

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

**Université Saâd DAHLAB, Blida**

N° D'ordre:.....



Faculté des sciences

**Département d'informatique**

Présenté par :

ALLOUCHE Sonia      et      SAADI Sihem

**En vue d'obtenir le diplôme de master**

Domaine : Mathématique et informatique

Filière : Informatique

Spécialité : Informatique

Option : Génie des systèmes informatique

Sujet :

**Systeme d'aide à la décision pour l'évaluation de l'intelligibilité  
de la parole dysarthrique**

Soutenu le : 21/06/2016

Mme Boumahdi.F

Président

Mr Nahal.D

Examineur

Proposé et dirigé par

M. YKHLEF Fayçal

Suivi par

M. OUKID Lamia

Promotion 2015/2016

# *Remercîment*

*Nous remercions tout d'abord ALLAH tout puissant de nous avoir donné le courage, la force et la patience d'achever ce modeste travail.*

*Nous voudrions adresser toute notre gratitude au directeur de ce mémoire, Dr YKHLEF Fayçal, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.*

*Nous remercions aussi Dr OUKID Lamia de nous avoir aidés à réaliser ce mémoire avec beaucoup d'efforts et de patience.*

*Nos plus profonds remerciements vont à nos parents. Tout au long de notre cursus, ils nous ont toujours soutenu, encouragé et aidé. Ils ont su nous donner toutes les chances pour réussir. Qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de notre plus affectueuse gratitude.*

*Merci.*

*Melle Sihem et Mlle Sonia.*

# *Dédicace*

*Je dédie ce mémoire à mes parents pour leur soutien, leur confiance  
et sur tout leur amour inestimable j'espère qu'ils trouvent en moi la  
source de leur fierté à qui je dois tout*

*À mes très chers frères et sœurs Sarah, Selma, Abdelkader et Anes*

*À qui je souhaite un avenir radieux plein de réussite*

*À mes chères cousine Nesrine et Hadjer*

*Et à toutes mes amies*

*Saadi Sihem*

# *Dédicace*

*Je dédie ce mémoire à :*

*Mes parents :*

*Ma mère, qui a œuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.*

*Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi.*

*Mes frères Sidali et Ilyes et ma sœur Romaiissa.*

*A mes chers grands parents, mes tantes, mes oncles et toute ma famille.*

*A mes chers amis.*

*ALLOUCHE Sonia*

# Résumé

La mesure du degré d'intelligibilité de la parole dysarthrique permettra aux médecins de prendre les précautions nécessaires pour lutter contre la progression de la dysarthrie.

Le développement d'un système objectif de mesure de l'intelligibilité de la parole produite par des patients dysarthriques est une bonne alternative à l'approche subjective de diagnostic. Nous proposons dans ce mémoire une méthode automatique de mesure objective de la maladie de la dysarthrie. Nous nous focalisons sur la dysarthrie spastique. La méthode proposée est composée de deux étages de traitements : (i) l'extraction d'attributs et (ii) la classification des niveaux d'intelligibilité. Nous avons exploité quatre différents groupes d'attributs : (1) l'entropie des coefficients en sous bandes selon l'échelle audiométrique, (2) les coefficients Cepstraux de MEL, (3) le taux de silence à l'intérieur d'un mot isolé et (4) le niveau vocal actif. La classification est basée sur les mélanges de gaussiens. Deux niveaux d'intelligibilité ont été considérées dans notre application : niveau très faible et niveau élevé. Pour évaluer les performances de la méthode proposée, nous avons utilisé la base UA speech. Les résultats que nous avons obtenus montrent que notre système donne un taux de bonne classification de 96%.

# Abstract

The measurement of the degree of dysarthric speech intelligibility allows doctors to take the necessary precautions to fight against the spread of dysarthria. The development of an objective system for the assessment of the intelligibility of speech uttered by patients suffering from dysarthria is a good alternative to subjective diagnosis.

We propose in this thesis an automatic method of objective measurement which consists of two processing stages: (i) feature extraction and (ii) intelligibility classification. We focus in this study on Spastic dysarthria.

We have exploited four different groups of features: (1) entropies of the sub-band coefficients based on the audiometric scale, (2) the Mel Frequency Cepstral Coefficients, (3) the silence rate within an isolated word and (4) the active speech level. The classification is based on the Gaussian mixture models. Two intelligibility levels were considered in our application: very low and high. We have used the Universal Access Speech (UA Speech) database for the performance evaluation of the proposed method. According to the experimental results, we have found that the classification accuracy is equal to 96%.

# ملخص

قياس وضوح لفظة الديزارثري تسمح للأطباء بأخذ الاحتياطات اللازمة لمحاربة تطور مرض الديزارثري, برمجة نظام موضوعي لتقييم نطق الشخص الديزارثري هي مبادرة جيدة للنهج الموضوعي نقترح في هذه المذكرة طريقة البية للقياس الموضوعي المكون من مستويين من المعالجة.

1-استخراج المعاملات

2-تصنيف مستويات درجة الوضوح

سنقوم بتشغيل 4 أفواج مختلفة من المعاملات

\*معاملات الانتروبيا حسب السلم السمعي

\*معاملات سبسترومال

\* نسبة الصمت داخل الكلمة

\*تصنيف قوسيان

هناك مستويات من درجة الوضوح نأخذ بعين الاعتبار مستويين من درجة الوضوح في تطبيقنا

\*المستوى الضعيف جدا

\*لمستوى المرتفع

لتقييم أداء الطريقة المقترحة، استخدمنا قاعدة البيانات

النتائج التي حصلنا عليها تبين أن نظامنا يعطي نسبة فعالية مرتفعة تصل إلى 96 بالمائة

# *Table des matières*

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

## **Chapitre 1 : Troubles d'articulation de la parole: Cas de la dysarthrie**

1.1. Introduction.....	5
1.2. Dysarthrie.....	5
1.2.1. Définition .....	5
1.2.2. Symptômes généraux.....	5
1.2.3. Types de la maladie de dysarthrie .....	6
1.2.4. Causes .....	8
1.3. Intelligibilité de la parole dysarthrique.....	8
1.3.1. Définition de l'intelligibilité.....	8
1.3.2. Méthodes d'évaluation de l'intelligibilité .....	8
1.3.2.1. Evaluation subjective .....	9
1.3.2.2. Evaluation objective.....	9
1.3.3. Systèmes d'aide à la décision en médecine.....	10
1.4. Analyse acoustique de la parole.....	11
1.4.1. Voyelles isolées .....	11
1.4.1.1. Pente mélodique du second formant(intonation) .....	12
1.4.1.2. Espace de la zone-voyelle .....	12
1.4.1.3. Rapport de centralisation du formant (RCF) .....	14
1.4.2. Consonnes isolées.....	14
1.5. Conclusion .....	15

## **Chapitre 2 : Evaluation objective de l'intelligibilité de la parole dysarthrique**

2.1. Introduction.....	17
2.2. Evaluation objective de l'intelligibilité .....	17
2.2.1. Méthodes basées sur des références.....	17
2.2.2. Méthodes sans références.....	19
2.3. Corpus et bases de données de test.....	19
2.3.1. Nemours.....	19
2.3.2. Torgo.....	23
2.3.3. UA speech.....	23
2.4. Conclusion .....	23

## Chapitre 3 : Système automatique proposé

3.1.	Introduction.....	28
3.2.	Approche proposée.....	28
3.2.1.	Collection des données .....	30
3.2.2.	Prétraitement .....	30
3.2.2.1.	Elimination du silence.....	31
3.2.2.2.	Segmentation en trames stationnaires .....	32
3.2.3.	Extraction d'attributs .....	32
3.2.3.1.	Entropies des coefficients en sous bande selon l'échelle audiométriqu .....	33
3.2.3.2.	Coefficients Cepstraux de MEL.....	34
3.2.3.3.	Taux de silence à l'intérieur d'un mot isolé .....	36
3.2.3.4.	Niveau vocal actif .....	38
3.2.4.	Normalisation d'attributs .....	39
3.2.5.	Classification.....	40
3.2.5.1.	Pourquoi nous avons utilisé les modèles GMMs ? .....	40
3.2.5.2.	Description du modèle .....	41
3.2.5.3.	Estimation des paramètres.....	41
3.2.5.4.	Idée de base de l'algorithme EM.....	41
3.2.6.	Mesure de performance.....	43
3.3.	Conclusion.....	46

## Chapitre 4: Expérimentations, résultats et interprétations

4.1.	Introduction.....	48
4.2.	Logiciels de développement .....	48
4.2.1.	Environnement MATLAB .....	48
4.2.2.	Langage JAVA .....	48
4.3.	Corpus de tests .....	49
4.3.1.	Evaluation du système globale .....	49
4.3.2.	Partie apprentissage .....	51
4.3.3.	Partie de test .....	51
4.4.	Mesure de performance.....	52
4.4.1.	Matrice de confusion .....	52
4.4.2.	Variation du nombre de gaussiennes .....	52
4.4.2.1.	Matrice de confusion avec nombre de gaussiennes N=2.....	52
4.4.2.2.	Matrice de confusion avec nombre de gaussiennes N =3.....	54

4.4.2.3.	Matrice de confusion avec nombre de gaussiennes $N=4$ .....	55
4.4.2.4.	Matrice de confusion avec nombre de gaussiennes $N=5$ .....	55
4.5.	Résultats .....	58
4.5.1.	Etapes du traitemants .....	59
4.5.1.1.	Interface d'accueil.....	59
4.5.1.2.	Interface du menu principale .....	59
4.5.1.3.	Traitement .....	60
4.5.1.4.	Résultats .....	60
4.6.	Conclusion .....	62
	Conclusion Générale.....	64

## Références

## Annexes

Annexe A.....	70
Annexe B.....	71

# Liste des figures

Figure 1.1 : Triangle vocalique.....	13
Figure 1.2 : Quadrilatère vocalique.....	13
Figure 2.1 : Méthodes d'estimation du niveau d'intelligence basé sur une.....	19
Figure 2.2 : Méthodes d'estimation du niveau d'intelligibilité sans références. ....	21
Figure 2.3 : Equipements utilisés pour l'enregistrement des mots de la base de.....	24
Figure 3.1 : Etapes de réalisation du système.....	29
Figure 3.2 : Exemple d'utilisation de l'algorithme EPD.....	32
Figure 3.3 : Analyse multi résolution. ....	34
Figure 3.4 : Extraction des MMFCC.....	36
Figure 3.5 : Enveloppe d'un signal. ....	37
Figure 3.6 : Taux de silence à l'intérieur d'un mot isolé.....	38
Figure 3.7 : Représentation des modèles gaussiens. ....	43
Figure 4.1 : Exemple d'un son plein de bruit.....	50
Figure 4.2 : Exemple d'un son où le médecin parle au cours de l'enregistrement. ....	51
Figure 4.3 : Schéma simplicité de la méthode d'évaluation du système global. ....	52
Figure 4.4 : Graphe de changement du TBC selon la variation du nombre de gaussiennes.....	57
Figure 4.5 : Interface d'accueil.....	58
Figure 4.6 : Interface du menu principale.....	60
Figure 4.7 : Lancement du traitement d'un son.....	61
Figure 4.8 : Exemple d'un patient avec degré d'intelligibilité élevé.....	62
Figure 4.9 : Exemple d'un patient avec degré d'intelligibilité très faible.....	62

# *Liste des tableaux*

<i>Tableau 1.1: Les différents types de la dysarthrie et leurs caractéristiques.....</i>	<i>7</i>
<i>Tableau 2.1 : Lase de données UA speech (22). .....</i>	<i>25</i>
<i>Tableau 3.1: Caractéristiques des patients dysarthriques spastiques .....</i>	<i>30</i>
<i>Tableau 3.2: Matrice de confusion. ....</i>	<i>44</i>
<i>Tableau 4.1 : Matrice de confusion avec un nombre de Gaussiennes <math>N=2</math>. ....</i>	<i>53</i>
<i>Tableau 4.2 : Taux de performance ave un nombre de Gaussiennes <math>N=2</math>. ....</i>	<i>53</i>
<i>Tableau 4.3 : Matrice de confusion avec un nombre de Gaussiennes <math>N=3</math>. ....</i>	<i>54</i>
<i>Tableau 4.4 : Taux de performance avec un nombre Gaussiennes <math>N=3</math>. ....</i>	<i>54</i>
<i>Tableau 4.5 : Matrice de confusion avec un nombre de Gaussiennes <math>N=4</math>. ....</i>	<i>55</i>
<i>Tableau 4.6 : Taux de performance avec un nombre de Gaussiennes <math>N=4</math>. ....</i>	<i>55</i>
<i>Tableau 4.7 : Matrice de confusion avec le nombre de Gaussiennes <math>N=5</math>. ....</i>	<i>56</i>
<i>Tableau 4.8 : Taux de performance avec un nombre de Gaussiennes <math>N=5</math>. ....</i>	<i>56</i>

# *Liste des abréviations*

*DS : Dysarthrie Spastique*

*DCT: Discret Cosine Transform*

*EM: Expectation Maximization*

*EPD: End Point Detection*

*FFT: Fast Fourier Transform*

*GMMs: Gaussien Mixture Models*

*HMMs: Hidden Markov Models*

*LDA: Linear Discriminative Analyse*

*MFCC : Mel-Frenquency Cepstral Coeficients*

*ML: Maximum Likelihood*

*PC: Paralyse Cérébrale*

*RAP : Reconnaissance Automatique de la Parole*

*RCF : Rapport de Centralisation du Formant*

*SIT :Synthese Intelligibility Test*

*SLA : Sclérose Latérale Amyotrophique*

*SVM: Support Vector Machine*

*TBC : Taux de Bonne Classification*

*TFP : Taux de Faux Positif*

*TFN : Taux de Faux Négatif*

*TVP : Taux de Vrai Positif*

*TVN : Taux de Vrai Négatif*

---

---

***INTRODUCTION GENERALE***

---

---

## *Introduction générale*

---

La dysarthrie, est un trouble de l'articulation vocalique causé par des lésions du cerveau, ou une atteinte au niveau des différents organes, ayant un rôle dans la production de la parole comme la langue, ou le voile du palais. Il existe plusieurs types de dysarthrie, nous citons : la dysarthrie paralytique, corticale, lacunaire, Hypokinétique, et spastique [1]. La différence entre ces types réside dans la manière de coordination des muscles de phonation.

Pour la majorité des types de dysarthries, la voix apparaît trop grave ou trop aiguë, rauque, scandée, explosive, et parfois complètement éteinte.

De nos jours, les traitements cliniques de cette maladie cognitive ne permettent pas de guérir définitivement les patients dysarthriques, cependant ils permettent d'empêcher la maladie d'évoluer. L'estimation précoce du degré d'intelligibilité de la parole dysarthrique permettra aux médecins d'agir à temps et de prendre les précautions nécessaires pour lutter contre la progression de la maladie.

La mesure du degré de l'intelligibilité de la parole dysarthrique peut s'effectuer par des méthodes subjectives et objectives. Les méthodes subjectives nécessitent l'intervention des cliniciens et spécialistes en orthophonie. Il s'agit en générale des tests d'écoutes. Ces méthodes prennent beaucoup de temps et sont souvent partiales et incompatibles en raison de leur nature perceptive. Cependant, les méthodes objectives visent à mesurer automatiquement le degré d'intelligibilité sans avoir recours à des interventions humaines. Elles sont généralement basées sur des techniques acoustiques, mathématiques et informatiques, c'est ce qui rend cette méthode d'évaluation moins coûteuse et plus rapide.

À notre connaissance, il n'y a aucun système de mesure objective de l'intelligibilité de la parole dysarthrique qui a été commercialisé dans le marché des dispositifs médicaux. Seulement des propositions de laboratoires cliniques et des équipes de recherche ont été reportées dans la littérature.

L'évaluation objective de l'intelligibilité de la parole est donc un sujet d'actualité qui est toujours dans le stade de la recherche scientifique.

Quelques solutions de recherche ont été proposé dans ce domaine.

L'identification des mesures acoustiques (ou attributs) pertinentes qui peuvent révéler le degré d'intelligibilité de la parole dysarthrique est l'objectif ultime de cette recherche. Les chercheurs dont la formation de base est en phonétique clinique médicale ont proposé des mesures robustes d'identification et de mesure de l'intelligibilité de la parole dysarthrique. Nous citons les mesures qui découlent de l'analyse acoustique des voyelles à

## *Introduction générale*

---

savoir : la pente mélodique du premier Formant, l'espace des zones voyelles et le rapport de centralisation des formants [2]. L'inconvénient majeur de ces mesures est lié directement à la difficulté de l'identification des voyelles dysarthriques à l'intérieur d'un message vocal continu (mot isolé ou parole continue) qui est une étape nécessaire avant l'estimation propre des mesures. A cet effet, cette approche reste seulement valable pour les voyelles isolées dans un contexte bien déterminé. De plus, ces études ont été conduites pour une langue bien précise (souvent l'Anglais Américain). D'autres chercheurs ont proposé d'exploiter un ensemble d'attributs développés initialement pour d'autres applications en traitement de la parole et de les combiner avec des modèles de prédiction mathématiques (souvent des machines d'apprentissages) [3]. Les mesures choisies caractérisent « grossièrement » le degré d'intelligibilité de la parole dysarthrique. La limite principale de ces méthodes est reliée à l'étape de l'extraction d'attributs. Les mesures acoustiques proposées sont souvent destinées à une ou plusieurs catégories acoustiques spécifiques du signal vocal (les sons voisés, non-voisés, mixtes, transitoires, chuchotés, impulsionnels et autres). Alors, la séparation de l'onde acoustique du signal en composantes connexes est obligatoire. La difficulté de séparation automatique de ces zones, surtout dans le cas des voix pathologiques, diminue les performances des méthodes proposées.

Dans notre étude, nous nous sommes focalisés sur la dysarthrie spastique. Nous nous sommes basés sur les critères suivants pour le choix préliminaire des attributs :

- L'opération d'extraction d'attributs doit être automatique et ne nécessite aucune intervention humaine ;
- La nature acoustique des sons de la base de données (mots isolés) doit être prise en considération dans le choix des mesures spécifiques (sons voisés, non voisés, mixtes, impulsionnels et autres) ;
- Toutes les étapes nécessaires à l'accomplissement de cette tâche doivent être prises en considération dans notre application (Fenêtrage, élimination du silence, rehaussement du signal et autres).

Nous avons exploité dans notre projet quatre différents groupes d'attributs :

- Entropies des coefficients en sous bandes selon l'échelle audiométrique ;
- Coefficients Cepstraux de MEL ;
- Taux de silence à l'intérieur d'un mot isolé ;
- Niveau vocal actif.

## *Introduction générale*

---

La modélisation mathématique (pour réaliser l'opération de classification) est basée sur les mélanges des gaussiens. Deux niveaux d'intelligibilité ont été considérées dans notre application : intelligibilité très faible et élevée.

Notre mémoire est scindé en quatre chapitres, deux chapitres théoriques, et deux pratiques. Dans le premier chapitre, nous introduisons la maladie de la dysarthrie ainsi que les différents types de son évaluation (subjective et objective). Dans le second chapitre, nous décrivons les méthodes d'évaluation objectives. Les principales bases de données des voix dysarthriques sont aussi citées dans le même chapitre. Le troisième chapitre est dédié à la présentation de notre système ainsi qu'aux différentes étapes de sa construction. Le quatrième chapitre est consacré aux résultats expérimentaux. Pour clore ce mémoire, une conclusion générale sera présentée à la fin.

## **1.1. Introduction**

La dysarthrie est un trouble d'articulation de la parole d'origine neurologique [1]. Les patients atteints de dysarthrie ont une incoordination au niveau de la production de la parole.

Dans ce chapitre, nous présentons un aperçu général sur cette maladie en se focalisant sur quatre points essentiels : (i) les symptômes, (ii) les types, (iii) les causes (iv) et l'évaluation de l'intelligibilité de la parole.

## **1.2. Dysarthrie**

L'étude de pathologies motrices de la parole nécessite des connaissances cliniques préalables. C'est l'objectif principal de cette section.

### **1.2.1. Définition**

La dysarthrie est un trouble de l'articulation causé par des lésions du cerveau ou une atteinte au niveau des différents organes ayant un rôle dans la parole comme la langue ou le voile du palais. Elle touche principalement les adultes. La dysarthrie se manifeste par des difficultés d'articulation, une altération de la voix et du débit de parole. Des pathologies peuvent être à l'origine de la dysarthrie, comme la maladie de Parkinson, un accident vasculaire cérébral, une paralysie faciale ou la sclérose en plaques. Une rééducation orthophonique peut améliorer l'état de la personne atteinte par la dysarthrie [4].

### **1.2.2. Symptômes généraux**

La dysarthrie peut apparaître soudainement ou bien elle peut se développer lentement. Les symptômes se remarquent donc surtout au niveau de la parole [1].

- Au niveau de la parole :
  - Mauvaise articulation ;
  - Rythme irrégulier, trop lent ou trop rapide ;
  - Marmonnée ;
  - Chuchotée.
  
- Au niveau de la voix :
  - Nasillarde ;
  - Nez bouché ;
  - Rauque ;
  - Tendue ;

- Soufflée.

### 1.2.3. Types de la maladie de dysarthrie

En fonction de la lésion qui altère le langage, plusieurs formes de dysarthrie ont été établies. Le tableau ci-dessous donne un aperçu sur les formes les plus courantes de cette pathologie :

Type	Caractéristiques	Voix/parole/discours
Dysarthrie flasque ou paralytique	<ul style="list-style-type: none"> <li>• Hypotonie</li> <li>• Faiblesse des muscles</li> </ul>	La voix peut être : <ul style="list-style-type: none"> <li>• Essoufflée</li> <li>• Nasale</li> </ul>
Dysarthrie spastique (DS)	<ul style="list-style-type: none"> <li>• Augmentation du tonus musculaire (par atteinte bilatérale du premier motoneurone central).</li> <li>• Parésie / spasticité des autres parties du corps</li> <li>• Réflexes hyperactifs</li> <li>• Réflexes pathologiques</li> <li>• Dysphagie</li> </ul>	<ul style="list-style-type: none"> <li>• La voix est :                             <ul style="list-style-type: none"> <li>○ Rauque râpeuse</li> <li>○ Étranglée</li> <li>○ Faible</li> </ul> </li> <li>• Tonalité cassée</li> <li>• Discours accéléré</li> <li>• Émission de phrases courtes</li> <li>• Parole laborieuse</li> <li>• Distorsions faciales</li> </ul>
Dysarthrie corticale	Perte soudaine de la parole et paralysie bilatérale des muscles	Articulation impossible mais compréhension normale
Dysarthrie lacunaire	Association d'une : <ul style="list-style-type: none"> <li>• Dysphagie</li> <li>• Parésie faciale</li> <li>• Maladresse de la main</li> </ul>	<ul style="list-style-type: none"> <li>• Troubles de la phonation et de la déglutition</li> </ul>
Dysarthrie hypokinétique	L'hypokinésie affecte : <ul style="list-style-type: none"> <li>• Les muscles respiratoires</li> <li>• Les muscles laryngés</li> <li>• Les muscles palatopharyngés et buccaux</li> </ul>	<ul style="list-style-type: none"> <li>• La voix :                             <ul style="list-style-type: none"> <li>○ Lente et monotone</li> <li>○ Devient sourde</li> <li>○ Perd son volume et ses inflexions</li> </ul> </li> <li>• L'élocution est :                             <ul style="list-style-type: none"> <li>○ Inintelligible</li> <li>○ Marquée par des répétitions</li> </ul> </li> </ul>

Dysarthrie hyperkinétique rapide	<ul style="list-style-type: none"> <li>• Atteinte du système extrapyramidal, dysfonctionnement des noyaux gris centraux</li> <li>• Cette dysarthrie se caractérise par des mouvements rapides et involontaires</li> </ul>	<ul style="list-style-type: none"> <li>• Élocution rapide</li> <li>• La voix est : <ul style="list-style-type: none"> <li>○ Forcée</li> <li>○ Rauque</li> <li>○ Irrégulière</li> <li>○ Mal coordonnée par rapport à la respiration</li> </ul> </li> </ul>
Dysarthrie hyperkinétique lente	Atteinte du système extrapyramidal entraînant une akinésie	<p>La voix est :</p> <ul style="list-style-type: none"> <li>• Étranglée</li> <li>• Stridente</li> </ul> <p>L'élocution est :</p> <ul style="list-style-type: none"> <li>• Sourde</li> <li>• Lente</li> <li>• Laborieuse</li> </ul>
Dysarthrie ataxique	<ul style="list-style-type: none"> <li>• Lésion du cervelet et des voies cérébelleuses qui affecte la coordination et le contrôle des muscles de la phonation</li> <li>• Elle se rencontre en cas de sclérose en plaques notamment</li> </ul>	<ul style="list-style-type: none"> <li>• Élocution lente et irrégulière : <ul style="list-style-type: none"> <li>○ Bredouillement</li> <li>○ Séparation anormale des mots</li> </ul> </li> <li>• Ton impossible à maintenir à une même hauteur et à une même intensité</li> <li>• Impression que la personne est en colère</li> <li>• Chuchotement impossible</li> </ul>
Dysarthrie parkinsonienne	<p>Rencontrée au cours de la maladie de Parkinson, elle se caractérise par :</p> <ul style="list-style-type: none"> <li>• Une perte d'intensité de la voix ;</li> <li>• Une perte des modulations.</li> </ul>	<ul style="list-style-type: none"> <li>• Réduction de l'intelligibilité globale de la parole</li> <li>• Voix basse malgré des efforts importants</li> <li>• Timbre monotone, accents intonatifs inexistantes</li> <li>• Débit variable : <ul style="list-style-type: none"> <li>○ Accélération subites</li> <li>○ Silences inappropriés</li> </ul> </li> </ul>

Tableau 1.1 : Les différents types de la dysarthrie et leurs caractéristiques [1].

#### **1.2.4. Causes**

La dysarthrie affecte le cerveau. Cela peut se produire à la naissance, comme dans le cas où la personne atteint une paralysie cérébrale, ou elle peut se produire plus tard dans la vie en raison de l'une des nombreuses causes suivantes [5]:

- Un Accident Vasculaire Cérébral (SLA);
- Une commotion cérébrale ;
- Une tumeur ;
- La maladie de Parkinson [6];
- La maladie de Lou Gehrig [7];
- La sclérose latérale amyotrophique (SLA);
- La maladie de Huntington [8];
- La maladie de sclérose en plaque.

Chez l'enfant, les causes ne sont pas nombreuses :

- Les atteintes du tronc cérébral, le plus souvent par des problèmes traumatiques ou tumoraux ;
- La myasthénie ;
- Certaines myopathies ;
- Les tumeurs du cervelet, très rares.

### **1.3.Intelligibilité de la parole dysarthrique**

#### **1.3.1. Définition de l'intelligibilité**

L'intelligibilité est définie comme le degré de précision avec lequel le message est compris par l'auditeur. Elle s'obtient en déterminant le nombre d'unités de parole reconnus par l'auditeur. Dans le sens le plus général, l'intelligibilité est la capacité d'un individu à être compris verbalement par un autre [9].

Dans le cas de notre étude, mesurer le degré de la sévérité de la dysarthrie revient à mesurer l'intelligibilité de cette maladie.

#### **1.3.2. Méthodes d'évaluation de l'intelligibilité**

L'évaluation des compétences verbales et communicationnelles est indispensable afin de déterminer le niveau d'intelligibilité et de comprendre les troubles du patient dysarthrique, dans le souci d'élaborer un plan thérapeutique adapté afin de réduire l'aggravation de la

maladie. La démarche de questionnement du thérapeute doit mettre en lien la perception et la sévérité des troubles [10].

Il existe deux formes d'évaluation :

- Evaluation subjective,
- Evaluation objective.

### **1.3.2.1.Evaluation subjective**

L'évaluation subjective de l'intelligibilité de la parole dysarthrique se fait à l'écoute de la parole du patient. Elle s'attache particulièrement à :

- La netteté et la précision des voyelles et des consonnes, à la réalisation de groupe de consonnes (coarticulation) ;
- L'intonation est étudiée à travers la facilité du discours ;
- La longueur des pauses ;
- Les changements de rythme et dans son maintien au niveau mélodique ou accentuel [11].

Elle nécessite la présence de deux personnes : (i) le clinicien et (ii) le transcripneur. Le rôle du clinicien est de sélectionner les mots et les phrases de test qui seront prononcés par le patient. Le transcripneur (une personne peu familière avec ces mots et des phrases) écoute les enregistrements des patients, et les transcrit en phrases phonémiques.

L'évaluation consiste généralement à mesurer les capacités vocales des patients. Il existe deux tests publiés pour évaluer l'intelligibilité des troubles moteurs de la parole :

- Evaluation de l'intelligibilité des locuteurs dysarthriques (*Assessment of Intelligibility in Dysarthric Speakers (AIDS)*) [12];
- Test de l'intelligibilité de la phrase (*the Sentence Intelligibility Test (SIT)*).

La différence entre ces deux méthodes réside dans le choix des unités phonétiques de test et le nombre de points de l'échelle d'évaluation. Pour plus de détails sur ces tests subjectifs, le lecteur peut consulter les références.

### **1.3.2.2.Evaluation objective**

Le meilleur standard pour l'évaluation de l'intelligibilité des troubles dysarthriques est l'évaluation subjective réalisée par un expert. Les tests subjectifs sont souvent coûteux, partiels, en raison de leur nature perceptive. En revanche, les méthodes d'évaluation objective, sont automatiques, répétables et relativement pas chers.

L'évaluation objective de l'intelligibilité des troubles de la parole repose sur deux approches complémentaires [13]:

○ Analyse acoustique :

Elle est basée sur l'exploration des mesures acoustiques qui révèlent les signes de pathologie. Dans le cas où les mesures identifiées sont jugées suffisamment robustes, elles peuvent être utilisées directement pour générer une estimation quantitative de l'intelligibilité perçue. Des exemples de ces mesures sont donnés dans la section 1.3.

○ Modélisation mathématique des données acoustiques :

Si on ne dispose d'aucune information a priori sur la fiabilité des mesures acoustiques extraites à partir de l'onde acoustique de la voix, on ne peut pas quantifier le degré de l'intelligibilité d'une manière directe. La modélisation mathématique de ces mesures est nécessaire pour extraire les connaissances à partir des données acoustiques. Le processus d'extraction de connaissances est accompli par le biais des machines d'apprentissage. Des exemples pratiques sur ces modèles seront donnés dans le chapitre suivant.

### **1.3.3. Systèmes d'aide à la décision en médecine**

Les systèmes d'aide à la décision en médecine permettant de mieux apprécier l'état du patient. Ils visent à diminuer l'incertitude sur la situation du patient. En outre, ils cherchent à proposer une meilleure stratégie de traitement [13].

Plusieurs systèmes ont été conçus, parmi eux nous pouvons citer :

- Un système d'aide au diagnostic des douleurs abdominales.

C'est un système de nature probabiliste, a comme objectifs l'aide pragmatique dans le cadre de la prise en charge des douleurs abdominales afin d'éviter les gestes inutiles ou les complications. L'évaluation montre que le médecin sans l'aide du système pose moins souvent le bon diagnostic que le médecin assisté du système. Complications et gestes inutiles sont également moins fréquents lorsque le système est utilisé.

- Un système d'aide au diagnostic et à la thérapeutique en médecine interne :
  - Interniste aux Etat Unis et ADM en France ;
  - Remarquable quant au domaine couvert et aux résultats proposés.

En ce qui concerne le diagnostic des troubles de la parole, il n'existe aucun système préétabli dans le marché. La plupart des travaux qui ont été reportés dans la littérature sont souvent des tests de laboratoire (voir section 1.4 et Chapitre 1).

#### **1.4. Analyse acoustique de la parole**

Bien que l'analyse subjective (auditive-perceptive) reste la base de l'évaluation de la dysarthrie, l'analyse acoustique de la parole a une valeur clinique (médicale) potentielle, car elle fournit des informations quantitatives qui contribue à la parole une description et une identification objective et mesurable.

La première étape d'une telle analyse est le choix de l'unité acoustique adéquate au type de maladie. Les unités acoustiques peuvent être des :

- Voyelles isolées ;
- Consonnes isolées ;
- Diphonies / Diphtongues ;
- Mots isolés
- Parole continue.

Par la suite, l'analyse des variations d'un ensemble de paramètres acoustiques susceptibles d'estimer le degré de l'intelligibilité de l'unité acoustique choisie est effectué.

Nous présentons dans les sections suivantes, un petit aperçu sur les méthodes d'analyse acoustiques qui ont été reportées dans la littérature.

##### **1.4.1. Voyelles isolées**

Les voyelles produites par des personnes atteintes d'une dysarthrie et qui souffrent d'une déformation tempo-fréquentielles. Elles sont caractérisées principalement par des articulations projetées (les voyelles produites ne peuvent pas atteindre la fréquence des formants<sup>1</sup> canoniques) [2]. Les mouvements d'articulation influencés par les déficits de production des voyelles dans un discours dysarthrique sont :

- La vitesse de la langue et des lèvres ;
- Les mouvements de la mâchoire ;
- La synchronisation anormale des mouvements.

Les mesures acoustiques des voyelles qui ont été largement étudiés dans la littérature sont :

- La pente mélodique du second formant ( $F_2$ ) ;
- L'espace de la zone-voyelle ;
- Le rapport de centralisation du formant (RCF) ;

---

<sup>1</sup> Les formants sont les fréquences de résonance du conduit vocal.

#### 1.4.1.1. Pente mélodique du second formant (intonation)

Weismer et Martin [2], ont fait une étude sur le comportement de la pente mélodique du deuxième formant ( $F_2$ ) capturée au cours de la production de la même voyelle dans une variété de contextes :

- Transitions entre voyelles et consonnes ;
- Voyelles complexes<sup>2</sup> ;
- Voyelles pures<sup>3</sup>.

D'après les résultats obtenus, la trajectoire de la pente mélodique de  $F_2$  est assez profonde pour les locuteurs masculins atteints d'une dysarthrie secondaire « La sclérose latérale amyotrophique (SLA) » que pour les locuteurs masculins sains.

Des résultats similaires ont été découverts pour les locuteurs atteints d'une dysarthrie Parkinsonienne.

#### 1.4.1.2. Espace de la zone-voyelle

Tjaden et Wilding [2] ont étudié la surface totale d'espace de la zone-voyelle comme indicateur de pathologie.

Les valeurs moyennes des deux premiers Formants ( $F_1$  et  $F_2$ ) de toutes les voyelles Anglaises sont calculées. Les valeurs de  $F_1$  et  $F_2$  sont représentées respectivement sur l'axe des ordonnées et des abscisses. Chaque voyelle est symbolisée sur le graphe par un point ( $F_1$ ,  $F_2$ ).

Après la représentation des formes triangulaires et quadrilatérales des formants obtenus pour des locuteurs sains et dysarthriques (sexe : masculin, type de la dysarthrie : SLA) (voir Figures 1.1 et 1.2), les remarques suivantes ont été reportées :

- La forme triangulaire [2] : La surface de l'espace de la zone-voyelle du coin (/i/, /a/, et /u/) à l'intérieur du triangle formé par les deux premiers Formants des locuteurs dysarthriques est inférieure par rapport à celle formée par les Formants de locuteurs sains.
- La forme quadrilatérale [2]: La surface de l'espace de la zone-voyelle de coin (/i/, /a/, /ae/ et /u/) à l'intérieur du quadrilatère formé par les deux premiers Formants des locuteurs dysarthriques est inférieure par rapport à celle formée par les Formants de locuteurs sains.

---

<sup>2</sup>Les voyelles qui change de timbre en cours d'émission

<sup>3</sup>Constituée d'un seul élément vocalique. Son articulation commence et se finit au même endroit

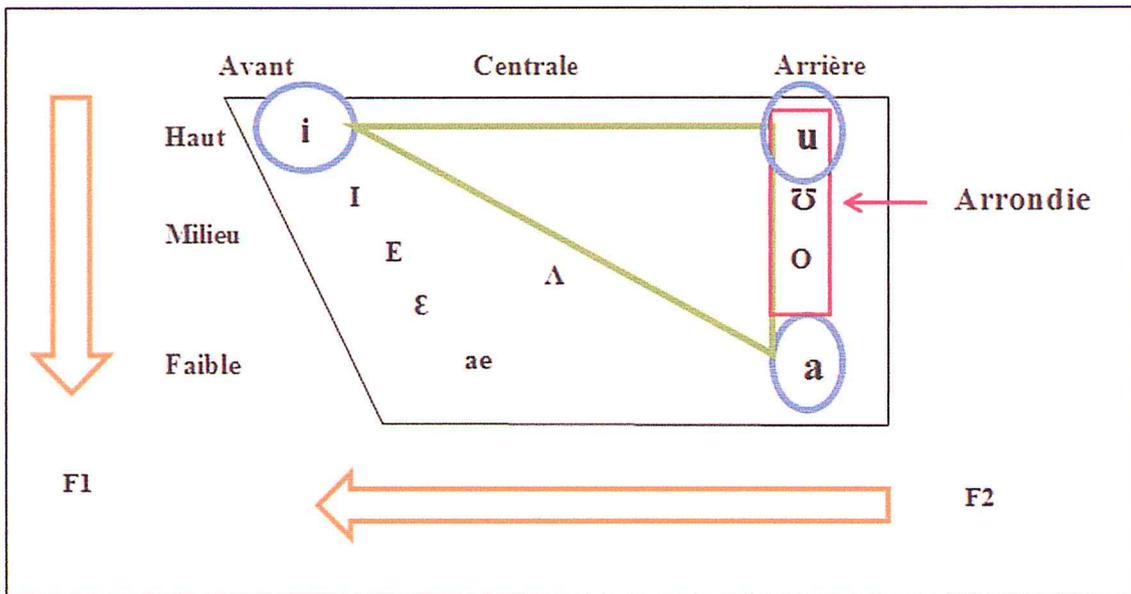


Figure 1.1 : Triangle vocalique [2].

D'après les résultats obtenus dans la même étude, aucune différence n'a été révélée entre les locuteurs dysarthriques (SLA et Parkinsoniens) féminins et les locuteurs sains.

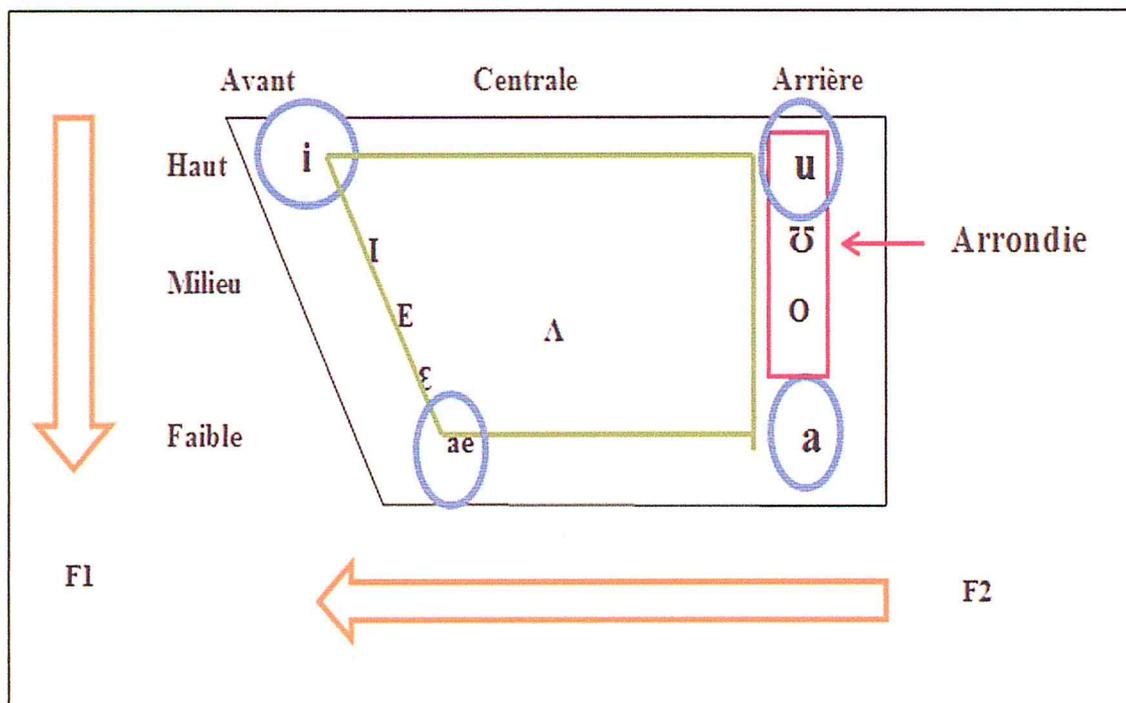


Figure 1.2 : Quadrilatère vocalique[2].

### 1.4.1.3. Rapport de centralisation du formant (RCF)

Le rapport de centralisation du Formant (RCF) a été proposé par **Sapir et RamigFox**. Cette mesure est une métrique d'espace voyelle qui maximise la sensibilité des voyelles centrales en minimisant la variance des fréquences de formants. Il est donné comme suit :

$$RCF = \frac{F2u+F2a+F1i+F1a}{F2i+F1a} \quad (1.1)$$

$F2u$  : Deuxième formant de la voyelle /u/;

$F2a$  : deuxième formant de la voyelle /a/ ;

$F1i$  : Premier formant de la voyelle /i/ ;

$F1a$  : Premier formant de la voyelle /a/ ;

$F2i$  : Deuxième formant de la voyelle /i/.

Il a été prouvé que le RCF peut distinguer de manière fiable entre les espaces de voyelles du type Hypokinétiques<sup>4</sup> et celles des locuteurs sains [2].

Une grande variété de métriques acoustiques a été calculée pour différencier entre un discours dysarthrique et un discours sain.

Le choix des métriques se fait en se basant sur le type de la dysarthrie diagnostiquée.

### 1.4.2. Consonnes isolées

Très peu d'études ont été reportées dans la littérature sur l'emploi des consonnes isolées en vue de l'estimation du degré d'intelligibilité des voix dysarthriques. Nous citons l'exemple des travaux de **Tanja KocjancicAntolík et al** [14].

Dans cet article, les auteurs ont étudié la distorsion des consonnes chez les personnes dysarthriques en se basant sur le calcul de leur fréquence d'apparition.

Le nombre total des consonnes examinées est de 4990 consonnes. Trois groupes de locuteurs sont construits pour représenter trois types différents de dysarthries : SLA, dysarthrie ataxique et, dysarthrie parkinsonienne.

L'unité acoustique utilisée dans cette expérimentation est la CV (Consonne/Voyelle). Le corpus est extrait à partir des mots prononcés par des locuteurs français et en langues

---

<sup>4</sup>Perturbations de l'activité et de l'attention

Française. Néanmoins, le nombre de consonnes prononcées par locuteur est très limité (environ 48 consonnes).

D'après les résultats obtenus, le nombre de consonnes déformées qui ont été prononcées par des patients dysarthriques est limité. Seulement 10% de consonnes par locuteur ont été jugées anormales [14].

Selon notre analyse, le nombre limité de consonnes rend cette étude non concluante. L'utilisation d'un corpus plus large est très nécessaire pour dresser des conclusions plus convaincantes.

### **1.5.Conclusion**

Dans ce chapitre, nous avons tout d'abord présenté les caractéristiques générales de la dysarthrie, ainsi que des définitions de l'intelligibilité de la parole dysarthrique, les méthodes d'évaluation : subjective et objective, et finalement, les méthodes d'analyse de la parole.

La proposition de nouvelles mesures acoustiques nécessite des connaissances approfondies en phonétiques et en médecine. L'utilisation des mesures acoustiques standards comme entrée aux machines d'apprentissage statistiques est une bonne démarche de traitement surtout pour des concepteurs des systèmes d'aide à la décision dont la formation de base est en informatique et traitement de signal.

Dans le chapitre suivant, nous présenterons les principales méthodes de l'évaluation objective de l'intelligibilité de la parole dysarthrique.

### **1.1. Introduction**

La dysarthrie est un trouble d'articulation de la parole d'origine neurologique [1]. Les patients atteints de dysarthrie ont une incoordination au niveau de la production de la parole.

Dans ce chapitre, nous présentons un aperçu général sur cette maladie en se focalisant sur quatre points essentiels : (i) les symptômes, (ii) les types, (iii) les causes (iv) et l'évaluation de l'intelligibilité de la parole.

### **1.2. Dysarthrie**

L'étude de pathologies motrices de la parole nécessite des connaissances cliniques préalables. C'est l'objectif principal de cette section.

#### **1.2.1. Définition**

La dysarthrie est un trouble de l'articulation causé par des lésions du cerveau ou une atteinte au niveau des différents organes ayant un rôle dans la parole comme la langue ou le voile du palais. Elle touche principalement les adultes. La dysarthrie se manifeste par des difficultés d'articulation, une altération de la voix et du débit de parole. Des pathologies peuvent être à l'origine de la dysarthrie, comme la maladie de Parkinson, un accident vasculaire cérébral, une paralysie faciale ou la sclérose en plaques. Une rééducation orthophonique peut améliorer l'état de la personne atteinte par la dysarthrie [4].

#### **1.2.2. Symptômes généraux**

La dysarthrie peut apparaître soudainement ou bien elle peut se développer lentement. Les symptômes se remarquent donc surtout au niveau de la parole [1].

- Au niveau de la parole :
  - Mauvaise articulation ;
  - Rythme irrégulier, trop lent ou trop rapide ;
  - Marmonnée ;
  - Chuchotée.
  
- Au niveau de la voix :
  - Nasillarde ;
  - Nez bouché ;
  - Rauque ;
  - Tendue ;

- Soufflée.

### 1.2.3. Types de la maladie de dysarthrie

En fonction de la lésion qui altère le langage, plusieurs formes de dysarthrie ont été établies. Le tableau ci-dessous donne un aperçu sur les formes les plus courantes de cette pathologie :

Type	Caractéristiques	Voix/parole/discours
Dysarthrie flasque ou paralytique	<ul style="list-style-type: none"> <li>• Hypotonie</li> <li>• Faiblesse des muscles</li> </ul>	La voix peut être : <ul style="list-style-type: none"> <li>• Essoufflée</li> <li>• Nasale</li> </ul>
Dysarthrie spastique (DS)	<ul style="list-style-type: none"> <li>• Augmentation du tonus musculaire (par atteinte bilatérale du premier motoneurone central).</li> <li>• Parésie / spasticité des autres parties du corps</li> <li>• Réflexes hyperactifs</li> <li>• Réflexes pathologiques</li> <li>• Dysphagie</li> </ul>	<ul style="list-style-type: none"> <li>• La voix est :                             <ul style="list-style-type: none"> <li>○ Rauque râpeuse</li> <li>○ Étranglée</li> <li>○ Faible</li> </ul> </li> <li>• Tonalité cassée</li> <li>• Discours accéléré</li> <li>• Émission de phrases courtes</li> <li>• Parole laborieuse</li> <li>• Distorsions faciales</li> </ul>
Dysarthrie corticale	Perte soudaine de la parole et paralysie bilatérale des muscles	Articulation impossible mais compréhension normale
Dysarthrie lacunaire	Association d'une : <ul style="list-style-type: none"> <li>• Dysphagie</li> <li>• Parésie faciale</li> <li>• Maladresse de la main</li> </ul>	<ul style="list-style-type: none"> <li>• Troubles de la phonation et de la déglutition</li> </ul>
Dysarthrie hypokinétique	L'hypokinésie affecte : <ul style="list-style-type: none"> <li>• Les muscles respiratoires</li> <li>• Les muscles laryngés</li> <li>• Les muscles palatopharyngés et buccaux</li> </ul>	<ul style="list-style-type: none"> <li>• La voix :                             <ul style="list-style-type: none"> <li>○ Lente et monotone</li> <li>○ Devient sourde</li> <li>○ Perd son volume et ses inflexions</li> </ul> </li> <li>• L'élocution est :                             <ul style="list-style-type: none"> <li>○ Inintelligible</li> <li>○ Marquée par des répétitions</li> </ul> </li> </ul>

Dysarthrie hyperkinétique rapide	<ul style="list-style-type: none"> <li>• Atteinte du système extrapyramidal, dysfonctionnement des noyaux gris centraux</li> <li>• Cette dysarthrie se caractérise par des mouvements rapides et involontaires</li> </ul>	<ul style="list-style-type: none"> <li>• Élocution rapide</li> <li>• La voix est : <ul style="list-style-type: none"> <li>○ Forcée</li> <li>○ Rauque</li> <li>○ Irrégulière</li> <li>○ Mal coordonnée par rapport à la respiration</li> </ul> </li> </ul>
Dysarthrie hyperkinétique lente	Atteinte du système extrapyramidal entraînant une akinésie	<p>La voix est :</p> <ul style="list-style-type: none"> <li>• Étranglée</li> <li>• Stridente</li> </ul> <p>L'élocution est :</p> <ul style="list-style-type: none"> <li>• Sourde</li> <li>• Lente</li> <li>• Laborieuse</li> </ul>
Dysarthrie ataxique	<ul style="list-style-type: none"> <li>• Lésion du cervelet et des voies cérébelleuses qui affecte la coordination et le contrôle des muscles de la phonation</li> <li>• Elle se rencontre en cas de sclérose en plaques notamment</li> </ul>	<ul style="list-style-type: none"> <li>• Élocution lente et irrégulière : <ul style="list-style-type: none"> <li>○ Bredouillement</li> <li>○ Séparation anormale des mots</li> </ul> </li> <li>• Ton impossible à maintenir à une même hauteur et à une même intensité</li> <li>• Impression que la personne est en colère</li> <li>• Chuchotement impossible</li> </ul>
Dysarthrie parkinsonienne	<p>Rencontrée au cours de la maladie de Parkinson, elle se caractérise par :</p> <ul style="list-style-type: none"> <li>• Une perte d'intensité de la voix ;</li> <li>• Une perte des modulations.</li> </ul>	<ul style="list-style-type: none"> <li>• Réduction de l'intelligibilité globale de la parole</li> <li>• Voix basse malgré des efforts importants</li> <li>• Timbre monotone, accents intonatifs inexistant</li> <li>• Débit variable : <ul style="list-style-type: none"> <li>○ Accélération subites</li> <li>○ Silences inappropriés</li> </ul> </li> </ul>

Tableau 1.1 : Les différents types de la dysarthrie et leurs caractéristiques [1].

#### **1.2.4. Causes**

La dysarthrie affecte le cerveau. Cela peut se produire à la naissance, comme dans le cas où la personne atteint une paralysie cérébrale, ou elle peut se produire plus tard dans la vie en raison de l'une des nombreuses causes suivantes [5]:

- Un Accident Vasculaire Cérébral (SLA);
- Une commotion cérébrale ;
- Une tumeur ;
- La maladie de Parkinson [6];
- La maladie de Lou Gehrig [7];
- La sclérose latérale amyotrophique (SLA);
- La maladie de Huntington [8];
- La maladie de sclérose en plaque.

Chez l'enfant, les causes ne sont pas nombreuses :

- Les atteintes du tronc cérébral, le plus souvent par des problèmes traumatiques ou tumoraux ;
- La myasthénie ;
- Certaines myopathies ;
- Les tumeurs du cervelet, très rares.

### **1.3.Intelligibilité de la parole dysarthrique**

#### **1.3.1. Définition de l'intelligibilité**

L'intelligibilité est définie comme le degré de précision avec lequel le message est compris par l'auditeur. Elle s'obtient en déterminant le nombre d'unités de parole reconnus par l'auditeur. Dans le sens le plus général, l'intelligibilité est la capacité d'un individu à être compris verbalement par un autre [9].

Dans le cas de notre étude, mesurer le degré de la sévérité de la dysarthrie revient à mesurer l'intelligibilité de cette maladie.

#### **1.3.2. Méthodes d'évaluation de l'intelligibilité**

L'évaluation des compétences verbales et communicationnelles est indispensable afin de déterminer le niveau d'intelligibilité et de comprendre les troubles du patient dysarthrique, dans le souci d'élaborer un plan thérapeutique adapté afin de réduire l'aggravation de la

maladie. La démarche de questionnement du thérapeute doit mettre en lien la perception et la sévérité des troubles [10].

Il existe deux formes d'évaluation :

- Evaluation subjective,
- Evaluation objective.

### 1.3.2.1. Evaluation subjective

L'évaluation subjective de l'intelligibilité de la parole dysarthrique se fait à l'écoute de la parole du patient. Elle s'attache particulièrement à :

- La netteté et la précision des voyelles et des consonnes, à la réalisation de groupe de consonnes (coarticulation) ;
- L'intonation est étudiée à travers la facilité du discours ;
- La longueur des pauses ;
- Les changements de rythme et dans son maintien au niveau mélodique ou accentuel [11].

Elle nécessite la présence de deux personnes : (i) le clinicien et (ii) le transcripteur. Le rôle du clinicien est de sélectionner les mots et les phrases de test qui seront prononcés par le patient. Le transcripteur (une personne peu familière avec ces mots et des phrases) écoute les enregistrements des patients, et les transcrit en phrases phonémiques.

L'évaluation consiste généralement à mesurer les capacités vocales des patients. Il existe deux tests publiés pour évaluer l'intelligibilité des troubles moteurs de la parole :

- Evaluation de l'intelligibilité des locuteurs dysarthriques (*Assessment of Intelligibility in Dysarthric Speakers (AIDS)*) [12];
- Test de l'intelligibilité de la phrase (*the Sentence Intelligibility Test (SIT)*).

La différence entre ces deux méthodes réside dans le choix des unités phonétiques de test et le nombre de points de l'échelle d'évaluation. Pour plus de détails sur ces tests subjectifs, le lecteur peut consulter les références.

### 1.3.2.2. Evaluation objective

Le meilleur standard pour l'évaluation de l'intelligibilité des troubles dysarthriques est l'évaluation subjective réalisée par un expert. Les tests subjectifs sont souvent coûteux, partiels, en raison de leur nature perceptive. En revanche, les méthodes d'évaluation objective, sont automatiques, répétables et relativement pas chers.

L'évaluation objective de l'intelligibilité des troubles de la parole repose sur deux approches complémentaires [13]:

- Analyse acoustique :

Elle est basée sur l'exploration des mesures acoustiques qui révèlent les signes de pathologie. Dans le cas où les mesures identifiées sont jugées suffisamment robustes, elles peuvent être utilisées directement pour générer une estimation quantitative de l'intelligibilité perçue. Des exemples de ces mesures sont donnés dans la section 1.3.

- Modélisation mathématique des données acoustiques :

Si on ne dispose d'aucune information a priori sur la fiabilité des mesures acoustiques extraites à partir de l'onde acoustique de la voix, on ne peut pas quantifier le degré de l'intelligibilité d'une manière directe. La modélisation mathématique de ces mesures est nécessaire pour extraire les connaissances à partir des données acoustiques. Le processus d'extraction de connaissances est accompli par le biais des machines d'apprentissage. Des exemples pratiques sur ces modèles seront donnés dans le chapitre suivant.

### 1.3.3. Systèmes d'aide à la décision en médecine

Les systèmes d'aide à la décision en médecine permettant de mieux apprécier l'état du patient. Ils visent à diminuer l'incertitude sur la situation du patient. En outre, ils cherchent à proposer une meilleure stratégie de traitement [13].

Plusieurs systèmes ont été conçus, parmi eux nous pouvons citer :

- Un système d'aide au diagnostic des douleurs abdominales.

C'est un système de nature probabiliste, a comme objectifs l'aide pragmatique dans le cadre de la prise en charge des douleurs abdominales afin d'éviter les gestes inutiles ou les complications. L'évaluation montre que le médecin sans l'aide du système pose moins souvent le bon diagnostic que le médecin assisté du système. Complications et gestes inutiles sont également moins fréquents lorsque le système est utilisé.

- Un système d'aide au diagnostic et à la thérapeutique en médecine interne :
  - Interniste aux Etat Unis et ADM en France ;
  - Remarquable quant au domaine couvert et aux résultats proposés.

En ce qui concerne le diagnostic des troubles de la parole, il n'existe aucun système préétabli dans le marché. La plupart des travaux qui ont été reportés dans la littérature sont souvent des tests de laboratoire (voir section 1.4 et Chapitre 1).

#### **1.4. Analyse acoustique de la parole**

Bien que l'analyse subjective (auditive-perceptive) reste la base de l'évaluation de la dysarthrie, l'analyse acoustique de la parole a une valeur clinique (médicale) potentielle, car elle fournit des informations quantitatives qui contribue à la parole une description et une identification objective et mesurable.

La première étape d'une telle analyse est le choix de l'unité acoustique adéquate au type de maladie. Les unités acoustiques peuvent être des :

- Voyelles isolées ;
- Consonnes isolées ;
- Diphonies / Diphtongues ;
- Mots isolés
- Parole continue.

Par la suite, l'analyse des variations d'un ensemble de paramètres acoustiques susceptibles d'estimer le degré de l'intelligibilité de l'unité acoustique choisie est effectué.

Nous présentons dans les sections suivantes, un petit aperçu sur les méthodes d'analyse acoustiques qui ont été reportées dans la littérature.

##### **1.4.1. Voyelles isolées**

Les voyelles produites par des personnes atteintes d'une dysarthrie et qui souffrent d'une déformation tempo-fréquentielles. Elles sont caractérisées principalement par des articulations projetées (les voyelles produites ne peuvent pas atteindre la fréquence des formants<sup>1</sup> canoniques) [2]. Les mouvements d'articulation influencés par les déficits de production des voyelles dans un discours dysarthrique sont :

- La vitesse de la langue et des lèvres ;
- Les mouvements de la mâchoire ;
- La synchronisation anormale des mouvements.

Les mesures acoustiques des voyelles qui ont été largement étudiés dans la littérature sont :

- La pente mélodique du second formant ( $F_2$ ) ;
- L'espace de la zone-voyelle ;
- Le rapport de centralisation du formant (RCF) ;

---

<sup>1</sup> Les formants sont les fréquences de résonance du conduit vocal.

#### 1.4.1.1. Pente mélodique du second formant (intonation)

Weismer et Martin [2], ont fait une étude sur le comportement de la pente mélodique du deuxième formant ( $F_2$ ) capturée au cours de la production de la même voyelle dans une variété de contextes :

- Transitions entre voyelles et consonnes ;
- Voyelles complexes<sup>2</sup> ;
- Voyelles pures<sup>3</sup>.

D'après les résultats obtenus, la trajectoire de la pente mélodique de  $F_2$  est assez profonde pour les locuteurs masculins atteints d'une dysarthrie secondaire « La sclérose latérale amyotrophique (SLA) » que pour les locuteurs masculins sains.

Des résultats similaires ont été découverts pour les locuteurs atteints d'une dysarthrie Parkinsonienne.

#### 1.4.1.2. Espace de la zone-voyelle

Tjaden et Wilding [2] ont étudié la surface totale d'espace de la zone-voyelle comme indicateur de pathologie.

Les valeurs moyennes des deux premiers Formants ( $F_1$  et  $F_2$ ) de toutes les voyelles Anglaises sont calculées. Les valeurs de  $F_1$  et  $F_2$  sont représentées respectivement sur l'axe des ordonnées et des abscisses. Chaque voyelle est symbolisée sur le graphe par un point ( $F_1$ ,  $F_2$ ).

Après la représentation des formes triangulaires et quadrilatérales des formants obtenus pour des locuteurs sains et dysarthriques (sexe : masculin, type de la dysarthrie : SLA) (voir Figures 1.1 et 1.2), les remarques suivantes ont été reportées :

- La forme triangulaire [2] : La surface de l'espace de la zone-voyelle du coin (/i/, /a/, et /u/) à l'intérieur du triangle formé par les deux premiers Formants des locuteurs dysarthriques est inférieure par rapport à celle formée par les Formants de locuteurs sains.
- La forme quadrilatérale [2]: La surface de l'espace de la zone-voyelle de coin (/i/, /a/, /ae/ et /u/) à l'intérieur du quadrilatère formé par les deux premiers Formants des locuteurs dysarthriques est inférieure par rapport à celle formée par les Formants de locuteurs sains.

---

<sup>2</sup>Les voyelles qui change de timbre en cours d'émission

<sup>3</sup>Constituée d'un seul élément vocalique. Son articulation commence et se finit au même endroit

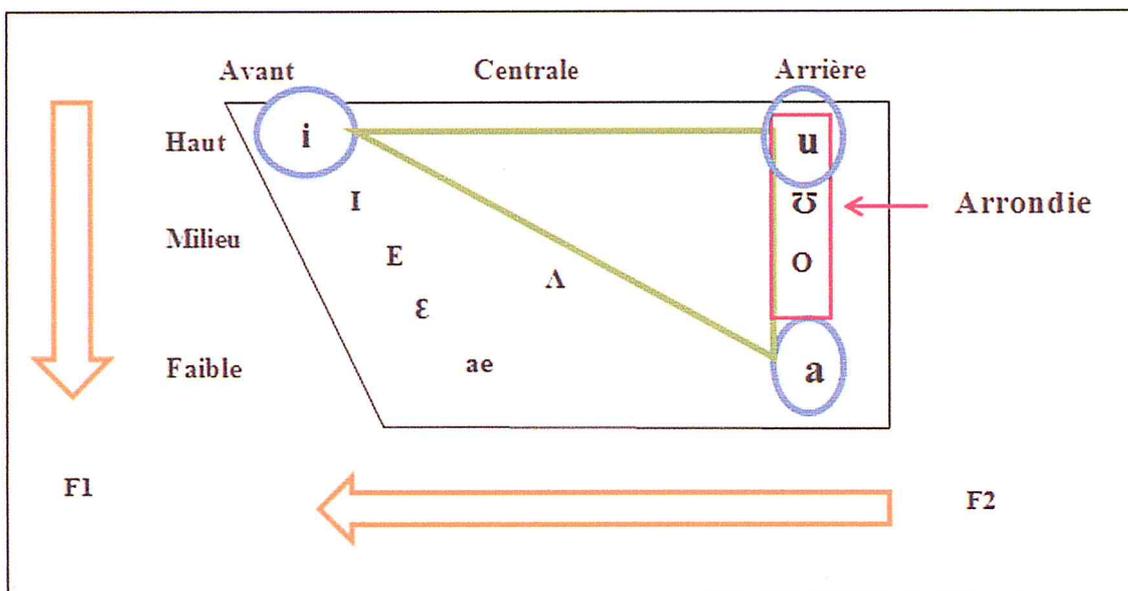


Figure 1.1 : Triangle vocalique [2].

D'après les résultats obtenus dans la même étude, aucune différence n'a été révélée entre les locuteurs dysarthriques (SLA et Parkinsoniens) féminins et les locuteurs sains.

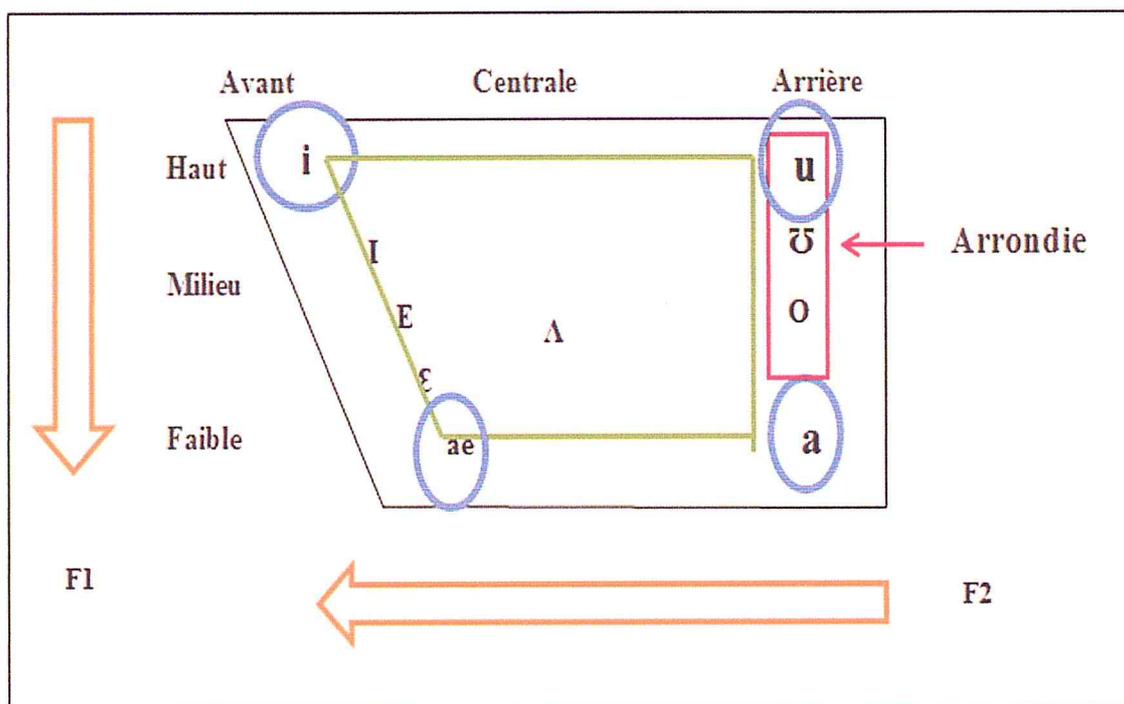


Figure 1.2 : Quadrilatère vocalique[2].

### 1.4.1.3. Rapport de centralisation du formant (RCF)

Le rapport de centralisation du Formant (RCF) a été proposé par **Sapir et RamigFox**. Cette mesure est une métrique d'espace voyelle qui maximise la sensibilité des voyelles centrales en minimisant la variance des fréquences de formants. Il est donné comme suit :

$$RCF = \frac{F2u+F2a+F1i+F1a}{F2i+F1a} \quad (1.1)$$

$F2u$  : Deuxième formant de la voyelle /u/;

$F2a$  : deuxième formant de la voyelle /a/ ;

$F1i$  : Premier formant de la voyelle /i/ ;

$F1a$  : Premier formant de la voyelle /a/ ;

$F2i$  : Deuxième formant de la voyelle /i/.

Il a été prouvé que le RCF peut distinguer de manière fiable entre les espaces de voyelles du type Hypokinétiques<sup>4</sup> et celles des locuteurs sains [2].

Une grande variété de métriques acoustiques a été calculée pour différencier entre un discours dysarthrique et un discours sain.

Le choix des métriques se fait en se basant sur le type de la dysarthrie diagnostiquée.

### 1.4.2. Consonnes isolées

Très peu d'études ont été reportées dans la littérature sur l'emploi des consonnes isolées en vue de l'estimation du degré d'intelligibilité des voix dysarthriques. Nous citons l'exemple des travaux de **Tanja KocjancicAntolík et al** [14].

Dans cet article, les auteurs ont étudié la distorsion des consonnes chez les personnes dysarthriques en se basant sur le calcul de leur fréquence d'apparition.

Le nombre total des consonnes examinées est de 4990 consonnes. Trois groupes de locuteurs sont construits pour représenter trois types différents de dysarthries : SLA, dysarthrie ataxique et, dysarthrie parkinsonienne.

L'unité acoustique utilisée dans cette expérimentation est la CV (Consonne/Voyelle). Le corpus est extrait à partir des mots prononcés par des locuteurs français et en langues

---

<sup>4</sup>Perturbations de l'activité et de l'attention

Française. Néanmoins, le nombre de consonnes prononcées par locuteur est très limité (environ 48 consonnes).

D'après les résultats obtenus, le nombre de consonnes déformées qui ont été prononcées par des patients dysarthriques est limité. Seulement 10% de consonnes par locuteur ont été jugées anormales [14].

Selon notre analyse, le nombre limité de consonnes rend cette étude non concluante. L'utilisation d'un corpus plus large est très nécessaire pour dresser des conclusions plus convaincantes.

### **1.5.Conclusion**

Dans ce chapitre, nous avons tout d'abord présenté les caractéristiques générales de la dysarthrie, ainsi que des définitions de l'intelligibilité de la parole dysarthrique, les méthodes d'évaluation : subjective et objective, et finalement, les méthodes d'analyse de la parole.

La proposition de nouvelles mesures acoustiques nécessite des connaissances approfondies en phonétiques et en médecine. L'utilisation des mesures acoustiques standards comme entrée aux machines d'apprentissage statistiques est une bonne démarche de traitement surtout pour des concepteurs des systèmes d'aide à la décision dont la formation de base est en informatique et traitement de signal.

Dans le chapitre suivant, nous présenterons les principales méthodes de l'évaluation objective de l'intelligibilité de la parole dysarthrique.

---

---

## *Chapitre 2*

### *Evaluation objective de l'intelligibilité de la parole dysarthrique*

---

---

### **2.1. Introduction**

Les évaluations subjectives sont souvent coûteuses, partiales et incompatibles en raison de leur nature perceptive, car elles nécessitent le passage par les médecins et les orthophonistes et faire plusieurs radios, en revanche les évaluations objectives (automatiques) sont relativement pas chères et ne prennent pas du temps le résultat s'affiche sur place. Les méthodes objectives peuvent être décomposées en deux sous-catégories : des méthodes basées sur des références et des méthodes sans références.

Dans ce chapitre nous allons décrire en détail les différentes méthodes d'évaluation ainsi que la définition de quelques bases de données existantes.

### **2.2. Evaluation objective de l'intelligibilité**

Comme nous l'avons indiqué dans le chapitre précédent, une partie importante du diagnostic de la dysarthrie comprend l'évaluation de l'intelligibilité de la parole. L'analyse subjective de la parole dysarthrique a longtemps été le premier standard d'évaluation de l'intelligibilité. Cependant, il existe actuellement plusieurs méthodes informatiques pour l'estimation automatique de l'intelligibilité dans la littérature. Il s'agit des méthodes objectives. Elles sont subdivisées en deux grandes familles :

- Méthodes basées sur des références ;
- Méthodes sans références.

#### **2.2.1. Méthodes basées sur des références**

L'idée de base derrière toute méthode fondée sur une référence est de procéder par une transcription orthographique-phonétique d'un mot prononcé par un patient dysarthrique, puis de comparer le résultat trouvé à un modèle de référence.

Le modèle référentiel est souvent obtenu par des systèmes de Reconnaissance Automatique de la Parole (RAP) [15].

Le système est entraîné (apprentissage supervisé) sur des messages vocaux sains (normaux) avec un vocabulaire bien ciblé et cela en utilisant les traits acoustiques appropriés (Souvent, les Coefficients MFCCs (Mel-FrequencyCepstralCoefficients)). La parole dysarthrique est testée à travers le système de RAP ; la transcription fournie par ce système est par la suite comparée à celle d'une référence (obtenu par des locuteurs normaux). Le taux de reconnaissance obtenu représente l'évaluation objective de l'intelligibilité. Cette approche donne une assez bonne estimation du niveau d'intelligibilité. Le système de RAP joue le rôle d'une oreille artificielle (Figure 2.1).

Plusieurs approches d'estimation du niveau d'intelligibilité ont été reportées dans la littérature. Nous citons les travaux ci-dessous.

**Carmichael et Green [15]** ont construit un système de RAP à base des HMMs (Hidden Markov Models), qui a pour objectif l'amélioration de l'évaluation de l'intelligibilité. Les mesures de l'intelligibilité sont basées sur les résultats des probabilités de vraisemblance (Likelihood) dérivées de l'alignement d'une parole dysarthrique à un modèle HMM entraînés sur les mots prononcés par des locuteurs normaux.

**Catherine Middag et al [16]** ont construit un système automatique de calcul de l'intelligibilité en se basant sur :

- Les caractéristiques phonologiques ;
- L'alignement automatique de la parole à partir des modèles acoustiques qui ont été extraits à partir des paroles normales.

Après l'acquisition de la parole, le système détecte les paramètres phonologiques de la parole en entrée. Ensuite un alignement (ajustement) des paramètres phonologiques de la parole acquise avec ceux existants dans un système (référence) est effectué. Finalement, l'intelligibilité est estimée à partir de mesures acoustiques appropriées de la parole de test.

La figure 2.1 illustre la manière de fonctionnement des méthodes d'estimation du niveau d'intelligence basé sur une référence.

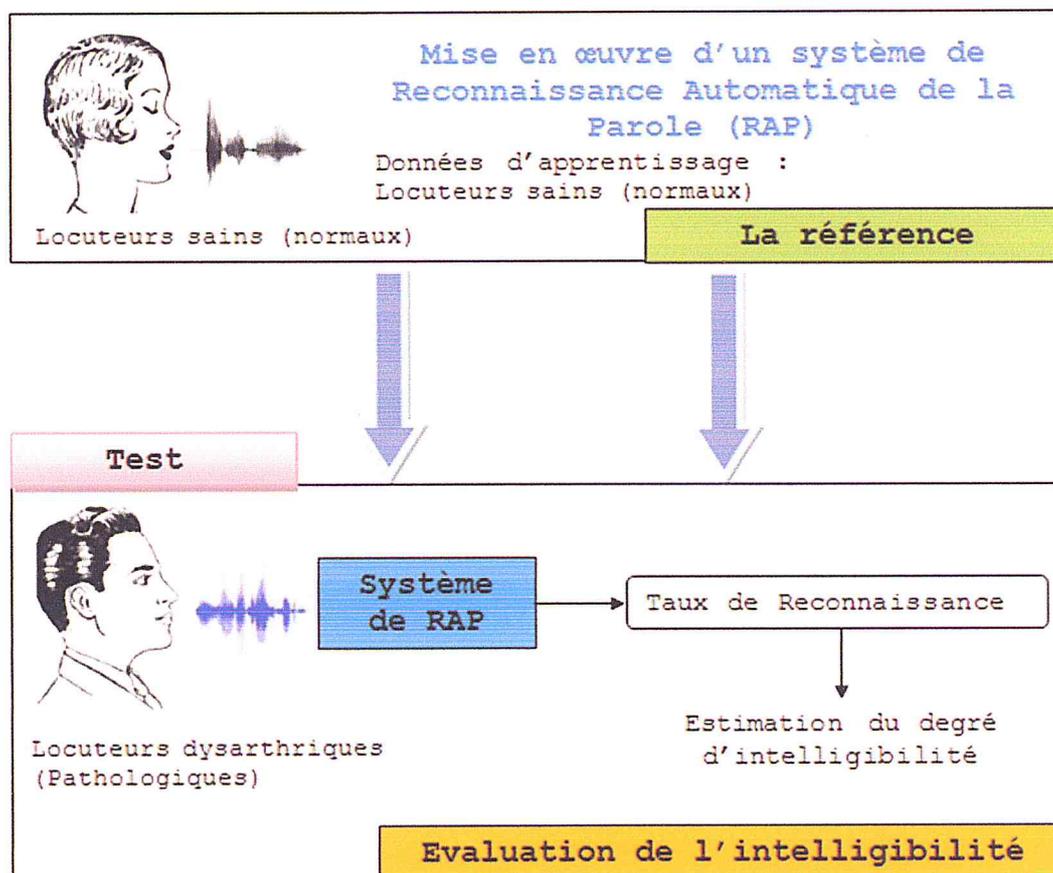


Figure 2.1 : Méthodes d'estimation du niveau d'intelligence basé sur une référence.

### 2.2.2. Méthodes sans références

L'objectif principal de cette catégorie de méthodes est d'estimer l'intelligibilité de la parole des patients dysarthriques sans avoir recours aux systèmes de RAP. Cela signifie que l'intelligibilité n'est pas mesurée à partir du taux de bonne reconnaissance d'un message vocal en entrée. Les systèmes sans références ont l'avantage de s'appliquer potentiellement à de nouveaux types de données de la parole, par exemple parole spontanée, ou messages vocaux d'une langue différente [7].

Les méthodes d'évaluation de l'intelligibilité sans-référence peuvent être implémentées suivant une variété d'approches. Nous citons les deux formes suivantes :

- (i) Une approche d'estimation de l'intelligibilité fondée sur une analyse des paramètres acoustiques qui peut révéler des signes de la dysarthrie. Le choix des paramètres acoustiques est très crucial pour ce type d'approche.

Il y a plusieurs approches qui ont été décrites dans la littérature, nous avons cité quelques-unes dans le chapitre 1 : le triangle vocalique, quadrilatère vocalique, les formants et la pente mélodique.

(ii) Une autre approche consiste à construire des modèles mathématiques distincts pour chaque niveau d'intelligibilité et cela à partir des données d'apprentissage (dites références supervisées). Ces données sont généralement représentées sous formes de mesures acoustiques extraites par le biais des techniques de traitement de signal. Dans la plupart des cas, le choix de ces mesures n'est pas vraiment critique du fait que la mesure du niveau d'intelligibilité se base sur deux étages de traitement : (i) extraction d'attributs et (ii) construction des modèles de prédiction.

La modélisation mathématique est basée sur les machines d'apprentissage statistiques de type classification ou régression. Par la suite, la mesure du niveau d'intelligibilité d'un locuteur de test est effectuée suivant deux étapes essentielles :

- Mesure des paramètres acoustiques du message vocal ;
- Exploitation des modèles mathématiques estimés dans l'étape d'apprentissage pour fournir une mesure quantitative du niveau d'intelligibilité du locuteur.

La figure 2.2 représente la méthode d'estimation d'intelligibilité sans références basés sur des modèles mathématiques.

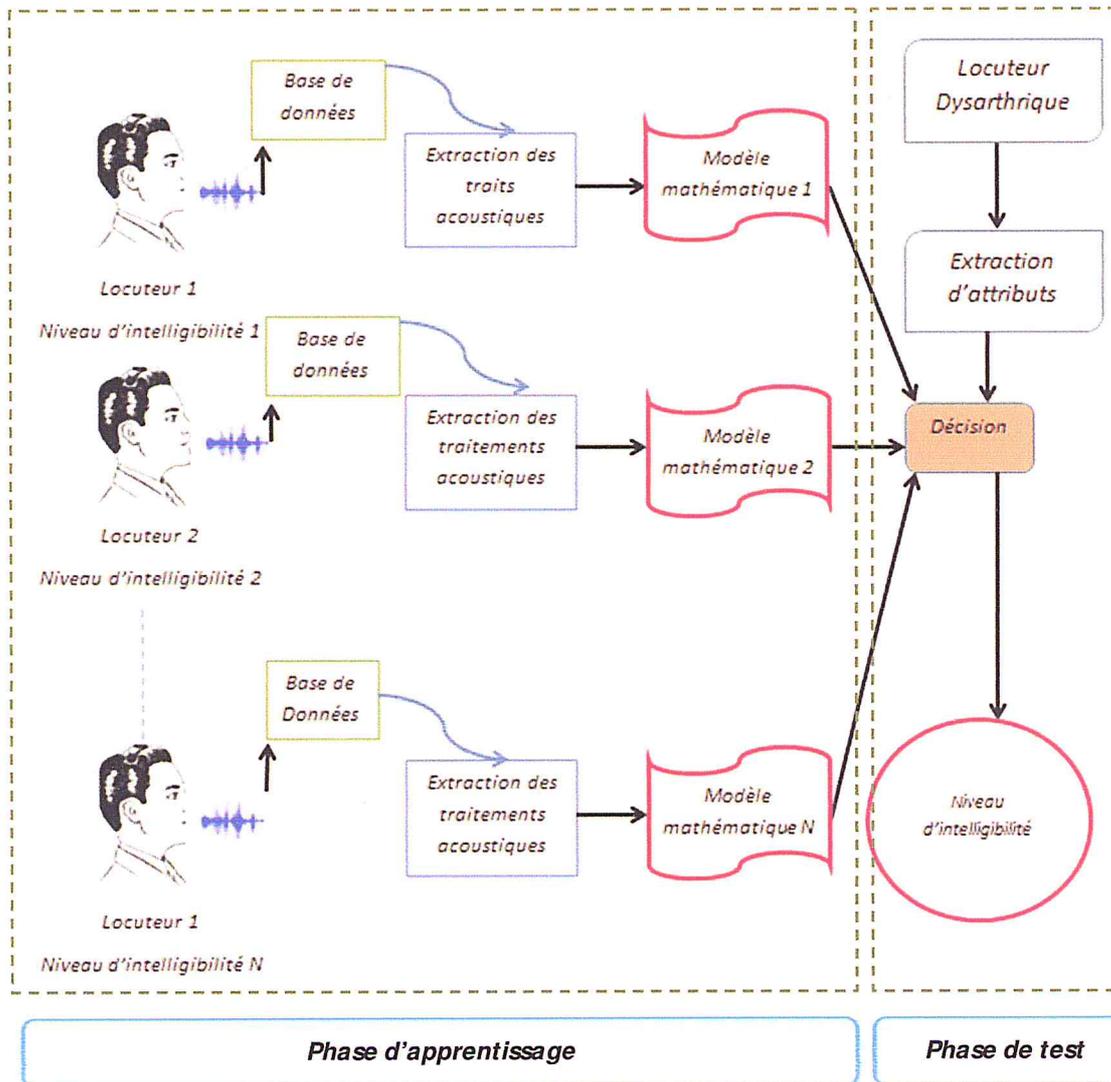


Figure 2.2 : Méthodes d'estimation du niveau d'intelligibilité sans références.

Plusieurs approches ont été reportées dans la littérature. Nous citons les travaux ci-dessous :

**Tiago H. Falk et al [17]** ont proposé plusieurs mesures objectives en vue d'une meilleure estimation de l'intelligibilité des paroles dysarthriques. Les mesures utilisées sont : la source d'excitation, les dynamiques temporelles et les paramètres prosodiques. Par la suite, une mesure composée a été développée par la combinaison linéaire d'un sous-ensemble de paramètres. Les expériences ont démontré que la mesure composée est corrélée avec des estimations subjectives de l'intelligibilité. La base de données UA speech a été utilisée dans cet article.

**Myung Jong Kim et Hoirin Kim [18]** ont construit un système automatique dont l'objectif est de régler le problème de l'évaluation de l'intelligibilité via les informations

spectrales déformées. Le système proposé extrait plusieurs mesures liées à la qualité phonétique ensuite ils sélectionnent les mesures qui ne produisent pas une erreur de prédiction minimale. Finalement, les mesures sélectionnées sont linéairement combinées en utilisant un modèle de régression multiple. L'évaluation de la méthode proposée est faite sur une base de données (UA speech) de 94 patients dysarthriques. Les résultats obtenus prouvent l'efficacité de la méthode proposée.

**K.L. Kadi et al [19]** ont construit un système d'évaluation automatique de l'intelligibilité de la parole dysarthrique basé sur un ensemble de paramètres prosodiques réduits via la technique LDA (*Linear Discriminative Analysis*). Deux modèles mathématiques différents : GMMs (*Gaussian Mixture Model*) et SVM (*Support Vector Machine*) ont été utilisés pour la classification.

La base de données Nemours des locuteurs dysarthriques Américains est utilisée tout au long des expériences. Les résultats de ces expériences montrent un meilleur taux de classification avec le système LDA/SVM sur quatre niveaux de sévérité de la dysarthrie. Les résultats de ces expériences montrent que le système LDA/SVM donne un meilleur taux de classification par rapport au système GMM/SVM sur quatre niveaux de sévérité de la dysarthrie.

### 2.3. Corpus et bases de données de test

Plusieurs bases de données des voix dysarthriques ont été reportées dans la littérature. Nous citons les trois bases de données suivantes :

- Nemours ;
- Universal Access (UA) speech ;
- Torgo.

#### 2.3.1. Nemours

Cette base de données a été construite par **Xavier Menéndez-pidal et al [20]**, c'est une collection de : 814 phrases courtes et 74 phrases prononcées par 11 locuteurs masculins (des enregistrements audio) avec divers niveaux de dysarthrie.

En plus, la base de données contient deux paragraphes de parole prononcés aussi par les mêmes locuteurs.

Cette base de données est dédiée à la maladie de la dysarthrie donc elle permet d'examiner l'intelligibilité de la parole dysarthrique avant et après l'amélioration de la maladie

par de diverses méthodes de traitement. L'enregistrement a été fait en Anglais Américain. Cette base de données est commercialement distribuée.

### **2.3.2. Torgo**

C'est une base de données d'articulation de la dysarthrie. Elle a été développée par le département de l'informatique et de l'orthophonie de l'université de Toronto en collaboration avec l'hôpital de réadaptation pour enfants **Holland-Bloorview** à Toronto, Canada [21].

Elle contient des données vocales en Anglais de 8 locuteurs (5 mâles, 3 femelles) atteints de paralysie cérébrale (PC) ou la sclérose latérale amyotrophique (SLA) et de 7 locuteurs (4 mâles, 3 femelles) à partir d'un groupe témoin non dysarthrique.

La base de données Torgo est principalement une ressource pour le développement de méthodes de RAP adaptées aux besoins des personnes atteintes de dysarthrie. Les enregistrements de cette base de données ont été faits en Anglais Américain. Cette base de données est payante donc elle n'est pas à la portée de tout le monde.

### **2.3.3. UA speech**

La base de données UA speech fournit une ressource fondamentale pour le développement des systèmes de reconnaissance automatique de la parole pour les personnes ayant un handicap neuro-moteur [22].

Cette base de données est construite par 19 locuteurs dysarthriques. Elle est composée de 765 mots isolés.

Cinq différents auditeurs ont été choisis pour chaque locuteur sur la base des critères suivants :

Les auditeurs devraient :

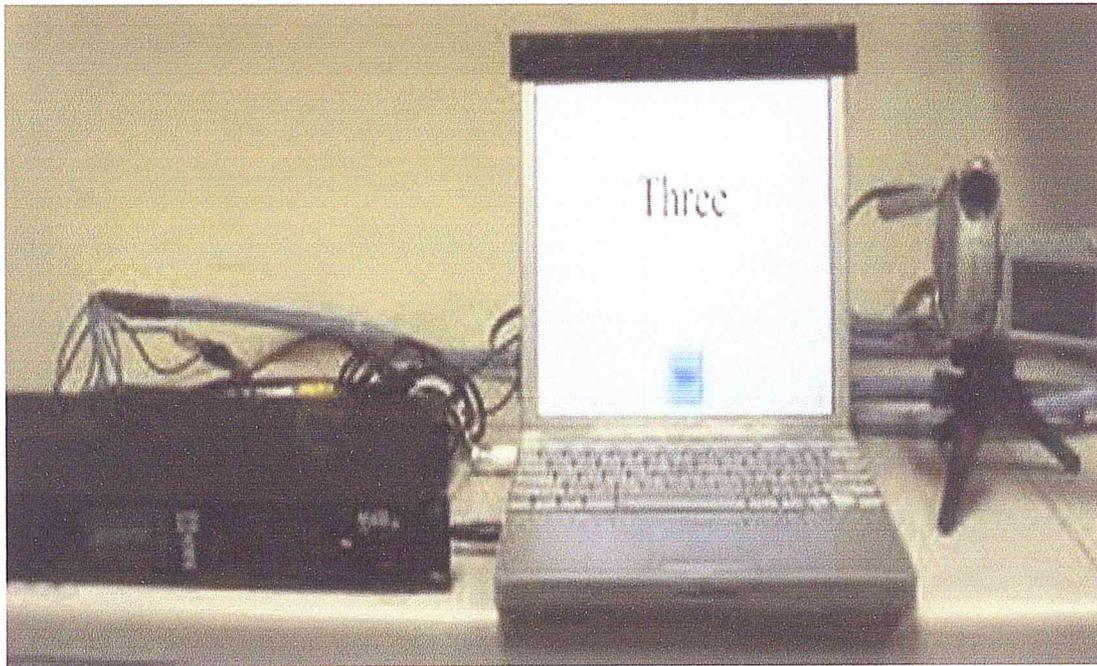
- Etre entre 18 - 40 ans ;
- Bien maîtriser l'Anglais américains ;
- Ne pas avoir une formation en troubles du langage [22].

Les locuteurs lisent des mots isolés à partir d'un écran d'ordinateur, les mots inclus sont :

- ✓ Digits (one, two, three, ...) : 10 mots répété 3 fois ;
- ✓ Lettres radio (alpha, bravo, charlie, ...) : 26 mots répété 3 fois ;
- ✓ Commandes informatiques et les commandes de traitement de texte (command, line, paragraph, enter,) : 19 mots répétés 3 fois ;
- ✓ Mots ordinaires (the, of, and,) : 100 mots répétés 3 fois;

- ✓ Mots peu fréquents (naturalization, faithfulness, frugality,) : 300 mots répétés une seule fois [23].

Les données audio ont été enregistrées à l'aide d'un microphone monté sur un écran d'ordinateur. La figure suivante représente les différents équipements utilisés pour l'enregistrement des mots de la base de données [23].



**Figure 2.3 :** *Equipements utilisés pour l'enregistrement des mots de la base de données UA Speech [23].*

Le tableau 2.1 récapitule les caractéristiques de dix-neuf locuteurs dysarthriques qui ont participé à l'enregistrement des sons de la base de données.

Les lettres M et F représentent le sexe des participants :

- M « Masculin » et F « Féminin ».

Patient	Age	Type de la dysarthrie	Intelligibilité
M01	>18	Spastique	Très faible
M02	>18	Spastique	Elevée
M03	>18	Spastique	Faible
M04	>18	Spastique	Très Faible
M05	21	Spastique	Moyen
M06	18	Spastique	Faible
M07	58	Spastique	Fiable
M08	40	Spastique	En cours de classement
M09	28	Spastique	En cours de classement
M10	18	Mixed	En cours de classement
M11	21	Athedoid	Moyen
M12	19	Mixed	En cours de classement
M13	40	Spastique	En cours de classement
M14	44	Spastique	En cours de classement
F01	30	Spastique	Faible
F02	51	Spastique	Faible
F03	18	Spastique	Très Faible
F04	22	Athedoid	Moyen
F05	>18	Spastique	Elevée

*Tableau 2.1 : Base de données UA speech [22].*

Nous avons utilisé cette base de données car elle est open-source et contient beaucoup de données

#### 2.4. Conclusion

Dans ce chapitre nous avons présenté en détails deux méthodes d'évaluation objectives de l'intelligibilité des paroles dysarthriques : (i) la méthode basée sur des références et (ii) la méthode non-basée sur des références .Nous avons également cité les différentes bases de données des paroles dysarthriques existantes dans la littérature.

La méthode d'évaluation sans références à l'avantage de s'appliquer potentiellement à de nouveaux types de données de la parole, c'est la raison pour laquelle nous avons choisi d'utiliser la méthode sans référence.

## *Chapitre 2 : Evaluation objective de l'intelligibilité de la parole dysarthrique*

---

Dans le prochain chapitre nous allons décrire en détails le système de mesure objectif de l'intelligibilité que nous proposons.

---

---

*Chapitre3*  
*Système automatique proposé*

---

---

### 3.1.Introduction

La voix est le résultat d'une coordination du larynx, conduit vocal, la langue et les lèvres. Chez une personne atteinte de la maladie de la dysarthrie, cette coordination est altérée. La voix a des rigidités, des faiblesses et des tremblements. Ces altérations représentent des indices de la maladie.

Dans ce chapitre, nous présentons le système d'aide à la décision en vue de l'estimation objective de l'intelligibilité de la parole dysarthrique. Nous présentons aussi une description des étapes de traitement et une méthode d'évaluation globale des performances.

### 3.2.Approche proposée

Le but de notre travail est la réalisation d'un système d'aide à la décision pour évaluer objectivement l'intelligibilité de la parole dysarthrique. L'approche proposée est basé sur l'utilisation de (i) l'analyse acoustique pour l'extraction des attributs et (ii) les machines d'apprentissage pour l'estimation des modèles de décision.

La figure3.1 présente l'architecture globale de notre système.

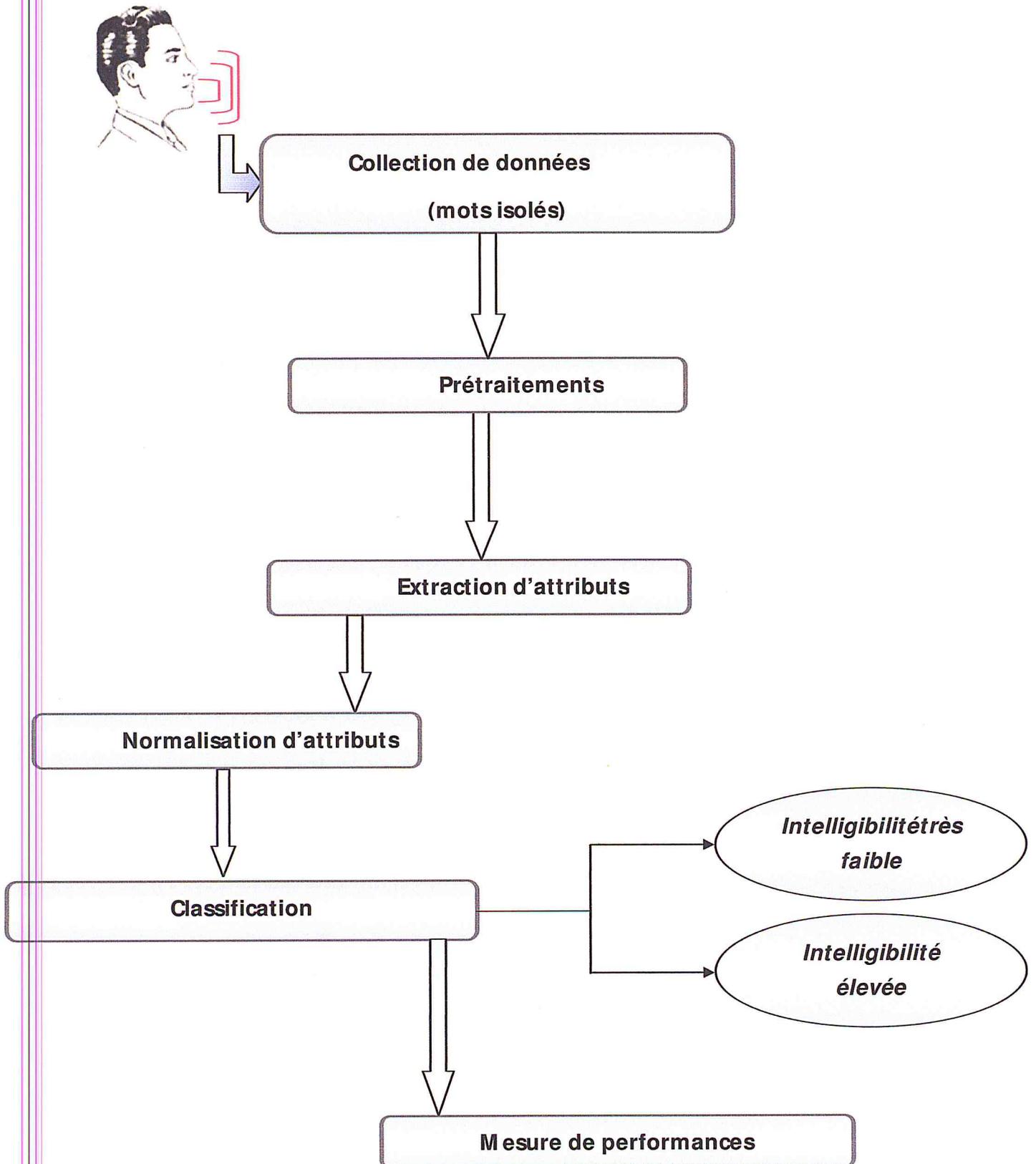


Figure 3.1. : Architecture globale de notre approche.

### 3.2.1. Collection des données

L'acquisition de la voix se fait à partir des bases de données internationales. Une base de données est un ensemble de fichiers des sons homogènes, cohérents, et représentatifs. Elle doit être facilement utilisable en alliant la souplesse, la simplicité, la convivialité et la fiabilité.

La base de données que nous avons utilisée dans notre projet est : UA Speech (voir le chapitre 2). Nous nous sommes basés sur un corpus de sons réduit qui inclut seulement la dysarthrie spastique (Tableau 3.1).

Cette base de données est composée d'un ensemble de mots isolés (455 mots), et elle se compose de deux parties, la première partie contient un ensemble de mot répétés utilisés pour l'apprentissage, et l'autre partie contient un ensemble de mot non-répétés utilisés pour le test.

Patient	Genre	Age	Intelligibilité	Classe
1	Male	18	2	Très faible
2	Male	18	15	Très faible
3	Male	28	93	Elevée
4	Femelle	51	6	Très faible
5	Male	40	91	Elevée
6	Femelle	22	95	Elevée

*Tableau3.1 : Caractéristiques des patients atteints d'une dysarthrie spastique [24].*

### 3.2.2. Prétraitements

Les bons résultats qu'un classifieur automatique peut fournir, dépendent fortement de la phase de prétraitements. Les données issues d'un mauvais prétraitement vont mettre en risque la qualité du classifieur. Cette phase consiste en une succession de traitements sur les données brutes afin d'extraire un ensemble d'informations et de garder que celles qui sont utiles à la classification.

Les phases de prétraitements que nous avons utilisés sont :

- L'élimination du silence ;
- L'extraction des fenêtres stationnaires.

### 3.2.2.1. Elimination du silence

Les sons enregistrés dans la base de données sont des signaux de parole (mots isolés). Le silence est présent au début et à la fin de chaque mot.

Pour notre cas, le silence a un effet perturbateur s'il n'est pas à l'intérieur du mot prononcé, mais il est un bon indicateur d'intelligibilité s'il y est inclus.

Pour cela, nous avons choisi d'éliminer le silence qui est à l'extérieur du mot prononcé avant d'extraire ces attributs acoustiques. La méthode la plus appropriée pour cette tâche est la détection des points du début et de fin des mots, appelés en Anglais « *End Point Detection* (EPD) ».

Plusieurs algorithmes d'EPD ont été proposés dans la littérature, parmi eux l'algorithme de **Rabiner** [25]. Cet algorithme se base sur l'énergie spectrale à courte durée et le taux de passage par zéros. Ces deux mesures sont utilisées comme paramètres de caractéristiques primaires. Si la valeur de l'énergie dans chaque segment de voix ne dépasse pas le seuil nominatif, le segment est considéré comme silence, sinon il est considéré comme signal utile.

Bien que cet algorithme contient de petites erreurs pour trouver les paramètres exacts, il a été conçu pour minimiser le nombre d'erreurs pures dans l'analyse. L'algorithme a été jugé suffisamment fiable pour la reconnaissance de parole [26]. C'est pour cette raison que nous avons choisi de l'utiliser dans notre application.

La figure 3.2 démontre un exemple d'élimination du silence d'un mot isolé.

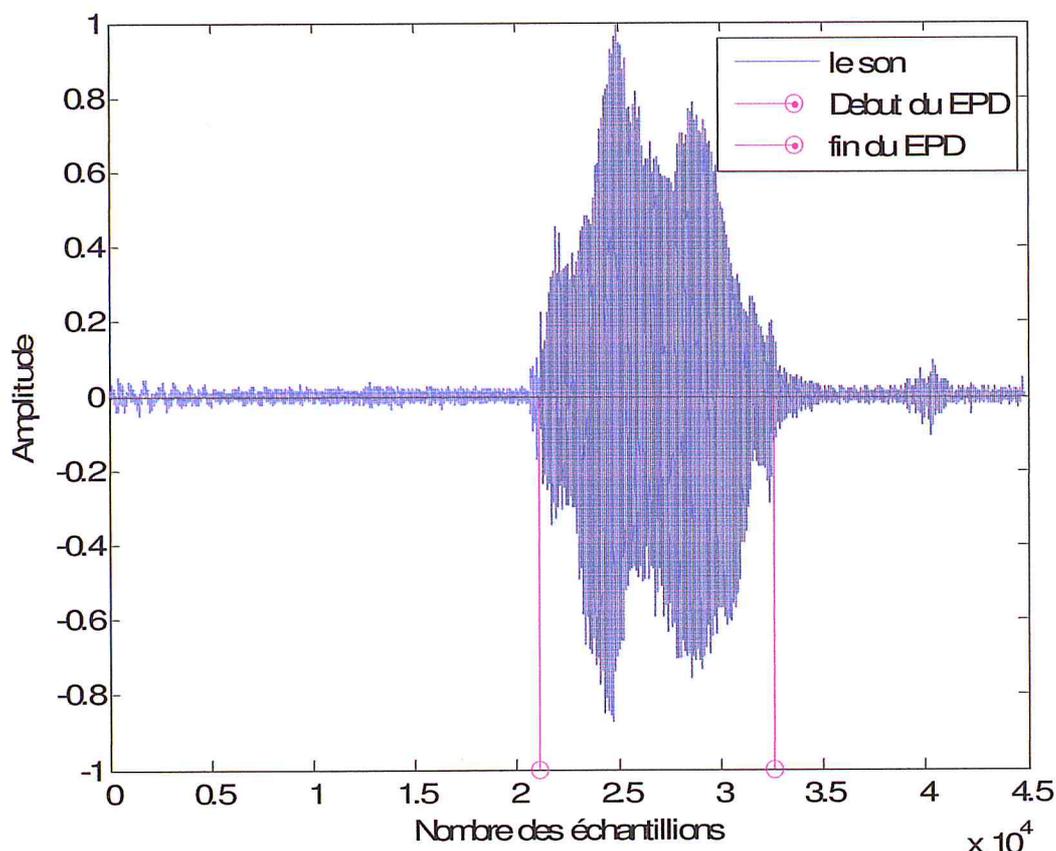


Figure 3.2 : Exemple d'utilisation de l'algorithme EPD.

### 3.2.2.2. Segmentation en trames stationnaires

La parole est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps.

La théorie de traitement de la parole suppose généralement que les propriétés du signal changent relativement et lentement avec le temps. L'analyse du signal à base des fenêtres (trames) temporelles à courte durée permet d'extraire des paramètres acoustiques qui peuvent être invariables pendant toute la durée de la fenêtre. Dans notre application, nous avons utilisé des trames de 40 ms décalée de 20ms.

#### Remarque :

**L'élimination du bruit :** Il y'a des cas de traitements où le bruit est un isolant d'où son élimination devient nécessaire, par contre dans notre cas le bruit est parfois un indicateur de la maladie, donc son élimination peut influencer le résultat final.

### 3.2.3. Extraction d'attributs

L'extraction d'attributs a pour but de mesurer un ensemble de traits acoustiques représentatifs qui caractérisent la pathologie du son. Cette étape est obligatoire avant d'entamer la phase de classification.

Nous nous sommes basés sur les critères suivants pour le choix préliminaire des attributs :

- L'opération d'extraction d'attributs doit être automatique et ne nécessite aucune intervention humaine ;
- La nature acoustique des sons de la base de données (mots isolés) doit être prise en considération dans le choix des mesures spécifiques (sons voisés, non voisés, mixtes, impulsionnels et autres) ;
- Toutes les étapes nécessaires à l'accomplissement de cette tâche doivent être prises en considération dans notre application (Fenêtrage, élimination du silence, rehaussement du signal et autres).

Nous avons exploité dans notre projet quatre différents groupes d'attributs :

- Entropies des coefficients en sous bandes selon l'échelle audiométrique ;
- Coefficients Cepstraux de MEL (MFCCs);
- Taux de silence à l'intérieur d'un mot isolé ;
- Niveau vocal actif.

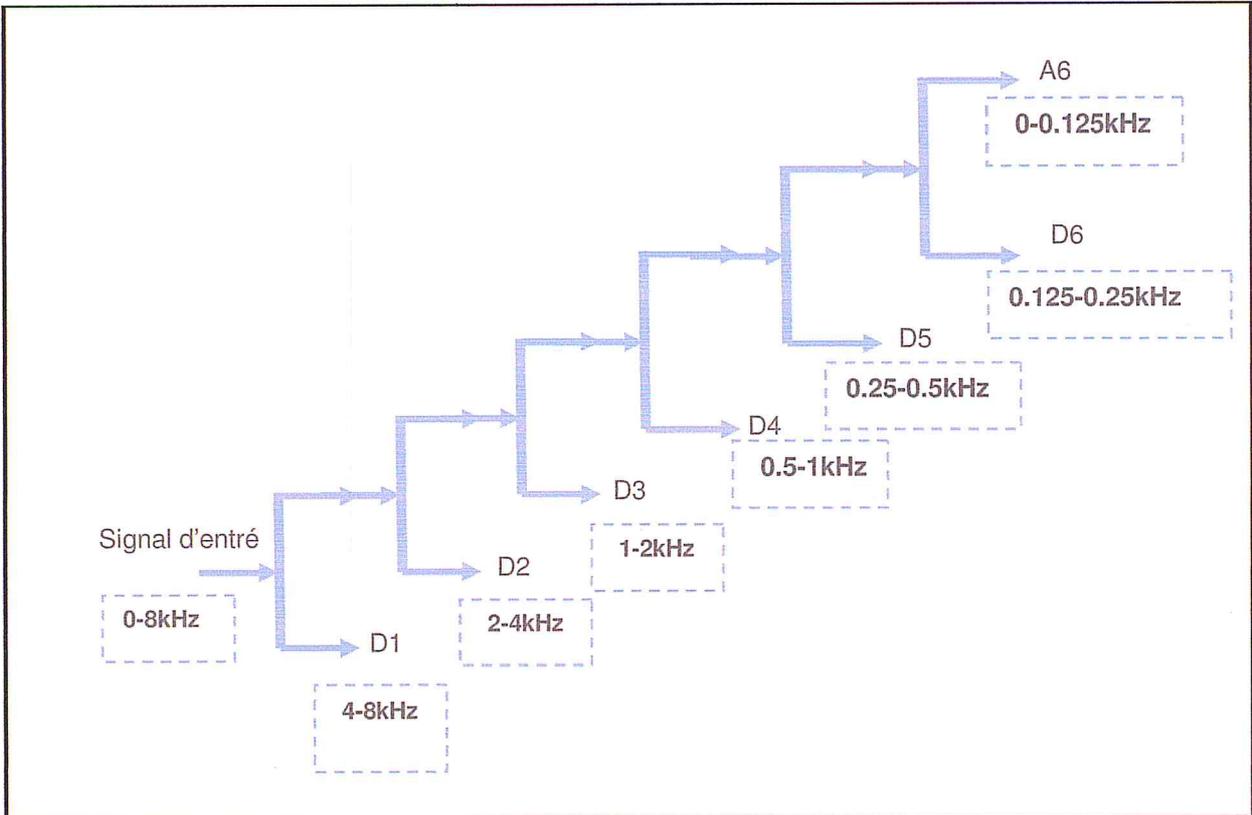
#### 3.2.3.1. Entropies des coefficients en sous bande selon l'échelle audiométrique

L'audiométrie est un test clinique qui consiste à déterminer le profil perceptuel d'une personne, c'est-à-dire à fournir un état précis sur son audition. Il s'agit de l'examen de base de l'audition humaine. Les signaux de tests sont des sons purs munis des fréquences, suivant une échelle semi logarithmique, équivalente à celle de l'oreille humaine. L'échelle des fréquences est générée par bande d'octave [26,27] : 125, 250, 500, 1000, 2000, 4000 et 8000 Hz.

Nous proposons dans notre application d'extraire un ensemble d'attributs pertinents en calculant les entropies des coefficients en sous bande selon l'échelle audiométrique.

Nous divisons le spectre du mot isolé (échantillonné à 16kHz) en plusieurs sous bandes représentatives de l'échelle audiométrique. Nous mesurons par la suite l'entropie des formes d'onde à chaque bande issue de cette décomposition.

La décomposition est achevée par une analyse multi-résolution à six niveaux de décomposition en utilisant la transformée en ondelettes discrète (Voir Annexe A pour plus de détails sur cette transformée). La Figure 3.3 représente le processus de décomposition.



**Figure 3.3 : Analyse multi résolution [48].**

L'ondelette de Daubechies d'ordre 15 a été utilisée dans notre application.

$A_6$  : Sixième coefficient d'approximation,

$D_i$  :  $i^{\text{ème}}$  coefficient de détail ( $i=1 :6$ )

Nous formons sept attributs différents en calculant l'entropie de chaque coefficient :

$$EA_{6j} = - \sum_{j=0}^{29} A_{6j}^2 \log(A_{6j}^2)$$

$$ED_i = - \sum_{ij=0}^{29} D_{6ij}^2 \log(D_{6ij}^2)$$

EA : Entropie d'approximation.

ED<sub>i</sub> :  $i^{\text{ème}}$  Entropie de détail ( $i=1 :6$ ).

### 3.2.3.2. Coefficients Cepstraux de MEL

Après la génération d'un ensemble d'attributs par le biais d'une échelle audiométrique, nous proposons aussi d'inclure d'autres attributs fréquentiels à base d'une autre échelle de perception auditive appelée échelle de MEL [28]:

$$\text{Mel}(f) = 2,595 * \log_{10}(1 + f / 700) \quad (3.1)$$

f : est la fréquence instantanée en Hz

La perception auditive humaine n'est pas aussi sensible à toutes les bandes de fréquences. En fait, elle est moins sensible à des fréquences élevées à peu près au-dessus de 1000 Hz.

Nous avons utilisé cette deuxième échelle dans notre projet pour extraire un ensemble d'attributs en calculant cette fois les énergies des bandes fréquentielles, selon l'échelle MEL. La différence entre les deux échelles est que l'échelle MEL se base sur le calcul des énergies fréquentielles tandis que l'échelle audiométrique se base sur le calcul des entropies.

Le mot entropie est parfois confondu avec l'énergie. Bien qu'ils soient des quantités connexes, ils sont distincts ;

- ✓ L'énergie permet de mesurer la capacité d'un objet ou d'un système pour faire le travail par exemple : les systèmes de reconnaissance de la parole.
- ✓ Entropique, est une mesure du "trouble" d'un système : détection de pathologie [29].

Pour représenter cette échelle nous avons utilisé les coefficients MFCCs. La procédure de calcul des MFCCs est représentée sur la Figure 3.4 et elle est décrite comme suit :

- ✓ **Préaccentuation:** Le signal vocal numérisé à l'instant  $n$ , est rehaussé pour devenir moins sensible aux effets de précision.
- ✓ **Segmentation en trames stationnaires :** section 3.2.2(B) du chapitre 3
- ✓ **Multiplication par une fenêtre de pondération :** L'étape suivante consiste à multiplier chaque trame d'analyse par une fenêtre de pondération (généralement de type Hamming ou Hanning) de manière à minimiser les discontinuités du signal au début et à la fin de la trame.
- ✓ **FFT (Fast Fourier Transform) :** les coefficients spectraux de fenêtrage des trames sont calculés en utilisant la transformée de Fourier rapide. Les résultats de cette transformée fournissent des informations sur la quantité d'énergie à chaque bande de fréquence.
- ✓ **Banc de filtre Mel :** Chaque transformée est multipliée par le gain de filtre correspondant et les résultats sont accumulés dans le filtre.
- ✓ **Log :** Le logarithme est appliqué à la sortie de l'étage précédent.
- ✓ **DCT :** Ensuite, la DCT est appliquée sur le logarithme de l'étage précédent pour obtenir les coefficients spectraux (MFCCs).

(Voir Annexe B pour plus de détails sur cette transformée).

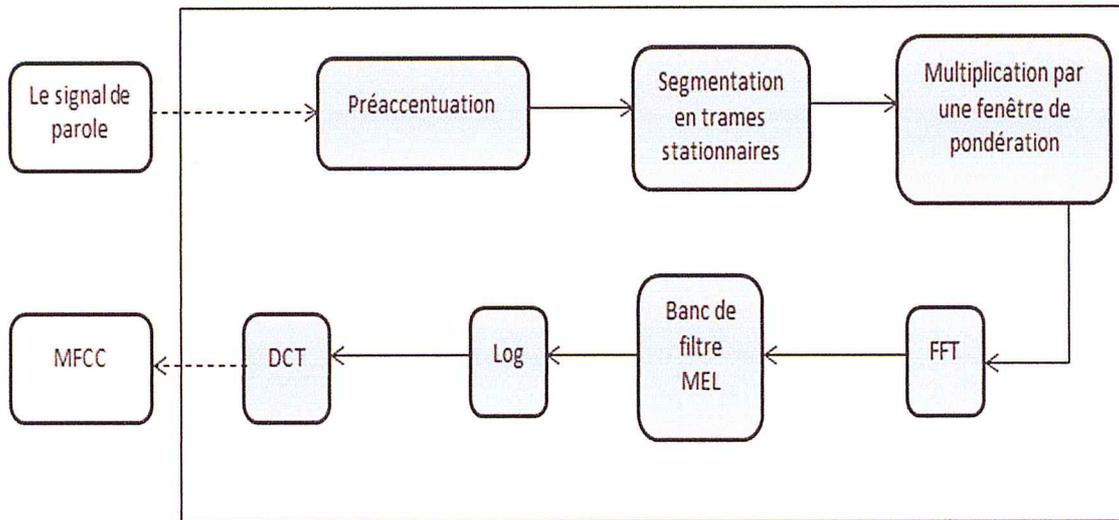


Figure 3.4 : Extraction de MFCC [30].

Le nombre de coefficients a été limité à 16 dans notre application.

MFCC(i) ; i=1 :16.

### 3.2.3.3. Taux de silence à l'intérieur d'un mot isolé

Le taux de silence à l'intérieur d'un mot isolé est un indicateur de pathologie vocale.

Pour calculer cette entité, nous avons utilisé la transformé de **Hilbert** [31].

En mathématiques et en traitement du signal, la transformée de Hilbert d'une fonction de la variable réelle est une transformation linéaire qui permet d'étendre un signal réel dans le domaine complexe, de sorte qu'il vérifie les équations de **Cauchy-Riemann** [32].

L'enveloppe du signal peut être obtenue en appliquant cette transformation.

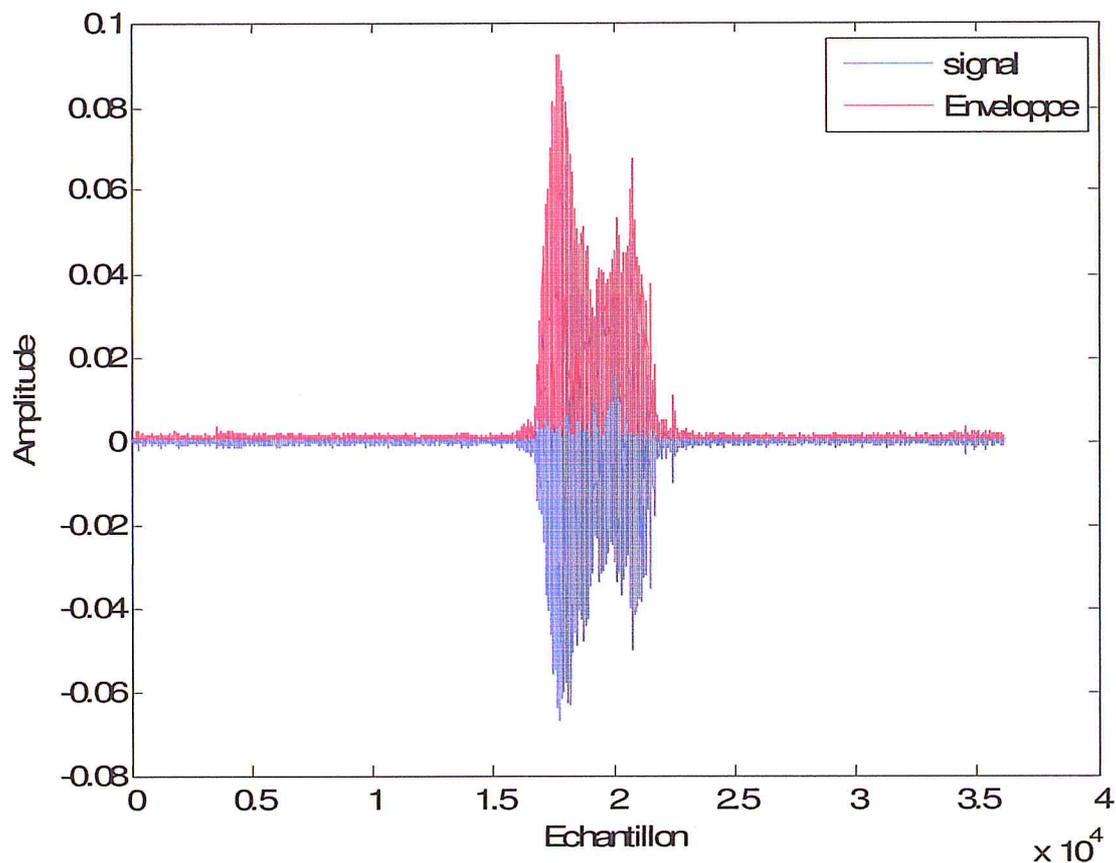


Figure3.5 : Enveloppe d'un signal.

L'estimation du taux de silence à l'intérieur du mot isolé est accomplie par l'algorithme suivant :

- (1) nous calculons l'enveloppe d'un signal X, après nous récupérons le module de cette enveloppe.
- (2) nous décomposons ensuite le module à des fenêtres à courtes durées.
- (3) nous définissons un seuil de comparaison (entropie de la première fenêtre multiplié par un entier)
- (4) nous faisons appel à l'algorithme EDP pour localiser le mot.
- (5) l'étape suivante consiste à faire l'analyse sur les segments en comparant l'entropie de chaque segment (issue du calcul du module d'enveloppe) avec le seuil défini ;

*SI entropie (frame) =>seuil*

*Zone active*

$$\text{Sinon Taux de Silence} = \frac{DA}{DS}$$

DA : durée de la zone active à l'intérieur du mot

DS : sur la durée du mot

La figure 3.6 montre un exemple de calcul du silence à l'intérieur d'un mot isolé.

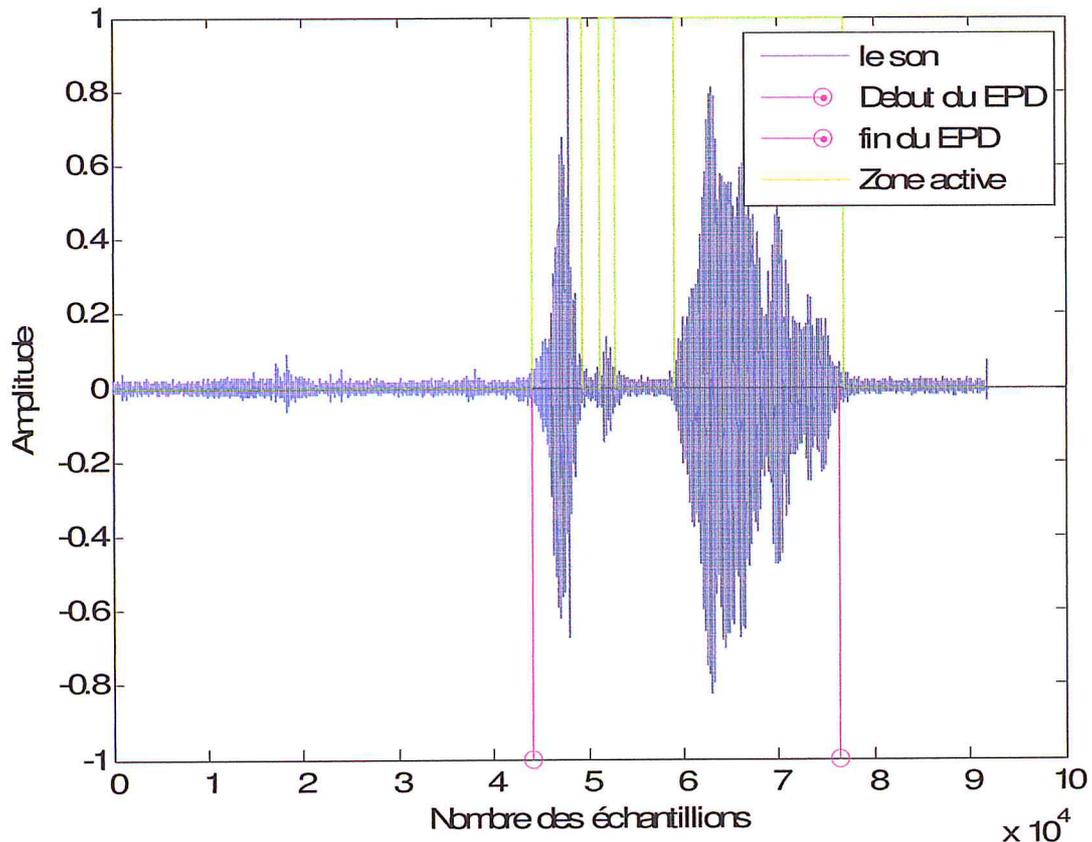


Figure 3.6 : Taux de silence à l'intérieur d'un mot isolé.

#### 3.2.3.4. Niveau vocal actif

Le CCIT (*Comité consultatif international télégraphique*) estime qu'il est important de disposer d'une méthode normalisée pour la mesure objective du niveau de la parole, cette organisation permet de mesurer la qualité des systèmes téléphoniques. Pour notre cas, la qualité vocalique est un bon indicateur de pathologie.

✓ **La norme UIT-T (P.56)** permet de calculé le niveau vocal actif.

La puissance moyenne d'un signal vocal que l'on sait être présent peut s'exprimer avec une grande précision d'après des échantillons. Cependant, la question la plus importante est de savoir quel critère doit être utilisé pour déterminer la présence de la parole.

L'idéal serait que le critère indique la présence de la parole pendant le même pourcentage de temps où cette présence est perçue par une personne qui écoute, non compris

le bruit qui ne fait pas partie de la parole (par exemple, impulsions, bruit stable pendant les périodes de silence), mais y compris les courtes périodes pendant lesquelles la puissance est faible ou nulle et qui ne sont pas perçues comme des interruptions dans l'écoulement de la parole [33].

Tous les attributs précédents sont utilisés pour la construction d'une machine d'apprentissage qu'on va utiliser dans l'étape suivante (la classification).

#### 3.2.4. Normalisation d'attributs

La normalisation permet d'ajuster une série de valeurs suivant une fonction de transformation pour les rendre comparables avec quelques points de références spécifiques. La normalisation des données est nécessaire quand l'incompatibilité des unités de mesures entre les variables peut affecter les résultats [34].

Pour notre cas nous avons des données d'apprentissage et des données de test, la normalisation est appliquée sur les données de test pour les rendre comparables avec le point de référence (données d'apprentissage).

La fonction de normalisation remet les valeurs de chaque entité dans les données qui ont une moyenne nulle (lorsque la soustraction de la moyenne dans le recenseur) et la variance unitaire [35].

Cette méthode est très utilisée pour la normalisation dans de nombreux algorithmes de machine d'apprentissage (par exemple, machines à vecteurs supports, régression logistique, et les réseaux de neurones). Cela se fait habituellement en calculant les scores standards. La méthode générale de calcul est de déterminer la moyenne de la distribution et l'écart type pour chaque fonction. Ensuite, nous soustrayons la moyenne de chaque fonction. Après, nous divisons les valeurs (la moyenne est déjà soustraite) de chaque fonction par son écart type.

$$x' = \frac{x - \bar{x}}{\sigma} \quad (3.2)$$

Où :

$x$  Est le vecteur de caractéristiques d'origine ;

$x'$  Est la moyenne de ce vecteur de caractéristique,

$\sigma$  Est son écart-type.

### 3.2.5. Classification

Les méthodes de classification ont pour but d'identifier les classes auxquelles appartiennent des objets à partir de certains paramètres descriptifs. Elles s'appliquent à un grand nombre d'activités humaines et conviennent en particulier au problème de la prise de décision automatisée. La classification est effectuée par le biais des machines d'apprentissages.

La procédure de classification sera extraite automatiquement à partir d'un ensemble d'exemples. Un exemple consiste en la description d'un cas avec la classification correspondante. Un système d'apprentissage doit alors, à partir de cet ensemble d'exemple, extraire une procédure de classification, il s'agit en effet d'extraire une règle générale à partir des données observées. La procédure générée devra classer correctement les exemples de l'échantillon et avoir un bon pouvoir prédictif pour classer correctement de nouvelles descriptions [36].

Il existe plusieurs méthodes de classification (supervisées, non supervisées ou semi-supervisés) dans la littérature, nous citons :

- ✓ HMMs [37];
- ✓ GMMs (Gaussian Mixture Models) [38];
- ✓ SVMs (*Support Vector Machine*) [39];
- ✓ Réseaux de neurones [40];

La méthode classification que nous avons exploitée dans notre application est à base des mélanges de Gaussiennes.

#### 3.2.5.1. Pourquoi nous avons utilisé les modèles GMMs ?

Les GMMs sont souvent utilisés dans les systèmes biométriques, notamment dans les systèmes de reconnaissance du locuteur, en raison de leur capacité de représenter une grande classe de distributions d'échantillons. Un des avantages puissants du modèle GMM est sa capacité à former des approximations de forme arbitraire des densités [41].

Le problème d'estimation du degré d'intelligibilité de la parole pathologique est très semblable au problème d'identification et reconnaissance du locuteur. Vu que les GMMs ont donné de très bons résultats dans ce genre d'applications, nous estimons que ces modèles sont adéquats à notre problématique.

### 3.2.5.2. Description du modèle

Un modèle GMM est une fonction de probabilité de densité paramétrique représentée comme une somme équilibrée des poids d'éléments de Gauss. Il s'agit des modèles paramétriques de représentation des distributions de probabilité [41].

Un modèle de mélanges de gaussiennes est une somme pondérée des densités de M composantes :

$$p(x|\lambda) = \sum_{i=1}^N w_i g(x|u_i, \Sigma_i) \quad (3.3)$$

$x$  : Vecteur de données.

$w_i$  : sont les poids du mélange gaussiens.

$g(x|u_i, \Sigma_i)$  : sont les densités des composantes gaussiennes.

$u_i$  : Vecteur des moyennes

$\Sigma_i$  : La  $i^{\text{ème}}$  Matrice de covariance

$N$  : le nombre de Gaussiennes

Le modèle ( $\lambda$ ) de mélange de gaussiennes est paramétré par :

- $u_i$
- $w_i$
- $\Sigma_i$

### 3.2.5.3. Estimation des paramètres

Pour estimer les paramètres d'une GMM, Il existe deux techniques disponibles. Parmi eux : Le maximum de vraisemblance (en Anglais : Maximum Likelihood « ML »). C'est la méthode la plus populaire et bien établie. Le but de l'estimation du ML est de trouver les paramètres du modèle qui maximisent la densité de probabilité (Vraisemblance) en utilisant les données d'apprentissage.

La maximisation de vraisemblance peut être achevée en utilisant de manière itérative un cas particulier de l'algorithme Expectation-Maximisation [42].

### 3.2.5.4. Idée de base de l'algorithme EM

Commencer par un modèle initial  $M_1$  Pour estimer un nouveau modèle  $M_2$  telle que :

$$p(X|M_2) > p(X|M_1)$$

Le nouveau modèle devient alors le modèle initial pour la prochaine itération et le processus est répété jusqu'à ce qu'un certain seuil de convergence est atteint [42].

### Formalisation de l'algorithme EM [42]

Nous disposons d'observations  $X = (X_1, \dots, X_n)$  de vraisemblance notée  $p(X|M)$  ;

- ✓ maximiser  $\log p(X|M)$  est impossible ;
- ✓ on considère des données cachées  $Z = (Z_1, \dots, Z_n)$  dont la connaissance rendrait possible la maximisation de la «vraisemblance des données complètes»,  $\log p(X, Z|M)$  ;
- ✓ comme on ne connaît pas ces données  $Z$ , on estime la vraisemblance des données complètes en prenant en compte toutes les informations connues : l'estimateur est naturellement  $E_{Z|X, M_m}(\log p(X, Z|M))$  « étape E » de l'algorithme EM;
- ✓ et on maximise enfin cette vraisemblance estimée pour déterminer la nouvelle valeur du paramètre : « étape M » de l'algorithme EM.

Ainsi, le passage de l'itération  $m$  à l'itération  $m+1$  de l'algorithme consiste à déterminer :

$$M_{m+1} = \text{Arg max}_M \{E_{Z|X, M_m}(\log p(X, Z|M))\} \quad (3.4)$$

La figure 3.7 représente la création de deux modèles différents.

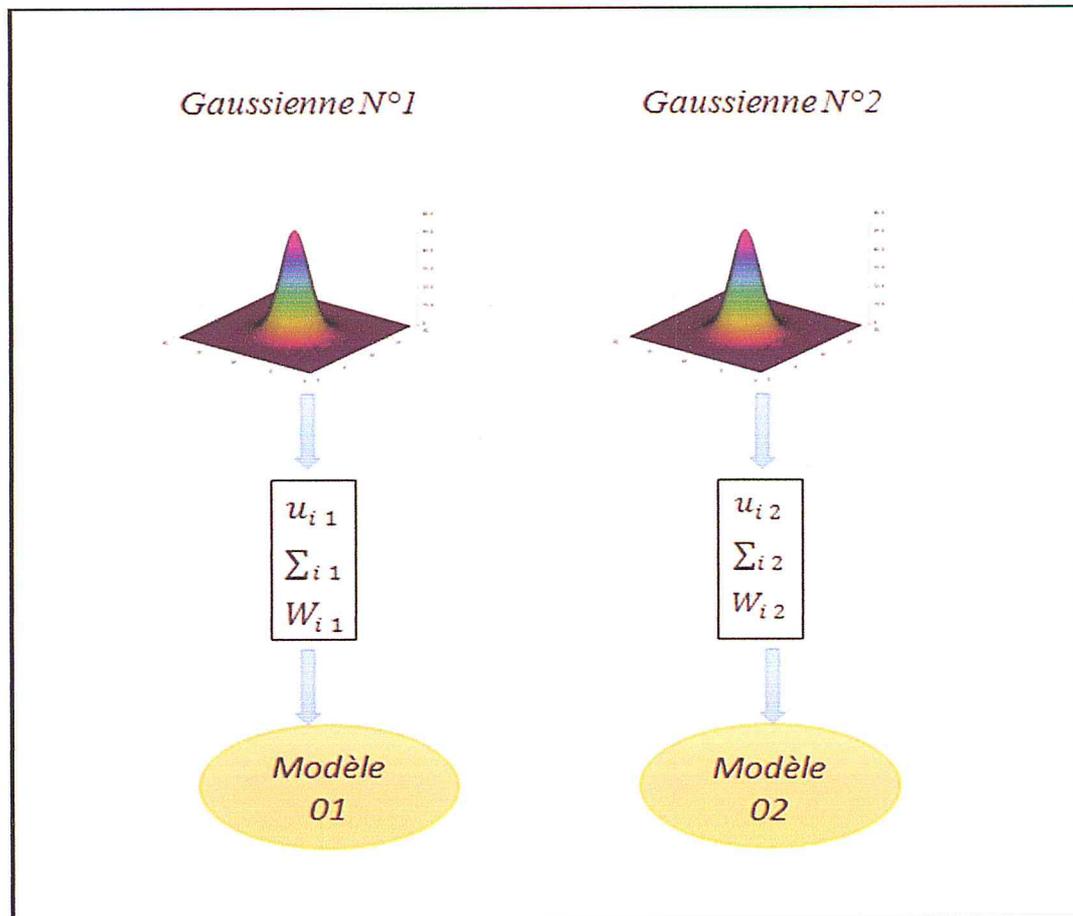


Figure 3.7 : Représentation des modèles gaussiens.

### 3.2.6. Mesure de performance

Evaluer les performances d'un classifieur est un enjeu de grande importance. Les performances globales sont utilisées dans l'étape d'optimisation des paramètres du classifieur. Pendant longtemps, le critère retenu pour évaluer ses performances a été le taux de bonne classification, c'est-à-dire le nombre d'éléments d'une base de tests correctement classés [36].

L'évaluation globale est basée sur un tableau à deux dimensions, appelé matrice de confusion. Cette matrice trie tous les cas du modelé en catégorie, en déterminant si la valeur prédite correspondait à la valeur réelle. Tous les cas dans chaque catégorie sont ensuite comptés et les totaux sont affichés dans la matrice. Les mesures que nous allons évoquer dans cette section utilisent la matrice de confusion.

Le tableau 3.2 représente une illustration de la matrice de confusion.

	Décision positive	Décision négative	
Etiquette Positive	Vrais Positifs (VP)	Faux Négatifs (FN)	Pos
Etiquette Négative	Faux Positifs (FP)	Vrais Négatifs (VN)	Neg
Total (T)	Ppos	Pneg	N

**Tableau3.2 : Matrice de confusion.**

Les paramètres de cette matrice sont décrits comme suit [36] :

Etiquettes positive : niveau d'intelligibilité faible.

Etiquettes négative : niveau d'intelligibilité élevé.

Décision positive : résultat positif.

Décision négatif : résultat négatif.

Vrais positifs (VP) : individus avec un niveau d'intelligibilité élevé classé élevé.

Vrais négatifs (VN) : individus avec un niveau d'intelligibilité élevé classé très faible

Faux positifs (FP) : individus avec un niveau d'intelligibilité très faible très faible.

Faux négatifs (FN) : individus avec un niveau d'intelligibilité faible classé élevé.

❖ Taux de Vrais Positif (TVP)

$$TVP = \frac{VP}{Pos} = \frac{VP}{VP+FN} \quad (3.5)$$

❖ Taux de Vrais Négatifs (TVN)

$$TVN = \frac{VN}{Neg} = \frac{VN}{VN+FN} \quad (3.6)$$

❖ Taux de Faux Positifs (TFP)

$$TFP = \frac{FP}{Neg} = \frac{FP}{VN+FP} \quad (3.7)$$

❖ Taux de Faux Négatifs (TFN)

$$TFN = \frac{FN}{Pos} = \frac{FN}{VP+FN} \quad (3.8)$$

❖ Taux de bonne classification ou l'exactitude (TBC)

$$TBC = \frac{VP+VN}{VP+FN+VN+FP} \quad (3.9)$$

❖ Précision (P)

$$P = \frac{VP}{PPos} = \frac{VP}{VP+FP} \quad (3.10)$$

- **Stratégie de validation**

Il s'agit de calculer les erreurs décrites après l'exécution du programme en se basant sur l'une des stratégies de la validation croisée.

- **Validation croisée**

C'est une méthode d'estimation de fiabilité d'un modèle fondé sur une technique d'échantillonnage. En fait, il y a au moins trois techniques de validation croisée : « *testset validation* » ou « *holdout method* », « *k-fold cross-validation* » et « *leave-one-out cross-validation* » et nous allons utiliser la technique Holdout [43].

- ✓ **Holdout** : Cette méthode a été développée à partir des techniques de validation croisée, mais elle n'utilise qu'un seul ensemble de données pour la validation et n'utilise pas de « croisement » lors de la validation. La différence entre la validation croisée et la validation de type hold-out est importante car la validation croisée est supérieure pour des petits ensembles de données. La validation hold-out implique la séparation de la base de données d'apprentissage en deux parties distinctes et fixes : une base d'apprentissage et

une base de validation. La base de validation servira de test après chaque époque d'apprentissage.

On divise l'échantillon de taille  $N$  en un échantillon d'apprentissage et un échantillon de test. Ici il faut faire attention, car on doit réserver la plus grande partie pour l'apprentissage. Généralement cet échantillon contient plus que 60% de l'échantillon principal [44].

### **3.3. Conclusion**

Dans ce chapitre, nous avons présenté en détail une approche de diagnostic des troubles de la dysarthrie à partir du signal vocal. Les étapes de traitement sont : l'acquisition de la voix, le prétraitement, l'extraction d'attributs, la normalisation d'attributs, la classification, et enfin les mesures de performances.

Pour la partie d'extraction d'attributs, nous nous sommes intéressés aux: Entropies des coefficients en sous bandes selon l'échelle audiométrique, les coefficients Cepstraux de MEL (MFCCs), le taux de silence à l'intérieur d'un mot isolé et le niveau vocal actif. La théorie de la classification à base des GMMs a été présentée dans la partie de classification de notre travail. La dernière section de ce chapitre a été consacrée aux méthodes d'évaluation des classifieurs. Le prochain chapitre va être consacré aux résultats et aux interprétations.

---

---

*Chapitre 4*  
*Expérimentations, résultats et interprétations*

---

---

### 4.1. Introduction

Dans ce chapitre nous discuterons les résultats du test de notre système, ainsi que les performances en faisant le calcul des mesures de performance pour voir à quel point il est fiable, et nous allons aussi réaliser les interfaces graphiques qui vont permettre l'interaction entre le système et le malade.

### 4.2. Logiciels de développement

Nous allons définir les deux langages que nous avons utilisés pour le développement.

#### 4.2.1. Environnement MATLAB [36]

C'est en exploitant les fonctionnalités de l'environnement de Développement Intégré(EDI) que les principaux travaux de ce projet ont été réalisés.

C'est la version 2015 qui a été utilisé (MATLAB R2015a). Un grand nombre de fonctions. Le choix de l'environnement MATLAB a été fait en raison de la rapidité des calculs et la facilité de développement sous ce dernier.

##### ✓ StatisticsToolbox

Les toolboxes sont réellement des caisses à outils comportant une collection de fonctions relatives à plusieurs domaines scientifiques et techniques, nous avons utilisé la toolboxstatistics. Cette toolbox fournit des algorithmes et des outils d'apprentissage statistique pour l'organisation, l'analyse et la modélisation des données.

Pour l'analyse des données multidimensionnelles, StatisticsToolbox inclut des algorithmes permettant la mise en œuvre des modèles.la boite à outils fournit aussi des algorithmes d'apprentissage supervisées ou non.

#### 4.2.2. Langage JAVA

Le langage **Java** est un langage de programmation informatique orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982).

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes\_d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et frame works associés visent à guider, sinon garantir, cette portabilité des applications développées en Java [45].

### ✓ L'environnement Netbeans [46]

NetBeans est un environnement de développement intégré (EDI), placé en *open source* par Sun en juin 2000 sous licence CDDL (Common Development and Distribution License) et GPLv2. En plus de Java, NetBeans permet également de supporter différents autres langages, comme C, C++, JavaScript, XML, Groovy, PHP et HTML de façon native ainsi que bien d'autres (comme Python ou Ruby) par l'ajout de greffons. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web). Conçu en Java, NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et SPARC), Mac OS X ou sous une version indépendante des systèmes d'exploitation (requérant une machine virtuelle Java). Un environnement Java Development Kit JDK est requis pour les développements en Java.

NetBeans constitue par ailleurs une plateforme qui permet le développement d'applications spécifiques (bibliothèque Swing (Java)). L'IDE NetBeans s'appuie sur cette plateforme. L'IDE Netbeans s'enrichit à l'aide de greffons.

### 4.3. Corpus de tests

#### 4.3.1. Evaluation du système globale

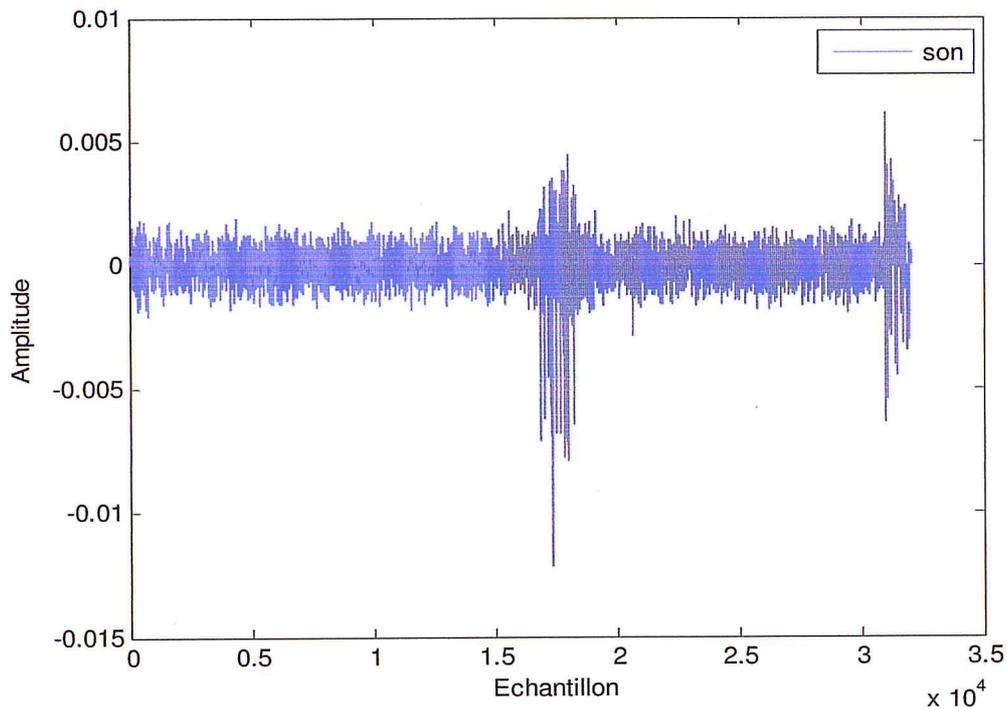
Notre base de données est construite à partir de deux types de sons : les sons « *Repeated* » et les sons « *Not repeated* ».

Les sons « *repeated* » sont les mots qui se répètent dans la base de données, et les « *not repeated* » sont les mots qui ne se répètent pas.

Les sons « *repeated* » sont dédiés au traitement et à l'extraction des attributs pour construire les modèles d'apprentissage. Les sons « *not repeated* » sont consacrés pour le test du système.

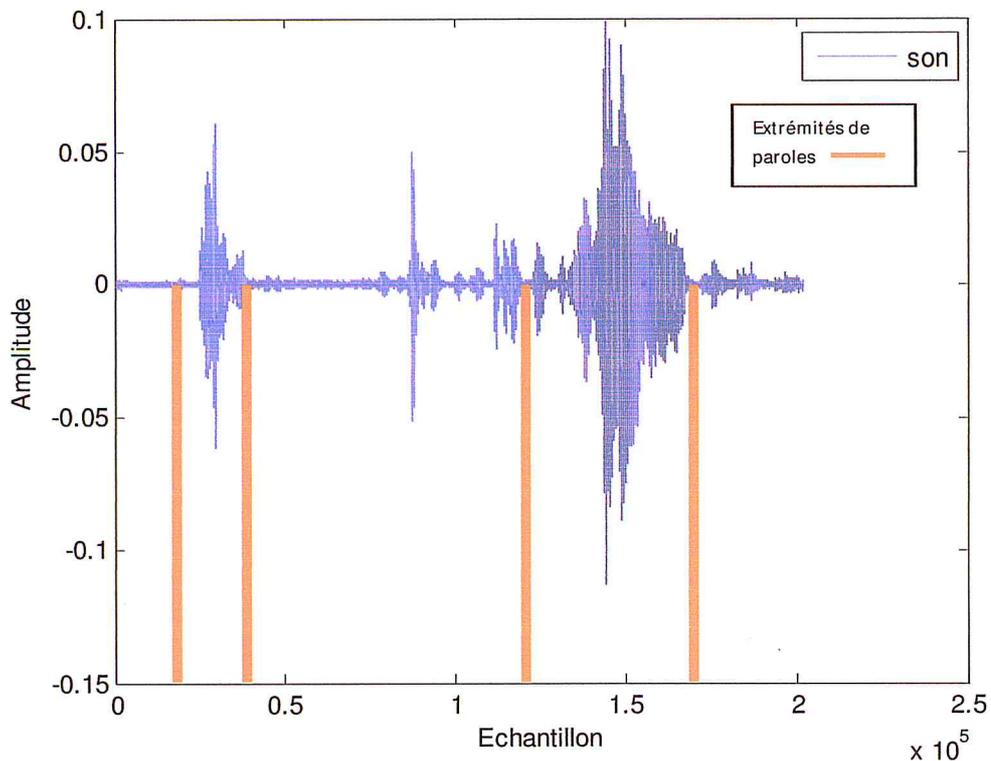
Lors du développement de notre application, nous avons rencontré des problèmes au niveau de la base de données. Nous avons trouvé quelques sons complètement noyés dans le bruit au niveau de la base d'apprentissage. Il y a même parfois des sons où nous pouvons remarquer qu'il y a deux personnes (le médecin et le patient) qui parlent au cours de l'enregistrement du son.

Ce genre d'enregistrement ne peut pas être utilisé pour construire les modèles de prédiction. Ce qui nous a poussé vers la suppression de ces sons et avoir par la suite une base de données ajustées, sachant que le nombre de ces sons n'était pas grand seulement une dizaine de sons.



*Figure4.1 : Exemple d'un son plein de bruit.*

La figure 4.2 montre un son où il y a l'apparition de deux personnes qui parlent au cours de l'enregistrement, les extrémités en couleur orange montre le début et la fin des deux messages vocaux.



*Figure 4.2 : Exemple d'un son où le médecin parle au cours de l'enregistrement.*

L'application de la technique Hold-out permet la séparation de la base de données en deux parties (apprentissage et test).

#### 4.3.2. Partie apprentissage

La partie d'apprentissage « les sons repeated » (62% de l'ensemble de données représentés sur la figure 4.3).

La partie d'apprentissage consiste à :

- Extraire des attributs à partir des sons de la base de données.
- Remplir la matrice d'apprentissage par ces attributs calculés.
- Estimer les modèles de prédiction à partir de la matrice d'apprentissage.

#### 4.3.3. Partie de test

Le test est effectué « les sons not repeated » (38% de l'ensemble de données représentées sur la figure 4.3).

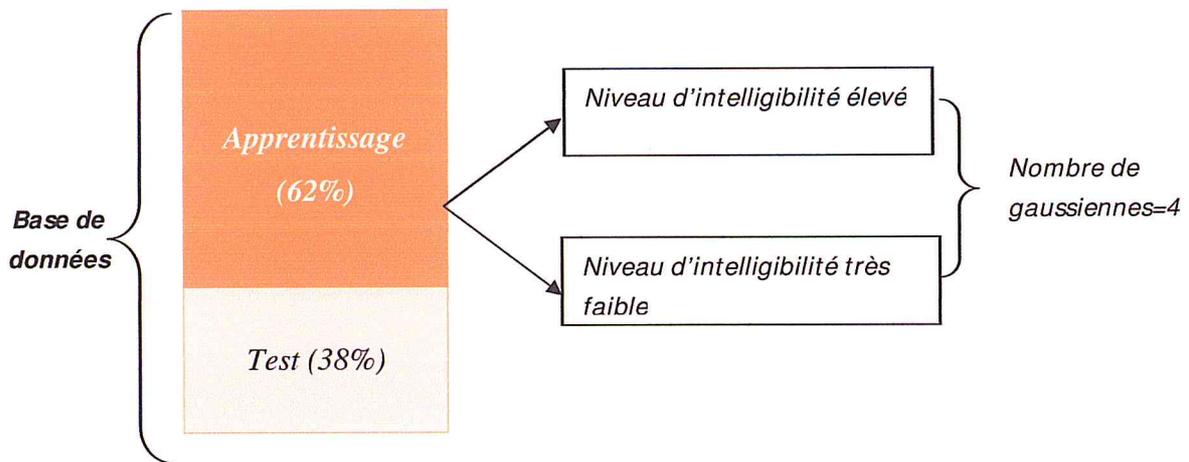


Figure 4.3 : Schéma simplicité de la méthode d'évaluation du système global.

#### 4.4. Mesures de performances

Dans une problématique de classification, pour évaluer les performances des systèmes de décision, nous utilisons les mesures suivantes : taux de classification, matrice de confusion, le taux de vrai positif, le taux de vrai négatif, le taux de faux positif, le taux de faux négatif, et la précision.

##### 4.4.1. Matrice de confusion

Une matrice de confusion ou tableau de contingence sert à évaluer la qualité d'une classification. Elle est obtenue en comparant les données classées avec des données de référence qui doivent être différentes de celles ayant servi à réaliser la classification [43].

##### 4.4.2. Variation du nombre de gaussiennes

Pour voir l'influence des nombres de gaussiennes sur les performances du système, nous allons varier le nombre de gaussiennes de 2 jusqu'à 5.

Nous avons construit la matrice de confusion en changeant à chaque fois le nombre de gaussiennes pour ensuite calculer le taux de bonne classification (TBC).

Le nombre de gaussiennes qui donne le TBC le plus élevé va être appliqué dans le programme.

##### 4.4.2.1. Matrice de confusion avec nombre de gaussiennes égale à 2

Les deux tableaux 4.1 et 4.2 montrent respectivement la matrice de confusion et les taux de performances pour un nombre de gaussiennes égale à 2.

	Décision positive	Décision négative
Etiquette Positive	754 (VP)	100 (FN)
Etiquette Négative	47 (FP)	831 (VN)
Total (T)	801	931

*Tableau4.1 : Matrice de confusion avec un nombre de gaussiennes N=2.*

	Décision positive (Taux)	Décision négative (Taux)
Etiquette Positive	0.88 (TVP)	0.11 (TFN)
Etiquette Négative	0.05 (TFP)	0.94 (TVN)

*Tableau4.2 : Taux de performance avec un nombre de gaussiennes N=2.*

La précision  $P = \frac{754}{754+47} = 0.94$

$TBC1 = \frac{VP+VN}{VP+FN+VN+FP} = \frac{754+831}{754+100+831+47} = 0.915$

Nous allons refaire la même chose en prenant le nombre de gaussiennes à 3.

4.4.2.2. Matrice de confusion avec nombre de gaussiennes égale à 3

Le tableau suivant montre la matrice de confusion en mettant le nombre de gaussiennes égale à 3.

	Décision positive	Décision négative
Etiquette Positive	832 (VP)	22 (FN)
Etiquette Négative	47 (FP)	831 (VN)
Total (T)	879	853

Tableau4.3 : Matrice de confusion avec un nombre de gaussiennes  $N=3$ .

Le tableau 4.4 montre les taux de performance en prenant le nombre de gaussiennes égale à 3.

	Décision positive (Taux)	Décision négative (Taux)
Etiquette Positive	0.97 (TVP)	0.02 (TFN)
Etiquette Négative	0.05 (TFP)	0.94 (TVN)

Tableau4.4 : Les taux de performance avec un nombre de gaussiennes  $N=3$ .

La précision  $P = \frac{832}{832+47} = 0.94$

$TBC2 = \frac{VP+VN}{VP+FN+VN+FP} = \frac{832+831}{832+22+47+831} = 0.960$

Dans ce cas, nous avons trouvé un TBC de 0.960. Ce résultat est légèrement supérieur par rapport à TBC1 (TBC pour le nombre de gaussiennes égale à 2). Nous allons refaire la même chose en prenant le nombre de gaussiennes égale à 4 pour calculer le TBC3.

#### 4.4.2.3. Matrice de confusion avec nombre de gaussiennes égale à 4

Les deux tableaux suivants (4.5 et 4.6) illustrent respectivement la matrice de confusion et les taux de performance avec un nombre de gaussiennes égale à 4.

	Décision positive	Décision négative
Etiquette Positive	835 (VP)	19 (FN)
Etiquette Négative	48 (FP)	830 (VN)
Total (T)	883	849

Tableau4.5: Matrice de confusion avec un nombre de gaussiennes N=4.

	Décision positive (Taux)	Décision négative (Taux)
Etiquette Positive	0.97 (TVP)	0.02 (TFN)
Etiquette Négative	0.05 (TFP)	0.94 (TVN)

Tableau4.6 : Les taux de performance avec un nombre de gaussiennes N=4.

La précision  $P = \frac{835}{835+48} = 0.94$

$$TBC3 = \frac{VP+VN}{VP+FN+VN+FP} = \frac{835+830}{835+19+48+830} = 0.961$$

Le TBC3 dans ce cas est de 0.961, il y'a une différence de 0.001 avec le TBC2.

#### 4.4.2.4. Matrice de confusion avec nombre de gaussiennes égale à 5

Les deux tableaux suivants 4.7 et 4.8 montrent respectivement la matrice de confusion et les taux de performance.

	Décision positive	Décision négative
Etiquette Positive	829 (VP)	25 (FN)
Etiquette Négative	50 (FP)	828 (VN)
Total (T)	879	853

*Tableau4.7 : Matrice de confusion avec un nombre de gaussiennes N=5.*

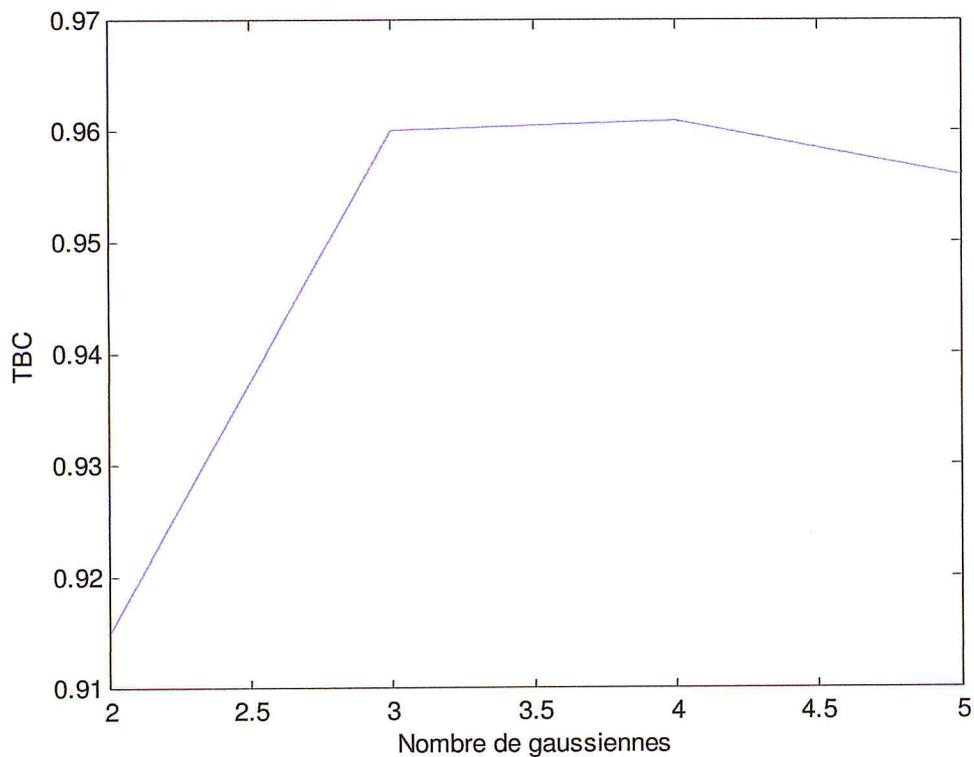
	Décision positive (Taux)	Décision négative (Taux)
Etiquette Positive	0.97 (TVP)	0.03 (TFN)
Etiquette Négative	0.05 (TFP)	0.94 (TVN)

*Tableau4.8 : Les taux de performance avec un nombre de gaussiennes N=5.*

La précision  $P = \frac{829}{829+50} = 0.94$

$$TBC4 = \frac{VP+VN}{VP+FN+VN+FP} = \frac{829+828}{829+828+50+25} = 0.956$$

Pour résumer ces résultats, nous proposons une courbe qui décrit l'évolution des TBC selon les différents nombres de gaussiennes (voir figure 4.4).



*Figure 4.4 : Graphe de changement du TBC selon la variation du nombre de gaussiennes.*

✓ **Discussion**

Après la variation du nombre de gaussiennes, nous avons remarqué que la valeur de TBC dans le cas de gaussiennes =2 est légèrement inférieur par rapport aux gaussiennes=3 et 4. Par contre, nous avons obtenu pratiquement le même résultat pour les gaussiennes=3 et 4. A partir du nombre de gaussiennes=5, la valeur du TBC diminue progressivement. Nous remarquons aussi que la valeur de la précision est la même dans tous les cas.

Pour notre application, nous avons donc choisi la valeur du gaussiennes 4, qui donne le meilleur résultat selon les expérimentations réalisées.

### 4.5. Résultats

Après programmation des fonctions du système sous Matlab et la confirmation de la fiabilité de notre système en utilisant les mesures de performance, nous avons développé des interfaces graphiques sous Java qui vont permettre au décideur d'interagir avec notre système.

#### 4.5.1. Etapes du traitement

Nous allons montrer les différentes étapes de l'exécution du système

##### 4.5.1.1. Interface d'accueil

L'interface d'accueil permet à l'utilisateur d'accéder à l'interface principale en appuyant sur le bouton « démarrer ».



Figure4.5: Interface d'accueil.

#### 4.5.1.2. Interface du menu principale

Cette interface permet d'afficher la décision finale sur le niveau d'intelligibilité de la parole dysarthrique de la personne qui a enregistré sa parole.

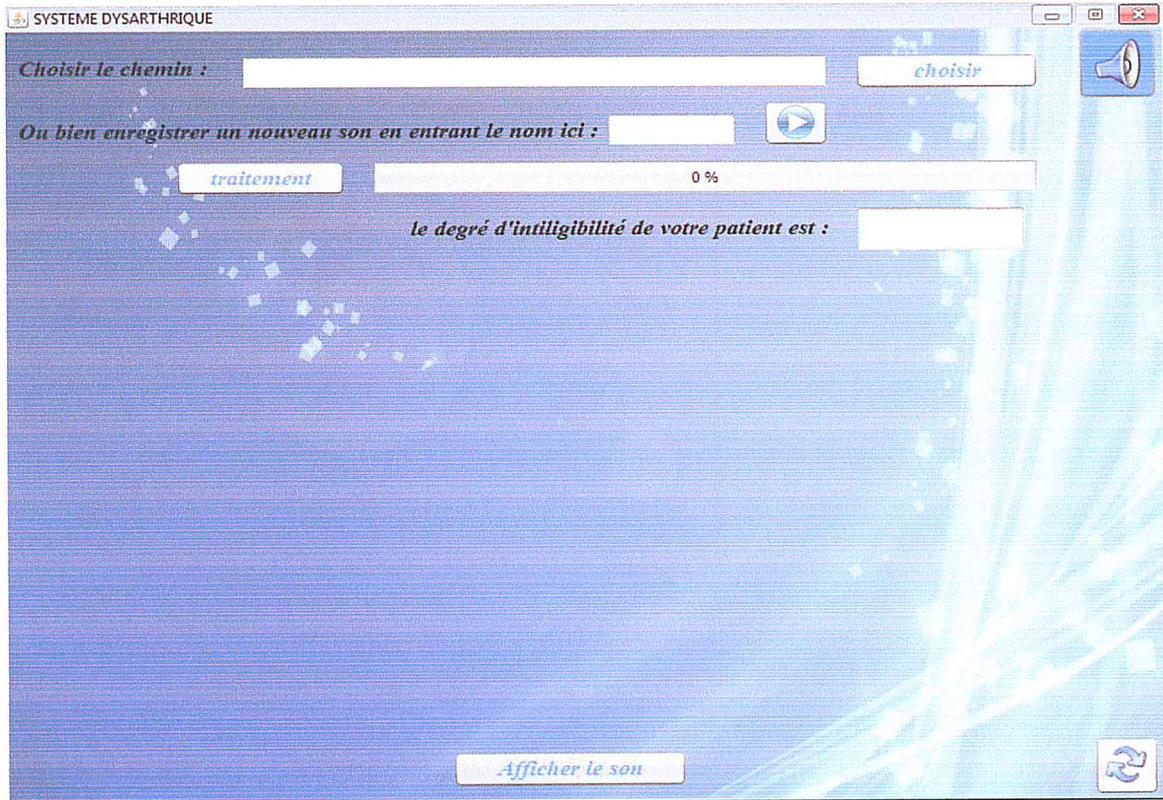


Figure4.6 : Interface du menu principale.

Cette interface contient plusieurs fonctionnalités représentées par les boutons suivants :

- 1- le bouton choisir : il permet de récupérer un fichier de n'importe quel emplacement sur le disque ;
- 2- le bouton  permet l'enregistrement manuel ;
- 3- le bouton traitement : il permet de lancer le traitement d'un son choisi ;
- 4- le bouton  sert à écouter le son du test ;
- 5- le bouton  sert à actualiser la fenêtre pour lancer un nouveau test ;
- 6- le bouton afficher le son : permet d'afficher le graphe qui représente le son.

### 4.5.1.3. Traitement

La barre du traitement nous indique que le traitement du son enregistré est en cours d'exécution pour afficher la décision finale, et cela peut prendre quelques secondes.

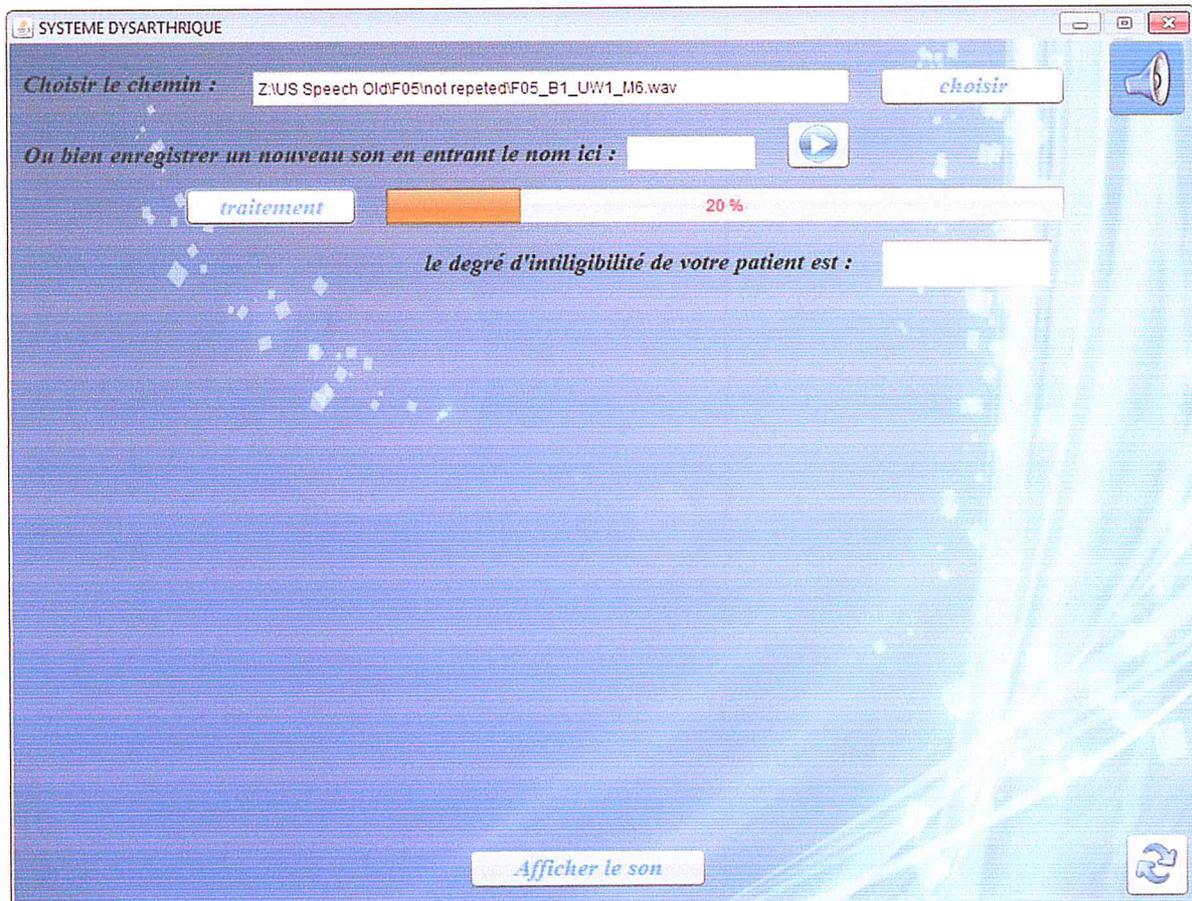


Figure 4.7 : Lancement du traitement d'un son.

### 4.5.1.4. Résultats

Lorsque le traitement se termine, si :

- ✓ un patient à un degré d'intelligibilité élevé le résultat est affiché en vert (Figure 4.8) ;
- ✓ s'il a un à un degré d'intelligibilité très faible le résultat est affiché en rouge (Figure 4.9).

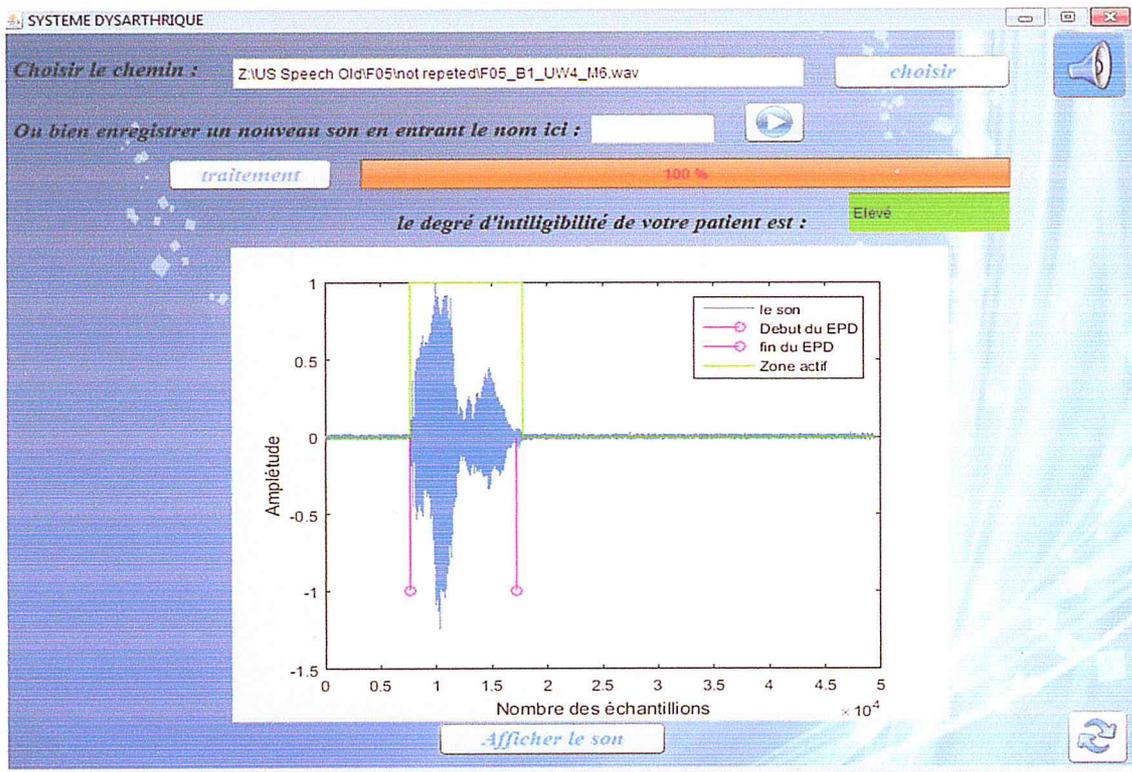


Figure4.8 : Un exemple d'un patient avec degré d'intelligibilité élevé.

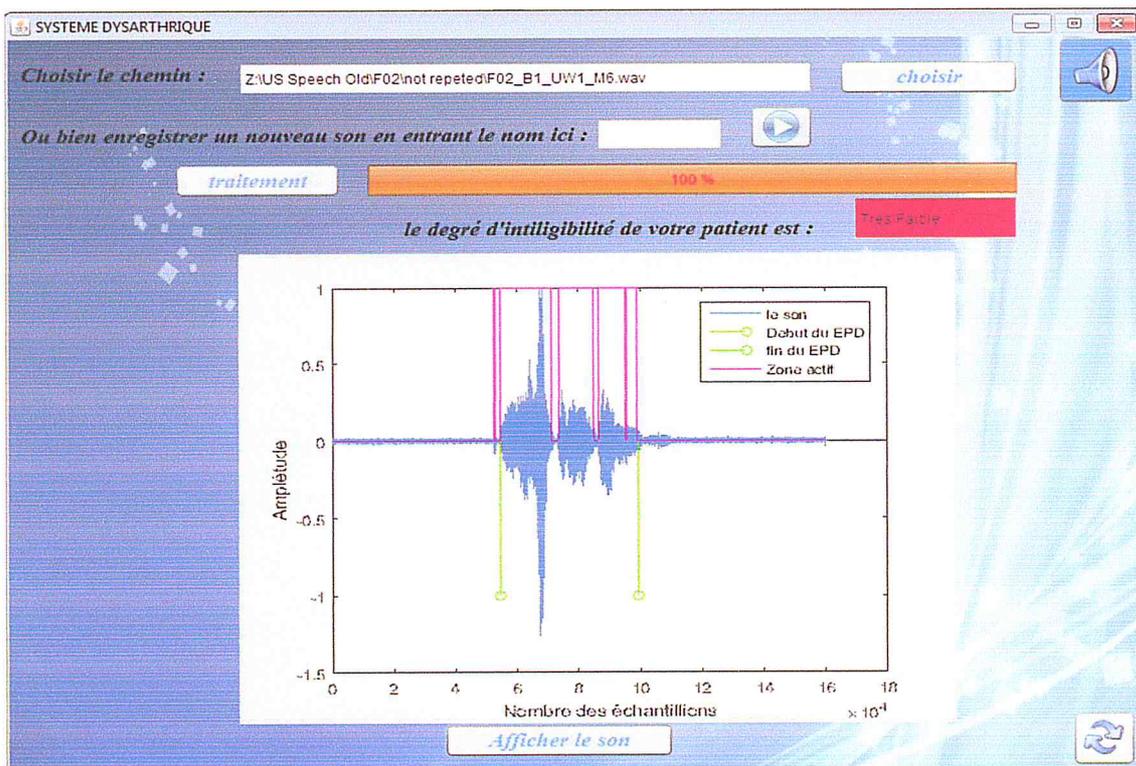


Figure4.9 : Un exemple d'un patient avec degré d'intelligibilité très faible.

#### **4.6. Conclusion**

Dans ce chapitre nous avons mis en fonctionnement notre système, le choix du résultat final était révélé après une certaine variation du nombre de gaussiennes. Le nombre de gaussiennes qui a donné le taux de bonne classification le plus élevé a été choisis. Pour permettre l'interaction entre le système et le décideur nous avons développé une application en utilisant le langage java.

---

---

***CONCLUSION GENERALE***

---

---

## Conclusion Générale

---

Le travail que nous avons présenté s'inscrit dans le cadre de l'apprentissage supervisé et s'intéresse essentiellement au problème d'extraction d'attributs robustes en vue de la mesure du degré d'intelligibilité de la parole dysarthrique.

Notre objectif était d'estimer le degré d'intelligibilité à partir de l'onde acoustique de la parole. Dans un premier temps nous avons extrait les attributs à partir des sons de la base de données, ensuite nous avons appliqué le model de gaussien sur ces attributs et enfin nous avons effectué une classification des sons selon leurs degrés d'intelligibilité c.-à-d. un son appartient soit à la classe « très faible » ou bien à la classe « élevée ».

La méthode proposée est composée de deux parties : (i) l'extraction d'attributs et (ii) la classification. Pour l'extraction d'attributs, nous avons exploité quatre différents groupes : (1) l'entropie des coefficients en sous bandes selon l'échelle audiométrique, (2) les coefficients Cepstraux de MEL, (3) le taux de silence à l'intérieur d'un mot isolé et (4) le niveau vocal actif. L'étape de classification utilise les mélanges de gaussiens. Deux niveaux d'intelligibilité ont été considérés dans notre application : niveau très faible et niveau élevé.

Pour les futures perspectives nous visons :

- L'estimation d'autres niveaux intelligibilité (moyen et faible).
- La comparaison avec d'autres méthodes qui ont été proposées dans la littérature.
- La transformation de notre système sous format d'une application androïde afin de permettre la détection des premiers signes de cette maladie.

---

---

## *Références*

---

---

## Références

- [1]. Ooreka. [En ligne] <http://orthophonie.ooreka.fr/comprendre/dysarthrie>.(accédé le 12-2015).
- [2]. Liss, **Kaitlin L Lansford et Julie M.** *Vowel Acoustics in Dysarthria: Speech Disorder Diagnosis and Classification* , NIH Public Access. Journal of Speech Language and Hearing Research; vol.57, pp.57-67. Arizona, 2014.
- [3]. **Moutarde, Fabien.** Apprentissage Artificiel. CAOR, Mines Paris-Tech, Mai 2011.
- [4]. **HORDÉ, Pierrick.** Sante-medecine. *Journaldesfemmes*. [En ligne] Juin 2014. <http://sante-medecine.journaldesfemmes.com/faq/8685-dysarthrie-definition>. (accédé le 01-2016).
- [5]. Bioam. [En ligne] <http://www.bioam.fr/encyclopedie-medicale/dysarthrie-causes-symptomes-traitements>.(accédé le 01-2016).
- [6].**Antoine-Flavien Eger, Christophe Gaudet-Blavignac, Arthur Hammer.** *Maladie de Parkinson*Antoine. Rapport interne, Université de Genève, 2009.
- [7]. **Patrick Aebischer and Ann Kato .** *Playing defense against Lou gehrig's disease.* USA , raport de recherche: SCIENTIFIC AME RICAN: Vol.297,pp.86 - 93, 2007.
- [8]. *La maladie de Huntington. : Encyclopédie Orphanet Grand public.* [En ligne] <https://www.orpha.net/data/patho/Pub/fr/Huntington-FRfrPub118.pdf> , 2006.
- [9]. **Hummel, Richard.** Objective Estimation of Dysarthric Speech Intelligibility. Master thesis, Kingston, Ontario, Canada , 2011.
- [10]. **Julie, CLAUDE Charlotte et MIQUEL.** *Intelligibilité de sujets dysarthrique chroniques.* Mémoire: Certificat de Capacité d'orthophoniste, Université Claude Bernard Lyon1, 2012.
- [11]. **Teston, Bernard.** Evaluation objectives des dysarthries. *méthodes acoustiques et aérodynamique.Rapport interne*, Université de Provence, 2007.
- [12]. **Yorkston, K.Miller.** Speech and swallowing disorders in degenerative disease. *HHS Public Access*, Journal literature et the U.S. National Institutes of Health's National Library of Medicine san Antonio, vol.24, pp.115–126.2008
- [13]. **KOHLER, F.** *Systeme d'aide à la décision en medecine.* Cours en informatique. Université de Paris 5. 2013.

- [14]. **Tanja Kocjancic Antolik, Cécile Fougéon.** *Consonant distortions in dysarthria due to Parkinson's disease, Amyotrophic Lateral Sclerosis and Cerebellar Ataxia.* In Proc. of InterSpeech, 2013.
- [15]. **J. Carmichael and P. Green.** Revisiting dysarthria assessment intelligibility metrics, In Proc. of InterSpeech, 2004.
- [16]. **G. Van Nuelen and M. De Bodt.** Automated intelligibility assessment of pathological speech using phonological features, EURASIP Journal on Advances in Signal Processing, 2009.
- [17]. **Tiago H. Falk, Wai-Yip Chan et Fraser Shein** Characterization of atypical vocal source excitation, *temporal dynamics and prosody for objective measurement of dysarthric word intelligibility*; Speech Communication, Vol. 54, pp. 622-631, 2012.
- [18]. **Kim, Myung Jong Kim and Hoirin.** Automatic Assessment of Dysarthric Speech Intelligibility Based on Selected Phonetic Quality Features. LNCS: Computers Helping People with Special Needs, pp. 447-450, 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012.
- [19]. **K.L. Kadi, S-A. Selouani, B. Boudraa, and M. Boudraa.** Discriminative Prosodic Features to Assess the Dysarthria Severity Levels. *Proceedings of the World Congress on Engineering.* London, U.K., Vol. 3, pp. 3-5, 2013.
- [20]. **Xavier Menéndez-Pidal, James B. Polikoff, Shirley M. Peters, Jennie E. Leonzio, H. T. Bunnell.** The Nemours Database of Dysarthric Speech; In the Proc. of the Fourth International Conference on Spoken Language (ICSLP 96), 1996.
- [21]. **Frank Rudzicz, Aravind Kumar Namasivayam, Talya Wolff.** The TORGO database of acoustic and articulatory speech from speakers with dysarthria. ; Language Resources and Evaluation, Vol. 46, Issue 4, pp. 523-541, December 2012.
- [22]. **Heejin Kim, Mark Hasegawa-Johnson, Adrienne Perlman, Jon Gunderson, Thomas Huang, Kenneth Watkin, Simone Frame,** Dysarthric Speech Database for Universal Access Research; In Proc of InterSpeech 2008.
- [23]. **Heejin Kim.** *Isle. Illinois.* [En ligne] <http://www.isle.illinois.edu/sst/data/UASpeech> (accédé le 05-2016).
- [24]. **Kinfe T. Mengistu, Frank Rudzicz, Tiago H. Falk:** Using Acoustic Measures to Predict Automatic Speech Recognition Performance for Dysarthric Speakers, In Proc. of the 7th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA'11), 2011.

- [25]. **L. R. Rabiner and M. R. Sambur**, An Algorithm for Determining the Endpoints of Isolated Utterances, Vol. 54, No. 2, pp. 297-315, February 1975.
- [26]. **F. Ykhlef et L.Bendaouia et W. Benzaba** Region-based amplification system of speech for listeners with hearing impairments, IEEE System Man and Cybernetics. San Diego, California, USA : October, 2014.
- [27]. Mthworks.[En ligne]<http://www.mathworks.com/help/wavelet/ref/dwt.html>(accédé le 03-2016).
- [28]. Outilsrecherche.[En ligne][http://outilsrecherche.overblog.com/pages/Presentations\\_3\\_Echelle\\_de\\_Mel-3122565.html](http://outilsrecherche.overblog.com/pages/Presentations_3_Echelle_de_Mel-3122565.html). (accédé le 03-2016).
- [29]. Nmsea.[En ligne] [http://www.nmsea.org/Curriculum/Primer/what\\_is\\_entropy.html](http://www.nmsea.org/Curriculum/Primer/what_is_entropy.html). (accédé le 04-2016).
- [30]. **Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi**; *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*. Journal Of Computing, vol.2, pp.2151-9617, March2010.
- [31]. **DRIVER., BRUCE K.** Hilbert Spaces Hilbert Spaces, Integration by Parts and Quasi-Invariance for Heat Kernel Measures on Loop Groups, journal of functional, vol.149, 1997.
- [32]. **Frédéric Hélein**; Fonctions holomorphes Résumé du cours, Université Paris Diderot - Paris 7, 2009.
- [33]. **Helsinki**, Union Internationale des Télécommunications Secteur de la Normalisation (03/93) des Télécommunications de L'UIT ; Mars 1993.
- [34]. Site web de Dell Statistica. *Normalisation des données*. [En ligne] <http://www.statsoft.fr/concepts-statistiques/glossaire/n/normalisation.html>(accédé le 5-2016).
- [35]. **Bin Mohamad, Ismail et Usman, Dauda**, *Standardization and Its Effects on K-Means Clustering Algorithm*. Research Journal of Applied Sciences, Engineering and Technology, 2013.
- [36]. **R.Abdelaziz et Z.Nab**, *Système d'aide à la décision pour le diagnostic de la maladie de parkinson à partir de la voix*. Mémoire de Master.. Université Saad Dahleb Blida, 2015.

- [37]. **Juang, L. R. .Rabiner B. H.** An Introduction to Hidden Markov Models, IEEE ASSP MAGAZINE, 1986.
- [38]. **Drukker, David M. L. R. .Rabiner B. H. Juang.** *An introduction to GMM estimation using Stata StataCorp* . Berlin: June 2010.
- [39]. **Weston, Jason.** Support Vector Machine (and Statistical Learning Theory) . Tutorial, University of Colombia, USA, 2012.
- [40]. **M.T. Hagan, H.B. Demuth, M. Beale,** Neural Network Design, PWS Publishing Compagny, 1995.
- [41]. **Reynolds, Douglas.** Gaussian Mixture Models. Encyclopedia of Biometrics, pp 659-663, 2009.
- [42]. **Santos, Frédéric.** L'algorithme EM ; une courte présentation au CNRS, Aout 2015.
- [43]. **Payam Refaeilzadeh, Lei Tang, Huan Liu.** Cross-Validation, Encyclopedia of Database Systems, pp. 532-538, 2009.
- [44]. **HENNIGE, PHILIPPE.** Apprentissage Supervisé par Reseaux de Neurones. Mémoire de Master, Ecole de technologie supérieure, Université du Quebec, 2006.
- [45]. **Ken Arnold, James Gosling et David Holmes,** *Le Langage Java*, Vuibert, 2001. 2001.
- [46]. netbeans [En ligne]<http://netbeans.org/community/articles/interviews/yarda-tulach.html>. (accédé le 05-2016).
- [47]. **Rene Alt,** Transformation en Ondelettes, Cours. Université Pierre et Marie Curie 4, Paris 2006.
- [48]. **Adhemar Bultheel.** Learning to swim in a sea of wavelets, Bull. Belg. Math. Soc. Vol.2, pp1-46, 1995.
- [49]. **Sahidullah, Md et Saha, Goutam.** Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. Speech Communication, Vol. 54, Issue 4, pp. 543-565,2012..

---

---

*Annexes*

---

---

## Annexe A: Ondelette

- **Naissance des ondelettes [47]**

- Morlet en 1983 utilise la transformée de Fourier à fenêtre glissante (c'est à dire nulle en dehors d'un intervalle donné) pour l'analyse de signaux sismiques ;
- Méthode de Fourier insuffisante ;
- Nécessité d'analyser le signal en temps et en fréquence ;
- Utilisation d'une fenêtre dont la longueur est dilatée ou contractée ;
- → Naissance de l'idée des ondelettes.

- **Qu'est-ce qu'une ondelette ? [47]**

- Une petite onde (ou vague) qui a un début et une fin
- Utilisation : Représenter une fonction (ou un signal) comme une somme pondérée de ces petites ondes translatées ou dilatées.

La transformée en ondelettes est similaire à la transformée de Fourier. La différence principale est que la transformée de Fourier décompose le signal en sinus et en cosinus, c'est-à-dire en fonctions localisées dans l'espace de Fourier ; contrairement à la transformée en ondelettes qui utilise des fonctions localisées à la fois dans l'espace réel et dans l'espace de Fourier. De manière générale, la transformée en ondelettes peut être exprimée avec l'équation suivante :

$$F(a, b) = \int_{-\infty}^{\infty} f(x) \varphi_{(a,b)}^*(x) dx$$

où le symbole \* désigne le conjugué complexe et  $\varphi$  est une fonction donnée. Cette fonction peut être choisie arbitrairement à condition qu'elle obéisse à certaines règles.

Comme nous pouvons le voir, la transformée en ondelettes est en fait un ensemble infini de diverses transformées. C'est la raison pour laquelle nous pouvons entendre parler de la « transformée en ondelettes » dans des contextes et applications très divers. Il y a deux transformées en ondelettes qui possèdent les propriétés suivantes :

1. La transformée en ondelettes discrète retourne un vecteur de données de même longueur que les données d'entrée. Généralement la plupart des valeurs de ce vecteur

sont pratiquement nulles. Cela vient du fait qu'elle décompose en un ensemble d'ondelettes (fonctions) orthogonales par translation et par homothétie. On décompose ainsi le signal en un nombre égal ou inférieur de coefficients du spectre d'ondelettes que le nombre de points de données du signal. Un tel spectre d'ondelettes est très bon pour le traitement du signal et la compression, par exemple, car nous n'obtenons ici aucune information redondante.

2. La transformée en ondelettes continue retourne quant à elle un vecteur de dimension supérieure de un à celle des données d'entrée. Pour des données 1D nous obtenons une image du plan temps-fréquence. On peut ainsi aisément visualiser l'évolution fréquentielle pendant la durée du signal et comparer son spectre avec ceux d'autres signaux. Comme on utilise ici un ensemble d'ondelettes non-orthogonales, les données sont hautement corrélées, on a donc dans ce cas une redondance élevée. Ceci permet de voir le résultat sous une forme plus « humaine » [48].

## *Annexe(B) : MFCC*

Cette méthode d'extraction de caractéristiques a été mentionné par Bridle et Brown en 1974 et développé par Mermelstein en 1976.

La méthode MFCC a pour objectif de trouver une représentation fréquentielle d'échelle MEL (une échelle de fréquences basée sur la perception humaine. Elle se mesure en *mels.*) Mel-fréquence coefficients cepstraux (MFCC) sont des coefficients qui constituent collectivement un MFC. Ils sont dérivés d'un type de cepstral de représentation du clip audio (un «spectre-of-a-spectre" non linéaire). La différence entre le cepstre (une transformation d'un signal du domaine temporel vers un autre domaine analogue, cest le résultat de la transformée de Fourier appliquée au logarithme naturel de la transformée de Fourier du signal dont la phase est ignorée) et le cepstre de mel-fréquence est que : dans le MFC, les bandes de fréquence sont équidistantes ( distance égale ) sur l'échelle mel, qui se rapproche plus étroitement la réponse du système auditif humain que les bandes de fréquences linéairement espacées utilisées dans le cepstre normal.

Cette déformation de fréquence peut permettre une meilleure représentation du son, par exemple, dans la compression audio [49].