Audio Signal Processing : I. Introduction

Emmanuel Bacry

bacry@ceremade.dauphine.fr
http://www.cmap.polytechnique.fr/~bacry

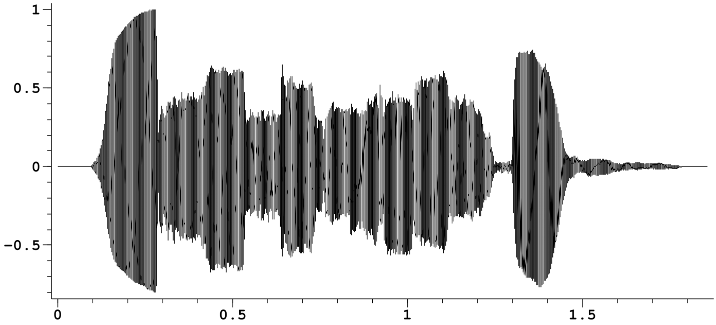
What is an Audio wave ?

- rarefaction/compression of molecules
- Progressive pressure wave
- different velocity depending on material
  - air : $340m.s^{-}1$
  - water : $1420m.s^{-}1$
  - steel : $6000m.s^{-}1$

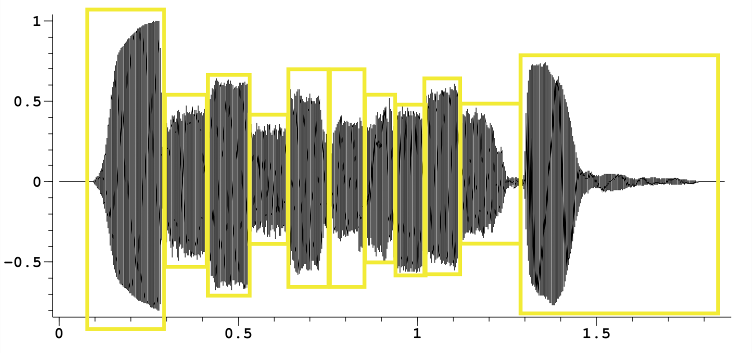
What is an Audio signal ?

- It is a 1D representation of the wave
- Audio signal $s(t)$: Measured pressure at a given location as a function of time
  - A microphone captures the wave with its membrane, and converts its movement into an electric (analog) signal
  - $s(t) = 0 \Rightarrow$ no sound, membrane is at rest
  - $\Rightarrow < s(t) >= 0$ in general

A first audio signal example

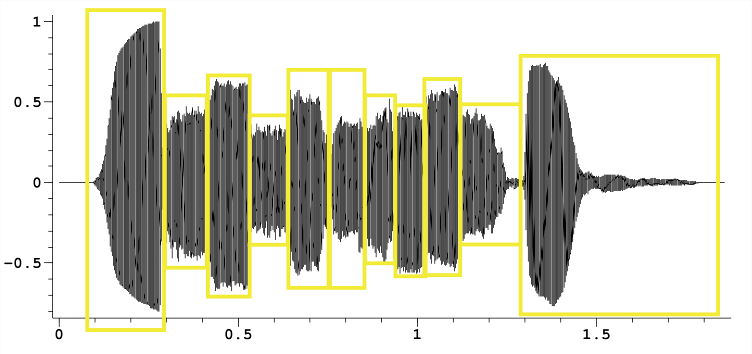


A first audio signal example



We can identify the notes on the signal representation !

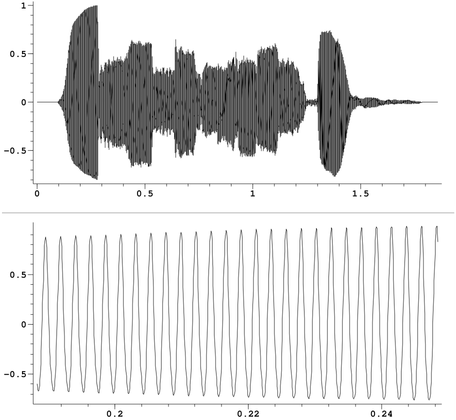
A first audio signal example



In first approximation, each note has a

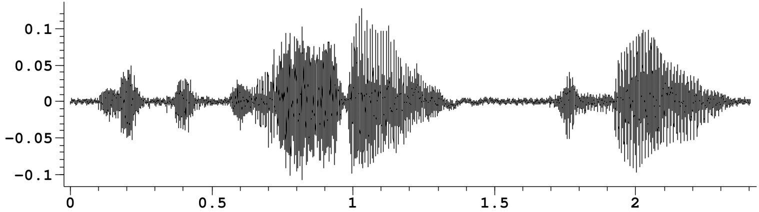
- attack phase
- sustain phase
- decay phase

Let's zoom on a sustain phase

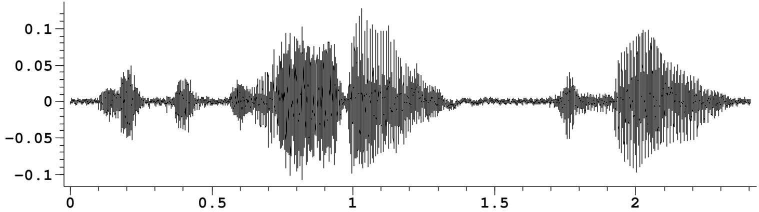


$\Longrightarrow$ **It is $\simeq$ Periodic !**

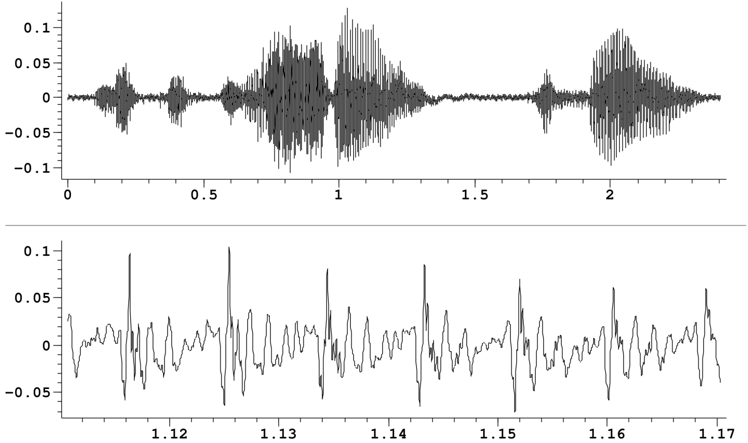
A second audio signal example



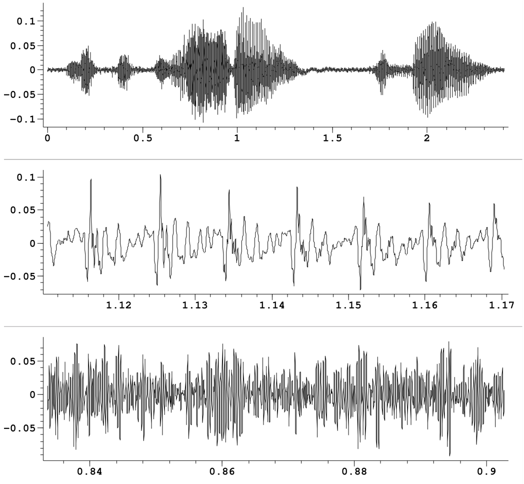
Well ... it is harder to "read" isn't it ?



Let's zoom on the "sustain" part of the 'a' vowel

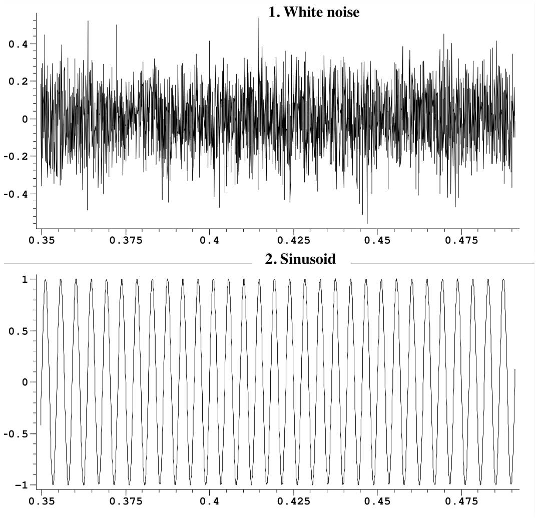


Let's zoom on the "noisy" part : the 'ch'



⟹ It is .... noisy !

Two "building bocks"



Two "building bocks"

1. White noise

$$s(t) = W(t)$$

2. Sinusoid

$$s(t) = A\sin(\omega t + \phi)$$

Pure sound

$$s(t) = A \sin(\omega t + \phi)$$

1. A is the **amplitude** and $I = A^2$ is the **intensity**
    - It is link to the volume of the sound
    - unit measure : deciBel

$$10 \log_{10} \frac{I}{I_0} = 20 \log_{10} \frac{A}{A_0} =$$

where I_0 is a seuil decoute

Pure sound

$$s(t) = A\sin(\omega t + \phi)$$

2. ω is the **pulsation**,and $F = \frac{\omega}{2\pi}$ is the **frequency**)
    - It is link to the pitch of the sound
    - unit measure : Hertz : number of oscillations per seconds

Pure sound

$$s(t) = A \sin(\omega t + \phi)$$

3. ϕ is the **phase**
   - On pure sound $\simeq$ translation in time

Periodic sound

$$s(t) = s(t+T)$$

where

- T is the period
- $F = \frac{1}{T}$ is the fundamental frequency
- $\omega = \frac{2\pi}{T}$ is the pulsation

$\Longrightarrow$ Let's use Fourier series !

Periodic sound

$$s(t) = s(t+T)$$

Fourier series Theorem :

$$s(t) = \sum_{n=-\infty}^{+\infty} c_n e^{in\omega t} \quad \text{with} \quad c_n = \frac{1}{T}\int_0^T s(t) e^{-in\omega t} dt$$

$$s(t) = \sum_{n=-\infty}^{+\infty} c_n e^{in\omega t} \quad \text{with} \quad c_n = \frac{1}{T}\int_0^T s(t) e^{-in\omega t} dt$$

- $s(t)$ is a real-valued signal

$$\Longrightarrow s(t) = \sum_{n=0}^{+\infty} a_n \sin(n\omega t + \phi_n)$$

- $< s(t) >= 0$

$$\Longrightarrow a_0 = 0$$

$$s(t) = \sum_{n=1}^{+\infty} a_n \sin(n\omega t + \phi_n), \quad \omega = \frac{2\pi}{T}$$

The n^{th} component $a_n \sin(n\omega t + \phi_n)$ corresponds to

- frequency $F_n = \frac{n}{T} = n\frac{\omega}{2\pi}$
  - $F_1 = \frac{1}{T}$ is the fundamental frequency ($\simeq$ pitch of the sound)
  - $F_n = nF_1$ are the harmonics of the sound
- amplitude a_n

Some very basic musical notions

The keys (on the piano)

- The white keys :
  - English notation : *C D E F G A B* . . . (cyclic)
  - French notation : *do ré mi fa sol la si* . . . (cyclic)
- cyclic $\rightarrow$ octave periodic (octave = frequency ratio of 2)
- General notation : *do4* = the *do* of the 4th octave
- The black keys
  do do# ré ré# mi fa fa# sol sol# la la# si . . . (cyclic)

A dive into harmonics . . .

$F_n = nF_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	?	?	?	?	?	?	?	?	?	?	?

A dive into harmonics . . .

$F_2 = 2F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	?	?	?	?	?	?	?	?	?	?

A universal frequency ratio : the octave = a ratio of 2

A dive into harmonics . . .

$F_3 = 3F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	*sol2*	?	?	?	?	?	?	?	?	?

Another universal frequency ratio : the fith = a ratio of 3/2 = *sol2/do2*

A dive into harmonics . . .

$F_4 = 4F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	*sol2*	*do3*	?	?	?	?	?	?	?	?

The octave again $do3/do2 = 4/2 = 2$

A dive into harmonics . . .

$F_5 = 5F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	?	?	?	?	?	?	?

Another important ratio : the third *mi3/do3* = 5/4

A dive into harmonics . . .

$F_6 = 6F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	?	?	?	?	?	?

The octave again $sol3/sol2 = 6/3 = 2$

A dive into harmonics . . .

$F_7 = 7F_1$

n	1	2	3	4	5	6	7	8	9	10	11	12
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	?	?	?	?	?

The seventh : *sib* is a note slightly above the *la#* of the piano

A dive into harmonics . . .

$F_8 = 8F_1$

n	1	2	3	4	5	6	7	8	9	10	11
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	*do4*	?	?	?

The octave again : $do4/do3 = 8/4 = 2$

A dive into harmonics . . .

$F_9 = 9F_1$

n	1	2	3	4	5	6	7	8	9	10	11
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	*do4*	*ré4*	?	?

The ninth : *ré4*

A dive into harmonics . . .

$F_1 0 = 10F_1$

n	1	2	3	4	5	6	7	8	9	10	11
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	*do4*	*ré4*	*mi4*	?

The octave again : $mi4/mi3 = 10/5 = 2$

A dive into harmonics . . .

$F_11 = 11F_1$

n	1	2	3	4	5	6	7	8	9	10	11
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	*do4*	*ré4*	*mi4*	*fa*

A dive into harmonics . . .

$F_1 1 = 11F_1$

n	1	2	3	4	5	6	7	8	9	10	11
note	*do1*	*do2*	*sol2*	*do3*	*mi3*	*sol3*	*sib3*	*do4*	*ré4*	*mi4*	*fa*

The octave again : $sol4/sol3 = 12/6 = 2$

Some important chords/intervals . . .

- the octave : *do1+do2*
- the fifth :*do+sol*
- the third :*do+mi*
- the perfect chord :*do+mi+sol*
- the seventh chord :*do+mi+sol+sib*

$$s(t) = \sum_{n=1}^{+\infty} a_n \sin(n\omega t + \phi_n), \quad \omega = \frac{2\pi}{T}$$

What about the amplitudes a_n ?

In a physical system, generally a_n is decreasing with n

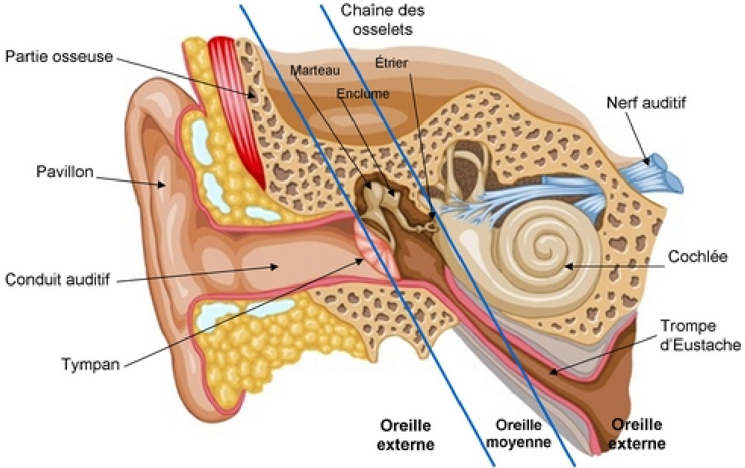
- pinched string (harp, pizzicatti) : $a_n = \frac{1}{n^2}$
- struck string (piano) : $a_n = \frac{1}{n}$

$$s(t) = \sum_{n=1}^{+\infty} a_n \sin(n\omega t + \phi_n), \quad \omega = \frac{2\pi}{T}$$

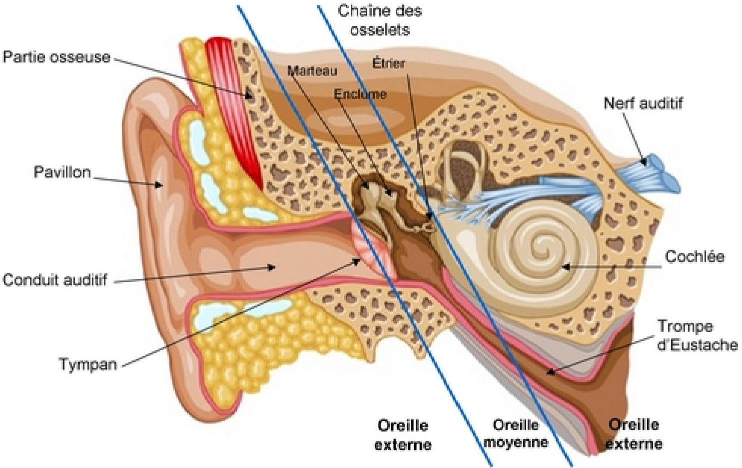
A "miracle" : when adding up harmonics we do not hear the different pitches, we still here the pitch corresponding to Fundamental frequency $F_1 = \frac{1}{T}$

The a_n are $\simeq$ responsible for the timber (more later)

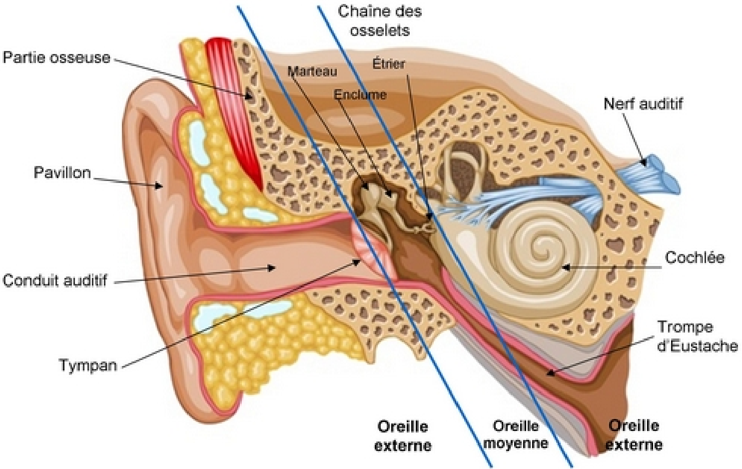
- The pure sound
- The triangle sound
- The sawtooth sound
- The square sound



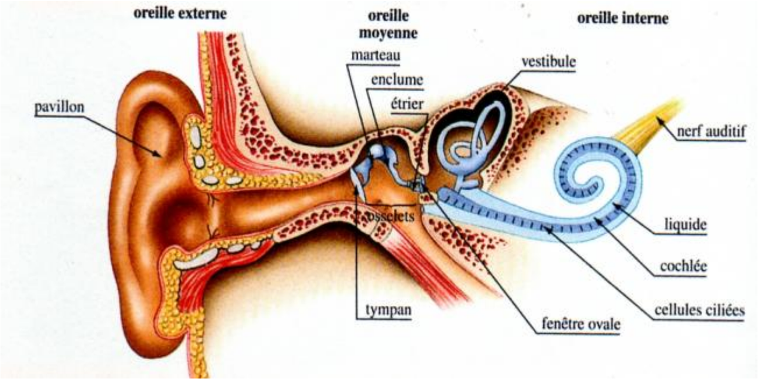
External ear : ear canal + pinacle



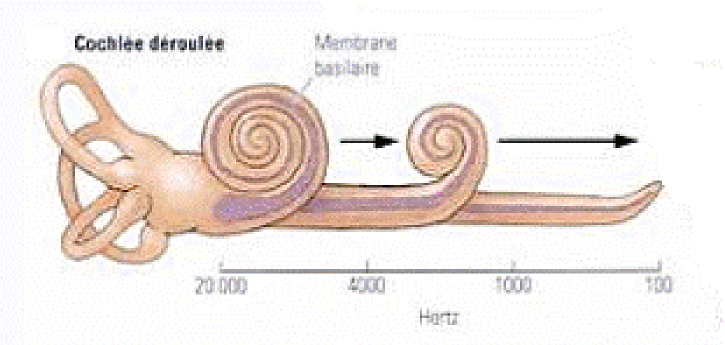
Middle ear : the eardrum operates the ear hammer which strikes on the anvil



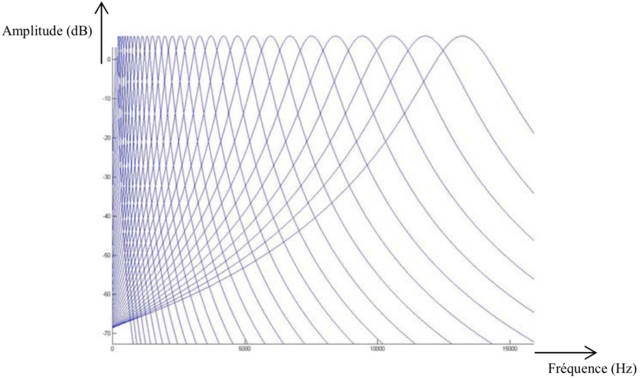
Internal ear : the anvil makes the stirrup vibrate at the entrance to the spiral shaped cochlea



Cochlea : has "hairy" cells that direcly connected to nerve cells



Cochlea : has "hairy" cells that are activated by different frequencies and direcly linked to nerve cells ... (and then it goes to the brain !)



The hairy cells act like a filter bank, with a better resolution in low frequency than high frequency (wavelets ?).

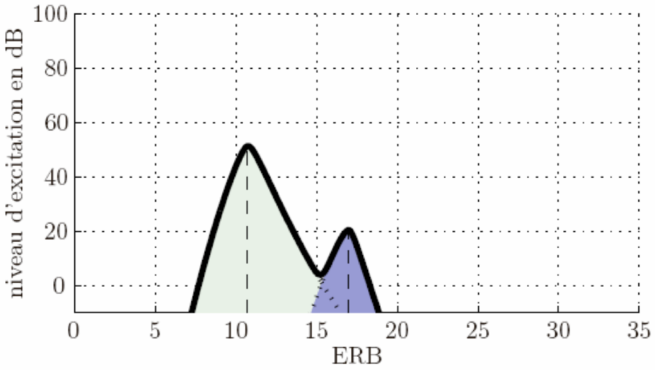
Psycho-acoustics : *the branch of psychology concerned with the perception of sound and its physiological effects*

Perception of Intensity

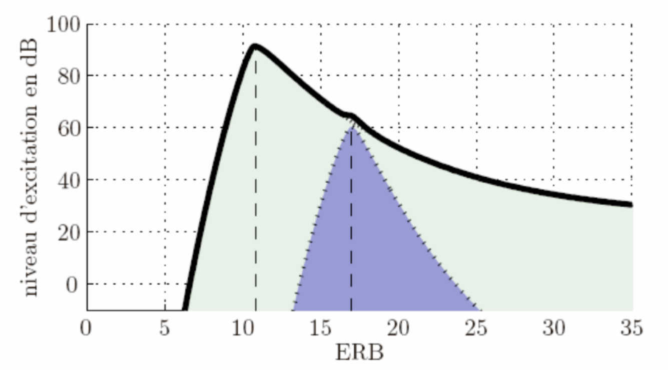
- Perceptively, the intensities of several sounds do not simply add up.
- The resulting perceived instensity is the result of a very complex process.
- Perceptive unit : Sonie (1 sone = 1000Hz at 40dB)
- In practice : at 100Hz with more or less perceive the real intensity, at 1000Hz the perceived intensity is much smaller
- Model ?

$$sonie =: \sum_i hairy\_cell\_responses_i$$

  And there is a masking phenomenon within each hairy cell (it "keeps" only the strongest stimulus)
- It results in a global **masking phenomenon**



Response of two pure tones not too loud : 500Hz (50dB) and 1200Hz (20dB)
The Equivalent Rectangular Bandwidth (ERB) model by Moore and Glassberg (1983)



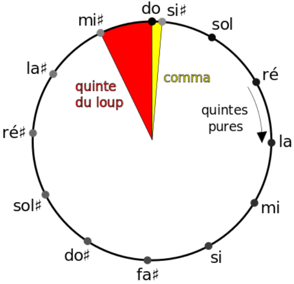
Response when tones are louder : 500Hz (90dB) and 1200Hz (60dB)
The 1200Hz tone is masked

Perception of Pitch

- Perceptive unit : Tonie
- The high picth sounds are narrower than the low pitch ones : octaves are less than 2 in hte high frequencies.
- The tuning of a piano is a complex process
  - in the mid-pitches there are 3 strings per key (potential interferences)
  - Which scale to choose ?
- Musical scales : the obsession of several centuries
  - Octave = 2 : OK ! (shares many harmonics)
  - then what ? fifth ?

Defining a musical scale . . .

1. The fifth circle of pythagore : using the $\frac{3}{2}$ ratio



→ : The pythagorician comma : $\frac{531441}{524288}$
→ : "The perfect chord" is .... awful

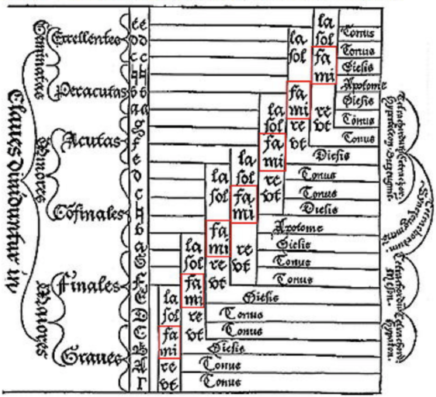
Defining a musical scale . . .

2. The Zarlino solution (16th century)

→ The perfect chord in doM sounds nice
→ But there are different types of thirds
→ Don't even dare transposing too far from the original tone . . .

Defining a musical scale . . .

3. 4. 5. . . . So many definitions . . .



→ Don't even dare transposing too far from the original tone . . .

Defining a musical scale . . .

- The tempered scale : A revolution (Simon Stevin 1548-1620, Werckmeister 1691)

The idea : All the notes will be equally out of tune !

Let's divide the octave in twelve equal intervals

Equal division between all the 12 keys of the piano using a single ration : $2^{1/12}$

Back to pitch . . .

- Is pitch $\Longrightarrow$ frequency ?
- Nope : pitch is about periodicity
- Pitch detection is a difficult task : all the more difficult if you are dealing with consonnant music (sixth, octaves and fifths)

What about timber ? . . .

- Timber : Very hard to define notion
  timbre is what makes a particular musical sound have a different sound from another
- Is there such a thing as the timber of a synthethizer ? of an organ ?

What about timber ? . . .

Case of a periodic sound :

$$s(t) = \sum_{n=1}^{+\infty} a_n \sin(n\omega t + \phi_n), \quad \omega = \frac{2\pi}{T}$$

$\longrightarrow$ Timber is given by the a_n

WARNING : Subtle difference between a chord of pure sounds and a monophonic periodic sound

What about timber ? . . .

Case of a periodic sound :

$$s(t) = \sum_{n=1}^{+\infty} a_n \sin(n\omega t + \phi_n), \quad \omega = \frac{2\pi}{T}$$

What about the phases ϕ_n ?

When exactly are we sensitive to the phases ?