

Avancement Projet BIOBIMO - LIA 24/01/08 - 30/06/08



Ajout d'une information au système Audio

- description du système
- segmentation phonétique
- synchronisation vidéo

(Publications)

Description du Système EBD Audio-Vidéo



OBJECTIF:

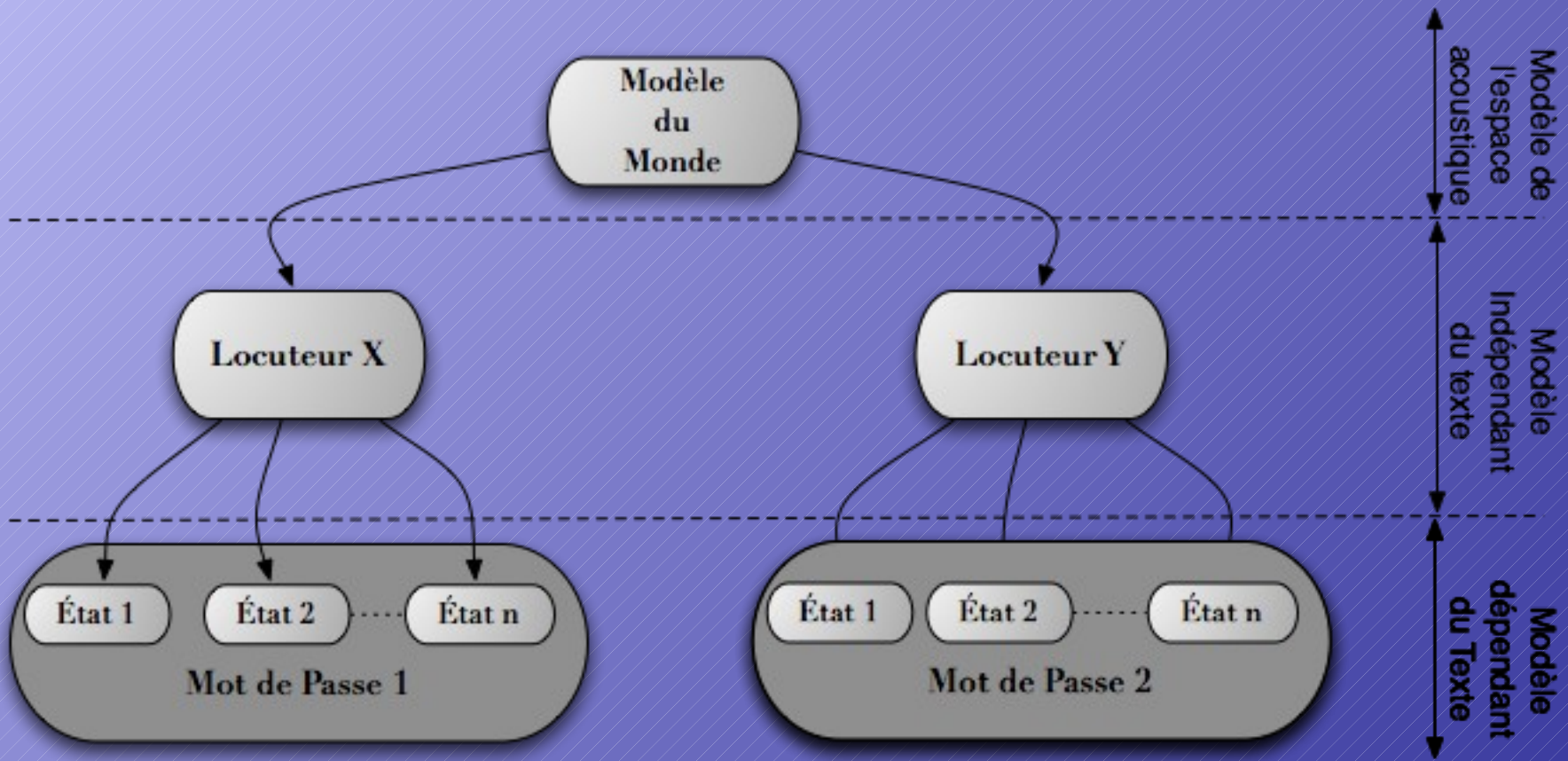
utiliser une information temporelle pour renforcer la robustesse du système de reconnaissance du locuteur.

MISE EN OEUVRE:

le système audio possède déjà une structure qui lui permet de capter une information temporelle

Renforcer cette structure audio en utilisant des points de synchronisation

Description du Système EBD Audio-Vidéo

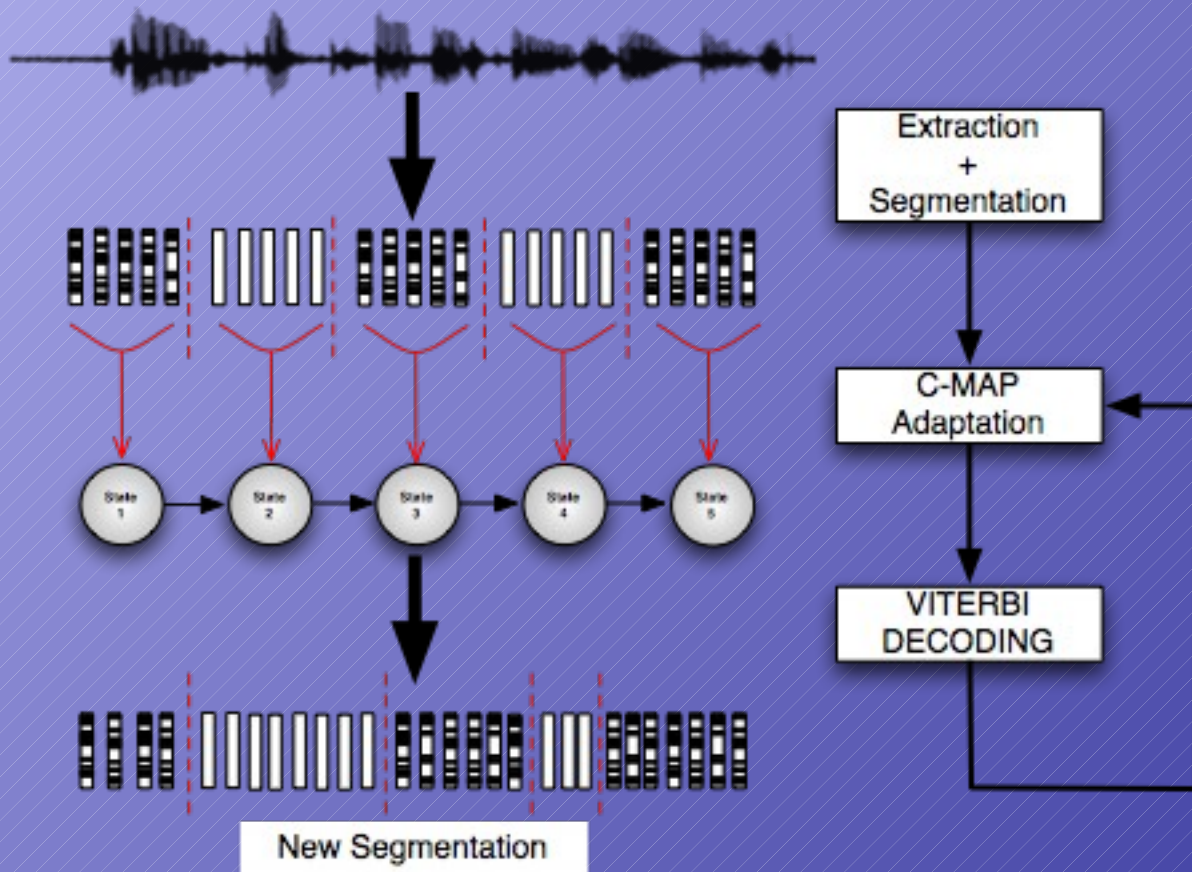


Description du Système EBD Audio-Vidéo

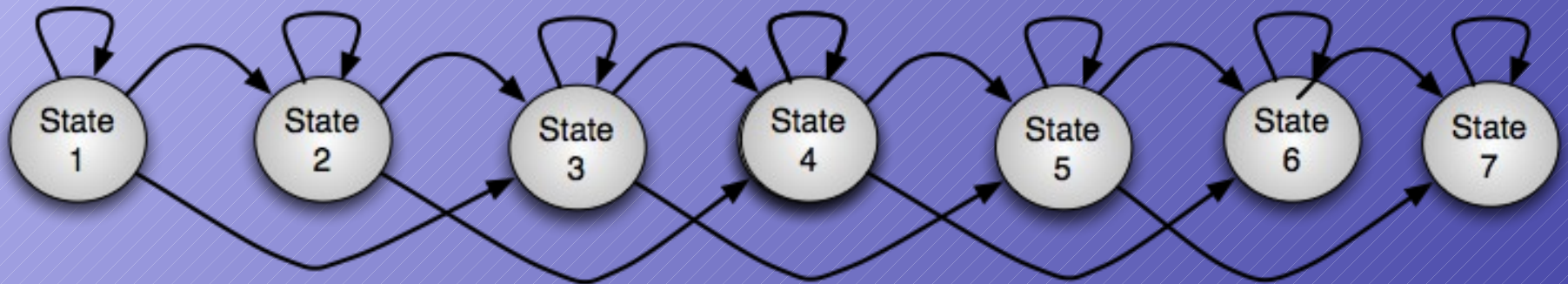


Modèle
dépendant
du Texte

Description du Système EBD Audio-Vidéo



Description du Système EBD Audio-Vidéo

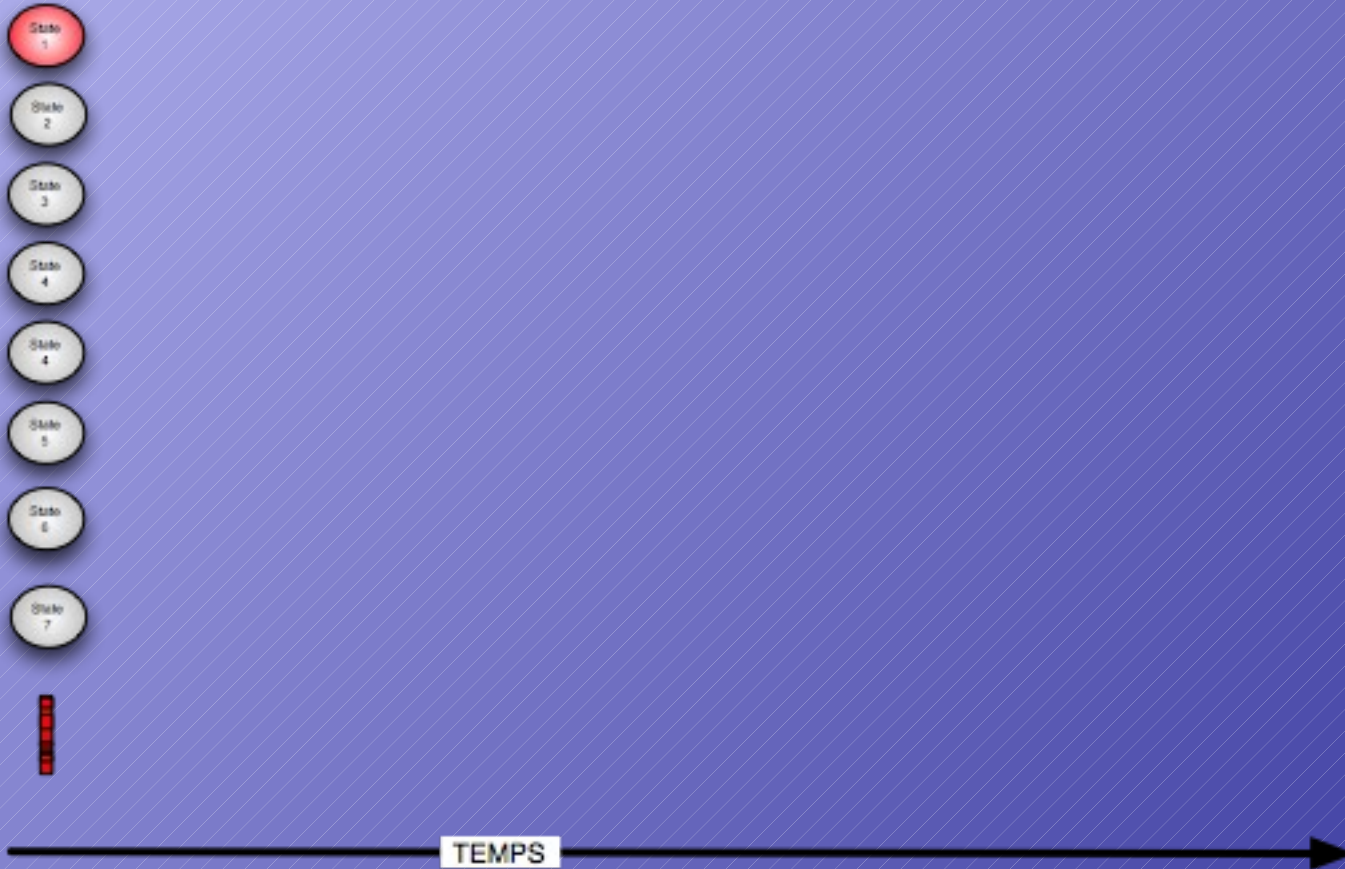


Annie

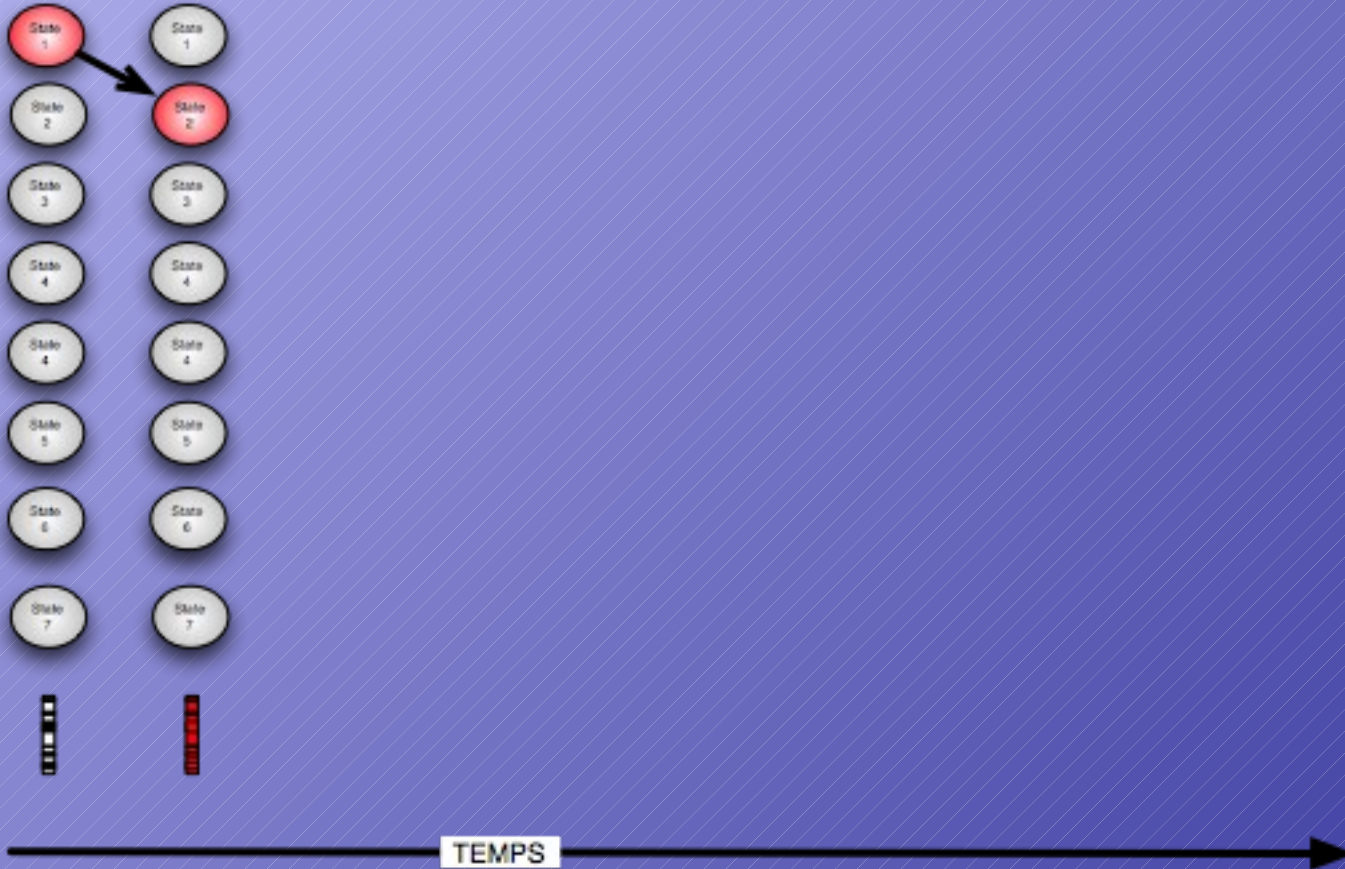
s'ennuie

loin...

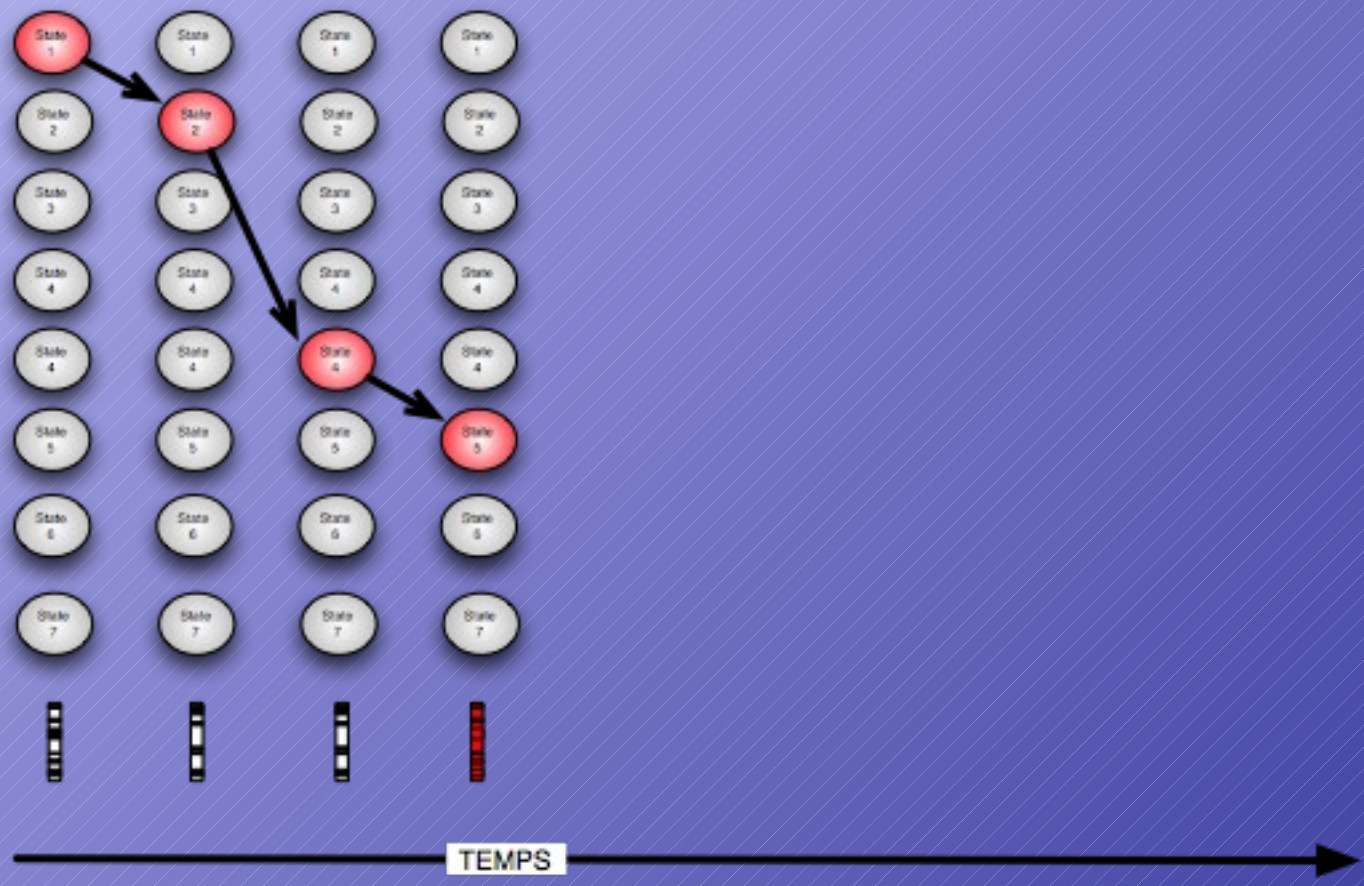
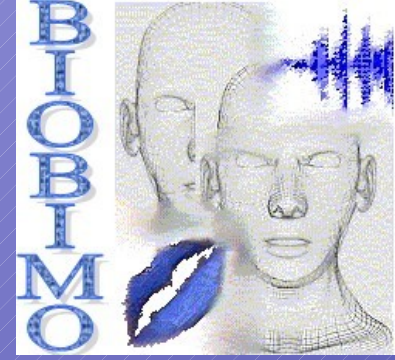
Description du Système EBD Audio-Vidéo



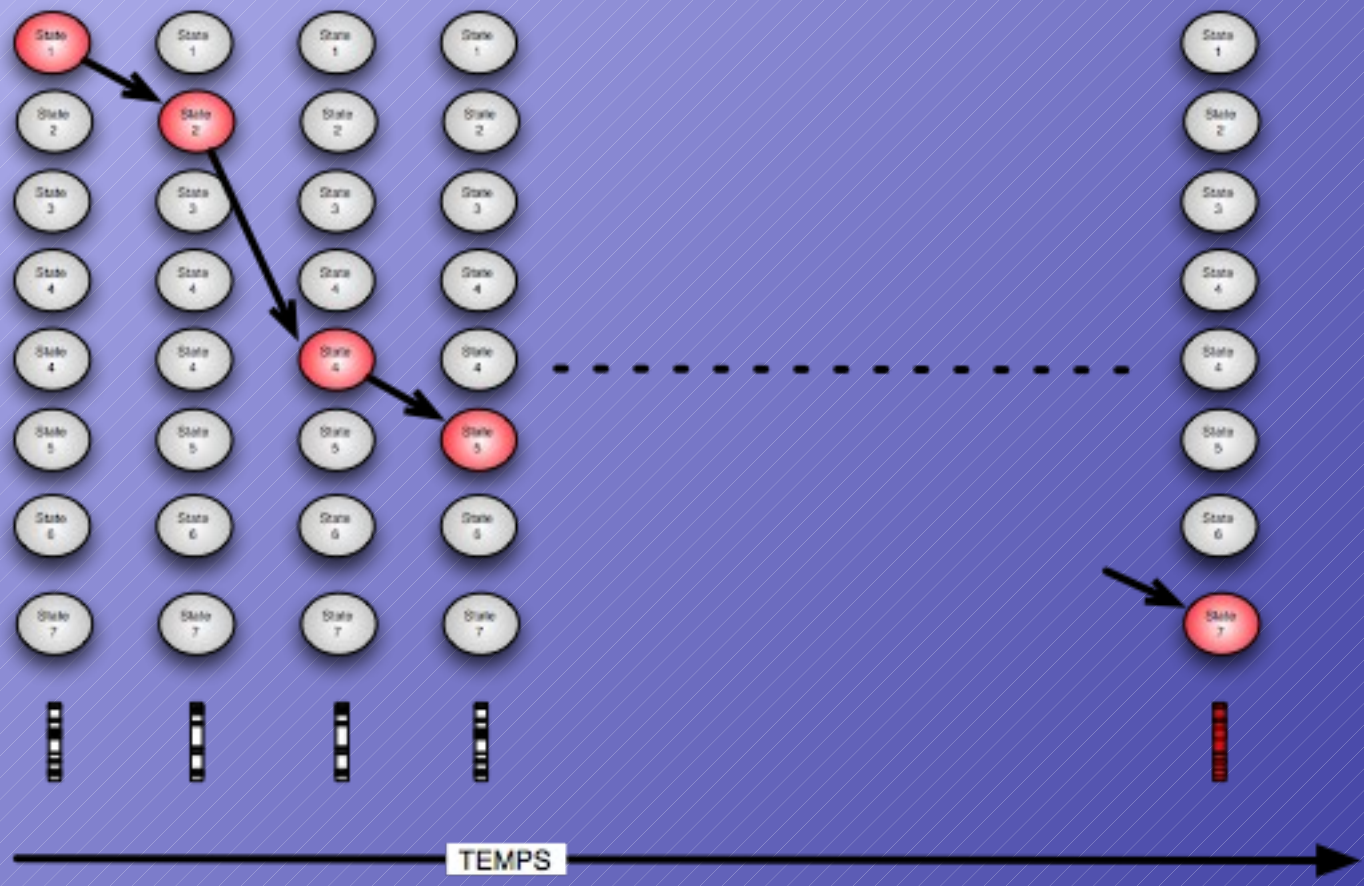
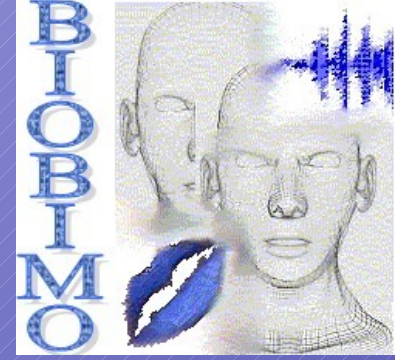
Description du Système EBD Audio-Vidéo



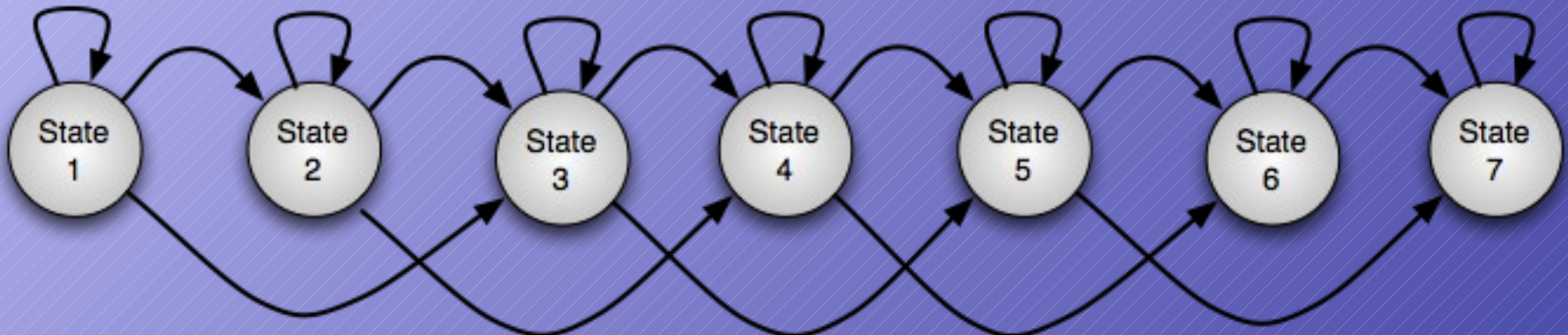
Description du Système EBD Audio-Vidéo



Description du Système EBD Audio-Vidéo



Description du Système EBD Audio-Vidéo

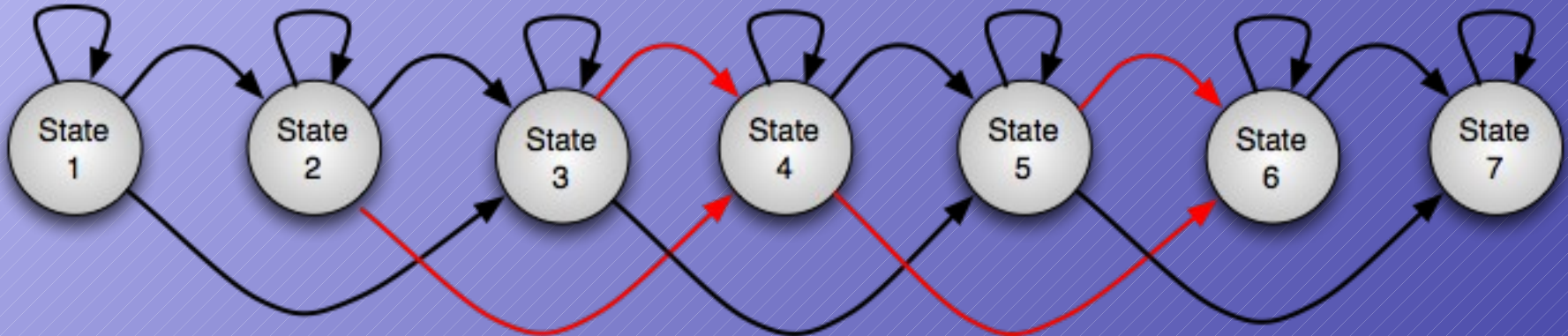


Annie

s'ennuie

loin...

Description du Système EBD Audio-Vidéo



Annie

s'ennuie

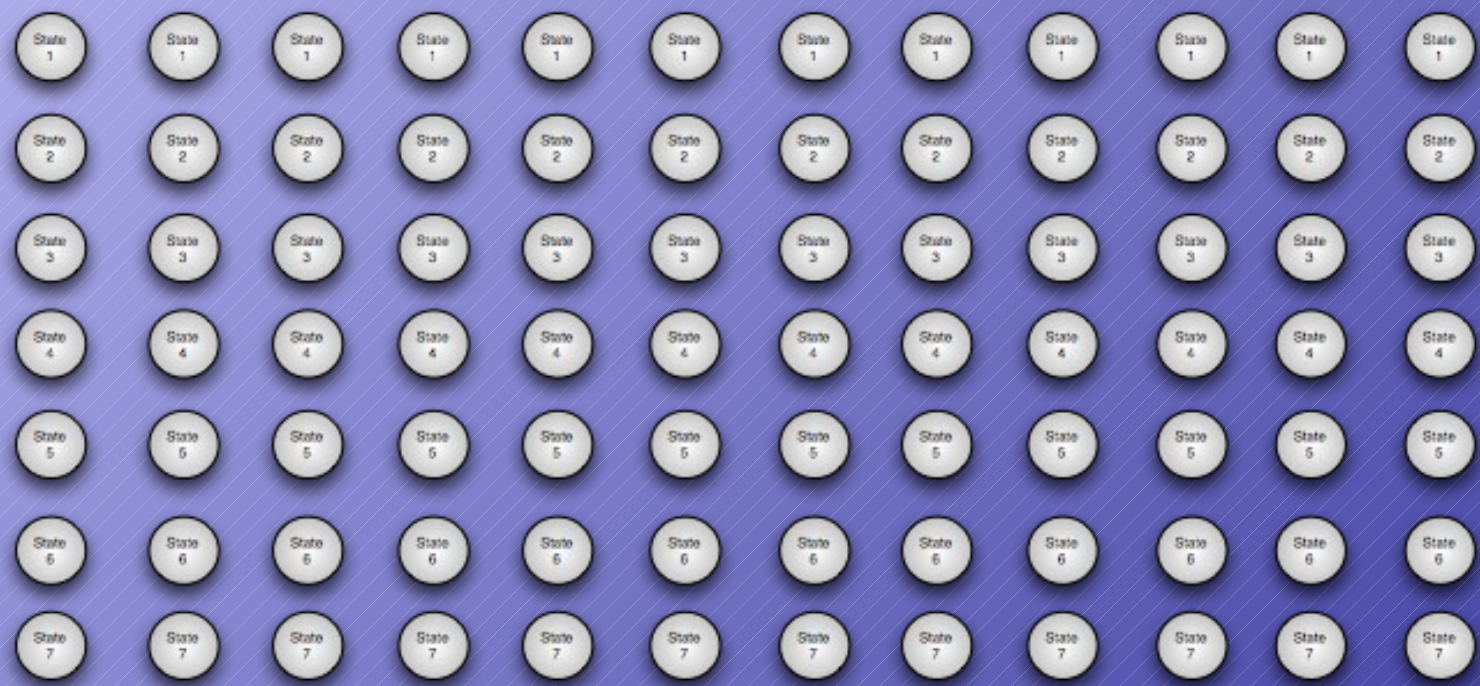
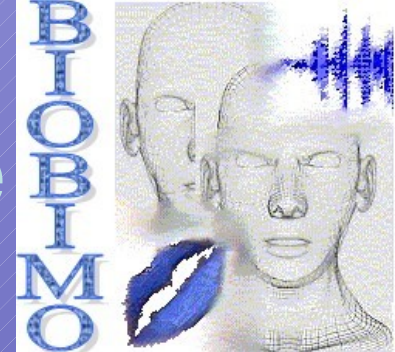
loin...

SEGMENT 1

SEGMENT 2

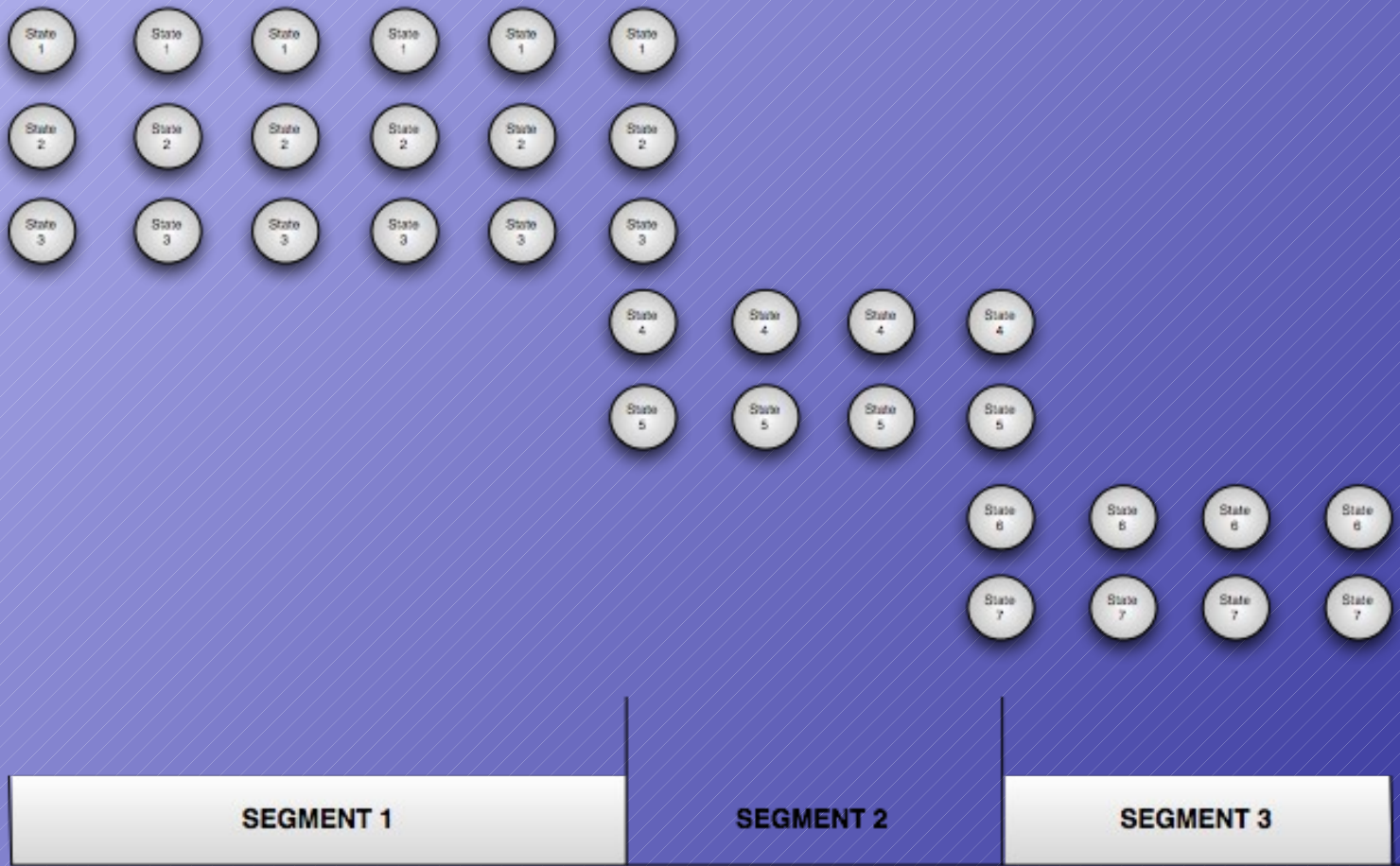
SEGMENT 3

Description du Système EBD Audio-Vidéo

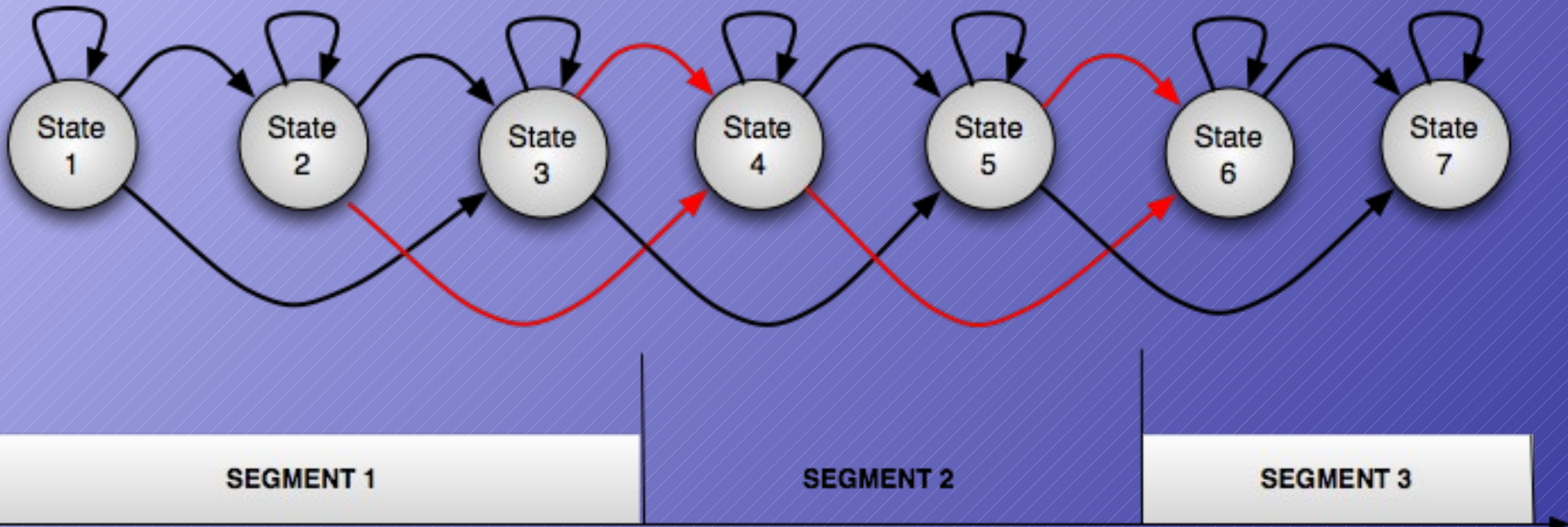




Description du Système EBD Audio-Vidéo



Obtention des points de synchronisation



Obtention des points de synchronisation: Segmentation Phonétique



MÉTHODE:

- GValign (développé au LIA).
- 5 segments par phrase.
- mêmes segments pour tous les locuteurs.

EXEMPLES:

Il se | garantira | du froid | avec un bon | capuchon.

Ce | petit | canard | apprend | à nager.

Les deux | camions | se sont | heurtés | de face.

Obtention des points de synchronisation: Segmentation Phonétique



Résultats publiés à Interspeech en septembre 2008

"Reinforced Temporal Structure Information For Embedded Utterance-Based Speaker Recognition"

Configuration		GMM/UBM Baseline	External Information Condition	
			NONE	ACTIVE
UNKNOWN	1-occ	2.72	2.07	1.89
UNKNOWN	2-occ	0.87	0.56	0.47
KNOWN	1-occ	4.49	4.61	4.39
KNOWN	2-occ	2.22	2.89	2.78
ALL	1-occ	3.72	3.66	3.33
ALL	2-occ	1.58	2.09	1.77

Obtention des points de synchronisation: Segmentation Vidéo



MÉTHODE:

- passage de la vidéo en niveau de gris
- différence d'images 2 à 2
- somme des pixels des "images différences"

on obtient un signal temporel mono-dimension

- récupération des maxima locaux grâce à un algorithme de fenêtre glissante.

Chaque maxima est utilisé comme point de
synchronisation

Obtention des points de synchronisation: Segmentation Vidéo



Résultats publiés à MMSP en octobre 2008
"Short Utterance-based Video Aided Speaker Recognition"

Configuration		GMM/UBM Baseline	Video learnt Information	
			NONE	ACTIVE
UNKNOWN	1-occ	2.72	2.00	1.78
UNKNOWN	2-occ	0.87	0.55	0.43
KNOWN	1-occ	4.49	4.72	4.44
KNOWN	2-occ	2.22	2.87	2.33
ALL	1-occ	3.72	3.55	3.16
ALL	2-occ	1.58	1.97	1.57

Autres articles publiés



Publié aux JEP 2008:

“Utilisation de la structure de mots de passe personnalisés pour la reconnaissance de locuteurs embarquée”

Publié à EUSIPCO en août 2008:

“From GMM to HMM for Embedded Password-Based Speaker Recognition”

Travaux en cours



- ✓ robustesse aux impostures (image fixe, vidéo ne correspondant pas à l'audio...)
- ✓ test du renforcement de l'influence de la vidéo
- ✓ amélioration des modèles audio
- ✓ affinement de la modélisation temporelle

Avancement des Sous-Projets



✓ SP3 :

- Accent mis sur le développement de l'algorithme de reconnaissance du locuteur dépendant du texte
 - Exigence atteinte en termes de ressources CPU et mémoire
 - Exigence atteinte en termes de fusion de modalités
 - Exigence atteinte en termes d'adaptation en ligne

Avancement des Sous-Projets



✓ SP3 (suite):

– Module de robustesse (Prévue à T0+30)

- Développement de techniques utilisées en RAL (feature warping/factor Analysis) pour le mismatch entre sessions d'apprentissage et de test
- Tests réalisés sur des données téléphoniques campagne d'évaluation NIST SRE08 (impossibilité de tests sur Myldea)

– Gestion en flux en suspend au profit du SP4 (module de qualité) (Prévue à T0+30)

Avancement des Sous-Projets



- ✓ SP4 :
 - Audio synchronisé par la vidéo (délivré à T0+24)
 - Tests d'Impostures (prévu à T0+36)