PROPOSITION DE SUJET POUR LE TRAVAIL PRATIQUE 3
COURS IFT-7022 :

Idée du projet :

« Speech to text : conversion des courts audios (à un mot) en texte. »

Membre de l'équipe :

- Kaba Sekou
- Khalil Salah Eddine

Problématique :

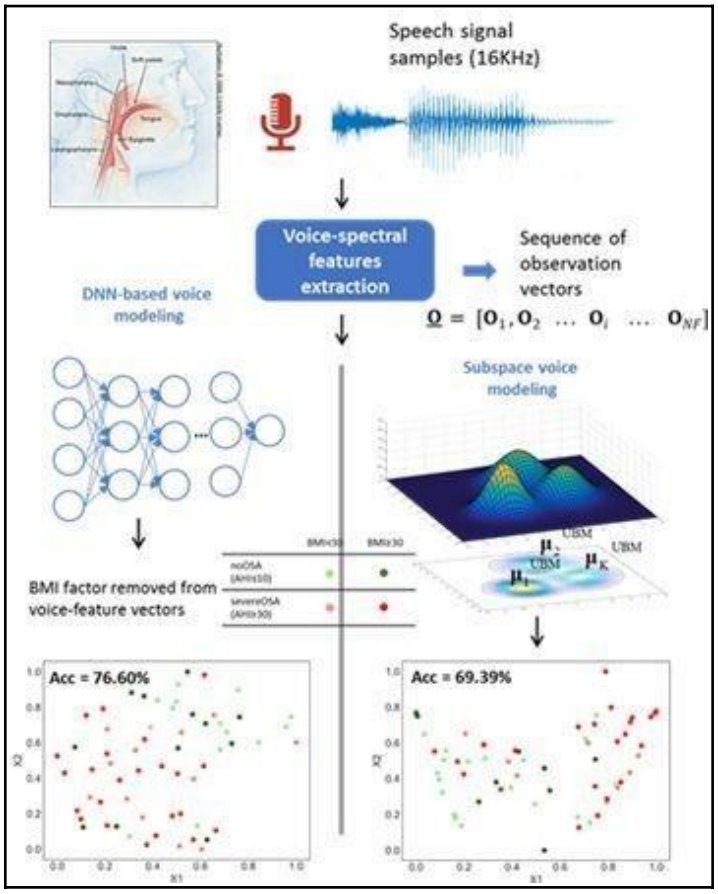
Nous participons de nos jours, à l'avènement d'outils technologiques de plus en plus sophistiqués. Le traitement automatique de la langue naturelle se trouve au cœur de ces innovations et qui, grâce à des algorithmes, gère pratiquement notre quotidien. D'autant essentielle est sa présence, partout où s'immiscent des données colossales, le traitement automatique de la langue naturelle a su opérer dans divers domaines fondamentaux : La synthèse vocale par exemple dans le but d'automatisation des tâches. Comment Google (ou tout autre système de commande vocale) comprend-il ce que je dis ? Et comment le système de Google convertit-il ma requête en texte sur l'écran de mon appareil ?

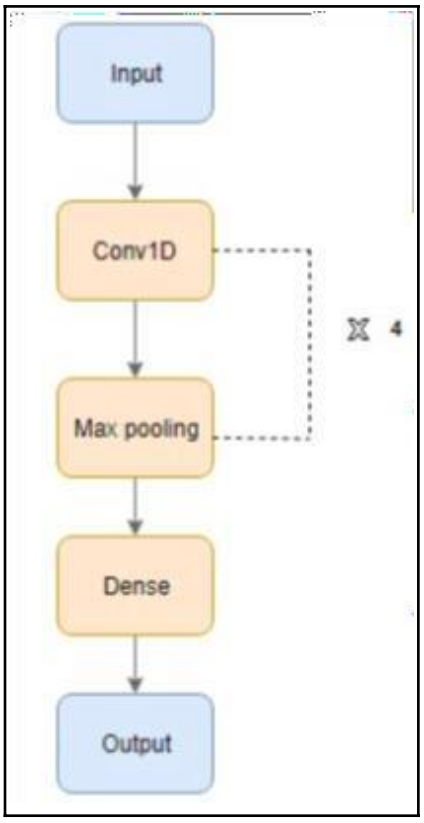
Idée Expliquée :

L'importance indéniable des traitements de signal audio dans le domaine du traitement automatique de la langue naturelle, impose des questionnements, sur les techniques d'étude, de traitement et de modélisation des sons. Dans une démarche progressiste, notre sujet nous permettra une étude approfondie des fondamentaux d'un signal vocal, puis une implémentation d'un modèle approprié de synthèse vocale, sur de courts audios, accompagné d'une petite application bureautique de traitement de son.

Pour traiter le sujet, nous utiliserons des données déjà disponible (voire section source de données). Par la suite échantillonner les audios, dans le but d'avoir une structure, plus maniable de données. Ainsi nous appliquerons des algorithmes d'apprentissage supervisée, à réseau de neurone profond

(Voire l'architecture photo droite ci-dessous), pour détecter la structure de chaque classe de son. Ceci grâce à des packages python, dont le principal est sckit-learn. On mettra en place, au final, un modèle de meilleure performance qui puisse exister permettant de recueillir un son et de prédire le mot qui est énoncé dans celui-ci. Le travail se voit comme une sorte de classification aussi, car l'algorithme essaye de prédire la classe d'un audio donné. Pour un peu résumer ces étapes, nous utilisons cette photo illustrative (gauche).





Source de données :

Il s'agit d'un ensemble de fichiers audio.wav d'une seconde (en moyenne), provenant de la source Kaggle et chacun contenant un seul mot anglais parlé. Ces mots sont prononcés par une variété de locuteurs différents. Les fichiers audios sont organisés en train, dossiers triés en fonction du mot que les audios contiennent, et test contenant un ensemble randomisé d'audio. Ce qui nous ferait un ensemble de données conçu pour former des modèles d'apprentissage automatique supervisé. Cliquez ici pour consulter la source des données.