

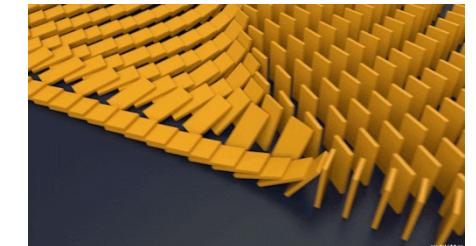
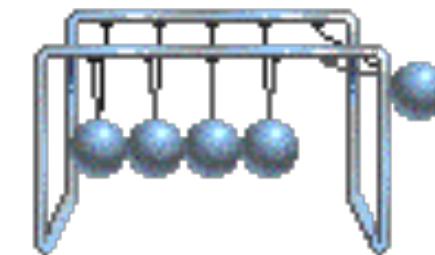
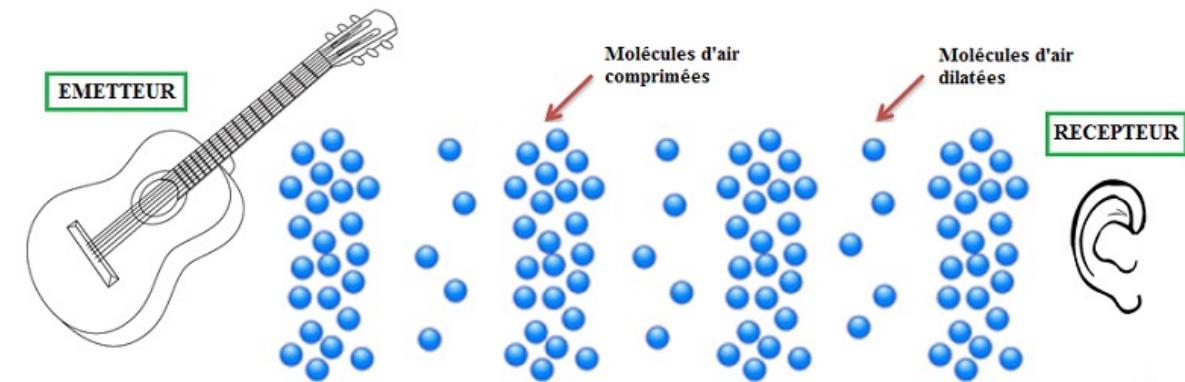
Représentations audionumériques

Louis Bigo

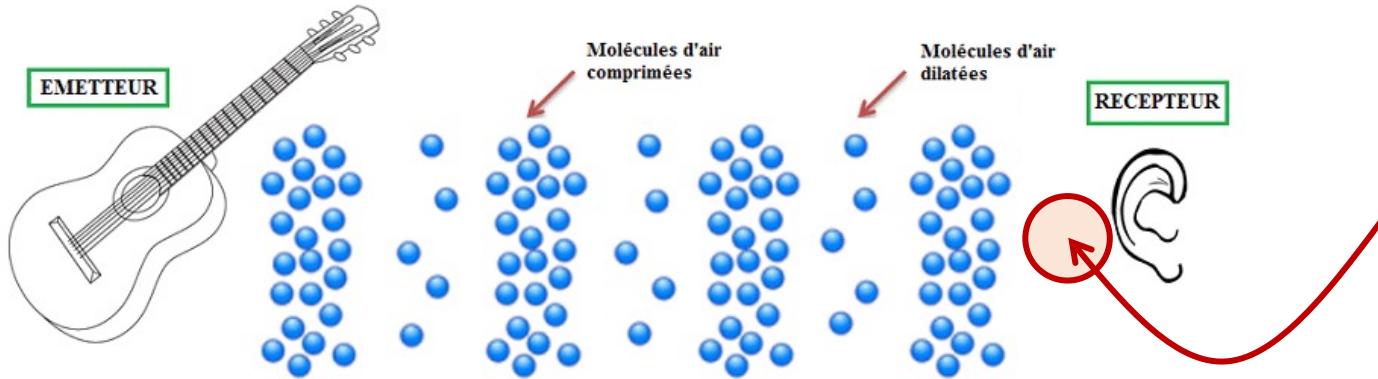
Intelligence artificielle musicale - 2024
ENSEIRB-MATMECA, 3A parcours IA

Le son

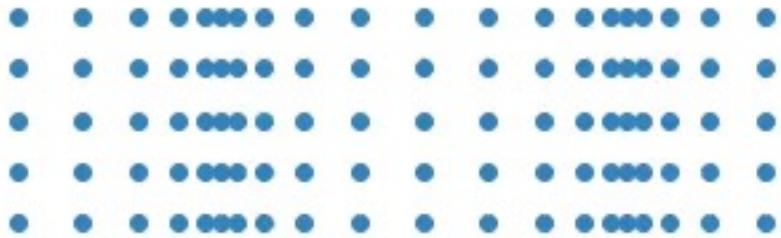
- Le son est une *onde de pression*
 - créé par des molécules qui vibrent
 - pas de déplacement de molécules (analogie avec des vaguelettes)
 - Vitesse du son dans l'air = 340m/s



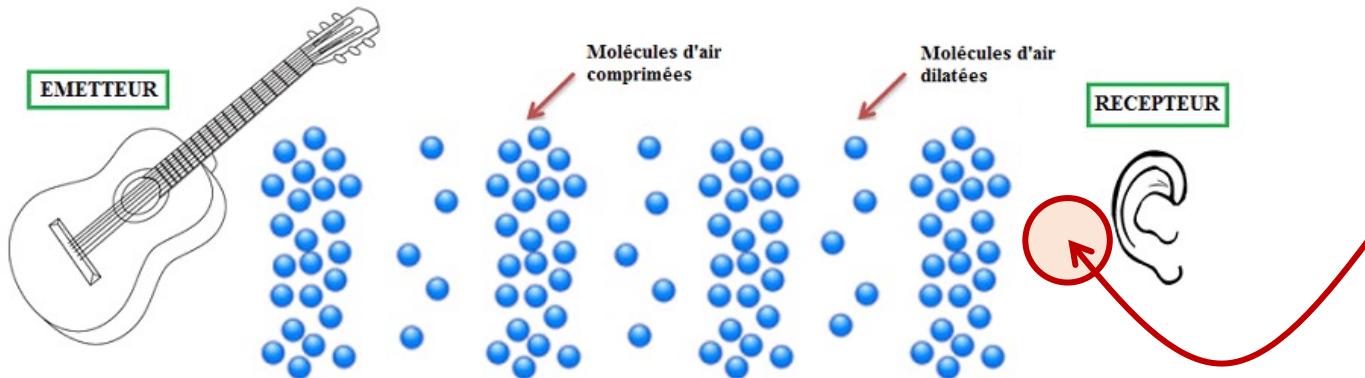
La « forme d'onde »



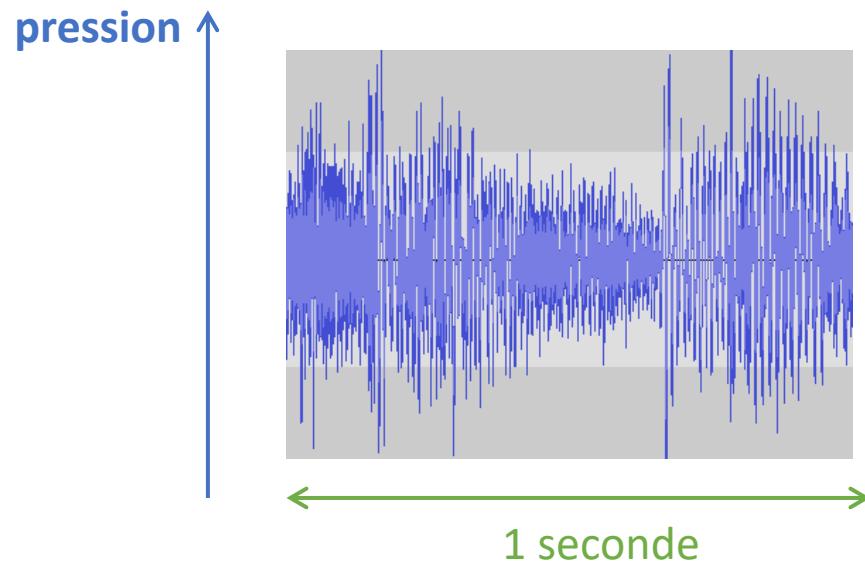
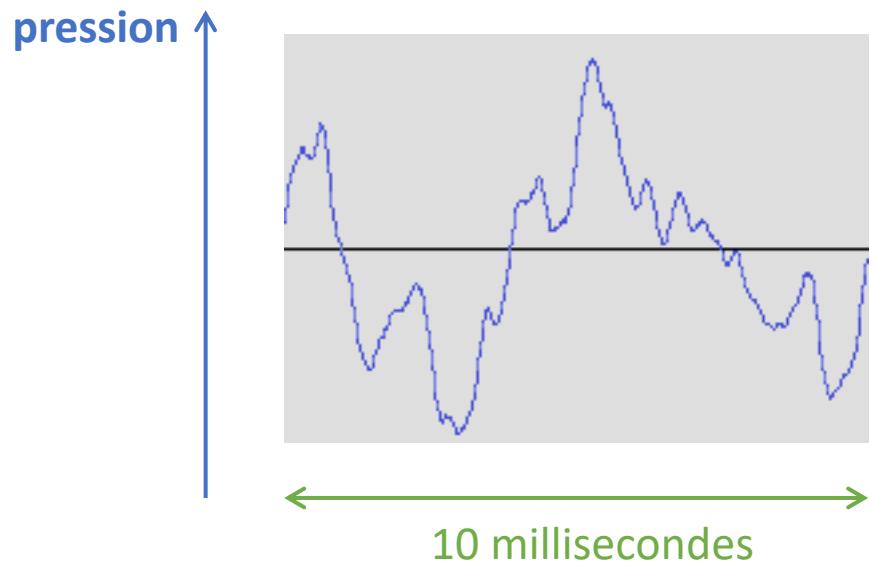
Ce que l'oreille perçoit :
Comment varie la **pression de l'air**
à cet endroit **au cours du temps** ?



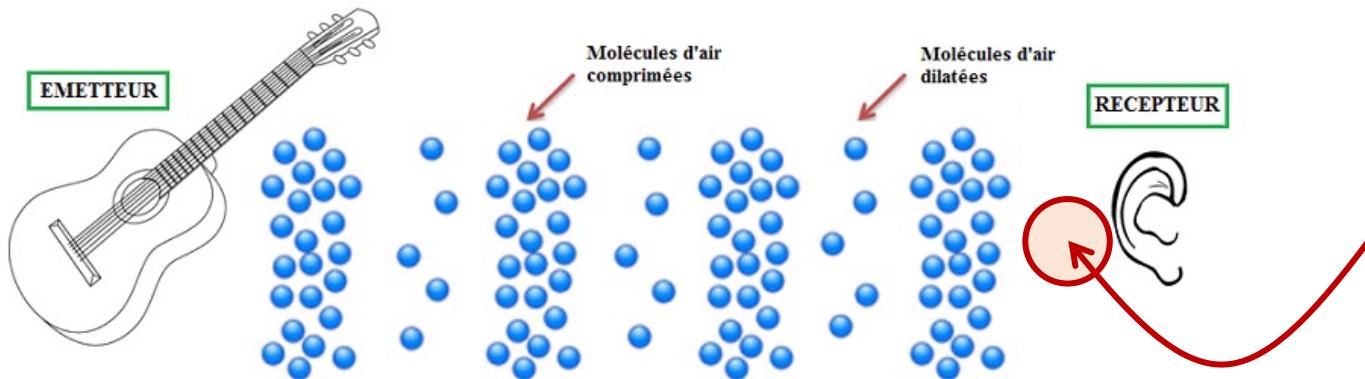
La « forme d'onde »



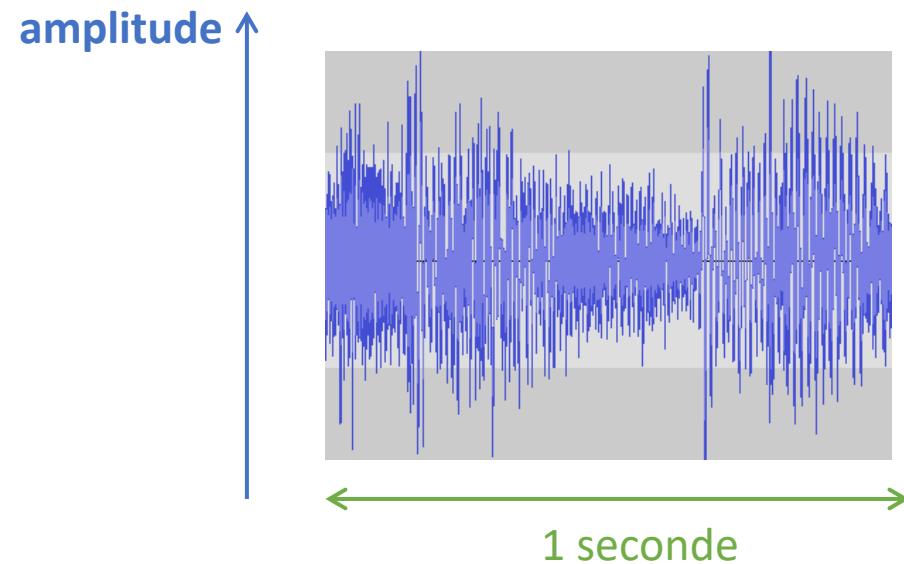
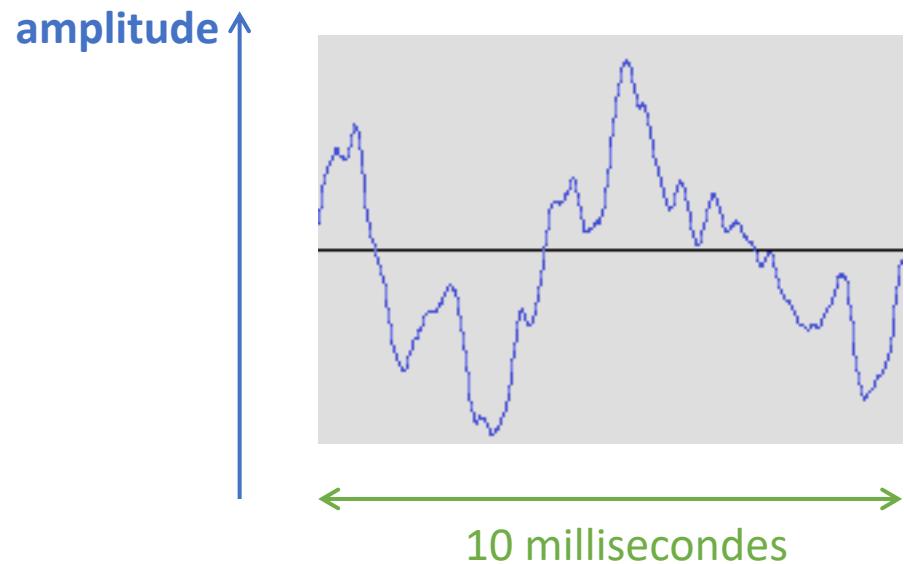
Ce que l'oreille perçoit :
Comment varie la **pression de l'air**
à cet endroit **au cours du temps** ?



La « forme d'onde »

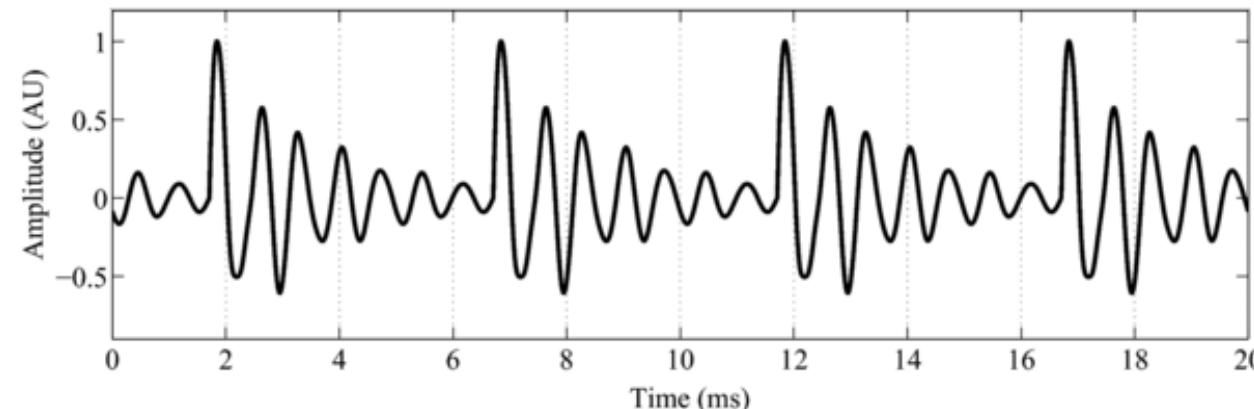


Ce que l'oreille perçoit :
Comment varie la **pression de l'air**
à cet endroit **au cours du temps** ?



Les types de sources sonores

- À impulsion initiale (percussion)
- À excitation entretenue (voix, violon, etc.)
 - Variations répétitives de la pression
 - Forme d'onde dite *périodique*
 - *Fréquence fondamentale* : nombre de cycles en 1 seconde
 - Fréquence : perception de *hauteur* (au sens musical)



La fréquence

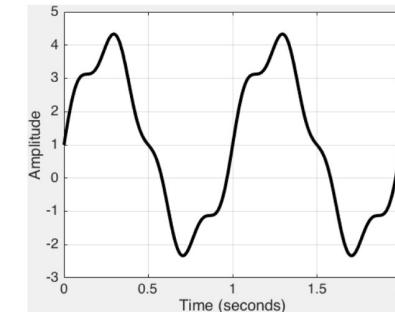
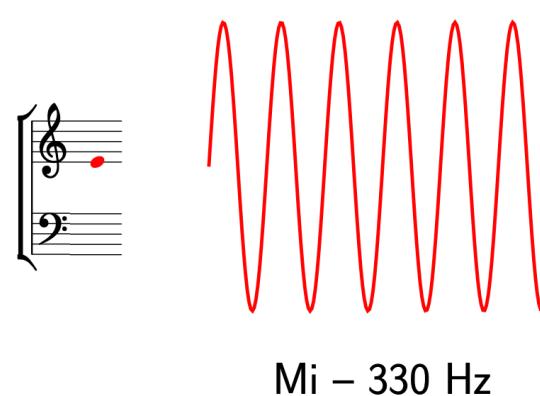
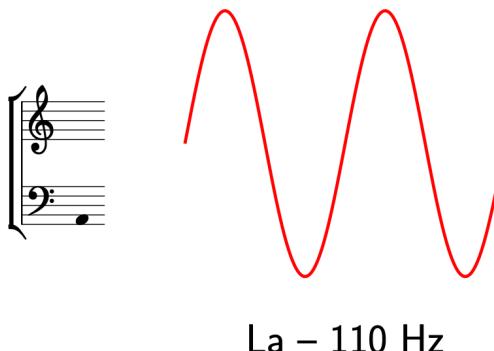
- La fréquence résulte d'un *schéma* qui se répète dans la forme d'onde

- La valeur de la fréquence est liée à la longueur de ce schéma :

Combien de fois ce schéma se répète t-il en une seconde ? 440 fois → la fréquence est égale à 440 Hz.

- La **fréquence fondamentale** induit une perception de **hauteur** dans un son

(à condition que cette fréquence se situe dans le domaine audible : 20Hz → 20 000Hz)

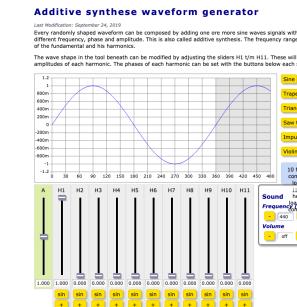
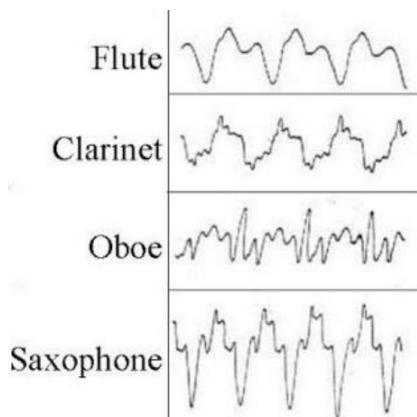


La fréquence

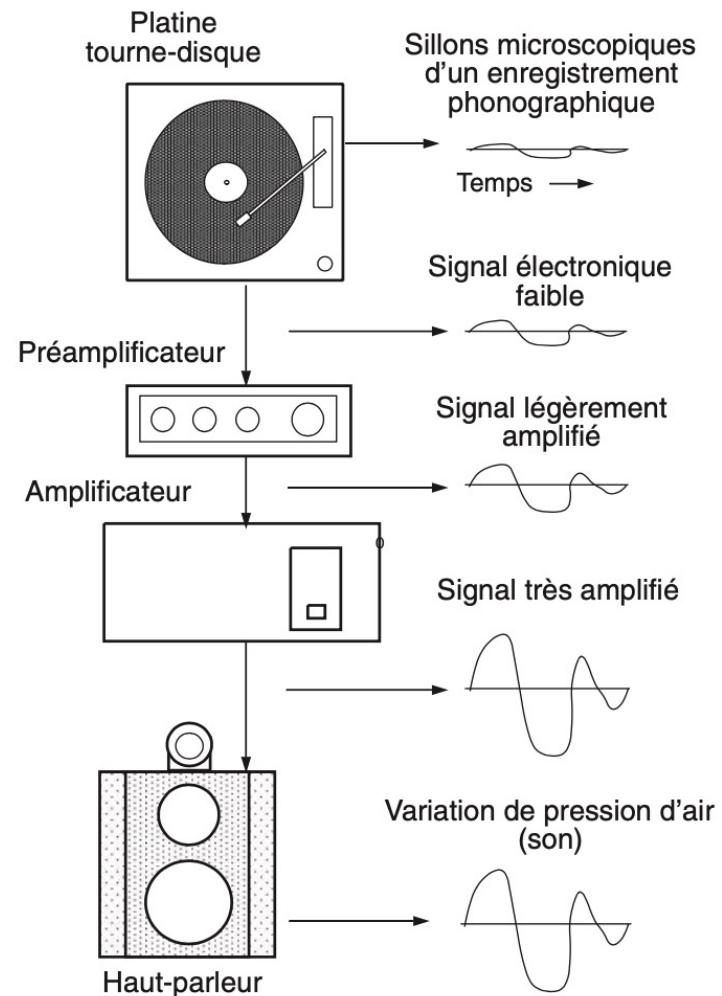
- La fréquence résulte d'un *schéma* qui se répète dans la forme d'onde
- La valeur de la fréquence est liée à la longueur de ce schéma :

Combien de fois ce schéma se répète t-il en une seconde ? 440 fois → la fréquence est égale à 440 Hz.

- La **fréquence fondamentale** induit une perception de **hauteur** dans un son
(à condition que cette fréquence se situe dans le domaine audible : 20Hz → 20 000Hz)
- Le **timbre** d'un son résulte de la forme de l'onde au cours d'une *période* de la fréquence fondamentale



Représentation analogique du son



Représentation numérique du son

- Le son est un phénomène **continu**.

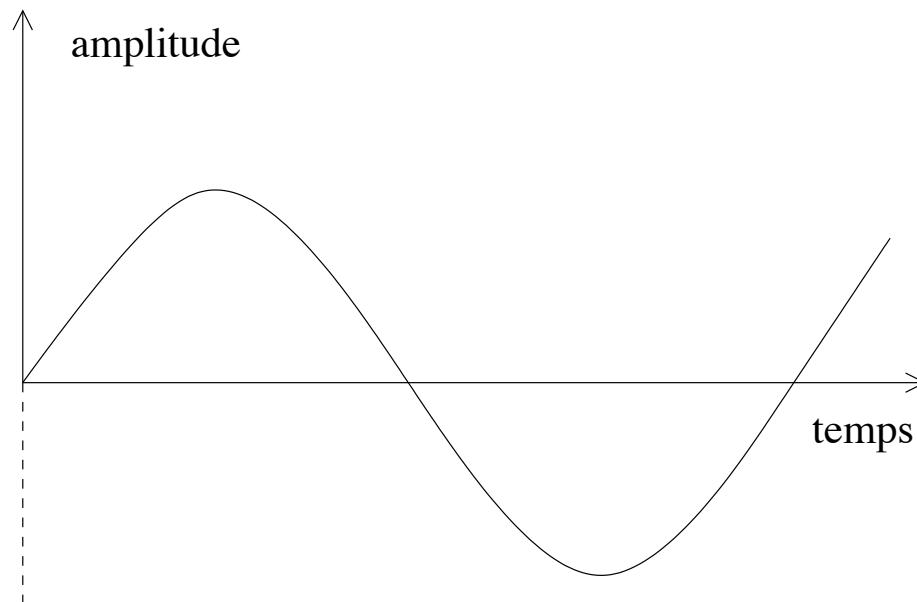
Entre deux instants précis, il correspond à une *infinité de valeurs*, chacune étant *infiniment précises*

→ *Comment un ordinateur peut-il mémoriser une telle information ?*

- *Problème n°1* : un ordinateur ne peut pas stocker une infinité de valeurs
 - Solution : ne mémoriser le son que toutes les x secondes (=échantillonnage)
 - On échantillonne couramment à 44100Hz (soit tout de même 44100 valeurs par secondes !)
 - Analogie avec le cinéma (la bobine de film) : au delà d'un certain nombre d'éléments pas secondes, nous percevons un *phénomène continu*

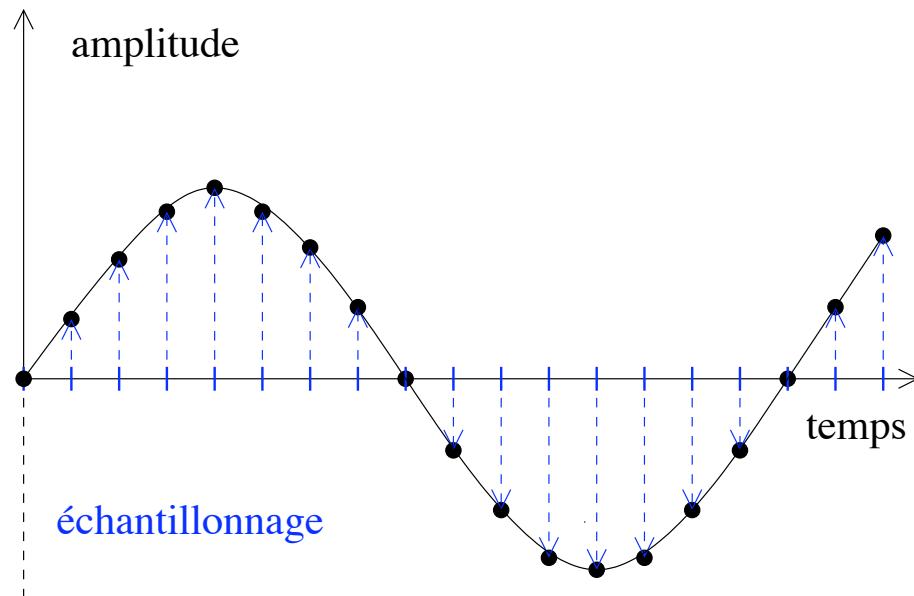
Représentation numérique du son

Problème n°1 : un ordinateur ne peut pas stocker une infinité de valeurs



Représentation numérique du son

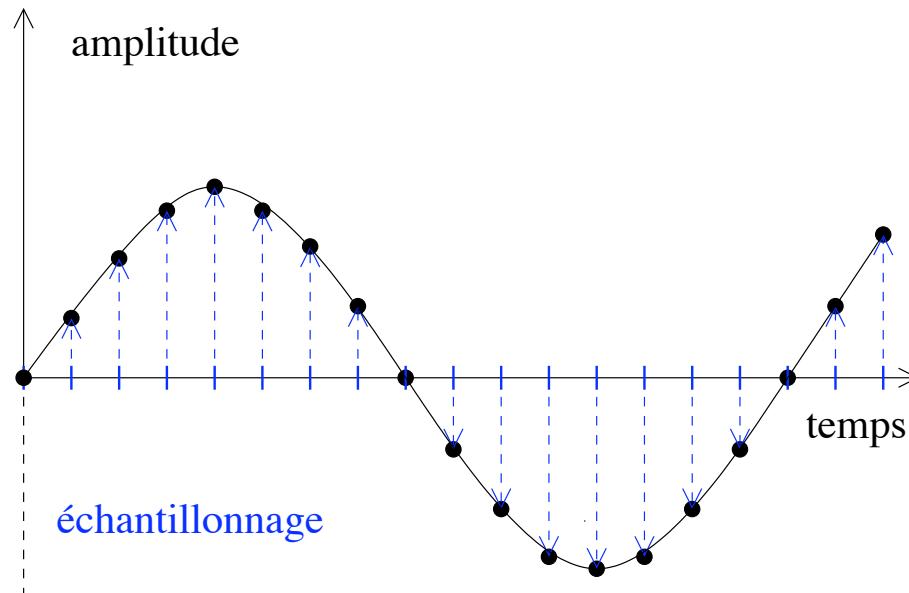
Problème n°1 : un ordinateur ne peut pas stocker une infinité de valeurs



Échantillonnage : on ne garde qu'une valeur d'amplitude toutes les xxx millisecondes

Représentation numérique du son

Problème n°1 : un ordinateur ne peut pas stocker une infinité de valeurs



Échantillonnage à 44100Hz : on ne garde qu'une valeur d'amplitude toutes les 0,02 millisecondes
(= on prend 44100 valeurs par seconde)

Représentation numérique du son

- Le son est un phénomène **continu**.

Entre deux instants précis, il correspond à une *infinité de valeurs*, chacune étant *infiniment précises*

→ *Comment un ordinateur peut-il mémoriser une telle information ?*

- *Problème n°1 : un ordinateur ne peut pas stocker une infinité de valeurs*

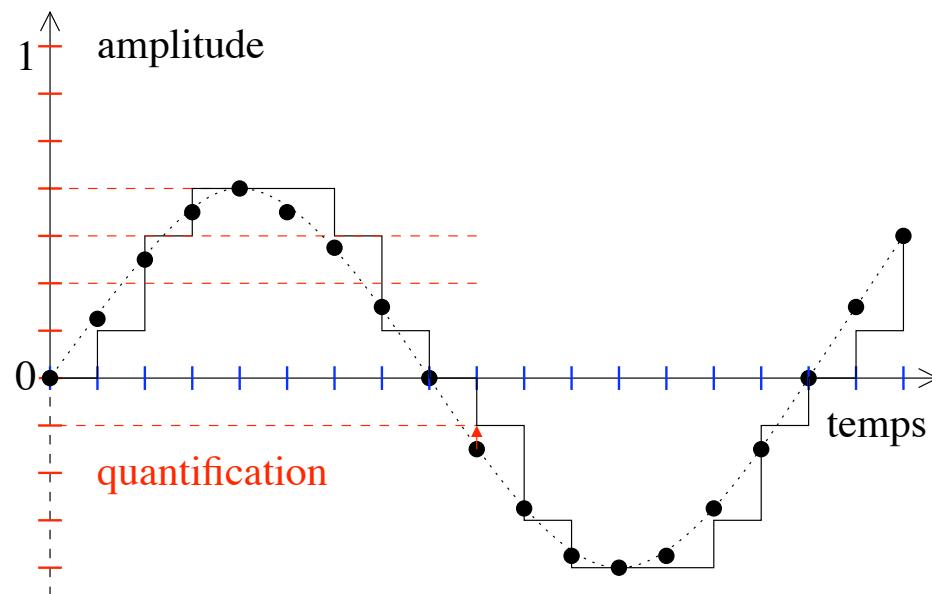
- Solution : ne mémoriser le son que toutes les x secondes (=échantillonnage)
- On échantillonne couramment à 44100Hz (soit tout de même 44100 valeurs par secondes !)
- Analogie avec le cinéma (la bobine de film) : au delà d'un certain nombre d'éléments pas secondes, nous percevons un *phénomène continu*

- *Problème n°2 : un ordinateur ne peut pas stocker une valeur avec une précision exacte*

- Solution : l'approximation (=quantification)

Représentation numérique du son

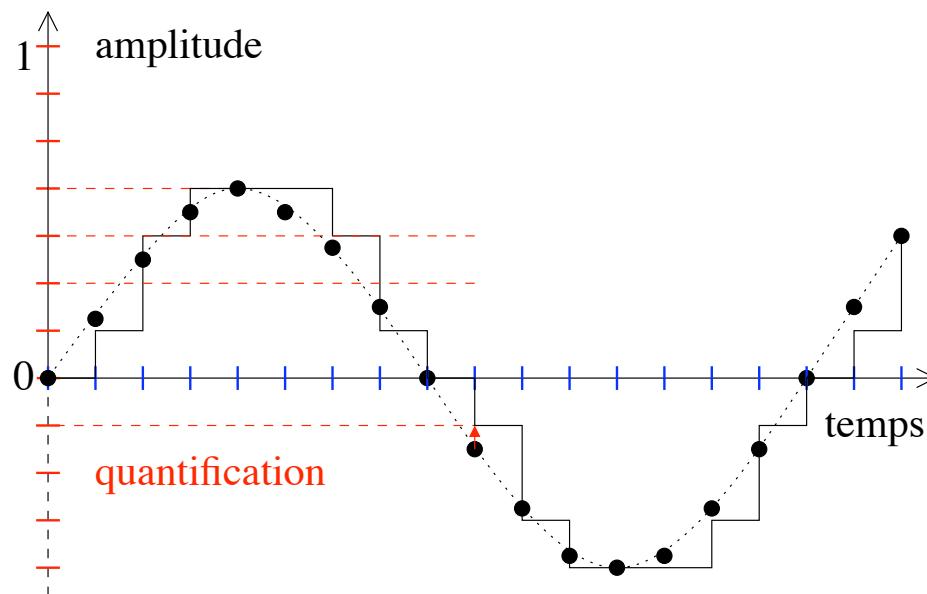
Problème n°2 : un ordinateur ne peut pas stocker une valeur avec une précision exacte



Quantification : on arrondit la valeur de l'amplitude

Représentation numérique du son

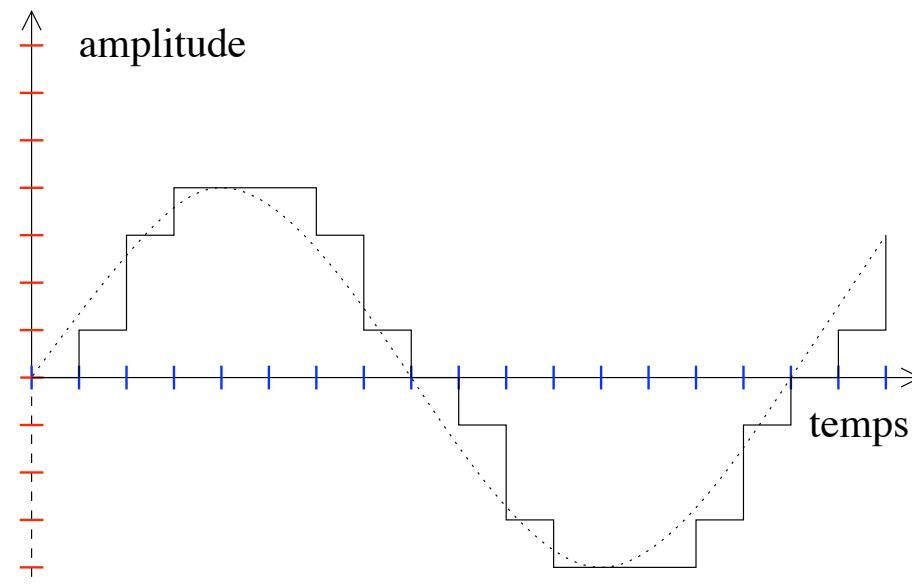
Problème n°2 : un ordinateur ne peut pas stocker une valeur avec une précision exacte



Quantification : on arrondit la valeur de l'amplitude
Pour un CD : encodage 16-bit (= 65 536 valeur d'amplitude possibles)

Représentation numérique du son

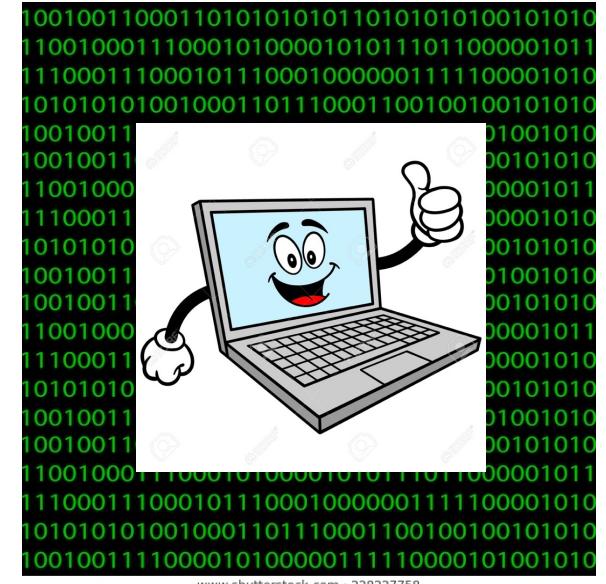
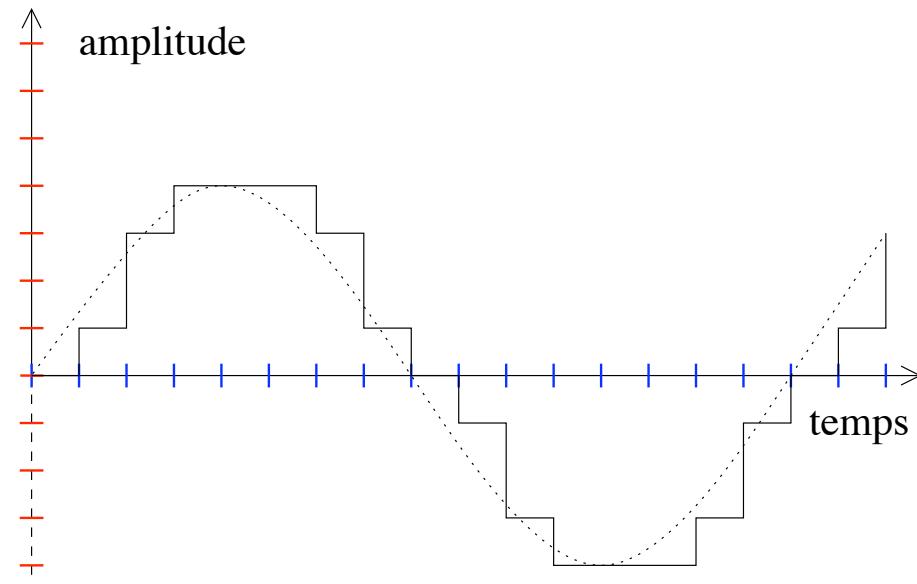
- Numérisation = échantillonnage + quantification
- La numérisation permet de représenter une forme d'onde sous la forme d'une liste de nombres



[0,1,3,4,4,4,3,1,0,-1,-3,-4,-4,-4,-3,-1,0,1,3]

Représentation numérique du son

- Numérisation = échantillonnage + quantification
 - La numérisation permet de représenter une forme d'onde sous la forme d'une liste de nombres



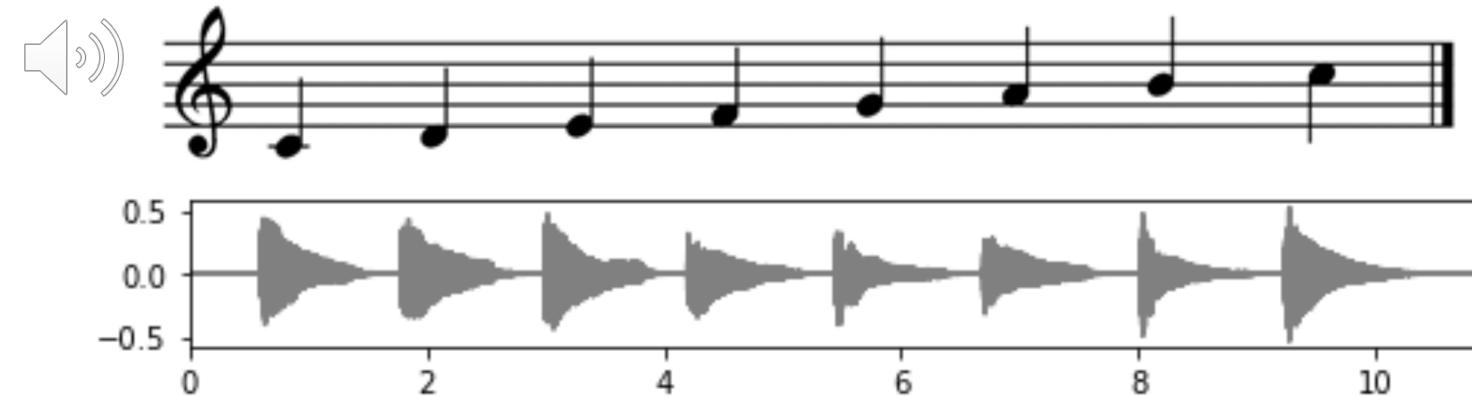
[0,1,3,4,4,4,3,1,0,-1,-3,-4,-4,-4,-3,-1,0,1,3]

Représentation numérique du son

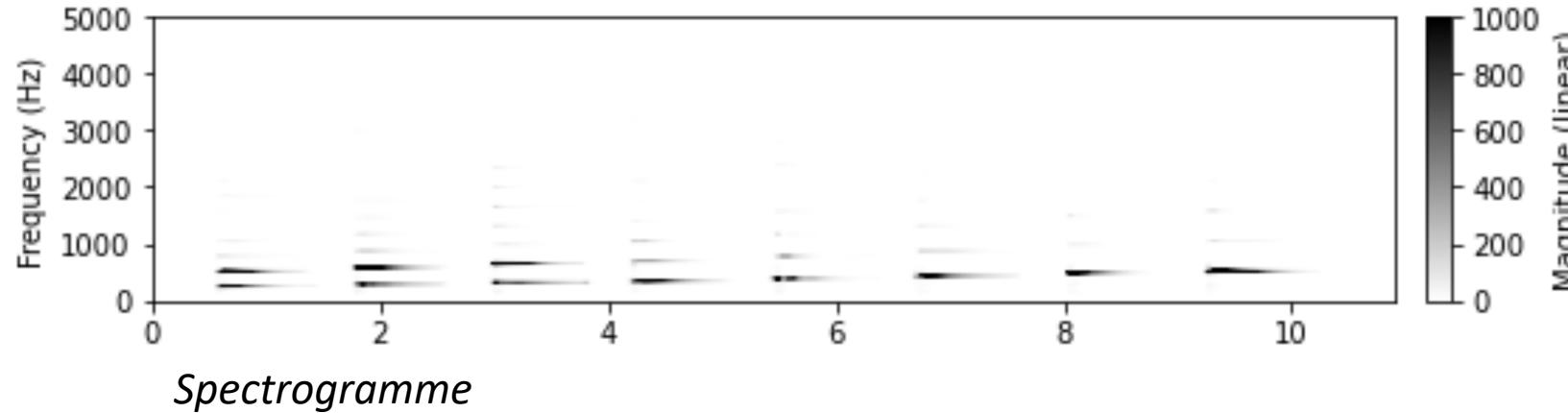
- Numérisation = **échantillonnage** + **quantification**
- La numérisation permet de représenter une forme d'onde sous la forme d'une liste de nombres
Cela facilite notamment :
 - Le **stockage numérique** (sur disque dure, clé usb, CD, ...)
 - La **transmission numérique** (cable réseau, Wifi, Bluetooth, xG, cable USB, etc.)
Copie sans « ajout de bruit » contrairement aux supports analogiques
 - Le **traitement numérique** :
 - Montage (couper, copier, coller, etc.)
 - Mixage (volumes, fondus, normalisation, compression, égalisation, etc. de forme d'ondes mélangées)
 - Application d'effets (réverbération, délai, chorus, distorsion, réduction du bruit de fond, *pitch-shifting* etc.)

Représentations fréquentielles

Spectrogrammes

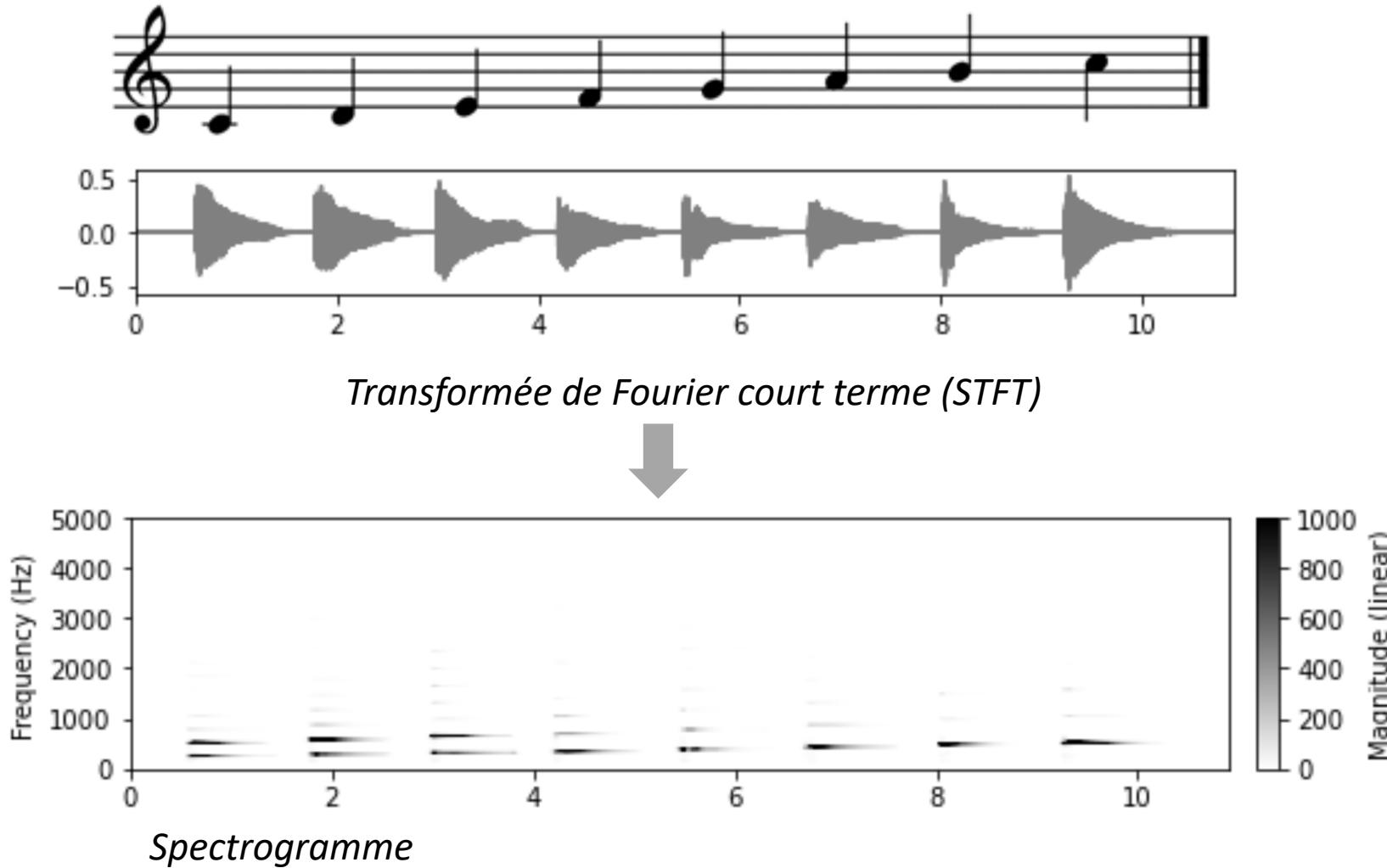


Transformée de Fourier court terme (STFT)

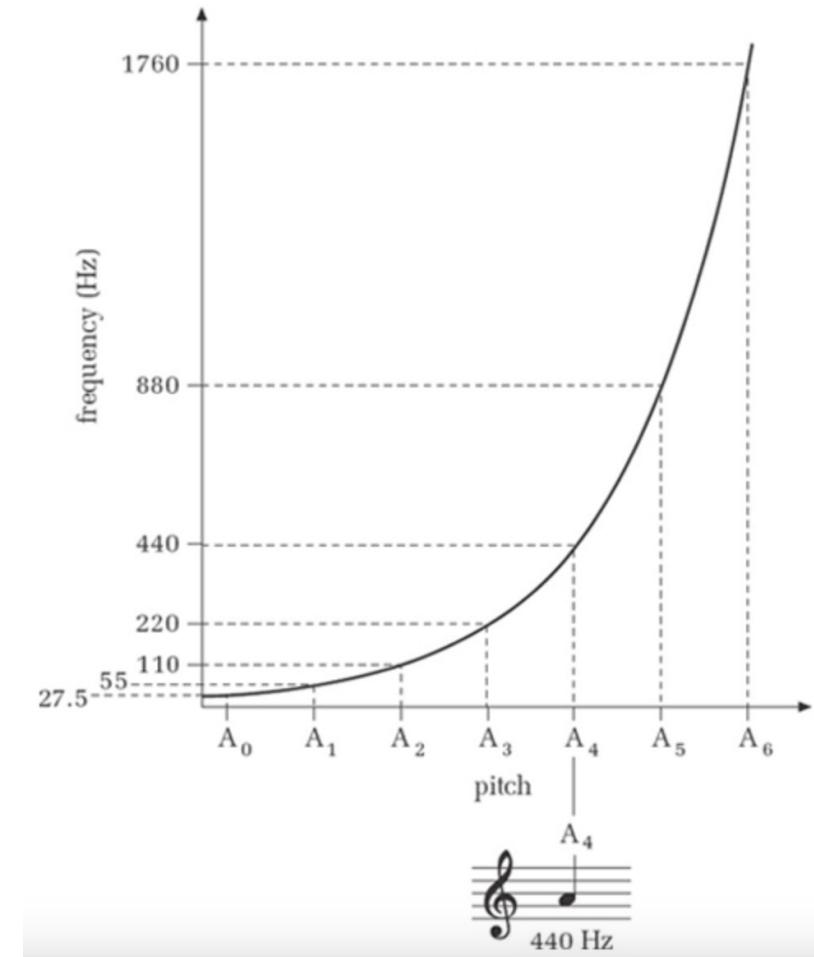


Représentations fréquentielles

Spectrogrammes



fréquence \leftrightarrow hauteur de note (=pitch)



→ log-frequency spectrogram

Représentations fréquentielles

- Fréquence de la n -ième touche d'un piano

$$f = 440 \text{ Hz} \cdot 2^{[(n - 49)/12]}$$

- Fréquence de la hauteur MIDI N (ou *pitch MIDI*)

$$f = 440 \text{ Hz} \cdot 2^{[(N_{\text{MIDI}} - 69)/12]}$$

- Pitch MIDI de la fréquence f

$$N_{\text{MIDI}} = 69 + 12 \cdot \log_2 \left(\frac{f}{440 \text{ Hz}} \right)$$

- Classe de hauteur*

$$C = N_{\text{MIDI}} \% 12$$

MIDI number	Note name	Keyboard	Frequency Hz	Period ms
21	A0		27.500	36.36
23	B0		30.868	32.40
24	C1		32.703	30.58
26	D1		36.708	28.86
28	E1		41.203	25.71
29	F1		43.654	22.91
31	G1		48.999	21.62
33	A1		55.000	19.26
35	B1		61.735	18.18
36	C2		65.406	17.16
38	D2		73.416	16.20
40	E2		82.407	14.29
41	F2		87.307	12.13
43	G2		97.999	10.81
45	A2		110.00	9.631
47	B2		123.47	8.581
48	C3		130.81	7.645
50	D3		146.83	7.216
52	E3		164.81	6.068
53	F3		174.61	5.727
55	G3		196.00	5.102
57	A3		220.00	4.816
59	B3		246.94	4.290
60	C4		261.63	3.822
62	D4		293.67	3.608
64	E4		329.63	3.214
65	F4		349.23	2.863
67	G4		392.00	2.703
69	A4		440.00	2.273
71	B4		493.88	2.025
72	C5		523.25	1.910
74	D5		587.33	1.703
76	E5		659.26	1.607
77	F5		698.46	1.432
79	G5		783.99	1.351
81	A5		880.00	1.204
83	B5		987.77	1.073
84	C6		1046.5	0.9556
86	D6		1174.7	0.8513
88	E6		1318.5	0.8034
89	F6		1396.9	0.7159
91	G6		1568.0	0.6378
93	A6		1760.0	0.5682
95	B6		1975.5	0.5062
96	C7		2093.0	0.4778
98	D7		2349.3	0.4257
100	E7		2637.0	0.3792
101	F7		2793.0	0.3580
103	G7		3136.0	0.3189
105	A7		3520.0	0.2841
107	B7		3951.1	0.2481
108	C8		4186.0	0.2120

J. Wolfe, UNSW

MuseScore

Représentations fréquentielles

Chromagrammes

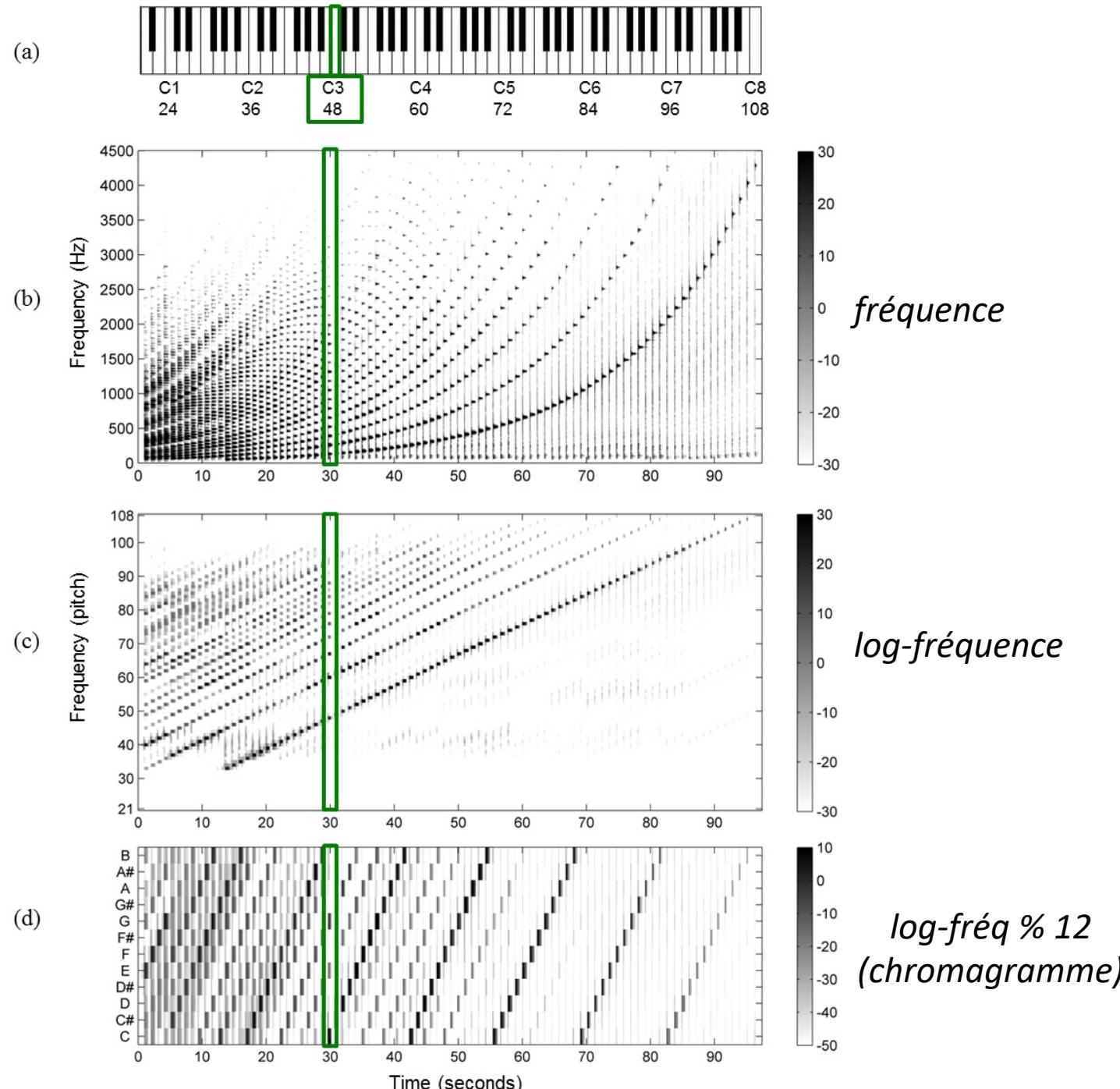
- Pitch MIDI de la fréquence f

$$N_{\text{MIDI}} = 69 + 12 \cdot \log_2 \left(\frac{f}{440 \text{ Hz}} \right)$$

- Classe de hauteur*

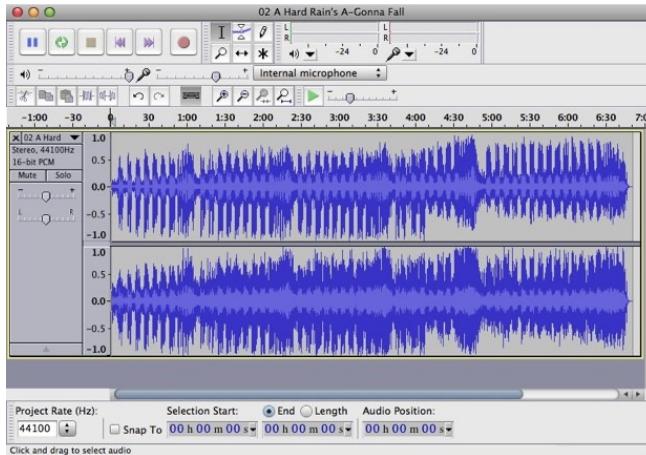
$$C = N_{\text{MIDI}} \% 12$$

Chromagramme : agréger l'information spectrale d'une *classe de hauteur* (= pitch % 12)

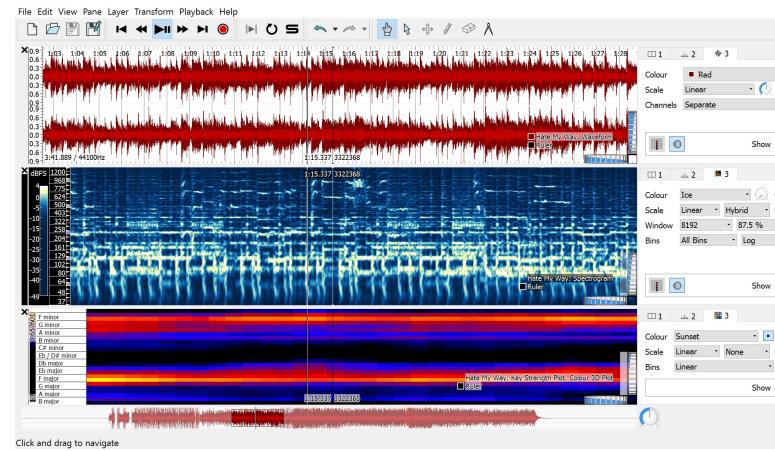


Outils pour l'analyse sonore (logiciels et programmation)

Visualisation, annotations, extraction d'information



Audacity



Sonic Visualizer

```
50 y,sr = librosa.load(filename_tab)
51 print("nombre de samples : {}".format(len(y)))
52 print("Fe : {}".format(sr))
53 # Set the hop length; at 22050 Hz, 512 samples ~ 23ms
54 #hop_length = 512
55
56 # Separate harmonics and percussives into two waveforms
57 y_harmonic, y_percussive = librosa.effects.hpss(y)
58
59
60 # Compute chroma features from the harmonic signal
61 chromagram = librosa.feature.chroma_cqt(y=y_harmonic,sr=sr)
62 print('chromagram :\n',chromagram)
63
64 print('Shape={}'.format(chromagram.shape))
```



Python +
- *Librosa*
- *Essentia*