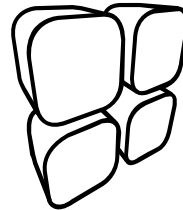


**Travail de Séminaire**  
**DIVA Research Group**  
**University of Fribourg**

**Etude de Projets Actuels**  
**sur l'enregistrement et**  
**l'analyse de Réunions**



# **Meeting Room : An Interactive Systems Laboratories Project**

**Canergie Mellon University – Karlsruhe University**

**Skultety Christophe**

**Article : Multimodal  
Meeting Tracker, 2000**

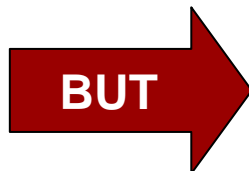
**Janvier 2005**

# Tables des Matières

- Description du Projet
- Vue générale du système
- Reconnaissance de la parole
- Identification des personnes
- Browser de réunion
- Travail futur
- Conclusion

# Introduction : Description du Projet

- ❑ Interrogations à propos d'une réunion
  - Qui a dit quoi ? et Quand ?
  - Était-ce une question ? une réponse ? une remarque ?
  - Quelle réaction ? colère, approbation ?
  - Niveau d'attention des participants ?
  
- ❑ Provenance des Informations
  - La Vison (gestes, expression du visage, langage corporel)
  - Le Son (paroles, intonation)



**Création d'un système d'enregistrement de réunion multimodale et automatisé**

# Caractéristiques du Projet

- ❑ Système d'enregistrement de réunions
  - Non Intrusif
  - Résolution du problème d'assignement
  - Système automatisé
  - Système multimodale
  - Fonctionne pratiquement en temps réel
  
- ❑ Doit pouvoir permettre
  - La visualisation de la Vidéo
  - L'écoute de l'Audio
  - La visualisation des Transcriptions

# Vue Générale du Système

Une salle de conférence à été spécialement équipée pour le projet

## ❑ Capacité du Système

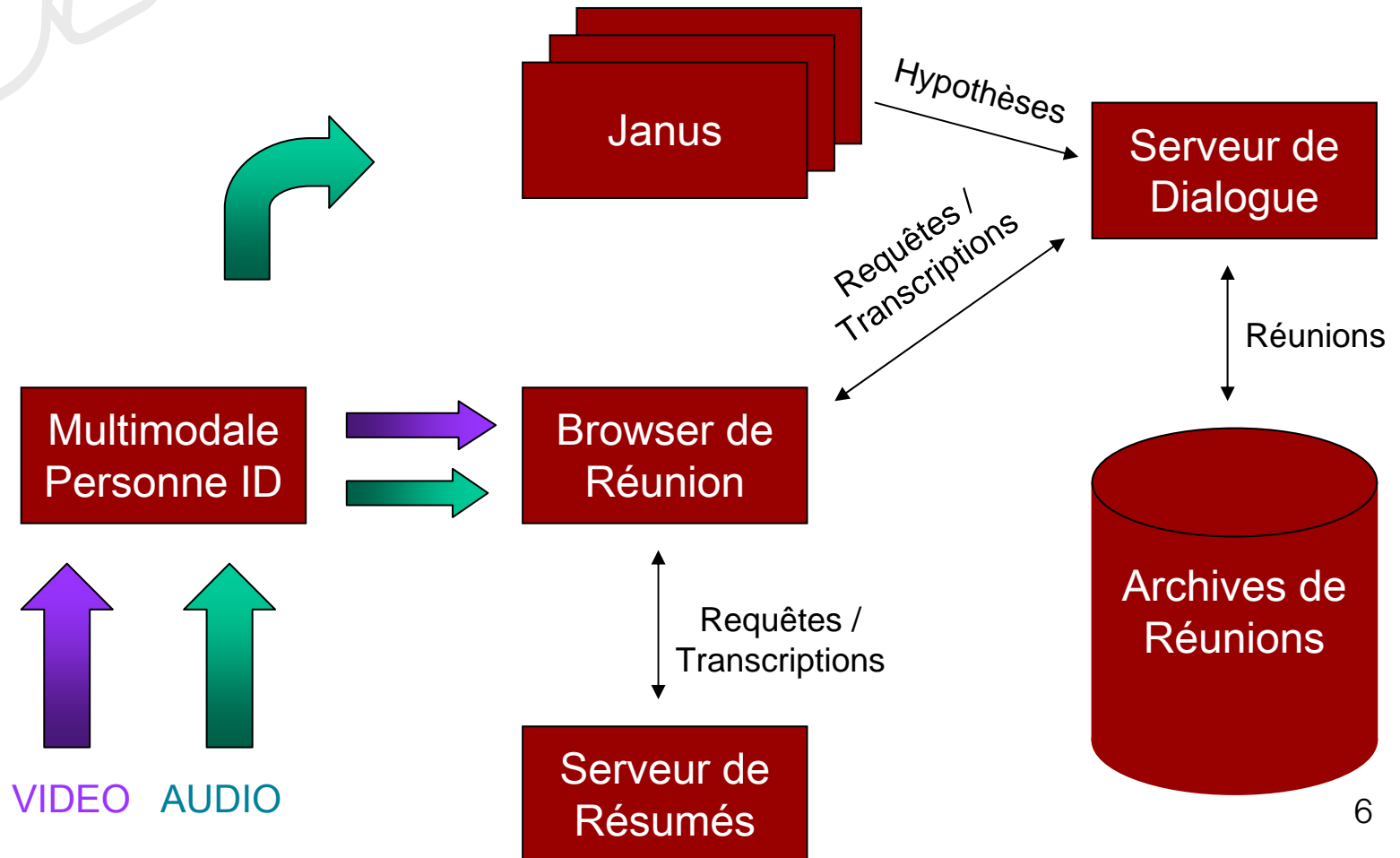
- Détection automatique pour un maximum de six orateurs
- Création automatique d'une transcription de la parole
- Création automatique de résumé(s)

## ❑ Le Système comporte trois composants principaux

- Système de reconnaissance de la parole
- Identificateur de personnes multimodale
- Browser de réunions



# Vue Générale du Système: Architecture



# Reconnaissance de la Parole

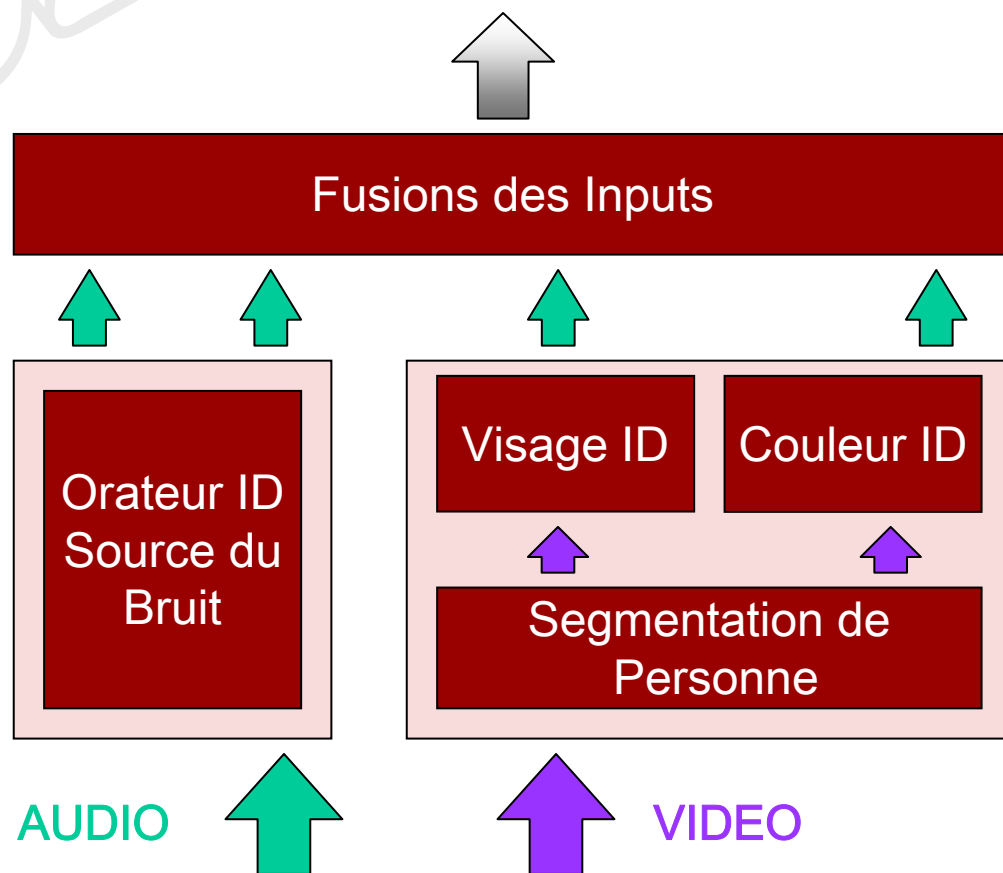
- ❑ Système : basé sur le système de reconnaissance Janus
- ❑ Données : enregistrement continu avec de multiples orateurs
- ❑ Principe de la Reconnaissance :
  - partitionnement des données en segments homogènes
  - assignation d'un label orateur à chaque segment
- ❑ Segmentation : chaque segment doit contenir les paroles d'un unique orateur
  - difficultés : sur-segmentation et sous-segmentation
- ❑ Assignation : deux techniques utilisées
  - Vocal Tract Length Normalization : estimation de facteurs de déformation
  - Speaker adaptation : modèle acoustique d'un orateur



RESULTAT

Des segments assignés à un orateur contenant uniquement les paroles de l'orateur

# Identificateur de Personne : Schéma



But de l'identificateur de personne multimodale :

Traque et identification des participants de manière continue.



# ID Personne : Les Identificateurs 1/2

## ❑ Segmentation de Personne

- étape de pré-calcul pour les autres identificateurs
- segmentation des personnes par rapport au fond
- basée sur la soustraction du fond
- modèle du fond évolutif à cause du déplacement des objets

## ❑ Identification du Visage

- en cours d'intégration au Système
- données : plusieurs visages avec plusieurs tailles et orientations
- localisation du visage dans l'image avec la couleur du visage
- techniques : Eigenface et DSW (Dynamic Space Warping)

## ❑ Identification de l'orateur et source du Bruit

- but : identifier l'orateur à tous moment
- le spectre de la parole de chaque orateur permet de les différencier
- l'entraînement pour les modèles des orateurs se fait offline
- permet l'association avec les labels orateurs de la recon. de la parole
- la source du bruit permet de combiner l'audio et la vidéo (quel micro)

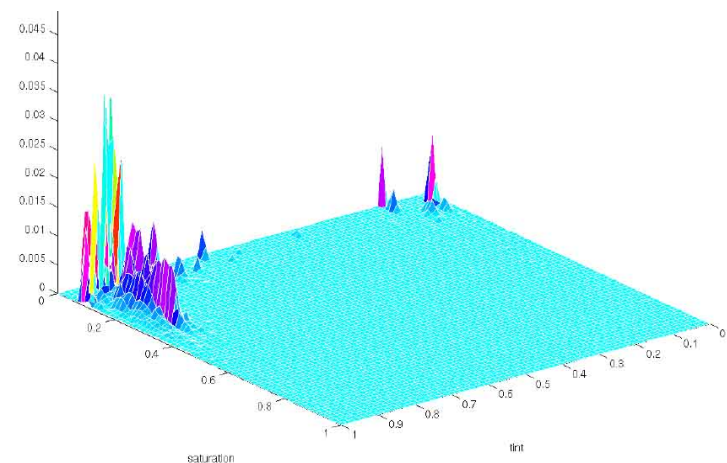


## ID Personne : Les Identificateurs 2/2

### □ Identification de la couleur

- but : identification des participants à tout moment
- principe : modèle d'histogramme de couleur pour chaque participant
- avantage : pas occlusion et de problème de changement de vue
- problème : la couleur varie avec l'illumination
- technique : histogrammes de teinte-saturation et probabilité de la fonction de distribution

→ Test réalisé avec des personnes circulant dans un vestibule. Résultats: 16 modèles distincts sur environ 5000 images.



Taille histogramme	25x 25	50x 50	75x 75	100x 100	125x 125	150x 150	175x 175	200x 200
Précision	63.2	62.3	67.8	65.8	63.4	59.1	56.4	53.3

## ID Personne : Fusion des Inputs

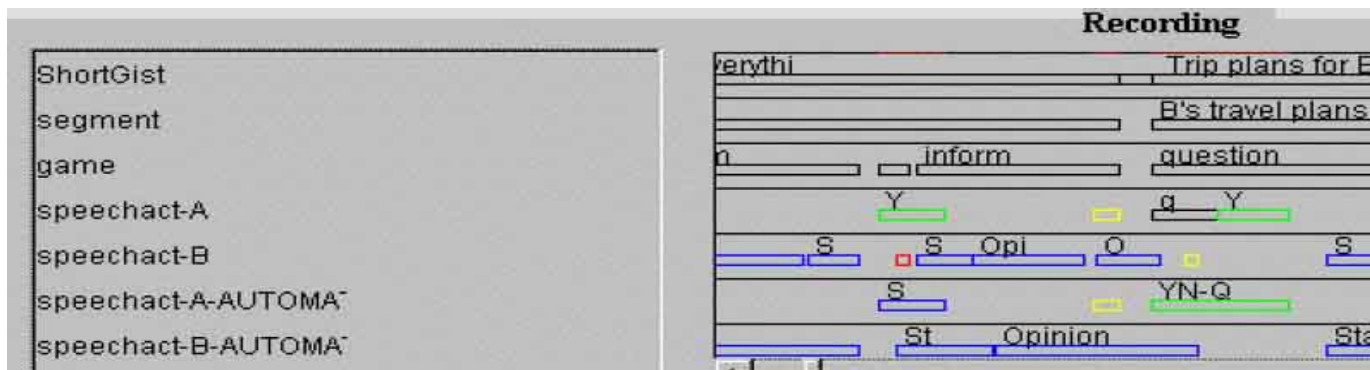
- ❑ Trouver la configuration la plus probable pour :
  - La localisation des personnes dans la salle
  - L'identité des personnes
  - L'identité de l'orateur
  
- ❑ Test réalisé avec : ( ~ 4 minutes)
  - trois participants
  - deux caméra, deux microphones



	Taux d'erreur
Sans Fusion	12.51 %
Avec Fusion	10.67 %

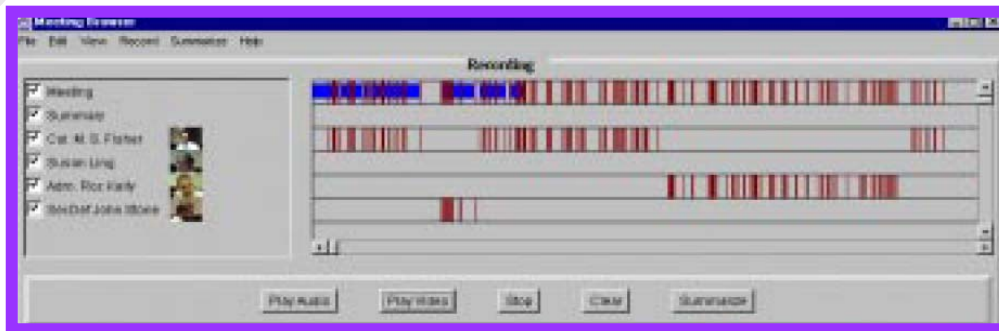
## Browser de Réunion 1/2

- ❑ Le Browser permet les possibilités suivantes
  - Accès rapide aux enregistrements (vidéo, audio, transcription)
  - Archivage des réunions dans une base de donnée
  - Création automatique de résumés
  - Corrections et/ou annotations possible des participants
  - Détection des émotions
    - o deux Systèmes développés à l'ISL
    - o fonctionnent offline
    - o résultats visualisés dans le browser



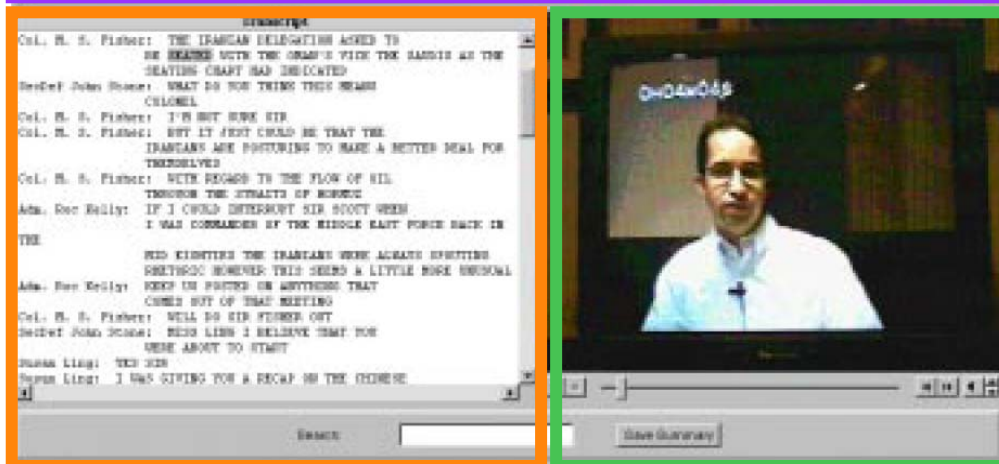
# Browser de Réunion 2/2

1



1) Déroulement et participants

2) Transcription de la réunion

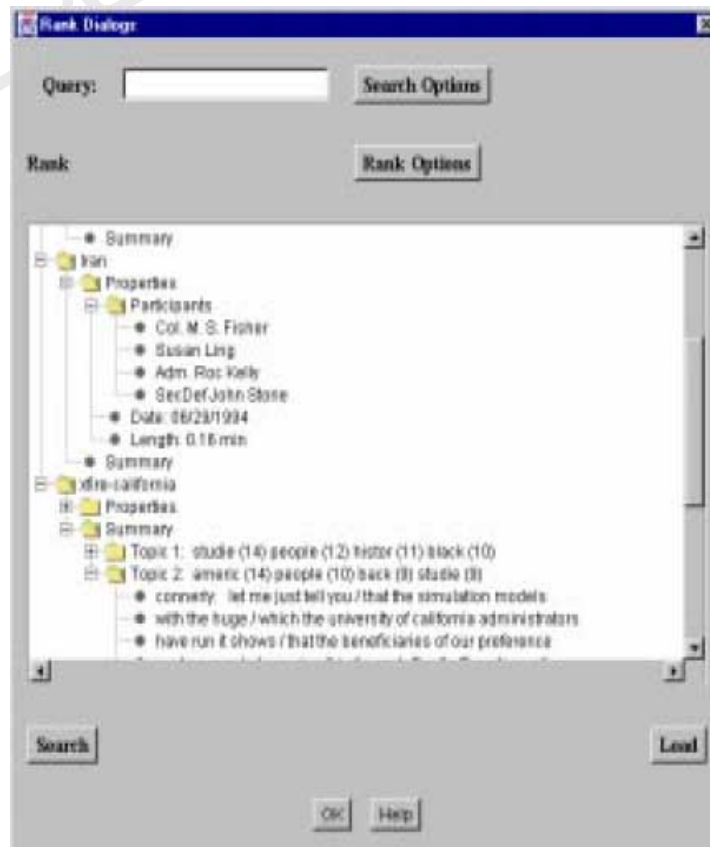


3) Vidéo ou Résumé de la réunion

2

3

# Browser de Réunion : Archivage



Les Réunions sont présentées dans un format en arbre.

Recherche dans la base de donnée :

- Participant
- Combinaison de participants
- Sujets discutés
- Mots-clés
- Durée de la réunion
- Date de la réunion

## Browser de Réunion : Résumés

- ❑ L'utilisateur spécifie la taille et le thème central du résumé
- ❑ Le résumé se fait toujours sur la base de la transcription
- ❑ L'algorithme (A. Weibel)
  1. Élimination des mots outils « stopwords »
  2. Phase de stemming et identification du stem le plus fréquent
    - simplification: quatre premières lettres des mots
  3. Pondération de chaque tour de parole (nbr d'occurrences du stem)
  4. Inclusion du tour de parole « le plus lourd » dans le résumé
  5. Les mots contenant le stem le plus fréquent sont désormais ignorés
  6. Si le résumé est plus petit que la taille spécifiée, on retourne à l'étape 2

# Travail Futur

## □ Travail Futur

- Amélioration des systèmes de reconnaissance
- Intégration de la détection du visage
- Développement d'un module de détection d'attention
- Test en dehors des laboratoires ISL
- Résumés sur de multiples réunions
- But ultime : Meeting Room sur un Laptop

# Conclusions

## ❑ Le Système d'Enregistrement

- Avantage de l'Automatisation
- Liberté maximum aux participants
- Aspect documents absent

## ❑ Les Publications et autres sources d'informations

- informations générales sur le système très brèves
- description des aspects techniques très condensée
- illustrations de mauvaise qualité
- browser très rapidement détaillé
- peu de détails sur les inputs manuels nécessaires