

## TP14 - Reconnaissance de phonèmes

Dans ce TP, nous nous intéressons à la reconnaissance de phonèmes, en l'occurrence de neuf phonèmes de type « voyelle » (a, e, é, è, i, o, o « ouvert », ou, u), à partir de fichiers sonores au format WAV. Ces fichiers sont contenus dans le répertoire /mnt/n7fs/ens/tp\_queau/Apprentissage, à raison de cinq enregistrements par voyelle : pour chaque voyelle, la hauteur et la durée du signal changent d'un enregistrement à l'autre.

Sur la figure 1, la partie **A** représente le **signal acoustique d'une voyelle [o]** extraite d'un mot prononcé « Aurore », tandis que la partie **B** représente les **spectres de la voyelle [o]** obtenus grâce à l'analyse spectrale (par transformation de Fourier) et à l'analyse cepstrale.

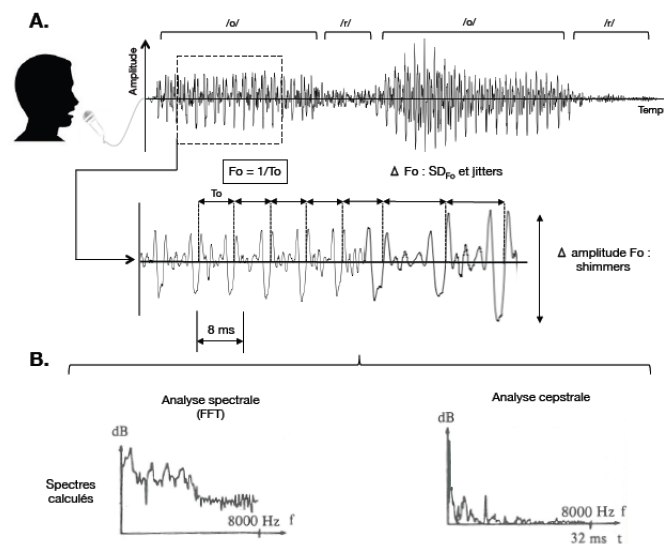


FIGURE 1 – Analyse d'un signal vocal.

La **fréquence fondamentale**, notée  $F_o$  et exprimée en Hertz, est un indicateur de la hauteur de la voix (aigüe ou grave). Il s'agit d'une mesure de la fréquence de vibration des cordes vocales lors de la production des sons voisés, c'est-à-dire du nombre de cycles d'ouverture et de fermeture de la glotte par seconde.

### Analyse spectrale

On appelle **transformée de Fourier discrète** (TFD) d'une suite de  $N$  termes  $x(0), \dots, x(N-1)$ , la suite de  $N$  termes  $X(0), \dots, X(N-1)$ , définis par :

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-2i\pi \frac{nk}{N}}, \quad k \in \{0, \dots, N-1\} \quad (1)$$

En pratique, les  $N$  termes  $x(n)$  peuvent être  $N$  échantillons d'un signal analogique échantillonné :  $x_n = x(nT_o)$  et les  $N$  termes  $X(k)$  peuvent correspondre à une approximation (à un facteur multiplicatif  $T_o$  près) de la transformée de Fourier de ce signal aux  $N$  fréquences :  $F_k = k F_o / N$ ,  $k \in \{0, \dots, N-1\}$ .

La **transformée de Fourier rapide** (FFT, pour *fast Fourier transform*) est une TFD calculée selon un algorithme permettant de réduire le nombre d'opérations et, en particulier, le nombre de multiplications à effectuer.

## Analyse cepstrale

Le cepstre d'un signal  $x(t)$ , notion apparue en 1963 (ceps = spec lu à l'envers), est obtenu par transformation de ce signal du domaine temporel vers un autre domaine analogue au domaine temporel. Le *cepstre réel* est défini comme la transformée de Fourier inverse du logarithme du module de la transformée de Fourier du signal acoustique. Il présente l'avantage de permettre la séparation des contributions respectives de la source et du conduit vocal.

Par analogie avec un sonagramme, il est possible de calculer un cepstre pour chaque position d'une fenêtre glissante. La figure 2 montre la chaîne complète de calcul des coefficients cepstraux. La *fenêtre de Hamming* est une fonction positive, à support borné de largeur  $T$ , égale à  $0,54 - 0,46 \cos(2\pi \frac{t}{T})$  pour  $t \in [0, T]$ .

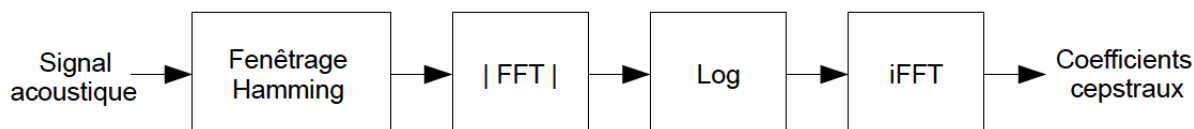


FIGURE 2 – Chaîne de calcul des coefficients cepstraux.

## Travail demandé

Écrivez un script qui, pour chaque enregistrement, charge le signal vocal à l'aide de la fonction `wavread`, puis :

1. Complétez le fichier `spec_ceps.m` qui permet de calculer les coefficients spectraux et cepstraux du signal passé en paramètre, en utilisant les fonctions `hamming`, `fft` et `ifft` de Matlab. Ces coefficients, qui caractérisent le signal vocal, sont obtenus en moyennant les coefficients correspondant aux différentes positions de la fenêtre de Hamming. Chaque signal est caractérisé par 882 coefficients spectraux et autant de coefficients cepstraux.
2. Effectuez une classification non supervisée par l'algorithme des  $k$ -moyennes (fonction `kmeans` de Matlab), pour les deux types de coefficients (spectraux et cepstraux). Lisez la documentation de cette fonction. Il est conseillé d'utiliser l'option `'emptyaction'` avec la valeur `'error'`, et l'option `'start'` avec comme valeur une matrice contenant les moyennes des coefficients correspondant aux différents enregistrements (il y en a cinq) de chaque voyelle. Calculez les taux de bonnes classifications : l'utilisation des cepstres est-elle pertinente ?
3. Analysez les résultats des classifications en vous référant au triangle voyellique de la figure 3.

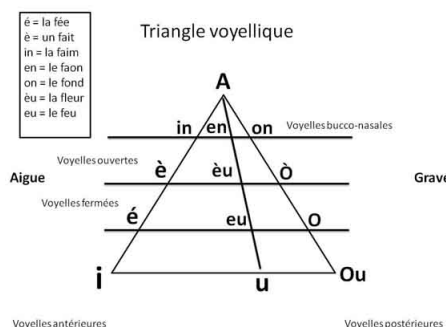


FIGURE 3 – Triangle voyellique.

4. Afin de pouvoir visualiser le résultat de ces classifications, réduisez la dimension des coefficients par ACP, en ne conservant que les deux ou trois premières composantes principales. Affichez les données dans ce repère, avec une couleur par classe, et en affichant par exemple en noir les enregistrements mal classés (cf. TP13).
5. Les fichiers du répertoire **Apprentissage** vous ont permis de réaliser un *classifieur*. On dit que ces données constituent des *données d'apprentissage*. Vous pouvez maintenant utiliser ce classifieur pour d'autres enregistrements, en l'occurrence ceux du répertoire **Tests**, qui constituent des *données de test*.
6. Enfin, vous pouvez effectuer une classification bayésienne inspirée du TP13. Les séparations entre classes ne seront plus des hyperplans, mais rien ne dit que le taux de bonnes classifications ne sera pas meilleur.