ECS7006 Music Informatics

Week 1 - Music Representations

School of Electronic Engineering and Computer Science
Queen Mary University of London

prepared by Emmanouil Benetos
adapted from material by Meinard Müller and Simon Dixon

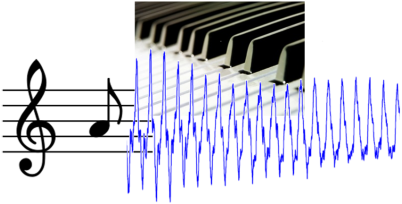
emmanouil.benetos@qmul.ac.uk

2023

Motivation and Aims

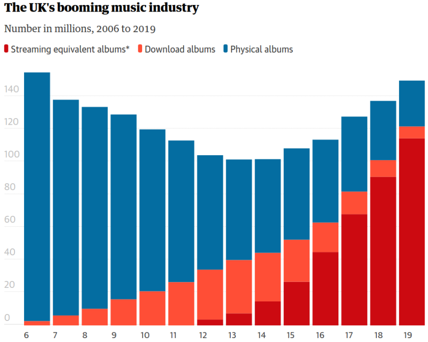
Setting the Stage

- Digital revolution in music distribution and storage
- Computers are involved in every aspect of music:
  - Education
  - Composition
  - Performance
  - Production
  - Consumption



State of the music industry

- Since 2015, digital exceeds "physical" revenue
- Hundreds of music streaming and downloading services
- Tens of millions of tracks available
- UK creative industries (a large part of which is music) generate similar income to financial services sector



Source: The Guardian 1/1/2020

Music Informatics / Music Information Retrieval (MIR)

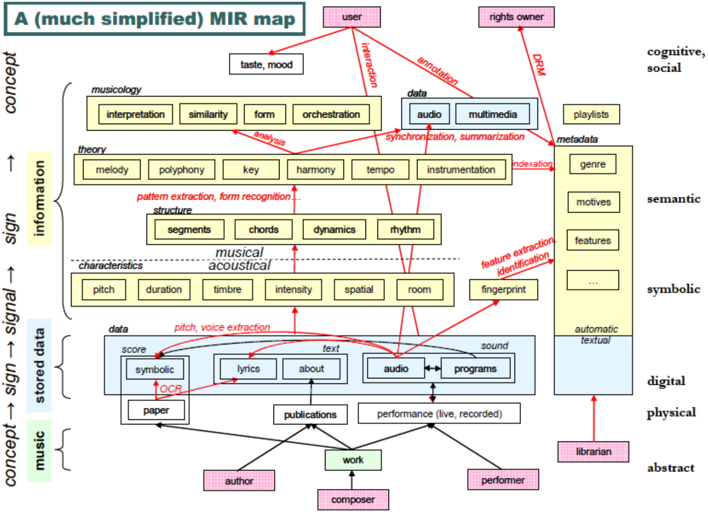
The field of research concerned with extending the understanding and usefulness of music data.

- Young discipline, compared e.g. with speech processing or computer vision
- International Society for Music Information Retrieval (ISMIR): formed in 2009

ISMIR

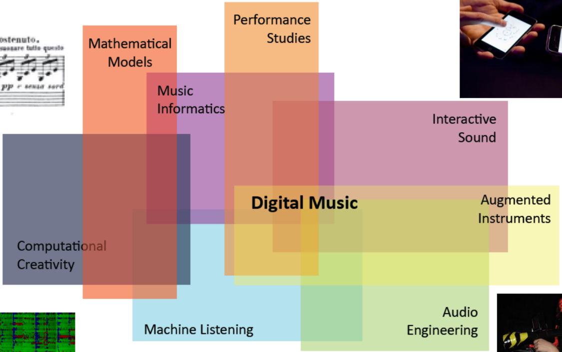
http://ismir.net/

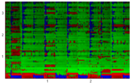
Music Informatics / Music Information Retrieval (MIR)



(source: Michael Fingerhut)

Centre for Digital Music





centre for digital music

http://c4dm.eecs.qmul.ac.uk/

Aims of the Module

To introduce students to state-of-the-art methods for the analysis of music data, with a focus on music audio.

- Objectives:
  - Music representations as analysis tools
  - Specific issues involving the analysis of music data
  - Music analysis techniques for rhythm, melody, harmony, and timbre
  - Use of presented techniques in real-world applications
  - Further topics: sound source separation, multimodal music tasks
- What this module is NOT about:
  - How to USE existing technologies / software
  - Music production
  - Music perception and cognition
  - Musical acoustics / psychoacoustics

Module Organisation

Module Organisation

- Lecturers: Simon Dixon, Emmanouil Benetos
- Demonstrators: Lele Liu, Saurjya Sarkar
- Lectures: Fridays 11:00-12:00 and 14:00-16:00 in Eng216 and MB204 respectively
  - Approximately one lecture topic per week
  - Lectures will be recorded
- Labs: Thursdays 10:00-12:00 (weeks 2,4,6,9,10) in the ITL 2nd floor
- Communication via lectures, student forum, and email
- Assessment:
  - 40%: Assignments (2 $\times$ 20% due weeks 7 and 12)
  - 60%: Written examination

Lecture Topic Outline

- Weeks 1-2: Representations
- Weeks 3-4: Rhythm
- Weeks 5-6: Pitch, melody, and harmony
- Week 7: Reading week
- Week 8: Structure and source separation & Part 1 revision
- Weeks 9-10: Content-based audio retrieval
- Week 11: (Bank holiday)
- Week 12: Multimodal music informatics & Part 2 revision

Resources: Reference Books

- *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*, M. Müller (Springer, 2015)
- *An Introduction to Audio Content Analysis: Applications in Signal Processing and Music Informatics*, A. Lerch (Wiley, 2012)
- *Music Data Analysis: Foundations and Applications*, ed. C. Weihs et al. (CRC, 2016)
- *DAFX - Digital Audio Effects*, ed. U. Zölzer (Wiley, 2002)
- *Spectral Audio Signal Processing*, J. O. Smith (2010)
  Available at `http://ccrma.stanford.edu/~jos/sasp/`
- *Music Information Retrieval: Recent Developments and Applications*, M. Schedl, E. Gómez, J. Urbano (Now Publishers, 2014)

Other Resources

- Conference proceedings:
  - ISMIR (http://www.ismir.net/conferences/)
  - ICASSP (http://www.ieeexplore.ieee.org)
  - SMC (http://www.smcnetwork.org)
  - ICMC (http://www.icma.org)
  - DAFx (http://www.dafx.de)
- Journals:
  - IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)
  - Transactions of the International Society for Music Information Retrieval (TISMIR)
  - Journal of New Music Research (JNMR)
  - Computer Music Journal (CMJ)
  - Journal of the Acoustical Society of America (JASA)

Assumed Knowledge

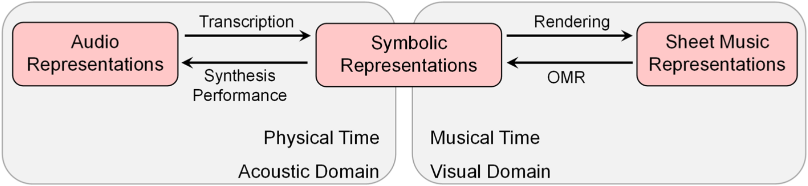
- Basic DSP: linear time/shift-invariant (LTI/LSI) systems
  - superposition, impulse response, convolution, transfer functions, frequency response, causality, stability, FIR and IIR filters
- Digital representation of audio signals
  - time domain: sampling theorem, Nyquist frequency, aliasing, quantisation
  - frequency (spectral) domain: DFT, zero-padding, window functions, spectrogram
- Basics of music theory and music acoustics: desirable but not essential
  - pitch, semitones, octaves, harmonics, instrument mechanics, harmony, rhythm, timbre
- Python programming

Music Representations

This week's content

Music Representations

- Sheet music representations
- Symbolic representations
- Audio representations



Definitions of Music

- “Organised sound”
- “The science or art of ordering tones or sounds in succession, in combination, and in temporal relationships to produce a composition having unity and continuity.”
- “The art of arranging sounds in time so as to produce a continuous, unified, and evocative composition, as through melody, harmony, rhythm, and timbre.”

Music Representations

Sheet music representations

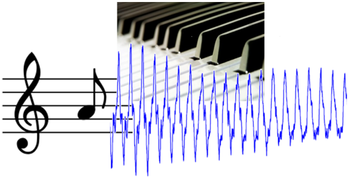
Visual representations of a score in printed form or in digitised images.

Symbolic music representations

Machine-readable data formats that represent musical entities.

Audio representations

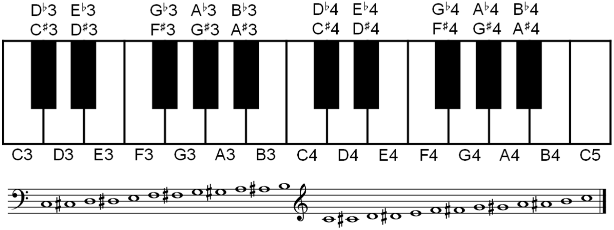
Representations of acoustic sound waves.



- Sheet music describes a musical work using a formal language based on musical symbols and letters, depicted in a graphical-textual form.
- A guide for performing a piece of music (“prescriptive notation”)

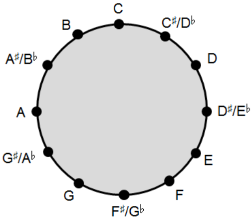
Musical notes and pitches

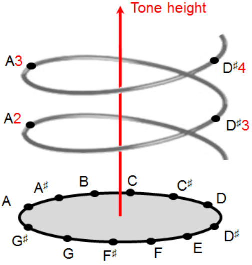
- Note: A musical symbol (in the context of scores), or a pitched sound (in the context of audio).
- Pitch: A perceptual property that allows a listener to order sounds on a frequency-related scale.
- Discretising pitches leads to the notion of a musical scale



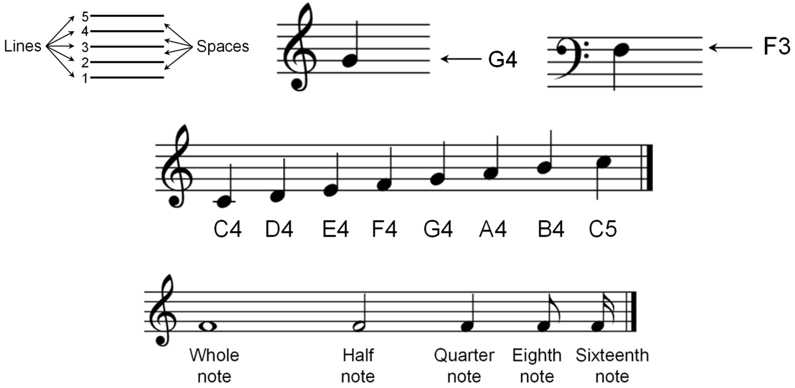
Musical notes and pitches

- Playing a note results in a periodic sound of a certain fundamental frequency.
- Octave: the interval between one note and another with half or double its fundamental frequency.
- Twelve-tone equal tempered scale: one octave is subdivided into 12 equally spaced scale steps.
- Chromatic scale: ordering all notes in the equal tempered scale according to their pitches.





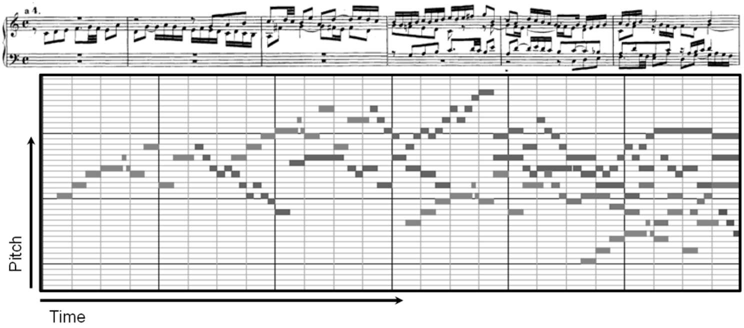
Western music notation



Western music notation

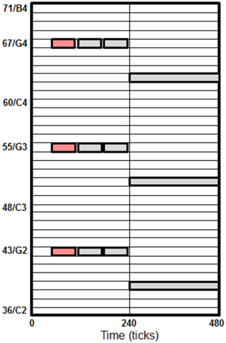
Symbolic representations

- A piano roll is a continuous roll of paper with perforations punched into it.
- A piano-roll representation is a two-dimensional representation of pitch over time.



- MIDI: Musical Instrument Digital Interface
- Industry standard for digital electronic musical instruments, adopted as a symbolic music format
- A MIDI file contains a list of MIDI messages
- MIDI messages: note-on, note-off. Message parameters: note number, velocity, channel.

Time (Ticks)	Message	Channel	Note Number	Velocity
60	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0



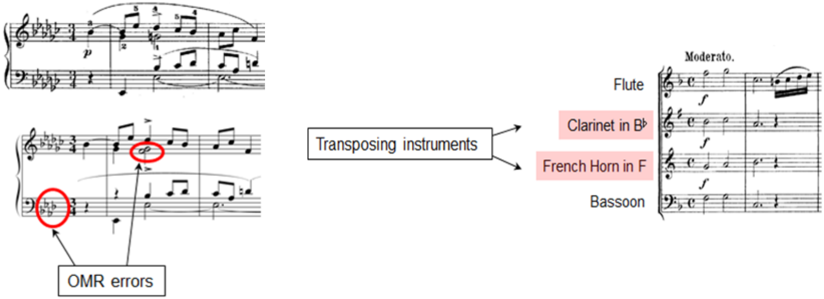
Symbolic representations

- Score representation: machine-readable symbolic representation yielding explicit information on musical symbols (e.g. staff system, clefs, time signatures, notes, rests, accidentals etc).
- Common symbolic score formats: musicXML, MEI, Lilypond, ABC...

```
<note>
  <pitch>
    <step>E</step>
    <alter>-1</alter>
    <octave>4</octave>
  </pitch>
  <duration>2</duration>
  <type>half</type>
</note>
```

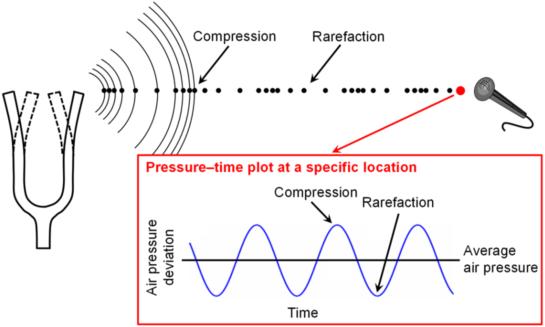
Optical music recognition

- Optical music recognition: the process of converting digital images of sheet music into symbolic music representations.
- Musical equivalent to optical character recognition (OCR).
- Pixels have to be grouped and interpreted as musical symbols.
- Currently open problem; common errors: missing/extra notes, incorrect key signature, transposing instruments...



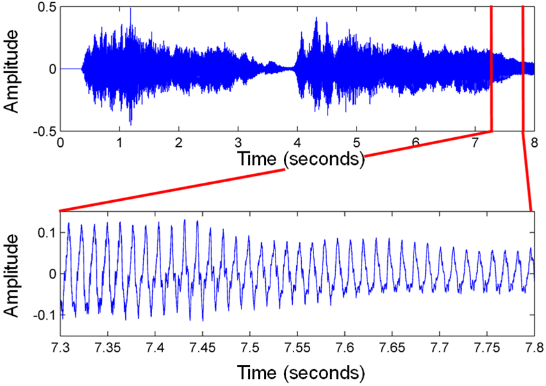
Audio representations

- Performing music results in sounds or acoustic waves, which are transmitted through the air as pressure oscillations.
- The term audio refers to the transmission, reception or reproduction of sounds that lie within the limits of human hearing.



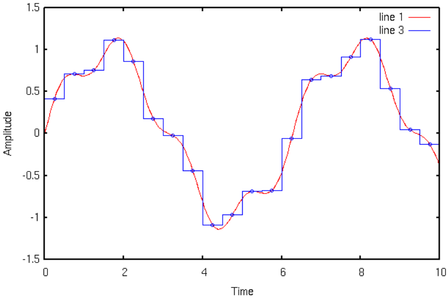
Waveform

- The change in air pressure at a certain location can be represented by a pressure-time plot, also referred to as the waveform of the sound.



Digitisation and Sampling

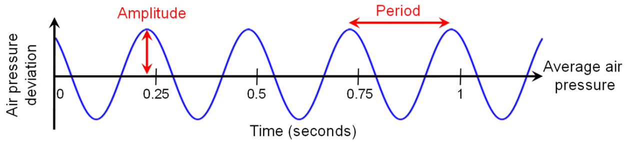
- A sound is represented as a digital signal by sampling the pressure at a point in space at regular intervals in time



- Digitisation involves quantisation in time and amplitude, resulting in two main types of reconstruction errors

Frequency and pitch

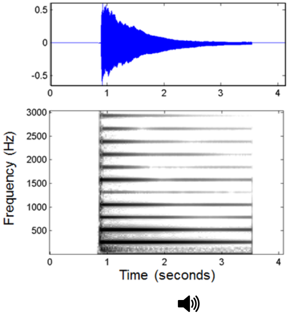
- If the points of high and low air pressure repeat in a regular fashion, the resulting waveform is called periodic.

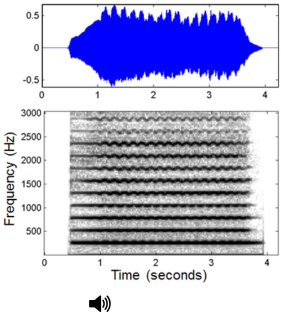


- The period of the wave is defined as the time required to complete a cycle.
- The frequency (measured in Hz) is the reciprocal of the period.

Frequency and pitch

- The notion of frequency is what determines the pitch of a sound.
- A musical tone can be described as a superposition of sinusoids, also called partials.
- The frequency of the lowest partial is called the fundamental frequency (f0) of the sound.
- A harmonic is a partial that is an integer multiple of a fundamental frequency.





Frequency and pitch

- The human perception of pitch is logarithmic in nature.
- Associating MIDI note numbers ($p \in [0:127]$) to frequencies:

$$F_{pitch}(p) = 2^{(p-69)/12} \cdot 440$$

- The frequency ratio of two subsequent pitches is constant:

$$F_{pitch}(p+1)/F_{pitch}(p) = 2^{1/12} \approx 1.059463$$

  This difference between two subsequent scale steps is also called a semitone.
- An octave is divided into 1200 cents, so that each semitone corresponds to 100 cents.

Frequency and pitch

- The term inharmonicity is used to denote the measure of the deviation of a partial from the closest ideal harmonic.
- Most instruments have very low inharmonicity; other instruments such as the marimba, vibraphone, and bells contain nonharmonic partials - yet still give a good sense of pitch.
- Nonpitched instruments such as cymbals or gongs make sounds rich in inharmonic partials.

Dynamics, intensity and loudness

- Dynamics refer to the volume of a sound as well as to the musical symbols that indicate the volume.
- Dynamics correlate with a perceptual property called loudness, by which sounds can be ordered on a scale from quiet to loud.
- Loudness also correlates with the objective measures of sound intensity and sound power. Sound intensity denotes the sound power per unit area.
- Threshold of hearing: the minimum sound intensity of a pure tone that a human can hear: $I_{TOH} := 10^{-12}$W/m^2

Dynamics, intensity and loudness

- Power and intensity are expressed in the decibel (dB) scale, which is a logarithmic unit expressing the ratio between two values.
- Intensity measured in dB:

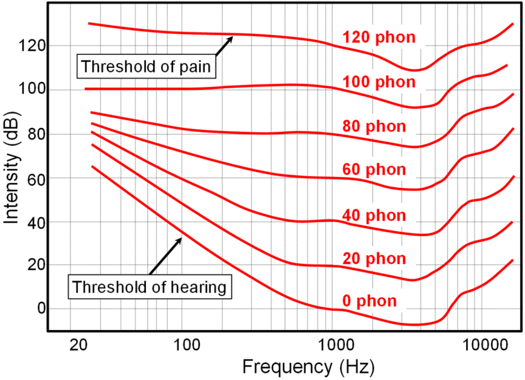
$$dB(I) := 10 \cdot \log_{10}\left(\frac{I}{I_{TOH}}\right)$$

Source	Intensity	Intensity level	× TOH
Threshold of hearing (TOH)	10^{-12}	0 dB	1
Whisper	10^{-10}	20 dB	10^{2}
Pianissimo	10^{-8}	40 dB	10^{4}
Normal conversation	10^{-6}	60 dB	10^{6}
Fortissimo	10^{-2}	100 dB	10^{10}
Threshold of pain	10	130 dB	10^{13}
Jet take-off	10^{2}	140 dB	10^{14}
Instant perforation of eardrum	10^{4}	160 dB	10^{16}

Table: intensity levels.

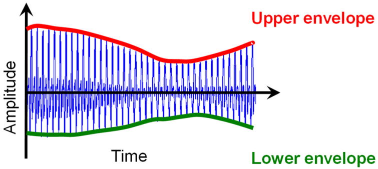
Dynamics, intensity and loudness

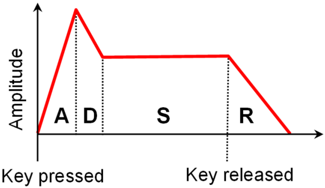
- Loudness is affected by several factors (e.g. age, sound duration).
- Two sounds with the same intensity but different frequencies are not perceived to have the same loudness.
- The perceived loudness of pure tones depending on the frequency is expressed by the unit phon.



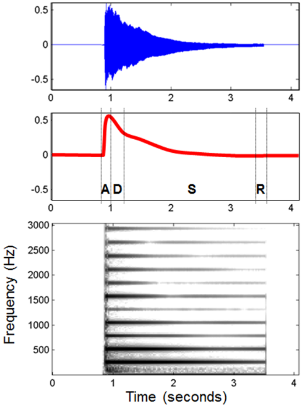
Timbre

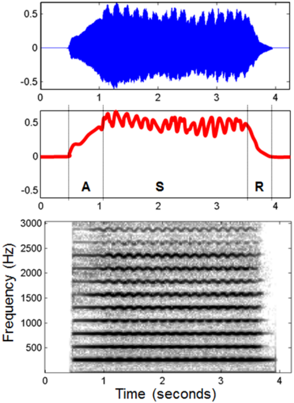
- Besides pitch, loudness, and duration, there is another fundamental aspect of sound called timbre or tone colour.
- Timbre is the attribute whereby a listener can judge two sounds as dissimilar using any criterion other than pitch, loudness, and duration.
- Researchers have tried to link timbre with more objective sound characteristics.





- Tremolo: periodic modulations in amplitude (i.e. amplitude modulation).
- Vibrato: regular, pulsating change of frequency (i.e. frequency modulation).





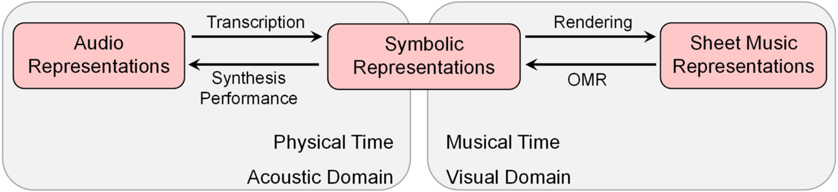
Audio Formats, Memory and Bandwidth

- Most audio file formats (e.g. WAV, AIFF) consist of a simple header specifying the encoding type and parameters, followed by the encoded data
- Although requirements for high fidelity audio storage and transmission are minimal compared to video, they have (had) a large impact on digital audio technologies
- e.g. "CD quality" audio:
  - Sampling frequency f_s: 44100 Hz
  - Channels: 2 (stereo)
  - Word size (bits per sample): 16 bits (2 bytes)
  - Memory = f_s $\times$ channels $\times$ wordsize $\times$ duration
    e.g. a 3-minute song: 44100 $\times 2 \times 2 \times$ 180 = 31.8MB
  - Bandwidth = 44100 $\times 2 \times$ 16 = 1.4Mbit/s

Audio Data Compression

- Compression is used to reduce bandwidth requirements (e.g. for streaming, storage)
- Not to be confused with dynamic range (level) compression
- Standard lossless data compression algorithms (e.g. ZIP) compress audio to around 87% of its original size
- Lossless audio compression algorithms (e.g. FLAC) reduce file sizes to around 50–60%
- Lossy audio compression algorithms (e.g. MP3, Ogg Vorbis) take advantage of psychoacoustic models to reduce data sizes to under 10% with little loss in audio quality

Recap


Audio Representations
Transcription
Synthesis
Performance
Symbolic Representations
Rendering
OMR
Sheet Music Representations
Physical Time
Acoustic Domain
Musical Time
Visual Domain

Python Music Informatics Fundamentals

Useful Resources

FMP Notebooks:
`https://www.audiolabs-erlangen.de/FMP`

Notes on Music Information Retrieval:
`https://musicinformationretrieval.com/`

LibROSA - audio and music processing in Python
`https://librosa.github.io/`

pretty_midi - MIDI file handling in Python
`https://github.com/craffel/pretty-midi`

Audio Objects

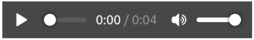
Integrating audio objects into a Jupyter notebook:

- HTML <audio> tag

```
<audio src="../note.wav" type="audio/mpeg"
controls="controls"></audio>
```

- Using IPython.display.Audio

```
import os
import IPython.display as ipd
path_filename = os.path.join('..','note.wav')
audio_element = ipd.Audio(filename=path_filename)
ipd.display(audio_element)
```



Plotting a waveform

- Prerequisites:

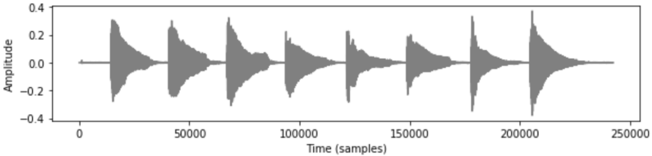
```
from matplotlib import pyplot as plt
import librosa
```

- Loading audio using librosa:

```
x, sr = librosa.load(os.path.join('..', 'scale.wav'))
```

- Plotting the waveform of the signal:

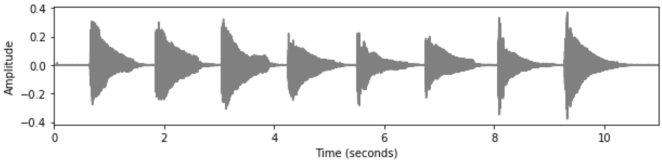
```
plt.figure(figsize=(10, 2))
plt.plot(x, color='gray')
plt.xlabel('Time (samples)')
plt.ylabel('Amplitude')
```



Plotting a waveform

- Converting sample indices to physical time positions:

```
import numpy as np
t = np.arange(x.shape[0]) / sr
plt.plot(t, x, color='gray')
plt.xlim([t[0], t[-1]])
```



- Or you can use librosa:

```
import librosa.display
librosa.display.waveshow(x, sr=sr, color='gray')
```

Audio read/write

- Librosa:

```
x, sr = librosa.load(path)
```

- PySoundFile:

```
import soundfile as sf
x, sr = sf.read(path)
sf.write(path, x, Fs)
```

- SciPy:

```
sr, x = wavfile.read(path)
```

Processing MIDI files

PrettyMIDI: python package for handling MIDI data

- Load MIDI file:
```
import pretty_midi
pm = pretty_midi.PrettyMIDI('motif.mid')
```
- Time signature changes: `pm.time_signature_changes`
- Instruments: `pm.instruments`
- Notes per instrument `pm.instruments[i].notes`
- The PrettyMIDI class also includes pitch bends, control changes...

Plotting MIDI files

Plotting MIDI as a piano-roll:

```
def plot_piano_roll(pm, pitch1, pitch2, fs=100):
# Using librosa's specshow function
librosa.display.specshow(pm.get_piano_roll(fs)[pitch1:pitch2],
  hop_length=1, sr=fs, x_axis='time', y_axis='cqt_note',
  fmin=pretty_midi.note_number_to_hz(pitch1))
plot_piano_roll(pm, 24, 84)
```