FACULTE DES SCIENCES
DEPARTEMENT DE PHYSIQUE

MASTER EN PHYSIQUE FONDAMENTALE ET APPLICATION

OPTION OPTOELECTRONIQUE ET PHOTONIQUE.

RAPPORT SUR LE PROJET EN TRAITEMENT DE SIGNAL.

TITRE : SEPARATION DE LA VOIX DES INSTRUMENTS DE MUSIQUE DANS UN EXTRAIT AUDIO.

Etudiant : MAGNANGO Malaki
Responsable de l'UE : Prof APEKE

Lomé, le 24 Juin 2025

RAPPORT : SEPARATION DE LA VOIX DES INSTRUMENTS DANS UN EXTRAIT AUDIO

Malaki MAGNANGO

Résumé

Ce rapport présente une méthode de séparation de la voix des instruments de musique à partir d'un signal audio monaural, basée sur la décomposition harmonique/percussive (HPSS Harmonic Percussive Source Separation). L'approche exploite la nature distincte des composants harmoniques (associés généralement à la voix et aux instruments soutenus) et percussifs (associés aux transitoires rapides des instruments) du signal audio. En utilisant la Transformée de Fourier à court Terme (STFT) pour obtenir une représentation temps-fréquence et des filtres médians appliqués sur le spectrogramme, le signal original est décomposé en ses composantes harmoniques et percussives. Les signaux séparés sont ensuite reconstruits via la Transformée de Fourier Inverse à Court Terme (ISTFT). Ce rapport détaille la méthodologie implémentée en Python avec la bibliothèque librosa, les paramètres clés, et discute des résultats et des limitations de cette technique.

1. Introduction

La séparation de source audio, et en particulier la distinction entre la voix chantée et les instruments de musique dans un morceau est un problème fondamental et complexe en traitement de signal. Cette tache connue sous le nom de « Voice/Music Separation ou Vocal Séparation », trouve des applications dans de nombreux domaines tels que le karaoké, le mixage, la transcription musicale automatique, la suppression de voix pour le doublage, ou encore l'analyse et l'indexation de contenu audio.

Historiquement, diverses approches ont été explorées pour résoudre ce problème. Les premières méthodes reposaient sur des techniques de filtrage spectral, d'analyse en composantes

principales (PCA) ou de factorisation matricielle non négative (NMF). Récemment, les avancées dans l'apprentissage profond ont conduit à des solutions très performantes, souvent basées sur des réseaux neuronaux récurrents ou convolutés.

Ce travail se concentre sur une méthode plus classique et computationnellement efficace : la Décomposition Harmonique-Percussive (HPSS). Cette technique repose sur l'hypothèse que la voix et les instruments peuvent être largement caractérisés par leurs propriétés harmoniques (variations lentes en fréquence) et percussives (transitoires et rapides). L'objectif de ce projet est de décrire en détail l'implémentation de cette méthode pour la séparation de la voix et des instruments, d'analyser son fonctionnement et d'évaluer les performances et les limites.

Pour la réalisation de notre projet, nous avons écrit un script afin de séparer automatiquement un signal audio en deux composantes :
- La voix correspondant à la partie harmonique du signal ;L
- Les instruments qui correspondent à la partie percussive.

L'approche utilisé repose entièrement sur l'analyse temps-fréquence du signal via la Transformée de Fourier à court Terme (STFT), suivi d'une décomposition harmonique/percussive (HPSS) fournie par la bibliothèque librosa.

2. Structure du code

Le code est structuré en plusieurs blocs fonctionnels clairement définis :

a. paramètres globaux

FILENAME = "téléchargement (2).wav"
DUREE_AUDIO = 10
N_FFT = 2048
HOP_LENGTH = 512

Ces variables permettent de personnaliser :
- Le fichier audio à traiter ;
- La durée d'analyse (en secondes) ;
- La résolution de la STFTr (n_fft et hop_length).

b. Chargement du signal

Chargement du signal audio : Le signal audio est chargé depuis un fichier (.wav) à l'aide de librosa.load ; nous avons défini une durée spécifique pour charger un segment du fichier.

y,sr=librosa. load(fichier, duration=duree_audio)

Ou y est le signal audio et sr la fréquence d'échantillonnage.

c. Analyse Spectral (STFT + HPSS)

D = librosa.stft(y, n_fft=n_fft, hop_length=hop_length)

S_full, phase = librosa.magphase (D)

S_harmonic, S_percussive = librosa.decompose.hpss (S_full)

- librosa.stft calcul la STFT du signal.
- librosa.magphase sépare magnitude et phase.
- librosa.decompose.hpss applique le filtrage médian pour obtenir les composantes harmoniques et percussives.

Calcul de la STFT : Dans cette partie, nous avons effectué un calcul STFT où le signal temporel y est transformé en sa représentation temps-fréquence D (spectrogramme complexe) en utilisant `librosa.stft()`.

```
D = librosa.stft(y, n_fft=n_fft, hop_length=hop_length)
```

Extraction de la Magnitude et de la Phase: Le spectrogramme complexe D est séparé en sa magnitude S_full et sa phase. La magnitude contient l'information sur l'énergie des fréquences, tandis que la phase est essentielle pour la reconstruction correcte du signal temporel.

S_full, phase = librosa.magphase(D)

S_harmonic,S_percussive = librosa.decompose.hpss(S_full)
Application de la HPSS: La fonction librosa.decompose.hpss () est appliqué à la magnitude du spectrogramme S_full. Ce qui renvoie deux (02) matrices : S_harmonic (magnitude de la composante harmonique) et S_percussive (magnitude de la composante percussive).

S_harmonic, S_percussive =librosa.decompose.hpss(S_full)

d. Reconstruction Temporelle

y_harm = librosa.istft(S_harmonic * phase)
y_perc = librosa.istft(S_percussive * phase)

Pour reconstruire les signaux harmoniques et percussifs dans le domaine temporel, les magnitudes séparées (S_harmonic et S_percussive) sont multipliées par la phase originale du signal. Ensuite nous avons appliqué la Transformée de Fourier Inverse à Court Terme (ISTFT) est appliquée à chaque spectrogramme complexe résultant pour obtenir les signaux temporels y_harmonic et y_percussive.

y_harm = librosa.istft(S_harmonic * phase)

y_perc = librosa.istft(S_percussive * phase)

e. Affichage Temporel et Spectral

Deux fonctions nous a permis d'afficher :
- Les formes d'ondes (temporel) ;
- Les spectrogrammes en décibel (fréquentiel).

f. Lecture audio

Le module sounddevice est utilisé pour jouer les signaux directement dans l'ordinateur.

g. Sauvegarde

sf.write("voix.wav", normaliser(y_harmonic), sr)
sf.write("instruments.wav", normaliser(y_percussive), sr)

Les signaux reconstruits sont normalisés pour éviter l'écrêtage et assurer une lecture audio confortable. Ils sont ensuite sauvegardés sous forme de fichier .wav et leurs spectrogrammes sont affichés pour une analyse visuelle.

3. Environnement de développement

Le code a été développé en Python 3, en utilisant les bibliothèques principales suivantes :

- **`librosa`** : Pour le chargement audio, la STFT, l'ISTFT et la décomposition HPSS ;
- **`numpy`** : Pour les calculs numériques ;
- **`sounddevice`** : Pour la lecture audio des signaux séparés ;
- **`soundfile`** : Pour la sauvegarde des signaux audio en fichier .wav ;
- **`matplotlib`** : Pour la visualisation des signaux temporels et des spectrogrammes.

4. Résultats et Discussions

L'évaluation des résultats de la séparation voix-instruments a été réalisé à la fois par une analyse quantitative des propriétés de signal et par une inspection qualitative des spectrogrammes et une écoute subjective des pistes séparées. L'affichage graphique montre clairement la distinction entre les spectres :

- Le spectre complet présentant une superposition de toutes les composantes ;
- Le spectre harmonique est constitué de lignes stables dans le temps, typiques de fréquences vocales ;
- Le spectre percussif montre des structures verticales brèves, associées aux impulsions rythmiques.

Les formes d'ondes temporelles révèlent également les différences de densité et de régularité entre les signaux séparés.

La méthode de décomposition HPSS bien que robuste présente des limites importantes en raison de ses hypothèses simplificatrices. Notamment dans sa supposition que la voix est uniquement harmonique et les instruments uniquement percussifs. Or de nombreux instruments ont des composantes harmoniques fortes et la voix elle peut contenir des éléments percussifs. En cas de chevauchement important entre la voix et les instruments, la HPSS peinera à les séparer distinctement.

5.Conclusion

Ce projet a démontré la mise en œuvre d'une technique de séparation de la voix et des instruments de musique basée sur la décomposition harmonique –percussive (HPSS) à l'aide de la bibliothèque librosa en Python. L'approche utilisée est relativement simple à comprendre et à implémenter, et elle fournit une première séparation utile des composantes harmoniques(tendant à contenir la voix) et percussives (tendant à contenir les instruments rythmiques) d'un signal audio. Les visualisations des spectrogrammes et l'écoute subjective confirment la capacité de la HPSS à distinguer ces deux grandes catégories de sons.