

Modifier la voix d'un enregistrement

Antonin CHIRON, Valentin HAMELIN, Jules TOUPIN

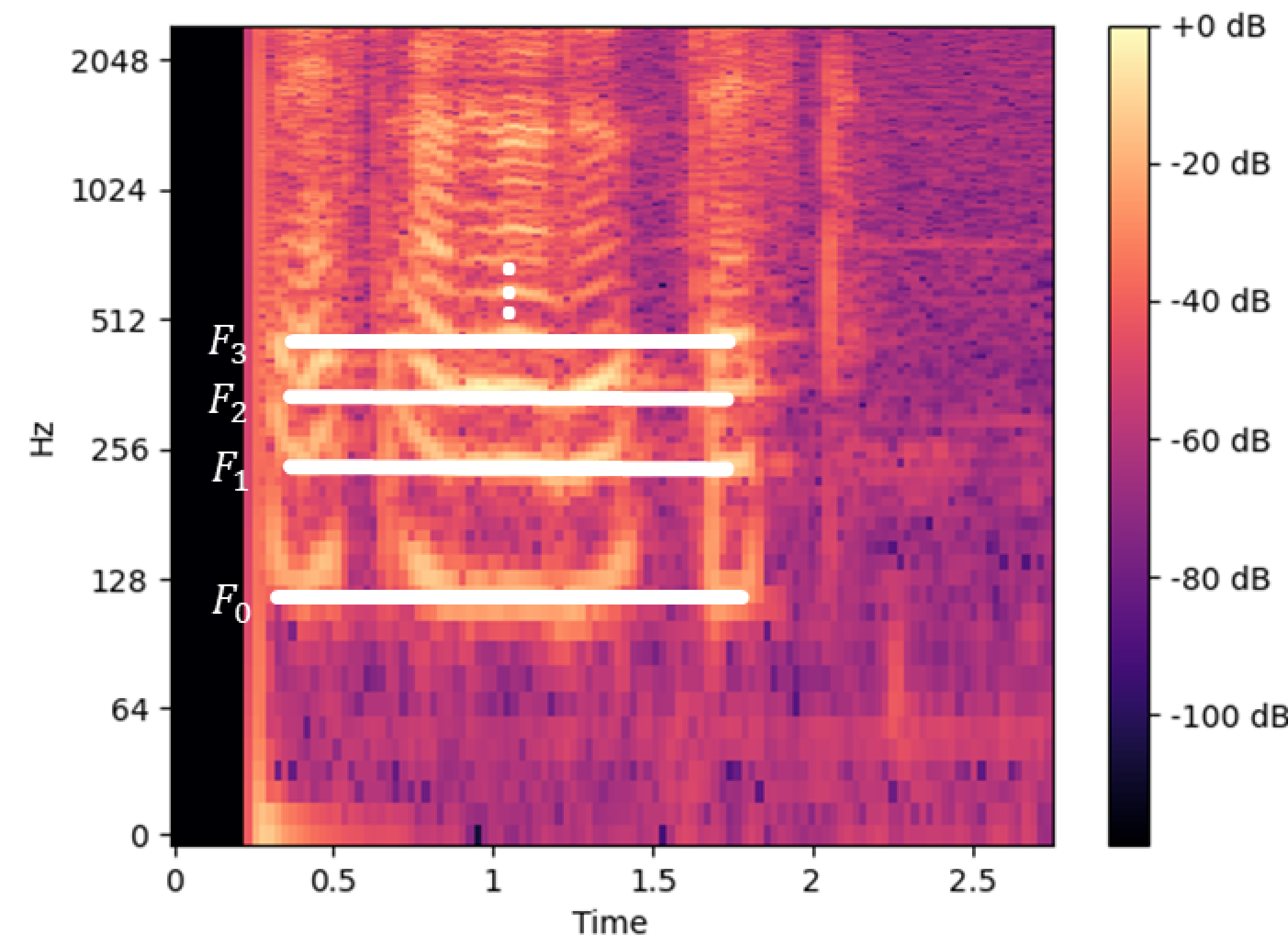
Caractéristique fréquentielle de la voix humaine

Avant de vouloir modifier la voix, intéressons nous d'abord à son traitement.

La voix humaine produit un fondamental (F_0) entre 80Hz et 400Hz, cette fréquence dépend de l'âge et du sexe de la personne. Comme n'importe quel instrument la voix créer aussi plusieurs harmoniques (F_1 , F_2 , ...).

Le timbre de la voix va justement être caractérisé par la fréquence du fondamentale et de ses harmoniques.

On peut notamment retrouver le fondamental et les harmoniques sur une représentation temps fréquence par des lignes lumineuses représenter ci-dessous.



ici, et pour le reste de l'étude, la graduation de l'échelle des intensités sonores a été mise arbitrairement et ne représente en rien une mesure représentative ou des ordres de grandeurs des intensités sonores mise en jeu dans l'enregistrement.

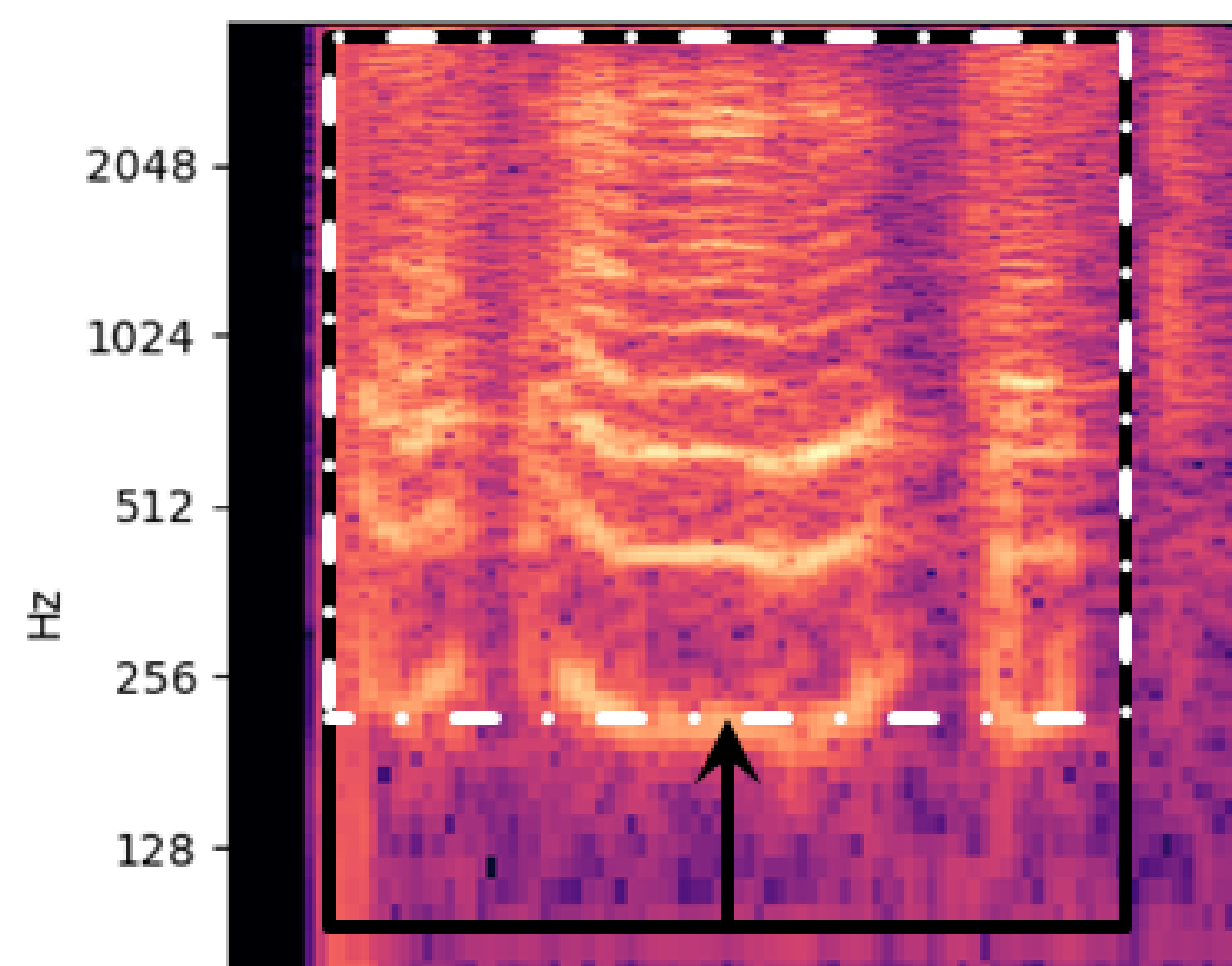
La figure ci-dessus représente le diagramme temps fréquence d'un enregistrement avec une voix et sans (sans que ce soit significatif) autre son enregistrer.

Modification du timbre de la voix

Pour réaliser les prochaines modifications nous allons utiliser la bibliothèque Librosa sur python qui va nous faciliter la manipulation de l'enregistrement dans le domaine fréquentiel. On utilisera en particulier la commande : `librosa.effects.pitch_shift`.

Rendre la voix plus aigu :

Plus la fréquence d'un son est élevée plus le son est aigu. On pourrait isoler le fondamental avec un algorithme de détection de F_0 et se limiter à le translaté seulement lui dans les aigus mais le résultat donne une voix très synthétisée. Pour réussir à obtenir un résultat convenable on va plutôt multiplier toutes les fréquences par un facteur plus ou moins important. Graphiquement cela ressemble à une compression du spectre vers les hautes fréquences :

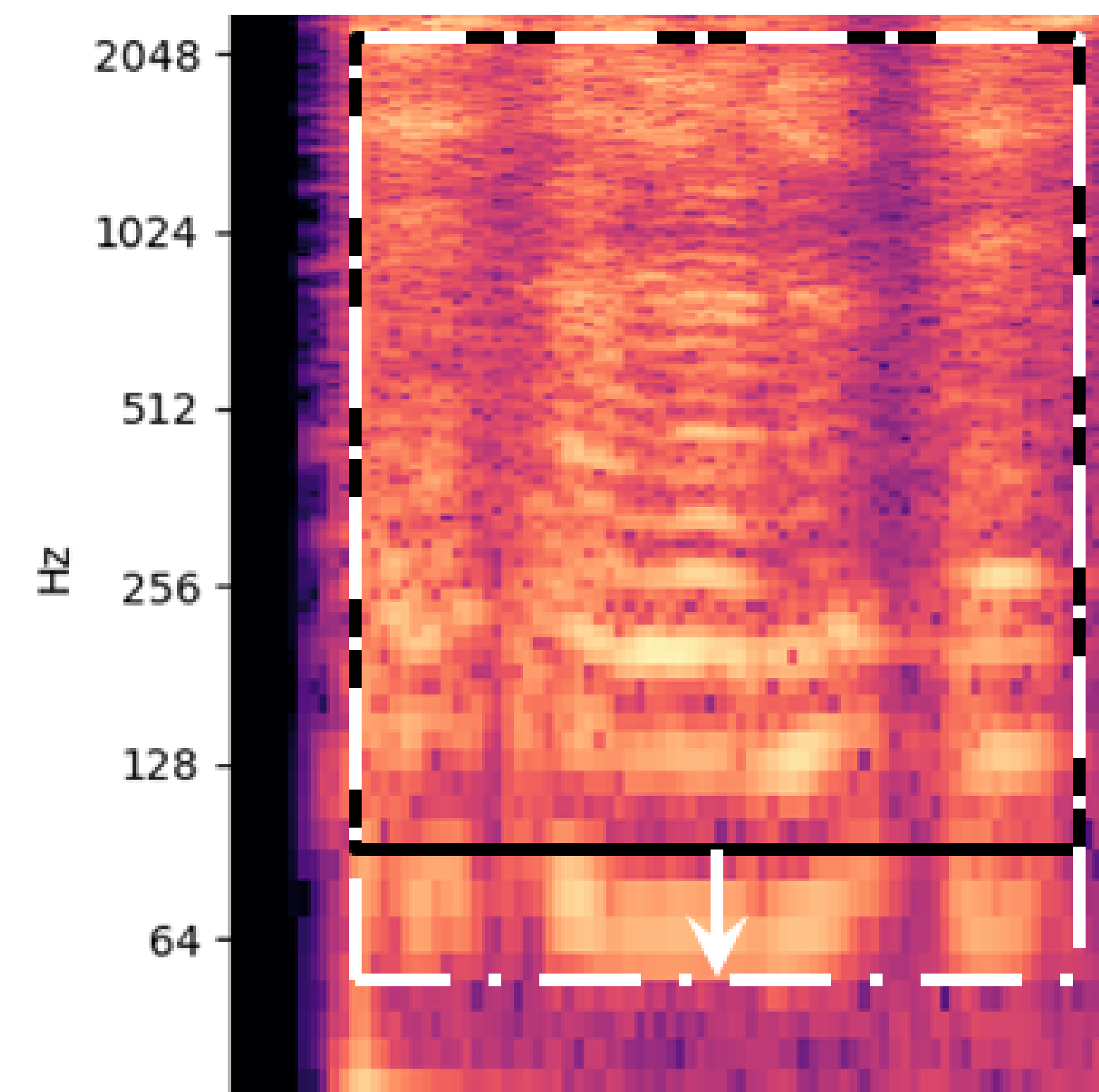


Cette méthode a l'avantage de déplacer tous les harmoniques avec le fondamental ce qui rend le résultat plus humain et moins synthétique à partir du moment que l'on décide d'être raisonnable sur la modification de la voix.

Attention : avec cette méthode c'est tous les sons de l'enregistrement qui sont modifiés et pas seulement la voix ! Ce qui veut dire que chaque son audible sur l'enregistrement seront plus aigus.

Rendre la voix plus grave :

De la même manière que pour rendre la voix plus aigu, on va multiplier chaque fréquence par un facteur variable afin de les rendre plus basse. Graphiquement on peut assister à une extension du spectre vers les basses fréquences :

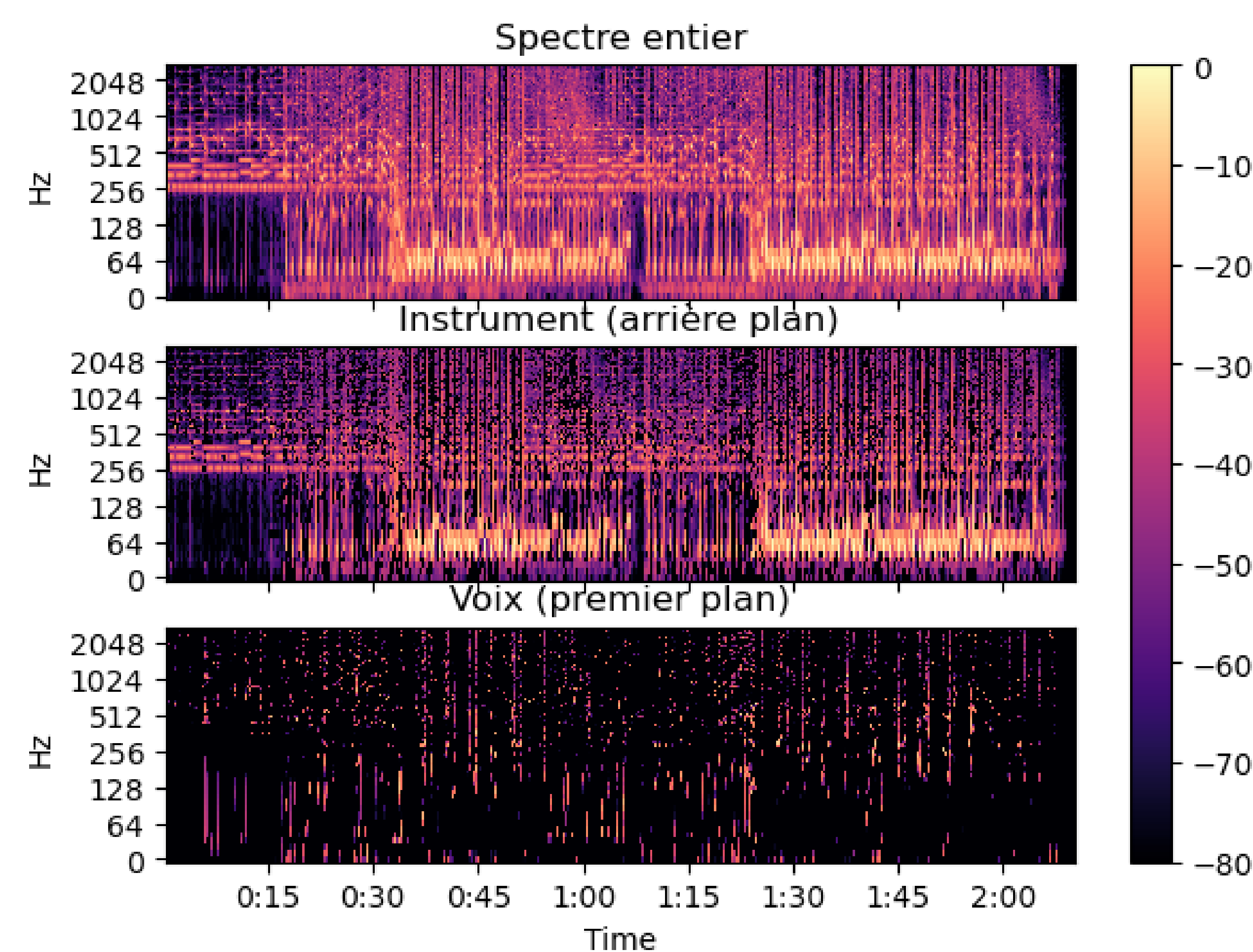


Attention : cette méthode étant la même que pour rendre la voix plus aigu, seul le résultat change, elle comporte les mêmes défauts !

Exemple d'utilisation sur une musique avec instrument

Séparation de la voix à l'enregistrement :

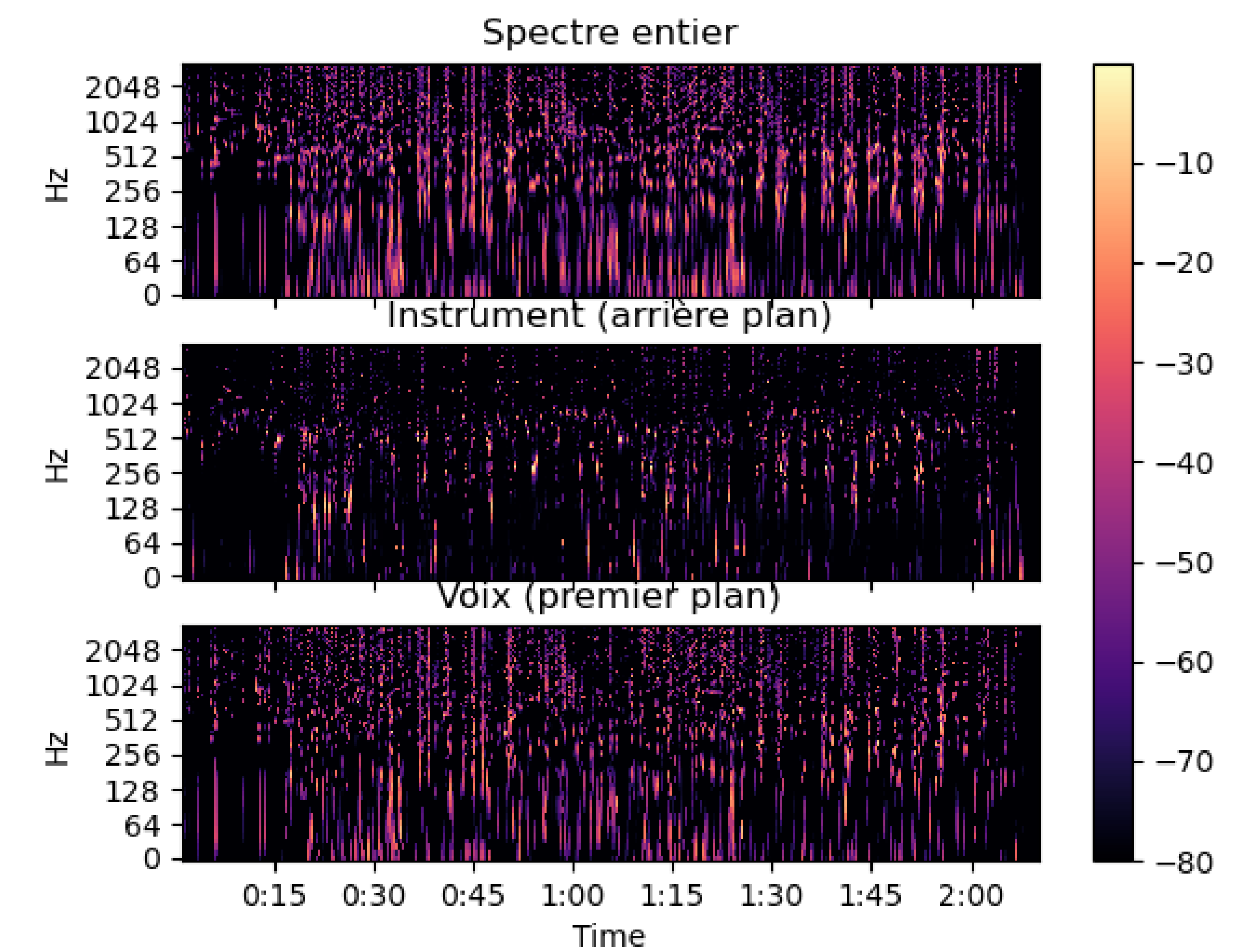
Les méthodes vues précédemment s'appliquent sur tout l'enregistrement et non que la voix, nous devons donc procéder à une séparation de la voix à l'audio d'origine. Pour cela nous utiliserons une méthode qui emploie des filtres "soft mask" (que nous ne détaillerons pas ici) qui a l'avantage de prendre en compte le caractère probabiliste du signal ce qui rend le résultat plus satisfaisant qu'un simple filtre passe-bande qui fait un découpage beaucoup plus grossier.



Remarque : il est beaucoup plus dur de repérer le fondamental et ses harmoniques de l'enregistrement que l'on traite ici car il y a plusieurs voix et que l'enregistrement est d'une taille beaucoup plus grande que l'exemple précédent.

Séparation supplémentaire de la voix : (optionnel)

Si la première séparation de voix n'est pas satisfaisante (on entend encore de manière non-négligeable des instruments en fond), on peut répéter l'opération de séparation de manière à mieux isoler la voix. Ce qui nous donne :



Résultat final :

Une fois que la voix a bien été séparée il faut lui appliquer l'une des deux méthodes vues précédemment en fonction du résultat que l'on veut obtenir et la rajouter au fond extrait lors de l'étape d'avant.

Le QR code ci-dessous mène à un dossier box (alternative à google drive) où il y a 2 exemples de ce que l'on peut obtenir en modifiant la voix comme énoncé.

