

La voix

ses caractéristiques et sa reconnaissance automatique

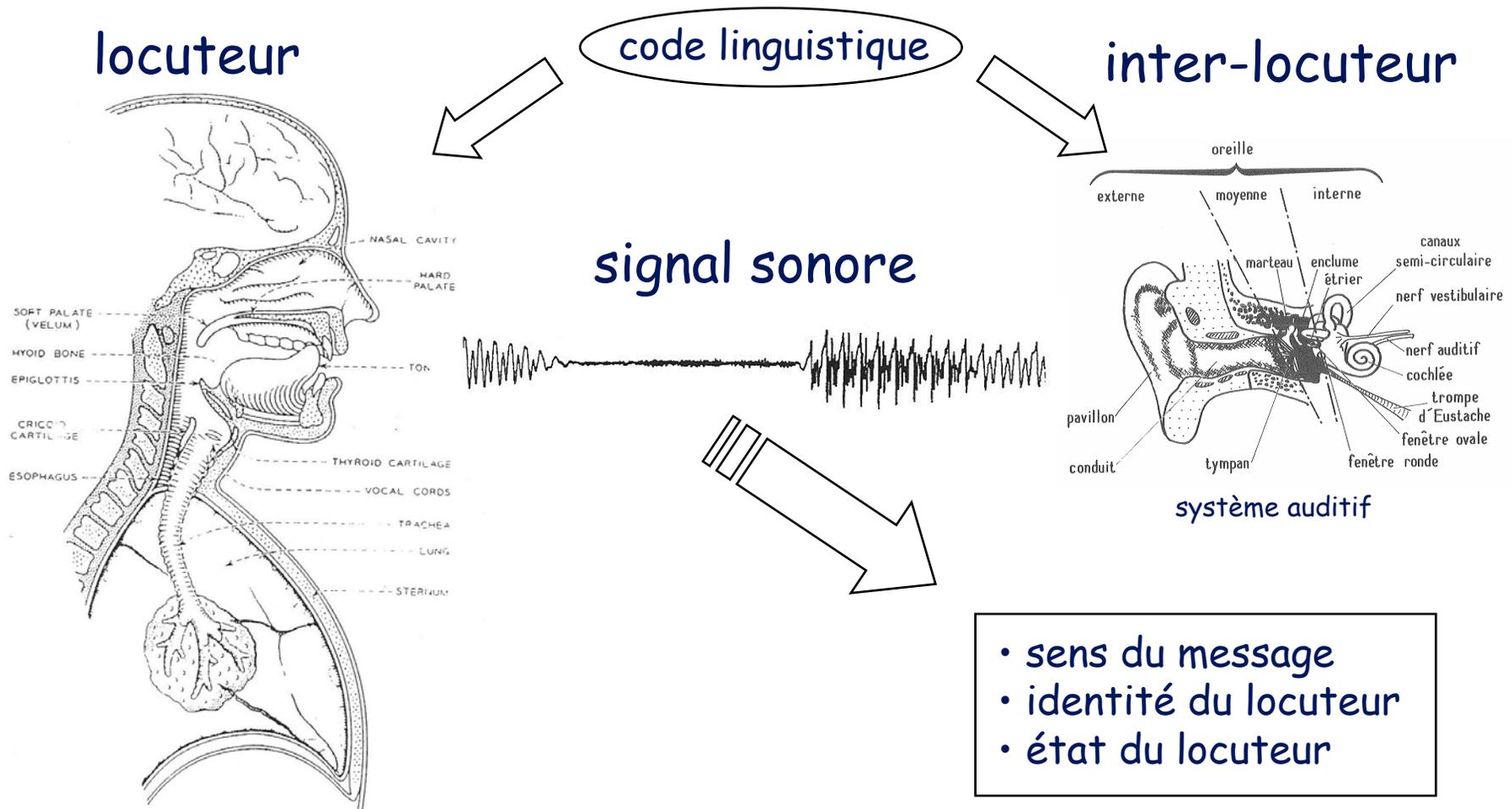
Frédéric BIMBOT

IRISA (CNRS & INRIA) - METISS

Plan de l'exposé

- La communication parlée, le locuteur et sa voix
- Fonctionnement d'un système de reconnaissance automatique
- Applications pratiques
- Questions

La communication vocale



Reconnaissance du locuteur

Processus de décision utilisant des caractéristiques du signal de parole pour déterminer des éléments concernant l'identité de la personne qui a prononcé un énoncé donné

Bonjour... Ça va ?

A 

B 

C 

D 

E 

Petit Quizz

Dans ces 5 enregistrements,

y a-t-il 2 fois la même personne ?

Ça va bien, merci !

A 

B 

C 

D 

E 

Réponse

NON

il s'agit de 5 locuteurs différents

→ *comparaison de voix*

Qui suis-je ?

A 

B 

C 

D 

E 

**Une de ces voix est
la mienne**

Laquelle d'après vous ?

Qui suis-je ?



Réponse

A: Mikaël



B: Sylvain



La voix C

C: Frédéric



D: Gilles



→ *identification de voix*

E: Boris

Spécificité et variabilité de la voix

Facteurs de spécificités qui caractérisent une voix

- la forme du conduit vocal (morphologie)
- la façon dont on s'en sert (neuro-motricité, état psychologique)
- le contexte d'apprentissage (facteurs socio-culturels)

→ SIGNATURE VOCALE

Sources de variabilité pour la voix d'une même personne

- la situation personnelle (état de santé, émotionnel, etc...)
- le contexte d'élocution (niveau de formalité → soin de l'articulation)
- l'environnement (bruits ambiants, distorsions, etc...)
- les modifications intentionnelles (masquage, imitation, ...)

→ PAS D'EMPREINTE VOCALE

La reconnaissance automatique du locuteur

- **Vérification / authentification du locuteur**
→ est-ce bien « untel » ?
- **Identification du locuteur**
→ qui est-ce (1 parmi N) ?
- **Segmentation en locuteur**
→ où parle qui dans un enregistrement
- ***Caractérisation du locuteur***
→ *à quelle catégorie appartient le locuteur ?*

Fonctionnement d'un système

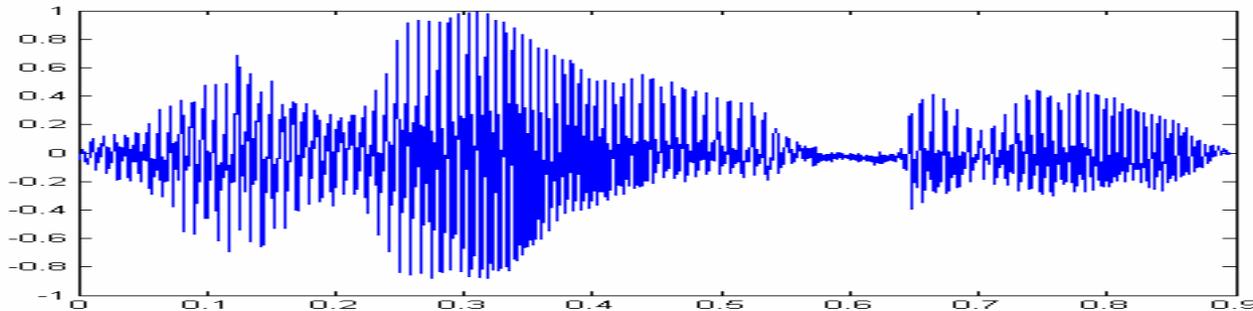
3 modes de fonctionnement

- initialisation du système (mise en service)
- apprentissage des caractéristiques de la voix des locuteurs
- test de la voix d'un énoncé dont le locuteur est inconnu

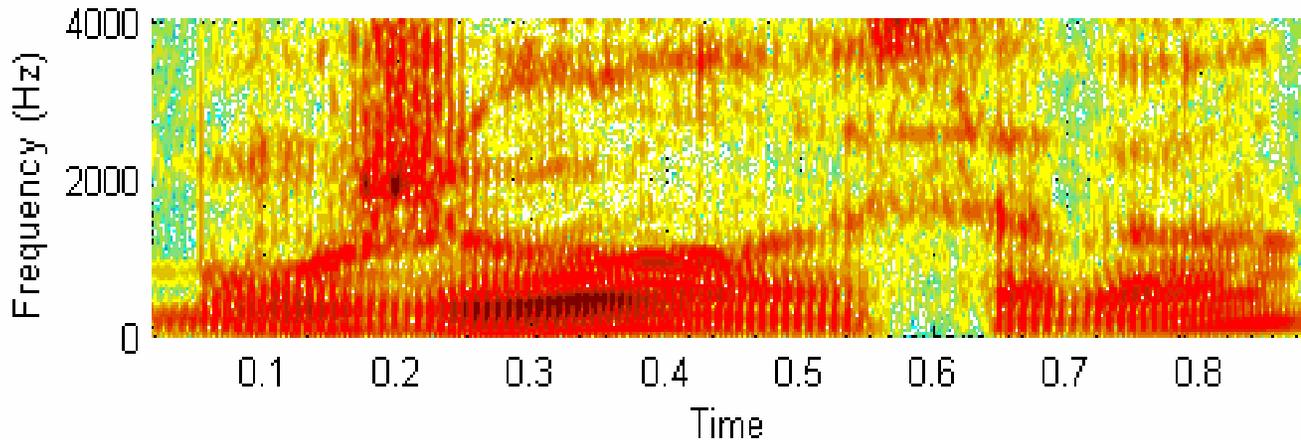
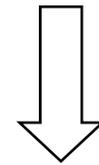
4 modules de traitement

- analyse et paramétrisation
- modélisation du locuteur (apprentissage)
- calcul de score de comparaison (test)
- décision

Analyse de la parole



signal sonore

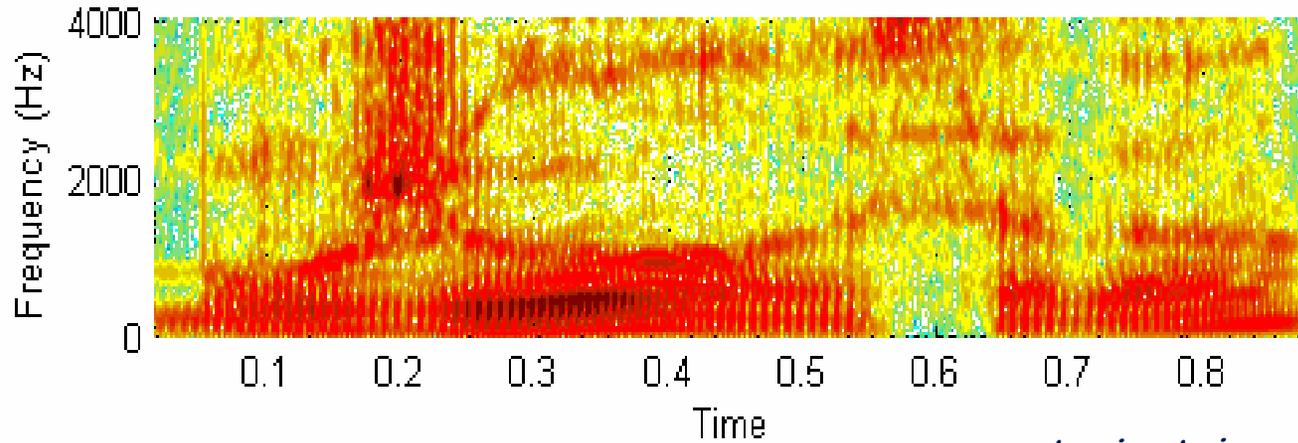


spectrogramme

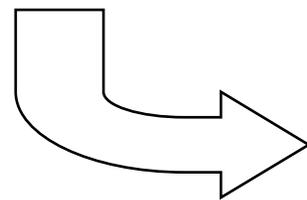
*représentation
temps-fréquence
amplitude*

Transformée de Fourier

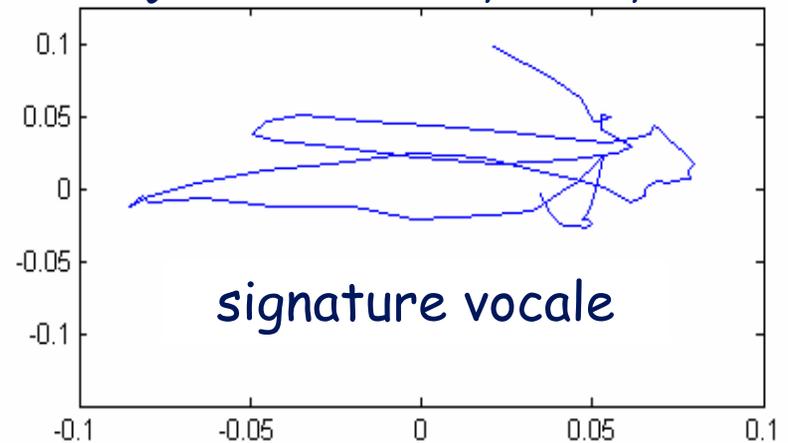
Paramétrisation



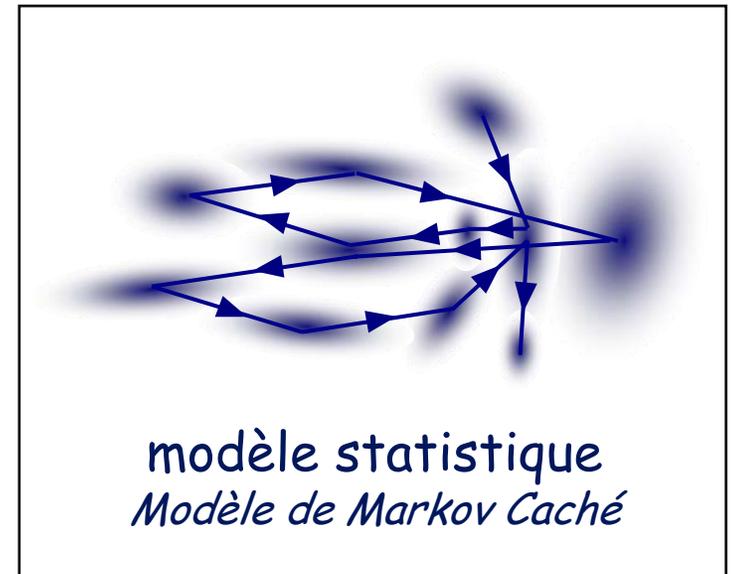
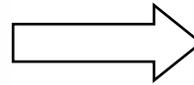
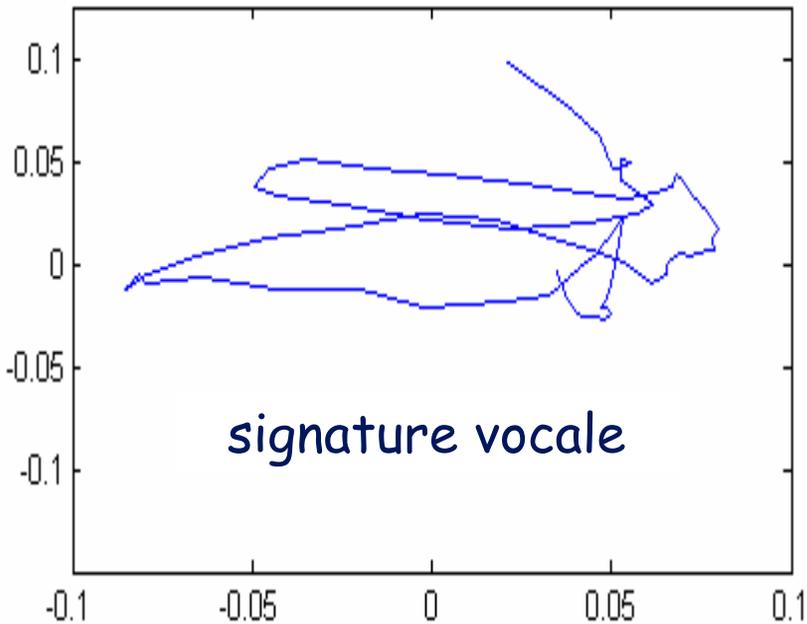
spectrogramme



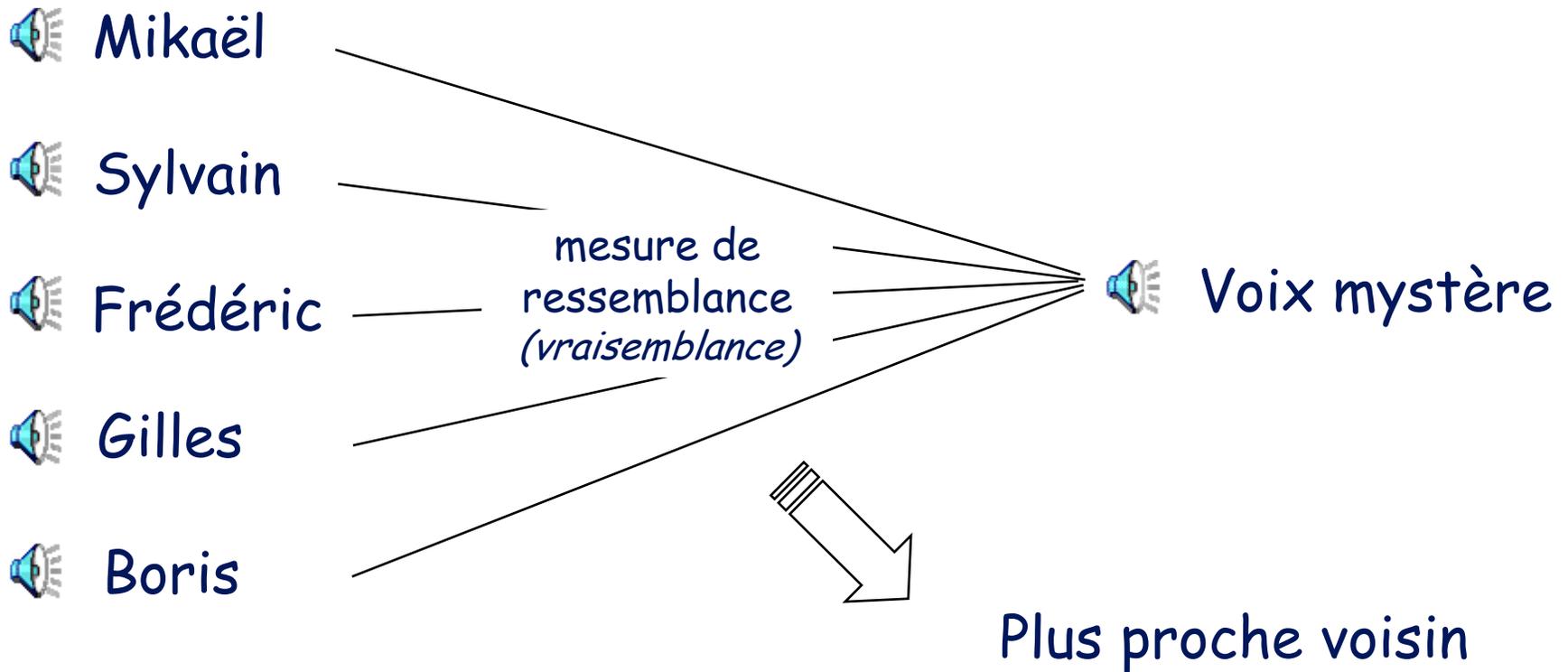
trajectoire dans l'espace cepstral



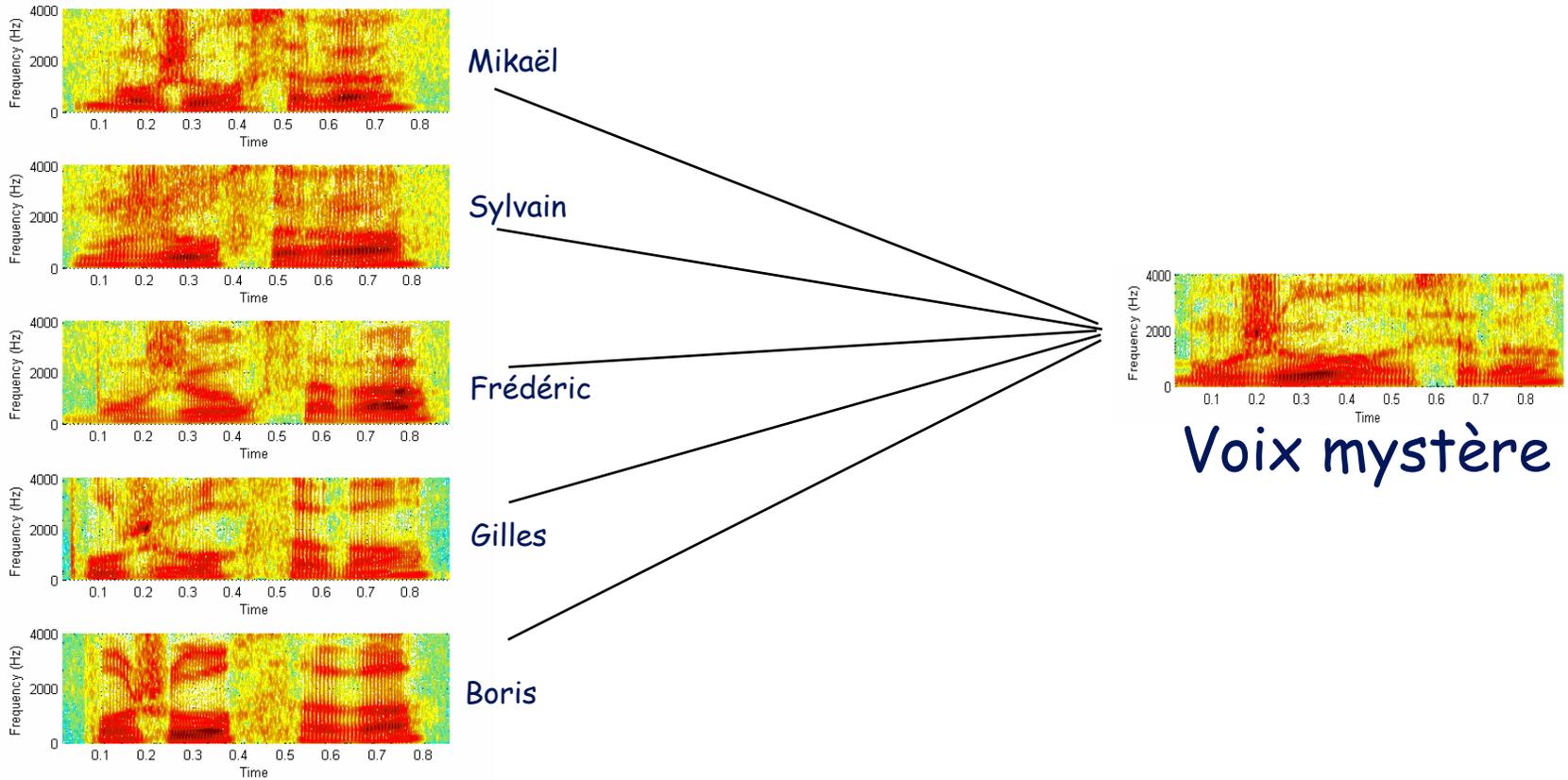
Modélisation



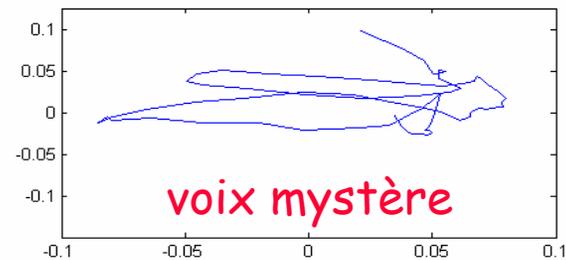
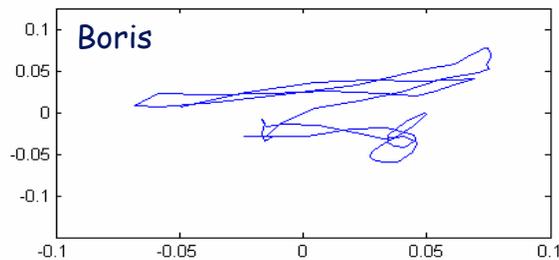
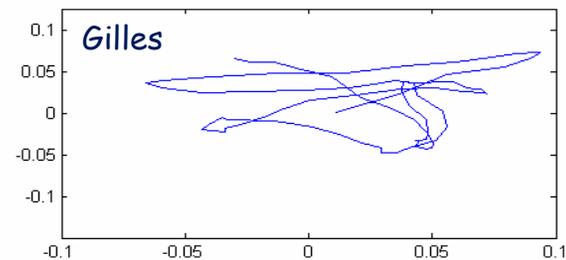
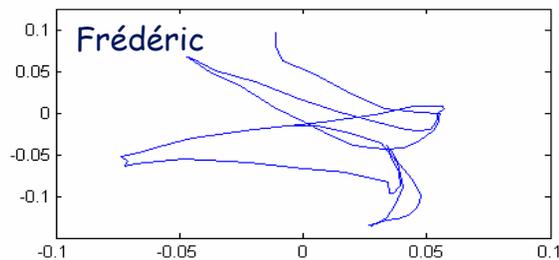
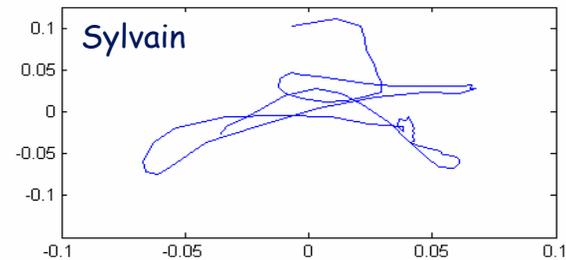
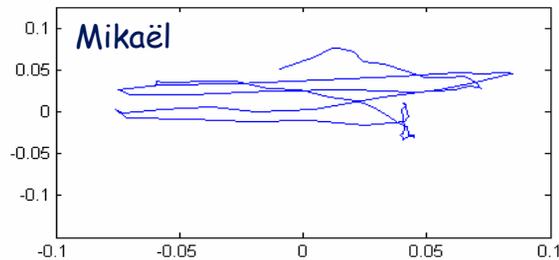
Comparaison et décision



Comparaison de spectrogrammes

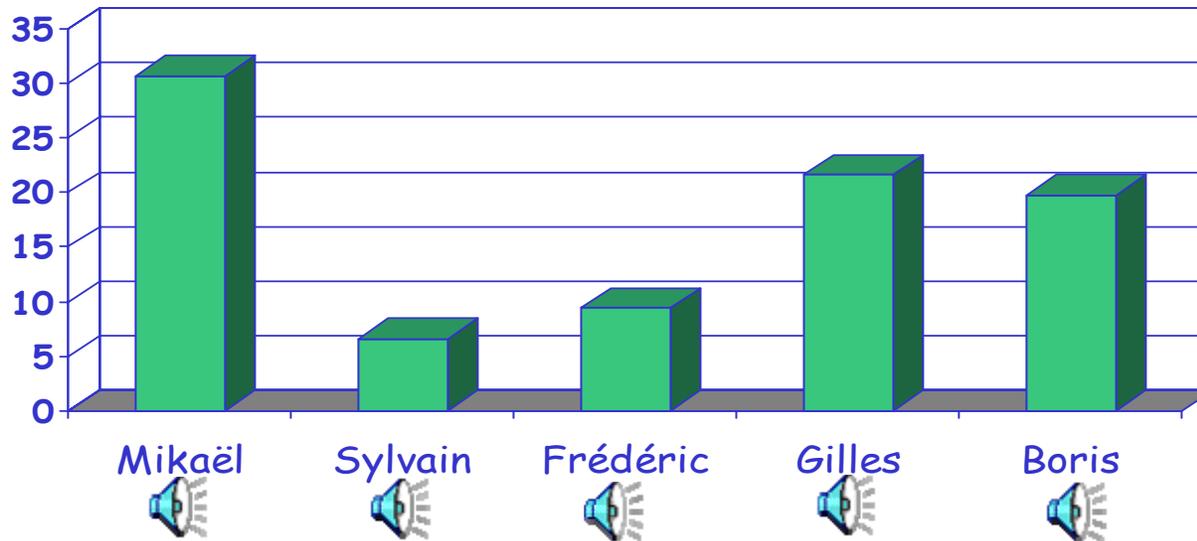


Comparaison de signatures vocales



Décision

Scores de ressemblance
entre la voix mystère
et les 5 voix de référence



 Identité la plus probable → **Mikaël**

De la théorie à la pratique...

- On n'a jamais assez d'exemples de la voix de quelqu'un pour faire des modèles statistiques vraiment fiables
- Les conditions de prise de son ou de transmission sont souvent très variables d'un enregistrement à l'autre
- La voix d'une personne évolue au cours du temps
- Difficultés à se prémunir contre :
 - la parole pré-enregistrée
 - l'imitation
 - les ressemblances dans une même famille

Performances actuelles

en authentification du locuteur

- En conditions « idéales » : 0,1 % d'erreur
- En conditions contrôlées : 1 à 2 % d'erreur
- En conditions quelconques : 5 à 10 % d'erreur

La reconnaissance de la voix n'est pas une façon très fiable d'identifier un individu

Applications

5 PROFILS TYPES :

- 1) Contrôle d'Accès Physique
- 2) Sécurisation de Transactions à Distance
- 3) Organisation de l'Information
- 4) Jeux, Jouets
- 5) Criminalistique et Renseignement

Contrôle d'accès physique

→ *L'utilisateur doit être présent en un lieu précis*

- Exemples :
- protection de locaux
 - mot de passe vocal sur ordinateur
 - accès à des valeurs ou des biens
 - passeport biométrique

Forte compétition d'autres techniques biométriques, plus fiables

Peu de systèmes envisagés sur cette base

Sécurisation de transactions à distance

→ *Reconnaître un utilisateur à travers un réseau de communication*

Exemples :

- accès à des services télécom
- opérations bancaires au téléphone
- commerce électronique sur internet

Au téléphone, la voix est la principale source d'information sur l'identité de l'utilisateur d'un service

**Objectif : diminuer la fraude
sans contrarier les clients**

Existence de quelques prototypes

Annotation automatique

→ *Transcription automatique de programmes radio et TV*

- Exemples :
- archivage d'enregistrements anciens
 - sous-titrage automatique de programmes
 - navigation dans les BD sonores
 - comptes-rendus automatiques
 - moteurs de recherche audio-visuels
 - ...

**Secteur en plein essor car les
besoins sont énormes**

Jeux et Jouets

Exemples :

- jouets pour enfants
- compagnons artificiels
- consoles de jeux
- téléphones mobiles
- jeux sur CD-ROM

*Reconnaissance du joueur
Imitation / Jeux de Rôle
Détection de l'état d'esprit*

Secteur en friche, à explorer

Criminalistique et renseignement

Exemples :

- incarcération à domicile
- orientation d'enquête, recherche de suspects
- surveillance des communications sensibles

La prudence est nécessaire car
il n'existe pas à ce jour de technique infallible
permettant de reconnaître avec certitude
la voix d'un individu

Et maintenant...

**C'est VOUS
qui avez la parole !**