

# TP Biométrie

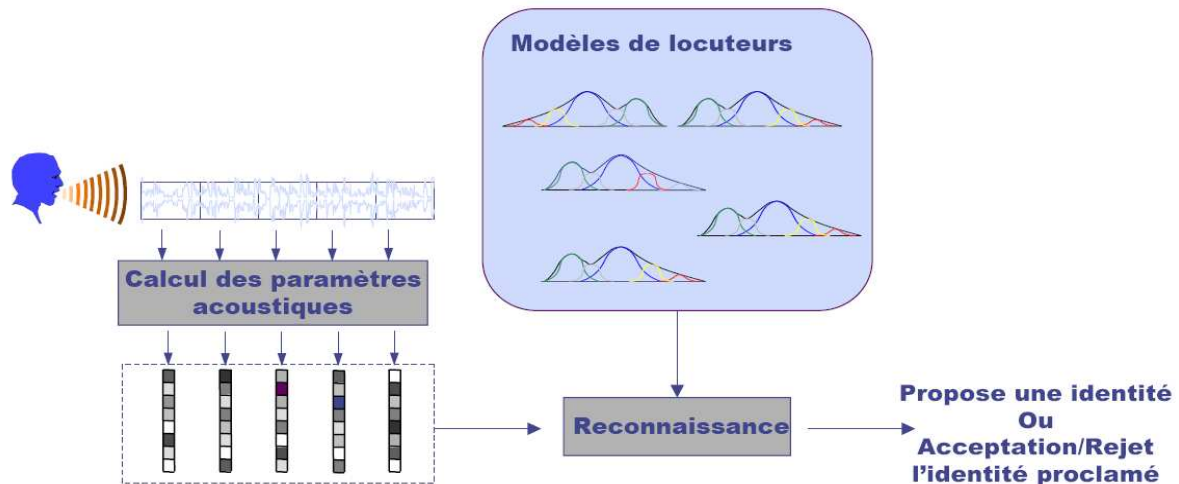
## Reconnaissance du locuteur

*D. Istrate (ESIGETEL)*

Ce TP vise à montrer un exemple de reconnaissance du locuteur par la voix en utilisant des GMM (modèles de mélange de distributions de Gauss). Nous utilisons le logiciel « open-source » ALIZE réalisé par le laboratoire LIA.

La base de données est Validb.

Le schéma d'un système de reconnaissance du locuteur est présenté dans la figure suivante :



Nous allons partir des fichiers wav qui contiennent l'enregistrement des différentes personnes pour construire notre système de reconnaissance du locuteur.

**La base Validb.** VALIDDB est une base de données multimodale constituée de cinq sessions d'enregistrement de deux textes en anglais par 106 personnes.

La première session est faite dans des conditions de studio sans bruit et les quatre autres avec du bruit réel de bureaux. Les deux textes enregistrés sont :

-1<sup>er</sup> texte : la série de chiffres « 5 0 6 9 2 8 1 3 7 4 ».

-2<sup>ème</sup> texte : la phrase « Joe took father's green shoe bench out ».

Nous disposons donc de 1060 fichiers audio (106 x 2 x 5) que nous allons utiliser pour l'évaluation des performances de quelques systèmes de vérification du locuteur.

Le format des noms des fichiers est le suivant : lll\_c\_p.wav où :

- lll est le nombre du locuteur entre 000 et 122
- c représente la condition d'enregistrement ; entre 1 et 5
- p représente la phrase prononcée ; entre 1 et 2

Nous avons divisé la base en deux parties ; une pour les clients (de 000 à 061) et l'autre pour les imposteurs (062 à 122).

**Paramétrisation.** La première étape consiste dans l'extraction des paramètres spectraux des signaux sonores pour obtenir les vecteurs acoustiques stockés dans des fichiers .prm.

Nous allons utiliser 16 paramètres LFCC et l'énergie du signal.

**Les fichiers de paramètres ont été déjà calculés.**

Pour calculer, vous-même, les paramètres :

Vous devez démarrer le logiciel « RecoLocuteur ».

Avant d'actionner le bouton Paramétrisation il faut préparer les fichiers de configuration qui se trouvent dans le répertoire cfg et s'appellent lfcc\_Clients.cfg et lfcc\_Imposteurs.cfg.

La première ligne du fichier contient le chemin où se trouve les fichiers wav à traiter et la deuxième le chemin du repertoire où sauvegarder les fichiers prm. Le paramètre suivant indique le nombre de filtres à utiliser (24) et le quatrième combien de coefficients nous allons utiliser (16). Le paramètre suivant si est égal à 1 indique qu'on rajoute l'énergie du signal comme paramètre. Les deux derniers coefficients indiquent la taille de la fenêtre de calcul (16 ms) et le recouvrement entre 2 fenêtres (50%).

Une fois paramétrer ces fichiers on peut lancer la paramétrisation des fichiers wav dans les répertoires PRM\clients et PRM\imposteurs.

Visualisez un fichier de paramètres (prm) en utilisant la fonctionnalité « Visualisation paramètres ». Cela vous permet de voir à la fois le signal, son énergie et un des paramètres LFCC (1 à la fois). En appuyant sur le bouton « Spectrogramme » vous allez obtenir la variation du spectre du signal dans le temps. La spectrogramme est calculé en utilisant une fenêtre de 1024 échantillons avec un recouvrement de 512 échantillons.

**Analysez plusieurs fichiers.**

**Tenant compte de la méthode de calcul des paramètres LFCC pour un exemple analysez les courbes affichés (tenez compte de la bande de fréquence de chaque coefficient et rapportez vous à la spectrogramme).**

**Détection d'énergie (Energy Detector).** Cette étape vise à détecter les périodes de silence dans les fichiers. Le résultat est constitué d'un fichier d'étiquettes (.lbl) qui indique les zones avec signal. L'algorithme utilise 3 distributions de Gauss sur le seul paramètre énergie.

**Les fichiers lbl sont déjà calculés.**

Pour générer ces fichiers, vous-même, :

Les fichiers de configuration sont « EnergyDetector\_LFCC\_Clients.cfg » et « EnergyDetector\_LFCC\_Imposteurs.cfg ». Les paramètres à modifier sont :

- featureFilePath = le chemin des fichiers prm
  - labelFilePath = le chemin où nous allons sauvegarder les fichiers .lbl (un seul répertoire LBL pour les clients et les imposteurs)
  - inputFeatureFilename = le fichier contenant la liste des fichiers à analyser
- Pour obtenir la liste des fichiers à analyser utiliser la commande « Création Liste simple » en indiquant le répertoire les prm des clients et des imposteurs.

**Analysez le résultat en utilisant la fonctionnalité « Visualisation étiquettes ».**

**Normalisation des paramètres.** Cette étape effectue une normalisation des paramètres acoustiques. Le résultat de la normalisation est constitué des nouveaux fichiers .norm.prm qu'on va stocker dans les mêmes répertoires prm.

**La normalisation a été déjà effectuée.**

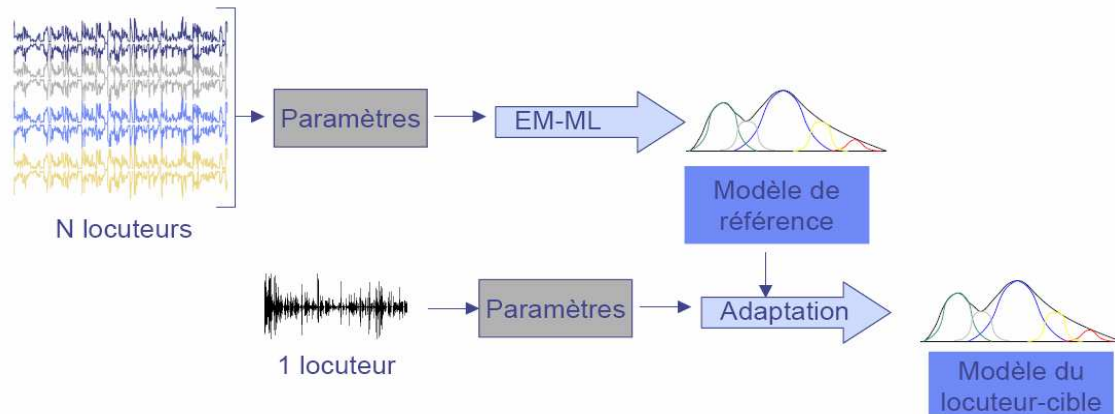
Pour la réaliser vous même, les 2 paramètres à modifier :

- `featureFilesPath` = le chemin des fichiers prm
  - `inputFeatureFilename` = le fichier contenant la liste des fichiers à analyser
- Nous allons utiliser les mêmes fichiers de liste que pour EergyDetector.

**Pour un même fichier sonore analyser le fichier .prm et le .norm.prm correspondant.**

**Apprentissage du modèle du monde.** Pendant cette étape nous allons créer un modèle du locuteur en générale. Nous allons utiliser tous les fichiers disponibles dans notre base de données.

Rappel :



**Le fichier correspondant au modèle du monde a été déjà créé.**

Pour le re-générer, le fichier de configuration pour TrainWorld est « TrainWorld\_LFCC.cfg » et les paramètres à modifier sont :

- `featureFilesPath` = le chemin des fichiers prm
- `mixtureFilesPath` = le chemin où nous allons sauvegarder le gmm
- `labelFilesPath` = le chemin où nous allons sauvegarder les fichiers .lbl (un seul répertoire LBL pour les clients et les imposteurs)
- `inputFeatureFilename` = le fichier contenant la liste des fichiers nécessaires pour le modèle du monde
- `outputWorldFilename` = le nom du fichier de sortie (le gmm du monde).

Le fichier liste on l'obtient avec la même commande « Création liste simple » sur le répertoire contenant tous les prm.

**Le fichier modèle est sauvegardé dans un fichier de type XML. Analysez le modèle du monde. (Combien de distributions de Gauss ? Quelle est celle la plus représentative ?)**

Nous allons étudier deux systèmes d'identification du locuteur en mode indépendant du texte et en mode dépendant du texte.

### 1. Système indépendant du texte

Dans le cas de ce système pour l'apprentissage nous utilisons le 1<sup>er</sup> texte avec les 5 conditions d'enregistrement et pour le test le 2<sup>ème</sup> texte avec les 5 conditions d'enregistrement.

**Apprentissage des modèles clients.** Cette étape réalise la modélisation GMM des modèles des clients utilisant le modèle du monde créé à l'étape précédente et les fichiers correspondants de chaque client.

**Les fichiers des clients ont été déjà réalisés.**

Pour le générer vous même, le fichier de configuration de TrainTarget est « TrainTarget\_LFCC\_Clients.cfg » avec les paramètres à modifier :

- featureFilesPath = le chemin des fichiers prm
- mixtureFilesPath = le chemin où nous allons sauvegarder le gmm
- labelFilesPath = le chemin où nous allons sauvegarder les fichiers .lbl (un seul répertoire LBL pour les clients et les imposteurs)
- targetIdList = un fichier ndx qui indique les fichiers d'apprentissage de chaque client
- inputWorldFilename = le nom du fichier gmm du monde

Le fichier ndx a la forme suivante :

mmm f1 f2 f3

où

- mmm = le nom du modèle à créer
- f1, f2, f3,... = les noms des fichiers d'apprentissage du modèle mmm

Ce fichier peut être créer en utilisant la commande « Création Train ndx » qui va demander : le chemin des fichiers prm clients, un fichier de configuration et le nom du fichier de sauvegarde.

A la fin de l'exécution du TrainTarget vous pouvez retrouver les modèles des clients dans le répertoire gmm.

**Analysez le modèle gmm d'un client (toujours en format XML) par rapport au modèle du monde. (Comparez les distributions de même rang)**

**Calcul des vraisemblances des fichiers de test.** Cette étape calcul pour chaque fichier de test sa vraisemblance par rapport à tous les modèles des clients.

**Les calculs ont été déjà effectués.**

Pour refaire les calculs de reconnaissance, les fichiers de configuration de ComputeTest sont « ComputeTest\_LFCC\_Clients.cfg » et « ComputeTest\_LFCC\_Imposteurs.cfg ». Les paramètres à modifier :

- featureFilesPath = le chemin des fichiers prm
- mixtureFilesPath = le chemin où nous allons sauvegarder le gmm
- ndxFilename = une liste avec les fichiers à tester par rapport à quel modèle
- outputFilename = le nom du fichier des résultats
- worldModelFilename = le nom du fichier du monde

Le fichier du monde est nécessaire pour normaliser les vraisemblances.

Le fichier ndx peut être obtenu avec la commande « Création Test ndx » qui demande le chemin du répertoire où se trouvent soit les fichiers prm des clients soit ceux des imposteurs, un fichier de configuration, le nom du fichier à sauvegarder et le répertoire avec les modèles gmm. Le fichier de configuration a la structure suivante :

- Un nombre indiquant le nombre de configurations
- Un nombre indiquant les nombres des conditions utilisés suivi des codes de ces conditions (1 à 5)

- Un nombre indiquant le nombre de phrase suivi du code des phrases

ComputeTest va créer en sortie un fichier .res qui indique la vraisemblance de chaque fichier à tester par rapport à chaque modèle.

**Traçage des courbes ROC.** Un fichier est reconnu comme appartenant à un locuteur si la vraisemblance obtenu par rapport au modèle de ce locuteur est la plus grande et plus grande qu'un seuil. Les paramètres utilisés pour caractériser un système de reconnaissance du locuteur sont FAR (le taux de fausses acceptations) et FRR (le taux de fausses rejections) et EER qui l'intersection de la première bissectrice avec la courbe ROC (FAR en fonction de FRR). La commande « Visu résultats Ind Texte » part des fichiers .res obtenus par ComputeTest et génère dans une première étape des nouveaux fichiers qui contient pour chaque fichier à tester seulement le modèle avec la vraisemblance la plus grande et cette valeur. En suite en faisant varier un seuil calcule FAR et FRR.

Si un client est reconnu comme locuteur de la base c'est OK sinon on comptabilise comme FRR. Si un imposteur est reconnu comme un client alors on comptabilise comme FAR.

**Analysez les résultats obtenus. Puis-je avoir un FAR de 10% ? Quel serait le FRR ? Quels sont les limites de variations du seuil pour obtenir la courbe ROC ?**

## **2. Système dépendant du texte**

Dans le cas de ce système pour l'apprentissage nous utilisons le 1<sup>er</sup> texte avec les 2 premières conditions d'enregistrement et pour le test le 1<sup>er</sup> texte avec les 3 autres conditions et le 2<sup>ème</sup> texte avec les 5 conditions. Un locuteur n'est accepté par le système que s'il correspond à l'identité proclamée et s'il prononce le 1<sup>er</sup> texte.

Réalisez les étapes précédents pour ce système dépendant du texte (vous allez démarrer avec l'apprentissage des modèles des clients et continuer avec le calcul des résultats).

**Donnez les résultats et comparez par rapport au système indépendant du texte.**