Titre

R. PETIT

MPI
Lycée Carnot

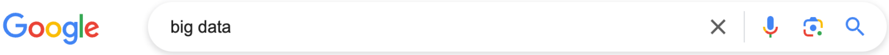
3 février 2025

Le labyrinthe du Big Data

Toute cette mutation en présence d'informations de plus en plus diverses et volumineuses (Big Data) requiert des besoins de mesure nouveaux, tant en termes de bande passante que de performances d'où la nécessité d'appareils de mesure, de développement et d'analyse qui répondent à ces exigences technologiques.

Un random.

En cours de chargement... Veuillez patientez...

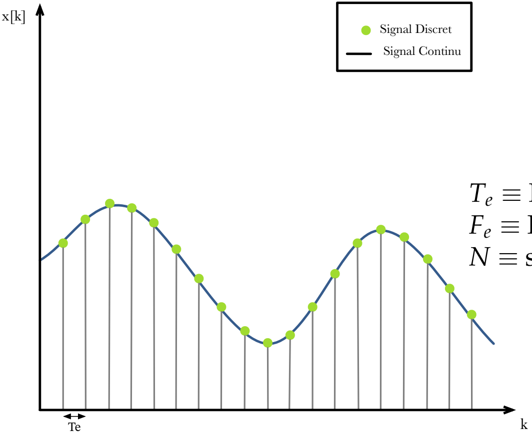


Environ **6 220 000 000** résultats (**0,28** secondes)

Problématique

Comment identifier des signaux numériques de manière efficace ?

Représentation du signal numérique



$T_e \equiv$ Période d'échantillonage
$F_e \equiv$ Fréquence d'échantillonage.
$N \equiv$ série de point composant.

Figure – (Signal discret et signal continu)

Représentation de la collection des musiques

- On modélise la bibliothèque de musique sous forme de table de hashage
- On note $\mathcal{C}$ l'ensemble des clés de la table.
- On note $\mathcal{N}$ l'ensemble des éléments contenu dans le tableau
- Chaque musique est représenté par un couple (cle, nom) avec $cle \in \mathcal{C}$ et $nom \in \mathcal{N}$.
- *nom* correspond au nom de la musique
- *cle* correspond au *fingerprint* de la musique. [1]
- On note S le signal numérique associé à une musique donné.
- On note $f : S \to C$ la fonction de hachage associant à un signal numérique donné, son *fingerprint*

1. Le terme *fingerprint* est un anglicisme voulant littéralement dire "empreinte de doigt", celle-ci étant unique pour chaque humain, on utilise ce terme pour parler d'un identifiant unique.

Définition

Définition 1

La *Fast Fourier Transform* abrégé *fft* est un algorithme permettant de déterminer les coefficients de Fourier d'un signal numérique.

Spécification

- Il est récursif
- Il est basé sur le paradigme de "diviser pour régner"
- Sa complexité est en $O\left(n \log\left(n\right)\right)$ avec n le nombre de coefficient de Fourier.
- Entrée : Un signal sous forme de décomposition de fourrier et la liste des racine n-ième de l'unité $\mathbb{U}_n$.
- Sortie : La liste des coefficients de fourrier.

Principe

- Soit $x[k]$ l'amplitude du signal audio dans l'espace temporel.

Principe

- Soit $x[k]$ l'amplitude du signal audio dans l'espace temporel.
- On définit ses coefficients de Fourier $X[n]$ selon la relation suivante avec $n \in [\![0, N-1]\!]$:

$$X[n] = \sum_{k=0}^{N-1} x[k] e^{-2\pi i \frac{kn}{N}}$$

Principe

- Soit $x[k]$ l'amplitude du signal audio dans l'espace temporel.
- On définit ses coefficients de Fourier $X[n]$ selon la relation suivante avec $n \in [\![0, N-1]\!]$:

$$X[n] = \sum_{k=0}^{N-1} x[k] e^{-2\pi i \frac{kn}{N}}$$

On décompose ensuite le polynôme selon le degré paire $(k = 2m)$ et impaire $(k = 2m + 1)$,

Principe

- Soit $x[k]$ l'amplitude du signal audio dans l'espace temporel.
- On définit ses coefficients de Fourier $X[n]$ selon la relation suivante avec $n \in [\![0, N-1]\!]$:

$$X[n] = \sum_{k=0}^{N-1} x[k] e^{-2\pi i \frac{kn}{N}}$$

On décompose ensuite le polynôme selon le degré paire $(k = 2m)$ et impaire $(k = 2m + 1)$,

$$\begin{aligned} X[n] &= \sum_{m=0}^{\lfloor \frac{N}{2} - 1 \rfloor} x[2m] e^{-2\pi i \frac{(2m)n}{N}} + \sum_{m=0}^{\lfloor \frac{N}{2} - 1 \rfloor} x[2m+1] e^{-2\pi i \frac{(2m+1)n}{N}} \\ &= \underbrace{\sum_{m=0}^{\lfloor \frac{N}{2} - 1 \rfloor} x[2m] e^{-2\pi i \frac{(2m)n}{N}}}_{E_n} + \underbrace{e^{-2\pi i \frac{n}{N}} \sum_{m=0}^{\lfloor \frac{N}{2} - 1 \rfloor} x[2m+1] e^{-2\pi i \frac{(2m+1)n}{N}}}_{O_n} \end{aligned}$$

Principe

Ainsi,

$$X[n] = E_n + e^{-\frac{2\pi i}{N}n} O_n$$

$$X\left[n + \frac{N}{2}\right] = E_n - e^{-\frac{2\pi i}{N}n} O_n$$

Les résultats intermédiaires E_n et O_n sont réutilisé. Le temps de calcul est alors réduit.

Principe

Algorithm 1 Algorithme de Cooley-Tukey

function FFT(T,N,s)
  if $N = 1$ **then**
    return $T[0]$
  end if
  partie_paire $\leftarrow$ FFT(T,$\frac{N}{2}$,2s)
  partie_impaire $\leftarrow$ FFT($T[:s]$,$\frac{N}{2}$,2s)
  resultat $\leftarrow$ []
  for i $\leftarrow$ 0 **to** $\frac{N}{2}$ **do**
    t $\leftarrow e^{-2j \cdot \pi \times \frac{k}{N}} \times$ partie_ impaire $[k]$
    resultat$[k]$ $\leftarrow$partie_paire$[k] + t$
    resultat $[k + \frac{N}{2}]$ $\leftarrow$ partie_paire $[k] - t$
  end for
  return resultat
end function

Échantillons

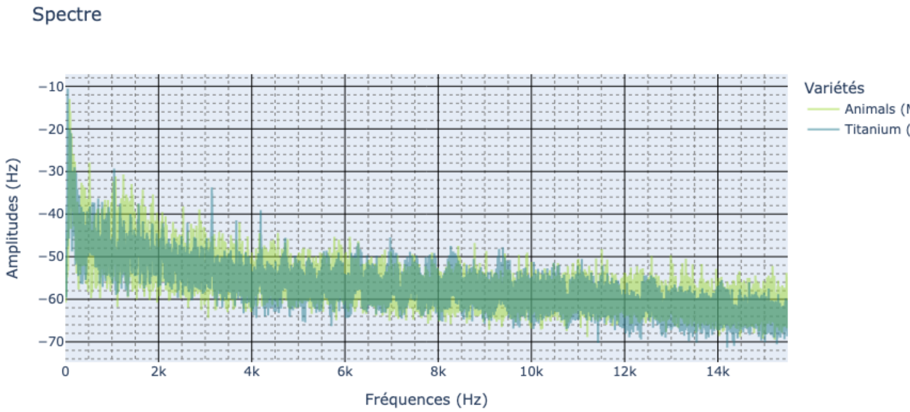
Animals (Martin Garrix)

$\approx$ 15 secondes
65535 points

Titanium (David Guetta)

$\approx$ 15 secondes
65535 points (faire différament)

Résultats


Spectre
Amplitudes (Hz)
−10
−20
−30
−40
−50
−60
−70
0
2k
4k
6k
8k
10k
12k
14k
Fréquences (Hz)
Variétés
Animals (N
Titanium (

Plusieurs problèmes

Figure – svg image

Le risque