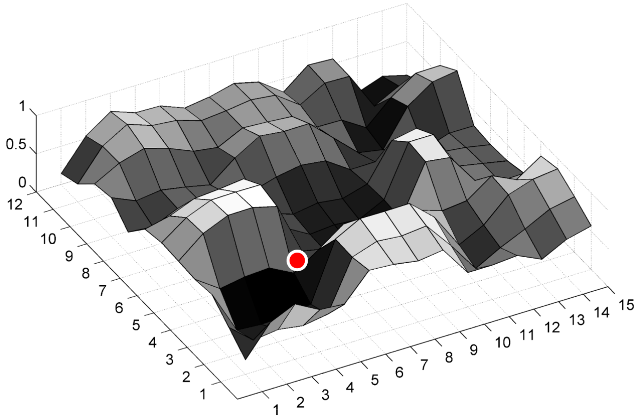
Temporal Alignment

Compare similarity between time-frames of each song (columns of each chromagram) to produce a **cost matrix**.



Temporal Alignment

"Walk through the valley" of the cost matrix!



...and that's not all!

Many more methods and applications:

- audio decomposition (source separation)
- audio thumbnailing
- chord recognition
- instrument recognition
- onset detection
- pitch estimation (monophony, polyphony)
- structure analysis (ABAB, ABBA, ...)
- tempo estimation

Interested in music signal processing?
6.300: Signal Processing (pre-req for ↓)
6.302: Fundamentals of Music Processing

Further Reading

Müller, Meinard. *Fundamentals of Music Processing: Using Python and Jupyter Notebooks*. Springer, 2021.

- available to MIT affiliates online for free
- official course textbook of 6.302
- FMP Notebooks provide examples in Python

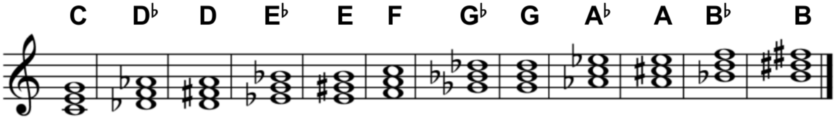
Textbook

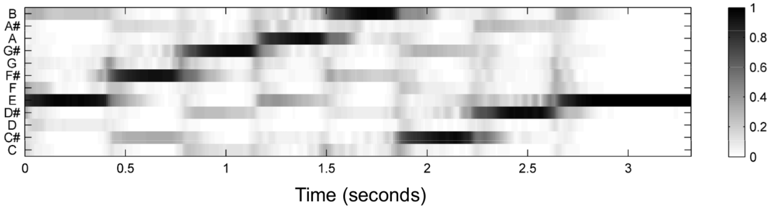
`link.springer.com/book/10.1007/978-3-030-69808-9`

FMP Notebooks

`audiolabs-erlangen.de/resources/MIR/FMP/C0/C0.html`

Music Information Retrieval





Signal processing techniques enable us to **extract features** from audio signals to understand **higher-level musical meaning**.

Signal Processing at MIT

6.300	Signal Processing (D. Freeman)	FA/SP
6.301	Signals, Systems, and Inference (Zheng)	SP
6.302	Fundamentals of Music Processing (Egozy)	FA
6.310	Dynamical Systems and Control (White)	FA/SP
6.700	Discrete-Time Signal Processing (Ward)	FA
6.701	Digital Image Processing (Rachlin, Lim)	SP
6.741	Digital Communication (Chan, Médard)	FA
6.862	Spoken Language Processing (Glass)	SP
6.880	Biomedical Signal and Image Processing	SP
6.C27	Computational Imaging (Barbastathis, You)	FA
ES.S31	ESG Special Seminar (Roesler)	FA/IAP

Al Oppenheim: "There will always be **signals**, and they will always need to be **processed**."