**Traitement du signal**

**et**

**méthodes d'analyse**

**INTRODUCTION**

Comme nous l'avons mentionné, même le plus élémentaires des systèmes de reconnaissance vocale utilise des algorithmes au carrefour d'une grande diversité de disciplines : reconnaissance de motifs statistiques, théorie de l'information, traitement du signal, analyse combinatoire, linguistique entre autres. Le dénominateur commun étant le traitement du signal qui transforme l'onde acoustique de la parole en une représentation paramétrique plus apte à l'analyse automatisée. Le principe est simple : garder les traits distinctifs du signal et s'absoudre au maximum de tout ce qui pourra en parasiter l'étude. Cette conversion ne se fait donc pas sans perte d'information, et la délicatesse de la discipline tient en la sélection judicieuse des outils les plus adaptés afin de trouver le meilleur compromis entre perte d'information et représentation fidèle du signal.

**PRÉREQUIS**

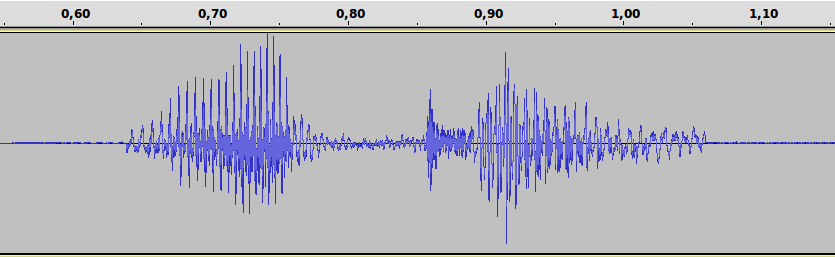
**Qu'est-ce que le son ?**

Le son est une onde mécanique se traduisant par une variation de la pression au cours du temps. Cette onde est caractérisée par différents facteurs comme son amplitude à chaque instant, qui est en d'autres termes la valeur de la dépression à cet instant, et par les fréquences qui la compose et qui changent au cours du temps.

**Comment le son est-il représenté dans l'ordinateur ?**

En se propageant, l'onde mécanique qu'est le son fait vibrer la membrane du micro. L'amplitude de la vibration dépend directement de l'amplitude du son. La position de la membrane est enregistrée à intervalles de temps réguliers définis par l'échantillonnage. L'échantillonnage correspond au nombre de valeurs prélevées en une seconde. Par exemple un échantillonnage à 44100 Hz correspond à relever la position de la membrane 44100 fois par secondes. La valeur de la position de la membrane est alors enregistrée sous la forme d'un entier signé codé sur n bits (n valant généralement 8,16,32 ou 64). Plus n est grand, plus la position de la membrane sera stockée de manière précise, et donc plus la qualité du son sera bonne. Grâce à l'échantillonnage et à, on définit aisément le bitrate, qui correspond au débit d'information par seconde, de la façon suivante : bitrate = n\*échantillonnage.

Ce dont nous disposons donc pour analyser un signal, est la donnée de l'amplitude en fonction du temps la caractérisant.

Exemple audiogramme prononciation du mot « VICA »

**PRE ACCENTUATION DES AIGUS**

Il a donc fallu pondérer l’influence des différentes bandes de fréquences pour rester le plus fidèle possible à la perception de l’oreille humaine. Ainsi, le poids des hautes fréquences (supérieures à 1000 Hz) a été diminué tandis que le poids des basses fréquences (inférieur à 1000 Hz) a été augmenté. Pour cela, on affecte l’amplitude d’une fréquence f à une nouvelle fréquence mel(f) selon la règle :

**ETALONNAGE DES FREQUENCES**

L’oreille humaine perçoit les fréquences de façon non linéaire. Ainsi, l’augmentation d’une octave correspond en réalité à la multiplication de la fréquence par 2 : par exemple, le la3 correspond à une fréquence de 440 Hz tandis que le la4 correspond à une fréquence de 880 Hz.

Il a donc fallu pondérer l’influence des différentes bandes de fréquences pour rester le plus fidèle possible à la perception de l’oreille humaine. Cela implique que le poids des hautes fréquences (supérieures à 1000 Hz) est diminué tandis que le poids des basses fréquences (inférieur à 1000 Hz) est augmenté. Pour cela, on affecte l’amplitude d’une fréquence f à une nouvelle fréquence mel(f) selon la règle :

A description...