Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences Western Switzerland

MASTER OF SCIENCE IN ENGINEERING

# Multimodal Processing, Recognition and Interaction

Practical Information
Introduction

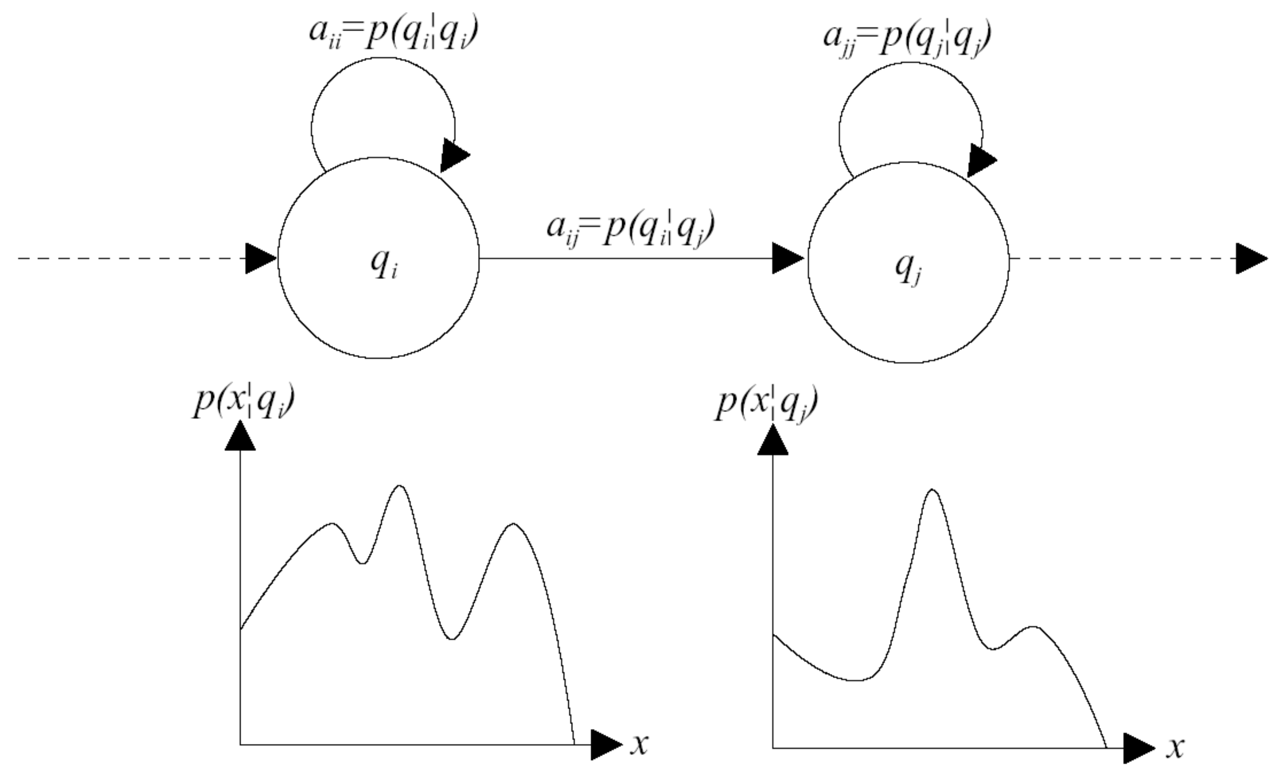
Elena Mugellini, Jean Hennebert, Stefano Carrino

# Modèles de Markov Cachés (HMMs) Partie 2

- Rappels
- Entraînement des HMMs
  - Réajustement des paramètres du HMM (A,B, π) pour maximiser P(X |M)
- Types de probabilités d'émission
  - Probabilités d'émission discrètes
  - Probabilités d'émission continues
    - Mixture de gaussiennes
- Applications
- TP

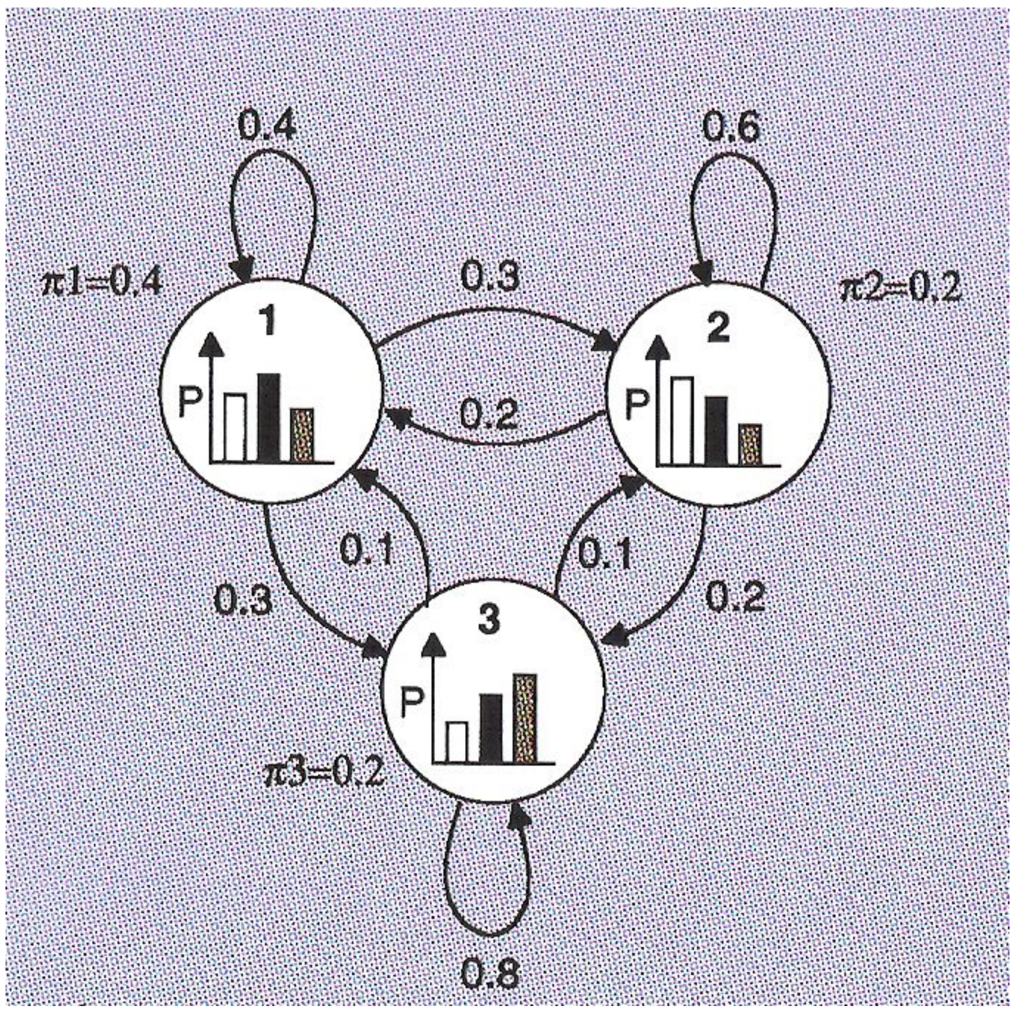
# RAPPELS

# Rappels



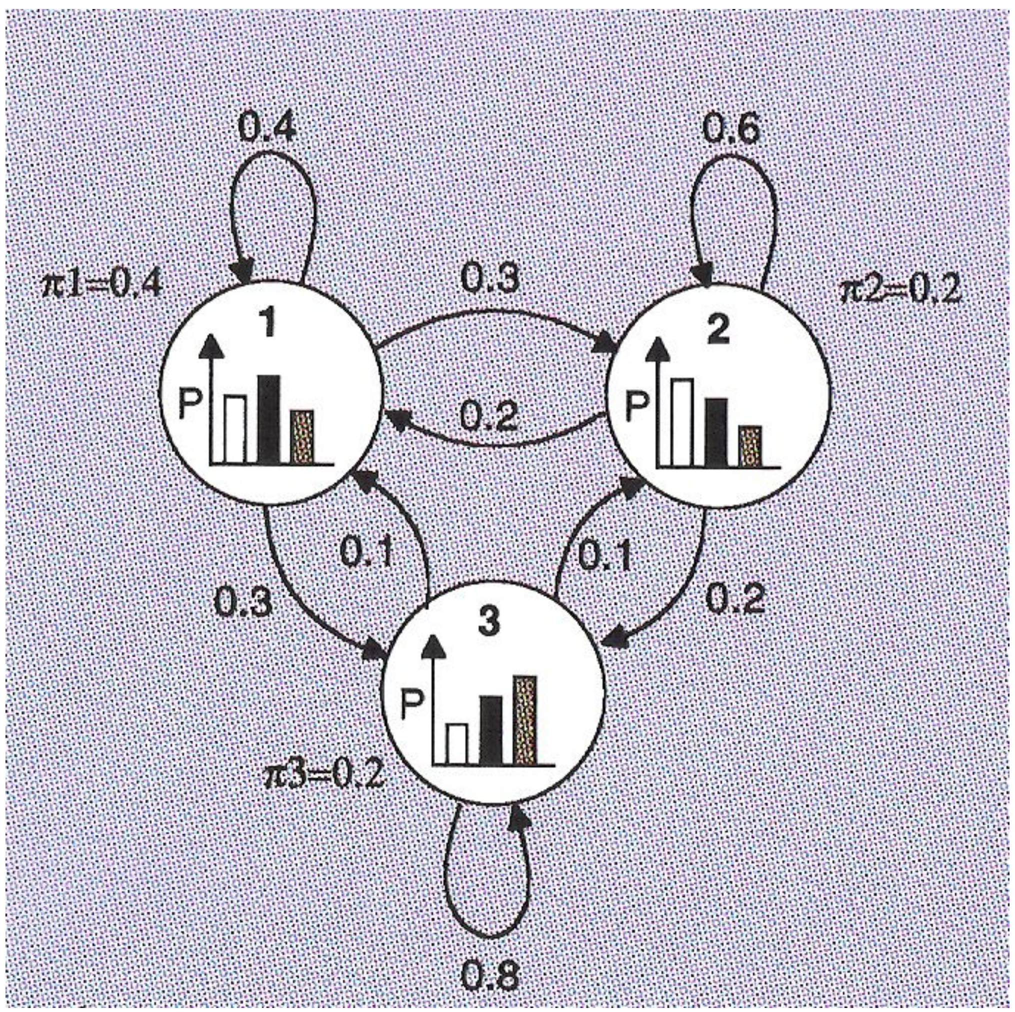

- Formulation du problème: une séquence observée est émise par des lois probabilistes qui dépendent d'un état particulier du système
- Dans la plupart des cas, l'état dans lequel se trouve le système est inconnu. On dit que la séquence des états est « cachée ».
- L'observation $x_n$ a une certaine probabilité d'être « émise » dans un état q
  - On parle de probabilités d'émission p(x|q)
- Le système va d'un état i à un état j avec une certaine probabilité de transition $a_{ij}$

# Les éléments d'un HMM




- *A*, l'ensemble des probabilités de transition
- *B*, l'ensemble des densités de probabilité d'observation
- $\pi$, l'ensemble des probabilités d'état initial
- $M=(A, B, \pi)$

# Les éléments d'un HMM




$q_t$ = état au temps $t$

$A = \{a_{ij}\}$ avec $a_{ij} = P(q_t = j \mid q_{t-1} = i)$

$B = \{b_j(x)\}$ avec $b_j(x) = P(x \mid q_t = j)$

$\pi = \{\pi_i\}$ avec $\pi_i = P(q_1 = i)$

$M = (A, B, \pi)$

Séquence d'observations

$X = \{x_1, \ldots, x_n\}$

Séquence d'états particulière

$\mathrm{q} = \{q^1, \ldots, q^n\}$

# Problèmes résolus par les HMMs

- Génération d'une séquence d'observations X
  - « on peut générer une séquence d'observations avec un modèle donné »
- Reconnaissance
  - Probabilité de la génération d'une séquence
    - connaissant X et q (simple multiplication des probabilités de transition et d'émission)
    - connaissant X (algorithme forward)
    - « on peut calculer une probabilité (ou 'vraisemblance') d'une observation X étant donné un HMM »
    - autrement dit, « on peut calculer la 'proximité' (ou 'correspondance') entre une observation et un modèle donné »

# Problèmes résolus par les HMMs - New

- Reconnaissance
  - Alignement: Séquence d'état optimale connaissant X
    - Algorithme de Viterbi
    - « on peut associer un état à chaque observation »
    - autrement dit, « on peut savoir où on passe dans le HMM et à quel moment on y passe »
- Entraînement
  - Réajustement des paramètres du HMM (A,B, π)
  - « comment trouver (A,B, π) pour maximiser $P(X_i | M)$ sur un ensemble d'observations $\{X_1, \ldots, X_M\}$ »
  - « comment trouver (A,B, π) pour faire coller au mieux le HMM par rapport au set d'entraînement »

# Alignement: séquence d'états optimale connaissant $X$

- Hypothèse
  - La séquence d'observations est connue
  - La séquence d'états est inconnue
- On veut déterminer la séquence d'états qui maximise la probabilité que le modèle génère l'observation: $P(X,Q_{max}|M)$
- Algorithme de Viterbi:

On considère la variable "forward"

$$\delta_{\mathrm{n}}(i) = \max_{q_1, q_2, \ldots, q_{n-1}} P(x_1 x_2 \ldots x_n, q_1 q_2 \ldots q_{n-1}, q_n = i \mid M)$$

1. Initialisation

$$\delta_1(i) = \pi_i b_i(x_1)$$

2. Récursion

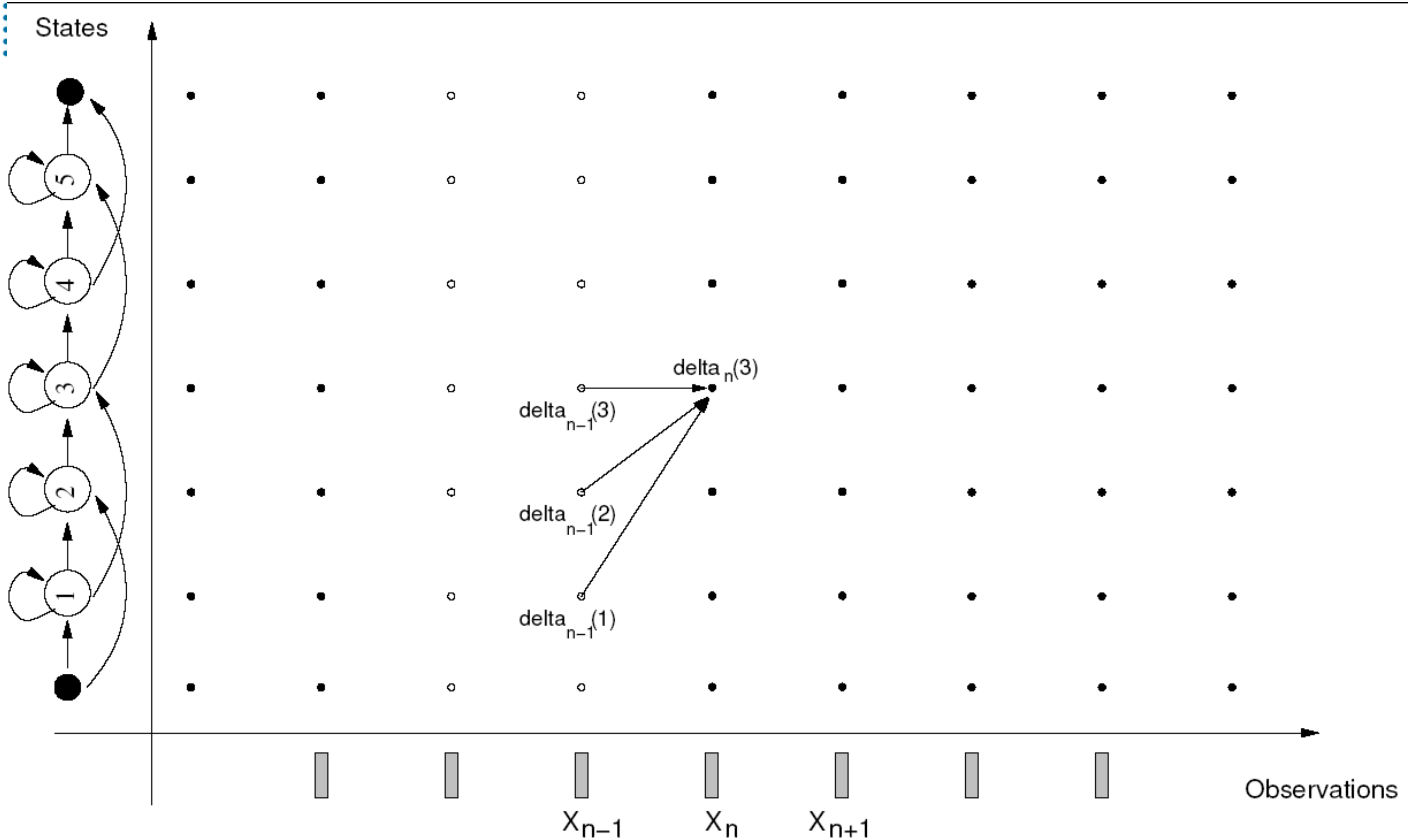
$$\delta_{\mathrm{t}}(j) = \max_{1 \le i \le Q} \left[ \delta_{t-1}(i) a_{ij} \right] b_j(x_t)$$

3. Terminaison

$$\mathrm{P}(\mathrm{X}, \mathrm{Q}_{\max} \mid \mathrm{M}) = \max_{1 \le i \le Q} [\delta_{\mathrm{N}}(i)]$$

# Alignement: séquence d'état optimale connaissant $X$

$$\delta_{\underset{\text{Pas}}{\mathrm{n}}}(\overset{\text{Etat}}{i}) = \max_{q_1, q_2, \ldots, q_{n-1}} P(\overbrace{x_1 x_2 \ldots x_n}^{\text{Observations}}, \overbrace{q_1 q_2 \ldots q_{n-1}, q_n}^{\text{Séquence d'état}} = i \mid \overbrace{M}^{\text{Modèle}})$$



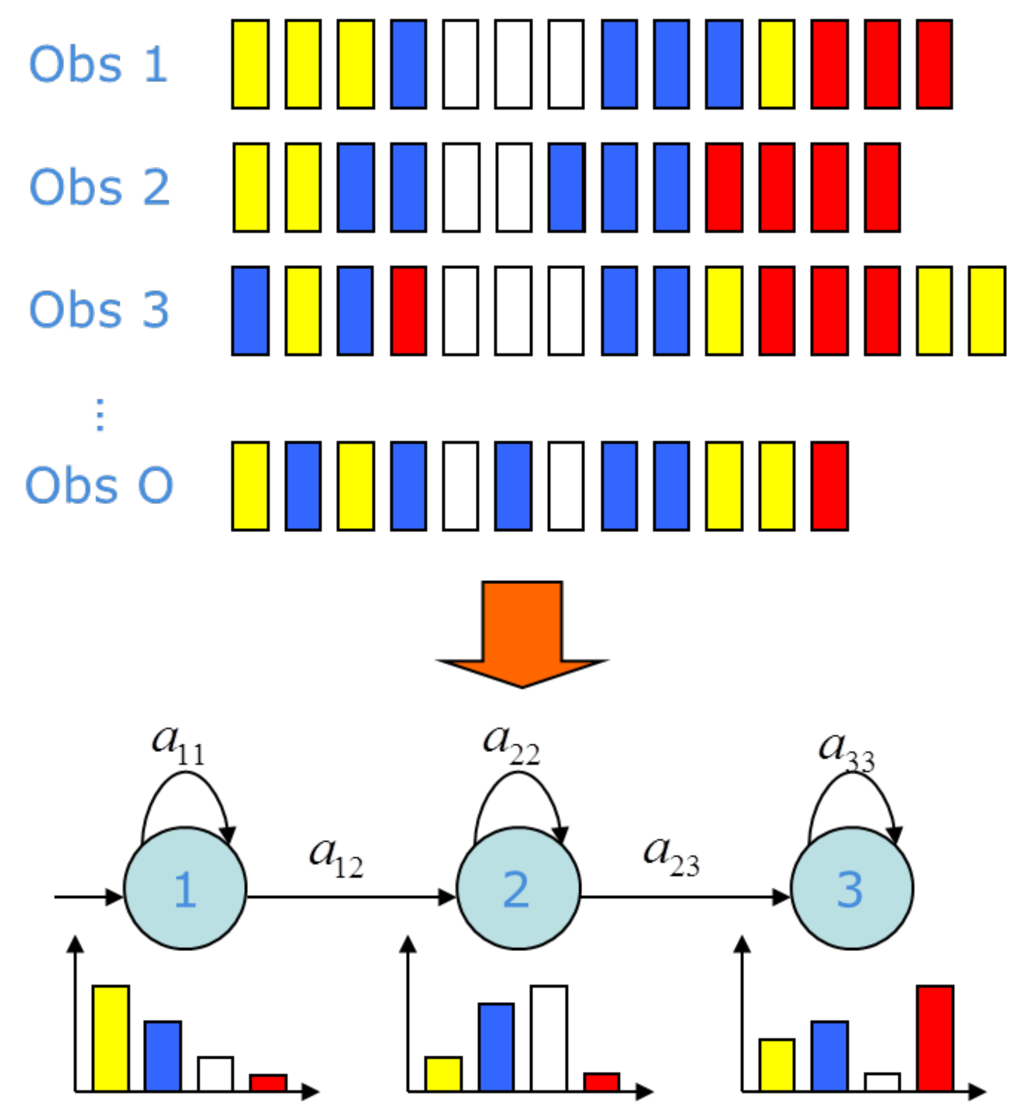

**Viterbi algorithm** propagates from left to right to compute the deltas. In this figure, $delta_n(3)$ is computed by looking at the highest value of the potential preceding deltas times the corresponding transition probabilities $a_{ij}$. In the example illustrated above, $delta_{n-1}(2).a_{23} > delta_{n-1}(3).a_{33}$ and $delta_{n-1}(2).a_{23} > delta_{n-1}(1).a_{13}$ . The best transitions leading from one point of the lattice to the next are then kept in memory to allow backtracking from the last state to the first.

Réajustement des paramètres du HMM (A,B, $\pi$) pour maximiser P(X |M)

# ENTRAÎNEMENT DES HMMS

# Enoncé du problème

- On dispose d'un ensemble de séquences d'observations
- On veut se construire un HMM qui « explique » au mieux ces observations



# Enoncé du problème

- Mathématiquement, il faut déterminer les différents paramètres du HMM qui maximisent l'ensemble des probabilités d'observer les séquences.

$$M = (A, B, \pi)$$

avec

$$A = \{a_{ij}\}$$

$$B = \{b_j(x)\}$$

$$\pi = \{\pi_i\}$$

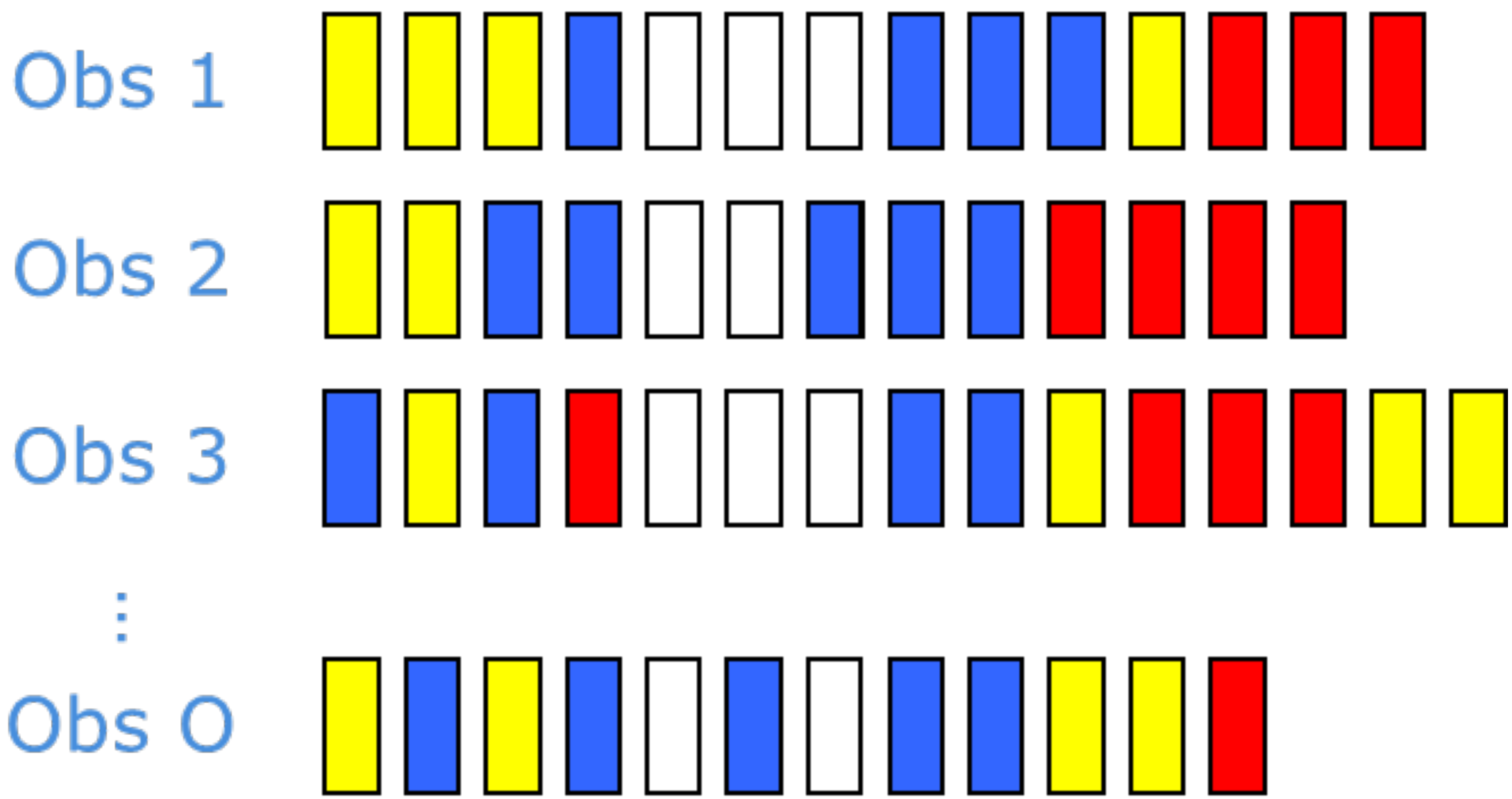
# Réajustement des paramètres du HMM (A,B, π) pour maximiser P(X |M)

Méthode d'entraînement utilisant le critère Viterbi:

1. On aligne l'ensemble des observations avec Viterbi
2. On accumule les ensembles d'observations dans les états
3. On ré-estime
   - Les probabilités d'émission
   - Les probabilités de transition
4. Retour en 1 et on itère jusqu'à la convergence

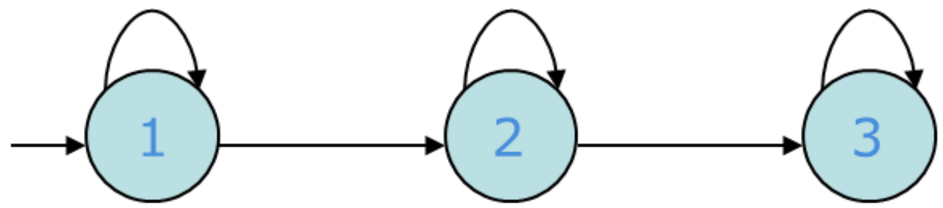
Remarques:
- Il faut définir une topologie initiale du HMM
- Comment démarrer?
  - E.g. Alignement linéaire

# Exemple: Lego – 3 Urnes

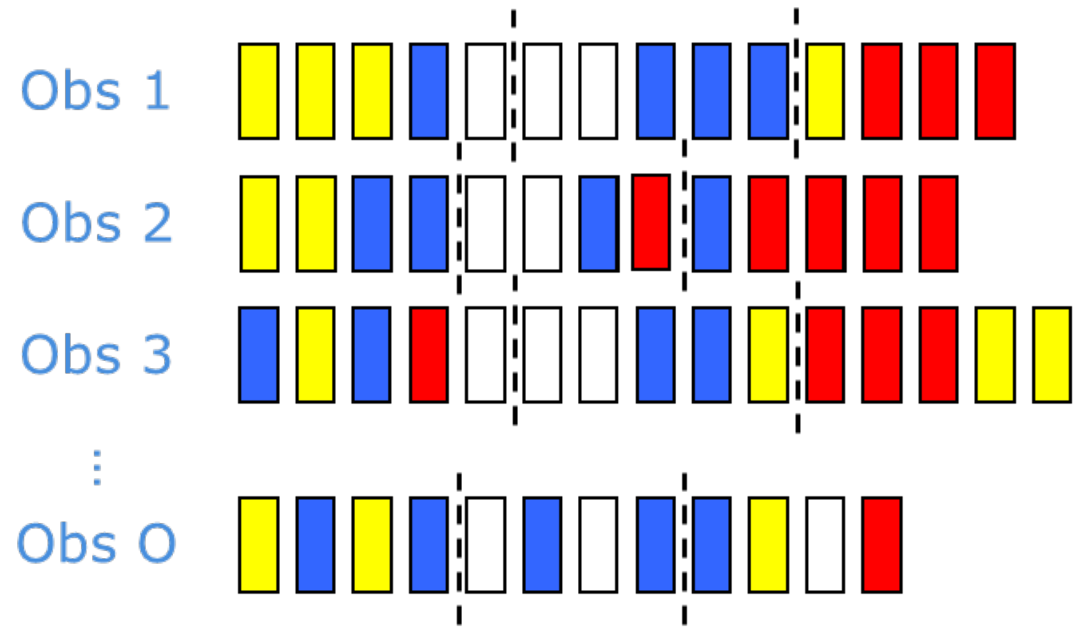


Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences Western Switzerland

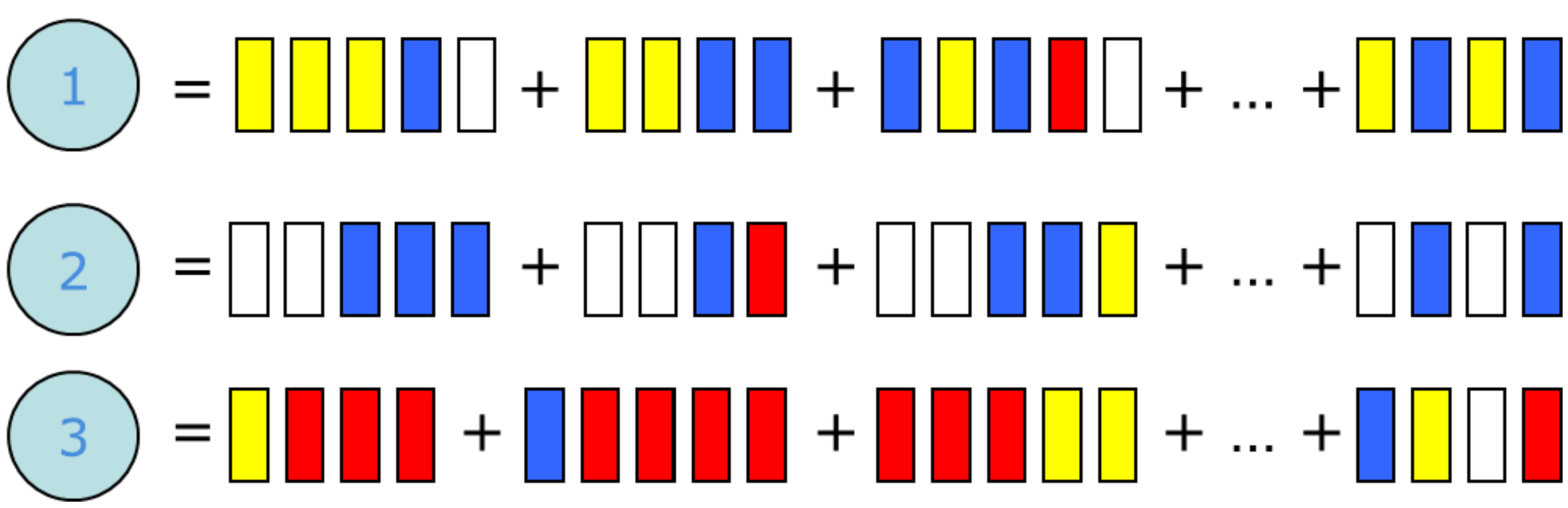
## 0.0 choix d'une topologie



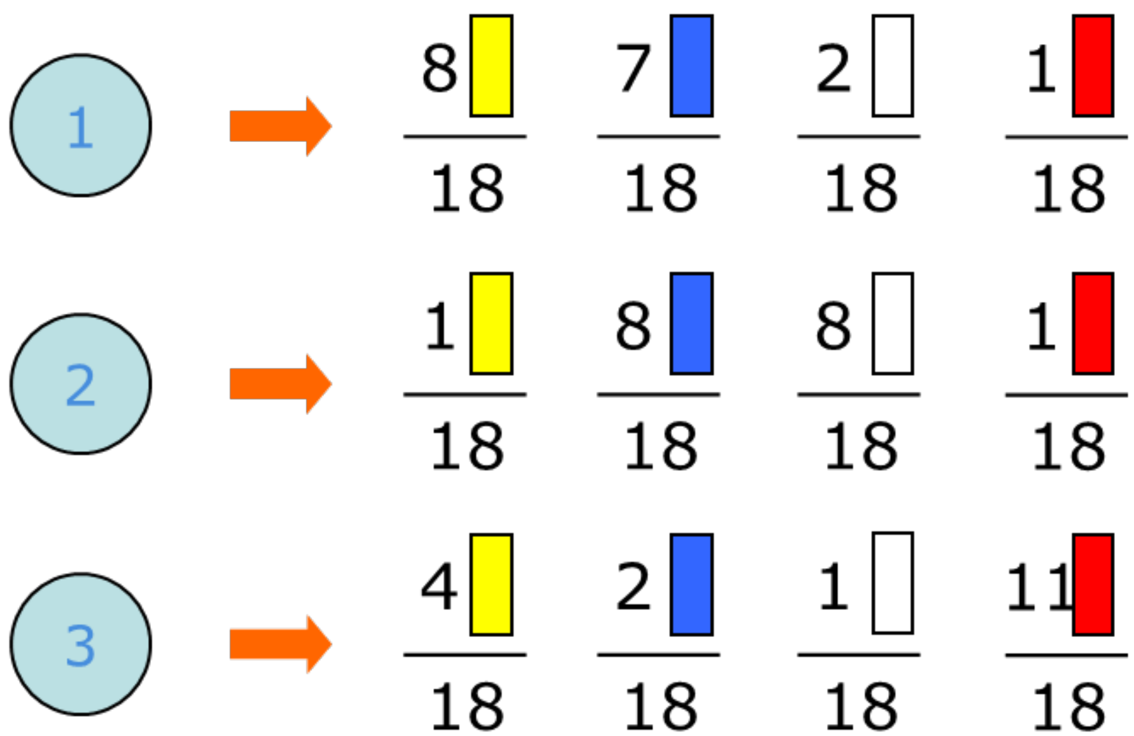

## 1* alignement linéaire




## 2. accumulation

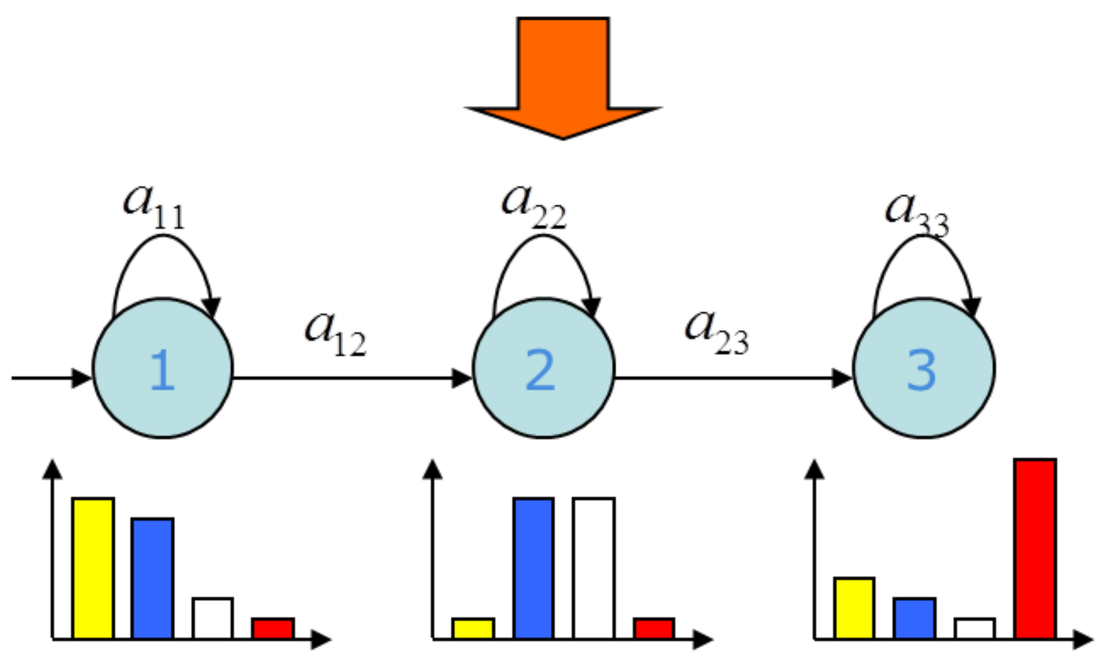


# 3. ré-estimation

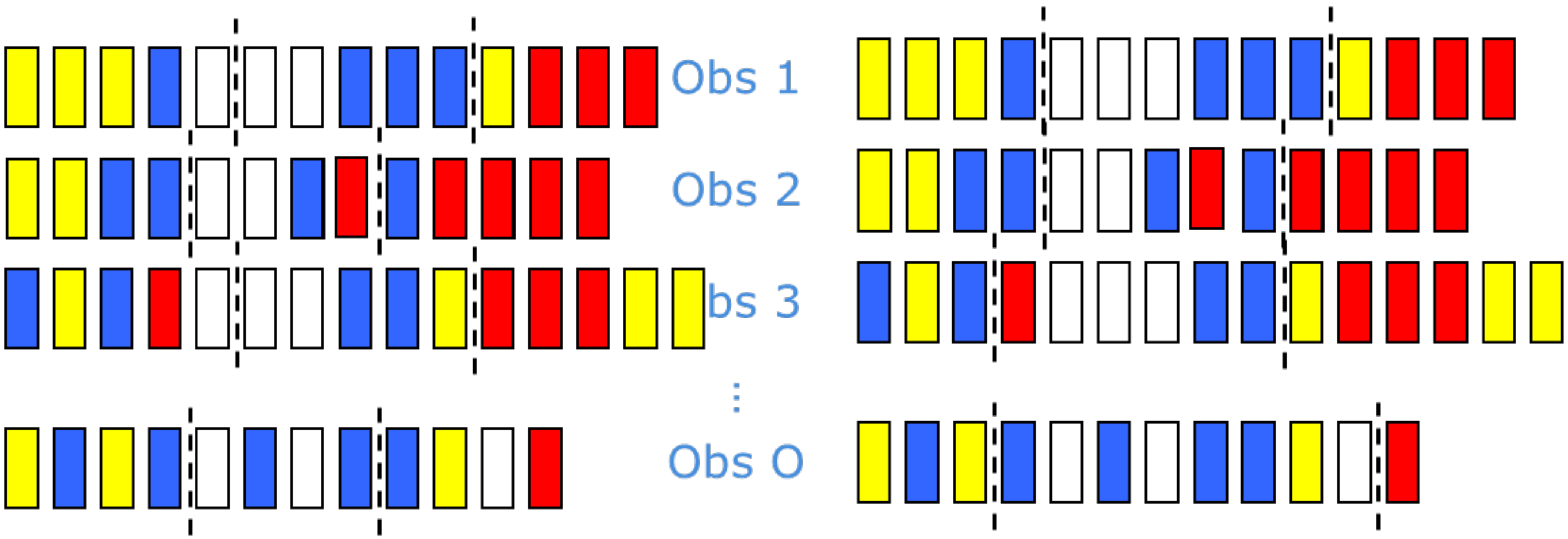



$$a_{11} = \frac{\# trans\, 1 \rightarrow 1}{\# trans\, 1}, \quad a_{12} = \frac{\# trans\, 1 \rightarrow 2}{\# trans\, 1}$$

$$a_{ij} = \frac{\# trans\, i \rightarrow j}{\# trans\, i}$$



## 1 alignement viterbi




## 2. accumulation

## 3. ré-estimation

## 1 alignement viterbi

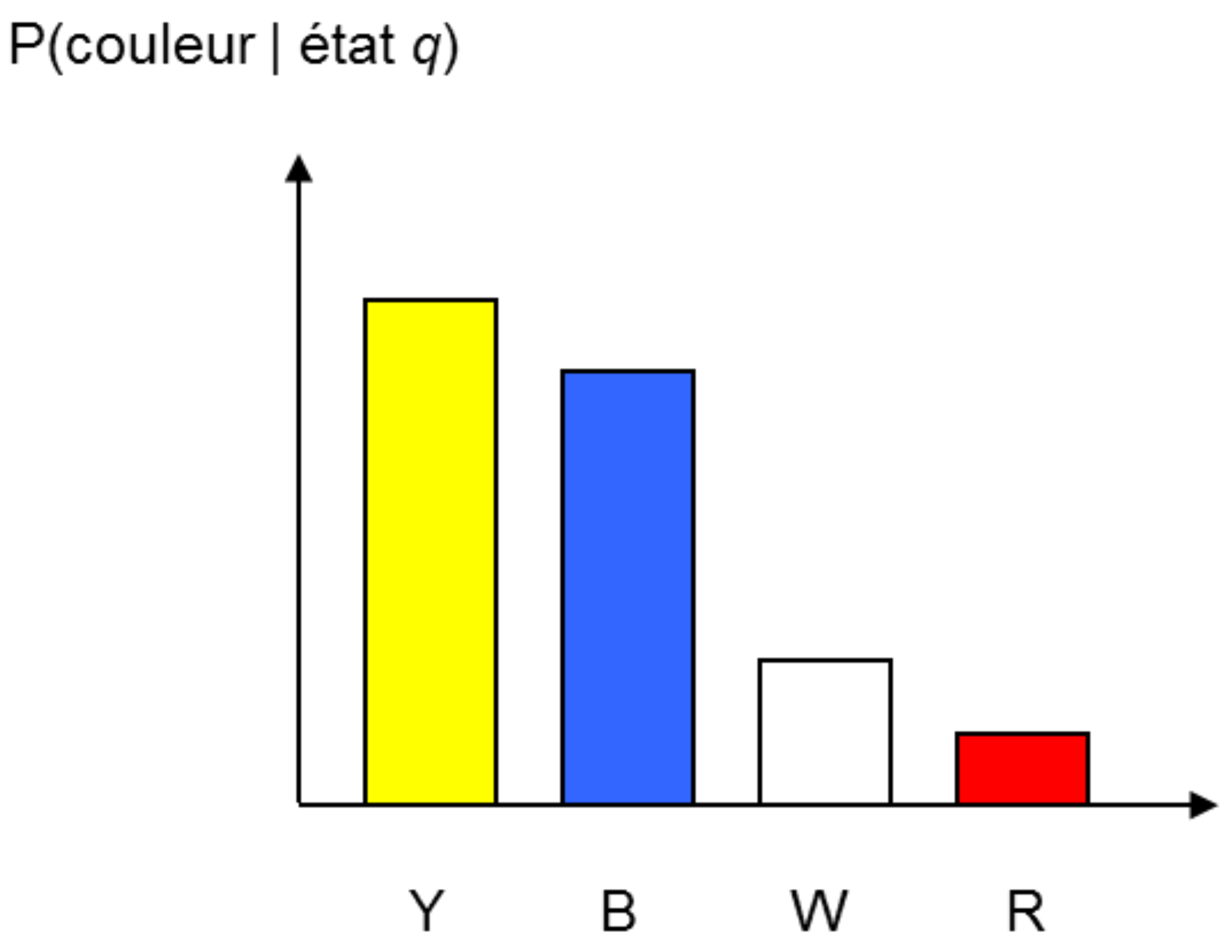
On itère jusque la convergence de la valeur accumulée

$$\sum_{i=1}^{O} P(X_i, Q_{\max} \mid M)$$

Probabilité d’émission de type mixture de gaussiennes

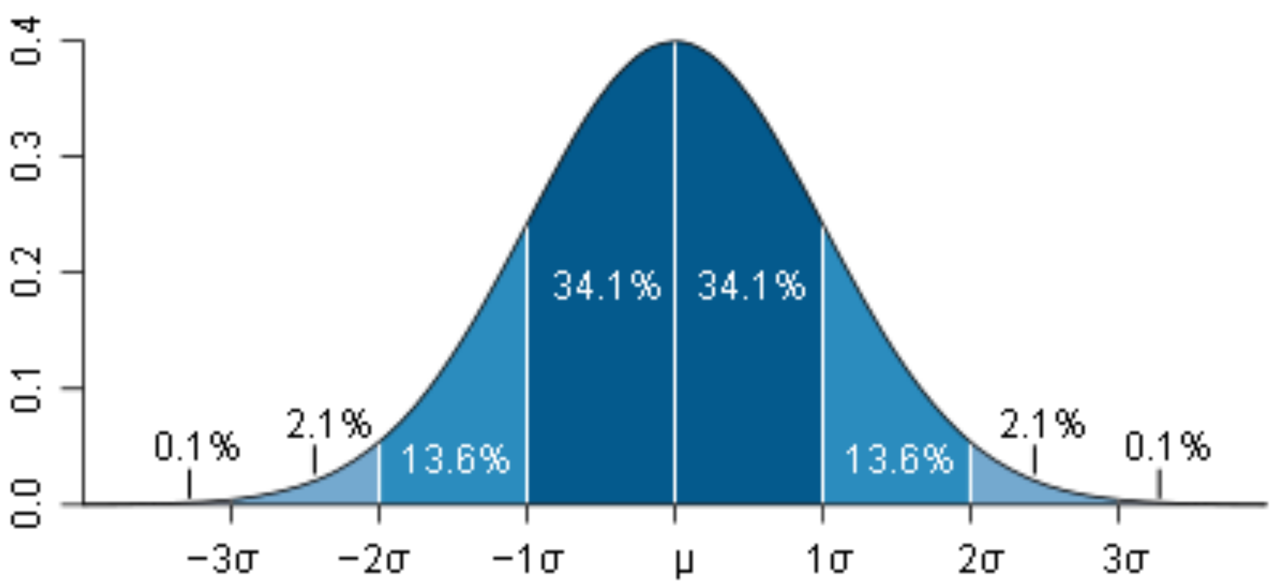
# TYPES DE PROBABILITÉS D’ÉMISSION

# Probabilités d'émission discrète




- Les observations sont **discrètes**
- Les probabilités d'émissions sont de simples **histogrammes**
- Une observation peut devenir discrète suite à une opération de **quantification**
  - La quantification est dite **vectorielle** lorsque l'observation est un vecteur

# Probabilités d'émission continue




$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

- Les observations sont des valeurs **continues** (scalaires ou vecteurs)
- On fait souvent l'hypothèse d'une distribution **normale** ou **Gaussienne**
- Cette hypothèse peut ne pas être juste !!!
- L'entraînement revient à calculer la moyenne $\mu$ et l'écart-type $\sigma$.

# Densités de probabilités d'émission continue : GMM

- On utilise souvent comme forme de distribution une somme pondérée de gaussiennes (« Gaussian mixture model»)

$$P(x_n|q_k, M_i) = \sum_{j=1}^{J} w_{kj} \mathcal{N}(x_n, \mu_{kj}, \Sigma_{kj})$$

- Avec:

$$\mathcal{N}(x_N, \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}(x_n - \mu)^T \Sigma_k^{-1}(x_n - \mu)\right)$$

# Densités de probabilités d'émission continue : GMM

- Les coefficients des vecteurs étant décorrélés, la matrice de covariance peut être considérée diagonale:

$$\mathcal{N}(x_n, \mu, \Sigma) = \prod_{i=1}^{D} \frac{1}{\sqrt{2\pi\sigma_i^2}} exp\left(-\frac{(x_n i - \mu_i)^2}{2}\right)$$

# Densités de probabilités d'émission continue : GMM

- Dans le cas des *Gaussian mixture models*, il existe des algorithmes itératifs qui permettent de calculer les valeurs des différentes moyennes $\mu_j$, écart-types $\sigma_j$ et poids $\omega_j$ de chaque gaussienne
  - Algorithme Expectation – Maximisation (EM)
  - Algorithme MAP
    - Adaptation des paramètres ($\mu_j$ , $\sigma_j$ , $\omega_j$ ) à partir de valeurs initiales

Hes·so
Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences Western Switzerland

MSE | MASTER OF SCIENCE IN ENGINEERING

# Multimodal Processing, Recognition and Interaction

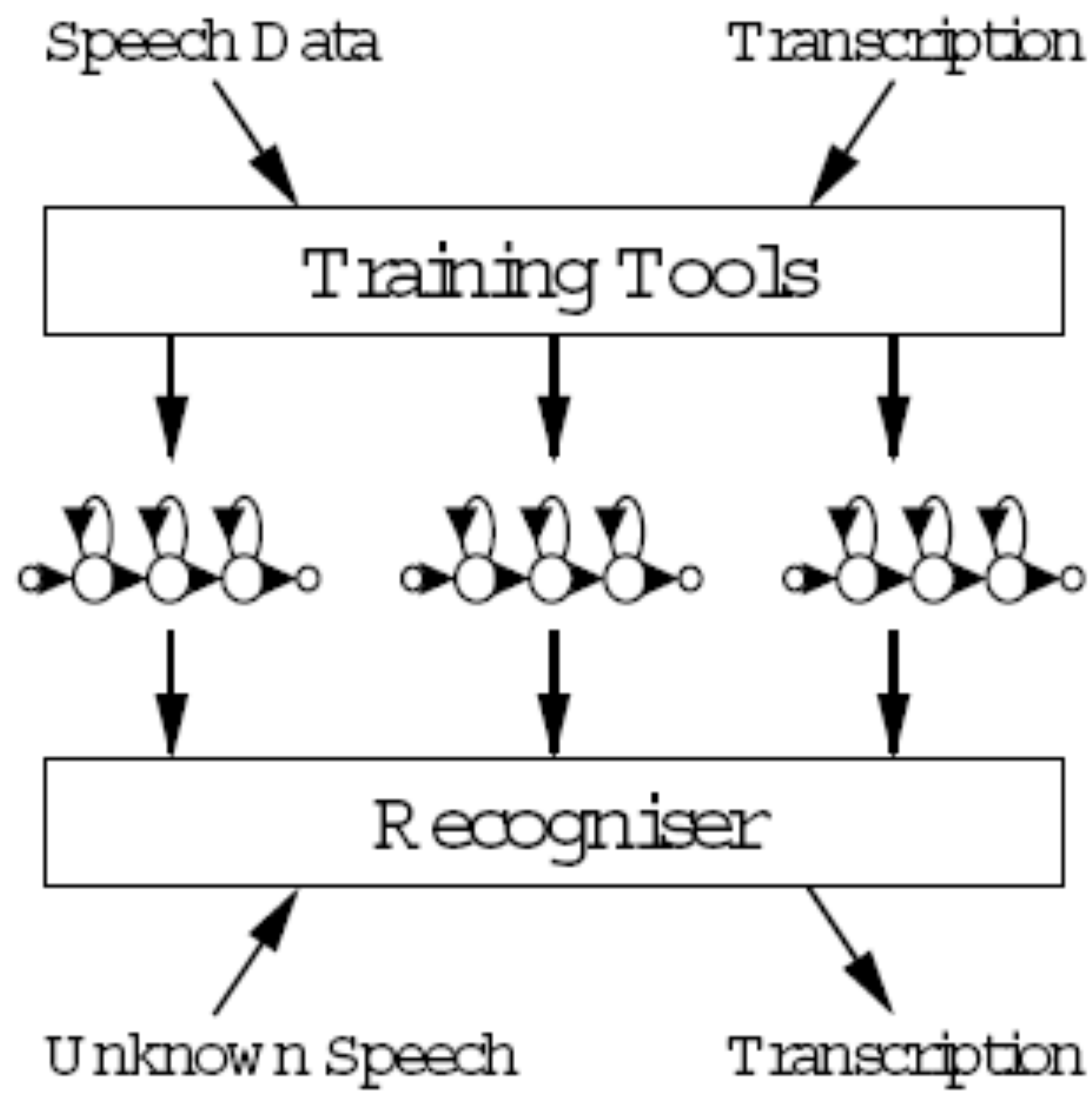
Practical Information
Introduction

Elena Mugellini, Jean Hennebert, Stefano Carrino

# APPLICATIONS

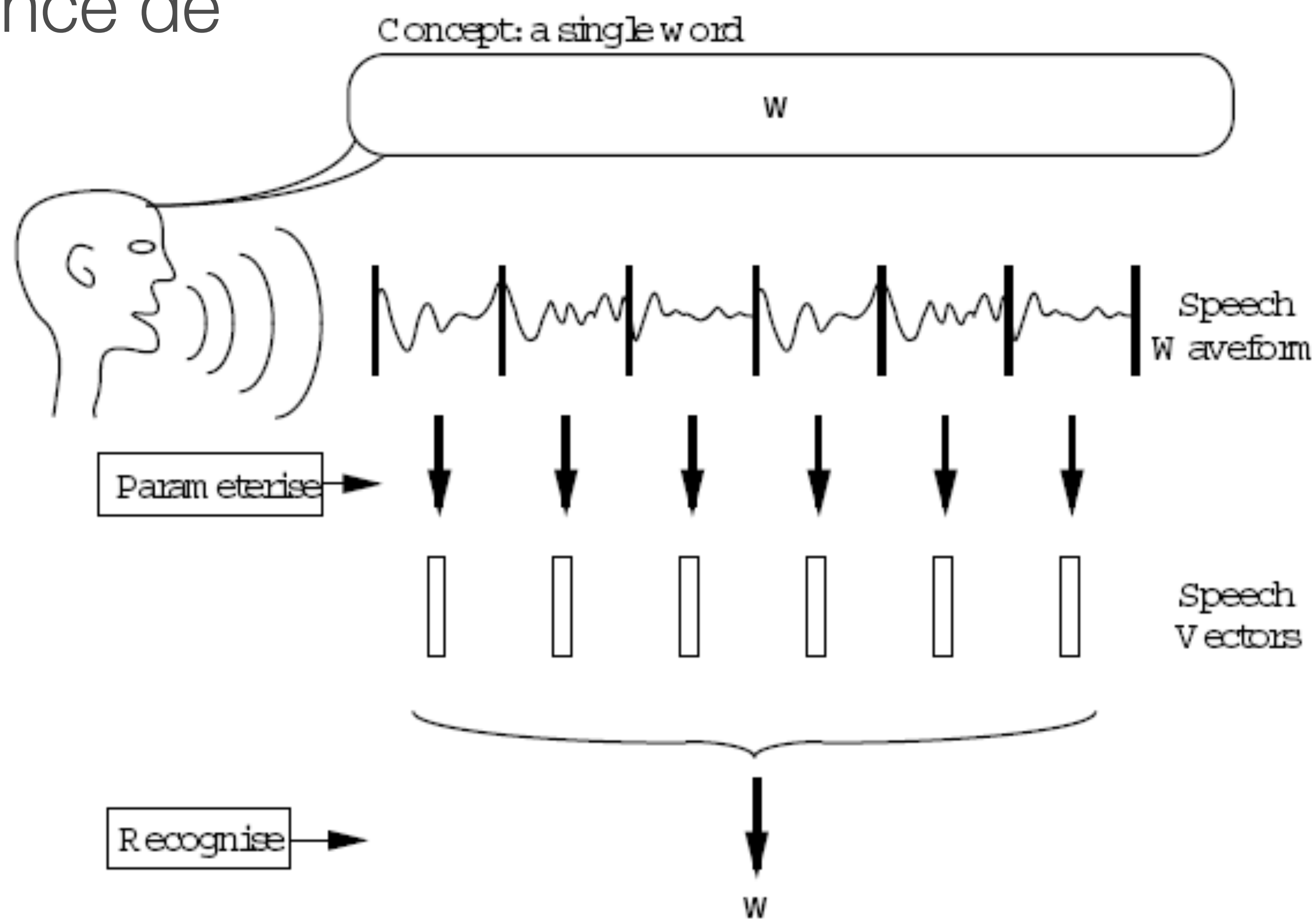
# Reconnaissance de la parole

- Vue d'ensemble



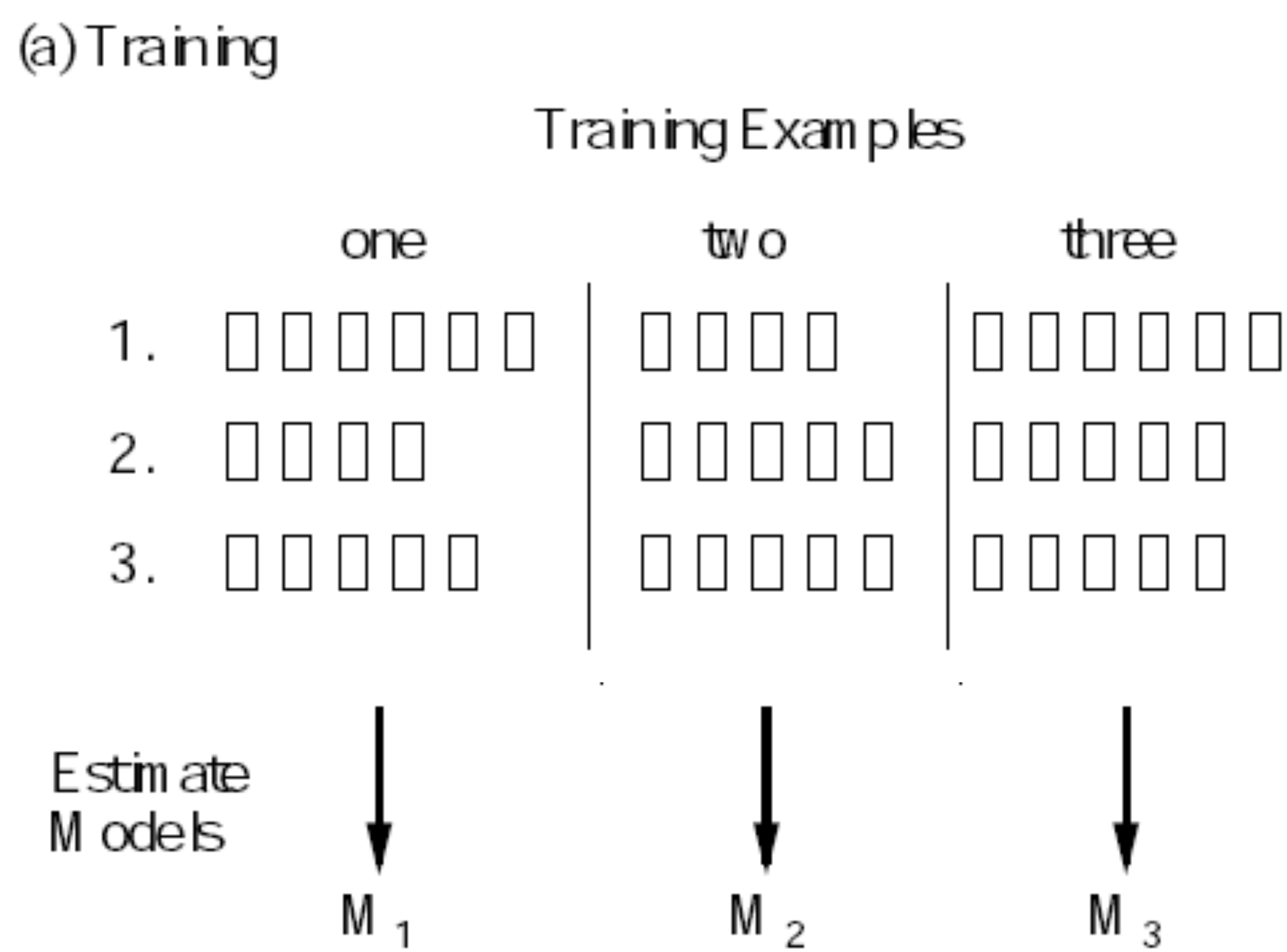
# Reconnaissance de la parole

- Reconnaissance de mots isolés



# Reconnaissance de la parole

- Training
  - Estimation des modèles



# Reconnaissance de la parole

- Reconnaissance