République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE DE M'SILA

FACULTE DE TECHNOLOGIE

DEPARTEMENT : ELECTRONIQUE

MEMOIRE DE MASTER

DOMAINE : SCIENCES ET TECHNOLOGIE

FILIERE : GENIE ELECTRIQUE

OPTION : CONTROLE INDUSTRIEL

Thème

UTILISATION DES APPROCHES EVOLUTIONNAIRES POUR LA SELECTION DES PARAMETRES PERTINENTS : APPLICATION A L'AUTHENTIFICATION DES PERSONNES

Présenté par :

Kheira. BOUKHAROUBA

Encadré par :

Dr. Djamel SAIGAA

Dr. Abdelghani HARRAG

N°d'ordre:2011/05/22/62

Promotion : JUIN 2011

Remerciements

Avant tout, nous remercions le bon dieu tout puissant qui nous donne de la foi, du courage et de patience, qu'il nous a données durant toutes ces années d'étude.

Ainsi, nous tenons également à exprimer nos vifs remerciements à mes encadreurs ***Mrs : DJ. SAIGAA*** *et* ***A.HARRAG*** *pour avoir d'abord proposée ce thème, pour suivi continuel durant toute cette période. Qui n'a pas cessée de nous donner ses conseils.*

Nos remerciements vont aussi à tous les enseignants et le ***chef de département d'Electronique*** *qui a contribué à notre formation par ailleurs, ainsi à tous* ***les membres du jury*** *qui ont accepté de juger notre travail.*

En fin nous tenons à exprimer notre reconnaissance à tous nos amis et collègues pour le soutient moral et matériel.

DEDICACE

Je dédie ce modeste travail

A mes chers parents mon père Brahim et ma mère (aichouche)

AMes frères Khaled et aissa.

Mes soeurs houria et ses fils, merieme et madjda, A tous mes ancles et ses femmes, mes cousines, mes cousins et toute les membres de ma grande famille.

A tout mes profs qu'ils m'ont appris durant toutes mes années d'étude

A mes encadreurs Dj. Saigaa et A.Harrag.

A tous les enseignants qui m'ont aidé de proche ou de loin pour terminer ce travail.

Et bien sur à mes collègues nassima, Samia, Asma, Amira, Souad et Asma qui m'a accompagné pendant le long de cette période pour réaliser ce modeste travail.

A tous mes collègues sans exception & à toutes les promos DE MASTER 2011.

Je voux remercie tous

Sommaire

Introduction générale…….1-3

Chapitre 1 : La reconnaissance automatique du locuteur

1.1. Introduction……...……..4

1.2. Terminologie……4

1.2.1. Identification et vérification ……4

1.2.2. Typologie des erreurs ……...……7

1.2.3. Dépendance du texte ……...……8

1.3. Structure d'un système de RAL……..……...……..…9

1 .3.1.Paramétrisation…….……9

1.3.2. Modélisation de locuteurs……...…10

1.3.3. Modes de décision ……..……13

1.3.3.1. Pour l'identification……...…13

1.3.3.2. Pour la vérification ……...…14

1.4. Evaluation des performances en reconnaissance vocale …….……15

1.5-Conclusion…….……..……...…....17

Chapitre 2 : Paramétrisation du signal parole

2.1. Introduction ……...18

2.2. Le signal de la parole …….…...19

2.2.1. Redondance du signal …….……19

2.2.2. Variabilité du signal…….……19

2.2.3. Les effets de coarticulation…….….20

2.2.4. Les facteurs extérieurs…….….20

2.3. Les différents paramètres extraits du signal parole ……20
2.3.1. Les principaux paramètres de l'analyse spectrale 22
2.3.1.1. Coefficients issus d'une analyse par prédiction linéaire……... 22
2.3.1.1.1. Les coefficients LPC……...……22
2.3.1.1.2. Les coefficients cepstraux(LPCC)……..…25
2.3.1.2. Coefficients issus d'une analyse par banc de filtres ……..26
2.3.1.2.1 Les coefficients MFCC……..…….....26
2.3.1.2.2. Les coefficients PLP……... ...27
2.3.2.Paramètres prosodiques……29
2.3.2.1. L'énergie……30
2.3.2.2.Fréquence fondamentale(ou pitch)…….30
2.4. Conclusion……

Chapitre 3 : Extraction et sélection des paramètres

3.1. Introduction……...…32
3.2. Sélection et Extraction des caractéristiques……....34
3.2.1 .La recherche exhaustive…….……..…….35
3.2.2. Meilleure Caractéristique Individuelle……..……35
3.2.3. Algorithme de recherche séquentielle…….…….…....35
3.2.4. Analyse en composantes principales…….…...36
3.2.5. L'analyse discriminante linéaire…….….37
3.2.6. Les algorithmes génétiques…….…….….38
3.3. Analyse et sélection des paramètres…….…….…38
3.3.1. Sélection par F-ratio…….....….39
3.3.2. Sélection basée sur les performances de reconnaissance……..……39
3.4. Conclusion……...…..41

Chapitre 4 : les algorithmes génétiques

3.1. Introduction ……..…42

3.2. Définition ……..…42

3.3. Caractéristique de l'AG ……...….43

3.4. Applications……..….43

3.5. Principe…….….44

3.6. Le Codage ……...46

3.6.1. Codage binaire……...…….46

3.6.2. Codage réel……...……47

3.7. Génération aléatoire de la population initiale……...……47

3.8. Opérateurs de sélection…….….48

3.9. Opérateurs de croisement……..….49

3.9.1. Croisement en un point (par découpage) …….....….49

3.9.2. Croisement en deux points :… …….……...…….50

3.9.3. Croisement uniforme …….…….....……50

3.10. Opérateur de mutation……..……51

3.11. Gestion des contraintes……...…….52

3.12. Un exemple simple…….…..52

3.12.1. Tirage et évaluation de la population initiale……..….53

3.12.2. Sélection…….…..53

3.12.3. Le croisement…….…..54

3.12.4. La mutation……..54

3.12.5. Retour à la phase d'évaluation……..……..……..55

3.13. Conclusion……...…..55

Chapitre 5 : simulations et résultats

5.1. Introduction…….……..56

5.2. Setup expérimental …….....…...56

5.2.1. Les paramètres de l'algorithme génétique……...…56

5.2.2. Classification K-NN (k-nearest neighbor algorithm) ……...……57

5.2.3. La base de données utilisée…….…...…61

5.2.4. Résultat de la simulation……..……...…61

5.2.4.1. Coefficients MFCC……61

5.2.4.2. Coefficients ΔMFCC…….....…64

5.2.4.3. Coefficients ΔΔMFCC……...…..……..68

5.2.4.4. Coefficients MFCC+ΔMFCC+ΔΔMFCC ……...… .71

5.3. Discussion des résultats…….....….75

5.4. Conclusion…….…75

Conclusion générale …….76

Références

Chapitre 1 : La reconnaissance automatique du locuteur

1.1. Introduction

La reconnaissance automatique du locuteur est une technique informatique qui permet d'analyser un mot ou une phrase captée au moyen d'un microphone pour la transcrire sous la forme d'un texte exploitable par une machine.

Les recherches en reconnaissance automatique du locuteur (RAL) font partie du domaine plus large de la communication Homme-Machine (Figure 1.1). Dans ce contexte, en effet il souhaitable qu'une machine puisse identifier automatiquement la personne qui lui parle, comme un locuteur le fait naturellement au cours d'une conversation. Cela peut-être nécessaire pour une authentification vocale (mot de passe vocal par exemple), ou pour aider à d'autres tâches (reconnaissance de la parole, synthèse de la parole, …).

1.2. Terminologie

Nous définissons dans cette section un certain nombre de termes fréquemment utilisés dans le domaine de la RAL.

1.2.1. Identification et verification

Commençons par définir les deux principales tâches que l'on distingue en reconnaissance automatique du locuteur, ainsi que les différentes phases qui constituent chacune de ces tâches.

La première phase d'un système de reconnaissance du locuteur est la phase d'apprentissage (Figure 1.2). Au cours de cette phase, on construit une base de référence contenant des données (signaux, paramètres, modèles) relatives à un nombre de locuteurs fixé S. La phase de test dépend alors de la tâche qui est réalisée. On en distingue essentiellement trois :

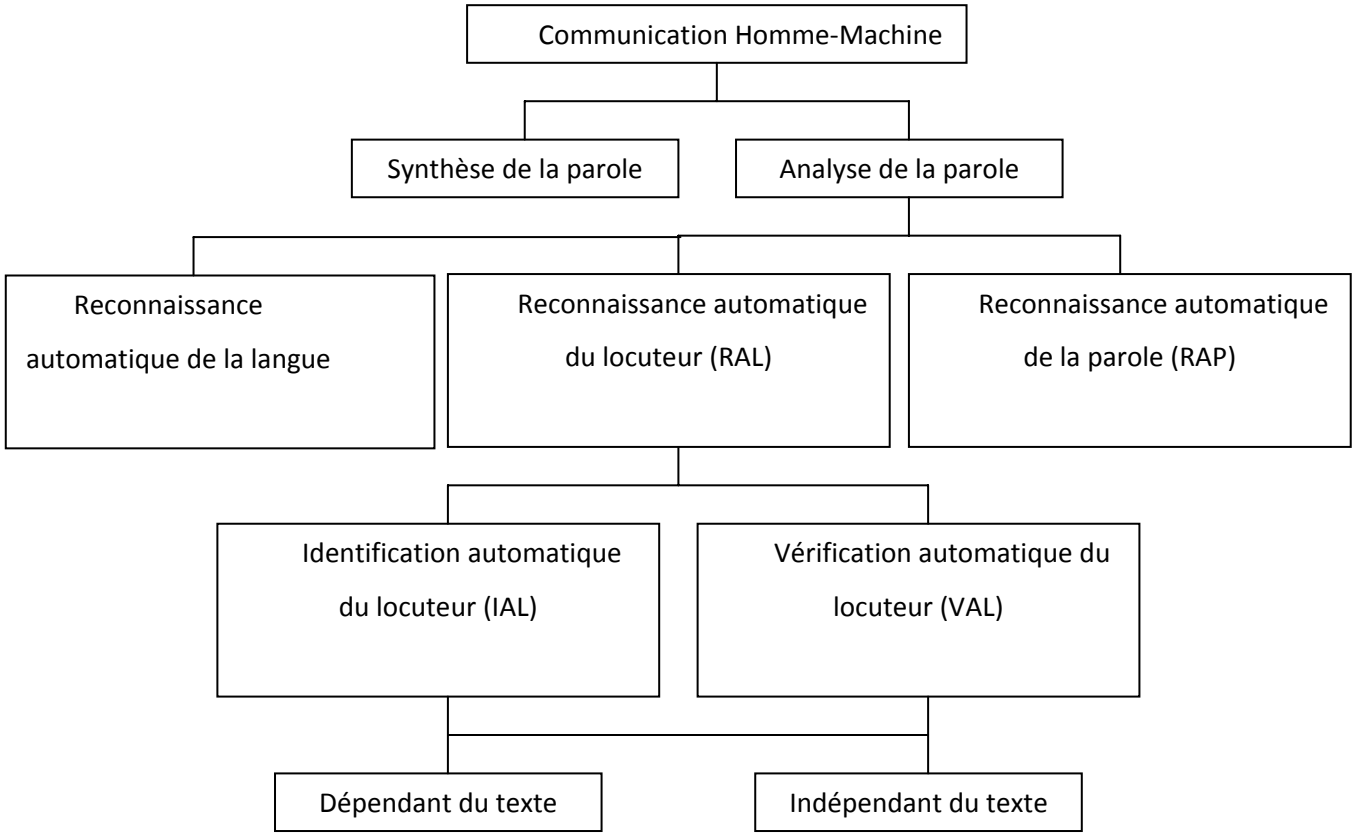

Figure 1.1: La reconnaissance automatique du locuteur dans le contexte de la communication Homme-Machine.

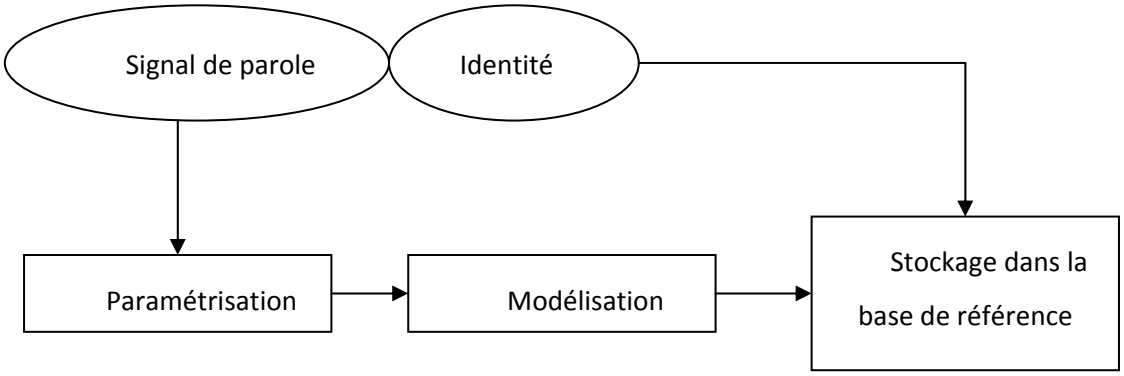

Figure 1.2 : Schéma modulaire de la phase d'apprentissage en reconnaissance du locuteur.

Identification en ensemble fermé (Atal, 1976 ; Doddington, 1985) : On dispose uniquement d'un échantillon de parole du locuteur inconnu. Le système fournit en sortie l'identité du locuteur de la base de référence dont le locuteur inconnu est le plus « proche ». Cette décision est prise après avoir comparé le locuteur inconnu à tous les locuteurs de la base de référence. Il s'agit du type 1 parmi *S*. On suppose en fait que le locuteur inconnu fait nécessairement partie de la base de référence (Figure 1.3)

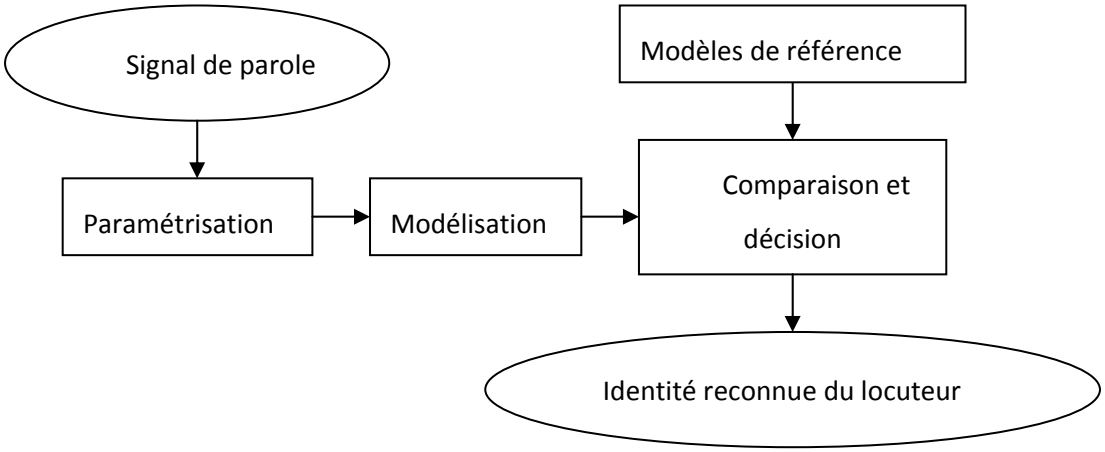

Figure. 1.3 : Schéma modulaire de l'identification du locuteur en ensemble fermé.

Vérification (Atal, 1976 ; Rosenberg, 1976 ; Doddington, 1985) : on dispose d'un échantillon de parole du locuteur inconnu ainsi que d'une identité proclamée, identité qui est celle de l'un des locuteurs de la base de référence. Le système doit alors vérifier si cette identité est correcte. On dit que le système rejette le locuteur si cette identité est considérée comme erronée, et qu'il accepte s'il juge cette identité correcte. Il s'agit cette fois d'une décision binaire (Figure 1.4).

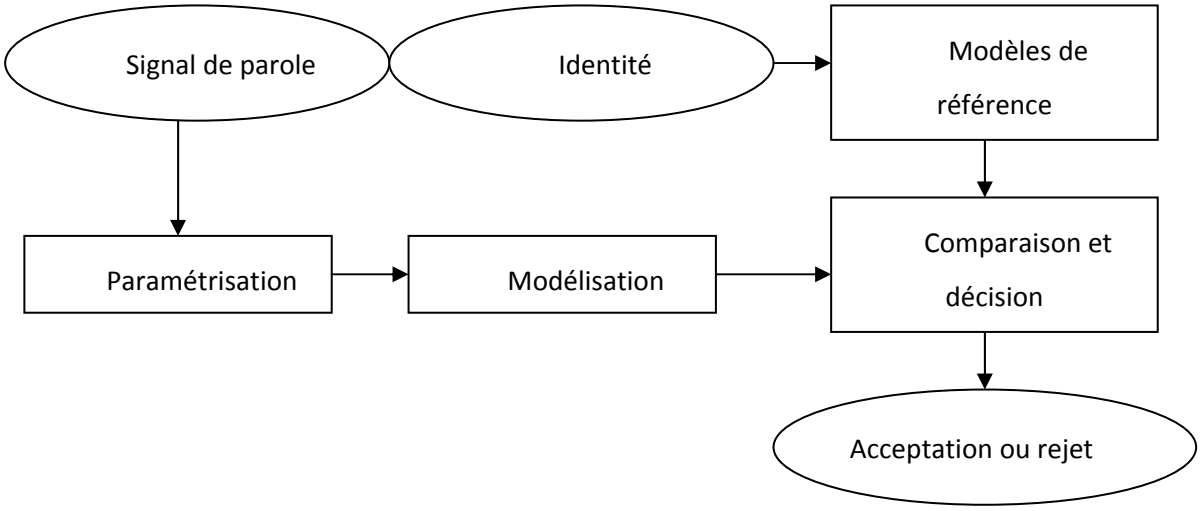

Figure 1.4 : Schéma modulaire de la vérification du locuteur.

Identification en ensemble ouvert (Doddington, 1985) : c'est une combinaison des deux tâches précédentes. Le système commence par faire une identification, et choisit donc le locuteur de la base de référence qui est le plus proche du locuteur inconnu. Puis il décide finalement si c'est bien ce locuteur-là (Figure 1.5).

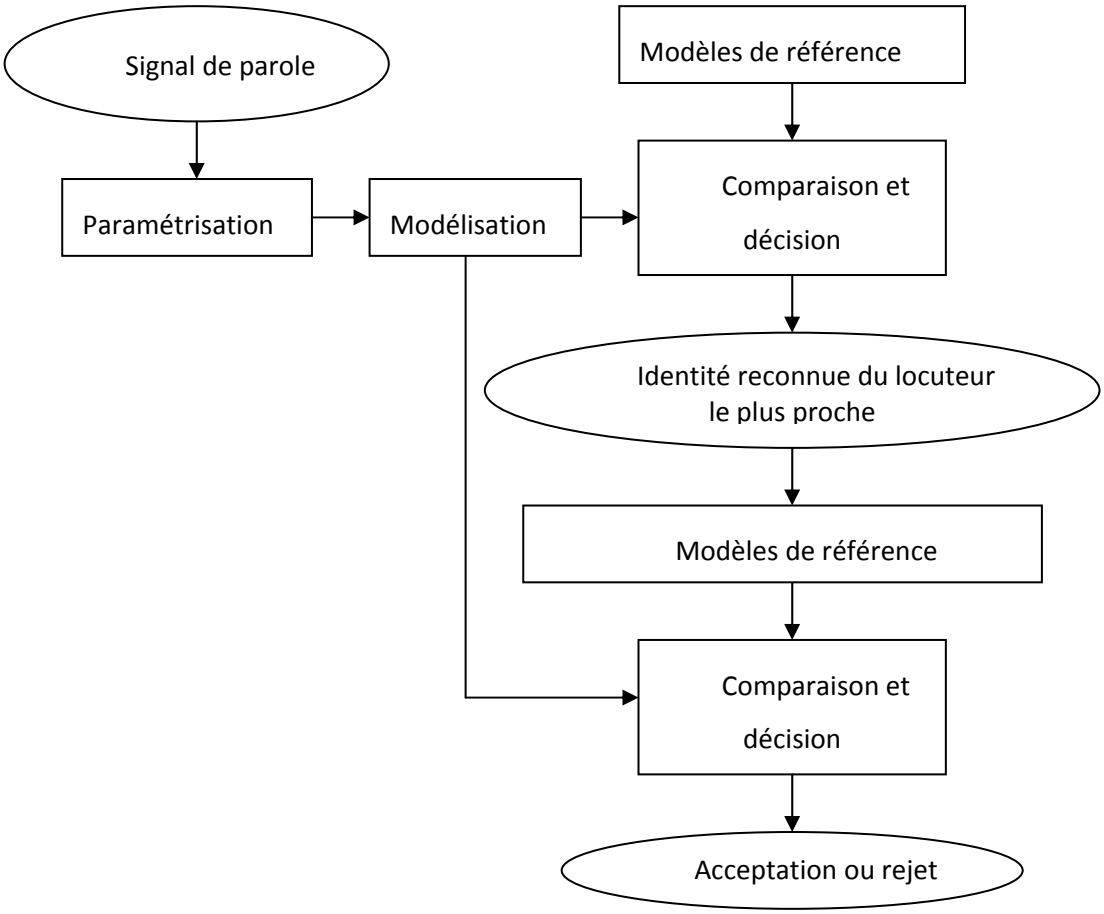

Figure. 1.5 : Schéma modulaire de l'identification du locuteur en ensemble ouvert.

1.2.2. Typologie des erreurs

Chaque tâche possède ses propres erreurs. Nous rappelons ici la typologie de chacune d'elles. Les performances d'un système d'identification en ensemble fermé se mesurent par son **taux de mauvaise identification**.

Celles d'un système de vérification se mesurent par son **taux de fausse acceptation** et par son **taux de faux rejet**. La fausse acceptation correspond au cas où le système accepte le locuteur inconnu alors que celui-ci n'est pas la personne qu'il prétend être. Le faux rejet correspond au cas où le système rejette le locuteur inconnu alors qu'il est vraiment la personne dont il a donné l'identité au système.

Enfin, les performances d'un système d'identification en ensemble ouvert se mesurent par son **taux de mauvaise identification**, c'est-à-dire un locuteur faisant partie de la base de référence est reconnu comme un autre locuteur de cette base, son **taux de fausse acceptation**, c'est-à-dire un imposteur est accepté comme l'un des locuteurs de la base de référence, et son **taux de faux rejet**, dans le cas où un locuteur faisant partie de la base de référence est rejeté.

1.2.3. Dépendance au texte

On distingue classiquement en reconnaissance automatique du locuteur deux types de contraintes par rapport au texte, l'une que l'on appelle **dépendante du texte**, et l'autre que l'on nomme **indépendante du texte** (Atal, 1976 ; Rosenberg, 1976 ; Doddington, 1985). Mais cette terminologie ne prend pas bien compte des différentes dépendances au texte possible, comme le remarquent les auteurs du rapport du projet Européen SAM-A (Bimbot, 1993). Les différents systèmes y sont classés, du plus contraignant au moins contraignant, de la façon suivante :

Système à texte fixé dépendant du locuteur : pour un locuteur donné, le texte est toujours le même d'une session à l'autre. Mais chaque locuteur à un texte différent.

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

Système dépendant d'événements phonétiques : le vocabulaire n'est pas directement imposé, mais certains événements phonétiques doivent être présents dans la séquence de parole prononcée (présence de certaines voyelles, de certaines nasales, …).

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

Système indépendant du texte : le locuteur est entièrement libre de ce qu'il dit à chaque session.

Cette classification prend bien mieux compte des différents systèmes que l'on peut effectivement trouver dans des articles, ou dans des applications, et cela sans ambiguïté.

1.3. Structure d'un système de RAL

La reconnaissance automatique du locuteur peut aussi être interprétée comme une tâche particulière de reconnaissance des formes. Un système de RAL se divise généralement en quatre modules: un **module de paramétrisation** du signal de parole, qui est généralement constitué d'une analyse spectrale vectorielle ; un **module de modélisation**, qui détermine les caractéristiques d'un modèle à partir des paramètres extraits ; un **module de comparaison**, qui consiste à utiliser des mesures de similarité entre modèles ; et enfin un **module de décision**, qui fournit finalement la réponse du système (Figure 1.6).

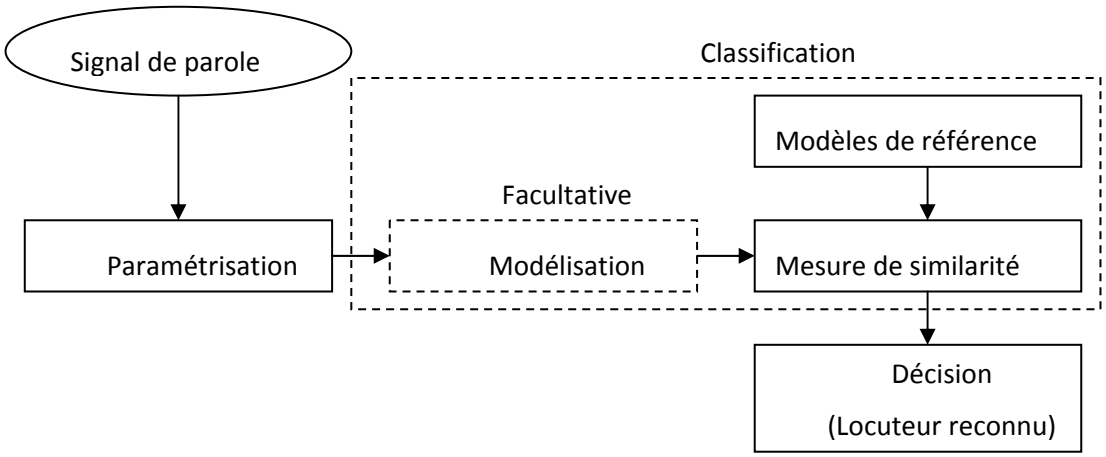

Figure 1.6 : Schéma modulaire d'un système de RAL (exemple d'un système d'identification du locuteur en ensemble fermé).

1.3.1. Paramétrisation

Un premier module de traitement du signal réalise l'analyse acoustique du signal de parole sur des fenêtres temporelles de courte durée, de 10 à 30 ms. A l'issue de cette étape, le signal est représenté par des vecteurs acoustiques de paramètres qui permettent de discriminer différents locuteurs, ce qui permet de réduire l'information en quantité et en redondance. Idéalement, les paramètres (ou traits acoustiques) doivent être :

- fréquents (p.exp. ne pas correspondre à des événements ne survenant que très rarement dans le signal de parole d'un locuteur),
- facilement mesurables,

- ne pas être trop sensibles à la variabilité intra-locuteur,
- ne pas être affectés par le bruit ambiant ou par les variations dues au canal de transmission,
- robustes face aux imitateurs.

En pratique, il est très difficile de réunir tous ces attributs en même temps. La sélection de traits acoustiques pertinents pour la RAL est donc un sujet largement traité dans la littérature ; sélection de paramètres séparant les locuteurs en terme de F-ratio (ou ses variantes) (Sambur, 1975 ; Bonastre, 1992) ; sélection par programmation dynamique (Cheung, 1978) ; sélection suivant le taux d'identification (Atal, 1974). Finalement, il ressort que les seuls types de paramètres vraiment pertinents et utilisables efficacement sont les paramètres de l'analyse spectrale (coefficients de prédiction linéaire et leurs transformations LPC, LPCC, PARCOR, … ainsi que les coefficients de l'analyse issus de l'analyse en banc de filtres et leurs transformations MFCC, MFC, …) et éventuellement les paramètres prosodiques (l'énergie, la durée (Van den Heuvel, 1994) et la fréquence fondamentale (Atal, 1972)). Nous pouvons noter qu'ils sont respectivement corrélés à la forme du conduit vocal et à la source de l'appareil de production de la parole.

1.3.2. Modélisation de locuteurs

Que ce soit pour reconnaître le message prononcé par un locuteur ou son identité, il nous est nécessaire de modéliser les entités que nous voulions reconnaître automatiquement. Notre connaissance du cerveau humain ne nous aide pas beaucoup ici, car si l'analyse du signal effectuée par l'oreille humaine semble plus ou moins connue, il en va tout autrement de ce que fait le cerveau avec les informations reçues par la cochlée, de leur stockage et de leur interprétation.

Les systèmes de reconnaissance automatique de la parole et du locuteur actuels utilisent pour la plupart des algorithmes de comparaison de motifs (« patterns matching » en anglais). Dans le cadre de la reconnaissance de locuteurs, nous modéliserons les différentes prononciations qu'un locuteur peut avoir effectuées pour le même motif. En étudiant la parole de locuteurs sur plusieurs prononciations des mêmes motifs, nous pouvons distinguer des variabilités caractéristiques du signal de parole nous permettant de séparer les locuteurs les

uns des autres (variabilités inter-locuteurs) et d'autres, intrinsèques au locuteur (variabilités intra-locuteur).

De manière à mémoriser des caractéristiques qui dépendent du locuteur, nous utilisons des algorithmes capables de capturer les points communs entre différentes représentations de motifs spectraux issus du même locuteur (constituant ainsi un **modèle du locuteur**), tout en ayant la possibilité de s'adapter aux variations d'échelles fréquentielles et temporelles liées au signal de parole. Ces motifs peuvent être soit des segments de parole déterminés (mots, phonèmes) si nous travaillons en mode dépendant du texte, soit des segments de parole dont on ne connaît pas le contenu phonétique si l'application fonctionne en mode indépendant du texte. Ces algorithmes doivent être couplés avec une mesure qui permettra de donner une valeur de distorsion (ou de similitude) entre le modèle du locuteur et un motif inconnu dont on cherche à déterminer la provenance (Ng, 1995).

Spectres et cepstres moyens : la première approche automatique très largement répandue a été l'utilisation du spectre moyen à long terme. Pour un locuteur de référence donné, on extrait d'une phrase prononcée un ensemble de vecteurs de paramètres (spectraux, cepstraux, ...), et on les modélise par leur vecteur moyen. Chaque locuteur de référence est ainsi modélisé par un spectre (ou cepstre) moyen global, ou à long terme. On calcule alors un vecteur moyen pour le locuteur de test, puis une distance spectrale (ou cepstrale) entre ce vecteur et un vecteur de référence.

La programmation dynamique : la programmation dynamique est une technique exclusivement utilisée en mode dépendant du texte. Elle permet d'aligner temporellement une phrase de test avec une phrase d'apprentissage, ce qui permet de prendre en compte les différences de débit qui peuvent survenir entre deux énoncés d'une même phrase prononcée par un même locuteur. Le premier algorithme de programmation dynamique a été proposé par (Sakoe, 1978). Dans cet article, l'algorithme est appliqué en reconnaissance de mots isolés. La programmation dynamique trouve parfaitement son utilité dans ce type d'application, puisqu'elle permet d'aligner temporellement des mots les uns avec les autres. On doit l'une des premières utilisations de la programmation dynamique en reconnaissance du locuteur à (Furui, 1981a).

La quantification vectorielle : la quantification vectorielle peut être utilisée indifféremment en mode dépendant ou indépendant du texte. En mode dépendant du texte, elle représente une alternative intéressante à la programmation dynamique. Elle a été utilisée un peu plus tard en mode indépendant du texte. Puis elle a été un peu mise à l'écart comme méthode en tant que telle, mais intervient quelques fois dans la phase de paramétrisation, suivie de l'utilisation d'une modélisation statistique. On l'utilise en particulier comme initialisation pour l'algorithme EM, lorsque l'on utilise les mélanges de Gaussiennes.

La quantification vectorielle consiste à représenter l'ensemble des vecteurs de paramètres extraits le long d'une phrase par un petit nombre de vecteurs représentatifs, appelés généralement centroïdes. On appelle dictionnaire l'ensemble des centroïdes extraits le long d'une phrase. Il existe plusieurs algorithmes pour établir ces dictionnaires. Cette méthode est particulièrement avantageuse quand le signal sur lequel on l'applique présente naturellement une structure en « segments », ce qui est justement le cas du signal de parole.

Les modèles de Markov cachés : l'inconvénient majeur de toutes les techniques déjà présentées est quelles ne prennent pas en compte la façon dont les vecteurs de paramètres se succèdent. Les modèles de Markov sont l'une des premières tentatives pour résoudre ce problème. Ce modèle a été initialement introduit en reconnaissance de la parole. Puis son utilisation s'est étendue peu à peu au domaine de la reconnaissance du locuteur (Poritz , 1982).

Un modèle de Markov caché est constitué de plusieurs états, chaque état étant caractérisé par une distribution de probabilité. On connaît en outre les probabilités de passage d'un état à l'autre. Enfin, les vecteurs de paramètres sont en fait les observations de ce modèle probabiliste, c'est-à-dire chaque état possède une densité de probabilité d'émission de ces différents vecteurs de paramètres. On caractérise alors entièrement un modèle de Markov caché par la donnée des différentes probabilités de se trouver à l'instant initial dans chaque état, par la donnée des différentes probabilités de transitions entre les différents états, et par la donnée des différentes densités de probabilités d'émissions.

Les mélanges de Gaussiennes : Les modèles de Markov cachés sont l'une des nombreuses méthodes statistiques pour modéliser les vecteurs de paramètres. Parmi ces méthodes, on trouve aussi la modélisation des vecteurs par une densité de probabilité

Gaussienne multidimensionnelle. L'une des extensions de cette modélisation Gaussienne est la modélisation par un mélange de densités Gaussiennes. Cette technique a été utilisée assez récemment en reconnaissance du locuteur, et elle fournit actuellement les meilleurs résultats en reconnaissance du locuteur indépendante du texte. On utilise en général un algorithme EM pour estimer les différents paramètres du mélange.

Les réseaux de neurones : les réseaux de neurones ont été assez largement utilisés en reconnaissance du locuteur. Ils offrent en effet une bonne alternative au problème de la discrimination entre les locuteurs. Ces outils de classification permettent en effet de séparer des classes, dans un espace de représentation donné, de façon non linéaire.

L'inconvénient important de l'application de cette technique en reconnaissance du locuteur est le coût important dû à l'ajout d'un nouveau locuteur dans la base de référence. On peut aussi utiliser les réseaux de neurones en les couplant à d'autres techniques, comme par exemple les modèles de Markov cachés (Bourlard, 1994). On parle alors de méthodes hybrides.

1.3.3. Modes de décision

La stratégie mise en jeu dans cette partie dépend essentiellement du type d'application : identification ou vérification.

1.3.3.1. Pour l'identification

Dans une application de l'identification du locuteur, la stratégie est assez simple puisqu'il s'agit d'évaluer la similarité des caractéristiques mesurées avec toutes les références correspondant à chacun des locuteurs autorisés. Le locuteur identifié est celui pour lequel la similarité est la plus grande. Notons que le coût de calcul de cette opération d'identification, ainsi que le volume des données qu'il est nécessaire de stocker, croissent linéairement avec la taille du groupe de locuteurs autorisés. La situation est plus complexe lorsqu'on a affaire à un ensemble ouvert car il est en plus nécessaire de rejeter les locuteurs n'appartenant pas au groupe de locuteurs autorisés. En général, la démarche adoptée consiste à effectuer d'abord l'identification, puis à utiliser une stratégie de vérification pour rejeter les éventuels

imposteurs en considérant que l'identité revendiquée est celle déterminée lors de la phase d'identification.

1.3.3.2. Pour la vérification

L'attitude adoptée en général consiste à fixer un seuil sur la mesure de similarité : au dessus, le locuteur est rejeté, en dessous, le locuteur est accepté (comme étant celui dont l'identité est revendiquée) (Furui, 1994). Toutefois, comme le montre de manière très claire (Noda, 1989), l'utilisation d'un seuil fixé, identique pour tous les locuteurs, conduit à des taux d'erreur variables en fonction du locuteur.

Une réponse très courante à ce problème consiste à fixer les seuils individuels de vérification a posteriori. Les seuils sont alors évalués grâce à des tests systématiques avec tous les locuteurs et tous les enregistrements disponibles pour l'apprentissage. On fixe en général le seuil de façon à obtenir des taux de rejets erronés et d'acceptations erronées de même valeur. (Furui, 1981b) présente une stratégie empirique alternative visant à remplacer cette procédure très coûteuse, et surtout peu réaliste compte tenu du fait qu'elle suppose que tous les imposteurs potentiels soient connus.

Récemment, une autre solution a été proposée qui consiste à normaliser les mesures de similarité, le seuil de décision restant lui fixé indépendamment du locuteur (Furui, 1994). L'intérêt principal de cette normalisation est de réduire les différences individuelles intrinsèques qui rendent nécessaires l'usage de seuils individuels. De plus, cette méthode semble aussi limiter les conséquences de la variabilité "accidentelle" introduite par le canal de transmission (Furui, 1994). Le principe de la normalisation de la mesure de similarité consiste à soustraire, à la mesure de similarité obtenue pour le locuteur dont l'identité est revendiquée, une mesure de similarité moyenne obtenue pour un groupe de locuteurs représentatifs (dit cohort speakers). Le débat reste ouvert quant à la manière dont il convient de composer ce groupe représentatif (Furui, 1994). (Matsui, 1994) présente une nouvelle approche, qui semble être assez efficace, consistant à effectuer la normalisation, non plus avec un groupe de locuteurs représentatifs, mais directement avec un modèle représentatif (dans le cadre du modèle de mélange de densités Gaussiennes).

1.4. Evaluation des performances en reconnaissance vocale

Dans le domaine de la reconnaissance du locuteur, une des principales difficultés réside dans l'évaluation de l'efficacité des techniques employées. D'une manière générale, la phase d'évaluation est souvent plus coûteuse, en termes de moyens techniques et de quantité de travail nécessaires, que la phase de mise au point (Doddington et al., 2000).

Il est possible d'évaluer la fiabilité d'une technique par une démarche empirique en constituant une base de données d'enregistrements de parole, puis en effectuant des tests systématiques. L'évaluation empirique constitue une méthode de validation très satisfaisante car elle permet d'obtenir directement une estimation de la fiabilité en situation réelle. Il faut bien avoir conscience du fait que l'évaluation empirique est, en général, une démarche très lourde car l'estimation des performances n'est significative que si le nombre d'enregistrements disponibles est très important. Le dimensionnement et la composition de la base de données utilisée pour l'évaluation empirique doivent en effet vérifier un ensemble de contraintes qui sont liées, soit à des considérations statistiques, soit à la nature du signal de parole.

Tout résultat obtenu à partir d'une série d'expériences ne représente qu'une estimation. Il est nécessairement entaché d'une certaine incertitude. Il est particulièrement important de noter que le nombre d'expériences nécessaires est inversement proportionnel aux taux d'erreur. Pour des systèmes appelés à être très fiable (taux d'erreur de quelques pour-cent), il est donc nécessaire d'organiser un très grand nombre d'essais (de l'ordre de 1000 à 5000). On peut déjà noter à ce stade que la constitution de la base de données implique l'enregistrement et le stockage d'un grand nombre d'enregistrements de parole ce qui constitue déjà en soi en travail considérable.

Pour rendre compte de la variabilité des caractéristiques de la parole il est nécessaire d'enregistrer chaque locuteur en plusieurs occasions afin d'intégrer la variabilité intra-locuteur. Il est fortement conseillé d'enregistrer chaque locuteur lors d'au moins quatre à cinq séances distinctes séparées dans le temps du plus grand délai possible (sur une période de plusieurs mois). Ces contraintes ne doivent pas être sous-estimées car la variabilité intra-locuteur des caractéristiques mesurées influe très notablement sur les performances de la reconnaissance. En particulier, si chaque locuteur est enregistré au cours d'une seule séance, les performances de la reconnaissance se trouvent artificiellement sur-évaluées. Il est par ailleurs indispensable d'enregistrer un nombre suffisant de locuteur. Le résultat de

l'évaluation empirique est d'autant plus significatif que le nombre de locuteurs est assez important et couvre un ensemble suffisamment représentatif de voix. Enfin selon l'application envisagée, il est nécessaire d'enregistrer les locuteurs dans des conditions plus ou moins particulières (à travers le téléphone, en présence de bruit, etc.) ou en simulant artificiellement ces conditions.

La performance des systèmes de vérification du locuteur chute d'une manière dramatique dans des applications réelles malgré le fait qu'elle soit acceptable dans des conditions contrôlées de laboratoire. Les bruits liés à l'environnement ainsi que le canal de transmission sont les principaux facteurs qui dégradent les informations dépendantes du locuteur contenues dans le signal de parole (Drygajlo, 2004).

A l'inverse du système auditif humain, les systèmes classiques de reconnaissance vocale échouent à produire une information fiable sur le signal propre de parole sérieusement affecté par le bruit. Les techniques qui exploitent seulement l'information fiable en utilisant la distribution marginale augmentent d'une manière significative

La performance des systèmes de reconnaissance vocale en comparaison avec les systèmes classiques. Ceci démontre que quand les paramètres retenus par leur fiabilité en temps et en fréquence détiennent une information suffisante sur le locuteur, alors les paramètres non-fiables peuvent être écartés du processus de reconnaissance.

On ne peut pas comparer deux systèmes à partir des seuls taux d'erreur qui dépendent de multiples facteurs. Ainsi, les éléments suivants doivent également être pris en compte :

– **qualité de la parole** : enregistrements en studio ou via le canal téléphonique ; environnement calme ou bruyant ; type de réseau téléphonique,

– **quantité de la parole** : durée de parole pour l'apprentissage des références de chaque locuteur ; durée de parole des sessions de test,

– **variabilité intra-locuteur** : la voix d'un locuteur dépend de son état physique et émotionnel ; de plus, le comportement d'un locuteur se modifie lorsque celui-ci s'habitue à un système,

– population de la base de locuteurs : en identification du locuteur, la taille de la population a une influence directe sur les performances ; la qualité de la population

(proportion hommes/femmes, bonne répartition géographique des locuteurs parlant une même langue) est également un facteur à intégrer,

Intention des locuteurs : la distinction est faite entre les locuteurs coopératifs (qui veulent être reconnus par le système) et les locuteurs non-coopératifs qui modifient leur voix pour ne pas être reconnus (cas de certaines applications judiciaires par exemple) (Drygajlo et al., 2003) Enfin, certains locuteurs imitent la voix d'une autre personne pour être reconnus à sa place : ce sont des imposteurs. A ce propos, lors de l'évaluation d'un système, les imposteurs sont en général d'autres locuteurs de la base de référence ce qui n'est pas très réaliste. En effet, en pratique, un imposteur réel qui tentera d'imiter la voix du locuteur pour lequel il voudra être reconnu, n'existera pas forcément dans la base de référence.

1.5. Conclusion

Ce premier chapitre donne une bonne introduction au domaine de reconnaissance vocale. La taxonomie, les principes de fonctionnement et les différentes techniques les plus utilisées y sont présentés. Il expose aussi les différentes étapes constituant un système RAL.

Chapitre 2 : Paramétrisation du signal parole

2.1. Introduction

La production de la parole est effectuée par deux fonctions mécaniques de bases : la phonation et l'articulation. La phonation est la production du signal acoustique par le mouvement du larynx. L'articulation est la modulation du signal acoustique par les articulateurs (les lèvres et la langue) et la résonance de ce signal dans les cavités (la bouche et le nez) (Figure 2.1) (Meuwly, 2000). L'appareil vocal humain est constitué d'un excitateur, le complexe glotte-cordes vocales, et d'un ensemble de résonateurs de l'appareil phonatoire : le pharynx, la cavité buccale, la cavité labiale, les fosses nasales. Lorsqu'un excitateur entre en vibration, il fournit un signal, dont le résonateur va amplifier certaines composantes. La présence ou l'absence d'obstacles sur le parcours de la colonne d'air modifie la nature du son produit. Le domaine de la phonétique articulatoire distingue les différentes classes de sons en classant ces obstacles éventuels. Pour distinguer la formation d'une voyelle et d'une consonne, il suffit de déterminer si le passage de l'air se fait librement à partir de la glotte ou non. Si tel est le cas, une voyelle est formée, alors que si le passage est partiellement totalement obstrué, c'est une consonne qui est prononcée.

Ce chapitre traite dans un premier la production et l'analyse du signal parole. Il expose les problèmes de variabilités de ce dernier et détaille les différentes modélisations ainsi que les paramétrisations possibles.

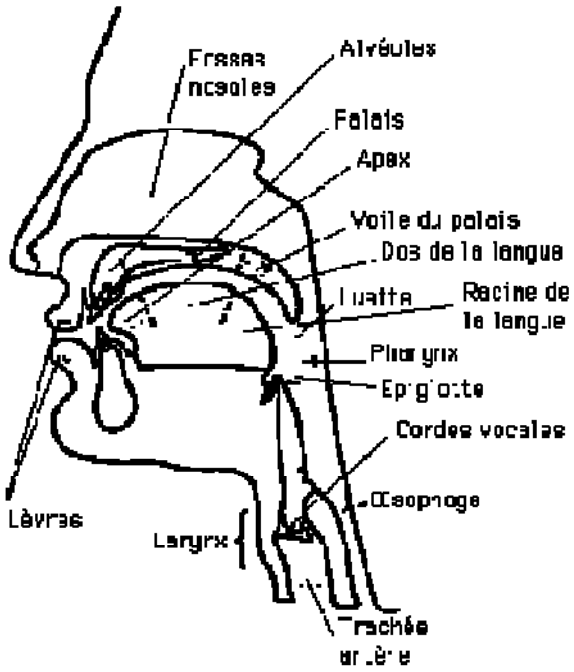

Figure 2.1 : Vue d'ensemble des organes de la parole

2.2. Le signal de la parole

Le signal parole n'est pas un signal ordinaire. Il est le vecteur d'un phénomène complexe : la communication parlée. D'un point de vue mathématiques, il est difficile de modéliser le signal de parole, compte tenu de sa variabilité. Nous allons ici tenter de mettre en évidence quelques caractéristiques importantes de ce dernier.

2.2.1. Redondance du signal

Le signal parole est extrêmement redondant. Cette grande redondance lui confère une robustesse à certains types de bruits. De nombreuses recherches sont menées afin de rendre les systèmes de reconnaissance robustes aux bruits, mais les performances humaines sont encore loin d'être atteintes.

2.2.2. Variabilité du signal

Le signal parole possède une très grande variabilité. Une même personne ne prononce jamais un mot deux fois de façon identique. La vitesse d'élocution peut varier, la durée du

signal est alors modifiée. Toute altération de l'appareil phonatoire peut modifier la qualité de l'émission (exemple : rhume, fatigue…). L'ensemble de ces altérations est connu comme la variabilité intra-locuteur. De plus, la diction évolue dans le temps. La voix est modifiée au cours des étapes de la vie d'un être humain (enfance, adolescence, âge adulte…). La variabilité inter-locuteur est encore plus accentuée. La hauteur de la voix, l'intonation et l'accent différent selon le sexe, l'origine sociale, régionale ou nationale.

2.2.3. Les effets de coarticulation

La production parfaite d'un son suppose un positionnement précis des organes phonatoires. Le déplacement de ces organes est limité par une certaine inertie mécanique. Les sons émis subissent alors l'influence de ceux qui les précèdent ou les suivent. Ces effets de coarticulation est un facteur de variabilité supplémentaire important du signal de parole.

2.2.4. Les facteurs extérieurs

En plus des variabilités déjà discutés, le signal est soumis aux variabilités liées aux facteurs extérieurs tels que les canaux de transmission, le bruit, le type de microphone, …etc. Tous ces facteurs génèrent des altérations en plus qui rendent le signal parole de plus en plus difficile à analyser.

2.3. Les différents paramètres extrait du signal parole

Pour résoudre les problèmes liés à la complexité de la parole, il est possible de calculer des coefficients représentatifs du signal traité. Ces coefficients sont calculés à intervalles temporels réguliers, chose qui est due à la non-stationnarité du signal parole, qui peut l'être sur de courtes durées de 20 à 30 ms. En simplifiant les choses, le signal de parole est transformé en une série de vecteurs de coefficients. Ces coefficients doivent représenter au mieux le signal à modéliser, et extraire le maximum d'informations nécessaires à la reconnaissance. Nous étudierons dans ce paragraphe les coefficients les plus utilisés en reconnaissance de la parole. Nous commencerons par les coefficients issus de l'analyse spectrale. Les coefficients cepstraux sont répandus comme les plus utilisée en reconnaissance

vocale. Suivent après les coefficients de prédiction linéaire et leurs transformations LPC, LPCC, …etc, ainsi que les coefficients issus de l'analyse par banc de filtres MFCC et PLP.

Avant tout calcul, il est nécessaire de mettre en forme le signal de parole. Pour cela, quelques pré-traitements sont effectués:

Le signal est tout d'abord filtré puis échantillonné à une fréquence permise. Suivit par une Pré-accentuation. La pré-accentuation est un exemple d'utilisation de connaissances sur la perception humaine. Elle consiste en un filtrage du signal de parole par le filtre suivant :

$$Y(z) = (1 - az^{-1})\, X(z) \qquad (2.1)$$

Le filtre passe-haut a pour effet de rehausser les composantes spectrales de haute fréquence. C'est un filtre de compensation des effets de filtrage des procédés d'acquisition qui sont assimilables à des filtres passe-bas (Gold et Nelson, 2000). Puis le signal est segmenté en trames. Chaque trame est constituée d'un nombre N fixe d'échantillons de parole. En général, N est fixé de telle manière que chaque trame corresponde à environ 20 ms de parole (durée pendant laquelle la parole peut être considérée comme stationnaire).

Enfin, un fenêtrage est effectué. Parmi les fenêtres utilisées, on peut citer les fenêtres de Hamming, Hanning, Blackman ou de Kaiser. Le choix se fait le plus souvent en fonction de l'application car les fenêtres présentent différentes atténuations à des fréquences bien précises. Cependant, il faut noter que la plupart des systèmes sont directement conçus sur des fenêtres de Hamming afin de limiter les effets du phénomène de Gibss. Ce traitement implique une hypothèse importante du fait des limitations postérieures qu'elle occasionne : Le signal vocal est supposé stationnaire sur une période limitée.

Après cette mise en forme du signal (commune à la plupart des méthodes d'analyse de la parole), une transformée de Fourier (algorithme de FFT Transformée de Fourier Rapide) est appliquée pour représenter le signal dans le domaine fréquentiel. Le spectre de puissance à court terme P(w) est calculé selon:

$$P(w) = Re\,[X(w)]^2 + Im\,[X(w)]^2 = |X(w)|^2 \qquad (2.2)$$

Où X(w) est le spectre de signal temporel et w représente la fréquence angulaire en rad.s^{-1}.

2.3.1. Les principaux paramètres de l'analyse spectrale

2.3.1.1. Coefficients issus d'une analyse par prédiction linéaire

La prédiction linéaire est une technique issue de l'analyse de la production de la parole permettant d'obtenir des coefficients de prédiction linéaire (Linear Prediction Coefficients - LPC). Des paramètres cepstraux LPCC (Linear Prediction Cepstral Coefficients), LAR (Log Area Ratio) et PARCOR sont ensuite calculés à partir de ces coefficients.

2.3.1.1.1. Les coefficients LPC

C'est une méthode d'analyse du signal qui a été largement utilisée dans les systèmes de reconnaissance de la parole. Elle fournit une bonne représentation du signal vocal et permet de déduire les paramètres de bases tels que la fréquence fondamentale, les formants, le spectre, et la fonction de transfert du conduit vocale.

Cette méthode connue sous le sigle LPC (Linear Predictive Coding) est fondé sur les connaissances de la production de la parole et suppose que le modèle de production est linéaire. Le modèle se décompose en une source active et un conduit passif, on modélise l'onde vocale comme la sortie d'un filtre dont la fonction de transfert est

$$H(z)= \frac{S(z)}{U(z)} = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-k}} \qquad (2.3)$$

L'idée de base de codage par prédiction linéaire est qu'un échantillon du signal peut être approximé par une combinaison linéaire des P échantillons précédents. Le système est excité, soit par un train d'impulsions pour générer un son voisé, soit par une séquence aléatoire pour produire un son non voisé. Ainsi, les paramètres du modèle sont l'indicateur du voisement, la période fondamentale, le gain G et les coefficients (a_k) du filtre numérique.

L'équation aux différences caractérisant le système et reliant l'entrée u(n) à la sortie s(n) est donnée par :

$$s(n)= \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \qquad (2.4)$$

Un prédicteur linéaire est défini par les coefficients α_k comme étant un système dont la sortie est :

$$\bar{s}(n)=\sum_{k=1}^{p} \alpha_k s(n-k) \quad (2.5)$$

L'erreur de prédiction e(n) est définie par :

$$e(n) = s(n) - \bar{s}(n) = \quad s(n) - \sum_{k=1}^{p} \alpha_k s(n-k) \quad (2.6)$$

Où les constantes $\alpha_k$sont les coefficients de prédiction, P l'ordre du prédicteur, et e(n) l'erreur de prédiction.

L'équation ci-dessous, permet de conclure que l'erreur de prédiction est la sortie d'un système dont la fonction de transfert est :

$$A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \quad (2.7)$$

En comparant les équations (2.4) et (2.6), nous pouvons conclure que si le signal vocal obéit au modèle (2.4), et que si $\alpha_k = a_k$, nous aurons donc *e(n)=G.u(n).* Dans ce cas *A(z)* est un filtre inverse du système et H(z) de l'équation (2.3) est identique à :

$$A(z)=G/A(z) \quad (2.8)$$

Cette méthode d'analyse est directement issue du modèle de production de la parole de la figure ci-dessous :

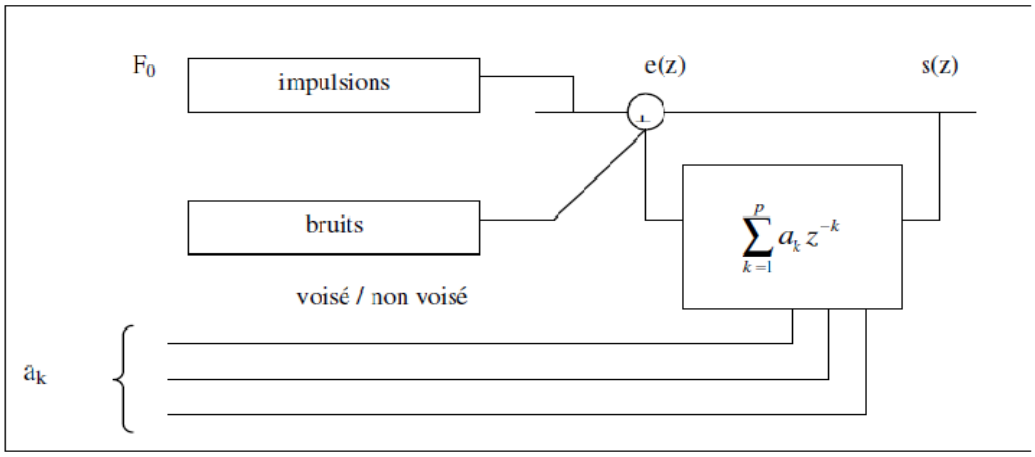

Figure 2.2 : Modèle de phonation

On voit sur ce modèle que l'erreur de prédiction e(n) représente à la synthèse l'excitation du conduit vocal. En fait, à l'analyse, e(n) contient l'information relative à l'excitation et les erreurs dues au modèle (nombre de coefficients insuffisants, précision des calculs, ..).

Les coefficients de prédiction sont choisis de façon à minimiser l'erreur quadratique moyenne de prédiction linéaire En définie par :

$$En = \sum_{k=1}^{p} e_n^2(m) \tag{2.9}$$

Minimiser l'erreur quadratique moyenne revient à annuler son gradient par rapport aux coefficients α_k .

Posant:

$$\Phi_n(i, k) = \sum_n s(m+n-i)(m+n-k) \tag{2.10}$$

$$\sum_{k=1}^{p} \alpha_k \Phi_n(i, k) = \Phi_n(i, 0) \qquad \text{avec } i=1,.., p \tag{2.11}$$

Pour avoir les coefficients $\alpha_k$de prédiction optimaux, il suffit de résoudre le système de p équations à p inconnues (2.11) minimisant l'erreur quadratique moyenne.

2.3.1.1.2. Les coefficients cepstraux (LPCC)

Les coefficients cepstraux (en anglais, « Cesptral coefficients ») proviennent de l'analyse homomorphique du signal de parole. Cette technique de traitement non-linéaire du signal est appropriée au signal de parole. En effet, la parole est la convolution temporelle de la réponse impulsionnelle du conduit vocal et de la fonction d'excitation.

Cette convolution devient une multiplication dans le domaine fréquentiel. Si le logarithme du spectre est considéré, cette multiplication devient une addition. Étant donné que l'oreille humaine est pratiquement insensible à la phase relative entre deux composantes sonores, on peut seulement se limiter au module

$$\log\left(\left|S(e^{jw})\right|\right)=\log\left(\left|H(e^{jw})\right|\right)+\log\left(\left|U(e^{jw})\right|\right) \tag{2.12}$$

où $S(e^{jw})$ est le spectre de la parole, $H(e^{jw})$ est le spectre du conduit vocal et $U(e^{jw})$ est le spectre de l'excitation. La transformée de Fourier inverse du logarithme de la norme du spectre du signal de parole définit le cepstre réel (Bogert, 1963). Ce cepstre se compose des coefficients, appelés coefficients cepstraux. La référence (Furui, 1989) montre qu'il existe des relations récursives entre ces coefficients et les coefficients de prédiction :

$$\begin{aligned}
c_0 &= E(0) = \phi(0)\\
c_1 &= -a_1\\
c_i &= -a_i - \sum_{k=1}^{i-1} \frac{i-k}{i} . c_{i-k} . a_k, \quad 1 < i \le P\\
c_i &= -\sum_{k=1}^{P} \frac{i-k}{k} . c_{i-k} . a_k, \quad 1 > P
\end{aligned} \tag{2.13}$$

Le premier coefficient, c_0, représente l'énergie de la tranche de parole analysée. Ces coefficients cepstraux présentent plusieurs propriétés. Ils permettent de calculer la distorsion spectrale entre deux filtres autorégressifs par une simple distance euclidienne. De plus, les coefficients CMS (de l'anglais, « Cesptral Mean Substraction ») dérivés des coefficients cepstraux sont insensibles aux distorsions linéaires engendrées par le canal de transmission ou par le microphone (Mammone, 1996).

Les coefficients CMS sont les coefficients cepstraux centrés sur leurs moyennes respectives :

$$c_i^{cms} = c_i - E(c_i) \tag{2.14}$$

Ces propriétés justifient l'utilisation des coefficients cepstraux dans les méthodes de reconnaissance du locuteur.

2.3.1.2. Coefficients issus d'une analyse par banc de filtres

L'analyse par banc de filtres est une technique initialement utilisée pour le codage du signal de parole. Le signal de parole est analysé à l'aide des filtres passe-bande permettant d'estimer l'enveloppe spectrale en calculant l'énergie dans les bandes de fréquences considérées.

Les bandes de fréquences des filtres sont espacées logarithmiquement selon une échelle perceptive afin de simuler le fonctionnement du système auditif humain. Les échelles perceptives les plus utilisées sont celles de Mel et de Bark. Plus la fréquence centrale du filtre est basse, plus la bande passante du filtre est étroite. Augmenter la résolution pour les basses fréquences permet d'extraire plus d'information dans ces zones où elle est plus dense.

2.3.1.2.1. Les coefficients MFCC

Le codage MFCC (Mel Frequency Cepstral Coding) est sûrement la technique de codage la plus utilisée en traitement de la parole. C'est une représentation que l'on retrouve dans des applications très diverses comme la reconnaissance du locuteur, la reconnaissance de la parole ou bien de la langue ou encore dans la discrimination parole/musique (Ezzaidi, 2002).

Le codage MFCC intègre deux notions importantes. La première est la notion de bancs de filtres qui modélisent la membrane basilaire .Ces bancs de filtres sont déployés non pas sur une échelle en Hertz mais sur une échelle non-linéaire : l'´echelle Mel. Cette échelle est issue de connaissances sur la perception humaine. La résolution perceptive des fréquences diffère selon que l'on écoute des sons de basses ou hautes fréquences.

Le procédé d'extraction de caractéristiques des coefficients MFCC consiste à la projection des log énergies du banc de filtres déployés sur une échelle Mel. La projection est réalisée par la projection en cosinus discrète (DCT : Discrete Cosinus Transform) (cf. figure 2.3). Le codage MFCC représente la référence des procédés d'extraction de caractéristiques. Toutes les méthodes proposées doivent donc se comparer au codage MFCC. Une des raisons de ce succès est la décorrélation des coefficients produit par la projection DCT.

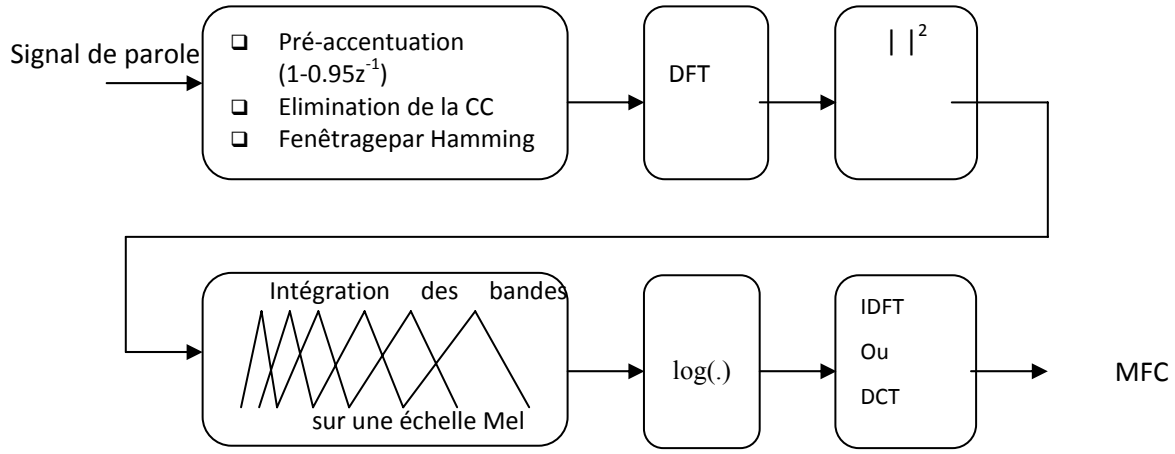

Figure 2.3 : Calcul des coefficients MFCC.

2.3.1.2.2. Les coefficients PLP

L'analyse par Prédiction Linéaire Perceptuelle (Perceptual Linear Prediction -PLP) repose sur un modèle de perception de la parole. Les différentes étapes de l'analyse PLP sont détaillées dans la figure (2.5).

Elle est basée sur le même principe que l'analyse prédictive et intègre trois caractéristiques de la perception (Hermansky, 1990) :

Intégration des bandes critiques : la prédiction linéaire produit la même estimation de l'enveloppe spectrale pour toute la zone de fréquences utiles, ce qui est en contradiction avec le fonctionnement de l'appareil perceptif humain. En effet, l'oreille humaine a la faculté d'intégrer certaines zones de fréquences en bande appelées bandes critiques. Les bandes critiques sont réparties selon l'échelle de Bark, dont la relation avec la fréquence est définie par :

$$f = 600 \sinh(z/6) \qquad (2.15)$$

Avec f la fréquence en Hertz et z la fréquence en Bark. La nouvelle densité spectrale est échantillonnée selon cette nouvelle échelle, ce qui augmente la résolution pour les basses fréquences.

Préaccentuation pas courbe d'isosonie : cette caractéristique provient de la psychoaccoustique qui a montré que l'intensité sonore d'un son pur perçue par l'appareil auditif varie avec la fréquence de ce son. Ainsi, dans l'analyse PLP, afin de prendre en compte la manière dont l'appareil auditif perçoit les sons, la densité spectrale doit être multipliée par une fonction de pondération non linéaire. Cette fonction peut être estimée en utilisant l'abaque sur laquelle sont reportées les lignes isosoniques (figure 2.4). Ces lignes correspondent à la trajectoire d'égale intensité sonore pour différentes fréquences d'un son pur. En pratique, cette préaccentuation est remplacée par l'application du filtre passe-haut dont la transformée en Z est $(1 - 0.95z{-}1)$.

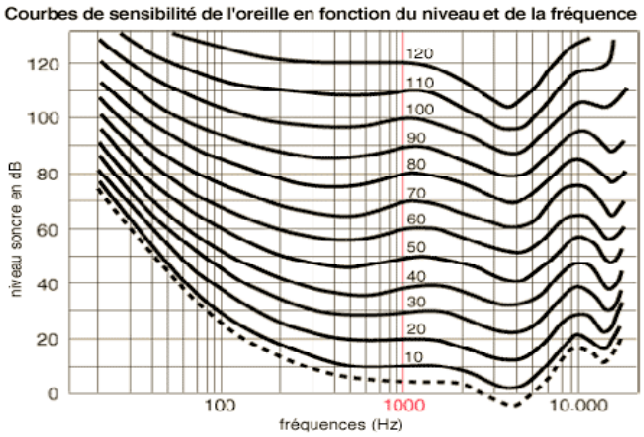

Figure 2.4 : Abaque des courbes d'égale intensité sonore (Stevens, 1957).

Loi de Stevens : l'intégration des bandes critiques et la pré-accentutation ne suffisent pas à faire correspondre l'intensité mesurée et l'intensité subjective (appelée sonie). La loi de Stevens donne la relation entre ces deux mesures

$$\text{sonie} = (\text{intensité})^{0.33} \qquad (2.16)$$

Les PLP sont basés sur le spectre à court terme du signal de parole, comme les coefficients LPC. Cela signifie que le signal est analysé sur une fenêtre glissante de courte durée. En général, on utilise une fenêtre de longueur 10 à 30 ms. que l'on décale de 10 ms pour chaque trame.

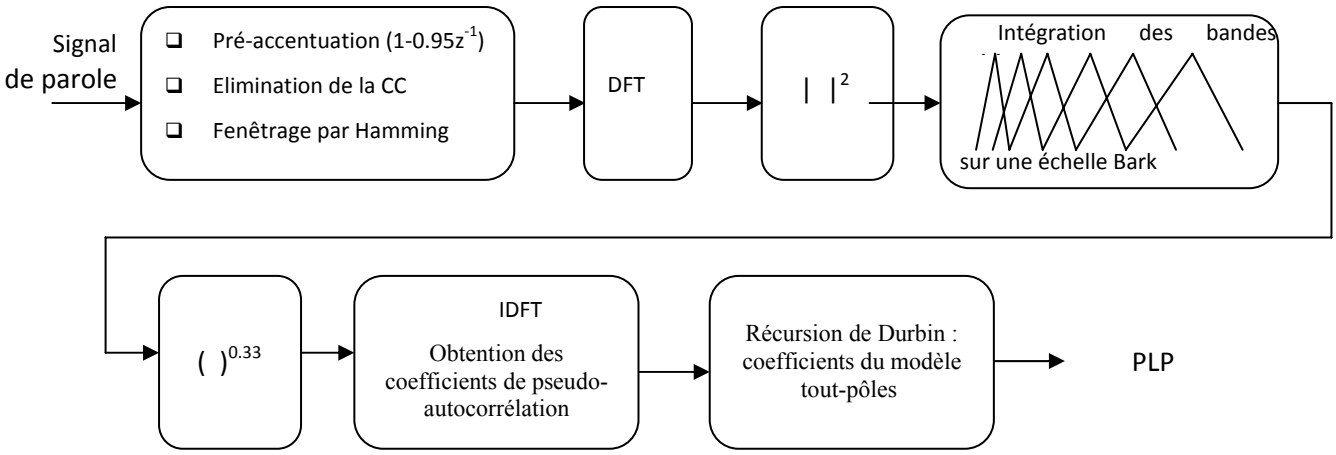

Figure 2.5 : Calcul des coefficients PLP.

L'oreille humaine est également plus sensible aux variations relatives de la valeur d'un signal acoustique qu'à ses variations absolues. L'intégration de cette nouvelle propriété dans la technique PLP a conduit à l'élaboration des paramètres RASTA-PLP (en anglais « RelAtive SpecTrAl ») (Hermansky, 1991). Elle consiste en un filtrage passe-haut des sorties d'un banc de filtres dans le domaine du logarithme du spectre afin de supprimer les variations lentes du signal, puis à appliquer un opérateur exponentiel pour retourner dans le domaine du spectre de puissance. Ces paramètres bénéficient de très bonnes propriétés de robustesse qui permettent de compenser des variations de microphone, mais ne luttent pas contre les perturbations provoquées par un bruit. Pour cela, (Hermansky, 1993) a développé un filtrage passe-bande d'une fonction du spectre, égale à log(1+Jx), approximativement logarithmique lorsque l'amplitude x du spectre est élevée, et approximativement linéaire lorsque l'amplitude du spectre est faible. Cette méthode appelée J-RASTA, permet de compenser principalement les effets du bruit lorsque x est faible, et les effets d'un filtrage linéaire lorsque x est élevé.

2.3.2. Paramètres prosodiques

Le terme « paramètres prosodiques » réunit l'énergie, la durée et la fréquence fondamentale (ou pitch) (Atal, 1972). Ces paramètres s'avèrent cependant fragiles en pratique et ne permettent pas, à eux seuls, de discriminer les locuteurs. En conséquence, il sont souvent associés aux paramètres de l'analyse spectrale.

2.3.2.1. L'énergie

L'énergie du signal est un indice qui peut par exemple contribuer à la détection du voisement d'un segment de parole. L'énergie totale E_0 est calculée directement dans le domaine temporel sur une trame de signal s(n) avec n entre 0 et N-1comme :

$$E_0 = \frac{1}{N}\sum_{n=0}^{N-1} s(n)^2 \qquad (2.17)$$

L'énergie ainsi obtenue est sensible au niveau d'enregistrement; on choisit en général de la normaliser, et d'exprimer sa valeur en décibels par rapport à un niveau de référence.

2.3.2.2. Fréquence fondamentale (ou pitch)

Le pitch est un paramètre très important pour l'étude acoustique et la synthèse de la parole, l'oreille est très sensible à ses variations lesquelles constituent la prosodie, l'évolution de la fréquence en fonction du temps au niveau du phonème constitue la micromélodie, par contre son évolution au niveau des groupes syntaxiques de la phrase est la macromélodie, l'intonation du message est directement liée au pitch. Plusieurs techniques en vue de l'extraction du fondamental peuvent être employées, parmi celles-ci on peut citer la méthode d'AMDF (en anglais « Autocorrelation Mean Difference Function ») qui est donnée par :

$$AMDF(k) = \frac{1}{N}\sum_{n=0}^{N-1} \left|a(n) - a(n-k)\right| \qquad (2.18)$$

Cette fonction présente un minimum aux multiples de la période fondamentale.

D'une manière générale, les méthodes d'estimation de pitch comportent trois étapes :

- Le prétraitement du signal parole pour l'adaptation du signal (filtrage, fenêtrage, pré-accentuation, …).
- Le traitement pour l'extraction de la fréquence fondamentale.
- Le post-traitement pour corriger les erreurs de calcul surtout pour les transitions voisé/non voisé.

2.4. Conclusion

Dans ce chapitre, nous avons passé en revue les deux types de modélisation d'un signal parole, à savoir : l'analyse spectrale par prédiction linéaire et l'analyse par banc des filtres. Il introduit quelques types de paramètres issus des deux types de modélisation tels que les LPC, LPCC, MFCC et PLP. Il introduit aussi les paramètres prosodiques : énergie et pitch.

Chapitre 3 : Extraction et sélection des paramètres

3.1. Introduction

L'objectif de l'extraction des caractéristiques est d'estimer les paramètres représentants la spécificité du locuteur, qui est la combinaison des différences physiques du système vocal et habitudes linguistiques ainsi que du style de parler. En conséquence, les informations spécifiques au locuteur contenues dans le signal de parole peuvent être classées en deux catégories : 1) l'information de bas niveau, qui est liée à la structure anatomique de l'appareil vocal et 2) l'information de haut niveau, qui est liée aux habitudes linguistiques et aux styles. Le tableau 3.1 présente les caractéristiques hiérarchiques pour la reconnaissance du locuteur par l'homme et la machine.

Tableau 3.1 Caractéristiques hiérarchiques pour la reconnaissance du locuteur par l'homme et la machine.

Affiliation physique/Sociale	Indices perceptifs pour l'homme	Caractéristiques pour la RAL	Faisabilité dans la RAL
Status socio-économique, éducation, lieu de naissance, …etc.	Accent, diction, sémantique, idiosynchrasies, …etc.	Mots, phrase, syntaxe, …etc.	Caractéristiques de haut-niveau, la représentation effective est en attente
Type de personnalité, l'influence parentale, …etc.	Style de parler, prosodie, rythme, intonation, modulation du volume, débit de parole, …etc.	Contour de F_0, les fluctuations d'énergie, les pauses, les durées, …etc.	Caractéristiques de niveau modéré, elles sont utilisées pour compléter les caractéristiques bas-niveau
Structure anatomique de l'appareil vocal	Les aspects acoustiques de la parole, les nasales, profondeur, souffle, raideur, la dureté, …etc.	F_0, les harmoniques de l'enveloppe spectrale, énergie, …etc.	Caractéristiques bas-niveau, largement et effectivement dans les systèmes RAL actuels.

Les caractéristiques acoustiques représentants les informations bas-niveau ont été largement appliquées dans la reconnaissance de la parole et la reconnaissance du locuteur. Ces caractéristiques bas-niveau révèlent les configurations du conduit vocal liées à la parole/au locuteur. Les caractéristiques bas-niveau les plus répandues sont celles basées sur l'analyse cepstrale de la parole, tels que les Coefficients Cepstraux issus de la Prédiction Linéaire (LPCC) (Atal, 1974 ; Furui, 1981) et les Coefficients Cepstraux à Fréquence Mel (MFCC) (Davis et Mermelstein, 1980). D'autres caractéristiques qui visent à capter les informations liées aux vibrations des cordes vocales, telles que la fréquence fondamentale (F_0) (Atal, 1972 ; Harrag et al., 2005 ; Sonmez et al., 1998) et les informations de l'intensité des harmoniques (Imprel et al., 1997) ont été utilisées. Contrairement à la reconnaissance de la parole, où l'on pense que les différences sont principalement liées à la structure des formants du système du conduit vocal, dans la reconnaissance du locuteur un certain nombre d'expérimentations ont montré que le style de vibration des cordes vocales comporte des informations riches en spécificité du locuteur et sont utiles pour sa reconnaissance. Cette thèse se concentre sur le développement de techniques efficaces pour extraire l'information spécifique au locuteur issue de l'excitation de la source vocale pour améliorer la performance du système de reconnaissance du locuteur conventionnel utilisant uniquement les caractéristiques du conduit vocal.

La sélection est la transformation des vecteurs de caractéristiques pour réduire la dimension tout en conservant les informations pertinentes. Ceci est particulièrement nécessaire pour des applications réelles où les données d'apprentissage disponibles sont généralement limitées. Une technique utile pour réduire la dimension du vecteur des caractéristiques est l'Analyse en Composantes Principales (ACP) (Jollife, 2002). Dans l'ACP, le vecteur de caractéristiques original est transformé dans un autre espace de représentation avec des coordonnées orthogonales. La sélection des caractéristiques est basée sur les vecteurs propres de la matrice de covariance des données fournies. En fait, les composantes dans l'espace orthogonal correspondent aux plus grandes valeurs propres restantes, tandis que celles correspondantes aux petites valeurs propres sont rejetées. Ainsi, les vecteurs de caractéristiques transformés retiennent les informations les plus importantes, donnant une représentation optimale des caractéristiques originales. En outre, l'orthogonalité entre les composantes des caractéristiques est particulièrement adaptée pour la modélisation des données par distribution gaussienne multi-variables à covariance diagonale, qui est une hypothèse nécessaire à la modélisation des données et à l'estimation des paramètres.

Une autre technique largement utilisée et l'Analyse Discriminante Linéaire (ADL)(Mclachlan, 1992). La sélection des caractéristiques par ADL est basée sur un critère discriminant. Seulement les composantes des caractéristiques avec les plus larges variations inter-classes et des faibles variations intra-classes sont retenues. Ces critères discriminants sont particulièrement compatibles avec la reconnaissance du locuteur, qui est un problème de discrimination plutôt qu'un problème de représentation. Un certain nombre d'articles ont démontré l'application de la méthode ACP et l'ADL pour la sélection des caractéristiques dans les domaines de la reconnaissance de la parole et la reconnaissance du locuteur (Jin et Waibel, 2000 ; Thyes et al., 2000).

3.2. Sélection et Extraction des caractéristiques

Les front-ends des systèmes de reconnaissance de la parole et du locuteur utilisent des caractéristiques spectrales court-terme, car non seulement elles sont porteuses de la distribution fréquentielle qui permet d'identifier les sons, mais aussi les informations liées à la source glottale et à la forme et la longueur du conduit vocal, qui sont des informations spécifiques au locuteur. Selon le front-end des caractéristiques concaténées, les vecteurs de caractéristiques qui en résultent peuvent avoir une dimension de 20 à 50 paramètres.

Dans des applications temps réelles utilisant des dispositifs à faibles ressources, par exemple, les services d'accès par téléphone portable ou des dispositifs embarqués avec de faibles tailles de stockage et de faibles capacités de calcul, un vecteur de caractéristiques avec 50 paramètres ne semble pas approprié, ce qui nécessite une réduction du jeu des paramètres.

Le problème de l'extraction des caractéristiques est parfois établi comme une transformation linéaire qui projette les vecteurs de caractéristiques sur le sous-espace transformé défini par les directions concernées. Etant donné un vecteur de caractéristiques X d'une dimension D, une matrice K*D est appliquée pour obtenir un vecteur Y des caractéristiques transformées de dimension K (K<D). La matrice est estimée de sorte que, du point de vue de la classification, la redondance est supprimée et les caractéristiques transformées ne retiennent que les informations pertinentes, ce qui a pour effet, en théorie, d'optimiser les performances pour les valeurs cibles de K, et devrait surpasser les performances des caractéristiques de base, vu qu'on a supprimé les éléments nuisibles ou qui prêtent à confusion et, plus probablement, de mieux estimer le modèle de paramètres (plus

robuste). Plusieurs méthodes d'extraction sont discutées dans la littérature de la reconnaissance des formes, parmi elles:

3.2.1 La recherche exhaustive

La recherche exhaustive est une méthode optimale pour la sélection d'un sous-ensemble de caractéristiques d'une dimension k parmi l'ensemble de caractéristiques de dimension plus grande K. La recherche exhaustive considère toutes les combinaisons possibles de (k, K). Une implémentation de ce type de recherche nécessite une énorme quantité de calcul, à savoir :

$$\binom{\mathbf{K}}{\mathbf{k}} = \frac{\mathbf{K!}}{\mathbf{k!(K-k)!}}\mathbf{recherches} \qquad (3.1)$$

Par exemple, pour k = 20 et K = 50, le nombre de recherches est d'environ $4.712x10^{13}$. Par conséquent, il est nécessaire de trouver d'autres procédures plus efficaces pour éviter ce type de recherche.

3.2.2. Meilleure Caractéristique Individuelle

Connue en anglais sous le nom de « Best Individual Feature » (BIF), la performance de classification de chaque caractéristique est calculée séparément, c'est-à-dire, sur une base individuelle, et les caractéristiques donnant lieu au plus haut taux de reconnaissance sont sélectionnées. Le meilleur sous-ensemble de caractéristiques k est composé des meilleurs éléments k considéré un à un. Cependant, un ensemble des meilleures caractéristiques k prise une à une n'est pas forcément le meilleur ensemble de caractéristiques k.

3.2.3. Algorithme de recherche séquentielle

Le célèbre algorithme connu en anglais sous le nom « Sequential Forward Search » (SFS) et son homomogue (SBS) (B pour backword) (Withney, 1997) sont des méthodes qui obtiennent une chaine de sous-ensembles de caractéristiques imbriquées d'une manière directe, soit par l'addition (soustraction dans le cas du SBS) de la meilleure (la mauvaise) caractéristique dans l'ensemble. Cet effet de nidification constitue un des principaux

inconvénients de ces méthodes. Les deux algorithmes ne peuvent corriger des additions (soustractions) précédentes de caractéristiques.

Dans la méthode SFS, les caractéristiques sont sélectionnées successivement en ajoutant la meilleure caractéristique locale, la caractéristique qui offre la meilleure information discriminante incrémentale au sous-ensemble des caractéristiques existantes. La technique SFS agit comme la technique BIF en identifiant la première caractéristique ayant le plus haut pouvoir discriminant. Elle procède toutefois, en ajoutant successivement au sous-ensemble sélectionné, les éléments qui contribuent le plus à la performance de classification au-dessus de ceux déjà sélectionnés. Ainsi, à partir d'une caractéristique singulière BIF, le sous ensemble SFS passe une paire, puis un triplet et ainsi de suite.

3.2.4. Analyse en composantes principales

L'Analyse en Composantes Principales (ACP) est une technique ancienne de l'analyse statistique multi-variable (Jolliffe, 2002). Dans cette méthode, la réduction de la dimension est obtenue par la projection de l'espace d'origine caractérisé par D paramètres vers un espace caractérisé par un sous ensemble caractérisé par d paramètres et qui préserve au mieux l'information contenue dans l'espace d'origine. La première composante principale est la projection des données dans la direction de la plus grande variance qui contient le plus d'information. Cependant, cette méthode souffre de plusieurs problèmes décrits ci-dessous :

- Les composantes principales sont variables suivant l'échelle des paramètres. Si un paramètre dans le vecteur d'entrée est multiplié par une grande constante, sa variance sera très grande et par conséquent la première composante sera approchée par ce paramètre.
- L'ACP est basée sur la plus grande matrice de covariance des données, c'est-à-dire la matrice est calculée sans avoir d'information concernant les différentes classes existantes.

Les composantes principales définies par les variations maximales n'impliquent pas que ces dernières contiennent plus d'information qui aide à la discrimination des classes. La Figure 3.1 explique ce dernier problème. On voit bien que la direction de la séparation maximale des classes est perpendiculaire à la première composante principale.

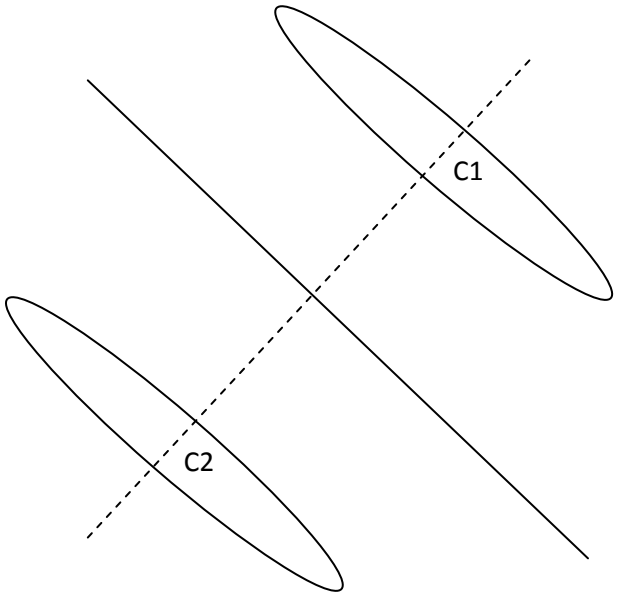

Figure 3.1 Classes Gaussiennes avec des matrices de covariance égales.

3.2.5. L'analyse discriminante linéaire

L'Analyse Discriminante Linéaire (ADL) est une technique statistique utilisée dans la reconnaissance des formes (Duda et al., 2000). Supposons que l'espace acoustique est réparti en un ensemble de classes, chaque classe étant représentée par une gaussienne de matrice de covariance W. Supposons que toutes les classes aient la même matrice de covariance. La technique (ADL) consiste à appliquer une transformation (rotation + dilatation) à tout l'espace acoustique de manière à rendre la variance intra-classe égale à l'identité. En supposant que les nouveaux centres des classes soient distribués selon une loi gaussienne de matrice de covariance inter-claases B, une deuxième rotation est appliquée à l'espace acoustique de manière à dégager les axes principaux. Ces axes sont les vecteurs propres de la matrice B. Ainsi un sous-ensemble de ces axes correspondant aux directions des grandes variances (les grandes valeurs propres) est utilisé pour former les nouveaux vecteurs de paramétrisation du signal.

Elle est équivalente à l'obtention d'une transformation linéaire qui maximise les mesures de discrimination des classes J_1 et J_2.

$$\begin{cases} \mathbf{J_1} = \mathbf{tr}(\mathbf{W^{-1}}.\mathbf{B}) \\ \mathbf{J_2} = (\mathbf{W^{-1}}.\mathbf{B}) \end{cases} \quad (3.2)$$

Plusieurs groupes de chercheurs (Aubert, 1993 ; Siohan, 1995) ont montré que l'analyse discriminante linéaire permettait d'améliorer les performances des systèmes de reconnaissance ainsi que leur robustesse à certains types de bruits. On peut montrer que l'optimisation des critères J_1 et J_2 conduit aux mêmes paramètres discriminants et que cette optimisation est indépendante du choix des matrices de variance.

L'analyse discriminante linéaire peut être résumée comme une procédure en deux phases : 1) dans la première phase, une fonction de normalisation dépendante des classes recueille les informations statistiques, 2) et dans la phase deux, une fonction de discrimination est dérivée des classes afin que les éléments résultants par (ADL) sont moins corrélés, et classés en fonction d'un critère d'objectif.

3.2.6. Les algorithmes génétiques

L'Algorithme Génétique (AG)(Holland, 1975) constitue une autre et nouvelle approche pour la recherche des caractéristiques, car elle permet une recherche aléatoire guidée par une mesure de fitness. Les AGs sont une classe de méthodes de recherche profondément inspirée du processus naturel d'évolution. A chaque itération de l'algorithme (correspondant à une génération), un nombre fixe (population) de solutions possibles (chromosomes) est générée par l'application de certains opérateurs génétiques dans un processus stochastique guidé par une mesure de fitness. Les plus importants opérateurs génétiques et couramment utilisés sont la recombinaison, le croisement et la mutation. Le résultat est un algorithme probabiliste qui a obtenu de bonnes, presque optimales, solutions aux problèmes dans lesquels les méthodes classiques ont échoué ou ne sont pas applicables. Un AG particulier est identifié par une forme particulière du codage des solutions en chaines d'un certain alphabet (généralement binaire), une forme particulière des opérateurs génétiques adoptés et une définition particulière de la fonction de fitness ou fonction d'objectif.

3.3. Analyse et sélection des paramètres

Il est difficile de juger a priori de l'efficacité des paramètres issus d'une analyse. Une première possibilité est de comparer des taux de reconnaissance obtenus avec un système de classification commun aux différents paramètres. Mais, dans ce cas, les résultats peuvent être biaisés par l'adéquation du système de classification à l'un ou l'autre des paramètres. Une

autre possibilité est d'estimer le degré de séparation des locuteurs dans l'espace de paramètres.

3.3.1. Sélection par F-ratio

Une mesure qui peut être utilisée pour évaluer un paramètre particulier est le F-ratio (Campbell, 1997). Il est défini comme le rapport entre variance inter-classe et variance intra-classe.

$$\mathbf{F-ratio} = \frac{\sigma^2_{\mathbf{inter}}}{\sigma^2_{\mathbf{intra}}} \tag{3.3}$$

Dans le cadre de sélection des paramètres qui caractérisent au mieux les locuteurs, le F-ratio peut être utilisé pour sélectionner les paramètres qui maximisent la séparation des classes de locuteurs et minimisent la dispersion à l'intérieur de chaque classe. En utilisant le F-ratio, les assomptions suivantes sont faites :

- Le vecteur des paramètres dans chaque classe a une distribution Gaussienne.
- Les paramètres sont non corrélés.
- Les variances dans chaque classe sont égales.

En pratique, les conditions précédentes sont rarement satisfaites, et dans ce cas on ne peut pas évaluer plus d'un seul paramètre à la fois, vu que généralement dans la parole, les paramètres sont souvent corrélés. Pour remédier à ce dernier problème, on peut transformer l'espace des paramètres corrélés en un espace de paramètres non corrélés et ceci en utilisant par exemple une analyse en composantes principales.

3.3.2. Sélection basée sur les performances de reconnaissance

Cette méthode consiste à calculer la contribution de chaque paramètre dans les performances du système de reconnaissance, c'est-à-dire le taux d'erreur, et à utiliser cette

dernière pour sélectionner ce paramètre ou non. Paliwal (Paliwal, 92) a utilisé chaque paramètre pour reconnaître toutes les réalisations du corpus d'apprentissage (Tableau 3.2).

Tableau 3.2 Coefficients LPCC rangés en utilisant plusieurs critères.

F-ratio		**Taux d'erreur**	
1 à 19	20 à 38	1 à 19	20 à 38
C_2	Δ^2C_1	ΔE	C_8
ΔE	ΔC_9	Δ^{2E}	Δ^2C_6
C_1	ΔC_{11}	C_4	Δ^2C_4
C_3	ΔC_8	C_6	Δ^2C_1
C_4	ΔC_{10}	C_1	C_9
C_{10}	Δ^2C_3	ΔC_4	ΔC_8
C_{11}	ΔC_6	C_5	Δ^2C_5
ΔC_2	Δ^2C_2	ΔC_6	Δ^2C_7
C_5	ΔC_7	ΔC_5	Δ^2C_9
C_8	ΔC_{12}	C_2	ΔC_{10}
ΔC_1	Δ^2C_4	ΔC_3	Δ^2C_8
ΔC_3	Δ^2C_{10}	C_3	C_{11}
C_9	Δ^2C_9	C_7	ΔC_{11}
Δ^2E	Δ^2C_5	ΔC_2	C_{10}
C_7	Δ^2C_{11}	ΔC_1	Δ^2C_{10}
C_6	Δ^2C_8	ΔC_7	C_{12}
C_{12}	Δ^2C_7	ΔC_9	ΔC_{17}
ΔC_5	Δ^2C_6	Δ^2C_3	Δ^2C_{12}
ΔC_4	Δ^2C_{12}	Δ^2C_2	Δ^2C_{12}

Cela est équivalent à faire tourner le système de reconnaissance N fois, avec N le nombre de paramètres dans l'espace d'origine, puis ranger les paramètres selon leurs performances et ne maintenir qu'un sous ensemble qui contient les premiers coefficients qui correspondent à ceux qui ont un faible taux d'erreur. Cette méthode a l'‘avantage de considérer les paramètres individuellement ce qui assure l'indépendance (non corrélation des paramètres). Cependant,

les résultats obtenus peuvent dépendre du système de reconnaissance utilisé et non pas des paramètres choisis.

3.4. Conclusion

Ce chapitre donne quelques méthodes de sélection et d'extraction des paramètres (ACP, ALD, SFS, génétique, …etc.) sans s'attarder sur une méthode en particulier. Il développe ensuite les deux techniques d'analyse de la pertinence des paramètres : analyse par F-ratio et analyse par calcul du taux de reconnaissance. Le chapitre suivant sera consacré aux algorithmes génétiques en donnant le principe, la taxonomie puis une première application qui sert d'un tremplin pour la mise en place de l'algorithme utilisé dans notre travail.

Chapitre 4 : Les algorithmes génétiques

4.1. Introduction

Les algorithmes génétiques (AG) sont des méthodes utilisées dans les problèmes d'optimisation. Les AG tirent leur nom de l'évolution biologique des êtres vivants dans le monde réel. Ces algorithmes cherchent à simuler le processus de la sélection naturelle dans un environnement défavorable en s'inspirant de la théorie de l'évolution proposée par C. Darwin.

Cette méthode a été mise œuvre par J.H.Holland dans les années 70 il introduit le premier modèle formel des algorithmes génétiques (*the canonical genetic algorithm AGC*) dans son livre *Adaptation in Natural and Artificial Systems*. Il expliqua comment ajouter de l'intelligence dans un programme informatique avec les croisements (échangeant le matériel génétique) et la mutation (source de la diversité génétique). Les Algorithmes génétiques (AG) représentant une stratégie de recherche réalisent un compromise équilibre entre l'exploration de l'espace de recherche et l'exploitation des meilleures solutions des analyses théoriques ont montre que les Algorithmes génétiques gèrent ce compromise de façon optimale, (Renders, 1995).

4.2. Définition

L'optimisation par algorithme génétique prend son origine dans les mécanismes de la sélection naturelle et la génétique de l'évolution. Comme son nom l'indique, elle est basée sur la traduction mathématique descensionnels naturels qui sont la reproduction des espace, la suivie et l'adaptation des individus, cette traduction est exploitée pour la résolution de problèmes nécessitant l'optimisation d'une fonction ou d'un système dépendant de plusieurs paramètres et qui ont besoin d'être calculés pour un critère bien défini (maximisation, minimisation,….)

Cette technique constant une méthode d'optimisation robuste L'AG peut résoudre, avec fiabilité, des fonctions représentant des reliefs de solution réputées très difficiles pour les méthodes d'optimisation classique (simplex, le plus fort gradient …).les fonction réputées sont des fonctions a plusieurs dimensions ou les méthodes ordinaires ne peuvent pas prendre en compte l'effet d'interaction entre tous les paramètres, (Goldberg, 1980).

4.3. Caractéristique de l'AG

Les principales caractéristiques relatives à cette technique se concentrent autour des trois points :

- Le parallélisme : l'algorithme génétique travaille en parallèle sur un certain nombre de candidats et non pas sur un candidat unique .la méthode de recherche est globale et couvre tout l'espace de recherche ;
- L'utilisation minimale d'informations: il n'a besoin que de la mesure d'adéquation (la qualité d'une solution), il ne repose sur aucune autre information, par exemple des dérives ou hypothèses telles que la continuité et la différentialiste il ne requiert qu'une capacité à classer les solutions entre elles ;
- L'utilisation règles probabilités plutôt que déterministes dans l'exploration de l'espace de recherche. L'introduction du hasard est très bénéfique pour l'optimisation de fonction présentant plusieurs optima et aussi en cas de fonction non permanente (déplacement ou changent des optima au cours du temps).

4.4. Applications

Les algorithmes génétiques, qui décrivent l'évolution d'une population d'individus en réponse à un environnement, sont utilisés pour créer des méthodes d'optimisation exploitées pour résoudre des problémes.ils sont par exemple utilisés pour optimiser la performance des outils de datamining.

Les algorithmes génétique résolvent des problèmes n'ayant aucun méthode de résolution décrite précisément ou dont la solution exacte, si elle est connue, et très complexe pour êtres calculée en un temps raisonnable, c'est notamment le cas quand des contraintes multiples, complexes et parfois même en partie contradictoire doivent êtres satisfaites simultanément.

Pour résumé, on peut dire que les algorithmes génétiques sont essentiellement utilisés pour traiter les deux cas suivants :

- L'espace de recherche est vaste où le problème possède énormément des paramètres devant êtres optimisés simultanément ;
- Le problème ne peut pas êtres facilement décrit par une modèle mathématique précis.

Parmi les avantages des AG, on peut mentionner la facilité de travailler avec des paramètres discrets ou continus (ou sur des mélanges des deux types de paramètres), et le fait qu'ils n'ont pas besoin d'information sur le gradient de la fonction objectif ; les éventuelles discontinuités de la fonction objectif ont peu d'effet sur la performance de ces algorithmes ; ils se laissent difficilement piéger par des optimums locaux ; ils peuvent traiter un grand nombre de paramètres, et sont très adaptés au calcul parallèle ; ils génèrent une liste de solutions semi-optimales et non une seule solution (ce qui est d'une grande importance pour l'optimisation multi objectif, comme mentionné auparavant) ; ils travaillent de la même façon, que les données soient générées numériquement, expérimentalement ou analytiquement ; etc. Toutes ces caractéristiques contribuent à ce que les AG soient efficaces pour une grande variété de problèmes d'optimisation.

4.5. Principe

Un algorithme génétique recherche le ou les extrema d'une fonction définie sur un espace de données. Pour l'utiliser, on doit disposer des cinq éléments suivants:

- **Un principe de codage de l'élément de population:** Cette étape 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. Le choix du codage des données conditionne le succès des algorithmes génétiques. Les codages binaires ont été très employés à l'origine. Les codages réels sont désormais largement utilisés, notamment dans les domaines applicatifs, pour l'optimisation de problèmes à variables continues ;
- **Un mécanisme de génération de la population initiale:** Ce mécanisme doit être capable de produire une population d'individus non homogène qui servira de base

pour les générations futures. Le choix de la population initiale est important car il peut rendre plus ou moins rapide la convergence vers l'optimum global. Dans le cas où l'on ne connaît rien du problème à résoudre, il est essentiel que la population initiale soit répartie sur tout le domaine de recherche ;

- **Une fonction à optimiser:** Celle-ci prend ses valeurs dans R et est appelée fitness ou fonction d'évaluation de l'individu. Celle-ci est utilisée pour sélectionner et reproduire les meilleurs individus de la population ;
- **Des opérateurs permettant de diversifier la population au cours des générations et d'explorer l'espace d'état:** L'opérateur de croisement recompose les gènes d'individus existant dans la population, l'opérateur de mutation a pour but de garantir l'exploration de l'espace d'état ;
- **Des paramètres de dimensionnement:** taille de la population, nombre total de générations ou critère d'arrêt, probabilités d'application des opérateurs de croisement et de mutation.

Le principe général du fonctionnement d'un algorithme génétique est représenté sur la figure 4.1, L'algorithme commence avec un ensemble de solutions possibles du problème (individus), constituant une population. Les individus sont formés par des variables, qui sont les paramètres à ajuster dans un dispositif (par exemple, la longueur et la largeur du système à optimiser). Cette population est conçue aléatoirement à l'intérieur de limites prédéfinies (par exemple, les limites dictées par les aspects constructifs).

Certaines solutions de la première population sont utilisées pour former, à partir d'operateurs génétiques (croisement, mutation, etc.), une nouvelle population. Ceci est motivé par l'espoir que la nouvelle population soit meilleure que la précédente. Les solutions qui serviront à former de nouvelles solutions sont sélectionnées aléatoirement d'après leurs mérites (représentés par une « fonction objectif » spécifique au problème posé, qui devra être minimisée ou maximisée) : meilleur est l'individu, plus grandes seront ses chances de se reproduire (c'est-à-dire, plus grande sera sa probabilité d'être sélectionné pour subir les opérateurs génétiques). Ceci est répété jusqu'à ce qu'un critère de convergence soit satisfait (par exemple, le nombre de générations ou le mérite de la meilleure solution).

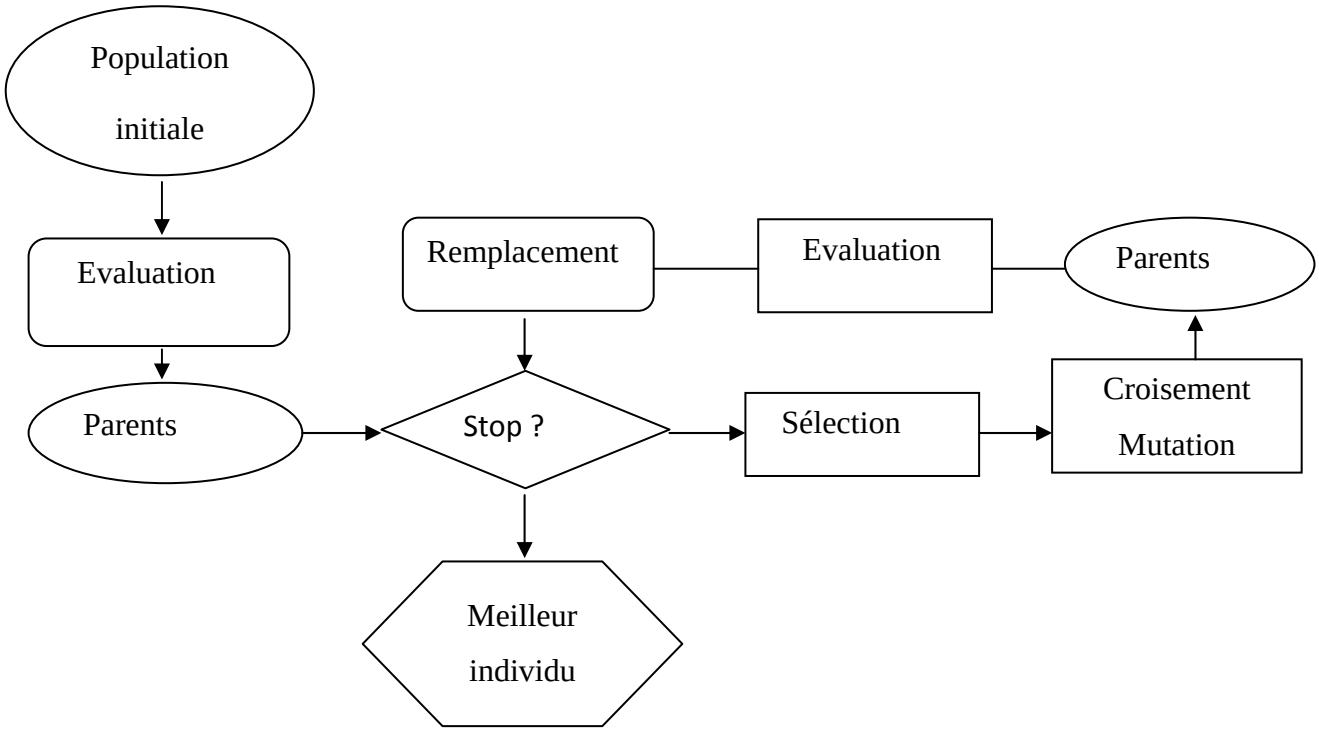

Figure 4.1 : Squelette d'un algorithme génétique.

4.6. Le Codage

Premièrement, il faut représenter les différents états possibles de la variable dont on cherche la valeur optimale sous forme utilisable pour un AG: c'est le codage. Cela permet d'établir une connexion entre la valeur de la variable et les individus de la population, de manière à imiter la transcription génotype-phénotype qui existe dans le monde vivant. Il existe principalement deux types de codage : le codage binaire, le codage réel.

4.6.1. Codage binaire

Ce codage a été le premier à être utilisé dans le domaine des AG. Il présente plusieurs avantages: alphabet minimum {0,1}, facilité de mise en point d'opérateurs génétiques et existence de fondements théoriques (théorie sur les schémas). Néanmoins ce type de codage présente quelques inconvénients :

Les performances de l'algorithme sont dégradées devant les problèmes d'optimisation de grande dimension à haute précision numérique. Pour de tels problèmes, les AG basés sur les chaînes binaires ont de faibles performances comme le montre michalewicz (Michalewicz, 1992).

La distance de Hamming entre deux nombres voisins (nombre de bits différents) peut être assez grande dans le codage binaire : l'entier 7 correspond à la chaîne 0111 et la chaîne 1000 correspond à l'entier 8. Or la distance de hamming entre ces deux chaînes est de 4, ce qui crée bien souvent une convergence, et non pas l'obtention de la valeur optimale *Figure 4.2 (1.b).*

4.6.2. Codage réel

Il a le mérite d'être simple. Chaque chromosome est en fait un vecteur dont les composantes sont les paramètres du processus d'optimisation. Par exemple, si on recherche l'optimum d'une fonction de n variables /(x_1, x_2, x_3, x_4,…,x_n), on peut utiliser tout simplement un chromosome ch contenant les n variables: Avec ce type de codage, la procédure d'évaluation des chromosomes est plus rapide vu l'absence de l'étape de transcodage (du binaire vers le réel). Les résultats donnés par Michalewicz (Michalewicz, 1992) montrent que la représentation réelle aboutit souvent à une meilleure précision et un gain important en termes de temps d'exécution *Figure 4.2 (1.a).*

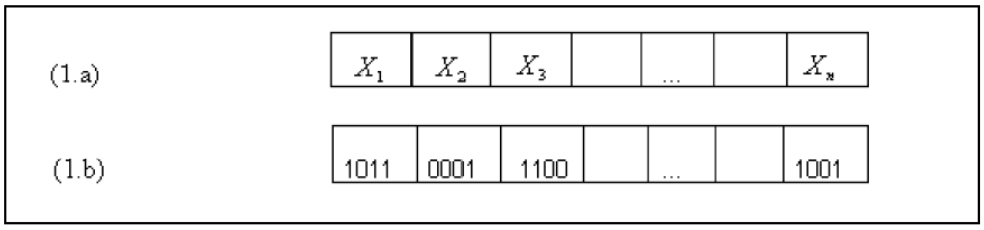

Figure 4.2 : représentation d'un individu; codage réel (1.a); codage binaire (1.b)

4.7. Génération aléatoire de la population initiale

Le choix de la population initiale d'individus conditionne fortement la rapidité de l'algorithme. Si la position de l'optimum dans l'espace d'état est totalement inconnue, il est naturel d'engendrer aléatoirement des individus en faisant des tirages uniformes dans chacun des domaines associés aux composantes de l'espace d'état, en veillant à ce que les individus produits respectent les contraintes(Michalewicz, 1991). Si par contre, des informations a

priori sur le problème sont disponibles, il paraît bien évidemment naturel d'engendrer les individus dans un sous-domaine particulier afin d'accélérer la convergence. Dans l'hypothèse où la gestion des contraintes ne peut se faire directement, les contraintes sont généralement incluses dans le critère à optimiser sous forme de pénalités.

4.8. Opérateurs de sélection

La sélection permet d'identifier statistiquement les meilleurs individus d'une population et d'éliminer les mauvais. On trouve dans la littérature un nombre important de principes de sélection plus ou moins adaptés aux problèmes qu'ils traitent.

La manière la plus classique d'implémenter l'opérateur de sélection est la méthode de la roulette biaisée. Ce principe se base sur l'image d'une roulette de casino telle que la probabilité de sélectionner un individu particulier soit proportionnelle à la valeur de sa fonction d'adaptation. Ainsi, les meilleurs individus auront une probabilité plus importante d'être sélectionnés pour la reproduction. Cependant, des tests empiriques ont montré qu'il était plus efficace de se baser sur le classement des individus dans la population, appelé leur rang, plutôt que sur la valeur de la fitness.

Dans ce dernier cas, si un individu obtient une fitness nettement meilleure que les autres, il risque d'envahir rapidement la population. Cette réduction rapide de la diversité génétique mène facilement à une convergence prématurée vers un unique minimum local.

Pour sélectionner les individus en fonction de leur rang, on commence par les classer en fonction de leur fitness par ordre décroissant puis on définit la probabilité de tirer un rang i comme le multiple de la probabilité de tirer le rang du dessous par un facteur multiplicatif. On aboutit à la formule suivante, permettant de tirer un index i :

$$i = n - \frac{n}{\log(k+1)} . \log(k.rand(1) + 1) \qquad (4.1)$$

Où $k = c^{n+1} - 1$, n est le nombre d'individus que l'on souhaite prendre en compte, c un coefficient par exemple égal à 1, 1 et rand(l) un générateur aléatoire renvoyant un nombre réel compris entre 0 et 1.

Une autre méthode classique de sélection est celle du tournoi. On tire successivement des couples d'individus et on sélectionne celui qui possède la meilleure fitness. On itère le processus jusqu'à ce que l'on ai sélectionné suffisamment de parents.

Ces deux approches peuvent être combinées au principe de l'élitisme. Afin d'éviter la perte des meilleurs individus, on peut choisir de les recopier directement vers la génération suivante. On évite ainsi que la meilleure fitness puisse décroître et que les très bonnes solutions soient éliminées par la nature stochastique (aléatoire) de l'opérateur de sélection. On conserve généralement 5 à 10 % des meilleurs individus.

Comme ce sera le cas pour les autres opérateurs, il existe très peu de base théorique pour guider le choix de la méthode de sélection. Seuls des tests concernant votre application spécifique vous permettront de déterminer l'approche la plus efficace.

4.9. Opérateurs de croisement

Le croisement a pour but d'enrichir la diversité de la population en manipulant la structure des chromosomes. Classiquement, les croisements sont envisagés avec deux parents et génèrent deux enfants.

4.9.1. Croisement en un point (par découpage)

Initialement, le croisement associé au codage par chaînes de bits est le croisement à découpage de chromosomes (slicing crossover). Pour effectuer ce type de croisement sur des chromosomes constitués de M gènes, on tire aléatoirement une position dans chacun des parents. On échange ensuite les deux sous-chaînes évolutionnaires comme montre figure 4.3.

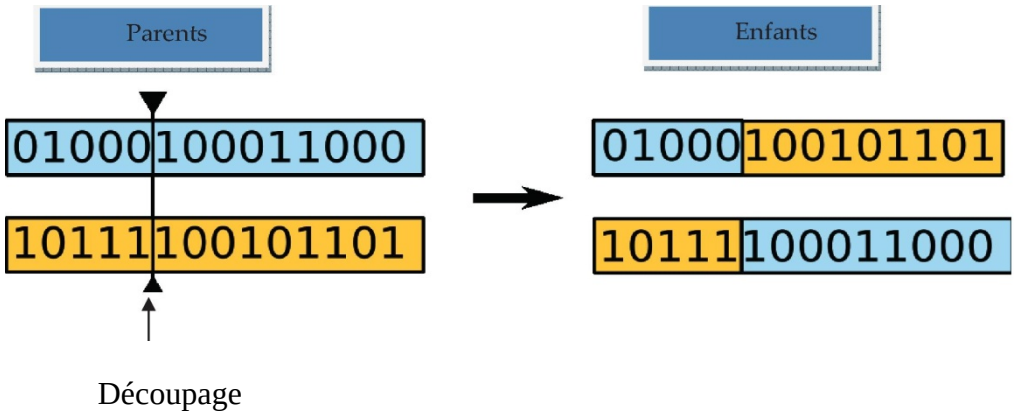

Figure 4.3: Croisement par découpage sur une chaîne binaire.

Ce type de croisement à découpage de chromosomes est très efficace pour les problèmes discrets. Pour les problèmes continus, un croisement « barycentrique » est souvent utilisé : Pour générer deux enfants el et e2 à partir des parents p1 et p2, on commence par tirer un nombre *a* au hasard dans l'intervalle [-0.5; 1.5], on applique ensuite les formules :

$$el = a\ p1 + (1 - a)p2 \qquad (4.2)$$

$$e2 = a\ p2 + (1 - a)pl$$

4.9.2. Croisement en deux points :

On choisit au hasard deux points de croisement et on échange les parties de chaîne situées entre ces deux points *Figure 4.4.*

	Parents		Enfants
A1	00|0100|111		**A3** 00|1011|111
A2	11|1011|000	⟶	**A4** 11|0100|000
	Avant		**après**

Figure 4.4 : représentation d'un croisement en deux points

4.9.3. Croisement uniforme

Dans ce type de croisement, on utilise un masque de croisement (*mask*), qui consiste en un vecteur généré aléatoirement, de longueur identique aux chaînes parents, et composé de 0 et 1. Lorsque le bit du masque vaut 0, l'enfant hérite le bit du premier parent, sinon il hérite de celui du second parent. Le second enfant est le complémentaire du premier. Ce croisement peut être considéré comme une généralisation du croisement multipoint sans connaissance préalable du point de croisement *Figure 4.5.*

A1 001010 (**Parent 1**)

A2 011111 (**Parent 2**)

Masque 001101

A3 001111 (**Enfant 1**)

A4 011010 (**Enfant 2**)

Figure 4.5 : Représentation d'un croisement uniforme

4.10. Opérateur de mutation

L'opérateur de mutation apporte aux algorithmes génétiques la propriété d'ergodicité de parcours d'espace. Cette propriété indique que l'algorithme génétique sera susceptible d'atteindre tous les points de l'espace d'état, sans pour autant les parcourir tous dans le processus de résolution. Ainsi en toute rigueur, l'algorithme génétique peut converger sans croisement, et certaines implémentations fonctionnent de cette manière. Les propriétés de convergence des algorithmes génétiques sont donc fortement dépendantes de cet opérateur sur le plan théorique.

La définition de la mutation dans le cadre du codage binaire est particulièrement simple. Dans un premier temps, une ou plusieurs positions sont tirées. Le ou les bits correspondants sont lors remplacés par une valeur tirée aléatoirement (en l'occurrence 0 ou 1) *Figure 4.6.*

Comme dans le cas du croisement, cette approche fonctionne mal sur des nombres réels. La stratégie la plus courante est d'ajouter du bruit à certains membres du vecteur, tirés aléatoirement. Différentes distributions de probabilité, centrées sur la valeur du gène avant la mutation, peuvent être envisagées. Dans le cas d'une loi uniforme, une variation est choisie uniformément dans un intervalle fixé autour de la valeur actuelle du gène. Une distribution gaussienne peut aussi être utilisée.

La largeur de l'intervalle choisi dans le cas de la loi uniforme et le paramètre d'écart type dans celui de la loi gaussienne peuvent être délicate à choisir. Si celui-ci est trop faible, l'espace de recherche risque ne pas être suffisamment exploré. À l'inverse, s'il est trop important, l'algorithme peut avoir des difficultés à affiner les solutions alors qu'il est proche de l'optimum. Une approche intéressante pour résoudre ce problème est de coder ces paramètres dans le génotype. Ainsi, les solutions explorant rapidement l'espace seront favorisées au début de l'algorithme et auront naturellement tendance à s'éteindre au fur et à mesure que le processus converge vers l'optimum.

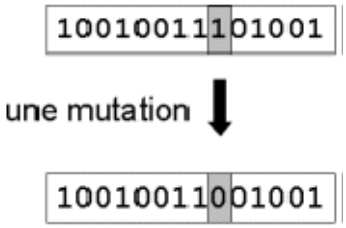

Figure 4.6 : Représentation d'une mutation

4.11. Gestion des contraintes

Un élément de population qui viole une contrainte se verra attribuer une mauvaise fitness et aura une probabilité forte d'être éliminé par le processus de sélection.

Il peut cependant être intéressant de conserver, tout en les pénalisant, les éléments non admissibles car ils peuvent permettre de générer des éléments admissibles de bonne qualité. Pour de nombreux problèmes, l'optimum est atteint lorsque l'une au moins des contraintes de séparation est saturée, c'est-à-dire sur la frontière de l'espace admissible.

Gérer les contraintes en pénalisant la fonction fitness est difficile, un « dosage » s'impose pour ne pas favoriser la recherche de solutions admissibles au détriment de la recherche de l'optimum ou inversement.

Disposant d'une population d'individus non homogène, la diversité de la population doit être entretenue au cours des générations, afin de parcourir le plus largement possible l'espace d'état. C'est le rôle des opérateurs de croisement et de mutation.

4.12. Un exemple simple

Nous reprenons ici l'exemple de Goldberg (Goldberg, 1989). Il consiste à trouver le maximum de la fonction f (x)=x sur l'intervalle [0;31] où x est un entier. La première étape consiste à coder la fonction. Par exemple, nous utilisons un codage binaire de x, la séquence (chromosome) contenant au maximum 5 bits. Ainsi, nous avons x = 2(0,0,0,1,0}, de même x = 31 {1,1,1,1,1}. Nous recherchons donc le maximum d'une fonction de fitness dans un espace de 32 valeurs possibles de x.

4.12.1. Tirage et évaluation de la population initiale

Nous fixons la taille de la population à N = 4. Nous tirons donc de façon aléatoire 4 chromosomes sachant qu'un chromosome est composé de 5 bits, et chaque bit dispose d'une probabilité - d'avoir une valeur 0 ou 1.

Nous observons que le maximum 16, est atteint par la deuxième séquence. Voyons comment l'algorithme va tenter d'améliorer ce résultat.

Tableau 4.1:résultats après tirage et évaluation de la population initiale

Numéro	Séquence	Fitness	% du Total
1	00101	5	14.3
2	10000	16	45.7
3	00010	2	5.7
4	00110	12	34.3
Total		**35**	**100**

4.12.2. Sélection

Une nouvelle population va être créée à partir de l'ancienne par le processus de sélection de la roue de loterie biaisée.

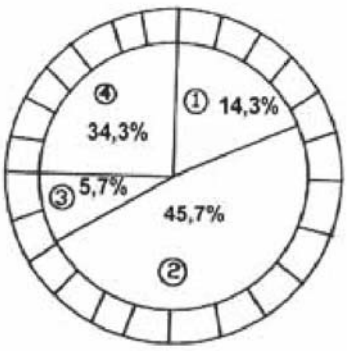

Figure 4.7 : La roue de loterie biaisée : opération de sélection.

Nous tournons cette roue 4 fois et nous obtenons au final la nouvelle population

Tableau 4.2: nouvelle population après l'opération de sélection

Numéro	Séquence
1	10000
2	01100
3	00101
4	10000

4.12.3. Le croisement

Les parents sont sélectionnés au hasard. Nous tirons aléatoirement un lieu de croisement dans la séquence. Le croisement s'opère alors à ce lieu avec une probabilité pc. Le tableau suivant donne les conséquences de cet opérateur en supposant que les chromosomes 1 et 3, puis 2 et 4 sont appariés et qu'à chaque fois le croisement s'opère (par exemple avec pc = 1).

Tableau 4.3.résultat du croisement

1=3	1=2
100|00	01|100
001|01	10|000
10001	**01000**
00100	**10100**

4.12.4. La mutation

Dans cet exemple à codage binaire, la mutation est la modification aléatoire occasionnelle (de faible probabilité) de la valeur d'un bit (inversion d'un bit). Nous tirons ainsi pour chaque bit un chiffre aléatoire entre 0 et 1 et si ce chiffre est inférieur à pm alors la mutation s'opère. Le tableau suivant, avec pm = 0,05, met en évidence ce processus.

Tableau 4.4 : processus de la mutation

Chromosome	Tirage	Nouveau Bit	Nouveau Chromosome
10001	**15 25 36 04 12**	1	10011
00100	**26 89 13 48 59**	-	00100
01000	**32 45 87 22 65**	-	01000
10100	**47 01 85 62 35**	1	11100

Maintenant que la nouvelle population est entièrement créée, nous pouvons de nouveau l'évaluer.

4.12.5. Retour à la phase d'évaluation

Le maximum est maintenant de 28 (séquence 4). Nous sommes donc passés de 16 à 28 après une seule génération. Bien sûr, nous devons recommencer la procédure à partir de l'étape de sélection jusqu'à ce que le maximum global, 31, soit obtenu, ou bien qu'un critère d'arrêt ait été satisfait.

Tableau 4.5: résultats de nouvelle évaluation

Numéro	Séquence	Fitness	% du Totale
1	10011	19	32.2
2	00100	4	6.8
3	01000	8	13.5
4	11100	28	47.5
Total		59	**100**

4.13. Conclusion

Ce chapitre décrit le fonctionnement, la taxonomie et les différents opérateurs d'un algorithme génétique standard. Il présente aussi une simple application afin de faciliter la compréhension. Ces principe de l'algorithme génétique qui seront utilisé pour extraite ou sélection les paramètres pertinents dans le cas d'un système de reconnaissance du locuteur (RAL).

Chapitre 5 : Simulations et résultats

5.1. Introduction

Ce chapitre présente l'étude expérimentale qui consiste à utiliser un algorithme génétique pour la sélection des paramètres pertinents parmi un jeu de paramètres disponible. Le but est de réduire la dimension des données tout en minimisant le taux d'erreur (l'erreur de classification). Pour le calcul du taux d'erreur, un classificateur K-NN *(k*-Nearest Neighbor algorithm) est utilisé.

5.2. Setup expérimental

5.2.1. Les paramètres de l'algorithme génétique

En réalité, il n'existe pas de paramétrage universel pour la quantification de ces paramètres de dimensionnement. Néanmoins, nous devons choisir ces paramètres avec soin pour résoudre concrètement un problème donné, dans notre cas c'est la réduction du jeu de paramètre et l'amélioration du taux de reconnaissance, comme suit :

- Le nombre des coefficients MFCC est 10 (c_1 … c_{10}). Ces coefficients sont codés sur 10 bits en utilisant un codage binaire (0,1). Un bit à 1 veut dire que le coefficient est présent. Tandis qu'un bit à 0 indique l'absence de ce dernier.
- Nombre de population égale à 50 (npop = 50).
- Nombre de génération égale à 100.
- Taux de croisement variable (% des individus dont la fitness > moyenne (fitness))
- Taux de mutation égale à 0.2 (mutrate = 0.2).
- Et la fonction d'objectif est la minimisation du taux d'erreurs.

5.2.2. Classification K-NN (k-nearest neighbor algorithm)

- **Notion d'une classification**

Une classification correspond à représenter un objet quelconque au moyenne d'un vecteur de caractéristique $X=[x_{1...}x_d]^T$. tous les vecteurs qui représentent l'ensemble des objets peuvent être positionnées dans l'espace Euclidien R^d. où ils correspondent chacun à un point, ceux-ci peuvent alors être regroupés en amas, chacun de ces amas étant associé à une classe particulière. Un exemple pour un problème à deux classes est illustré à la figure 5.1.

Le rôle d'un classificateur est de déterminer, parmi un ensemble fini de classes, à laquelle appartient un objet donné.

Un classificateur doit être capable de modéliser au mieux les frontières qui séparent les classes les unes des autres. Cette modélisation fait appel à la notion de *fonction discriminante,* qui permet d'exprimer le critère de classification de la manière suivante:

Assigner la classe W_i à l'objet représenté par le vecteur X si, et seulement si, la valeur de la fonction discriminante de la classe W_i est supérieure à celle de la fonction discriminante de n'importe quelle autre classeW_j "

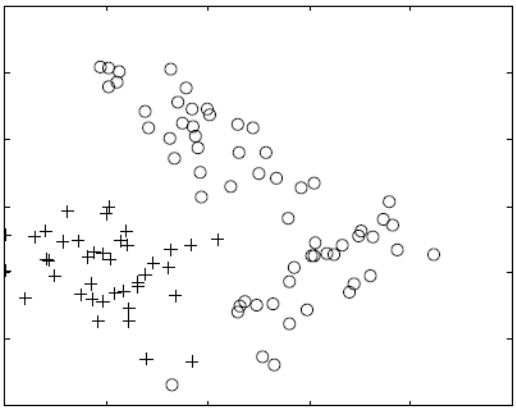

Figure 5.1 : Représentation d'objets appartenant à deux classes distinctes, dans un espace à deux dimensions

- **Présentation et principe**

La classification par méthode du *Plus Proches Voisins* est une extrapolation du classificateur Euclidien. Pour rappel, le classificateur euclidien l'un des plus simples classificateurs qui puissent être conçus. La classe dont le vecteur de caractéristiques moyen est le plus proche, au sens de la distance Euclidienne, du vecteur de caractéristiques de l'objet à classifier est assignée à ce dernier. Les fonctions discriminantes utilisées sont donc de la forme suivante :

$$\Phi_i(X) = -\frac{1}{2}(X-M_i)^T(X-M_i) \qquad (5.1)$$

Où $M_i = E\{X|W_i\}$ est le vecteur de caractéristiques moyen des éléments qui appartiennent à la classe W_i , $E\{.\}$ Désignant l opérateur d'espérance mathématique, et $(.)^T$ celui de transposition.

Le terme quadratique X^TX est indépendant de la classe de l'objet, et les fonctions discriminant discriminantes peuvent également s'écrire:

$$\Phi_i(X) = M_i^T X - \frac{1}{2} M_i^T Mi \qquad (5.2)$$

Les frontières qui séparent les classes dans l'espace R^d sont ici linéaires. Un exemple en est donné à la figure 5.2.

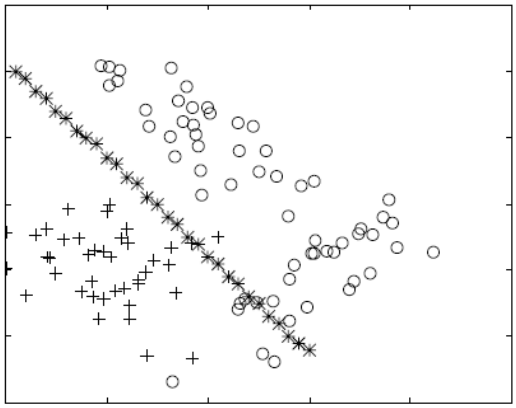

Figure 5.2 : Frontière fournie par le classificateur Euclidien dans le cas d'un problème à deux classes

La différence est que au lieu d'utiliser le vecteur de caractéristiques moyen $\mathbf{M}_i$ comme unique prototype d'une classe comme le cas d'un classificateur Euclidien , la méthode du plus proche voisin fait intervenir tous les exemplaires des vecteurs de caractéristiques disponibles. La distance Euclidienne entre chacun de ceux-ci et celui de l'objet à classifier est calculée, et la classe assignée à l'objet est alors celle du prototype le plus proche de celui-ci. Les fonctions discriminantes sont donc de la forme :

$$\Phi_i(\mathrm{X}) = \text{-}\min_{x_k \,\epsilon w_i} \frac{1}{2}(X - X_K)^{\mathrm{T}}(X - X_K) \tag{5.3}$$

Le terme quadratique pouvant être omis, ces fonctions se réduisent à :

$$\Phi_i(\mathrm{X}) = \text{-}\min_{x_k \,\epsilon w_i}(X_k^T X - \frac{1}{2} X_k^T X_K) \tag{5.4}$$

Les frontières de décision entre classes sont linéaires par morceaux, c'est-à-dire constituées de nombreux petits polygones convexes, chacun contenant un seul prototype d'une seule classe. Chaque classe est alors délimitée par un polygone très complexe (*figure 5.3*), qui n'est pas nécessairement convexe, ni même d'une seule pièce. Ce classificateur permet ainsi d'établir des frontières de décision relativement complexes, lorsque suffisamment d'exemplaires de chaque classe sont disponibles.

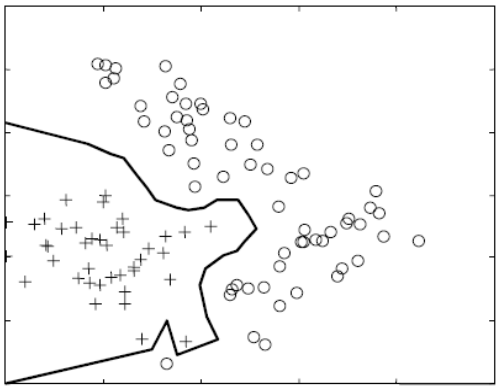

Figure 5.3 : Frontières fournies par le classificateur du Plus Proche Voisin.

Un des inconvénients majeurs de la méthode du Plus Proche Voisin est, qu'en pratique, elle peut présenter une sensibilité élevée aux abords des frontières entre classes. Le plus proche voisin d'un objet peut être d'une classe incorrecte, alors que la majorité de ses voisins ne le sont pas. Afin de contrer cet effet, la classe assignée à un objet peut être celle qui est la plus représentée parmi les *k* plus proches prototypes trouvés. La méthode porte dans ce cas le nom de " *k Plus Proches Voisins* " en anglais K-NN (k-nearest neighbor algorithm). Cette méthode est en fait motivée par les considérations précédentes, relatives à l'estimation des probabilités *à posteriori.* La fonction discriminante d'une classe est à présent simplement le nombre de prototypes de cette classe qui se situent parmi les *k* plus proches voisins de l'objet à classifier :

$$\Phi(X) = \sum_{x_j \in w_i} | X_j \ \Psi_k(x) | \qquad (5.5)$$

Il est possible de montrer que, plus *k* est élevé (le plus souvent impair, afin de limiter le risque d'une indétermination éventuelle), plus la borne supérieure de la probabilité d'erreur associée à la règle de décision des *k* Plus Proches Voisins se rapproche de la borne inférieure, la probabilité optimale de Bayes (Duda et Hart, 1973). Ceci conduit donc à chercher à utiliser une valeur de *k* aussi élevée que possible. En pratique, cependant, afin d'assurer que $p(\theta'=w_i|X')$ soit approximativement la même que $p(\theta=w_i|X)$, il est nécessaire d'avoir l'ensemble des *k* Plus Proches Voisins *X* ' très proches de *X*. Cela implique de faire un compromis, et de choisir une valeur de *k* qui ne demeure qu'une faible fraction du nombre total *N de* prototypes disponibles. Ce n'est que lorsque *N* tend vers l'infini que l'on peut être assuré du comportement optimal de la règle de classification des *k* Plus Proches Voisins.

Le volume de calcul, ainsi que la quantité de mémoire, exigés par les classificateurs du type " *k Plus Proches Voisins* ", sont cependant souvent prohibitifs, au vu du grand nombre de prototypes à prendre en considération et de distances à calculer. Bien qu'une recherche exhaustive puisse être évitée en tenant compte des propriétés triangulaires de la distance Euclidienne, ou du fait que seuls les prototypes particuliers qui déterminent effectivement les frontières entre classes soient réellement déterminants, la mise en application pratique de tels classificateurs requiert souvent des ressources de calcul très élevées.

5.2.3. La base de données utilisée

Pour cette étude nous avons utilisé la base de données calculée à partir des fichiers sonores extraits de la base BDSONS (Carrey, 1986). Cette base contient des coefficients MFCC (Mel-Frequency Cepstrum Coefficients) correspondants à 10 locuteurs. Chaque locuteur a prononcé les cinq voyelles à quatre reprises, ce qui nous donne cinq matrices de 40 vecteurs (5x40x10 = 2000 valeurs).

5.2.4. Résultat de la simulation

5.2.4.1. Coefficients MFCC

Les figures (Figure 5.4 à 5.8) et les tableaux (5.1. et 5.2) ci-dessous représentent les résultats obtenus pour les voyelles /a/, /e/, /i/, /o/ et /u/ respectivement.

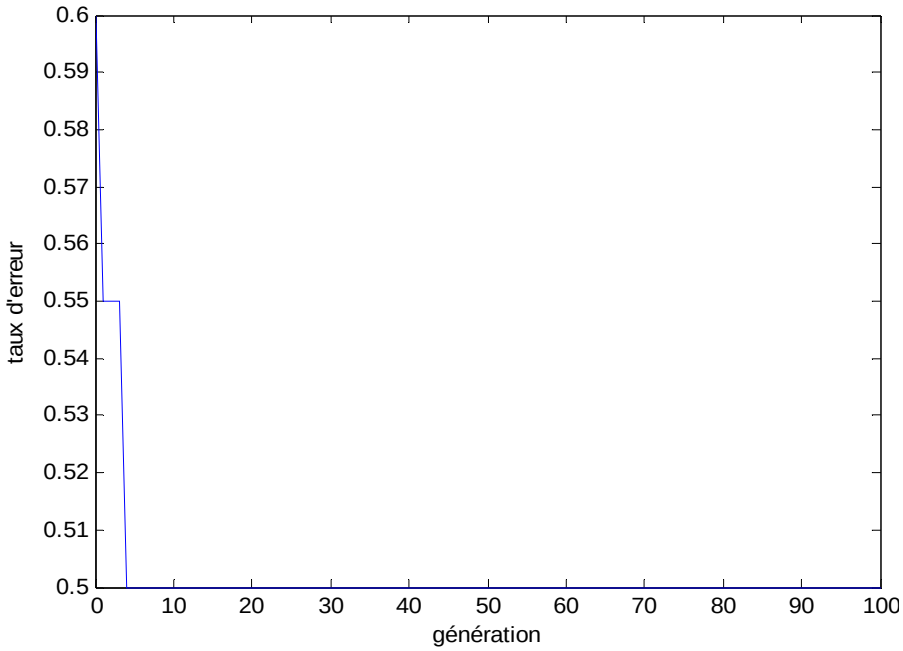

Figure 5.4 : variation du taux d'erreurs pour la voyelle «a»

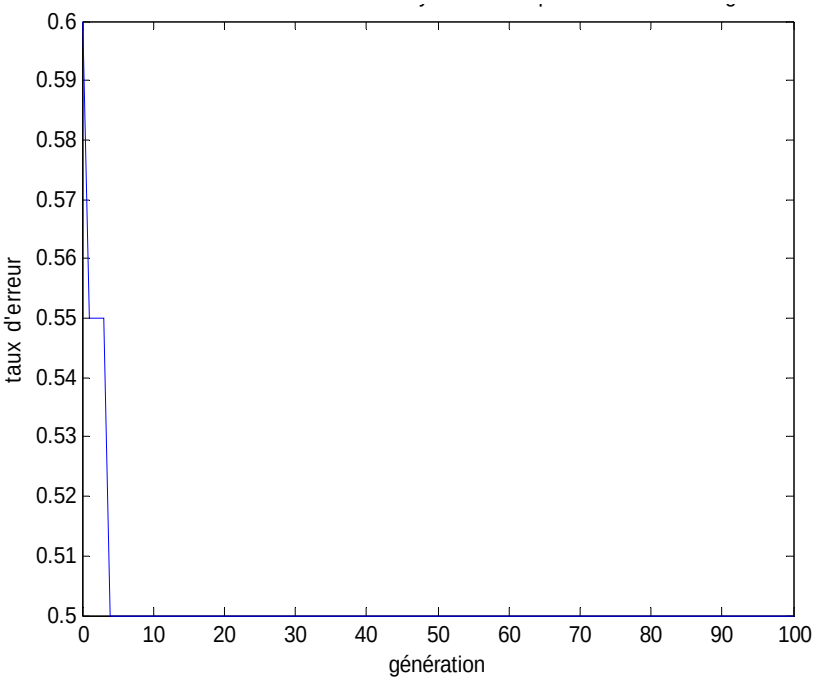

Figure 5.5 : variation du taux d'erreurs pour la voyelle «e»

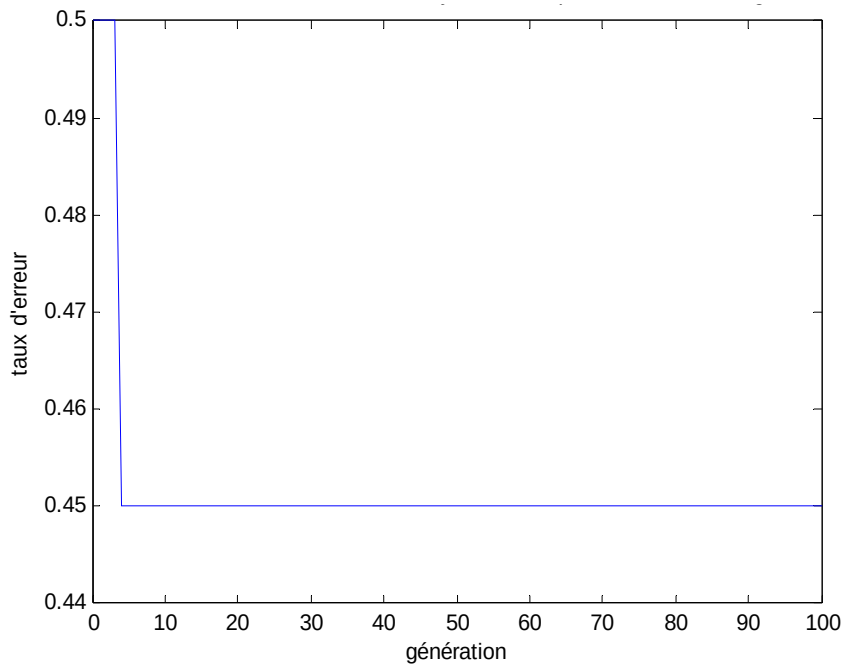

Figure 5.6 : variation du taux d'erreurs pour la voyelle «i»

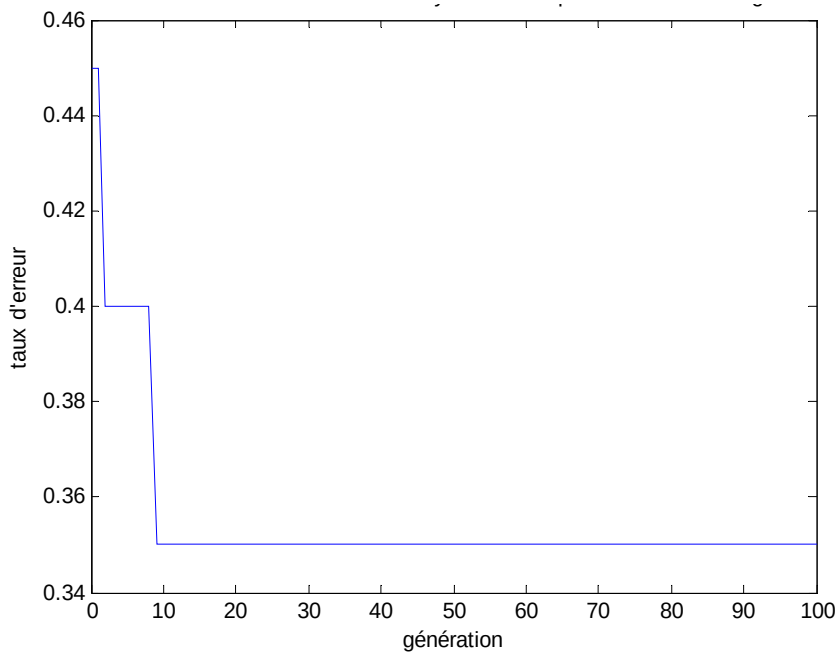

Figure 5.7 : variation du taux d'erreurs pour la voyelle «o»

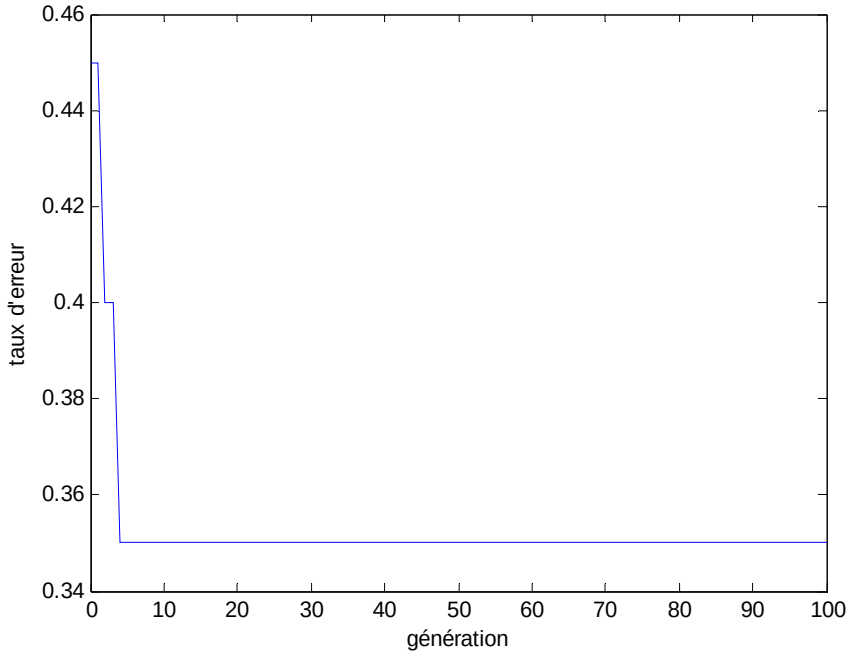

Figure 5.8 : variation du taux d'erreurs pour la voyelle «u»

Tableau 5.1 résultat de la sélection des paramètres MFCC par AG

Les voyelles	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
voyelle a	0	0	0	1	1	0	1	1	1	0
voyelle e	1	1	0	0	0	0	1	1	0	1
voyelle i	1	1	1	0	1	0	0	0	1	1
voyelle o	0	0	1	0	0	0	1	1	0	1
voyelle u	0	1	0	0	1	1	1	0	1	0

Tableau 5.2 résultats obtenus après simulation pour les coefficients MFCC

resultats / les voyelles	Le nombre des paramètres sélectionnées parmi les 10 paramètres	le taux d'erreurs avant application du AG(en %)	le taux d'erreurs après application du AG(en %)
A	5	30	20
E	5	60	50
I	6	50	45
O	4	45	35
U	5	45	35

5.2.4.2. Coefficients ΔMFCC

Les figures (Figure 5.9 à 5.13) et les tableaux (5.3 et 5.4) ci-dessous représentent les résultats obtenus pour les voyelles /a/, /e/, /i/, /o/ et /u/ respectivement pour les coefficients ΔMFCC :

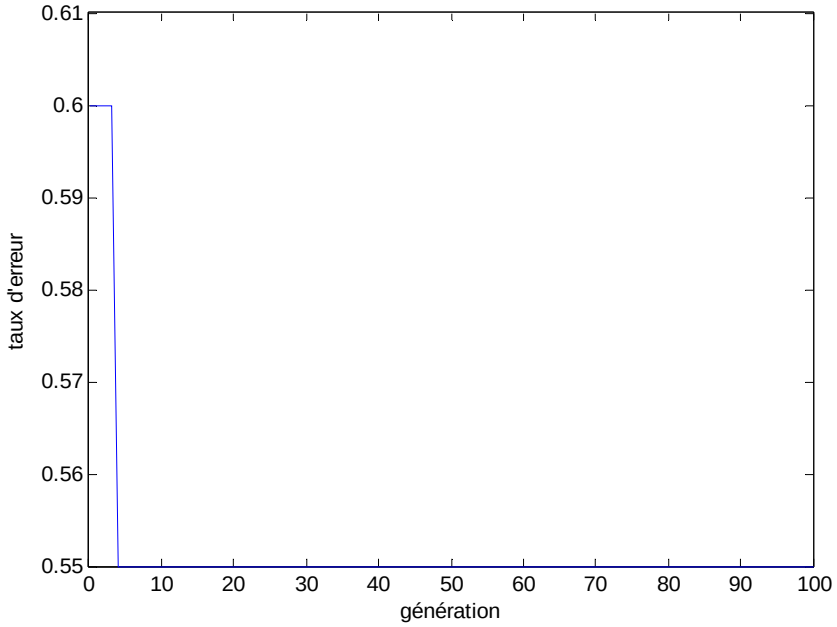

Figure 5.9 : variation du taux d'erreurs pour la voyelle «a»

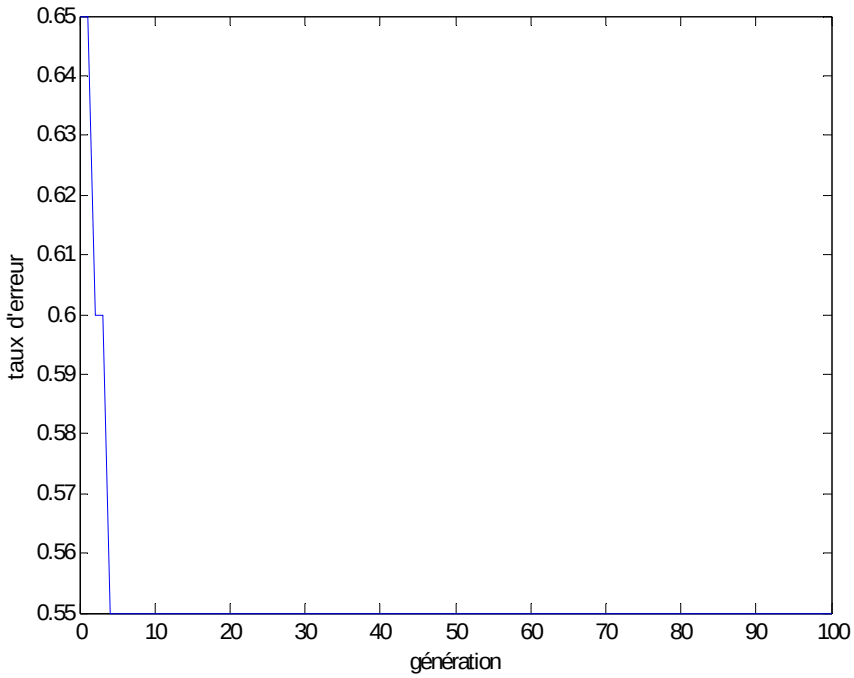

Figure 5.10 : variation du taux d'erreurs pour la voyelle «e»

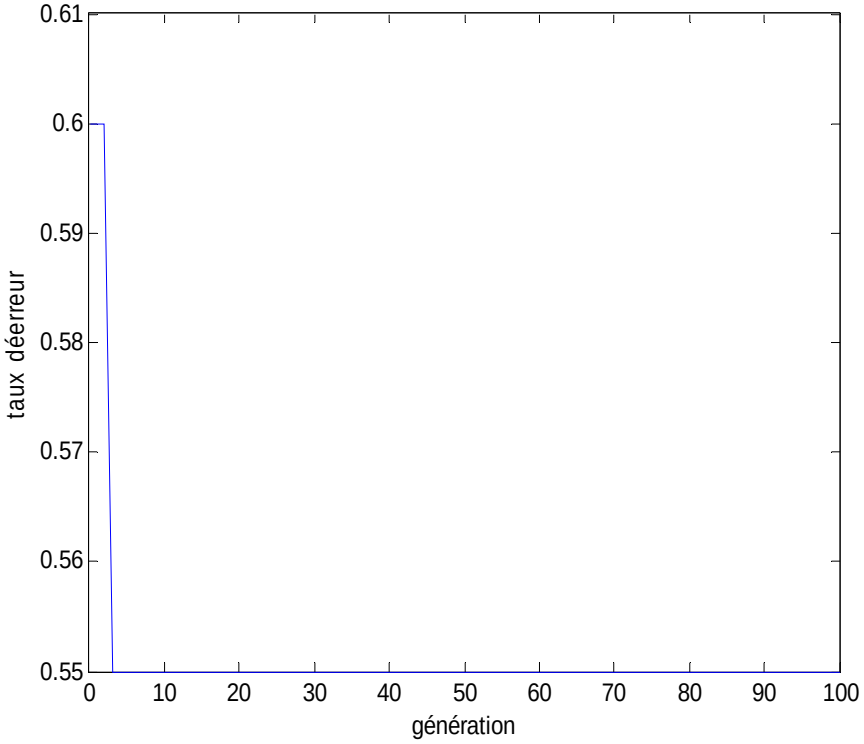

Figure 5.11 : variation du taux d'erreurs pour la voyelle «i»

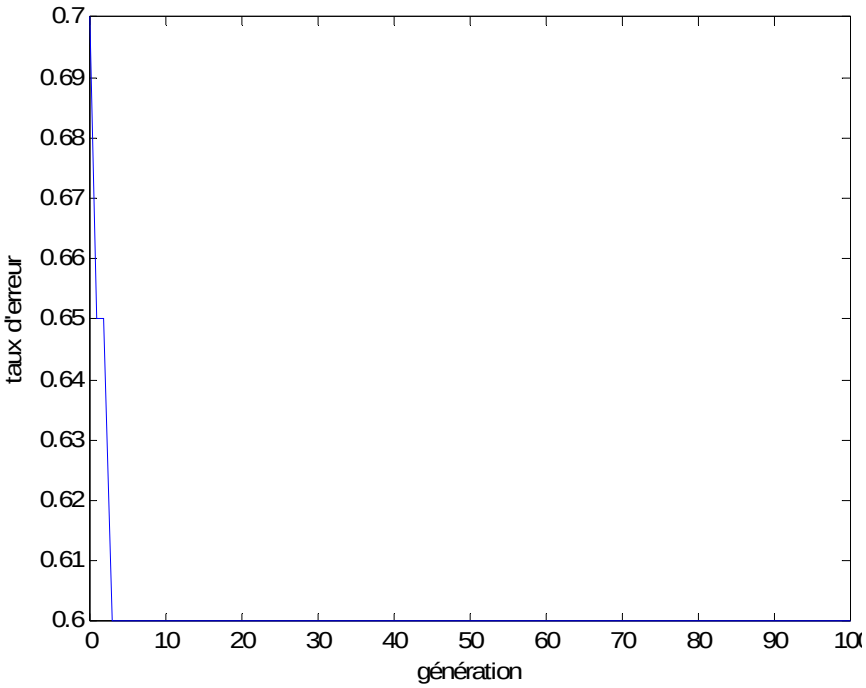

Figure 5.12 : variation du taux d'erreurs pour la voyelle «o»

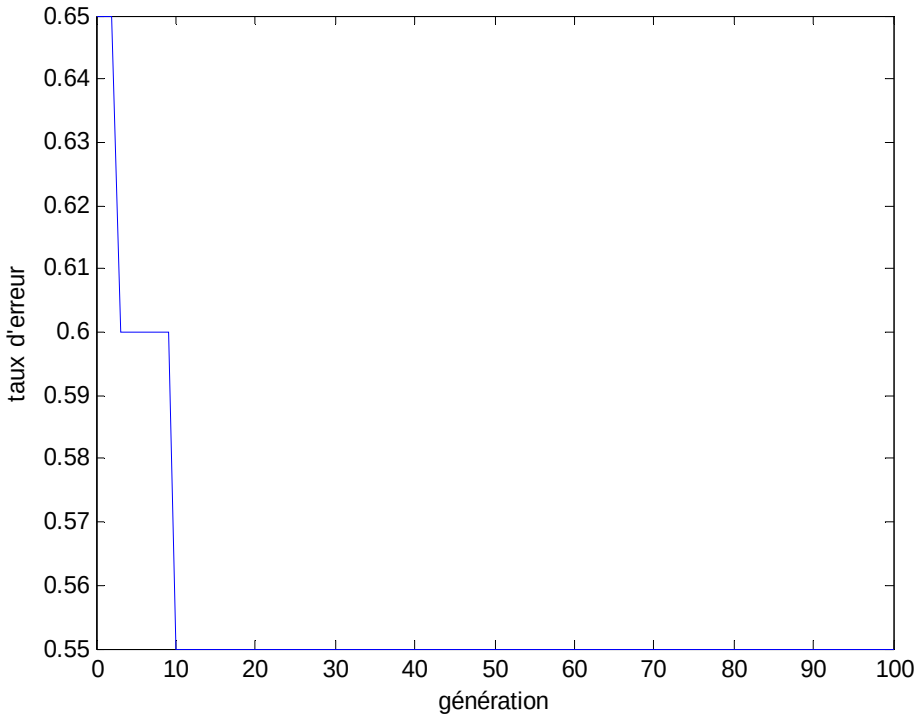

Figure 5.13 : variation du taux d'erreurs pour la voyelle «u»

Tableau 5.3 résultat de la sélection les des paramètres ΔMFCC par AG

Les voyelles	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
voyelle a	1	1	1	0	0	1	1	1	0	0
voyelle e	0	0	1	1	0	1	0	1	1	1
voyelle i	1	1	1	0	0	0	0	0	1	1
voyelle o	0	1	0	0	1	0	0	1	1	1
voyelle u	1	1	0	0	0	1	1	0	1	0

Tableau 5.4 résultats obtenus après simulation pour les coefficients ΔMFCC

resultats / les voyelles	Le nombre des paramètres sélectionnées parmi les 10 paramètres	le taux d'erreurs avant application du AG(en %)	le taux d'erreurs après application du AG(en %)
A	6	60	55
E	6	65	55
I	5	60	55
O	5	70	60
U	5	65	55

5.2.4.3. Coefficients ΔΔMFCC

Les figures (Figure 5.14 à 5.18) et les tableaux (5.5 et 5.6) ci-dessous représentent les résultats obtenus pour les voyelles /a/, /e/, /i/, /o/ et /u/ respectivement pour les coefficients ΔΔMFCC :

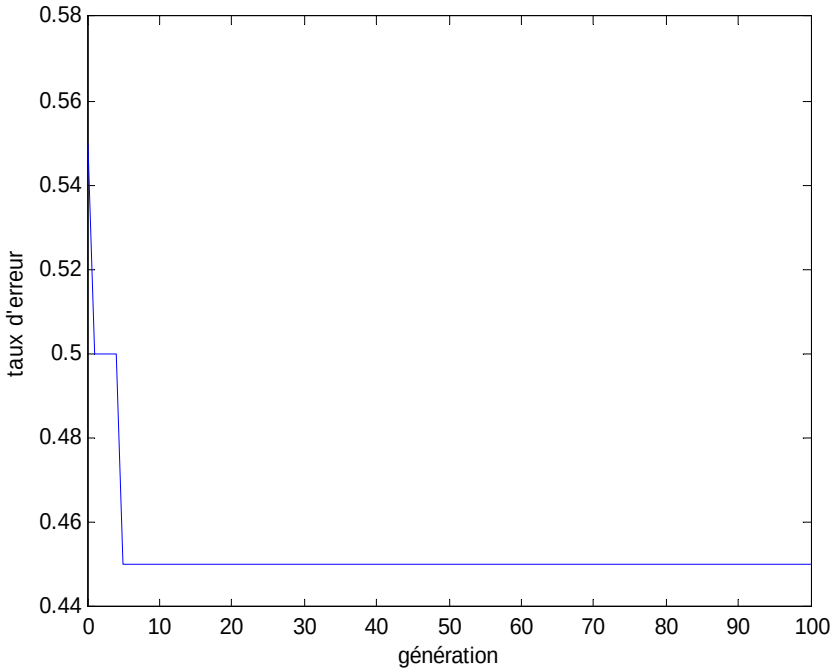

Figure 5.14 : variation du taux d'erreurs pour voyelle «a»

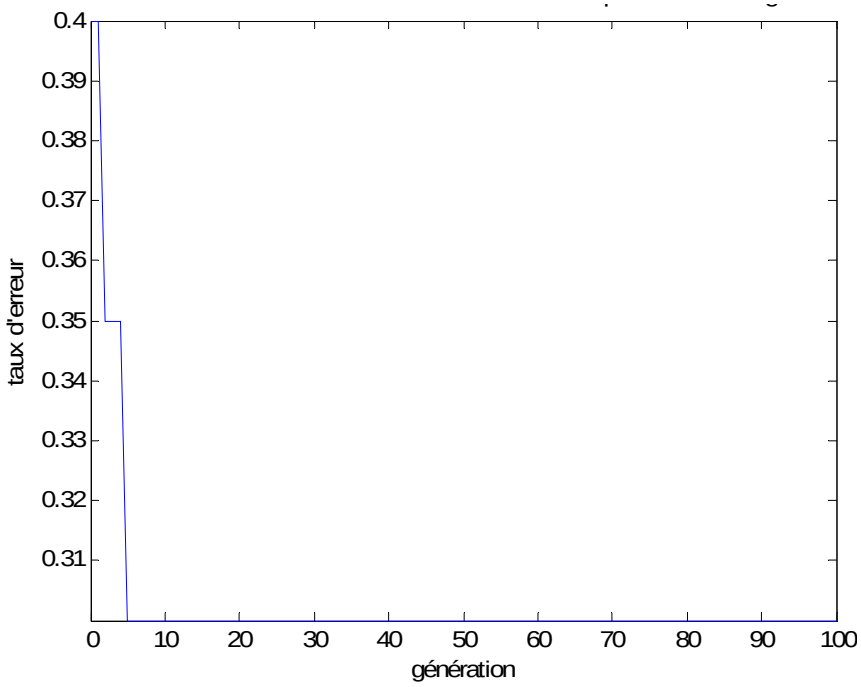

Figure 5.15 : variation du taux d'erreurs pour voyelle «e»

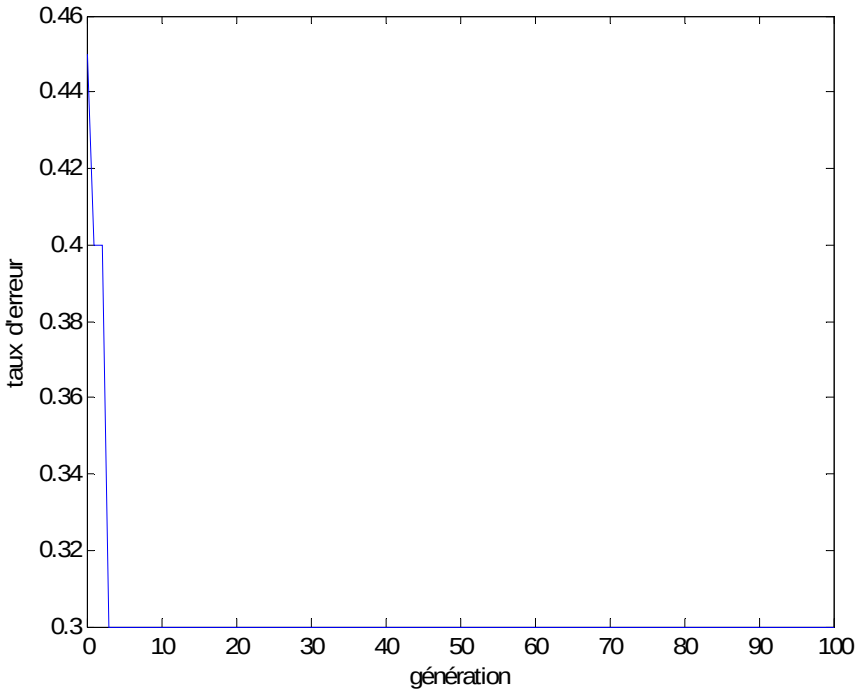

Figure 5.16 : variation du taux d'erreurs pour voyelle «i»

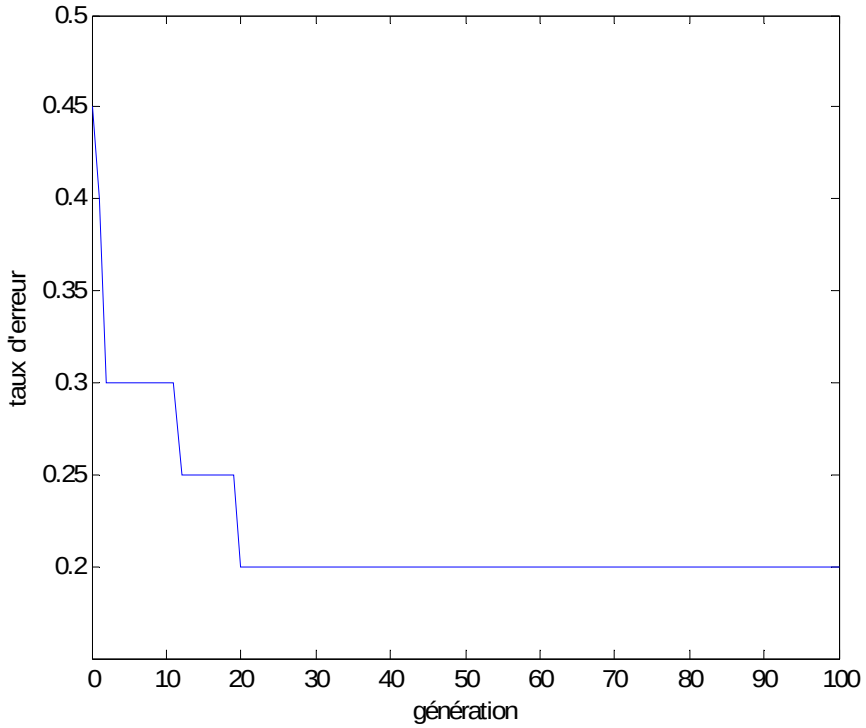

Figure 5.17 : variation du taux d'erreurs pour voyelle «o»

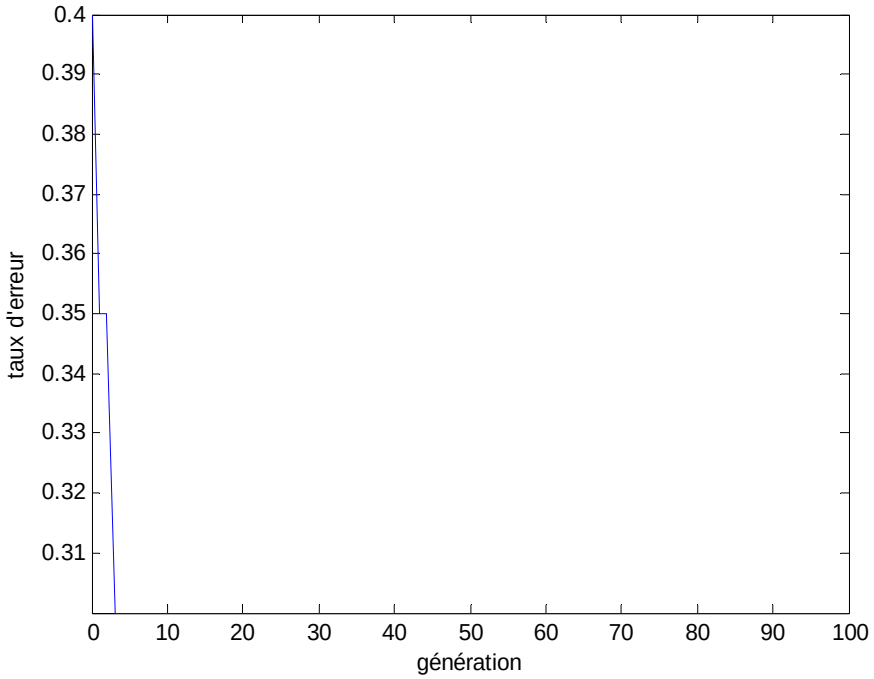

Figure 5.18 variation du taux d'erreurs pour voyelle «u»

Tableau 5.5 résultat de la sélection les paramètres ΔΔMFCC par AG

Les voyelles	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
voyelle a	0	0	1	0	1	1	1	0	1	1
voyelle e	0	1	1	0	0	1	1	1	1	1
voyelle i	0	1	1	1	1	0	0	0	1	1
voyelle o	1	1	0	1	1	0	0	1	0	1
voyelle u	0	1	0	1	0	0	0	0	1	1

Tableau 5.6 résultats obtenus après simulation pour les coefficients ΔΔMFCC

Résultat / les voyelles	Le nombre des paramètres sélectionnées parmi les 10 paramètres	le taux d'erreurs avant application du AG(en %)	le taux d'erreurs après application du AG(en %)
A	6	55	45
E	7	40	30
I	6	45	30
O	6	45	20
U	4	40	30

5.2.4.4. Coefficients MFCC+ΔMFCC+ΔΔMFCC

Les figures (Figure 5.19 à 5.23) et les tableaux (5.7 et 5.8) ci-dessous représentent les résultats obtenus pour les voyelles /a/, /e/, /i/, /o/ et /u/ respectivement pour les coefficients MFCC, ΔMFCC, ΔΔMFCC ensemble c'est-à-dire que obtenons cinq matrices de 40 vecteurs chaque matrice contient (5x40x30 = 6000 valeurs).

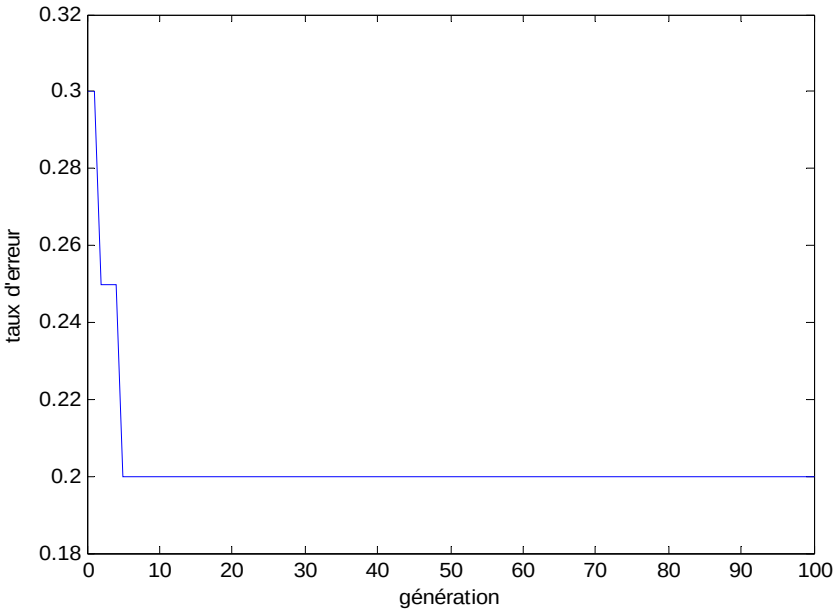

Figure 5.19 : variation du taux d'erreurs pour voyelle «a»

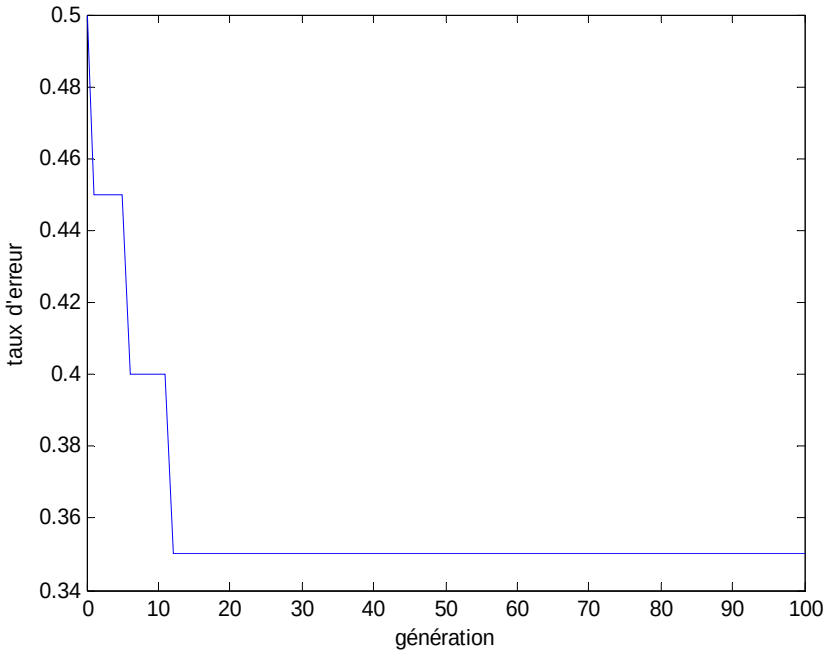

Figure 5.20 : variation du taux d'erreurs pour voyelle «e»

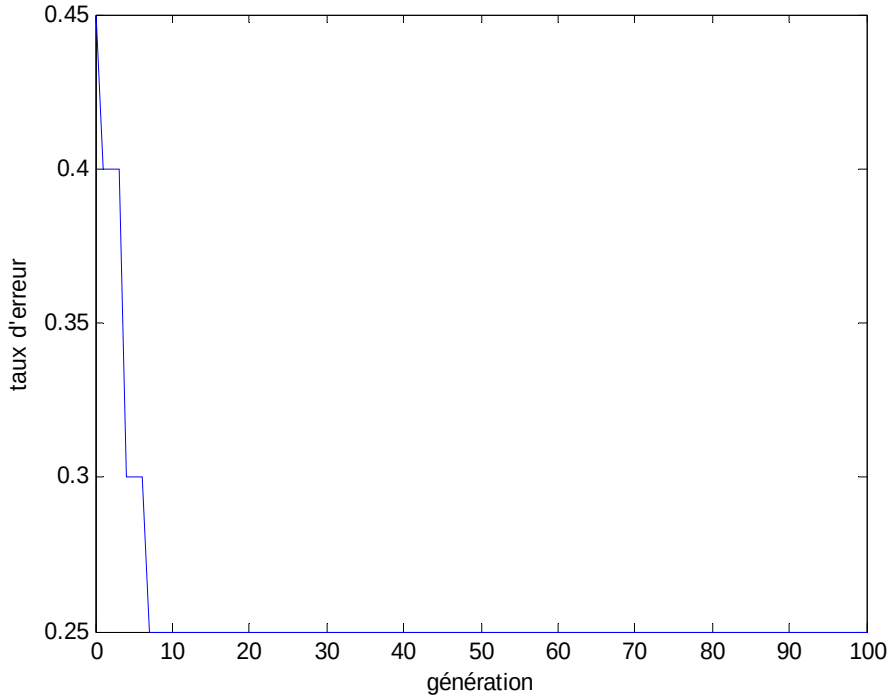

Figure 5.21 : variation du taux d'erreurs pour voyelle «i»

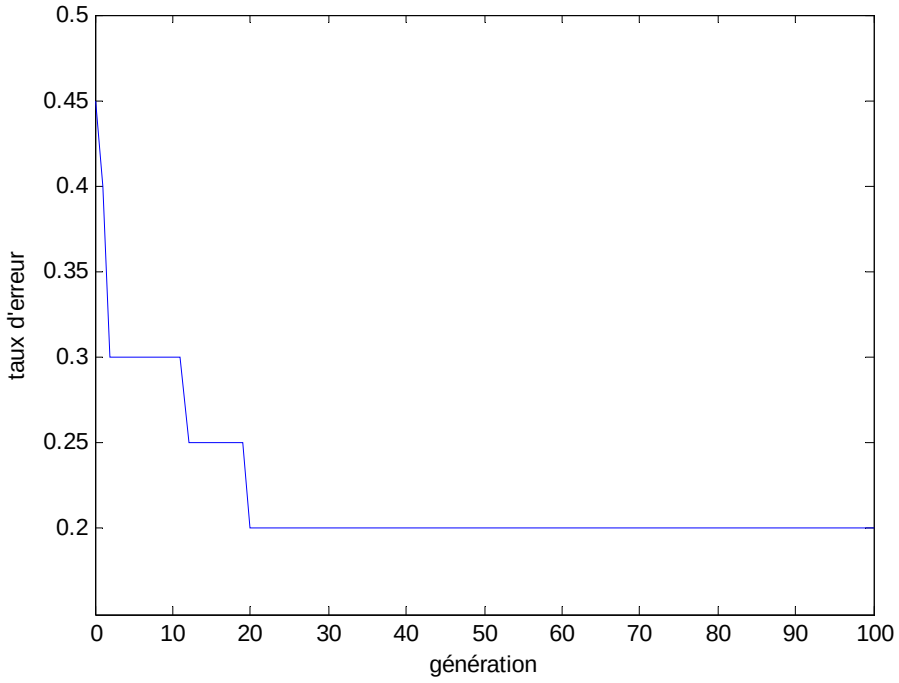

Figure 5.22 : variation du taux d'erreurs pour voyelle «o»

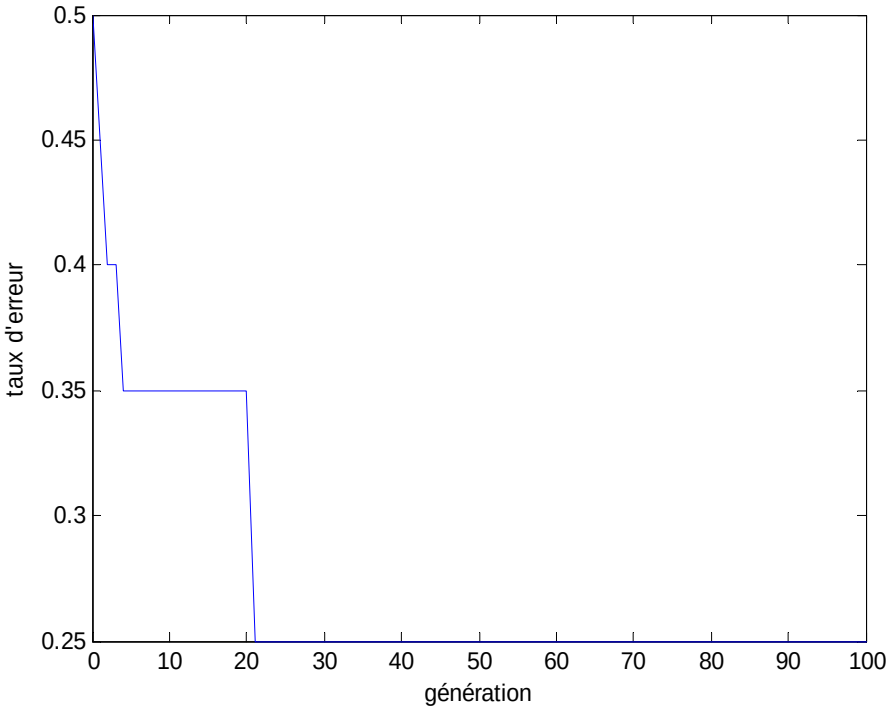

Figure 5.23 : variation du taux d'erreurs pour voyelle «u»

Tableau 5.7 résultat de la sélection les paramètres MFCC, ΔMFCC, ΔΔMFCC par AG

Coefficients / Les voyelles	Les coefficients sélectionnés parmi le vecteur des coefficients MFCC + ΔMFCC + ΔΔMFCC (c1…c30)
a	0 0 0 1 1 0 1 1 1 0 0 0 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 1 0 0
e	1 0 0 1 0 1 1 1 1 0 0 1 0 1 0 1 1 1 0 0 1 1 0 1 1 0 0 0 1 1
i	0 1 1 1 0 1 1 0 1 0 1 1 0 1 1 0 0 0 1 0 1 1 1 0 1 0 1 0 1 1
o	1 0 0 1 1 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0 1 1 1 1 0 1 1 1 0 0
u	0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 1 1 1 0 1 1

Tableau 5.8 résultats obtenus après simulation pour MFCC, ΔMFCC, ΔΔMFCC

resultats / les voyelles	Le nombre des paramètres sélectionnées parmi les 30 paramètres	le taux d'erreurs avant application du AG(en %)	le taux d'erreurs après application du AG(en %)
A	17	30	20
E	17	50	35
I	18	45	25
O	14	45	20
U	12	50	25

5.3. Discussion des résultats

A partir les figures et les tableaux présentés ci-dessus, nous constatons que les valeurs des taux d'erreurs pour les cinq voyelles diminuent (ex. la figure 5.4 qui représente la variation du taux d'erreurs pour la voyelle « a » : la valeur du taux est de 0.6 (60%), elle passe à 0.55 (55%) pour se stabiliser à la valeur 0.5 (50%)). Nous constatons aussi qu'il y a une réduction dans l'espace des coefficients MFCC. Les mêmes conclusions sont constatées pour les ΔMFCC, ΔΔMFCC ou pour les trois types de coefficients concaténés.

La combinaison des coefficients MFCC+ΔMFCC+ΔΔMFCC donne de meilleurs résultats par rapport aux MFCC, ΔMFCC et ΔΔMFCC pris séparément. Les coefficients ΔΔMFCC semblent donner de meilleurs résultats par rapport aux coefficients MFCC et ΔMFCC.

5.4. Conclusion

Ce cinquième et dernier chapitre a été consacré à l'implémentation d'un algorithme génétique pour la sélection des paramètres pertinents du locuteur et qui permettent de réduire non seulement la taille du vecteur de données mais en plus améliore les performances en réduisant les taux d'erreurs. Les résultats de simulation montrent que l'application de l'algorithme génétique améliore les taux de classification de 10 à 25% ce qui n'est pas négligeable. En plus, l'algorithme génétique permet une réduction de la dimension qui peut atteindre 60% (12 contre 30 coefficients).

Conclusion générale

Le travail présenté dans ce mémoire a été consacré à la mise en oeuvre d'une nouvelle technique de sélection des paramètres pertinents au locuteur par le biais d'le biais d'un algorithme génétique (AG) pour améliorer la performance et réduire la dimension. L'usage d'un algorithme génétique est adapté à une exploration rapide et globale d'un espace de recherche de taille importante fournissant un ensemble de solutions et non pas une solution unique.

Cette étude a été structurée suivant un fil conducteur qui sert de lien entre les différents parties en commençant par introduire le lecteur au domaine de la reconnaissance du locuteur, puis en lui fournissant les outils théorique et mathématiques nécessaires à l'analyse du signal parole ainsi que les méthodes de paramétrisation et d'extraction des paramètres pertinents. Cette première partie a été complétée par un chapitre détaillant la taxonomie des algorithmes génétiques, le principe de fonctionnement et les différents opérateurs tout en donnant au lecteur l'implémentation d'un algorithme simple illustrant les différentes notions et principes d'un AG. Une fois le lecteur bien armés nous avons détaillé l'implémentation d'un algorithme génétique pour la sélection des paramètres dans le cas d'un système RAL.

Les résultats de simulation ont montré une nette amélioration du taux de reconnaissance en plus d'une réduction de l'espace des données ce qui cruciale pour des applications à faibles ressources de calcul et de stockage tels que les PDA, les téléphone mobile, …etc. Les résultats valident que l'utilisation de l'AG minimise le taux d'erreurs de 5% à 25% avec une réduction de la dimension de 40% à 70%, ce qui permet une utilisation en temps réels des paramètres sélectionnés sans perte au niveau de la performance.

Références

(Atal, 1976)B. S. Atal. Automatic recognition of speakers from their voices. Proceedings of the IEEE, vol. 64, no. 4, pp. 460-475, Apr. 1976.

(Atal, 1974) B. S. Atal. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. JASA, vol. 55, pp. 1304-1312, Jun. 1974.

(Atal, 1972) B. S. Atal. Automatic speaker recognition based on pitch contours. The Journal of the acoustical society of America, no. 52, pp. 1687-1697, 1972.

(Aubert, 1993) X. Aubert & al. Improvement in connected digit recognition using linear discriminant analysis and mixture densities. Proceedings of the ICASSP, vol. 2, pp. 648-651, 1993.

(Bimbot, 1993) F. Bimbot & al. Assessment methodology for speaker identification and verification systems. Technical report – Task 2500 – Report 19, SAM-A ESPRIT Project 6819, 1993.

(Bogert, 1963). B. Bogert & al. The quefrency analysis of time series for echoes. Proceedings of Sumposium on Time Series Analysis, 1963.

(Bonastre, 1992) J. F. Bonastre & H. Meloni. A study of spectral variability for speaker characterization. 19ème JEP, p. 555, Juin 1992, Bruxelles, Belgique

(Bourlard, 1994) H. Bourlard & N. Morgan. Connectionist speech recognition : a hybrid approach. Kluwer Academic Press, 1994

(Campbell, 1997) J. P. Campbell, “Speaker Recognition: A Tutorial”, in Proceedings of the IEEE, 85(9)(1997), pp. 1437–1462.

(Cheung, 1978) R . S. Cheung & B. A. Esenstein. Feature selection via dynamic programming for text-independent speaker identification. IEEE Transactions on speech and audio processing, vol. 25, no. 6, pp. 397-403, Oct. 1978.

(Davis et Mermelstein, 1980) S. B. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans. Acoust., Speech, Signal Processing, 28(4):357-366, 1980.

(Doddington et al., 2000) G. R. Doddington, M. A. Przybocki, A F Martin, D. A. Reynolds, The NIST speaker recognition evaluation—overview, methodology, systems, results, perspective, Speech Commun (2000).

(Doddington, 1985) G. R. Doddington. Speaker recognition, identifying people by their voices. Proceedings of the IEEE, vol. 73, no. 11, pp. 1651-1664, Nov. 1985.

(Drygajlo, 2004) A. Drygajlo, Speech Coding and Recognition in Noisy Environments for Communication Terminals, Intelligent Integrated Media Communication Techniques, 2004, Part IV, 337-357.

(Drygajlo et al., 2003) A. Drygajlo, D. Meuwly et A. Alexander, "Statistical methods and Bayesian interpretation of evidence in forensic automatic speaker recognition", In EUROSPEECH-2003, 689-692.

(Duda et Hart, 1973) R.O. Duda & P.E. Hart.Pattern Classification and Scene Analysis John Wiley & Sons, 1973.

(Duda et al., 2000). Duda R.O., Hart P.E. and Stork D.G.(2000), “Pattern Classification”, Wiley Interscience,2000.

(Ezzaidi, 2002) H. Ezzaidi. Discrimination Parole/Musique et étude de nouveaux paramètres et modèles pour un système d’identification du locuteur dans le contexte de conférences téléphoniques. PhD thesis, Université du Québec, 2002.

(Furui, 1994) S. Furui. An overview of speaker recognition technology. Proceedings of the ESCA Workshop on Automatic Speaker Recognition Identification and Verification, pp. 1-9, 1994.

(Furui, 1989) S. Furui. Digital speech processing, synthesis and recognition. Marcel Dekker, N-Y, 1989.

(Furui, 1981b) S. Furui. Comparison of speaker recognition methods using static features and dynamic features.

(Furui, 1981) S. Furui. Cepstral analysis technique for automatic speaker verification. IEEE Trans. Acoust., Speech, Signal Processing, ASSP-29(2):254-272, 1981. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 3, pp. 342-350, Jun. 1981.

(Furui, 1981a) S. Furui. Cepstral analysis techniques for automatic speaker verification. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 29, no. 2, pp. 254-272, Aug. 1981.

(Goldberg, 1989) D.E Goldberg. Genetic Algorithms in Search, Optimization and Machine Learning. Reading MA Addison Wesley, 1989.

(Goldberg,1980) D.E Goldberg, " Algorithmes Génétiques : Exploration, Optimisation et Apprentissage Automatique"addition – wesley France (2 février 1980)

(Gold et Nelson, 2000) B. Gold and N. Nelson. Speech and Audio Signal Processing : Processing and Perception of Speech and Music. John Wiley and Sons, INC, 2000.

(Harrag et al., 2005) Harrag A., Mohamadi A., Serignat J.F. (2005), "LDA Combination of Pitch and MFCC Features in Speaker Recognition", INDICON, Chennai, India, 11–13, 237–240.

(Hermansky, 1993) H. Hermansky & al. Recognition of speech in additive and convolutional noise based on RASTA spectral processing. Proceedings of the IEEE, ICASSP, vol. 2, pp. 83-86, 1993.

(Hermansky, 1991) H. Hermansky & al. Compensation for the effect of the communication channel in auditory-like analysis of speech (Rasta-PLP). Proceedings of EUROSPEECH, pp. 1367-1370, 1991

(Hermansky, 1990) H. Hermansky, 1990. Perceptual linear predictive (PLP) analysis of speech. The Journal of the Acoustical Society of America 87, 1738–1752.

(Holland, 1975) Holland J. (1975), "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor, MI.

(Imprel et al., 1997) B. Imperl, Z. Kacic, and B. Horvat. A study of harmonic features for speaker recognition. Speech Communication, 22(4):385-402, 1997.

(Jin et Waibel, 2000) Q. Jin and A. Waibel. Application of LDA to speaker recognition. In Proc. Int. Conf. on Spoken Language Processing (ICSLP), 2000.

(Jollife, 2002) I. T. Jolliffe. Principal Component Analysis. New York : Springer-Verlag, 2002.

(Mammone,1996) R. J. Mammone & al. Robust speaker recognition. IEEE Signal Processing Magazine, pp. 58-71, Sept. 1996

(Matsui, 1994) T. Matsui & S. Furui. Similaritynormalization method for speaker verification based on a posteriori probability. Proceedings of the ESCA, Workshop on Automatic Speaker Recognition Identification and Verification, pp. 59-62, 1994.

(Mclachlan, 1992) G. J. McLachlan. Discriminant Analysis and Statistical Pattern Recognition. New York : Wiley,, 1992.

(Meuwly, 2000) Meuwly D. Reconnaissance de locuteur: Travail pour l'homme ou l'ordinateur? Crimiscope, Institut de Police Scientifique, Université de Lausanne, n°8, avril 2000.

(Michalewicz, 1992) Z.Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs. Springer-Verlag , (1992).

(**Michalewicz, 1991)** Z. Michalewicz and C.Z. Janikov. Handling constraints in genetic algorithms. In Proceedings of the Fourth International Conference on Genetic Algorithm. ICGA, 1991.

(Ng, 1995) K. T. Ng & al. Some no-parametric distance measures in speaker verification. EUROSPEECH 95, pp. 317-320, 1995.

(Noda, 1989) H. Noda. On the use of the information on individual speaker's position in the parameter space for speaker recognition. Proceedings of the ICASSP, pp. 516-519, 1989.

(Paliwal, 1992) K. K. Paliwal. Dimensionality reduction of the enhanced feature set for the HMM-based speech recognizer. Digital Signal Processing, pp. 157-173, 1992.

(Poritz , 1982) A. B. Poritz. Linear predictive hidden Markov models and the speech signal. Proceedings of the ICASSP 82, pp. 1291-1294, May 1982, Paris, France.

(Renders, 1995) J.M.Renders**,** "Algorithmes Génétiques et Réseaux de Neurones : Application a la commande de Processus "Edition, Paris, 1995

(Rosenberg, 1976) A. E. Rosenberg. Automatic speaker vérification: a review. Proceedings of the IEEE, vol. 64, no. 4, pp. 475-487, Apr. 1976.

(Sakoe, 1978) H. Sakoe & S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on acoustics, speech, and signal processing, vol. 26, pp. 43-49, 1978.

(Sambur, 1975) M. R. Sambur. Selection of acoustic features for speaker identification. IEEE Transactions on ASSP, vol. 2, no. 23, pp.176-182, Apr. 1975.

(Siohan, 1995) O. Siohan. On the robustness of linear discriminant analysis as a preprocessing step for noisy speech recognition. Proceedings of the ICASSP, pp. 125-128, 1995

(Sonmez et al., 1998) K. Sonmez, E. Shriberg, L. Heck, and M. Weintraub. Modeling dynamic prosodic variation for speaker verification. In Proc. Int. Conf. on Spoken Language Processing (ICSLP), pages 3189-3192, 1998.

(Stevens, 1957) S. S. Stevens, 1957. On the psychophysical law. Psychological Review 64, 153–181.

(Thyes et al., 2000) O. Thyes, R. Kuhn, P. Nguyen, and J.-C. Junqua. Speaker identification and verification using eigenvoices. In Proc. Int. Conf. on Spoken Language Processing (ICSLP), 2000.

(Van den Heuvel, 1994) Van Den Heuvel & al. Methodology aspects of segment speaker-related variability : a study of segmental durations in Dutch. Journal of Phonetics, no. 22, pp. 389-406, 1994.

(Withney, 1997) A. Whitney, A direct method of nonparametric measurement selection, IEEE Trans. Comput., (1997), 20, pp. 1100-1103.

MEMOIRE DE FIN D'ETUDES POUR L'OBTENTION DU DIPLOME DE MSTER EN GENIE ELECTRONIQUE

OPTION : CONTROLE INDUSTRIEL

Proposé et dirigé par : Dr. Dj. SAIGAA & Dr. A. HARRAG

Présenté par : BOUKHAROUBA. Kheira

THEME : UTILISATION DES APPROCHES EVOLUTIONNAIRES POUR LA SELECTION DES PARAMETRES PERTINENTS : APPLICATION A L'AUTHENTIFICATION DES PERSONNES

Résumé :
Les caractéristiques acoustiques représentant le conduit vocal ont été largement appliquées pour la reconnaissance du locuteur. Bien qu'il ait été révélé que la phonation glottique joue un rôle important dans la caractérisation du locuteur, l'utilité des caractéristiques de la source vocale pour la reconnaissance automatique du locuteur, ainsi que sa technique efficace d'extraction des caractéristiques, n'a pas été pleinement exploitée. L'objectif de la thèse est de proposer une nouvelle méthode d'extractions des traits du locuteur à partir d'une base de données parole, cette extraction centrée sur la recherche d'une meilleure représentation de l'information pertinente spécifique au locuteur est basée sur l'utilisation des algorithmes génétiques.

Mots clés : caractéristiques du locuteur, algorithmes génétiques, fusion des données.

Abstract :
The acoustic features representing the vocal tract have been widely applied to speaker recognition. Although it was revealed that glottal phonation plays an important role in speaker characterization, the usefulness of voice source characteristics for speaker recognition, and its efficient technique for feature extraction 'has not been fully exploited. The objective of this thesis is to propose new extraction method of speaker features from a speech database, this extraction focused on finding a better representation of relevant information specific to the speaker based on genetic algorithms.

Keywords: speaker features, acoustic and prosodic features, genetic algorithms, feature fusing.

الخلاصة:
شكلت الخصائص الصوتية التي تمثل القناة الصوتية تطبيقها على نطاق واسع في التعرف على الأشخاص .وعلى الرغم من كشف النقاب عن أن معالجة تردد الحبال الصوتية تلعب دورا هاما في تحديد خصائص الناطق، وفائدة خصائص مصدر صوت في التعرف على الأشخاص، ولها أسلوب فعال لاستخراج الخصائص،لم تستغل بالكامل .الهدف من هذه الرسالة هو اقتراح أسلوب جديد لاستخراج سمات الناطق من قاعدة بيانات صوتية، وهذا الاستخلاص تركز على إيجاد تمثيل أفضل للمعلومات ذات الصلة المحددة الناطق باستعمال الخوارزميات الجينية.

الكلمات الشائعة: تحديد هوية الناطق، قاعدة بيانات صوتية، الخوارزميات الجينية، خصائص المسالك الصوتية.