

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

**Ecole Nationale Polytechnique**



المدرسة الوطنية المتعددة التقنيات  
Ecole Nationale Polytechnique

**Département d'Electronique**

**Laboratoire Signal & Communications**

**Thèse de Doctorat**

Présentée par :

**AISSIOU Mohamed**

**Thème**

**Application des Algorithmes Génétiques au  
Décodage Acoustico- Phonétique de la parole en  
Arabe Standard**

Devant le Jury :

<b>A.ZERGUERRAS</b>	<b>Professeur</b>	<b>ENP</b>	<b>Président</b>
<b>M.GUERTI</b>	<b>Professeur</b>	<b>ENP</b>	<b>Rapporteur</b>
<b>A.GUESSOUM</b>	<b>Professeur</b>	<b>USD-Blida</b>	<b>Examineur</b>
<b>M.DEBYECHE</b>	<b>Maître de Conférence</b>	<b>USTHB</b>	<b>Examineur</b>
<b>H. SAYOUD</b>	<b>Maître de Conférence</b>	<b>USTHB</b>	<b>Examineur</b>
<b>F.Z. ABDI</b>	<b>Orthophoniste</b>	<b>Ecole des sourds El-Mohammadia Alger</b>	<b>Invitée</b>

**30 Juin 2008**

# *Dédicaces*

*Je dédie cet humble travail à mes Parents,  
épouse et mes deux enfants, amis,  
professeurs et collègues, à qui je suis  
sincèrement reconnaissant pour tout ce  
qu'ils ont fait pour moi.*

# **Au Nom du Bon Dieu le Miséricordieux le tout Miséricordieux**

## Remerciements

*Ce travail n'aurait sans doute jamais eu lieu sans le concours volontaire ou involontaire de tous ceux qui m'ont formé, soutenu, encouragé ou aidé.*

*Je tiens tout spécialement à :*

*Remercier le Bon Dieu pour tout ;*

*remercier vivement le Professeur M. GUERTI, enseignante à l'ENP, pour avoir bien voulu me proposer un sujet, m'encadrer, et surtout pour sa précieuse collaboration, sa grande disponibilité et l'aide constante qu'elle m'a apportée durant mes travaux ;*

*exprimer ma profonde gratitude à Monsieur Ahmed Zerguerras, Professeur à l'ENP, pour l'honneur qu'il me fait en acceptant de présider le jury de cette thèse ;*

*remercier vivement Monsieur A. GUESSOUM, Professeur à l'USD de Blida, pour m'avoir fait le plaisir d'accepter de participer à mon jury de thèse en tant qu'examineur ;*

*remercier Messieurs M. DEBYECHE et H. SAYOUD, Maîtres de conférence à l'USTHB, pour avoir accepté de participer à mon jury de thèse en tant qu'examineurs ;*

*remercier Madame Abdi Fatma Zohra, orthophoniste à l'Ecole des Sourds d'El-Mohammadia, El-harrach, pour avoir acceptée de participer en tant qu'Invitée.*

*Que ceux qui m'ont aidé de près ou de loin trouvent ici l'expression de mes profonds remerciements.*

إن هذا العمل يهدف إلى تطبيق الخوارزميات الوراثة في مرحلة حل التشفير أو للاتشفير اللفظي الصوتي التي تعتبر أول زريدة الاستكشاف الآلي للكلام. كل المعالجات اللاحقة مرتبطة ارتباطاً قوياً بنوعية النسيج اللفظي إن أدق مرحلة تتمثل في تحقيق نموذج وراثي قادر على التصنيف الآلي لأشعة الملفات الصوتية للكلام، حيث كان لزاماً علينا إيجاد أحسن طريقة لتطبيق الخوارزميات الوراثة للاستكشاف الآلي للصوتيات العربية وتحديد المستوى الذي تتم فيه مرحلة القولية الوراثة للمسألة. المدون المختار متكوّن من مجموعة من جمل مسجلة من طرف متكلم باللغة العربية. تماشياً مع ذلك، قمنا بدراسة عامة للخوارزميات الوراثة ليفهم سيرها وبالتالي ضمان حسن استعمالها في مجال استكشاف الصوتيات.

للحصول على الأشعة الصوتية للكلام قمنا بتحليل صوتي لإشارة الكلام باستعمال الطريقة الكبسترية معتمدين في ذلك على سلم ميل. و بإدخال ثابت الطاقة وكذا المشتق الأول والثاني للمعامل الكبستري. إن الدالة التكييفية التي يجب أن تكون أكثر تمييز ممكن تساوي معيار دالة البعد. النتائج المتحصل عليها مشجعة وقدرت نسبة التصنيف العام ب 81,17%. ومن أجل تحسين فعالية الخوارزميات الوراثة، استعملنا خصية التوازي بتحقيق نموذج وراثي منعدم الهجرة.

### الكلمات المفاتيح :

الاتشفير اللفظي الصوتي، الاستكشاف الآلي للكلام، التحليل الكبستري، الخوارزميات الوراثة، التصنيف الصوتي، صوتيات اللغة العربية.

### Résumé :

Dans ce travail de thèse, nous avons appliqué les Algorithmes Génétiques (AGs) au Décodage Acoustico-Phonétique (DAP), qui constitue une phase importante du domaine de la Reconnaissance Automatique de la Parole (RAP), afin de classer les séquences vocales. La RAP est considérée comme étant un problème de classification de données en attribuant à un signal vocal d'entrée une et une seule classe des phonèmes possibles. Aussi, nous avons formulé la classification acoustique supervisée comme étant un problème d'optimisation de fonctions. Par conséquent, nous avons cherché à classer les quarante phonèmes de l'Arabe Standard (AS) de la parole naturelle en utilisant les AGs en raison de leurs capacités de résoudre les problèmes d'optimisation les plus compliqués. Ainsi, nous avons analysé un corpus composé d'une centaine de phrases enregistrées par un seul locuteur adulte arabophone dans un milieu ambiant et qui contient tous les types de phonèmes dans les différents contextes vocaliques possibles. Cette analyse qui est de type cepstral, nous a permis de représenter le signal vocal par les premiers coefficients cepstraux à l'échelle de Mel, plus l'énergie et leurs dérivées premières et secondes. Nous avons choisi parmi plusieurs classificateurs celui qui se base sur le critère décisionnel de la fonction de mesure de distance, en choisissant comme fonction d'évaluation ce type de critère. Dans ce cadre, plusieurs fonctions de distance ont été utilisées. Pour rendre plus performant notre modèle génétique de la classification qui est de type séquentiel, nous avons opté pour un parallélisme de son fonctionnement en utilisant le modèle parallèle de type à sous-populations isolées sans migration. Il en résulte un taux moyen global de classification des phonèmes du corpus, de 81.17%.

**Mots Clés :** Décodage Acoustico-Phonétique, Reconnaissance Automatique de la Parole, Analyse cepstrale, Algorithmes Génétiques, Classification Supervisée, Phonèmes de l'Arabe Standard.

---

**Abstract :**

In this thesis, we have applied the Genetic Algorithms (GAs) to the acoustic-phonetic decoding step of the Automatic Speech Recognition (ASR) domain at the acoustic classification level. Speech recognition was been cast as a pattern classification problem where we would like to classify an input acoustic signal into one of all possible phonemes. Also, the supervised classification has been formulated as functions optimization problem. Thus, we have looked for classifying the forty Standard Arabic (SA) phonemes of naturally spoken speech, by using GAs which have several advantages in resolving complicated optimization problems. We have analyzed a corpus that contains several sentences composed of the whole SA phonemes types in the initial, medium and final positions, recorded by an alone and male speaker. We have used the Mel-scale Frequency Cepstral analysis method to extract vocal tract coefficients from the speech signal. Furthermore, the acoustic segments classification and the GAs have been explored. Among a set of classifiers, we have used the distance one. It is based on the classification measure criterion. Therefore, we have used several decision rules based on distance functions as the fitness functions for our GA evaluation. To perform our acoustic classification genetic model which is sequential, we have parallelize its' work by using the isolated Multi-demes or islands without migration as topology. The Corpus phonemes were extracted and classified successfully with a well overall accuracy equals to 81.17 %.

**Key words :** Acoustic-Phonetic Decoding, Automatic Speech Recognition, Cepstral Analysis, Genetic Algorithms, Supervised Classification, Standard Arabic phonemes.

## Liste des Abréviations

<b>AGPGG</b> :	<b>A</b> lgorithmes <b>G</b> énétiq <u>ue</u> s <b>P</b> arallèles à <b>G</b> ros <b>G</b> rain
<b>AGs</b> :	<b>A</b> lgorithmes <b>G</b> énétiq <u>ue</u> s
<b>API</b> :	<b>A</b> lphabet <b>P</b> honétique <b>I</b> nternational
<b>AR</b> :	<b>A</b> uto <b>R</b> égressif
<b>ARMA</b> :	<b>A</b> uto <b>R</b> égressif à <b>M</b> oyenne <b>A</b> justée
<b>AS</b> :	<b>A</b> rabe <b>S</b> tandard
<b>BF</b> :	<b>F</b> réquence du <b>B</b> urst
<b>DAP</b> :	<b>D</b> écodage <b>A</b> coustico- <b>P</b> honétique
<b>DEV</b> :	<b>D</b> élai d' <b>E</b> tablishement du <b>V</b> oisement
<b>ECT</b> :	<b>E</b> nergie à <b>C</b> ourt <b>T</b> erme
<b>EM</b> :	<b>E</b> xpectation <b>M</b> aximization
<b>emph.</b> :	consonne <b>em</b> phatique
<b>FFT</b> :	<b>F</b> ast <b>F</b> ourier <b>T</b> ransform
<b>GMM</b> :	<b>G</b> aussian <b>M</b> ixture <b>M</b> odels
<b>HMM</b> :	<b>H</b> idden <b>M</b> arkov <b>M</b> odels
<b>IA</b> :	<b>I</b> ntelligence <b>A</b> rtificielle
<b>KPPV</b> :	<b>K</b> - <b>P</b> lus <b>P</b> roches <b>V</b> oisins
<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>MA</b> :	<b>M</b> oyenne <b>A</b> justée
<b>MFCCs</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>MGS</b> :	<b>M</b> odèle <b>G</b> énétiq <u>ue</u> s <b>S</b> équentiel
<b>MGP</b> :	<b>M</b> odèle <b>G</b> énétiq <u>ue</u> s <b>P</b> arallèle
<b>MLG</b> :	<b>M</b> elange des <b>L</b> ois <b>G</b> aussiennes
<b>MLP</b> :	<b>M</b> ulti <b>L</b> ayers <b>P</b> erceptron
<b>PE</b> :	<b>P</b> rogrammation <b>E</b> volutive
<b>PMC</b> :	<b>P</b> erceptron <b>M</b> ulti <b>C</b> ouches
<b>PLP</b> :	<b>P</b> rédition <b>L</b> inéaire <b>P</b> erceptuelle
<b>RAP</b> :	<b>R</b> econnaissance <b>A</b> utomatique de la <b>P</b> arole
<b>RNA</b> :	<b>R</b> éseaux de <b>N</b> eurones <b>A</b> rtificiels

<b>RNTD :</b>	<b>R</b> éseaux de <b>N</b> eurones à <b>D</b> écalage <b>T</b> emporel
<b>SE :</b>	<b>S</b> tratégies d' <b>E</b> volution
<b>SGBD</b>	<b>S</b> ystèmes de <b>G</b> estion de <b>B</b> ases de <b>D</b> onnées
<b>SI :</b>	<b>S</b> ystèmes <b>I</b> mmunitaires
<b>SMVSV :</b>	<b>S</b> ous- <b>P</b> opulation des <b>V</b> oyelles et leurs variantes et les <b>S</b> emi- <b>V</b> oyelles
<b>SPDFN :</b>	<b>S</b> ous- <b>P</b> opulation des consonnes <b>F</b> ricatives et <b>N</b> asales
<b>SPDSL :</b>	<b>S</b> ous- <b>P</b> opulation des consonnes <b>O</b> cclusives et de la <b>L</b> iquide
<b>TF :</b>	<b>T</b> ransformée de <b>F</b> ourie
<b>TCDi :</b>	<b>T</b> ransformée en <b>C</b> osinus <b>D</b> iscrete <b>i</b> nverse
<b>TDNN :</b>	<b>T</b> ime <b>D</b> elay <b>N</b> eural <b>N</b> etworks
<b>TFD :</b>	<b>T</b> ransformée de <b>F</b> ourier <b>D</b> iscrete
<b>TFR :</b>	<b>T</b> ransformée de <b>F</b> ourier <b>R</b> apide
<b>TFI :</b>	<b>T</b> ransformée de <b>F</b> ourier <b>I</b> nverse
<b>TPZ :</b>	<b>T</b> aux de <b>P</b> assage par <b>Z</b> éro
<b>VOT :</b>	<b>V</b> oice <b>O</b> nset <b>T</b> ime

## Liste des figures

<b>Figure 1.1 :</b> Anatomie de l'appareil phonatoire Humain.....	5
<b>Figure 1.2 :</b> Classification des sons du langage.....	8
<b>Figure 1.3 :</b> Relation acoustico-articulatoire des voyelles orales du Français.....	9
<b>Figure 1.4 :</b> Système vocalique de l'Arabe Standard du point de vue articulatoire.....	12
<b>Figure 1.5 :</b> Analyse numérique du signal parole par FFT.....	20
<b>Figure 1.6 :</b> Modèle général de production de la parole .....	22
<b>Figure 1.7 :</b> Processus de calcul des coefficients LPC ET PLP.....	24
<b>Figure 1.8 :</b> Transformation schématisée pour l'obtention de la structure formantique à partir du cepstre.....	26
<b>Figure 1.9 :</b> Schéma d'extraction des coefficients MFCCs .....	28
<b>Figure 2.1 :</b> schéma de la structure générale d'un système de reconnaissance vocale.....	33
<b>Figure 2.2 :</b> Structure d'un système de reconnaissance vocale par l'approche globale.....	34
<b>Figure 2.3 :</b> Schéma de la structure d'un système de reconnaissance analytique.....	36
<b>Figure 2.4 :</b> Schéma de principe des premières étapes du traitement d'un système de RAP	38
<b>Figure 2.5 :</b> Schéma modulaire d'un système de classification acoustique.....	40
<b>Figure 3.1 :</b> différentes étapes de fonctionnement des Algorithmes Génétiques.....	57
<b>Figure 3.2 :</b> les cinq niveaux d'organisation d'un Algorithme Génétique.....	61
<b>Figure 3.3 :</b> Croisement en un point .....	69
<b>Figure 3.4 :</b> Mutation d'un gène.....	72
<b>Figure 4.1 :</b> Forme d'onde et schématisation des coefficients MFCCs, l'énergie et leurs dérivées premières et seconde de la voyelle [a].....	82
<b>Figure 4.2 :</b> Forme d'onde, spectrogramme à bande large de la semi-voyelle [y] et schématisation des coefficients MFCCs, l'énergie et leurs dérivées premières et secondes .....	83
<b>Figure 4.3 :</b> Forme d'onde, spectrogramme à bande large de la consonne nasale [m] et schématisation des coefficients MFCCs, l'énergie et leurs dérivées premières et secondes.....	84
<b>Figure 4.4 :</b> forme d'onde, spectrogramme à bande large de la consonne occlusive [θ] et schématisation des coefficients MFCCs, l'énergie et leurs dérivées premières et secondes.....	84

<b>Figure 4.5 :</b> Chromosome représentative d'une consonne occlusive à classer.....	85
<b>Figure 4.6 :</b> Chromosome représentative d'une consonne occlusive ou nasale à classer.....	86
<b>Figure 4.7 :</b> Algorithme Génétique de classification.....	90
<b>Figure 4.8 :</b> Application de l'inverse de la fonction de distance d'Euclide pour la recherche de la voyelle [i] dans la phrase p1.....	90
<b>Figure 4.9 :</b> Application de l'inverse de la fonction de distance de Manhattan pour la recherche de la consone [ʃ] dans la mot isolé [ʃ a ʃ i ʃ u] .....	91
<b>Figure 4.10 :</b> Convergence de l'AG vers l'individu qui représente l'optimum global au fil des générations.....	92
<b>Figure 4.11 :</b> Déroulement de l'Algorithme Génétique Parallèle sans migration pour la classification acoustique des phonèmes.....	95
<b>Figure 5.1 :</b> forme d'onde et spectrogramme à bande large du corpus p1.....	98
<b>Figure 5.2:</b> Comparaison des délais d'exécution entre les Algorithmes séquentiel et parallèle.....	101
<b>Figure 5.3 :</b> Evolution de la diversité des individus en fonction des générations.....	101
<b>Figure 5.4 :</b> Evolution du taux de classification global des phonèmes en fonction de la probabilité de mutation .....	102
<b>Figure 5.5 :</b> Evolution du taux de classification global des phonèmes en fonction de la probabilité de Croisement.....	102
<b>Figure 5.6 :</b> Influence du taux de croisement sur la convergence de l'AG.....	103
<b>Figure 5.7 :</b> Recherche de la voyelle[a] du corpus p1 par l'inverse de la fonction de distance de Manhattan.....	104
<b>Figure 5.8 :</b> Recherche de la voyelle[a] du corpus p2 par l'inverse de la fonction de distance de Manhattan.....	104
<b>Figure 5.9 :</b> Recherche de la voyelle[a] du corpus p3 par l'inverse de la fonction de distance Euclidienne.....	105
<b>Figure 5.10 :</b> Meilleure image acoustique représentative de la voyelle [a] après 500 Génération.....	106
<b>Figure 5.11 :</b> Organigramme du modèle génétique séquentiel de classification acoustique.....	107
<b>Figure 5.12 :</b> Critère d'arrêt de l'AG selon le nombre maximum de générations.....	108
<b>Figure 5.13 :</b> Evolution du taux de classement des différentes catégories des sons de l'Arabe Standard au fil des générations.....	109

---

<b>Figure 5.14</b> : Recherche des occlusives [t] et [q] du corpus1 par l'inverse de la fonction de distance de Manhattan.....	109
<b>Figure 5.15</b> : Recherche des fricatives [h] et [ʃ] du corpus1 par l'inverse de la fonction de distance de Manhattan.....	110
<b>Figure 5.16</b> : Localisation des voyelles de l'AS dans la phrase p3 par l'inverse de la fonction de distance de Manhattan : fonction objectif du MGS.....	111
<b>Figure 5.17</b> : Recherche des voyelles de l'AS dans la phrase p3 par l'application de l'inverse de la fonction de distance de Manhattan.....	111

## Liste des tableaux

<b>Tableau 1.1</b> : Transcription des consonnes de l'Arabe Standard en API.....	12
<b>Tableau 1.2</b> : Valeurs relatives des trois premiers formants et de la durée des voyelles de l'Arabe Standard en état isolée .....	13
<b>Tableau 1.3</b> : Modes et lieux d'articulation des phonèmes de l'Arabe Standard.....	14
<b>Tableau 1.4</b> : Les consonnes d'arrière de l'AS du point de vue articulatoire.....	18
<b>Tableau 3.1</b> : Analogie de la terminologie génétique avec la génétique.....	56
<b>Tableau 4.1</b> : Vecteurs acoustiques de référence des phonèmes de l'AS.....	87
<b>Tableau 5.1</b> : Evolution du taux global de classification des phonèmes en fonction de la taille de la population.....	100
<b>Tableau 5.2</b> : Comparaison des taux de classification suivant le type de mutation.....	101
<b>Tableau 5.3</b> : Evolution du taux global de classification des phonèmes en fonction du type de la fonction d'évaluation choisie.....	103
<b>Tableau 5.4</b> : Comparaison des taux globaux de classification des différents modèles de classification.....	106
<b>Tableau 5.5</b> : Evolution du taux de reconnaissance des phonèmes au fil des générations.....	110
<b>Tableau 5.6</b> : Taux de classification des voyelles de de l'Arabe Standard .....	110
<b>Tableau 5.7</b> : Taux de classification des consonnes occlusives en contexte vocalique.....	112
<b>Tableau 5.8</b> : Taux de classification des consonnes fricatives en contexte vocalique.....	112
<b>Tableau 5.9</b> : Taux de classification des semi-voyelles et des consonnes liquide, nasales et vibrantes en contexte vocalique.....	113

---

<b>INTRODUCTION</b> .....	1
<b>CHAPITRE 1 : Analyse Acoustico- Phonétique des sons de l'Arabe standard</b>	
1.1. Fonctionnement du système phonatoire humain.....	4
1.1.1. Production de l'onde glottique.....	5
1.1.2. Fonction résonateur du conduit vocal.....	5
1.1.3. Fonction générateur de bruit du conduit vocal.....	6
1.2. Complexité du signal vocal .....	6
1.2.1. Continuité.....	6
1.2.2. Variabilités.....	6
1.2.2.1. Variabilité intra-locuteur.....	6
1.2.2.2. Variabilité inter-locuteur.....	7
1.2.2.3. Variabilité contextuelle.....	7
1.2.3. Coarticulation.....	7
1.2.4. Redondance.....	7
1.3. Classification des sons du langage.....	8
1.3.1. Voyelles.....	9
1.3.2. Consonnes.....	10
1.3.3. Semi-voyelles.....	11
1.4. Description des sons de l'Arabe Standard.....	11
1.4.1. Système vocalique de l'Arabe Standard.....	12
1.4.2. Consonnes de l'Arabe Standard.....	14
1.4.3. Aspect acoustique des consonnes.....	15
1.4.3.1. Consonnes Occlusive.....	15
1.4.3.2. Consonnes fricatives.....	16
1.4.3.3. Consonnes nasales.....	16
1.4.3.4. Semi-voyelles et Consonnes liquides.....	16
1.4.4. Indices acoustiques des consonnes.....	16
1.4.5. Particularités des son de l'Arabe Standard.....	17
1.5. Techniques d'analyses du signal vocal.....	18
1.5.1. Méthodes non paramétriques.....	19
1.5.2. Méthodes paramétriques.....	20
1.5.2.1. Codage prédictif Linéaire.....	21
1.5.2.2. Prédiction Linéaire Perceptuelle.....	24

---

1.5.2.3. Analyse cepstrale.....	25
1.6. Vecteur à court terme image acoustique du signal vocal.....	26
1.6.1. Coefficients MFCCs à court terme.....	26
1.6.2. Energie à court terme .....	29
1.6.3. Taux moyen de Passage par Zéro à court terme.....	29
1.6.4. Paramètres dynamiques.....	30
1.7. Conclusion.....	30

## **CHAPITRE 2 : Etude du Décodage Acoustico-Phonétique**

2.1. Approches de la Reconnaissance Automatique de la Parole.....	32
2.1.1. Approche globale.....	34
2.1.2. Approche analytique.....	35
2.2. Présentation du Décodage Acoustico-Phonétique.....	36
2.2.1. Extraction des paramètres utiles du signal vocal.....	38
2.2.2. Segmentation du signal vocal.....	39
2.2.3. Etiquetage phonétique.....	40
2.3. Différents modèles de la Reconnaissance Automatique de la Parole.....	41
2.4. Classification supervisée.....	42
2.5. Méthodes de Classification.....	43
2.5.1. Mélanges des lois Gaussiennes.....	44
2.5.2. Modèles de Markov Cachés.....	44
2.5.3. Réseaux de Neurones Artificiels.....	45
2.5.4. K- Plus Proches Voisins.....	46
2.5.5. Quantification Vectorielle.....	46
2.5.6. Fonctions de Mesure Distance.....	46
2.6. Algorithmes biomimétiques de classification.....	48
2.6.1. Approches Evolutionnaires.....	48
2.6.2. Fourmis Artificielles.....	50
2.6.3. Intelligence en essaim.....	51
2.6.4. Systèmes Immunitaires.....	52
2.7. Conclusion.....	53

---

## CHAPITRE 3 : Notions de base sur les Algorithmes Génétiques

3.1. Concepts de base Algorithmes Génétiques .....	54
3.2. Fonctionnement des Algorithmes Génétiques.....	56
3.2.1. Etapes de fonctionnement.....	57
3.2.2. Réglage des paramètres des Algorithmes Génétiques.....	58
3.2.3. Avantages et limites des Algorithmes Génétiques.....	59
3.3. Codage et espace de recherche.....	60
3.3.1. Codage.....	60
3.3.1.1. Codage binaire .....	61
3.3.1.2. Codage de Gray.....	62
3.3.1.2. Codage en réel des paramètres .....	63
3.3.2. Espaces de recherche.....	63
3.4. Sélection.....	63
3.4.1. Loterie biaisée ou Roulette.....	64
3.4.2. Méthode élitiste ou par rang.....	65
3.4.3. Sélection par tournoi.....	65
3.4.4. Sélection uniforme.....	65
3.5. Techniques d'amélioration de la sélection.....	66
3.5.1. Technique du changement d'échelle.....	66
3.5.1.1. Technique du changement d'échelle linéaire.....	66
3.5.1.2. Technique du changement d'échelle exponentiel .....	66
3.5.2. Technique de partage.....	67
3.6. Reproduction.....	68
3.6.1. Croisement.....	68
3.6.2. Mutation.....	71
3.6.2.1. Mutation aléatoire.....	73
3.6.2.2. Mutation uniforme.....	73
3.6.2.3. Mutation non uniforme.....	73
3.7. Remplacement.....	74
3.7.1. Remplacement stationnaire.....	74
3.7.2. Remplacement élitiste.....	74
3.8. Conclusion.....	75

## **CHAPITRE 4 : Modèle Génétique de classification acoustique**

4.1 Méthodes d'optimisation.....	76
4.1.1. Recherche stochastique globale.....	77
4.1.1.1. Perturbation aléatoire.....	78
4.1.1.2. Critères d'acceptation.....	78
4.1.1.3. Critères d'arrêt.....	78
4.1.2. Les principales méta heuristiques modernes.....	78
4.2. Modélisation Génétique de classification vocale.....	80
4.2.1. Population initiale.....	81
4.2.2. Segmentation acoustique à court terme du continuum vocal.....	81
4.2.3. Vecteur paramétrique à court terme.....	82
4.2.4. Encodage Génétique du continuum vocal.....	85
4.2.4.1. Consonnes occlusives.....	85
4.2.4.2. Consonnes fricatives et nasales.....	85
4.2.4.3. Voyelles et semi voyelles.....	86
4.2.5. Données acoustiques de référence.....	86
4.2.6. Population des vecteurs acoustiques.....	87
4.2.7. Choix de la Fonction d'évaluation.....	88
4.2.8. Convergence des Algorithmes Génétiques.....	91
4.3. Parallélisme des Algorithmes Génétiques .....	92
4.3.1. Parallélisme globale.....	93
4.3.2. Algorithmes Génétiques Parallèles à Gros Grains.....	93
4.3.3. Algorithmes Génétiques parallèles à Sous-populations isolées.....	93
4.4. Conclusion.....	96

## **CHAPITRE 5 : : Résultats de la classification par Algorithmes Génétiques des phonèmes de l'Arabe Standard**

5.1. Implémentation.....	97
5.2. Paramètres du Modèle Génétique Séquentiel.....	98
5.2.1. Taille de la population et parallélisme.....	99
5.2.2. Diversité des individus et convergence.....	100
5.2.3. Opérateurs Génétiques de Reproduction.....	101

---

5.2.4. Fonction d'Evaluation.....	103
5.3. Résultats et commentaires de la classification génétique .....	105
5.4. Conclusion.....	114
<b>Perspectives et Conclusion Générale .....</b>	<b>115</b>
<b>Recherche Bibliographique.....</b>	<b>117</b>
Annexe 1.....	124
Annexe 2.....	130
Annexe 3.....	133

## Introduction

L'Homme a toujours manifesté le besoin de communiquer avec ses semblables. La communication a de tout le temps joué un rôle déterminant dans l'évolution des sociétés humaines. Pouvoir communiquer vite et loin, a toujours été un des plus vieux rêves de l'Homme. Ce rêve est en phase de réalisation grâce à l'avancée fulgurante de la recherche dans le domaine du dialogue Homme-Machine. Dès le début de l'informatique, la nécessité d'un tel dialogue à l'aide du Langage Naturel a vu le jour. Reconnaître la parole, vecteur d'informations, privilégié de notre société, constitue donc un défi à relever pour permettre le passage à la prochaine génération d'interfaces Homme-Machine.

Notre travail concerne l'application des Algorithmes Génétiques (AGs) au Décodage Acoustico-Phonétique (DAP) des phonèmes de l'Arabe Standard (AS) en vue de leur reconnaissance. Le DAP constitue une étape très importante dans la conception d'un système de Reconnaissance Automatique de la Parole (RAP) puisqu'il représente l'ensemble des processus de transformation du signal acoustique continu en une description linguistique discrète sous forme d'unités tels que les phonèmes, les diphonèmes ou diphones, les syllabes, etc. La classification acoustique joue un rôle essentiel dans le DAP.

Le Traitement Automatique de la Parole (TAP) est un domaine de recherche actif, au croisement du traitement du signal numérique et du traitement symbolique du langage. Depuis les années 60, il bénéficie d'efforts de recherche très importants, liés au développement des moyens et techniques de télécommunications. Un thème important de la recherche actuelle dans le domaine du TAP, est la réalisation de véritables systèmes de dialogue oral entre l'Homme et la Machine dont la Reconnaissance Automatique de la Parole (RAP) joue un rôle très important. C'est dans cette optique que se situe notre travail.

La RAP s'applique à ce jour sur de nombreux signaux de qualités différentes (fréquence d'échantillonnage, quantification, codage, conditions d'enregistrement, etc.). Par conséquent, le problème de l'adaptation des modèles acoustiques à un nouvel environnement, se pose avec acuité.

Pour cela, nous avons mis l'accent sur la classification supervisée des vecteurs représentatifs des segments acoustiques en utilisant la technique des Algorithme Génétiques

(AGs). Les segments acoustiques à classer, sont le résultat de la segmentation automatique du continuum vocal du corpus à traiter.

Le corpus que nous avons utilisé comporte une centaine de phrases comprenant toutes les combinaisons phonémiques de l'AS, enregistré par un locuteur masculin dans un milieu ambiant.

L'utilisation des AGs en tant que modèle de classification acoustique est très récente et fait l'objet de plusieurs travaux de recherche, à travers le monde, qui commencent à donner des résultats, de plus en plus en conformité avec ceux attendus.

Les AGs font partie d'une nouvelle génération d'algorithmes appelés algorithmes évolutionnaires. Ils ont été conçus en se basant sur les principes de la génétique et de la reproduction naturelle. Ils sont généralement utilisés pour résoudre les problèmes, là où les méthodes classiques ont échoué. Ils sont très efficaces dans l'optimisation des systèmes dynamiques complexes et qui évoluent rapidement avec le temps. En outre, ils sont caractérisés par leur parallélisme. En RAP, les AGs ont été utilisés, d'une manière auxiliaire, pour optimiser les **Modèle de Markov Cachés (MMC)** Cachés, les modèles de Réseaux de Neurones, etc.

Nous avons opté pour l'utilisation des AGs à des fins de DAP au lieu des modèles classiques connus comme les MMC ou les **Réseaux de Neurones Artificiels (RNA)**, vue leur capacité de traitement de grandes bases de données, leur robustesse aux données bruitées ou manquantes et leur puissance ainsi que leur rapidité de discrimination pour diviser l'espace des paramètres acoustiques en classes phonétiques. Ainsi, l'utilisation des AGs comme modèle de classification vocale consiste à optimiser les classificateurs ayant comme fonction d'évaluation celle qui se base sur les fonctions de mesure de distance (Euclidienne, Manhattan, etc.).

Dans notre travail nous avons focalisé notre étude sur les différentes étapes du DAP en insistant sur l'étape de la classification acoustique supervisée des différents phonèmes de l'AS. Aussi, nous avons présenté les AGs, le principe et les étapes de leur fonctionnement, les différents opérateurs de la phase de reproduction et le rôle de chacun d'eux. Sans oublier leur dimensionnement et leurs critères d'arrêt. Nous dégageons l'importance des choix de l'espace de représentation, du type de codage des paramètres du problème à résoudre et de la nature de la fonction objective ou d'évaluation. Les différentes techniques d'amélioration des AGs ont été abordées.

Dans le quatrième chapitre nous avons expliqué comment nous avons appliqué les AGs au DAP en insistant sur le choix de la fonction d'évaluation de notre modèle génétique et celui de ses différents paramètres, la nature des individus de sa population et le choix de ses opérateurs de reproduction ;

Le cinquième chapitre décrit notre modèle génétique de classification acoustique qu'il soit séquentiel ou parallèle tout en expliquant le choix des paramètres de réglage, des opérateurs de reproduction et de la fonction d'évaluation pour atteindre les meilleurs taux de classification. En interprétant les résultats obtenus, il a été démontré que le choix du modèle génétique parallèle de type à gros grains ou en îlots sans migration permet d'atteindre les meilleurs scores de performance.

Durant notre travail, nous nous sommes confrontés à des difficultés concernant le choix approprié des paramètres de notre modèle génétique, relatifs aux divers opérateurs (mutation, croisement, sélection et remplacement). Nous avons remarqué que le taux de rejet est contrôlé indirectement par le choix de ces paramètres. Par conséquent, il faudra se pencher sur la meilleure manière de les sélectionner, qui reste un problème ouvert.

## **Chapitre 1 :      Analyse Acoustico- Phonétique des sons de l'Arabe Standard**

Dans le processus de la communication parlée, pour une langue donnée, les sons permettent de distinguer les différentes unités de signification du langage. Pour cela, une étude préalable du signal vocal ainsi que son système de production doit être faite, afin de donner à la machine toutes les propriétés de la parole : linguistiques, acoustiques et prosodiques [1].

Nous présentons dans ce chapitre, les principales caractéristiques phonético-articulatoires des sons de la parole et de l'Arabe Standard en particulier. En outre, nous étudions différentes techniques d'analyse du signal vocal tout en insistant sur l'analyse cepstrale afin de déterminer les paramètres pertinents nécessaires à une meilleure représentation du signal de parole.

### **1.1. Fonctionnement de l'appareil phonatoire humain**

Comprendre le mécanisme de production de la parole est un aspect d'une grande importance. En effet, c'est l'étude du système de phonation qui va nous permettre d'identifier et de caractériser les grandes classes de sons élémentaires et d'expliquer les variations de ces derniers dans les différents contextes. De plus, les algorithmes de paramétrisation du signal vocal sont obtenus à partir de modèles du conduit vocal.

Les traits acoustiques du signal de parole sont évidemment liés à sa production. L'intensité du son dépend de la pression de l'air en amont du larynx. Sa fréquence, qui n'est rien d'autre que celle du rythme d'ouverture/fermeture des cordes vocales, induit par la tension de muscles qui les contrôlent. Son spectre résulte du filtrage du signal glottique (impulsions, bruit, ou combinaison des deux) par le conduit vocal, qui peut être considéré comme une succession de tubes ou de cavités acoustiques de sections diverses [2].

La parole est articulée en interrompant et en modulant le flux d'air à l'aide des lèvres, de la langue, des dents, de la mâchoire inférieure et du palais (Fig. 1.1).

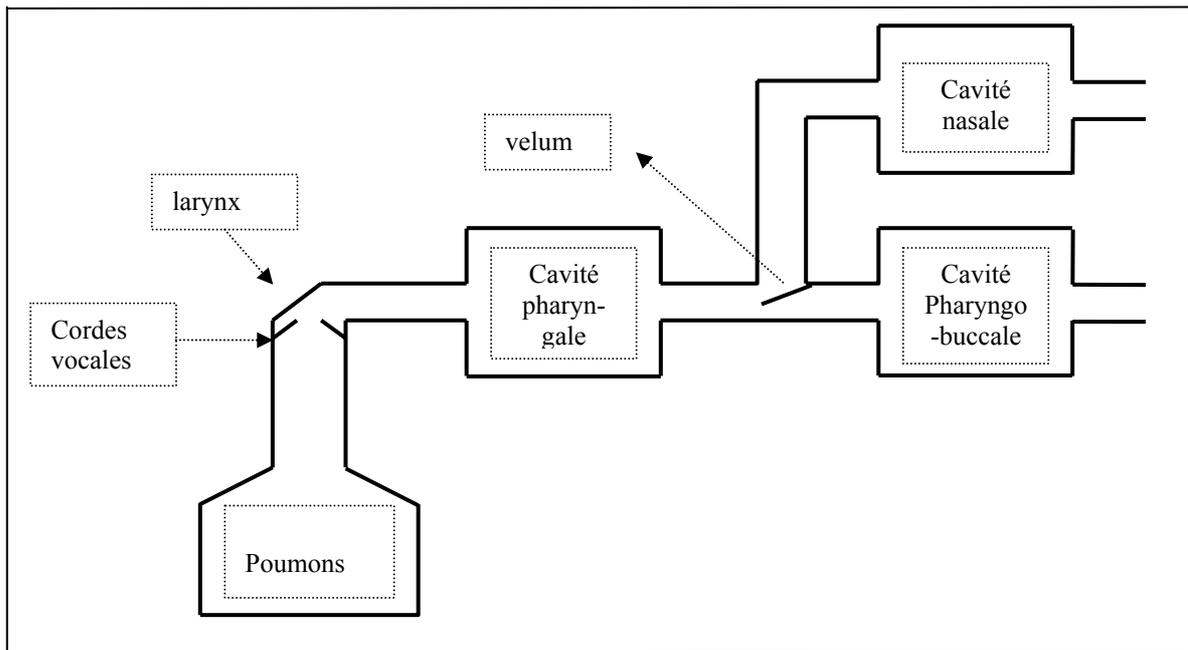


Figure 1.1 : Anatomie de l'appareil phonatoire Humain

### 1.1.1. Production de l'onde glottique

L'air produit par excès de pression dans les poumons rencontre un premier obstacle qui sont les cordes vocales (source d'excitation). Ces dernières accolées, sous l'effet de la pression sub-glottique se mettent à vibrer laissant passer l'air par impulsions. C'est ainsi que se forme l'onde glottique dont la fréquence d'oscillations notée  $F_0$  (fréquence fondamentale ou pitch), est déterminée par la masse et la tension des cordes vocales ainsi que la pression sub-glottique. Quand elles vibrent, il y a émissions de sons dits sons voisés ou sonores par opposition aux sons non voisés ou sourds qui sont assimilables à un bruit blanc.

### 1.1.2. Fonction résonateur du conduit vocal

Le conduit vocal imprime au son émis les caractéristiques spécifiques permettant de distinguer les différents phonèmes et ceci selon deux fonctions, en tant que :

- résonateur de l'onde glottique pour la production des phonèmes sonores ;
- générateur de bruit pour la production des consonnes sourdes.

En effet, l'onde glottique est modifiée lors de son passage à travers le conduit vocal. Les positions de la mâchoire et de la langue déterminent les cavités qui jouent le rôle de caisses de résonance en renforçant certaines régions du spectre acoustique. Les maxima de la courbe de réponse en fréquences du conduit vocal sont appelés formants

### **1.1.3. Fonction générateur de bruit du conduit vocal**

Le flux d'air créé peut rencontrer soit, un obstacle partiel tel un rétrécissement du conduit vocal pour générer un bruit caractéristique des sons fricatifs ou constrictifs, soit un obstacle total produisant une augmentation de la pression en amont de l'obstacle (lieu d'articulation) suivi d'un relâchement brusque. Ce phénomène engendre la formation des sons occlusifs.

## **1.2. Complexité du signal vocal**

La grande difficulté du TAP et en particulier celui de la RAP provient du caractère du processus de la communication parlée et des caractéristiques intrinsèques du signal vocal. La parole est un signal continu d'énergie finie, non stationnaire. Sa structure est complexe et variable dans le temps ; tantôt périodique ou plus exactement pseudo périodique pour les sons voisés, tantôt aléatoire pour les sons fricatifs et tantôt impulsionnel pour les sons occlusifs.

### **1.2.1. Continuité**

Le langage oral est une suite continue de sons sans séparation entre les mots. Les silences correspondent en général à des pauses de respiration dont l'occurrence est aléatoire. Il peut très bien y avoir des intervalles de silence au milieu d'un mot et aucun intervalle entre deux mots successifs. Par conséquent, il est très difficile de déterminer le début et la fin des mots composant la phrase.

### **1.2.2. Variabilités**

La parole présente une très grande variabilité qui résulte de plusieurs facteurs et ceci que ce soit pour un même ou plusieurs locuteurs. Parmi ces facteurs, les perturbations apportées par le microphone (selon le type, la distance et l'orientation) et l'environnement (bruit et réverbération). De telles variations ne donnent pas naissance à de nouveaux phonèmes, puisqu'elles ne portent aucune information sémantique. Ainsi, les phonèmes apparaissent sous une multitude de formes articulatoires, appelées *allophones* ou *variantes*.

#### ***1.2.2.1. Variabilité intra-locuteur***

La variabilité intra-locuteur concerne les différences de production du signal parole chez un même locuteur. Plusieurs critères peuvent être responsables de ces différences :

- la fatigue ;
- l'état émotionnel du sujet qui affecte le timbre et le rythme de la voix ;

- les maladies affectant les organes de la voix.

### **1.2.2.2. Variabilité inter-locuteur**

Des différences acoustiques apparaissent dans un mot prononcé par plusieurs locuteurs. En effet, des contrastes considérables peuvent se manifester suivant l'âge, le sexe, l'origine géographique et le milieu social.

### **1.2.2.3. Variabilité contextuelle**

Les mouvements articulatoires peuvent en effet être modifiés de façon à minimiser l'effort à produire pour les réaliser à partir d'une position articulatoire donnée, ou pour anticiper une position à venir. Ces effets sont connus sous le nom de *réduction*, *d'assimilation* et de *coarticulation* [3].

Les phénomènes coarticulatoires sont dûs au fait que chaque articulateur évolue de façon continue entre les positions articulatoires. Ils apparaissent même dans le parlé le plus soigné. Au contraire, la réduction et l'assimilation prennent leur origine dans des contraintes physiologiques et sont sensibles au débit de la parole. L'assimilation est causée par le recouvrement de mouvements articulatoires et peut aller jusqu'à modifier un des traits phonétiques du phonème prononcé. La réduction est plutôt due au fait que les cibles articulatoires sont moins atteintes dans le parler rapide.

Ces phénomènes sont en grande partie responsables de la complexité des traitements réalisés sur les signaux de parole.

### **1.2.3. Coarticulation**

La coarticulation est l'effet contextuel d'un son sur son voisin. Les contraintes introduites par les mécanismes de production créent ce genre de phénomènes. La production d'un son est fortement influencée par les sons qui le précèdent mais aussi par ceux qui le suivent en raison de l'inertie du geste articulatoire.

### **1.2.4. Redondance**

Le signal de la parole est très redondant. Son traitement automatique nécessite, en effet, de réduire au maximum cette redondance afin de diminuer l'encombrement en mémoire et de limiter les durées du traitement, lequel doit se faire en temps réel. A l'inverse, le débit ne doit

pas être trop faible pour conserver un bon rapport signal/ bruit. En effet, Il existe une grande disproportion entre le débit du signal enregistré et la quantité utile pour une tâche de reconnaissance.

### 1.3. Classification des sons du langage

D'un point de vue linguistique, la production des sons ou d'un mot réside dans la production en série de tous les phonèmes constituant ce mot. Ces phonèmes forment les unités phonétiques qui sont classées en voyelles, consonnes et semi-voyelles, etc. [ 2, 4].

Il est intéressant de grouper les sons de parole en classes phonétiques, en fonction de leur mode et lieu d'articulation. Dans la cavité buccale, le point d'articulation est l'endroit où se trouve un obstacle au passage de l'air. D'une manière générale, nous pouvons dire que le point d'articulation est l'endroit où vient se placer la langue pour obstruer le passage du canal d'air. Nous distinguons généralement trois classes principales : les voyelles, les semi-voyelles et les consonnes (Fig. 1.2).

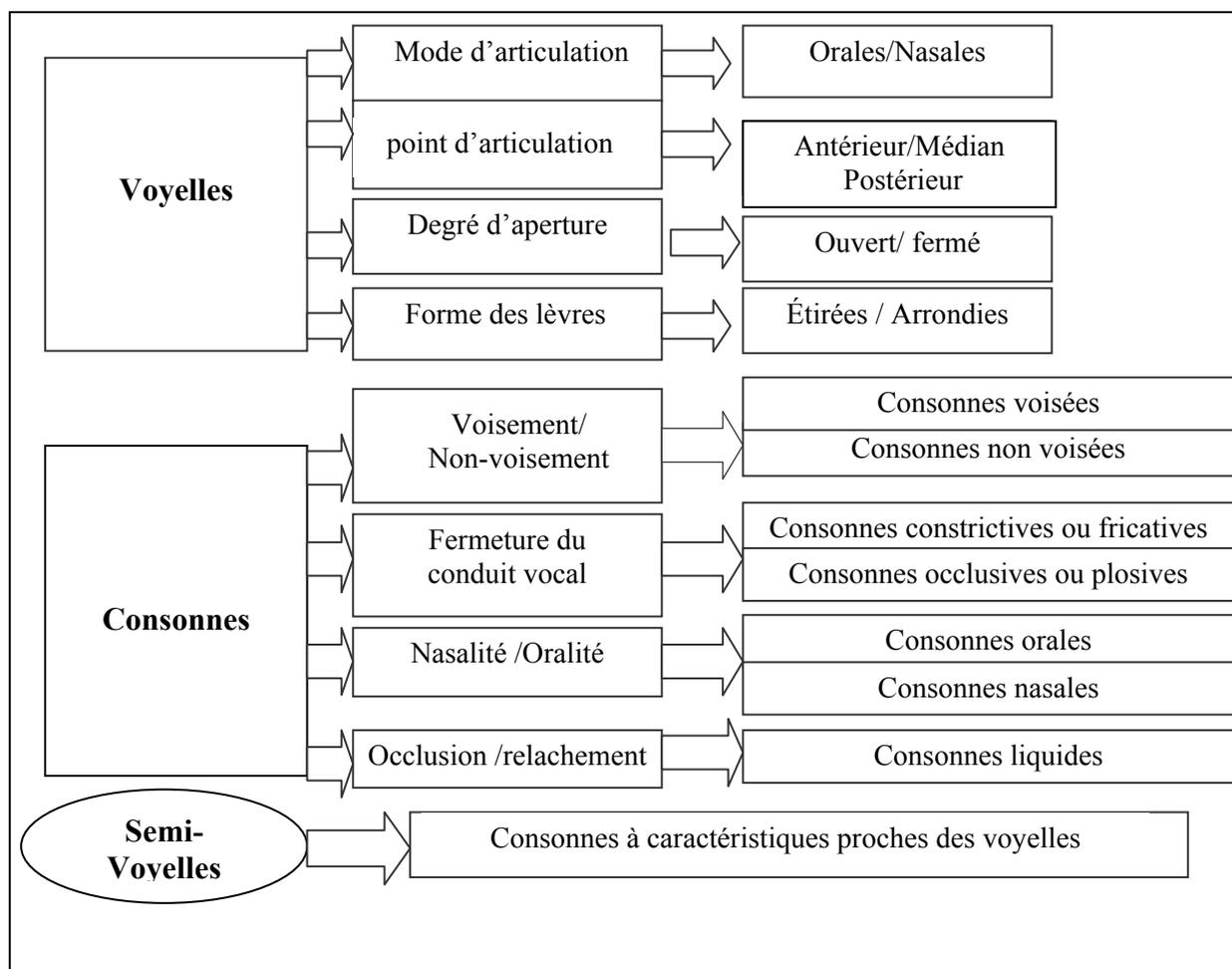


Figure 1.2 : Classification des sons du langage

### 1.3.1. Voyelles

Les voyelles diffèrent de tous les autres sons par le degré d'ouverture du conduit vocal. Quand ce dernier est suffisamment ouvert pour que l'air expiré par les poumons le traverse sans obstacle, il y a production d'une voyelle. Le rôle de la cavité buccale se réduit alors à une modification du timbre vocalique. Si, au contraire, le passage se rétrécit par endroit, ou même s'il se ferme temporairement, le passage forcé de l'air donne naissance à un bruit : une consonne est produite. Une voyelle se caractérise par un passage libre de l'air dans le conduit vocal et par la vibration des cordes vocales.

Elles se différencient principalement les unes des autres par leur lieu d'articulation (position de la langue), leur degré d'ouverture (espace compris entre la pointe de la langue et le palais), et leur nasalisation. Nous distinguons ainsi, selon la localisation de la masse de la langue, les antérieures, les moyennes, et les voyelles postérieures, et, selon l'écartement entre l'organe et le lieu d'articulation, les voyelles fermées et ouvertes.

Les voyelles orales sont dues à une élévation du palais qui détermine la fermeture des fosses nasales ainsi qu'à l'écoulement de l'air expiratoire à travers la cavité buccale. Par contre les voyelles nasales sont caractérisées par l'écoulement d'une partie de l'air à travers la cavité nasale. L'AS ne possède pas de voyelles nasales. Elles sont représentées sur un plan dont les axes sont les formants  $F_1$  et  $F_2$ . Elles tracent alors un triangle dont les extrémités sont occupées par les voyelles [i, u, a]. Ce triangle représente également les positions de la langue dans la cavité buccale selon deux axes : antérieur à postérieur (avant et arrière) et fermé à ouvert (Fig1.3).

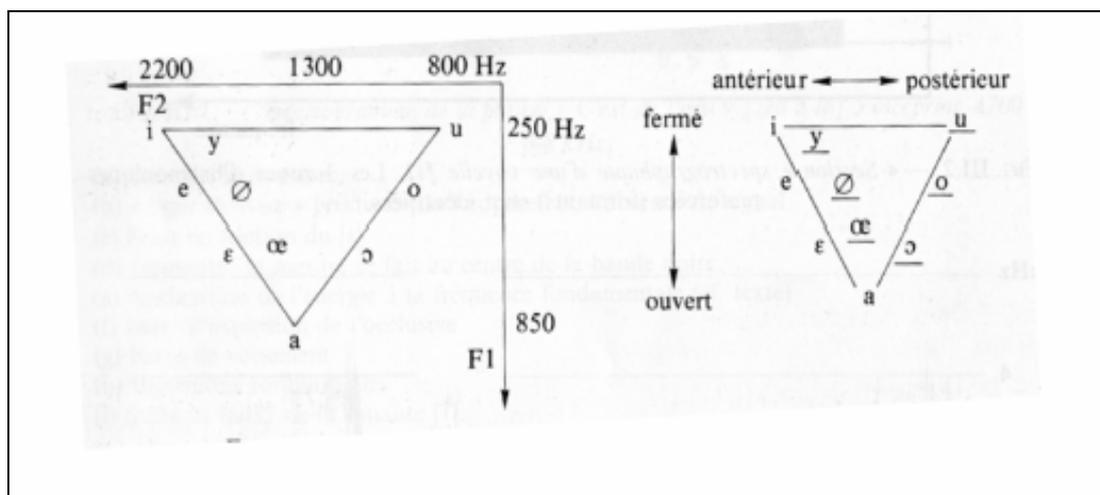


Figure 1.3 : Relation acoustico-articulatoire des voyelles orales du Français [1]

Les voyelles soulignées sont labialisées

### 1.3.2. Consonnes

Les consonnes se caractérisent par une fermeture partielle du conduit vocal ou constriction (constrictives ou fricatives) ou totale du conduit vocal (occlusion) : occlusives ou plosives. Nous classons principalement les consonnes en fonction de leur mode d'articulation, de leur lieu d'articulation, et de leur nasalisation. Le mode d'articulation est défini par un certain nombre de facteurs qui modifient la nature du courant d'air expiré :

- intervention ou mise en vibrations des cordes vocales : articulation sonore ;
- fermeture momentanée du passage de l'air suivie d'une ouverture brusque (explosion): articulation occlusive ;
- rétrécissement du passage de l'air qui produit un bruit de friction ou de frôlement : articulation fricative ;
- position abaissée du voile du palais: articulation nasale ;
- contact de la langue au milieu du canal buccal; l'air sort des deux côtés ;
- une série d'occlusions brèves ; séparées de la luvette: articulation vibrante.

La distinction du mode d'articulation conduit à deux classes : les fricatives ou constrictives et les occlusives ou plosives. Les consonnes fricatives appelées également spirantes sont créées par une constriction du conduit vocal au niveau du lieu d'articulation, qui peut être le palais, les dents ou les lèvres. Les fricatives non voisées sont caractérisées par un écoulement d'air turbulent à travers la glotte, tandis que les fricatives voisées combinent des composantes d'excitation périodique et d'autres turbulentes : les cordes vocales s'ouvrent et se ferment périodiquement, mais la fermeture n'est jamais complète.

Les consonnes occlusives ou plosives sont reconnues grâce au silence provenant de la fermeture totale du conduit vocal ou occlusion. Cette dernière comporte trois phases :

- l'implosion ou fermeture ;
- l'occlusion proprement dite tenue de la fermeture ;
- l'explosion ou détente.

Les consonnes liquides combinent une occlusion et une ouverture simultanée du conduit vocal. Elles sont caractérisées par un degré de sonorité proche de celui des voyelles. Enfin, les consonnes nasales font intervenir la cavité nasale par abaissement du voile du palais. Elles sont produites par l'écoulement de l'air phonatoire dans le conduit nasal.

### 1.3.3. Semi-voyelles

Les semi-voyelles, quant à elles, combinent certaines caractéristiques des voyelles et des consonnes. Comme les voyelles, leur position centrale est assez ouverte, mais le relâchement soudain de cette position produit une friction qui est typique des consonnes. Enfin, elles sont assez difficiles à classer.

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

L'Arabe est la langue du Coran, des médias, de la science, de l'enseignement, de la littérature, etc. Elle est structurée d'une manière différente relativement aux autres langues : les consonnes ou [huru:f] et les voyelles ou [haraka:t]. L'Arabe n'a pas une écriture strictement phonétique, car à une même graphie correspondent plusieurs images phoniques selon le contexte. Son écriture, comme celle des autres langues sémitiques, est consonantique. Les sons de l'Arabe se composent de :

- vingt huit consonnes qui peuvent prendre des formes légèrement différentes selon qu'elles sont situées en position isolée ou initiale, médiane ou finale dans le mot ;
- trois voyelles brèves kasra, damma et fatha [i, u, a]. ces voyelles ne sont pas notées. Mais pour faciliter la lecture et la compréhension d'un texte, les voyelles brèves sont représentées par les signes diacritiques ;
- trois voyelles longues [i:, u:, a:], appelées [huru:f ε l madd] ;
- le silence, appelé [suku:n].

Les Phonéticiens symbolisent les sons du langage au moyen de signes divers auxquels nous attribuons une valeur conventionnelle. Selon les auteurs la transcription varie beaucoup. Nous choisissons la transcription de l'Alphabet Phonétique International (API) pour des raisons de simplicité (Tab.1.1).

L'originalité de la phonétique arabe se fonde, sur les consonnes emphatiques, pharyngales et laryngales et huru:f ε l mad, car elles donnent une valeur particulière à la langue [4, 5].

Tableau 1. 1 : Transcription des consonnes de l'Arabe Standard en API.

Phonèmes de l'AS (API)	Phonèmes de l'AS (Arabe)	Phonèmes de l'AS (API)	Phonèmes de l'AS (Arabe)
[ʔ]	[ء]	[d̥]	[ض]
[b]	[ب]	[t̪]	[ط]
[t]	[ت]	[z̪]	[ظ]
[θ]	[ث]	[ɛ]	[ع]
[ʒ̪]	[ج]	[ɣ]	[غ]
[h]	[ح]	[f]	[ف]
[x]	[خ]	[q]	[ق]
[d]	[د]	[k]	[ك]
[ð]	[ذ]	[l]	[ل]
[R]	[ر]	[m]	[م]
[Z]	[ز]	[n]	[ن]
[S]	[س]	[h]	[ه]
[ʃ]	[ش]	[w]	[و]
[ʕ]	[ص]	[y]	[ي]

**1.4.1. Système vocalique de l'Arabe Standard**

Le système vocalique de l'AS ne possède que des voyelles orales dont les caractéristiques sont étudiées selon deux points de vue : articuloire et acoustique. Sur le plan articuloire, l'AS n'a que deux degrés d'aperture pertinents (Fig.1.4) :

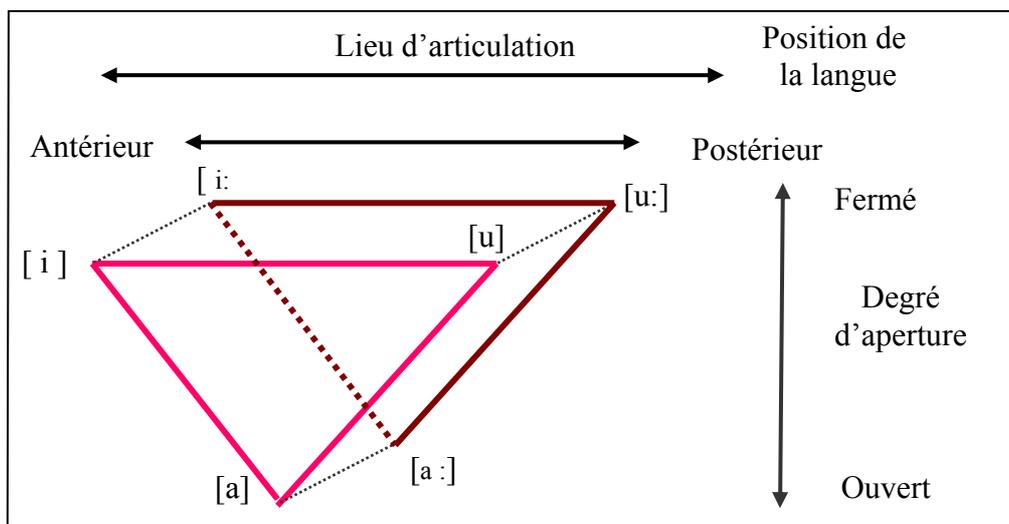


Figure 1. 4 : Système vocalique de l'Arabe Standard du point de vue Articuloire [5]

- [i, i:] et [u, u:] sont des voyelles fermées ;
- [a, a:] sont des voyelles ouvertes.

L'AS possède également trois lieux d'articulation. Les voyelles :

- [i, i:] sont postérieures (et arrondies) ;
- [u, u:] sont antérieures (et étirées) ;
- [a, a:] sont médianes.

Sur le plan acoustique, les voyelles se caractérisent par la présence d'énergie substantielle dans les régions de basses et moyennes fréquences qui caractérisent principalement les valeurs des trois premières fréquences formantiques (Tab.1.2).

Tableau 1. 2 : Valeurs relatives des trois premiers formants et de la durée des voyelles de l'Arabe Standard à l'état isolée [5]

Voyelles	Durée (ms)	F1 (Hz)	F2 (Hz)	F3 (Hz)
[i]	300	290	2200	2700
[i:]	600	285	2200	2700
[u]	300	290	800	2150
[u:]	600	285	775	2050
[a]	300	600	1500	2100
[a:]	600	675	1200	2150

Lors d'une réalisation vocalique, le conduit vocal se déforme très peu, les fréquences des formants sont stables, elles sont spécifiques à la voyelle émise [6].

Dans la réalisation acoustique, les trois phonèmes fondamentaux [i, u, a] comportent diverses variantes :

- [ˈima:la] est une prononciation du [a], son point d'articulation se rapproche de [e] ;
- le [tafɣ i:m], la présence des consonnes [mufaɣ ɣ ama] empêche [ˈima:la].

Ces consonnes reportent en arrière le point d'articulation des voyelles suivantes de sorte qu'elles deviennent [ɑ, o, e] [7, 8].

### 1.4.2. Consonnes de l'Arabe Standard

L'AS comprend sept classes de consonnes relatives au mode d'articulation. En effet, il existe quatorze consonnes fricatives dont cinq ont pour lieu d'articulation des positions très en arrière aux niveaux uvulaire, glottal et pharyngal. Aussi, l'AS comprend huit consonnes plosives (Tab.1.3).

Tableau 1.3 : Modes et lieux d'articulation des phonèmes de l'Arabe Standard

\* : consonne voisée.

emph. : consonne emphatique

Lieux Modes	Occlusives		Fricatives		Nasales	Vibrante	Liquide	Semi-Voyelles
	emph.	Non emph	emph.	Non Emph				
Bilabiale		[b*]			[m*]			
Labiodentale				[f]				
Interdentale			[z]	[θ] [θ*]				
Dentale	[t*]	[t] [d*]						
Apico-dentale								
Alvéo-dentale	[d]		[s]	[s] [z*]	[n*]	[r*]	[l*]	
Palatale				[j]				[y*]
Palato-alvéolaire				[ʒ*]				
Post-palatal		[k]						
Vélaire				[x]				
Labio-vélaire								[w*]
Uvulaire		[q]		[ɣ*]				
Pharyngale				[h] [ε*]				
Glottale		[ʔ]		[h*]				

Nous distinguons les occlusives aspirées et celles qui ne le sont pas. Lors de l'aspiration nous remarquons que :

- l'existence d'une composante de bruit de friction forte au niveau de la glotte ou du larynx ;

- la présence de formants proches de ceux des voyelles voisines.

Les fricatives font partie, avec les occlusives, de la classe des obstruantes puisqu'elles présentent toutes une obstruction à des degrés divers du courant d'air expiré.

L'AS comprend deux consonnes nasales ; la bilabiale [m] et l'alvéodentale [n]. Ces deux consonnes sont sonores. Dans la production de ce type de consonnes, deux cavités sont normalement utilisées ; les cavités orale et nasale.

Il comprend, aussi, deux semi-voyelles qui sont la palatale [ɣ] et l'uvulaire [w]. Elles sont toutes les deux sonores.

### 1.4.3. Aspect acoustique des consonnes

Généralement, les consonnes ont moins d'énergie acoustique que les voyelles. Les caractéristiques des voyelles sont relativement plus proéminentes et plus stables que celles des consonnes. Ces dernières varient individuellement, c'est pour cette raison que nous allons les étudier par classes [9, 10].

#### 1.4.3.1. Consonnes Occlusives

Il existe respectivement des consonnes occlusives sourdes ou non voisées et des occlusives sonores. Lorsque la sourde est suivie d'une voyelle, l'évolution acoustique présente :

- un silence momentané dû à l'occlusion ;
- une explosion ;
- un bruit de friction, produit au niveau de la constriction ; le spectre est celui d'un bruit à bande aiguë ;
- un bruit d'écoulement glottal ; le spectre est celui d'un bruit de bande très aiguë ;
- des vibrations de cordes vocales dues à la voyelle suivant l'articulation consonantique qui produisent un spectre harmonique. L'intensité décroît régulièrement du grave vers l'aigu et ce signal apparaît avec un certain retard par rapport à l'explosion. Ce retard est appelé **Délai d'Etablissement du Voisement (DEV)** ou **Voice Onset Time (VOT)** en Anglais.

Quant aux consonnes occlusives sonores, les étapes de leur réalisation sont identiques à celles qui précèdent à la production des occlusives sourdes, mais elles sont accompagnées de vibrations des cordes vocales. Le voisement a pour conséquences :

- une atténuation des bruits d'explosion et de friction ;

- la disparition du bruit d'écoulement glottal, caractéristique des occlusives sourdes.

#### **1.4.3.2. Consonnes fricatives**

Pour ce type de consonnes, nous pouvons faire les remarques suivantes :

- les fricatives sonores subissent une double excitation, c'est-à-dire un mélange d'impulsions périodiques (du larynx) et de bruits au lieu de constriction ;
- la concentration énergétique se fait dans les mêmes régions pour les sourdes et les sonores, avec surimposition de l'énergie de basses fréquences pour les sonores.

#### **1.4.3.3. Consonnes nasales**

Comme l'air peut s'écouler librement par les fosses nasales, l'occlusion peut être maintenue très longtemps. Pendant l'occlusion, le son émis possède des antiformants de nasalité, dûs aux anti-résonances propres aux cavités nasales. Ces Anti-résonances sont plus fortement amorties que celles du conduit oral quand il est ouvert. La zone 500-2000 Hz, notamment, est fortement atténuée lors de la réalisation de [n], un peu moins lorsque nous produisons [m]. L'explosion est extrêmement brève, voire presque absente : l'air passe librement par la cavité nasale.

#### **1.4.3.4. Semi-voyelles et Consonne liquide**

Le bruit des consonnes est essentiellement produit par le passage de l'air à travers le conduit vocal, alors que les voyelles sont le fruit des propriétés de résonateurs du conduit vocal. Les semi-voyelles sont produites par le passage de l'air à travers le conduit vocal qui fonctionne également et simultanément en mode résonnant.

#### **1.4.4. Indices acoustiques des consonnes**

La perception phonétique et la production de la parole constituent chacune une synthèse dynamique opérée sur la base de l'information extraite du signal. Le signal acoustique est porteur d'indices. Les traits sont des constructions qui ne résident pas dans le signal. Ils se manifestent par un ensemble d'indices acoustiques organisés sur le plan de la perception.

Le sujet sait extraire du signal acoustique les indices qui lui permettent de reconnaître les traits correspondants. Les indices sont des événements concrets organisés sur l'axe temporel. Ils ne peuvent être reconnus à mesure qu'ils arrivent, sinon on devrait admettre que la perception est immédiate, ce qui est démenti par la mesure du temps de réaction. Le

stimulus auditif est stocké dans une mémoire auditive préperceptuelle ou précatégorielle. L'image auditive persiste après présentation du stimulus pendant 200 ms environ. La mémoire à court terme intervient ensuite pour retenir l'information de la mémoire précatégorielle et élaborer les traits distinctifs [10].

Les phonèmes qui possèdent le trait consonantique sont acoustiquement caractérisés par la présence de zéros acoustiques qui affectent le spectre [10, 11].

Parmi les indices acoustiques qui jouent un rôle dans la production et la perception des consonnes, nous citons en ce qui concerne les :

- occlusives : la direction et la pente des transitions, la fréquence des bruits d'explosion, la fréquence terminale des transitions, etc ;
- fricatives : la durée des bruits, la direction et la pente des transitions, la fréquence des bruits ;
- consonnes nasales : la fréquence des anti-résonances ou zéros acoustiques, la réduction d'intensité des formants, etc ;
- voyelles : la durée des tenues et des transitions, la fréquence des formants et des transitions, la réduction d'intensité des formants, etc.

Nous citerons également les indices qui interviennent dans l'opposition consonne voisée-non voisée, c'est-à-dire la présence ou l'absence de la barre de voisement sur la tenue, l'intensité des bruits d'explosion ou des bruits de friction, la coupure (cutback) de  $F_1$ , le délai d'établissement du voisement, etc.

#### **1.4.5. Particularités des sons de l'Arabe Standard**

L'originalité de la phonétique Arabe se fonde, pour une grande partie sur la pertinence de la durée dans le système vocalique avec l'existence des huru:f  $\mathcal{E}1$  mad et sur la présence de consonnes emphatiques, qui sont marquées par plusieurs caractéristiques qui se superposent les unes aux autres dont le recul de la racine de la langue et la pharyngalisation. En présence des consonnes emphatiques toutes les voyelles de l'AS sont influencées énergiquement d'où l'obtention de variantes voyelles emphatisées.

Une autre caractéristique déterminante est la gémiation. Celle-ci joue un rôle fondamental dans le développement morphologique nominal et verbal. La gémiation correspond au dédoublement de deux consonnes identiques en une seule dite gémignée. C'est aussi la

production d'une consonne avec une concentration d'énergie très intense. Toutes les consonnes arabes sont susceptibles d'être gémées sauf la glottale hamza.

L'AS se caractérise, aussi, par la présence des consonnes arrières glottales, pharyngales et vélares qui sont au nombre de huit. Les pharyngales et les glottales se distinguent des autres consonnes par leur lieu d'articulation qui est définie comme un ensemble de localisations anatomiques allant du palais jusqu'à la glotte [5, 12].

Les consonnes pharyngales, glottales et uvulaires sont appelées consonnes arrières par rapport aux lieux d'articulation (Tab. 1.4).

Mode articulaire	Lieu d'articulation des consonnes			
	Post-palatal	Vélaire uvulaire	pharyngales	Laryngales ou glottales
Occlusif ou plosif	[k]	[q]		[ʔ]
Constrictif ou fricatif		[x]	[ħ] sonore	[h] sonore
		[ɣ] sonore	[ʕ] sonore	

Tableau 1.4 : Les consonnes arrières de l'AS du point de vue articulaire [13].

## 1.5. Techniques d'analyse du signal vocal

Une fois que le son a été émis par le locuteur, il est capté par un microphone. Le signal vocal est ensuite numérisé à l'aide d'un Convertisseur Analogique/Numérique. Comme la voix humaine est constituée d'une multitude de sons, souvent répétitifs, le signal peut être compressé pour réduire le temps de traitement et l'encombrement en mémoire. Ainsi comme prétraitement, nous échantillons et préaccentuons le signal vocal. Pour les techniques de reconnaissance, d'analyse ou de synthèse de la parole, la fréquence d'échantillonnage peut varier de 08 jusqu'à 16 kHz. Le filtre de préaccentuation qui est souvent non récursif du premier ordre, permet d'égaliser les aigus toujours plus faibles que les graves. Aussi et vu qu'il est non stationnaire, nous réalisons un fenêtrage avec une fenêtre glissante; chaque trame couvrant une durée de 20 à 30 ms sur laquelle le signal est supposé quasi-stationnaire. Le pas d'analyse entre deux trames successives est de l'ordre de quelques dizaines de ms. Le découpage du signal en trames produit des discontinuités aux frontières des trames, qui se manifestent par des lobes secondaires dans le spectre. Pour compenser ces effets de bord,

nous multiplions en général préalablement chaque tranche d'analyse par une fenêtre de pondération de type fenêtre de Hamming.

Le signal vocal peut être analysé soit, en tenant compte des mécanismes de production en utilisant les méthodes paramétriques, soit en les ignorant en utilisant les méthodes non paramétriques.

Dans la plupart des méthodes d'analyse vocale, nous supposons que le signal de parole est localement stationnaire car les propriétés de ce signal varie très doucement en fonction du temps, d'où le recours aux méthodes d'analyse à court terme. Ainsi de courts segments de la parole sont analysés, on les appelle les trames d'analyse temporelle.

Les mesures comme l'énergie, le Taux de Passage par Zéro (TPZ) et la fonction de l'autocorrélation font partie des méthodes temporelles.

Les coefficients les plus utilisés en RAP sont certainement les cepstres. Ils peuvent être extraits de deux façons : soit par l'analyse paramétrique, à partir du Codage Linéaire Prédicative ou Linear Predictive Coding (LPC) en Anglais, soit par l'analyse spectrale. Dans le premier cas, nous parlerons de coefficients cepstraux prédictifs linéaires ou Linear Predictive Cepstral Coefficients (LPCC) en Anglais , et dans le deuxième cas de coefficients cepstraux à échelle de Mel ou en Anglais Mel scaled Frequency Cepstral Coefficients (MFCCs).

On utilise entre huit et seize coefficients cepstraux plus l'énergie, ensemble avec leurs dérivées temporelles premières et secondes, constituent le vecteur de paramètres élargi appliqué à un système moderne de reconnaissance vocale.

### **1.5.1. Méthodes non paramétriques**

Le signal de parole peut être analysé dans le domaine temporel ou dans le domaine spectral par des méthodes non paramétriques, sans faire l'hypothèse d'un modèle pour rendre compte du signal observé. Les méthodes spectrales sont fondées sur la décomposition fréquentielle du signal sans connaissance a priori de sa structure fine. Une analyse spectrale du signal permet de mettre en évidence certaines caractéristiques de la production de la parole qui peuvent contribuer à l'identification phonétique. L'articulation des phonèmes a une influence directe sur la forme du conduit vocal et des cavités, et donc sur les résonances qui apparaissent dans l'enveloppe du spectre [14].

L'analyse fréquentielle de la parole se ramène aux opérations de la Transformée de Fourier (TF) et n'a d'intérêt que si elle s'applique à une période du signal vocal, donc sur une période assez courte [15].

En RAP, il est important de connaître l'évolution de ce spectre dans le temps. Actuellement, les spectres sont obtenus numériquement par la Transformée de Fourier Discrète (TFD), en particulier grâce à l'algorithme de la Transformée de Fourier Rapide (TFR) ou Fast Fourier Transform (FFT) en Anglais. Cependant, le nombre de paramètres spectraux calculés sur une trame par FFT reste trop élevé pour un traitement automatique ultérieur (Fig. 1.5).

Pour une analyse très fine de la parole, la fenêtre de Hamming est déplacée à chaque fois de 128 points environ 10 ms.

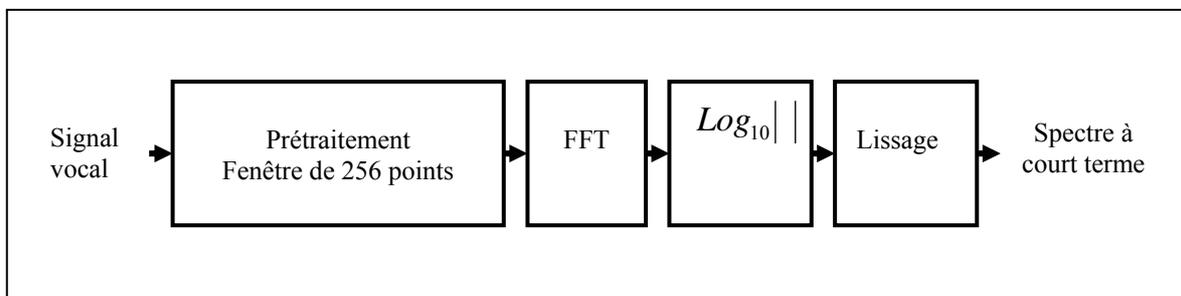


Figure 1.5 : Analyse numérique du signal parole par FFT

### 1.5.2. Méthodes paramétriques

Les méthodes paramétriques appelées aussi méthodes d'identification sont fondées sur une connaissance des mécanismes de production de la parole (ex : le conduit vocal). Les plus utilisées sont celles basées sur l'analyse prédictive linéaire et l'analyse cepstrale. L'hypothèse de base est que le canal buccal est constitué d'un tube cylindrique de section variable. L'ajustement des paramètres de ce modèle permet de déterminer à tout instant sa fonction de transfert. Cette dernière fournit une approximation de l'enveloppe du spectre du signal à l'instant d'analyse.

Ces méthodes consistent à ajuster un modèle aux données observées. Les paramètres du modèle, en nombre faible, caractérisent le signal, nous pouvons ainsi injecter des connaissances a priori sur le processus physique qui a engendré ce signal [16].

Les avantages de cette approche sont la souplesse de l'analyse, l'introduction naturelle de l'information et les choix variés des espaces de représentations paramétriques.

Dans le cas de la modélisation du signal parole, nous n'avons accès qu'à une seule sortie du système alors que l'entrée n'est pas mesurée. Il en résulte un problème d'estimation non linéaire car nous ne disposons pas d'observation de l'onde glottique d'excitation. En conséquence, nous en sommes limités à faire quelques hypothèses relativement neutres sur l'entrée ; par exemple, bruit blanc à moyenne nulle et reporter tout l'effort de modélisation sur le système.

### 1.5.2.1. Codage prédictif Linéaire

La méthode de la prédiction linéaire est particulièrement bien adaptée à l'étude de la parole. Elle consiste à prédire approximativement la valeur d'un échantillon de parole à partir d'un développement limité des échantillons précédents sur une base polynomiale orthogonale. Elle permet d'obtenir des coefficients de prédiction et la valeur des formants.

Nous partons de l'hypothèse qu'un échantillon de parole peut être prédit à partir d'une pondération linéaire d'un nombre fini d'échantillons précédents. Cette hypothèse se justifie par le fait que la forme du conduit vocal n'évolue pas rapidement.

Les coefficients de pondération des sections représentent le modèle du conduit vocal entre les lèvres et la glotte. Par conséquent, coder la parole par cette méthode, c'est en fait découper artificiellement la parole en échantillons sonores et représenter chacun d'eux par le fondamental ou pitch, l'énergie acoustique et un nombre retreint de coefficients de pondérations appelés aussi formants [16].

Cette méthode se base sur les connaissances de la production de la parole en représentant le système phonatoire par un filtre numérique. Elle suppose que le modèle de production est linéaire. Globalement ce modèle peut se décomposer en deux parties : la source active, le conduit passif (Fig.1.6).

Pour cela, nous faisons l'hypothèse que le signal vocal noté  $y_n$  est produit par un signal excitateur (source glottique) noté  $u_n$  traversant un système linéaire de réponse impulsionnelle noté  $b_n$ . Avec ces hypothèses nous pouvons écrire :

$$y_n = u_n * b_n \quad (1.01)$$

Le modèle du conduit vocal est un filtre tout pôle ou **Auto Régressif (AR)** d'ordre 2 décomposable en deux pôles en série (tuyaux résonnants).

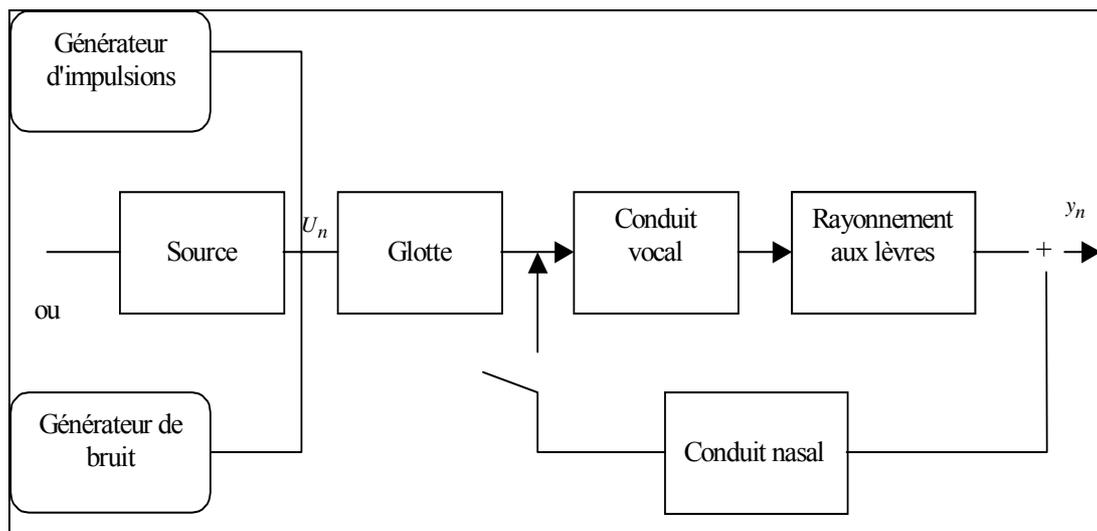


Figure 1.6 : Modèle général de production de la parole

Le modèle du conduit nasal est un filtre pôle zéro appelé modèle **Auto Régressif à Moyenne Ajustée (ARMA)** et le rayonnement aux lèvres peut se modéliser par un filtre tout - zéro ou **Moyenne Ajustée (MA)**.

L'ensemble du conduit se comporte comme un système ARMA qui est un *filtre discret linéaire invariant causal* caractérisé par une équation aux différences :

$$\sum_{n=0}^N a_n y(k-n) = \sum_{m=0}^M b_m u(k-m) \quad (1.02)$$

L'équation (1.01) peut se mettre sous la forme suivante, avec l'hypothèse que :  $a_0 \neq 0$

$$y(k) = \sum_{m=0}^M \frac{b_m}{a_0} u(k-m) - \sum_{n=1}^N \frac{a_n}{a_0} y(k-n) \quad (1.03)$$

La fonction de transfert notée  $G(z)$  de ce filtre à coefficients réels comportant  $P$  pôles et  $\varphi$  zéros est de la forme suivante :

$$G(Z) = \frac{\prod_{i=1}^{\varphi} (1-Q_i Z^{-i})}{\prod_{i=1}^P (1-P_i Z^{-i})} \quad (1.04)$$

Q : Zéros du filtre.

P : Pôles du filtre.

Le filtre ARMA étant stable (tous les pôles se trouvent à l'intérieur du cercle unité sur le plan des  $z$  en coordonnées polaires), sa réponse harmonique notée  $G(f)$  est donnée par :

$$G(f) = \frac{\sum_{n=0}^{\varphi} b(n) e^{-j2\pi n f}}{\sum_{n=0}^P a(n) e^{-j2\pi n f}} = \sum_{k=0}^{\alpha} g(k) e^{-j2\pi k f} \quad (1.05)$$

$g(k)$  : réponse impulsionnelle.

$G(f)$  : fonction définie, quelle que soit la fréquence  $f$ , elle est périodique et paire. Il suffit de l'évaluer entre 0 et  $\frac{1}{2}$  pour réaliser son implantation sur machine en discrétisant la variable  $f$ .

Appelons  $N$ , le nombre d'évaluations à effectuer dans l'intervalle  $[0, \frac{1}{2}]$ . On obtient :

$$G(f) = G\left(\frac{k}{2\pi}\right) = \frac{\sum_{n=0}^{\varphi} b(n) e^{-j2\pi \frac{nk}{2N}}}{\sum_{n=0}^P a(n) e^{-j2\pi \frac{nK}{2N}}} \quad k = 1, 2, \dots, N-1 \quad (1.06)$$

Le modèle ARMA est plus délicat à estimer que le modèle AR. Cela nous amène à préférer, pour une qualité donnée de la modélisation, un modèle AR avec un ordre un peu surestimé [16].

La difficulté essentielle dans l'estimation des modèles ARMA vient de la non linéarité du problème à résoudre résultant de la partie MA. Mais, la principale limitation réside dans l'hypothèse de stationnarité du signal acoustique qui est faite. Il faut réaliser un compromis entre la longueur de la fenêtre d'analyse et la durée pendant laquelle l'hypothèse de stationnarité est raisonnable. Ce compromis est réalisable pendant les zones stables (voyelles), mais il n'est pas satisfaisant durant les phases transitoires et injustifiées sur les plosives.

Ce modèle est appelé *prédictif linéaire* en raison du fait qu'il correspond à une régression linéaire simple entre le signal d'excitation et le signal vocal produit. Les coefficients de cette régression linéaire sont ceux du filtre numérique récursif. Elle permet d'évaluer la fonction de transfert du conduit vocal, en optant pour un modèle exclusivement récursif [17].

L'analyseur fournit un jeu de 08 à 16 coefficients de prédiction. Cette technique est une alternative intéressante au calcul du spectre à court terme pour plusieurs raisons :

- elle est basée sur un modèle de production simple, mais cohérent de la parole ;
- ce modèle est facilement inversible. Le cadre théorique est celui d'une modélisation source/filtre.

### 1.5.2.2. Prédiction Linéaire Perceptuelle

La Prédiction Linéaire Perceptuelle (PLP) est une exploitation des connaissances du système auditif humain pour paramétrer la parole : il s'agit de rendre compte des mécanismes psychoacoustiques de l'oreille humaine. Elle repose sur l'analyse par prédiction linéaire où la matrice d'autocorrélation est calculée par Transformée de Fourier Inverse (TFI) du module au carré de la TF du signal (Fig.1.7).

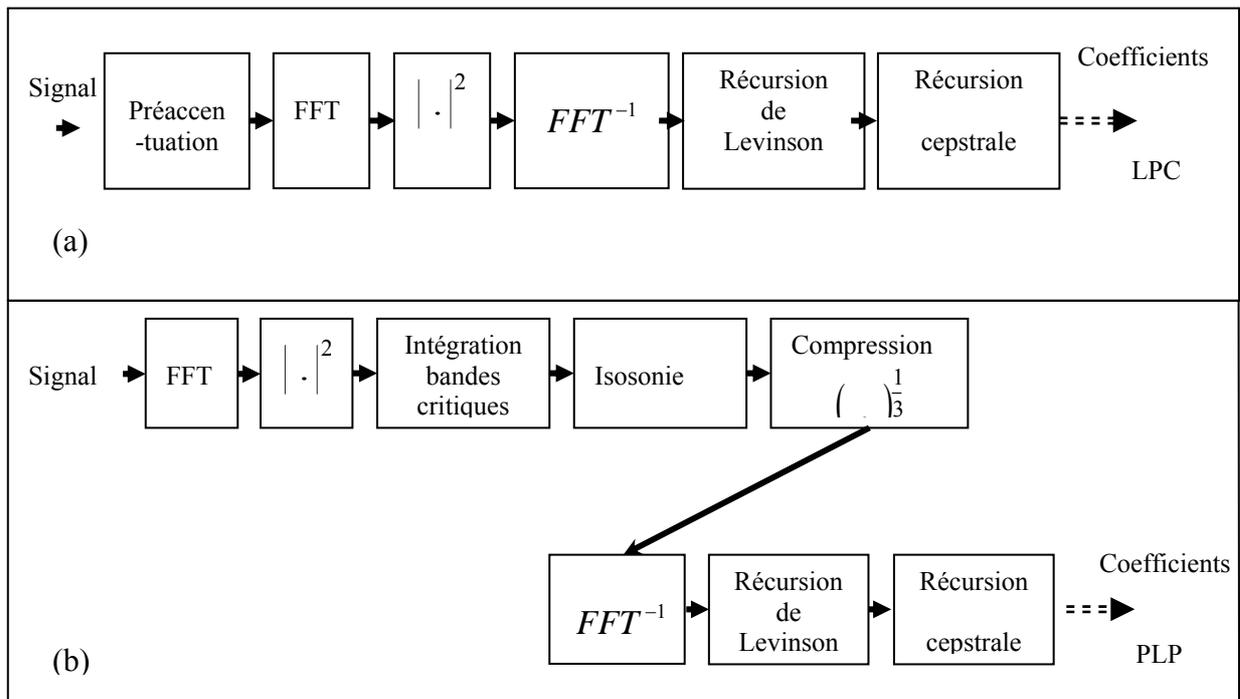


Figure 1.7 : Processus de calcul des coefficients : a) LPC, b) PLP [17]

Nous introduisons au niveau du spectre de puissance des bandes critiques. Cette intégration se fait avec un banc d'un certain nombre de filtres dont les fréquences centrales sont espacées linéairement selon l'échelle Bark. Cette fréquence perceptuelle est donnée par :

$$Bark = 13 \tan^{-1}\left(\frac{0.76f}{1000}\right) + 3.5 \tan^{-1}\left(\frac{f^2}{7500^2}\right) \quad (1.07)$$

Cette intégration se justifie par le fait que le système auditif se comporte comme un banc de filtres dont les bandes, appelées bandes critiques, se chevauchent et dont les fréquences centrales s'échelonnent continûment. Chaque bande critique correspond à l'écartement en fréquence nécessaire pour que deux harmoniques soient discriminées dans un son complexe périodique. L'échelle Bark, comme l'échelle Mel, reproduit approximativement la sensibilité de l'oreille [18].

### 1.5.2.3. Analyse cepstrale

Le défaut majeur des méthodes d'analyse, comme la FFT, pour le calcul du spectre réside dans l'intermodulation source/conduit vocal, qui rend difficile la mesure du fondamental  $F_0$  et des formants.

Le lissage cepstral est une méthode qui vise à séparer la contribution du conduit vocal de l'excitation glottique. Cette séparation est réalisée par un homomorphisme qui transforme la convolution des signaux dans le domaine temporel en une addition dans le domaine cepstral. En outre, cette méthode permet de fournir un vecteur spectral des MFCCs pour des fins de la RAP et de lisser le spectre de parole pour trouver les formants [19, 20].

Pour cela, nous faisons l'hypothèse que le signal vocal  $y_n$  est produit par le signal excitateur  $u_n$  traversant un système linéaire de réponse impulsionnelle  $b_n$ .

Le but du cepstre est de séparer ces deux contributions par déconvolution. Il est fait l'hypothèse que  $u_n$  est soit une séquence d'impulsions (périodiques, de période  $T_0$ , pour les sons voisés), soit un bruit blanc pour les sons non voisés, conformément au modèle de production de la parole. Une transformation en  $Z$  permet de transformer la convolution en produit.

$$Y(z) = B(z).U(z) \quad (1.08)$$

Le logarithme du module uniquement (car nous ne nous intéressons pas à l'information de phase) transforme le produit en somme. Nous obtenons alors :

$$\text{Log } |Y(z)| = \text{Log } |U(z)| + \text{Log } |B(z)| \quad (1.09)$$

Par transformation inverse, nous obtenons le cepstre. Dans la pratique, la transformation en  $Z$  est remplacée par une TFR. L'expression du cepstre est donc :

$$C(n) = FT^{-1} \{ \log (FT \{y(n)\}) \} \quad (1.10)$$

Le cepstre qui ne fait appel à aucune information a priori sur le signal acoustique, est basé sur une connaissance du mécanisme de production de la parole. L'espace de représentation du cepstre ou espace quéfrentiel est homogène par rapport au temps. Les premiers coefficients cepstraux contiennent l'information relative au conduit vocal. Cette contribution devient négligeable à partir d'un échantillon  $n_0$  qui correspond à la fréquence fondamentale  $F_0$ . Les pics périodiques visibles au-delà de  $n_0$ , reflètent les impulsions de la source.

Le spectre du cepstre pour les indices inférieurs à  $n_0$  permet d'obtenir un spectre lissé, débarrassé des lobes dûs à la contribution de la source. Ces deux contributions peuvent être séparées par une simple fenêtre temporelle notée F (lifrage) telle que le filtre adouci ou le filtre rectangulaire [21].

La présence d'un pic important dans le cepstre renseigne d'une part sur le caractère voisé ou non du son et d'autre part constitue une bonne indication sur la fréquence fondamentale. L'enveloppe spectrale du conduit vocal (structure formantique) est obtenue par une transformation supplémentaire (Fig. 1.8).

Le spectre lissé débarrassé théoriquement de la contribution de la source ne contient que des informations sur le conduit vocal et en particulier ses extréma (Formants).

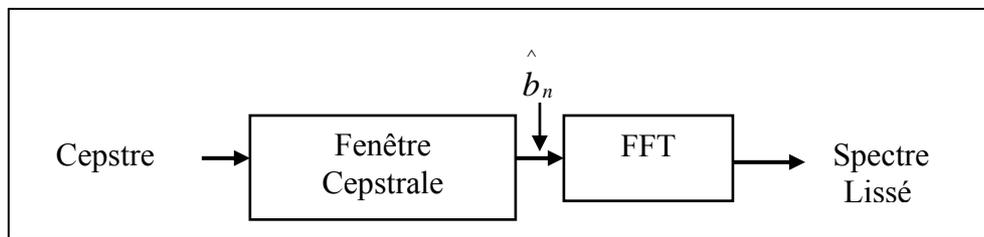


Figure 1.8 : Transformation schématisée pour l'obtention de la structure formantique à partir du cepstre

## 1.6. Vecteur à court terme image acoustique du signal vocal

Nous choisissons de représenter chaque trame du signal par un vecteur constitué de treize coefficients cepstraux plus l'énergie, ensemble avec leurs dérivées temporelles premières et secondes. Cette représentation paramétrique est bien adaptée à un système moderne de RAP.

### 1.6.1. Coefficients MFCCs à court terme

L'analyse menant aux paramètres MFCCs est dite homomorphique c'est-à-dire que la représentation dans le temps des paramètres spectraux et celle des paramètres cepstraux ont la même forme. Ces coefficients sont considérés comme robustes à une grande gamme de variabilité. D'une part, ils assurent une séparation entre les deux composantes du signal vocal : la fonction d'excitation qui est caractérisée par le pitch et la fonction de transfert du conduit vocal. La contribution du conduit vocal se retrouve dans les faibles quéfrenes (premiers coefficients cepstraux) alors que la contribution de l'excitation se localise dans les quéfrenes élevées. Un lifrage basse-bas permet donc de séparer ces deux composantes.

D'autre part, ils sont très peu sensibles à la puissance acoustique du signal vocal. En effet, cette puissance est concentrée dans le premier coefficient cepstral. L'adjonction des dérivées premières et secondes par rapport au temps des coefficients cepstraux rend ces derniers encore plus résistants aux fluctuations dues au locuteur ou à l'environnement [22].

Le paramètre le plus communément utilisé est le coefficient cepstral à fréquence de Mel. Les coefficients MFCCs sont une extension des cepstres qui sont utilisés pour mieux représenter les modèles de l'audition humaine. Le principe de calcul des MFCCs est issu des recherches psychoacoustiques sur la tonie et la perception des bandes de fréquences par l'oreille humaine. La résolution spectrale du système d'audition humaine n'est pas constante le long de l'axe de fréquence. La contribution à la perception des sons de la parole des hautes fréquences est plus faible que celle des basses fréquences. Afin d'imiter l'analyse spectrale de l'audition humaine, nous devons procéder à une analyse spectrale avec une résolution fixe le long d'une échelle de fréquence subjective appelé échelle de fréquence Mel. Cette échelle est linéaire en basses fréquences et logarithmique en hautes fréquences. Il existe une relation monotone entre l'échelle de fréquence Mel et celle de la fréquence physique (Hz).

$$f_{Mel} \cong 2595 \log_{10} \left( 1 + \frac{f_{Hz}}{700} \right) \quad (1.11)$$

L'échelle Mel est non linéaire. Elle tient principalement compte du fait que la perception des intervalles change suivant la zone du spectre à laquelle les hauteurs qui les composent appartiennent. Le principal intérêt de ces coefficients est d'extraire des informations pertinentes en nombre limité en s'appuyant à la fois sur la production (théorie cepstrale) et à la fois sur la perception de la parole (échelle de Mel).

Une fois le signal vocal prétraité, la TFD à court terme équivalente à la FFT pour un signal réel est calculée sur chaque fragment ou trame du signal. Cette dernière est ensuite filtrée par un banc de filtres triangulaires repartis le long de l'échelle de Mel. Les bancs de filtres sont ceux des filtres passe-bandes. Ils sont utilisés pour mesurer l'énergie dans différentes bandes de fréquences. Le module du logarithme de l'énergie de sortie du banc de filtres noté  $X_k$  est calculé avant que la Transformée en Cosinus Discrète inverse (**TCDi**) ne soit appliquée. Elle transforme les vecteurs paramétriques dans l'espace cepstral. Nous utilisons N filtres passe-bandes tels que la bande large de chaque filtre soit choisie égale à celle en échelle Mel de la fréquence centrale du filtre correspondant. Les coefficients cepstraux notée  $c_n$  sont calculés comme suit [23] :

$$c_n = \sum_{k=1}^N X_k \cos\left(n \cdot \frac{\Pi(k - 0.5)}{N}\right) \quad \forall 0 \leq n \leq N/2 \quad (1.12)$$

Il en résulte le vecteur à coefficients MFCCs noté  $V_k$  tel que :

$$v_k = (c_1, c_2, \dots, c_M) \quad (1.13)$$

Seuls les premiers coefficients (8 à 16) de la séquence cepstrale sont conservés pour lisser le spectre et minimiser l'influence du pitch non utile pour le processus de la RAP (Fig.1.9).

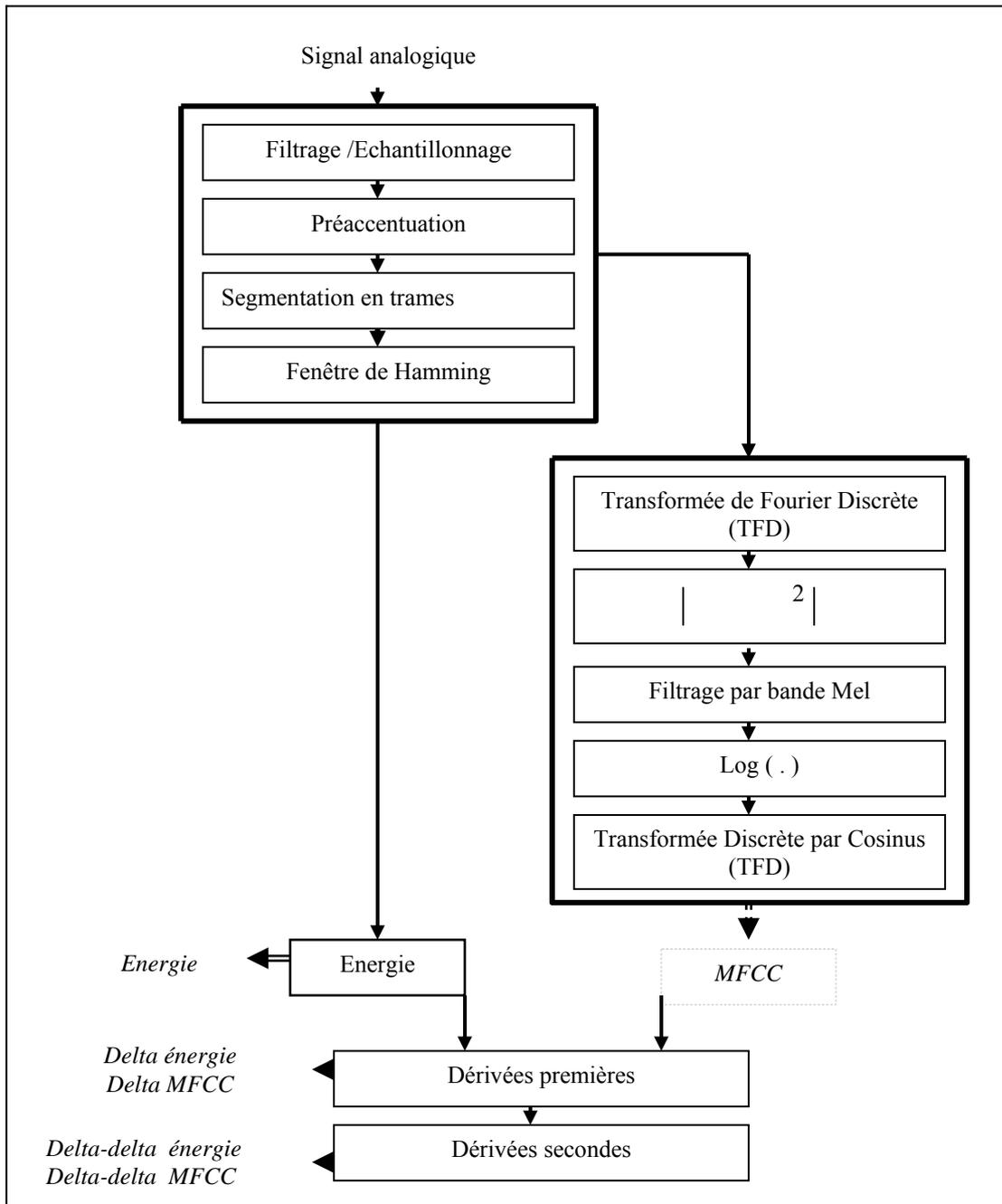


Figure 1.9 : Schéma d'extraction des coefficients MFCCs [24]

### 1.6.2. Energie à court terme

La différence en énergie entre les différents phonèmes démontre l'importance de ce paramètre pour distinguer entre les sons du langage. Le logarithme normalisé de la série de l'énergie du signal est usuellement utilisé comme étant le coefficient énergie. Il est calculé comme étant le logarithme de l'énergie du signal. Pour un signal noté  $x$  de longueur  $N$ , sa formule est donnée par :

$$E = \log \sum_{n=1}^N x_n^2 \quad (1.14)$$

Le logarithme d'énergie normalisé est obtenu par la soustraction à  $E$  la valeur maximum de l'énergie  $E_{\max}$  de la suite.

$$E_{NOR} = E - E_{\max} \quad (1.15)$$

L'amplitude du signal vocal varie en fonction du temps d'une manière appréciable. En particulier, l'amplitude des segments sonores est généralement plus basse que celle des segments sourds.

L'énergie à court terme du signal vocal nous offre une représentation convenable qui reflète ces variations d'amplitudes. La fonction d'énergie peut être utilisée aussi pour localiser approximativement le temps auquel la parole sonore devient sourde. Elle peut être utilisée pour distinguer entre la parole et le silence.

### 1.6.3. Taux moyen de Passage par Zéro à court terme

Le TPZ est un indice très simple à calculer qui donne une indication sur le taux de bruit contenu dans le signal. Il représente le nombre de fois que le signal, dans sa représentation amplitude/temps, passe par la valeur centrale de l'amplitude, généralement égale à zéro. Du fait de sa nature aléatoire, le bruit possède un TPZ supérieur à celui des sons voisés. Pour un signal  $x$  de longueur  $N$ , sa formule est donnée par :

$$TPZ(x) = \frac{1}{N} \sum_{n=1}^{N-1} \begin{cases} 1 & \text{si } x_{n+1} - x_n < 0 \\ 0 & \text{SINON} \end{cases} \quad (1.16)$$

Dans le cas des signaux discrétisés dans le temps, le passage par zéro peut avoir lieu si les échantillons successifs ont des signes algébriques différents. Cette mesure est une bonne manière pour estimer la fréquence de la courbe sinusoïdale. Les hautes fréquences impliquent un taux élevé de passages par zéro et les basses fréquences impliquent un TPZ bas, Il existe

une corrélation importante entre le taux de passage par zéro et la distribution d'énergie en fréquences. L'énergie et le TPZ peuvent aider à localiser le silence de la parole.

#### 1.6.4. Paramètres dynamiques

Les paramètres acoustiques décrits plus hauts sont appelés paramètres statiques du signal vocal puisqu'en représentant le spectre de puissance sur un court intervalle de parole, ils ont pour objectif la représentation d'un signal stationnaire. Il est intéressant de combiner ces paramètres avec les paramètres dynamiques qui traduisent l'évolution temporelle du spectre de puissance sur un intervalle plus long. La trajectoire temporelle du premier coefficient cepstral est un exemple de ce type de paramètres.

Ainsi, le vecteur des paramètres pertinents doit inclure des informations sur l'évolution temporelle du spectre. Cette information est incluse dans l'ensemble des coefficients à travers les dérivées premières et secondes du cepstre et celles de l'énergie qui sont les paramètres dynamiques les plus connus. La dérivée première de ces coefficients est appelée delta coefficients et la dérivée seconde est appelé delta-delta coefficients. Ces dérivées temporelles des paramètres statiques correspondants sont calculées sur trois à cinq trames successives. Les deltas coefficients sont calculés en utilisant une régression linéaire :

$$\Delta x(m) = \frac{\sum_{i=1}^k (i)[c(m+i) - c(m-1)]}{2 \sum_{i=1}^k i^2} \quad (1.17)$$

Où  $(2k + 1)$  est la taille de la fenêtre de régression ;

C : représente le cepstre.

Les delta-delta coefficients, les dérivées du second ordre sont calculées en utilisant la même régression linéaire appliquée à une fenêtre des delta coefficients.

### 1.7. Conclusion

Dans ce chapitre, nous avons, présenté les mécanismes de production des sons en général et les propriétés acoustico-phonétiques des phonèmes de l'Arabe Standard, en particulier en étudiant les spécificités des sons de l'AS sur les deux plans articulatoire et acoustique. En outre, Nous avons présenté une étude acoustique du signal vocal en commençant par un prétraitement de ce dernier. Dans ce contexte, il existe plusieurs types de représentations

paramétriques du signal qui dépendent des mécanismes de production. Notre intérêt s'est porté sur la méthode cepstrale à court terme qui se base sur le modèle de l'audition humaine afin de nous fournir les paramètres statiques et dynamiques nécessaires pour une meilleure représentation du signal vocal.

## Chapitre 2 : Etude du Décodage Acoustico-Phonétique

La Reconnaissance automatique de la parole est le domaine de recherche dont l'objectif est l'extraction de l'information contenue dans un signal de parole. Le processus d'interprétation d'une phrase est en général décrit comme une succession d'étapes depuis le niveau acoustique jusqu'au niveau sémantique.

Le DAP représente le premier maillon de reconnaissance analytique. Tous les traitements ultérieurs sont largement tributaires de la qualité de la transcription phonétique. Il concerne l'ensemble des processus de transformation du signal acoustique continu en une description linguistique discrète sous forme d'unités plus fines.

Le problème du DAP de la parole continue n'a toujours pas reçu de solution satisfaisante malgré les travaux considérables qui lui ont été consacrés depuis de nombreuses années. A cause des variations aléatoires, contextuelles et linguistiques des unités phonétiques. Une solution fréquemment utilisée consiste à ne retenir lors d'un premier traitement que les informations sûres identifiables sans ambiguïté dans le signal vocal [25].

Dans ce chapitre, nous avons étudiée les différentes étapes du DAP en insistant sur l'étape de la classification acoustique des différents phonèmes de l'AS.

### 2.1. Approches de la Reconnaissance Automatique de la Parole

La RAP est une discipline récente. Vers 1950 apparut le premier système de reconnaissance de chiffres, appareil entièrement câblé et très imparfait. Vers 1960, l'introduction des méthodes numériques et l'utilisation des ordinateurs changent la dimension des recherches. Néanmoins, les résultats demeurent modestes, car la difficulté du problème avait été largement sous-estimée, en particulier en ce qui concerne la parole continue.

La RAP est le processus qui consiste à décoder un signal acoustique de parole en une suite de mots effectivement prononcés. Compte tenu des caractéristiques de la parole, on distingue plusieurs démarches pour aborder les problèmes de reconnaissance. Ces démarches portent sur les points suivants :

- le mode d'élocution : mots isolés ou parole continue ;
- l'indépendance vis-à-vis du locuteur : systèmes mono ou multilocuteurs ;

- la complexité du langage autorisé : qui porte à la fois sur la taille du vocabulaire et sur la difficulté de la grammaire autorisée ;
- la reconnaissance ou compréhension : qui distingue entre une simple reconnaissance au niveau des mots ou l'interprétation de tout le message émis.

Tout système de reconnaissance de la parole résulte d'un compromis entre ces différentes démarches. Il est fonction du but et des objectifs à atteindre. Une large variété de techniques est utilisée dans le domaine de la RAP. Toutes ces techniques débutent par un échantillonnage du signal de la parole, suivie par un traitement acoustique (Fig.2.1).

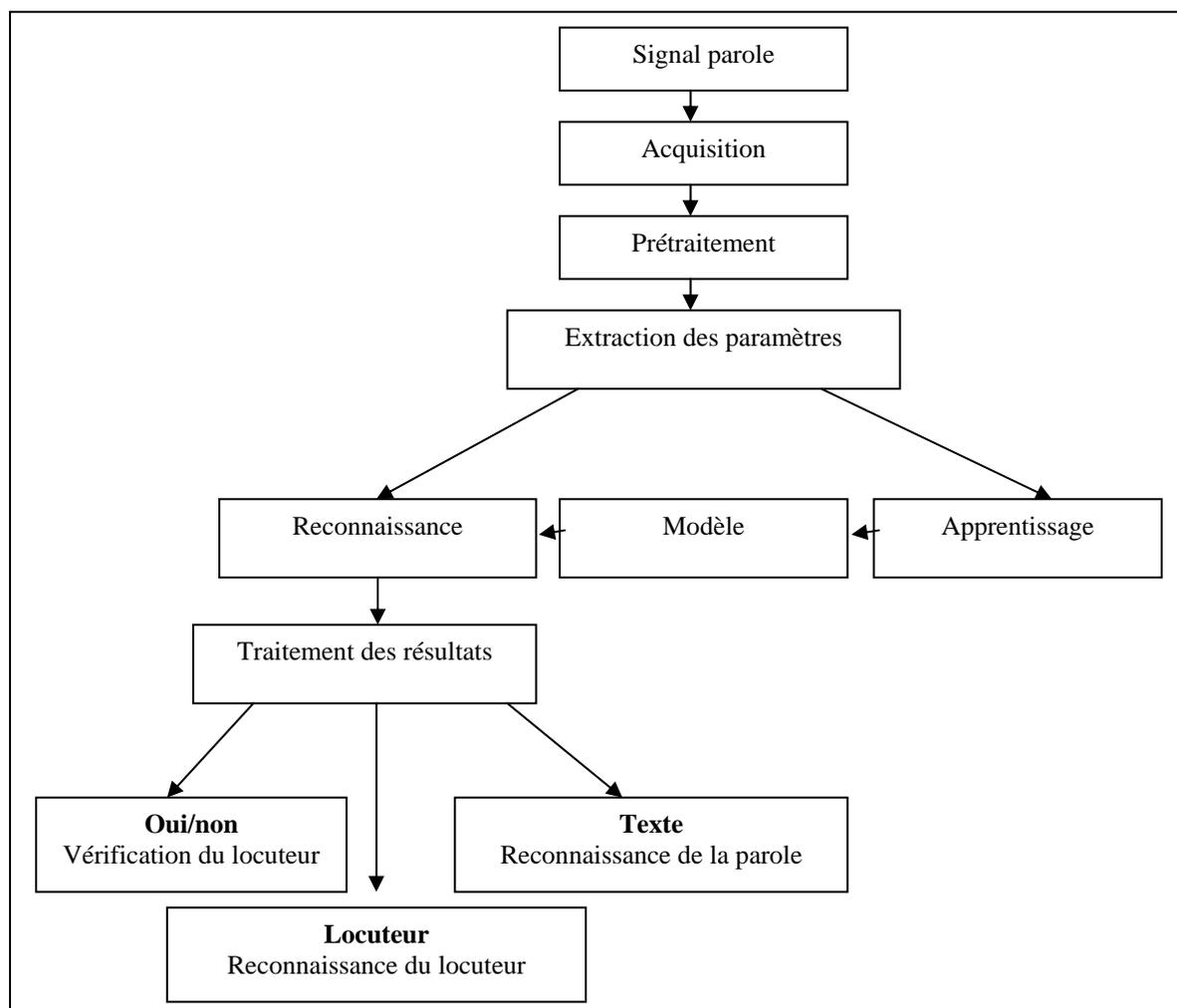


Figure 2.1 : Schéma de la structure générale d'un système de reconnaissance vocale [26]

Nous distinguons, usuellement, en RAP plusieurs approches dont l'approche analytique et l'approche globale. Ce qui les différencie est la manière d'appréhender le problème de la RAP [26].

### 2.1.1. Approche globale

Le terme reconnaissance globale s'applique aux systèmes pour lesquels l'unité de décision est l'entité lexicale qui peut être un phonème, un mot ou un groupe de mots. Cette méthode s'applique le mieux pour le cas de la reconnaissance de mots isolés à vocabulaire réduit. En effet, les images acoustiques des mots peuvent être facilement isolées les unes des autres car tous les mots prononcés sont supposés séparés par des silences.

Cependant, pour la reconnaissance de mots isolés à grand vocabulaire, cette méthode ne convient plus car la machine nécessite une mémoire et une puissance considérable pour respectivement stocker les images acoustiques de tous les mots du vocabulaire et comparer un mot inconnu à l'ensemble des mots du dictionnaire.

Le processus de reconnaissance globale ne tient pas compte de la structure et des propriétés phonétiques des mots, seules les formes acoustiques sont considérées (Fig.2.2).

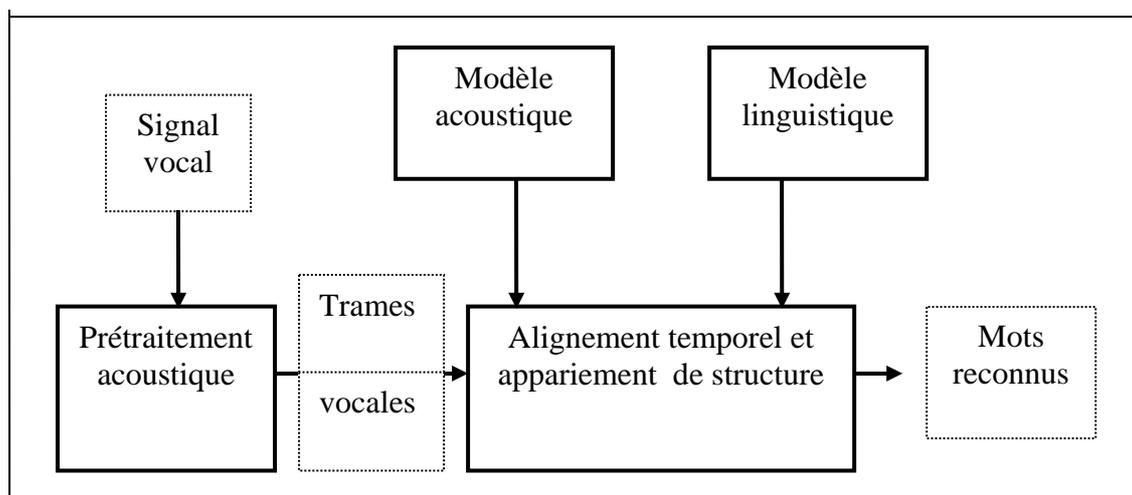


Figure 2.2 : Structure d'un système de reconnaissance vocale par l'approche globale [27]

L'idée de base de cette méthode consiste à donner au système au moins une image acoustique de chacune des entités lexicales qu'il devrait identifier par la suite. Cette opération est faite lors de la phase d'apprentissage. La reconnaissance d'une entité lexicale inconnue se fait alors par comparaison de sa forme aux différentes formes de référence. Le système va chercher la référence qui est la plus proche de l'image acoustique de l'entité à identifier, au sens d'un indice de dissemblance [28 , 29].

Cependant cette méthode fait face à deux problèmes majeurs :

- la difficulté d'extraire les paramètres pertinents du signal vocal afin de construire les images acoustiques des entités lexicales ;

- la définition de l'indice de dissemblance qui doit prendre en compte les différences entre les échelles temporelles des images acoustiques à comparer. Une technique couramment utilisée pour calculer la distance, ou plutôt la mesure de dissemblance, entre les formes de tests et de références, est la méthode de comparaison dynamique qui permet de tenir compte de la variabilité dans la vitesse de prononciation.

Son principal inconvénient est qu'elle ne permet pas de traiter de grands vocabulaires et son inadaptation au traitement des phrases. En effet, elle nécessite un énoncé mot à mot des phrases avec un silence entre chaque mot de l'ordre de quelques dizaines de ms.

### 2.1.2. Approche analytique

Cette approche cherche à résoudre le problème de la parole continue en isolant des unités acoustiques courtes comme les phonèmes, c'est-à-dire, reconnaître la parole continue en procédant à une segmentation en unités de base et à leur identification. Toutefois la coarticulation a pour conséquence un taux d'erreurs assez important (Fig. 2.3).

La structure d'un système de reconnaissance analytique se compose de deux parties principales :

- le système acoustico-phonétique qui, à partir du signal vocal, génère des unités phonétiques. C'est le Décodage Acoustico-Phonétique ;
- le système linguistique qui utilise en entrée la suite d'unités phonétiques générée par le système précédant, pour donner comme résultat l'interprétation sémantique du message vocal.

Cette méthode est plus générale et présente l'avantage de permettre la reconnaissance de la parole continue pour de grands vocabulaires, en ne mémorisant qu'un nombre restreint d'éléments, indépendamment de la taille du vocabulaire. Mais elle est aussi beaucoup plus délicate à mettre en œuvre, en particulier au niveau des algorithmes de segmentation du message.

Dans ce schéma général, le niveau du décodage acoustico-phonétique constitue une étape importante et une difficulté majeure dans la conception d'un système de reconnaissance.

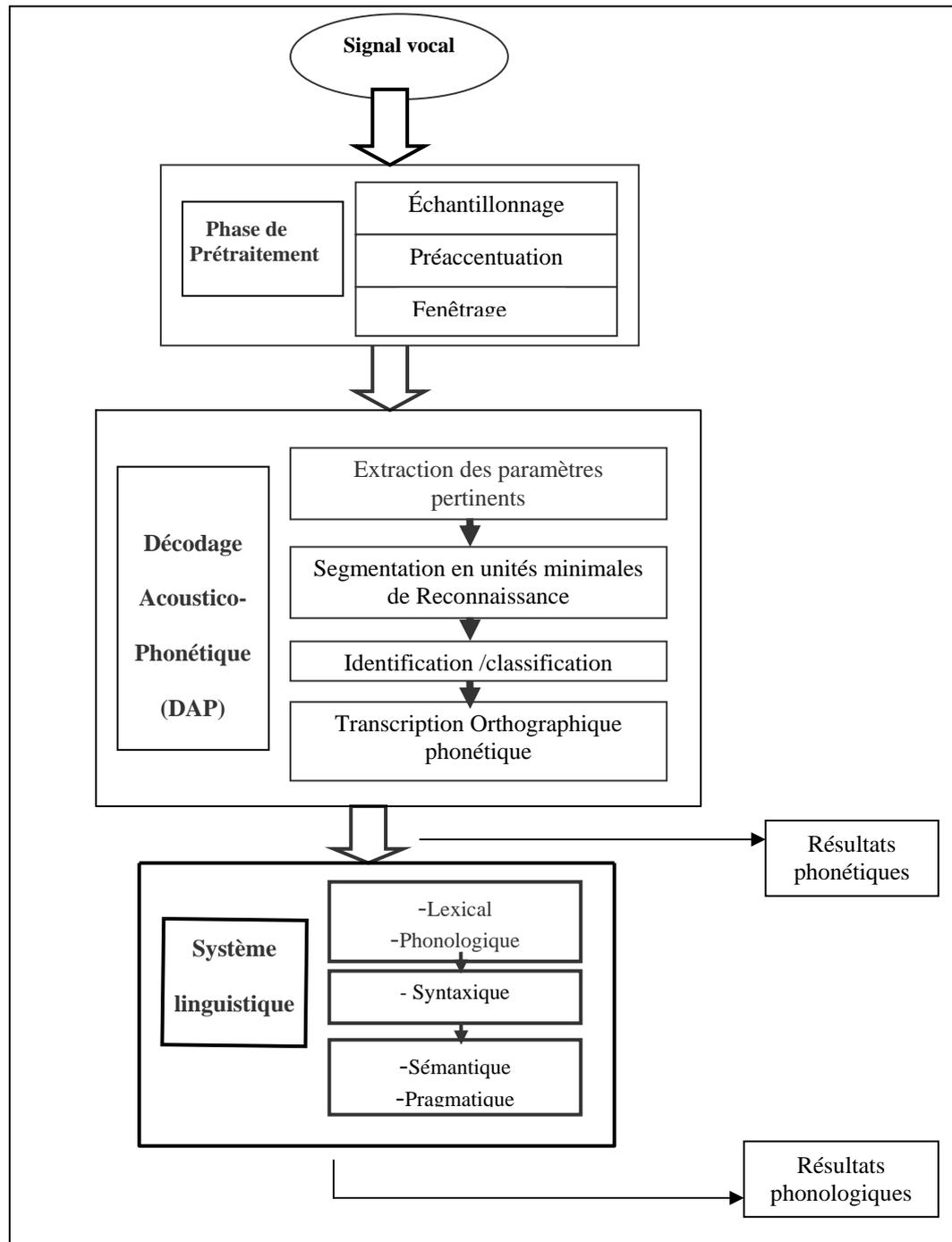


Figure 2.3 : Schéma de la structure d'un système de reconnaissance analytique de la parole [30]

## 2.2. Présentation du Décodage Acoustico-Phonétique

Les performances d'un système de RAP continue sont directement dépendantes de celles de l'étape de DAP. Il importe donc de concevoir des décodeurs phonétiques robustes et efficaces. Un facteur primordial en ce qui concerne le DAP est l'importance de la variabilité contextuelle des sons qui tire sa source d'un ensemble de phénomènes divers : la

coarticulation (labiale, nasale, etc. ), les différences dans les longueurs du conduit vocal, variations du spectre de source, phénomènes de dévoisement, etc. [30].

Le processus de DAP englobe deux tâches différentes qui peuvent être menées séquentiellement ou en parallèle, selon la méthode adoptée :

- une tâche de segmentation du signal vocal en unités élémentaires ;
- une tâche d'étiquetage phonétique de ces segments.

Le choix d'une unité phonétique est d'une grande importance. Diverses unités ont été utilisées, parfois concurremment dans le même système : syllabes, diphonèmes, triplets, phonèmes, etc. Syllabes, diphonèmes et triplets présentent l'avantage d'intégrer des informations sur les transitions, parties les plus difficiles à identifier dans la parole. L'utilisation d'allophones peut, dans une certaine mesure, faciliter la tâche d'étiquetage automatique mais il est nécessaire de trouver un compromis performance/complexité car le nombre d'allophones est potentiellement infini.

De nombreux systèmes font intervenir le phonème comme unité, à un moment ou à un autre du DAP. Plusieurs raisons optent pour cela. Le phonème est l'unité de base de description de la parole en terme acoustique. De plus, cette unité nécessite moins de temps d'apprentissage et moins d'espace mémoire de stockage. Enfin, les variations phonologiques des phonèmes peuvent être prédites à l'aide de règles contextuelles intra- et inter phonèmes. En fait, aucune unité de décodage n'est entièrement satisfaisante.

Comme nous l'avons déjà souligné précédemment, deux méthodes conceptuellement différentes permettent d'aborder la reconnaissance de la parole. Toutes les deux sont fondées sur la segmentation du signal en entités lexicales.

Pour la reconnaissance de vocabulaires limités, l'extraction directe de mots dans le signal a été utilisée avec succès. Cependant, quand nous envisageons de reconnaître de grands vocabulaires ou de la parole continue, il devient nécessaire d'extraire du signal des unités plus fines : C'est le rôle du DAP. Les différentes étapes du DAP sont :

- l'extraction des paramètres pertinents ou utiles (indices, corrélats, etc.) du signal parole à partir de son traitement acoustique ;
- la segmentation de la parole en unités définies ;
- l'identification ou classification de ces unités ou du moins leur donner des attributs significatifs (étiquettes de classes ou de macro-classes, traits, etc.).

Ces différentes fonctions interagissent entre elles et avec les modules linguistiques environnants.

### 2.2.1. Extraction des paramètres utiles du signal vocal

Cette étape consiste à extraire du signal vocal un nombre restreint de paramètres pertinents, tout en éliminant les informations redondantes. L'extraction des caractéristiques vise à réduire la dimension de l'espace de données tout en conservant le maximum d'informations pertinentes pour les niveaux supérieurs du traitement, c'est-à-dire, des informations discriminantes (Fig.2.4).

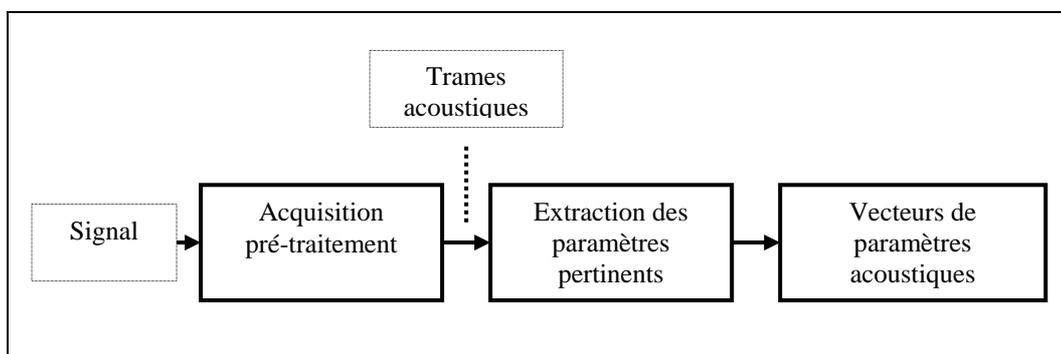


Figure 2.4 : Schéma du principe des premières étapes du traitement d'un système de RAP

Les principales méthodes d'analyse utilisées sont des méthodes temporelles ou spectrales. Les systèmes de la RAP utilisent pour la plupart les caractéristiques du signal estimées sur de courtes périodes de temps, pendant lesquelles le signal est supposé stationnaire. Certains paramètres dérivés soit, du signal vocal lui-même soit de ses représentations diverses (LPC, Cepstre, etc.) sont d'une pertinence vérifiée pour l'analyse phonétique.

La prise en compte de l'information temporelle, qui est largement utilisée pour la reconnaissance de la parole et du locuteur, consiste à inclure dans le vecteur de paramètres, des dérivées des vecteurs spectraux ou cepstraux [31].

L'utilisation de ces dérivées, appelées encore coefficients de delta et delta-delta, reste l'approche la plus utilisée actuellement en raison de la simplicité pour sa mise en oeuvre et l'amélioration de performances que nous pouvons observer [32].

### 2.2.2. Segmentation du signal vocal

La segmentation a pour but de fragmenter ou de décomposer automatiquement le continuum vocal en une suite de segments tels que chacun d'eux correspond à la réalisation d'un phonème, d'un mot, etc. A ce niveau, deux problèmes se présentent : d'une part, choisir l'unité de décision, d'autre part, opérer une segmentation correcte, ou du moins cohérente.

Nous ne nous intéressons qu'à la segmentation en phonèmes. Malheureusement, il n'existe pas de liens directs et simples entre un phonème et sa représentation acoustique, à cause de la continuité et la variabilité du signal vocal.

Diverses méthodes de segmentation sont employées, parmi lesquelles, celle dite explicite qui utilise les paramètres telles que l'énergie du signal, la variation spectrale, la fonction fondamentale et la densité de passage par zéro. Elle permet de trouver les frontières entre segments. L'avantage de ce type de segmentation est de diminuer la complexité de la reconnaissance en permettant un positionnement immédiat sur le segment à identifier, avec des frontières bien localisées.

Le principe de segmentation se fonde en général sur l'étude des changements d'une fonction mesurant les variations et les discontinuités de l'onde vocale et de son spectre. Des critères articulatoires ont également été utilisés. La segmentation repose dans tous les cas sur une connaissance acoustico-phonétique qu'il est possible de rendre explicite. Cette démarche conduit à des systèmes de segmentation par règles de réécriture ou à base de connaissances. Cette dernière approche permet d'intégrer des connaissances heuristiques que possède le phonéticien et d'améliorer ainsi les performances [33].

La plupart des segmenteurs utilisent une stratégie ascendante partant des données brutes du signal pour remonter vers une représentation plus symbolique. Une stratégie descendante fondée sur la prédiction de la succession de phonèmes a été également proposée. Toutes ces méthodes introduisent des erreurs de sur- ou sous-segmentation qu'il importe de détecter et de corriger au mieux. Cela implique en particulier deux démarches largement répandues dans les systèmes actuels : une interaction aussi étroite que possible entre segmentation et étiquetage phonétique et une stratégie de décision retardée consistant à repousser des décisions définitives de segmentation mais aussi d'étiquetage le plus tard possible de façon à rassembler le maximum d'éléments décisifs [34].

Toutes les méthodes de segmentation de parole sont fondées sur l'hypothèse implicite que, représenté dans un certain espace paramétrique, le signal montre une transition entre deux

unités phonétiques consécutives. Or, cette hypothèse n'est pas toujours vérifiée dans la parole continue ; lorsque la transition n'est pas nette, le positionnement précis de marques devient ambigu voire impossible. Cette situation apparaît fréquemment dans les phrases où la réalisation acoustique ne contient ni fricatives ni plosives.

Une alternative consiste à concevoir des modèles de reconnaissance sans segmentation explicite. La segmentation reste donc difficile et pose souvent des problèmes de substitution, insertion et omission de symboles phonémiques dont la modélisation n'est pas encore satisfaisante. Pratiquement, des expériences ont montré que les résultats obtenus par des systèmes sans segmentation sont actuellement meilleurs que ceux avec segmentation [35].

### 2.2.3. Etiquetage phonétique

L'étiquetage phonétique de la parole fait appel à des techniques le plus souvent issues de la classification ou de la reconnaissance de formes. La classification est l'élaboration d'une règle de décision qui transforme les attributs caractérisant les phonèmes en appartenance à une classe (passage de l'espace de codage vers l'espace de décision). Avant qu'un modèle de décision ne soit intégré dans un système de reconnaissance vocale, il faut avoir procédé auparavant à deux étapes : l'étape d'*apprentissage* et l'étape de *test* [36].

L'étape d'apprentissage consiste à caractériser les classes des phonèmes de manière à bien distinguer les familles homogènes de phonèmes. C'est une étape clé dans le système de reconnaissance (Fig.2.5).

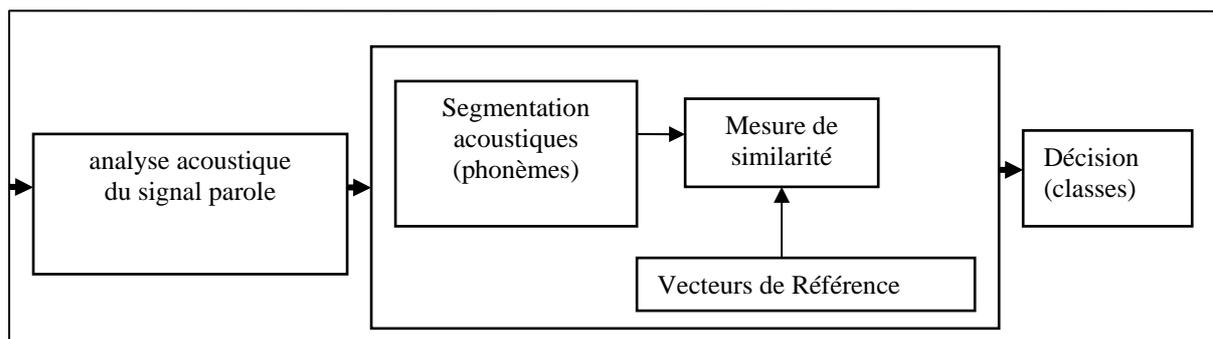


Figure. 2.5 : Schéma modulaire d'un système de classification acoustique [36].

Nous distinguons deux types d'apprentissage : *supervisé* et *non supervisé*. Dans le cas de l'apprentissage supervisé, un échantillon représentatif de l'ensemble des phonèmes à reconnaître est fourni au module d'apprentissage. Chaque phonème est étiqueté par un opérateur, cette étiquette permet d'indiquer au module d'apprentissage la classe dans laquelle l'opérateur souhaite que le phonème soit rangé. Cette phase consiste à analyser les

ressemblances entre les éléments d'une même classe et les dissemblances entre les éléments de classes différentes pour en déduire la meilleure partition de l'espace de représentation. Les paramètres décrivant cette partition sont stockés dans une table d'apprentissage à laquelle le module de décision se réfèrera ensuite pour classer les phonèmes qui lui sont présentés [36].

Dans le cas de l'apprentissage non supervisé, nous fournissons au système de reconnaissance un grand nombre de phonèmes non étiqueté. L'étape de la classification va se charger d'identifier automatiquement les entités appartenant à une même classe.

L'étape de test permet d'évaluer la performance du classificateur pour un apprentissage donné. C'est une étape importante car elle peut mettre en cause le choix des paramètres ou le choix de la méthode d'apprentissage.

Dans le domaine de la RAP, ce sont les méthodes basées sur un apprentissage supervisé qui sont le plus souvent utilisées.

### **2.3. Différents modèles de la Reconnaissance Automatique de la Parole**

Le choix d'un modèle adéquat pour représenter les diverses sources de connaissances disponibles est déterminant pour ce qui est des performances d'un système de reconnaissance de la parole. Les connaissances actuelles dans le domaine de la linguistique, de la psychologie et de la perception ne permettent pas de s'appuyer sur un modèle du cerveau humain pour tenter d'implanter un système de RAP. Parmi les principaux modèles existants, nous citons :

- les modèles stochastiques : actuellement en RAP, une des directions de recherche les plus prometteuses consiste à modifier l'ensemble des opérations de reconnaissance en termes de processus stochastique. Ce modèle purement statistique permet d'intégrer les niveaux acoustique et linguistique dans un seul processus de décision. Ces niveaux sont classiquement représentés par les MMC ou **Hidden Markov Models (HMM)** en Anglais. Les unités acoustiques modélisées peuvent être des mots comme dans l'approche globale ou, des unités plus courtes telles que le phonème dans l'approche analytique. Le temps de calcul avec ce modèle est important, mais les performances figurent parmi les meilleures obtenues jusqu'à présent ;
- les modèles connexionnistes : ce type de modèle est fondée sur une modélisation plus ou moins réaliste du cortex cérébral humain. Il est constitué par l'interconnexion d'un très grand nombre de processeurs élémentaires inspirés du fonctionnement des neurones. Un neurone est modélisé par une fonction d'activation des sommes des

signaux d'entrées, similaires aux signaux électrochimiques du système biologique. Un réseau de neurones est modélisé par les connexions entre les neurones. Chaque liaison entre deux neurones a un certain poids, nommé poids synaptique, exprimant l'importance de lier les neurones en question. Pendant la phase d'apprentissage, les poids synaptiques changent afin d'aboutir à une solution du problème et la conservation de ces poids constitue l'expérience du réseau de neurones ;

- les modèles à base de connaissances et systèmes experts : ils mettent à profit l'expertise humaine par l'acquisition et la formalisation des connaissances. Dans de tels systèmes, la représentation des connaissances utilise souvent le formalisme des règles de production ;
- les modèles biologiques basés sur les algorithmes évolutionnaires, s'appuient sur des techniques dérivées de la génétique et de l'évolution naturelle. Les AGs que nous adopterons pour notre modèle, sont les plus connus. Ils feront l'objet d'une étude détaillée ultérieurement.

## 2.4. Classification supervisée

Classifier des formes ou des individus (objets, images, phonèmes, etc.) décrits par un ensemble de grandeurs caractéristiques (taille ou masse de l'objet, pixels de l'image numérisée, spectre acoustique des phonèmes, etc.), c'est les ranger en un certain nombre de catégories ou classes définies à l'avance.

Le type d'une méthode de classification se décline généralement en deux familles : le mode supervisé et le mode non supervisé. Si nous disposons d'un ensemble de points étiquetés, nous parlerons de classification supervisée. Dans le cas contraire, nous effectuons une classification non supervisée appelée également classification automatique. Nous distinguons deux approches duales ; la recherche :

- de frontières et de fonctions discriminantes ;
- d'une partition des données en sous-ensembles appelées classes.

Dans le cas supervisé, les classes d'appartenance des données sont connues. La recherche des frontières entre les classes peut être effectuée par la recherche des fonctions discriminantes.

Les modèles les plus classiquement utilisés en classification sont, sans conteste, les partitions et les hiérarchies de parties. Dans les deux cas, les objets qui se ressemblent sont regroupés en classes. Pour les partitions, les classes sont deux à deux disjointes ; pour les hiérarchies, elles peuvent être emboîtées. Dans les deux cas, elles ne sont pas empiétantes au sens où l'intersection de deux d'entre elles n'en produira jamais de troisième. Une des vertus de la non-empiétance est de doter la classification de solides assises mathématiques.

Lorsque les objets à classer sont eux-mêmes décrits par une dissimilarité, le problème devient complètement homogène : transformer une dissimilarité en une dissimilarité d'un type donné. La classification s'inscrit alors dans le champ de l'approximation mathématique [36].

La résolution des problèmes de classification par apprentissage se déroule donc en plusieurs étapes :

- faire classer un échantillon d'individu par un expert, cet échantillon est désigné ;
- concevoir et mettre en œuvre un algorithme appelé classificateur qui parvient à reproduire la classification de l'échantillon d'apprentissage ;
- évaluer la qualité du classificateur en l'appliquant à un ensemble d'individus classés par l'expert, mais qui n'ont pas été utilisés durant la phase d'apprentissage (cet ensemble est la base de test) ;
- si le test est satisfaisant, appliquer la méthode de la deuxième étape à l'ensemble de la population à classer.

C'est cette procédure qui est appliquée lorsque nous confions la résolution d'un problème de classification à une machine. Elle porte alors le nom de classification supervisée car elle requiert l'intervention d'un superviseur ou expert. Afin de concevoir les classificateurs nous devons :

- choisir des variables descriptives (descripteurs ou facteurs) pertinentes, c'est-à-dire les variables qui déterminent la classe de l'objet à classer. Nous supposons qu'il existe une relation, entre ces variables et la classe désirée ;
- chercher une famille de fonctions, celle qui permet d'estimer au mieux la valeur de la sortie mesurée du phénomène à partir des valeurs des descripteurs.

## 2.5. Méthodes de Classification

Les méthodes de classification ont pour objectif commun la découverte d'un estimateur assurant l'affectation d'une classe parmi  $c$  classes disponibles à un individu inconnu sur la

base de la connaissance d'un ensemble de  $m$  caractères le décrivant (attributs). Cet estimateur est généralement construit au départ d'un jeu de  $n$  individus représentatifs de la population globale, décrits par ces mêmes caractères mais dont l'appartenance de classe est connue, et qui va servir d'exemple lors de la construction des règles d'attribution.

Les méthodes de classification qui se distinguent essentiellement par la forme conceptuelle de ces règles et par les procédures de construction qui en découlent, sont issues des statistiques et de l'Intelligence Artificielle (IA). Parmi les principales méthodes existantes, nous avons :

### 2.5.1. Mélanges des Lois Gaussiennes

Si la distribution de probabilité pour chaque classe est connue dans l'espace de représentation, alors le calcul de la probabilité d'appartenance à une classe est simple. Néanmoins, en pratique la distribution de probabilité n'est pas connue d'avance et une estimation de celle-ci est nécessaire à partir des données d'apprentissage de chaque classe.

Les **Mélanges des Lois Gaussiennes (MLG)**, ou **Gaussian Mixture Models (GMM)** en anglais, sont une famille de classificateurs où on suppose que la distribution des données pour chaque classe est une combinaison de plusieurs distributions Gaussiennes dans l'espace de représentation. L'apprentissage des GMM se fait généralement par l'algorithme d'Espérance Maximale (EM) qui garantit théoriquement une convergence vers une solution optimale. L'algorithme EM permet un calcul itératif des paramètres de chaque Gaussienne.

### 2.5.2. Modèles de Markov Cachés

Les MMC permettent de modéliser l'évolution temporelle du signal vocal. Ils sont basés sur une théorie probabiliste. Un MMC est caractérisé par la matrice de transition entre états, la distribution de probabilité d'émission des vecteurs de caractéristiques pour chaque état, et les états initiaux du système. Les paramètres des MMC sont estimés par apprentissage sur des données manuellement classifiées. Le MMC donne pendant la phase de reconnaissance, le chemin le plus probable des états et donc des classes sonores à chaque instant. Un algorithme efficace, largement utilisé, permet la détermination de la séquence d'état. Il est basé sur la programmation dynamique et est appelé algorithme de Viterbi.

Les MMC ont été largement utilisés en RAP. L'information temporelle de la parole est parfaitement modélisée. Un avantage majeur en est la possibilité de rendre compte de la variabilité du signal vocal par le traitement de très gros corpus de données acoustiques,

nécessaires à un bon ajustement du modèle Markovien. Ces modèles sont actuellement parmi les plus performants pour le DAP. En revanche, s'appuyant sur une modélisation purement mathématique, ils ne permettent pas d'introduire de façon explicite des connaissances phonétiques ou phonologiques [37, 38].

### 2.5.3. Réseaux de Neurones Artificiels

Les Réseaux de Neurones Artificiels sont des classificateurs inspirés par la manière dont nous supposons que le cerveau humain fonctionne.

Sur le plan pratique, une des architectures des réseaux de neurones les plus célèbres et la plus classique est le **Perceptron Multi Couches (PMC)** ou **Multi Layers Perceptron (MLP)** en Anglais. S'ils sont intéressants à cause leur pouvoir discriminant, ils demandent une adaptation pour intégrer la notion de temps, comme la normalisation à une longueur fixe, les MLP contextuels ou les **Réseaux de Neurones à Décalage Temporel (RNTD)** ou **Time Delay Neural Networks (TDNN)** en Anglais.

Cette architecture consiste en un réseau de neurones constitué de plusieurs couches, une couche d'entrée recevant du signal d'entrée, une ou plusieurs couches cachées responsables de la résolution des problèmes de classification, et une couche de sorties exposant les sorties du réseau [39].

Tout en ayant peu d'opérations de calcul pendant la phase de classification, les PMCs permettent la résolution des problèmes de classification non linéaires avec des bordures de décision complexes. En plus, dans la mesure où un PMC est caractérisé par son architecture et les poids synaptiques, la représentation du classificateur est très compacte.

Le connexionnisme a naturellement été appliqué à la RAP pour leurs bonnes capacités en classification de forme. Les RNA présentent beaucoup d'avantages potentiels mais sont cependant mal adaptés à traiter les signaux séquentiels. Il a été démontré que ces réseaux de neurones y compris les réseaux récurrents complexes ne sont pas capables de modéliser les dépendances à long terme, ce que fait par contre très bien un MMC par l'intermédiaire de ces contraintes topologiques traitant les contraintes phonologiques, lexicales et syntaxiques.

Les modèles connexionnistes n'ont pour l'instant pas surpassé les MMC. Ceci est, en grande partie, dû à leur relative inadéquation au problème du traitement séquentiel de l'information. C'est pourquoi une grande tendance dans le domaine du neuromimétisme

consiste à étendre les systèmes classiques ou à développer de nouveaux modèles pour prendre en compte les variabilités inhérentes à la parole.

Ces dernières années, nous constatons une volonté de dépasser les limitations actuelles du connexionisme. Ainsi, de nouveaux types de systèmes voient le jour, inspirés de la neurobiologie, de la psychologie ou mixant des techniques connexionnistes avec d'autres symboliques, ces modèles sont couramment appelés *modèles hybrides* ou stochastiques [40].

Plusieurs résultats récents (obtenus sur différentes bases de données allant de petits lexiques aux très grands lexiques) ont montré que les systèmes MMC/RNA conduisent généralement à des performances de reconnaissance équivalentes ou meilleures que celles des systèmes MMC utilisés dans les mêmes conditions (taux d'erreurs inférieur à 05%), avec cependant plusieurs avantages supplémentaires au niveau des besoins en CPU et mémoire. [41, 42].

#### **2.5.4. K- Plus Proches Voisins**

La méthode des **K- Plus Proches Voisins (KPPV)** offre l'avantage d'être très simple et néanmoins efficace. Son principe consiste à calculer une distance entre l'individu à classer et les individus connus, puis à attribuer le premier à la classe présentant le plus grand effectif parmi ses k plus proches voisins. Outre sa simplicité, cette méthode est couramment employée en reconnaissance des formes, parce qu'elle s'y prête à de nombreux titres : tout d'abord, elle ne nécessite pas de connaître la distribution de probabilité des classes de la population, ce qui est rarement le cas. Ensuite, comme nous fixons le nombre k de voisins, la méthode ne dépend pas de la densité de probabilité.

#### **2.5.5. Quantification Vectorielle**

Cette technique s'appuie sur les propriétés statistiques des sons dans un certain espace de représentation. Elle est largement utilisée en codage et en synthèse de la parole et présente aussi un intérêt en reconnaissance, notamment pour effectuer un premier étiquetage en grandes classes phonétiques. Une telle méthode est utilisée dans le système de reconnaissance sans segmentation [43].

#### **2.5.6. Fonctions de Mesure Distance**

Pour analyser des données, nous commençons par choisir les caractéristiques des objets que nous voulons analyser en les plaçant dans un espace de représentation. Ensuite, nous

devons se doter d'outils métriques permettant de mesurer des distances (des ressemblances, des dissemblances, etc.) entre les dits objets. Ainsi, dans un espace métrique, la distance entre deux vecteurs notés  $x$  et  $y$  doit satisfaire à l'expression suivante :

$$\begin{cases} d(x, y) \geq 0 \\ d(x, y) = d(y, x) \\ d(x, u) \leq d(x, u) + d(u, y) \end{cases} \quad (2.1)$$

$u$  : vecteur

Or, ces conditions ne sont pas toujours respectées par les distances utilisées en traitement de la parole. Par conséquent, il est plus correct de parler de mesure de distorsion ou de mesure de dissemblance. Ces distances sont intimement liées à la nature des paramètres retenus.

Il existe plusieurs systèmes d'apprentissage dont la performance dépend du choix de la fonction de distance.

Une mesure assez générale de la distance dans un espace métrique est donnée par l'indice de Minkowski. Celui-ci englobe, sous forme générale, le cas particulier des mesures Euclidiennes. Dans l'espace défini par  $j$  variables, l'indice Minkowski entre les individus  $x$  et  $y$  se définit de la manière suivante :

$$d(x, y) = \left( \sum_{k=1}^K w(x_k - y_k)^\nu \right)^{\frac{1}{\mu}} \quad (2.2)$$

Pour  $\nu = \mu = 2$  et si les poids sont égaux à 1. Cette mesure donne la mesure Euclidienne traditionnelle qui est très utilisée pour les données quantitatives. La distance de Manhattan est équivalente à celle de Minkowski pour  $\mu = 1$

Dans l'espace défini par  $j$  variables, la distance Euclidienne entre les individus  $x$  et  $y$  est :

$$d^2(x, y) = \sum_j (x_j - y_j)^2 \quad (2.3)$$

La distance de Mahalanobis du vecteur  $x$  à la classe 2 caractérisé par son vecteur moyen  $y$ , permet de mesurer la distance à échantillon défini par sa position (vecteur moyen) et sa dispersion (matrice de covariance). Elle est donnée par l'expression suivante :

$$d(x, y) = \sqrt{(x - y)^T \Lambda (x - y)^{-1}} \quad (2.4)$$

Afin d'éviter les effets dus à l'inégalité du nombre d'occurrences de chaque classes, le statisticien P.C. Mahalanobis a mis au point un calcul des distances basé sur la variance des registres ou clusters en Anglais. Pour chaque cluster, la moyenne et la matrice de covariance sont calculées. Ces données définissent la transition vers l'espace centré sur la moyenne du cluster, et dont la base est normée selon les variances des directions du registre. Ainsi nous pouvons calculer pour tout point test sa distance au registre en calculant dans l'espace d'arrivée la distance entre son origine (correspondant à la moyenne du registre dans l'espace des paramètres) et le point test. Cette distance est bien adaptée à la RAP, puisque nous sommes contraint à travailler avec de grandes bases de données.

## 2.6. Algorithmes biomimétiques de classification

Le problème de la classification de données est identifié comme une des problématiques majeures en extraction des connaissances à partir de données. Depuis des décennies, de nombreux sous problèmes ont été identifiés, comme par exemple la sélection des données ou des descripteurs, la variété des espaces de représentation (numérique, symbolique, etc.), l'incrémentalité, la nécessité de découvrir des concepts, d'obtenir une hiérarchie, etc.

La popularité, la complexité et toutes ces variantes du problème de la classification de données ont donné naissance à une multitude de méthodes de résolution. Ces méthodes peuvent à la fois faire appel à des principes heuristiques ou encore mathématiques. Parmi celles-ci, il existe une branche qui s'inspire plus spécialement de principes issus de la biologie. Les motivations des chercheurs sont d'une part de tester de nouveaux algorithmes sur le problème de la classification et de connaître leurs apports. Mais elles sont aussi de proposer de nouvelles sources d'inspiration, car le problème de la classification se rencontre souvent chez les espèces vivantes animales et végétales [44].

Nous allons donc donner un aperçu de ces méthodes. Nous ne traiterons pas ici les approches neuronales mais plutôt les approches à base de population d'agents ou algorithmes évolutionnaires, fournis artificielles, intelligence en essaim, systèmes immunitaires).

### 2.6.1. Approches Evolutionnaires

Pendant les années 70, les premiers travaux sur l'évolution artificielle ont concerné les AGs, les **S**tratégies d'**E**volution (**SE**) et la **P**rogrammation **E**volutive (**PE**). Ces trois types d'algorithmes sont basés sur des principes globalement communs inspirés du néo-darwinisme:

- utilisation d'une population d'individus ;
- évaluation des individus par une fonction ;
- sélection des meilleurs et génération d'une nouvelle population avec des opérateurs de croisement et de mutation.

Cependant, des choix méthodologiques ont initialement opposé ces méthodes. Ainsi, les premiers AGs utilisaient plutôt un codage binaire des individus alors que les SE utilisaient un codage réel. Ensuite, dans les années 90 est apparue la **Programmation Génétique (PG)** qui introduit notamment des représentations arborescentes [45, 46].

Les premiers travaux proposant un AG et plus généralement un algorithme évolutionnaire pour le problème de la classification ont été effectués en 1979. Le nombre de classes est fixé à l'avance et la représentation de longueur  $n$  associe une classe à chaque donnée. Les opérateurs génétiques sont une adaptation directe des opérateurs génétiques binaires [47, 48].

En 2003, Il a été développé à notre connaissance le seul AG apprenant une classification hiérarchique présentée sous la forme d'un arbre de centroïdes. Cet algorithme est restreint aux données numériques mais ne fait pas d'hypothèses sur le nombre de classes [49].

Les autres catégories d'algorithmes évolutionnaires ont été nettement moins développés que les AGs. Il est certain que les AGs pour la classification à eux seuls ont fait l'objet d'un volume de travaux plus important que toutes les autres méthodes réunies, cela étant certainement dû à leur popularité mais aussi aux succès rencontrés en tant que méthode globale d'optimisation [50].

Cependant, ces algorithmes n'ont pas nécessairement tous beaucoup d'avantages. Le problème du choix des paramètres reste difficile (cela ne concerne pas le choix du nombre de classes mais plutôt des paramètres liés à la méthode comme la taille de la population, les opérateurs, etc.). La diversité des codages utilisés montre par ailleurs que les AGs sont sensibles aux choix de la représentation et des opérateurs de reproductions [51].

Il faut noter également que les principales différences entre les AGs et les autres méthodes biomimétiques viennent du fait que dans les AGs, un individu représente généralement une partition entière et la population est un ensemble de partition, alors que dans les autres algorithmes, un individu représente une donnée et la population dans son ensemble représente la classification. Dans un AG, la solution au problème est le meilleur individu de la population, alors que dans les autres algorithmes, la solution est l'ensemble des individus. Cette différence est fondamentale puisque elle va obliger l'AG à centraliser son

fonctionnement. Les autres algorithmes vont au contraire utiliser des principes heuristiques plutôt locaux et agissant en parallèle sur toute la classification. Sans doute que cela a des répercussions sur le temps d'apprentissage dans les AGs, ce qui justifie l'étude d'approches génétiques hybrides.

Plusieurs approches hybrides ont cependant été proposées en utilisant conjointement les AG avec des approches plus classiques comme K-Means ou encore Fuzzy-C-Means. Ces heuristiques sont utilisées par exemple juste après l'AG qui sert donc à trouver une bonne partition initiale. Elles peuvent également servir au même titre que les opérateurs génétiques dans la boucle de l'AG ; elles sont appliquées sur chaque individu. Cela permet d'accélérer la convergence des AGs tout en conservant les avantages d'une méthode globale. Ces hybridations qui restent liées aux données numériques, ont données de très bons résultats avec des taux d'erreur moins de 05 % [52, 53, 54].

### **2.6.2. Fourmis Artificielles**

Les travaux fondateurs datent des années 1990. Il s'agit d'abord des travaux de biologistes s'intéressant de près à la modélisation des fourmis en termes mathématiques et informatiques, et à l'utilisation concrète de ces modèles. Deneubourg apparaît donc comme un pionnier dans le domaine du tri d'objets par des fourmis artificielles.

Il propose avec ses collègues les principes suivants : des fourmis artificielles se déplacent sur un plan. Les objets à rassembler sont repartis sur ce plan. Une fourmi ne dispose que d'une perception locale de ces objets et ne communique pas avec les autres. Compte tenu de cela, la configuration des objets sur le sol va influencer leurs actions. Lorsqu'une fourmi rencontre un objet, elle le ramasse avec une certaine probabilité qui dépend de la fréquence de rencontre d'objets dans un passé récent. Autrement dit, plus une fourmi rencontre d'objets, moins elle a de chance d'en prendre un. Ensuite, une fois un objet ramassé, la fourmi se déplace au hasard dans le plan, et elle le dépose avec une certaine probabilité d'autant plus grande que la fourmi a rencontré récemment des objets.

Ces principes relativement simples font qu'il apparaît des regroupements d'objets. L'approche peut être généralisée à plusieurs types d'objets: cet algorithme permet alors de trier des objets.

Le passage du tri d'objets à la classification a eu lieu en adaptant l'algorithme présenté précédemment : les données sont initialement reparties aléatoirement sur une grille à deux dimensions. Chaque fourmi est située dans une case de cette grille et ne perçoit que les

données situées dans son voisinage. Ensuite, la fréquence utilisée dans l'algorithme de tri vu précédemment peut être remplacée par une moyenne des similarités entre une donnée déportée par une fourmi et les données notées  $d_i$  situées dans son voisinage. Une donnée  $d_i$  sur la grille est ramassée avec une probabilité d'autant plus grande qu'elle est peu similaire aux données voisines. De la même manière, une donnée  $d_i$  portée par une fourmi est plus facilement déposée dans une région comportant des données qui lui sont similaires. Cet algorithme a été depuis étendu à d'autres applications comme le partitionnement de graphes ou la classification de sessions sur des sites Web [55].

Ultérieurement, il a été introduit un nouveau modèle à base de fourmis pour la classification utilisant le système d'identification chimique des fourmis. Celui-ci est fondé sur la construction d'une odeur coloniale qui est le fruit des apports génétiques, environnementaux et comportementaux. Cette odeur est construite par les individus pour identifier qui fait partie du groupe et qui doit être rejeté. A partir de ce modèle, un nouvel algorithme de classification a été proposé dans lequel chaque donnée est une fourmi dont l'odeur est déterminée par les valeurs prises par les attributs décrivant cette donnée. Les fourmis effectuent des rencontres aléatoires et décident d'appartenir au même groupe ou non. Il en résulte l'établissement d'une classification [56].

Enfin, il a été introduit un nouveau modèle permettant d'effectuer rapidement une classification hiérarchique. Il s'agit de copier la manière dont les fourmis construisent des structures vivantes en s'accrochant les unes aux autres en fonction de critères locaux (la forme de la structure influençant le comportement d'accrochage ou de décrochage).

Dans ce modèle, chaque fourmi artificielle représente une donnée. Les fourmis sont placées initialement à la racine de l'arbre et vont pouvoir se déplacer dans cet arbre et s'accrocher afin de construire une structure hiérarchique dont chaque noeud représente une donnée. L'objectif est de construire automatiquement un site portail (données textuelles) et d'obtenir la propriété suivante : chaque noeud noté  $o$  de l'arbre est une catégorie composée de toutes les données portées par les sous-arbres de ce noeud. Les sous catégories (représentées par les noeuds connectés à  $o$ ) doivent être très similaires à leur mère dans l'arbre, mais également les plus dissimilaires entre elles [57].

### 2.6.3. Intelligence en essaim

L'Intelligence en Essaim ou Swarm intelligence en Anglais regroupe de nombreux algorithmes à base de population d'agents. Les fourmis artificielles en font partie comme

d'autres algorithmes plus spécifiques qui utilisent les déplacements d'un essaim d'agents pour résoudre un problème.

Dans la biologie, de nombreux chercheurs se sont intéressés à la manière dont les animaux se déplacent en groupe. Aucun individu ne contrôle les autres mais pourtant des formes et des comportements complexes peuvent apparaître lors de ces déplacements. Des règles comportementales généralement simples permettent aux individus de se déplacer en groupe, d'éviter des obstacles, etc. En 1998, ces principes ont été appliqués pour la première fois à un problème de classification.

Les agents représentent chacun une donnée. Un agent réagit aux autres agents présents dans son voisinage en tenant compte de la similarité des données. Un agent se déplacera plutôt vers des données qui lui sont similaires. Cette règle comportementale permet donc de former des groupes de données similaires. Cet algorithme a été amélioré et évalué d'une manière plus systématique. Une distance idéale entre individus est définie, distance qui dépend de la similarité entre les données. Un critère d'arrêt est utilisé également en mesurant l'entropie spatiale du nuage d'agents. Cet algorithme a été intégré dans un système de fouille visuelle de données utilisant la réalité virtuelle [46].

#### **2..4. Systèmes Immunitaires**

Les **Systèmes Immunitaires (SI)** sont un ensemble de modélisations du système immunitaire humain et animal appliqués à différents problèmes en informatique. Ils utilisent les principes suivants :

Des agents ou lymphocytes qui génèrent des anticorps vont apprendre à reconnaître le soi du non-soi ou les antigènes. Pour cela, ces agents doivent d'abord être engendrés en utilisant un principe de composition de briques élémentaires. Ensuite, ils subissent un test de sélection. Les agents rejetant le soi sont éliminés, et les autres, qui vont rejeter le non-soi, sont gardés. Chaque fois qu'il y a reconnaissance d'un antigène par un anticorps, la présence des lymphocytes générant ces anticorps est favorisée par un processus de sélection par clonage et par la disparition des lymphocytes non stimulés par les antigènes. Ce clonage donne donc lieu à des interactions entre les lymphocytes et peut mettre en oeuvre des mutations. Certains lymphocytes, lorsqu'ils sont souvent utilisés, prennent un rôle d'élément de mémorisation à long terme.

Ces systèmes disposent de propriétés complexes car ils sont capables de générer des solutions et de les sélectionner en fonction de leur efficacité selon des heuristiques originales.

En ce qui concerne la classification, les principes des systèmes immunitaires sont à un niveau général [58].

Un anticorps est également représenté par un vecteur de dimensions  $n$ . Les anticorps suffisamment proches de l'antigène, au sens de la distance euclidienne, vont subir des clonages avec mutation ou interaction anticorps/antigènes afin d'amplifier et d'affiner la réponse du système. Egalement, ces anticorps vont subir une sélection ou interaction anticorps/anticorps : ceux qui sont trop proches les uns des autres seront diminués en nombre. Après ces itérations, le système converge en plaçant des anticorps de manière judicieuse et en nombre adapté aux données [59].

## **2.7. Conclusion**

Dans ce chapitre, nous avons présenté dans ce chapitre les principes du Décodage Acoustico-Phonétique et le rôle qu'il joue dans les systèmes de RAP. En outre, nous avons étudié les différentes étapes du DAP ainsi que la nature des paramètres utiles à extraire du continuum vocal.

Nous avons focalisé notre étude sur la classification acoustique supervisée et explicité la manière dont elle sera utilisée pour classer les phonèmes de l'AS de la parole continue. Pour cela, différents modèles de classification ont été cités en insistant sur les algorithmes évolutionnaires et plus précisément les AGs.

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

Dans ce chapitre, nous présentons les principes généraux de fonctionnement des AGs , ainsi que le rôle des différents opérateurs génétiques utilisés, y compris leur dimensionnement et leurs critères d'arrêt. Nous mettrons en relief l'importance du choix de l'espace de représentation, du type de codage des paramètres du problème à résoudre et de la nature de la fonction objective. En outre, nous avons étudié la sélection naturelle et les différentes techniques d'amélioration des AGs.

### **3.1. Concepts de base des Algorithmes Génétiques**

Les AGs sont des méthodes de recherche stochastique reposant sur les mécanismes de la sélection naturelle et le principe darwinien de survie du plus robuste. Ce sont des méthodes d'optimisation de fonctions. Ils sont utilisés comme alternative lorsque les méthodes classiques ou déterministes ne sont pas applicables ou ont échoué. Ils sont désormais utilisés dans les problèmes d'optimisation complexe et permettent de réaliser une recherche multi-dimensionnelle afin de trouver une valeur optimale. Ils utilisent un parallélisme implicite en ce sens qu'ils manipulent simultanément de multiples solutions dont ils calculent les valeurs d'adaptation. Ainsi, la résolution d'un problème d'optimisation par un AGs s'effectue sur une population d'individus dont chacun est une solution potentielle au problème. Ces individus sont encodés par une chaîne binaire, entière ou réelle de longueur finie, mais qui peut être variable dans le temps, appelée chromosome [60].

Ces algorithmes s'inspirent de l'évolution génétique des espèces naturelles. Schématiquement, ils copient de façon extrêmement simplifiée certains comportements de ces populations dont l'évolution est assurée au moyen d'opérateurs tels que la sélection, le croisement et la mutation qui permettent de combiner les chromosomes. Ainsi, les AGs se basent sur l'évolution d'une population de solutions qui sous l'action de règles précises optimisent un comportement donné, exprimé sous forme d'une fonction, dite fonction objectif ou fitness function en Anglais ou d'adaptation à l'environnement [61].

Les AGs sont des méthodes heuristiques puisque il n'existe pas de résultats théoriques prouvant qu'un AG conduit de manière certaine à un extremum global. Actuellement, ils sont utilisés dans plusieurs domaines comme celui de la programmation automatique, de

l'apprentissage et de la classification et celui du monde réel ; marchés économiques et comportement sociaux [62].

Un Algorithme Génétique est un algorithme itératif de recherche d'optimum, il manipule une population de taille constante. Cette population est formée de points candidats appelés chromosomes. La taille constante de la population entraîne un phénomène de compétition entre les chromosomes. Chaque chromosome représente le codage d'une solution potentielle au problème à résoudre, il est constitué d'un ensemble d'éléments appelés gènes, pouvant prendre plusieurs valeurs appartenant à un alphabet non forcément numérique.

A chaque itération, appelée génération, est créée une nouvelle population avec le même nombre de chromosomes. Cette génération consiste en des chromosomes mieux adaptés à leur environnement tel qu'il est représenté par la fonction objectif. Au fur et à mesure des générations, les chromosomes vont tendre vers l'optimum de la fonction objective. La création d'une nouvelle population à partir de la précédente se fait par application des opérateurs génétiques de reproduction. Ces opérateurs sont stochastiques [63].

Cette technique est d'application générale. En effet, quand nous utilisons les AGs, aucune connaissance de la manière dont nous devons résoudre le problème n'est requise, il est seulement nécessaire de fournir une fonction permettant de coder une solution sous forme de gènes, de faire le travail inverse et de fournir une fonction permettant d'évaluer la pertinence d'une solution au problème donné. Par conséquent, nous avons un modèle minimal et canonique pour n'importe quel système évolutionnaire et pour n'importe quel problème pouvant être abordé sous cet angle et ce paradigme. Cette représentation nous permet donc d'étudier des propriétés quasiment impossibles à étudier dans leur milieu naturel, et de résoudre des problèmes n'ayant pas de solutions calculables en temps raisonnables si nous les abordons sous d'autres paradigmes, avec des performances quantifiables, facilement mesurables et que nous pouvons confronter aux autres stratégies de résolution. Les principales caractéristiques des AGs par rapport aux autres paradigmes sont les suivantes [63]:

- nous utilisons un codage des informations : nous représentons toutes les caractéristiques d'une solution par un ensemble de gènes, c'est-à-dire un chromosome, sous un certain type de codage (binaire, réel, code de Gray, etc.), valeurs que nous concaténons pour obtenir une chaîne de caractères spécifique à une solution bien particulière ;
- nous traitons une population d'individus : cela introduit donc du parallélisme ;

- l'évaluation de l'optimalité du système n'est pas dépendante vis-à-vis du domaine ;
- nous utilisons des règles probabilistes : il n'y a pas d'énumération de l'espace de recherche, nous en explorons une certaine partie en étant guidé par un semi hasard.

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

- les chromosomes sont les éléments à partir desquels sont élaborés les solutions ou individus ;
- la population est l'ensemble des chromosomes ;
- la reproduction est l'étape de combinaison des chromosomes. La mutation et le croisement génétiques sont des méthodes de reproduction.

Tableau 3.1: Analogie de la terminologie génétique avec la nature

<b>Nature</b>	<b>Algorithmes Génétiques</b>
génom	individu
Chromosome	Chaîne, séquence
Gènes	Traits, caractéristiques
Allèle	Valeur de la caractéristique
Locus	Position dans la chaîne
Génotype	Structure
Phénotype	Ensemble de paramètres Une structure décodée

D'autres notions sont propres au domaine des AGs :

- l'indice de qualité ou fitness, aussi appelé indice de performance, est une mesure abstraite permettant de classer les chromosomes ;
- la fonction d'évaluation ou la fonction coût est la formule théorique qui permet de calculer l'indice de qualité d'un chromosome.

### 3.2. Fonctionnement des Algorithmes Génétiques

Les AGs fournissent des solutions aux problèmes n'ayant pas de solutions calculables en temps raisonnable de façon analytique ou algorithmique. Selon cette méthode, des milliers de solutions ou génotypes plus ou moins bonnes sont créées au hasard et sont soumises à un procédé d'évaluation de la pertinence de la solution mimant l'évolution des espèces les plus adaptées, c'est-à-dire, les solutions au problème qui sont les plus optimales survivent

davantage que celles qui le sont moins ; ainsi la population évolue par générations successives en croisant les meilleures solutions entre elles et en les faisant muter, puis en relançant ce procédé un certain nombre de fois afin d'essayer de tendre vers la solution optimale.

### 3.2.1. Etapes de fonctionnement

Le fonctionnement des AGs est caractérisé par plusieurs phases qui succèdent les unes aux autres : l'étape de reproduction consiste à sélectionner les individus les plus adaptés dans la population au sens de la fonction de mérite, et à les reproduire tels quels dans la génération suivante. L'évolution de la population est assurée au moyen des opérateurs de la sélection, de croisement et de mutation qui permettent de combiner et de modifier les chromosomes [58].

La sélection des meilleurs chromosomes est la première opération dans un AG. Au cours de cette opération l'algorithme sélectionne les éléments pertinents qui optimisent mieux la fonction. L'étape de croisement permet de créer deux individus appelés enfants en combinant les gènes des deux parents obtenus à l'étape précédente. Cet opérateur permet d'augmenter en moyenne la qualité d'une population. Enfin, l'étape de mutation permet d'introduire une variation aléatoire dans le génotype de l'individu. Cet opérateur permet donc d'explorer de nouvelles régions de l'espace de recherche, diminuant ainsi les risques de converger vers des minima locaux (Fig.3.1).

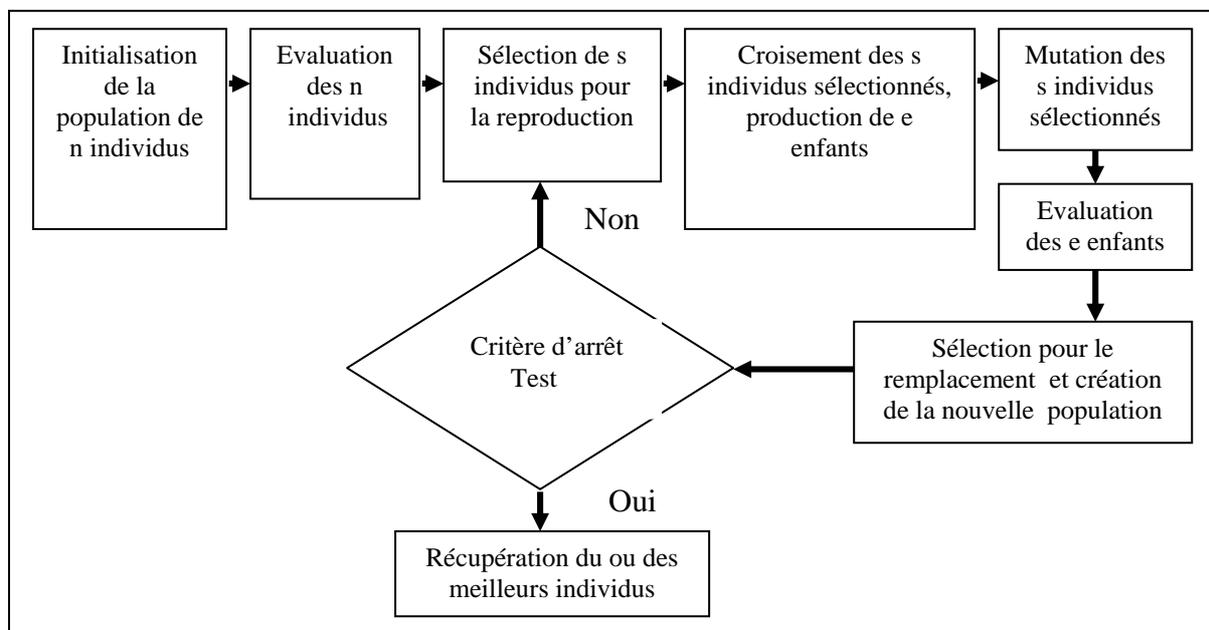


Figure 3.1 : Différentes étapes de fonctionnement des Algorithmes Génétiques [63]

Le mécanisme d'évolution et de sélection est indépendant du problème à résoudre. Seules trois fonctions changent :

- la fonction qui s'occupe de représenter le problème en codant chaque information caractérisant une solution possible selon un codage bien particulier. Chaque information représente alors un gène et toutes les valeurs que peuvent prendre cette caractéristique représentent les allèles possibles pour ce gène et en concaténant tous ces gènes pour obtenir un chromosome qui lui représente une solution dans son intégralité ;
- la fonction inverse qui à partir d'un chromosome permet d'obtenir une solution par décodage du génome ;
- la fonction qui évalue l'adaptation d'une solution à un problème, sa pertinence.

### 3.2.2. Réglage des paramètres des Algorithmes Génétiques

Une des difficultés dans l'utilisation de ces métaheuristiques est de trouver une configuration au problème à traiter. Cette dernière inclue la taille de la population et les critères d'arrêt, les taux respectifs de la mutation, du croisement et de la sélection, ainsi que le type de ces opérateurs de reproduction [64].

Nous touchons là au problème délicat du réglage des paramètres de l'algorithme. Celui-ci doit être optimisé pour chaque type de problème traité, ce qui constitue une part importante du travail de l'utilisateur. Les caractéristiques de l'algorithme doivent être adaptées à la topologie du paysage dessiné par la fonction fitness. L'ensemble problème - méthodes - paramètres constitue donc un tout. Certaines études en témoignent où les paramètres d'un AG sont réglés et optimisés par un autre AG. L'AG est efficace si :

- l'évaluation est juste : un individu évalué fort l'est effectivement ;
- la méthode de génération d'individus (croisement, mutation, etc.) doit permettre une bonne exploration de toutes les possibilités et être potentiellement améliorant ;
- tous les individus possibles et imaginables doivent pouvoir être créés ;
- la probabilité de produire un individu meilleur que ses parents ne doit pas être trop faible. Si elle est nulle, aucune amélioration n'est possible et si elle n'est pas assez forte, l'apparition d'un individu intéressant ne pourra pas avoir lieu dans un délai raisonnable.

Une population trop petite évoluera probablement vers un optimum local peu intéressant. Une population trop grande sera inutile car le temps de convergence sera excessif. La taille de

la population doit être choisie de façon à réaliser un bon compromis entre le temps de calcul et la qualité du résultat [65].

Aussi, il faut être conscient que la taille de la population dépend de la puissance de calcul disponible, des méthodes utilisées (sélection et opérateurs génétiques), du nombre de variables considérées et de la fonction d'adaptation. Si la fonction à optimiser comporte peu d'optima locaux et un optimum global net, la population nécessaire sera plus petite que dans le cas d'une fonction beaucoup plus compliquée comportant de nombreux optima locaux.

La probabilité de croisement dépend en général de la forme de la fonction de performance. Son choix est bien souvent heuristique. Plus elle est élevée, plus la population subit de changements importants. Les valeurs généralement admises sont comprises entre 0.5 et 0.9. 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 [62].

Dans la pratique, les méthodes et paramètres des AGs sont tous réglés approximativement par tâtonnement avec des fonctions de ' $n$ ' variables couramment utilisées pour tester les algorithmes d'optimisation. Le temps de calcul de ces fonctions étant minime, nous pouvons ainsi régler rapidement les paramètres [63].

### **3.2.3. Avantages et limites des Algorithmes Génétiques**

Les AGs présentent plusieurs avantages :

- aucune hypothèse à faire sur l'espace de recherche ;
- nombreuses méthodes disponibles ;
- solutions intermédiaires interprétables ;
- adaptation rapide à de nouveaux environnements ;
- parallélisme et distribution aisés ;
- les représentations facilitent la compréhension.

Quant aux limites des AGs, nous les résumons comme suit [62] :

- Ils sont le plus souvent difficiles à mettre en oeuvre: des paramètres comme la taille de la population ou le taux de mutation sont parfois difficile à déterminer. Le succès de l'évolution en dépend et plusieurs essais sont donc nécessaires, ce qui limite encore l'efficacité de l'algorithme. En outre, choisir une bonne fonction d'évaluation est aussi critique. Celle-ci doit prendre en compte les bons paramètres du problème. Elle doit donc être choisie avec soin ;

- aucune garantie de solution optimale en un temps fini ;
- initialisation de plusieurs paramètres, importance du choix des méthodes ;
- coût d'exécution important. Ils nécessitent de nombreux calculs, en particulier au niveau de la fonction d'évaluation ;
- possibilité de convergence vers des optimums locaux.

### 3.3. Codage et espace de recherche

Le codage associe à chacun des points de l'espace d'état une structure de données. Elle se place généralement après une phase de modélisation mathématique du problème traité. La qualité du codage des données conditionne le succès des AGs. Le codage binaire a été très utilisé à l'origine. Les codages réels sont désormais largement utilisés, notamment dans les domaines applicatifs pour l'optimisation de problèmes à variables réelles ;

Le programmeur doit dans un premier temps choisir le génotype de chacun des individus de la population. Chaque paramètre d'une solution est assimilé à un gène, toutes les valeurs qu'il peut prendre sont les allèles de ce gène, nous devons trouver une manière de coder chaque allèle de façon unique en établissant une bijection entre l'allèle réel et sa représentation codée.

#### 3.3.1. Codage

Explicitement, l'AG manipule des populations d'individus qui est le point auquel nous appliquons la fonction fitness. Chaque individu est représenté par une chaîne de caractères. Cependant, nous pouvons considérer qu'implicitement, il manipule des schémas. Ce concept introduit une nouvelle notion, la brique élémentaire ou schème. Un schème est un fragment d'individu de longueur donnée comportant un certain nombre de positions instanciées. A chaque variable d'optimisation notée  $x_i$  ou paramètre du dispositif, nous faisons correspondre un gène [66].

Un chromosome est une suite de gènes. Nous pouvons choisir de regrouper les paramètres similaires dans un même chromosome appelé chromosome à un seul brin et chaque gène sera repérable par sa position appelée locus sur le chromosome. Chaque individu est représenté par un ensemble de chromosomes et chaque population est un ensemble d'individus. Ainsi, nous aboutissons à une structure présentant cinq niveaux d'organisation, d'où résulte le comportement complexe des AGs (Fig3.2).

Par exemple, si une population a une taille égale à 100 et si le nombre de variables de la fonction fitness est égal à 42, alors elle sera représentée par une matrice de 100 lignes et 42 colonnes. Il y a trois principaux types de codage utilisables, et nous pouvons passer de l'un à l'autre.

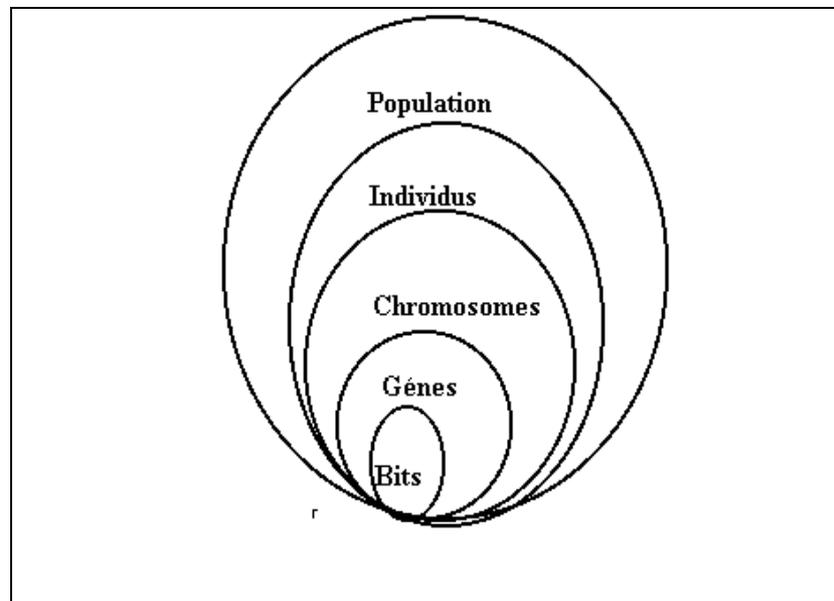


Figure 3.2 : Les cinq niveaux d'organisation d'un Algorithme Génétique [63]

### 3.3.1.1. 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 sont représentés par des tableaux de gènes et les individus de notre espace de recherche sont représentés par des tableaux de chromosomes.

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 [63]:

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

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 :

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

Soit :

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

Le codage binaire présente les inconvénients suivants [67] :

- il peut être difficile d'adapter ce codage à certains problèmes. La représentation binaire traditionnelle utilisée pour les AGs 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 d'une 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 AGs 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.

### 3.3.1.2. Codage de Gray

Dans le cas d'un codage binaire, nous utilisons souvent la distance de Hamming comme mesure de la dissimilarité entre deux éléments de la population. Cette mesure compte les différences de bits de même rang de ces deux séquences. La distance de Hamming entre deux chaînes de bits est le nombre de bits qui diffèrent de l'une à l'autre. C'est à ce niveau que le codage binaire commence à montrer ses limites. En effet, deux éléments voisins en terme 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 diffère.

### 3.3.1.3. 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 [68, 69].

### 3.3.2. Espace de recherche

Il s'agit de la composante principale de l'algorithme et qui est préalable aux autres. Dans de nombreux cas, l'espace de recherche est totalement déterminé par le problème à résoudre. Cet espace, où sont appliqués les opérateurs génétiques, est appelé espace génotypique et l'espace de recherche initial, dans lequel est calculée la performance des individus, est appelé espace phénotypique. Parmi les espaces de recherche les plus utilisés, nous citons :

- l'espace de recherche noté  $\Omega$  tel que :

$$\Omega = \{ 0, 1 \}^N \quad (3.06)$$

La justification de l'utilisation intensive de cet espace de recherche particulier est fondée à la fois sur des considérations théoriques et sur un parallèle encore plus précis avec la biologie puisqu'une chaîne de bits est assimilée à un chromosome ;

- les vecteurs réels : c'est le cas le plus fréquent en calcul numérique.  $\Omega$  est un sous-ensemble de  $\mathbb{R}^n$ , borné ou non. Nous parlons alors d'optimisation paramétrique.

## 3.4. Sélection

La sélection des individus a pour but d'éliminer les plus mauvais et recombinaison les survivants par croisement puis mutation pour en créer de nouveaux qui seront à leur tour évalués. La sélection des meilleurs individus suppose que nous disposons d'un indice de la valeur de chacun. Tout repose donc sur la qualité de la fonction d'évaluation et les possibilités de recombinaisons.

Cet opérateur est chargé de définir quels seront les individus de la population des parents notée  $P$  qui vont être dupliqués dans la nouvelle population notée  $P'$ . Soit  $n$  le nombre d'individus de  $P$ , nous devons en sélectionner  $n/2$  (l'opérateur de croisement nous permet de repasser à  $n$  individus). Cet opérateur est peut-être le plus important puisqu'il permet aux individus d'une population de survivre, de se reproduire avant de mourir. En règle générale, la probabilité de survie d'un individu sera directement reliée à son efficacité relative au sein de la population [70].

Pour sélectionner les meilleurs chromosomes dans une population, certaines techniques sont proposées par différents auteurs. Essentiellement, ces techniques sont en nombre de quatre :

- la méthode de la loterie biaisée (roulette en Anglais) ou probabilité de sélection proportionnelle à l'adaptation ;
- la méthode élitiste ou par rang (elitist en Anglais) ;
- la sélection par tournoi (tournament en Anglais) ;
- la sélection uniforme (uniform en Anglais).

### 3.4.1. Loterie biaisée ou Roulette

Cette méthode est la plus connue et la plus utilisée. Avec cette méthode chaque individu a une chance d'être sélectionné proportionnellement à sa performance; donc plus les individus sont adaptés au problème (probabilité d'adaptation élevée), plus ils ont de chance d'être sélectionnés.

Pour utiliser l'image de la "roue du Forain", chaque individu se voit attribué un secteur dont l'angle est proportionnel à son degré d'adaptation, sa fitness. Nous faisons tourner la roue et quand elle cesse de tourner nous sélectionnons l'individu correspondant au secteur désigné par une sorte de curseur en pointant sur un secteur particulier de celle-ci après qu'elle soit arrêtée de tourner. Cette méthode, bien que largement répandue, a plusieurs inconvénients :

- elle a une forte variance. Il n'est pas impossible que sur  $n$  sélections successives destinées à désigner les parents de la nouvelle génération  $P'$ , la quasi-totalité, voire la totalité des  $n$  individus sélectionnés soient des individus ayant une fitness mauvaise et donc aucun individu à forte fitness ne fasse partie des parents de la nouvelle génération. Ce phénomène est très dommageable car cela va complètement à

l'encontre du principe des AGs qui veut que les meilleurs individus soient sélectionnés de manière à converger vers une solution la plus optimale possible ;

- à l'inverse, nous pouvons arriver à une domination écrasante d'un individu localement supérieur. Ceci entraînant une grave perte de diversité, d'où la convergence prématurée. Ainsi, nous constatons la stagnation de l'évolution sans jamais atteindre l'optimum global.

### 3.4.2. Méthode élitiste ou par rang

Cette méthode consiste à sélectionner les  $n$  individus dont nous avons besoin pour la nouvelle génération  $P'$  en prenant les  $n$  meilleurs individus de la population  $P$  après les avoir triés de manière décroissante selon leur fitness. Donc, elle choisit toujours les individus possédant les meilleurs scores d'adaptation.

Cette méthode amènera à une convergence prématurée encore plus rapidement et surtout de manière encore plus sûre que la méthode de sélection de la loterie biaisée. En effet, la pression de la sélection est trop forte, la variance nulle et la diversité inexistante.

### 3.4.3. Sélection par tournoi

Cette technique utilise la sélection proportionnelle sur des paires d'individus, puis choisit pour ces paires, l'individu qui a le meilleur score d'adaptation. Le principe de cette méthode est le suivant : nous effectuons un tirage avec remise de deux individus de  $P$ , et nous les faisons "combattre". Celui qui a la fitness la plus élevée l'emporte avec une probabilité notée  $p_s$  comprise entre 0.5 et 1. Nous répétons ce processus  $n$  fois de manière à obtenir les  $n$  individus de  $P'$  qui serviront de parents. La variance de cette méthode est élevée et le fait d'augmenter ou de diminuer la valeur de  $p_s$  permet respectivement de diminuer ou d'augmenter la pression de la sélection.

### 3.4.4. Sélection uniforme

La sélection se fait aléatoirement, uniformément et sans intervention de la valeur d'adaptation. Chaque individu a donc une probabilité  $1/I$  d'être sélectionné, où  $I$  est le nombre total d'individus dans la population ou taille de la population.

### 3.5. Techniques d'amélioration de la sélection

Les processus de sélection présentés sont très sensibles aux écarts des valeurs de la fonction fitness et dans certains cas, un très bon individu risque d'être reproduit trop souvent et peut même provoquer l'élimination complète de ses congénères. Pour éviter ce comportement, il existe d'autres techniques de sélection dont la technique de changement d'échelle ou scaling et celle du partage ou sharing, qui empêchent les individus forts d'éliminer complètement les plus faibles [63].

#### 3.5.1. Technique du changement d'échelle

Le changement d'échelle, modifie les fitness afin de réduire ou d'amplifier artificiellement les écarts entre les individus. Le processus de sélection n'opère plus sur l'aptitude réelle notée  $f_r$ , mais sur son image après changement d'échelle.

Cette technique permet une bonne exploration de l'espace de recherche mais ne favorise pas la répartition des individus sur les différents modes de la fonction objective. Il existe plusieurs types du changement d'échelle dont le partage linéaire et le changement d'échelle exponentiel

##### 3.5.1.1. Technique du changement d'échelle linéaire

Cette technique est définie de la façon suivante :

$$f_s = a.f_r + b \quad (3.07)$$

Avec  $a = \frac{\max' - \min'}{\max - \min}$  ; et  $b = \frac{\min' . \max - \min . \max'}{\max - \min}$

En règle générale,  $a < 1$ , ce qui permet de réduire les écarts de fitness et de favoriser l'exploration de l'espace. Ce type de changement d'échelle pénalise la fin de convergence si nous désirons favoriser les modes dominants.

- 
- 3.5.1.2. Technique du changement d'échelle exponentiel
- Il est défini de la façon suivante :

$$f_s = (f_r)^{k(n)} \quad (3.08)$$

n : génération courante.

- $k \approx 0$  : nous réduisons fortement les écarts de fitness. Aucun individu n'est vraiment favorisé et l'AG se comporte comme un algorithme de recherche aléatoire et permet d'explorer l'espace de recherche.
- $k \approx 1$  : le changement d'échelle est inopérant.
- $k > 0$  : les écarts sont exagérés et seuls les bons individus sont sélectionnés, ce qui produit l'émergence des modes.

Dans la pratique, nous faisons généralement varier  $k$  des faibles valeurs vers les fortes valeurs au cours des générations. Pour cela nous pouvons utiliser la formule suivante :

$$k = \left[ \tan \left[ \left( \frac{n}{N+1} \right) \frac{\pi}{2} \right] \right] P \quad (3.09)$$

$n$  : la génération courante,

$N$  : le nombre total de générations prévues,

$P$  : un paramètre à choisir. Le choix de  $p = 0.1$  s'est avéré pertinent dans les applications.

Ce dernier principe donne de meilleurs résultats que le changement d'échelle linéaire et sera donc systématiquement utilisé. Dans le cas des fonctions objectives multi-modes présentant des optimaux quasi-équivalents, cette technique du changement d'échelle, en amplifiant les écarts de fitness en fin de convergence, va effectivement favoriser le mode dominant mais aussi masquer les modes sous optimaux qui peuvent tout de même présenter un intérêt.

### 3.5.2. Technique de partage

De la même façon que le changement d'échelle, le partage consiste à modifier la fitness utilisée par le processus de sélection. Pour éviter le rassemblement des individus autour d'un sommet dominant, le partage pénalise les fitness en fonction du taux d'agrégation de la population dans le voisinage d'un individu. Il requiert l'introduction d'une notion de distance. Dans la pratique, il faut définir une distance indiquant la dissimilarité entre deux individus. Cette distance est alors utilisée pour calculer la nouvelle fitness de la façon suivante :

$$f_i' = \frac{f_i}{m_i}; \quad m_i = \sum_{j=1}^{j=N} S(d(x_i, x_j)) \quad (3.10)$$

$$S(d) = 1 - S\left(\frac{d}{\sigma_{share}}\right) \cdot \alpha \quad \text{si} \quad d < \sigma_{share}$$

$$S(d) = 0 \quad \text{si} \quad d > \sigma_{share}$$

Le paramètre  $\sigma_{share}$  délimite le voisinage d'un point et dépend du problème traité. Dans la pratique, cette technique donne effectivement de bons résultats mais au prix de  $N^2$  calculs de distances entre chromosomes à chaque génération pour une population de taille  $N$ .

### 3.6. Reproduction

Durant la phase de reproduction de l'AG, des individus de la population sont sélectionnés d'après la méthode choisie et sont recombinaison 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 [71].

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  $\{0, \dots, 2L_i\}$  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 palier à ce genre de problème, la représentation réelle sera donc utilisée au lieu de la représentation binaire [67].

#### 3.6.1. Croisement

Le crossover ou recombinaison utilisé par les AGs est la transposition informatique du mécanisme qui permet, dans la nature, la production de chromosomes qui héritent partiellement des caractéristiques des parents. Son rôle fondamental est de permettre la recombinaison des informations présentes dans le patrimoine génétique de la population. Aussi, il favorise l'exploration de l'espace de recherche.

Cet opérateur est utilisé après l'application de l'opérateur de sélection sur la population  $P$ ; nous nous retrouvons donc avec une population  $P'$  de  $n/2$  individus et nous devons doubler ce nombre pour que notre nouvelle génération soit complète. Nous allons donc créer de manière aléatoire  $n/4$  couples et nous les faisons se reproduire. Les chromosomes ou ensembles de paramètres des parents sont alors copiés et recombines de façon à former deux descendants possédant des caractéristiques issues de ces deux parents.

Le choix de la technique de croisement dépend de la façon dont le programmeur a choisi de représenter les solutions. La notion de chromosome est indissociable de ces techniques de reproduction, dont les plus connues sont le croisement :

- en un point ou simple : nous sélectionnons les génotypes de deux individus sur lesquels nous choisissons au hasard le même point de coupure ou de croisement appelé loci. Nous échangeons alors les fragments ou segments situés après ce point pour donner naissance à deux nouveaux génotypes selon une probabilité de croisement  $p_c$  (Fig. 3.3) ;
- en deux points ou double: de la même façon que précédemment mais avec deux points de coupure et en échangeant les fragments situés entre ces deux points ;
- en  $k$  points ou multipoints : généralisation des méthodes précédentes à  $k$  points de coupure ;
- uniforme : assimilable à un croisement multipoints dont le nombre de points est inconnu ;
- flottants : uniquement applicable à des individus à représentation flottante. Il en existe un certain nombre dont : le croisement arithmétique, volumique, etc.

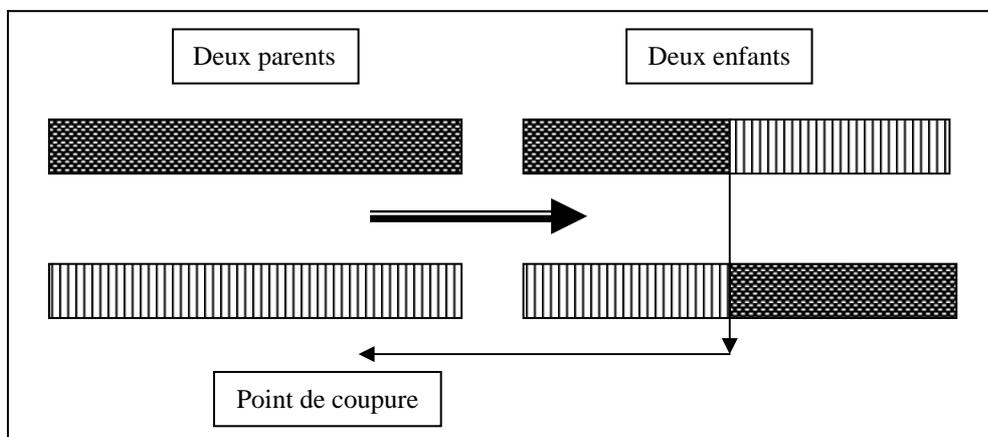


Figure 3.3 : Croisement en un point

Les avis des auteurs divergent en ce qui concerne le choix d'une technique de croisement lorsque la représentation du chromosome est binaire. Certains préfèrent le croisement multipoints ou multiple et d'autres sont portés sur le croisement uniforme. Les conclusions sont assez diversifiées et dues à la nature des différents problèmes étudiés [71].

A partir de l'ensemble des techniques de croisement citées précédemment, nous constatons que peu importe la technique utilisée, il faut respecter le concept de base du croisement. Ce concept consiste à permuter les valeurs des gènes de deux parents pour former des progénitures et d'explorer le domaine de recherche dans l'espace des solutions. Pour faire un bon choix d'une technique de croisement, il suffit d'effectuer plusieurs expériences avec différentes techniques. Pour raffiner cette recherche, les techniques de sélection standard (par exemple la roulette biaisée) et l'opérateur de mutation vont nous permettre de converger plus rapidement vers des solutions.

Cependant, nous pouvons noter que le nombre de points de croisements ainsi que la probabilité de croisement  $p_c$ , permettent d'introduire plus ou moins de diversité. En effet, plus le nombre de points de croisements est grand et plus la probabilité de croisement sera élevée plus il y aura d'échange de segments, donc d'échange de paramètres, d'informations, etc. Plus le nombre de points de croisements est petit et plus la probabilité de croisement sera faible, moins le croisement apportera de la diversité. Malgré tout, il est possible que l'action conjointe de la sélection et du croisement ne permette pas de converger vers la solution optimale du problème, raison pour laquelle l'opérateur de mutation est utilisé [72].

Certaines remarques sur le croisement sont illustrées dans les points suivants :

- le croisement est la clef de la puissance des AGs. Il est directement lié à l'aptitude d'une population d'individus à explorer son espace de recherche et à combiner entre eux les meilleurs résultats. Grâce au croisement, les AGs se concentrent sur les parties les plus prometteuses de l'espace des solutions du fait que cet opérateur de croisement combine des chaînes contenant des solutions partielles ;
- le croisement n'est habituellement pas appliqué à toutes les paires d'individus choisis aléatoirement lors de la reproduction. La probabilité du croisement appliquée est comprise entre 0.6 et 1.0. Dans le cas où le croisement ne s'appliquerait pas, alors les enfants sont semblables aux parents ;

- dans le cas particulier d'un chromosome matriciel constitué par la concaténation de vecteurs, nous pouvons étendre ce principe de croisement aux vecteurs constituant les gènes.

### 3.6.2. Mutation

L'opérateur de mutation modifie donc de manière complètement aléatoire les caractéristiques d'une solution, ce qui permet de maintenir la diversité au sein de la population de solutions. Il consiste à changer la valeur allélique d'un gène avec une probabilité notée  $p_m$  très faible, généralement comprise entre 0.01 et 0.001. Nous pouvons aussi prendre :

$$p_m = \frac{1}{l_g} \quad (3.11)$$

où  $l_g$  est la longueur de la chaîne de bits codant notre chromosome.

La mutation consiste simplement en l'inversion d'un bit se trouvant en un locus bien particulier et lui aussi déterminé de manière aléatoire (Fig. 3.4).

Cet opérateur introduit du bruit au sein de la population. Il dispose de quatre grands avantages [73] :

- il garantit la diversité de la population, ce qui est primordial pour les AGs ;
- il permet d'éviter le phénomène connu sous le nom de dérive génétique. Nous parlons de dérive génétique quand certains gènes favorisés par le hasard se répandent au détriment des autres et sont ainsi présents au même endroit sur tous les chromosomes. Le fait que l'opérateur de mutation puisse entraîner de manière aléatoire des changements au niveau de n'importe quel locus, il permet d'éviter l'installation de cette situation défavorable ;
- il permet de limiter les risques d'une convergence prématurée causée par exemple par une méthode de sélection élitiste imposant à la population une pression sélective trop forte. En effet, dans le cas d'une convergence prématurée nous nous retrouvons avec une population dont tous les individus sont identiques mais ne sont que des optimums locaux. Tous les individus étant identiques, le croisement ne changera rien à la situation. En effet, l'échange d'informations par croisement entre des individus identiques est sans conséquences et la population n'évoluera pas. L'évolution se retrouvant bloquée, l'optimum global ne sera jamais atteint. La mutation apporte la

solution en entraînant des inversions de bits de manière aléatoire et en réintroduisant des différences entre les individus ;

- il permet d'atteindre la propriété d'ergodicité garantit que chaque point de l'espace de recherche puisse être atteint. Nous avons la certitude mathématique que n'importe quel permutation de notre chaîne de bits peut apparaître au sein de la population et donc que tout point de l'espace de recherche peut être atteint. Grâce à cette propriété, nous sommes sûrs de pouvoir atteindre l'optimum global.

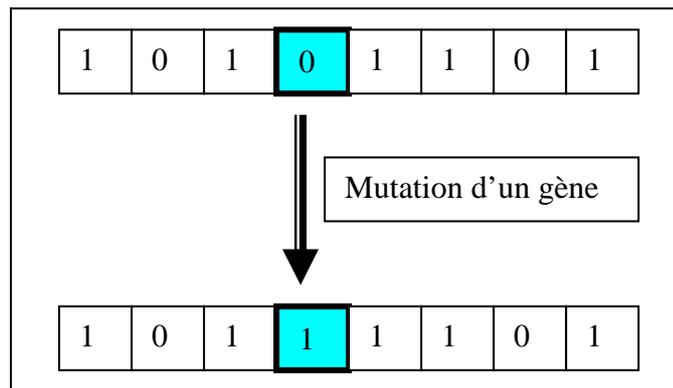


Figure 3.4 : Mutation d'un gène

Les propriétés de convergence des AGs sont donc fortement dépendantes de cet opérateur sur le plan théorique. Cependant, plusieurs techniques de mutation ont été développées dans la littérature. Certaines d'entre elles s'appliquent sur des gènes dont la représentation est binaire et d'autres sur des gènes de type réel. Le type de représentation choisi par le programmeur a une influence sur le choix la technique de mutation dont nous citons les plus utilisées [73; 74] :

- bitflip : nous choisissons au hasard un bit dans la chaîne d'une représentation binaire, et nous l'invertissons ;
- permutation : permutation de deux éléments quand les représentations sont de type chaîne de caractère ; ce qui inclut les représentations binaires ;
- addition : uniquement pour les représentations flottantes, nous ajoutons ou retirons une certaine quantité au flottant. Nous pouvons aussi retirer au hasard un individu de la population ;
- mutation Gaussienne : utilisé pour la représentation flottante à laquelle nous ajoutons une variable aléatoire gaussienne.

Pour déterminer le nombre de positions dont les gènes doivent subir un changement, il suffit de connaître la taille du chromosome notée  $L$  et la probabilité de mutation  $p_m$ . Ce nombre est défini par le produit  $L \times P_m$ . Posons la notation suivante du chromosome qui doit subir la mutation :

$$C = (c_1, c_2, \dots, c_i, \dots, c_L) \quad (3.12)$$

Tels que :  $c_i \in [a_i, b_i]$

Avec :  $a_i$  et  $b_i$  les bornes inférieures et supérieures pour  $c_i$ .

où,  $c_i \in \{1, 2\}$

Le résultat obtenu par l'opérateur de mutation est un chromosome  $C'$  tels que :

$$C' = (c_1', c_2', \dots, c_i', \dots, c_L') \quad (3.13)$$

### 3.6.2.1. Mutation aléatoire

Dans le cas binaire, si la valeur du gène à muter est égale à 1, alors elle est inversée à 0 et si la valeur du gène est égale à 0, alors elle est inversée à 1.  $c_i$  est modifié par le gène  $c_i'$ . Cette valeur du gène  $c_i'$  est choisie aléatoirement de l'intervalle  $[a_i, b_i]$ .

### 3.6.2.2. Mutation uniforme

La mutation uniforme à codage réel est identique à celle du codage binaire. Ainsi, chaque variable  $c_i$  est changée selon une certaine probabilité  $p_m$  en un nombre aléatoire tiré dans une distribution uniforme sur l'intervalle  $[a_i, b_i]$ .

### 3.6.2.3. Mutation non uniforme

La technique de la mutation non uniforme est appliquée en fonction de la génération  $t$  courante et le nombre maximum de générations notée  $gen\_max$ . Le gène  $c_i'$  est défini comme suit :

$$\begin{aligned} c_i' &= c_i + (t, b_i \cdot c_i) & \text{si} & \quad c_i = 0 \\ c_i' &= c_i - (t, c_i \cdot a_i) & \text{si} & \quad c_i = 1 \end{aligned} \quad (3.14)$$

$c_i'$  : généré aléatoirement tels que:

$a \in \{0, 1\}$

$$(t - y) = y(1 - r^{(\frac{1-i}{gen - max})}) \quad (3.15)$$

r : nombre aléatoire de l'intervalle [0,1]

b : paramètre choisi par l'utilisateur permettant de déterminer le degré de dépendance sur le nombre d'itérations.

### 3.7. 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 P. 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 [70] :

#### 3.7.1. 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.

#### 3.7.2. 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. Donc, 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 AGs dans certains cas. Mais présente le 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 [75].

### **3.8. Conclusion**

L'intérêt des Algorithmes Génétiques est de produire des solutions diversifiées notamment en utilisant des heuristiques, en temps contrôlable. Afin de les appliquer à un problème particulier, il est nécessaire de définir convenablement la fonction d'évaluation et le codage d'une solution. Ces algorithmes nécessitent également la définition d'opérateurs, aussi bien de mutation que de croisement, en adéquation avec le problème. Dans de nombreux cas, ils peuvent être difficiles à paramétrer. Ils peuvent être efficaces dans la résolution d'un problème en apportant assez rapidement une solution acceptable.

## Chapitre 4 :           Modèle Génétique de classification acoustique

Dans ce chapitre, nous allons expliquer les méthodes d'optimisation classiques et expliquer comment nous avons appliqué les AGs à la classification acoustique supervisée des sons de l'AS en insistant sur le choix de la fonction d'évaluation et les différents paramètres de l'AG, la nature des individus et le choix des opérateurs de reproduction.

Le problème que nous avons à traiter consiste à concevoir un modèle génétique capable d'optimiser les classificateurs que nous avons choisis à base de mesure de distance en utilisant les AGs. Ainsi, notre modèle génétique de classification a pour fonction objectif ou fonction d'évaluation, la fonction de mesure de distance choisie. Cette dernière, présente des optimaux multiples et aucune des méthodes classiques d'optimisation ne s'adaptent vraiment à notre application. Ceci nous a obligés à nous orienter vers une autre classe d'algorithmes stochastiques qui sont les AGs. Ces derniers travaillent sur des populations de points de l'espace d'état et sont capables d'isoler plusieurs optimaux du critère souhaité. Les AGs optimisent une population tandis que beaucoup d'autres méthodes optimisent un point unique ou individu. Ainsi, nous avons choisi de les utiliser car ils sont bien adaptés à notre problème : espace d'état discret, optimaux multiples ou population, de grande dimension et faible temps de calcul.

### 4.1. Méthodes d'optimisation

Les méthodes d'optimisation permettent de déterminer plusieurs solutions dans l'espace d'état qui maximisent ou minimisent un critère. Suivant la nature de la fonction à optimiser, la structure de l'espace d'état et les modes d'exploration de ce dernier, certaines méthodes sont plus adaptées que d'autres pour résoudre un problème donné. Nous définissons l'espace d'état par la modélisation mathématique des paramètres d'entrée du système. L'espace objectif représente les sorties du système. Suivant la nature des variables, ces espaces qui peuvent être continus ou discrets, déterminent des méthodes d'optimisation différentes.

Les différentes méthodes d'optimisation peuvent se décomposer selon le mode d'exploration de l'espace d'état : les méthodes déterministes et les méthodes stochastiques. Les méthodes d'optimisation les plus courantes sont caractérisées par une exploration

déterministe de l'espace d'état. Elles permettent de déterminer la direction de l'optimum par rapport à chacun des points de cet espace. Le principal inconvénient de cette méthode est qu'elle reste limitée à des espaces de faibles dimensions et qu'elle ne permet de trouver qu'un seul optimum. Par conséquent, elle ne permet pas de résoudre notre problème [76].

Lorsqu'une exploration déterministe n'est pas possible, nous parcourons aléatoirement l'espace de recherche. Ces méthodes stochastiques sont caractérisées par un processus de génération aléatoire de points de l'espace d'état qui s'oriente progressivement vers la solution optimale en évaluant les valeurs successives du critère associé. Ce type de méthodes assure qu'après un nombre suffisamment grand d'exécutions de l'algorithme, l'espace mathématique des solutions fournies converge vers la solution optimale.

Parmi ces méthodes nous citons le recuit simulé. Cette méthode ne travaille que sur un seul point de l'espace d'état et ne produira donc qu'une solution quasi-optimale. Si le problème présente plusieurs solutions équivalentes au sens du critère, une seule sera sélectionnée et ce choix dépendra essentiellement du point choisi pour initialiser le processus.

#### **4.1.1. Recherche stochastique globale**

La recherche globale fait référence à la recherche des optima globaux de la fonction objective d'un problème donné. De ce point de vue, la méthode d'optimisation globale vise la détermination des optima globaux du problème, en évitant le piègeage dans l'un de ses optima locaux.

Nous pouvons partager les méthodes en deux catégories. Celles qui permettent de déterminer un minimum local, ces méthodes sont appelées méthodes locales, et celles qui s'efforcent de déterminer un optimum global, ces méthodes sont appelées méthodes de recherche globale. Ces dernières ne s'excluent pas mutuellement. Afin d'améliorer les performances d'une recherche, plusieurs auteurs combinent les deux types d'algorithmes. Une recherche globale permet de bien explorer l'espace de recherche. Cette phase est appelée phase de diversification; et une recherche locale qui permet de bien exploiter une zone prometteuse susceptible de contenir un minimum global, localisée lors de l'exploration du domaine de recherche. Cette phase est appelée phase d'intensification.

Selon les critères de robustesse, ces méthodes d'optimisation sont subdivisées en trois types : déterministes, énumératives et les méthodes stochastiques *qui* utilisent un processus aléatoire guidé.

La plupart des algorithmes stochastiques sont itératifs et comportent trois éléments principaux : un mécanisme de perturbation, un critère d'acceptation et un critère d'arrêt. Ils sont appliqués à partir d'un ou plusieurs points de la fonction objectif, choisis aléatoirement.

#### ***4.1.1.1. Perturbation aléatoire***

La perturbation aléatoire est définie par :

- toutes les coordonnées du vecteur de solutions, ou une partie seulement, qui sont perturbées ;
- la transformation qui suit une loi de distribution particulière dans un voisinage du point courant, par exemple une loi de distribution uniforme ;
- le ou les nouveaux points qui sont générés selon une loi qui dépend des points précédents et/ou des valeurs précédentes de la fonction objective.

#### ***4.2.1.2. Critères d'acceptation***

Les Critères d'acceptation sont définis par :

- le ou les nouveaux points qui sont acceptés selon une certaine loi de probabilité ;
- le ou les nouveaux points qui sont acceptés selon certaines conditions ;
- le ou les meilleurs points qui sont conservés pour l'étape suivante ;
- une recherche locale qui est effectuée à partir du ou des nouveaux points.

#### ***4.1.1.3. Critères d'arrêt***

Ils peuvent être liés à la qualité du minimum obtenu ou à des coûts de calcul. La procédure est arrêtée :

- quand il n'y a plus d'amélioration de la solution après un certain nombre d'itérations ;
- quand les perturbations ne dépassent plus un certain seuil ;
- pour un nombre maximal d'évaluations de la fonction objective ;
- pour un nombre d'itérations fixé ;
- après un certain temps de calcul fixé.

#### **4.1.2. Méta heuristiques modernes**

Parmi les différentes méthodes stochastiques d'optimisation globale, nous allons uniquement nous intéresser aux heuristiques modernes. Le mot heuristique vient du grec heurein (découvrir) et qualifie tout ce qui sert à la découverte, à l'invention et à la recherche.

Pour l'algorithmique, les heuristiques sont des méthodes qui cherchent à approcher une solution optimale. On les appelle parfois méthodes approchées. Une heuristique peut être conçue pour résoudre un type de problème donné, ou bien être conçue comme une méthode générale, qui peut être adaptée à divers problèmes d'optimisation. Dans le second cas, elle est désignée sous le terme de métaheuristique.

Les métaheuristiques sont à l'origine dédiées aux problèmes combinatoires, où les paramètres ne peuvent prendre que des valeurs discrètes. Par exemple, dans le cas du problème du voyageur de commerce, la permutation de deux villes voisines est un moyen d'engendrer les tournées voisines d'une tournée donnée et la distance entre deux configurations peut être considérée comme le nombre minimal de transformations élémentaires nécessaires pour passer de l'une à l'autre. Cette distance, ainsi définie, induit une topologie dans l'espace des configurations.

Pour l'optimisation d'une fonction continue, ces méthodes d'optimisation peuvent être adaptées moyennant des transformations plus ou moins aisées, en inventant une nouvelle topologie. Chaque paramètre doit être discrétisé de façon individuelle. La difficulté majeure réside dans la détermination de la taille optimale du pas de discrétisation et de sa direction qui résulte des variables sur lesquelles nous agissons. Le choix de la loi de discrétisation est un compromis entre deux situations extrêmes ; si le pas est trop :

- petit, nous n'explorons qu'une région limitée de l'espace des configurations, et l'algorithme risque d'être piégé dans un minimum local ;
- grand, la recherche devient quasiment aléatoire.

La meilleure solution peut consister à élaborer une topologie adaptative.

Les heuristiques comportent souvent plusieurs paramètres contrôlant les différents opérateurs et l'influence du ou des processus stochastiques. L'efficacité d'une heuristique dépend du choix de ses paramètres de contrôle. Ce réglage est complexe, surtout quand le nombre de paramètres est élevé et quand la plage de variation de chacun de ces paramètres est étendue. Les différents paramètres sont généralement dépendants à des degrés divers, ce qui rend encore plus difficile leur réglage. Enfin pour un jeu de paramètres de contrôle donnés, l'aspect stochastique fait que les résultats varient d'une exécution à l'autre.

## 4.2. Modélisation Génétique de classification vocale

L'application des AGs à la classification acoustique des séquences vocales consiste à regrouper de manière optimale des segments acoustiques qui sont, des vecteurs ou ensemble de vecteurs acoustiques, dans des classes. Pour associer les entités à reconnaître à des classes, il suffit de trouver la correspondance analytique entre les deux ensembles des vecteurs respectivement des séquences vocales et des entités à reconnaître de référence. A ce niveau, nous ne parlons pas d'unités phonétiques, mais uniquement de deux ensembles de vecteurs acoustiques de même dimension [77].

La conception d'un modèle génétique pour la classification des segments acoustiques du continuum vocal à des classes de phonèmes nécessite la définition de plusieurs concepts essentiels pour le bon fonctionnement de l'AG :

- nous définissons la nature de l'individu qui va composer la population de l'algorithme au fil des générations, le choix de son encodage génétique ainsi que la manière de segmentation du continuum vocal ;
- nous déterminons la nature et la taille de la population initiale, la nature de la fonction d'évaluation et le choix des opérateurs de reproduction ;
- nous choisissons les paramètres de réglage de l'AG ainsi que ses critères d'arrêt.

Etant donné que les vecteurs de la population initiale représentent l'espace de recherche des solutions potentielles au problème, l'AG aura pour tâche d'effectuer sur ces vecteurs initiaux un grand nombre de modifications, réorganisations, mutations, pour tenter de produire, au sens de la fonction d'adaptation, le meilleur vecteur de solutions potentielles ou les principaux meilleurs vecteurs. Ces derniers étant produits par l'AG, sont des tentatives de modèles de classement, c'est-à-dire, des tentatives d'attribution de chaque vecteur ou ensemble de vecteurs acoustiques représentant un phonème à une classe unique [78].

L'un des effets de la recherche génétique est de distribuer les segments parmi un nombre réduit de ces classes, et de laisser les autres vides, afin de classer tous les phonèmes du continuum vocal. L'alphabet de l'Arabe Standard est composé de quarante phonèmes dont vingt huit consonnes, six voyelles et six variantes vocaliques en contexte emphatique. Chacun de ces phonèmes représente une classe [79].

### 4.2.1. Population initiale

La population initiale peut être construite d'un ensemble de solutions approchées déjà connues, mais d'une manière générale, il est souvent avantageux de partir d'une population aléatoire dont la taille peut aller de quelques dizaines à quelques centaines. A noter que la vitesse de convergence de l'AG est de l'ordre de la taille de la population.

A l'initialisation de l'AG, nous avons choisi comme nombre initial de classes, noté  $N$ , le nombre  $n$  de phonèmes de l'alphabet multiplié par leurs différents contextes dans le continuum vocal, de manière à donner au processus la possibilité de construire toutes les classes possibles. Il y a toujours  $N$  classes potentielles, mais un certain nombre d'entre elles, que nous souhaitons le plus grand possible, reste vide. C'est là où réside l'importance de la fonction d'adaptation; est de réduire le nombre de classes effectives à une valeur optimale au cours des cycles de l'AG dont le nombre est fixé arbitrairement à une valeur assez grande.

### 4.2.2. Segmentation acoustique à court terme du continuum vocal

L'introduction du concept de population et en particulier de la population initiale dans le contexte de la RAP, passe par la définition préalable des deux ensembles suivants :

- l'ensemble noté  $T$  des  $M$  segments du continuum vocal tel que :

$$T = \{S_1, \dots, S_m, \dots, S_M\} \quad (4.01)$$

La durée de chaque segment est choisie égale à 20 ms. La dimension du vecteur de paramètres notée  $l$  correspondant de chaque segment acoustique dépend du type de phonème à classer. La position n°  $i$  contient le coefficient  $cx_{im}$  du segment n°  $m$ . Chaque segment acoustique est représenté par un vecteur paramétrique à court terme noté  $Vx_{lm}$  tel que :

$$Vx_{lm} = (cx_{1m}, cx_{2m}, \dots, cx_{im}, \dots, cx_{lm}) \quad (4.02)$$

Avec  $m = 1, \dots, M$

Et  $M = T/10$

T est la durée du continuum vocal.

- l'ensemble  $K$  de tous les phonèmes est donné par l'équation :

$$K = \{C_1, \dots, C_k, \dots, C_{40}\} \quad (4.03)$$

### 4.2.3. Vecteurs paramétriques à court terme

Le changement temporel, dans le spectre vocal, joue un rôle important dans la perception. Cette information est capturée sous forme de coefficients de vitesse et d'accélération connus sous le nom de coefficients différentiels ou dynamiques. Ces caractéristiques apportent plus d'informations relatives aux propriétés phonétiques des sons vocaux. Aucune information temporelle n'est incluse dans les MFCCs contrairement à leurs dérivées premières et secondes.

Aussi, la différence d'énergie entre les phonèmes permet de mieux les différencier et donc de les séparer. L'**Energie à Court Terme (ECT)** est simplement le logarithme d'énergie normalisé de chaque trame du signal vocal.

Concernant les voyelles, il est connu qu'elles sont stables. Ainsi et avec les semi voyelles, elles peuvent être représentées à court terme rigoureusement par un vecteur paramétrique noté  $V_{v_{42m}}$ , constitué des treize premiers coefficients MFCCs, du coefficient d'énergie ECT et de leurs dérivées premières et secondes (Fig.4.1) et (Fig.4.2).

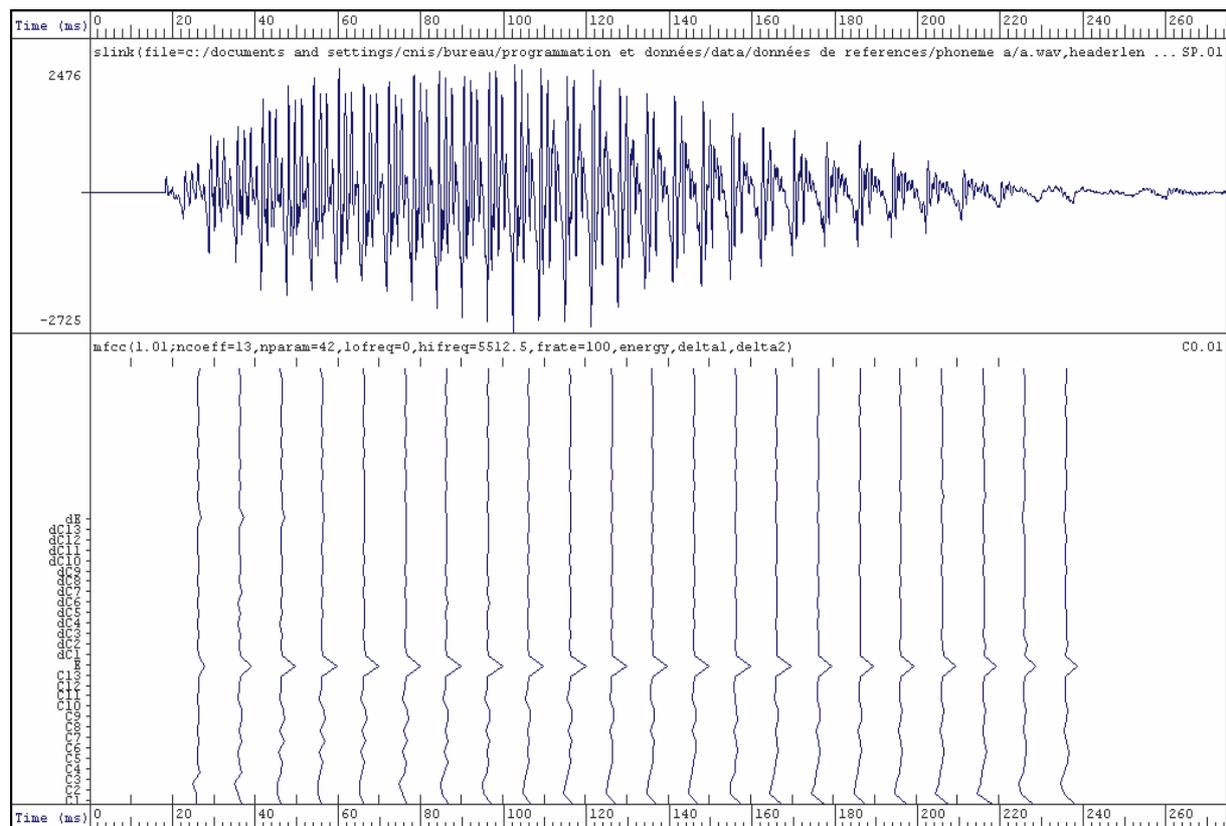


Figure 4.1 :Forme d'onde et schématisation des coefficients MFCCs, l'énergie et leurs dérivées premières et seconde de la voyelle [a]

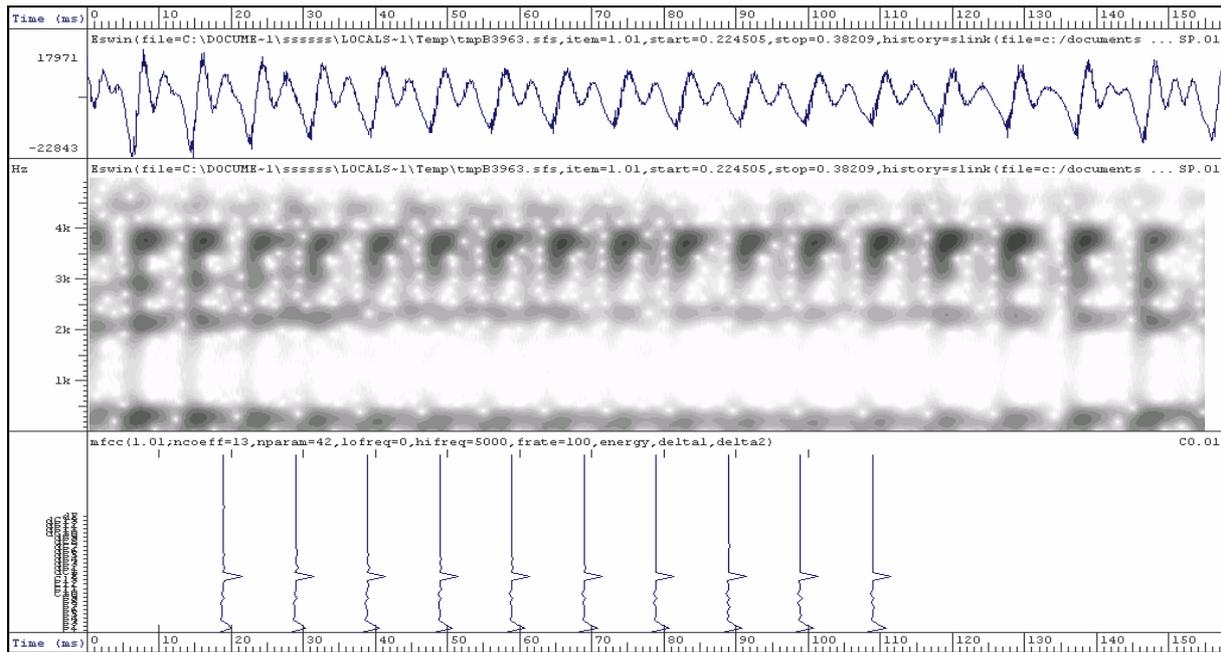


Figure 4.2 : Forme d’onde, spectrogramme à bande large de la semi-voyelle [y] et schématisation des coefficients MFCCs, l’énergie et leurs dérivées premières et secondes

Selon l’équation (4.02),  $Vv_{42m}$  est donnée par :

$$Vv_{42m} = (cv_{1m}, cv_{2m}, \dots, cv_{im_c}, \dots, cv_{42m}) \quad (4.04)$$

Pour les consonnes fricatives et nasales, le segment acoustique n°  $m$  est représenté par un vecteur paramétrique noté  $Vcf_{42m}$  qui contient les treize coefficients MFCCs, le coefficient SCT et leurs dérivées premières et secondes.  $Vcf_{42m}$  (Fig.4.3).

Il est donné par l’expression suivante :

$$Vcf_{42m} = (cf_{1m}, cf_{2m}, \dots, cf_{42m}) \quad (4.05)$$

Pour chercher les autres types de consonnes comme les occlusives, chaque vecteur paramétrique à court terme noté  $Vcs_{42m}$  représentatif du segment acoustique n°  $m$  est composé de treize coefficients MFCCs, le coefficient ECT et leurs dérivés premières et secondes. En outre, nous pouvons ajouter à ce vecteur, La Fréquence du **B**urst (**BF**) et le paramètre VOT. Le coefficient BF est un paramètre perceptuel très important pour la classification des consonnes occlusives. Il permet de trouver la fréquence qui correspond à la concentration d’énergie des occlusives. Cependant, nous nous sommes contentés de représenter ce type de consonnes de la même manière que les autres phonèmes pour des raisons de calculs (Fig.4.4).

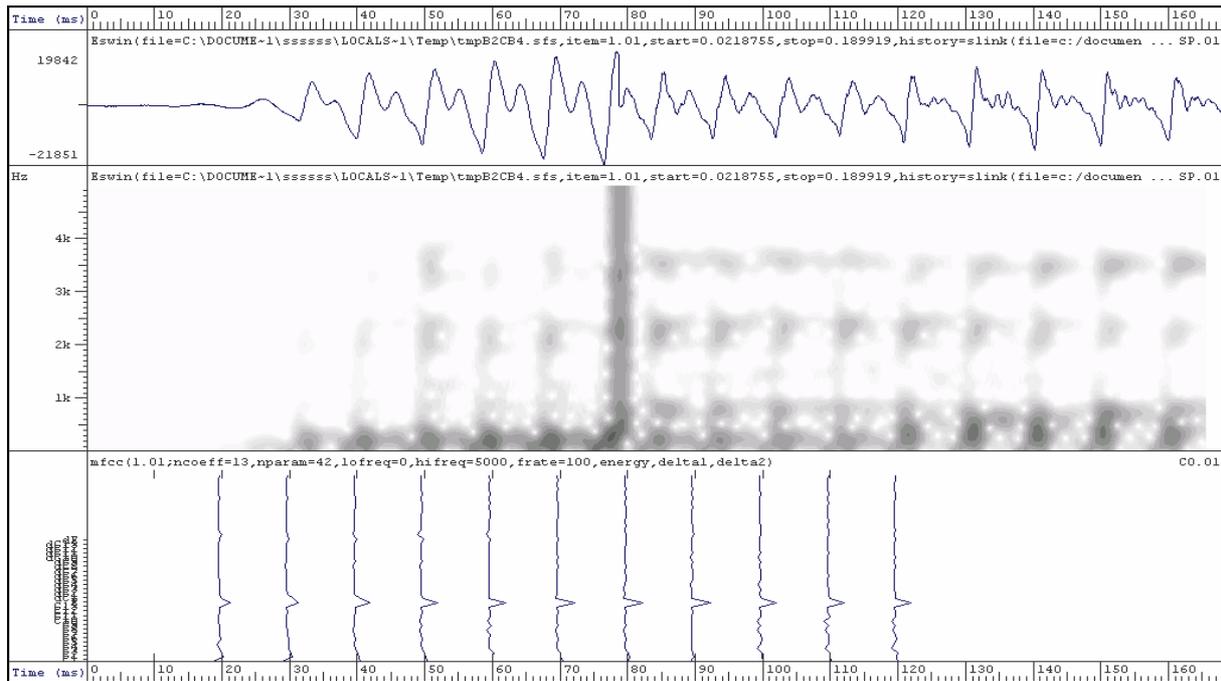


Figure 4.3 : Forme d’onde, spectrogramme à bande large de la consonne nasale [m] et schématisation des coefficients MFCCs, l’énergie et leurs dérivées premières et secondes

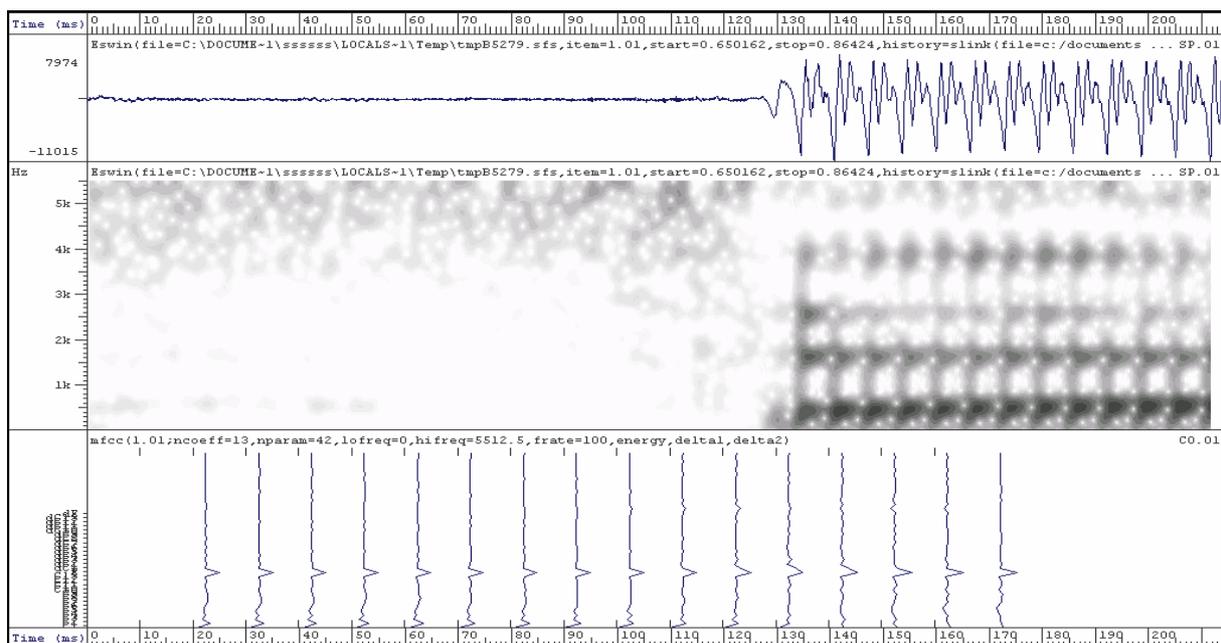


Figure 4.4 : Forme d’onde, spectrogramme à bande large de la consonne occlusive [θ] et schématisation des coefficients MFCCs, l’énergie et leurs dérivées premières et secondes

La description de ce vecteur est donné par :

$$Vcs_{42m} = (cs_{1m}, cs_{2m}, \dots, cs_{42m}) \tag{4.06}$$

#### 4.2.4. Encodage Génétique du continuum vocal

L'encodage génétique du continuum vocal passe inévitablement par la définition de l'individu de la population. L'individu choisi est un chromosome dont les gènes contiennent les coefficients  $cx_{im}$  de plusieurs segments  $S_m$  du continuum vocal du corpus choisi. L'individu est élaboré relativement au phonème à classer. Si par exemple, l'AG cherche à classer la consonne fricative [h], l'individu correspondant est défini comme étant le chromosome qui contient les coefficients de plusieurs segments acoustiques successifs et dont la taille est égale à l'ordre du vecteur de référence de cette consonne. Le processus de concaténation de ces segments successifs se fait du début jusqu'à la fin du continuum vocal avec un pas de 10 ms.

Ces vecteurs acoustiques qui sont notés  $Vx_{r, xm}$  ont une taille égale à l'ordre du vecteur de référence du phonème à classer noté  $vo_k$ . Ils résultent de la concaténation de  $r$  segments acoustiques. Ils sont assimilés à des individus, et par conséquent, la base de connaissances, l'ensemble de ces vecteurs, est assimilable à une population. Ces vecteurs acoustiques représentatifs des individus de la population des AGs sont donnés par les expressions suivantes selon le type du phonème à classer :

##### 4.2.4.1. Consonnes occlusives

$$Vx_{r, xm} = Vcs_{r, 42m} = (cs_{1m}, \dots, cs_{42m}, \dots, cs_{1(m+1)}, \dots, cs_{42(m+2)}, \dots, cs_{vo_k}) \quad (4.07)$$

Avec  $vo_k = r \times \eta \times 42$

et  $\eta$  : nombre d'échantillons par segments à court terme.

Et  $r$  est le nombre des segments acoustiques concaténés pour former les vecteurs acoustiques des individus (fig.4.5).

$cs_{1m}$	...	$cs_{42m}$	$cs_{1(m+1)}$	.....	$cs_{42(m+1)}$	...	$cs_{42(m+n)}$	...	$cs_{42(m+r)}$
-----------	-----	------------	---------------	-------	----------------	-----	----------------	-----	----------------

Figure 4.5 : Chromosome représentatif d'une consonne occlusive à classer

##### 4.2.4.2. Consonnes fricatives et nasales

$$Vx_{r, xm} = Vcf_{r, 42m} = (cf_{1m}, \dots, cf_{42m}, \dots, cf_{1(m+1)}, \dots, cf_{42(m+2)}, \dots, cf_{vo_k}) \quad (4.08)$$

Avec  $vo_k = r \times \eta \times 42$

Le vecteur acoustique de chaque consonne fricative et nasale peut être représenté par un chromosome ou individu (Fig.4.6).

$cf_{1m}$	...	$cf_{42m}$	$cf_{1(m+1)}$	.....	$cf_{42(m+1)}$	...	$cf_{42(m+n)}$	...	$cf_{42(m+r)}$
-----------	-----	------------	---------------	-------	----------------	-----	----------------	-----	----------------

Figure 4.6 : Chromosome représentatif d'une consonne fricative ou nasale à classer

#### 4.2.4.3. Voyelles et semi voyelles

$$Vx_{r, xm} = Vv_{r, 42m} = (cv_{1m}, \dots, cv_{im}, \dots, cv_{42m}, \dots, cv_{1(m+1)}, \dots, cv_{vo_k}) \quad (4.09)$$

Avec  $vo_k = r \times \eta \times 42$

Chaque solution potentielle est représentée sous forme d'une suite de bits ou bit string en Anglais. La taille de cette suite dépend de la précision numérique demandée. La précision de la solution est liée à celle des données utilisées dans l'implémentation de l'AG. Nous avons utilisé une double précision data type de 64 bits selon les spécifications IEEE Standard for Binary Floating-Point Arithmetic. Ce type de données a une précision de 15 - 17 digits [80].

Le continuum vocal noté  $c_r$  subit une segmentation acoustique dépendamment du phonème à classer. Les vecteurs représentatifs  $Vx_{r, xm}$  des segments obtenus ont une taille égale à l'ordre du vecteur qui représente l'image acoustique de référence du phonème objet de notre recherche. Pour obtenir ces segments, nous procédons à un balayage système de tout le continuum vocal avec un pas de 10 ms. Nous répétons cette opération pour chaque phonème de l'alphabet de l'AS.

Nous obtenons un ensemble de vecteurs acoustiques pour chaque balayage. Ces vecteurs vont constituer les individus ou chromosomes de la population représentative de l'espace de recherche. Nous pouvons mener cette opération en parallèle pour tous les phonèmes de l'alphabet et obtenir quarante sous populations qui vont constituer l'espace de recherche d'un seul AG séquentiel ou de quarante AGs en parallèles.

#### 4.2.5. Données acoustiques de référence

Le choix du type, la fiabilité des indices et les paramètres acoustiques utiles des phonèmes à classer sont très importants. Ces derniers représentés sous formes de vecteurs acoustiques, constituent l'ensemble des données de référence pour l'AG. L'AG se réfère à cet ensemble, durant la phase d'évaluation des solutions potentielles au problème [81].

Chaque phonème est représenté par un vecteur centi-seconde qui représente un segment acoustique du continuum vocal de quelques centièmes de seconde de longueur temporelle. Chacun des quarante phonèmes de l'AS qui constitue une classe, est représenté par un vecteur acoustique de référence d'ordre spécifique noté  $\mu_{kj}$  (tab.4.1).

Ces vecteurs paramétriques moyens de référence sont donnés par l'expression suivante :

$$\mu_{Kj} = (\mu_{K1}, \mu_{K2}, \dots, \mu_{Kj}) \quad (4.10)$$

Ces vecteurs acoustiques qui sont des images de référence sont obtenus lors de la phase d'apprentissage en utilisant la relation suivante :

$$\mu_{kj} = \frac{1}{N_k} \sum_{z=1}^{N_k} v_{kz}, \quad 0 \leq k \leq 40 \quad (4.11)$$

Où  $N_k$  est le nombre des vecteurs pré classés de la classe k ;

et  $v_{kz}$  est le vecteur numéro z de la classe k.

Tableau 4.1 : Vecteurs acoustiques de référence des phonèmes de l'AS

Phonèmes de l'Arabe	vecteur de référence $\mu_{Kj}$	ordre du vecteur de référence ( $v_{o_k}$ )
[ b ]	$(\mu_{11}, \dots, \mu_{1bb})$	bb
....	.....	...
[ d ]	$(\mu_{k1}, \dots, \mu_{kdd})$	dd
....	.....	...
[ a ]	$(\mu_{401}, \dots, \mu_{40m})$	m

#### 4.2.6. Population des vecteurs acoustiques

Les vecteurs acoustiques sont assimilés à des individus, et par conséquent, l'ensemble de ces vecteurs est assimilable à une population. Cette population va subir une pression sélective qui va, augmenter la qualité moyenne des individus qui la composent. Il faut continuellement renouveler la population, traquer les individus inadaptés pour les remplacer par de nouveaux.

Comme pour une population réelle, des individus naissent et meurent, les effectifs varient. Il faudra donc porter son attention au contrôle de la taille de la population. Si trop d'individus meurent, l'espèce s'éteint, la base de connaissances se vide. Si le taux de natalité est trop important, la population croît exponentiellement. La base de connaissances devient vite ingérable et il est alors difficile de reconnaître les bons individus et la mémoire de la machine sera vite saturée.

Une première méthode simpliste est simplement d'empêcher toute variation de la taille de la population. Elle est fixée dès le départ, toute naissance entraîne une mort et inversement. Tout le problème est alors de choisir une dimension adaptée : une population trop petite ne permet pas une exploration efficace. Par contre une population trop grande peut ne pas se stabiliser vers un optimum global et sa gestion consomme inutilement des ressources [82].

#### 4.2.7. Choix de la Fonction d'évaluation

La fonction d'évaluation ou fonction fitness quantifie la qualité de chaque chromosome par rapport au problème. La valeur de cette fonction pour un individu est appelée score de ce dernier. Cette fonction est celle que nous voulons optimiser. Pour les algorithmes standards d'optimisation, elle est appelée fonction objectif. Elle est utilisée pour sélectionner les chromosomes pour la reproduction. Les chromosomes ayant une bonne qualité ont plus de chance d'être sélectionnés pour la reproduction et donc plus de chance pour que la population suivante hérite de leur matériel génétique. La fonction d'évaluation produit la pression qui permet de faire évoluer la population de l'AG vers des individus de meilleure qualité. Clairement, le choix de la fonction d'évaluation va fortement influencer sur le succès de l'AG.

La matrice composée des vecteurs acoustiques des séquences vocales et de leurs vecteurs acoustiques de référence, est utilisée par la fonction objectif qui doit être la plus discriminante possible. Cette discrimination porte sur deux aspects complémentaires [81]:

- la cohésion interne ou le degré d'affinité des segments à l'intérieur de chaque classe ;
- la dissimilarité des classes entre elles.

La fonction d'évaluation notée  $F_k$  peut être représentée de plusieurs manières selon le choix du critère décisionnel de classification, choisi. Parmi ces critères les plus utilisés nous avons le critère Bayésien et celui qu'est à la base de mesure de distance. Ce dernier est le plus simple et le plus direct. L'idée de base du critère de distance pour la classification est qu'une donnée est affectée à la classe qui lui est la plus proche. Une variété de fonctions de

distance sont valables parmi lesquelles nous avons la fonction de Minkowsky et celle de Mahalanobis.

Dans notre cas, la fonction d'évaluation  $F_k$  est représentée par la fonction de la distance choisie entre les vecteurs images des segments acoustiques du continuum vocal et le vecteur image acoustique de référence de chaque phonème de l'AS (Fig.4.7).

Pour évaluer les solutions potentielles à notre problème, il faut minimiser au maximum cette fonction d'évaluation pour pouvoir décider du type de classes d'appartenance des segments acoustiques ou des ensembles de ces segments. Pour cela, il suffit de maximiser la fonction inverse de la fonction d'évaluation, c'est-à-dire, chercher les extrêma de la fonction inverse notée  $F_e$  [81, 82]:

$$F_e = 1 / F_k \quad (4.12)$$

La recherche de ces extrêma peut donner deux types de résultats :

- l'obtention d'un maximum signifie l'existence du phonème recherché dont la classe est bien définie. Dans ce cas, nous continuons la recherche du même phonème pour le reste du continuum vocal ;
- la fin du continuum vocal signifie qu'il n'y a pas de phonème de la classe k dans ce continuum. Dans ce cas, nous répétons le travail effectué pour les autres phonèmes.

Dans le cas de la fonction Euclidienne notée  $F_k e$ , la distance entre les vecteurs paramétriques des segments acoustiques  $Vx_{r_k \times m}$  et le vecteur de référence de chaque classe de phonèmes  $\mu_{kj}$  est donnée par (Fig.4.8) :

$$F_k = F_k e = \sqrt{\sum_{j=1}^{v_{o_k}=42 \times r \times \eta} (\vec{V}x_{r, j} - \vec{\mu}_{kj})^2}$$

(4.13)

Avec :  $j = x \times m \times \eta$  ;

et :  $x = 1, \dots, 42$ .

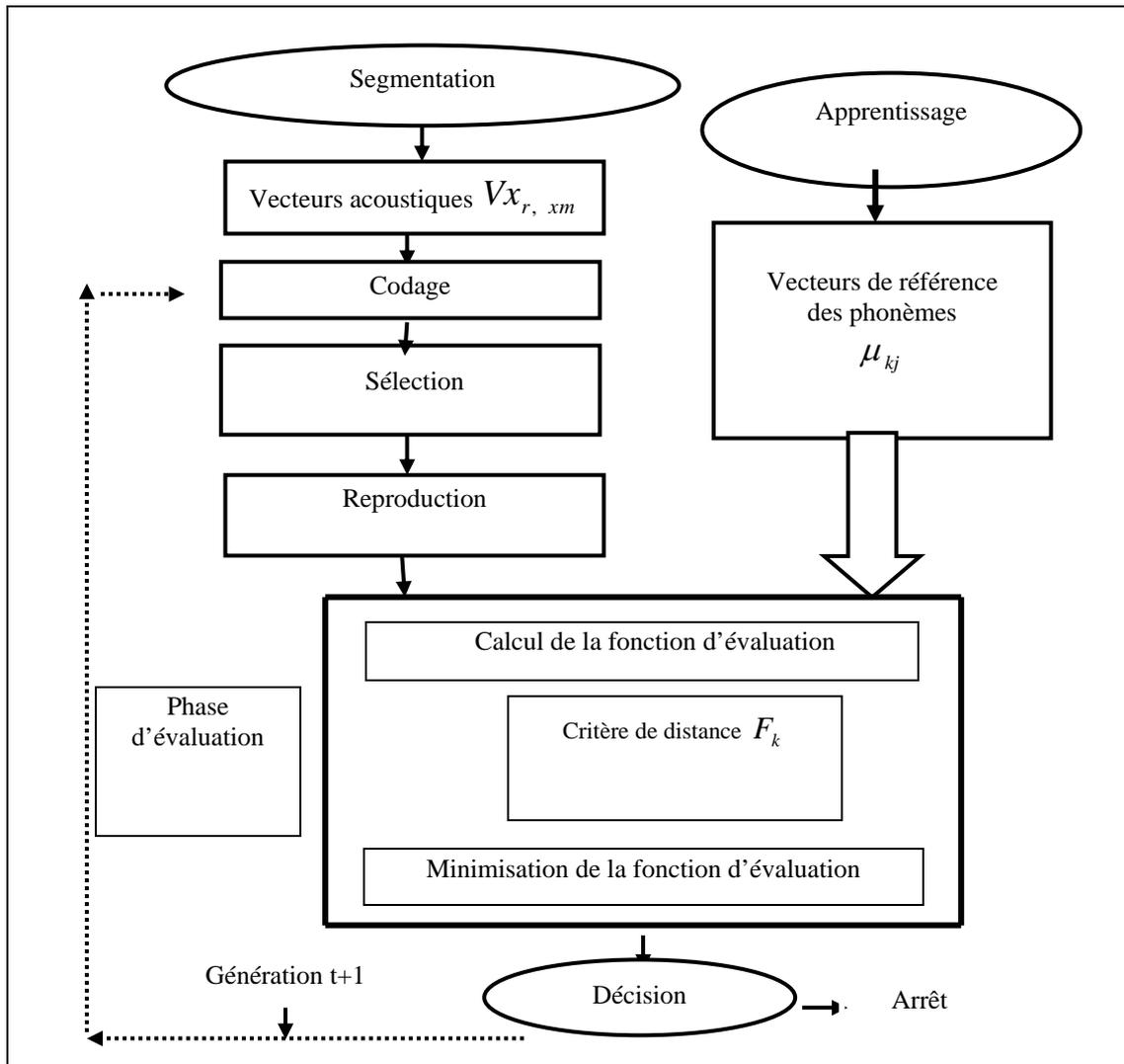


Figure 4.7 : Algorithme Génétique de classification acoustique [81]

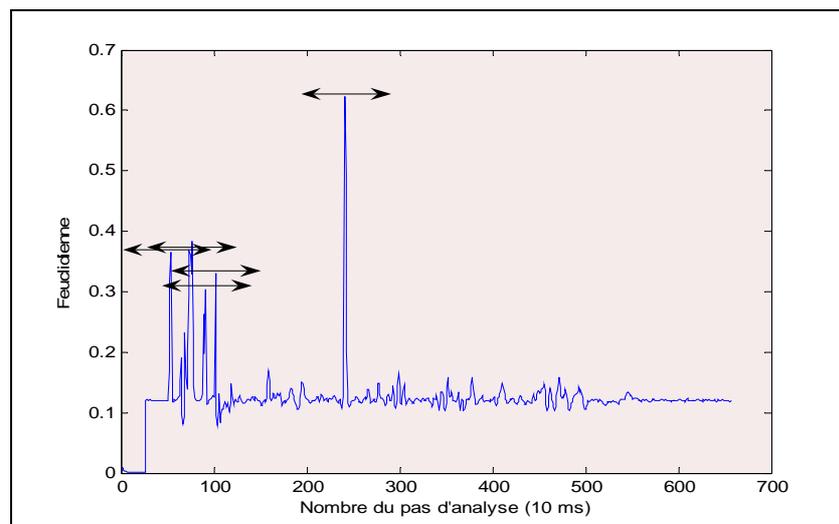


Figure 4.8 : Application de l'inverse de la fonction de distance d'Euclide pour la recherche de la voyelle [i] dans la phrase p1 (Annexe 2)

Les extrema de l'inverse de la fonction Euclidienne correspondent aux voyelles [i] existantes dans la phrase p1.

Nous pouvons aussi représenter  $F_k$  par la fonction de distance de Manhattan notée  $F_k mt$ , dont l'expression est la suivante (Fig.4.9) :

$$F_K = F_k mt = d(cx_j, \mu_j) = \sum_{j=1}^{v_{o_k}} |cx_j - \mu_j| \quad (4.14)$$

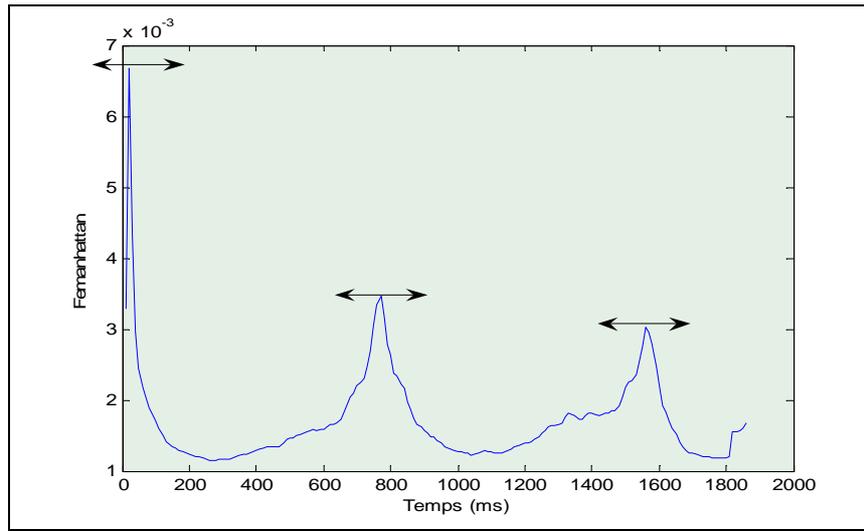


Figure 4.9 : Application de l'inverse de la fonction de distance de Manhattan pour la recherche de la consonne [ʃ] dans le mot isolé [ʃ a ʃ i ʃ u]

Les pics de l'inverse de la fonction de Manhattan traduisent l'existence de trois consonnes [ʃ] dans le mot isolé [ʃ a ʃ i ʃ u].

Dans le cas de la fonction de distance de Mahalanobis notée  $F_k mb$ , la fonction d'évaluation s'écrit comme suit :

$$F_k mb = d(x, y) = \sqrt{(x - y) \text{cov}(D) (x - y)} \quad (4.15)$$

où :  $\text{cov}(D)$  est la matrice de covariance. Si les éléments de  $x$  et de  $y$  sont indépendants, la matrice de covariance sera l'identité et la distance de Mahalanobis serait égale à la distance Euclidienne.

#### 4.2.8. Convergence des Algorithmes Génétiques

Si l'AG a été correctement implanté, la population évoluera au fur et à mesure des générations de telle sorte que la qualité moyenne des individus se rapproche du maximum global (Fig. 4.10).

La convergence est une progression vers l'uniformité de la population puisque les meilleures caractéristiques ou gènes auront tendance à apparaître de plus en plus dans les individus alors que les mauvaises auront tendance à disparaître [83].

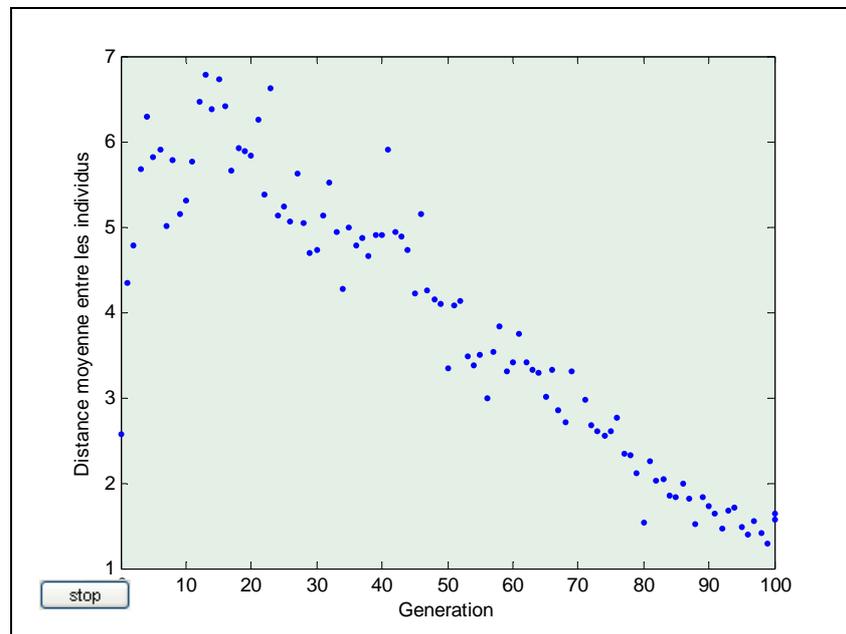


Figure 4.10 : Convergence de l'AG vers l'individu qui représente l'optimum global au fil des générations

### 4.3. Parallélisme des Algorithmes Génétiques

Il a été établi que l'efficacité des AGs à trouver les solutions optimales est largement tributaire de la taille de la population. Avec une large population, la diversité génétique augmente, ainsi l'algorithme devient plus apte à trouver les optima globaux. Cependant, une large population nécessite plus de mémoire pour son stockage; il a été prouvé qu'elle nécessite plus de temps pour converger [84].

Actuellement, l'utilisation en parallèle de nouveaux ordinateurs nous permet non seulement de gagner en espace de stockage mais aussi d'utiliser plusieurs processeurs pour produire et évaluer plus de solutions en un temps très réduit.

Il existe plusieurs méthodes de parallélisme des AGs comme celle dite maître esclave AGs parallèle ou Master Slave Parallel GAs en Anglais ou encore parallélisme globale et celle dite AGs parallèle de sous-populations ou Multiple-demes parallel GAs en Anglais ou encore parallélisme à gros grains ou coarse-grained parallelisation en Anglais

### **4.3.1. Parallélisme global**

Selon cette méthode, l'AG parallèle fait évoluer une seule population, comme dans l'algorithme séquentiel, mais l'évaluation des individus et/ou l'application des opérateurs génétiques se fait en parallèle. Le calcul de la qualité des individus de la population est parallélisé en assignant à chaque processeur une fraction de la population qu'il sera chargé d'évaluer. Ce calcul étant indépendant du reste de la population, aucune communication entre les processeurs n'est nécessaire entre le moment où un processeur reçoit ses individus et celui où il renvoie leurs qualités.

### **4.3.2. Algorithmes Génétiques Parallèles à Gros Grain**

Les Algorithmes Génétiques Parallèles à Gros Grains (AGPGG) sont aussi appelés modèle en îlots ou island model en Anglais, car ils ressemblent à un modèle utilisé pour décrire des populations naturelles isolées par une distance entre elles comme les îlots. Pour ce modèle, la population est divisée en plusieurs sous-populations qui vont évoluer séparément dans différents processeurs, mais qui échange des individus occasionnellement. Cet échange est appelé migration. Il existe différentes stratégies pour contrôler la migration. Ces algorithmes introduisent des paramètres supplémentaires par rapport aux algorithmes séquentiels comme le nombre et la taille des sous-populations, le taux de migration ou le nombre des individus échangés et la fréquence de migration. Ils nécessitent également des choix de la stratégie pour la sélection des individus échangés, leur remplacement dans la sous-population qui les accueille et le choix d'une topologie pour leur migration en déterminant les connexions entre les sous-populations. Ces paramètres ont un impact important sur l'efficacité de l'algorithme, tant du point de vue de la rapidité que de la qualité des solutions.

Ce type d'AG introduit donc des changements fondamentaux par rapport à l'algorithme séquentiel et possède un comportement différent [84].

### **4.3.3. Algorithmes Génétiques parallèles à Sous-populations isolées**

Pour ce type d'AG, les sous-populations évoluent en complète isolation entre elles. Sans communication, le taux de migration est nul. Dans ce cas, la performance du parallélisme est simplement le rapport entre le temps d'évaluation des individus du cas séquentiel et celui du cas parallèle. Cette topologie que nous avons choisie pour notre AG parallèle (Fig. 4.11).

Pour chaque classe de phonème, nous avons une sous-population dont les individus sont définis comme étant les vecteurs acoustiques  $Vx_{r, xm}$ . Nous divisons notre population en trois groupes de sous-populations relativement à la nature de leurs individus [85].

Le premier groupe est appelé **Sous-Population de Voyelles et Semi-Voyelles (SPVSV)**. Il contient comme individus, les vecteurs acoustiques  $Vv_{r, 42m}$  (équat.4.09). Ce groupe est divisé en seize sous-populations ou demes en Anglais relativement au nombre des voyelles et semi-voyelles de l'AS. Chaque sous-population qui correspond à un seul phonème constitue la population initiale de l'AG séquentiel de base.

Le second groupe est appelé **Sous-Population des Occlusives et des Liquides (SPOL)**. Il contient comme individus, les vecteurs acoustiques  $Vcs_{r, 42m}$  (équat.4.07). Ce groupe est divisé en neuf demes qui correspondent au nombre d'occlusives et des consonnes liquides.

Le troisième groupe qui est appelé **Sous-Population des consonnes Fricatives et Nasales (SPFN)** contient comme individus, les vecteurs acoustiques  $Vcf_{r, 42m}$  (équat.4.08). Il est divisé en quinze demes relativement au nombre des consonnes fricatives et nasales.

Il résulte que pour chaque deme, un AG séquentiel de base évolue en parallèle sans aucune migration des individus.

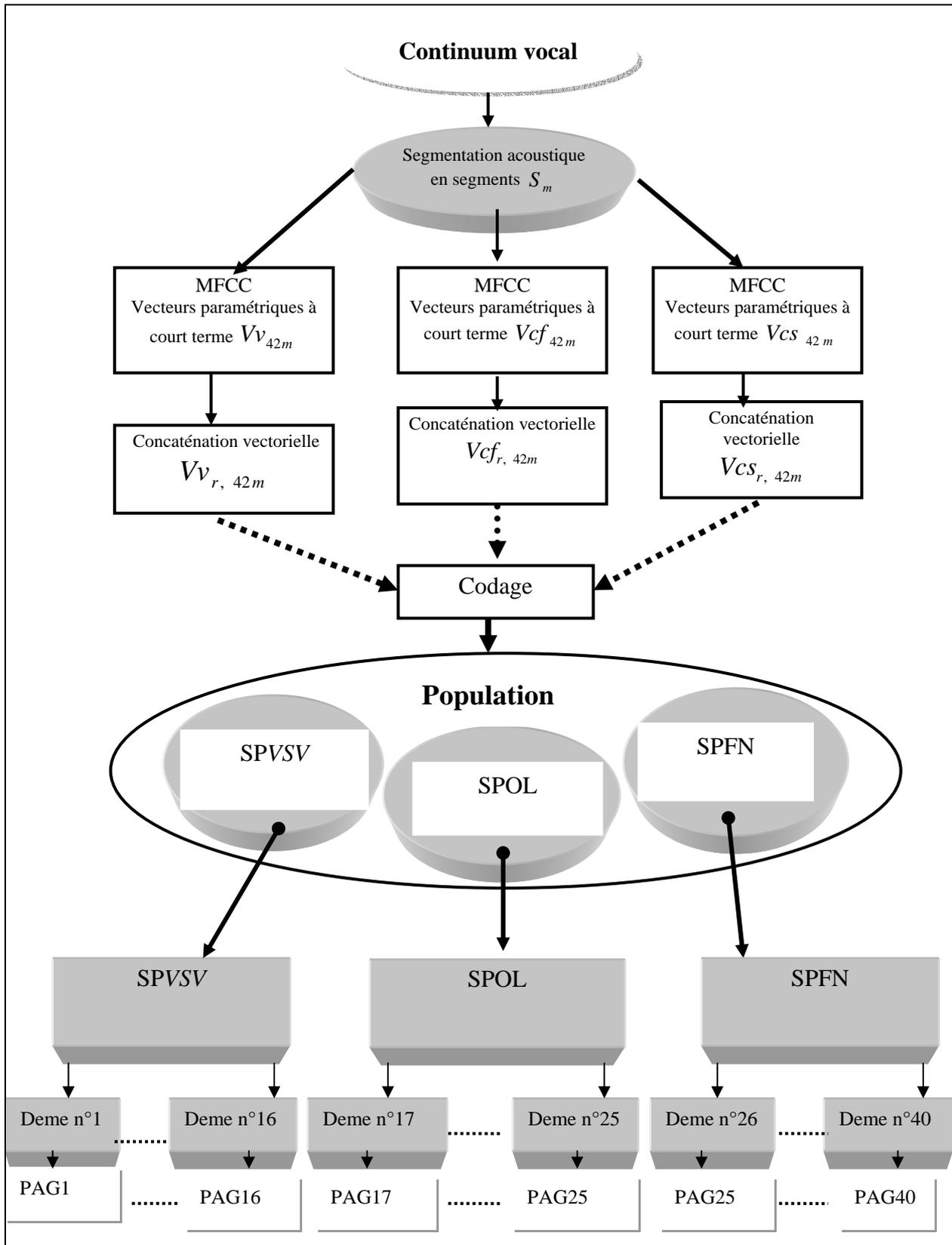


Figure 4.11 : Déroulement de l'Algorithme Génétique Parallèle sans migration pour la classification acoustique des phonèmes

#### **4.4. Conclusion**

Dans ce chapitre, nous avons explicité l'usage des Algorithmes Génétiques comme modèle de classification acoustique des phonèmes de l'Arabe Standard. Ce modèle se base sur la notion de mesure de distance pour évaluer la qualité des individus qui sont candidats à la solution globale. Notre AG procède à l'optimisation de la fonction d'évaluation qui est choisie comme étant un critère de distance bien défini. Nous avons choisi plusieurs types de fonctions de distance comme celle de Manhattan ou d'Euclide et chercher celle qui est la plus adaptée à notre problème.

Afin d'accroître les performances de notre modélisation génétique et de réduire le coût de calculs, nous avons utilisé un modèle génétique parallèle ayant comme topologie celle de la Sous-population isolée qui se base sur le fonctionnement parallèle de plusieurs AGs séquentiels sans migration.

## Chapitre 5 : Résultats de la classification par Algorithmes Génétiques des phonèmes de l'Arabe Standard

Nous commençons ce chapitre par un ajustement des paramètres de réglage de notre **Modèle Génétique Séquentiel** noté **MGS** et par le choix de ses opérateurs de reproduction et de sa fonction d'évaluation pour atteindre les meilleurs taux de classification. Nous montrons que les résultats obtenus sont influencés par les valeurs respectives de la probabilité de croisement, celles de la mutation et de la sélection. En outre, nous remarquerons que la nature de la fonction d'évaluation ainsi que la méthode de remplacement jouent un rôle crucial dans les performances de notre MGS.

Nous montrons que le MGS donne de meilleurs résultats de classification des phonèmes de l'AS que les classificateurs à distance seuls.

Nous terminons ce chapitre par une interprétation des résultats obtenus.

### 5.1. Implémentation

Nous avons implémenté notre modèle ou algorithme génétique en Matlab version 7.0, en utilisant les fonction de l'Algorithme Génétique de recherche Directe et d'Optimisation ou Genetic Algorithm of direct Search and Optimisation Toolbox en Anglais. Ces fonctions concernent toutes les étapes de fonctionnement d'un AG séquentiel ou parallèle. Donc, nous avons fait appel à ces fonction pour développer et exécuter notre application (voir annexe 1).

Les données utilisées par notre application génétique sont divisées en deux groupes. Celui qui représente les données de référence des phonèmes à classer et celui qui constitue le corpus à reconnaître (voir annexe 2).

Notre corpus est un ensemble de plusieurs sous corpus de différentes tailles qui représentent des phrases, des mots isolés ou une succession de paragraphes. Ils sont enregistrés sous forme de fichiers wave. La taille de chacun d'eux ne dépasse pas les 2,5 Mo pour des raisons de mémoire de calcul (Fig. 5.1).

L'expérimentation a porté sur un corpus d'une centaine de phrases de la parole continue de l'Arabe Standard, enregistrées, dans un milieu ambiant, par un seul locuteur qui est un

chercheur Jordanien. Nous avons obtenu des taux de classification de chaque type de phonèmes dans différents contextes vocaliques.

Les données de ces deux groupes sont sous la forme de vecteurs dont les coefficients sont les MFCCs, l'énergie et leurs dérivées premières et secondes. Ces coefficients sont obtenus en utilisant le logiciel Speech Filing System (SFSwin) [86].

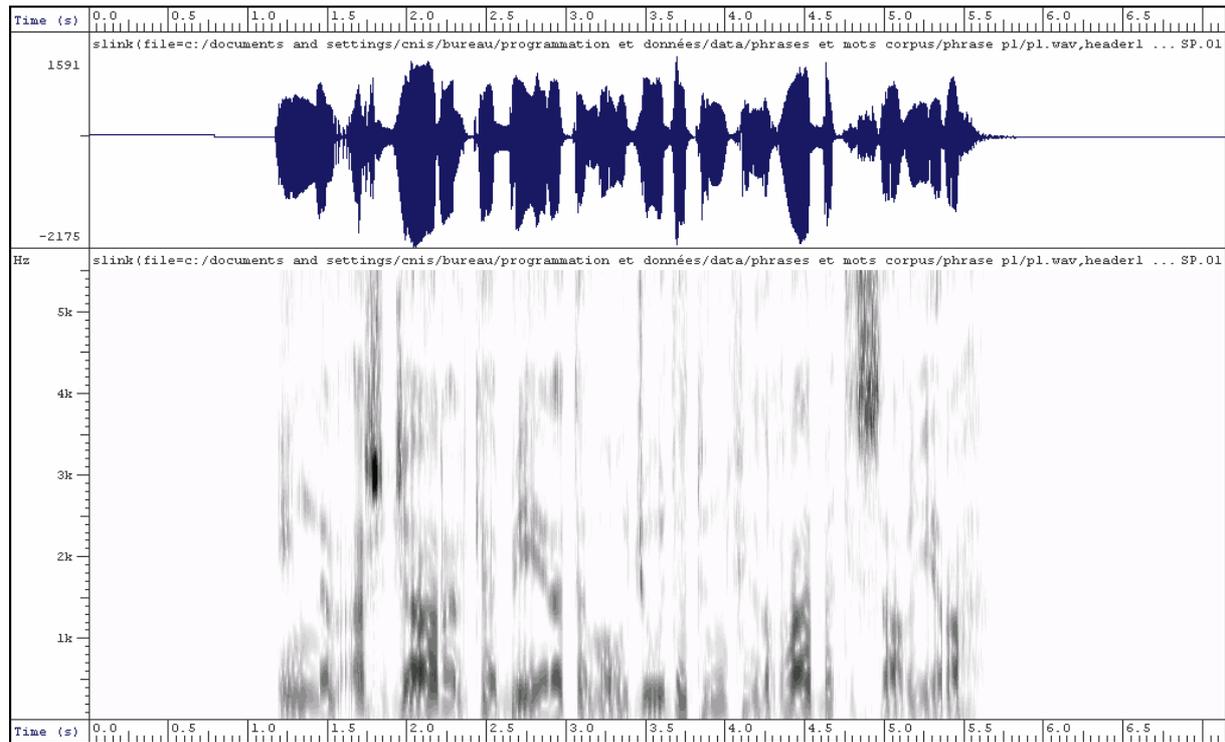


Figure 5.1 : Forme d'onde et spectrogramme à bande large de la phrase p1 (Annexe 2)

## 5.2. Paramètres du Modèle Génétique Séquentiel

Il est nécessaire d'effectuer des choix concernant la manière de faire évoluer les individus. Partant d'une méthode basée sur les AGs, nous désirons déterminer quelle mutation et quel croisement choisir, quels taux (fréquences d'agissement) leur associer ? Tout d'abord, il faut décider des types de mutation et de croisement, ainsi que de leurs taux. La littérature dans le domaine donne une idée de l'ordre de grandeur des probabilités ou taux de mutation  $p_m$  ou de croisement  $p_c$ . En effet, celui de la mutation est rarement au-dessus de 10%, celui du croisement souvent supérieur à 60%.

Nous avons fixé le taux de croisement à 80% et celui de la mutation à 10%. Par la suite, nous avons étudié l'influence de ces taux pour quelques valeurs choisies afin de déterminer

celui qui donne les meilleures performances du modèle. Ensuite, nous avons présenté l'influence du taux de croisement sur le bon fonctionnement du MGS.

Les opérateurs de reproduction opèrent sur une population d'individus de taille choisie égale à 1000 durant toutes les générations afin d'accroître les performances de notre modèle génétique. Aussi, nous avons choisi le nombre maximal de générations égale à 500 d'une manière empirique. Le remplacement des individus d'une génération à l'autre se fait avec un taux ou probabilité de remplacement égal à 0.5.

Nous avons effectué une série d'expériences afin de trouver les meilleures valeurs des différents paramètres de l'évolution du MGS. .

### 5.2.1. Taille de la population et parallélisme

Nous pouvons constater que les performances des populations augmentent avec leur taille. En effet, une population de 100 individus obtient des performances de 65,16% pour le meilleur individu, alors qu'une population de 500 individus obtient des performances de 75,57% pour le meilleur individu. Grâce aux nouvelles performances des circuits intégrés, nous pouvons nous permettre de travailler sur de grandes populations. Nous avons choisi une population de 1000 individus car ses performances sont raisonnables plus de 80% pour le meilleur individu.

Cependant, avec une taille supérieure à mille, les performances du MGS diminuent rapidement. Donc, plus la taille de la population augmente et plus la convergence de l'AG vers un extremum est compromise. Pour y remédier nous avons choisi un modèle parallèle de type à gros grains ou en îlots sans migration. Pour chaque sous population de taille égale à mille, nous appliquons l'AG séquentiel. Ainsi, nous pouvons travailler sur n'importe quelle population avec des tailles aussi grande que possible. Il suffit de diviser la population mère en  $n$  sous population ayant chacune une taille de 1000 individus et appliquer l'AG séquentiel sur chacune d'elles (Fig.5.2).

Les résultats montrent que 100 individus ne sont pas en nombre suffisant pour faire évoluer les performances du meilleur individu, alors que 500 individus permettent d'améliorer les performances de plus de 25 % en 1000 générations (Tab.5.1 ).

L'utilisation du **Modèle Génétique Parallèle** noté (**MGP**) au lieu du modèle séquentiel a permis d'atteindre les meilleurs scores au bout de moins de 500 générations. Nous avons

choisi le modèle parallèle de type à gros grains ou en îlots sans migration pour lequel chaque sous population a une taille identique égale à 1000. A chacune d'elles, est appliqué un MGS séquentiel.

Tableau : 5.1: Evolution du taux global de classification des phonèmes  
en fonction de la taille de la population

Type de l'AG	Modèle Génétique Séquentiel						
Taille de la population (individus)	100	200	500	750	1000	1100	1500
Taux global de Classification des Phonèmes (%)	17.25	35.12	55.45	67.85	81.17	72.14	41.12

Ces résultats montrent que notre AG parallèle est performant puisque le taux moyen global de classification des phonèmes de l'AS noté  $T_c$  peut atteindre les 81.20 %. Ce taux est obtenu en choisissant pour chaque MGS la technique de la mutation uniforme, le croisement bi-points, la sélection par tournoi et par la roulette biaisée et la technique du remplacement stationnaire avec un taux ou probabilité de remplacement notée  $p_r$  égal à 0.5.

### 5.2.2. Diversité des individus et convergence

La mesure de diversité est calculée comme le rapport du nombre d'individus distincts sur le nombre total d'individu de la population. Ainsi, une diversité égale à '1' correspond à l'initialisation de l'algorithme (tous les individus ont été générés aléatoirement), alors qu'une diversité égale à zéro correspond à la convergence de l'algorithme vers une solution optimale.

$$diversité = \frac{\text{nombre d'individus distincts}}{\text{nombre d'individus}} \quad (5.1)$$

Lorsqu'une mutation pertinente se produit, il se produit un pic de la fonction de diversité. Par conséquent, si nous supposons que les phases d'explorations et d'exploitation sont complètement disjointes, les pics de la fonction diversité correspondent à la phase d'exploitation de l'algorithme de recherche, alors que les plateaux montrent la phase d'exploration (Fig. 5.3).

Une meilleure diversité des individus permet de faciliter la convergence rapide vers la solution.

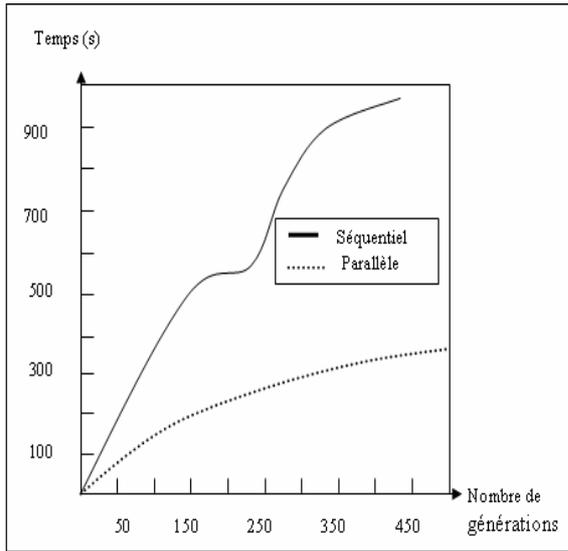


Figure 5.2: Comparaison des délais d'exécution entre les Algorithmes séquentiel et parallèle

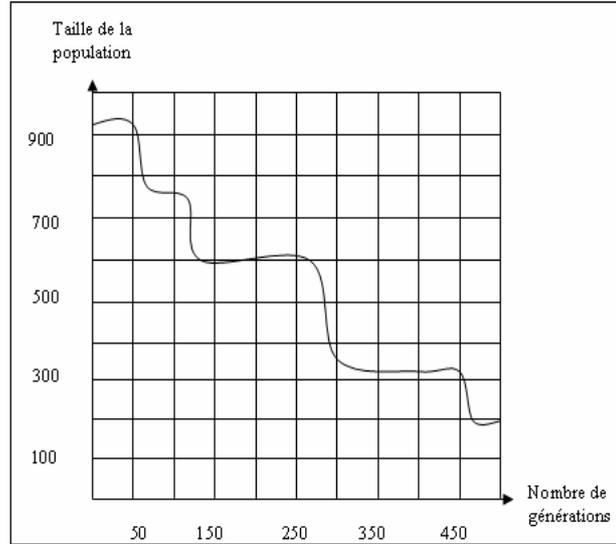


Figure 5.3 : Evolution de la diversité des individus en fonction des générations

Nous constatons que l'utilisation du modèle génétique parallèle au lieu du modèle séquentiel a permis d'atteindre les meilleurs scores au bout de moins de 500 générations. Cette rapidité d'exécution de l'AG a garanti sa convergence vers la solution en le rendant plus performant. Quant à la taille de la population nous constatons que plus elle est grande et plus le facteur de diversité est grand. Ainsi, nous avons choisi cette taille égale à 1000, en adaptation avec notre problème, ce qui permet d'éviter la convergence prématurée qui affectera les performances du MGS.

### 5.2.3. Opérateurs Génétiques de Reproduction

Concernant la mutation, il est nécessaire de prendre une mutation adaptée aux valeurs réelles que nous manipulons. Nous avons constaté, lors de nos simulations, que les résultats sont assez mauvais avec la mutation Gaussienne ; le meilleur individu a un taux de reconnaissance alors que la mutation uniforme a plus de 80%. Les mutations non uniforme et aléatoire obtiennent, quant à elles, des performances non concluantes (Tab. 5.2).

Tableau : 5.2 : Comparaison des taux de classification suivant le type de mutation

Type de mutation	Gaussienne	uniforme	non uniforme	aléatoire
Taux global de Classification des Phonèmes	67.25 %	81.20 %	64.12 %	63.15 %

Quant au taux ou probabilité de mutation  $p_m$ , les résultats montrent, que lorsqu'il est fixé dans une fourchette de 0.08 à 0.2, il nous permet d'éviter de tomber dans le piège des extrema locaux et donc d'éviter de réduire les performances du système. Cependant, les résultats obtenus montrent que lorsque le taux de mutation est inférieur à 0.08 ou supérieur 0.2, notre modèle MGS converge prématurément (Fig. 5.4).

Concernant l'opérateur de recombinaison, un seul type a été implémenté pour cette expérience qui est le crossover bi-points. Les simulations sont effectuées sur une population de 1000 individus soumis à une mutation dont la probabilité  $p_m$  est égale à 0.1. Les résultats montrent que les meilleurs taux de croisement  $p_c$  qui influent sensiblement sur les performances des individus sont compris entre 40 et 80% (Fig. 5.5).

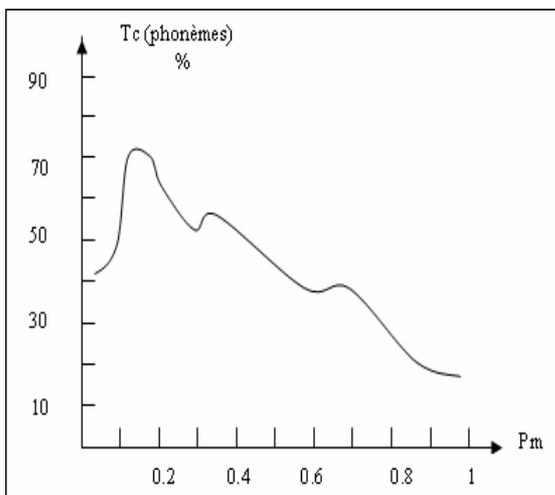


Figure 5.4 : Evolution du taux de classification global des phonèmes en fonction de la probabilité de mutation

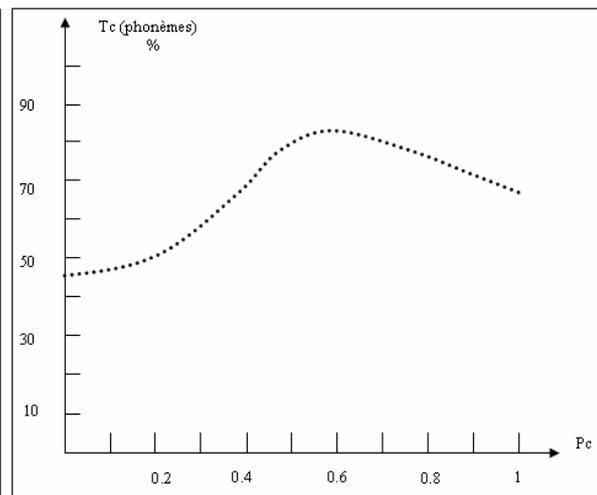


Figure 5.5 : Evolution du taux de classification global des phonèmes en fonction de la probabilité de Croisement

Nous déduisons que le taux de classification est influencé par les valeurs respectives de la probabilité de croisement et celle de la mutation. La valeur du taux de croisement égale à 0.8 permet la convergence du MGS vers un optimum global après un certain nombre d'itérations. Nous constatons que les meilleures valeurs de la fonction fitness sont celles obtenues avec une probabilité de croisement qui avoisine la valeur 0.8 (Fig. 5.6).

Cette influence devient de plus en plus grande à chaque fois que la valeur de probabilité de croisement diminue. La recherche des solutions se limite à une partie de l'espace de recherche, quand la valeur de probabilité de croisement n'est pas bien choisie.

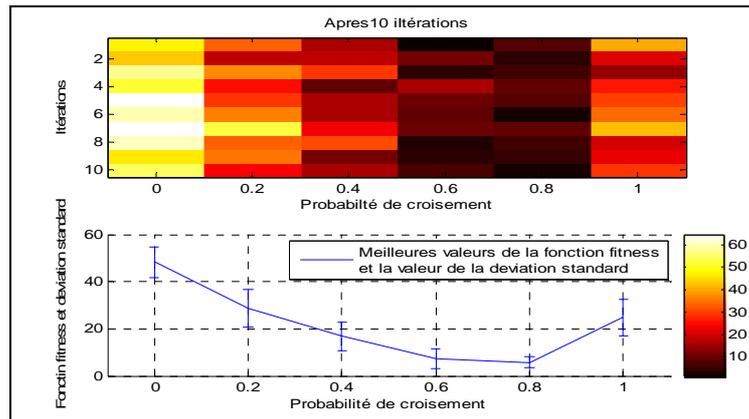


Figure 5.6 : Influence du taux de croisement sur la convergence de l'AG

### 5.2.4. Fonction d'Evaluation

Chaque chromosome apporte une solution potentielle au problème à résoudre. Néanmoins, ces solutions n'ont pas toutes le même degré de pertinence. Cette fonction fait évoluer la population dans un sens bénéfique pour la recherche de la meilleure solution. Autrement dit, elle doit pouvoir attribuer à chaque individu un indicateur positif représentant sa pertinence pour le problème que nous cherchons à résoudre. Par conséquent, l'influence de cette fonction sur les performances de l'AG est importante. Aussi, il est impératif de faire un choix judicieux de cette fonction.

Avec un croisement bipoints et une mutation uniforme, le taux de classification des phonèmes  $T_c$  varie selon la nature de la fonction d'évaluation choisie (Tab. 5.3).

Tableau 5.3 : Evolution du taux global de classification des phonèmes en fonction du type de la fonction d'évaluation choisie.

Type de la fonction d'adaptation	$F_{ke}$	$F_{kmb}$	$F_{kmt}$
Taux global de Classification des Phonèmes	71.50 %	65.75 %	81.20 %

Nous constatons que le choix de la distance de Manhattan  $F_{kmt}$ , comme fonction d'évaluation donne les meilleurs scores de classification (Fig. 5.7) et (Fig. 5.8).

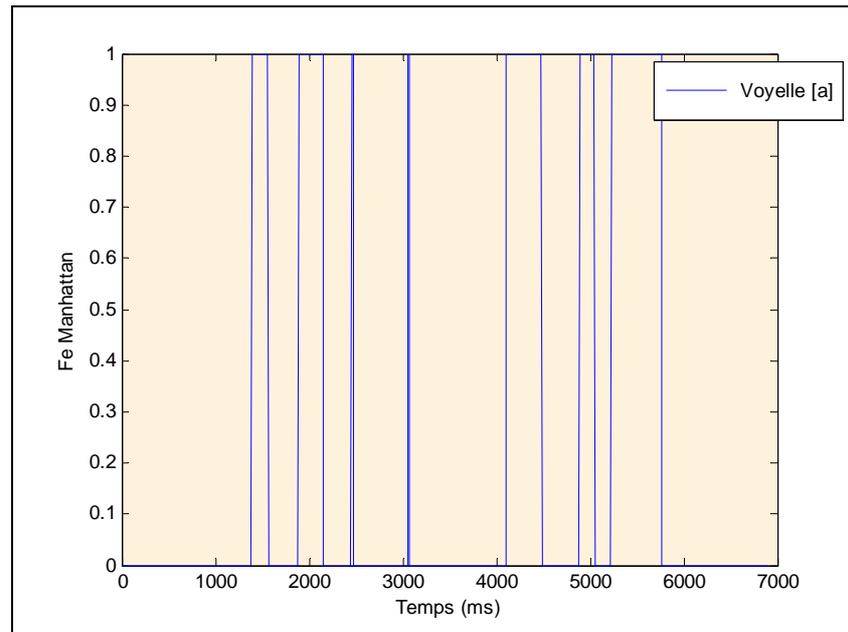


Figure 5.7 : Recherche de la voyelle [a] dans la phrase p1 par l'inverse de la fonction de distance de Manhattan

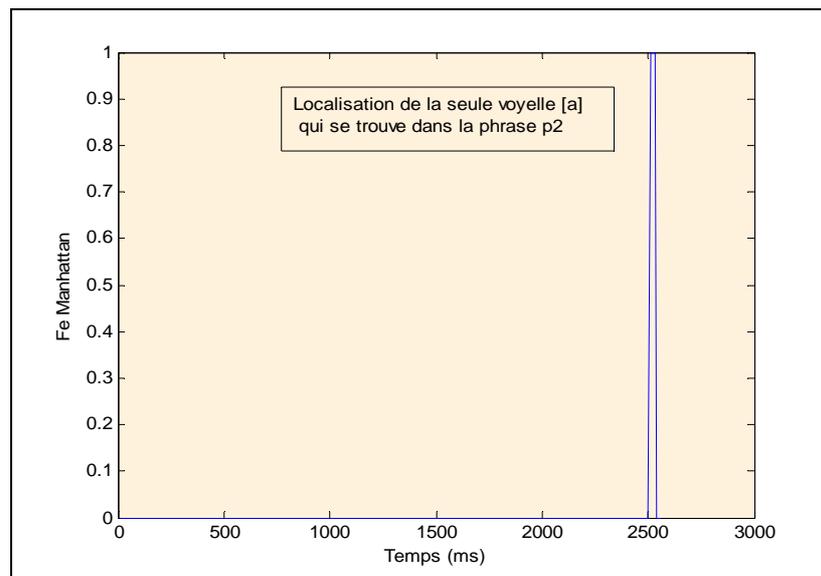


Figure 5.8 : Recherche de la voyelle [a] dans la phrase p2 par l'inverse de la fonction de distance de Manhattan (Annexe 2)

Nous constatons à travers ces deux graphes que le choix de la fonction de distance de Manhattan en tant que fonction objectif du MGS, nous a permis de localiser et de classer douze parmi les quatorze voyelles [a] qui se trouvent dans la phrase p1 ; et la seule voyelle [a] se trouvant dans la phrase p2. Cependant, la fonction de distance d'Euclide n'est pas aussi performante (Fig. 5.9).

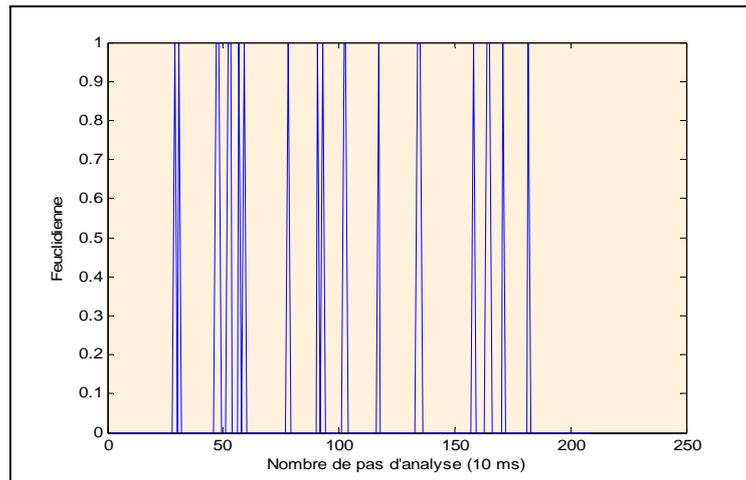


Figure 5.9 : Recherche de la voyelle [a] dans la phrase p3 par l'inverse de la fonction de distance Euclidienne (Annexe 2)

A travers cette figure, nous constatons que le nombre des voyelles [a] localisées et classées, en utilisant l'inverse de la fonction de distance de Manhattan, ne correspond pas au nombre réel existant dans la phrase p3.

### 5.3. Résultats et commentaires de la classification génétique

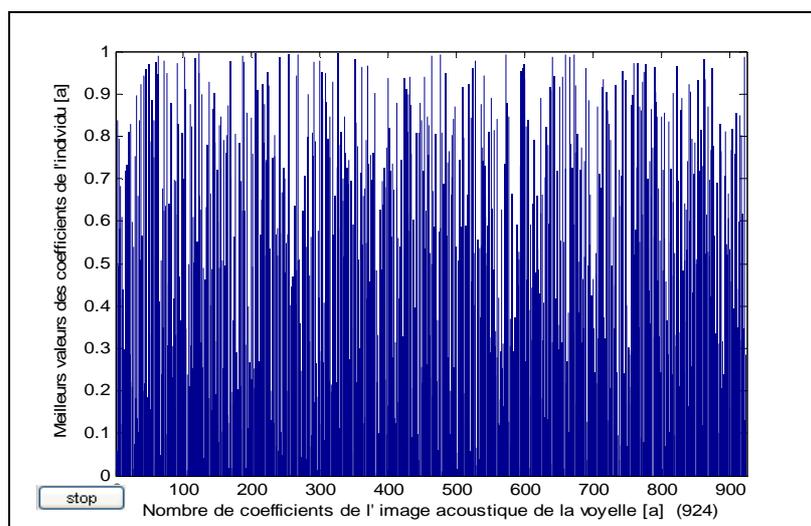
L'expérimentation a porté sur un corpus d'une centaine de phrases de la parole continue de l'Arabe Standard, enregistrées, dans un milieu ambiant, par un seul locuteur (un chercheur Jordanien). La segmentation acoustique du continuum vocal a été faite d'une manière automatique en balayant ce continuum de gauche à droite. Avec un pas de 10 ms et une fenêtre de segmentation de 20 ms, nous avons découpé le signal vocal du corpus à traiter en un ensemble de segments acoustiques de même taille auxquels nous avons appliqué l'analyse cepstrale. Ainsi, nous avons obtenu une représentation plus compacte de notre corpus vocal. L'ensemble de ces segments a été représenté par un autre ensemble équivalent constitué de vecteurs à court terme, images acoustiques de ces segments. Chacun d'eux est composé de 42 coefficients pertinents qui sont les treize premiers MFCCs, l'énergie et leurs delta et delta-delta coefficients.

En comparant les taux globaux de classification des phonèmes de l'AS obtenus en utilisant respectivement, les classificateurs à fonction de distance, le modèle génétique MGS et le modèle parallèle MGP ; nous avons remarqué que pour le même type de fonction de distance choisie qui est la distance de Manhattan, l'utilisation des AGs a rendu les classificateurs plus performants (Tab.5.4).

Tableau 5.4 : Comparaison des taux globaux de classification des  
différents modèles de classification

Type du modèle De classification	classificateur à fonction de distance	MGS	MGP
Taux global de Classification des Phonèmes %	75.25	80.13	81.20

Les individus qui vont composer la population de chaque MGS ne sont autres que les images codées des vecteurs acoustiques obtenus par concaténations successives de  $r$  vecteurs à court terme représentant les segments qui se suivent dans le continuum vocal. Le choix de la valeur de  $r$  dépend de l'ordre ou de la taille du vecteur image acoustique de référence du phonème à reconnaître ou à classer. Ainsi pour chaque phonème à classer, nous segmentons le signal vocal en un ensemble de vecteurs qui sont le résultat de la concaténation successive d'un certain nombre de segments qui dépend de l'ordre du vecteur image acoustique de référence du phonème à classer (Fig. 5.10)

Figure 5.10 : Meilleure image acoustique représentative de la voyelle [a]  
après 500 générations

Une fois codés, ces vecteurs vont constituer la population d'un MGS. Ainsi, pour chaque phonème à classer, nous obtenons des individus et une population spécifiques et différents. Pour notre modèle, nous avons utilisé respectivement le codage binaire et réel [20].

Nous avons opté pour le modèle génétique parallèle de type à gros grains ou en îlots sans migration pour lequel chaque sous population a une taille identique égale à 1000. Cette taille

est invariante au fil des générations pour tous les MGS. A chacune d'elles, est appliqué un AG séquentiel auquel nous avons choisi la technique (Fig. 5.11) :

- du croisement bi-points avec une probabilité  $p_c$  égale à 0.8 ;
- de la mutation uniforme avec une probabilité  $p_m$  égale à 0.1 ;
- de la sélection par tournoi et par la roulette biaisée avec une probabilité de sélection  $p_s$  égale à 0.75 ;
- du remplacement stationnaire avec une probabilité de remplacement  $p_r$  égale à 0.5.

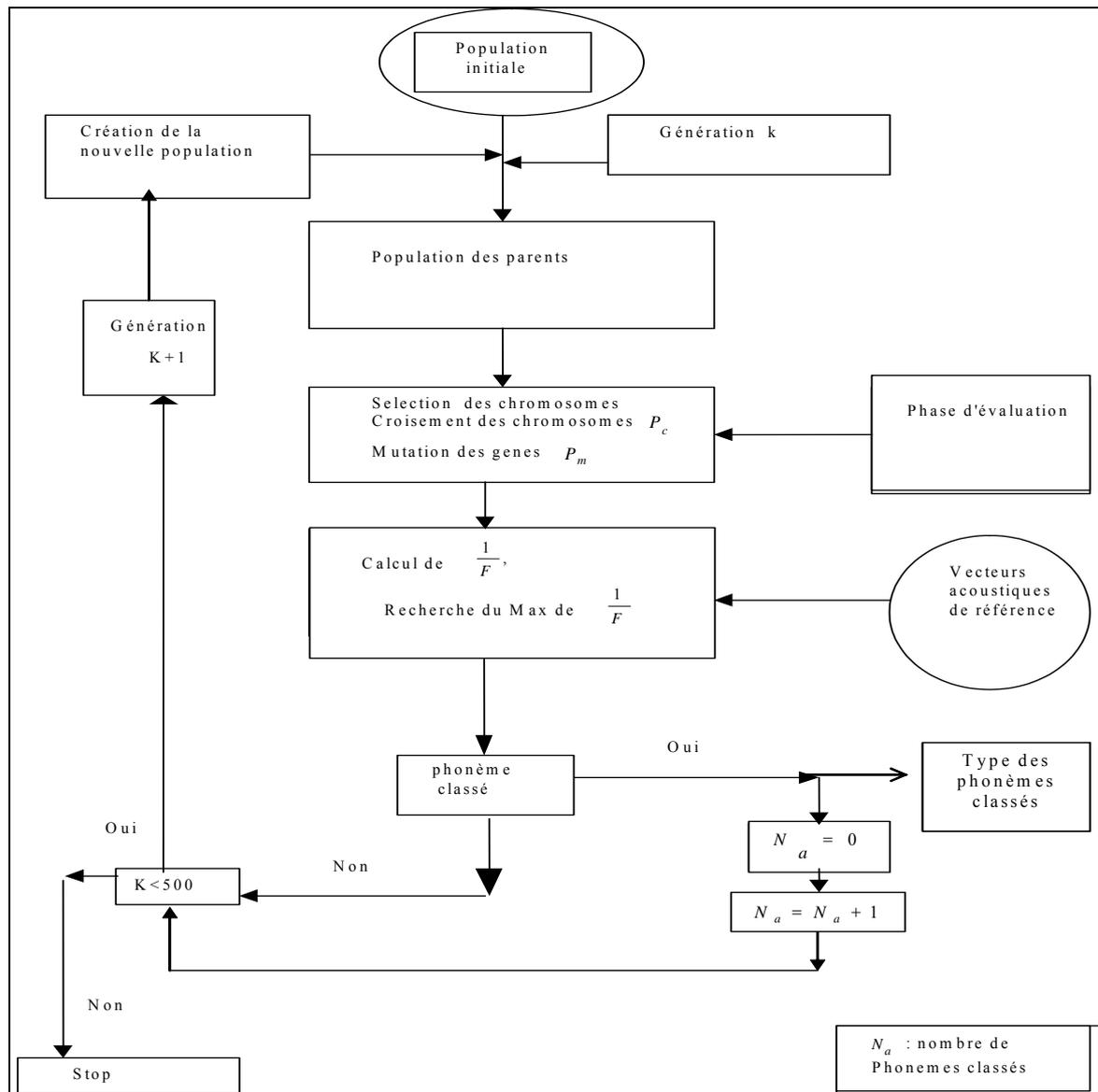


Figure 5.11 : Organigramme du modèle génétique séquentiel de classification acoustique

Les valeurs de ces paramètres sont choisies d'une manière empirique afin d'assurer le bon fonctionnement de l'AG. Quand ces paramètres sont mieux réglés, notre modèle génétique converge rapidement vers un optimum global (voir Annexe 3).

En outre, nous avons choisi la fonction de distance de Manhattan  $F_k mt$ , comme fonction d'évaluation. Dans notre cas et pour des raisons de convergence, le critère d'arrêt du MGS est choisi comme étant le nombre de générations maximum ou itérations qui a été fixé à 500 (Fig.5.12).

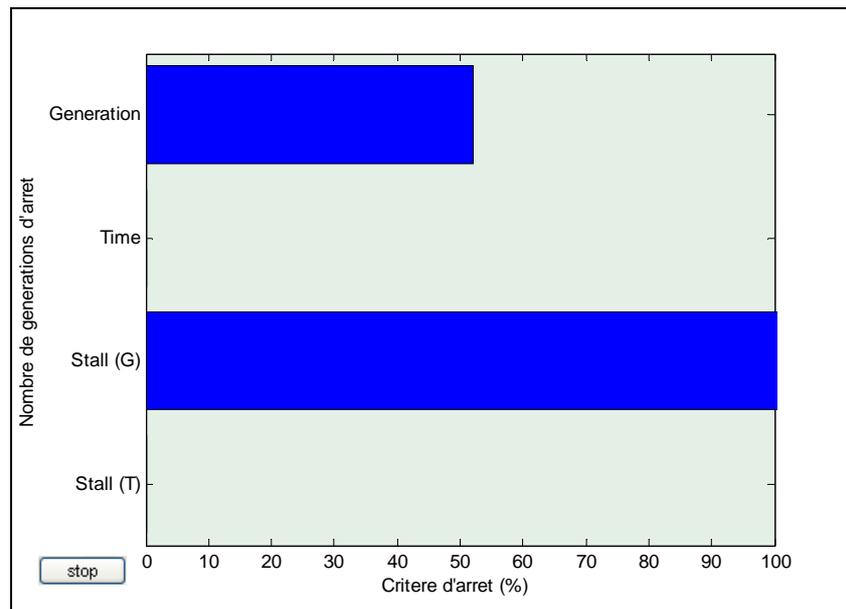


Figure 5.12 : Critère d'arrêt de l'AG selon le nombre maximum de générations

Les résultats obtenus (Fig.5.13) concernent le taux total de classification des :

- phonèmes noté  $T_c p$  égal à 81.20 % (Tab.5.5) ;
- voyelles noté  $T_c V$  égal à 83.50 (Tab.5.6) ;
- consonnes occlusives noté  $T_c Co$  égal à 84.34% (Tab.5.7) ;
- consonnes fricatives noté  $T_c Cf$  égal à 76.79% (Tab.5.8) ;
- des autres consonnes de l'AS noté  $T_c C$  égal à 80.674 % (Tab.5.9).

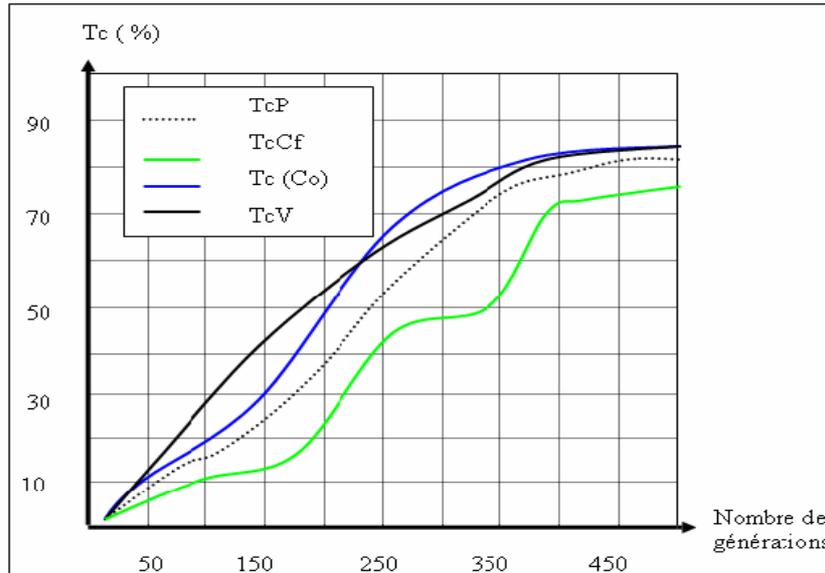


Figure 5.13 : Evolution du taux de classement des différentes catégories des sons de l'Arabe Standard au fil des généralisations

Nous remarquons que le taux moyen total de classification des consonnes fricatives est le plus bas et ceux des consonnes nasales et occlusives sont les plus élevés. Cela s'explique par le fait que les caractéristiques acoustiques des fricatives sont moins stables et sont plus affectées par la nature de l'entourage et le bruit environnant que les autres types de consonnes ou les voyelles (Fig.5.14) et (Fig.5.15).

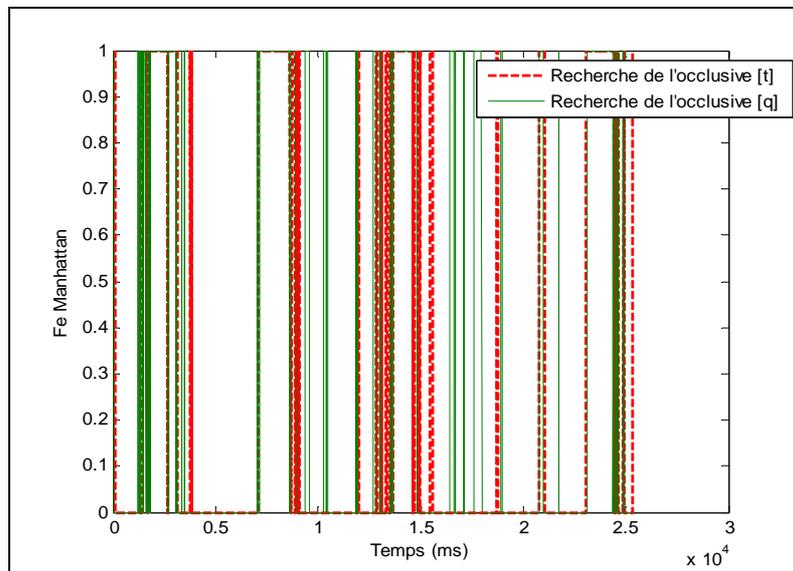


Figure 5.14 : Recherche des occlusives [t] et [q] du corpus1 par l'inverse de la fonction de distance de Manhattan (Annexe 2)

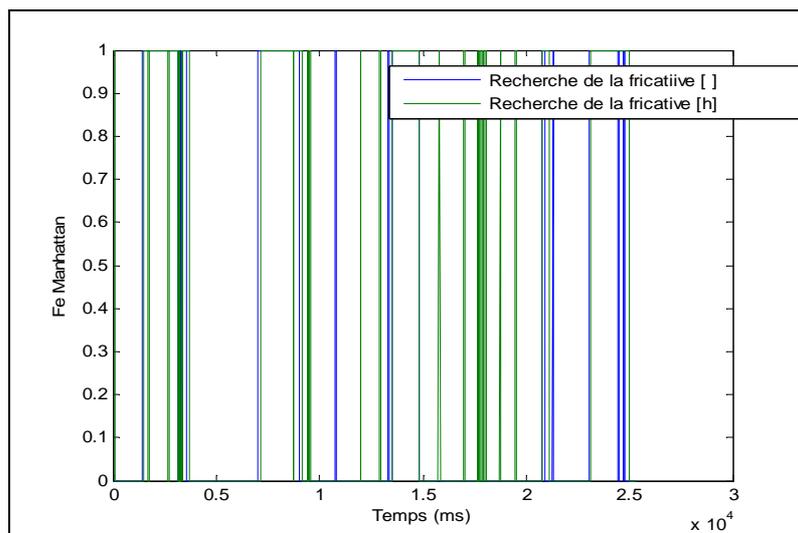


Figure 5.15 : Recherche des fricatives [ʃ] et [h] du corpus1 par l'inverse de la fonction de distance de Manhattan (Annexe 2)

Aussi, nous pouvons dire que notre modèle génétique de classification n'est pas très robuste au bruit environnant.

Tableau 5.5: Evolution du taux de classification des phonèmes au fil des générations

Nombre de Générations	50	100	150	200	250	300	350	400	450	500
Taux des phonèmes classés (%)	9.25	16.40	25.81	39.62	52.69	67.75	75.90	79.45	80.75	81.17

Nous remarquons que le taux moyen de classification des phonèmes de l'AS évolue au fil des générations pour atteindre le meilleur score après 500 générations. Cette évolution s'accélère à partir 200 générations.

Tableau 5.6 : Taux de classification des voyelles de l'Arabe Standard

Voyelles	[i]	[i:]	[u]	[u:]	[a]	[a:]
Taux de Classification (%)	83.25	85.00	82.00	84.25	83.75	82.25

Quant aux taux moyens de classification des voyelles de l'AS, nous constatons qu'ils diffèrent d'une voyelle à l'autre et ceux des voyelles longues sont légèrement plus importants que ceux des voyelles courtes (Tab.5.6) et (Fig.5.16) et (Fig.5.17).

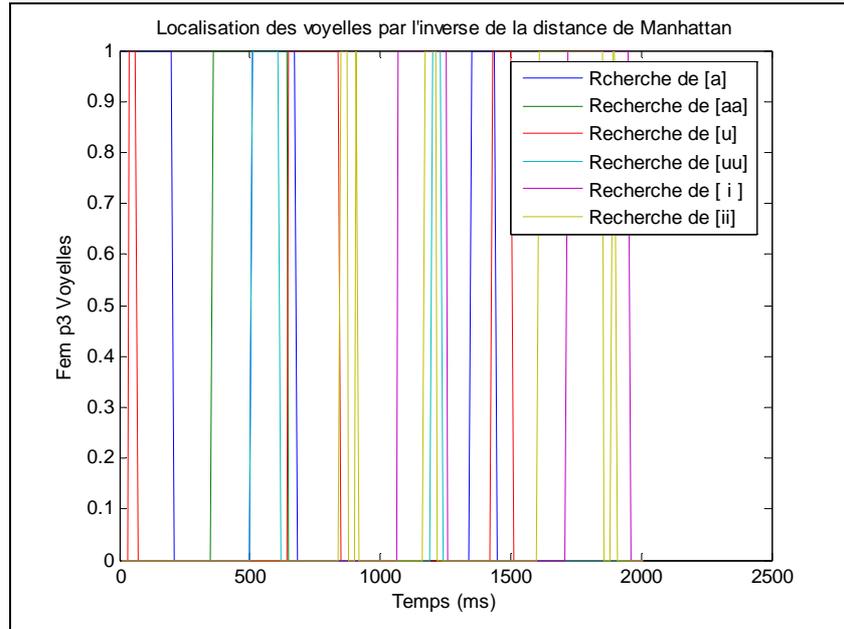


Figure 5.16 : Localisation des voyelles de l'AS dans la phrase p3 par l'inverse de la fonction de distance de Manhattan : fonction objectif du MGS

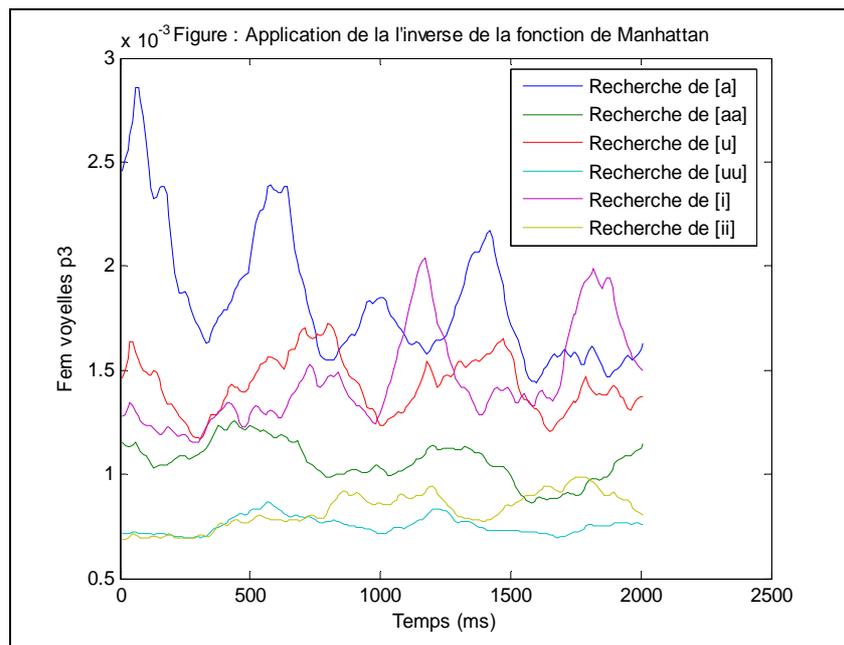


Figure 5.17 : Recherche des voyelles de l'AS dans la phrase p3 par l'application de l'inverse de la fonction de distance de Manhattan

Cela permet de conclure que des contraintes indépendantes à ce modèle génétique comme les effets de la coarticulation influent sur ces performances. Cela peut s'expliquer par l'importance de la partie stable des voyelles longues par rapport à leurs opposées les voyelles courtes.

Tableau 5.7 : Taux de classification des consonnes occlusives  
en contexte vocalique

Consonnes occlusives	Taux de Classification en contexte vocalique (%)						Taux moy.
	[a]	[a:]	[i]	[i:]	[u]	[u:]	
[k]	83.25	83.50	89.00	89.50	80.25	80.00	<b>84.25</b>
[q]	82.75	82.25	91.25	90.50	79.75	78.50	<b>84.08</b>
[t]	83.75	84.25	88.75	89.00	82.00	83.25	<b>85.16</b>
[t̤]	80.25	81.75	86.25	87.00	84.25	83.00	<b>83.75</b>
[d]	82.25	80.00	82.25	83.50	81.00	82.75	<b>81.95</b>
[ʔ]	88.25	90.75	88.00	89.75	87.00	87.00	<b>88.66</b>
[b]	84.25	86.75	85.00	84.25	84.00	83.50	<b>84.62</b>
[d̤]	83.25	80.00	82.75	81.25	83.00	83.25	<b>82.25</b>

Les taux moyens de classification des consonnes occlusives sont très intéressants comparés à ceux des autres types de consonnes. Cela signifie que les propriétés acoustiques des occlusives sont en adéquation avec la représentation paramétrique choisie.

Tableau 5.8 : Taux de classification des consonnes fricatives  
en contexte vocalique

Consonnes Fricatives	Taux de Classification en contexte vocalique (%)						Taux Moy.
	[a]	[a:]	[i]	[i:]	[u]	[u:]	
[f]	73.00	73.50	78.00	79.25	75.25	75.25	<b>79.70</b>
[θ]	72.75	72.50	78.25	78.50	74.75	75.50	<b>79.37</b>
[z]	82.00	82.25	81.50	82.00	81.75	80.25	<b>80.62</b>
[ɣ]	72.25	73.75	78.25	79.00	73.25	72.50	<b>72.66</b>
[ʒ]	81.25	82.00	81.25	81.50	80.00	79.25	<b>80.87</b>
[ʃ]	72.25	72.00	79.25	77.50	74.00	72.75	<b>79.54</b>
[x]	71.75	72.75	80.00	78.50	75.00	74.00	<b>71.30</b>
[s]	82.25	80.75	82.00	81.25	80.00	81.50	<b>80.95</b>
[ʂ]	83.25	80.00	82.75	85.25	83.00	83.25	<b>80.91</b>
[ʐ]	72.25	71.00	70.75	72.00	71.50	70.25	<b>78.29</b>
[ʕ]	83.50	84.75	85.50	87.75	86.00	84.25	<b>80.29</b>
[h]	75.50	76.00	73.25	72.75	74.75	75.25	<b>71.58</b>
[ħ]	74.50	75.75	74.00	73.25	75.00	75.50	<b>71.44</b>
[ε]	75.50	76.50	70.75	71.50	72.00	73.50	<b>71.62</b>

Nous constatons que le taux moyen de classification des consonnes fricatives diffère d'une consonne à l'autre. Si le taux moyen de la fricative [x] est de 71.62 %, celui de la fricative [s] est de 80.95 %. Cette différence est liée aux particularités de chacune d'elles. Les scores les plus faibles concernent les cinq fricatives arrières qui sont l'uvulaire [ɣ], la vélaire [x], les deux pharyngales ([ħ], [ʕ]) et la glottale [h].

Tableau 5.9 : Taux de classification des semi-voyelles et des consonnes liquides, nasales et vibrantes en contexte vocalique

Consonnes vibrantes liquide et nasales et semi-voyelles	Taux de Classification en contexte vocalique (%)						Taux Moy.
	[a]	[a :]	[i]	[i :]	[u]	[u:]	
[m]	85.25	84.75	89.75	89.00	86.25	87.50	<b>85.41</b>
[n]	85.75	85.25	90.00	90.75	87.50	86.75	<b>87.66</b>
[l]	82.25	82.75	83.00	84.25	81.00	81.25	<b>82.41</b>
[r]	84.25	85.00	84.25	83.75	84.00	82.75	<b>84.00</b>
[w]	83.75	82.00	84.25	83.50	83.50	83.25	<b>83.37</b>
[y]	82.25	82.75	81.75	81.50	83.25	82.75	<b>82.29</b>

A partir de ces résultats, nous remarquons que :

- les résultats obtenus sont satisfaisants puisque le taux de classification global des phonèmes de l'AS du corpus peut atteindre dans les meilleures conditions d'ajustement des paramètres de l'AG le seuil des 81% ;
- le taux de classification dépend de deux types de facteurs essentiels : le premier a trait aux caractéristiques acoustiques des unités vocales à classer et de leur environnement phonétique; le second dépend du choix et de la manière avec laquelle les paramètres de l'AG sont ajustés ;
- le taux moyen total de classification des voyelles et semi voyelles est plus important relativement à ceux de plusieurs types de consonnes. En effet, les caractéristiques des voyelles sont relativement plus proéminentes et plus stables que celles des consonnes. L'existence des variantes [ɑ, o, e] dans ces réalisations acoustiques des voyelles orales de l'AS explique le taux des voyelles non classées avoisinant les 17.50% ;

- le milieu vocalique adjacent aux consonnes influe différemment sur leur taux de classification selon la nature de ce milieu. Presque toutes les consonnes voient leur taux de classification amplifiés légèrement en présence des voyelles [i] et [i :]. Les voyelles constituent le point d'ancrage privilégié des modèles de coarticulation. Les gestes consonantiques rapides se superposent au geste vocalique lent ;
- l'utilisation de différentes fonctions de distance en tant que fonctions d'évaluation de l'AG a donné ses fruits puisque, nous avons obtenu des taux de classification très satisfaisants en comparaison à d'autres méthodes de classifications connues [87].

#### 5.4. Conclusion

L'application des Algorithmes Génétiques pour la RAP, en tant que modèle génétique de classification des vecteurs acoustiques des séquences vocales, est très récente à travers le monde. Dans beaucoup de cas, elle est en stade d'expérimentation. Le recours aux AGs pour la RAP ne s'est généralisé que pour optimiser des modèles déjà existants. Pour cela, Nous les avons appliqués pour classer les phonèmes de l'Arabe Standard.

A partir de cette étude, nous avons présenté une méthode de classification acoustique des séquences vocales, basée sur les AGs et appliquée aux phonèmes de l'Arabe Standard à des fins de la RAP. Ainsi, nous avons choisi le modèle génétique parallèle à gros grains sans migration afin d'optimiser l'efficacité de notre AG. Pour chaque sous-population de même taille, nous avons appliqué un AG séquentiel de mêmes caractéristiques MGS et dont les paramètres de réglages ont été choisis judicieusement. Le choix des fonctions de distance en tant que fonction d'évaluation et le codage réel utilisé ont eu un impact plus que positif sur le taux moyen global de classification des phonèmes qui a atteint le score de 81.20%.

Cependant, des facteurs indépendants au modèle génétique comme la coarticulation ou le contexte vocalique ainsi que le bruit environnant ont influé négativement sur le taux de classification de certaines catégories de sons de l'AS.

Aussi, nous nous sommes aperçus des difficultés pour choisir pertinemment les bons paramètres pour les divers opérateurs (mutation, croisement, sélection et remplacement). Nous avons remarqué que le taux de rejet est contrôlé indirectement par le choix de ce type de paramètres.

## Perspectives et Conclusion Générale

Ce travail de thèse s'inscrit dans le cadre du Décodage Acoustico-Phonétique pour reconnaître les phonèmes de l'AS en parole naturelle. Son principal but est de réaliser un modèle robuste de DAP qui permet de traiter un corpus de grandes dimensions, enregistré dans un milieu ambiant en minimisant les effets néfastes du phénomène de la coarticulation ou variabilité contextuelle. Il se positionne comme une contribution au développement des AGs pour des problèmes de classification automatique des données et leur application à la RAP. Notre contribution se situe dans la modélisation du problème en choisissant comme fonction objectif de notre modèle génétique, le critère à base de fonctions de distance.

Nous avons utilisé les AGs pour affecter des objets qui sont des segments acoustiques à des classes bien définies, en utilisant un modèle génétique de classification supervisée mono-objectif.

Faisant partie des metaheuristiques évolutives, les AGs ont été utilisés, avec succès, dans plusieurs domaines dont l'Intelligence Artificielle, l'optimisation combinatoire mono et multi-objectifs (à cause de la nature multi-critères des problèmes réels) et pour résoudre les problèmes de l'ordonnement, de la sélection et du partitionnement [88, 89].

Aussi, ils ont été utilisés dans la Classification Automatique ou Clustering en Anglais et plus généralement dans l'Analyse Combinatoire et Statistique des données complexes ou Combinatorial Data Analysis en Anglais et pour l'extraction de connaissances et leur gestion connues sous le terme de Fouille de Données ou Data Mining en Anglais [90].

Les plus récents domaines de leurs applications sont la Bioinformatique, l'Imagerie Numérique et le Traitement Automatique des Langues Naturelles.

Plusieurs travaux de recherches réalisés qui ont traité les problèmes de la segmentation, de la classification et de la reconnaissance automatique des données en utilisant les AGs, ont donnés de très bon résultats [91, 92].

Actuellement, on a recours à des algorithmes dits hybrides pour leurs performances. Il existe plusieurs types d'hybridations possibles :

- nous pouvons utiliser une Recherche Locale (Recuit Simulé et Recherche Tabou) dans les méthodes évolutives ou AGs. L'avantage est que la Recherche Locale réduit le

danger de passer à côté d'une solution optimale sans la voir. En règle générale, une méthode évolutive est excellente pour détecter de bonnes régions dans l'espace de recherche alors qu'une Recherche Locale explore efficacement les régions prometteuses [93].

- nous pouvons exécuter en parallèle diverses métaheuristiques, voire même plusieurs à la fois la même métaheuristique mais avec divers paramètres. Ces processus parallèles communiquent entre eux régulièrement pour échanger de l'information sur leurs résultats partiels. La co-évolution de plusieurs agents de recherche permet entre autre une meilleure diversité et aussi une bonne exploration de l'espace de recherche. [94, 95].
- une troisième forme d'hybridation consiste à combiner les AGs avec des méthodes exactes ou déterministes (Réseaux de Neurones, MMC , etc...). Une métaheuristique peut par exemple fournir des bornes à une méthode de type branch-and-bound. Au contraire, une méthode exacte peut donner lieu à une technique efficace pour la détermination du meilleur voisin d'une solution (ce qui peut s'avérer plus judicieux que de choisir la meilleure solution parmi un échantillon de voisins) [96, 97, 98].

D'autres metaheurtistiques à base de population ont été utilisés avec succès pour résoudre les problèmes de la classification des données; parmi lesquelles, nous avons la programmation Génétique et le Système Immunitaire [99,100] .

Il résulte que l'utilisation des metaheuristiques évolutives basées sur l'hybridation et le parallélisme est un axe de recherche privilégié car très prometteur pour les domaines de la classification des données et de la Reconnaissance Automatique.

Ainsi et pour accroître les performances des AGs dans le domaine de la RAP, nous proposons de tester d'autres types de topologies du parallélisme et la mise en place de schémas d'hybridation avec d'autres algorithmes déterministes ou stochastiques ayant prouvé leur efficacité dans le domaine général de la RAP. Ces algorithmes peuvent être de nature multi-objectifs pour une représentation multi-critères du signal vocal.

Pour de meilleurs résultats, la constitution d'une base de donnée fiable et normalisée des sons de l'Arabe est nécessaire. Aussi, le choix univoque des divers paramètres sous les AGs de classification reste un problème ouvert. C'est une des pistes qu'il convient maintenant explorée.

## Références Bibliographiques

- [1] Calliope, "la Parole et son Traitement Automatique", Edition Masson, 717 pages. 1989.
- [2] R. Boitier & M. Kunt, "Traitement de la parole". Presse Polytechniques, Romandes. Lausanne, 1987.
- [3] J. Haton, J. Caelen, & al, "La Reconnaissance Automatique de la Parole", Dunod, Paris, France, 1991.
- [4] J.F. Bonnot " Etude expérimentale de certains aspects; de la gémination et de l'emphase en Arabe" Actes du 9<sup>ème</sup> ICPS Copenhague, Vol.1, 1979.
- [5] S. h. Al Ani, " Arabic Phonology an acoustical and physiological investigation", Edition Mouton, 1970.
- [6] A. Hadj-Salah, "La notion de syllabe et la théorie cinético-impulsionnelle des phonéticiens phonéticiens arabes". Al-Lissaniyat. Revue Algérienne de Linguistique.Vol.1, pp.: 63-83, I.L.P Alger , Algerie, 1971.
- [7] A. Haidar, "Variabilité et Invariance du système vocalique de l'Arabe Standard", PhD thesis, Université de Franche-Comté. 1991.
- [8] A.H..Salah, "Applied Arabic Linguistic and Signal Information Processing"; pp. 3-22, Alger.1987.
- [9] P. Ladefoged, "A Course in Phonetics". Fort Worth TX., Harcourt Brace College Publishers. 1993.
- [10] P. Delattre, "Des indices acoustiques aux traits pertinents", Proceeding of the 6<sup>th</sup> International Congress of Phonetic Sciences, Prague, pp. 46, 1970.
- [11] B. David Pisoni, "Auditory Short-Term Memory and Vowel Perception", Memory and Cognition, Vol 3, pp.7-18, 1975.
- [12] M. Debyeche "Reconnaissance automatique des glottales et pharyngales de l'Arabe Standard en parole continue multilocuteurs". Thèse de magistère. USTHB. Algérie. 1991.
- [13] J. Cantineau, " Cours de Phonétique Arabe", Klincksiek, Paris, 1960.
- [14] L.R. Rabiner & R.W. Shafer, "Digital Processing of Speech Signal"», Printice Hall, Ing, Englewoodchiffes, New Jersey, 1978.
- [15] J.F. Bonnot, "Recherche expérimentale sur la nature des consonnes emphatiques de l'Arabe classique", Institut de phonétique de Strasbourg, France, Vol.9, pp. 47-88. 1979.
- [16] B.S. Atal & S.L. Hanauer, "Speech analysis and synthesis by linear prediction of speech wave", J. Acous. Soc. Am. VOL. 50(2), pp. 637-655. 1971.
- [17] J. Makhoul, "Linear Prediction : A tutorial Review".Proc.IEEE 63, 4, pp. 561-580. 1975.

- [18] H. Hermansky, "Perceptual Linear Predictive (PLP) Analysis of Speech", *J. Acous. Soc. Am.* 87(4); pp. 1738-1752, 1990.
- [19] J.C. Junqua, "Evaluation of ASR Front-ends in Speaker-dependent and Speaker-independent Recognition", *J. Acous. Soc. Am.*, 81 S1:S93, 1987.
- [20] J. Rosca, & A. Kofmehl, "Cepstrum-like ICA Representations for Text Independent Speaker Recognition", *ICA2003*, pp. 999-1004, 2003.
- [21] Hesham Tolba, Sid-Ahmed Selouani and Douglas O'Shaughnessy, "Auditory-based Acoustic Distinctive Features and Spectral Cues for Automatic Speech Recognition Using a Multi-Stream Paradigm. *IEEE-ICASSP'2002*: 837-840.2002.
- [22] J. Antoni & al, "Blind system identification of non-minimum phase systems using the mean differential cepstrum". *European signal processing conference, Tampere, 2000.*
- [23] J-P. Haton, "Reconnaissance de la parole", chapitre méthodes Robustes, Mariani, 2002.
- [24] S. Molau & al, "Computing Mel-Frequency Cepstral Coefficients of the power cepstrum", *ICASSP*, 2001.
- [25] J. Caelen & G. Caelen-Haumont, "Indices et propriétés dans le projet ARIAL II"; *Processus d'Encodage et de Décodage Phonétique, GALF*, pp. 128-139, Toulouse, France, 1981.
- [26] P. Kevin, "The Listening Telephone - Automating Speech Recognition over the PSTN", *BT Technology Journal*. Vol.14, pp. 112-126, janvier, 1996.
- [27] L. R. Rabiner & B. H. Juang., "Fundamentals of Speech Recognition", Englewood Cliffs, New Jersey, Prentice Hall, 1993.
- [28] J.C. Junqua, "Utilisation d'un modèle d'audition et de connaissances phonétiques en Reconnaissance Automatique de la Parole", *Traitement du signal*, Vol. 7 , pp. 275-284, 1990.
- [29] J. Mariani, " Reconnaissance Automatique de la Parole, progrès et tendances, *Traitement du signal*", Vol 7, pp. 239-266, 1990.
- [30] C. Myers & L. R. Rabiner, "A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition", *IEEE Trans. Acoust., Speech and Signal Processing*, ASSP-29 (2), pp. 284, 1981.
- [31] F. Lonchamp, "Reading Spectrograms : the View of the Expert, in *Fundamentals in Computer Understanding : Speech and Vision*", J. P. Haton, editor, Cambridge University Press, 1987.
- [32] C. Fredouille, "Approche statistique pour la Reconnaissance Automatique du Locuteur : Informations dynamiques et normalisation Bayésienne des vraisemblances", *Thèse de Doctorat, Université d'Avignon et des Pays de Vaucluse*, 2000.

- [33] L. Lazli & M.T. Laskri, "Nouvelle méthode d'entraînement des systèmes hybrides HMM/ANN à base d'une segmentation floue. Application pour la Reconnaissance Automatique de la Parole", CARI 2004, 7<sup>ème</sup> 7ème Colloque Africain sur la Recherche en Informatique, pp. 331-338, Novembre 22-25, Hammamet, Tunisie, 2004.
- [34] V. K. Prasad & all, "Automatic Segmentation of Continuous Speech Using Minimum Phase Group Delay Functions", Speech Communication., vol. 42, no(3-4), pp. 429-446, 2004.
- [35] C. Scagliola, "Continuous Speech Recognition without Segmentation: Two Ways using Diphones as basic Speech Units", Speech Commun., Vol. 2, 2-3, pp. 199-201, 1983.
- [36] R.O. Duda & al, "Pattern Classification", 2nd Ed., Wiley-Interscience, 2001.
- [37] L. Lazli & M. Selami, "Proposition d'une Architecture d'un Système Hybride MMC - PMC pour la Reconnaissance de la Parole Arabe". MCSEAI 2002, the 7<sup>ème</sup> Magrebian Conference on Computer Sciences, Vol. 1, pp. 101-109, 6-8 Mai, Annaba, Algérie, 2002.
- [38] S. R. Levinson & al, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition", The Bell System Technical Journal, Vol. 62, 4, pp. 1035-1074, 1983.
- [39] S. Haykin, "Neural Networks A Comprehensive Foundation", Macmillan College Publishing Company, 1994.
- [40] L. Lazli & M. Sellami, "Système Neuro-symbolique pour la Reconnaissance de la parole Arabe", CGE'02 : Conférence Nationale sur le Génie Electrique, p. 64 (résumé), EMP : Ecole Militaire Polytechnique", 17-18 décembre, Alger, Algérie, 2002.
- [41] L. Lazli & M. Sellami. "Speaker independent isolated speech recognition for Arabic language using hybrid HMM-MLP-FCM system", AICCSA 2003 (ACS/IEEE): International Conference on Applications & Computer Systems, p. 108, Tunis, Tunisie, 2003.
- [42] L. Lazli et M. T. Laskri, " Nouvelle méthode de fusion de données pour l'apprentissage des systèmes hybrides MMC/RNA : Application pour la reconnaissance Automatique de la parole". Laboratoire de Recherche en Informatique (LRI). Groupe de Recherche en Intelligence Artificielle (GRIA). Département d'Informatique. Faculté des Sciences de l'Ingénieur. Université Badji Mokhtar d'Annaba. Algerie. 2005.
- [43] V. W. Zue, "The use of phonetic rules in automatic speech recognition", Speech Com., Vol. 2, pp.181-186, 1983.
- [44] N. Monmarché & al, "Fouille Visuelle et classification de données par nuage d'insectes volants, RSTI-RIA-ECA : Méthodes d'optimisation pour l'extraction de connaissances et l'apprentissage", N° 6, pp. 729-752 , 2002.

- [45] A.E. Eiben, & J.E. Smith, "Introduction to Evolutionary Computing", Springer, Berlin. 2003.
- [46] S. Gustafson & all, "Problem difficulty and code growth in genetic programming". Genetic Programming and Evolvable Hardware, 5(3):271-290, 2004.
- [47] J. Hansohm, "Two-mode Clustering with Genetic Algorithms, Classification, Automation, and New Media", Proceedings of the 24<sup>th</sup> Annual Conference of the Gesellschaft Fur Klassifikation E. Vol.1, pp. 87-94, 2000.
- [48] S Bandyopadhyay & C. A Murthy, "Pattern Classification Using Genetic Algorithms", Pattern Recognition Letters, Vol. 16, pp. 801—808. 1995.
- [49] W. Greene, "Unsupervised Hierarchical Clustering via a Genetic Algorithm", Press I., Ed., Proceedings of the 03 Congress on Evolutionary Computation, Canberra, Australia, pp. 998-1005, 2003.
- [50] A. Spalanzani & H. Kabré, "Perception Artificielle de la Parole par l'Utilisation des Algorithmes Génétiques", Journées Internationales d'Orsay sur les Sciences Cognitives, Orsay, France, 1997.
- [51] G. Lee & V. Bulitko. "GAMM: genetic algorithms with meta-models for vision", in Proceedings of the Genetic and Evolutionary Computation Conference (GECCO), pages 2029–2036, Washington, DC, 2005.
- [52] K. Krishna & M. Murty, "Genetic K-means algorithm", IEEE Transactions on Systems, Man and Cybernetics - Part B, Vol. 29, N° 3, pp. 433-439, 1999.
- [53] A. Spalanzani & Sélouani, "Improving Robustness of Connectionist Speech Recognition Systems by Genetic Algorithms", International Conference on Information, Intelligence and Systems, Washington, 1999.
- [54] M. Debyeche & all, "Phoneme recognition system based on HMM with distributed VQ codebook", In EUROSPEECH'99, 1315-1318. 1999.
- [55] P. Kuntz., "A Colony of Ant-like Agents for Partitioning in VLSI Technology", Husband P., Harvey I., Eds., Proceedings of the 4th European Conference on Artificial Life, pp. 417–424, 1997.
- [56] N. Labroche & al, "A new clustering algorithm based on the chemical recognition system of ants", Harmelen F., Ed., Proceedings of the 15<sup>th</sup> European Conference on Artificial Intelligence, Lyon, France, IOS Press, pp. 345–349, July, 2002.
- [57] H. Azzag & al, "A New Model for Clustering with Artificial Ants", IEEE Congress on Evolutionary Computation, Canberra, Australia, 08-12 December, 2003.

- [58] L. De Castro & F. Von Zuben, "An Evolutionary Immune Network for Data Clustering", In Proceedings of the IEEE SBRN'00 (Brazilian Symposium on Artificial Neural Networks), pp. 84–89, 2000.
- [59] T. Knight & J. Timmis, "On data clustering with artificial ants", Proceedings of the 4<sup>th</sup> International Conference on Recent Advances in Soft Computing, Nottingham, UK, pp. 266-271, December, 2002.
- [60] C. Jacob, "Stochastic Search Methods". In: Berthold, M. and Hand, D. J. (eds.) Intelligent Data Analysis, Springer, Berlin. 2003.
- [61] D. E Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", New York: Addison-Wesley, 1989.
- [62] Z. Sen & A. Oztopal, "Genetic algorithms for the classification and prediction of precipitation occurrence", Hydrological Sciences, 46(2), pp. 255-268, 2001.
- [63] M. D.Vose, "The Simple Genetic Algorithm: Foundations and Theory". MIT Press, Cambridge, MA.1999.
- [64] Z. Michalewicz & C.Z. Janikov, "Handling constraints in Genetic Algorithms". In Proceedings of the 4<sup>th</sup> International Conference on Genetic Algorithm. ICGA, 1991.
- [65] D.E. Goldberg & al, "Genetic algorithms, noise, and the sizing of population. Complex Systems", Vol.6, pp. 333-362, 1992.
- [66] J.H. Holland, "Adaptation in Natural and Artificial Systems", University of Michigan press, 1975.
- [67] D.E. Goldberg, "Algorithmes génétiques, exploration, optimisation et apprentissage automatique", Paris : Addison-Wesley, BU : 006.3 GOL, 1994.
- [68] A.H. Wright, "Genetic algorithms for real parameter optimization". In Proceeding of the Foundation of Genetic Algorithms". Foga, Itali,1991.
- [69] Z. Michalewicz & al, "A Modified Genetic Algorithm for Optimal Control problems", Computers and Mathematics with Applications. Vol. 23, pp. 83-94. 1992.
- [70] J.H. Holland, "Genetic algorithms", Sci.Amer., July, pp. 66-72, 1992.
- [71] Z. Michalewicz, & al, "Evolutionary operators for continuous convex parameters spaces". Proceedings of the 3<sup>rd</sup> Annual Conference on Evolutionary Programming, World Scientific Publishing, pp. 84-107, 1994.
- [72] C.L. Bridges, D.E. Goldberg, "An Analysis of Multipoint Crossover", In Proceedings of the Foundation Of Genetic Algorithms. Foga, 1991.

- [73] T. P. Hong & al, "Simultaneously applying multiple mutation operators in genetic algorithms". *Journal of Heuristics*, Vol. 6, pp. 439- 455, 2000.
- [74] R.L. Haupt & S.E. Haupt, "Optimum population size and mutation rate for a Simple Real Genetic Algorithm that optimizes array factors". *Applied Computational Electromagnetics. Society Journal*, Vol. 15, N°. 2, 2000.
- [75] K. A. De Jong & al, "Using genetic algorithms for concept of Machine Learning", Vol. 13 pp.161–188, 1993.
- [76] D. Delahaye, "optimisation de la sectorisation de l'espace aérien par Algorithmes Génétiques", Thèse ENSAE, 1995.
- [77] D. Whitley, "Foundations of Genetic Algorithms and Classifier Systems 2", Vail, Colorado. San Mateo, CA: Morgan Kaufmann Publishers. Inc, USA, 1992.
- [78] A. Guenoche, "Partitions optimisées selon différents critères : évaluation et comparaison", Technical report, Institut de Mathématique de Luminy, Marseille, 2003.
- [79] M.Aissiou & M.Guerti, "Application des Algorithmes Génétiques en vue de la Reconnaissance des voyelles de l'Arabe Standard", A.J.O.T., International Publication, Série A, Vol. 16, N°1, pp. : 103-120, 2004. ISSN : 1111-357X.
- [80] K. A. De Jong & W. M. Spears, "Learning concept classification rules using Genetic Algorithms", *International Joint Conference on Artificial Intelligence*, pp. 651-656, 1991.
- [81] M.Aissiou & M.Guerti, "Genetic Algorithms Application for the Automatic Recognition of the Arabic stop sounds", *Applied Science Research Journal*, pp. 358-366, issue Mai 2007. Print ISSN: 1816-157X. Online ISSN: 1819-544X.
- [82] J.R. Koza, "Genetic Programming II", the MIT Press, 1994.
- [83] D. E. Goldberg, "Design of Innovation: Lessons From and For Competent Genetic Algorithms", Kluwer, Boston, MA. 2002.
- [84] E. Cantu-Paz, "A survey of Parallel Genetic Algorithms. *Calculateurs Parallèles. Réseaux et Systèmes Repartis*", Vol. 10(2). pp. 141-171, 1998.
- [85] E. Cantu-Paz & D. E. Goldberg, "Predicting speed ups of idealized bounding cases of Parallel Genetic Algorithms", In T. Bäck. Editor. *Proceedings of the 7<sup>th</sup> International Conference on Genetic Algorithms*, pp. 113-121. San Francisco. Morgan Kaufmann, 1997.
- [86] <http://www.phon.ucl.ac.uk/resource/sfs/SFS@phon.ucl.ac.uk>.
- [87] M.Aissiou & M.Guerti, "Genetic Algorithms Application to the Standards Arabic phonemes classification", *Cybernetics and Systems*. Vol. 39. Issue 3, pp. 199-212, 2008. Print ISSN : 0196-9722. Online ISSN : 1087-6553.

- [88] L.Rokach, "Genetic algorithm-based feature set partitioning for classification problems", *PR*(41), No. 5, pp. 1693-1717, May 2008.
- [89] A. Blansch e & S. Iwata, "Discovering new orders of the chemical elements through genetic algorithms". In *Proceedings of the twelfth Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD'08*, Vol. 5012 of *LNAI*, pp. 849–857, 2008.
- [90] S. Bandyopadhyay & al, "Multiobjective Genetic Clustering for Pixel Classification in Remote Sensing Imagery", *GeoRS*(45), N . 5, pp. 1506-1511, May 2007.
- [91] A. Blansch e & P. Ga carski, "Algorithme g n tique de pond ration d'attributs pour une classification non supervis e d'objets complexes". *RNTI*, 2005.
- [92] E.G. Talbi & al, "A parallel peer to peer branch and bound algorithm for computational Grids", *Proc. of the Intl. Workshop on Peer to Peer, Parallel, Grid and Internet Computing (3PGIC)*, Vienna, Austria, Apr 2007.
- [93] J. Garcia-Nieto & al, "A comparison of PSO and GA approaches for gene selection and classification of microarray data", *GECCO '2007 Genetic and Evolutionary Computations Conference*, London, UK, July 2007.
- [94] P. Ga carski & A. Blansch e, "Approches Darwienienne, Lamarkienne et Baldwinienne pour la pond ration automatique d'attributs en classification non supervis e", *Revue I3 (HS 2007)*, pp. 207–230, 2007.
- [95] P. Ga carski & al, "Comparison between two co-evolutionary feature weighting algorithms in clustering. *Pattern Recognition*", 41(3): pp.983–994, 2008.
- [96] E. Alba & al, "Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms", *CEC'2007 Congress on Evolutionary Computation*, Singapore, Sept 2007.
- [97] M.J. Zhang & al, "Gaussian distribution to construct fitness functions in genetic programming for multiclass object classification", *PRL*(27), N . 11, pp. 1266-1274, August 2006.
- [98] M. Basseur & al, "Towards parallel design of hybrids between Metaheuristics and exact methods". In *Parallel combinatorial Optimization*, chapter 7, pp. 163-186, ISBN: 0-471-72101-8, November 2006.
- [99] J.Y. Lin & al, "Designing a classifier by a layered multi-population genetic programming approach", *PR*(40), N . 8, pp. 2211-2225, August 2007.
- [100] K. Leung & al, "Generating Compact Classifier Systems Using a Simple Artificial Immune System", *SMC-B*(37), N . 5, pp. 1344-1356, October 2007.

% SCRIPT en Matlab pour la classification des phonèmes de l'AS du corpus p1 en utilisant la distance de Manhattan ;

```

function yyy= my_fun(i,j)
global p1, d,a,aa,i,ii,u,uu,k,m,l,f,d,n,y,w,q,r,h,cha,s,kha,t,tha,gha,dta,w,hamza,z,x
[na,naa]=size(a) ; [naa,naaa]=size(aa); [nu, nuu]=size (u); [nuu, nuuu]=size (uu);
[nw,nww]=size(w); [ni, nii]=size (i); [nii, niia]=size (ii); [ny, nyy]= size(y);
[nk, nkk]=size (k); [nm, nmm]=size (m); [nl, nll]=size (l); [nd, ndd]=size (d);
[nn, nnn]=size (n); [nq, nqq]=size (q); [nr, nrr]=size (r); [nh, nhh]=size (h);
[ncha, nchacha]=size (cha); [ns, nss]=size (s); [nkha, nkhakha]=size (kha);
[nt, ntt]=size (t); [ntha, nthatha]=size (tha); [ngha, nghagha]=size (gha);
[ndta, ndtadta]=size (dta); [nhamza, nhamzahamza]=size (hamza);
[nz, nzz]=size (z); [nx, nxx]=size (x); [ncorpus, nccorpus]=size(p1);
for count=1:ncorpus-50;
    corpus11a=p1(count:count+na-1,:);
    AAAa(count)=1/(sum(sum(abs(p1(count:count+na-1,:)-a(1:na,:)))));
    Fea=AAAa';
    corpus11aa=p1(count:count+naa-1,:);
    AAAaa(count)=1/(sum(sum(abs(p1(count:count+naa-1,:)-aa(1:naa,:)))));
    Feaa=AAAaa';
    corpus11u=p1(count:count+nu-1,:);
    AAAu(count) =1/(sum(sum(abs(p1(count:count+nu-1,:)-u(1:nu,:)))));
    Feu=AAAU';
    corpus11uu=p1(count:count+nuu-1,:);
    AAAuu(count) =1/(sum(sum(abs(p1(count:count+nuu-1,:)-uu(1:nuu,:)))));
    Feuu=AAAuu';
    corpus11i=p1(count:count+ni-1,:);
    AAAi(count) =1/(sum(sum(abs(p1(count:count+ni-1,:)-i(1:ni,:)))));
    Fei=AAAI';
    corpus11ii=p1(count:count+nii-1,:);
    AAAii(count) =1/(sum(sum(abs(p1(count:count+nii-1,:)-ii(1:nii,:)))));
    Feii=AAAii';
    corpus11y=p1(count:count+ny-1,:);
    AAAy(count) =1/(sum(sum(abs(p1(count:count+ny-1,:)-y(1:ny,:)))));
    Fey=AAAY';
    corpus11w=p1(count:count+nw-1,:);
    AAAw(count) =1/(sum(sum(abs(p1(count:count+nw-1,:)-w(1:nw,:)))));
    Few=AAAw';
    corpus11k=p1(count:count+nk-1,:);
    AAAk(count) =1/(sum(sum(abs(p1(count:count+nk-1,:)-k(1:nk,:)))));
    Fek=AAAK';
    corpus11d=p1(count:count+nd-1,:);
    AAAd(count) =1/(sum(sum(abs(p1(count:count+nd-1,:)-d(1:nd,:)))));
    Fed=AAAd';
    corpus11f=p1(count:count+nf-1,:);
    AAAf(count) =1/(sum(sum(abs(p1(count:count+nf-1,:)-f(1:nf,:)))));
    Fef=AAAF';
    corpus11l=p1(count:count+nl-1,:);
    AAAl(count) =1/(sum(sum(abs(p1(count:count+nl-1,:)-l(1:nl,:)))));
    Fel=AAAl';
    corpus11m=p1(count:count+nm-1,:);
    AAAm(count) =1/(sum(sum(abs(p1(count:count+nm-1,:)-m(1:nm,:)))));
    Fem=AAAm';
    corpus11n=p1(count:count+nn-1,:);
    AAAn(count) =1/(sum(sum(abs(p1(count:count+nn-1,:)-n(1:nn,:)))));
    Fen=AAAn';
    corpus11r=p1(count:count+nr-1,:);
    AAAr(count) =1/(sum(sum(abs(p1(count:count+nr-1,:)-r(1:nr,:)))));
    Fer=AAAr';
    corpus11s=p1(count:count+ns-1,:);

```

```

        AAAs(count) =1/(sum(sum(abs(p1(count:count+ns-1,:)-s(1:ns,:)))));
        Fes=AAAs';
    corpus11t=p1(count:count+nt-1,:);
        AAAAt(count) =1/(sum(sum(abs(p1(count:count+nt-1,:)-t(1:nt,:)))));
        Fet=AAAAt';
    corpus11h=p1(count:count+nh-1,:);
        AAAh (count)=1/(sum(sum(abs(p1(count:count+nh-1,:)-h(1:nh,:)))));
        Feh=AAAh';
    corpus11cha=p1(count:count+ncha-1,:);
        AAACHa(count) =1/(sum(sum(abs(p1(count:count+ncha-1,:)-cha(1:ncha,:)))));
        Fecha=AAACHa';
    corpus11kha=p1(count:count+nkha-1,:);
        AAAkha(count) =1/(sum(sum(abs(p1(count:count+nkha-1,:)-kha(1:nkha,:)))));
        Fekha=AAAkha';
    corpus11tha=p1(count:count+ntha-1,:);
        AAAtha(count) =1/(sum(sum(abs(p1(count:count+ntha-1,:)-tha(1:ntha,:)))));
        Fetha=AAAtha';
    corpus11gha=p1(count:count+ngha-1,:);
        AAAgha(count) =1/(sum(sum(abs(p1(count:count+ngha-1,:)-gha(1:ngha,:)))));
        Fegha=AAAgha';
    corpus11dta=p1(count:count+ndta-1,:);
        AAAdta(count) =1/(sum(sum(abs(p1(count:count+ndta-1,:)-dta(1:ndta,:)))));
        Fedta=AAAdta';
    j=1:ncorpus-50;
        jj=j';
        t=jj*10 ;
        format long e
        resultMax= [Fea,Feaa,Feuu, Feu,Fei,Feii,Fey, Few, Feh, Fef, Fed,Fek, Fel,Fem, Fen,
    Feq,Fer,Fes,Fet,Fetha,Fecha,Fegha,Fekha,Fedta ];
    plot(jj,Fea,jj,Feaa,jj,Feu,jj,Feuu,jj,Fei,jj,Feii,jj,Fey,jj,Few,jj,Feh,jj,Fed,jj,Fef,jj,Fel,jj,Fem,jj,Fen,jj,Fer,jj,Fes,jj,Fe
    q,jj,Fek,jj,Fet,jj,Fetha,jj,Fegha,jj,Fecha,jj,Fedta,jj,Fekha)
end

```

% SCRIPT en Matlab pour la classification des voyelles et semi voyelles de l'AS du corpus p1 en utilisant la distance de Manhattan ;

```

function yyy= my_fun(i,j)
global corpus, a,aa,i,ii,u,uu,y,w
[na,naa]=size(a);
[naa,naaa]=size(aa);
[nu, nuu]=size (u);
[nuu, nuuu]=size (uu);
[ni, nii]=size (i);
[nii, niia]=size (ii);
[ny, nyy]= size(y);
[nw, nww]=size (w);
corpus1=corpus11(1:2583,2:43);
[ncorpus, nccorpus]=size(corpus1);
for count=1:ncorpus-51;
    corpus11a=corpus1(count:count+na-1,:);
        AAAa (count)=1/(sum(sum(abs(corpus1(count:count+na-1,:)-a(1:na,:)))));
        Fea=AAAa';
        yfa=Fea*1;
    Ma=mean(yfa);
    za=Ma<yfa;
    corpus11aa=corpus1(count:count+naa-1,:);
        AAAaa (count)=1/(sum(sum(abs(corpus1(count:count+naa-1,:)-aa(1:naa,:)))));
        Feaa=AAAaa';
        yfaa=Feaa*1;
    Maa=mean(yfaa);

```

```

zaa=Maa<yfaa;
    corpus11u=corpus1(count:count+nu-1,:);
        AAAu(count) =1/(sum(sum(abs(corpus1(count:count+nu-1,:)-u(1:nu,:))));
        Feu=AAAu';
        yfu=Feu*1;
Mu=mean(yfu);
zu=Mμ<yfu;
    corpus11uu=corpus1(count:count+nuu-1,:);
        AAAuu(count) =1/(sum(sum(abs(corpus1(count:count+nuu-1,:)-uu(1:nuu,:))));
        Feuu=AAAuu';
        yfuu=Feuu*1;
Muu=mean(yfuu);
zuu=Muu<yfuu;
    corpus11ii=corpus1(count:count+ni-1,:);
        AAAi(count) =1/(sum(sum(abs(corpus1(count:count+ni-1,:)-i(1:ni,:))));
        Fei=AAAi';
        yfi=Fei*1;
Mi=mean(yfi);
zi=Mi<yfi;
    corpus11iii=corpus1(count:count+nii-1,:);
        AAAii(count) =1/(sum(sum(abs(corpus1(count:count+nii-1,:)-ii(1:nii,:))));
        Feii=AAAii';
        yfii=Feii*1;
Mii=mean(yfii);
zii=Mii<yfii;
    corpus11iy=corpus1(count:count+ny-1,:);
        AAAy(count) =1/(sum(sum(abs(corpus1(count:count+ny-1,:)-y(1:ny,:))));
        Fey=AAAy';
        yfy=Fey*1;
Mw=mean(yfy);
zy=Mw<yfy;
    corpus11lw=corpus1(count:count+nw-1,:);
        AAAw(count) =1/(sum(sum(abs(corpus1(count:count+nw-1,:)-w(1:nw,:))));
        Few=AAAw';
        yfw=Few*1;
My=mean(yfw);
zw=My<yfw;
j=1:ncorpus-51;
        jj='j';
        t=jj*10 ;
        format long e
        resultMax= [Fea,Feaa,Feuu, Feu,Fei,Feii,Fey, Few];

end

plot(t,Fea,t,Feaa,t,Feu,t,Feuu,t,Fei,t,Feii,t,Fey,t,Few);
plot(t,za,t,zaa,t,zu,t,zuu,t,zi,t,zii,t,zy,t,zw);

```

% SCRIPT en Matlab pour la classification de la voyelle [a] de l'AS du corpus corpus111 en utilisant les distances de Manhattan et celle d'Euclide ;

```

function yy= my_fun(i,j)
[na,naa]=size(a);
corpus1=corpus111(1:2583,2:43);
[ncorpus, nccorpus]=size(corpus1);
for count=1:ncorpus-na+1;
corpus11=corpus1(count:count+na-1,:);
yy(count)=1/(sum(sum(abs(corpus11(1:na,:)-a(1:na ,:))));
Fea=yyy';
for k=1:42-na+1;

```

```

yy(:,:)= 1/(sqrtm(abs(sum(sum((corpus1(count:count+na-1,k:k+na-1))^2-(a(:,k:k+na-1))^2))))));
yyyy(count)=yy(1,1);
Feucla=yyyy';
end
end
plot(1:ncorpus-na+1,Fea,1:ncorpus-na+1,Feucla)
plot(1:500,Feucla(1:500,1))

```

% SCRIPT en Matlab pour la classification de la voyelle [a] de l'AS du corpus corpus111 en utilisant la distance d'Euclide ;

```

function yy= my_fun(i,j)
[na,naa]=size(a);
corpus1=p3;
[ncorpus, nccorpus]=size(corpus1);

for count=1:ncorpus-na+1;
for k=1:42-na+1;
yy(:,:)= 1/(sqrtm(abs(sum(sum((corpus1(count:count+na-1,k:k+na-1))^2-(a(:,k:k+na-1))^2))))));
yyy(count)=yy(1,1);
Feucla=yyy';
end
end
j=1:ncorpus-na+1;
jj=j';
t=jj*10 ;
[jn,in]=size(Feucla);
yf=Feucla*1;
Mm=mean(yf);
z=Mm<yf;
plot(t,z)

```

\$script pour le calcul de la fonction objectif qui est la fonction de distance de Manhattan de notre modèle génétique séquentiel sous Matlab pour classer la voyelle a du corpus p1

```

function yyy= my_fun(i,j)
global p1 a
P=0.1
[na,naa]=size(a);
[ncorpus, nccorpus]=size(p1);
for count=1:ncorpus-na+1;
corpus11=p1(count:count+na-1,:);
yyy(count)=P/(sum(sum(abs(corpus11(1:na,:)-a(:,:))))));
Fema=yyy';
end

```

%script pour le calcul de la fonction objectif qui est la fonction de distance d'Euclide de notre modèle génétique séquentiel sous Matlab pour classer la voyelle a du corpus p1

```

function yy= my_fun(i,j)
global a p1
[na,naa]=size(a);
[ncorpus, nccorpus]=size(p1);
for count=1:ncorpus-na+1;
for k=1:42-na+1;
yy(:,:)= 1/(sqrtm(abs(sum(sum((p1(count:count+na-1,k:k+na-1))^2-(a(:,k:k+na-1))^2))))));
yyy(count) =yy(1,1);
Feucla=yyy';
end
end

```

%script pour le calcul de la fonction objectif qui est la fonction de distance de Manhattan de notre modèle génétique séquentiel sous Matlab pour classer des consonnes occlusives et fericatives du corpus p1

```

function yyy= my_fun(i,j)
global corpus, m,n,r,t,q,h,tha,cha
[ntha,nthatha]=size(tha); [nt,ntt]=size(t); [nm, nmm]=size (m); [nn, nnn]=size (n);
[nr, nrr]=size (r); [nq, nqq]=size (q); [ncha, nchacha]= size(cha); [nh, nhh]=size (h);
corpus1=corpus11(1:2583,2:43);
[nrcorpus, nccorpus]=size(corpus1);
for count=1:nrcorpus-51;
    corpus11tha=corpus1(count:count+ntha-1,:);
    AAAta(count)=1/(sum(sum(abs(corpus1(count:count+ntha-1,:)-tha(1:ntha,:)))));
    Fetha=AAAta';
    yftha=Fetha*1;
Mtha=mean(yftha);
ztha=Mtha<yftha;
    corpus11cha=corpus1(count:count+ncha-1,:);
    AAAcha(count)=1/(sum(sum(abs(corpus1(count:count+ncha-1,:)-cha(1:ncha,:)))));
    Fecha=AAAcha';
    yfcha=Fecha*1;
Mcha=mean(yfcha);
zcha=Mcha<yfcha;
    corpus11h=corpus1(count:count+nh-1,:);
    AAAh(count) =1/(sum(sum(abs(corpus1(count:count+nh-1,:)-h(1:nh,:)))));
    Feh=AAAh';
    yfh=Feh*1;
Mh=mean(yfh);
zh=Mh<yfh;
    corpus11t=corpus1(count:count+nt-1,:);
    AAAt(count) =1/(sum(sum(abs(corpus1(count:count+nt-1,:)-t(1:nt,:)))));
    Fet=AAAt';
    yft=Fet*1;
Mt=mean(yft);
zt=Mt<yft;
    corpus11r=corpus1(count:count+nr-1,:);
    AAAr(count) =1/(sum(sum(abs(corpus1(count:count+nr-1,:)-r(1:nr,:)))));
    Fer=AAAr';
    yfr=Fer*1;
Mr=mean(yfr);
zr=Mr<yfr;
    corpus11q=corpus1(count:count+nq-1,:);
    AAAq(count) =1/(sum(sum(abs(corpus1(count:count+nq-1,:)-q(1:nq,:)))));
    Feq=AAAq';
    yfq=Feq*1;
Mq=mean(yfq);
zq=Mq<yfq;
    corpus11m=corpus1(count:count+nm-1,:);
    AAAm(count) =1/(sum(sum(abs(corpus1(count:count+nm-1,:)-m(1:nm,:)))));
    Fem=AAAm';
    yfm=Fem*1;
Mm=mean(yfm);
zm=Mm<yfm;
    corpus11n=corpus1(count:count+nn-1,:);
    AAAn(count) =1/(sum(sum(abs(corpus1(count:count+nn-1,:)-n(1:nn,:)))));
    Fen=AAAn';
    yfn=Fen*1;
Mn=mean(yfn);
zn=Mn<yfn;
    j=1:nccorpus-51;
    jj=j';

```

```

        t=jj*10 ;
        format long e
        resultMax= [Fetha,Fecha,Fem,Feq,Fer,Fet, Feh];

end

plot(t,Fetha,t,Fecha,t,Feq,t,Fer,t,Fet,t,Feh,t,Fem);
plot(t,zcha,t,zh,);

% script en Matlab pour executer l'AG ayant pour fonction object la fonction de Manhattan
function yyy= my_fun(i,j)
global a p1
    fitnessFunction=@Fmap1count;
    %%Fitness function
%%Number of Variables
GenomeLength = na*42;
nvars = na*42;
%Start with default options
options = gaoptimset;
%%Modify some parameters
options = gaoptimset(options,'PopulationType' , 'bitstring')
options = gaoptimset(options,'PopulationSize' ,1000);
options = gaoptimset(options,'MigrationFraction' ,0);
options = gaoptimset('SelectionFcn', { @selecttournament, siz});
options = gaoptimset(options,'StallTimeLimit' ,Inf);
options = gaoptimset(options,'Generations' ,500);
options = gaoptimset(options,'SelectionFcn' ,@selectionroulette);
options = gaoptimset(options,'CrossoverFcn' ,@crossovertwopoint);
options = gaoptimset(options,'MutationFcn' ,{ @mutationuniform 0.01 });
options = gaoptimset(options,'Display' , 'diagnose');
options = gaoptimset('InitialPop', final_pop);
options = gaoptimset('OutputFcn', @Fmap1)
options = gaoptimset(options,'PlotFcns' ,{ @gaplotbestf @gaplotbestindiv @gaplotdistance @gaplotexpectation
@gaplotgenealogy @gaplotrange @gaplotscorediversity @gaplotscores @gaplotselection @gaplotstopping });
%%Run GA
deterministicstudy
[X,FVAL,REASON,OUTPUT,final_pop,SCORES] = ga(@Fmap1count, na*42,options)

```

الجملة الأولى: إنّ الأشجار الكبيرة موجودة في الغابة الصغيرة. Phrase 1 (P1).

الجملة الثانية: ورود البستان في البيوت و القصور. Phrase 2 (P2).

الجملة الثالثة: رجالان نحيلان كئيبان. Phrase 3 (P3).

الجملة الرابعة: وعود اللّثيم كاذبة. Phrase 4 (P4).

## Corpus 1

## مدينة بغداد

بغداد مدينة عربية مهمّة، لا لأتّها عاصمة العراق اليوم فقط، بل لأنّها لها مكانتها في عهد التاريخ العربيّ الإسلاميّ أيضا. فقد كانت مركز الحضارة الإسلامية خلال مدّة طويلة من حكم العبّاسيّين من عام سبعمائة و خمسين إلى عام ألف و مئتين و خمسين.

أسّسها المنصور الخليفة العبّاسيّ الثاني بعد قيام الحكم العبّاسي بسنوات و سمّاها "دار السّلام" إلا أنّ اسم بغداد غلب عليها. و يقال أنّ اسم بغداد يعني عطية الله باللّغة الفارسية القديمة.

و من أسباب اختيار بغداد عاصمة الدّولة أنّها تقع على نهر الدجلة و من هنا كانت على اتصال بالبلاد التي يمر بها هذا النّهر. و هي في موقع عند أقصر مسافة بين الدجلة و الفرات. و بهذا كان الاتصال بالفرات و بالبلاد الواقعة عليه سهلا ميّسرا.

وهي فوق ذلك في موقع تحيط به الأنهار بحيث يصعب على العدوّ الوصول إليها، و بذلك يكون الدّفاع عنها سهلا، و هي أخيرا في مكان متوسط من العالم الإسلاميّ كما كان يعرف في ذلك الوقت.

و لم يمض على تأسيس بغداد وقت طويل حتى أصبحت مركزا حضاريا مشهورا، يقبل عليه المسلمون و غير المسلمين من مختلف مناطق العالم.

كما أصبحت ملقّى الشّعراء و العلماء و الأطباء و رجال الفلسفة و الدّين و المترجمين. و زاد عدد سگانها أيّام هارون الرّشيد و المأمون عن مليون نسمة في رأي بعض المؤرخين، و هذا عدد كبير بالنسبة لذلك العصر. و قد ذاع اسمها في العالم، لا بفضل مكنتاتها و مدارسها و دور التّرجمة فيها فقط بل بسبب مكانتها من قصص ألف ليلة و ليلة أيضا، التي تركت أثرها في الأدب و الموسيقى العالميّة.

بغداد في الوقت الحاضر من المدن الكبرى في العالم العربيّ، يجد السّائح فيها آثار الماضي و مظاهر المدينة الحديثة في شوارعها و أبنيتها و مصايفها و معاهدها العلميّة و متاحفها و مساجدها. و يبلغ عدد

سكانها اليوم أكثر من مليون نسمة، و بغداد مشهورة كذلك بأسواقها المسقوفة التي تحمل الطابع الشرقي. و فيها يجد الزائر كثيرا من الصناعات اليدوية.

و من أهم شوارعها شارع الرشيد و هو مركز الحركة التجارية، و شارع أبو نواس المعروف بمقاهيه و ملاهيه و مطاعمه الواقعة على نهر الدجلة.

و مما يزيدها جمالا أن فيها بساتين و حدائق عامة كثيرة، كما أن في بغداد أماكن مقدسة كقبر "الإمام ابن حنبل" و الشيخ "عبد القادر الكيلاني" و "الإمامين الكاظمين". و لا تزال هذه الأماكن مقصدا لجماهير المسلمين من مشارق الأرض و مغاربه.

## Corpus 2 :

### الدرس :

جان يستمع إلى الراديو فإذا بصوت جميل يقول : يوضع اللحم في القدر ثم يغطي بالبطاطس ويضاف إليه البصل إلى آخر ما جاء في برنامج المرأة ذلك اليوم

وكان ذلك المستمع الكريم يسمع بقلب يطير شوقا ولم يستطيع أن يطير فقام إلى أهله يقول لهم : لا بد من الزواج بهذه المرأة فسألوه هل تعرفها ؟

فأجاب لا اعرف إلا إذاعتها الذيزة في الراديو إنها ملكت قلبي

وكان صاحبنا هذا من أوائل الذين يخلطون بين القلب والمعدة ، فإذا سأله الطبيب يوما أين معدتك أشار إلى قلبه ، وإذا سأله أين قلبك أشار إلى معدته

وكان لابد للمرأة التي تريد أن تملك قلبه من أن تعرف طريقها إلى المعدة أولا ، فإذا ملكتها كل شيء وتم الزواج ومرت الأيام الأولى والعريس يطير شوقا منتظرا اليوم الذي تدخل فيه المطبخ ، تطبخ له الأكلات اللذيذة التي وصفتها في الراديو وأخيرا دخلت الزوجة إلى المطبخ وزوجها يقول في شوق ، صلوات الله على تلك التي ستجعلني سعيدا بالأكلة اللذيذ وانتظر ساعة ثم ساعة ثم ساعة حتى تعب من الانتظار فخرجت من المطبخ والعرق يسيل من وجهها وقالت له أسفة عملت لك بيض مقليا خوفا من التأخير

فاخفي الرجل شعوره بالخيبة وبدأ يأكل البيض المقلي وجده قد احترق وتحجر ولما كانت الساعة الرابعة  
أسرعت الزوجة إلى الزوج كأنها على موعد هام وفي الخامسة جلس الزوج وحده يستمع إلى الراديو فإذا  
بامرأته تذيع على المستمعين يوضع اللحم في القدر ثم يغطى بالبطاطس ويضاف إليه البصل إلى آخره  
وفكر الزوج المسكين قليلا ولم يعد يعرف ماذا يفعل هل يضحك، هل يبكي. ( توفيق إليكم امن الله القاهرة  
بدون تاريخ الصفحة مئة وستة وستين إلى مئة وتسع وستين )

---

Valeurs moyennes de l'inverse de FeManhattan pour les consonnes :

temps	[j]	[θ]	[t]	[r]	[q]	[h]	[m]	[n]
10.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0
20.0	0.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0
30.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
40.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
50.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
60.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
70.0	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0
80.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0
90.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0
100.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0
110.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	1.0
120.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
130.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
140.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
150.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
160.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
170.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
180.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
190.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
200.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
210.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
220.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
230.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
240.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
250.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
260.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
270.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
280.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
290.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
300.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
310.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
320.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
330.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
340.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
350.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
360.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
370.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
380.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
390.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
400.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
410.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
420.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
430.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
440.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
450.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
460.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
470.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
480.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
490.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
500.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
510.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
520.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
530.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
540.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
550.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
560.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
570.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
580.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
590.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
600.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
610.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
620.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
630.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
640.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
650.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
660.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
670.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
680.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
690.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0









				///				
				///				
				///				
9820.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
9830.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9840.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9850.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9860.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9870.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9880.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
9890.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9900.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9910.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9920.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9930.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9940.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9950.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9960.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9970.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9980.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9990.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10010.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10020.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10030.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10040.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10050.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10060.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10070.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10080.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10090.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
10100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
10110.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
10120.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10130.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10140.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10150.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10160.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10170.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10180.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10190.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
10200.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
10210.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
10220.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
10230.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
10240.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	0.0
10250.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
10260.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
10270.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	0.0
10280.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
10290.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
10300.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
10310.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0
10320.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0

Tableau : Valeurs prise par la moyenne de la fonction objectif qui est l'inverse de la fonction de distance de Manhattan de chacune des huit consonnes à classer du corpus 1

L'existence du phonème recherché est déterminée par les intervalles de valeurs successives égales à 1. Ces intervalles sont bien situés dans le continuum vocal qui peut être un corpus ou une phrase ou un mot isolé.

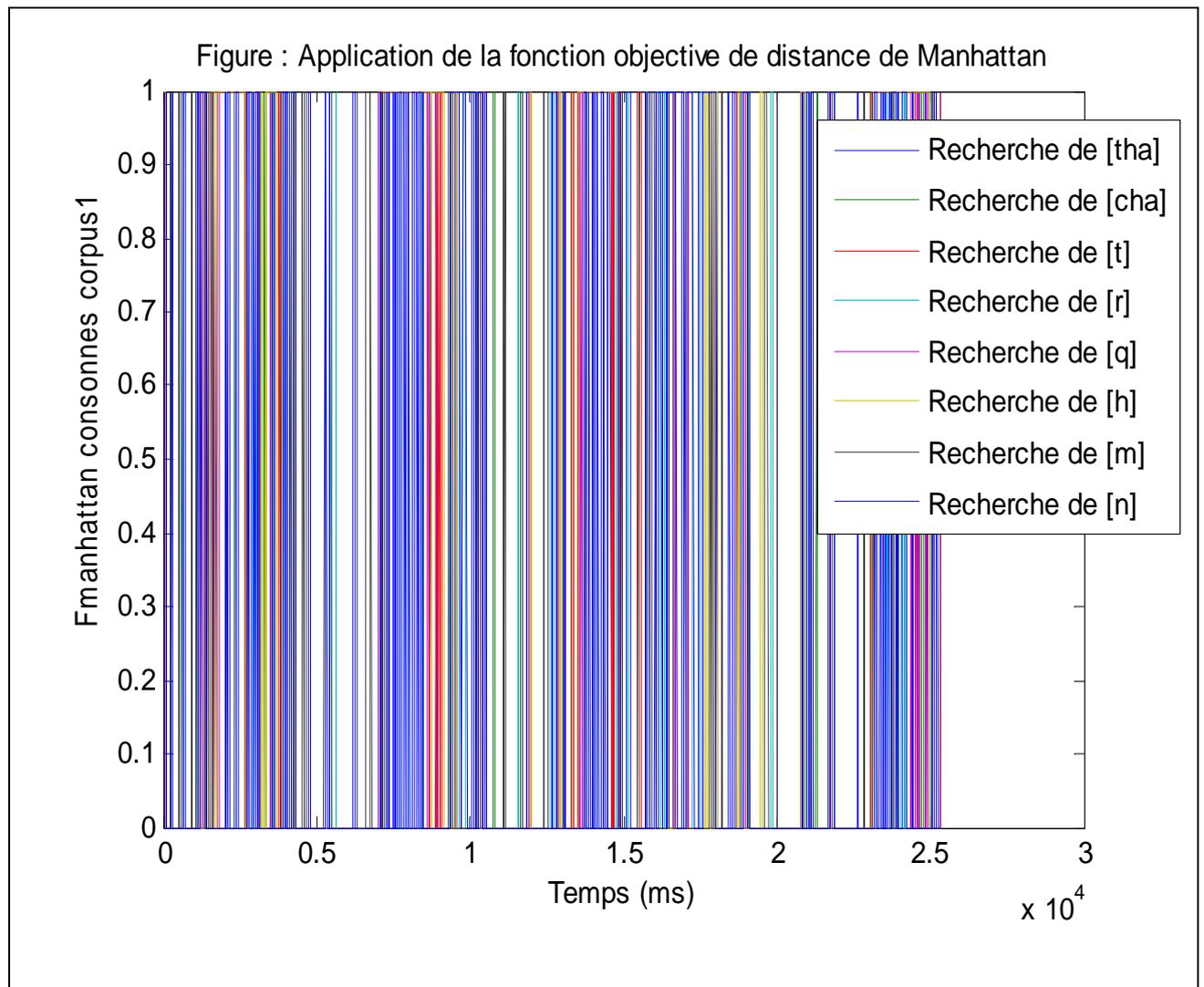


Figure Annexe 3. 1 : Moyenne de la fonction objectif qui est l'inverse de la fonction de distance de Manhattan de chacune des huit consonnes à classer du corpus 1

([j] [θ] [t] [r] [q] [h] [m] [n])

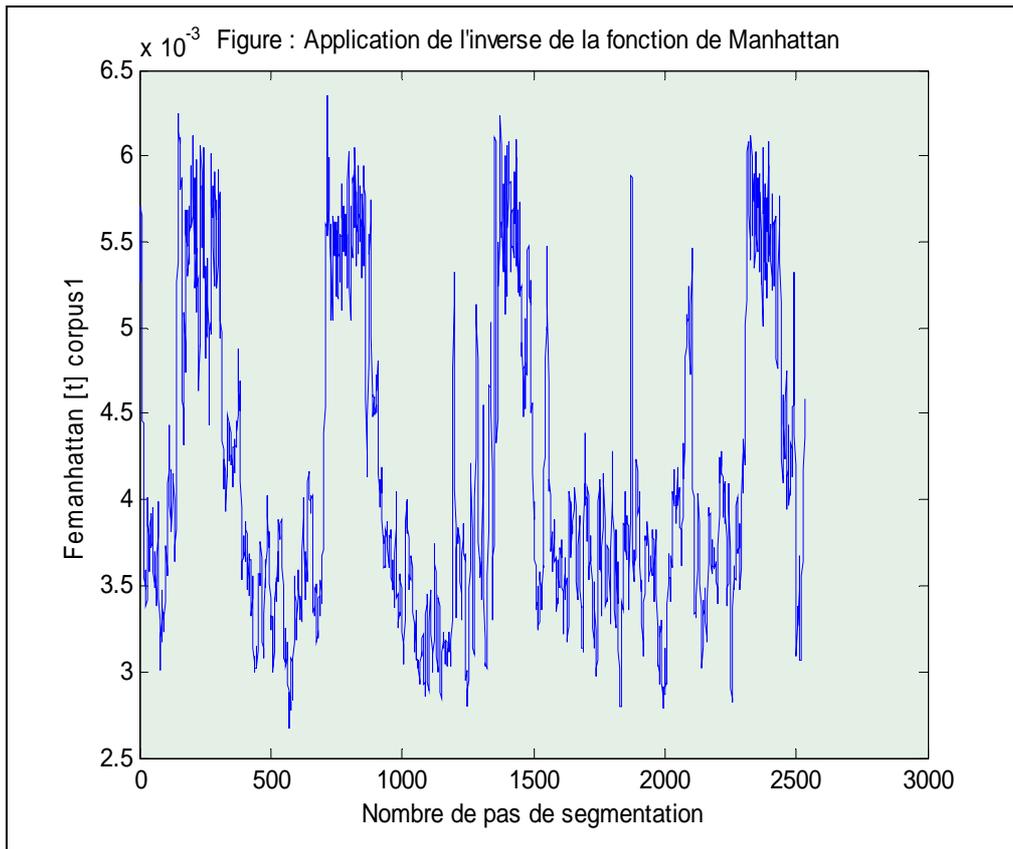


Figure Annexe 3.2 : Application de l'Inverse de la fonction de distance de Manhattan pour chercher la consonnes [ t ] dans le corpus 1

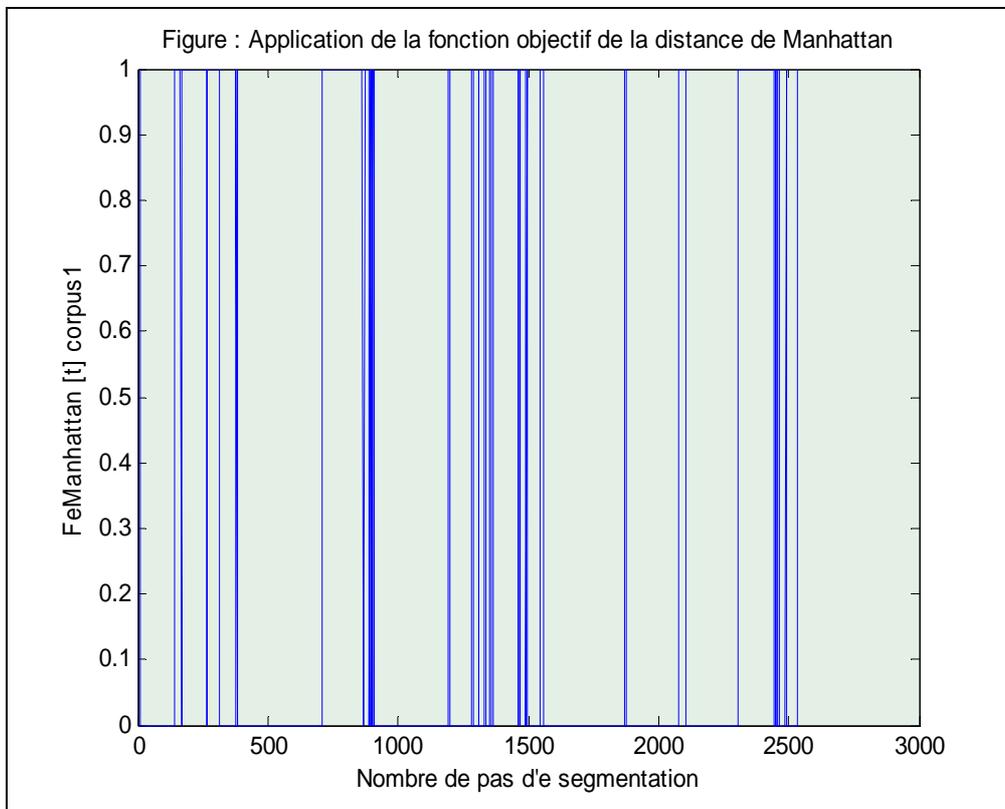


Figure Annexe 3.3 : Fonction objectif qui est l'inverse de la fonction de distance de Manhattan de la consonne [ t ]