



Reconnaissance vocale

Soutenance de ModEx
PHY 434
Mardi 19 avril 2005

BELARDI S.
ROBIN O.



Plan

- Intro
- Les méthodes
- Le LPC
- La distance
- Les applications

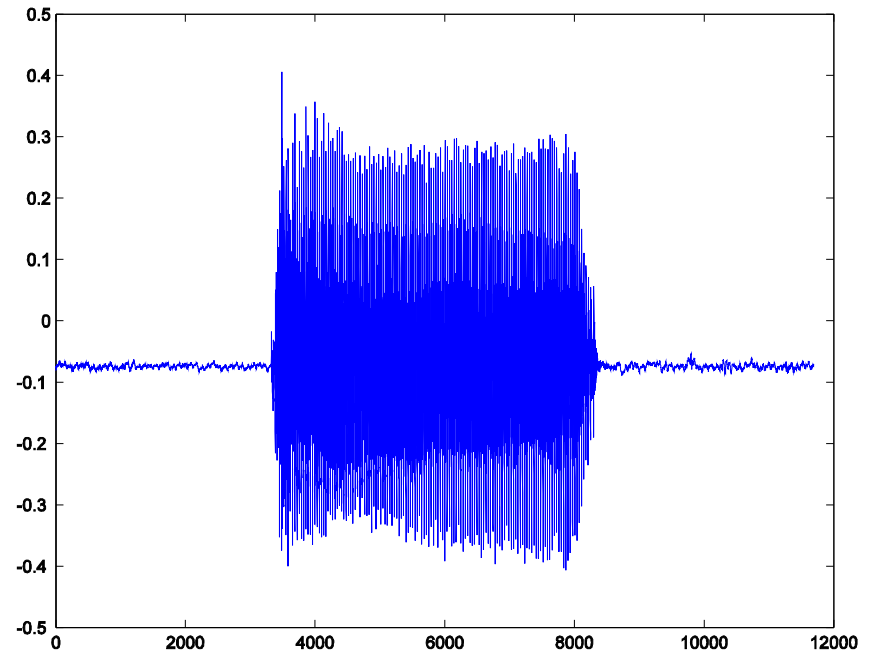


Introduction

- Recherche de nouveaux procédés d'identification
- Avantages de la reconnaissance vocale
- Existence de puces vocales
 - Reconnaissance bon marché
 - Compacte
 - Rapide

La voix

- Une onde mécanique
- Caractérisée
 - Amplitude
 - Fréquences
- Caractérisations du signal ?





Caractéristiques des sons

- Toutes nos opérations se font sur des intervalles d'étude d'environ 23ms
- Fréquence d'échantillonnage de 44100Hz
- Monophonique
- Microphones variables



Ce qu'on veut faire

- Contrôler le mot
- Reconnaître le locuteur
 - De manière relativement automatique et rapide
- Essayer d'autres applications



Les méthodes

- Recherche sur les caractéristiques
 - Distances amplitude
 - Distances spectre
- Approfondissements
 - Cepstre
 - Coefficients LPC

Méthode 0 : Amplitude

- Fonction non temporellement homogène

$$L(f(t)) = g(t) \Rightarrow l(f(t - \tau)) = g(t - \tau)$$

- Dépend de l'origine
- Dépend énormément des circonstances d'enregistrement (volume, fréquence).
- Étudiée, non testée

Spectre

- Transformée de Fourier

$$TF(f)(k) = \sum_{n=0}^{N-1} f(n) e^{\frac{-i2\pi n}{N}k}$$

- Indépendant de l'origine

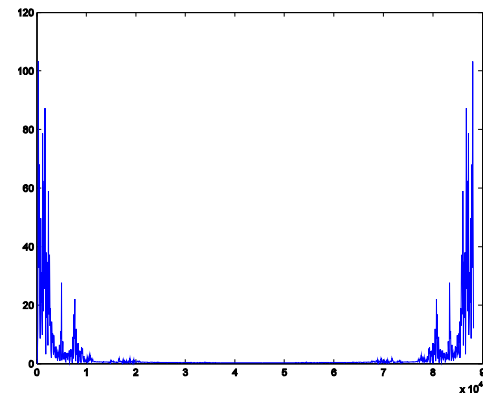
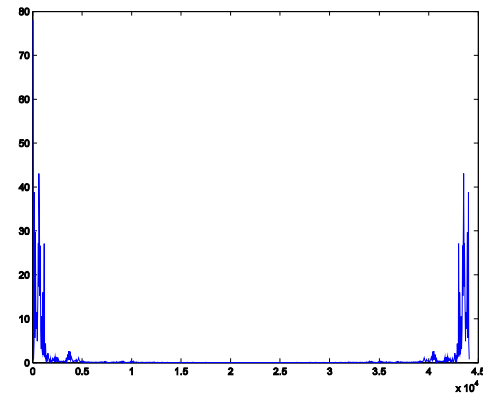
- Normalisable

- Séparation des éléments intéressants

- (40 – 4000 Hz) sur les (0 - 44100 Hz)

Spectre résultats

- Très peu d'informations importantes
- Une dépendance en hauteur
- Résultats mitigés (20% de réussite sur 2 personnes)
- Pas de fenêtrage





Décalage du spectre

- A partir de la fréquence fondamentale
- Résultats similaires
- Problèmes de filtres fixes

Cepstre

- Définition

$$Cepstre(k) = TF^{-1} \left(\log |TF(f)|^2 \right)$$

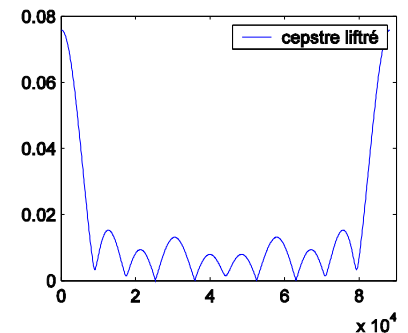
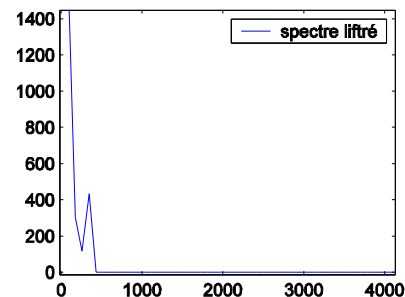
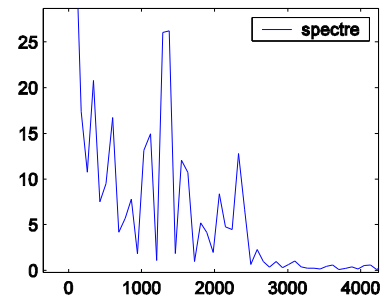
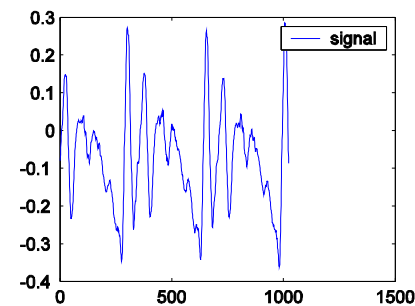
- Intérêt : Déconvolution du filtre

- Liftrage

- ☐ Suppression du terme d'excitation
- ☐ Suppression de certaines quéfrences

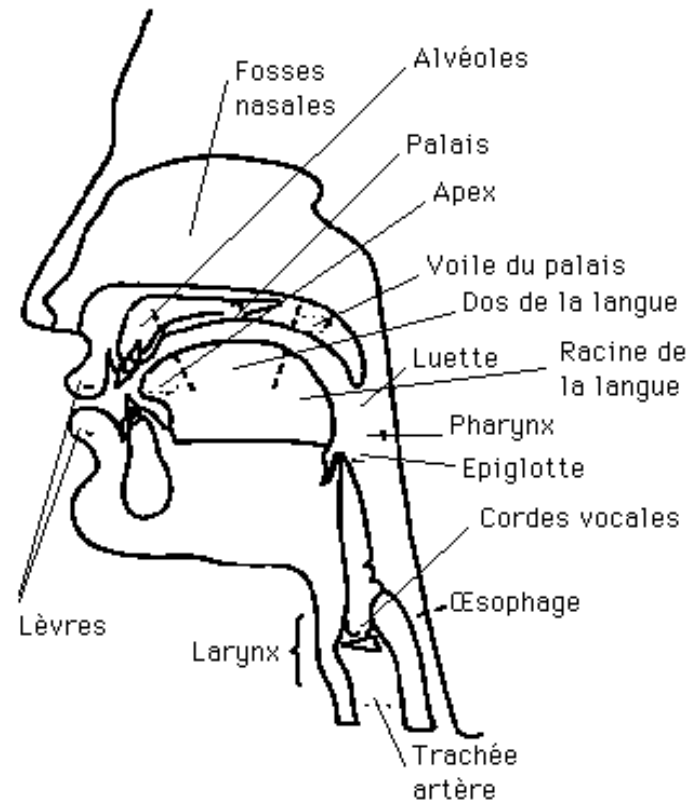
Implémentation

- Difficultés à obtenir des courbes satisfaisantes
- Abandon rapide pour des méthodes plus élaborées.
- Pas de réel test



La voix

- Cavités et filtres dépendant
 - De la personne
 - De la lettre
- Train d'onde excitateur variable
- Idée : analyser ces filtres



LPC (Linear Predictor Coefficients)

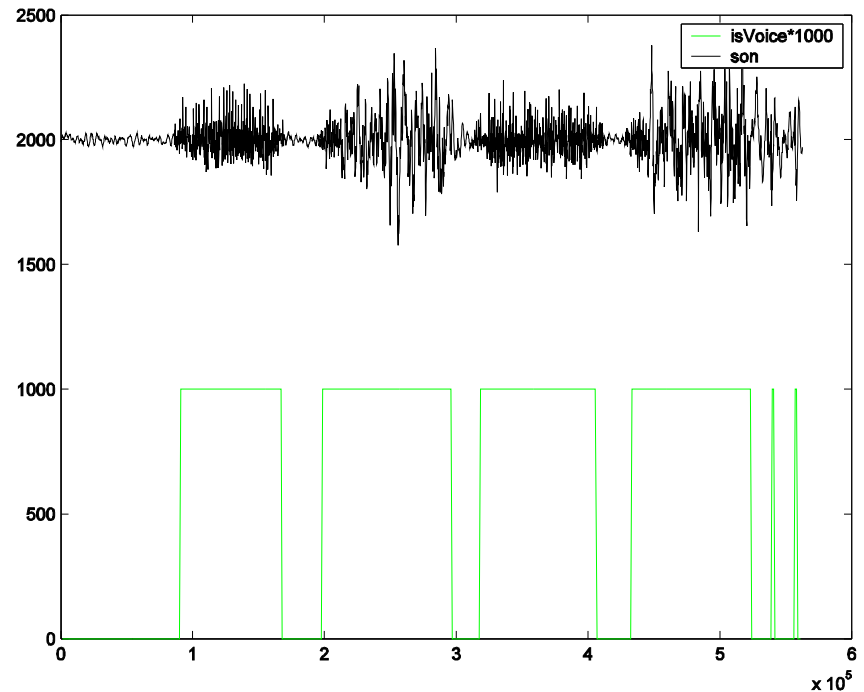
- Coefficients du filtre gorge / bouche

$$H(e^{j\omega}) = \left(\sum_{k=0}^N L(k) e^{-jk\omega} \right)^{-1}$$

- On utilise l'enveloppe de la réponse fréquentielle
- Théoriquement
 - Même personne + même lettre ! même filtre
 - En pratique, des différences apparaissent

Séparation voix / non-voix

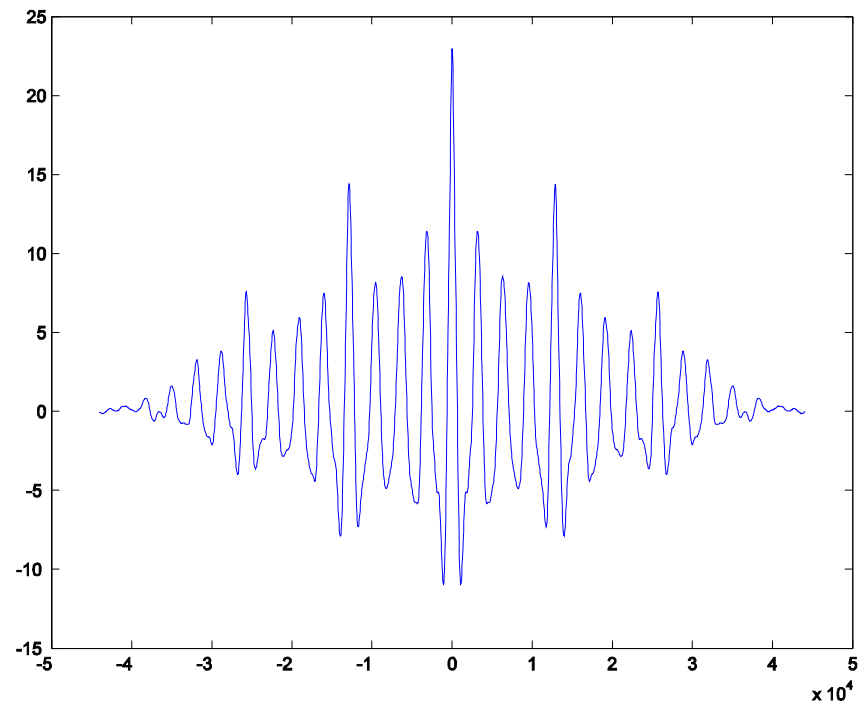
- Par l'amplitude seule
- Développement d'une méthode plus évoluée
 - Amplitude
 - Fréquence fondamentale
 - Variation de fréquence fondamentale



Par l'amplitude seule

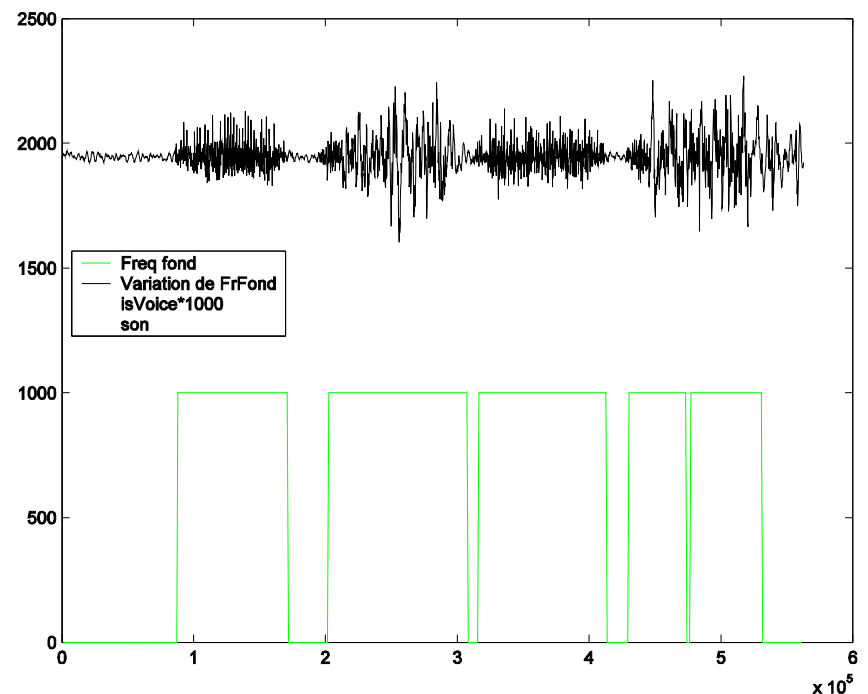
La fréquence fondamentale

- Par autocorrélation
- Problème de détection du 1^{er} pic
 - Détection fréquente d'harmoniques (ou de non harmoniques)



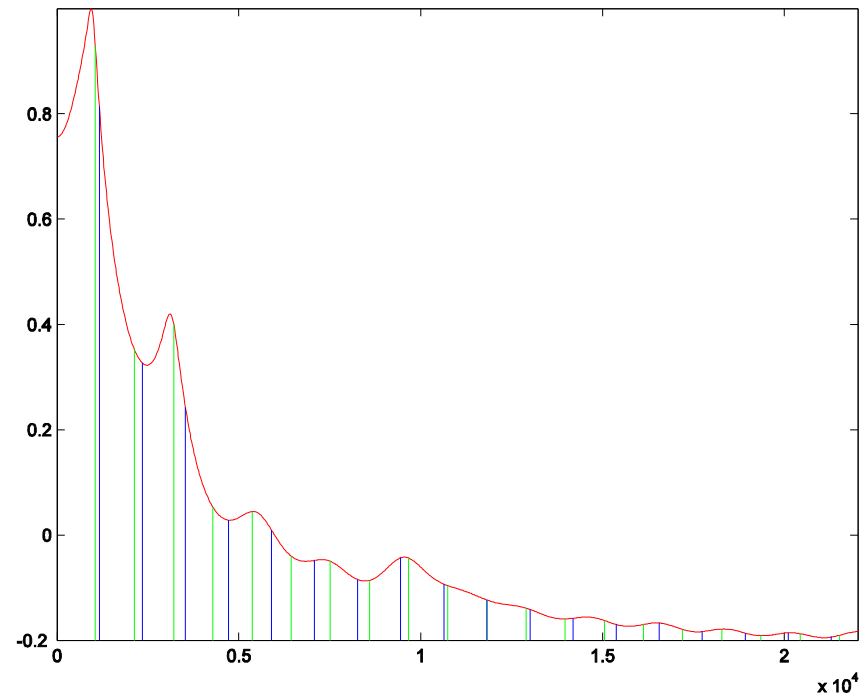
Séparation

- Fréquence fondamentale entre 100 et 1000 Hz
- Reconnaît les bruits
- Plus lent
- Problème de microcoupures néfastes pour les mots



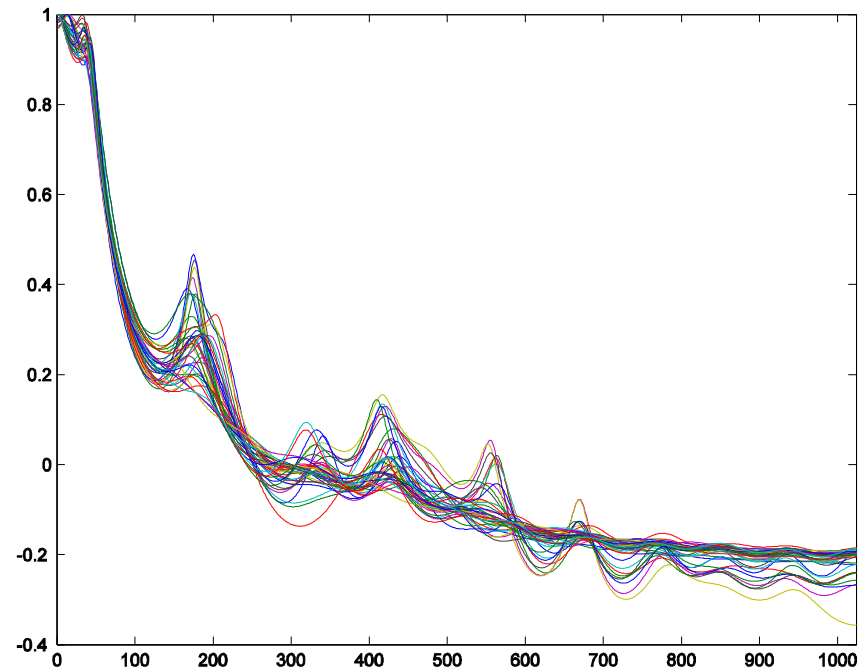
Le filtre du LPC

- MatLab® intègre une fonction *LPC* pré-programmée
- On a travaillé essentiellement sur la réponse fréquentielle de ce filtre



Moyenne ou pas ?

- Les références sont des sons de 0.5 à 2 secondes
- On effectue les études sur toutes les fenêtres d'analyse
- On effectue la moyenne
- On conserve les écarts types du dictionnaire
- Dans certains cas, on conserve toutes les données



Théorie : Les distances

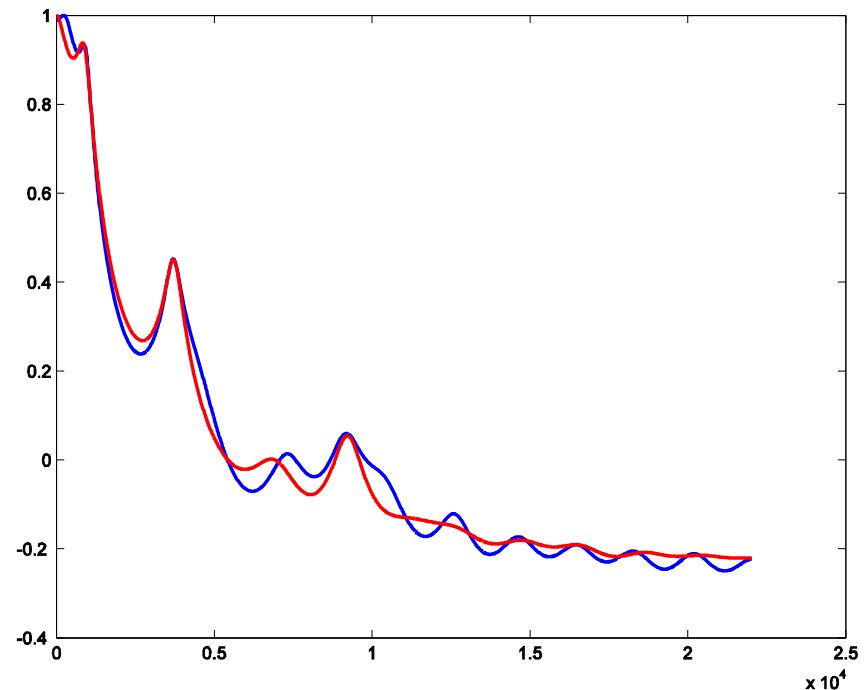
■ Base :

- Intégrale du carré de la différence

■ Mauvais résultats

- Les courbes LPC différent beaucoup pour la partie aiguë

$$d_1(f, g) = \sum_{k=1}^{1024} |f(k) - g(k)|^2$$



Théorie : Importance des graves

- Multiplication par un facteur en $1/k$

$$d_2(f, g) = \sum_{k=1}^{1024} \frac{|f(k) - g(k)|^2}{k}$$

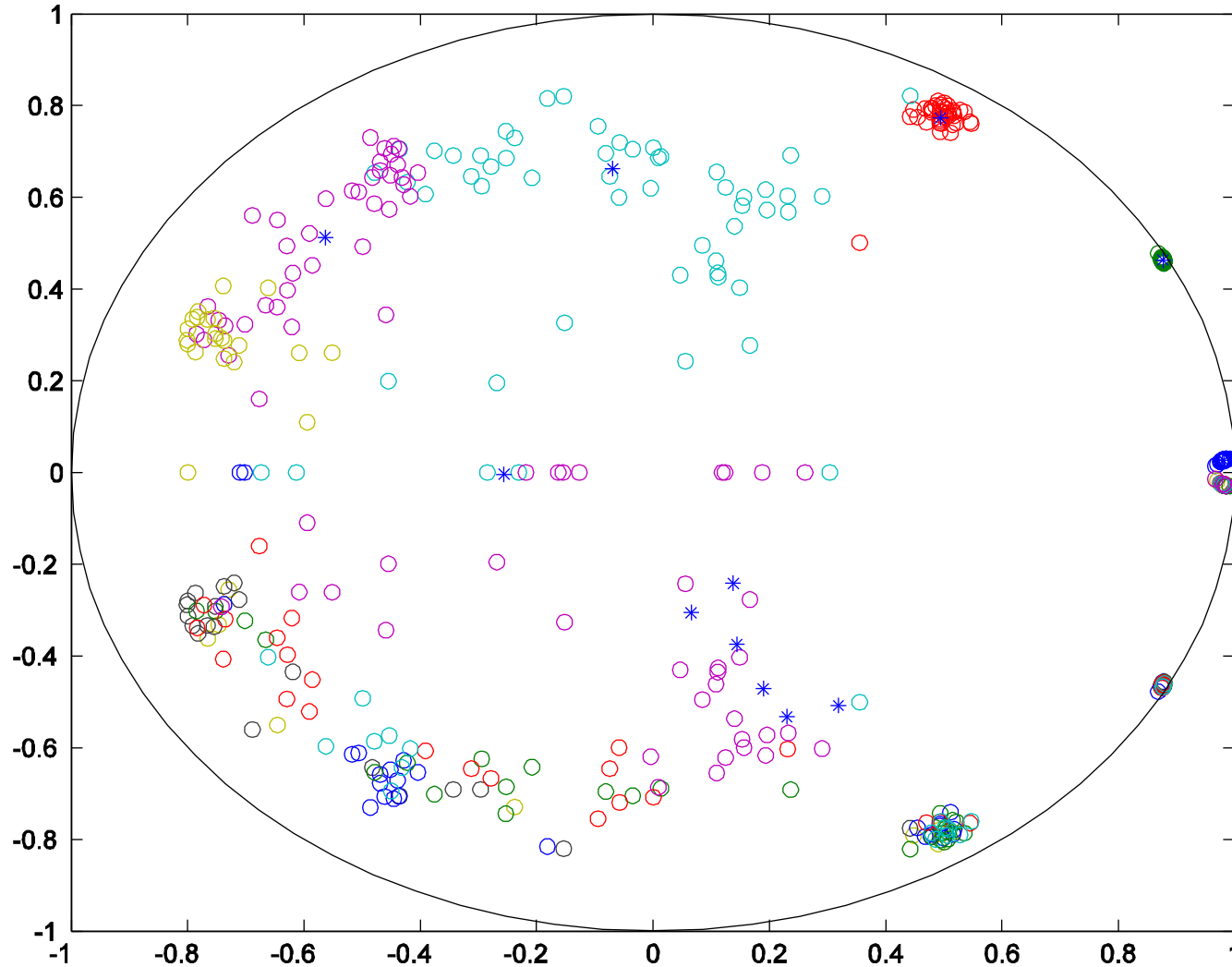
- Division par l'écart type

$$d_3(f, g) = \sum_{k=1}^{1024} \frac{|f(k) - g(k)|^2}{k \cdot \sigma(k)}$$

Théorie distance

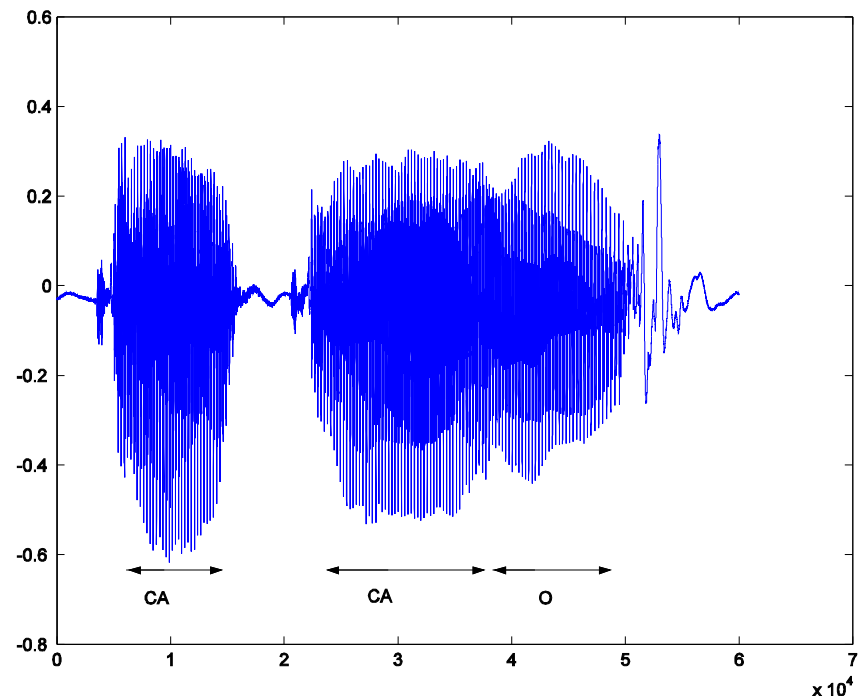
- D'autres distances ont été testées pour donner plus ou moins d'importance aux graves
 - Log, exp, polynomes
- Variation du poids en $1/N$

La distance sur les pôles



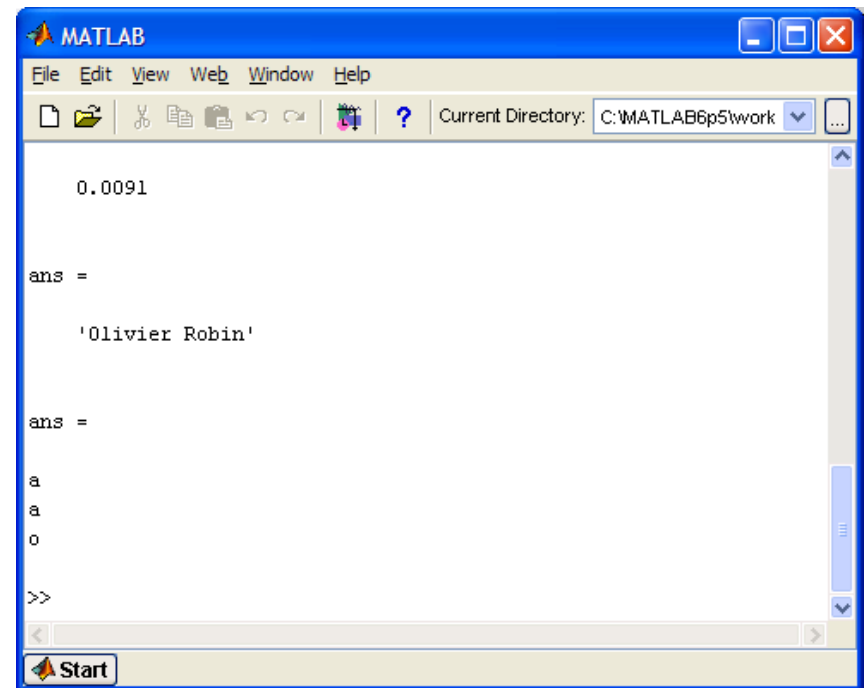
Les applications : mot de passe

- Séparation des syllabes
 - Par les blancs
 - Arbitrairement à partir d'une référence



Principes de la reconnaissance

- Pour chaque personne, on reconnaît la voyelle la plus probable
- On somme les distances
- La personne la plus proche emporte toutes les voyelles



A screenshot of the MATLAB command window. The window has a blue title bar with the MATLAB logo and the text 'MATLAB'. Below the title bar is a menu bar with 'File', 'Edit', 'View', 'Web', 'Window', and 'Help'. A toolbar with various icons is located below the menu bar. The 'Current Directory' is set to 'C:\MATLAB6p5\work'. The command window shows the following text:

```
0.0091  
  
ans =  
  
    'Olivier Robin'  
  
ans =  
  
a  
a  
o  
  
>>
```

The window also shows a 'Start' button at the bottom left.



La calculatrice vocale

- Problèmes de consonnes
 - Plus,...
- Problèmes de similarités
 - Un et Moins,...
- Mise en place d'une micro analyse syntaxique

Distance choisie

- Pour chaque voyelle, prend les 5 fenêtres de plus forte assurance
- On fait une moyenne sur ces 5 voyelles

```
Command Window
>> calculatrice('C:\Developpement\MatLab\work\calc\2f7-6.wav', bC, b0)

formule =

2*7-6

resultat =

8

~
```