

# BE reconnaissance des formes

Veysseire Daniel<sup>1</sup>

Fabre Michaël<sup>2</sup>

<sup>1</sup> Université Paul Sabatier

<sup>2</sup> Université Paul Sabatier

Mon adresse complète

Mon adresse électronique

3 novembre 2014

## Résumé

*Cet article vise à comparer l'efficacité de deux méthodes de classification (méthode de classification par loi normal multidimensionnel et méthode des K-Plus Proche voisin), ainsi que les choix de paramétrisation des données (FFT, cepstre, MFCC), principalement dans le cadre de la reconnaissance de la parole.*

*Dans un premier temps, nous ferons une présentation théorique de ces méthodes et paramétrisations. Dans un deuxième temps nous présenterons le protocole expérimental mis en place afin de comparer leurs efficacités.*

*Nous interpréterons ensuite les résultats obtenus puis nous finirons par une conclusion sur l'efficacité des différentes méthodes et paramétrisations.*

## Mots Clef

Méthodes de classification, reconnaissance de la parole, loi normale, K plus proche voisins, paramétrisation, FFT, Cepstre, MFCC, apprentissage supervisé.

## Abstract

*It's the English version of the abstract. Exactly as in French it must be short. It must speak of the same topics...*

## Keywords

Example, model, template.

## 1 Introduction

La reconnaissance automatique de la parole est une technique informatique qui permet d'analyser un signal de parole.

On se place ici dans le cas où on essaye de reconnaître chaque syllabe individuellement. On dispose d'une référence de 1000 éléments sonore de 64ms échantillonnés à 16KHz et quantifiés sur 16 bits. On a ainsi 100 échantillons pour chacune des dix syllabes suivantes :

[a],[e],[ɛ],[ə],[ɪ],[ø],[ɔ],[o],[u],[y]

correspondant aux classes :

'aa','ee','eh','eu','ii','oe','oh','oo','uu','yy' ;

## 1.1 Les différentes paramétrisations

Nous allons utiliser différentes paramétrisations des données et conserver celles qui offrent les meilleurs résultats.

### Transformé de Fourier Rapide (FFT)

La transformée de Fourier Rapide est un algorithme permettant de traiter un signal afin d'obtenir son spectre. Le spectre d'un signal nous fournit l'intensité de chacune des plages de fréquences pendant un intervalle de temps t. Elle s'effectue sur un certain nombre de point, et la taille des plages de fréquences diminue et le nombre de plages augmentent en augmentant ce nombre de points.

En générale on effectue plusieurs FFT sur signal, à l'aide de fenêtres glissantes, afin d'obtenir l'intensité des fréquences à plusieurs instants t, puis on utilise des algos comme le DTW (Dynamic time warping). Mais dans le cas qui nous intéresse ici, les échantillons sont extrêmement courts (64ms avec une fréquence d'échantillonnage de 16KHz). Utiliser une fenêtre glissante serait inefficace. On est donc dans un cas simplifié, on ne cherche qu'à comparer des voyelles prononcées dans un temps très court. Une simple FFT est donc suffisante, on obtient ainsi un vecteur de taille variable selon le nombre de point sur lesquels on a réalisé la FFT. Il faudra par la suite comparer ses vecteurs entre eux (e.g par distance euclidienne) On effectue souvent un lissage du signal par Hamming lorsque'il y a un recouvrement de fenêtre pour éviter de trop grandes discontinuités entre les fenêtres. Il serait donc une erreur de vouloir faire un lissage par Hamming ici car on a pas utilisé de fenêtres glissante. De plus on ne garde que la valeur absolue du résultat pour ne pas manipuler des nombres complexes.

### Le cepstre et les MFCC

Le cepstre est obtenu à partir du spectre. On effectue la transformée inverse du logarithme de la transformée de Fourier (ou spectre) obtenu précédemment. En pratique on ne garde que la valeur absolue du résultat. On obtient ainsi une transformation du signal dans un domaine analogue au domaine temporel. "Les MFCC (Mel-Frequency Cepstral

Coefficients) sont des coefficients cepstraux calculés par une transformée en cosinus discrète appliquée au spectre de puissance d'un signal. Les bandes de fréquence de ce spectre sont espacées logarithmiquement selon l'échelle Mel" (wikipédia). Les MFCC sont proches du cepstre, mais diffèrent par l'utilisation de l'échelle Mel, échelle basée sur la perception humaine.

## 1.2 Les différentes méthodes de classifications

Comme dit précédemment nous allons comparer les deux méthodes de classifications

### classification par loi normal multidimensionnel

lois normales

16	0	0	3	0	1	0	0	0	0
0	17	0	0	3	0	0	0	0	0
0	0	19	0	0	0	1	0	0	0
0	0	0	14	0	3	3	0	0	0
0	1	0	0	18	1	0	0	0	0
0	0	0	1	0	15	3	1	0	0
0	0	0	0	0	1	17	2	0	0
0	0	0	0	0	0	0	17	3	0
0	0	0	0	0	0	0	0	20	0
0	0	0	0	0	0	0	1	0	19

### classification par les K-plus proche voisin

Les en-têtes sont également en 12 points gras.

Il n'y a pas nécessairement d'espacement entre les paragraphes.

Les références à la Bibliographie peuvent être de la forme [2] où [1]. Les numéros correspondent à l'ordre d'apparition dans la bibliographie, pas dans le texte. L'ordre alphabétique est conseillé.

## 2 Le coin L<sup>A</sup>T<sub>E</sub>X

Pour les utilisateurs de L<sup>A</sup>T<sub>E</sub>X, ce patron est minimaliste et vous aurez besoin de votre manuel L<sup>A</sup>T<sub>E</sub>X pour insérer équations et images.

Pour les images le « paquet » `graphicx` est très bien.

Les fichiers de style nécessaires pour la compilation L<sup>A</sup>T<sub>E</sub>X sont :

- `a4.sty` (pour L<sup>A</sup>T<sub>E</sub>X, mais pas L<sup>A</sup>T<sub>E</sub>X 2<sub>ε</sub>)
- `french.sty` (ou Babel français)
- `rfia2000.sty`

Vous devriez avoir les deux premiers dans votre installation de L<sup>A</sup>T<sub>E</sub>X. Le dernier contient la définition des marges et vous devrez le récupérer.

### Annexe

Merci de votre participation.

### Références

- [1] U. Nexpert, *Le livre*, Son Editeur, 1929.
- [2] I. Troiseu-Pami, Un article intéressant, *Journal de Spi-rou*, Vol. 17, pp. 1-100, 1987.