Lab experience 3

Machine learning in acoustics

Audio content analysis

- Goal: analyse an audio signal to extract information on musical content
- Highly interdisciplinary: physics, psycho-acoustics, musicology, digital signal processing, machine learning...
- Active area of research!
  - song recommendation Spotify last.fm
  - fingerprinting SHAZAM gracenote A NIELSEN COMPANY
  - score following smartmusic.
  - pitch detection & correction melodyne

The sound

Physics

- **Vibration** that propagates as an **acoustic wave** in a **medium** (typically air)
- For gaseous and liquid media, the sound propagates through **longitudinal waves**, i.e. is associated to **variations of pressure**
- In a gaseous the sound speed is given by

$$c = \sqrt{\left(\frac{\partial p}{\partial \rho}\right)_s} = \sqrt{\frac{\gamma p}{\rho}}$$

- In air, in particular, we find $c = 20.05\sqrt{T}$ m/s $= 243.3$ m/s at $T = 20°$ C.

The sound

Perception

- All our senses are based on **logarithmic scales**
- This applies to sound in two ways:
    - We have a logarithmic perception of sound **frequences**
    - We have a logarithmic perception of sound **pressure** (or, equivalently, to its power-related quantity, the **sound intensity**)
- These facts have a huge impact on the production itself of the sound and thus on music

Basic properties of a single “note”

- **Duration**: the length (in seconds) a note is played
- **Intensity**: a measure of the power of the sound wave or, equivalently, of its squared amplitude
- **Pitch**: the main frequency we perceive, often called the “fundamental” note
- **Timbre**: a complex combination of “properties” of the note played (we could say: everything not included in the previous items!)

The pitch

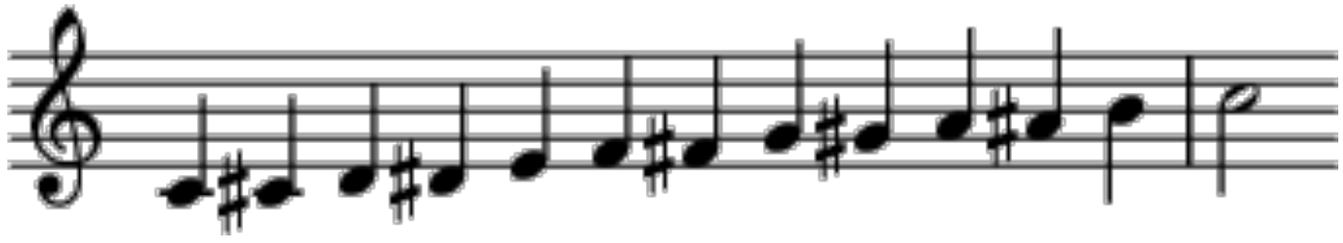
Physics of the musical scales

- Sounds that have fixed **frequency ratios** will be perceived as **equidistant in pitch**
    - Example: two notes that differ by an **octave** will have a frequency ratio of **2**, independently on the notes!
- **“Simple” ratios** of frequency produce musically **pleasant notes**: this can be traced to the periodic nature sound
    - For example, two notes whose frequencies are close to a 3:2 ratio, to a 4:3 ratio, or to a 5:4 ratio will produce “consonances”
- These two facts, together, essentially define the way a **musical scale** works!

The pitch

Physics of the musical scales

- Consider all notes in an octave: there are 12 of them



- An octave is associated to a doubling in frequency
- Then all note are equally spaced in pitch if their frequencies follow a simple exponential law, $f \propto 2^{p/12}$ (equal temperament)

C#/D♭	D	D#/E♭	E	F	F#/G♭	G	G#/A♭	A	A#/B♭	B	C
1.059	1.122	1.189	1.260	1.335	1.414	1.498	1.587	1.682	1.782	1.888	2.000
16/15	**9/8**	**6/5**	**5/4**	**4/3**	**10/7**	**3/2**	**8/5**	**5/3**	**16/9**	**15/8**	**2**

The pitch

Physics of the musical scales

- Modern tuning fixes the frequency of the central A around **440 Hz** and uses a **equal temperament**
- We can thus easily compute the frequency associated to each note...
- ...or, given a frequency, find the (approximate) note it corresponds
- **Warning**: different cultures have produced different scales in different times. However, the 12-tone temperament described here is by far the most commonly used

Sound production

Musical instruments

- In most musical instruments sound can be produced using two main means:
    - **chordophones**: by inducing vibrations on a string (piano, violin, guitar...)
    - **aerophones**: by inducing vibrations on an air column (flute, oboe, trumpet, organ...)
- For both these wide classes of instruments, the sound produced will be characterised by a **main frequency** (fundamental) together with a set of **overtones**

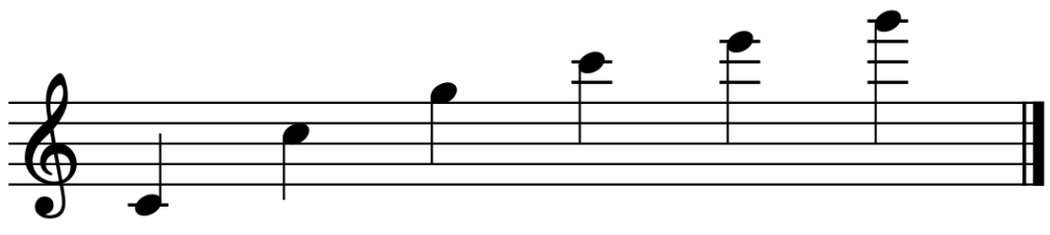
The timbre

- Each instrument produces a specific timbre, easily recognisable for a trained person
- The timbre depends on a number of factors
  - The **attack**, i.e. the way a note “sounds” in the first instants: this is strongly associated to the instrument family
  - The **overtones** produced by the instrument
  - To a lesser extent, to the **decay**, i.e. the way the sound disappears

The overtones

Relationships with the musical scale

- A (linear) instrument usually produces overtones for each note
- Overtones are frequencies that are integer multiples of the fundamental frequency
- The second overtone has frequency $2f$, and thus it is an octave above the fundamental note
- The third overtone has frequency $3f$, and thus it is a fifth above the first overtone
- Following overtones are a second octave, an upper third, fifth...



[see Wikipedia]

The overtones

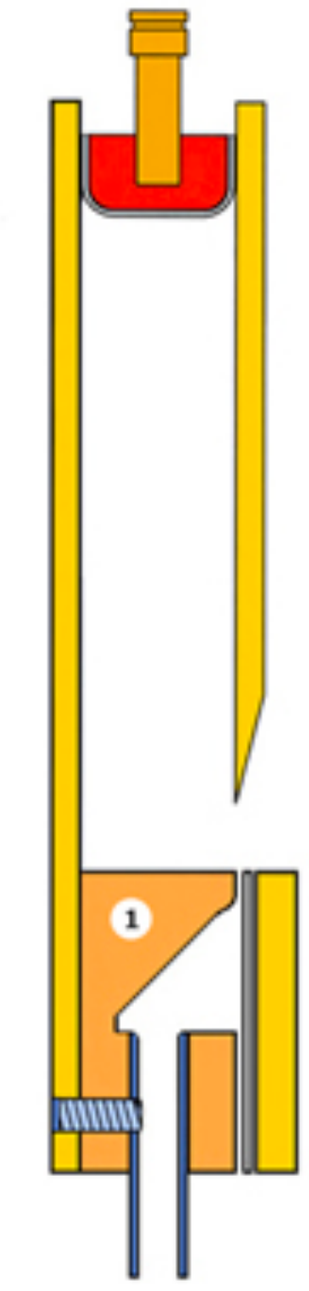
Production

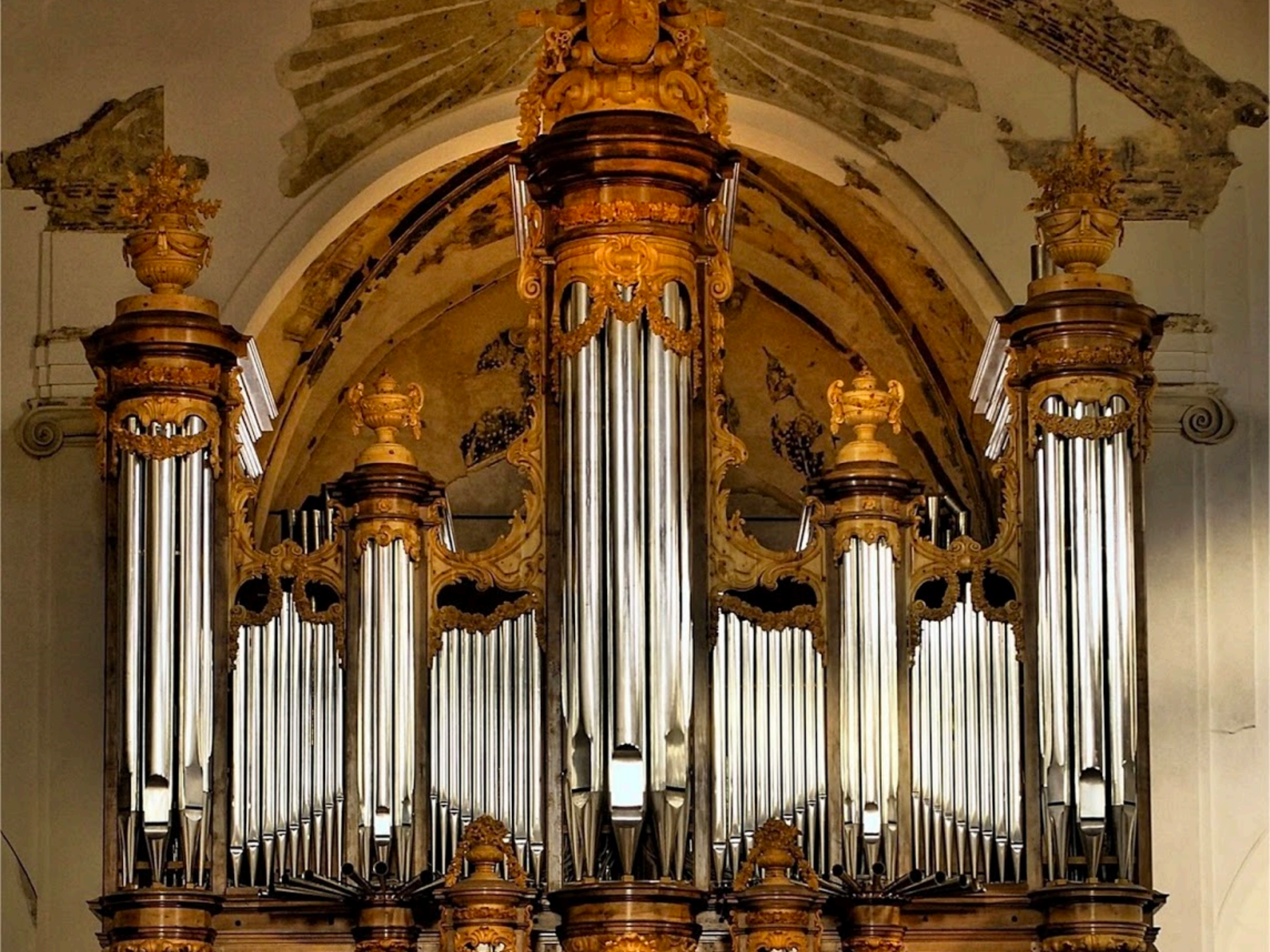
- Chordophones produces overtones when the chord vibrates with one or more **nodes** in the middle
- Similarly, aerophones produces overtones when there are **pressure nodes** within the resonating pipe
  - Pipes **open** on both sides (such as the flute or many organ pipes) can produce **all overtones**: $f = nc/2L$
  - Pipes **closed on one side** (such as the oboe, the clarinet, or a “bourdon” stop in the organ) can produce only **odd overtones**, with $f = nc/4L$

Pipe organs

Sound generation

- An air flux is directed against a “knife” (*labium*) in the pipe bottom part
- Swirls are created
- Because of Bernoulli’s law, air is sucked inside the pipe…
- …where another swirl is created
- The process repeats very rapidly






Flûte 8
Trompette 8
Flûte 16
Flûte 4
Clairon

Overtones and power spectrum

- The fundamental and the sequence of overtones can be easily discovered by computing the **power spectrum** of the sound pressure
- This operation, clearly, has to be carried out on a finite time interval
    - The shorter the interval, the higher our ability to trace frequency variations (such as quick notes)...
    - ...but also the poorer our frequency resolution!
- To perform this operation, we can use a (Gaussian) **window function** to isolate a specific time interval in the signal

Digital music

CDs, online services, MP3s...

- When an audio signal is digitalised, we are forced to perform two approximations
    - The signal is **sampled** at regular intervals: a consumer high-quality sampling is around 44-48 kHz
    - The signal is **quantised**: the sound pressure is at each sampling time is converted into a number with finite precision (16 bit or higher)
- Note that a sampling above 40 kHz is generally acceptable, since we cannot ear frequencies higher than 20 kHz

Discrete Fourier transform

- Is computed from a list of data (sound pressure) $\{x_n\}$ as

$$\hat{x}_k = \sum_{n=0}^{N} x_n \exp\left[-2\pi i \frac{kn}{N}\right]$$

- To compute it one can make use of scipy's fft or, better, rfft procedure: the latter does not return negative frequencies (which would be redundant)
- It is advisable to perform a DFT within a Gaussian window function of amplitude δ
- The corresponding DFT signal will be then convolved with a Gaussian of amplitude $1/2\pi\delta$

The experience

Data at https://tinyurl.com/yc7byrbk

- Compute the **power spectrum** of the signal within a window function (avoid attach and decay parts)
- Identify the **frequency** of the signal
    - Warning: the frequency is not necessary given by the highest peak in the power spectrum (in extreme cases, we might have a “missing fundamental”)
    - Rather, compute the maximum common divisor of the frequencies associated to the main peaks or use autocorrelation (be careful: approximations can lead to issues here…)
- Compute the amplitude of the **overtones**
- **Classify** the sound based on the relative amplitudes of the overtones using K-mean clustering