

**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et de la Recherche**  
**Scientifique**

**Ecole Nationale Polytechnique**



المدرسة الوطنية العليا المتعددة التقنيات  
Ecole Nationale Supérieure Polytechnique

**Département d'Electronique**  
**Laboratoire Signal & Communications**

**Mémoire de Magister**

**Présenté par : DJEMAI Mohamed**

**Thème**

**Vérification Automatique du Locuteur**  
**Arabophone par les Algorithmes**  
**Génétiques**

**Devant le Jury :**

<b>L. HAMAMI</b>	<b>Professeur</b>	<b>ENP</b>	<b>Président</b>
<b>M. GUERTI</b>	<b>Professeur</b>	<b>ENP</b>	<b>Rapporteur</b>
<b>L. MITICHE</b>	<b>Maître de Conférence</b>	<b>UZA-Djelfa</b>	<b>Examineur</b>
<b>N.BENBLIDIA</b>	<b>Maître de Conférence</b>	<b>USD-Blida</b>	<b>Examineur</b>

**30 Octobre 2011**

# *Dédicace*

Je dédie ce modeste travail à mes parents, mes frères et mes sœurs, mes amis et toutes personnes qui m'ont aidé de loin ou de près.

# Remerciements

*J'exprime mes sincères et profonds remerciements et reconnaissances au Professeur M. GUERTI, enseignante à l'ENP, qui m'a encadré durant ce travail. Je la remercie aussi pour ses conseils et pour m'avoir dirigé et encouragé pendant l'accomplissement de ce travail.*

*Mes vifs remerciements vont également aux membres du jury de soutenance de ce mémoire :*

*- Madame L. HAMAMI, professeur à l'ENP, d'avoir accepté la présidence du jury de soutenance.*

*- Madame N. BENBLIDIA, Maître de conférences à l'USD de Blida, d'avoir accepté d'être examinateur.*

*- Monsieur L. MITICHE, Maître de conférences à l'UZA de Djelfa, d'avoir accepté d'être examinateur.*

*Je tiens à exprimer mes remerciements à tous les enseignants du Département d'Electronique, ainsi qu'à toute personne ayant contribué de près ou de loin à l'élaboration de ce travail.*

## ملخص

يعتبر هذا العمل جزء من مجال الاستكشاف الآلي للكلام وبالتحديد الاستكشاف الآلي للمتكلم وهذا له عدة مجالات البحث، مثل: التتبع الآلي للمتكلم، التعرف الآلي على للمتكلم، والتحقق الآلي من للمتكلم،... وما إلى ذلك. وقد ركزنا على هذا الأخير من صنف المرتبط بالنص. لهذا، استخدمنا كلمات منطوقة من طرف متكلمين باللغة العربية مسجلة في وسط ملائم. وتستند غالبية الأنظمة الحالية، على النموذج المتعدد الغوصيات باستخدام طريقة التشابه الأقصى. ويستخدم هذا الأسلوب على نطاق واسع بسبب بساطته والتتبع الرياضي، ومع ذلك، فإنه سوف يؤدي عموماً إلى نموذج دون المستوى الأمثل لجميع النماذج الأولية الذي تم اختيارها. للتغلب على هذا المشكل، نقترح خوارزمية مهجنة بين الخوارزميات الوراثية و طريقة التشابه الأقصى لنمذجة المتكلم على أساس النموذج المذكور سابقاً. لقد قمنا بتطبيق الخوارزميات الوراثية - نظراً لقدرتها على تحسين مشاكل أكثر تعقيداً - في تدريب المتكلم من أجل التحقق من هويته، استخدمنا طريقة التشابه الأقصى للنمذجة كعامل مساعد لتحسين سرعة التقارب للخوارزميات الوراثية. وكانت نسبة التعرف الآلي في حدود 98% مما يثبت أن فعالية هذه الخوارزمية المهجنة أفضل من التقديرات التقليدية في نمذجة المتكلم على أساس النموذج المتعدد الغوصيات.

**مفاتيح الكلمات :** النموذج المتعدد الغوصيات، الاستكشاف الآلي للكلام، والتحقق الآلي من للمتكلم، الخوارزميات الوراثية، التشابه الأقصى.

## Résumé

Ce travail de Magister s'inscrit dans le domaine du Traitement Automatique de la Parole (TAP) et plus particulièrement dans la Reconnaissance Automatique du Locuteur (RAL). Cette dernière a plusieurs axes de recherche, tels que : l'Indexation Automatique du Locuteur (InAL), l'Identification Automatique du Locuteur (IAL), la Vérification Automatique du Locuteur (VAL), etc.

Nous avons mis l'accent sur la VAL en mode dépend du texte. Pour cela, nous avons utilisé un corpus comprenant des mots prononcés par des locuteurs arabophones et enregistrés dans un milieu ambiant sous forme Wav.

La majorité des systèmes actuels de la VAL, sont basés sur le Modèle de Mélange de Gaussiennes (GMM) en utilisant une méthode (ML). Cette méthode est largement utilisée à cause de sa simplicité et sa traçabilité mathématique, cependant, elle mène généralement à un modèle sous optimal pour tous les modèles initiaux arbitraires choisis. Pour pallier à cette limitation, nous proposons un algorithme hybride (AG-ML) pour la modélisation du locuteur basée sur les GMM. Nous avons appliqué les Algorithmes Génétiques (AG) à l'apprentissage du locuteur pour vérifier son identité en raison de leurs capacités concernant les problèmes d'optimisation les plus compliqués. Nous avons utilisé la réestimation ML comme un opérateur heuristique pour améliorer la vitesse de convergence de l'AG.

Les Taux de Reconnaissance (TR) sont égaux à 98%, ceux-ci montrent que la méthode hybride AG-ML proposée est meilleure que celle de l'estimation traditionnelle ML dans la modélisation du locuteur basée sur les GMM.

**Mots clés :** modèle GMM ; Reconnaissance Automatique du Locuteur ; Vérification Automatique du Locuteur ; Algorithmes Génétiques ; Maximum Likelihood.

## Abstract

This work is part of Magister in the field of Automatic Speech Processing (TAP), particularly in the Automatic Speech Recognition (ASR). This has several research areas, such as: Follow automatic speaker (FAS), the automatic speaker identification (ASI), the automatic speaker verification (VAL), etc..

We have focused on the VAL mode depends on the text. For this, we used a corpus consisting of words of Arabic speakers and recorded in an ambient medium as Wav.

The majority of current systems of VAL, are based on Gaussian Mixture Model (GMM) using a method (ML). This method is widely used because of its simplicity and mathematical traceability, however, it will generally lead to a suboptimal model for all arbitrary initial models chosen. To overcome this limitation, we propose a hybrid algorithm (GA-ML) for modeling the speaker based on GMM. We applied the genetic algorithms (GA) to learn the speaker to check his identity because of their capacity for optimization problems more complicated. We used the ML re-estimation as a heuristic operator to improve the convergence speed of GA.

The recognition rates (RR) are equal to 98%, they show that the hybrid method proposed GA-ML is better than the traditional ML estimation in the modeling of the speaker based on GMM.

**Keywords:** Gaussian Mixture Model, Automatic Speech Recognition; automatic speaker verification; genetic algorithms; Maximum Likelihood.

## Liste des Abréviations

<b>API</b> :	<b>A</b> lphabet <b>P</b> honétique <b>I</b> nternational
<b>AG</b> :	<b>A</b> lgorithmes <b>G</b> énétiq <u>u</u> es
<b>AS</b> :	<b>A</b> rabe <b>S</b> tandard
<b>AR</b> :	<b>A</b> uto <b>R</b> égressif
<b>BD</b> :	<b>B</b> ase de <b>D</b> onnées
<b>DET</b> :	<b>D</b> etection <b>E</b> rror <b>T</b> ra-deoff
<b>DCF</b> :	<b>D</b> écision <b>C</b> ost <b>F</b> unction
<b>CHM</b> :	<b>C</b> ommunication <b>H</b> omme- <b>M</b> achine
<b>DTW</b> :	<b>D</b> ynamic <b>T</b> ime <b>W</b> arping
<b>EM</b> :	<b>E</b> xpectation <b>M</b> aximization
<b>FA</b> :	<b>F</b> ausses <b>A</b> ceptations
<b>FR</b> :	<b>F</b> aux <b>R</b> ejets
<b>GMM</b> :	<b>G</b> aussian <b>M</b> ixture <b>M</b> odels
<b>HTER</b> :	<b>H</b> alf <b>T</b> otal <b>E</b> rror <b>R</b> ate
<b>HMM</b> :	<b>H</b> idden <b>M</b> arkov <b>M</b> odels
<b>IAL</b> :	<b>I</b> dentification <b>A</b> utomatique du <b>L</b> ocuteur
<b>LPC</b> :	<b>L</b> inear <b>P</b> redictive <b>C</b> oding
<b>LPCC</b> :	<b>L</b> inear <b>P</b> rediction <b>C</b> epstral <b>C</b> oefficients
<b>MFCC</b> :	<b>M</b> el scaled <b>F</b> requency <b>C</b> epstral <b>C</b> oefficients
<b>MMC</b> :	<b>M</b> odèles de <b>M</b> arkov <b>C</b> achés
<b>ML</b> :	<b>M</b> aximum <b>L</b> ikelihood
<b>PLP</b> :	<b>P</b> rédition <b>L</b> inéaire <b>P</b> erceptuelle
<b>RAL</b> :	<b>R</b> econnaissance <b>A</b> utomatique de la <b>L</b> ocuteur
<b>ROC</b> :	<b>R</b> eceiver <b>O</b> perating <b>C</b> haracteristic
<b>RN</b> :	<b>R</b> éseaux de <b>N</b> eurones
<b>RAP</b> :	<b>R</b> econnaissance <b>A</b> utomatique de la <b>P</b> arole
<b>TA</b> :	<b>T</b> raitement <b>A</b> utomatique
<b>TTS</b> :	<b>T</b> ext- <b>T</b> o- <b>S</b> peech
<b>VAL</b> :	<b>V</b> érification <b>A</b> utomatique du <b>L</b> ocuteur

# Sommaire

---

Introduction générale .....	1
-----------------------------	---

## **Chapitre 1 : Reconnaissance Automatique du Locuteur**

1.1. Introduction.....	3
1.2. Production et analyse de parole .....	3
1.2.1 Pré-traitement acoustique.....	4
1.2.2 La pré-accentuation.....	4
1.2.3 Le fenêtrage.....	5
1.3 Extraction de paramètres .....	5
1.3.1 Analyse cepstrale .....	5
1.3.2. Analyse par prédiction lineaire .....	9
1.4. Description des sons de l'Arabe Standard .....	12
1.4.1. Particularites des sons de l'Arabe Standard.....	12
1.4.2. Problèmes de la Langue Arabe en traitement automatique .....	13
1.4.2.1. Agglutination des mots .....	13
1.4.2.1. Voyellation.....	14
1.5. Système de reconnaissance automatique de locuteurs.....	14
1.6. Variabilité .....	14
1.6.1. Conditions d'enregistrement.....	15
1.6.2. Variabilité intra-locuteur.....	15
1.6.3. Variabilité inter-locuteur.....	16
1.7. Applications et tâches pour la RAL.....	16
1.7.1. Identification du locuteur.....	16
1.7.2. Vérification Automatique du Locuteur.....	17
1.8. Mode dépendant et indépendant du texte .....	17

1.9. Conclusion .....	18
-----------------------	----

## **Chapitre 2 : Vérification Automatique du Locuteur**

2.1. Introduction.....	19
2.2. Modélisation des locuteurs .....	19
2.2.1. L'approche vécatorielle.....	19
2.2.1.1. L'alignement temporel dynamique (DTW-Dynamic Time Warping).....	19
2.2.1.2. Quantification vécatorielle .....	20
2.2.2. L'approche statistique.....	20
2.2.2.1. Modèle de Markov caché.....	20
2.2.2.2. Les mélanges de gaussiennes.....	20
2.2.3. L'approche connexionniste .....	21
2.2.4. L'approche relative.....	21
2.3. Modules d'un système de vérification du locuteur.....	22
2.3.1. Le prétraitement.....	22
2.3.2. L'apprentissage.....	22
2.3.3. L'attribution de scores .....	22
2.3.4. La prise de décision .....	22
2.4. Modélisation par les GMM.....	23
2.5. Estimation des paramètres .....	26
2.5.1. Algorithme des k-moyennes .....	26
2.5.1.1. L'initialisation aléatoire.....	26
2.5.1.2. L'algorithme à seuil .....	27
2.5.1.3. La règle de centroïde.....	27
2.5.1.4. La règle de plus proche voisin .....	27

2.5.2. Algorithme EM (Expectation Maximisation).....	28
2.6. La phase d'apprentissage .....	29
2.7. Le rapport d'hypothèses bayésien.....	29
2.8. Evaluation d'un système de VAL.....	29
2.8.1. Mésures de performances .....	31
2.8.2. Les courbes DET .....	32
2.8.3. Point de fonctionnement .....	32
2.8.4. Choix du seuil de décision .....	33
2.9. Conclusion .....	34

### **Chapitre 3 : Notions de base sur les Algorithmes Génétiques**

3.1. Introduction.....	35
3.2. Concepts de base des algorithmes génétiques .....	35
3.3. Caractéristiques des algorithmes génétiques .....	35
3.4. Etapes de fonctionnement.....	36
3.4.1. Présentation d'AG.....	37
3.4.2. Codage .....	40
3.4.2.1. Codage binaire .....	40
3.4.2.2. Codage de gray .....	39
3.4.2.3. Codage en réel de paramètres .....	39
3.4.3. Génération de population initiale.....	41
3.4.4. La fonction d'évaluation .....	41
3.4.5. Sélection .....	41
3.4.5.1. Roulette de casino .....	41
3.4.5.2. Méthode élitiste ou par rang .....	41

3.4.5.3. Sélection par tournoi .....	42
3.5. Reproduction.....	42
3.5.1. Croisement .....	43
3.5.1.1. Croisement binaire .....	43
3.5.1.2. Croisement en un point .....	43
3.5.1.3. Croisement en deux points.....	44
3.5.1.4. Croisement réel .....	44
3.5.1.5. Croisement arithmétique.....	44
3.5.2. Mutation.....	45
3.5.2.1. Mutation binaire.....	47
3.5.2.2. Mutation réelle .....	47
3.5.2.3. Mutation uniforme .....	47
3.5.2.4. Mutation non uniforme .....	47
3.6. Remplacement .....	48
3.6.1. Remplacement stationnaire.....	48
3.6.2. Remplacement élitiste.....	48
3.7. Critère d'arrêt.....	49
3.8. Les paramètres d'un AG .....	49
3.8.1. La taille de la population .....	49
3.8.2. Le taux de croisement .....	50
3.8.3. Le taux de mutation .....	50
3.8.4. Le fosse des generations .....	50
3.9. Conclusion .....	50

## **Chapitre 4 : Utilisation des Algorithmes Génétiques dans la Modélisation du locuteur**

4.1. Introduction.....	51
4.2. Description de corpus utilisé et paramétrisation.....	51
4.3. Modèle de mélange de gaussiennes.....	55
4.3.1. Estimation des paramètres par l’algorithme EM .....	55
4.3.1.1. Initialisation des paramètres .....	55
4.3.1.2. Phase d’estimation .....	56
4.3.1.3. Phase de maximisation.....	56
4.3.1.4. Arrêt de l’algorithme .....	56
4.4. Optimisation des paramètres par AG.....	57
4.4.1. Mécanisme de l’encodage.....	57
4.4.2. Initialisation de la population .....	58
4.4.3. Fonction d’évaluation .....	59
4.5. Les opérateurs génétiques.....	59
4.5.1. Le croisement.....	59
4.5.2. La mutation .....	60
4.5.3. Le remplacement.....	60
4.6. Critère d’arrêt de l’AG.....	61
4.7. Conclusion .....	67
<b>Conclusions générales et perspectives .....</b>	<b>68</b>
<b>Références bibliographiques.....</b>	<b>70</b>

## Liste des figures

<b>Figure 1. 1</b> : Modèle de production de la parole.....	4
<b>Figure 1. 2</b> : Pré-traitement et extraction des paramètres .....	5
<b>Figure 1. 3</b> : Implémentation de bancs de filtres triangulaires selon l'échelle Mel avec des canaux répartis entre 0 et 4500 Hz. ....	6
<b>Figure 1. 4</b> : Schéma d'extraction des coefficients MFCC .....	8
<b>Figure 1. 5</b> : Représentation du canal buccal.....	9
<b>Figure 1. 6</b> : Les différentes étapes de l'analyse cepstrale du signal. ....	11
<b>Figure 1. 7</b> : Principe Général de système automatique du locuteur RAL.....	15
<b>Figure 2. 1</b> : Système de vérification du locuteur.....	23
<b>Figure 2. 2</b> : Les modèles probabilistes : exemples d'une distribution Gaussienne à 2 dimensions.....	25
<b>Figure 2. 3</b> : Les modèles probabilistes : exemple d'un modèle Gaussien à 2 dimensions. ....	25
<b>Figure 2. 4</b> : Evaluation des taux FA et FR.....	31
<b>Figure 2. 5</b> : Exemple de représentation des performances d'un système de vérification d'identité par une courbe DET .....	32
<b>Figure 3. 1</b> : Principe général des AG .....	36
<b>Figure 3. 2</b> : Les cinq niveaux d'organisation d'un AG.....	41
<b>Figure 3. 3</b> : Chaque gène est codé par un entier long (32 bits) .....	40
<b>Figure 3. 4</b> : Représentation schématique du croisement en 1 point. ....	43
<b>Figure 3. 5</b> : Représentation schématique du croisement en 2 points .....	44
<b>Figure 3. 6</b> : Représentation schématique d'une mutation dans un chromosome. ....	45
<b>Figure 3. 7</b> : Principe de l'auto-adaptation .....	46
<b>Figure 4. 1</b> : Forme d'onde et schématisation des coefficients MFCC, l'énergie et leurs dérivées premières du mot [wʔzʔna] prononcé par locuteur Fe1 .....	53
<b>Figure 4. 2</b> : Forme d'onde, spectrogramme à bande large du mot [wʔzʔna] prononcé par locuteur Fe1 et schématisation des coefficients MFCC, l'énergie et leurs dérivées premières.....	53

<b>Figure 4.3</b> : Forme d'onde et schématisation des coefficients MFCC, l'énergie et leurs dérivées premières du mot [wʔzʔna] prononcé par locuteur H1 .....	54
<b>Figure 4.4</b> : Forme d'onde, spectrogramme à bande large du mot [wʔzʔna] prononcé par locuteur F1 et schématisation des coefficients MFCC, l'énergie et leurs dérivées premières .....	54
<b>Figure 4.5</b> : Représentation du chromosome dans l'apprentissage AG-ML .....	57
<b>Figure 4.6</b> : Organigramme de l'AG .....	61
<b>Figure 4.7</b> : Comparaison entre l'apprentissage par AG-ML et ML avec $k=32$ .....	62
<b>Figure 4.8</b> : Convergence de l'AG vers l'individu qui représente l'optimum global $K=16$ .	63
<b>Figure 4.9</b> : Convergence de l'AG vers l'individu qui représente l'optimum global $K=32$ .	64
<b>Figure 4.10</b> : Convergence de l'AG vers l'individu qui représente l'optimum global $K=64$ .	65
<b>Figure 4.11</b> : Meilleur, mauvais chromosome et la moyenne des individus de chaque génération. ....	66
<b>Figure 4.12</b> : Critère d'arrêt de l'AG selon le nombre maximum de générations.....	66

## Liste des Tableaux

<b>Tableau 1. 1 :</b> Ambiguïté causée par l'absence de voyelles pour les unités lexicales <i>مدرسة</i> et <i>كتب</i> .	13
<b>Tableau 3. 1 :</b> Analogie de la terminologie génétique avec la nature. ....	37
<b>Tableau 4. 1 :</b> Le corpus utilisé dans la tâche de VAL.....	51
<b>Tableau 4. 2 :</b> Les paramètres du contrôle de l'apprentissage AG-ML .....	60
<b>Tableau 4. 3 :</b> Les valeurs finales de la fonction $f v_n$ .....	67
<b>Tableau 4. 4 :</b> Taux de la phase de vérification .....	67

# **Introduction Générale**

## **Introduction générale**

La Reconnaissance Automatique du Locuteur RAL est un domaine du Traitement Automatique de la Parole dont l'objectif est de déterminer l'identité d'un locuteur en analysant sa voix. Elle est largement utilisée dans plusieurs domaines par exemples, la sécurisation d'accès à des sites protégés, pour faire des transactions bancaires, ou pour répondre aux besoins d'un tribunal pour un crime ou un litige civil.

Ce travail s'inscrit dans le domaine du TAP et plus particulièrement dans la RAL, qui a plusieurs axes de recherche : indexation automatique du locuteur, IAL,..., VAL. Nous avons mis l'accent sur cette dernière.

L'utilisation d'informations dynamiques est considérée comme une source potentielle pour caractériser le locuteur, La RAL permet d'extraire, à partir des signaux de parole, des informations relatives à l'identité, et d'estimer avec ces dernières un modèle du locuteur suffisamment robuste afin de permettre son identité.

Un système de VAL permet de décider à partir d'un signal de parole, appelé segment de test, et une identité proclamée si le signal provient de l'identité ciblée ou non.

La majorité des systèmes de VAL actuels sont basés sur les modèles de mélange de Gaussiennes GMM en utilisant la méthode ML. Cette méthode est largement utilisée à cause de sa simplicité et sa traçabilité mathématique cependant, elle conduit généralement à un modèle sous optimal pour tous les modèles initiaux arbitraires choisis. Pour pallier à cette limitation, nous expliquons dans ce mémoire la nouvelle méthode d'apprentissage d'un locuteur arabophone pour vérifier son identité basée sur les GMM en appliquant un algorithme hybride AG-ML.

Les AG appartiennent à la famille des algorithmes évolutionnistes. Leur but est d'obtenir une solution approchée à un problème d'optimisation.

L'utilisation de la théorie de l'évolution comme modèle informatique pour trouver une solution optimale, peut se justifier par le fait que la théorie de l'évolution permet de chercher une solution parmi un très grand nombre de possibilités dans un laps temps raisonnable.

Notre mémoire est composée de 4 chapitres :

- le premier est consacré à la présentation des principes du Traitement Automatique de la Parole (TAP) et celui de l'Arabe Standard (AS). Une étude sur le système de RAL est ses principales tâches ;
- dans le deuxième, nous avons expliqué brièvement les différentes méthodes de la modélisation du locuteur, et en détaillant les techniques adoptant une modélisation générative du locuteur à base de modèles de GMM, et les principales étapes du système du VAL ;
- le troisième nous avons présenté les AG, le principe et les étapes de leur fonctionnement, les différents opérateurs de la phase de reproduction ainsi que leur rôle.
- le quatrième est consacré à l'utilisation des AG dans les modèles de GMM. Nous avons commencé par une population de solutions potentielles (chromosomes) qui a été choisie par l'algorithme k-means ensuite nous avons évalué leur performance relative (fitness).

Sur la base de ces performances nous avons créé une nouvelle population de solutions en utilisant des opérateurs évolutionnaires simples : la sélection, le croisement et la mutation pour aboutir à un meilleur modèle.

# **Chapitre 1 :**

## **Reconnaissance Automatique**

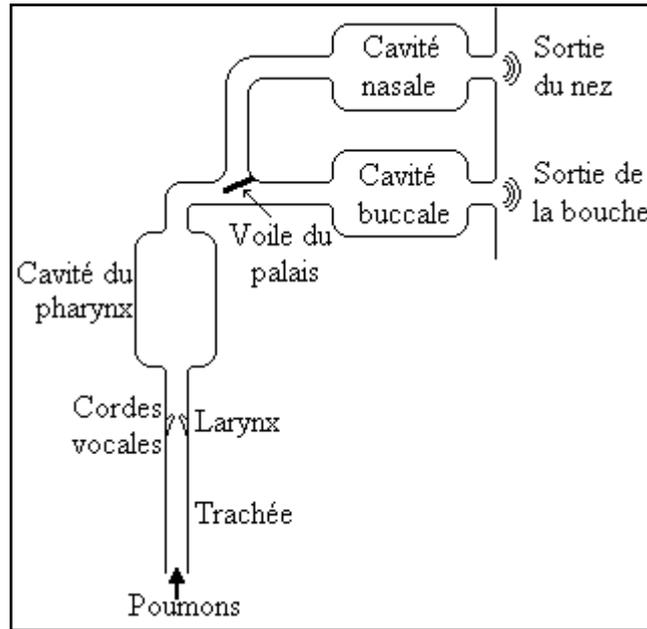
### **du Locuteur**

## 1.1. Introduction

La reconnaissance Automatique du Locuteur (RAL) s'inscrit dans le domaine plus général de la Communication Homme-Machine (CHM) s'agit de reconnaître automatiquement l'identité d'une personne prononçant une ou plusieurs phrases, comme un auditeur humain identifie son interlocuteur au cours d'une conversation. Quelques tâches voisines de la RAL sont l'adaptation au locuteur pour la Reconnaissance Automatique de la Parole et la détection de changement de locuteurs au cours d'un dialogue. Les applications directes de la RAL concernent les problèmes de confidentialité et d'authentification.

## 1.2. Production et Analyse de parole

Lors de la production de la parole, Le diaphragme expulse l'air des poumons, produisant ainsi l'énergie nécessaire à la parole. L'air arrive via la trachée- artère dans le larynx où se trouvent les cordes vocales. L'air arrive alors dans le conduit vocal proprement dit, avec d'abord la cavité pharyngienne, puis en parallèle la cavité buccale et la cavité nasale. Cette dernière peut être utilisée ou non grâce au voile du palais qui permet une isolation totale de la cavité nasale (Figure 1.1) [1]. Si le son à produire est non voisé ou sourds, les lèvres symétriques constituant les cordes vocales forment une grande ouverture triangulaire nommée glotte et l'air continue son chemin. Si le son est voisé ou sonores, les cordes vibrent périodiquement (ouverture brusque, fermeture plus progressive) ce qui module le son en impulsions périodiques de pression dont la fréquence est nommée fondamentale ou pitch (de 80 à 600 Hz) [1].



**Figure 1. 1 :** Modèle de production de la parole [1]

L'analyse acoustique du signal de parole consiste à extraire l'information pertinente et à réduire au maximum la redondance. Généralement, on calcule un jeu de coefficients acoustiques à des intervalles de temps réguliers, sur des blocs de signal de longueur fixe. Ce jeu de coefficients constitue un vecteur acoustique. Les techniques de paramétrisation acoustique sont nombreuses, néanmoins, on peut les regrouper en trois grandes familles, analyse par :

- bancs de filtres ;
- transformée de Fourier ;
- prédiction linéaire.

### 1.2.1. Pré-traitement acoustique

Un système de reconnaissance de la parole est capable de transcrire une voix humaine en informations numériques, compréhensibles et reconnaissables par l'ordinateur.

### 1.2.2. La pré-accentuation

L'onde acoustique sortante des lèvres subit, à cause de la désaccentuation de 6 dB octave. Pour pouvoir compenser cette distorsion, et accentuer les hautes fréquences, on applique un filtre de pré-accentuation passe haut de transmittance :

$$H(Z) = 1 - \alpha Z^{-1} \quad \text{avec} \quad 0.9 \leq \alpha \leq 1 \quad (1.01)$$

### 1.2.3. Le fenêtrage

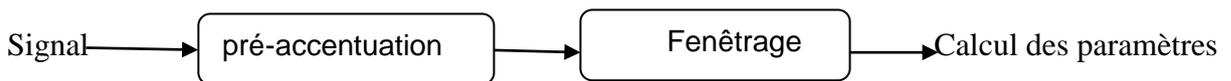
L'étape de fenêtrage consiste à appliquer au signal vocal une fenêtre glissante de durée limitée, et ce afin de limiter le nombre d'échantillons et de réduire les effets de bords (phénomène de Gibbs).

Parmi les différentes fenêtres de pondération, les plus utilisées sont : la fenêtre rectangulaire, Hamming , Hanning et de Blackmann .

La fenêtre de Hamming est donnée par l'expression :

$$w(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{N-1}\right) , \quad 0 \leq n \leq N - 1 \quad (1.02)$$

N : Le nombre d'échantillons dans une fenêtre.



**Figure 1. 2 :** Pré-traitement et extraction des paramètres

## 1.3 Extraction de paramètres

En reconnaissance du locuteur, nous voudrions plutôt extraire les caractéristiques plus spécifiques à l'identité du locuteur. Pour modéliser le signal de parole, un découpage du signal en trames de taille fixe environ 25 ms, prises toutes les 10 ms est effectué. De chaque trame est extrait un vecteur de paramètres caractérisant celle-ci.

Les méthodes d'analyse du signal de la parole couramment utilisées aujourd'hui se divisent en deux groupes: les méthodes basées sur une modélisation de la production de la parole (LPC, LPCC) ainsi que les méthodes modélisant le système auditif humain (PLP, MFCC) [2].

### 1.3.1 Analyse cepstrale

L'analyse cepstrale repose sur une analyse à court terme du signal par banc de filtres triangulaires qui sont linéaires espacés de 0 à 1 kHz, puis non-linéaire placé selon les approximations de l'échelle Mel.

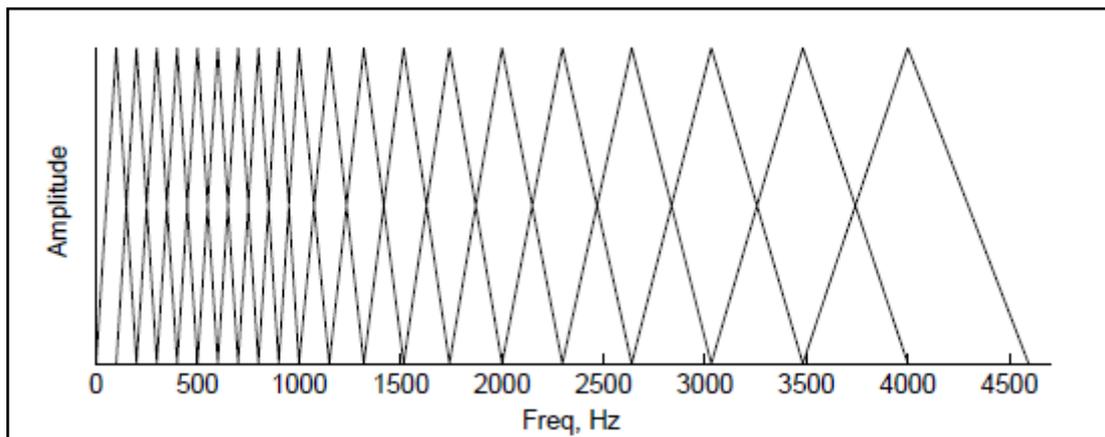
La formule qui établit la relation entre l'échelle Mel et la fréquence en Hz est la suivante:

$$f_{Mel} = 2595 \log \left( 1 + \frac{f_{Hz}}{7000} \right) \quad (1.03)$$

Ce processus d'extraction de caractéristiques comporte quatre étapes :

- Calculer le spectre en puissance de la trame analysée ;
- Calculer l'énergie correspondant à chaque filtre du banc ;
- Appliquer un logarithme aux coefficients obtenus ;
- Appliquer une transformée en cosinus discret (DCT) [3].

La répartition des fréquences des filtres est différente selon les échelles choisies soient linéaires ou logarithmiques. Il existe plusieurs implémentations de bancs de filtres comme l'échelle Mel (Figure 1.3).



**Figure 1.3 :** Implémentation de bancs de filtres triangulaires selon l'échelle Mel avec des canaux répartis entre 0 et 4500 Hz [4]

Le signal de la parole peut être représenté par la convolution de la source (cordes vocales) et du filtre (canal buccal) dans le domaine temporel.

$$s(t) = e(t) \otimes h(t) \quad (1.04)$$

On passe dans le domaine fréquentielle pour obtenir l'enveloppe spectrale qui permet de faire apparaître les différences de fréquences. La convolution devient donc une multiplication.

$$S(f) = E(f) \cdot H(f) \quad (1.05)$$

On souhaite séparer la source du filtre pour récupérer l'enveloppe spectrale du signal. Pour cela, on utilise la fonction log :

$$\log([S(f)]) = \log([E(f)]) + \log([H(f)]) \quad (1.06)$$

Pour obtenir les coefficients temporels appelés coefficients cepstraux MFCC (Mel-frequency Cepstral Coefficients), le module du logarithme de l'énergie de sortie du banc de filtres triangulaires repartis le long de l'échelle de Mel noté  $X_k$  est calculé avant que la Transformée en Cosinus Discrète (TCD) ne soit appliquée. Elle transforme les vecteurs paramétriques dans l'espace cepstral (Figure.1.4). Les coefficients cepstraux notée  $c_n$  sont calculés comme suit :

$$c_n = \sum_{k=1}^N X_k \cos\left(n \frac{\Pi(k-0.5)}{N}\right) \quad \forall \quad 0 \leq n \leq \frac{N}{2} \quad (1.07)$$

Il en résulte le vecteur à coefficients MFCC noté  $v_k$  tel que :

$$v_k = (c_1, c_2, \dots, c_M) \quad (1.08)$$

Les premiers MFCC donnent les paramètres de l'enveloppe spectrale, les coefficients les plus élevés fournissent les variations de l'excitation.

Cette analyse se base sur le système de perception humaine. Leur étagement en fréquence limite la répartition et la forme des filtres de la cochlée. Plusieurs jeu de paramètres peuvent être extraits selon qu'on ajoute l'énergie, les dérivés premières (appelés  $\Delta$  par la suite), ou seconde  $\Delta \Delta \dots$  (Figure 1.4)[5].

Il est très important d'avoir une information temporelle concernant les coefficients dérivés de la prédiction linéaire ou de l'analyse cepstrale. Pour cela, on peut calculer les coefficients de vitesse et les coefficients d'accélération qui correspondent respectivement aux première et deuxième dérivées des coefficients acoustiques. Une manière directe d'avoir la première dérivée des coefficients acoustiques est définie comme suit :

$$\Delta c_h(n; m) = c_h(n; m + 1) - c_h(n; m) \quad (1.09)$$

La deuxième dérivée peut être obtenue en appliquant l'équation (1.09) deux fois sur les paramètres  $c_h(n; m)$ .

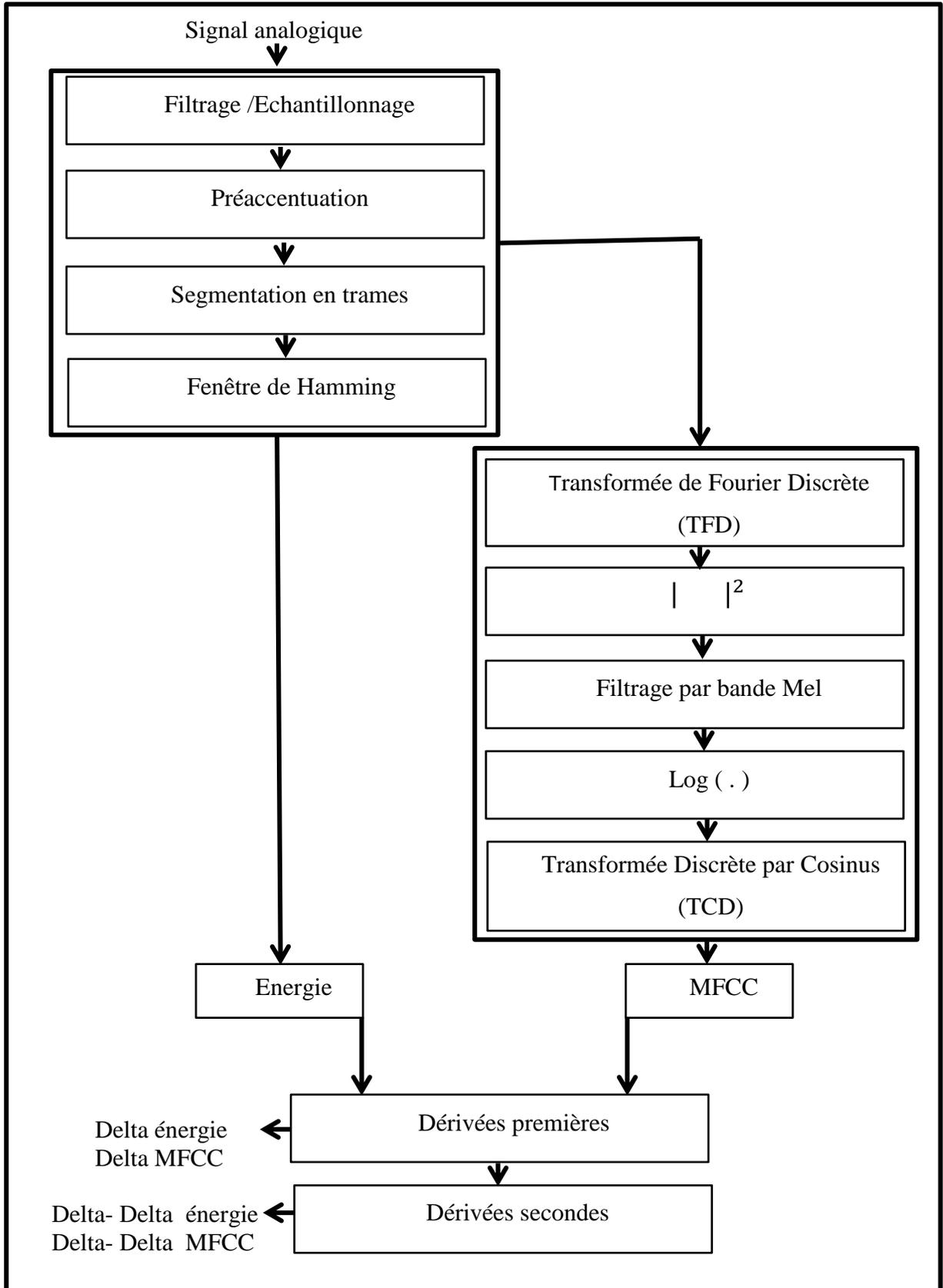


Figure 1.4 : Schéma d'extraction des coefficients MFCC[6]

### 1.3.2. Analyse par prédiction linéaire

C'est un modèle basé sur les corrélations entre les échantillons successifs du signal vocal. Cette méthode se base sur l'hypothèse que le canal buccal est constitué d'un tube cylindrique de section variable. La LPC (Linear Predictive Coding ) pour un ordre  $p$  se définit de la manière suivante :

$$\hat{s}(n) = \sum_{i=1}^p a_i s(n-i) \quad (1.10)$$

On considère que le signal de la parole à l'instant  $n$  peut être représenté par une combinaison linéaire des  $p$  échantillons précédents. Les coefficients de prédiction  $a_i$  sont supposés constants sur une fenêtre d'analyse. On introduit le terme d'excitation unitaire  $u(n)$  et un gain  $G$ .

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (1.11)$$

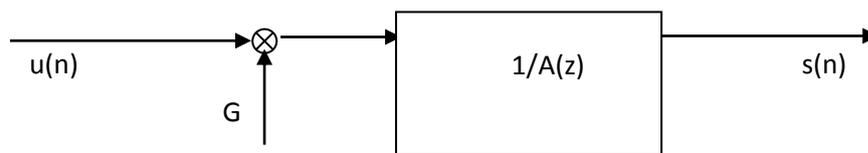
On réalise la transformée en  $z$  de l'expression, on obtient :

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z) \quad (1.12)$$

La fonction de transfert est donc la suivante :

$$H(z) = \frac{1}{G} \times \frac{S(z)}{U(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (1.13)$$

Le système peut être rapproché au modèle acoustique linéaire de production de parole. La fonction  $u(n)$  est soit un train d'impulsions quasi périodiques pour les sons voisés (produit par les cordes vocales) ou une source de bruit aléatoire pour les sons non voisés (Figure 1.5).



**Figure 1. 4:** Représentation du canal buccal

Pour définir le signal  $s(n)$ , on définit l'erreur de prédiction :

$$e(n) = s(n) - \hat{s}(n) = Gu(n) \quad (1.14)$$

Pour déterminer les coefficients  $a_i$ , on utilise la méthode des moindres carrés sur une fenêtre de temps de longueur  $m$  :

$$E_m = \sum_m e^2(m) = \sum_m [s(m) - \sum_{i=1}^p a_i s(m-i)]^2 \quad (1.15)$$

On cherche à minimiser  $E_m$ , deux méthodes peuvent être réalisées pour résoudre le système d'équation :

- la méthode de covariance ;
- la méthode d'autocorrélation.

Une résolution rapide des modèles Autorégressifs (AR) est donnée par l'algorithme de Levinson et de Schur.

Les coefficients en sortie des bancs de filtres ou les coefficients  $a_i$  peuvent être utilisés pour mesurer des différences entre deux trames comme dans l'algorithme Dynamic Time Warping (DTW). Ils présentent cependant des inconvénients comme il dépend de l'énergie du signal ou de l'excitation. La transformation cepstrale permet d'obtenir une information normalisée.

Les coefficients cepstraux LPCC (Linear Prediction Cepstral Coefficients)  $c_m$  peuvent être obtenus à partir des coefficients  $a_p$  de la LPC :

$$c_0 = \ln G \quad (1.16)$$

$$c_m = a_m + \sum_{j=1}^{m-1} \binom{j}{m} c_j a_{m-j}, \quad 1 \leq m \leq p \quad (1.17)$$

$$c_m = \sum_{j=1}^{m-1} \binom{j}{m} c_j a_{m-j}, \quad m > p \quad (1.18)$$

L'énergie du signal est un indice qui peut contribuer à augmenter les performances d'un système de reconnaissance, cette énergie correspond à la puissance du signal. Elle est souvent

évaluée sur plusieurs trames de signal successives pour pouvoir mettre en évidence ses variations. La formule de calcul de ce paramètre est :

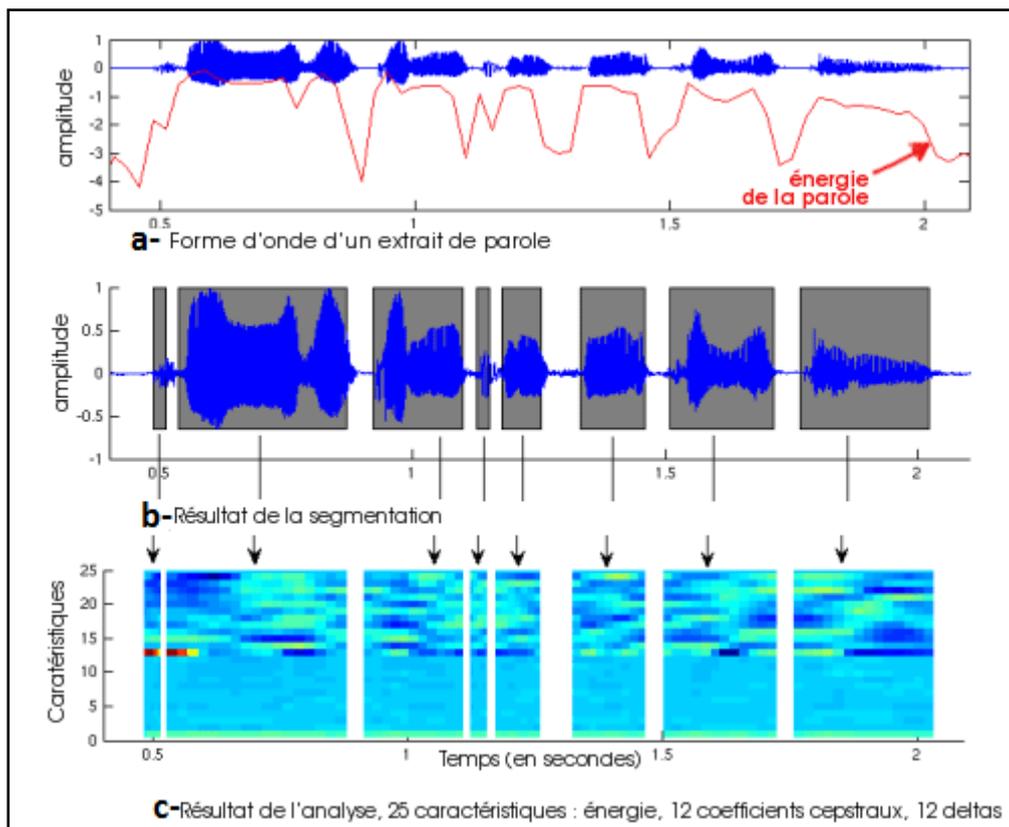
$$E = (\sum_{n=0}^{N-1} s^2(n)) \quad (1.19)$$

Comme paramètre acoustique, on peut aussi utiliser l'énergie logarithmique qui est définie comme suit :

$$E = \ln(\sum_{n=0}^{N-1} s^2(n)) \quad (1.20)$$

Où  $N$  est le nombre d'échantillons du signal, et les  $s(n)$  sont les échantillons du signal.

L'énergie ainsi obtenue est sensible au niveau des enregistrements. Généralement elle est normalisée, exprimée en décibels (par rapport à un niveau de référence).



**Figure 1. 5:** Les différentes étapes de l'analyse cepstrale du signal [7]

Sur Figure 1.6- a, on observe un exemple de signal de parole (courbe bleue) et la valeur de son énergie à chaque instant (courbe rouge).

La Figure 1.6- b montre les parties les plus énergétiques qui sont gardées pour l'analyse (partie grisées).

La Figure 1.6- c illustre le résultat de la transformation cepstrale. En plus de la valeur de l'énergie, on calcule 12 coefficients cepstraux et 12 dérivées de ces coefficients, ce qui constitue les 25 lignes représentées sur la figure. Les amplitudes de ces coefficients sont représentées par les variations de couleur, et chaque ligne horizontale représente l'évolution des valeurs d'un coefficient dans le temps.

## 1.4. Description des sons de l'Arabe Standard

La langue arabe est une langue du Coran, sémitique qui s'écrit de droite à gauche et dont l'alphabet est un abjad. L'écriture arabe comprend essentiellement les consonnes, et également les voyelles longues "أ", "ي" et "و". Les deux dernières sont des réalisations contextuelles des glides "ي" et "و". L'écriture arabe comporte également des voyelles qui ne sont pas essentielles à l'écriture ainsi qu'un certain nombre de signes annexes dont l'emploi est facultatif hormis pour le Coran servant à noter les trois voyelles brèves [a], [u] et [i]. Il existe, de plus, une série d'autres diacritiques de syllabation dont les plus courants sont l'indication de l'absence de voyelle "◌ْ" [suku:n] et la gémination des consonnes "◌ّ" [fada].

Notant aussi que si un mot arabe est indéfini (sans article ni complément de nom), il prend (sauf exception) les désinences, "◌ان" (an), "◌ان" (un) et "◌ان" (in), nommée nounation ou tanwin. Celles-ci sont notées par diacritiques spéciaux marqués par le redoublement du signe de la voyelle qui précède le suffixe (n) attendu en fin de mot [8].

### 1.4.1. Particularités des sons de l'Arabe Standard

Les particularités de la langue arabe que constituent par exemple, l'emphase, la gémination et l'allongement pertinent des phonèmes posent de nombreux problèmes aux constricteurs des systèmes de RAP pour cette langue [9].

L'emphase est un trait phonétique caractérisant les consonnes arabe, La gémiation correspond au dédoublement de deux consonnes identiques en une seule dite gémifiée. C'est aussi la production d'une consonne avec une concentration d'énergie très intense. Toutes les consonnes arabes sont susceptibles d'être gémifiées sauf la glottale hamza [6].

Les voyelles sont nécessaires à la lecture et à la compréhension correcte d'un texte et elles permettent de différencier des unités lexicales ayant la même représentation.

Pour mieux comprendre prenons l'exemple du tableau 1.1. Le dictionnaire nous renvoie les voyellations lexicales suivantes :

- كَتَبَ, Il a écrit ;
- كُتِبَ, Il a été écrit ;
- كُتُب, Livres.

**Tableau1.1:** Ambiguïté causée par l'absence de voyelles pour les unités lexicales مدرسة et كتب

Unité Lexicale	Interpretation	Interpretation	Interpretation
كتب	كَتَبَ IL a écrit	كُتِبَ IL a été écrit	كُتُب Livres
مدرسة	مَدْرَسَة Ecole	مُدْرِسَة Enseignante	مُدْرِسَة Enseignée

#### 1.4.2. Problèmes de la langue arabe en traitement automatique

La langue arabe rencontre deux principaux problèmes en traitement automatique : le premier, général, concerne l'agglutination des mots ; le second, spécifique, a trait à l'absence de voyelles à l'écrit.

##### 1.4.2.1. Agglutination des mots

La plupart des mots en AS sont composés par agglutinations d'éléments lexicaux de base (proclitique + base + enclitique). Par exemple, la détermination peut s'exprimer par agglutination de l'article ال [ʔal] avant le mot (الولد, [alwaladu], "l'enfant") ou par agglutination d'un pronom personnel après celui-ci (وَلَدُهُ, [waladuhu], "son enfant"). De même, les pronoms personnels peuvent se rattacher aux verbes (ضَرَبَهُ, [ḍarabahu], "il l'a frappé"), les particules régissant le cas indirect aux noms (كَدَارِهِ, [kadaarihi], "comme sa maison") et les conjonctions de coordination aux verbes (فَدَاهَبَ, [faḍahaba], "et il est parti"), etc.

Dans toute perspective de traitement automatique, le problème est donc de décomposer le mot en ces différentes parties. Cette décomposition nécessite des connaissances de niveau supérieur en cas d'ambiguïtés.

#### **1.4.2.2. Voyellation**

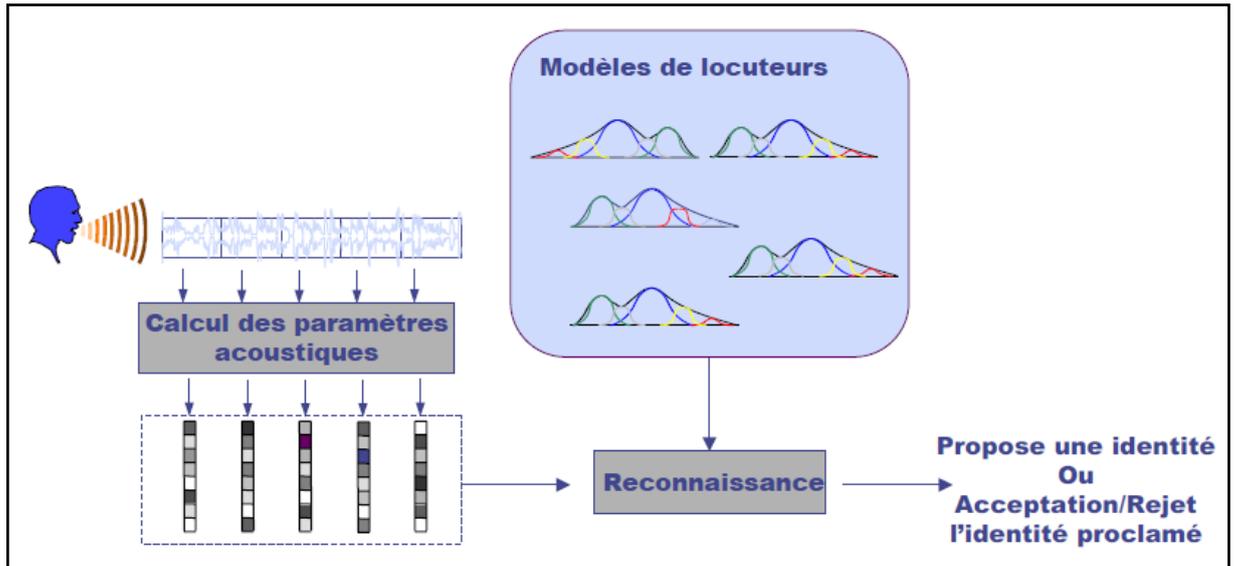
Les textes en AS sont ordinairement dépourvus de diacritiques. Pour les lire, tout un processus mental est nécessaire : identifier le mot comme appartenant au lexique puis lui attribuer ses voyelles dans son contexte, ce qui nécessite la compréhension du texte.

Pour le TA d'un texte, il est indispensable d'introduire les voyelles avant le traitement dans le cas d'une synthèse TTS ou après dans le cas d'une Reconnaissance Optique (RO). Cette opération, appelée voyellation ou vocalisation automatique, est effectuée par la machine et se déroule généralement en deux étapes, sous forme d'une analyse :

- morphologique qui va assigner à chaque mot non-voyellé l'ensemble des mots voyellés correspondants. Ce qui nécessite la présence d'un lexique total avec toutes les formes canoniques et fléchies des mots ;
- syntaxique pour réduire l'ambiguïté au vu du contexte grammatical ;
- sémantique qui est nécessaire pour réduire l'ambiguïté au vu du sens de la phrase [10].

### **1.5. Système de reconnaissance automatique de locuteurs**

La reconnaissance automatique du locuteur peut être comme une tâche particulière de reconnaissance de formes. Différents modules sont présents dans ce système. Tout d'abord, le message vocal, capté par un microphone, est converti en signal numérique. Il est ensuite analysé dans un étage d'analyse acoustique. A l'issue de cette étape, le signal est représenté par des vecteurs de coefficients pertinents pour la modélisation du locuteur. Dans l'étape d'apprentissage, on crée un modèle du locuteur. A la reconnaissance, un module de classification va mesurer la similarité entre les paramètres acoustiques du signal prononcé et les modèles de locuteurs présents dans la Base de Données(BD). En dernier lieu, un module de décision, basé sur une stratégie de décision donnée, fournit la réponse du système. On peut également introduire un module d'adaptation pour augmenter les performances du système de reconnaissance [11]. (Figure 1.7).



**Figure 1. 6:** Principe Général de système automatique du locuteur RAL [12]

## 1.6. Variabilité

La variabilité d'une personne à une autre (variabilité inter-locuteurs) démontre les différences du signal de parole en fonction du locuteur. Cette variabilité, utile pour différencier les locuteurs, est également mélangée à d'autres types de variabilité : variabilité intralocuteur, variabilité due aux conditions d'enregistrement et de transmission du signal de parole [13].

### 1.6.1. Conditions d'enregistrement

Le matériel d'enregistrement (en particulier les microphones) a des spécificités techniques propres qui influencent le signal. Il a été montré que ces conditions d'enregistrement ont une incidence sur les performances des systèmes de RAL.

Généralement, chaque locuteur utilise une combine différente. Les spécificités des microphones et des canaux de transmission facilitent la segmentation en locuteurs.

### 1.6.2. Variabilité intra-locuteur

La variabilité intra locuteur se présente lorsqu'un même locuteur prononce la même phrase ou le même mot à plusieurs reprises (parfois à différentes sessions). Plusieurs sources peuvent

être attribuées à cette variabilité telle que la fatigue, le stress, le sommeil, l'horaire de la journée (matin, soir), le débit d'élocution, l'état émotionnel, etc [14].

### **1.6.3. Variabilité inter-locuteur**

On définit par variabilité interlocuteur les caractéristiques qui sont propres à chaque locuteur et qu'on ne retrouve pas chez d'autres locuteurs (variabilité induite par le changement du locuteur). La grande variabilité entre les locuteurs est due, d'une part aux variations anatomiques des organes responsables de la production vocale et d'autre part à l'héritage linguistique et au milieu socioculturel de l'individu [14].

## **1.7. Applications et tâches pour la RAL**

La Reconnaissance Automatique du Locuteur consiste à reconnaître l'identité d'un individu à partir de sa voix. Elle est fondée sur la variabilité entre locuteurs et a pour tâche d'extraire du signal de parole les informations de nature à renseigner sur la spécificité d'un individu : identité, caractéristiques physiques, émotivité, état pathologique ou particularités régionales.

Les deux tâches pionnières des systèmes de RAL sont l'Identification Automatique du Locuteur (IAL) et la Vérification Automatique du Locuteur (VAL) [15].

Dans la suite de ce paragraphe, nous présentons brièvement les principales tâches associées à la RAL.

### **1.7.1. Identification du locuteur**

L'identification du locuteur est le processus qui consiste à déterminer, parmi une population de locuteurs connus, la personne ayant prononcé un message donné [15]. En identification, la réponse apportée n'est plus de type binaire (acceptation ou rejet) comme dans le cas de la vérification puisqu'il est nécessaire de désigner un locuteur parmi un groupe. Deux modes sont distingués : le fonctionnement en milieu fermé et le fonctionnement en milieu ouvert. En milieu fermé le locuteur est supposé faire partie de la population connue. Le système retourne l'identité du locuteur le plus probable parmi la population. En milieu ouvert le locuteur peut ne pas être connu du système. Dans ce cas le système associe au locuteur le

plus probable un indice de fiabilité. Le locuteur le plus probable peut aussi être le locuteur « inconnu ».

### **1.7.2. Vérification Automatique du Locuteur**

La Vérification Automatique du Locuteur (VAL) consiste à vérifier l'identité proclamée par un individu par la comparaison d'un signal vocal et d'un modèle de référence du locuteur présumé, préalablement appris par le système. Un système de VAL a donc deux entrées : une identité et un accès de test. Le résultat de cette comparaison est considéré comme une mesure de similarité avant d'être comparé à un seuil d'acceptation. Lorsque la mesure de similarité est supérieure à ce seuil, l'individu est accepté, il est rejeté dans le cas contraire [16].

## **1.8. Mode dépendant et indépendant du texte**

On parle de reconnaissance vocale en mode dépendant du texte lorsque le texte prononcé par le locuteur est fixé et connu à l'avance. A l'opposé, lorsque le texte prononcé par le locuteur n'est pas connu a priori, on parle de mode indépendant du texte. Mais cette terminologie ne rend pas bien compte des différentes dépendances au texte possible. Les différents systèmes peuvent être classés, selon le degré croissant d'indépendance au texte, de la façon suivante :

- **Système à texte fixé dépendant du locuteur**

Pour un locuteur donné, le texte est toujours le même d'une session à l'autre. Mais chaque locuteur a un texte différent.

- **Système dépendant du vocabulaire**

L'utilisateur du système prononce une séquence de mots, issus d'un vocabulaire limité (des séquences de chiffres par exemple), mais dont l'ordre peut varier d'une session à l'autre.

- **Système dépendant d'événements phonétiques**

Le vocabulaire n'est pas directement imposé, mais certains événements phonétiques doivent être présents dans la séquence de parole prononcée par exemple : présence de certaines voyelles nasales). Les phrases à prononcer peuvent éventuellement être affichées sur l'écran à chaque session.

- **Système à texte imposé par la machine**

Le texte est différent pour chaque session et pour chaque locuteur, mais affiché à chaque fois par la machine. Le texte est choisi de manière imprédictible pour éviter l'utilisation d'enregistrements par un imposteur.

- **Système indépendant du texte**

Le locuteur est entièrement libre de ce qu'il dit à chaque session.

## **1.9. Conclusion**

Dans ce chapitre, nous avons, présenté les mécanismes de production des sons en général et les différentes méthodes d'analyse acoustique du signal de parole les principales caractéristiques de système de reconnaissance automatique du locuteur et ses application et nous avons prés les différentes étapes du système .

# **Chapitre 2 :**

## **Vérification Automatique du Locuteur**

## 2.1. Introduction

La Vérification Automatique du Locuteur (VAL) est probablement la méthode la plus ergonomique pour résoudre les problèmes d'accès. Cependant, la voix ne peut être considérée comme une caractéristique biométrique d'une personne compte tenu de la variabilité intra-locuteur.

Dans ce chapitre, nous nous intéressons à la tâche de la VAL, et nous présentons que les différentes étapes du système.

## 2.2. Modélisation des locuteurs

Nous parcourons brièvement les techniques les plus couramment utilisées en reconnaissance du locuteur. Comme dans le cas de la reconnaissance de la parole, le problème de reconnaissance du locuteur peut se formuler selon un problème de classification. Différentes approches ont été développées, néanmoins on peut les classer en quatre grandes familles (approche vectorielle, approche statistique et approche connexionniste) [17].

### 2.2.1. L'approche vectorielle

Elle consiste à représenter un locuteur par un ensemble de vecteurs issus directement de la phase de paramétrisation. Cette approche comporte deux techniques principales: l'alignement temporel dynamique et la quantification vectorielle [17].

#### 2.2.1.1. L'Alignement Temporel Dynamique (DTW-Dynamic Time Warping)

L'alignement temporel dynamique est un modèle basé sur le calcul d'une distance entre deux vecteurs. Principalement, il fait la comparaison d'une séquence de  $M(X_1, \dots, X_M)$  vecteurs avec une autre séquence de  $N(X_1, \dots, X_N)$  vecteurs par le calcul de la distance accumulée entre ces deux séquences. Si les deux séquences sont identiques alors le chemin entre eux est diagonal, et par conséquent, la distance qui les sépare est minimale. Cette méthode est utilisée souvent dans les systèmes de reconnaissance automatique du locuteur dépendante de texte [18].

### **2.2.1.2. Quantification vectorielle**

Il s'agit de représenter l'espace acoustique par un nombre fini de vecteurs acoustiques. Cela consiste à faire un partitionnement de cet espace en régions, qui seront représentées par leur vecteur centroïde. Pour déterminer la distance d'un vecteur acoustique à cet espace, on effectue une mesure de distance avec chacun des centroïdes des régions et on retient la distance minimale. Si le vecteur acoustique provient du même locuteur pour lequel on a établi le dictionnaire de quantification, la distorsion sera en général moins grande que si ce vecteur provient d'un autre locuteur. Ainsi, on va représenter un locuteur par son dictionnaire de quantification [19].

### **2.2.2. L'approche statistique**

Une approche qui repose essentiellement sur des fondements mathématiques (probabilité & statistique). L'objet de cette approche est de décrire les formes à partir d'un modèle probabiliste simple à utiliser et de regrouper les formes dans des classes.

#### **2.2.2.1. Modèle de Markov Caché**

Les modèles de Markov (ou HMM pour Hidden Markov Models) ont été initialement introduits en reconnaissance de la parole, puis leur utilisation s'est étendue peu à peu au domaine de la reconnaissance du locuteur. Dans cette approche, il ne s'agit plus d'une mesure de distance d'une forme acoustique à une référence, mais de la probabilité que la forme acoustique a engendrée par le modèle de référence du locuteur. Le modèle d'un locuteur est constitué de l'association d'une chaîne de Markov, une succession d'états avec des probabilités de transition d'un état à l'autre, et des lois de probabilités (probabilités d'observation d'un vecteur acoustique dans un état) [11].

#### **2.2.2.2. Les Mélanges de Gaussiennes**

La reconnaissance du locuteur par Mélanges de Gaussiennes (ou GMM pour Gaussian Mixture Models) consiste à modéliser un locuteur par une somme pondérée de composantes gaussiennes. Ainsi une large gamme de distributions peut être parfaitement représentée. Chaque composante des gaussiennes est supposée modéliser un ensemble de classes acoustiques. L'utilisation de ce type de modèle semble être bien prometteuse. Il semble bien modéliser les caractéristiques spectrales des voix des locuteurs, et il est relativement simple à

mettre en œuvre. Les GMM sont considérés comme un cas particulier des HMM et une extension de la quantification vectorielle [11].

### 2.2.3. L'approche connexionniste

Les Réseaux de Neurones (RN) ont été assez largement utilisés en reconnaissance du locuteur. Ils offrent en effet une bonne alternative au problème de la discrimination entre les locuteurs. Ces outils de classification permettent de séparer des classes, dans un espace de représentation donnée, de façon non linéaire. L'inconvénient important de l'application de cette technique en identification du locuteur est le coût important lié à l'ajout d'un nouveau locuteur dans la base de référence (ce n'est pas le cas en vérification du locuteur). On peut aussi utiliser les RN en les couplant à d'autres techniques, comme par exemple les modèles de Markov cachés. On parle alors de méthodes hybrides [20].

### 2.2.4. L'approche relative

Cette nouvelle technique consiste à modéliser un locuteur non plus de façon absolue mais relativement à un ensemble de locuteurs bien appris [11].

L'avantage de ce type de modèle GMM se récapitule dans ce qui suit :

- L'utilisation d'un mélange de plusieurs densités gaussiennes multidimensionnelles a permis de donner une très bonne représentation des vecteurs acoustiques. Il est à noter que les vecteurs acoustiques suivent des densités de probabilité très complexes. La connaissance de la forme multidimensionnelle d'une densité gaussienne a simplifié beaucoup l'apprentissage de ce type de modèle ;
- l'utilisation du modèle GMM permet d'estimer fidèlement des densités de probabilités aléatoires telles que celle des vecteurs acoustiques ;
- le temps d'apprentissage est relativement petit par rapport à d'autres modèles tel que le modèle HMM [18].

## 2.3. Modules d'un système de vérification du locuteur

Comme le représente Figure 2.1, les systèmes de vérification du locuteur sont composés de quatre modules principaux interdépendants.

### 2.3.1. Le prétraitement

Il s'agit d'extraire du signal de parole "pur" (valeurs de l'amplitude échantillonnées à des fréquences de l'ordre de 8 KHz) des caractéristiques (paramètres numériques et/ou symboliques). Un bon prétraitement fournit des paramètres dépendants des variations interlocuteurs et peu sensibles aux variations extrinsèques à l'identité du locuteur (conditions d'enregistrement, variabilités intra locuteurs, etc.).

### 2.3.2. L'apprentissage

Il s'agit d'instancier des modèles à partir de paramètres extraits de locuteurs étiquetés ou non. L'apprentissage se fait souvent par des méthodes d'entraînement itératives.

### 2.3.3. L'attribution de scores

Ce module est étroitement lié à la façon dont ont été conçus et entraînés les modèles dans le module d'apprentissage. Alors que ce dernier s'applique à des séquences d'entraînement, l'attribution de scores s'applique à des séquences test. Notons que l'apprentissage peut tenir compte de plusieurs séquences train pour l'élaboration d'un modèle, alors que le module de scoring traite les séquences test indépendamment les unes des autres.

### 2.3.4. La prise de décision

Ce petit module vient directement après l'attribution de scores. Typiquement, il s'agit de comparer les scores à un seuil (fixé lors de la phase de développement) pour renvoyer une décision binaire. Un seuil pour la prise de décision joue un rôle très important dans la performance d'un système vérification du locuteur [21].

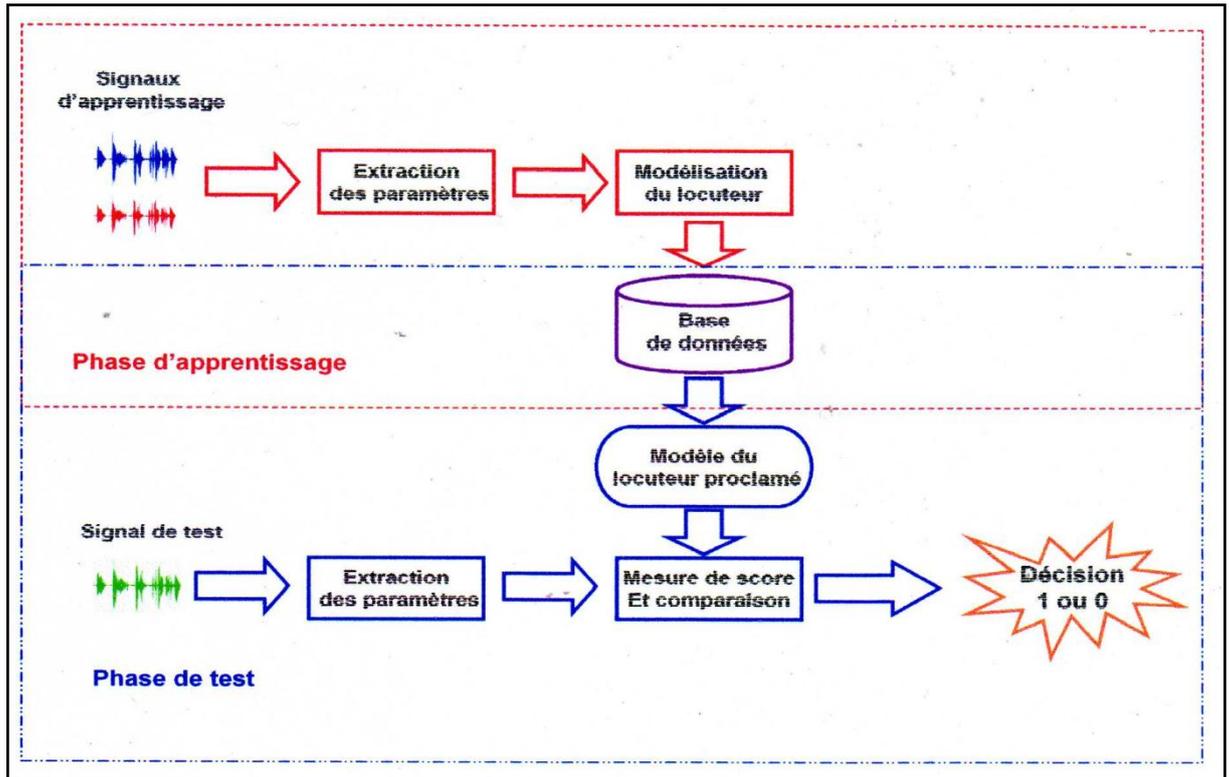


Figure 2. 1: Système de vérification du locuteur

## 2.4. Modélisation par les GMM

Un GMM est une fonction de densité de probabilité (de dimension variable) composée d'une somme de Gaussiennes utilisée pour approximer une fonction de densité de probabilité complexe en soulignant ses différents modes propres.

Cette technique, qui constitue l'état de l'art a été choisie pour sa flexibilité au type de signal et son bon compromis entre les performances du système en termes de précision et la vitesse et la complexité des algorithmes [21].

Pour obtenir une modélisation pertinente des caractéristiques d'un locuteur ce GMM est entraîné à partir des vecteurs issus du signal de parole de ce locuteur (on utilise par exemple les MFCC). S'il existe plusieurs techniques permettant de calculer les paramètres des GMM, la plus courante consiste à maximiser la vraisemblance en utilisant l'algorithme EM (Expectation- Maximization) couplé à une entité de maximisation de la vraisemblance (ML pour Maximum Likelihood). Les GMM consistent en la modélisation, pour chaque modèle,

des données  $x_t$  sous la forme d'une somme pondérée par les coefficients  $w_k$  de fonctions de densité de probabilité gaussiennes  $p(x_t, \mu_k, \Sigma_k)$  :

$$P(X/\lambda) = \sum_{k=1}^K w_k p(x_t, \mu_k, \Sigma_k) \quad (2.01)$$

Avec  $K$  est le nombre de composantes de densité considéré pour le modèle. Chaque composante s'exprime en fonction de sa moyenne  $\mu_k$  et de sa matrice de covariance  $\Sigma_k$  :

$$p(x_t, \mu_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp \left[ -\frac{1}{2} (x_t - \mu_k)' \Sigma_k^{-1} (x_t - \mu_k) \right] \quad (2.02)$$

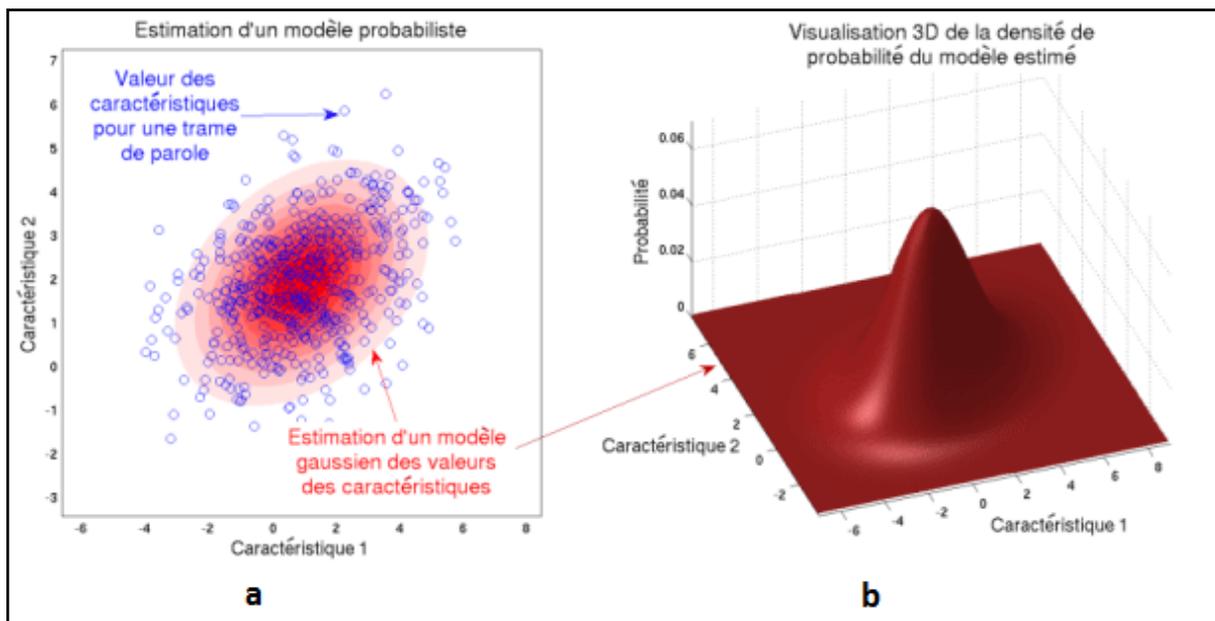
La matrice de covariance utilisée est diagonale, c-à-dire les modèles sont appris en considérant les observations associées à chacun des descripteurs de manière indépendante. Pour chaque modèle, chacune des composantes du mélange modélise une région différente de l'espace des données appelée aussi cluster.

L'apprentissage consiste en l'estimation à partir des observations d'une même classe des paramètres des gaussiennes qui composent le modèle de cette classe.

Les paramètres à estimer sont :

- les associés poids  $(w_k)_{k=1, \dots, K}$  à chacune des  $k$  composantes du mélange ;
- les moyennes et matrices de covariance de chacune des composantes du mélange :

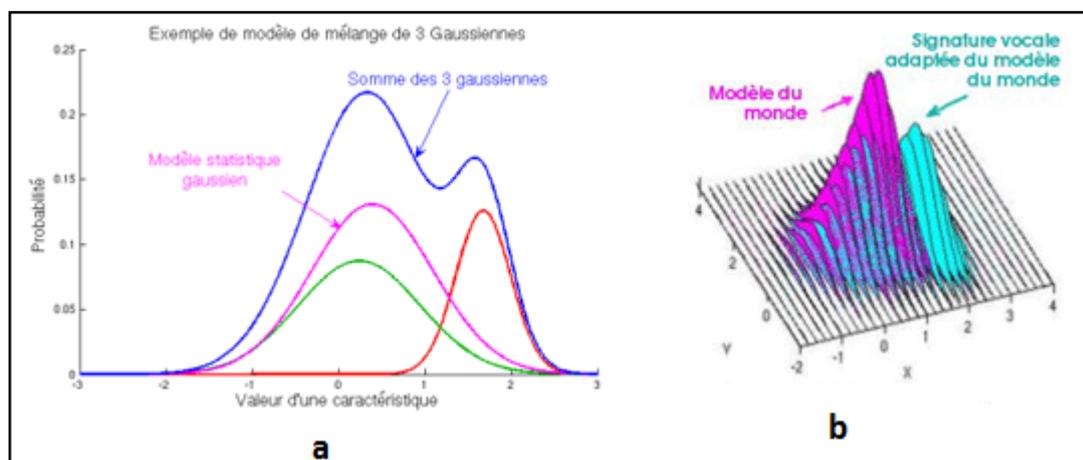
$$(\mu_k, \Sigma_k)_{k=1, \dots, K}$$



**Figure 2. 2 :** Les modèles probabilistes : exemples d'une distribution Gaussienne à 2 dimensions [7]

Sur Figure 2.2 -a, on observe en bleu des valeurs de 2 caractéristiques .Chaque rond bleu représente à un instant en abscisse et en ordonnée les valeurs respectives du premier et du second coefficient cepstral. La répartition des valeurs dans le plan peut alors être modélisée sous la forme d'une distribution Gaussienne, représentée en rouge.

Figure 2.2 -b montre la distribution Gaussienne estimée. La hauteur de la distribution représente la probabilité d'une caractéristique de prendre cette valeur. C'est pourquoi on parle de modèles probabilistes (ou statistiques).



**Figure 2. 3:**Les modèles probabilistes : exemple d'un modèle Gaussien à 2 dimensions [7]

Figure 2.3 -a montre un exemple de GMM à 3 gaussiennes en 1 dimension. Figure 2.3 -b illustre la manière dont peut être adaptée une signature vocale à partir du modèle du monde.

## 2.5. Estimation des paramètres

Choisi, vis à vis des données d'apprentissage, le critère le plus utilisé pour l'apprentissage des modèles GMM, est le critère de Maximum de Vraisemblance ML (maximum likelihood). L'estimation des paramètres des GMM consiste à trouver ceux qui maximisent la fonction de vraisemblance des données d'apprentissage.

$$\tilde{\lambda}_x = \operatorname{argmax}_{\lambda} p\left(\vec{x}_t / \lambda\right) \quad (2.03)$$

### 2.5.1. Algorithme des K-moyennes

L'algorithme des  $K$ -moyennes consiste à faire la répartition des vecteurs acoustiques  $x_i$  d'une classe (locuteur) en  $N$  sous-ensemble disjoints caractérisés par un centroïde. Le résultat de cette répartition est appelé dictionnaire. L'algorithme des  $K$ -moyennes n'est que localement optimal, par conséquent, il est influencé par ses conditions initiales. L'algorithme des  $K$ -moyennes est défini comme suit :

- La première étape est l'initialisation du dictionnaire, il existe plusieurs méthodes d'initialisation, nous pouvons citer :

#### 2.5.1.1. L'initialisation aléatoire

Le dictionnaire le plus simple est celui qui contient les  $L$  premiers vecteurs de la suite d'apprentissage, où ces  $L$  vecteurs sont extraits aléatoirement de cette suite. Ces vecteurs peuvent bien sûr ne pas être du tous des représentants de la suite d'apprentissage, et dans ce cas, on aboutit à des résultats très médiocres ;

### 2.5.1.2. L'algorithme à seuil

Au lieu de prendre  $L$  vecteurs aléatoirement, on fixe une distance minimale entre les éléments du dictionnaire initial. Cette méthode permet d'obtenir une meilleure représentativité que dans le cas précédent.

- la deuxième étape consiste à appliquer deux règles, tant qu'il y'a une amélioration importante de la distorsion moyenne donnée par la formule suivante :

$$D_m = \frac{1}{N} \sum_{K=1}^N d(\vec{X}_K, C(\vec{X}_K)) \quad (2.04)$$

$C(\vec{X}_K)$  est le centroïde de la région où  $\vec{X}_K$  est affecté.

$d(\vec{X}_K, C_i)$  est la distance euclidienne entre les vecteurs  $\vec{X}_K$  et  $C_i$ .

Les deux règles sont définies comme suit :

### 2.5.1.3. La règle de centroïde

Cette règle exige que tous les centroïde soient les moyennes des vecteurs acoustiques des régions représentées par ces centroïde. Cela peut être formulé comme suit:

$$C_i = \frac{1}{N} \sum_{K=1}^{N_i} \vec{X}_K \quad (2.05)$$

### 2.5.1.4. La règle de plus proche voisin

Le vecteur  $\vec{X}_K$  est affecté à la région  $i$  si la distance euclidienne entre ce vecteur et le centroïde de cette région est minimale. La formule suivante décrit explicitement cette règle.

$$region(\vec{X}_K) = i, C_i = \min(d(\vec{X}_K, C_i)) \quad (2.06)$$

$K$  est le nombre de régions.

$C_i$  est le centroïde de la  $i^{eme}$  région.

$d(\vec{X}_K, C_i)$  est la distance euclidienne entre les vecteurs  $\vec{X}_K$  et  $C_i$

### 2.5.2. Algorithme EM (Expectation Maximisation)

L'algorithme EM permet de régler les paramètres d'un modèle de distribution GMM pour atteindre un maximum de vraisemblance d'un ensemble d'observation. Ces observations sont typiquement des vecteurs d'apprentissage non étiquetés. Les paramètres libres sont constitués des poids, des vecteurs moyens et des matrices de covariance [23].

L'algorithme EM permet l'estimation de ces paramètres. Cet algorithme itératif garantit la croissance de la vraisemblance des données d'apprentissage avec les itérations. Chacune d'elle est formée de deux étapes :

- une étape E (Estimation) où la fonction vraisemblance des données complètes étant donnés les paramètres des modèles à l'itération précédente est estimée en commençant par les paramètres initiaux  $w_k, \mu_k, \Sigma_k$  du modèle initial  $\lambda$ , on estime les nouveaux paramètres  $\bar{w}_k, \bar{\mu}_k, \bar{\Sigma}_k$  telle que la vraisemblance du nouveau modèle soit supérieure ou égale à la vraisemblance du modèle initial.

Dans chaque itération de l'algorithme EM, pour tous les vecteurs acoustiques  $x_t: t = 1, \dots, T$ , il faut calculer la probabilité  $\gamma_{n,m}$  qui indique dans quelle proportion un vecteur  $x_t$  appartient à la gaussienne  $k$  tel que :

$$\gamma_t(k) = \frac{p(x_t, \mu_k, \Sigma_k)}{\sum_{k=1}^K p(x_t, \mu_k, \Sigma_k)} \quad (2.07)$$

- une étape M (Maximization) où une nouvelle estimation des paramètres du modèle est obtenue en maximisant la fonction de vraisemblance précédente. Les nouveaux paramètres sont définis comme suit :

$$\bar{w}_k = \frac{\sum_{t=1}^T \gamma_t(k)}{\sum_{k=1}^K \sum_{t=1}^T \gamma_t(k)} \quad (2.08)$$

$$\bar{\mu}_k = \frac{\sum_{t=1}^T \gamma_t(k) x_t}{\sum_{t=1}^T \gamma_t(k)} \quad (2.09)$$

$$\bar{\Sigma}_k = \frac{\sum_{t=1}^T \gamma_t(k) (x_t - \mu_k)(x_t - \mu_k)'}{\sum_{t=1}^T \gamma_t(k)} \quad (2.10)$$

Ce processus est répété plusieurs fois jusqu'à atteindre un seuil de convergence.

La qualité des paramètres estimés de la modélisation dépend de la quantité et de la représentativité des données d'apprentissage.

## 2.6. La phase d'apprentissage

Dans cette phase, on estime les paramètres des gaussiennes qui composent un modèle GMM en se basant sur les vecteurs acoustiques déterminés dans l'étape d'extraction de paramètres. L'apprentissage se fait en deux étapes:

- la première étape est l'initialisation des paramètres du modèle en utilisant l'algorithme K-moyennes (K-means) ou l'algorithme LBG ;
- la deuxième étape est l'optimisation des paramètres obtenus dans la première étape en utilisant l'algorithme EM.

## 2.7. Le rapport d'hypothèses Bayésien

En vérification du locuteur, le processus de décision est basé sur un test d'hypothèses.

Étant donné un signal de parole  $S$  et une identité  $I_X$  revendiquée par l'utilisateur, le système doit décider laquelle des deux hypothèses suivantes est la plus vraisemblable:

- $H_0$  : le signal  $S$  a été produit par  $I_X$  ;
- $H_1$  : le signal  $S$  n'a pas été produit par  $I_X$ .

Le rapport de vraisemblance (Likelihood Ratio - LR) entre les deux hypothèses  $H_0$  et  $H_1$  pour l'identité  $I_X$  est noté  $LR(X, H_0, H_1)$ . Le test bayésien est la comparaison du rapport de vraisemblance avec un seuil de décision  $\Omega$ .

$$\mathcal{L} = (S, H_0, H_1) = \frac{p(H_0/S)}{p(H_1/S)} \quad (2.11)$$

En pratique, il est plus facile d'estimer  $p(S/H_0)$  que  $p(H_0/S)$  et le théorème de Bayes permet d'écrire :

$$p(H_0/S) = \frac{p(S/H_0)p(H_0)}{p(S)} \quad (2.12)$$

Ainsi l'équation (2.11) devient :

$$\mathcal{L} = (S, H_0, H_1) = \frac{p(S/H_0)p(H_0)}{p(S/H_1)p(H_1)} \quad (2.13)$$

les probabilités a priori  $p(H_0)$  et  $p(H_1)$  sont incorporées au seuil de décision  $\Omega$  et finalement, le rapport entre les deux hypothèses  $H_0$  et  $H_1$  s'écrit :

$$\mathcal{L} = (S, H_0, H_1) = \frac{p(S/H_0)}{p(S/H_1)} \geq \Omega \begin{cases} H_1 \text{ est acceptée} \\ H_0 \text{ est acceptée} \end{cases} \quad (2.14)$$

La valeur  $p(H_0/S)$  est généralement la vraisemblance moyenne du signal  $S$  avec le modèle GMM de  $I_X$ . La valeur choisie au dénominateur du deuxième terme de l'équation (2.14) représente la vraisemblance du modèle de non-locuteur (tous les locuteurs hormis  $I_X$ ). Cette valeur n'est pas observable, elle est seulement la probabilité complémentaire de  $p(H_0/S)$ . Nous verrons par la suite que cette valeur, probabilité d'un événement non observable, est difficile à estimer et ne peut qu'être approximée [24].

Bien qu'ils soient capables de capturer les informations à plus long terme d'un locuteur, ils ne contiennent pas d'aspects dynamiques. Pour une bonne modélisation beaucoup de Gaussiennes nécessitent beaucoup de données.

## 2.8. Evaluation d'un système de VAL

L'évaluation de la qualité d'un système de VAL dépend de plusieurs facteurs. Ce sont les performances en termes de taux d'erreurs qui vont en déterminer la qualité. Cependant, un système d'authentification en phase d'exploitation dépend aussi des échecs d'apprentissage, c'est-à-dire si pour des raisons de défauts matériels, ou parce que l'enregistrement a de trop mauvaise qualité pour servir à l'authentification, le système décide de rejeter le signal et de procéder à une nouvelle phase d'entraînement. Ces mesures d'échecs sont courantes dans les modalités comme l'iris ou les empreintes digitales, où les techniques d'acquisition jouent un rôle primordial. Elles sont en revanche peu prises en compte pour la parole, dans le cadre de travaux laboratoire.

Un autre critère consiste à prendre en considération le corpus sur lesquels ces mesures ont été effectuées. En effet, un corpus comportant peu de variabilité ou un trop petit nombre de locuteurs peut guider à une mauvaise interprétation des résultats [16].

### 2.8.1. Mesures de performances

Les performances d'un système de VAL s'évaluent en fonction de deux taux d'erreurs. La probabilité de Faux Rejets (FR) ou de rejet du client à l'identité proclamée et la probabilité de Fausses Acceptations (FA) ou d'acceptations d'impostures. Ces taux sont étroitement liés. Au point de fonctionnement, pour un certain seuil de vérification, ces deux taux sont définis. En fonction du type d'application souhaitée, le seuil de vérification peut être choisi pour minimiser le taux de FA : application de sécurité, ou minimiser le taux de FR pour augmenter l'ergonomie d'utilisation.

Il n'est pas possible de minimiser conjointement ces deux taux :

$$P_{FA} = \frac{\text{Nombre d'imposteur acceptés}}{\text{Nombre d'accès imposteurs}} \quad (2.15)$$

$$P_{FR} = \frac{\text{Nombre de de client rejetés}}{\text{Nombre d'accès client}} \quad (2.16)$$

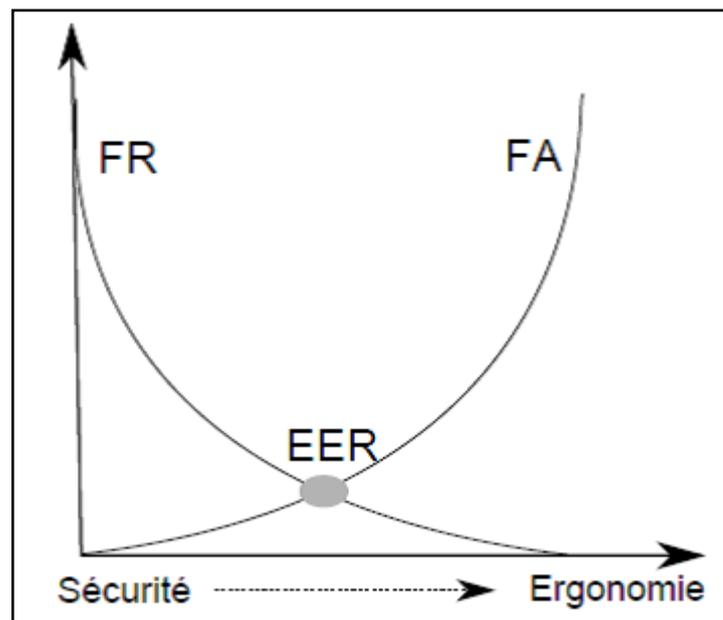
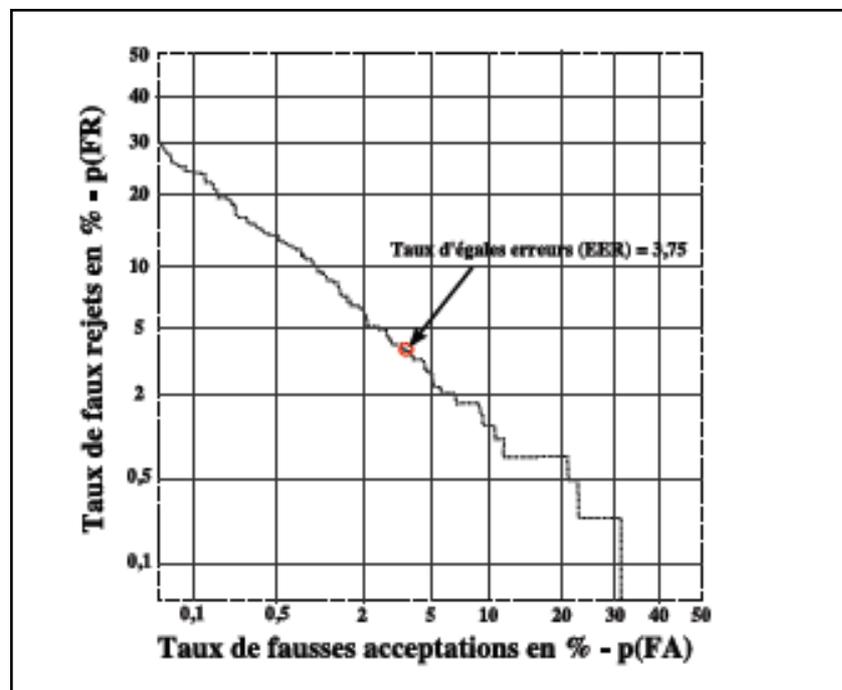


Figure 2.4: Evaluation des taux FA et FR [25]

### 2.8.2. Les courbes DET

La représentation la plus communément utilisée pour évaluer la pertinence du seuil de décision en fonction de ces deux taux d'erreurs est la courbe DET (Detection Error Tradeoff). Les échelles des axes suivent la répartition d'une loi normale contrairement à leurs prédécesseurs, les courbes ROC (Receiver Operating Characteristic), qui utilisent une échelle linéaire.

L'échelle logarithmique est utilisée pour rendre la courbe DET linéaire quand les scores des systèmes suivent une distribution Gaussienne. La courbe DET permet d'évaluer, pour chaque seuil de vérification, les valeurs du couple (FA, FR). La figure 2.5 illustre un exemple de courbe DET [25].



**Figure 2. 5:** Exemple de représentation des performances d'un système de vérification d'identité par une courbe DET [24]

### 2.8.3. Point de fonctionnement

L'évaluation d'un système en mode opérationnel doit être effectuée sur un point précis de fonctionnement, c'est-à-dire une valeur du seuil. Pour cela, une mesure pondérée des taux d'erreurs appelée fonction de coût de décision ou Décision Cost Function (DCF) est utilisée. Les pondérations relatives à cette mesure sont déterminées par l'application visée par le système. Les paramètres en jeu sont : les coûts associés à chaque taux d'erreur  $C_{FA}$ ,  $C_{FR}$  et les

probabilités a priori des populations imposteurs  $P_{IMP}$  et clients  $P_{CL}$ , cette fonction de coût s'exprime sous la forme :

$$DCF = C_{FR} \cdot P_{CL} \cdot P_{FR} + C_{FA} \cdot P_{IMP} \cdot P_{FA}. \quad (2.17)$$

La DCF est souvent normalisée pour obtenir des valeurs entre 0 et 1. Cette mesure est utilisée pour évaluer les systèmes en mode opérationnel, c'est-à-dire lorsque le réglage du seuil de décision a été effectué. Dans la phase de développement, les taux d'erreurs sont fonction du seuil. Dans les expériences en laboratoire, la mesure de performance utilisée est celle qui minimise la fonction de coût  $DCF_{min} = \min_{\theta} DCF$ . Nous assimilerons pour des notions de clarté la mesure  $DCF_{min}$  à la mesure DCF. Lorsque les coûts associés aux erreurs sont identiques et qu'il n'existe pas d'a priori sur la population ( $P_{CL} = P_{IMP} = 0.5$ ) alors ce point de fonctionnement correspond à l'HTER (pour Half Total Error Rate). Une autre mesure largement utilisée pour mesurer les performances est l'Equal Error Rate, EER où les taux d'erreurs sont identiques  $P_{FA} = P_{FR}$ . Ces différentes mesures sont illustrées sur une courbe DET présentée en exemple à la (figure 2.5) [16].

#### 2.8.4. Choix du seuil de décision

La décision d'acceptation ou de rejet d'une séquence test se fait généralement par comparaison du score final à un seuil. Si l'on dispose d'un corpus de développement étiqueté, on peut fixer ce seuil pour optimiser la mesure de performance considérée pour l'application visée. Sinon, on peut procéder par validation croisée sur les données d'apprentissage. Si l'on dispose de beaucoup de données d'apprentissage pour un locuteur, on peut même envisager une approche individuelle (ou lieu d'une approche globale), c'est-à-dire choisir un seuil de décision spécifique au locuteur. Mais les diverses normalisations appliquées aux scores ("cohorte normalisation" dans le rapport de vraisemblance, et Z- ou T-Norm) rendent légitime une approche globale, qui est non seulement moins lourde à mettre en œuvre, mais qui permet aussi de rendre cohérente la courbe DET pour une visibilité des performances. En effet, dans le cas où le seuil de décision est variable selon les locuteurs, le point de fonctionnement réel situe généralement hors de la courbe DET [16].

## **2.9. Conclusion**

Dans ce chapitre, nous avons présenté les différentes approches de modélisation du locuteur par le mélange de gaussiennes (GMM). Nous avons présenté ensuite la méthode d'estimation de l'ensemble de ces paramètres et enfin, les techniques d'évaluation du système de Vérification Automatique du Locuteur (VAL).

# **Chapitre 3 :**

**Notions de base sur les Algorithmes**

**Génétiques**

### 3.1. Introduction

Dans ce chapitre nous allons présenter le principe de fonctionnement des algorithmes génétiques (population, hybridation, mutation,...), ainsi que le rôle des différents opérateurs génétiques utilisés, y compris leur dimensionnement et leurs critères d'arrêt. Ces opérations sont imitées par les AG afin de faire évoluer les populations de solutions de manière progressive.

### 3.2. Concepts de base des algorithmes génétiques

Les Algorithmes Génétiques sont des méthodes d'optimisation s'appuyant sur des techniques dérivées de la génétique et de l'évolution naturelle. L'évolution biologique procède en sélectionnant des génotypes (intégrés aux chromosomes), sur la base de l'adaptation relative à leur environnement des phénotypes qu'ils génèrent (la qualité de cette adaptation est alors mesurée par la performance - fitness - relative de chaque génotype).

Chaque chromosome est muni d'une mesure d'adaptation (fitness). Un algorithme génétique choisit des paires de chromosomes-parents, en favorisant les plus adaptés (ceux de meilleur fitness) et engendre de nouvelles solutions (enfants) en appliquant des opérateurs de croisement (crossover) et de mutation [26].

L'évaluation de la population initiale ainsi qu'à ces progénitures est faite à l'aide des fonctions d'évaluation dites "objectives". Ensuite, le hasard guide la recherche vers un espace susceptible d'améliorer la solution [27].

### 3.3. Caractéristiques des algorithmes génétiques

Pour un problème particulier, un AG doit avoir les cinq composantes suivantes :

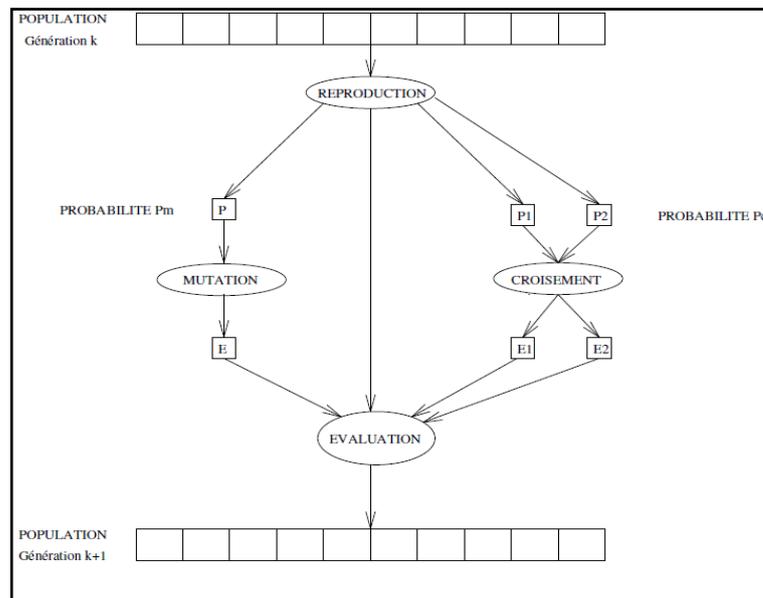
- une représentation génétique des solutions possibles au problème ;
- un moyen de créer une population initiale de solutions potentielles ;
- une fonction d'évaluation qui joue le rôle de l'environnement. Evaluation des solutions en termes de leur "fitness" ;
- les opérateurs génétiques qui altèrent la composition des enfants ;

- les valeurs des divers paramètres qu'emploie d'AG (taille de population probabilités d'appliquer les opérateurs génétiques, etc.) [28].

### 3.4. Etapes de fonctionnement

L'idée principale des AG est de faire évoluer un ensemble (population) de solutions possibles (individus ou chromosomes) d'un problème d'optimisation donne pour obtenir l'optimum global du dit problème [29].

On commence par générer une population d'individus de façon aléatoire. Pour passer d'une génération  $k$  à la génération  $k + 1$ , les trois opérations suivantes sont répétées pour tous les éléments de la population  $k$ . Des couples de parents  $P_1$  et  $P_2$  sont sélectionnés en fonction de leurs adaptations. L'opérateur de croisement leur est appliqué avec une probabilité  $P_c$  (généralement autour de 0.6) et génère des couples d'enfants  $E_1$  et  $E_2$ . D'autres éléments  $P$  sont sélectionnés en fonction de leur adaptation. L'opérateur de mutation leur est appliqué avec la probabilité  $P_m$  ( $P_m$  est généralement très inférieur à  $P_c$ ) et génère des individus mutés  $E$ . Le niveau d'adaptation des enfants  $E_1$  et  $E_2$  et des individus mutés  $P$  sont ensuite évalués avant insertion dans la nouvelle population. Différents critères d'arrêt de l'algorithme peuvent être choisis [30] (Figure 3.1).



**Figure 3. 1:** Principe général des AG [30]

### 3.4.1. Présentation d'AG

Un AG défini par :

- un Individu/chromosome/séquence : une solution potentielle du problème ;
- une Population : un ensemble de chromosomes ou de points de l'espace de recherche ;
- un Environnement : l'espace de recherche ;
- une Fonction de fitness : la fonction - positive - que nous cherchons à maximiser [26].

La terminologie employée est empruntée à la génétique est la suivante (Tableau.3.1) :

**Tableau 3. 1** : Analogie de la terminologie génétique avec la nature [6]

Algorithmes Génétiques	Nature
Individu	Génome
Chaîne, séquence	Chromosome
Traits, caractéristiques	Gènes
Valeur de la caractéristique	Allèle
Position dans la chaîne	Locus
Structure	Génotype
Ensemble de paramètres Une structure décodée	Phénotype

### 3.4. 2. Codage

La première étape dans le fonctionnement de l'algorithme génétique est le codage de l'ensemble des paramètres à optimiser en une chaîne de caractères de longueur finie. Pour une fonction (modèle d'un système) multivariable  $f(k_1, k_2, \dots, k_n)$ , le codage se déroule de la façon suivante:

- Coder chaque variable  $k_i$  ( $i = 1, \dots, n$ ) selon le choix de sa longueur.
- Construire les chaînes en concaténant les différents codes, c'est-à-dire juxtaposer ces derniers l'un après l'autre.

Plusieurs types du codage sont utilisés : Codage binaire, entier, réel et Codage Gray [31].

### 3.4.2.1. Codage binaire

Le codage binaire est le plus utilisé. Chaque gène dispose du même alphabet binaire 0 ou 1.

Un gène est alors représenté par un entier long de 32 bits, les chromosomes qui sont des suites de gènes, et Chaque individu est représenté par un ensemble de chromosomes et chaque population est un ensemble d'individus.

Un des avantages du codage binaire est que nous pouvons facilement coder toutes sortes d'objets dont des réels, des entiers, des valeurs booléennes et des chaînes de caractères. Cela, nécessite simplement l'usage de fonctions de codage et de décodage pour passer d'une représentation à une autre. Si, nous avons 'n' variable réel noté  $x_i$ , nous considérons un espace de recherche fini, tel que:

$$x_{i \min} \leq x_i \leq x_{i \max}, \quad \forall i \in [1, \dots, n] \quad (3.01)$$

Afin de coder les variables réelles en binaire, nous discrétisons l'espace de recherche. Ainsi un codage sur 32 bits implique une discrétisation des intervalles tel que :

$$g_{\max} = 2^{32} - 1 \quad (3.02)$$

Avec  $g_{\max}$  : nombre de valeurs discrètes.

A chaque variable réelle  $x_i$  nous associons un entier long noté  $g_i$  tel que :

$$0 \leq g_i \leq g_{\max}, \quad \forall i \in [1, \dots, n] \quad (3.03)$$

Les formules de codage et de décodage sont alors :

$$g_i = \frac{x_i - x_{i \min}}{x_{i \max} - x_{i \min}} g_{\max} \quad (3.04)$$

$$x_i = x_{i \min} + (x_{i \max} - x_{i \min}) \frac{g_i}{g_{\max}} \quad (3.05)$$

Soit :

Le codage binaire présente les inconvénients suivants :

- il peut être difficile d'adapter ce codage à certains problèmes. La représentation binaire traditionnelle utilisée pour les AG crée des difficultés pour les problèmes

d'optimisation de grandes dimensions à haute précision numérique. Par exemple, si nous devons utiliser 100 variables appartenant au domaine et si nous avons besoin précision de l'ordre de  $10^{-6}$ , nous devons utiliser des chromosomes d'une longueur de  $l=3000$ . Cela, en retour, génère un espace de recherche de l'ordre de  $10^{30}$ .

Pour de tels problèmes, les AG basés sur des représentations binaires ont de faibles performances ;

- suivant le problème, la résolution de l'algorithme peut être coûteuse en temps ;
- le croisement et la mutation peuvent être inadaptés pour la création, par exemple, d'individus n'appartenant pas à l'espace de recherche.

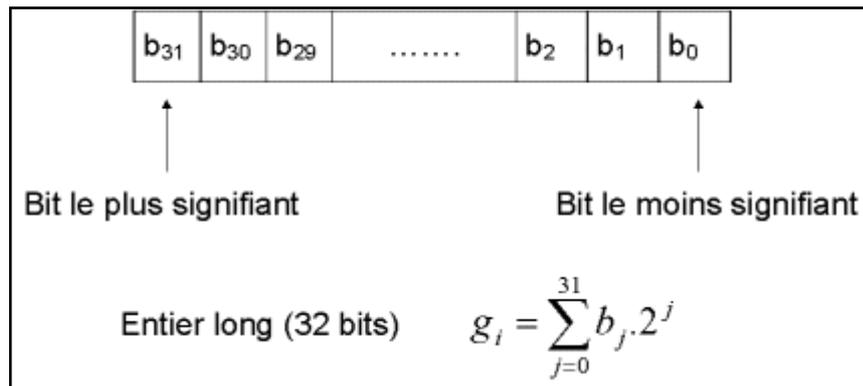
#### **3.4.2.2. Codage de Gray**

Dans le cas d'un codage binaire on utilise souvent la "distance de Hamming", deux chaînes (chromosomes) qui sont adjacentes au niveau phénotype comme mesure de la dissimilarité entre deux éléments de population, cette mesure compte les différences de bits de même rang de ces deux séquences. Et c'est là que le codage binaire commence à montrer ses limites. En effet, deux éléments voisins en termes de distance de Hamming ne codent pas nécessairement deux éléments proches dans l'espace de recherche. Cet inconvénient peut être évité en utilisant un "codage de Gray" : le codage de Gray est un codage qui a comme propriété qu'entre un élément  $n$  et un élément  $n + 1$ , donc voisin dans l'espace de recherche, un seul bit. L'utilisation du codage Gray peut produire une amélioration des performances GA dans certains cas [32].

#### **3.4.2.3. Codage en réel de paramètres**

Les études ont montré que la représentation réelle aboutit souvent à une meilleure précision et qu'en règle générale le gain en termes de temps de calcul est important. Ainsi, une représentation plus naturelle du problème offre des solutions plus efficaces. Cela, peut être utile notamment dans le cas où nous recherchons le maximum d'une fonction réelle. En utilisant le codage réel, l'individu n'est alors plus qu'un nombre réel dans l'espace des valeurs permises. L'opérateur de sélection reste identique à celui de la roulette biaisée ou du tournoi.

En revanche, nous utilisons de nouveaux opérateurs de croisement et de mutation.



**Figure 3. 2:** chaque gène est codé par un entier long (32 bits) [33]

### 3.4.3. Génération de population initiale

A l'initialisation de l'algorithme, une population de taille fixée est générée aléatoirement. Elle est en général répartie uniformément sur l'espace de recherche. Chaque individu est ensuite évalué sur la base de la fonction objective qui constitue la fonction « fitness » du chromosome[34].

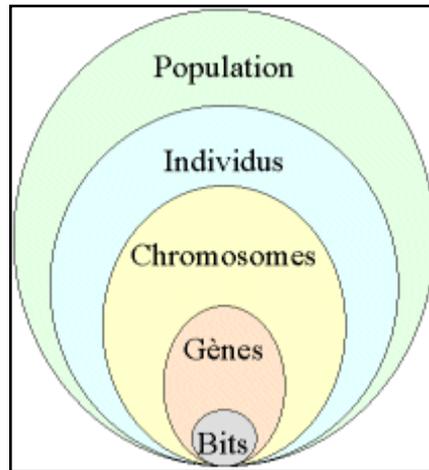
Les principales questions à examiner sont d'abord la taille de la population, et d'autre part la méthode par laquelle les individus sont choisis.

Les résultats empiriques de nombreux auteurs Suggèrent que les tailles de population plus petite que 30 sont tout à fait adéquates dans de nombreux cas. Une analyse ultérieure à partir d'une perspective différente conduit Goldberg et ses collègues à l'idée que la dépendance linéaire de la taille de la population sur la longueur de la chaîne était adéquate [35].

### 3.4.4. La fonction d'évaluation

Elle prend en argument l'individu et lui associe un coût de performance. Cette fonction doit guider l'algorithme vers l'optimum en réalisant implicitement une pression de sélection dans cette direction. De plus, l'efficacité d'un AG s'appuie pour une grande part, sur la qualité de cette fonction. La mise au point d'une bonne fonction d'adaptation doit respecter plusieurs critères qui se rapportent à sa complexité et à la satisfaction des contraintes du problème. Lorsque la fonction d'adaptation apparaît excessivement complexe, consommant une importante puissance de calcul, la recherche d'une approximation plus simple est souhaitable [36].

La population évolue jusqu'à satisfaction d'un critère d'arrêt : présence d'un individu satisfaisant ou nombre de générations fixé.



**Figure 3.3 :** Les cinq niveaux d'organisation d'un AG [33]

### 3.4.5. Sélection

Le procédé de la sélection, est celui, comme on le sait naturellement, chargé de la procréation des générations successives à partir de la population initiale, en sélectionnant les individus les plus adaptés à contribuer dans la création de la génération suivante [37].

Il existe plusieurs méthodes de sélection, citons quelques-unes :

#### 3.4.5.1. Roulette de casino

C'est la sélection naturelle la plus employée pour l'AG binaire. Chaque chromosome occupe un secteur de roulette dont l'angle est proportionnel à son indice de qualité. Un chromosome est considéré comme bon aura un indice de qualité élevé, un large secteur de roulette et alors il aura plus de chance d'être sélectionné.

#### 3.4.5.2. Méthode élitiste ou par rang

Les individus sont triés selon leur fonction d'adaptation, seul la moitié supérieure de la population correspondant aux meilleurs composants est sélectionnée, nous avons constaté que la pression de sélection est trop forte. Il est important de maintenir une diversité de gènes pour les utiliser dans la population suivante et avoir des populations nouvelles quand on les combine [33].

### 3.4.5.3. Sélection par tournoi

La sélection par tournoi qui consiste à prélever au hasard un échantillon de  $n$  individus (2 au minimum) à chaque tournoi. Ensuite, le meilleur de cet échantillon est choisi pour être parent [38].

Les individus qui gagnent à chaque fois, on peut les copier plusieurs fois ce qui favorisera la pérennité de leurs gènes.

## 3.5. Reproduction

Durant la phase de reproduction de l'AG, des individus de la population sont sélectionnés d'après la méthode choisie et sont recombinaisonnés produisant des enfants de la prochaine génération. Cette phase utilise des mécanismes de reproduction qui sont : le croisement ou (crossover en Anglais) et la mutation. Les transitions entre générations se font par ces opérateurs génétiques qui sont des fonctions aléatoires. Cette étape est stochastique puisque le résultat de l'application d'un opérateur dépend des tirages aléatoires. Les méthodes de reproduction proposées dans la littérature concernent les chromosomes dont la représentation des gènes est binaire. Dans le cas d'un problème d'optimisation où les gènes du chromosome sont des nombres réels, avant d'effectuer les opérations de reproduction, les gènes du chromosome sont convertis en bits (16 ou 32 bits), puis les opérations de reproduction sont effectuées sur cette représentation binaire [6].

Pour résoudre un problème dans un domaine de recherche continue, chaque paramètre (gène) noté  $C_i$  du chromosome noté  $C$  défini dans l'intervalle  $[a_i, b_i]$  est converti en code binaire tel que l'intervalle  $[a_i, b_i]$  soit transformé en un ensemble  $\{0, \dots, 2^{L_i}\}$  où  $L_i$  est le nombre de bits. Le chromosome ainsi converti aura une grande dimension répondant à la grande précision numérique requise pour la résolution du problème. Durant l'exécution de l'AG, plusieurs opérations inutiles seront effectuées entraînant une exploration trop étendue de l'espace de recherche. Cette exploration va balayer des solutions sans aucune signification impliquant la non convergence de l'algorithme. Pour pallier à ce genre de problème, la représentation réelle sera donc utilisée au lieu de la représentation binaire [6].

### 3.5.1. Croisement

Suite à la génération de la population initiale aléatoirement, nous procédons la phase de croisement qui assure la recombinaison de gènes parentaux pour former de nouveaux descendants [39].

Le croisement consiste à prendre pour chaque bit de l'enfant l'un des deux bits des parents, et plusieurs méthodes sont possibles pour cela : on peut, par exemple, décider de prendre les bits d'un parent jusqu'à une certaine position, puis ceux de l'autre parent (ce qui revient à couper le génome en un seul point) ; on peut aussi simplement décider pour chaque bit avec probabilité 1/2, indépendamment, si on prendra celui de l'un ou l'autre parent. Par la suite, pour la simplicité de l'exposé, on se tiendra à cette dernière possibilité.

Le phénomène de croisement est une propriété naturelle de l'ADN, et c'est analogiquement qu'on fait les opérations de croisement dans les AG.

#### 3.5.1.1. Croisement binaire

Le croisement est un processus aléatoire de probabilité  $P_C$  appliqué essentiellement à deux couples de parents pris au hasard dans la population. Il existe à échanger une partie de matériel génétique des parents pour former deux individus (enfants).

Plusieurs types de croisement binaires sont disponibles :

#### 3.5.1.2. Croisement en un point

On choisit au hasard un point de croisement, pour chaque couple. Notons que le croisement s'effectue directement au niveau binaire, et non pas au niveau des gènes. Un chromosome peut donc être coupé au milieu d'un gène (Figure 3.4).

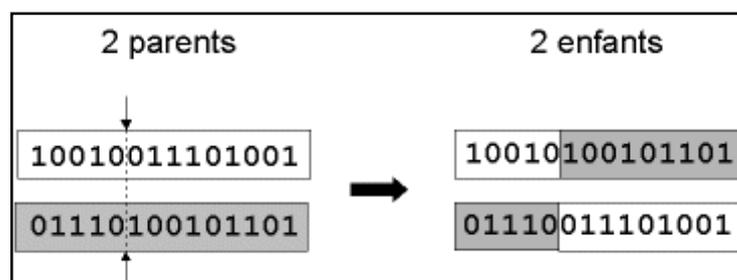
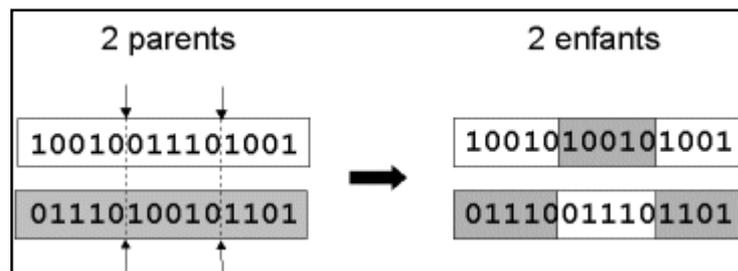


Figure 3. 4: Représentation schématique du croisement en 1 point

Les chromosomes sont bien sûr généralement beaucoup plus longs.

### 3.5.1.3. Croisement en deux points

On choisit au hasard deux points de croisement (Figure 3.5). Par la suite, nous avons utilisé cet opérateur car il est généralement considéré comme plus efficace que le précédent. Néanmoins nous n'avons pas constaté de différence notable dans la convergence de l'algorithme.



**Figure 3. 5:** Représentation schématique du croisement en 2 points

Notons que d'autres formes de croisement existent, du croisement en  $k$  points jusqu'au cas limite du croisement uniforme.

### 3.5.1.4. Croisement réel

Le croisement réel ne se différencie du croisement binaire que par la nature des éléments qu'il altère : ce ne sont plus des bits qui sont échangés à droite du point de croisement, mais des variables réelles.

### 3.5.1.5. Croisement arithmétique

Le croisement arithmétique est propre à la représentation réelle. Il s'applique à une paire de chromosomes et se résume à une moyenne pondérée des variables des deux parents.

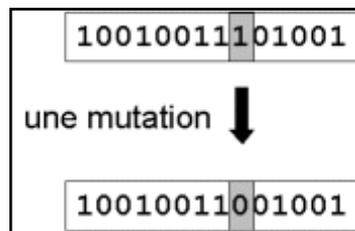
Soient  $[a_i, b_i, c_i]$  et  $[a_j, b_j, c_j]$  deux parents, et  $p$  un poids appartenant à l'intervalle  $[0, 1]$ , alors les enfants sont  $[pa_i + (1 - p)a_j, pb_i + (1 - p)b_j, pc_i + (1 - p)c_j]$ . Si nous considérons que  $p$  est un pourcentage, et que  $i$  et  $j$  sont nos deux parents, alors l'enfant  $i$  est constitué à  $p\%$  du parent  $i$  et à  $(100 - p)\%$  du parent  $j$ , et réciproquement pour l'enfant  $j$ .

### 3.5.2. Mutation

Le dernier opérateur dans l'AG est l'algorithme de mutation. L'effet de la mutation est d'empêcher la population de la stagnation à un optimum local.

Elle peut se produire à chaque position de bit dans une chaîne avec une certaine probabilité, généralement très faible (par exemple,  $10^{-3}$ ). Nous définissons une mutation comme étant l'inversion d'un bit dans un chromosome [40].

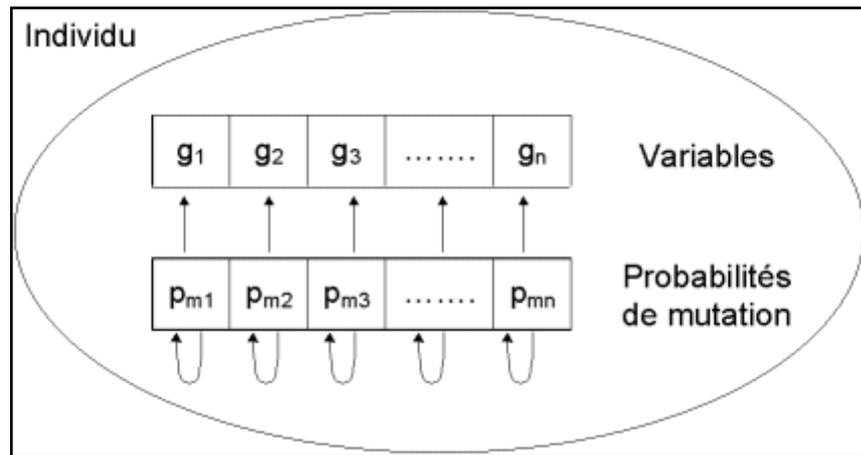
Cela revient à modifier aléatoirement la valeur d'un paramètre du dispositif. Les mutations jouent le rôle de bruit et empêchent l'évolution de se figer. Elles permettent d'assurer une recherche aussi bien globale que locale, selon le poids et le nombre des bits mutés. De plus, elles garantissent mathématiquement que l'optimum global peut être atteint (Figure 3.6).



**Figure 3. 6:** Représentation schématique d'une mutation dans un chromosome

D'autre part, une population trop petite peut s'homogénéiser à cause des erreurs stochastiques : les gènes favorisés par le hasard peuvent se répandre au détriment des autres. Cet autre mécanisme de l'évolution, qui existe même en l'absence de sélection, est connu sous le nom de dérivé génétique. Du point de vue du dispositif, cela signifie que l'on risque alors d'aboutir à des dispositifs qui ne seront pas forcément optimaux. Les mutations permettent de contrebalancer cet effet en introduisant constamment de nouveaux gènes dans la population.

Comment réaliser notre opérateur mutation ? De nombreuses méthodes existent. Souvent la probabilité de mutation  $P_m$  par bit et par génération est fixée entre  $10^{-3}$  et  $10^{-2}$ . On peut prendre également  $P_m = \frac{1}{l}$  où  $l$  est le nombre de bits composant un chromosome. Il est possible d'associer une probabilité différente de chaque gène, et ces probabilités peuvent être fixes ou évoluer dans le temps.



**Figure 3. 7:** principe de l'auto-adaptation [33]

A chaque variable est associée sa propre probabilité de mutation, qui est elle-même soumise au processus d'évolution. L'individu possède donc un second chromosome codant ces probabilités.

Après divers essais, ils ont abouti à la méthode d'auto-adaptation des probabilités de mutation. Si dans un environnement stable il est préférable d'avoir un taux de mutation faible, la survie d'une espèce dans un environnement subissant une évolution rapide nécessite un taux de mutation élevé permettant une adaptation rapide. Les taux de mutation d'une espèce dépendent donc de leur environnement [33].

Pour prendre en compte cette formulation biologique et l'adapter à notre cas, ils ont introduit dans chaque individu (dispositif) un second chromosome (ensemble de paramètres) dont les gènes (paramètres) représentent les probabilités de mutation de chaque gène du premier chromosome (Figure 3.8). Ce second chromosome est géré de façon identique au premier, c'est-à-dire qu'il est lui-même soumis aux opérateurs génétiques (croisement et mutation). Cela revient à fixer les probabilités assurant la modification des valeurs des paramètres du composant en fonction des valeurs d'un ensemble d'autres paramètres (les probabilités de mutation).

Lors de la genèse, les probabilités de mutation sont posées égales à 0,1 (valeur qui a paru la meilleure après plusieurs essais). Au cours du déroulement de l'algorithme, les gènes et les individus ayant des probabilités de mutation trop élevées ont tendance à disparaître. De même, les gènes ayant des probabilités de mutation trop faibles ne peuvent pas évoluer

favorablement et tendent à être supplantés. Les probabilités de mutation dépendent donc du gène considéré et de la taille de la population. De plus, elles évoluent au cours du temps. Il y a donc auto-adaptation des probabilités de mutation [33].

### **3.5.2.1. Mutation binaire**

La mutation binaire s'applique à un seul chromosome. Un bit du chromosome est tiré au hasard. Sa valeur est alors inversée.

Il existe une variante où plusieurs bits peuvent muter au sein d'un même chromosome. Un test sous le taux de mutation est effectué non plus pour le chromosome mais pour chacun de ses bits : en cas de succès, un nouveau bit tiré au hasard remplace l'ancien.

### **3.5.2.2. Mutation réelle**

La mutation réelle ne se différencie de la mutation binaire que par la nature de l'élément qu'elle altère : ce n'est plus un bit qui est inversé, mais une variable réelle qui est de nouveau tirée au hasard sur son intervalle de définition.

### **3.5.2.2. Mutation uniforme**

Il s'agit de modifier un paramètre en choisissant une valeur de façon uniformément aléatoire selon une certaine probabilité tiré dans une distribution uniforme sur l'intervalle définie par les contraintes du domaine [38].

### **3.5.2.3. Mutation non uniforme**

La mutation non uniforme possède la particularité de retirer les éléments qu'elle altère dans un intervalle de définition variable et de plus en plus petit. Plus nous avançons dans les générations, moins la mutation n'écarte les éléments de la zone de convergence. Cette mutation adaptative offre un bon équilibre entre l'exploration du domaine de recherche et un affinement des individus. Le coefficient d'atténuation de l'intervalle est un paramètre de cet opérateur.

### 3.6. Remplacement

Le remplacement consiste à réintroduire les descendants obtenus par application successive des opérateurs de sélection, de croisement et de mutation dans la population de leurs parents. Ils vont remplacer une certaine proportion de ceux-ci, proportion pouvant bien sûr être choisie. Le rapport entre le nombre d'individus nouveaux allant être introduits dans la population  $P$  et le nombre d'individus de cette population est connu sous le nom de saut de génération. Nous trouvons essentiellement deux méthodes de remplacement [6].

#### 3.6.1. Remplacement stationnaire

Dans ce cas, les enfants remplacent automatiquement les parents sans tenir compte de leurs performances, et le nombre d'individus de la population ne varie pas tout au long du cycle d'évolution simulé, ce qui implique l'initialisation de la population avec un nombre d'individus suffisant. Cette méthode peut être mise en oeuvre de deux façons différentes :

- la première se contente de remplacer la totalité de la population  $P$  par la population  $P'$ . cette méthode est connue sous le nom de remplacement générationnel et nous avons donc un saut de génération qui vaut 1 ;
- la deuxième méthode consiste à choisir une certaine proportion d'individus de  $P'$  qui remplaceront leurs parents.

Ce type de remplacement engendre une population ayant une grande variation et de ce fait favorise la dérive génétique qui se manifeste, d'autant plus que la population est de petite taille. De plus dans bien des cas, étant donné que même un enfant ayant une faible performance remplace forcément un parent, nous n'atteignons pas la meilleure solution mais nous nous en approchons seulement.

#### 3.6.2. Remplacement élitiste

Dans ce cas, nous gardons au moins l'individu possédant les meilleures performances d'une génération à la suivante. En général, nous pouvons partir du principe qu'un nouvel individu enfant prend place au sein de la population que s'il remplit le critère d'être plus performant que le moins performant des individus de la population précédente. Par conséquent, les enfants d'une génération ne remplaceront pas nécessairement leurs parents comme dans le remplacement stationnaire et par la même la taille de la population n'est pas figée au cours du

temps. Ce type de stratégie améliore les performances des AG dans certains cas. Mais présente l'inconvénient d'augmenter le taux de convergence prématuré ; Néanmoins, des implémentations plus fines procèdent de manière différente. Le taux de remplacement n'est pas de 100 %, la taille de la population augmente au cours des générations successives [6].

### **3.7. Critère d'arrêt**

Les étapes d'évaluation reproduction/sélection sont appliquées d'une manière itérative sur la population, jusqu'à satisfaction d'une condition Fin ; cette condition est soit un nombre maximum de générations, soit une valeur de fitness minimale, soit une convergence vers la meilleure solution.

Au cours de son évaluation, la population tend à converger, c.-à-d., que les individus tendent à se ressembler de plus en plus. Quand la population s'est uniformisée en grande partie, les individus fournissent une bonne approximation d'un optimum du problème. Si cet optimum n'est pas toujours l'optimum global, c'est généralement un optimum local proche (en qualité) de celui-ci.

### **3.8. Les paramètres d'un AG**

Les algorithmes génétiques peuvent être réglés à travers différents paramètres fixés à l'avance et dont dépend très fortement la bonne convergence de l'algorithme, les plus importants sont :

#### **3.8.1. La taille de la population**

Les conditions de convergence changent avec la taille de la population, lorsque celle-ci est grande, sa diversité augmente ce qui diminue la convergence vers un optimum local, le temps d'exécution de chaque génération augmente, la recherche risque de s'effectuer d'une façon redondante et l'efficacité de l'algorithme est globalement affectée. Par contre, si la taille de la population est petite, alors la probabilité de s'attarder sur des minima locaux est grande. La taille selon le cas se situe entre 25 et 100 [36].

### 3.8.2. Le taux de croisement

Le taux de croisement dépend en général de la forme de la fonction de performance. Son choix est bien souvent heuristique. Plus il est élevé, plus la population subit de changements importants. Les valeurs généralement admises sont comprises entre 0.5 et 0.9 [6].

### 3.8.3. Le taux de mutation

La probabilité de mutation est généralement faible puisqu'un taux élevé risque de conduire à une solution sous-optimale en perturbant celle qui est optimale [6].

### 3.8.4. Le fossé des générations

K. DE JONG (Syswerda, 1989) a proposé de ne pas remplacer l'ensemble de la population à chaque génération et pour cela, il a introduit la notion d'écart entre les générations, un nombre compris entre 0 et 1 qui indique la proportion de parents qui sont remplacés par leurs descendants. Si ce taux est à 1, l'ensemble de la population est remplacé [36].

## 3.9. Conclusion

Les algorithmes génétiques présentent un fort potentiel d'applications pratiques. D'ailleurs, ils sont de plus en plus utilisés dans multiples domaines. Ils peuvent être également utilisés pour l'optimisation des systèmes de RAL.

Il faut dire qu'ils fournissent d'excellentes performances à de faibles coûts. Les AG évitent un piège très souvent rencontré dans les algorithmes d'optimisation : ils ne s'arrêtent pas dans les extrema locaux, c'est-à-dire qu'ils essayent constamment de trouver de meilleures solutions, même s'ils semblent les avoir atteintes.

# **Chapitre 4 :**

**Utilisation des Algorithmes  
Génétiques dans  
la Modélisation  
du Locuteur**

#### 4.1. Introduction

Dans ce travail nous proposons un algorithme hybride AG-ML pour la modélisation du locuteur basée sur GMM, nous utilisons la ré-estimation ML comme un opérateur heuristique pour améliorer la vitesse convergent du AG.

#### 4.2. Description de corpus utilisé et Paramétrisation

Nous avons utilisé un corpus composé de 6 locuteurs 3 hommes ( $H_1, H_2, H_3$ ) et 3 femmes ( $Fe_1, Fe_2, Fe_3$ ) âgés entre 22 et 55 ans.

Notre corpus contient tous les énoncés des 10 mots suivants (Tableau 4.1).

**Tableau 4. 1:** le corpus utilisé dans la tache de VAL

Mot	API
وَزَنَ	[wʔzʔna]
الشَّمْسُ	[eʃʔmsu]
ضَرَبَ	[ðaraba]
هُتَافٌ	[hutʔfun]
عُرُوبٌ	[ʃuru:bun]
نَادَى	[nʔ:da]
دَحْرَجَ	[dʔhradʒa]
خُرَاعَةٌ	[xuzaʃa]
شَحَدَ	[ʃahaða]
مَكْتَبٌ	[mʔktʔbun]

Nous avons concaténé ces mots pour l'apprentissage et 3 mots sont utilisés pour le test.

Chaque locuteur doit prononcer chaque mot 3 fois par conséquent il y a 30 énoncés pour l'apprentissage et 9 pour le test pour chaque locuteur.

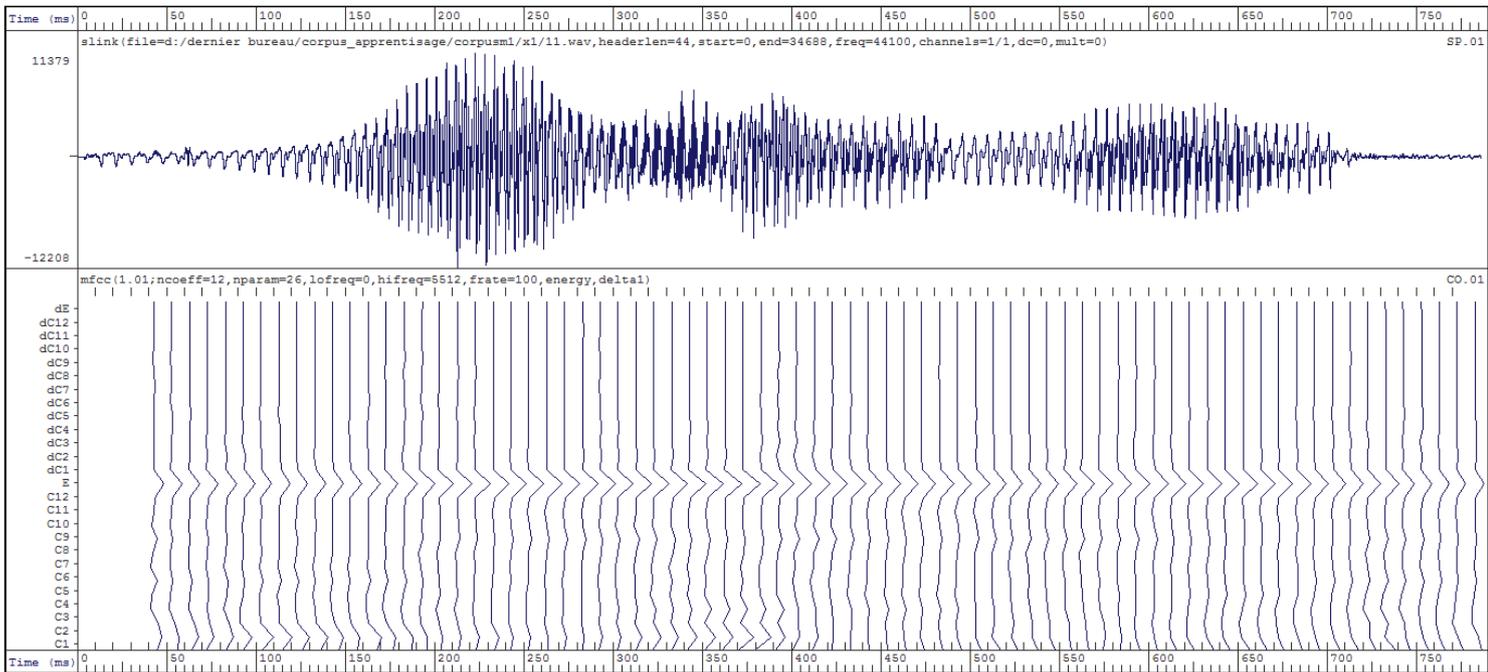
Les mots de notre corpus sont enregistrés dans un milieu ambiant, sous forme de fichiers .wave. Avec une fréquence d'échantillonnage 44100 Hz et codé e sur 16-bits.

La numérisation du signal a été faite par un logiciel professionnel "Sound Forge version 8", c'est le plus connu dans l'édition audio numérique.

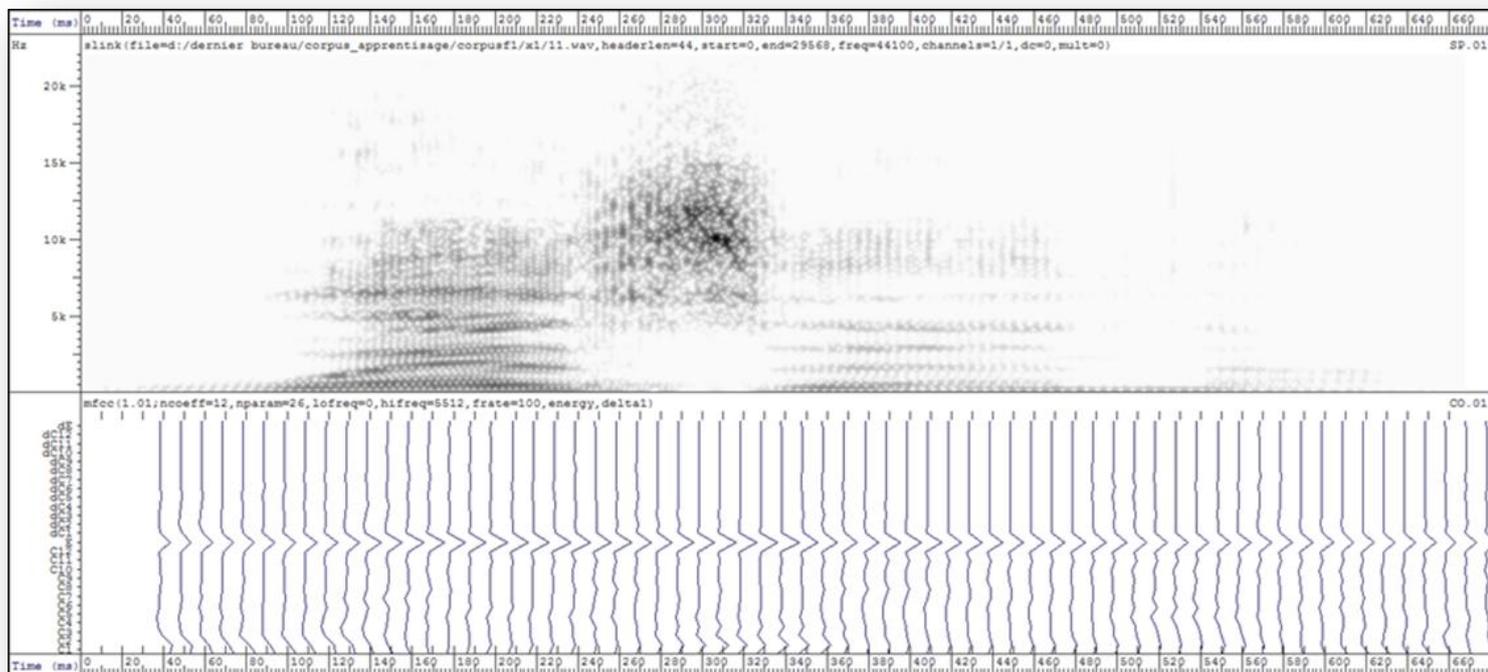
En phase de paramétrisation, nous avons spécifié l'espace de caractéristiques utilisé. Cet espace est défini par des vecteurs de taille fixe. En effet, comme le signal audio est dynamique et variable, nous avons présenté les séquences d'observations avec de tailles variables par des vecteurs de taille fixe : les MFCC, l'énergie et leurs dérivées premières.

La représentation du signal par les MFCC est très utilisée dans les tâches de reconnaissance du locuteur. Ces coefficients caractérisent bien la forme du spectre et permettent de séparer l'influence de la source du signal vocal de celle du conduit vocal. Cette séparation est rendue possible grâce à un filtre déconvolutif. Chaque vecteur est donné par la concaténation des 12 coefficients, de ces dérivées premières (12 coefficients) ainsi que l'énergie (2 coefficients), extraits de la fenêtre du Hamming toutes les 10 ms.

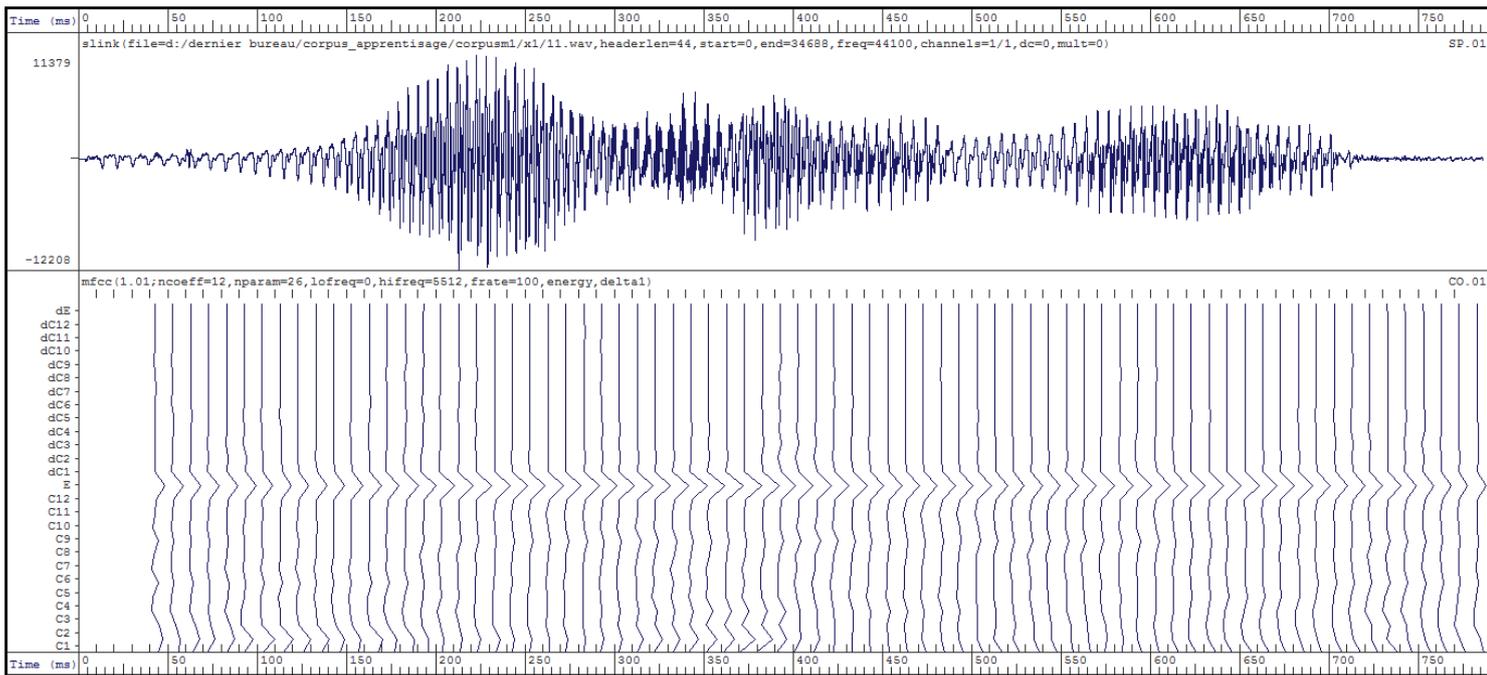
Ces coefficients sont obtenus en utilisant le logiciel Speech Filing System (SFSwin) [41].



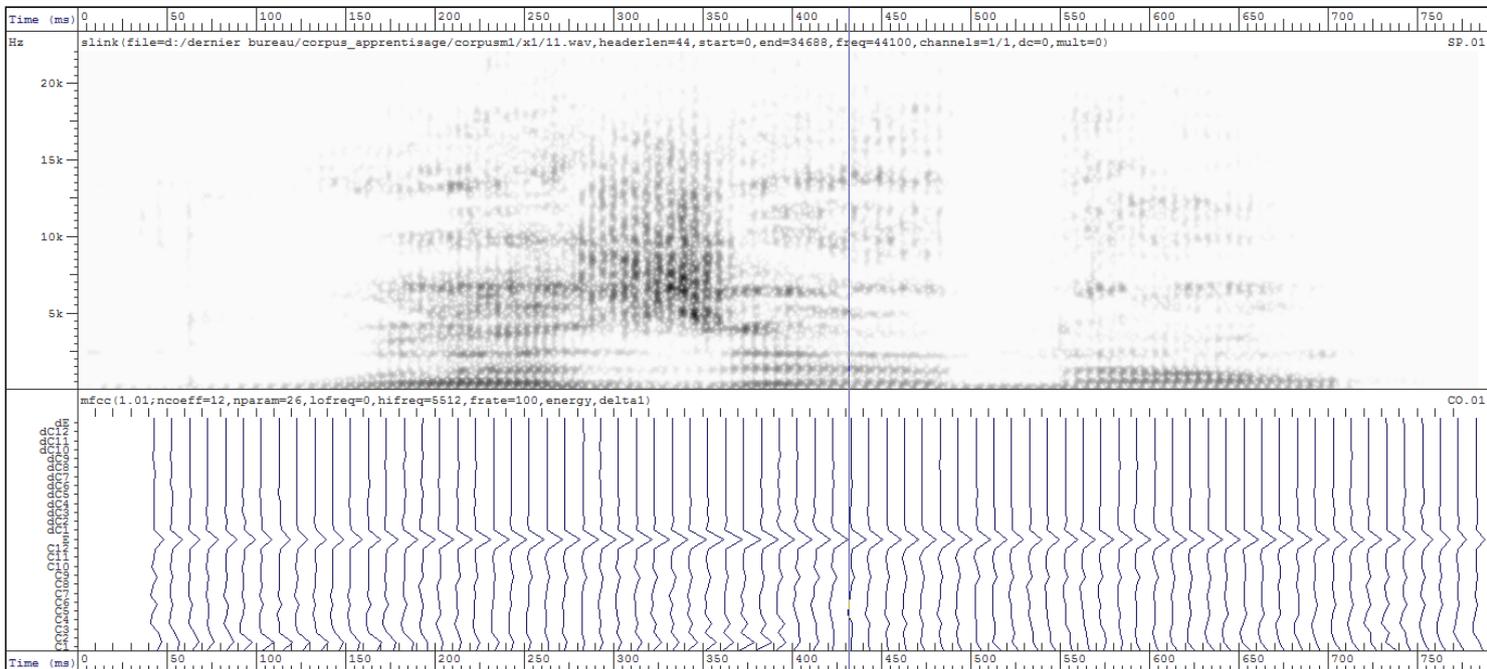
**Figure 4. 1:** Forme d’onde et schématisation des coefficients MFCC, l’énergie et leurs dérivées premières du mot [wʔzʔna] prononcé par locuteur Fe1



**Figure 4. 2:** Forme d’onde, spectrogramme à bande large du mot [wʔzʔna] prononcé par locuteur Fe1 et schématisation des coefficients MFCC, l’énergie et leurs dérivées premières



**Figure 4. 3:** Forme d’onde et schématisation des coefficients MFCC, l’énergie et leurs dérivées premières du mot [wʔzʔna] prononcé par locuteur H1



**Figure 4. 4:** Forme d’onde, spectrogramme à bande large du mot [wʔzʔna] prononcé par locuteur Fe1 et schématisation des coefficients MFCC, l’énergie et leurs dérivées premières

### 4.3. Modèle de mélange de gaussiennes

Nous proposons une représentation des données basées sur une modélisation Multi-Gaussiennes GMM des locuteurs. La distribution des vecteurs de caractéristiques tirée de la parole du locuteur est modélisée par une somme pondérée de  $k$  composantes de mélange,  $X = \{x_t, 1 \leq t \leq T\}$

Pour créer un modèle statistique du locuteur, il est nécessaire d'estimer les paramètres de ce modèle :

$$\lambda = \{w_k; \mu_k; \Sigma_k\} \quad (4.01)$$

Avec  $1 \leq k \leq K$

C'est à dire trouver ceux qui maximisent la fonction de vraisemblance des données d'apprentissage.

$$\lambda_{ML} = \operatorname{argmax}_{\lambda} P(X/\lambda) \quad (4.02)$$

Tel que  $P(X/\lambda)$  est la probabilité de la séquence de l'observation.

#### 4.3.1. Estimation des paramètres par l'algorithme EM

L'algorithme EM se compose de deux paliers comme nous l'avons expliqué (chapitre 2) :

- l'initialisation du modèle ;
- l'optimisation des paramètres du mélange par l'algorithme classique expectation maximisation.

##### 4.3.1.1. Initialisation des paramètres

Nous avons initialisé les paramètres de notre modèle comme suit :

- les moyennes et des  $K$  gaussiennes du modèle par l'algorithme K-means ;
- toutes les matrices de covariance  $\Sigma_k$  à la matrice unité  $I$  ;
- tous les poids des composantes :  $\omega_K = \frac{1}{K}$ .

Itération pour  $i = 1, \dots, N$

#### 4.3.1.2. Phase d'Estimation

Pour tous les vecteurs acoustiques

$$x_t: t = 1, \dots, T$$

Il faut calculer la probabilité  $\gamma_{n,m}$  : probabilité que le vecteur  $x_t$  soit généré par la loi gaussienne  $k$ .

En prenant des valeurs dans l'intervalle  $[0, 1]$ , qui donnent un étiquetage des données (vecteurs acoustiques) en indiquant dans quelle proportion un vecteur  $x_t$  appartient à la gaussienne  $k$ .

#### 4.3.1.3. Phase de Maximisation

La réestimation des paramètres  $w_k^i$ ,  $\mu_k^i$ ,  $\Sigma_k^i$  du modèle de  $\lambda^i$  a été faite à partir de  $\gamma_t(k)$  en utilisant les équations (2.07), (2.08), (2.09) et (2.10).

Tel que  $T$  est la durée de la séquence d'observation, pour le modèle de locuteur  $\gamma_t(k)$  est la probabilité sur le temps  $t$  avec la  $k^{eme}$  composante de mixture.

Dans le cas présent, tous les vecteurs de données participent à la mise à jour du modèle, mais leur participation est proportionnelle à la valeur  $\gamma_t(K)$

Incrémementation de  $i$  et retour à la phase d'estimation.

#### 4.3.1.4. Arrêt de l'algorithme

Les paramètres du modèle sont améliorés à chaque itération, cette estimation sera terminée jusqu'à ce que quelques critères de convergence qui se base sur un seuil soit retrouvés, par conséquent l'estimation est terminée sinon elle est reprise à l'itération suivante.

Dans cet algorithme nous avons besoin de calculer la matrice de covariance à chaque itération et pour passer à l'itération suivante nous sommes obligé de calculer l'inverse de cette matrice pour qu'elle soit non-singulière.

Pour éviter ces fausses singularités, nous appliquons la contrainte suivante ; dans toutes les matrices de covariances. Nous remplacerons les valeurs inférieures à une certaine valeur minimale par cette dernière. Pour une composante gaussienne arbitraire. Après chaque itération EM nous appliquons la contrainte suivante :

$$\begin{cases} \Sigma_i^2 & \text{si } \Sigma_i^2 > \Sigma_{min}^2 \\ \Sigma_{min}^2 & \text{si } \Sigma_i^2 < \Sigma_{min}^2 \end{cases} \quad (4.03)$$

Dans ce travail la valeur minimale a été fixée à  $10^{-4}$ .

#### 4.4. Optimisation des paramètres par AG

Nous avons implémenté notre modèle ou algorithme génétique sous Matlab version 7.10 ; nous allons introduire un apprentissage des GMM basé sur le AG-ML .l' AG fournit la capacité de la recherche globale pour l'optimisation du problème. L'apprentissage de GMM peut s'échapper de la supposition initiale et trouver la solution optimale si nous appliquons l'AG pour le processus de l'apprentissage.

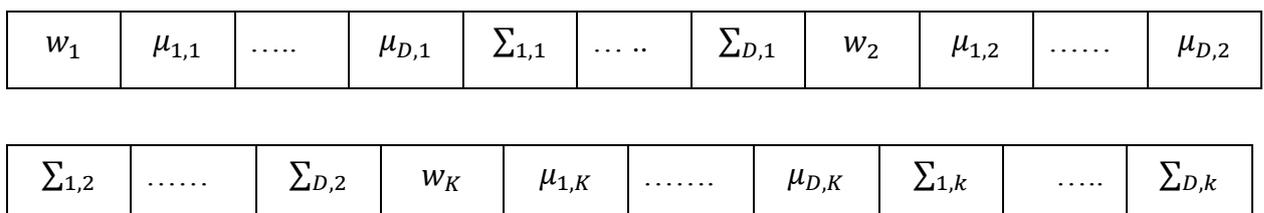
##### 4.4.1. Mécanisme de l'encodage

Dans ce travail, nous avons choisi le codage réel parce que les études ont montré que la représentation réelle aboutit souvent à une meilleure précision [42].

Chaque modèle est défini comme étant le chromosome qui contient les MFCC de plusieurs segments successifs.

Chaque segment présente une composante des GMM dont la taille est égale à  $k(2D + 1)$  qui contient le poids, le vecteur de moyennes et la matrice de covariance.

La structure de chromosome est illustrée (Figure 4. 5).



**Figure 4. 5:** Représentation du chromosome dans l'apprentissage AG-ML

tel que  $w_k$ ,  $\mu_{n,k}$ ,  $\Sigma_{n,k}$  sont le poids,  $n^{\text{ème}}$  moyenne et  $n^{\text{ème}}$  covariance diagonale respectivement de la composante  $k$  de mélange.

Nous avons fait appel à la fonction "ga" dans logiciel Matlab en tenant compte des contraintes qu'elles soient respectées dans notre chromosome qui va nous permettre d'exécuter notre algorithme correctement :

$x = \text{ga}(\text{fitnessfcn}, \text{nvars}, \text{LB}, \text{UB})$

$x$  : le chromosome qui a une meilleure valeur de la fonction de fitness .

fitnessfcn : la fonction de fitness  $f_v$  .

nvars : nombre de variable.

Pour vérifier les conditions suivantes :

- Non-singularité de la matrice de covariance.

- $\sum_{k=1}^K w_k = 1$  . (4.04)

- $0 \leq w_k \leq 1$  . (4.05)

Nous avons introduit dans la fonction 'ga' la contrainte suivante :

$$LB \leq x \leq UB$$
 (4.06)

Tel que :

LB : la bande inférieure de chromosome  $x$  ;

UB : la bande supérieure de chromosome  $x$ .

#### 4.4.2. Initialisation de la population

Il existe deux méthodes utilisées pour initialiser la population :

- l'initialisation des chromosomes d'une façon aléatoire ;
- l'initialisation des chromosomes par l'algorithme K means.

Nous avons choisi la deuxième méthode parce que les résultats expérimentaux prouvent que la deuxième stratégie est meilleure [42].

### 4.4.3. Fonction d'Evaluation

La fonction de fitness ou fonction d'évaluation de l'individu est utilisée pour sélectionner et reproduire les meilleurs individus de la population.

Chaque chromosome apporte une solution potentielle au problème à résoudre. Néanmoins, ces solutions n'ont pas toutes le même degré de pertinence.

Dans l'apprentissage par AG-ML, les valeurs correspondant à des résultats de la fonction objective. Cette dernière est définie comme une moyenne des logarithmes des probabilités des séquences d'observation  $X_1, X_2, \dots, X_C$  généralisée par le  $n^{\text{ème}}$  mélange.

$$fv_n = \left( \sum_{i=1}^C \log \left( p \left( X_i / \lambda_n \right) \right) \right) / C \quad (4.07)$$

Tel que  $fv_n$  est la valeur de  $n^{\text{ème}}$  chromosome. Cette fonction objective considère juste la fonction de vraisemblance de locuteur unique.

La probabilité  $p(X_i/\lambda_n)$  est calculée à partir les Eqs (2.01), (2.02).

## 4.5. Les opérateurs génétiques

Dans le processus d'apprentissage, trois opérateurs génétiques sont désignés :

- le croisement ;
- la mutation ;
- Le remplacement.

Sans aucun opérateur heuristique, la vitesse de convergence du cycle AG est lente.

Nous avons utilisé la ré-estimation ML comme un opérateur heuristique.

### 4.5.1. Le croisement

Dans cet opérateur, nous avons sélectionné deux parents du groupe de la population, basés sur le mécanisme de la roulette.

Cinq mélanges sont aléatoirement sélectionnés d'un parent et ils sont remplacés par les mélanges correspondants dans un autre parent. Les résultats expérimentaux prouvent que le 5<sup>ème</sup> est de meilleur choix, et le premier parent est réservé comme un offspring.

#### 4.5.2. La mutation

Cette étape permet d'introduire une variation aléatoire dans le génotype de l'individu (ici les enfants). Cet opérateur permet donc d'explorer de nouvelles régions de l'espace de recherche, diminuant ainsi les risques de converger vers des minima locaux. Chaque gène est donc susceptible de changer selon une probabilité donnée, il est aussi possible de ne choisir aléatoirement qu'une partie des gènes.

#### 4.5.3. Le remplacement

A chaque génération, tout chromosome dans l'offspring est ré-estimé cinq fois avec la formule ML. Eqs. (2.08), (2.09) et (2.10).

Dans l'étape du remplacement, nous avons généré la nouvelle population, en remplaçant les mauvais chromosomes dans la population actuelle avec la sous-population (groupe de l'offspring).

Les opérateurs de l'AG peuvent détruire les conditions contraintes du modèle thresholding doit être conduit la ré-estimation ML. Et la somme du poids de la mixture est normalisée à 1.0. Finalement, les paramètres du contrôle de l'apprentissage AG (Tableau 4.2).

**Tableau 4. 2:** Les paramètres du contrôle de l'apprentissage AG-ML

Paramètres	Valeurs
Taille de la population	30
Nombre de chromosomes d'offspring	5
Taux du mélange de la mixture	0.08
Taux de la mutation	0.01
Nombre maximum des générations	50

#### 4.6. Critère d'arrêt de l'algorithme

Arrêt de l'AG :

- après un certain nombre de générations ;
- lorsque le meilleur individu n'a pas été amélioré depuis un certain nombre de générations.

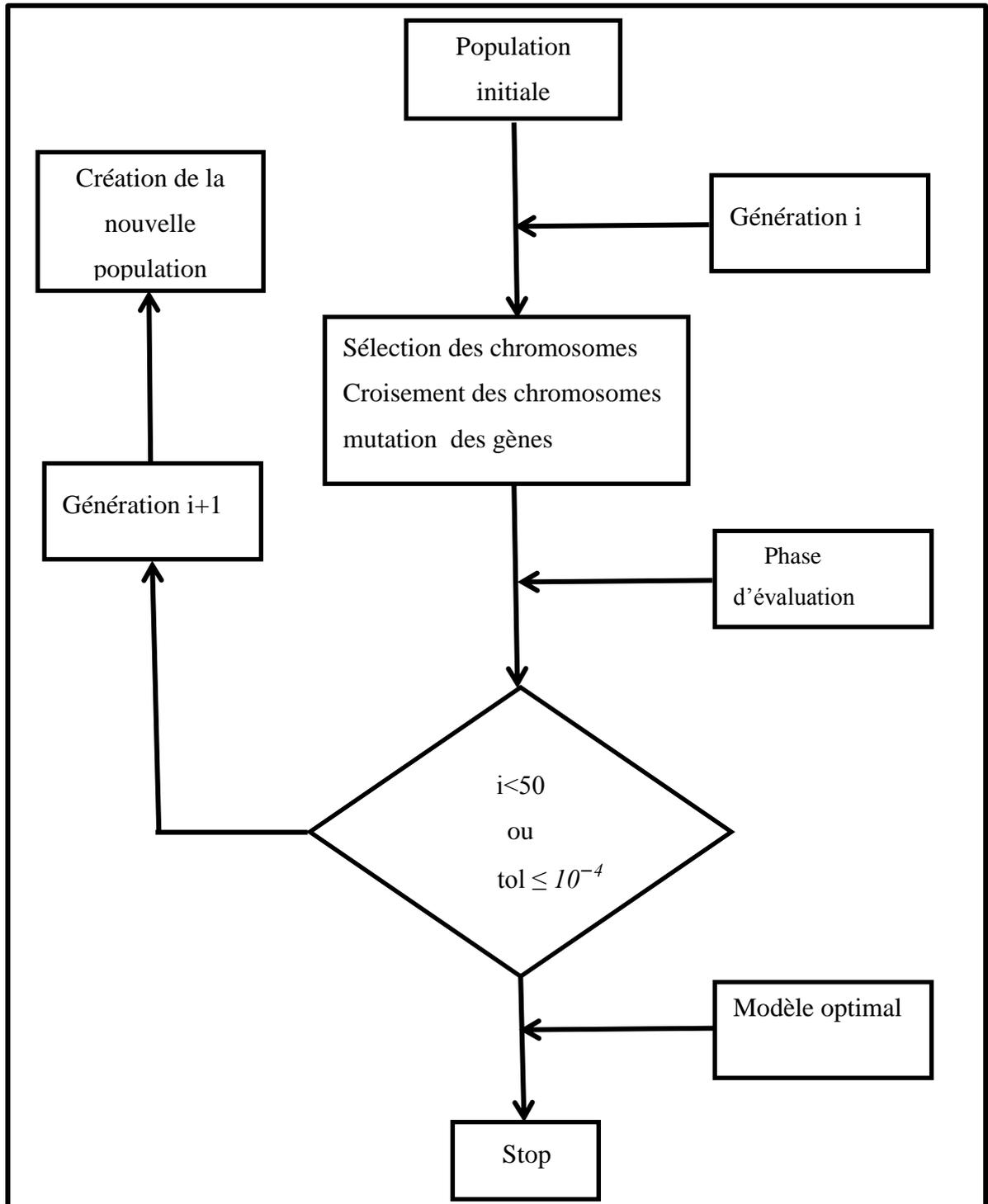


Figure 4. 6: Organigramme de l'AG

Les résultats de la méthode traditionnelle ML sont utilisés comme une base pour la comparaison. A l'apprentissage ML, les ré-estimations seront terminées quand l'augmentation de la probabilité de log moyen de l'apprentissage de données entre deux itérations successives est inférieure de  $10^{-4}$ .

Nous comparons la remise en forme finale de l'hybride AG-ML et ML simple après 50 générations desquels, on peut voir que la valeur de  $fv_n$  pour chaque locuteur est augmentée à travers l'opérateur heuristique (Figure 4.7).

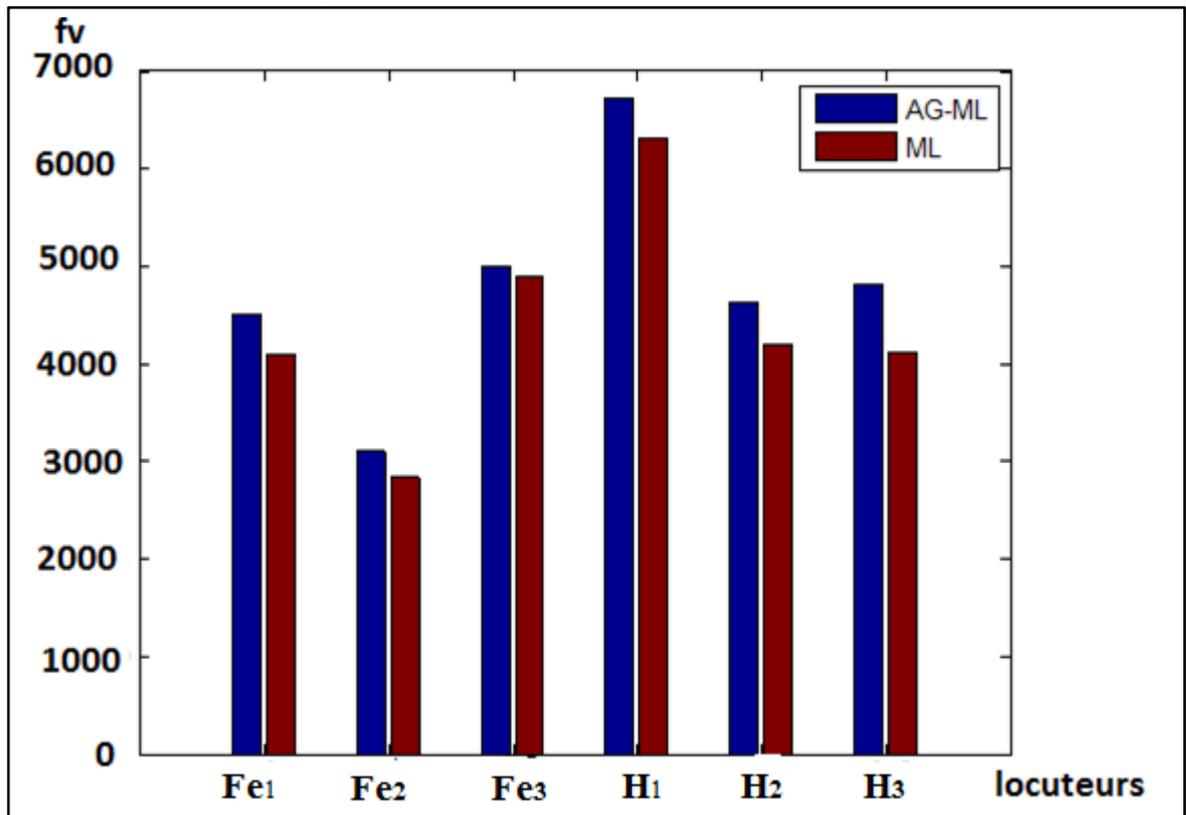


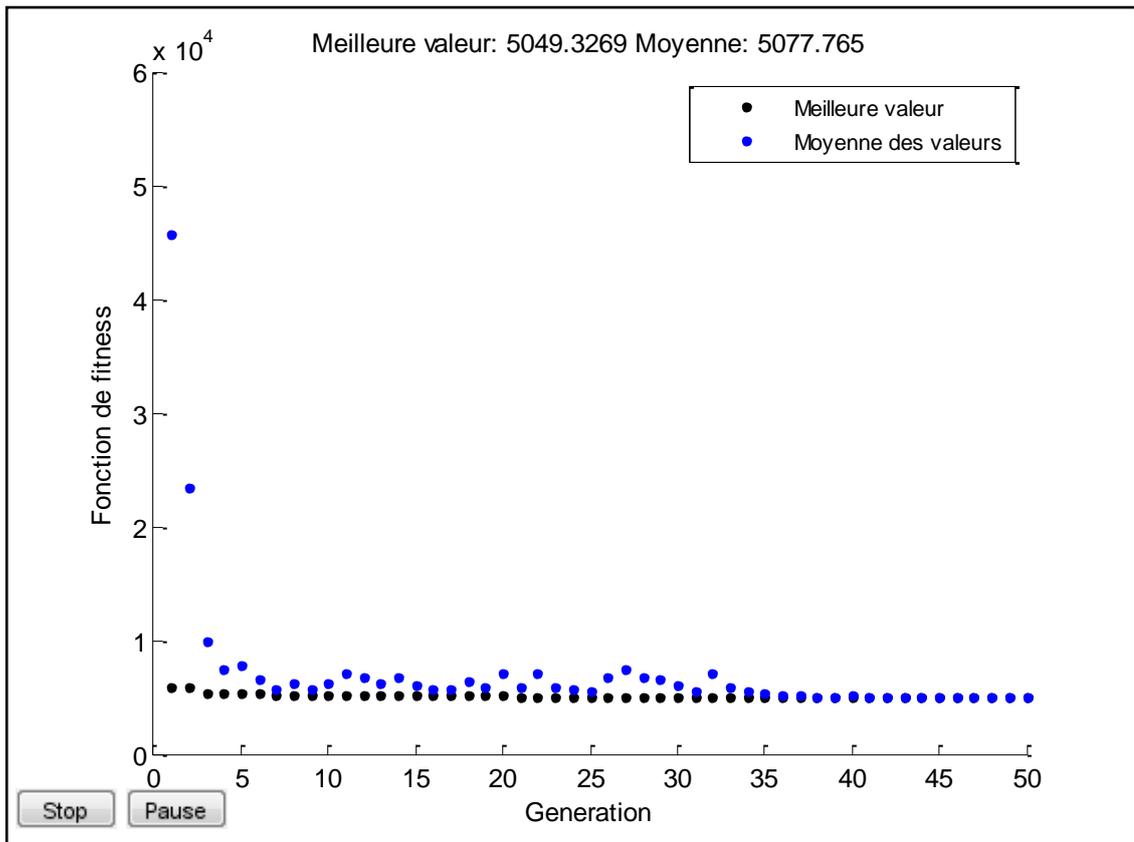
Figure 4. 7: comparaison entre l'apprentissage par AG-ML et ML avec k=32

Pour comparer la performance de ML et l'apprentissage AG-ML, les GMM sont entraînés respectivement avec 16,32 et 64 mélanges (Tableau 4.3).

**Tableau 4. 3 :** Les valeurs finales de la fonction  $fv_n$

	K=16		K=32		K=64	
	ML	AG-ML	ML	AG-ML	ML	AG-ML
Fe1	5326,7	5058,6	4253,5	3862,2	3883,7	3611
Fe2	4312,2	3994,9	3019,1	2613,7	2080,3	1914,8
Fe3	6022,8	5721,5	4887,5	4645,8	4491,6	4223,3
H1	7162,1	7111,1	6346,1	6149,1	6604,5	6031,2
H2	5461,4	5176	4369,1	4164,2	3924,3	3695,5
H3	5321	4995,8	4244,1	4087,4	4026,9	3792,3

La figure 4. 8 illustre l'évolution de valeurs de la fonction de fitness en fonction de générations.

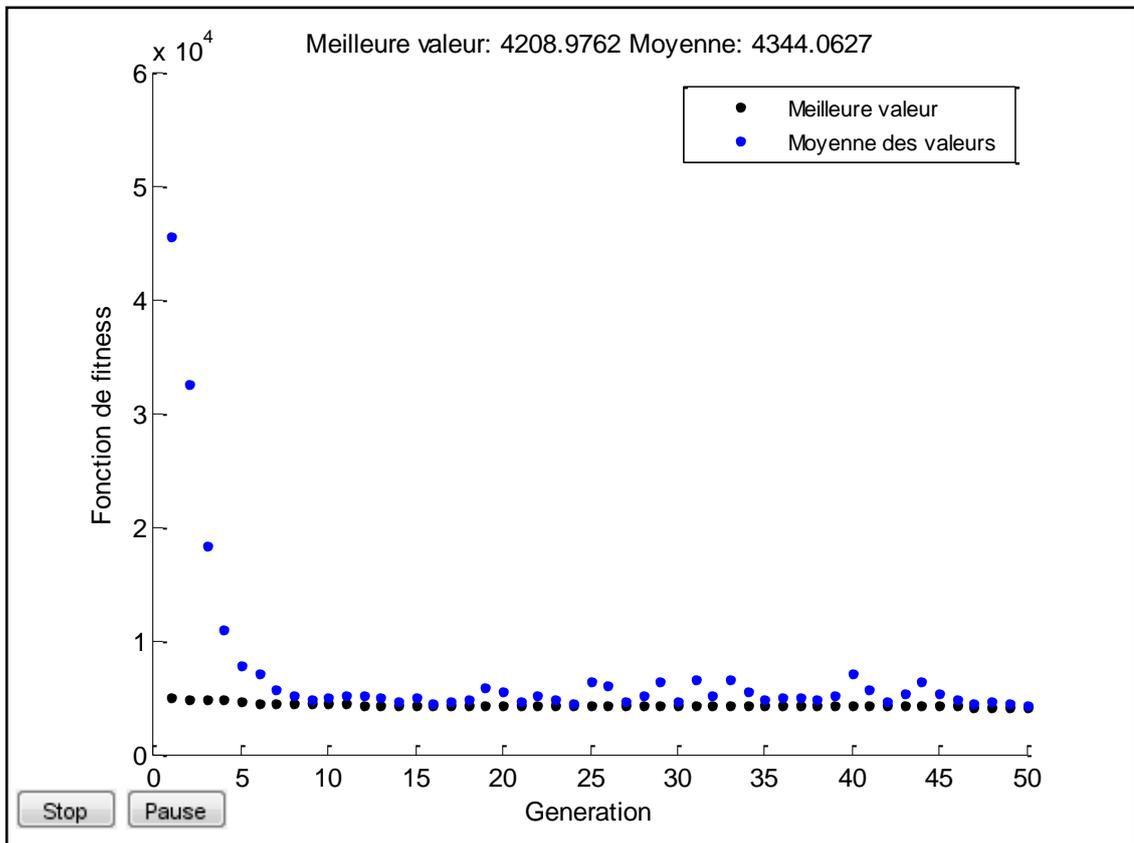


**Figure 4. 8 :** Convergence de l'AG vers l'individu qui représente l'optimum global au fil des générations K =16

Les points bleus présentent la moyenne des valeurs de fonction objective de chaque génération lors les points noirs présentent le chromosome qui a une meilleure valeur dans la génération entière. Nous constatons qu'il y a une convergence entre les points bleus et les points noirs dans les dernières générations à cause du remplacement des mauvais chromosomes par les bons.

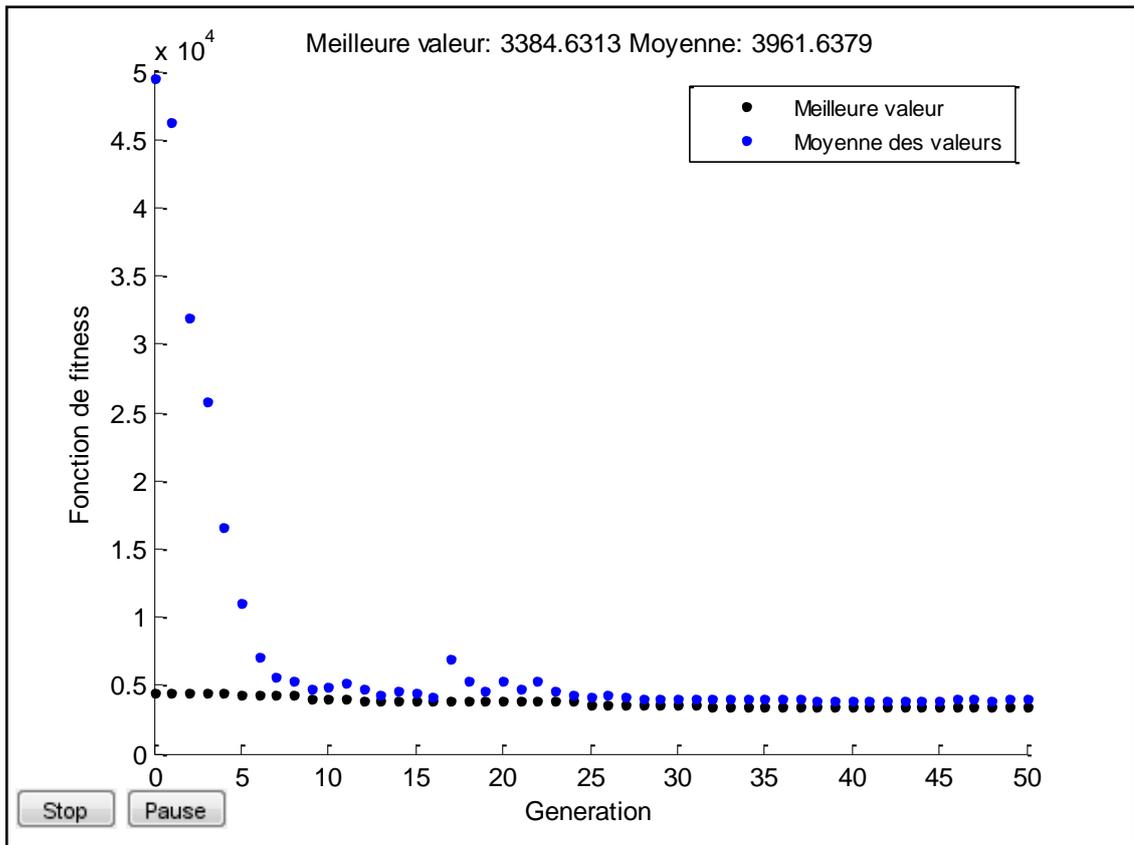
La petite divergence entre ces points indique l'intervention de l'opérateur de mutation pour but d'enrichir la diversité de la population.

Nous avons augmenté le nombre de composantes de GMM a 32, nous constatons que la valeur de  $f v_n$  de meilleur chromosome dans toutes les générations est mieux que celle de GMM a 16 composantes (Figure 4. 9).



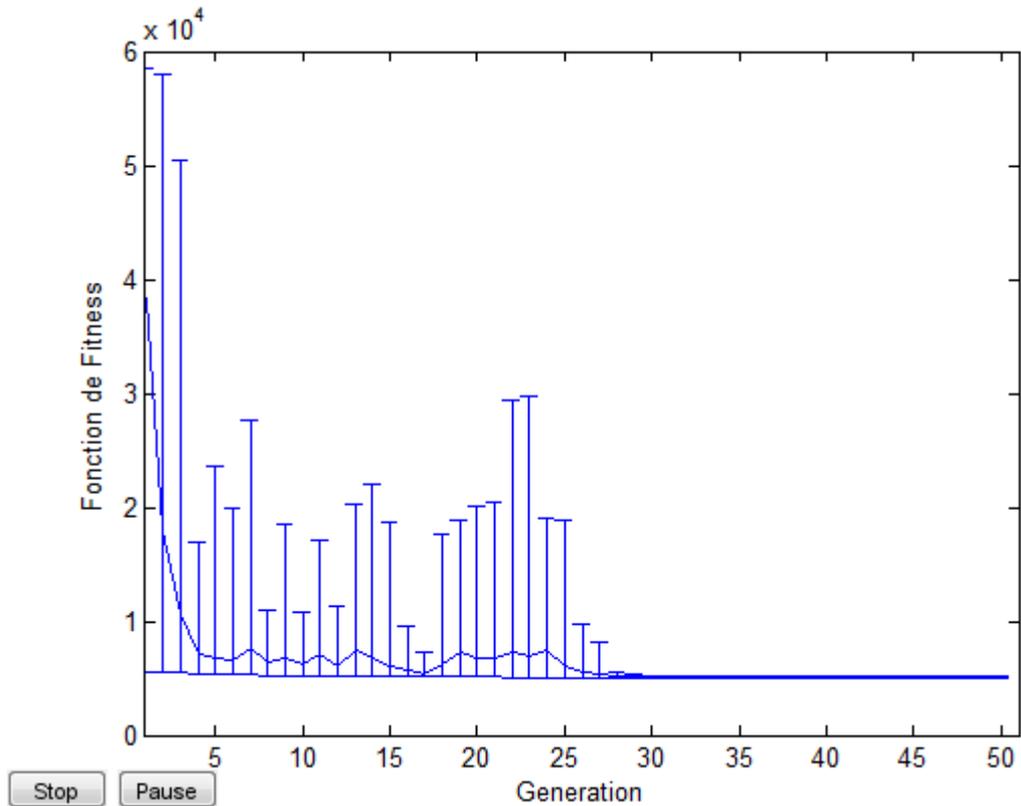
**Figure 4. 9:** Convergence de l'AG vers l'individu qui représente l'optimum global au fil des générations K =32

Les résultats obtenus lorsque nous avons utilisé 32 gaussiennes prouvent qu'il y a un taux plus élevé d'homogénéité des données avec le nombre des composantes des GMM (Figure 4. 9).



**Figure 4. 10:** Convergence de l'AG vers l'individu qui représente l'optimum global au fil des générations  $K = 64$

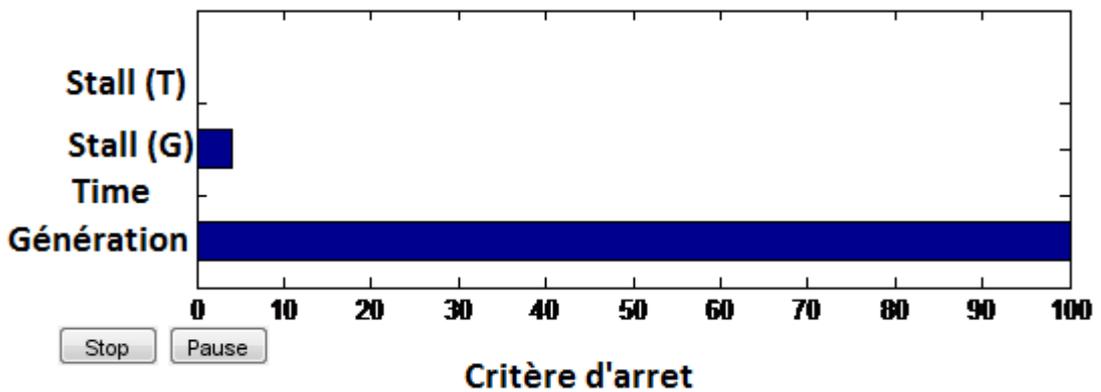
Nous pouvons aussi représenter  $f v_n$  de meilleur, mauvais chromosome et la moyenne des individus de chaque génération (Figure 4. 11).



**Figure 4. 11:** Meilleur, mauvais chromosome et la moyenne des individus de chaque génération

Les lignes verticales montrent la plus petite et la plus grande valeur de  $f v_n$  de remise en forme, La courbe en milieu présente la distance moyenne entre les individus, ce qui tend vers 0 à la dernière génération.

Le critère d'arrêt du AG est choisi comme étant le nombre de générations maximum ou itérations qui a été fixé à 50 (Figure 4. 12).



**Figure 4. 12:** Critère d'arrêt de l'AG selon le nombre maximum de générations

Tableau 3 donne les résultats de l'exactitude de remise en forme et de vérification (Tableau 4.4).

**Tableau 4. 4 :** Taux de la phase de vérification

Locuteur	k=16	k=32	k=64
Fe1	87.33	87.34	98.66
Fe2	96	97.33	98 ,01
Fe3	78.66	82.66	94.66
H1	70	75.33	84
H2	78	85.33	90.66
H3	92	92	89.33

L'estimation des paramètres de GMM par AG-ML a des meilleures valeurs que par la méthode ML simple, ce qui implique que les GMM entraînés par notre AG-ML sont plus optimisés que les GMM entraînés par la méthode ML.

#### 4.7. Conclusion

Dans cette étude, nous avons développé un algorithme hybride AG-ML pour l'apprentissage du locuteur basée sur les GMM, qui permet à ce processus de fournir la capacité de recherche globale pour trouver un GMM plus optimisé. En utilisant la re-estimation ML comme opérateur heuristique, les résultats obtenus prouvent qu'AG a une meilleure performance dans la phase de VAL.

# **Conclusions Générales et Perspectives**

## Conclusions Générales et Perspectives

Ce travail de mémoire s'inscrit dans le domaine la Vérification Automatique du Locuteur. Pour cela nous avons utilisé un corpus composé de mots en l'AS, prononcés par des locuteurs arabophones et enregistrés dans un milieu ambiant. Il s'agit donc de construire un système hybride AG-ML pour la vérification du locuteur en mode dépendant du texte permettant d'étudier la similarité des caractéristiques entre les différents locuteurs, afin d'apporter une décision de vérification. En effet, Nous proposons des nouvelles représentations des données basées sur une modélisation multi-gaussiennes "GMM" des locuteurs, en appliquant les AG.

La méthode de GMM a une capacité remarquable à modéliser les données irrégulières. Ces caractéristiques permettent à ce mélange d'être très convenable pour les GMM afin d'avoir une bonne estimation de la variabilité acoustique du locuteur (interlocuteur, intralocuteur, émotion, état de santé, etc.).

Un AG opère sur une population d'individus. Dans notre application l'individu choisi est un chromosome qui est constitué de plusieurs segments chacun d'eux contient une composante de GMM, définie par leurs paramètres : le poids, le vecteur des moyennes et la matrice de covariance dont les gènes contiennent les MFCC.

La taille de chaque chromosome est égale à  $K$  (nombre de composantes de GMM).

L'algorithme mis en œuvre est constitué des opérateurs : sélection, croisement, mutation et remplacement. Ces opérateurs sont appliqués à la population courante, produisant une nouvelle génération en évaluant chaque chromosome par la fonction de fitness.

Nous avons utilisé la ré-estimation ML comme un opérateur heuristique. Les résultats de la méthode traditionnelle ML sont utilisés comme une base pour la comparaison avec ceux de l'AG-ML.

Nous comparons la remise en forme finale de l'AG-ML et ML simple sous l'opérateur de la ré-estimation ML après 50 générations.

Nos résultats ont montré que la méthode hybride AG-ML proposée est meilleure que celle de l'estimation traditionnelle ML dans la modélisation du locuteur basée sur les GMM.

Les taux de reconnaissance sont égaux à 98%, ceux-ci montrent que la méthode hybride AG-ML proposée est meilleure que celle de l'estimation traditionnelle ML dans la modélisation du locuteur basée sur le GMM.

Dans le future :

- l'application de l'algorithme d'AG-ML dans une Base de Données qui contient un grand nombre de locuteurs ;
- nous étudierons la possibilité de l'application des autres discriminatives comme la MMD pour l'apprentissage par AG afin d'améliorer la performance de la vérification automatique du locuteur.

# Références Bibliographiques

## Références Bibliographiques

- [1] David Mercier, Renaud Séguier, Utilisation des STANN en audio illustration en reconnaissance de chiffre, Journées Valgo, France, 2001.  
[www.rennes.supelec.fr/ren/rd/scee/ftp/articles/mercier\\_jv2001.pdf](http://www.rennes.supelec.fr/ren/rd/scee/ftp/articles/mercier_jv2001.pdf)
- [2] C. Charbuillet, B. Gas, M. Chetouani & J.L. Zarader, Application d'un algorithme génétique à la synthèse d'un prétraitement non linéaire pour la segmentation et le regroupement du locuteur, Actes des XXVI<sup>es</sup> journées d'études sur la parole, Dinard, pp.227-230, juin 2006.
- [3] C. Charbuillet, B. Gas, M. Chetouani & J.L. Zarader, Combinaison de codeurs par algorithme génétique : Application à la vérification du locuteur, Colloque GRETSI, Troyes, France, pp.177-180, 11-14 septembre 2007.
- [4] Mark D. Skowronski & John G. Harris, Improving the Filter Bank of a Classic Speech Feature Extraction Algorithm, IEEE Intl Symposium on Circuits and Systems, Bangkok, ISBN: 0-7803-7761-31, vol IV, pp 281-284, Bangkok, Thailand, May 25 - 28, 2003.
- [5] A. Alami Masmoudi, A. Benabbou, M. Harti, Vérification du Locuteur en Lange Arabe à l'Aide des Réseaux de neurones, Proceedings des journées Scientifiques en Technologie de l'Information et de la Communication JOSTIC, pp. 90-92, Rabat, Maroc, 2et 3 Novembre 2008.
- [6] M. Aissou , Application des Algorithmes Génétiques au décodage acoustico-phonétique de la parole en arabe standard, Thèse de Doctorat, Ecole Nationale Polytechniques , Alger, Algérie, 2008.
- [7] [http://interstices.info/jcms/c\\_9758/de-la-reconnaissance-automatique-du-locuteur-a-la-signature-vocale](http://interstices.info/jcms/c_9758/de-la-reconnaissance-automatique-du-locuteur-a-la-signature-vocale)
- [8] S. Mesfar, Analyse Morpho-syntaxique Automatique et Reconnaissance des entités nommées en Arabe Standard, Thèse de Doctorat, Université de Franche-comte , 2008.
- [9] M. Kabache & M. Guerti, Application des Réseaux de Neurones á la Reconnaissance des Phonèmes Spécifiques á l'Arabe Standard, 3<sup>th</sup> International Conference, SETIT, Sfax, Tunisie, 27-31 Mars 2005.
- [10] A. Chentir, Etude de la Microprosodie en vue de la Synthèse de la parole en Arabe Standard, Thèse de Doctorat, Ecole Nationale Polytechnique, Alger, Algérie, 2009.

- [11] Y. Mami, Reconnaissance de Locuteurs par localisation dans un espace de locuteurs de référence, Thèse de Doctorat, Ecole Nationale Supérieure de télécommunication, spécialité : signal et communication, Paris, France, 2003.
- [12] J. Bonastre, l'Authentification Biométrique Vocale mars 2005.  
[http://pari-stic.labri.fr/TUTORIAL/tutorial\\_BIO\\_3\\_PARISTIC\\_05.pdf](http://pari-stic.labri.fr/TUTORIAL/tutorial_BIO_3_PARISTIC_05.pdf)
- [13] T. Bänziger, G. Klasmeyer, T. Johnstone, T. Kamceva & K. R. Scherer, Améliorer les systèmes de Vérification Automatique du Locuteur en intégrant la variabilité émotionnelle : Méthodes et premières données, *XXIII<sup>èmes</sup> Journées d'Etude sur la Parole*, Aussois, France, 19-23 juin 2000.
- [14] H. Ezzaidi, Discrimination parole/musique et étude de nouveaux paramètres et modèles pour un système d'identification du locuteur dans le contexte de conférences téléphoniques, Thèse de Doctorat, Université du Québec, Chicoutimi, Canada, 2002.
- [15] F. Debbeche, N. Ghoualmi-zine, Système Acoustico-Anatomique pour l'Identification des Locuteurs par Localisation dans un Espace de Locuteurs de Référence, *INFØDays'2008*, pp. 12-15, 15-16 Avril, Chlef, Algérie, 2008.
- [16] N. Scheffer, Structuration de l'Espace Acoustique par le Modèle Générique pour la Vérification du Locuteur, Thèse de Doctorat, Université d'Avignon et des Pays de Vaucluse, France, 2006.
- [17] J. Kharroubi, Etude de Techniques de Classement 'machines à vecteurs supports 'pour la Vérification Automatique du Locuteur, Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications, Algérie, 2002.
- [18] B. Tounsi, Inférence d'identité dans le domaine forensique en utilisant un système de reconnaissance automatique du locuteur adapté au dialecte algérien, Thèse de Magister, Institut National de Formation en Informatique (i.n.i) oued-smar, Alger, Algérie, 2008.
- [19] T. Matsui, S. Furui, Comparaison of tex-independent speaker recognition methods using VQ- distortion and discrete /contiuous hmms, *IEE international conference on acoustics, speech and signal processing, ICASSP*, vol 2, pp. 157-160, San Francisco, CA, USA, 1992.
- [20] J. Oglesby J.S. Mason, Speaker Recognition with neural classifier, *First IEEE International Conference*, pp. 306-309, 1989.
- [21] Ke. Chen, Towards better making a decision in Speaker Verification, *Pattern Recognition*, N° 36, pp. 329 – 346, 2003.

- [22] E. Tazi, A. Benabbou & M. Harti, Design of an Automatic Speaker Recognition System Based on Adapted MFCC and GMM Methods for Arabic Speech, IJCSNS International Journal of Computer Science and Network Security, VOL.10 N°.1, pp. 45-50, Fès, Maroc, 2010.
- [23] S. Zribi boujelbene, D. Ben ayed mezghani. & N. Ellouze, Identification du Locuteur par Système Hybride (GMM-SMO), 5<sup>th</sup> International Conference, SETIT, Sfax, Tunisie, 2009.
- [24] A. Larcher, Modèles acoustiques à structure temporelle Renforcée pour la vérification du locuteur embarquée, Thèse de Doctorat, Université d'Avignon et des Pays de Vaucluse, France, 2009.
- [25] A. Preti, Surveillance de réseaux professionnels de communication par la reconnaissance du locuteur, Thèse de Doctorat, l'Université d'Avignon et des Pays de Vaucluse, France, 2008.
- [26] A. Elfazziki, A. Nejeoui, M.Sadgal, Une Approche multi-agents pour la modélisation et l'optimisation des Systèmes de gestion de transport maritime, RIST, Vol.15 N° 01-02, Algérie, 2005.
- [27] A. Damak, J. Zrida & M. Annabi, Application de l'Algorithme Génétique pour l'homogénéisation de la qualité d'un mélange de céréales, 3<sup>th</sup> International Conference, SETIT, Tunisi, 2005.
- [28] Z. Michalewicz, Genetic Algorithms + Data structures = Evolution Programs , 3<sup>th</sup> Revised and Extended Edition, USA, pp. 17-18, 1999.
- [29] D. Coulot, A. Pollet & X. Collilieux, Application des Algorithmes Génétiques à la recherche de sous-réseaux de stations de télémétrie laser, Bulletin d'information scientifique et technique de l'IGN, N° 77, France, pp. 89-93, 2011.
- [30] <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.102.3787&rep=rep1&type=pdf>
- [31] A. Soukkou, A. khellaf & S. Leulmi, Supervision neuro-floue a apprentissage genetique d'un pid robuste, Sciences & Technologie, N°23, pp. 95-106, 2005.
- [32] Uday K. Chakraborty & Cezary Z. Janikow, An analysis of Gray versus binary encoding in genetic search, Information Sciences, N°.156, pp.253–269, 2003.
- [33] <http://magnin.plil.net/spip.php?article44>

- [34] N, Bel hadj Ali, Une approche globale d'optimisation des structures métalliques avec les Algorithmes Génétiques, *XXI<sup>èmes</sup> rencontres de genie civil*, universite de savoie, pp. 199-206, france 2003.
- [35] C. Reeves, J. Rowe, genetic algorithms principles and perspectives a guide to GA theory, kluwer Academic Publishers, pp. 25-26, USA, 2003 .
- [36] N.lassouaoui , L.Hamami, N .Nouali, Les Algorithmes Génétiques application à la segmentation des images, Laboratoire signal & communications, Ecole nationale polytechnique, RIST Vol.14 N°02, pp. 27-56, Algérie, 2004.
- [37] L.Oulladji, Al . Janka , J .Désidéri &A. Dervieux, Optimisation Aérodynamique par Algorithmes Génétiques hybrides: application à la réduction d'un critère de bang sonique, Rapport de recherche, INRIA, N° 4884, France, 2003.
- [38] W. Bahloul, M. Ouali, M. Chaabene & M. Kamoun, Identification des Paramètres Dynamiques de la Machine Synchrone par Algorithme Génétique, International Renewable Energy Congress, pp.172-177, Sousse, Tunisie, 2009.
- [39] I. Harbaoui Dridi, R. Kammarti, P. Borne & M. Ksouri, Un Algorithme génétique pour le problème de ramassage et de livraison avec fenêtres de temps à plusieurs véhicules, CIFA, hal-00521139, vol 1, Roumanie, 2010.
- [40] M. Mitchell, An introduction to genetic algorithms, A Bradford Book The MIT Press, Cambridge, UK, 1998.
- [41] <http://www.phon.ucl.ac.uk/resource/sfs/SFS@phon.ucl.ac.uk>
- [42] Q.Y. Hong & S. Kwong, A genetic classification method for speaker recognition, Elsevier Ltd, N°.18, pp. 13–19, 2005.