T.I.P.E. 2017

Identification vocale : analyse fréquentielle et réseaux de neurones

Plan:

I. La Voix en biométrie

II. Principes généraux

III. Mise en pratique

I. La Voix en biométrie :

→ C’est la reconnaissance de caractéristiques propres aux individus en vue d’une identification.

→ Parmi les plus utilisés : les empreintes digitales, la reconnaissance d’iris ou l’analyse comportementale, et **la *voix***.

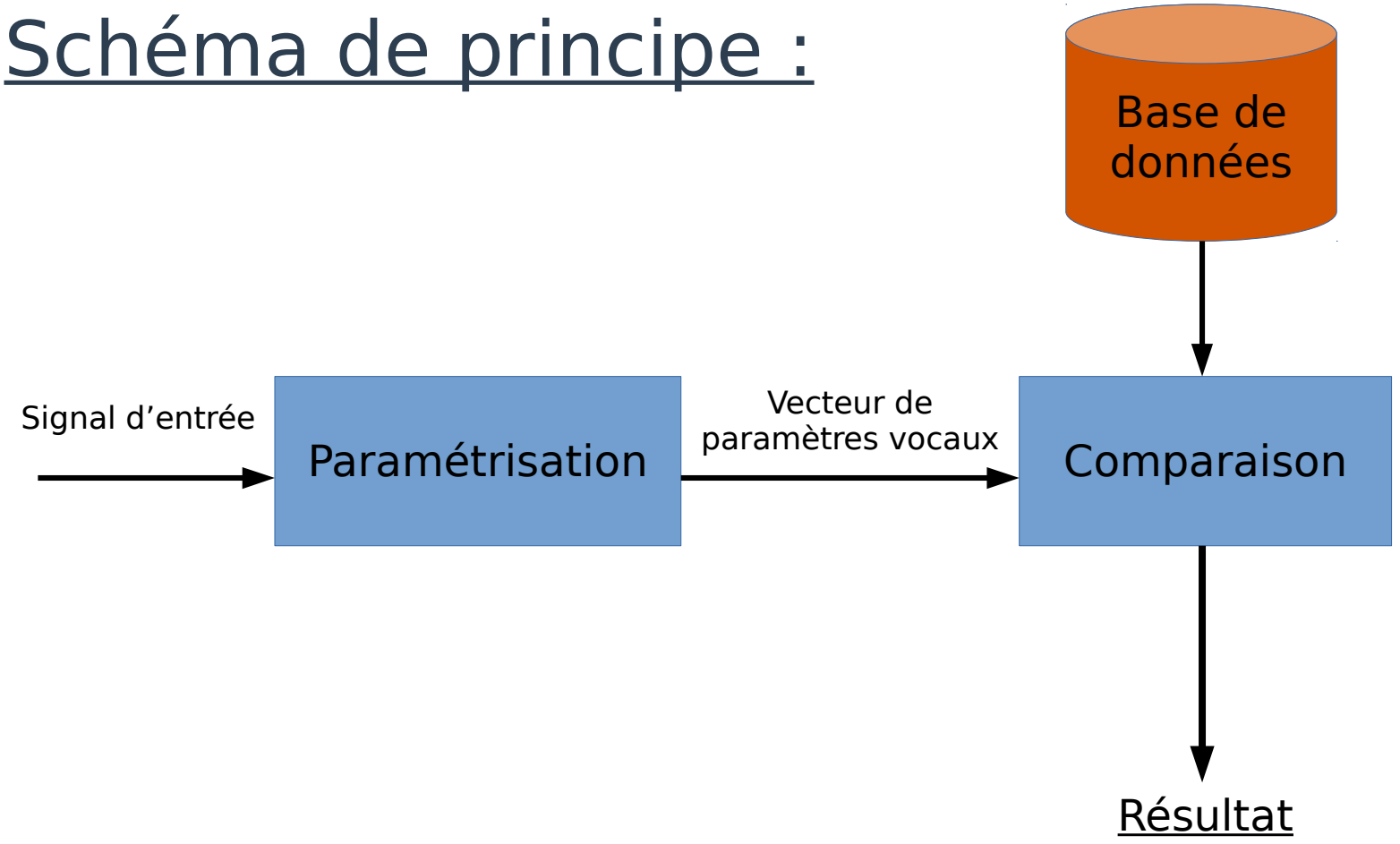
I. La Voix en biométrie :

Deux cas :

- **Vérification** : on veut savoir si celui qui parle est bien celui qu'il dit être.
- **Identification** : on veut identifier une voix parmi une liste de voix préenregistrées → **c'est notre objectif**

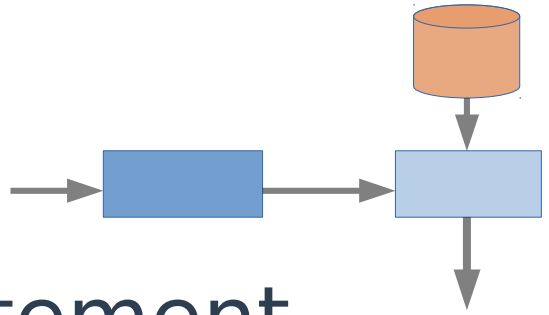
II. Principes généraux

Schéma de principe :



II. Principes généraux

A petite échelle, l’étape la plus importante est **la paramétrisation** : une application transformant un signal initial **lourd** en un vecteur sonore **léger.**



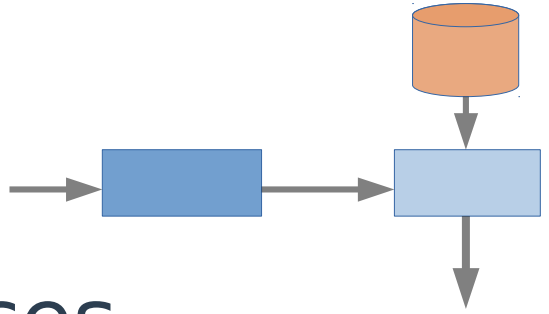
→ Gain en vitesse de traitement

→ Vecteur sonore caractéristique du locuteur

II. Principes généraux

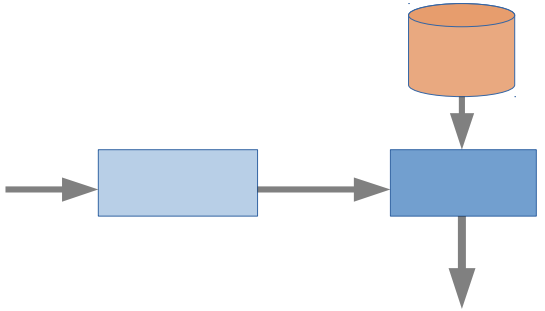
→ Se base sur un modèle:

Modèle de production de la voix,
avec des hypothèses simplificatrices



II. Principes généraux

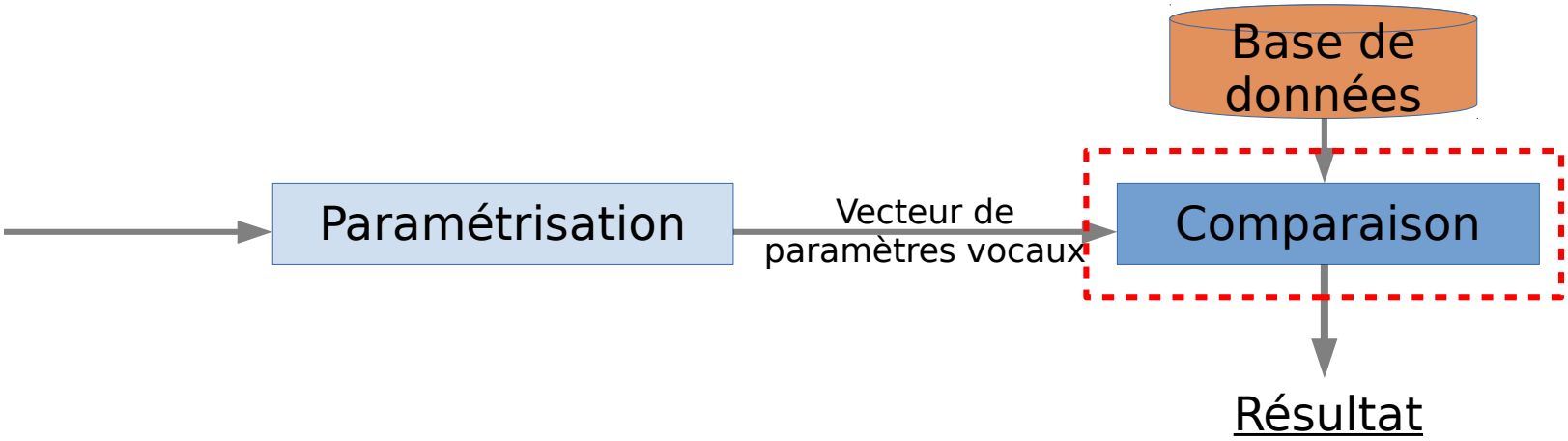
Comparaison:

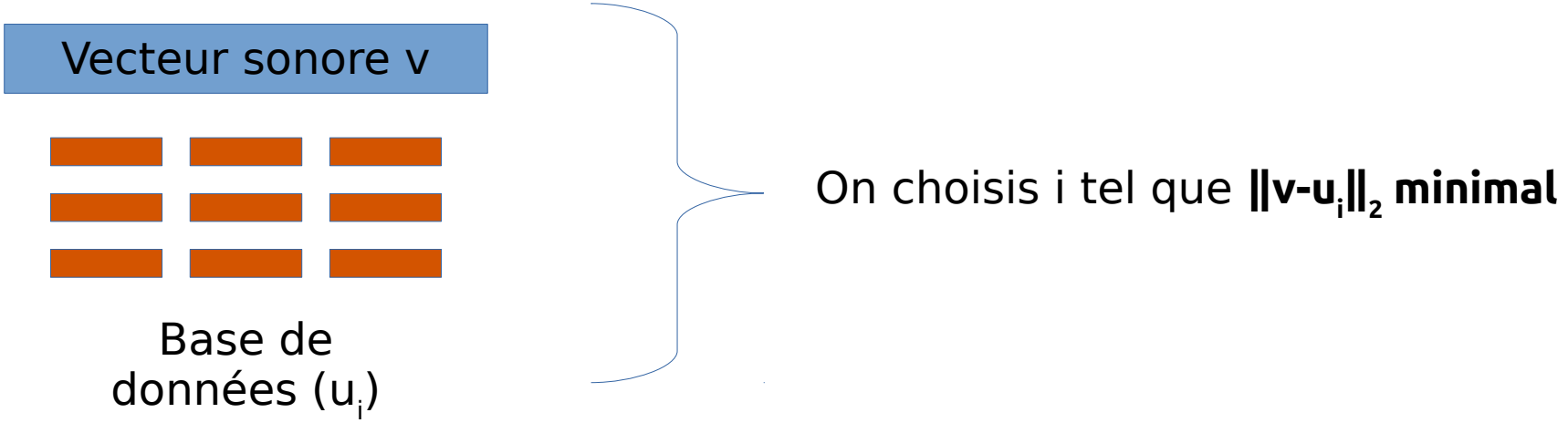


→ Déterminer quel vecteur sonore de la base de donnée est le plus proche du vecteur d’entrée.

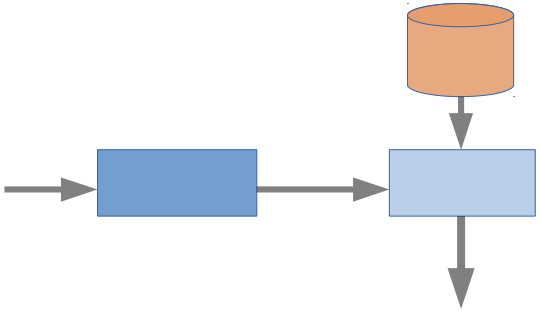
→ Renvoie ce vecteur résultat.

III. Mise en pratique





III. Mise en pratique



On utilise des “A”

Analyse spectrale simple :

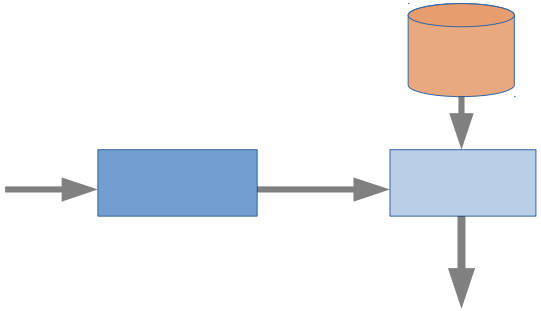
→ Décroissance en amplitude des harmoniques

Modèle :

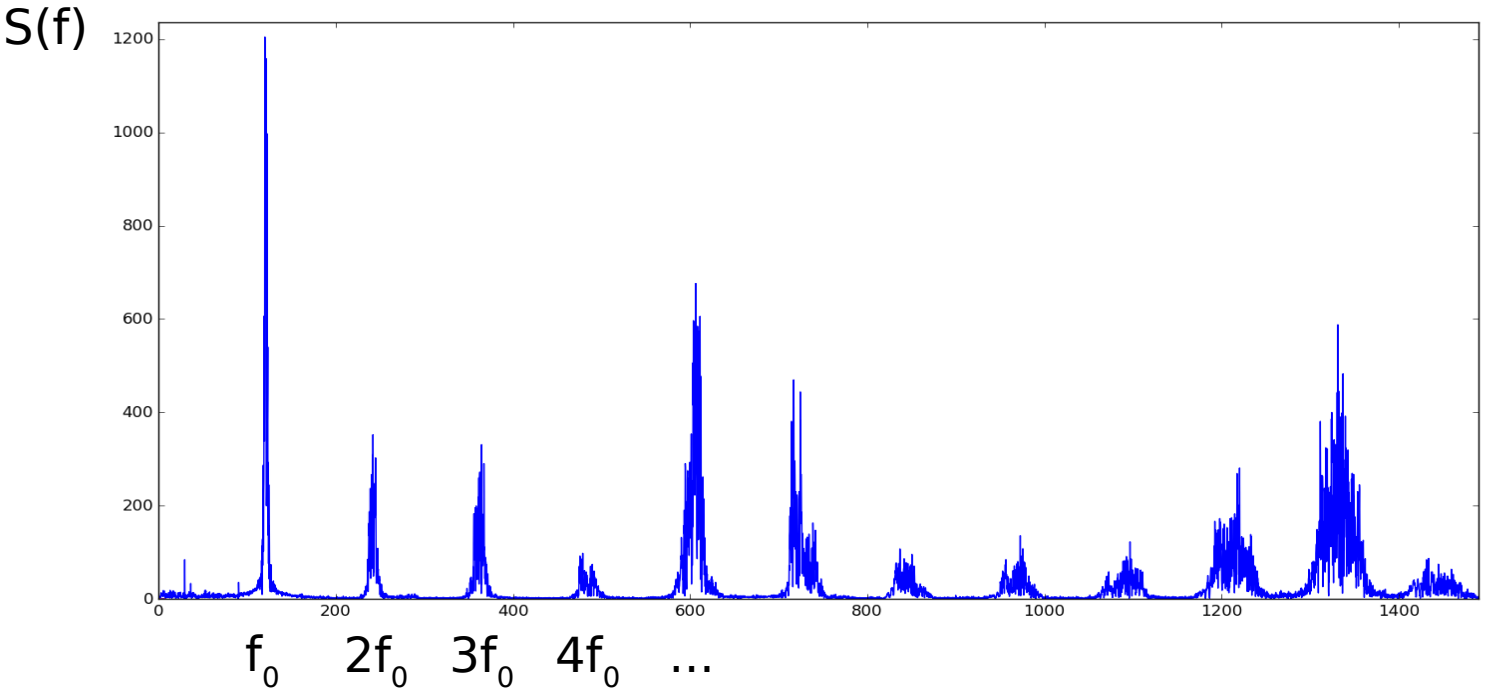
- On suppose que le spectre a la même forme quelle que soit la hauteur du son

III. Mise en pratique

“A”:



→ Décroissance des harmoniques

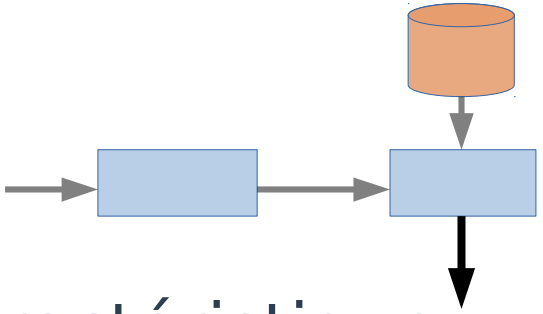


Vecteur sonore:

[$S(2f_0)/S(f_0)$, ... $S(kf_0)/S(f_0)$]

III. Mise en pratique

“A”: Résultats:



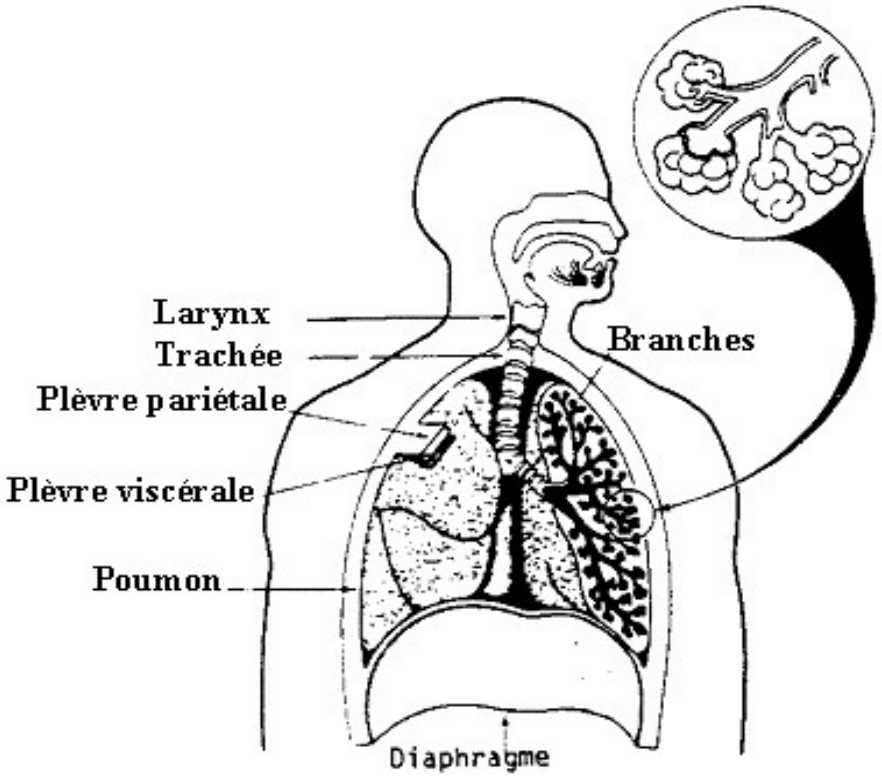
Paramétrisation mauvaise:

→ Décroissance des harmoniques : pas caractéristique du locuteur.

Il faut pousser plus loin l’analyse.

III.Mise en pratique

Nouveau modèle:



Son produit par les cordes vocales
Son modifié par la trachée et le larynx → **filtre**

D'où:

$$s(t) = h * e(t)$$

$$S(f) = H(f) \times E(f)$$

III. Mise en pratique

Analyse cepstrale:

S(f) = H(f) x E(f) :

→ On passe au log pour **séparer** l'influence de l'entrée et du filtre (en abandonnant la phase)

log(|S(f)|) = log(|H(f)|) + log(|E(f)|)

III. Mise en pratique

→ On repasse en temporel en appliquant une transformée de Fourier inverse (en pratique une DCT) :

s’(*ce t*) = h’(*ce t*) + e’(*ce t*)

→ C’est le **cepstre.**

***ce t* est homogène à un temps, dans un domaine différent du domaine temporel (espace *quéfrentiel*)**

III. Mise en pratique

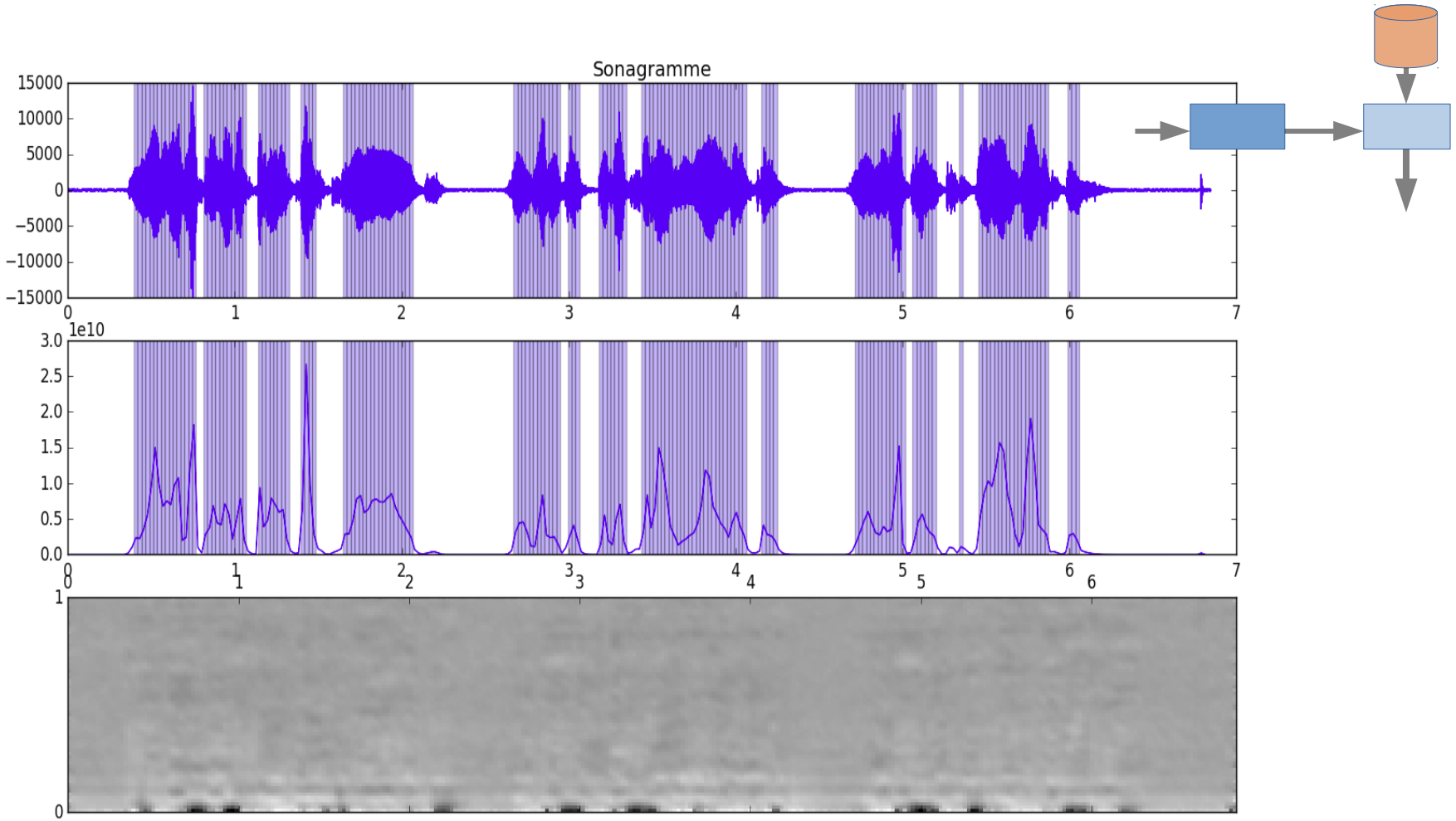
Cependant, le spectre varie dans le temps:

→ On découpe le signal en trames d’une longueur fixe

→ On calcule les énergies sur chaque trame

→ On calcule les CC sur chaque trame d’énergie suffisante

III. Mise en pratique



III. Mise en pratique

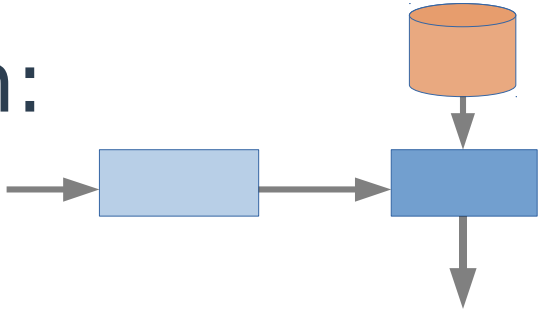
On le teste sur des “a”: on moyenne et on compare:

→ Résultats tout aussi mauvais que précédemment.

→ L’étape de comparaison doit être fausse

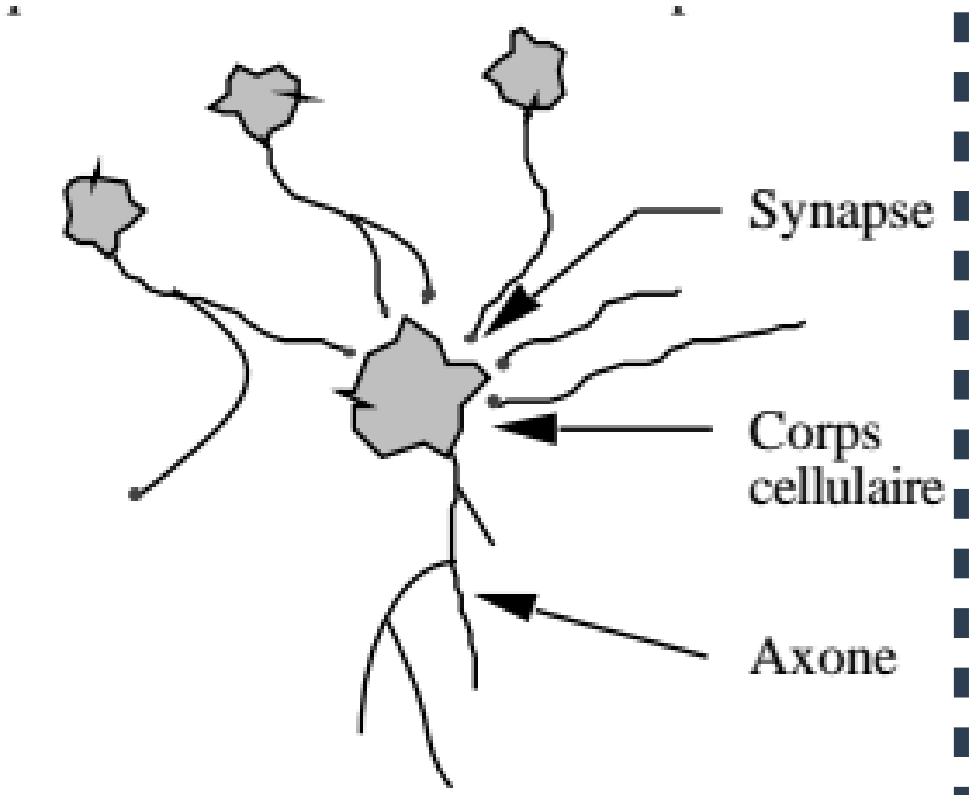
III. Mise en pratique

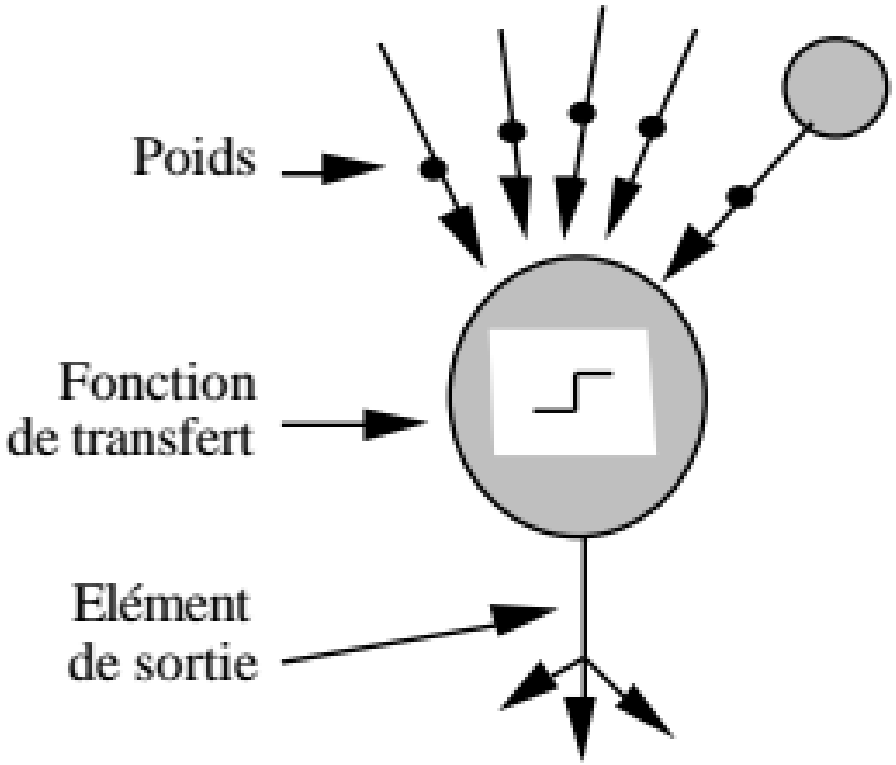
On change donc de comparaison:



→ On va utiliser un réseau de neurones

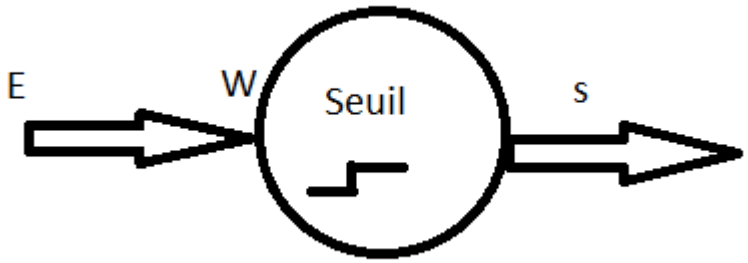
III. Mise en pratique





III. Mise en pratique

Le perceptron :



$E = (e_1,\ldots,e_n)$
$W = (w_1,\ldots,w_n)$

$s_b = \Sigma e_i.w_i$
$s = f(s_b)$
Fonction de seuil f

Perceptrons en parallèle → réseau à une couche

III. Mise en pratique

Apprentissage :

→ Faire évoluer les poids pour que la sortie se rapproche de la sortie voulue:

→ $\mathbf{w_{i+1} = w_i + (s - s_b) * h * e_i}$

(h un pas qu'on choisi pour faire varier la vitesse d'évolution de w)

→ On teste ensuite ce perceptron avec un autre "a" test.

III. Mise en pratique

Résultat :

→ Taux de reconnaissance de 100% à 16 locuteurs.

→ Le réseau est, pour chaque locuteur, sûr à 90% de sa réponse.

Conclusion

A partir de l’alanyse cepstrale d’un signal, on peut identifier un locuteur parmi une base de départ.

→ Ouverture : on peut améliorer les CC par fenêtrage, changement d’échelle (Mel), ou les réseaux de neurones (multicouche).

Merci !