



SP3 – Segmentation et Authentification de la parole

Description de l'algorithme d'authentification vocale dépendante du texte

LIA – Anthony Larcher



- Présentation des objectifs
- Présentation des outils et méthodes
- Présentation du système à trois niveaux
- Principe de l'ULT et WRE, sélection des gaussiennes
- Utilisation de la Vidéo
- Avantages du système et estimation de la mémoire



Objectifs

Reconnaissance vocale dépendante du texte :

- identification du locuteur en milieu ouvert
- biométrie audio-vidéo
- système de reconnaissance audio incluant une information temporelle issue de la vidéo.
- système embarqué → minimum de ressources (mémoire et temps de calcul)



Présentation des outils

Formalisme utilisé : théorie bayésienne de la décision



Présentation des outils

Formalisme utilisé : théorie bayésienne de la décision

Comparaison d'hypothèses par un rapport bayésien

$$\begin{array}{ccc} & \textit{accept} & \\ \log(P(S|H) - \log(P(S|\bar{H})) & > & \textit{Seuil} \\ & < & \\ & \textit{reject} & \end{array}$$



Présentation des outils

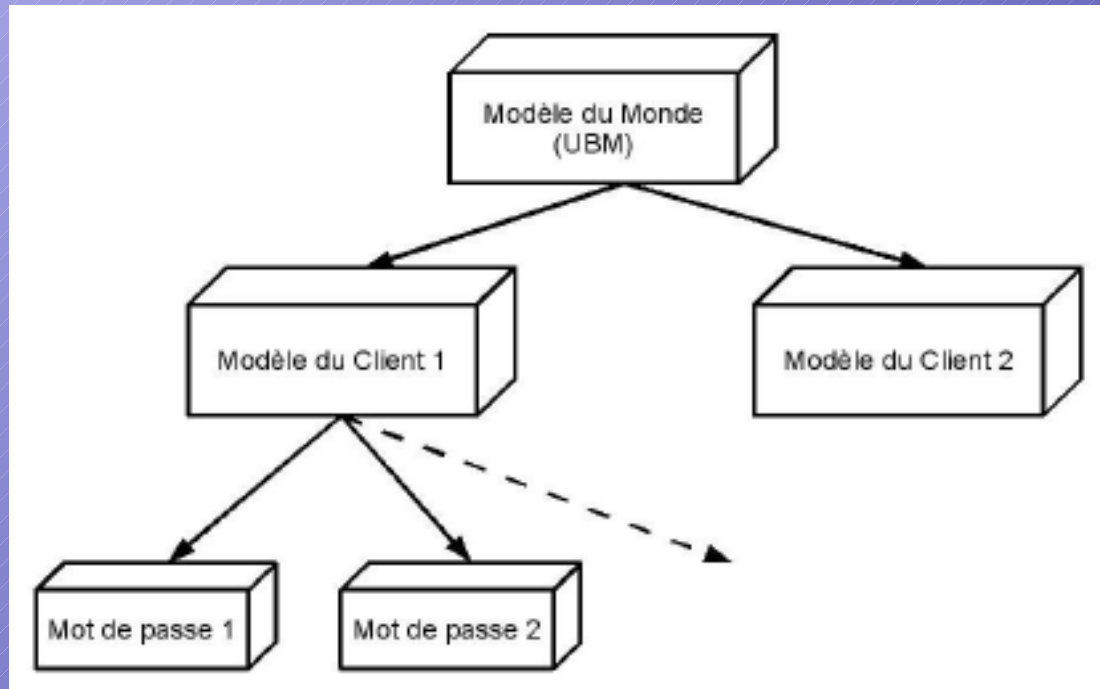
- Formalisme utilisé : théorie bayésienne de la décision
- Comparaison d'hypothèses par un rapport bayésien
- Modélisation des distributions par des GMM (Gaussian Mixture Model)



Présentation des outils

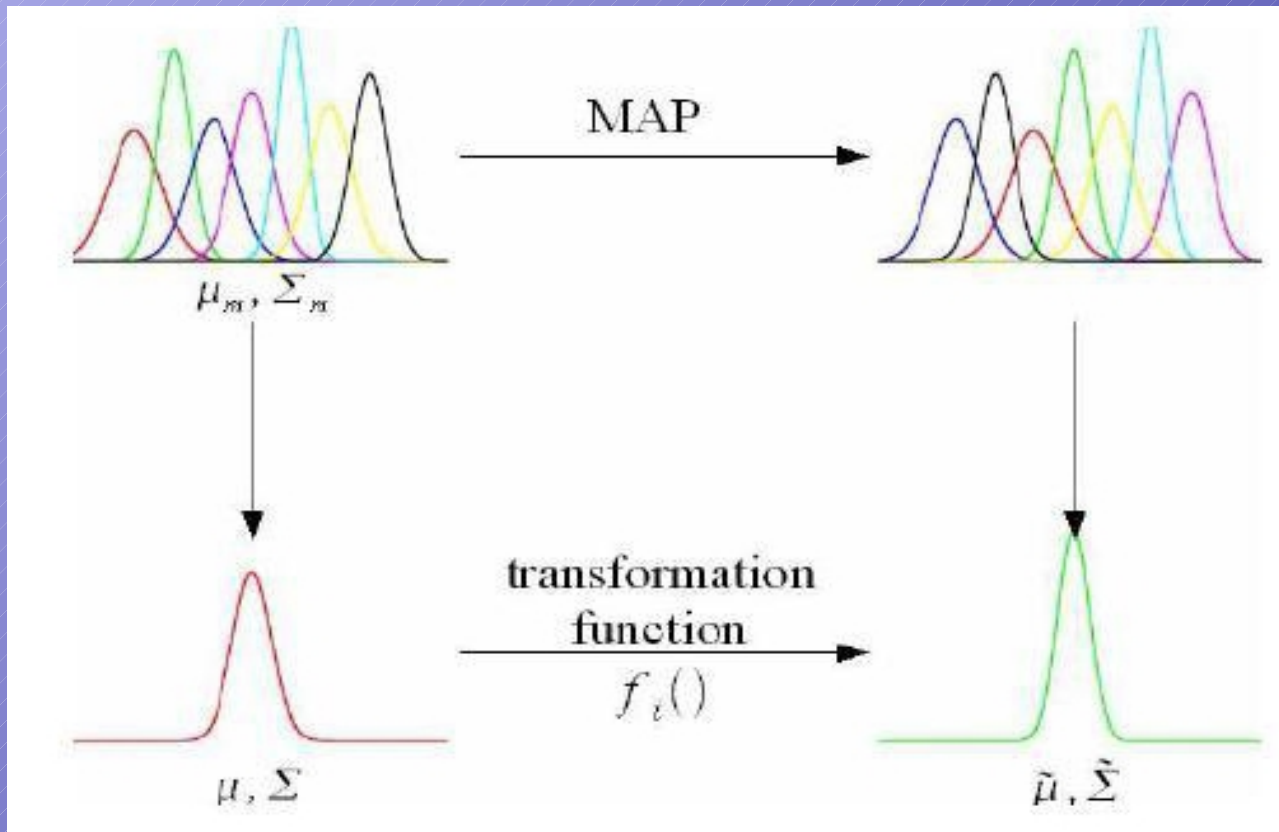
- Formalisme utilisé : théorie bayésienne de la décision
- Comparaison d'hypothèses par un rapport bayésien
- Modélisation des distributions par des GMM (Gaussian Mixture Model)
- Modélisation d'un mot de passe et de sa structure temporelle par un HMM (Hidden Markov Model)

Modèle à 3 Niveaux



LIAMAP

Obtention d'une transformation linéaire unique
(ULT)





Weight Re-Estimation

Méthode d'adaptation de modèle ne modifiant que les poids

Le poids des gaussiennes du modèle est ré-estimé avec un critère de maximum de vraisemblance

Avantage : la vraisemblance est calculée composante par composante et on fait ensuite une somme pondérée (donc un seul calcul de vraisemblance pour tous les états du HMM)



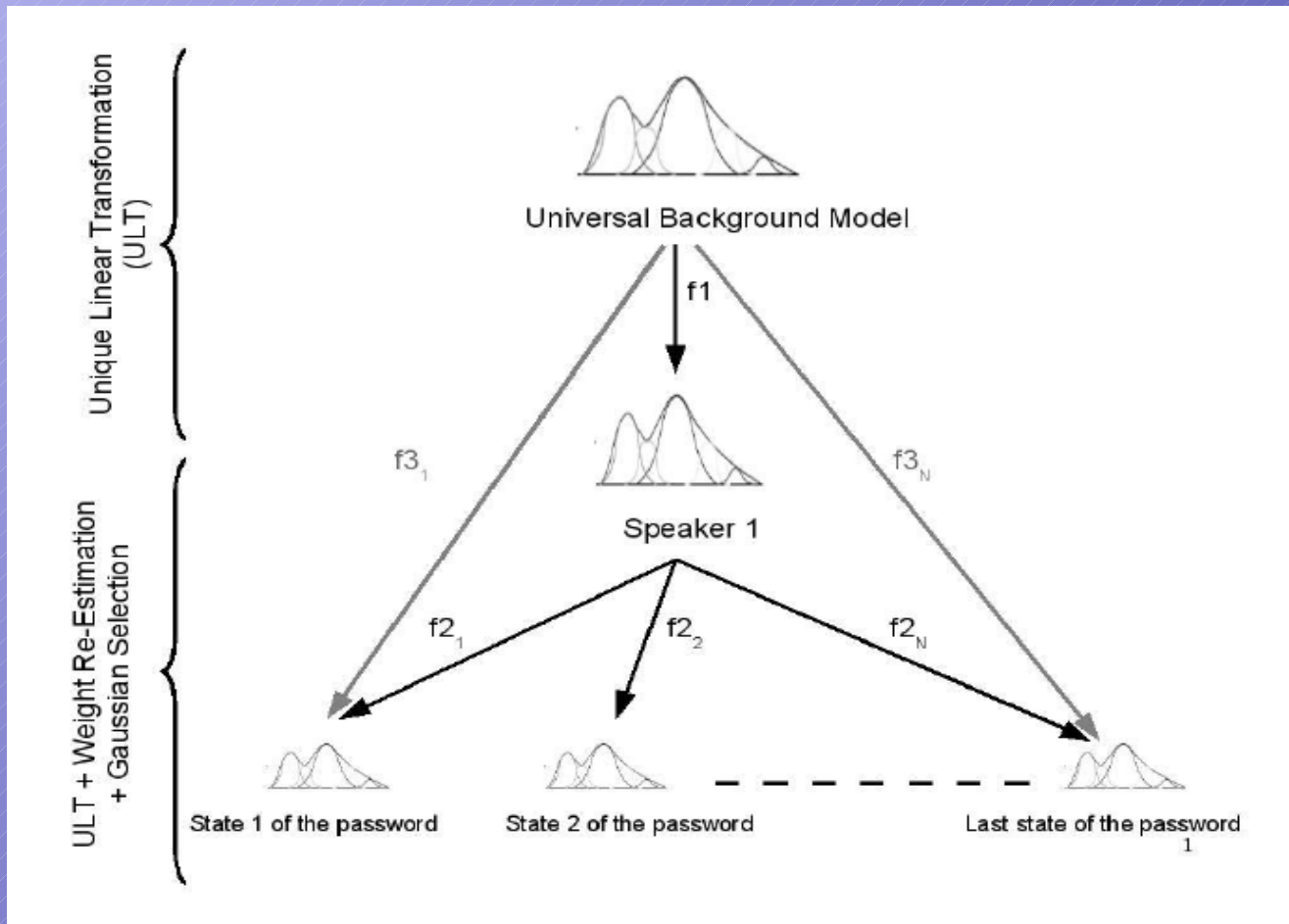
Sélection de Gaussiennes

Objectif : réduire la taille des GMM et donc le nombre de paramètres en mémoire)

En phase d'entraînement : calcul des poids (WRE)

On sélectionne les 20 gaussiennes qui ont les poids les plus importants (ce choix est donc dépendant des données d'entraînement)

Transformations





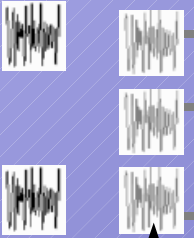
The L1 Model is built
The L1-Password Model is built

Universal Background Model

L1

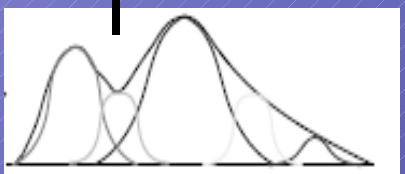
HMM model of the password

Training Material for a speaker



Repetitions of a target event (password)

Transitions between states



GMM of the images transitions

Memory

Complete GMM

Unique Linear Transformation

ULT + Weight Re-Estimation for each state

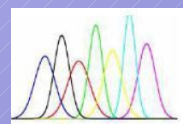
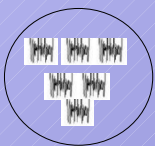
Transitions between states

Apprentissage non-embarqué

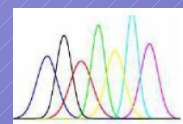
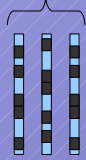
mémoire embarquée

Tests sur système embarqué

Enregistrements de différents locuteurs



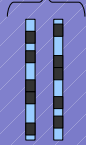
poids, moyennes et variances de l'UBM



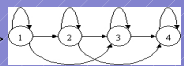
Enregistrements divers du locuteur L1



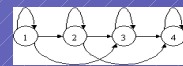
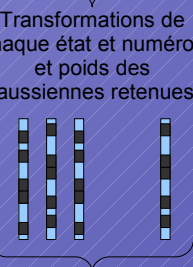
Coefficients α et β de la transformation linéaire



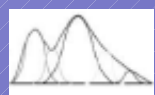
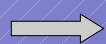
Répétitions du mot de passe 1



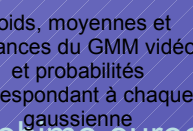
Transformations de chaque état et numéros et poids des gaussiennes retenues



Données vidéo



poids, moyennes et variances du GMM vidéo et probabilités correspondant à chaque gaussienne



TESTS LOCUTEUR ET DECODAGE VITERBI



Estimation de la Mémoire Audio

Modèle UBM :

- 512 gaussiennes
- Dimension des gaussiennes : 13

Modèle du locuteur :

- transformation ULT : coefficients α et β (3 X 13 coefficients)

Modèle de mot de passe :

- HMM à 20 états (57 transitions)
- Chaque état : 1 GMM à N gaussiennes
- Transformations ULT : 2 X 13 coefficients
- Numéros des N gaussiennes retenues et poids de ces gaussiennes



Estimation de la Mémoire

Modèle: tous les coefficients sont codés sur 4 octets

- Audio : 85 256 octets



Discussion