CERGY PARIS UNIVERSITÉ

Identification de titres musicaux

Méthode de fingerprinting

Othman EL HOUFI
Mohamed DIAWARA

M1 Ingénierie des Systèmes Complexes et Intelligents

16/06/2021

Rapporteur
Dan VODISLAV

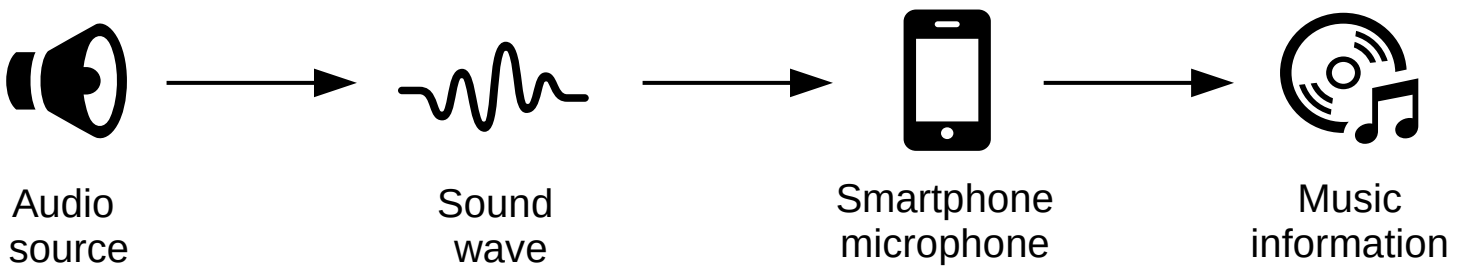
Tuteur technique
Dimitris KOTZINOS

Encadrant de Gestion de Projet
Tianxiao LIU

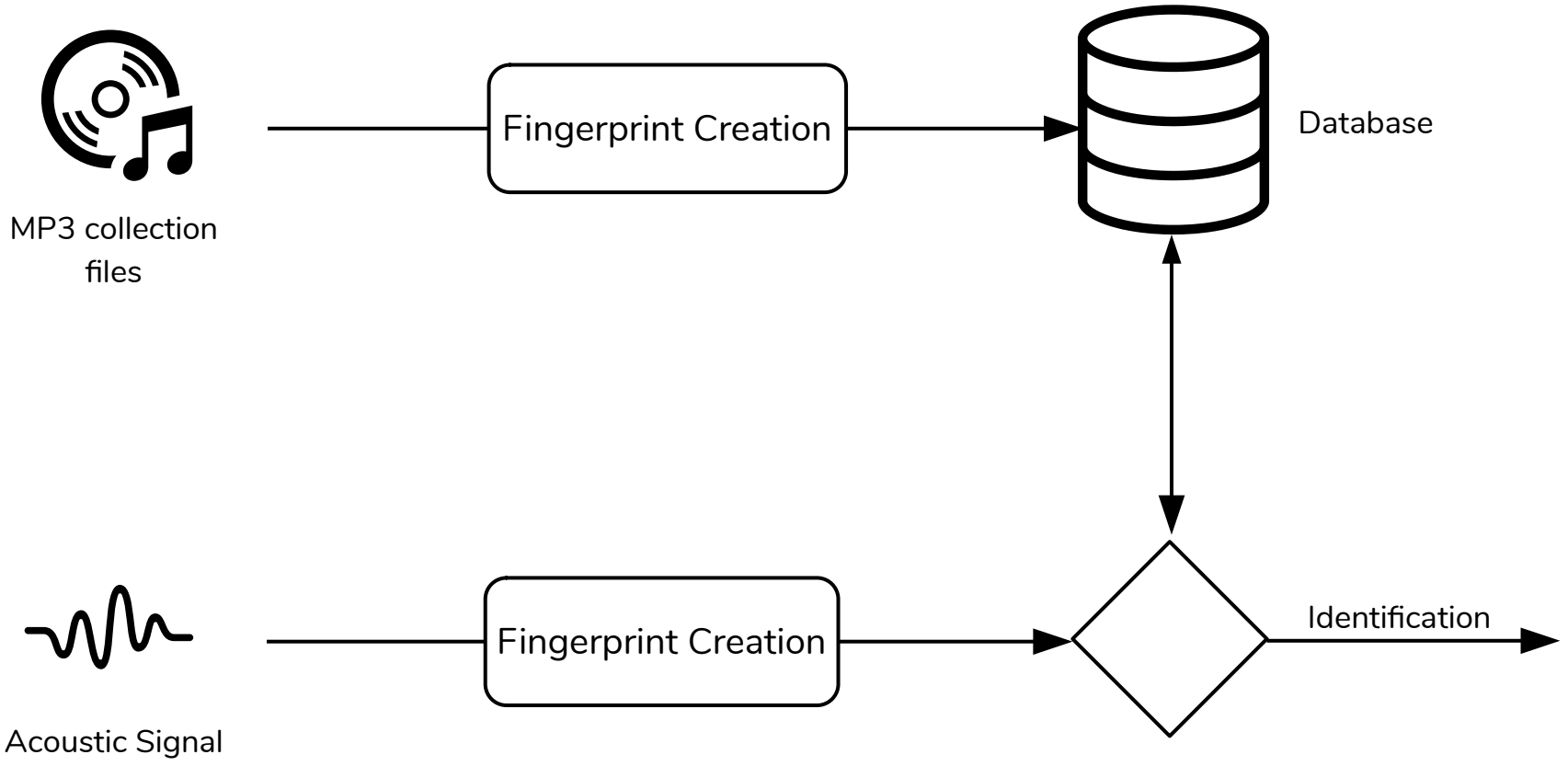
Sommaire

Mise en scénario et objectif du projet

- Identifier des titres musicaux en temps réel en utilisant un microphone.
- Utiliser peu de ressources computationnelles.
- La tâche doit pouvoir être effectuée dans des milieux très perturbés.



Fonctionnement général



Problématiques

Bruit

Les autres sources s'ajoutent linéairement au signal

Désynchronisation

La localisation de l'extrait dans le signal complet

Mémoire et temps de calcul

Recherche rapide et utilisation le minimal de mémoire possible

Besoins techniques

Traitement de Signal

- Choisir la bonne représentation de signal.
- Identifier les information nécessaires dans notre signal.
- Résoudre le problème du bruit et de désynchronisation.

Base de données avancée

- Une solution pour stocker l'information extraite du signal – Une empreinte.
- Recherche réalisable dans un temps raisonnable – millisecondes.
- La mémoire est précieuse !

Traitement du signal acoustique

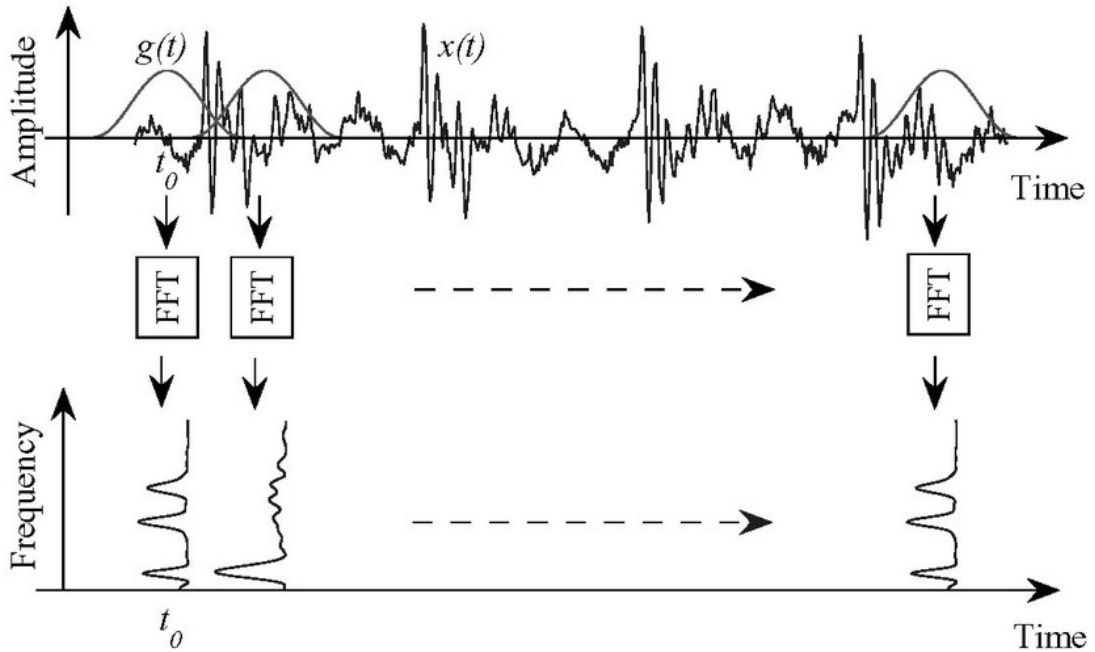
Différentes représentations d'un signal acoustique

Temporelle temps et amplitudes	Fréquentielle fréquences et amplitudes	Spectrogramme temps, fréquences et amplitudes
Simple et directe. Elle est peu robuste au bruit et aux distorsions.	Robuste au bruit. Peu robuste à une désynchronisation car elle ne représente aucune évolution temporelle.	Robuste au bruit. Robuste à la désynchronisation. Très coûteuse en mémoire.

Spectrogramme : création

On utilise une **STFT** (Short-Time Fourrier Transform, transformée de Fourrier à court terme) :

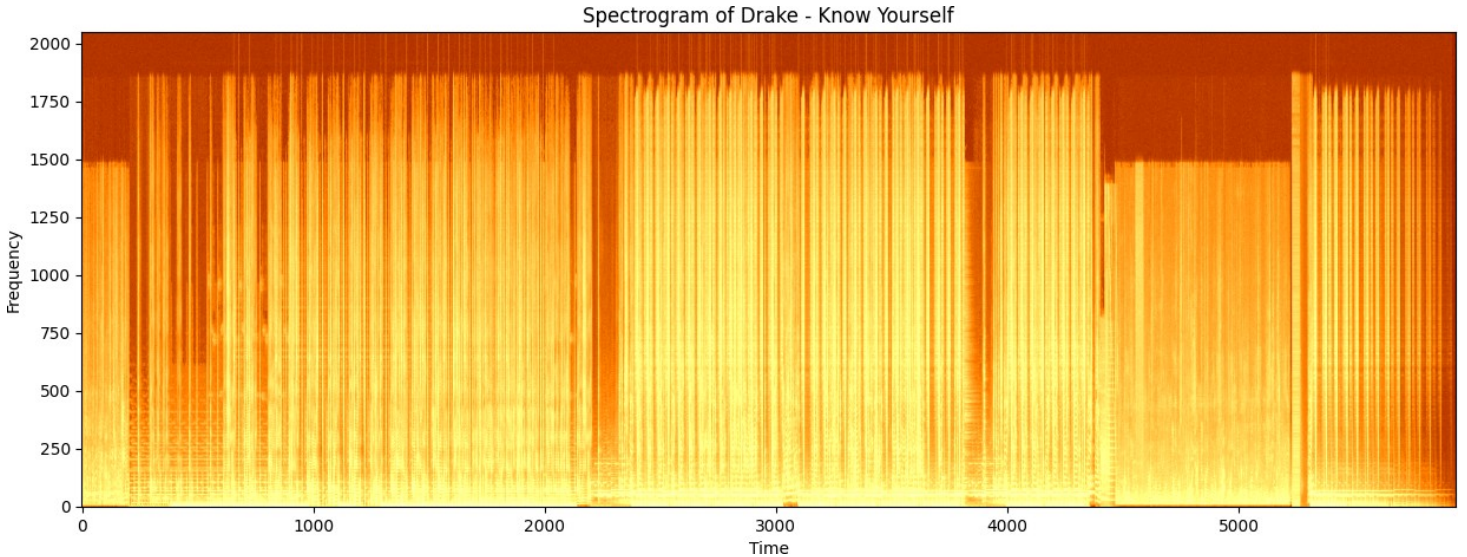
1) analyse le signal par fenêtres temporelles,
2) décompose le spectre localement sur un certain nombre de bandes fréquentielles.



Spectrogramme : création

On utilise une **STFT** (Short-Time Fourrier Transform, transformée de Fourrier à court terme) :

1) analyse le signal par fenêtres temporelles,
2) décompose le spectre localement sur un certain nombre de bandes fréquentielles.



exemple de Spectrogramme de la chanson « Drake – Know Yourself »

WINWOW_SIZE = 4069 OVERLAP_RATIO = 0.5

Spectrogramme : conclusion

Un Spectrogramme contient un nombre immense de points (informations) ce qui pose problème du point de vue de mémoire et vitesse de recherche.

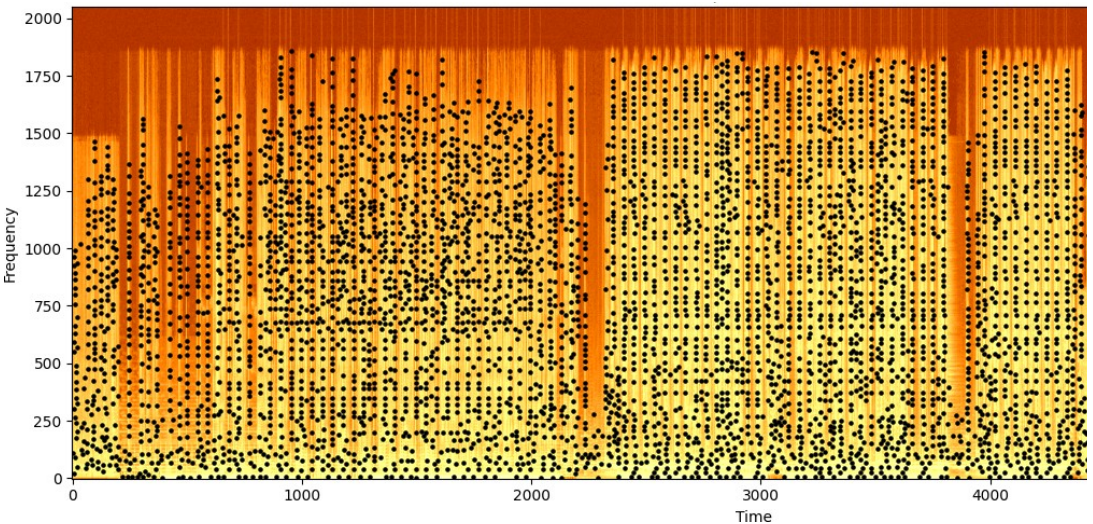
On stocke tout dans la base de données?

Extraction des pics spectraux : procédure

Un pic spectral : une paire (temps, fréquence) avec une amplitude localement supérieure à ses voisins.

➔ Donc résistible aux bruits.

1) Considérer le Spectrogramme comme une image,
2) Appliquer un filtre pour trouver les maximas locaux (amplitudes),
3) Garder les amplitudes supérieures à un seuil donné,
4) Retourner seulement les paires (temps, fréquences).



Constellation de la chanson « Drake – Know Yourself »

AMP_MIN = 10 **NEIGHBORHOOD_SIZE = 10**

Extraction des pics spectraux : résultats

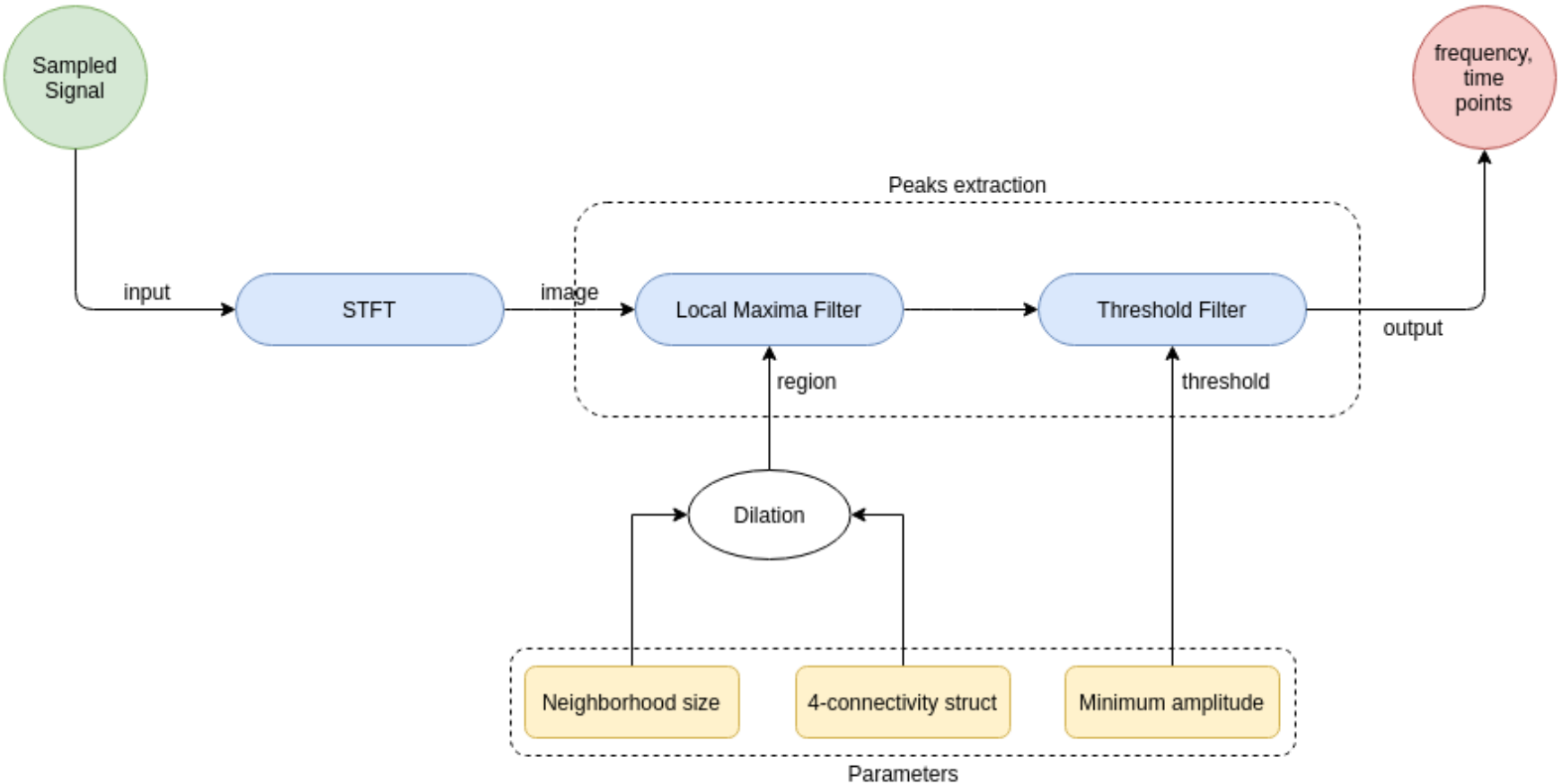
Différence de taille finale et de temps de traitement entre les deux opérations :
Spectrogramme et **Extraction des pics spectraux**

Toujours pour la même musique Drake – Know Yourself

operaton	execution time (sec)	output size
spectrogram	0.91859579	12 169 011
peaks extraction	11.18466210	6 128

la réduction de la taille finale est considérablement **bénéfique** car nous avons économisé **99.95%**

Traitement complet du signal

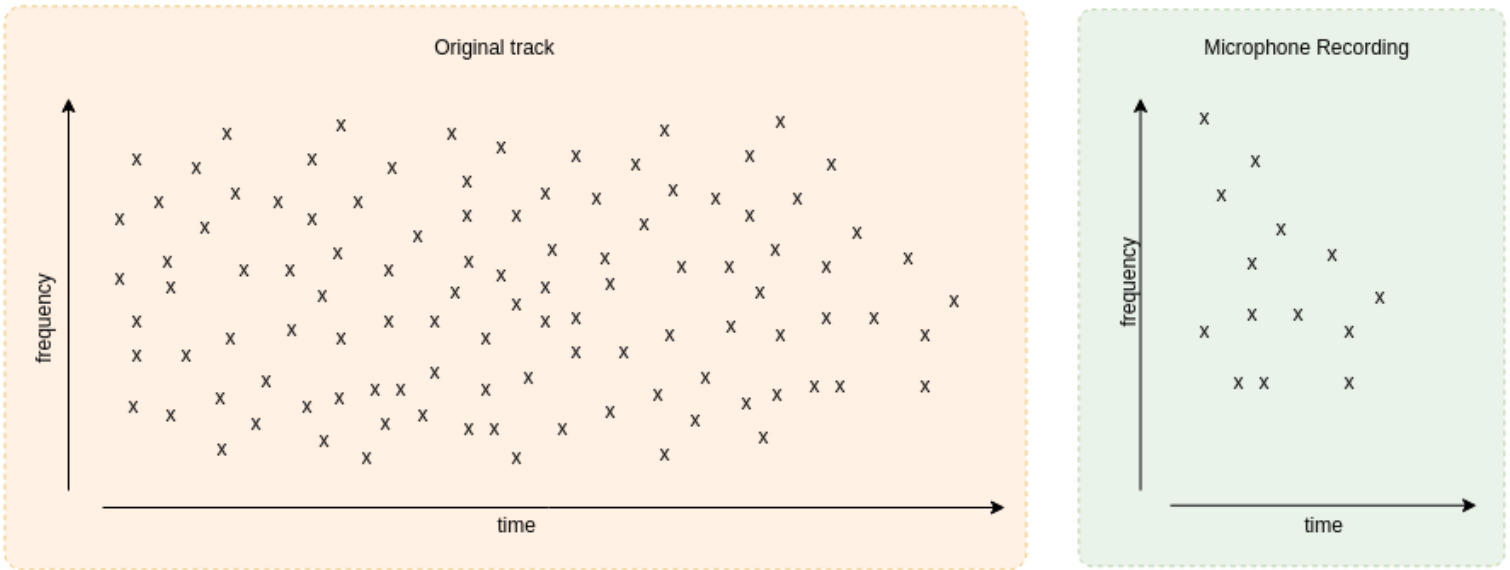


Création d’une Empreinte Acoustique

un résumé numérique qui peut être utilisé pour identifier un échantillon audio

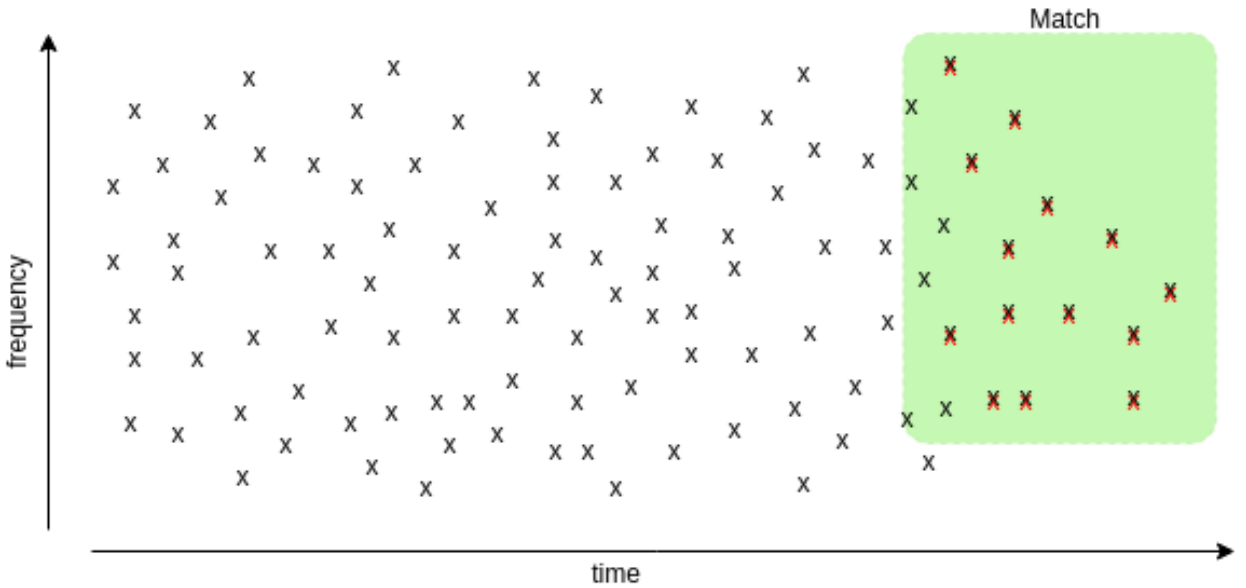
Superposition des constellations

- Les pics spectraux directement enregistrés dans la BDD.
- L'identification consiste sur une comparaison point à point.



Superposition des constellations

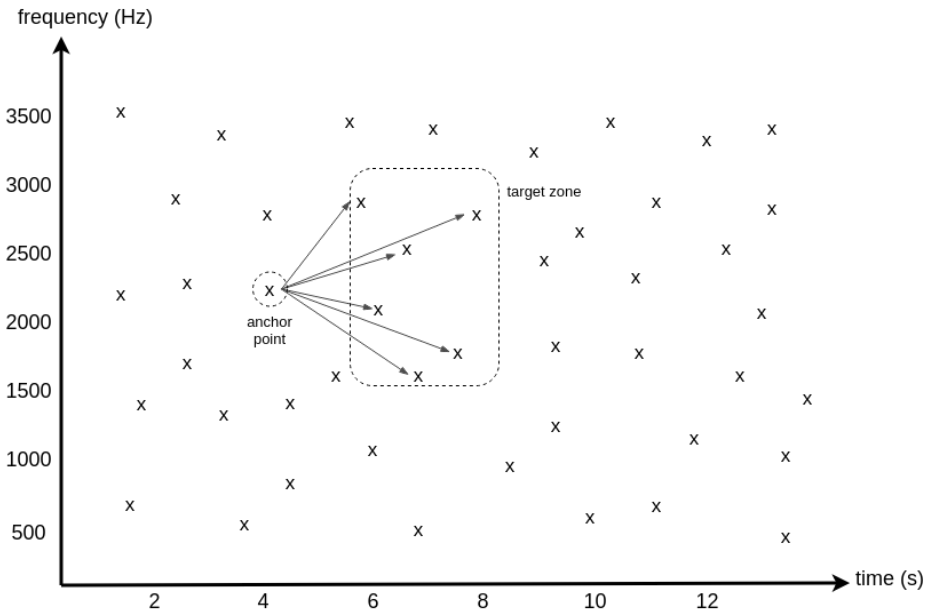
- Les pics spectraux directement enregistrés dans la BDD.
- L'identification consiste sur une comparaison point à point.

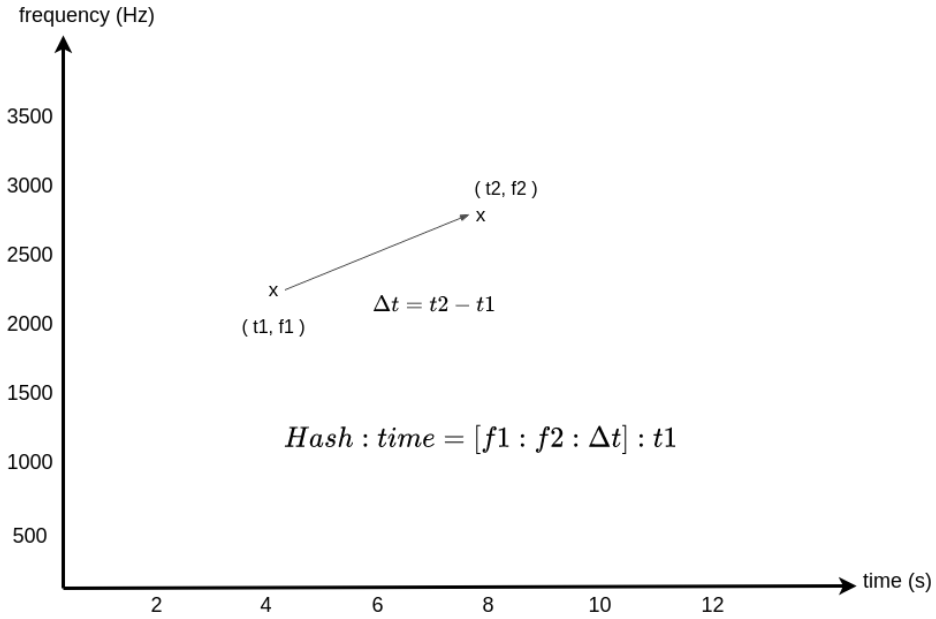


Comparaison individuelle : spécificité faible et recherche lente

Hachage Combinatoire Rapide

1) Choisir un point d'ancrage,
2) Définir une zone cible,
3) Itérer sur les paires (point ancrage, point cible)
4) Retourner le triple (fréquence ancrage, fréquence cible, différence temporelle)





Hachage Combinatoire Rapide

hash(frequencies of peaks, time difference between peaks) = fingerprint hash value

- Reproductibles même avec un bruit ou une compression.
- Chaque hache est associé à un point de début (temps) : pas de désynchronisation.
- On accélère la recherche.

Nous avons ainsi obtenu une Empreinte de notre signal acoustique !

table	number of lines	data size (Mo)	index size (Mo)	total size (Mo)
SONGS	50	0.016	0	0.016
FINGERPRINTS	7 761 966	459.0	289.0	748.0

la taille de mémoire occupée par 50 chansons

Tests et Certifications

Base de données : stockage et recherche

résultats en variant le nombre de secondes enregistré à travers le microphone

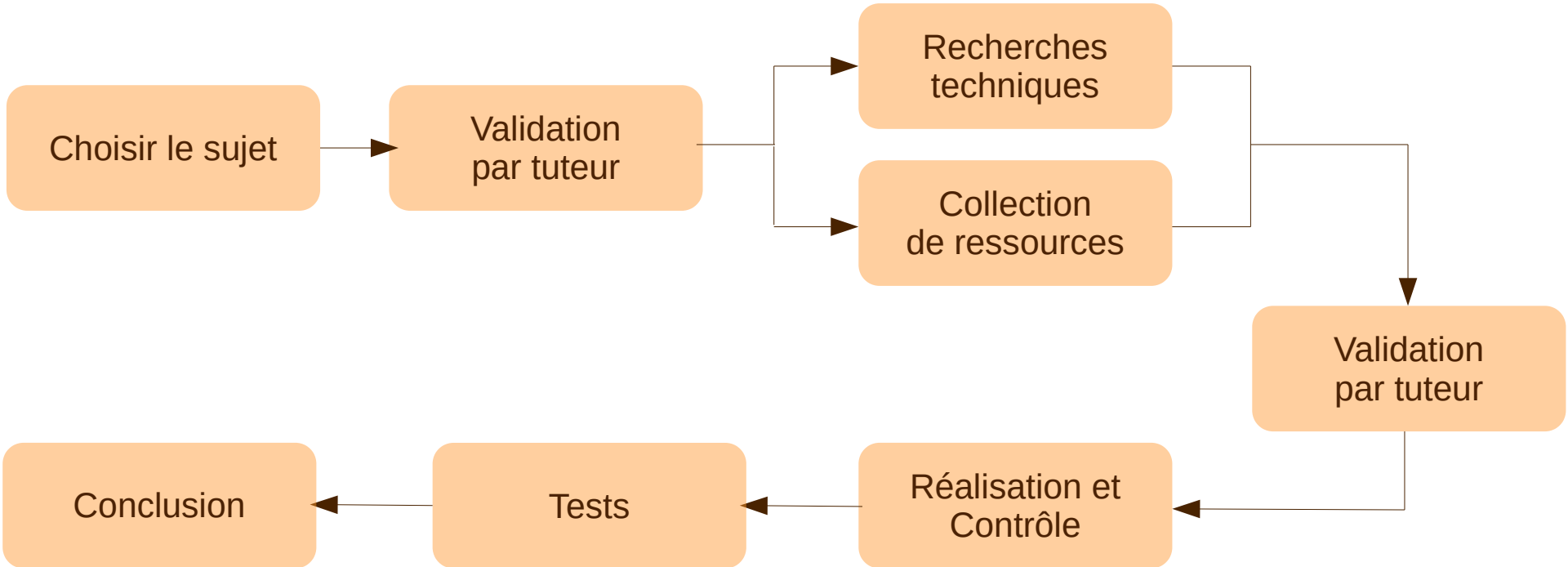
Number of Seconds	Number Correct	Percentage Accuracy	Average Execution Time (sec)
1	37/50	74 %	0.156
2	48/50	96 %	0.268
3	49/50	98 %	0.386
4	50/50	100 %	0.477
5	50/50	100 %	0.578
6	50/50	100 %	0.684

résultats obtenus en rajoutant un bruit réel à travers un extrait 5 sec en mp3

Crowd Noise Loudness (dBFS)	Number Correct	Percentage Accuracy	Average Execution Time (sec)
-17.22	49/50	98 %	0.610
-12.22	48/50	96 %	0.589
-7.49	46/50	92 %	0.571
-4.03	37/50	74 %	0.582
-2.10	25/50	50 %	0.599
-1.11	2/50	4 %	0.599

Gestion
de Projet

Gestion de projet



Répartition des tâches

tâche	réalisée par
Traitement de signal	Othman EL HOUFI
Hachage Combinatoire	Mohamed Diawara
Base de données	EL HOUFI et DIAWARA

Conclusion et Perspectives

Conclusion

De nombreuses techniques de création d'empreintes digitales et d'indexation ont été proposées et sont maintenant utilisées dans des produits commerciaux.

Plusieurs paramètres doivent être ajustés afin de trouver un bon compromis entre les différentes exigences : la robustesse, la spécificité, l'évolutivité et la compacité.

Les aspects importants à traiter/améliorer :

- les paramètres de la STFT
- la stratégie de sélection et d'extraction des pics spectraux
- la taille des zones cibles
- des structures de données appropriées pour le hachage

Perspectives

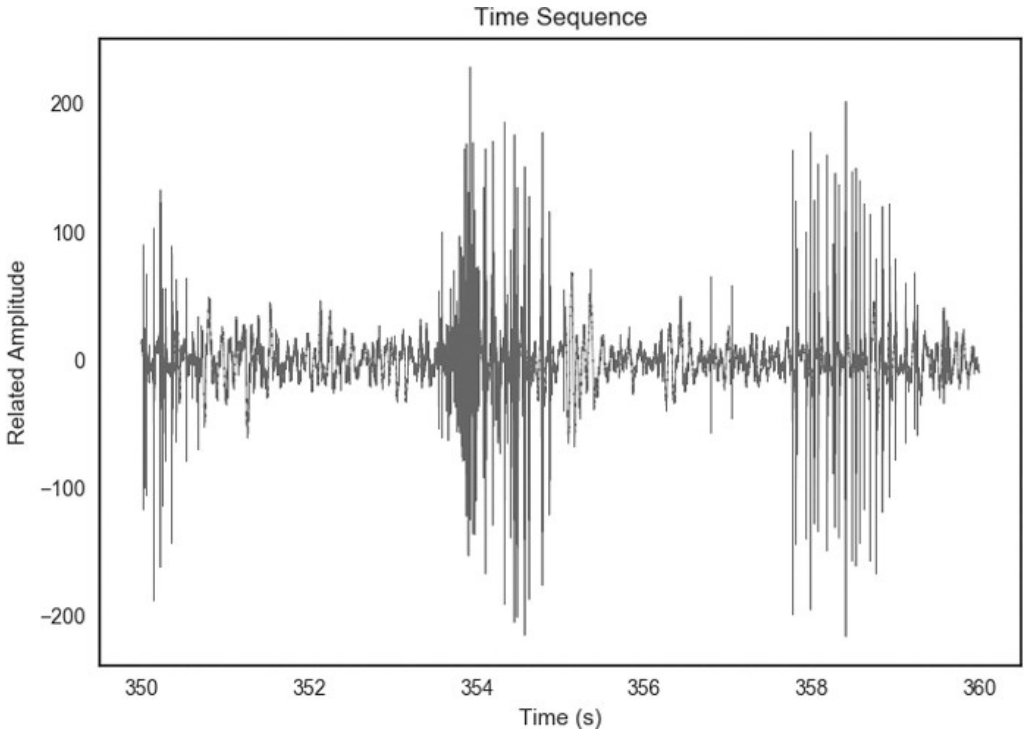
Utilisation d'un modèle de réseau neurones artificielles :

- en entrée les paramètres de notre application,
- en sortie les différentes exigences voulues tel que la robustesse, la mémoire, et le temps de recherche

Entraînement sur une large base qui provienne des tests déjà effectués.

On pourra ajuster les paramètres de notre application dynamiquement par rapport à chaque situation.

Différentes représentations d'un signal acoustique

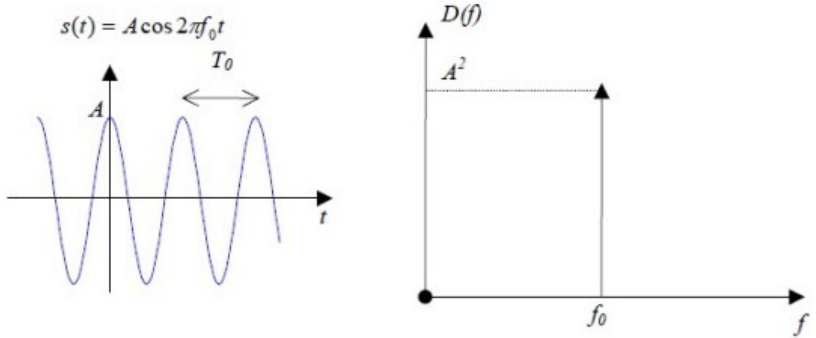


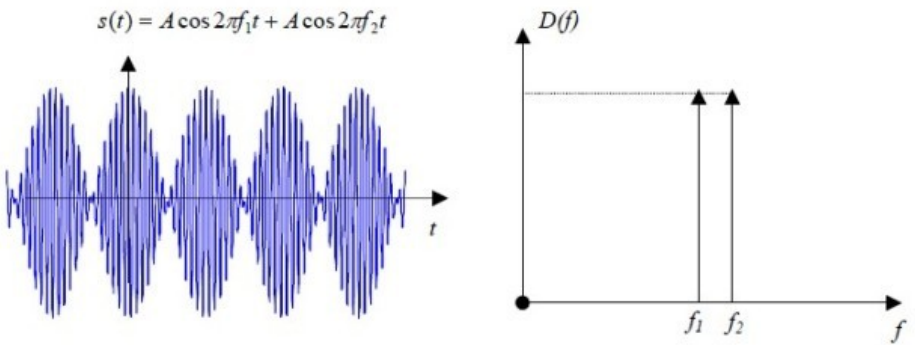
Représentation temporelle

Informations fournies : temps et amplitudes

- Elle est peu robuste au bruit et aux distorsions.

Différentes représentations d'un signal acoustique





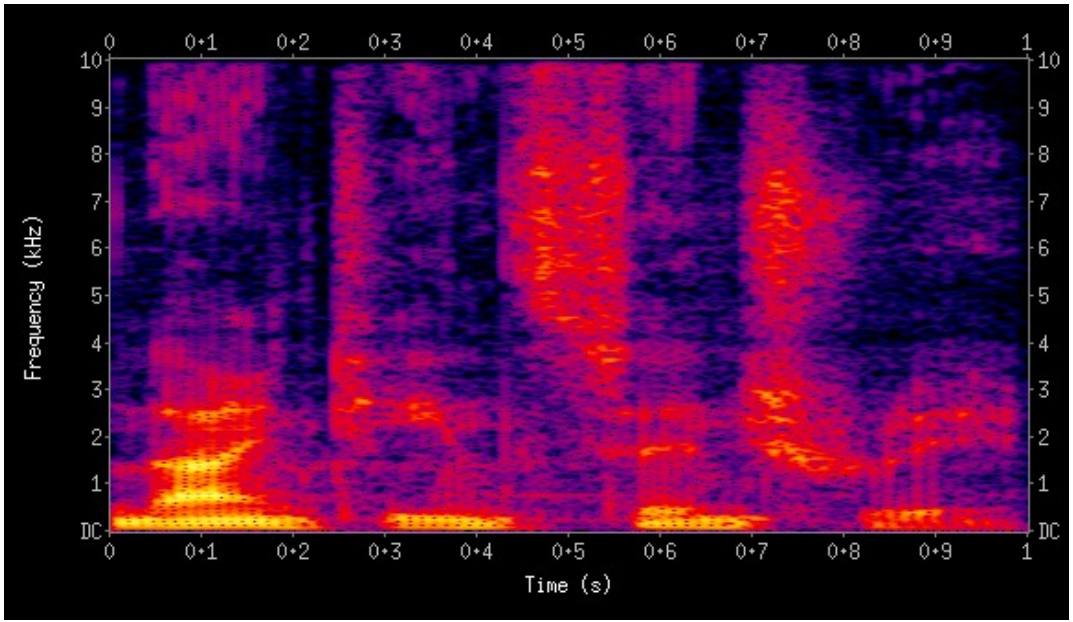
Représentation fréquentielle

Informations fournies : fréquences et amplitudes

- Robuste au bruit.
- Peu robuste à une désynchronisation car elle ne représente aucune évolution temporelle.

$$X(f) = \int_{-\infty}^{\infty} x(t) \times e^{-i2\pi ft}\, dt$$

Différentes représentations d'un signal acoustique



Représentation Spectrogramme

Informations fournies : temps, fréquences et amplitudes

- Robuste au bruit.
- Robuste à la désynchronisation.
- Très coûteuse en mémoire.

Spectrogramme : création

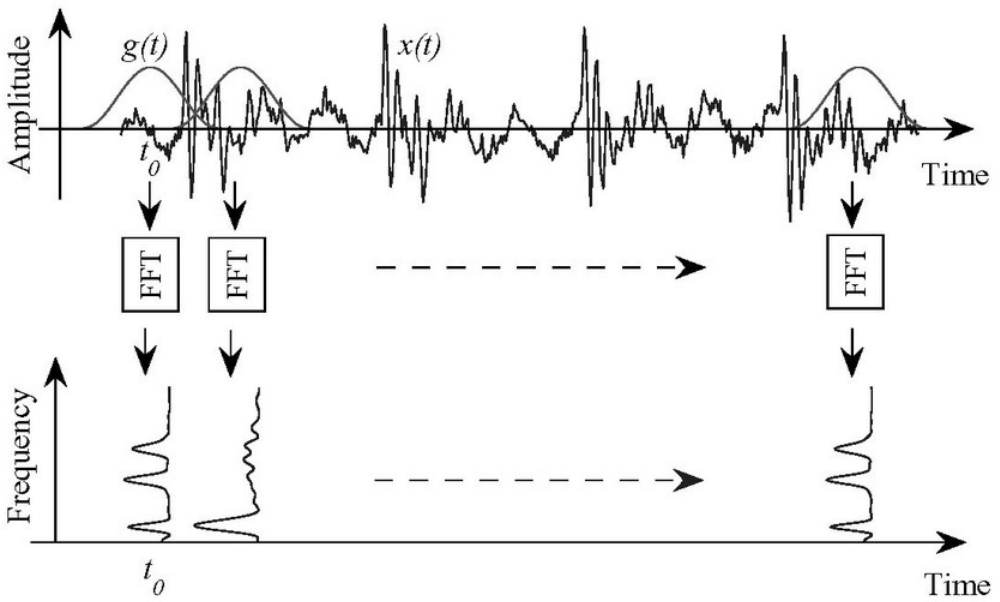
On utilise une **STFT** (Short-Time Fourrier Transform, transformée de Fourrier à court terme) :

1) analyse le signal par fenêtres temporelles,
2) décompose le spectre localement sur un certain nombre de bandes fréquentielles.

$$STFT = X(\tau, \omega) = \int_{-\infty}^{\infty} x(t) w(t-\tau) e^{-i\omega t} dt$$

Paramètres :

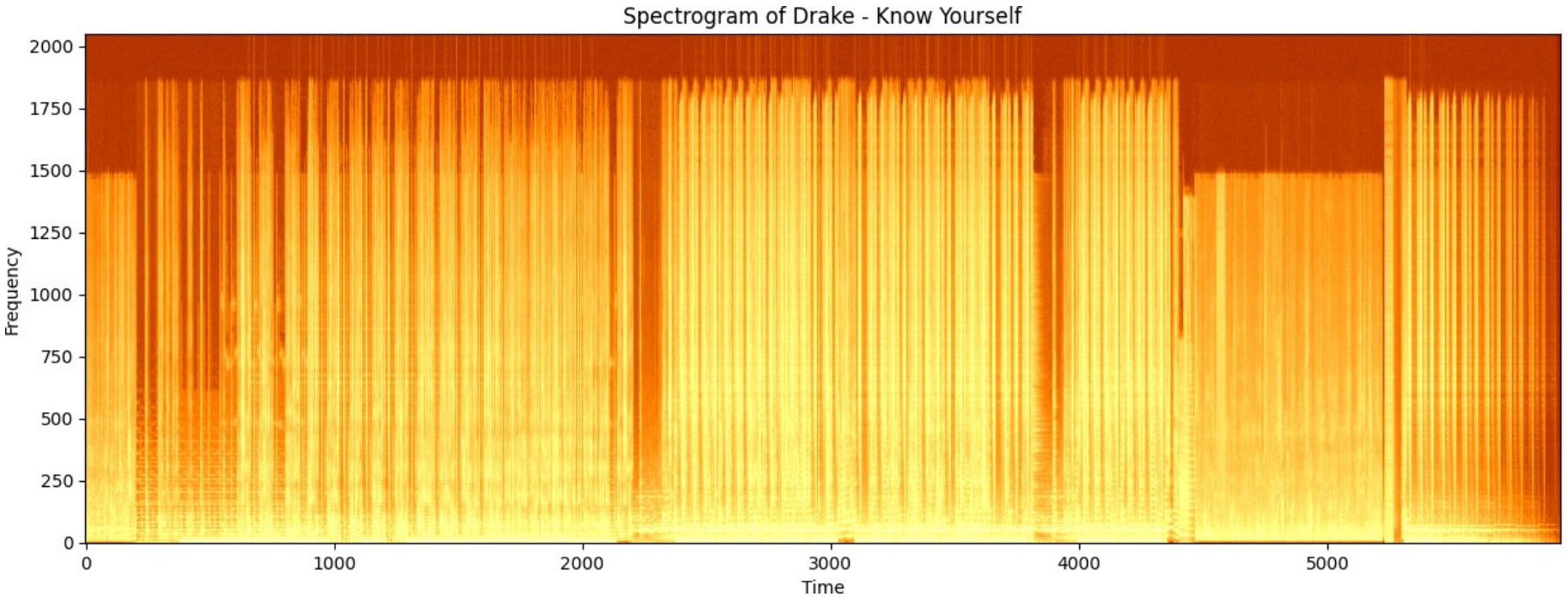
- Taille de la fenêtre temporelle (WINWOW_SIZE)
- Taille du chevauchement des fenêtres (OVERLAP_RATIO)



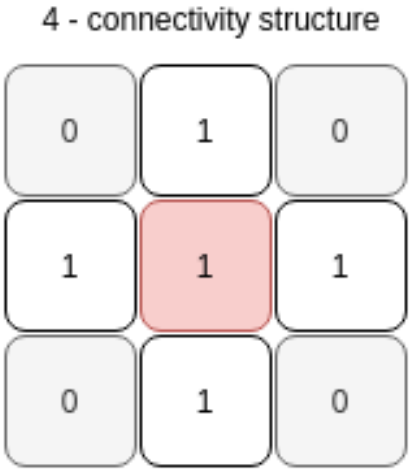
Spectrogramme : exemple

Voici un exemple de Spectrogramme de la chanson « Drake – Know Yourself » tel que :

- WINWOW_SIZE = 4069
- OVERLAP_RATIO = 0.5



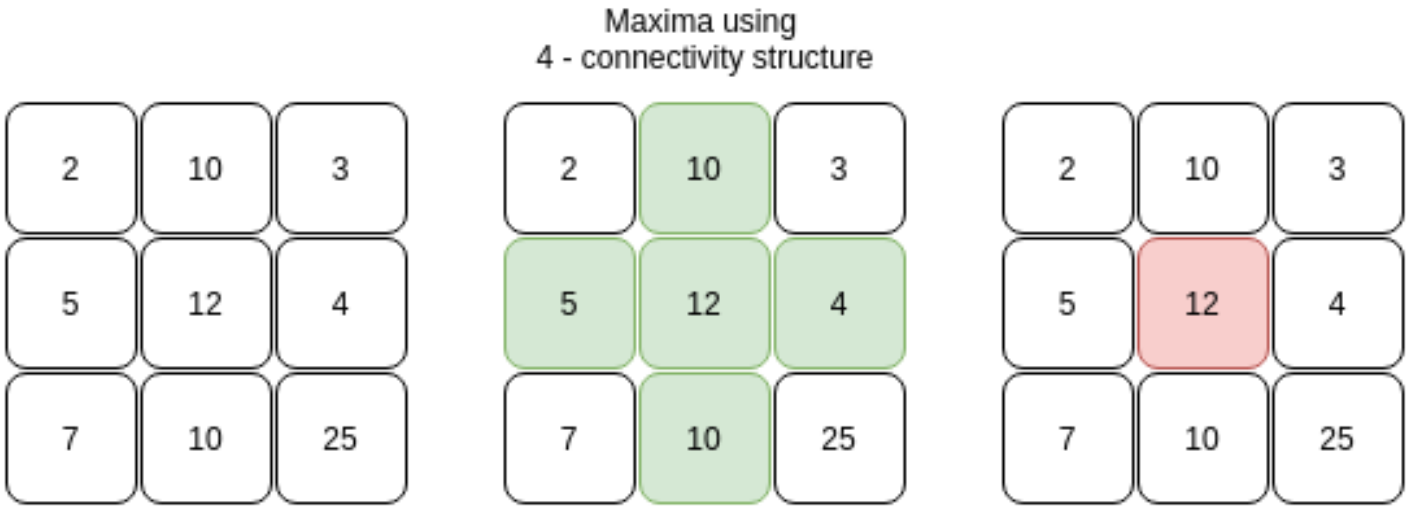
Peaks Extraction : Local Maxima Filter



On choisit d’abord une structure ou *kernel*

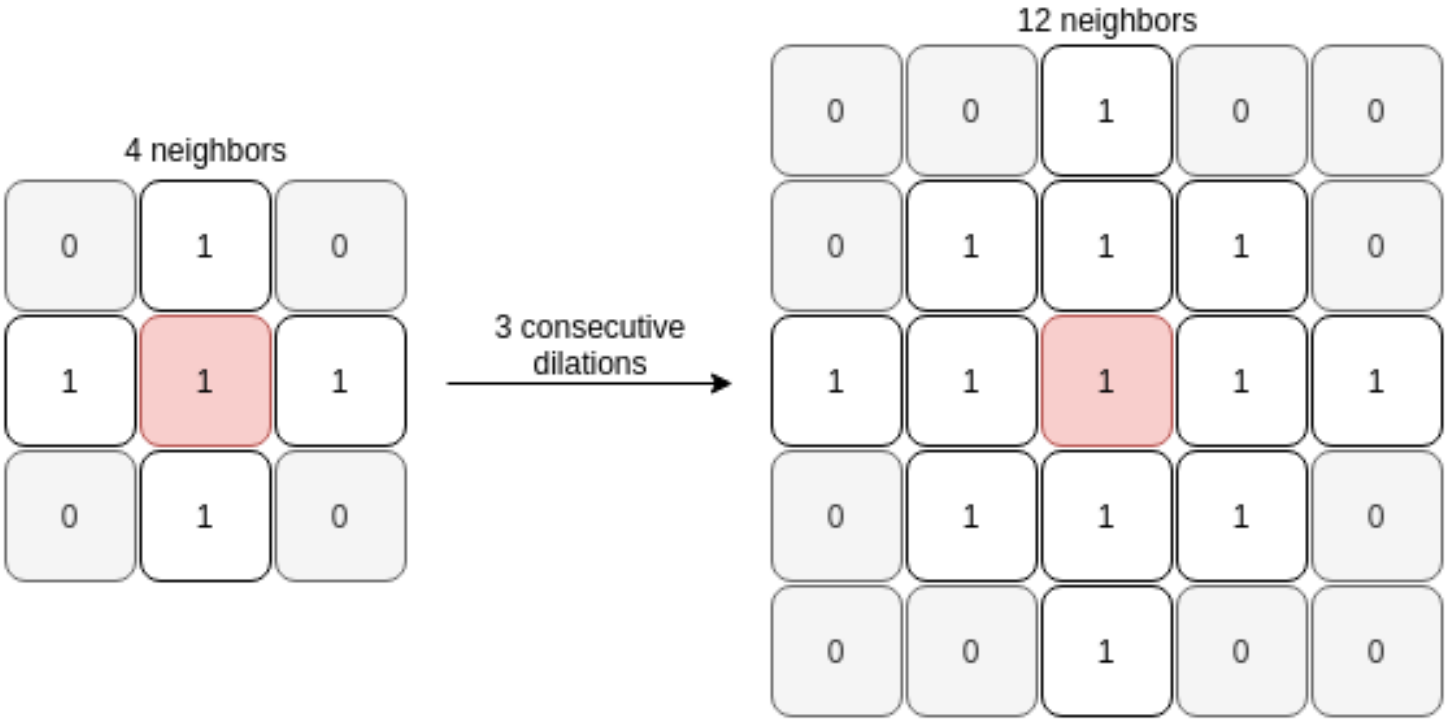
Peaks Extraction : Local Maxima Filter

On définit l'opération associée au maxima local



Ici le noyau doit être plus grand de ces 4 voisins pour qu'il soit considérer comme un maxima local

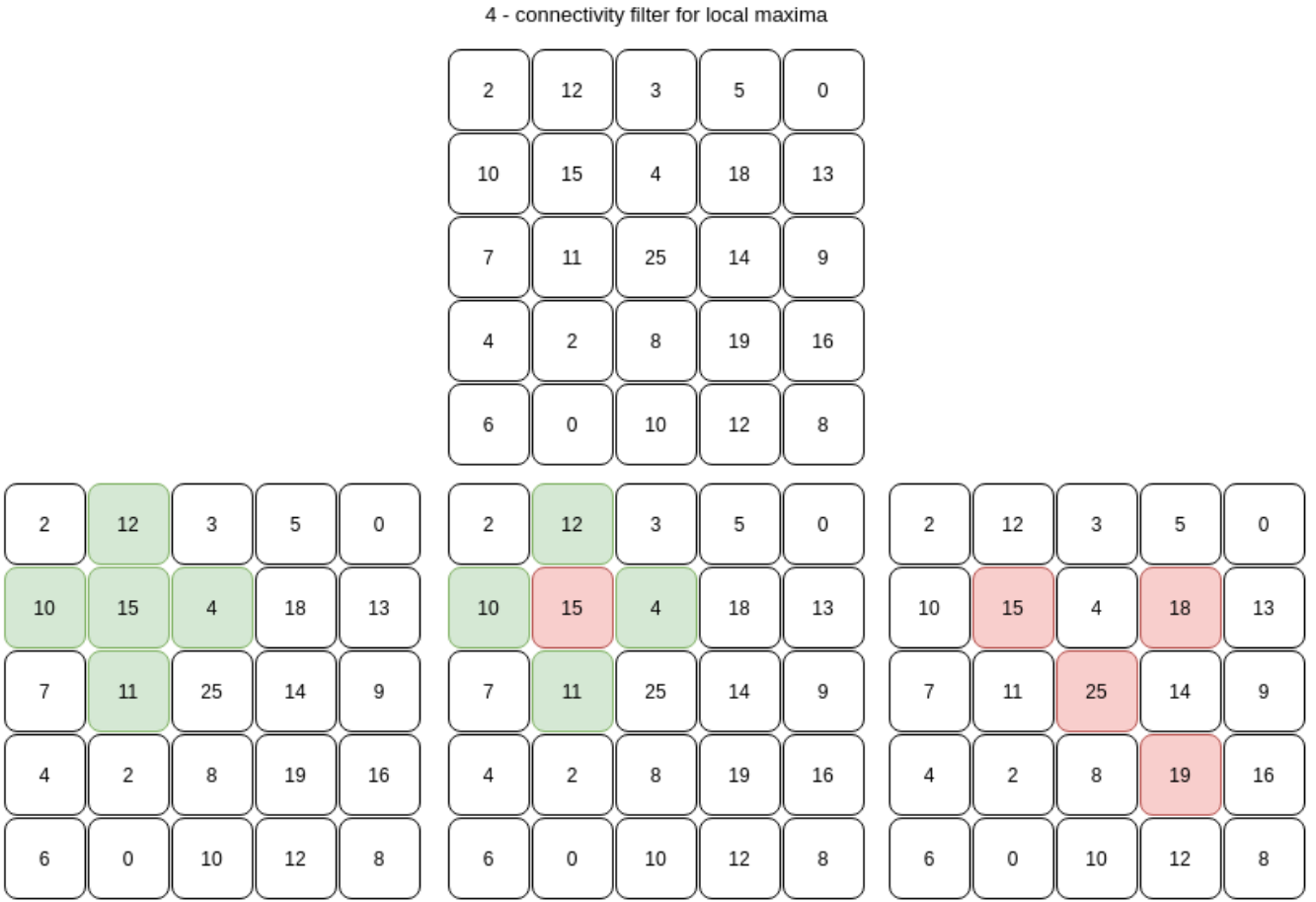
Peaks Extraction : Local Maxima Filter



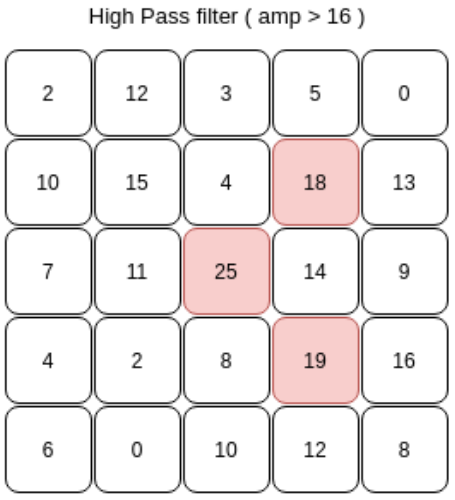
La taille de la région des voisin est **paramétrable**
on peut **étendre** cette région en appliquant une **dilatation morphologique**

Peaks Extraction : Local Maxima Filter

La procédure complète du filtre à maxima local

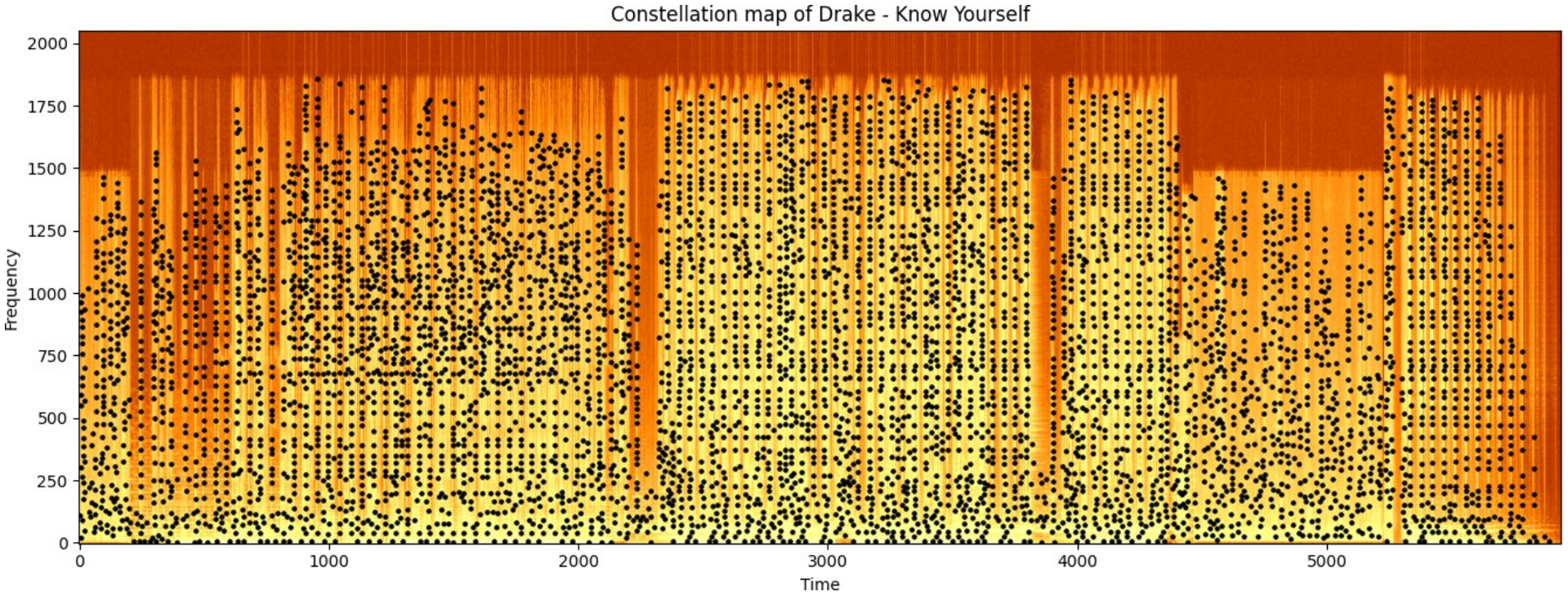


Extraction des pics spectraux (procédure)



Finalement on applique un filtre pour garder des amplitudes qui dépassent un certains **seuil** donné en paramètre.

Extraction des pics spectraux (résultat)



Constellation Map of Drake – Know Yourself