TRAITEMENT DE LA PAROLE

RECONNAISSANCE DU LOCUTEUR

Le Mans Université

Anthony Larcher - Le Mans Université - 2019

PLAN DU COURS

Contexte

- Rappel sur les super-vecteurs
- Compression d'information

Factor Analysis

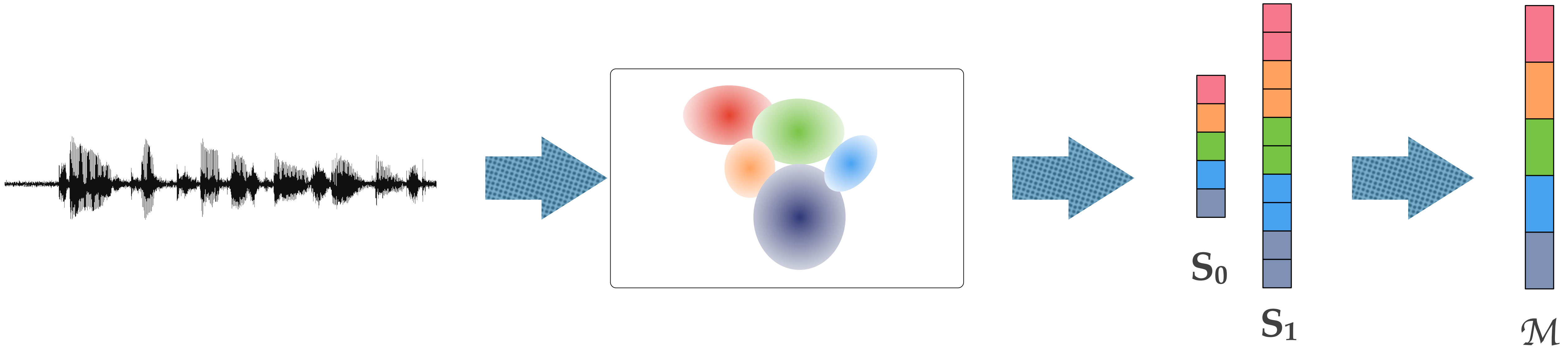
- Motivations
- Théorie
- Analyse Linéaire Discriminante Probabiliste (PLDA)

RAPPEL SUR LES SUPER-VECTEURS

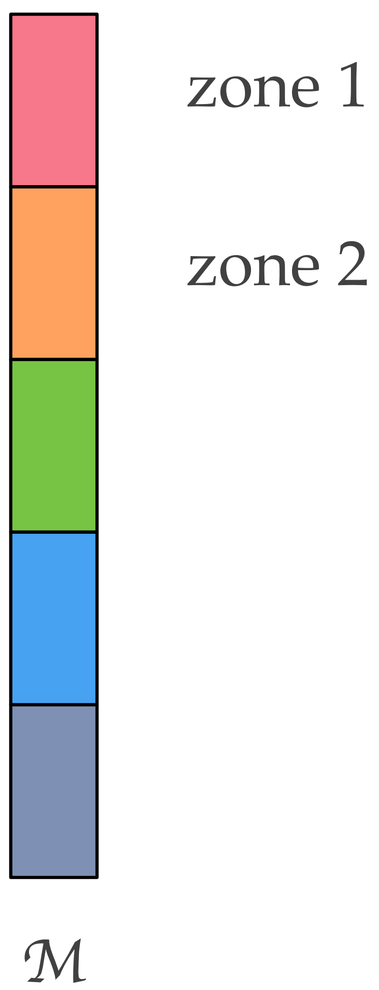
- on représenter un locuteur par un GMM
- on adapte seulement les moyennes
- un locuteur = un super-vecteur de moyennes
- on va classifier (séparer) les locuteurs dans un espace de très grande dimension (~20 000 dimensions)

RAPPEL SUR LES SUPER-VECTEURS

- Étant donné une partition de l'espace acoustique
- L'information portée par un échantillon de parole est représentée par les statistiques d'ordre 0 et 1
- … et condensée dans un super-vecteur

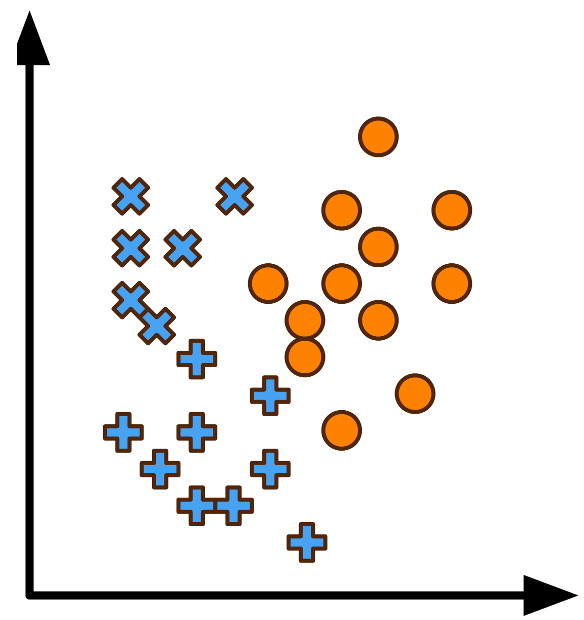


RAPPEL SUR LES SUPER-VECTEURS



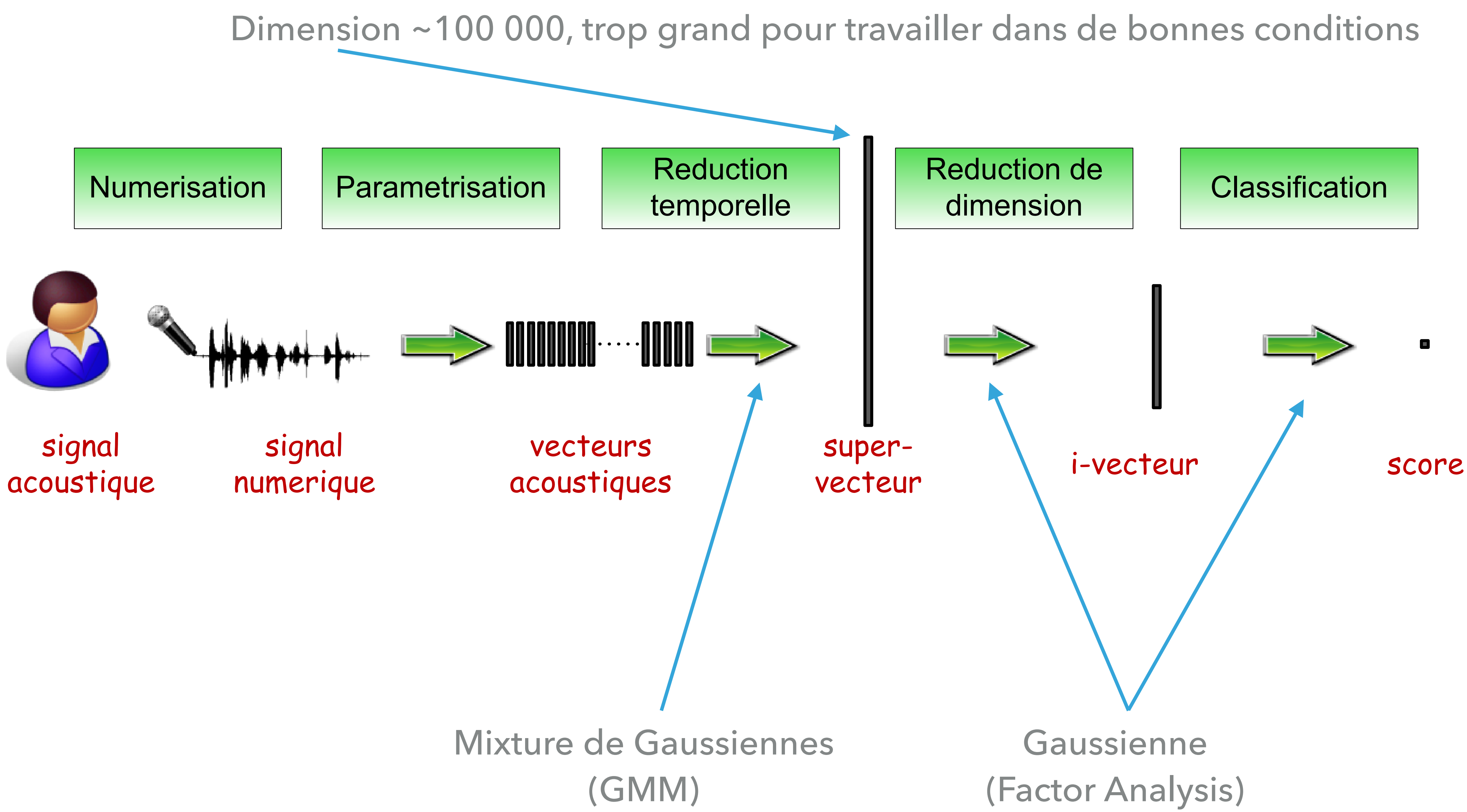
- 1 échantillon de parole = 1 super-vecteur
- 1 super-vecteur contient:
  locuteur + canal + bruit + langue + …

RAPPEL SUR LES SUPER-VECTEURS

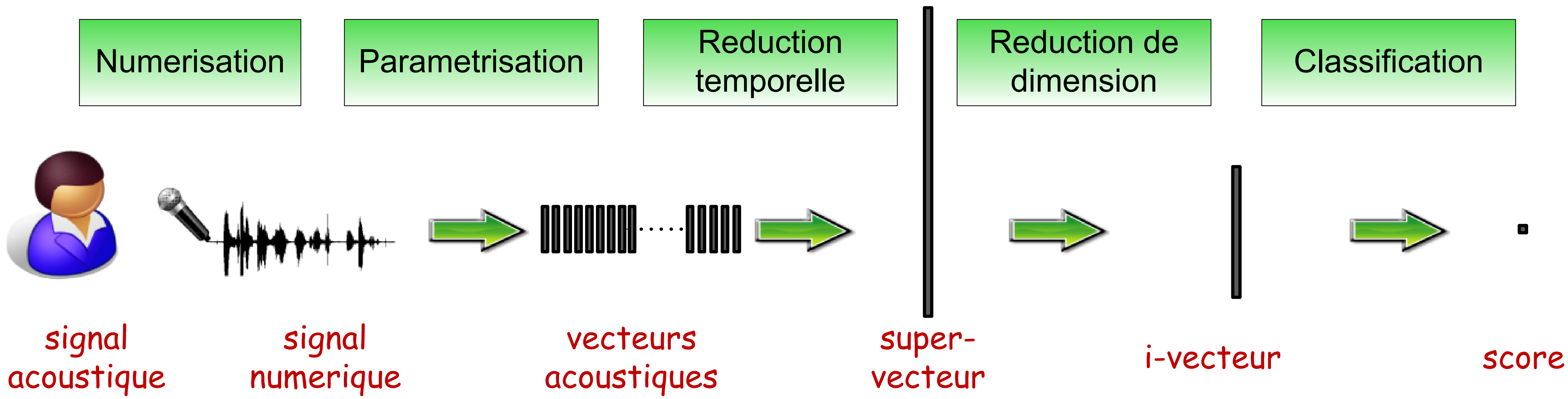


- Représentation de locuteurs dans un espace de très grande dimension
- 1 point = un enregistrement (de durée variable)
- Problématique: séparer les locuteurs
- ATTENTION: nous sommes dans l'espace des super-vecteurs: 1 point = 1 segment de parole >= 20 000 paramètres

COMPRESSION D'INFORMATION



COMPRESSION D'INFORMATION

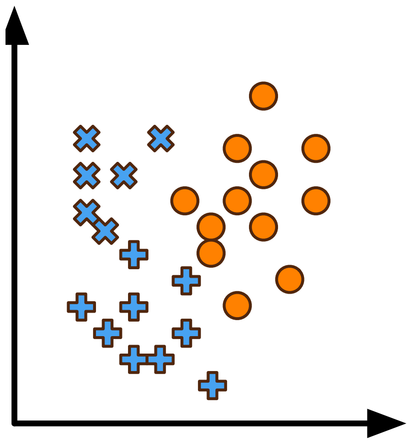


<u>Note:</u>

Le Factor Analysis est utilisé 2 fois

Les i-vecteurs ne sont pas détaillés ici. Leur extraction utilise les statistiques d'ordre 0 et 1 calculés avec un modèles GMM et un Factor Analysis multi-Gaussien.

LE FACTOR ANALYSIS POUR DISCRIMINER LES LOCUTEURS



Problème de la variabilité:

- Un super-vecteur contient: locuteur + canal + bruit…
- Hypothèse 1: les super-vecteurs suivent une loi Gaussienne*
- Hypothèse 2: l'espace des locuteurs est plus petit que l'espace des super-vecteurs
- On souhaite trouver le sous-espace des locuteurs qui maximise la séparabilité des locuteurs (maximise la variabilité inter-locuteurs)
- Hypothèse 3: la source principale de variabilité est le locuteur

* hypothèse simplificatrice pour introduire le Factor Analysis, sera remise en cause par la suite

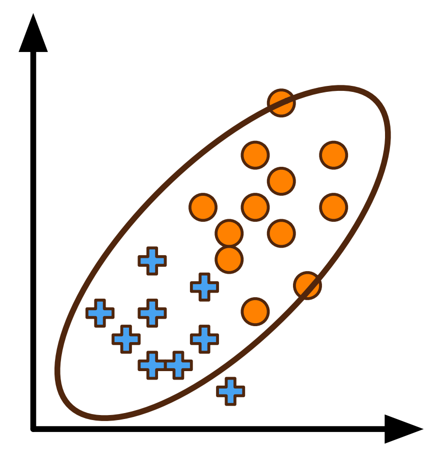
RECONNAISSANCE DU LOCUTEUR

FACTOR ANALYSIS

Le Mans Université

Anthony Larcher - Le Mans Université - 2019

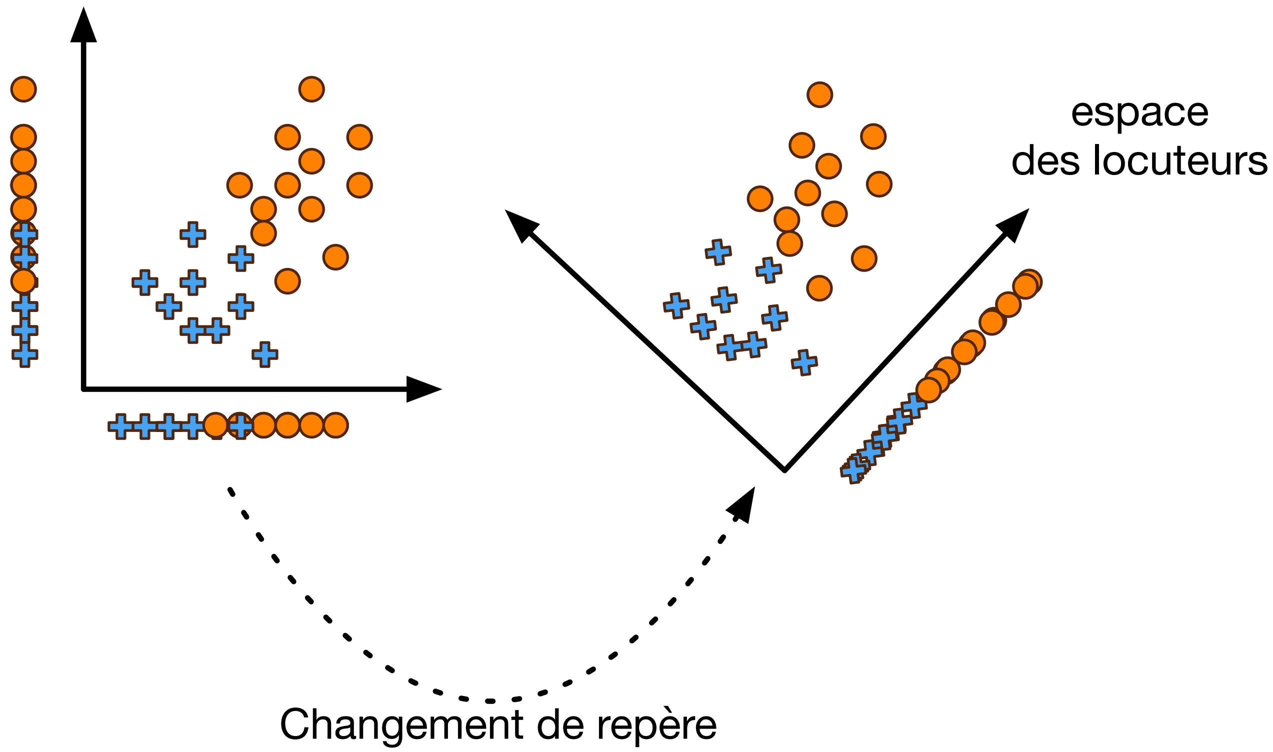
FACTOR ANALYSIS: MOTIVATIONS



Graphiquement:

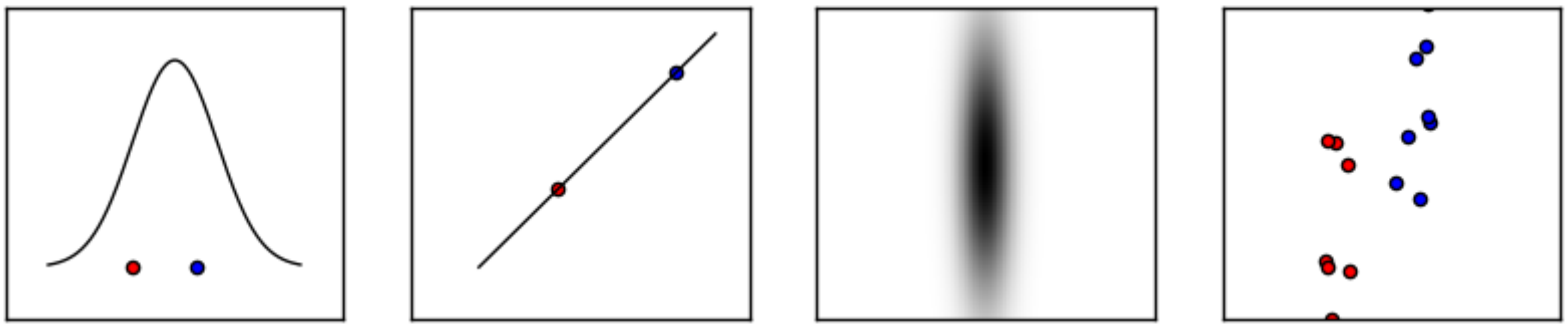
- Hypothèse 1: les super-vecteurs suivent une loi Gaussienne
- Hypothèse 2: l'espace des locuteurs est plus petit que l'espace des super-vecteurs: espace de dimension 1 (sur cet exemple)
- Hypothèse 3: la source principale de variabilité est le locuteur (« axe principaux » de la distribution)

FACTOR ANALYSIS: MOTIVATIONS



- Objectif: trouver le sous-espace qui maximise la variabilité interlocuteurs

FACTOR ANALYSIS

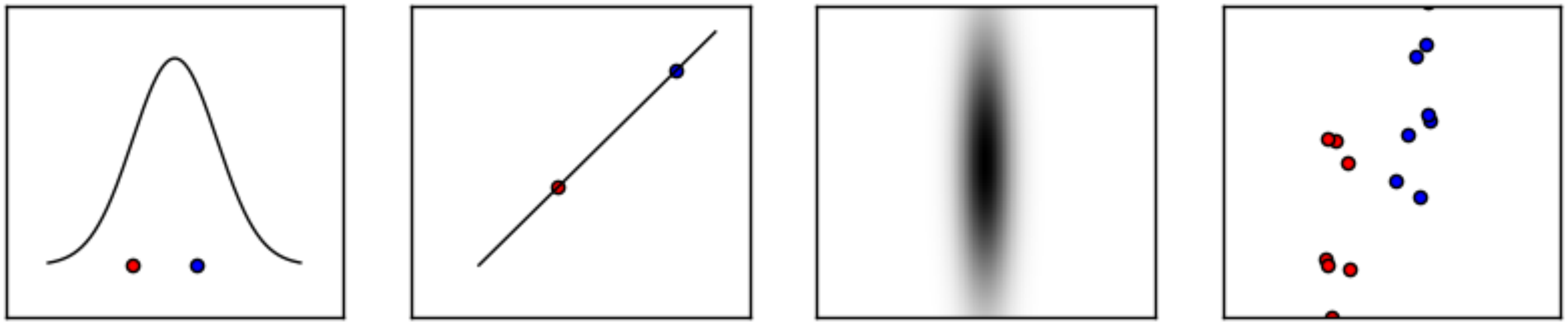


$\mathcal{N}_y(0, 1)$ $\quad m + \mathbf{V}y$ $\quad \mathcal{N}_\epsilon(0, \Sigma)$ $\quad m + \mathbf{V}y + \epsilon$

Génération d'un super-vecteur:

- Tirage aléatoire de 2 locuteurs
- Projection dans le sous-espace des locuteurs
- distribution du bruit (diagonale)
- Distributions de super-vecteurs finales pour 2 locuteurs

FACTOR ANALYSIS: EIGENVOICES

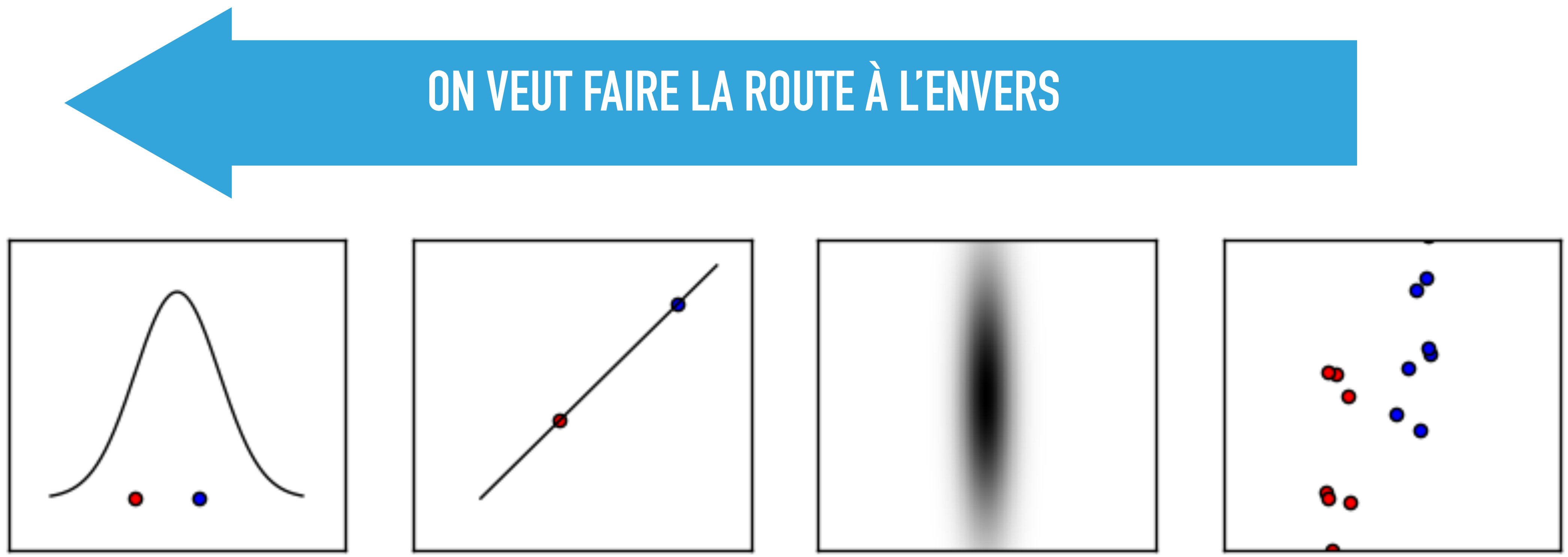


$\mathcal{N}_y(0, 1)$ $\quad m + \mathbf{V}y$ $\quad \mathcal{N}_\epsilon(0, \Sigma)$ $\quad m + \mathbf{V}y + \epsilon$

Interprétation:

Tous les super-vecteurs d'un même locuteur sont générés à partir d'une unique valeur de y: y est un vecteur qui caractérise le locuteur dans un espace de dimension réduite.

FACTOR ANALYSIS



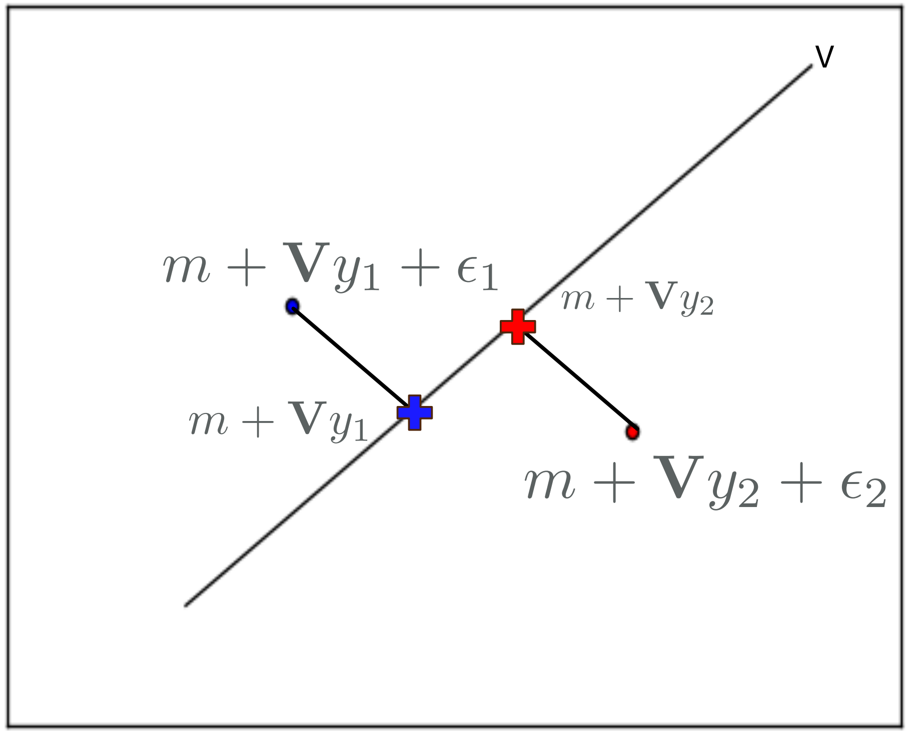
$\mathcal{N}_y(0,1)$ $\quad m + \mathbf{V}y$ $\quad \mathcal{N}_\epsilon(0,\Sigma)$ $\quad m + \mathbf{V}y + \epsilon$

Difficulté: on observe des vecteurs et on veut estimer y

(applicable aux super-vecteurs, i-vecteur, x-vecteurs…)

FACTOR ANALYSIS

- On estime m et V à partir de données d'apprentissage
- Lorsqu'on reçoit un échantillon (super-vecteur), on estime y et on l'utilise pour comparer les locuteurs



FACTOR ANALYSIS

THÉORIE

FACTOR ANALYSIS: THÉORIE

Une nouvelle étape de modélisation Gaussienne:

- 1 vecteur = 1 enregistrement
- Hypothèse: la distribution de **l'ensemble des sessions (de tous les locuteurs)** est Gaussienne
- On veut estimer une Gaussienne de **grande dimension**
- On suppose que l'information « locuteur » est portée par un sous espace de dimension plus réduite.

FACTOR ANALYSIS: THÉORIE

$$Pr(x) = \mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma)$$

- vecteur: x
- Vecteur moyen: μ
- Matrice de covariance composée de 2 termes:
  - matrice diagonale Σ de rang D (dimension de l'espace)
  - matrice $\Phi\Phi^T$ pleine de rang K. Φ est de dimension D x K

FACTOR FACTOR ANALYSIS: THÉORIE

$$Pr(x) = \mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma)$$

- Φ est de dimension D x K
- K << D (K ~ 100 et D ~ 5 000)
- Problème: estimer $\Phi\Phi^T + \Sigma$

FACTOR FACTOR ANALYSIS: THÉORIE

$$Pr(x) = \mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma)$$

Note:

si Σ est sphérique (multiple de l'identité), ce modèle s'appelle une Analyse en Composante Principale Probabiliste (PPCA) et les paramètres peuvent être calculés directement (sans algorithme EM).

FACTOR FACTOR ANALYSIS: THÉORIE

$$Pr(x) = \mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma)$$

Quel est le lien avec ce qu'on a vu précédemment?

$$\mathcal{M} = m + \mathbf{V}y + \epsilon$$

FACTOR ANALYSIS: THÉORIE

Le Factor Analysis comme une marginalisation:

$$\boxed{\mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma) = \int Pr(x|h)Pr(h)dh} = \int Pr(x,h)dh = Pr(x)$$

avec:

$$Pr(h) = \mathcal{N}_h(0, \mathbf{I})$$

$$Pr(x|h) = \mathcal{N}_x(\mu + \Phi h, \Sigma)$$

où I est la matrice identité.

(preuve fournie dans un document annexe)

FACTOR FACTOR ANALYSIS: THÉORIE

Le Factor Analysis comme une marginalisation:

$$\mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma) = \int Pr(x|h)Pr(h)dh = \int Pr(x,h)dh = Pr(x)$$

avec:

$$Pr(h) = \mathcal{N}_h(0, \mathbf{I})$$

$$Pr(x|h) = \mathcal{N}_x(\mu + \Phi h, \Sigma) \Rightarrow \boxed{x = \mu + \Phi h + \epsilon}$$

avec: $\epsilon \sim \mathcal{N}(0, \Sigma)$

(preuve fournie dans un document annexe)

FACTOR FACTOR ANALYSIS: THÉORIE

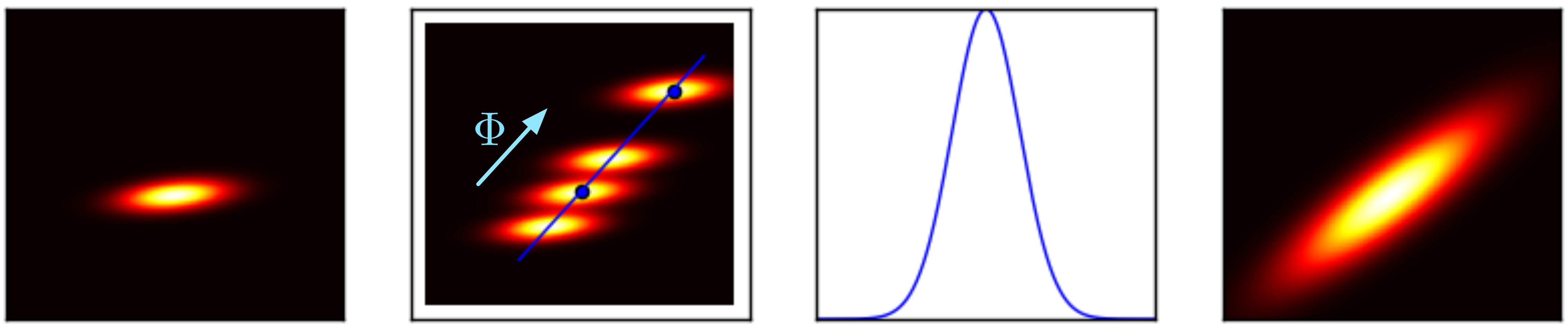
- Un locuteur = un point dans l'espace des locuteurs: h
- On passe d'un espace de faible dimension à l'espace observé (grande dimension): projection selon Φ
- Lorsqu'on observe, il y a du bruit: $+\epsilon$

$$x = \mu + \Phi h + \epsilon$$

FACTOR FACTOR ANALYSIS: THÉORIE

A quoi sert la variable latente « h »?

Que signifie vraiment: $\int Pr(x|h)Pr(h)dh$?



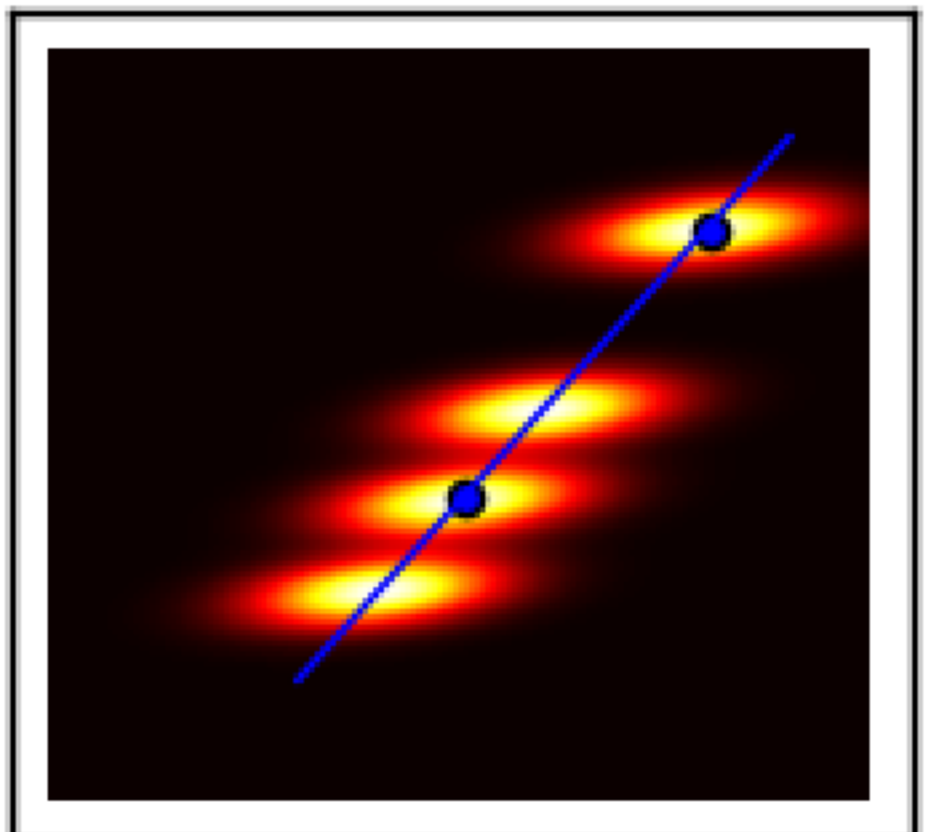
$Pr(x|h_1)$ $Pr(x|h_2)$ $Pr(x|h_3)$ $\mathcal{N}_h(0, \mathbf{I})$ $\int Pr(x|h)Pr(h)dh$

On calcule une somme pondérée infinie (intégrale) de distributions Pr(x|h)
Pour toutes les valeurs de « h » avec une probabilité (un poids) qui est Pr(h)

FACTOR FACTOR ANALYSIS: THÉORIE

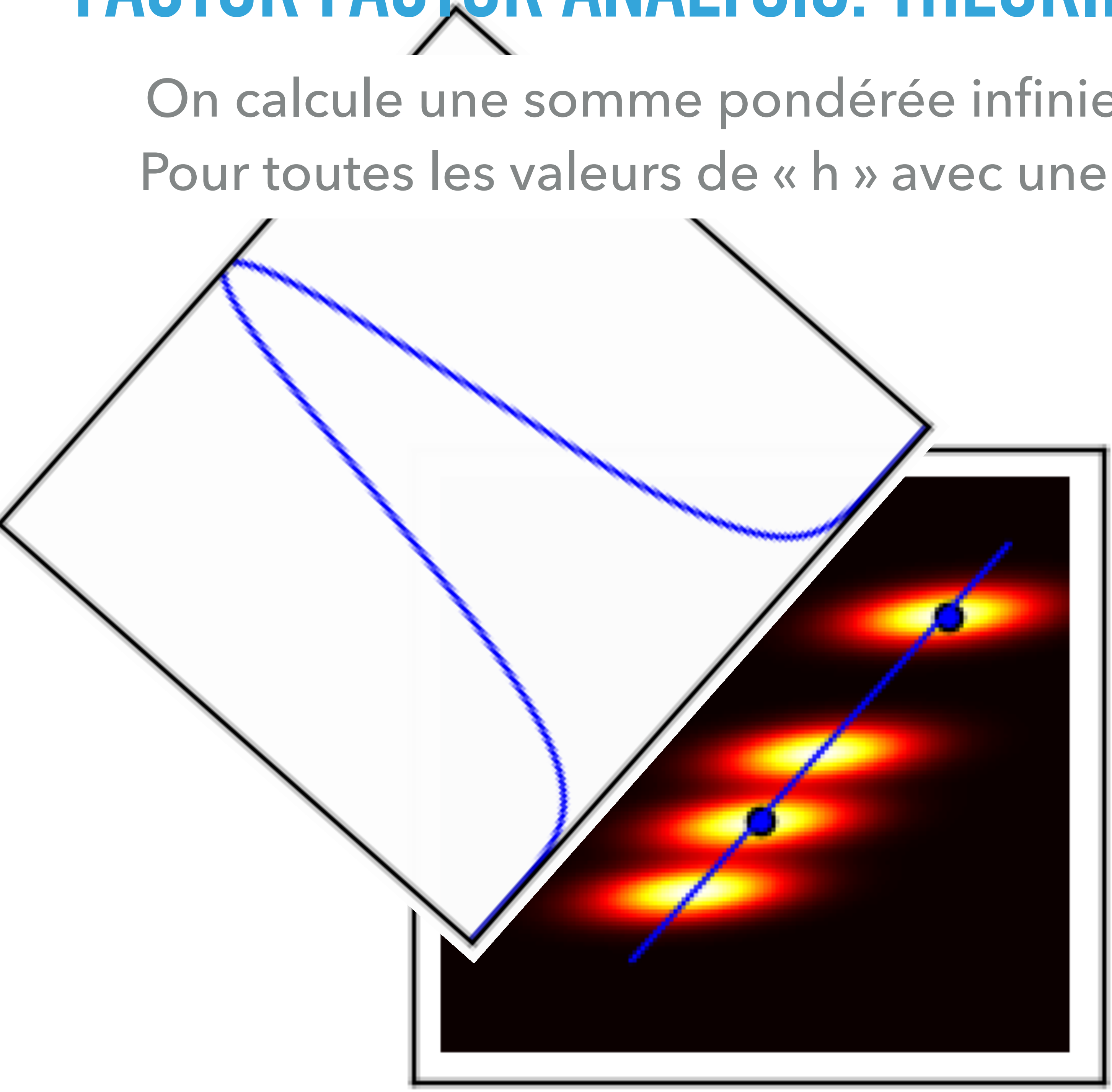
On calcule une somme pondérée infinie (intégrale) de distributions Pr(x|h)
Pour toutes les valeurs de « h » avec une probabilité (un poids) qui est Pr(h)

Si on choisit un locuteur dans l'espace de dimension réduite,
la distribution de ses observations est une Gaussienne dont la Covariance ne dépend pas du locuteur



FACTOR FACTOR ANALYSIS: THÉORIE

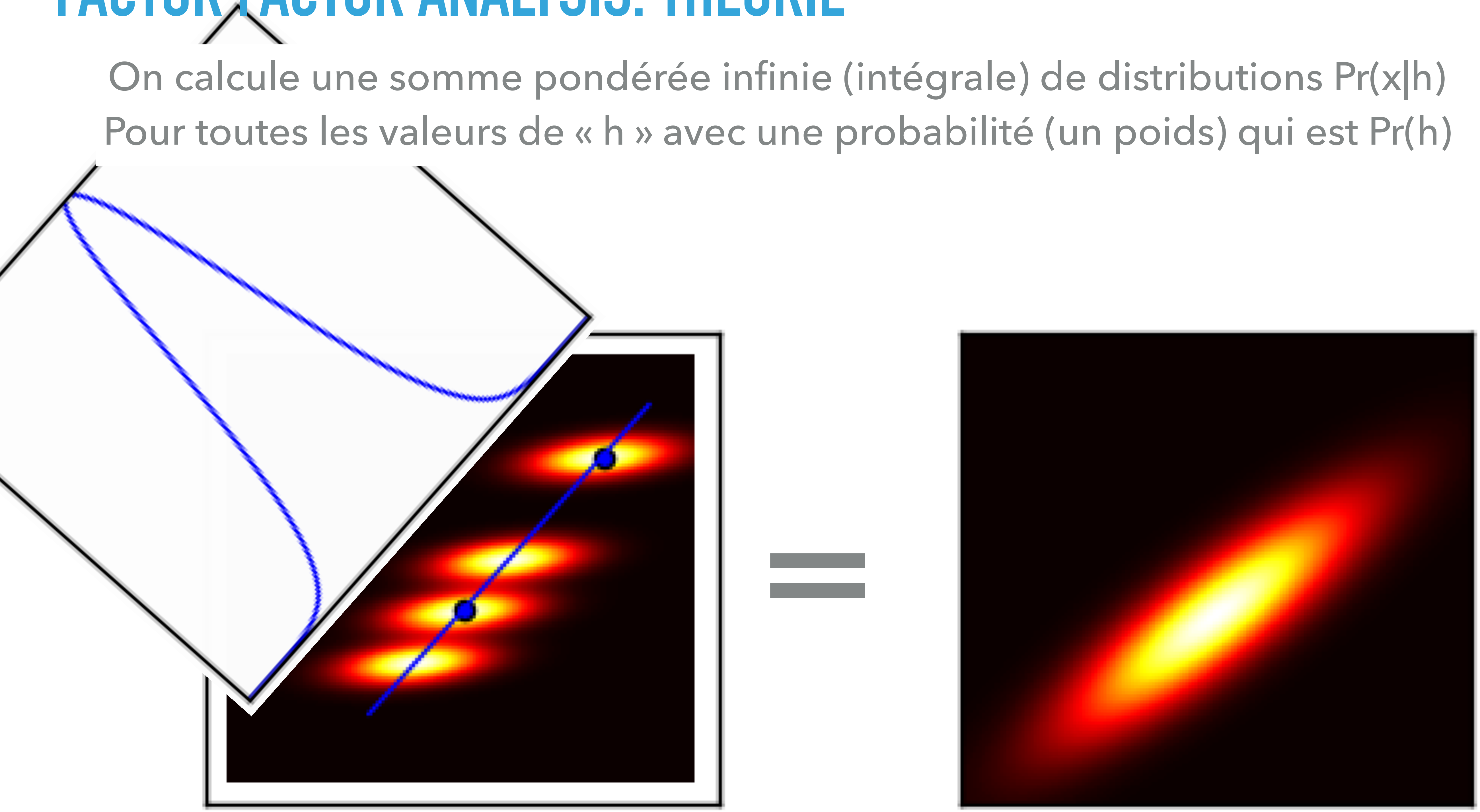
On calcule une somme pondérée infinie (intégrale) de distributions Pr(x|h)
Pour toutes les valeurs de « h » avec une probabilité (un poids) qui est Pr(h)



Dans l'espace des locuteurs, la distribution des locuteurs est Gaussienne (toutes les voix varient autour d'une voix moyenne).

FACTOR FACTOR ANALYSIS: THÉORIE

On calcule une somme pondérée infinie (intégrale) de distributions Pr(x|h)
Pour toutes les valeurs de « h » avec une probabilité (un poids) qui est Pr(h)



FACTOR ANALYSIS: APPRENTISSAGE

$$Pr(x) = \mathcal{N}_x(\mu, \Phi\Phi^T + \Sigma)$$

- Apprentissage d'un modèle complexe
- Utilisation de variables latentes

On estime Φ grâce à l'algorithme EM.

(preuve non fournie dans ce cours mais cas d'une seule Gaussienne multi-variée en annexe)

FACTOR ANALYSIS

SUITE DE L'HISTOIRE…

Le Mans Université

FACTOR ANALYSIS: SUITE DE L'HISTOIRE

Système	Equal Error Rate	Commentaire
GMM/UBM (MAP)	8.1 %	-
EigenChannel	5.22 %	Supprime l'effet canal
Joint Factor Analysis	3.11 %	Supprime l'effet canal Apprentissage discriminant

KINNUNEN, Tomi et LI, Haizhou. An overview of text-independent speaker recognition: From features to supervectors. *Speech communication*, 2010, vol. 52, no 1, p. 12-40.

FACTOR ANALYSIS: SUITE DE L'HISTOIRE

Le *Joint Factor Analysis* amène un gain conséquent:

-61% d'EER relatif

FACTOR ANALYSIS: I-VECTORS

$$\mathcal{M} = m + \mathbf{T}w + \epsilon$$

Super-vecteur

Résidu

Vecteur moyen
UBM

Variable latente

Matrice de Variabilité Totale
(Total Variability)

FACTOR ANALYSIS: I-VECTORS

$$\mathcal{M} = m + \mathbf{T}w + \epsilon$$

Super-vecteur
~50 000

Vecteur moyen
UBM
~50 000

Matrice de Variabilité Totale
(Total Variability)

500 x 50 000

Variable latente

~500

Résidu
~50 000

FACTOR ANALYSIS: I-VECTORS

Avantages et inconvénients des i-vecteurs:

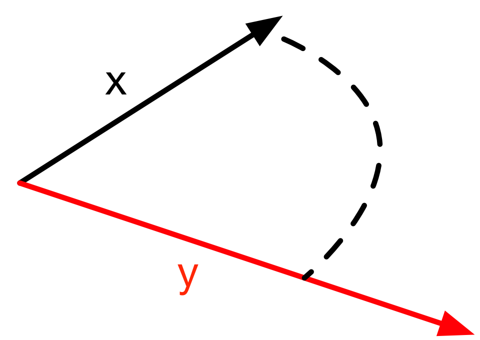
- représentation d'une « session » (locuteur, canal, langue, émotion, bruit…)
- dimension réduite et fixe

De nombreuses méthodes de classifications existaient dans la littérature pour des dimensions raisonnables
Les i-vecteurs ont ouverts de nombreuses possibilités

FACTOR ANALYSIS ET I-VECTEURS

Reconnaissance du locuteur dans l'espace des i-vecteurs

- Similarité cosine



$$score = \frac{< x, y >}{|x|_2 \times |y|_2}$$

- où alors on refait des Gaussiennes…? (PLDA, Gaussian backend)

$$x = \mu + \Phi h + \psi s + \epsilon$$

i-vecteur

moyenne

locuteur

canal

résidu

RÉFÉRENCES

Modèles Gaussiens:

BIMBOT, Frédéric, BONASTRE, Jean-François, FREDOUILLE, Corinne, *et al.* A tutorial on text-independent speaker verification. *EURASIP journal on applied signal processing*, 2004, vol. 2004, p. 430-451.

REYNOLDS, Douglas A. et ROSE, Richard C. Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE transactions on Speech and Audio Processing*, 1995, vol. 3, no 1, p. 72-83.

I-vecteurs:

DEHAK, Najim, DEHAK, Reda, KENNY, Patrick, *et al.* Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification. In : *Tenth Annual conference of the international speech communication association*. 2009.

DEHAK, Najim, KENNY, Patrick J., DEHAK, Réda, *et al.* Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, vol. 19, no 4, p. 788-798.

PLDA et Factor Analysis

PRINCE, Simon JD. *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.

Réseaux de neurons profonds pour la reconnaissance du locuteur

LEI, Yun, SCHEFFER, Nicolas, FERRER, Luciana, *et al.* A novel scheme for speaker recognition using a phonetically-aware deep neural network. In : *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014. p. 1695-1699.