Création d’un modèle de reconnaissance vocal

Le modèle utilisé se base d’un sequence modeling with CTC (source : https://distill.pub/2017/ctc/)

Dans ce code utilisé pour la création du modèle un dataset a été pris venant LJSpeech. Il se compose de courts extraits audios d'un seul locuteur lisant des passages de 7 livres de non-fiction.

Nous évaluerons la qualité du modèle en utilisant le Taux d'Erreur de Mot (WER). Le WER est obtenu en additionnant les substitutions, insertions et suppressions qui se produisent dans une séquence de mots reconnus. Divisez ce nombre par le nombre total de mots originellement prononcés. Le résultat est le WER.

Le WER nous permet ainsi de mesurer l’accuracy quant à une metrics élaborer à notre solution.

1 : Ce qui a été utiliser pour créer le model.

Jiwer : Outil utilisé pour calculer le Taux d'Erreur de Mot (WER).

Pandas : Bibliothèque Python utilisée pour la manipulation et l'analyse des données, offrant des structures de données flexibles et performantes.

Numpy : Bibliothèque Python spécialisée dans le calcul numérique, offrant des tableaux multidimensionnels et des fonctions mathématiques.

Tensorflow : Cadre de machine learning utilisé pour construire et former des modèles d'apprentissage automatique.

Matplotlib : Bibliothèque Python de visualisation de données, permettant la création de graphiques et de visualisations de manière efficace.

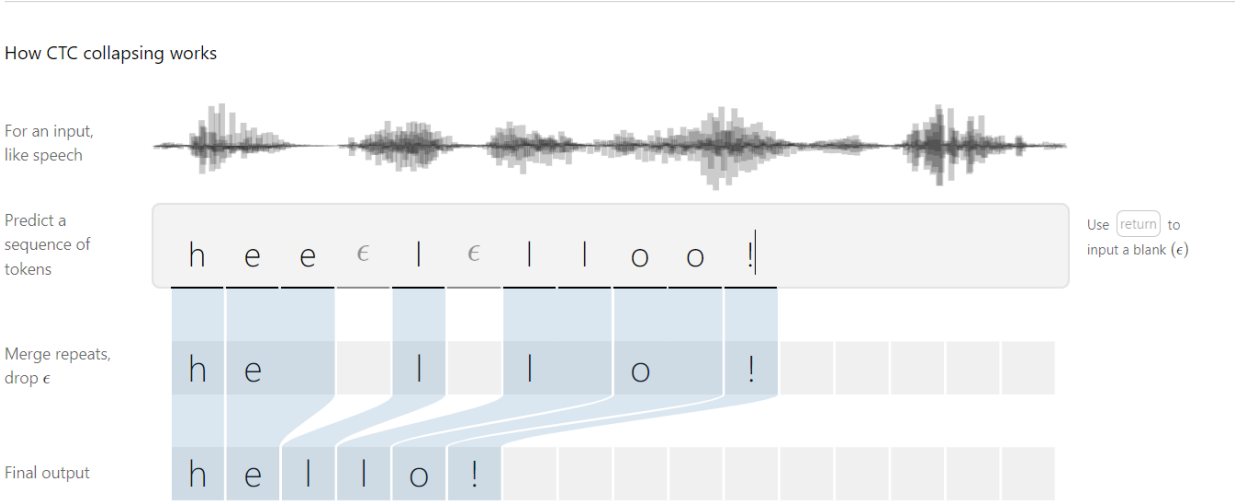
Cette IA construite autour de Tensorflow a permis de déduire une prédiction d’audio assez performante pour un dataset anglais.

```
Word Error Rate: 0.2562
-----------------------------------------------------------------------------------------------------
Target     : report of the president's commission on the assassination of president kennedy the warren commission report
by the president's commission on the assassination of president kennedy
Prediction: report of the president's commission on the assassination of president kennedy the warren commission report
by the president's commission on the assassination of president kennedy
-----------------------------------------------------------------------------------------------------
```

Par la suite du projet a donc été utilisé un model déjà tout fait.

2 : Fonctionnement du CTC

La manière de présenter le CTC est la page de présentation de la logic prise de https://distill.pub/2017/ctc/ .



Ici la fréquence est donc décortiquée pour venir abstraire quant à la magnitude, fréquence, timbre abstraite depuis un spectrogramme (voir transformée de Fournier).

Ainsi par liaison avec le datasetAnglais on est capable de déduire quant à la fréquence la lettre rattachée à l'instant d'un audio

```
Downloading data from https://data.keithito.com/data/speech/LJSpeech-1.1.tar.bz2
2748572632/2748572632 [==============================] - 19s 0us/step
```

	file_name	normalized_transcription
0	LJ043-0102	It is possible that his immediate supervisor n...
1	LJ030-0178	Looking over her right shoulder, she saw that ...
2	LJ011-0021	Further investigations brought other similar f...

Voilà quant à la présentation généraliste du fonctionnement de l'IA.
Quant aux fonctionnels il suffit d'aller voir le code dans speechrecognization.ipynb.