Introduction

Générale

Chapitre I

Généralités

Chapitre II

Pathologie du langage parlé

Chapitre III

Extraction des caractéristiques

Chapitre IV

Les algorithmes de classification

Chapitre V

Simulation

&

Expériences

Conclusion

Générale

Références

&

Bibliographie

ANNEXE

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne démocratique et populaire

وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

جامعة سعد دحلب البليدة
Université SAAD DAHLAB de BLIDA

كلية التكنولوجيا
Faculté de Technologie

قسم الإلكترونيك
Département d'Électronique

Mémoire de Projet de Fin d'Études

présenté par

Ounnoughi Imane

&

Haddad Affaf

Pour l'obtention du diplôme de Master II en Électronique option signaux en ingénierie des systèmes et informatique industrielles(SISII)

Thème

Classification des sons pathologiques utilisant l'ELM

Dirigé par : Mr Benselama & Mr Bencherif

Année Universitaire 2013-2014

ملخص

مشروعنا يندرج في مجال التعرف على الكلام وعلى نحو أدق تصنيف الأصوات المرضية، لذلك عملنا يكمن في تطبيق خوارزمية الELM على برنامج الMATLAB أولا ثم مقارنتها مع خوارزمية ال SVM وشجرة القرارات على برنامج ال Weka. لذالك استعملنا مستخرج الخصائص Open Smile .

كلمات المفاتيح:

تصنيف الكلام, التعرف على الكلام, الأصوات المرضية, ELM, مستخرج الخصائص.

Résumé

Notre projet s'inscrit dans le domaine de la reconnaissance de la parole et plus précisément la classification des sons pathologiques, pour cela, notre travail consiste à implémenter l'ELM en premier temps en Matlab et le comparer avec SVM et l'arbre de décision sous Weka. Nous utilisons dans ce cas l'extracteur de caractéristiques Open Smile.

Mots clés :

Classification de la parole, reconnaissance de la parole, sons pathologiques, ELM, extracteur de caractéristiques.

Abstract

Our project falls under the field of speech recognition and more precisely the classification of pathological sounds, for that, our work consist to implement the algorithm ELM first time in Matlab and compare it with SVM and decision tree in Weka. We use in this case the features extractor Open Smile.

Keywords:

Classification of speech, speech recognition, pathological sounds, ELM, features extractor.

Remerciements

*Nous remercions notre dieu **ALLAH** qui est toujours présent avec nous dans le meilleur et dans le pire.*

*Nous tenons à manifester nos sincères remerciements à nos directeurs de mémoire **Monsieur Benslama** et **Monsieur Bencherif** de nous avoir donnés la chance d'enrichir nos connaissances et d'acquérir plus d'expérience en ce domaine de recherche. Leurs conseils ont été très précieux pour mener à bien ce travail. L'expérience d'avoir travaillé avec eux nous a été très enrichissante.*

*Nous exprimons également nos remerciements à **Madame Akak** pour ses remarques et son encouragement.*

Nos remerciements sincères iront à Messieurs les membres du jury d'avoir bien voulu examiner et évaluer notre modeste contribution dans ce travail.

Pour finir, un grand merci à tous ceux qui nous ont aidés et encouragés.

Je dédie ce travail :

A mes très chers parents Saad et Chahla,.

A mes frères Kamel, Hichem, Fouzi et Marouane.

A mes soeurs Nassima, Nabila et Djamila.

A mon beau frère Boualèm et mes belles soeurs Saida et Hassiba

A mes neveux Yacine, Houcem et Mohamed

A mes nièces Inès, Aya, Chahrazed,

Maria, Malek et Nour elhayat.

A ma chère Imane, mon binôme

Ainsi qu'à tous mes amis

Spécialement

Katmir F.zohra, Djellab Yahia et Seddik Khadidja ....

Affaf

Je dédie ce travail :

A mon père Mahrez qui est mon premier maitre,

A ma mère Sabah : ma fierté,

Ma soeur Hadjer : mes souvenirs

Et à mon petit frère Zakaria qui est mon espoir

A mon marie Mohamed mon soutien et A ma belle famille

A ma grande famille mes oncles, mes tantes, mes cousins, mes cousines et bien sûr ma chère grand-mère

A mon binôme Affaf pour leurs efforts et sa patience,

A tous mes amis, mes collègues de travaille et tous mes professeurs.

"Avec tous mes condoléances à ma famille pour tous ses membres qui nous ont quitté, mon beau père, ma tante SAMIA et mon oncle EHMED récemment (ce moi), en laissant un grand vide . Qu'ils ont assisté à ma soutenance d'ingéniorat et qu'ils ont été fiers de moi et malheureusement qu'ils seront absents cette fois ci malgré qu'ils ont été parmi la liste des invités. "

Imane

Liste des figures

Liste des tableaux

Sommaire

Chapitre I : Généralités

Chapitre II : Pathologie du langage parlé

Chapitre III : L'extraction des caractéristiques

Chapitre IV : Algorithmes de classification

Chapitre V : Implémentation et résultats

Conclusion Générale

Il reste incontestablement que le domaine de reconnaissance de la parole est du point de vue recherche l'un des plus riche, les multiples et diverses simulations du fonctionnement du cerveau humain, démontré par la technologie neuronale permettent d'espérer beaucoup.

Ainsi, notre travail a consisté à mettre un moyen de classification des sons pathologiques a base de l'ELM (l'Extrême Learning Machine), en utilisant une extraction des caractéristiques du signal vocal qui repose sur « Open Smile ».

Ces diverses expériences auxquelles nous nous étions astreintes, nous ont permis d'obtenir les taux de reconnaissance, acceptables et peuvent être considérer comme satisfaisants. Cependant le taux de reconnaissance peut être amélioré en enrichissant la base de données et d'améliorer les conditions d'enregistrement au niveau du laboratoire.

Nous nous permettons donc de donner quelques perspectives auxquelles ce travail peut s'orienter vers la création d'un système de correction de certaines pathologies envers les personnes qui souffrent de problèmes de prononciation, ça sera certes un grand événement pour cette catégorie de personnes qui verront leur handicap surmonté et leurs difficultés aplanies, c'est d'ailleurs l'objectif attendu dans notre étude.

I.1 Introduction

La parole est la capacité de l'être humain de communiquer la pensée par l'intermédiaire desons articulés. Dû à son importance, la parole a préoccupé depuistoujours les scientifiques.Ainsi quelques-unes des sciences qui se préoccupent de l'étude de la parole ont déjà descentaines d'années.

Dans ce chapitre on va voir une approche générale sur l'identité du signal de parolecommençant par tout ce qui concerne sa production y compris l'architecture de l'appareilvocal, le mécanisme de la phonation et la classification des phonèmes seront aussi abordédans le contenu de ce chapitre. Nous allons aussi voir le niveau acoustique du signal deparole particulièrement son spectre sa fréquence fondamentale son énergie….etc.

I.2 Production de la parole

I.2.1 Architecture de l'appareil vocal

L'appareil vocal, ou système phonatoire, comprend quatre éléments fondamentauxfonctionnant en étroite synergie pour produire des signaux acoustiques. Ce sont, dans l'ordreoù ils s'élaborent :

- La soufflerie
- Le vibrateur
- Le corps sonore
- Le système articulateur

La soufflerieest constituée d'un réservoir d'air, les poumons, actionnés par les muscles du thorax et de l'abdomen, et d'un tube, la trachée artère, qui conduit l'air aux cordes vocales ; le vibrateur est le larynx, quiengendre les ondes aériennes; le corps sonore est constitué d'un ensemble complexe de résonateurs, dont le pharynx et la bouche sont les principaux;le système articulateur, enfin, se compose d'éléments fixes et mobiles permettant de modifier largement la forme de l'onde laryngée. Tous ces éléments sont placés sous ladépendance étroite du système nerveux central, qui en assure le synchronisme et la coordination [1].

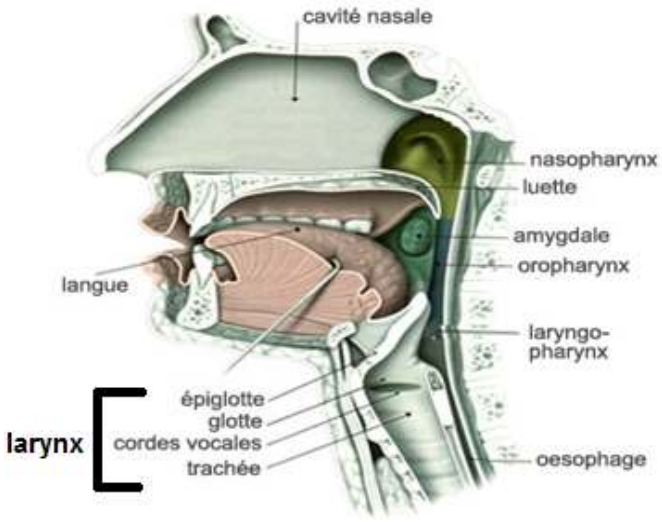


Figure I.1 : L'appareil vocal.

I.2.1.1 La soufflerie et Le vibrateur

L'air est la matière première de la voix. Si le fonctionnement de notre appareil vocal estsouvent comparé à celui d'un instrument de musique, il doit être décrit comme celui d'uninstrument à vent. En effet, en expulsant l'air pulmonaire à travers la trachée, le systèmerespiratoire joue le rôle d'une soufflerie. Il s'agit du « souffle phonatoire » produit, soit parl'abaissement de la cage thoracique, soit dans le cadre de la projection vocale par l'action des muscles abdominaux [2].

I.2.1.2 Le larynx

Le larynx constitue l'extrémité supérieure de la trachée artère, situé à la hauteur de lasixième vertèbre cervicale (chez l'adulte). C'est un assemblage de cartilages articulés, reliésentre eux par des ligaments et des muscles (dont les cordes vocales), l'ensemble étant tapisséd'une muqueuse [1].

I.2.1.3 Le corps sonore

Les résonateurs du système phonatoire sont, pour l'essentiel, responsables du timbre de lavoix. Leur originalité par rapport aux caisses de résonance des instruments de musiquetraditionnels est leur faculté de changer, grâce à un réseau musculaire dense et élaboré, —dans de larges proportions, et très rapidement — de forme et de volume, assurant ainsi au sonvocal une variété acoustique sans équivalent.

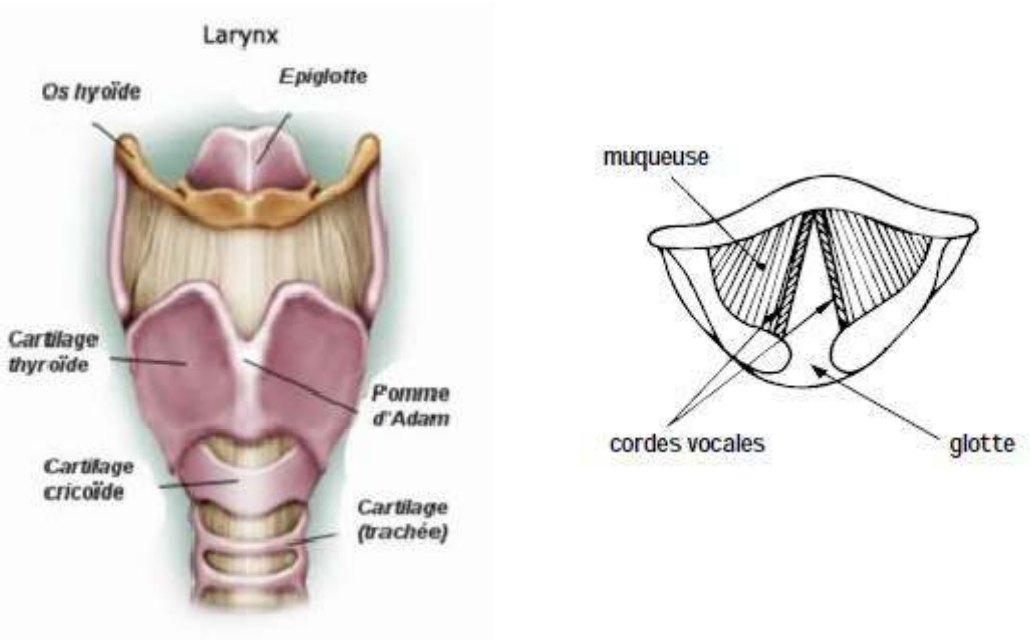


Figure I.2 : Vue schématique antérieure du larynx (Gauche) et sa section, vue de haut(droite).

Les résonateurs sont au nombre de cinq : le pharynx, la cavité buccale divisée en deux, lacavité labiale, et la cavité nasale. Tous communiquent entre eux par des ouvertures de taillesréglables. Tapissés de muqueuse, ils sont peu amortis.

Les sinus de la face, contrairement à une croyance largement répandue, sont trop petits pourse comporter en résonateurs, et n'ont d'ailleurs aucune fonction reconnue dans la phonation[1].

I.2.1.4 Le système articulateur

Il intègre un ensemble d'organes mobiles, le voile du palais, la mâchoire inférieure (oumandibule), la langue et les lèvres. Les mouvements de la mâchoire inférieure contribuentlargement aux variations de volume de la bouche. La langue, reliée par sa base à l'os hyoïde,est extrêmement mobile, car commandée par dix-sept muscles (huit pairs et un impair) ; elleprend appui sur différents points du conduit pharyngo-buccal pour articuler les phonèmes. Cespoints d'articulation sont [1] :

- les lèvres (articulations labiales ou bilabiales)
- les dents (articulations dentales)
- les alvéoles16 (articulations alvéolaires)
- le palais dur, ou partie osseuse de la voûte (articulations palatales)
- le voile du palais ou « palais mou » (articulations vélaires)

- la luette (articulations uvulaires)
- le pharynx (articulations pharyngales)
- la glotte (articulations glottales)

I.2.2 Mécanisme de la phonation

L'une de plus importantes caractéristiques du signal vocal est la nature de l'excitation. Ilexiste deux types élémentaires d'excitation qui produisent les sons voisés et non voisés.

I.2.2.1 Phonation de sons voisés

Les sons voisés sont produits à partir d'une excitation qui actionne sur le conduit vocal etqui consiste en une suite d'impulsions périodiques d'air fournies par le larynx. Les cordesvocales au début sont fermées. Sous la pression continue de 1'air qui vient des poumons elless'ouvrent graduellement délivrant cette énergie potentielle. Pendant cette ouverture la vitessede l'air et l'énergie cinétique augmentent jusqu'à ce que la tension élastique des cordes vocaleségale la force de séparation du courant d'air. A ce point l'ouverture de la glotte est maximale.L'énergie cinétique qui a été accumulée comme tension élastique dans les cordes vocalescommence à rétrécir cette ouverture et de plus la force de Bernoulli accélère encore lafermeture abrupte de la glotte [3]. Ce processus périodique est caractérisé par une fréquencepropre à chaque personne, connue sous le nom de fréquence du fondamental (Fo) où pitch et ildonne la hauteur normale de la voix. La fréquence fondamentale peut varier de 80 à 200 Hzpour une voix masculine, de 150 à 450 Hz pour une voix féminine et de 200 à 600 Hz pourune voix d'enfant [4].

Cette fréquence fondamentale peut varier suite à des facteurs liés au stress, intonation etémotions. Le timbre de la voix est déterminé par les amplitudes relatives des harmoniques dufondamental.

L'intensité du son émis est liée à la pression de l'air en amont de larynx. Tous ces aspectspour un son voisé peuvent être observés dans la figure I.3.

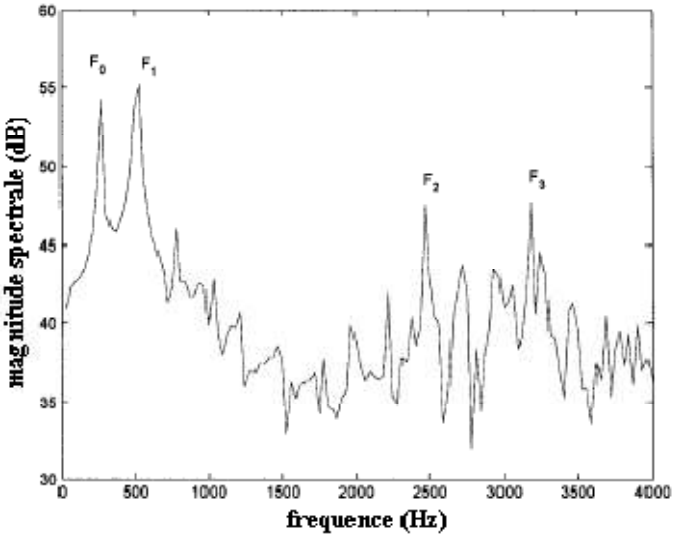


Figure I.3 : Le spectre du son voisé / i /

I.2.2.1 Phonation de sons non voisés

Les sons non voisés sont générés par le passage de l'air dans une constriction étroite situéen un point du conduit vocal. Ils sont générés sans l'apport du larynx et ne présentent pas destructure périodique [3]. Ces caractéristiques d'un son non voisé peuvent être observées dans la figure I.4.

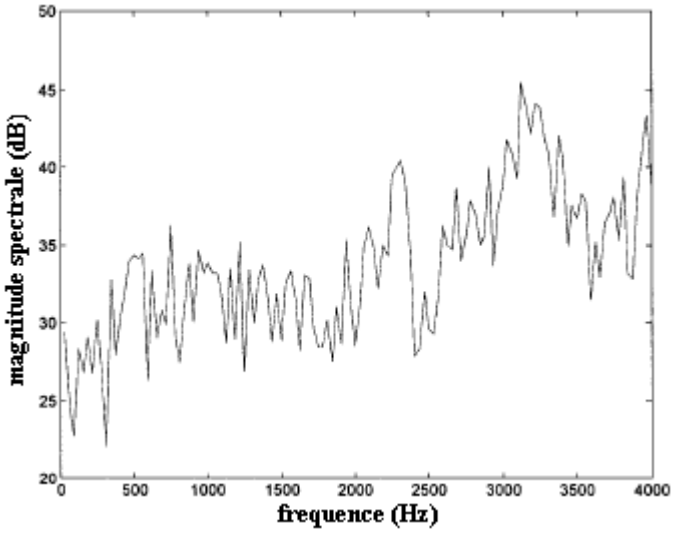


Figure I.4 : Le spectre du son non voisé / f /

I.2.3 Classification des phonèmes

Il y a plusieurs façons de classifier les phonèmes. Un phonème est stationnaire oucontinuant si la configuration du conduit vocal ne change pas pendant la production du son.Un phonème est non continuant si pendant sa production il y a des changements dans laconfiguration du conduit vocal [3].

On peut grouper les 36 phonèmes de la langue française en classes et sous classes d'aprèsle mode d'articulation de l'appareil de phonation (tableau I.1).

I.2.3.1 Les voyelles

Les voyelles sont des sons voisés, continus, normalement avec la plus grande amplitudeparmi tous les phonèmes et elles peuvent varier beaucoup en durée, entre 40 et 400 ms.

Les voyelles orales sont produites sans l'intervention de la cavité nasale pendant que pourles voyelles nasales, le conduit nasal est couplé à la cavité buccale et la production de son sefait par la bouche et par les narines en même temps. Les voyelles sont différentiées en troisgroupes d'après la position de la courbure de la langue et le dégrée de la constriction induitdans le conduit vocal.

Différentiées en trois groupes d'après la position de la courbure de la langue et le dégrée deconstriction induit dans le conduit vocal.

L'analyse dans le domaine temporel et fréquentiel révèle plusieurs caractéristiquesacoustiques qui aident à la classification de chaque son. L'analyse dans le domaine temporelmontre que les voyelles sont de sons quasi périodiques dus à l'excitation.

Les voyelles peuvent être identifiées par les locations de leurs fermants dans le domainefréquentiel. La position des deux premiers fermants est suffisante pour caractériser la majoritédes voyelles, le troisième formant est nécessaire juste pour quelques-uns. La position defermants de fréquence plus élevée reste presque inchangée et n'apporte pas d'information utilepour l'identification.

Phonèmes								
Voyelles		Semi-Consonnes	Consonnes					
					Fricatives		Occlusives	
Orales	Nasales		Liquides	Nasales	Voisées	Non-Voisées	Voisées	Non-Voisées
i(I) e(E) ɛ(AI) a(A) ɔ(O) u(OU) y(U) Ø(EU) œ(OE) ə(E) o(AU)	ɛ̃ (IN) œ̃(UN) ã (AN) ɔ̃(ON)	j (Y) w(W) ɥ(UI)	l(L) R(R)	m(M) n(N) ɲ(GN)	v(V) z(Z) ʒ(J)	f(F) s(S) ʃ (CH)	b(B) d(D) g(G)	p(P) t(T) k(K)

Tableau I.1 : Phonèmes de la langue française.

I.2.3.2 Les diphtongues

Les diphtongues impliquent un mouvement d'une voyelle initiale vers une autre voyellefinale. Donc les diphtongues sont essentiellement des sons non continus. La différence entreune diphtongue et les deux voyelles individuelles composantes est que la durée de latransition est plus grande que la duré de chaque voyelle. De plus la voyelle initiale est pluslongue que la voyelle finale. Dans la parole les deux voyelles composant une diphtonguepeuvent ne pas être réalisées entièrement ce qui accentue l'idée de non-stationnarité quicaractérise les diphtongues.

I.2.3.3 Les semi-consonnes

Les semi -consonnes sont des sons non continus et voisés qui possèdent descaractéristiques spectrales semblable aux voyelles. On peut voir les semi-consonnes commedes sons transitoires qui s'approchent, atteint et après s'éloignent d'une positionne cible. Ladurée des transitions est comparable à la durée passée en position cible.

I.2.3.4 Les consonnes

Les consonnes sont de sons pour lesquels le conduit vocal est plus étroit pendant laproduction, par rapport aux voyelles. Les consonnes impliquent les deux formes d'excitationpour le conduit vocal et elles peuvent être continuantes ou non.

- **Les consonnes fricatives**

Les fricatives non voisées résultent d'une turbulence créée par le passage de l'air dans uneconstriction du conduit vocal qui peut se trouver près des lèvres pour les labiales, au milieu duconduit vocal pour les dentales et au fond du conduit vocal pour le palatales.

Dans ce cas la constriction cause une source de bruit et aussi divise le conduit vocal en deuxcavités. La première cavité actionne comme une enceinte anti-résonante qui atténue les bassesfréquences d'où la concentration de l'énergie vers les hautes fréquences dans le domainespectral.

Pour les fricatives voisées l'excitation est mixte et à la source de bruit s'ajoutent lesimpulses périodiques crées par la vibration de cordes vocales.

- **Les consonnes occlusives**

Les consonnes occlusives sont des sons non continuants qui sont des combinaisons desons voisés, non voisés et de courtes périodes de silence. Une forte pression d'air s'accumuleavant une occlusion totale dans un point du conduit vocal qui après est relâché brusquement.Cette période d'occlusion s'appelle la phase de tenue.

Pour les occlusifs non voisés la phase de tenue est un silence et la période de friction qui suitest plus longue que pour les occlusives voisés. Pour les occlusives voisés, pendant la phase detenue, un son de basse fréquence est émis par vibration des cordes vocales.

- **Les consonnes nasales**

Les consonnes nasales sont des sons continus et voisés. Les vibrations produites par lescordes vocales excitent le conduit vocal que cette fois est formé de la cavité nasale ouverte etla cavité buccale fermée. Même fermée, la cavité buccale est couplée à la cavité nasale etinfluence la production de sons comme une enceinte anti-résonante qui atténue certainesfréquences, en fonction du point où elle est fermée. Les formes d'onde des consonnes nasalesressemblent aux celles des voyelles mais sont normalement plus faibles en énergie due à lacapacité réduite de la cavité nasale de radier des sons par rapport à la cavité buccale.

➢ **Les consonnes liquides**

Les consonnes liquides sont des sons non continus et voisés qui possèdent descaractéristiques spectrales similaires aux voyelles. Elles sont plus faibles en énergie due aufait que le conduit vocal et plus étroit pendant leur production.

I.3 Modélisation du processus de la perception de la parole

La seconde modélisation consiste naturellement à considérer non plus la productionmais la perception humaine [5]. Il s'agit donc de modéliser lescaractéristiques de l'oreille humaine [6]. Ces modèles sont appelés modèlesauditifs (AuditoryModels) [7], [8], [9].

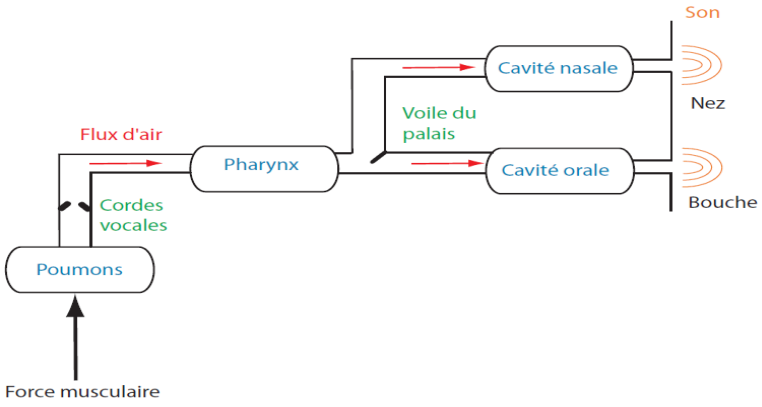


Figure I.5 : Modèle de production de la parole[7], [8], [9].

Depuis de nombreuses années, des études ont été menées pour essayer de "mimer" l'oreille humaine [10]. Cependant, la réplique des différents phénomènes n'améliore pas systématiquement les performances des systèmes de reconnaissance[11].

Les principales étapes de la perception humaine sont représentées sur la partiegauche de la figure I.6. On peut également trouver sur la partie droite de cettefigure la transcription de ces étapes au domaine du traitement du signal ou del'information.

On constate que les opérations de filtrage ou la mise en place d'échelle non linéaire (Mel ou Bark) sont des techniques largement utilisées dans les méthodesde codage traditionnelles. D'autres techniques, basées sur la connaissance du processus de perception, ont été développées pour la conception de modèles auditifs[12].

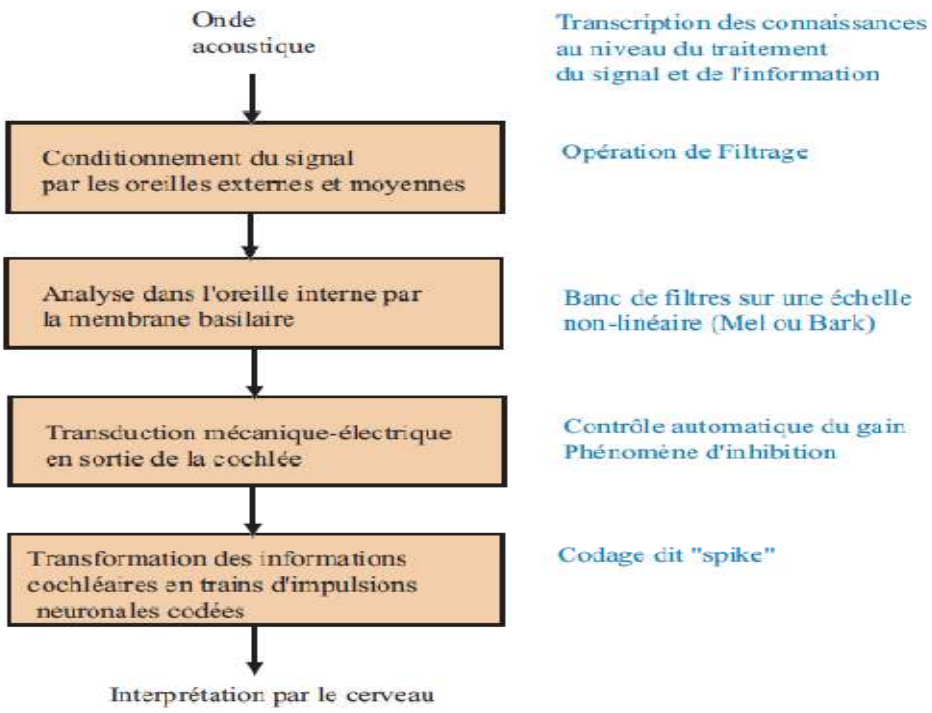


Figure I.6 : Modèle de la perception humaine[12].

Malgré le développement de ces nombreux modèles, il s'avère difficile de trouverun compromis entre la modélisation fidèle du processus de perception humaine etune extraction de caractéristiques efficace. Hermansky[13], qui précise que lacopie intégrale du processus de perception n'est sans doute pas la meilleure solutionpour améliorer les scores de reconnaissance, propose des voies intéressantes pourla conception d'extracteurs de caractéristiques. On peut citer comme exemplessignificatifs :

- Bandes critiques (banc de filtres).
- Echelle Mel (MFCC:Mel Frequency Cepstrum Coefficients) et Bark (PLP :Perceptual Linear Predictive) [14].
- Production de la parole (LPCC :LinearPredictiveCepstralCoding).
- OpenSmile (smileextract).

I.4 Analyse de la parole

L'analyse de la parole permet la mise en forme du signal mais aussi l'extraction de paramètres nécessaires pour les prochaines étapes telles que la reconnaissance.

Un des objectifs de cette analyse est d'obtenir une représentation compacte et informativedu signal. Le signal de parole est un signal redondant et non stationnaire mais il

peut être considéré comme localement stationnaire. L'analyse du signal de parole se fait pendant ces périodes stationnaires dont la durée varie de 10 à 30ms.

Cette durée correspond aussi à la durée de stabilité du modèle de production.

L'analyse de la parole (figure I.7) consiste à effectuer des prétraitements, nécessaires pour la mise en forme du signal, tels que le découpage en fenêtres. La seconde étape de l'analyse est l'extraction de caractéristiques qui est l'objet de ce mémoire.

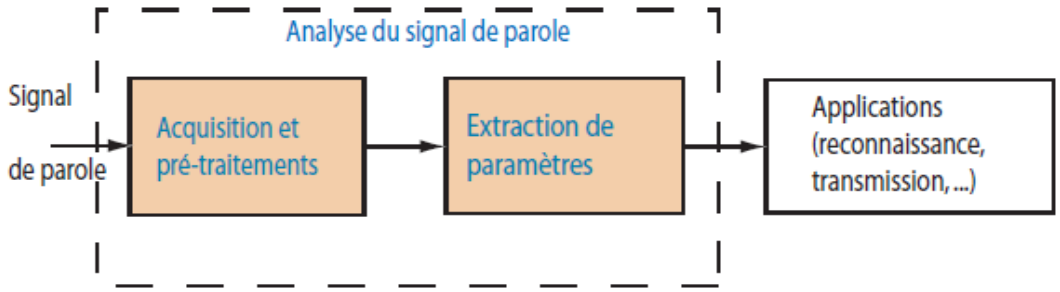


Figure I.7 : Extraction de paramètres dans le cadre du traitement de la parole.

I.4.1 Prétraitements

Les prétraitements débutent par un échantillonnage des signaux (figure I.8), suivit d'une préaccentuation. Le signal s(n) est divisé en fenêtres de longueurN (10-20ms). Lesignal final x(n) est obtenu par une multiplication du signal(n) par une fonction, ou encore fenêtre, de pondération non nulle w(n) :

$$X(n) = s(n)w(n) \qquad (1.1)$$

La préaccentuation est un exemple d'utilisation de connaissances sur la perceptionhumaine. Elle consiste en un filtrage du signal de parole par le filtre suivant :

$$Y(z) = (1 - a\,z^{-1})\,X(z) \qquad (1.2)$$

Le choix de la fenêtre est très important. Parmi les fenêtres utilisées, on peut citer les fenêtres de Hamming, Hanning, Blackman ou de Kaiser. Le choix se faitle plus souvent en fonction de l'application car les fenêtres présentent différentesatténuations à des fréquences bien précises. Cependant, il faut noter que la plupartdes systèmes sont directement conçus sur des fenêtres de Hamming.

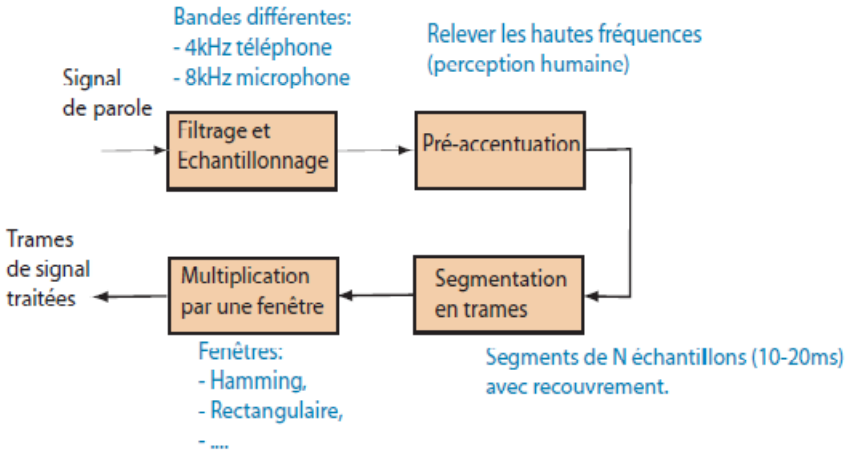


Figure I.8 : Prétraitements généralement réalisés en traitement de la parole.

I.4.2 Extraction de paramètres

L'extraction de paramètres est l'objet principal de l'analyse de la parole et c'est le passage obligé de toutes les applications en traitement de la parole (figure I.9).

Figure I.9 : Extraction de paramètres dans le cadre du traitement de la parole.

L'extraction de caractéristiques, avec ou sans apprentissage, est avant tout larecherche d'une représentation du signal de parole adaptée à l'application. Généralement,le vecteur caractéristique est formé de plusieurs caractéristiques telles que, par exemple :

- Vecteur code : LPCC, MFCC, PLP, SmileExtract, ...
- Paramètres Δ et ΔΔ : dérivées premières et secondes du vecteur code. On cherche à modéliser la dynamique du signal de parole,
- Energie et sa dérivée,
- Le taux de passage par zéros (ZCR : Zero Cross Rate) et sa dérivée,
- La fréquence fondamentale F0 (pitch),
- Le voisement.

Dans la suite, nous nous intéresserons principalement au vecteur code. En effet,une fois le vecteur code extrait, les autres caractéristiques peuvent être ajoutéespour former le vecteur caractéristique.

Le vecteur code est une représentation du signal de parole. Une des classifications possibles des méthodes de représentation des signaux de parole est lasuivante :

- Méthode paramétrique : utilisation d'un modèle (LPC,...). [15]
- Estimation de paramètres à partir de méthodes non-paramétrique(MFCC,...). [16]
- Extracteur de paramètres à grands échelles en temps réel (OpenSmile : sera détaillé dans le chapitre III).

I.5Synthèse de la parole

Dans le cas de la parole le terme «synthèse» implique une combinaison des sons et des bruits et de réalisé des mots ou des phrases.

L'objectif de la synthèse de la parole et de produire des sons de parole a partir d'une représentation phonétique du message. [15]

I.5.1 Les techniques de synthèse de la parole

Pour transmettre une information donnée entre un émetteur et un récepteur il faut utiliser des techniques qui permettent une représentation du signal sous forme réduit (codage) tel que, après transmission (stockage), on peut reconstituer le signal original (décodage) à partir de ces informations figure I.10.

Le principe de synthétiseur est de créer une analogie avec l'appareil phonatoire humain. Les différentes techniques offertes par les synthétiseurs de la parole sont :

- Synthétiseur a canaux.
- Synthèse par formant.
- Synthèse basée sur la prédiction linéaire.

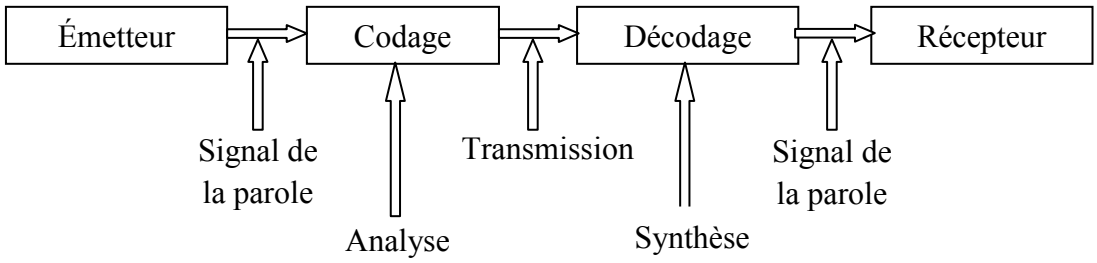


Figure I.10 : Synthèse et codage de la parole.

I.6 Reconnaissance de la parole

La reconnaissance automatique de la parole consiste à extraire l'information lexicale contenue dans un signal de parole (signal électrique obtenue à la sortie d'un microphone et typiquement échantillonné à 8khz dans le cas de ligne téléphoniques ou entre 10 et 16khz dans le cas de saisie par microphone). Bien que ceci soulève également le problème de la compréhension de la parole.

La compréhension de la parole est une étape supplémentaire qui consiste à exiger de la machine une action adéquata au contenu du message vocal [17].

Elle contient plusieurs système peut être classé on :

I.6.1 Reconnaissance de mot isolé

Elle est effectuée normalement au niveau acoustique. Les mots du vocabulaire constituent des références aux quelles un mot à reconnaitre doit être comparé.

Il faut donc définir d'une façon judicieuse une distance ou une mesure de dissemblance entre deux mots ; toute fois il est essentiel que cette distance soit peu sensible aux petites variations relatives des axes des temps. Apres l'alignement optimal des axes temporel, la référence la plus proche désigne le mot à identifier. Pour cela on utilise un algorithme basé sur la programmation dynamique appelé la DTW (Dynamique Time Warping).

I.6.2 Reconnaissance de mots enchainés

Le problème consiste à identifier une suite finie de mots prononcés d'une façon continue; ces mots appartiennent à un vocabulaire connu qui comporte un nombre limité de mots.

La méthode sera donc basée sur une comparaison globale de la suite à reconnaitre avec des suit de référence.

I.6.3 Reconnaissance du locuteur

C'est un terme générique pour discriminer parmi plusieurs personnes en fonction de leur voix (les émotions et les sons pathologiques : ce qui est utilisé dans notre projet) on distingue en général l'identification et la vérification du locuteur.

L'identification consiste à reconnaitre un locuteur appartenant à une population de N locuteur.

La vérification consiste à accepter ou à refuser une identité proclamée par un locuteur ; on compare la distance entre son expression vocale et sa référence [17].

I.7 conclusion

Dans ce chapitre nous avons décrit les principaux éléments caractérisant le signal de la parole. Nous avons aussi abordé l'analyse de la parole ainsi les différents types de reconnaissance de parole ; dans notre travail nous s'intéressons à la reconnaissance des sons pathologiques qui seront explicités dans le chapitre qui suit.

II .1 Introduction

Lorsqu'on parle de pathologies, une référence intuitive nous fait penser à un médecin, c'est un réflexe très logique, toutefois le domaine pathologie fait intervenir différents scientifiques, par exemple l'utilisation de l'effet doppler en médecine, l'utilisation, en imagerie médicale, du scanner en 3D ainsi que des rayons X. Ces appareils sont le fruit de la technologie qui implique d'autres disciplines telle l'électronique, la mécanique, la physique, etc.

Le traitement des pathologies langagières se situe essentiellement à détecter la zone à traiter, mesurer l'ampleur de la maladie, ou pathologie, intervenir en post ou en pré chirurgie, corriger par un orthodontiste ou par un orthophoniste. Tous ces spécialistes, différent par leur degré d'intervention relatif au stade d'évolution de la pathologie et de son emplacement.

Ce chapitre est une introduction aux pathologies relatives à la parole, ou en d'autres termes « au langage vocal » et aux pathologies subséquentes des différents « articulateurs » lors de la production de la voix et son altération au cours de son cheminement à travers le conduit buccal ou nasal.

II .2 Différentes pathologiques vocaux :

Les signaux pathologique vocaux sont des atteintes peuvent concerner les organes périphériques, atteintes qui gênent la production de la parole : le bec de lièvre, la division palatine, l'insuffisance vélaire, les malformations linguales, labiales ou laryngées. Il s'agit d'anomalies consistant en des erreurs mécaniques et constantes dans l'exécution du mouvement propre à un phonème. L'articulation est la capacité à articuler les sons de façon permanente et systématique, ce qui nécessite des mouvements précis de la mâchoire inférieure, de la langue, des lèvres, des joues, du voile du palais. Le trouble d'articulation isolé est donc l'incapacité à prononcer ou à former un certain phonème correctement. C'est une erreur constante, systématique et mécanique pour un phonème donné. Cette erreur est plutôt de type praxique. La Production de la voix normale est basée la qualité de sa production, son intensité, son débit. Une voix pathologique présente une altération d'un ou de plusieurs de ces paramètres.

On peut classifier les troubles du langage en régions pathologiques, c'est ce qui montre que la voix peut être altérée ou modifiée tout le long de sa production, voire disparaître, phénomène décrit par l'apparition d'une aphonie ou absence de voix complètement, surtout lors du cancer des cordes vocales.

II .2.1 Pathologie des cordes vocales :

Avant de commencer l'étude des pathologies affectant la voix, il serait important de voir l'apparence de cordes vocales saines, car elles sont l'organe le plus important dans la production vocale.

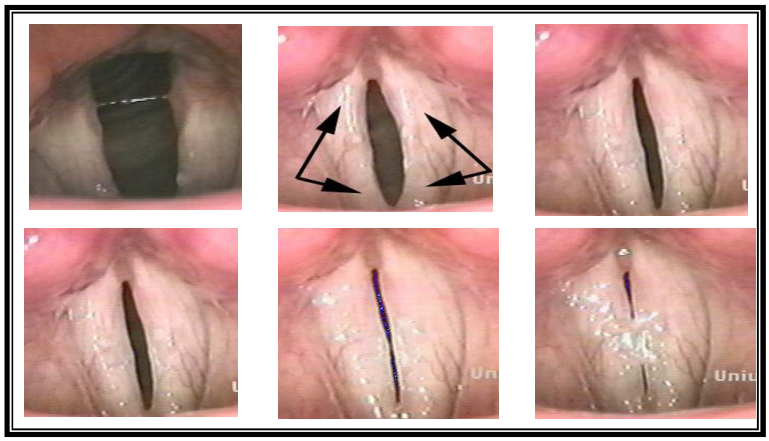


Figure II.1 : Cordes vocales saines avec différents degrés d'aperture.

L'atteinte des cordes vocales par n'importe quelle maladie ou par simple irritation, lorsqu'on crie très fort pendant un événement quelconque, agit essentiellement sur leur manière de vibrer, soit en atténuant le mode vibratoire par une paralysie ou en les rendant plus enroué ou plus âpre, comme illustré dans la figure II.2.

Les différentes pathologies les plus importantes concernant les cordes vocales sont illustrées ci-après :

Nous commençons notre étude sur les pathologies du langage par le lieu le plus important dans la phonation, les cordes vocales.

1-Nodules :

Ce sont des nœuds durs en forme de pointe de flèche situés sur la partie vibratoire de contact des deux cordes vocales, voir figure II.2.

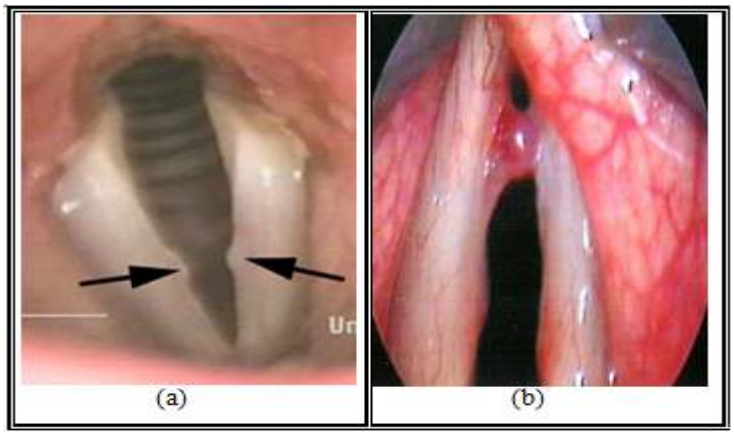


Figure II.2 : (a, b) Nodules sur les cordes vocales

2-Paralysie des cordes vocales :

Ce phénomène apparaît lorsque l'une des cordes, devient non élastique, comme illustré dans la figure II.3, ou s'arrête presque de bouger; Cette situation est généralement traitée par intervention chirurgicale soit par une lipo-injection ou thyroplastie de la corde paralysée pour permettre une fermeture complète.

Figure II.3 : Paralysie unilatérale des cordes vocales

3- Cordes vocales arquées :

Lorsque les deux cordes vocales ne se ferment pas complètement, La surface restante affecte la production phonatoire, ceci peut rendre la voix très fragile.

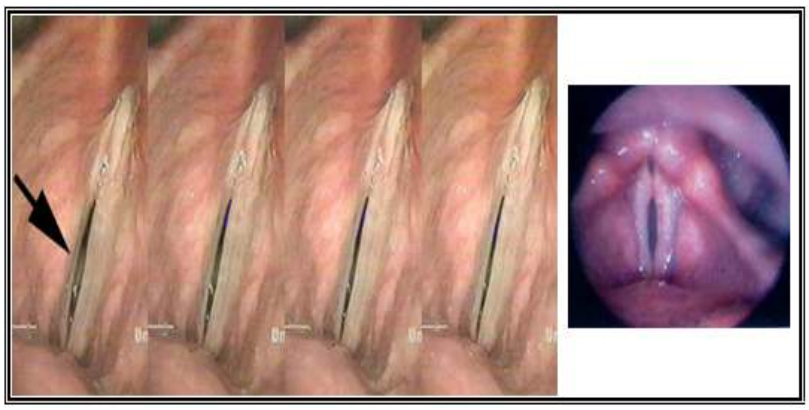


Figure II.4 : Cordes vocales pathologiques présentant un arc à la fermeture

4- Polypes dans les cordes vocales

Les polypes, voir figure II.5 dérangent la fermeture et les vibrations des cordes vocales, cette situation cause l'enrouement, la fatigue des cordes et une diminution du mode musicale.

5- Odeme de Reinke

L'oedème de Reinke est marqué par l'accumulation d'oedème et de fibrose dans la totalité de la corde vocale, voir figures II.5 et II.6. L'étiologie est essentiellement le tabac associé ou non à l'alcool. Il s'agit d'une lésion bénigne mais qui peut être associée à un cancer développé ailleurs dans les Voies Aéro digestives Supérieures "V.A.D.S." dans 5 à 10 % des cas. L'accumulation de l'oedème peut conduire à un véritable ballonnement des cordes accompagnées de dyspnée. La répartition est globalement identique suivant les deux sexes, mais la population féminine consulte plus facilement du fait de la répercussion de cet oedème chronique sur la voix. En effet le fait marquant de cette dysphonie est l'abaissement de la hauteur vocale. Ce trouble vocal est généralement bien accepté chez les hommes auxquels il donne un caractère "viril". A l'opposé chez la femme, la gêne est manifeste. La patiente est fréquemment appelée Monsieur au téléphone.

Le traitement est microchirurgical et consiste à inciser la corde vocale sur sa face supérieure et à aspirer la glue. L'abstention tabagique prévient la récidive.

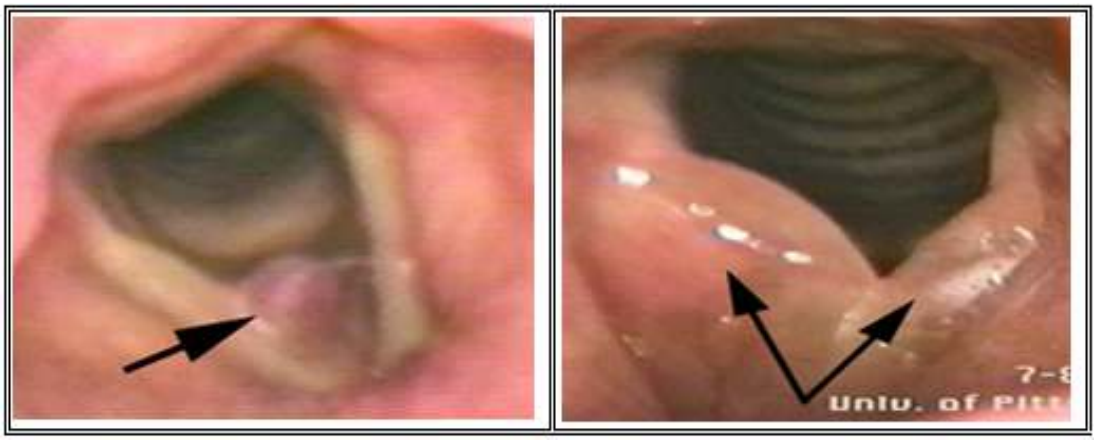


Figure II.5 : Polype très distingué sur l'une des cordes vocales à gauche

Figure II.6: Cordes vocales gonflées à gauche.

6- Kyste localisé au niveau des cordes vocales :

Le kyste muqueux, voir figures II.7, est formé par l'obstruction de canal excréteur d'une glande muqueuse de la corde vocale. Macroscopiquement, on observe une voussure plus ou moins allongée. Plus le kyste est ancien et plus le liquide paraît épais. Le traitement est microchirurgical.

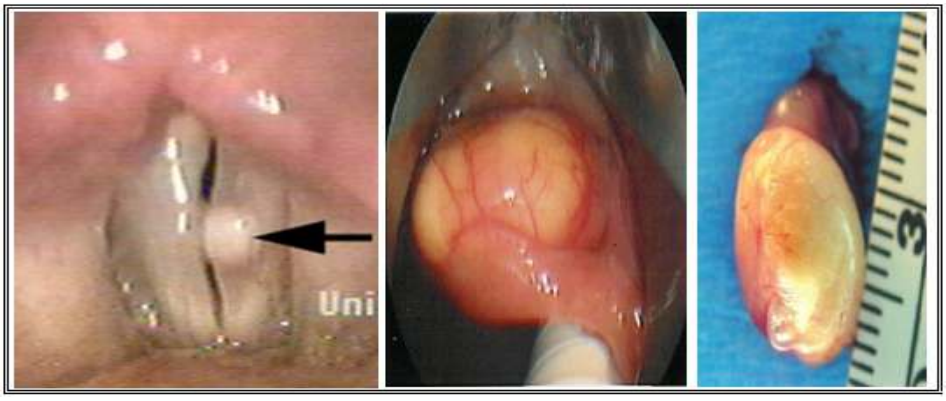


Figure II.7 : Détail d'un kyste de différents patients

7-Granulomes dans les cordes vocales :

Petits amas granulomateux, inflammatoire « constitué de chair », c'est-à-dire de tissu conjonctif se développant sur la muqueuse du larynx et à ses dépens. Quelquefois ils sont observés au niveau de la trachée après une intubation du larynx et de la trachée pour une ventilation assistée, le plus souvent ayant eu lieu au cours d'une anesthésie générale ou un coma.

La flèche sur la figure II.8. Indique un tissu épais et irrégulier sur les cordes vocales.

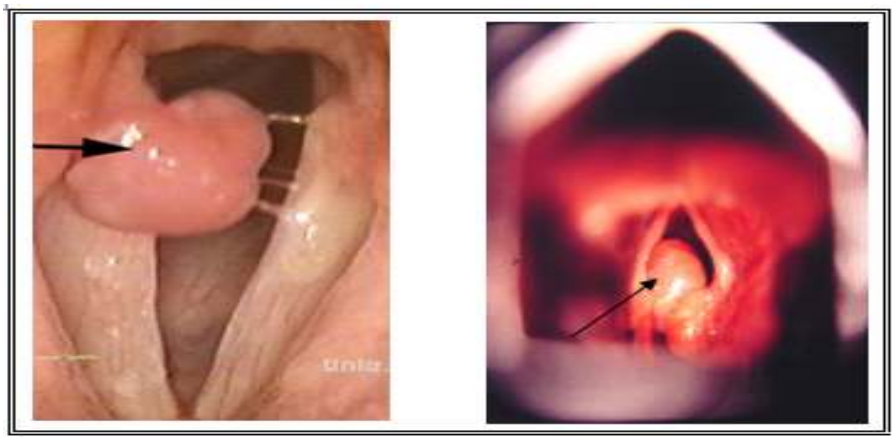


Figure II.8 : Granulomes de différents patients

8- Papillomes laryngés

Les différentes flèches sur la figure II.9 indiquent l'évolution des papillomes, dans le larynx, ceux-ci sont causés par une infection virale.

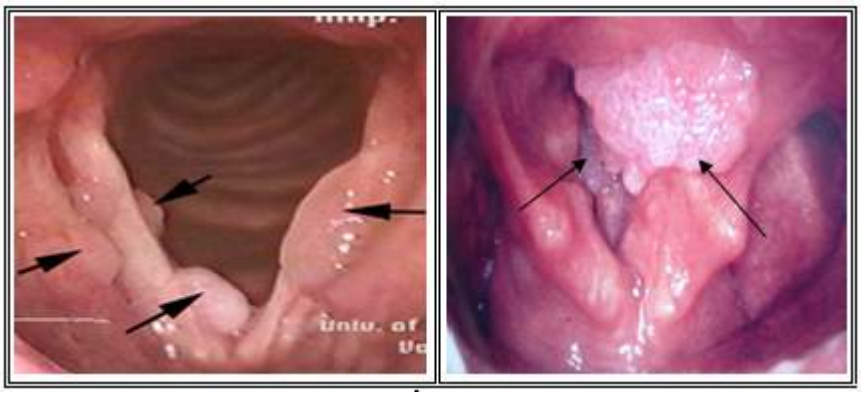


Figure II.9 : Papillomes chez différents patients

Des études bien détaillées des pathologies indiquées sont décrites par « The National Center for Voice and Speech » et par la clinique Otolaryngology-Head & Neck Surgery.

9- Cancers des Voies Aero-digestives Supérieures "V.A.D.S." :

Le cancer, maladie parfois non bénigne, détectable après biopsie, se localise dans différents points du corps humain, entre autres dans le larynx et les différentes cavités vocales et nasales, appelées Voies Aérodigestives Supérieures, voir figure II.10.

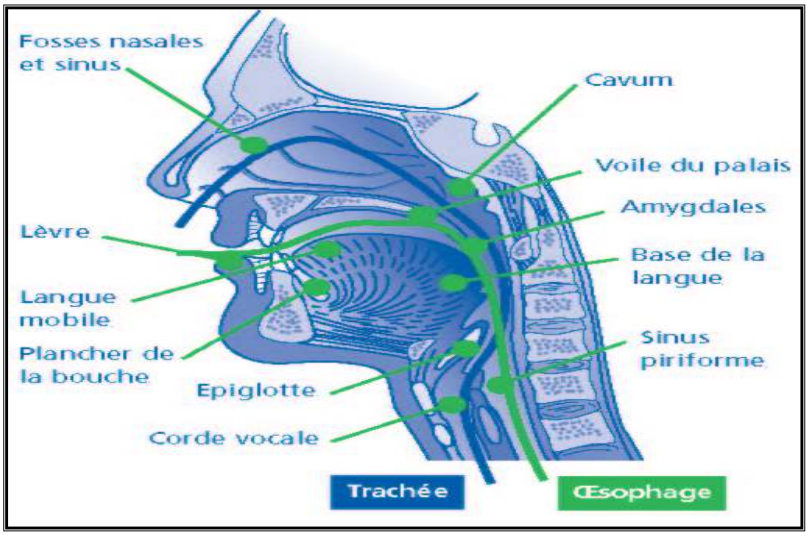


Figure II.10 : Sièges possibles des cancers.

10- Cancer du larynx :

Ce cancer, favorisé par l'alcool et le tabac, se voit surtout chez l'homme après 50 ans.

Symptômes

Le signe révélateur est une dysphonie progressive : le patient se plaint d'un enrouement. La dyspnée (gène à la respiration), la dysphagie (difficulté pour avaler), sont beaucoup plus tardives. Tout enrouement chronique nécessite un bon examen laryngologique direct.

L'ORL examine sous anesthésie locale le larynx grâce au miroir laryngé. Voir figure II.11. La tumeur est ainsi observée. Le médecin apprécie ensuite la mobilité du larynx et recherche des ganglions palpables. L'examen endoscopique (laryngoscopie) recherche une localisation cancéreuse oesophagienne et permet la biopsie à la pince de la lésion.

Diagnostic différentiel

- tumeurs bénignes du larynx (polypes des cordes vocales, nodules vocaux…);
- tuberculose laryngée;
- laryngite chronique (qui peut dégénérer) ;
- atteinte neurologique des cordes vocales.

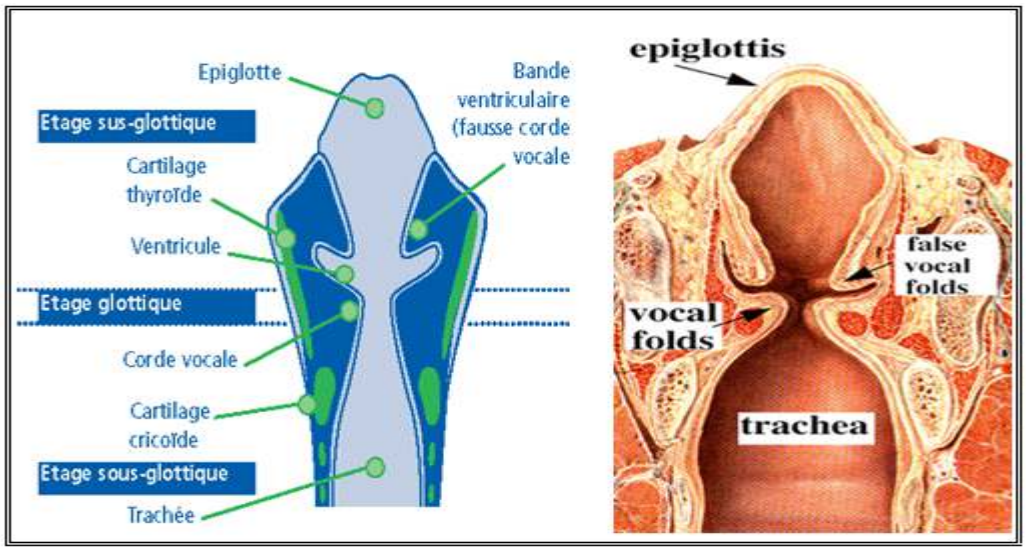


Figure II.11 : Vue en coupe du larynx.

11- Cancer Du Pharynx (Oropharynx et Hypopharynx) :

Le cancer du pharynx apparaît en général après les symptômes suivants :

- une gène ou une douleur d'un coté de la gorge;
- une sensation permanente d'un corps étranger ou d'angine traînante d'un seul coté ;
- une douleur à une oreille ;
- une difficulté à avaler.
- une sensation de brûlure d'un coté de la gorge ;
- une modification progressive de la voix qui devient couverte, voilée ou rauque ;
- apparition d'une boule dans le cou qui correspond à un ganglion.

12- Cancer De La Bouche :

- Le cancer de la bouche apparaît en général après les symptômes suivants :
- une douleur d'un coté de la bouche;
- une zone bourgeonnante ou creusée saignante, ne guérissant pas après un traitement d'une anomalie dentaire;

- un changement de la muqueuse persistant dans la bouche (tache rouge foncé ou blanche ressemblant à un aphte, mais à bords irréguliers);
- une gène au port d'un dentier;
- une douleur à une oreille;
- une difficulté à avaler;
- une sensation de chaud, au froid, au vinaigre, au citron.

13- Cancer des Cordes Vocales (Glottiques) :

Ces cancers se révèlent par une modification progressive de la voix qui devient couverte, voilée, rauque (dysphonie). Cette modification persiste et s'aggrave progressivement. Elle est parfois précédée d'épisodes transitoires de laryngite où complique une laryngite chronique ancienne, fréquente chez les fumeurs et/ou les personnes travaillant en atmosphère chaude et sèche, ou chargée de poussières.

14- Cancer Des Cordes Vocales (Sub - Glottiques) :

Ils siègent au niveau de l'épiglotte, par :

- une douleur d'un seul côté de la gorge,
- une difficulté à avaler,
- une sensation permanente de corps étranger ou d'angine d'un seul côté.
- une douleur à une oreille
- l'apparition d'une boule dans le cou qui correspond à un ganglion.

II .2.2 Pathologies des autres canaux vocaux :

1-Bec de lièvre :

C'est une déformation prénatale; voir figure II.12, s'attaquant à la lèvre supérieure, d'origine génitale, pouvant être corrigée par une intervention chirurgicale.

2-Palais enclavé :

L'une des conséquences directes de l'absence du palais, voir figure II.13 est l'hypernasalité, c'est une pathologie de la résonance de la voix, causé par le disfonctionnement du mécanisme vélopharyngeale, celle-ci provoque un :

- nasonnement ouvert ou hyperrhinophonie : le voile du palais ne ferme pas le passage de l'air à la cavité nasale dans le cas de division palatine ou d'opérations des végétations notamment ;
- nasonnement fermé ou hyporhinophonie : pas de nasalisation pour les consonnes et les voyelles nasales.
- Une intervention chirurgicale est à la base de la correction de cette pathologie;
- Une étude très intéressante, portant sur les remarques d'enfants parlant l'Arabe avec un palais enclavé, est développée dans.
- D'autres définitions, causes et traitements sont bien traités dans le guide vocologique ou « Guide to Vocolgy » émis par le Centre National de la Voix et de la Parole, le N.C.V.S.

Figure II.12 : Avant intervention / Après intervention.

Figure II.13: Palais totalement absent.

II .2.3 Classification des pathologies :

La première partie a concerné les pathologies des organes intervenant dans la production ou l'altération de la voix, celles-ci sont classées comme suit:

- **Disfonctionnement fonctionnelle :** L'organe existe mais, il y'a eu soit un mauvais apprentissage, soit une maladie en cours d'évolution ce qui présente un symptôme de pathologie de la parole, si la détection de la pathologie n'est pas effectuée à temps.

- **Disfonctionnement organique :** L'organe existe ou est absent, cas du palais enclavé ou laryngectomie, mais ne peut exécuter la tache préconçue, soit par atrophie cas de la langue trop courte, soit par surdimensionnement cas du volume du palais démesurée,

Les différents défauts émanant de ces pathologies sont classés selon le diagramme de la figure II.14.

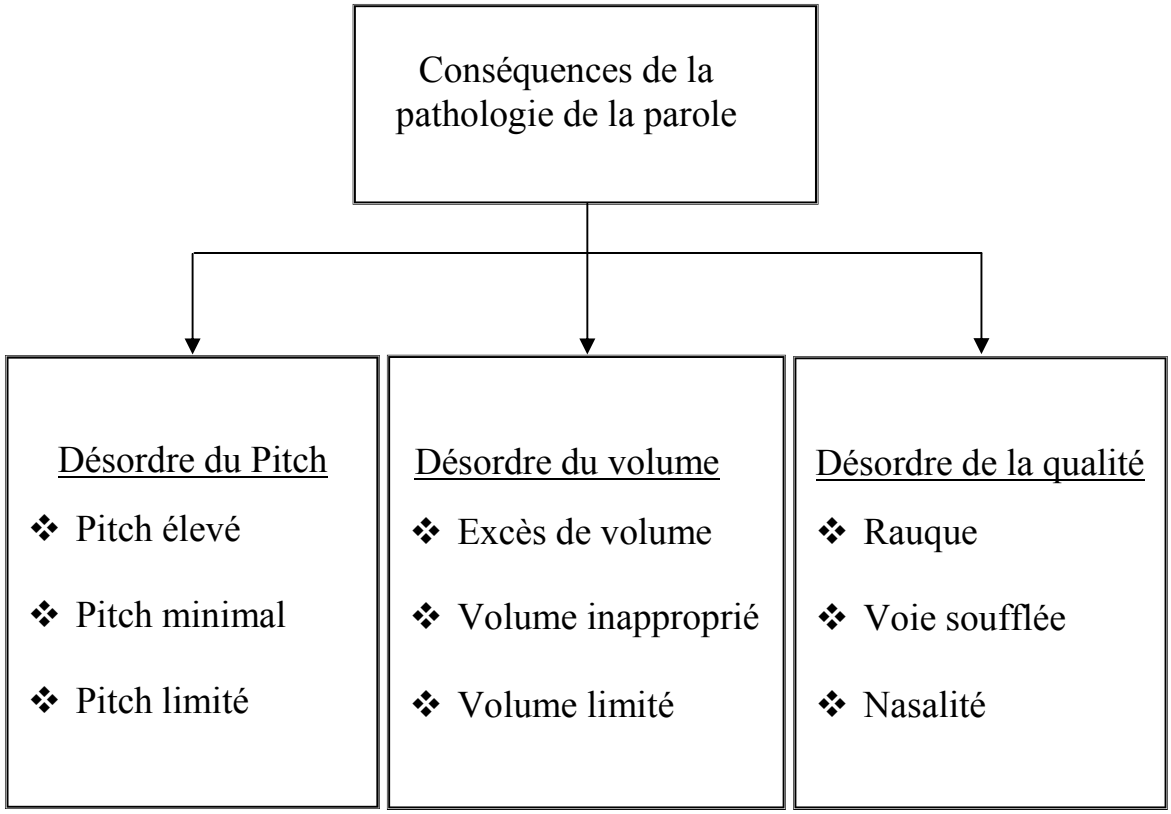


Figure II.14 : Diagramme de classement des pathologies.

II .2.4 Défauts de la voix détectés par l'oreille :

Les différents défauts de la voix détectés par l'oreille humaine sont :

1- Le blésement ou zézaiement :

Le blésement ou zézaiement est un défaut de prononciation qui consiste en la substitution de [ʃ] (une consonne chuintante) par [s] (une sifflante) et de [g] ou [j] (Consonnes chuintantes) par [z] (sifflante).

2- Chuintement :

Le chuintement est la prononciation du [s] et du [z] à la manière du [ʃ] et du [j] Français.

Exemple :

- J'ai pris l'autobus jusqu'à la gare Saint-Lazare
- J'ai pris l'autobuch juchqu'à la gare Chaint-Lajare.

3- Le rhotacisme :

Le rhotacisme (terme formé à partir du grec ρ, [r]) est une modification phonétique complexe, consistant en la transformation d'un phonème en [r]; Dans d'autres langues comme le Français c'est avec le [z] que le [r].

Pour la langue arabe c'est la confusion entre le [r] et le [ʁ].

Donc, au lieu de prononcer 'ريح ' [rihø].

La personne atteinte de rhotacisme prononce "غيح" [ʁihø],

4- Nasonnement :

C'est l'altération du son de la voix; le nasonnement provient de la diminution de la résonance nasale par suite de l'obstruction du nez, de la présence de végétations adénoïdes, etc., et produit une déformation des syllabes nasales, [an], [on], [in], et des consonnes nasales, telles que [m], que l'on prononce [b].

Exemple : En prononçant [pa] l'air ne doit pas passer par le nez. Le nasonnement est l'inverse du rhume, ou l'air ne peut pas passer par le nez.

5- Bégaiement :

C'est le trouble de la communication affectant le débit et le rythme de la parole se traduisant par:

- une forme clonique : répétition;
- une forme tonique : blocage;
- des troubles associés;

Si rien n'est entrepris, sur 4 enfants de 2 à 5 ans commençant à bégayer, 1 restera bègue à l'âge adulte. Il est nécessaire d'intervenir le plus tôt possible pour ne pas prendre le risque de la chronicisation.

Les signes d'appel et manifestations du bégaiement se présentent comme suit :

- répétition de sons ou syllabes supérieures à 3 (ex: tou tou tou toupie) ;
- prolongation de sons ;

- blocage de syllabes ;
- répétitions de mots, de parties de phrases ;
- reprise d'énoncés ;
- hypertonie, blocages respiratoires lors de la prise de parole;
- comportement ou modification du comportement : colères, retrait, timidité, énurésie;
- Comportement verbal: refus ou repli ;
- antécédents de bégaiement dans la famille.

6-Le clichement :

Le clichement est un défaut de prononciation se caractérisant par le fait d'additionner le son (double L) mouillé, positionné après certaines consonnes.

Une consonne mouillée est articulée avec le son j. Par exemple l dans grisaille.

Un exemple de clichement : prononcer chilluchoter au lieu de chuchoter.

7- Le gammacisme

Défaut de prononciation se caractérisant par la difficulté voir l'impossibilité de prononcer les consonnes gutturales, [k] à la place de [g].

8 - Retard de parole :

Le retard de la parole est l'altération de phonèmes ou de groupes de phonèmes, par leur mise en ordre séquentielle à l'intérieur d'un même mot, le stock phonétique étant acquis. C'est la forme du mot dans son ensemble qui ne peut être reproduite.

Un parler bébé qui perdure au-delà de 4 ans est caractérisé par :

- des omissions mots raccourcis ou élidés : fleur/feur, herbe/è;
- des inversions brouette / bourette;
- des assimilations : lavabo lalabo ou vavabo;

- des interversions : kiosque / kiokse;
- des simplifications : parapluie / papui ...;
- des substitutions : train / crain, fleur fieur;
- des élisions de syllabes finales : pelle pè, assiette assiè...

Et de façon plus globale :

- des problèmes de perception auditive;
- une mauvaise structuration de la perception du temps;
- une mauvaise structuration de la chronologie des sons;
- des difficultés motrices diverses;
- une attention auditive labile;
- une immaturité psychoaffective;
- un refus de grandir ...

Fréquemment, un retard de langage et/ou un trouble d'articulation peuvent être associés au retard de parole.

II.3 Conclusion

A travers ce chapitre, nous avons cité un éventail assez large de pathologies langagières, ayant trait aux variations phonétiques dû surtout à des pathologies anatomiques du conduit vocal et l'appareil phonatoire. Pour notre travail nous nous sommes intéressés à deux classes de sons, des sons éligibles et non éligibles sans se souciés du types de pathologie anatomique. Dans le chapitre suivant, différentes techniques d'analyse du signal vocal sont présentées, en vue de classifier les classes de sons cités précédemment.

III.1 Introduction

L'extraction de caractéristiques est une partie essentielle pour le traitement du signal, il précède souvent l'étape de la reconnaissance vocale. Il en existe un nombre important de techniques pour représenter le signal de la parole, parmi lesquels les paramètres LPC (Linear Prédictive Coefficients), les LPCC (Linear Prédictive Coefficients Cepstral) et NPC (Neural Prédictive Coding) ou bien encore les coefficients MFCC (Mel-FrequencyCepstral Coefficients).

Nous introduisons à travers notre travail un nouvel extracteur de caractéristiques « Open Smile » qui est un Open-Source pour le traitement incrémental. Smile est un acronyme pour la parole et l'interprétation de la musique par l'extraction à grande espace.

III.2 Définition d'Open Smile

La boite à outils de Munich Open Speech et l'interprétation de la musique par extracteur à grand espace (Open Smile), est un extracteur de caractéristiques modulaire et flexible pour le traitement du signal et les applications des machines d'apprentissage. L'objectif principal est clairement mis sur les caractéristiques du signal audio. En raison de leur haut degré d'abstraction, les composants d'Open Smile peuvent être utilisés pour analyser les signaux ded'autres modalités comme les signaux physiologiques, les signaux visuels et d'autres capteurs physiques. Il est écrit purement en C++ et dispose d'une architecture rapide, efficace et souple, il fonctionne sur diverses plates formes tels que Lunix, Windows et Mac OS. Open Smile est conçu pour le traitement en-ligne en temps réel, mais peut être utilisé hors-ligne avec la nécessité de la présence de touts les entrées. Open Smile peut extraire les caractéristiques progressivement quand les nouvelles données arrivent, il utilise la bibliothèque Port Audio [18] qui permet l'extraction en temps réel.

Pour facilité l'interopérabilité, Open Smile gère la lecture et l'écriture de différentes formats de données qui sont utilisées dans l'exploitation des données et l'apprentissage des machines ; ces formats sont PCM, WAVE pour les fichiers audio, CSV (Comma Separated Value : format de feuille de calcul) et SSLIA (Weka Data Mining) pour les fichiers de données de type texte, HTK (Hidden Markov Toolkit) des fichiers de paramètres et une matrice de format binaire simple flottant pour des données de caractéristiques binaires.

III.3 Domaine d'utilisation

Open Smile est utilisé par les chercheurs et les entreprises du monde entier qui travaillent dans le domaine de la reconnaissance vocale (extraction de caractéristique frontal, mot clé spotting…), le domaine de l'information affective (reconnaissance des émotions, affectation des agents virtuels sensibles…) et la récupération de l'information musicale (de l'étiquetage de la corde, la détection de l'apparition…).

III.4 Vue d'ensemble

Les capacités d'Open Smile sont distinguées par les catégories suivantes : les données d'entrées, traitement du signal, traitement général de données, caractéristiques audio bas niveau, fonctionnelles, les classifieurs et d'autres composants, les données de sortie et d'autres capacités.

III.4.1 Les données d'entrée

Open Smile peut lire les données à partir des formats de fichiers suivantes :

- RIFF-WAVE (PCM) (pour MP3, MP4, OGG, …un convertisseur doit être utilisé) ;
- Valeur séparées par des virgules (CSV) ;
- Fichier de paramètres HTK ;
- Formats ARFF de Weka.

En outre, l'enregistrement d'un audio en direct à partir de n'importe quelle carte son, est pris en charge par la bibliothèque Port Audio.

III.4.2 Traitement de signal

Cette fonctionnalité est fournie pour un traitement ou un pré-traitement (avant l'extraction) :

- Fonctions de fenêtrage (Rectangulaire, Hamming, Hann, Gauss, Sine, Triangulaire, Bartlett, Bartlett-Hann, Black man-Harris, Lonczos) ;
- Pré / De-emphasis (1^{ier} ordre pass haut / bas) ;
- FFT (amplitude, phase, complexe) et l'inverse ;

- Mise à l'échelle de l'axe spectral par l'interpolation spline(version Open Source uniquement) ;
- dbA pondération de spectre d'amplitude ;
- Fonction d'auto-corrélation (ACF) (via IFFT de spectre de puissance) ;
- Fonction de différence de la grandeur moyenne (AMDF).

III.4.3 Traitement des données

Open Smile peut effectuer un certains nombres d'opérations pour la fonction de normalisation, modification et de la différenciation :

- La normalisation de la variance moyenne (hors ligne et en ligne) ;
- L'égalisation d'histogramme (expérimental) ;
- Les coefficients de Delta régression (et simple différentiel) ;
- Différentiel pondérée ;
- Filtre à moyenne mobile pour le lissage des contours supplémentaires.

III.4.4 Caractéristiques audio (niveau bas)

Les distributeurs de bas niveau peuvent être calculés par Open Smile :

- Energie ;
- L'intensité de la trame (approximation) ;
- Le spectre de bande critique (Mel / Bark / Octave, filtre de masquage triangulaire) ;
- Les coefficients spectraux de fréquence Mel / Bark (MFCC) ;
- Spectre auditif ;
- Les coefficients de prédiction linéaire perceptuelle (PLP) ;
- Les coefficients de prédiction linéaire (LPC) ;
- Fréquence fondamentale ;
- Probabilité de sonorisation (voicing) ;
- Pitch ;
- Les fréquences formants et largeurs de bande (à partir de racine LPC) ;
- Taux de passage par zéro ;
- Caractéristiques spectrales (énergie arbitraire de la gravité, entropie, variance (=propagation), asymétrie, aplatissement, pente) ;
- Chroma (spectres d'une octave demi-ton déformé) et la caractéristiques de la dérivé de chroma pour CHORD et reconnaissance clef).

III.4.5 Fonctionnels

Afin de reporter les contours de descripteurs de bas niveau sur un vecteur de dimensionnalité fixe, les fonctionnelles suivantes peuvent être appliquées :

- Les valeurs et les positions extrêmes ;
- Les moyennes (arithmétique, quadratique, géométrique) ;
- Les moments (écart type, variance, coefficient d'aplatissement, asymétrie) ;
- Percentiles et limites de percentile ;
- Centre de gravité (centroid) ;
- Pics (peaks) ;
- Les valeurs des échantillons ;
- Temps / Durée ;
- Onsets ;
- Transformation de cosinus directe (DCT) ;
- Passage par zéro.

III.4.6 Les classifieurs et d'autres composantes

Souvent les démonstrateurs en direct pour les taches de traitement audio exigent la segmentation de l'audio au courant, pour cela Open Smile fournit des algorithmes pour la détection d'activité vocale et un détecteur de tour, pour classer les caractéristiques extraites des segments de façon incrémentielle, les machines à vecteurs de supportsont mis en œuvre en utilisant la bibliothèque Libsvm.

- Détection d'activité vocale basée sur la logique floue ;
- Classifieur NN (réseau de neurones) pour la détection d'activité vocale adaptatif ;
- Détecteur de tour ;
- Libsvm (en ligne).

III.4.7 Les données de sorties

Pour écrire des données dans des fichiers, les mêmes formats d'entrées sont pris en charge sauf le format de matrice binaire :

- RIFF-WAVE (PCM non compressé) ;
- Valeurs séparés par des virgules (CSV) ;
- Fichiers de paramètres HTK ;

- Fichier Weka ARFF ;
- Format de fichiers de caractères Libsvm ;
- Format de matrice binaire flottante.

En outre, la lecture d'audio en direct est pris en charge par les bibliothèques Port Audio.

Open Smile est livré avec d'autres capacités qui rendent son utilisation facile et polyvalente :

- **Multi – threading :**c'est un composant indépendant qui pet être exécuté en parallèle pour utiliser plusieurs CPU ou à fin d'accélérer l'extraction des paramètres quand le temps est critique.
- **Plugin – support :**c'est un composants supplémentaire peut être construit comme une bibliothèque partagé (ou DLL sous Windows) liée avec la bibliothèque d'API de base d'Open Smile.
- **Extensive logging.**
- **Configuration flexible d'Open Smile.**
- **Traitement incrémentiel.**

III.5 Utilisation d'Open Smile

III.5.1 Installation d'Open Smile

Un paquet de version binaire contient un exécutable principale SMILExtract qui est lié statiquement pour les systèmes Lunix et un SMILExtract.exe et openSmileLib.dll pour les systèmes Windows, la version stable d'Open Smile peut être trouvé sur le site web [19].

III.5.2 L'extraction des premiers caractéristiques

Pour commencer à utiliser Open Smile il faut assurer que les fichiers de configuration soient au même répertoire que l'exécutable SMILExtract.exe et OpenSmileLib.dll, ces fichiers peuvent être téléchargés ou générés selon nos besoins [20].

Dans notre travail nous avons utilisés Open Smile avec Matlabpour générer un fichier de configuration de type Arff qui comporte les caractéristiques extraites de l'audio souhaité. Afin d'assurer un démarrage rapide, il faut définir tous d'abord le chemin du fichier de configuration utilisé, le chemin de fichier wave ainsi le nom du fichier Arff qui va être généré en tapant la commande suivante :

'smilExtract -C ',PathToConf ,' -I ' ,PathToMyWavefile,' -O ',OutPutFile.arff

Mais si nous utilisons cette commande telle quelle est,Matlab va nous donner un message d'erreur, pour cela nous devons ajouter une expression qui permet l'exécution de cette commande sous dos.

III.5.3 L'ensemble de caractéristiques par défaut

Pour des tâches courantes dans le domaine de la récupération de l'information musicale et de traitement de la parole, ils ont fourni certains fichiers de configuration dans le répertoire config/ pour des ensembles de caractéristiques fréquemment utilisés :

- Les caractéristiques de Chroma pour la reconnaissance de clé et corde ;
- MFCC et PLP pour la reconnaissance vocale ;
- Prosodie (hauteur et volume) ;
- L'interspeech 2009, le défi pour l'ensemble de caractéristiques des émotions ;
- L'interspeech 2010, le défi pour l'ensemble de caractéristiques de paralinguistique;
- Trois références d'ensembles de caractéristiques pour la reconnaissance des émotions.

III.6 Extraction des caractéristiques pour la reconnaissance vocale

Depuis l'utilisation d'Open Smile dans le projet openEAR [21] pour la reconnaissance des émotions, divers fichiers de configurations sont disponibles que nous allons les utilisés mais pour la reconnaissance des sons pathologiques.

III.6.1 L'interspeech 2009

Le défi pour l'ensemble de caractéristiques des émotions (voir [22]) est représenté par le fichier de configuration config/emo_IS09.conf (voir figure III.1). Il contient 384 paramètres comme fonctionnelles statistiques appliquées aux contours de descripteur de bas niveau. Les caractéristiques sont enregistrées sous format Arff (pour Weka) de sorte que les nouveaux cas sont ajoutés à un fichier existant (il est utilisé pour le traitement par lots où open Smile est appelé à plusieurs reprises pour extraire les caractéristiques de plusieurs fichiers wave dans un seul fichier de paramètres). Les noms des 16 descripteurs de bas niveau, comme ils apparaissent dans le fichier Arff, sont présentés dans la liste suivante :

- **pcm_RMSenergy**racine carré de la moyenne d'énergie pour une trame de signal ;
- **mfcc**de 1 à 12 ;
- **pcm_zcr**taux de fois que le signal passe par zéro;
- **voiceProb**la probabilité de sonorisation (voicing) ;
- **F0** la fréquence fondamentale.

Le suffixe _smaest ajouté auxnoms des descripteurs bas niveau, indique qu'ils étaient lissés par un filtre à moyenne mobile avec une fenêtre de longueur 3. Le suffixe _dejoint au suffixe _sma, indique que la caractéristique actuelle est un coefficient de 1er ordre delta (différence) du descripteur lissé de bas niveau. Les noms des 12 fonctionnelles sont :

- **max** la valeur maximale du contour ;
- **min** la valeur minimale du contour ;
- **range**=max-min ;
- **maxPos**la position absolue de la valeur max (dans les trames) ;
- **minPos**la position absolue de la valeur min (dans les trames) ;
- **amean**la moyenne arithmétique de contour ;
- **linregc1** la pente (m) d'une approximation linéaire du contour ;
- **linregc2** le décalage (t) d'une approximation linéaire du contour ;
- **linregerrQ**l'erreur quadratique calculé comme la différence de l'approximation linéaire et le contour réel ;
- **stddev**l'écart type des valeurs dans le contour ;
- **skewness**l'asymétrie (moment d'ordre 3) ;
- **kurtosis**le coefficient d'aplatissement (moment d'ordre 4).

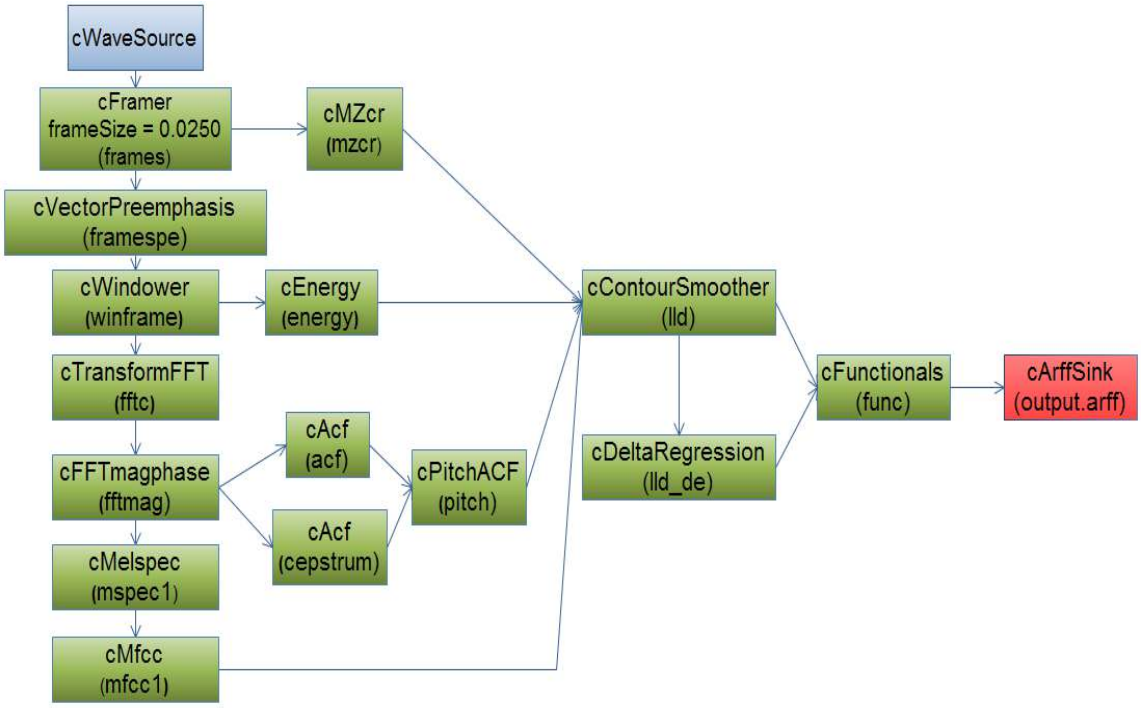


Figure III.1 : plan de configuration de emo_IS09.

III.6.2 L'interspeech 2010

Le défi pour l'ensemble des caractéristiques de paralinguistique est représenté par le fichier de configuration config/paraling_IS10.conf(voir figure III.2). L'ensemble contient 1582 paramètres qui résultent une base de 34 descripteurs de bas (LLD) avec 34 coefficients ajoutés qui correspond à delta et 21 fonctionnelles sont appliquées à chacun de ces 68 contours de LLD (1428 paramètres). En outre, 19 fonctionnelles sont appliquées au 4 pitch de base de LLD et leurs 4 contours de coefficient delta (152 paramètres). A la fin le nombre d'onsets de pitch (pseudo syllabe) et la durée totale de l'entrée sont ajoutés (2 paramètres).

Les caractéristiques sont enregistrées au format Arff (pour Weka), de sorte que les nouveaux exemples sont ajoutés à un fichier existant. Les noms des 34 descripteurs sont :

- **pcm_loudness**le volume sonore et l'intensité normalisée sont élevées à une puissance de 0.3 ;
- **mfcc**de 0 à 14 ;
- **logMelFreqBand**l'énergie logarithmique de fréquence Mel de bande 0-7 (distribués sur une plage de 0 à 8Khz) ;

- **IspFreq**les 8 paires de fréquences des lignes spectrales sont calculés à partir des 8 coefficients LPC ;
- **F0finEnv** l'enveloppe du contour de fréquence fondamentale lissée ;
- **voicingFinalUnclipped**la probabilité de sonorisation (voicing) du condidat final de la fréquence fondamentale ;

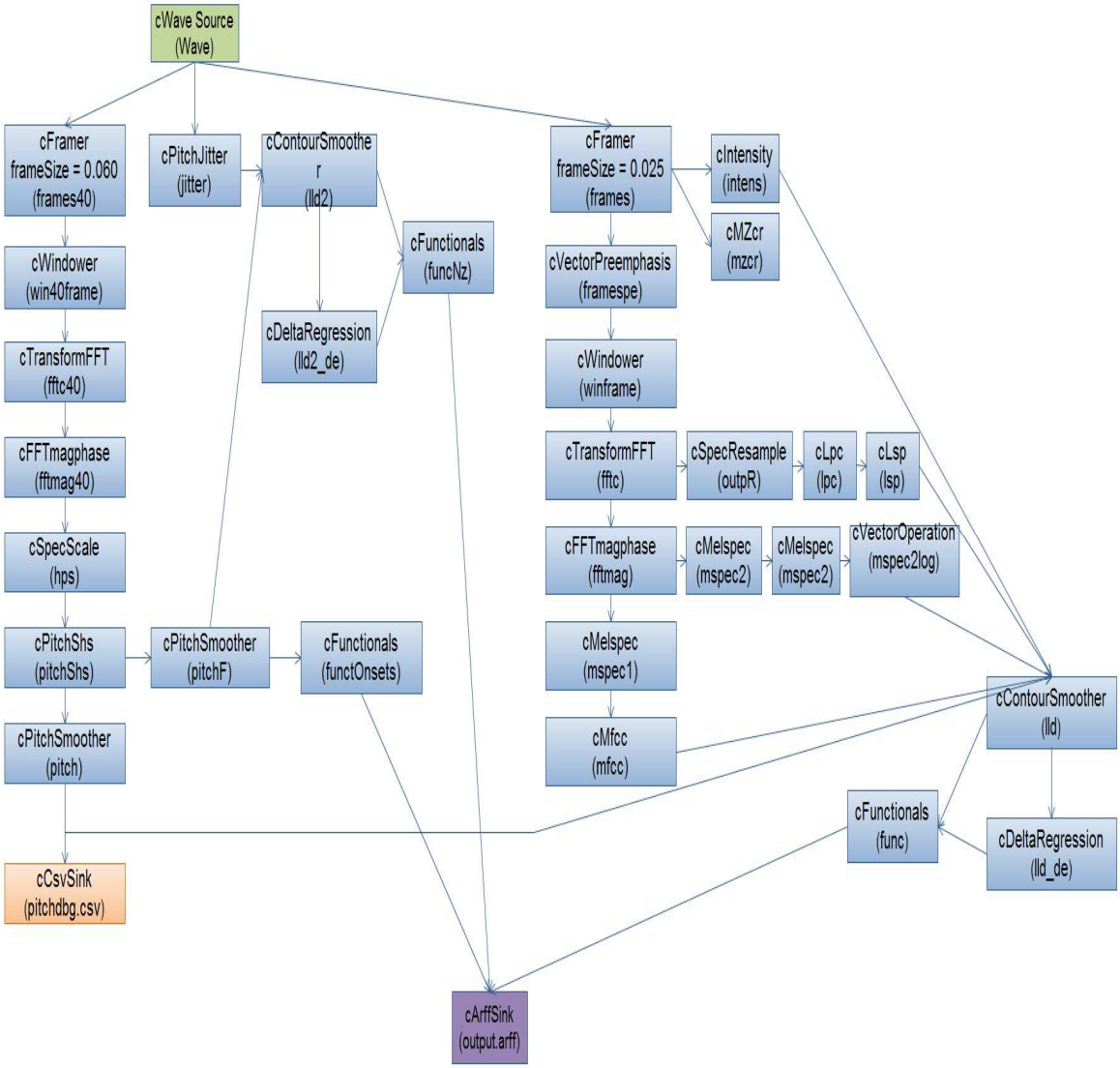


Figure III.2 : plan de configuration de paraling_IS10.

Quand les suffixes _smaet_de sont ajoutés aux noms des descripteurs, le nombre de fonctionnelles devient égale à 21, les 12 premiers sont identiques que celles de l'interspeech 2009, les noms de ce qui reste sont :

- **quartile1** le 1er quartile (25% percentile) ;
- **quartile2**le 2ème quartile (50% percentile) ;
- **quartile3**le 3ème quartile (75% percentile) ;

- **irq1-2** l'intervalle interquartile : quartile 2- quartile 1 ;
- **irq2-3** l'intervalle interquartile : quartile 3- quartile 2 ;
- **irq3-1** l'intervalle interquartile : quartile 1- quartile 3 ;
- **percentile1.0** la valeur minimale robuste des valeurs aberrantes du contour, représenté par 1% percentile ;
- **percentile99.0** la valeur maximale robuste des valeurs aberrantes du contour, représenté par 99% percentile ;
- **pctlrange0-1** l'aberrante robuste de l'intervalle de signal « max-min » est représenté par l'intervalle de 1% et 99% percentile ;
- **upleveltime75** le pourcentage de fois que le signal est au-dessus (75%*intervalle+min) ;
- **upleveltime90** le pourcentage de fois que le signal est au-dessus (90%*intervalle+min).

Les quatre pitch liés au LLD (et les coefficients delta correspondantes) sont :

- **F0final** le contour de fréquence fondamentale lissé ;
- **jitterLocal**le jitter local (trame à trame) (l'écart type de la période de pitch) ;
- **jitterDDP**la différence de jitter trame à tramme ;
- **shimmerLocal**le shimmer local (trame à trame) (l'écart type d'amplitude entre les périodes de pitch).

III.6.3 l'ensemble openSMILE/openEAR'emotion'

L'ancien jeu de base de 988 caractéristiques acoustiques pour la reconnaissance des émotions peut être extrait utilisant la commande suivante :

```
SMILExtract –C config/emobase.conf –I input.wav –O output.arff
```

Cela produira un fichier Arff avec un en-tête contenant tous les noms des caractéristiques et un vecteur de paramètres pour le fichier d'entrée. Pour ajoutés plusieurs exemples au même fichier Arff, il suffit d'exécuter la commande précédente une autre fois pour un autre fichier d'entrée. Le fichier Arff aura une étiquette de classe fictive appelée émotion, qui comporte une classe inconnue par défaut. Pour changer ce comportement et d'attribuer des classe et des étiquettes de classe personnalisées pour un cas particulier, il faut utiliser la ligne de commande suivante :

SMILExtract –C config/emobase.conf –I inputN.wav –O output.arff –instnameonputN –classes {anger, fear, disgust} –classlabelanger

Le paramètre –classes spécifie la liste des classes nominales qui sont entre les caractères {}, ou elles peuvent être un ensemble de chiffres pour des classes numériques (régression). Le paramètre –classlabelspécifie la valeur / l'étiquette de la classe de l'exemple calculé à partir de l'entrée actuellement donnée (-I).

L'ensemble de caractéristiques définies par emobase.confcontient les descripteurs bas niveau (LLD) suivantes (voir figure III.3): l'intensité, le volume sonore, 12 MFCC, pitch F0, probabilité de sonorisation (voicing), l'enveloppe de F0, LSF (fréquences des lignes spectrales), taux de passage par zéro.les coefficients de delta régression sont calculés à partir de ces LLD, et les fonctionnelles suivantes sont appliquées aux LLD et les coefficients delta : valeur Max/Min dans l'entrée, gamme, 2 linéaires quartile 1-3 et 3 gammes interquartile.

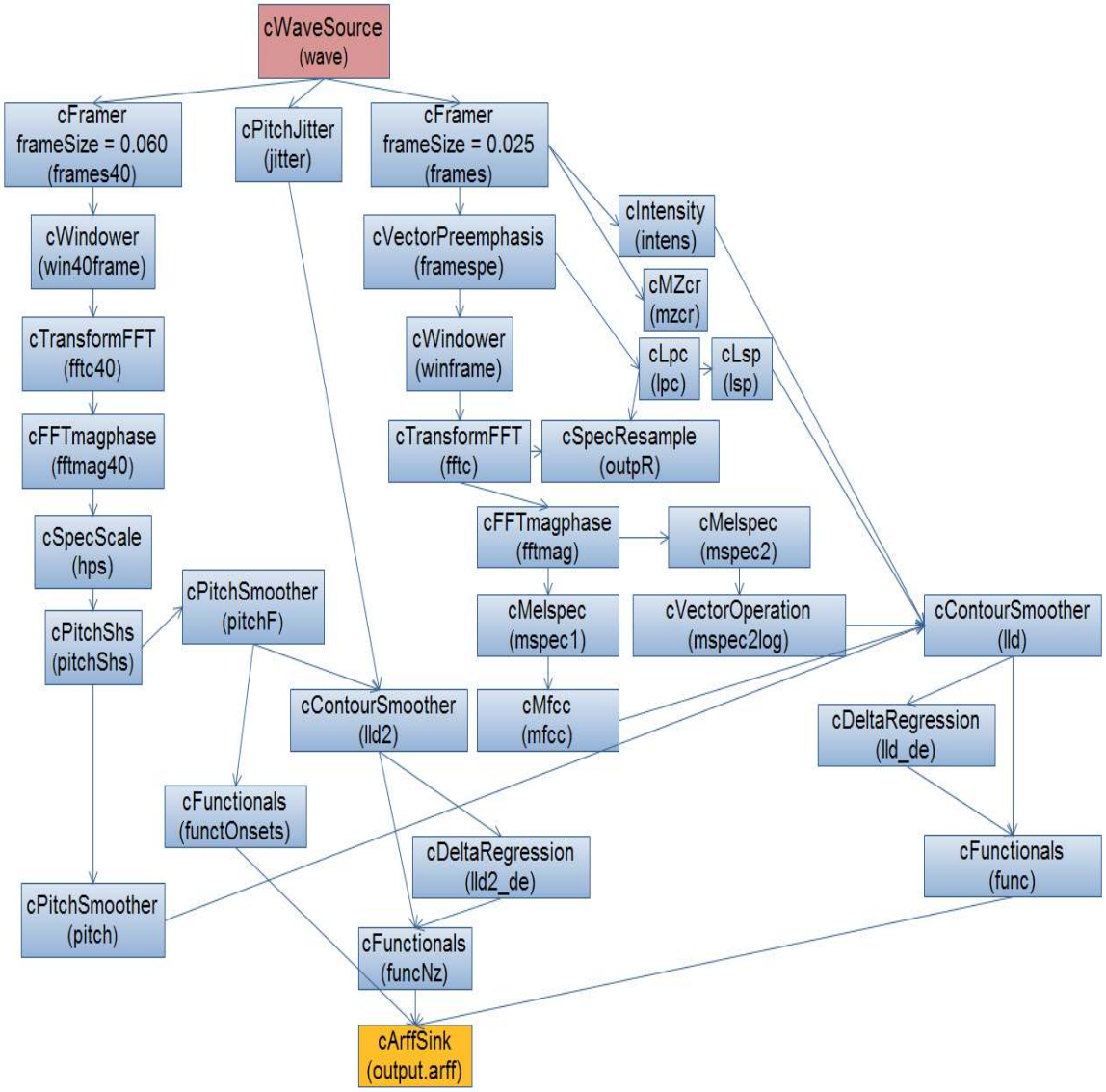


Figure III.3 : plan de configuration de l'emobase.

III.6.4 L'ensemble de références d'Open Smile 'emobase2010'

L'ensemble de caractéristiques est basé sur l'interspeech 2010 (même nombres de paramètres 1582), le défi de l'ensemble de paramètres paralinguistique (voir figure III.4). Il est représenté par le fichier config/emobase2010.conf. La seule différence de l'interspeech 2010 est la normalisation de maxPos et minPos qui sont normalisés par rapport à la longueur de segment dans l'ensemble actuel.

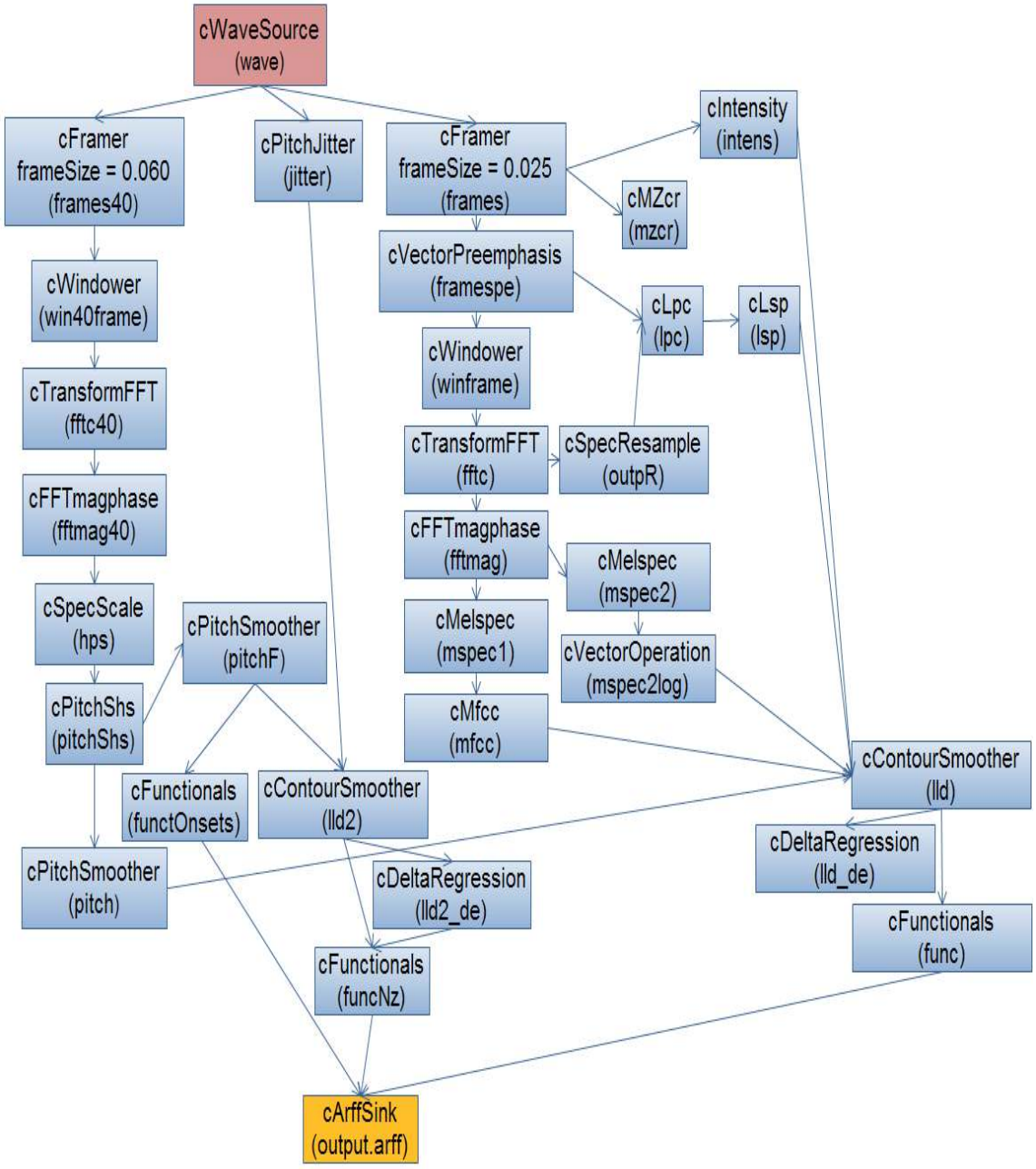


Figure III.4 : plan de configuration de l'emobase2010.

III.6.5 le grand ensemble de caractéristiques pour l'émotion d'Open Smile

Pour extraire un grand ensemble de paramètres avec plus de fonctionnelles et LLD (total 6552 paramètres), le fichier de configuration config/emo_large.conf est primordial (voir figure III.5).

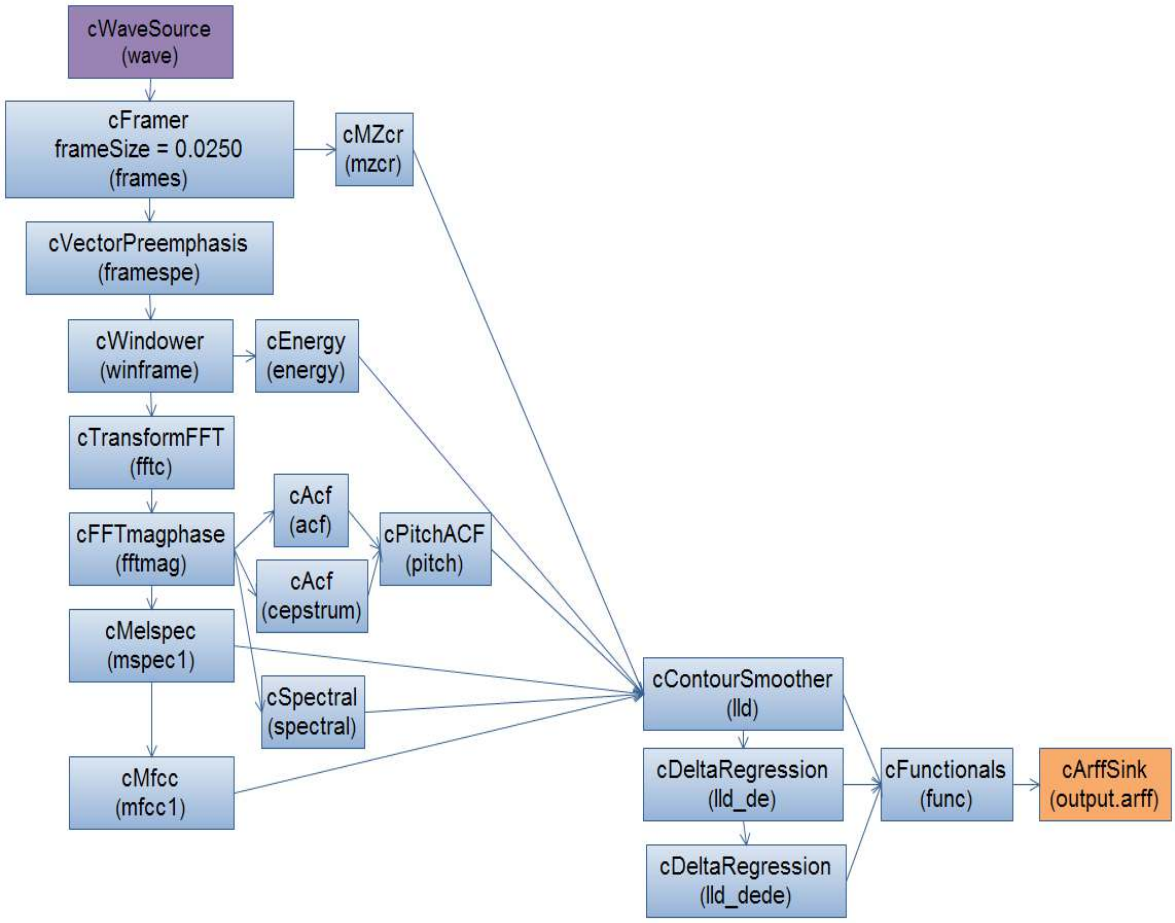


Figure III.5 : plan de configuration de l'emo_large.

III.7 Conclusion

L'extraction de caractéristiques du signal de parole est une étape fondamentale dans le traitement du signal vocal. Dans notre travail nous nous sommes intéressés à l'extracteur de paramètres à grand échelle Open Smile, ce qui améliore le taux de reconnaissance de la parole. Pour vérifier cela, dans le prochain chapitre, nous allons appliquer des classifieursà ce type de caractéristiques.

IV.1 Introduction

Après avoir expliqué et détaillé l'extracteurdes caractéristiques du signal vocal, nous aborderons dans ce chapitre quelques algorithmes de la classification tels que l'Extrême Learning Machine (ELM),le Support Vector Machine(SVM)et l'arbre dedécision.

Dans notre travailnousnousintéressons par l'ELM en premier lieu, puis nous allonsintroduitle SVM et l'arbre de décision pour comparer les résultats par la suite.

Avant d'entamer l'étude sur l'ELM, une petite étude sur les réseaux de neuronesest nécessaire.

IV.2Généralité sur les réseaux de neurones

Le réseau de neuronespeut être décrit par le nombre de couches et le nombre de neurones dans chaque couche [29],chaque neurone possède des entrées et une sortie. Onpeut retenir deux caractéristiques essentielles .La première consiste à la tâche effectué par le réseau, elle est décomposée en tâches élémentaires et réalisées par des neurones, ces derniers peuvent être organisés en des couches reliés entre eux. La seconde caractéristique c'est que le réseau de neurone est adaptatif, tel que chaque neurone contient des paramètres qui peuvent être modifiés, et servent à adapter le réseau à une tâche particulière. Ces modifications sont faites lors d'une phase appelée apprentissage du réseau.

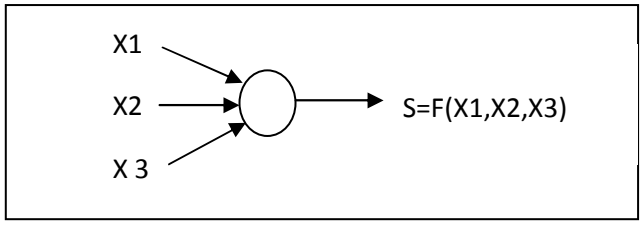
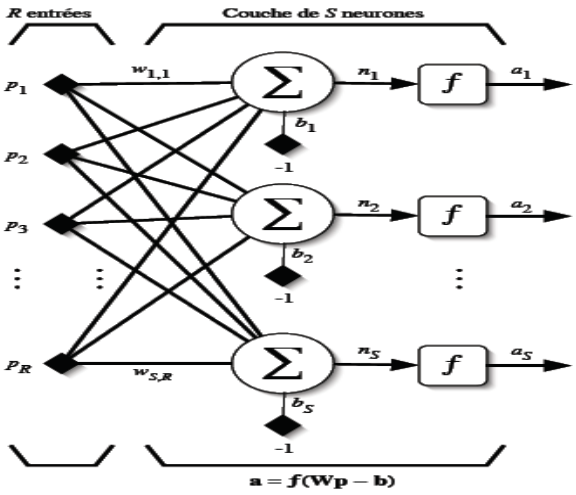


Figure IV.1: Model représentatif d'un neurone

Les réseaux de neurones sont utilisés dans des domaines très variés tels que la reconnaissance d'écriture manuscrite, le traitement d'images, et les télécommunications …. On peut distinguer deux types de réseaux : les réseaux monocouches et les réseaux multicouches :

- **Les réseaux monocouches**

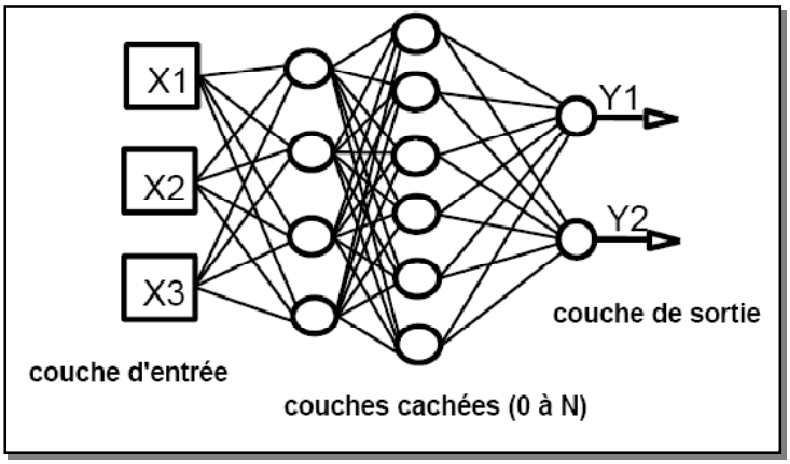
On dit qu'un réseau est monocouche si les neurones d'entrée sont entièrement connectés aux neurones de sorties, que l'on appelle aussi perceptron. Le réseau monocouche possède une structure comme celle représentée dans la figureIV.2.



FigureIV.2: Réseau monocouche.

- **Les Réseaux multicouches (ou MLP)**

Le model à multicouches est le plus simple et le plus connu des réseaux de neurones,il possède une structure comme celle représentée dans la figureIV.3.



FigureIV.3 :Réseau multicouches.

Les caractéristiques d'un tel réseau sont les suivantes :

- La topologie est formée de plusieurs couches de neurones sans communication à l'intérieur d'une même couche.
- Une couche d'entrée qui représente les données à traiter en provenance d'une source extérieure au réseau.
- Une ou plusieurs couches cachées effectuant le traitement spécifique du réseau.
- Une couche en sortie qui délivre les résultats.
- Chaque neurone de chaque couche possède une liaison avec tous les neurones de la couche suivante.

IV.3 L'étude de l'ELM

IV.3.1Généralité sur ELM

Le principe du réseau neuronal n'est pas modifié, mais le rôle del'adaptation est reconsidéré. Plutôt que d'ajuster tous lespoids d'un réseau relativement petit pour émuler unefonction, le réseau est constitué d'un grand nombre de neurones dans la couche interne. Les poids d'entrée sontinitialisés aléatoirement et restent constants.L'adaptation du réseau porte donc uniquement sur lespoids de la couche de sortie. La phase d'apprentissage estainsi grandement simplifiée puisque l'ajustement des poidsde sortie peut s'exprimer à l'aide de règles linéaires.Malgré la taille des réseaux d'adaptation est rapide etfournit d'excellentes performances.L'un des inconvénients de réseau neuronal est l'étude de temps d'apprentissage. Récemment, Huang et al [33], [34] ont proposé un nouvel algorithme d'apprentissage pour l'unité Couche d'architecture anticipatrice Neural Network appelé Extreme Machine Learning(ELM) qui élimine les problèmes causés par la descente de gradient basée sur un algorithmes tels que Retro propagation appliquées RNA, l'ELM peut réduire considérablement le temps nécessaire pour qu'il forme le réseau de neurones[36] si en comparant par d'autres méthodes.

La machine d'apprentissage extrême(ELM) est une technologie qui a une bonne performance dans des applications de régression et de classification.L'ELMprésente plusieurs avantages : La facilité d'utilisation, une grandevitesse d'apprentissage, la

performance supérieure de la généralisation, utilisé pour plusieurs fonctions d'activations non linéaires et les fonctions du noyau.

IV.3.2Principe de l'ELM :

L'ELM [24],[36] a été initialement proposé pour une seule couche cachée feedforward, la couche anticipatrice des réseaux de neurones, et a été étendu par la suite aux SLFNs généralisées où les neurones de la couche cachée ne pas être semblables [37], [38]. En ELM, la couche cachée n'a pas besoin d'être réglée[40]. La fonction de ELM de sortie pour généralisée SLFNs (dans le cas d'un seul nœud de sortie) est :

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i\,(x) = h(x)\beta \quad (4.1)$$

où : $\beta = [\beta_1, \ldots, \beta_L]^T$ est le vecteur des poids de sortie entre la couche cachée de nœud Let le nœud de sortie ,h (x) = [h_1 (x),. . . , H_L (x)] est la sortie (ligne) vecteur de la couche cachée par rapport à l'entrée x. Pour le classement binaire [40] Applications, la fonction de décision d'ELM est :

$$f_L(x) = sign(h(x)\beta) \quad (4.2)$$

Par rapport aux différents algorithmes d'apprentissages traditionnelles [43], l'ELM non seulement peut atteindre l'erreur de la formation la plus petite mais aussi la plus petite norme de pondération de sortie. Selon la théorie de Bartlett [44], pour les réseaux de neuronesatteint la plus petite erreur de formation, plus les normes de poids, mieux performance de généralisation des réseaux ont tendance à avoir[41], [42]. L'ELM minimise l'erreur de formation de même que la norme du signal de sortie poids [45], [46] donc il minimise :

$$\|\mathrm{H}\beta - T\|^2 \qquad and \qquad \|\beta\| \quad (4.3)$$

Où : H est la matrice de sortie de la couche cachée

$$\mathrm{H} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_N(x_N) & \cdots & h_L(x_1) \end{bmatrix} \quad (4.4)$$

Afin de minimiser la norme des coefficients de pondération de sortie$\|\beta\|$est en fait de maximiser la distance de séparation de la marges des deux classes différentes dans la fonction espace ELM: $^{2}/_{\|\beta\|}$

La norme minimale de la méthode des moindres carrés a été utiliséeà la place de la norme Procédé d'optimisation dans la mise en œuvre initiale ELM [45], [46] :

$$\beta = \mathrm{H}^{+}\,\mathrm{T} \tag{4.5}$$

OùH^{+} est l'inverse de Moore-Penrose généralisée de la matrice H [47], [48]. Différentes méthodes peuvent être utilisées pour calculer l'Inverse de Moore-Penrose généralisée d'une matrice: orthogonal Procédé de projection, un procédé d'orthogonalisation, méthode itérative, et la décomposition en valeurs singulières (SVD) [26]. La méthode de la projection orthogonale[48] peut être utilisée dans deux cas: lorsqueH^{T}Hest inversible$\mathrm{H}^{+} = (\mathrm{H}^{\mathrm{T}}\mathrm{H})^{-1}\mathrm{H}^{\mathrm{T}}$ oulorsque H^{T}Hest non inversible $\mathrm{H}^{+} = \mathrm{H}^{\mathrm{T}}(\mathrm{H}^{\mathrm{T}}\mathrm{H})^{-1}$

Selon la théorie de crête de régression [49], on peut ajouterune valeur positive à la diagonale de H^{T}H ouHH^{T}; lasolution résultanteest plus stable et tend à avoir une meilleure performance degénéralisation. [51],[50]

- **La fonction universelle d'approximation Capacité**:

Selon lathéorie d'apprentissage ELM, un type répandu de longs mappages h (x)peut être utilisé de sorte qu'il peut rapprocher tout les cibles des Fonctions continues (voir [24] ,[39] pour les détails). Autrement dit, pour donnern'importe quelle cible d'unefonction continue f (x), il existe une série de ß$_{\mathrm{i}}$ de telle sorte que :

$$\lim_{L\to\infty}\|\mathrm{f_L}(\mathrm{x}) - \mathrm{f}(\mathrm{x})\| = \lim_{L\to\infty}\left\|\sum_{\mathrm{i}=1}^{\mathrm{L}} \beta_i h_i\,(x) - \mathrm{f}(\mathrm{x})\right\| = 0 \tag{4.5}$$

- **La classification des fonctions généralisée:**

A partir de théorème de la classification des fonctions à simple-couche cachée de neurones feedforward réseaux [30], nous pouvons prouver la capacité de classification de les SLFNs généralisées avec la couche cachée cartographie h (x)satisfaisant à la condition de rapprochement universel.

Le théorème des fonctions de classification de Huang et al [52] peut être étendu à SLFNs généralisées les neurones qui ne sont pas forcement les mêmes.L'optimisationd'ELM proposée, avec un nœud unique en sortie peut être formulée sous la forme :

$$L_{P_{ELM}} = \frac{1}{2}\|\beta\|^2 + c\frac{1}{2}\sum_{i=1}^{N} \varepsilon_i{}^2 \quad (4.6)$$

$$h(x_i)\beta = t_i - \varepsilon_i i = 1, \ldots\ldots\ldots\ldots, N. \quad (4.7)$$

Nous nous basons sur le théorème de KKT, formant ELM est équivalent à résoudre le problème d'optimisation dual suivant:

$$L_{D_{ELM}} = \frac{1}{2}\|\beta\|^2 + c\frac{1}{2}\sum_{i=1}^{N} \varepsilon_i{}^2 - \sum_{i=1}^{N} \alpha_i (h(x_i)\beta - t_i + \varepsilon_i) \quad (4.8)$$

Où chaque multiplicateur de Lagrange α_i correspond à l'ième échantillon d'apprentissage. Nous pouvons avoir les conditions d'optimalité KKT [18]comme suit:

$$\frac{\partial L_{D_{ELM}}}{\partial \beta} = 0 \rightarrow \beta = \sum_{i=1}^{N} \alpha_i h(x_i)^T = H^T \alpha \quad (4.9)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \varepsilon_i} = 0 \rightarrow \alpha_i = C\varepsilon_i \quad , \ i = 1, \ldots, N \quad (4.10)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \alpha_i} = 0 \rightarrow h(x_i)\beta - t_i + \varepsilon_i = 0 \ , \ i = 1, \ldots, N \quad (4.11)$$

Où$\alpha = [\alpha_1, \ldots\ldots, \alpha_N]^T$

Lemulti classificateur Avec Multi-sortie: Une approche alternative pour les applications multi-classés c'est d'appliquer l'ELM ennœudsmulti-output au lieu d'un nœud unique en sortie. Avec m-class de classificateurs et m nœuds de sortie. Si le label de la classe d'origine est p, le vecteur de sortie prévue des nœuds m de sortie est :

ti = [0,. . . , 0, p 1, 0,. .0]^T. Dans ce cas, l'élémentPième du ti = [$t_{i,1}$,... , $t_{i,\,m}$] ^{T}c'est le seul qui égal à un, alors que le reste des éléments sont mis à zéro. Le problème de classification pour ELM avec multinœuds de sortie peuvent être formulé sous la forme :

$$L_{P_{ELM}} = \frac{1}{2}\|\beta\|^2 + c\frac{1}{2}\sum_{i=1}^{N}\|\varepsilon_i\|^2 \quad (4.12) h(x_i)\beta = t_i{}^T - \varepsilon_i{}^T i$$

$$= 1, \ldots\ldots\ldots, N. \quad (4.13)$$

Où $\varepsilon_i = [\varepsilon_{i,1}, \ldots\ldots, \varepsilon_{i,m}]^T$ est le vecteur d'erreur de formation des m nœuds de sortie par rapport à l'échantillon d'apprentissage x_i.

En utilisantsur le théorème KKT, formantl'ELM c'est équivalent à résoudre le problème d'optimisation duale suivant:

$$L_{P_{ELM}} = \frac{1}{2}\|\beta\|^2 + c\frac{1}{2}\sum_{i=1}^{N}\|\varepsilon_i\|^2 - \sum_{i=1}^{N}\sum_{j=1}^{m}\alpha_{i,j}\,(h\,(x_i)\beta_j - t_{i,j} + \varepsilon_{i,j}) \quad (4.14)$$

Où βj est le vecteur des coefficients de pondération de liaison de la couche cachée j-ièmenœud de sortie et la β = [β1,. . . , ßm]. Nous pouvons avoir la KKT conditions d'optimalité correspondant comme suit:

$$\frac{\partial L_{D_{ELM}}}{\partial \beta_j} = 0 \rightarrow \beta_j = \sum_{i=1}^{N}\alpha_{i,j}\,h\,(x_i)^T \rightarrow \beta = H^T\alpha \quad (4.15)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \varepsilon_i} = 0 \rightarrow \alpha_i = C\varepsilon_i \quad , \; i = 1, \ldots, N \quad (4.16)$$

$$\frac{\partial L_{D_{ELM}}}{\partial \alpha_i} = 0 \rightarrow h\,(x_i)\beta - t_i{}^T + \varepsilon_i{}^T = 0 \;, \; i = 1, \ldots, N \quad (4.17)$$

Où $\alpha_i = [\alpha_{i,1},. . . , \alpha_{i,m}]^T$ et $\alpha = [\alpha_1,. . . , \alpha_N]^T$.

Un seul nœud sortie considérée comme un cas particulier de multi-output nœuds lorsque le nombre de nœuds de sortie est fixé à m = 1, nous n'avons besoin que de considérer classificateur la multi classent de nœuds de multioutput. Pour les deux cas, la matrice H (4.4) reste le même, et sa taille est décidé par le nombre d'échantillons d'apprentissage et le nombre N des nœuds cachés L, ce qui est en rapport avec le nombrenœuds de sortie (nombre de classes).

Égalité à contrainte-optimisation basée ELM,différentes solutions aux conditions KKT précités peuvent être obtenus sur la base des préoccupations concernant l'efficacité en taille différente des ensembles de données d'entraînement.

1-Dans le cas où le nombre d'échantillons de formation n'est pas grand: en remplaçant (4.15) et (4.16) dans (4.17), les équations ci-dessus peuvent être écrites de manière équivalente Comme :

$$\left(\frac{1}{C} + H^T H\right)\alpha = T \quad (4.18)$$

$$\mathrm{T} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h_1(x_1) & \cdots & h_L(x_1) \\ \vdots & \ddots & \vdots \\ h_N(x_N) & \cdots & h_L(x_1) \end{bmatrix} \quad (4.19)$$

De (4.15) et(4.20) , nous trouvins:

$$\beta = \mathrm{H}^{\mathrm{T}} \left(\frac{1}{C} + \mathrm{H}\mathrm{H}^{\mathrm{T}}\right)^{-1} \mathrm{T} \quad (4.20)$$

La fonction de sortie d'ELM classificateur est :

$$f(x) = (h(x)\mathrm{H}^{\mathrm{T}} \left(\frac{1}{C} + \mathrm{H}\mathrm{H}^{\mathrm{T}}\right)^{-1} \mathrm{T}) (4.21)$$

Nœud simple sortie (m = 1): Pour des classifications multi-classés, parmi toutes les labels multi-classés, la classe prédite d'un échantillon d'essai donné est plus proche de la sortie d'un Classificateur ELM. Pour le cas binaire de classification, l'ELM doit avoir un seul nœud de sortie (m = 1), et la fonction de décision d'ELM classificateur est :

$$f(x) = sign(h(x)\mathrm{H}^{\mathrm{T}} \left(\frac{1}{C} + \mathrm{H}\mathrm{H}^{\mathrm{T}}\right)^{-1} \mathrm{T}) \quad (4.22)$$

2- Pour le cas où le nombre d'échantillons de formation est très grand: le nombre de données de formation est très important, par exemple, il est beaucoup plus grand que la dimension de l'espace des attributs. De (4.15) et (4.16), nous avons :

$$\beta = C\mathrm{H}^{\mathrm{T}}\varepsilon \quad (4.23)$$

$$\varepsilon = \frac{1}{C}(\mathrm{H}^{\mathrm{T}})^{+}\beta (4.24)$$

De (4.17),nous avons:

$$H\beta - \mathrm{T} + \frac{1}{C}(\mathrm{H}^{\mathrm{T}})^{+}\beta = 0 \quad (4.25)$$

$$H^T(H + \frac{1}{C}(H^T))^{+}\beta = H^T T (4.26)$$

$$\beta = \left(\frac{1}{C} + \mathrm{H}\mathrm{H}^{\mathrm{T}}\right)^{-1} \mathrm{H}^{\mathrm{T}}\mathrm{T} (4.27)$$

Dans ce cas, la fonction de sortie d'ELM classificateur est :

$$f(x) = h(x)\beta = h(x) \left(\frac{1}{C} + \mathrm{H}\mathrm{H}^{\mathrm{T}}\right)^{-1} H^T T (4.28)$$

IV.4 Les SVMs :

IV.4.1Généralité sur les SVM (Support Vector Machines):

Les machines à vecteurs de support ou séparateurs à vaste margesont des ensembles de techniques d'apprentissage supervisés destinées à résoudre des problèmes de discrimination[28] et de régression.

Les SVMs sont des généralisations des classifieurs linéaires ,ils ont été développés à partir des considérations théoriques de Vladimir Vapnik [27] sur le développement d'une théorie statistique d'apprentissage « la Théorie de Vapnik-Chervonenkis »[25],Ils ont été adoptés à travailler avec des données de grandes dimensions, leur faible nombre d'hyper paramètres, leurs garanties théoriques, et leurs bons résultats en pratique[25].

Les SVMsont été largement utilisés dans des applications très répondues. Cette façon de l'optimisation standard est utilisée pour trouver la solution de maximisation de marge pour la séparation de deux différentes classes tout en minimisant les erreurs d'apprentissage,ils ont été appliqués dansplusieurs domaines (bio-informatique, recherche d'information, vision par ordinateur, finance[23]...). Selon les données, la performance des machines à vecteurs de support est de même ordre, ou supérieure, à celle d'un réseau de neurones ou d'un modèle de mixture gaussienne.

IV.4.2 Principe des SVMs

Les SVMs sont des classificateurs qui reposent sur deux idées principales, ils permettent de traiter des problèmes de discrimination non-linéaire, et de reformuler le problème de classement comme un problème d'optimisation quadratique[24].
Les SVM peuvent être utilisés pour résoudre des problèmes de discrimination, c'est-à-dire décider à quelle classe appartient un échantillon, ou de régression, c'est-à-dire prédire la valeur numérique d'une variable. La résolution de ces deux problèmes passe par la construction d'une fonction h qui à un vecteur d'entrée x fait correspondre une sortie y :

$$y = h(x) \quad (4.29)$$

On se limite pour l'instant à un problème de discrimination à deux classes (discrimination binaire), c'est-à-dire $y \in \{-1,1\}$, le vecteur d'entrée x étant dans un espace $x$muni d'un produit scalaire. On peut prendre par exemple$x=R^N$ [25].

Le cas simple est le cas d'une fonction discriminante linéaire, obtenue par combinaison linéaire du vecteur d'entréex=(x1,.....,x_N)^T(4. 30)

avec un vecteur de poids :x=(x1,.....,x_N)^T(4.31)

$$h(x) = w^T x + w_0 \text{(4.32)}$$

Il est alors décidé que x est de classe 1 si $h(x) \geq 0$ et de classe -1 sinon. C'est un classifieur linéaire.

La frontière de décision $h(x) = 0$ est un hyperplan, appelé hyperplan séparateur, ou séparatrice. Le but d'un algorithme d'apprentissage supervisé est d'apprendre la fonction h(x) par le biais d'un ensemble d'apprentissage :

$$\{(x_1, l_1), (x_2, l_2), \dots\dots\dots, (x_p, l_p)\} \subset R^N \times \{-1,1\} \text{(4.33)}$$

Où les l_k sont les labels, p est la taille de l'ensemble d'apprentissage, N la dimension des vecteurs d'entrée. Si le problème est linéairement séparable, on doit alors avoir :

$l_k h(x_K) \geq 0 \quad 1 \leq k \leq p$Autrement dit :$l_k(w^T x_K + w_0) \geq 0 \quad 1 \leq k \leq p$(4.34)

Pour construire un bon modèle d'apprentissage, il est nécessaire de construire un système capable à la généralisation correcte et minimisation des erreursLe classifieur linéaire est donnée par :y(x) = signe(w · x + b)(4.35)

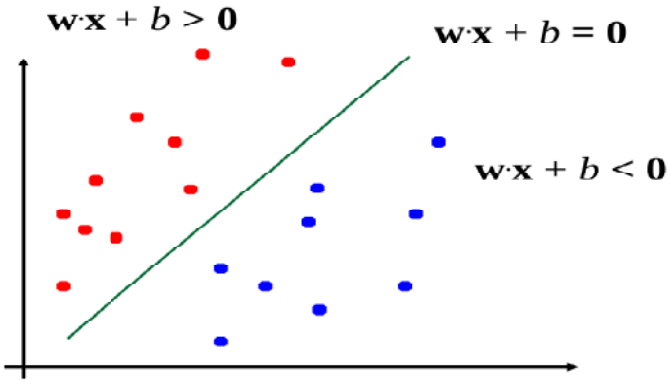
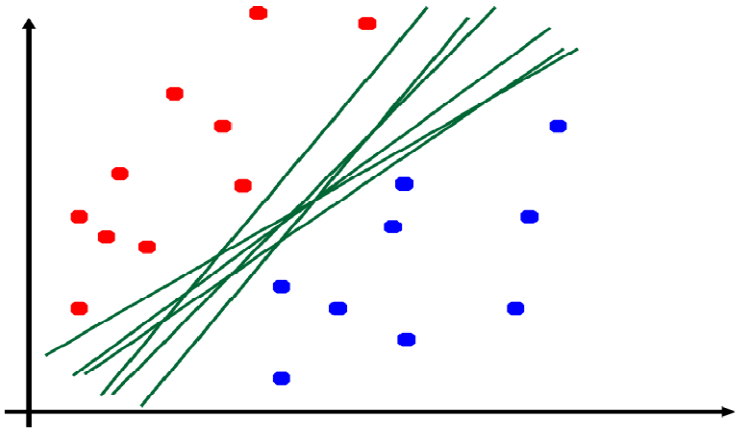


Figure IV.4 : La représentation graphiqued'un classifieurlinéair

Il existe de nombreux choix possibles pour w et b :

$$y(x) = \text{signe}(w \cdot x + b) = \text{signe}(kw \cdot x + k \cdot b) \text{(4.36)}$$



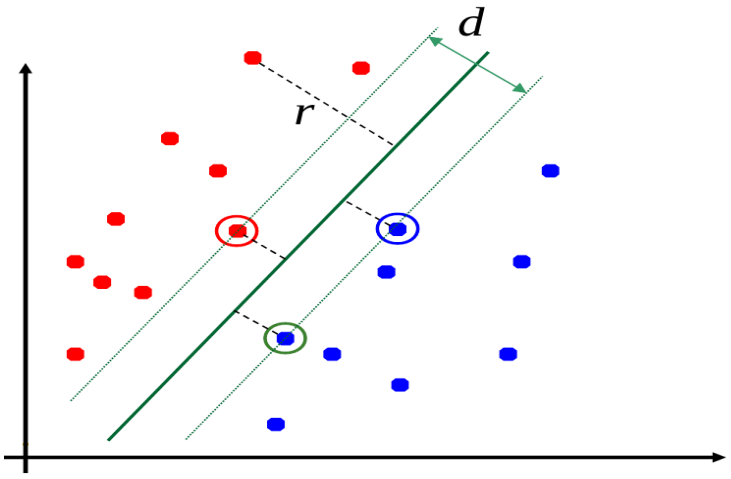
FigureIV.5 : Classifieur linéaire à plusieurs choix possible

La notion de marge maximale signifie la distance entre la frontière de séparation et les échantillons les plus proches. Ces derniers sont appelés *vecteurs supports*. Dans les SVM, la frontière de séparation est choisie comme celle qui maximise la marge. Ce choix est justifié par la théorie de Vapnik-Chervonenkis (ou théorie statistique de l'apprentissage), qui montre que la frontière de séparation de marge maximale possède la plus petite capacité [10].Le problème est de trouver cette frontière séparatrice optimale, à partir d'un ensemble d'apprentissage. Ceci est fait en formulant le problème comme un problème d'optimisationquadratique, pour lequel il existe des algorithmes connus.

Pour tout point de l'espace desexemples, la distance `al'hyperplan séparateur est donnéepar :

$$r = \frac{|w \cdot x + b|}{\|w\|} \quad (4.37)$$

On appelle marge d la distance entre les deux classes,c'est cette distance d qu'on souhaiterait maximiser.



FigureIV.6 : Représentation de la marged'un classifieur

Pour limiter l'espace des possibles on consid`ere que les points les plus proches sont situés sur les hyperplans canoniques données par : $w \cdot x + b = \pm 1$ (4.38)

Dans ce cas, la marge est $d = \frac{2}{\|w\|}$(4.39)

Les conditions d'une bonne classification sont :

$$\begin{cases} w \cdot x + b \geq 1, si\ y_i = 1 & (4.40) \\ w \cdot x + b < 1, si\ y_i = -1 & (4.41) \end{cases}$$

Le problème revient alors à trouver w et b tels que $d = \frac{2}{\|w\|}$est maximale ∀(xi , yi)

Sous les contraintes : (1)et (2)

De manière équivalent, le problème peut s'écrire tous simplement comme la minimisation de :$\frac{1}{2}\|w\|^2$(4.42)

Sous les contraintes :$y_i\ (w \cdot x + b) \geq 1, \forall i \in [1,N]$

Cette minimisation est possible sous la condition dites « Karush-Kuhn-Tuker (KKT) » [26] :

Soit le Lagrangien L :$\mathcal{L}(w,b,\lambda) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \lambda_i[(y_i\ w \cdot x + b) - 1]$(4.43)

Les conditions de KKT sont alors :

$$\frac{\delta \mathcal{L}}{\delta w} = 0 \quad , \quad \frac{\delta \mathcal{L}}{\delta b} = 0 \quad , \frac{\delta \mathcal{L}}{\delta \lambda_i} \geq 0 \quad , \lambda_i \geq 0$$

$\lambda_i[y_i(\mathrm{w}\cdot \mathrm{x}+\mathrm{b})-1]=0$(4.44)

Par ailleurs la dernière condition implique que pour tout point ne vérifiant pasy_i (w · xi + b) = 1 leλ_i est nul.Les points qui vérifient y_i (w · x_i + b) = 1, sont appelés "vecteurssupports". Ce sont les points les plus prés de la marge. Ils sontsensésà êtremoins nombreux par rapport à l'ensemble des exemples.

•Classification à marge souple :

En général, il n'est pas possible de trouver un séparateur linéaire dans l'espace de données. Il se peut trouver que des échantillons qui sont mal étiquetés, et que l'hyperplan séparateur ne soit pas la meilleure solution au problème de classement,et donc si les donnée ne sont pas linéairement séparable, L'idée est d'ajouter des variables d'ajustement $\varepsilon_i$dans laformulation pour prendre en compte les erreurs de classification ou le bruit

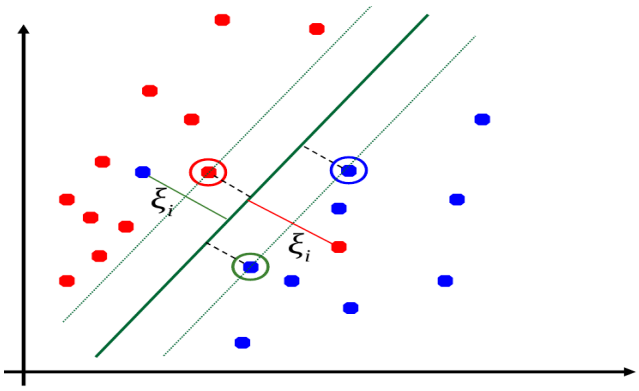


Figure IV.7 : la marge souple en prenant en compte les erreurs

Corinna Cortes et Vladimir Vapnik proposent une technique dite de marge souple[30][35], qui tolère les mauvais classements. La technique cherche un hyperplan séparateur qui minimise le nombre d'erreurs grâce à l'introduction de variables ressortε_k (*slack variables* en anglais), qui permettent de relâcher les contraintes sur les vecteurs d'apprentissage :

$$l_k(w^T x_K + w_0) \geq 1 - \varepsilon_k \varepsilon_k \geq 0, 1 \leq k \leq p \quad (4.45)$$

Avec les contraintes précédentes, le problème d'optimisation est modifié par un terme de pénalité, qui pénalise les variables ressort élevées :

$$\text{Minimiser} \frac{1}{2}\|w\|^2 + \mathrm{c}\sum_{\mathrm{k}=1}^{\mathrm{p}} \varepsilon_k \ , \quad \mathrm{C} > 0 \quad (4.46)$$

Il existe de nombreux algorithmes de résolution les problèmes d'optimisation quadratique [26] :

- **SMO :** résolution analytique (par 2 points), gestion efficace de lamémoire, mais converge en un nombre d'étapes indéterminé
- **SimpleSVM :** facilite de la reprise a chaud, converge en moins d'étapes maislimitation mémoire
- **LASVM :** utilisation en ligne, résolution analytique mais solution sous optimale, plusieurs passes nécessaires pour les petites bases dedonnées.

IV.5 Arbre de décision

IV.5.1Généralité sur l'arbre de décision

Un arbre de décision est un modèle de classification présente sous la forme graphique d'un arbre. L'extrémité de chaque branche est une feuille qui présente le résultat obtenu en fonction des décisions prises a partir de la racine de l'arbre jusqu'a cette feuille. Les feuilles intermédiaires sont appelées des nœuds. Chaque nœud de l'arbre contient un test sur un attribut qui permet de distribuer les données dans les différents sous-arbres. Lors de la construction de l'arbre un critère de pureté comme l'entropie est utilisépour transformer une feuille en nœud. L'objectif est de produire des groupes d'individus les plus homogènespossibles du point de vue de la variable àprédire. En prédiction, un exemple àclasser "descend" l'arbre depuis la racine jusqu'à une unique feuille. Son trajetdans l'arbre est entièrementdétermine par les valeurs de ses attributs. Il est alorsaffecte a la classe dominante de la feuille avec pour score la proportion d'individusdans la feuille qui appartiennent a cette classe. [9]

Les arbres de décisionpossèdent les avantages suivant :

• La lisibilité du modèle.

• La capacitéà trouver les variables discriminantes dans un important volume de données.
Les algorithmes de références de la littérature sont ID3, C4.5, CART mais ils nesont pas incrémentaux.

Des versions incrémentales des arbres de décision sont assez rapidementapparues. Schlemmer et Fisher, 1986 propose ID4 et Utgoff, 1989 propose ID5Rqui sont bases sur ID3 mais dont la construction est incrémentale. ID5R garantit laconstruction d'un arbre

similaire à ID3 alors qu'ID4 peut dans certains cas ne pasconverger et dans d'autre cas avoir une prédictionmédiocre.

La lisibilité des arbres ainsi que leur rapidité de classement en font un choixtrès pertinent pour une utilisation sur d'importantes quantités de données.

Cependant les arbres ne sont pas très adaptés aux changements de concept cardans ce cas d'importantes parties de l'arbre doivent êtreélaguées et réapprises. [31]

IV.5.2Construction de l'arbre :

Un arbre de décision est un arbre binaire dans lequel :

-Un nœud interne est associe a une variable, parmi un ensemble V de variables ;

- Une feuille est associéea un booléen (vrai ou faux).

Si chaque variable de l'ensemble V reçoit une valeur booléenne, un tel arbrepermet de prendre une décision en parcourant l'arbre :

-On part de la racine.

-Quand on arrive sur un nœud interne (racine comprise), on regarde quelle est lavaleur de la variable associée au nœud : si elle vaut vrai on poursuit le parcoursdans le sous-arbre gauche, sinon on poursuit le parcours dans le sous-arbre droit ;

-Quand on arrive sur une feuille, le booléen associe constitue la décision.

Les nœuds représentent les attributs, les branches portent les valeurs, et lesfeuilles nous donnent les décisions (classes)

La taille d'un arbre de décision est donnée par le nombre de nœuds qui lereprésentent.
Pour construire un nœud dans l'arbre, on cherche parmi les attributs celui qui a lameilleure entropie.

- **L'Entropie :** L'entropie est une mesure de l'incertitude associéea un échantillonaléatoire

variable, est mesurée par :

Entropie(S)=-$\sum p_i log_i p_i$(4.47)

- **Le gain:**Le gain nous donne une information de mesure, il est raccordé avec lesattribues, est calculé par :

$$\text{gain}(S ; A)=\text{Entropie}(S)-\sum \frac{|s_k|}{|s|}\text{Entropie}(s_k) \quad (4.48)$$

S_K : est la division d'un exemple dans S pour l'attribue A qui a la valeur K pour tester et arriver à la classification.

Nous nous sommes intéresses au logiciel "Weka" qui sera définit dans le prochain chapitre (Chapitre V)et qui nous a permets dans ce cas, de construirenotre arbre de décisionpour faire une classification pour but de comparer les résultats pour celle de l'ELM par la suite.

IV.6 Conclusion

Dans ce chapitre nous avons introduit trois Algorithmes de classification, deux d'entre eux seront utilisé grâce à l'outil « Weka » il s'agira du SVM et « l'Arbre de décision » et le troisième « l'ELM » qui sera implémenter sur MATLAB. Ces derniers technique seront simulés et étudiées sur une base de données enregistré au département de l'oncologie de la tête et du cou et à l'institut Néerlandais de la chirurgie du cancer [53] « NKI CCRT Speech Corpus » (NCSC).

V.1 Introduction

Dans ce chapitre, nous allons présenter en premier lieu le système de reconnaissance ainsi la base de données utilisées, et par la suite nous parlerons de l'extracteur de paramètres Open Smile sous Windows et les fichiers de types ARFF, à la fin nous parlerons de la comparaison entre la classification utilisant le logiciel Weka (SVM et l'arbre de décision) et celle avec l'ELM sous Matlab pour tirer lequel des deux est le meilleur.

V.2Architecture du système de reconnaissance

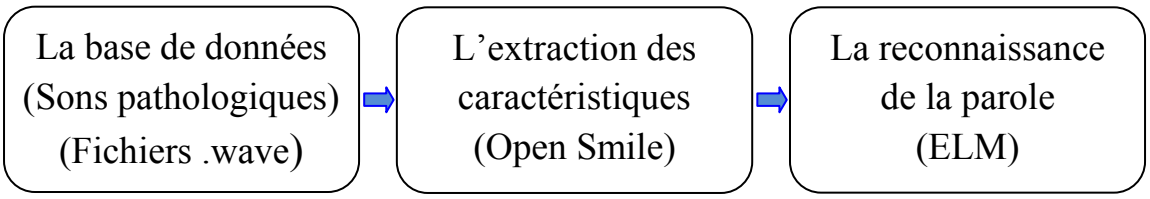


Figure V.1 : Schéma global du système de reconnaissance.

V.2.1 Présentation du logiciel

- La programmation a été faite sous MATLAB version 7.14.0.739 à 64 bits qui est un environnement de calcul technique conçu pour le calcul numérique et la visualisation à haute performance.

MATLAB offre des familles d'applications réunis ou sauvegardées dans des boites à outils, donc il s'avère être le logiciel le plus adéquat pour nos applications de par les avantages qu'il nous a offerts en termes de traitement et l'extraction des paramètres que nécessite notre étude ou par la facilité qu'elles engendrent les fonctions pour la classification.

- D'autre part, nous avons utilisés Weka (Waikato Environment for KnowledgeAnalysis) qui est un logiciel libre dédié au Data Mining(exploration de données). Parmi les fonctionnalités qu'il couvre, nous nous intéresserons aux arbres de décision ainsi que le SVM. Il permet de modéliser simplement, graphiquement et rapidement un phénomène mesuré plus ou moins complexe. Sa lisibilité, sa rapidité d'exécution et le peu d'hypothèses nécessaires à priori expliquent sa popularité actuelle.

L'installation de WEKA est facile et rapide, il suffit juste d'exécuter l'exe précédemment téléchargé. Aucune exigence n'est notée sur le lieu d'installation. Une fois installé, il peut être lancé à partir du menu démarrer, en cliquant sur Wekala fenêtre suivante s'ouvre :

Figure V.2 : Fenêtre de Weka.

Les formats de fichiers supportés par WEKA sont l'arff et le csv. Le format de fichier le plus utilisé sous WEKA étant l'arff. Généralement les données sont stockées dans une base de données relationnelle.

Pour se connecter au fichier arff, cliquer sur **Explorer** sur la première fenêtre apparue lors du lancement de WEKA.Ensuite, cliquer sur **Open file**... de l'onglet **Preprocess**.

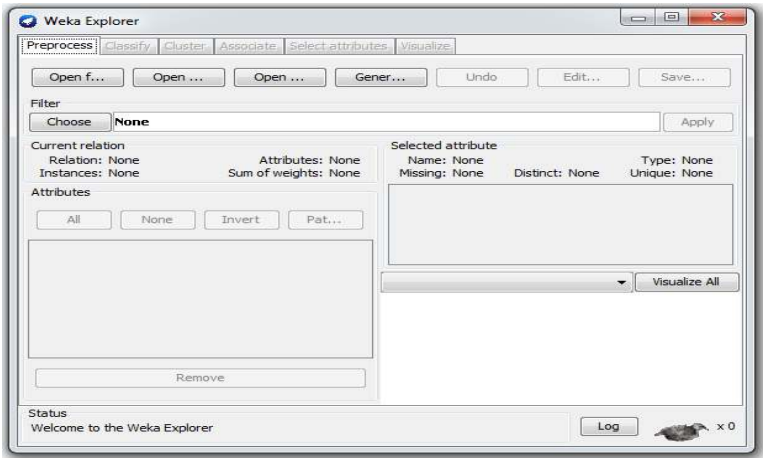


Figure V.3 : Fenêtre d'explorer de Weka

Dans la fenêtre qui s'ouvre choisir le fichier de type Arff et cliquer sur ouvrir.

Vous aurez un aperçu du genre :

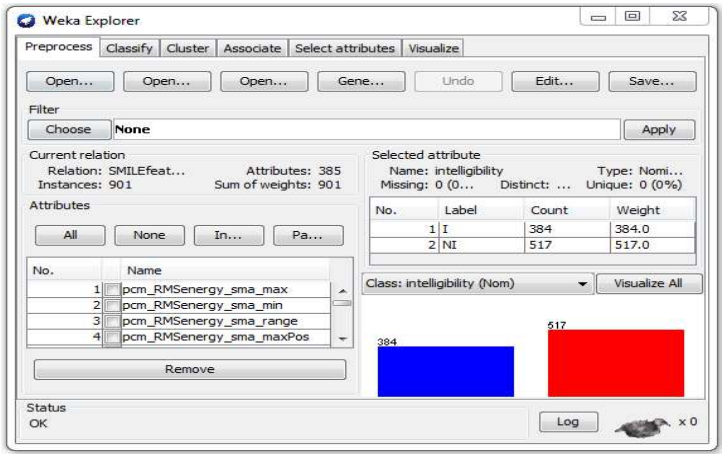


Figure V.4 : Fenêtre de process de Weka.

Pour établir l'arbre de décision cliquer sur l'onglet **Classify**, choisir l'option **Use training set** de **Test options** comme suit :

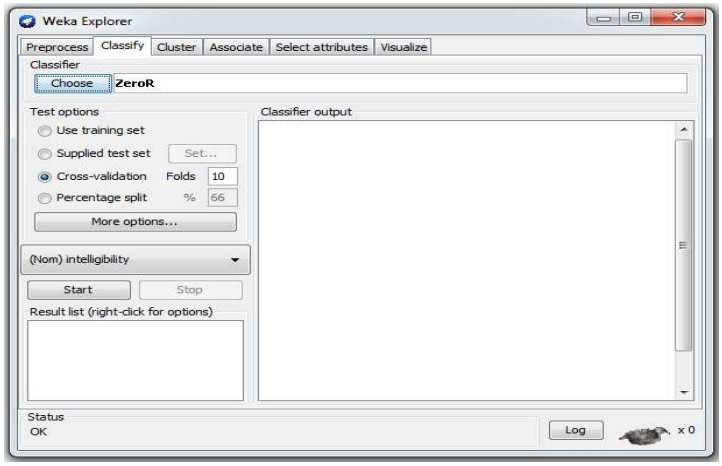


Figure V.5 : Fenêtre« classify » de Weka.

La zone Test options permet de choisir de quelle façon l'évaluation des performances du modèle appris se fera.

- L'option **Use training set** utilise l'ensemble d'entraînement pour cette évaluation.
- L'option **Supplied test set** va utiliser un autre fichier.
- Lorsque l'option **Cross-validation** est sélectionnée, l'ensemble d'apprentissage est coupé en 10 (si Folds vaut 10). L'algorithme va apprendre 10 fois sur 9 parties et le modèle sera évalué sur le dixième restant. Les 10 évaluations sont alors combinées.

- Avec l'option **Percentage split**, c'est un pourcentage de l'ensemble d'apprentissage qui servira à l'apprentissage et l'autre à l'évaluation.

Ensuite, cliquer sur le bouton **Choose** de **Classifier** pour choisir un algorithme parmi ceux proposés par WEKA et Cliquer sur**Start** pour effectuer l'analyse.

Dans la partie **Classifier output** vous avez des statistiques sur le fichier exploité, à savoir le nombre d'instances **Total Number of Instances** de votre fichier, le nombre d'instances correctement classifiées **CorrectlyClassified Instances** et incorrectement classifiées **IncorrectlyClassified Instances** et autres statistiques à découvrir.

Sur le même écran vous avez aussi la **matrice de confusion** : **Confusion Matrix** de cette analyse.

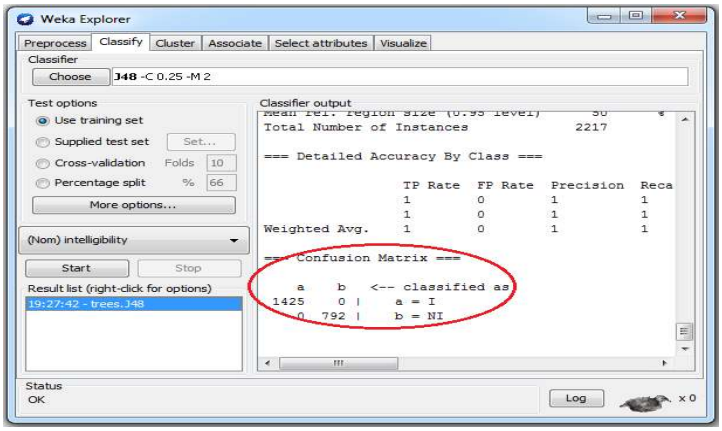


Figure V.6 : Les sorties de la classification.

Pour afficher l'arbre de décision, cliquer droit dans la partie **Resultlist,** lors de ce clique droit un menu d'options s'affiche, choisir l'option **Visualizetree**, L'arbre de décision s'affiche ressemblant à celle de la figure V.8.

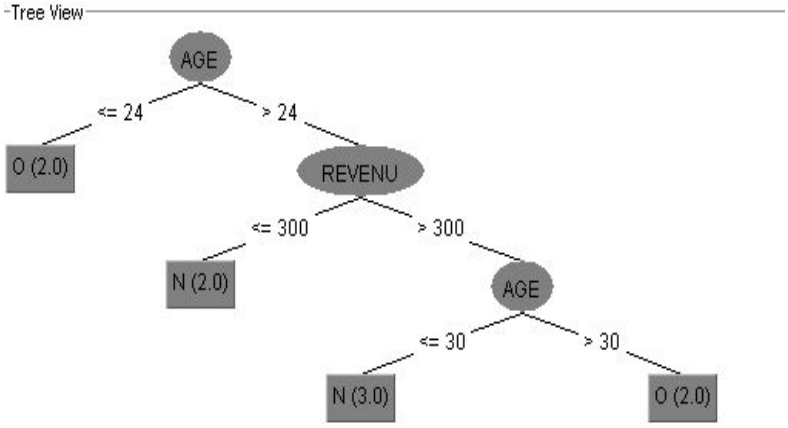


Figure V.7 :L'arbre de décision de Weka.

V.3 La base de données utilisée

Pour la pathologie sous défi, ils ont choisi le « NKI CCRT Speech Corpus » (NCSC) enregistré au département de l'oncologie de la tête et du cou et à l'institut Néerlandais de la chirurgie du cancer [53]. Le corpus contient des enregistrements et des évaluations de perception de 55 orateurs (10 femmes, 45 hommes) qui ont subi un traitement de radio-chimiothérapie concomitante (CCRT) pour les tumeurs inopérables de la tête et du cou. Les enregistrements et les évaluations dans le corpus ont été faits avant et après CCRT : avant CCRT (T0 ; 54 orateurs), 10 semaines après CCRT (T1 ; 48 orateurs) et 12 mois après (T3 ; 39 orateurs). L'âge moyen des orateurs était 57. Pas tous les intervenants n'étaient des néerlandophones. Tous les orateurs ont lus un texte néerlandais de contenu neutre. Treize récemment diplômé ou futurs diplômés sur les orthophonistes (tous de sexe féminin, orateurs d'origine néerlandaise, âge moyen 23,7 ans) ont évalués les enregistrements de la parole dans une expérience en ligne à l'échelle de l'intelligibilité de 1 à 7. Des participants ont été invités à compléter les évaluations dans un environnement calme. Tous les participants ont rempli un module de familiarisation en ligne. Tous les échantillons ont été transcrits manuellement et un alignement automatique de phonèmes a été généré par un dispositif de reconnaissance de la parole formé sur le discours néerlandais utilisant le corpus néerlandais parlé (CGN). La transcription et la phonémisation sont prévus pour les participants. Pour le défi, les échantillons originaux ont été segmentés aux limites de la phrase. Les partitions d'entrainement, de développement et de test ont été obtenues selon l'âge, le sexe et l'autochtonie des orateurs approximativement à un partitionnement de 40%, 30%, 30% (tableau V.1). La corrélation de rang moyen (le rho de spearmans) des évaluations individuelles avec la note moyenne est de 0,783.

NCSC	Train	Devel	Test	Σ
I	384	341	475	1200
NI	517	405	264	1186
Σ	901	746	739	2386

Tableau V.1 : le partitionnement du corpus de la parole NKI CCRT (I: intelligible / NL: non-intelligible).

Selon le sous défi de cordialité (likability), EWE (evaluatorweightedestimator : l'estimateur de l'évaluateur pondéré) a été calculé et discrétisé en étiquettes de classes binaires (intelligible, non intelligible), division à la médiane de la distribution. Noter que l'étiquette de classes des segments de parole ne sont pas exactement équilibrées (1200/1186) depuis la médiane à été prise à partir des notes de discours original non segmenté.

V.4 Extraction de caractéristiques avec Open Smile

Comme nous avons vus dans le chapitre III, Open Smile nécessite la présence de l'exécutable « SMILExtract.exe », la bibliothèque « openSmileLib.dll » ainsi que le fichier de configuration à utiliser, ce dernier définit le nombre de paramètres à extraire et leurs types.

Dans notre travail, nous avons utilisé emo_IS09.confqui donne 384 caractéristiques, a fin de créer trois fichiers de format arff suivant la division de la base de données en trois groupes: Devel, Train et test.

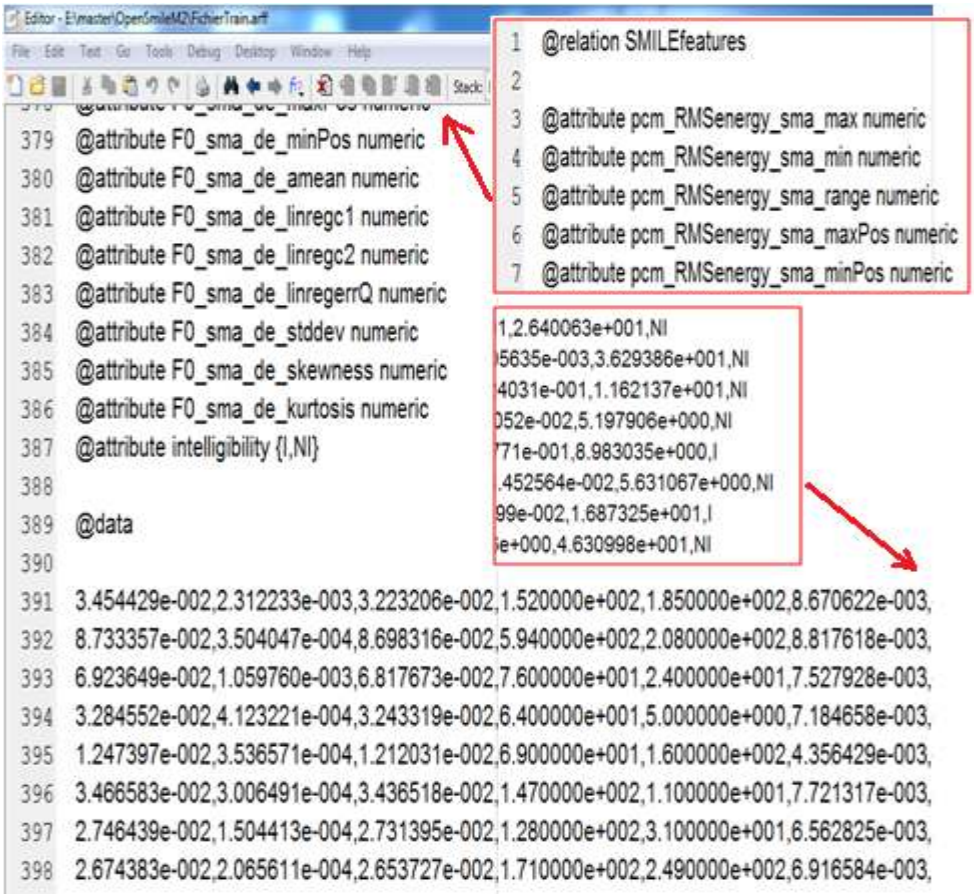


Figure V.8 : Exemple d'un fichier arff.

Les fichiers arff sont utilisés généralement par le logiciel Weka alors que notre but principal est d'injecter ce grand nombre de paramètres dans l'ELM, pour cela nous les avons convertis au format .MAT lus par Matlab.

V.5 L'interprétation des résultats

V.5.1 La classification utilisant l'ELM

Dans la série de tests suivante, nous faisons varier en premier lieu le facteur de régression C et le nombre de neurones pour trouver ces valeurs optimales, afin de les injectés une autre fois dans l'ELM en variant la fonction d'activation. Ces tests sont répétés pour les deux équations qui calculent les poids de sortie:

C optimal	Nbr de neurones optimal	La fonctiond'activation	Précisiond'entrainement (%)	Précision de test (%)
-	60	sigmoid	67,48	56,43
-	85	sine	60,60	50,34
-	60	hardlim	67,48	56,70
-	85	tribas	59,27	36,27
-	35	radbas	61,15	45,60

Tableau V.2 : Implémentation sans facteur de régularisation

C optimal	Nbr de neurones optimal	La fonctiond'activation	Précisiond'entrainement (%)	Précision de test (%)
100	60	sigmoid	67,70	56,29
0,1	85	sine	60,71	50,20
1	60	hardlim	67,92	57,37
10	85	tribas	59,27	36,27
1	60	radbas	62,49	48,17

Tableau V.3 : Implémentation avec facteur de régularisation (fastermethod).

V.5.2La classification utilisant SVM sous Weka

Pour relever les pourcentages de précision d'entrainement et de test, nous faisons varier le facteur de paramètre de complexité et relever les résultats suivants :

C	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1	10	100
Précisiond'entrainement (%)	54.3	54.3	60.9	57.8	57.4	53.2	54
Précision de test (%)	35.7	35.5	59.3	61.4	62.2	57.1	56.2

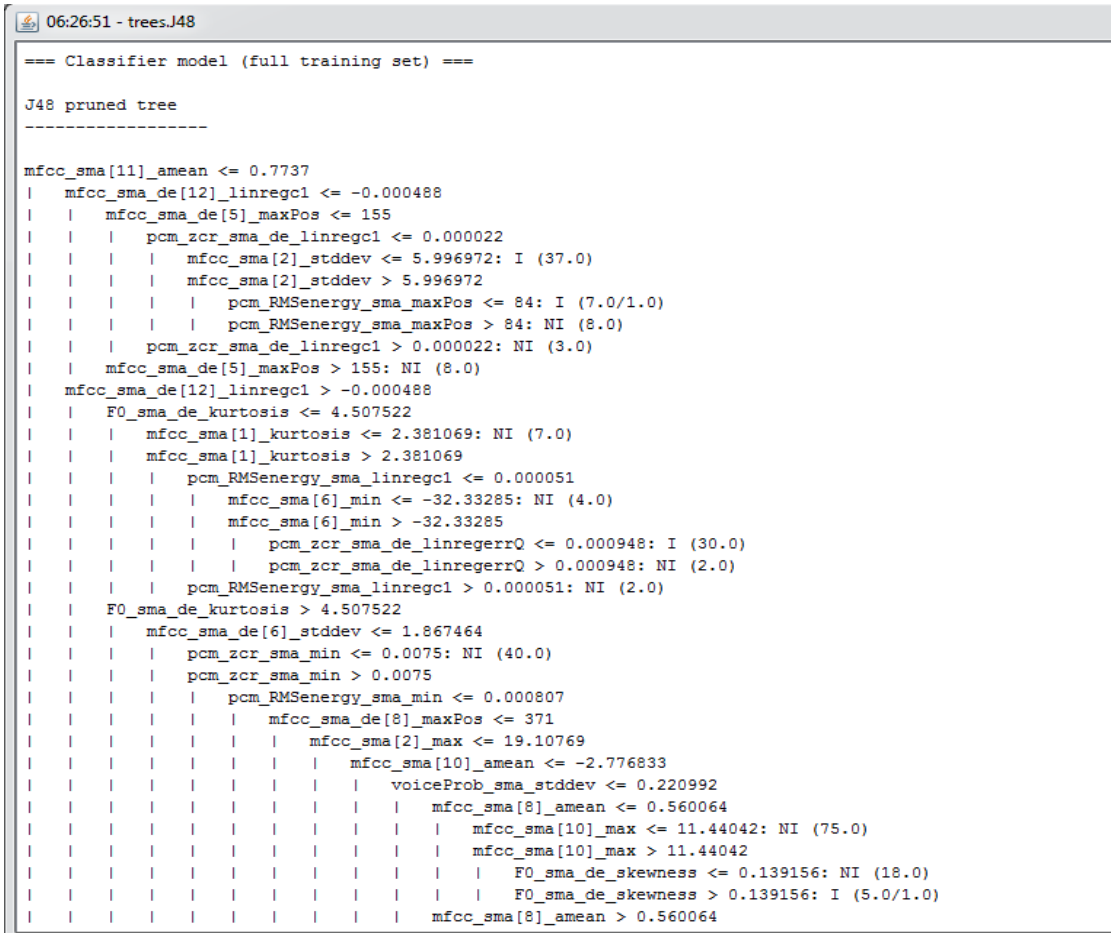
Tableau V.4 : Utilisation de SVM sous Weka.

V.5.3La classification utilisant l'arbre de décision (J48) sous Weka

Dans la série de tests suivante, nous faisons varier le nombre minimal d'occurrence par feuille (minNumObj), car nous avons remarqué qu'il est un élément principal dans la classification utilisant le J48.

minNumObj	10	50	100	105	110	120	130	140	150	160	190	200
Précisiond'entr (%)	57.1	58.6	57.1	60.1	53.1	53.1	53.1	53.1	53.1	53.9	59.8	44.9
Précision de test(%)	51.5	51.6	59.5	45.5	63.1	63.1	63.1	63.1	62.7	44.9	44.9	44.9

Tableau V.5 : Utilisation de J48 sous Weka.



```
06:26:51 - trees.J48
=== Classifier model (full training set) ===

J48 pruned tree
------------------

mfcc_sma[11]_amean <= 0.7737
|   mfcc_sma_de[12]_linregc1 <= -0.000488
|   |   mfcc_sma_de[5]_maxPos <= 155
|   |   |   pcm_zcr_sma_de_linregc1 <= 0.000022
|   |   |   |   mfcc_sma[2]_stddev <= 5.996972: I (37.0)
|   |   |   |   mfcc_sma[2]_stddev > 5.996972
|   |   |   |   |   pcm_RMSenergy_sma_maxPos <= 84: I (7.0/1.0)
|   |   |   |   |   pcm_RMSenergy_sma_maxPos > 84: NI (8.0)
|   |   |   pcm_zcr_sma_de_linregc1 > 0.000022: NI (3.0)
|   |   mfcc_sma_de[5]_maxPos > 155: NI (8.0)
|   mfcc_sma_de[12]_linregc1 > -0.000488
|   |   F0_sma_de_kurtosis <= 4.507522
|   |   |   mfcc_sma[1]_kurtosis <= 2.381069: NI (7.0)
|   |   |   mfcc_sma[1]_kurtosis > 2.381069
|   |   |   |   pcm_RMSenergy_sma_linregc1 <= 0.000051
|   |   |   |   |   mfcc_sma[6]_min <= -32.33285: NI (4.0)
|   |   |   |   |   mfcc_sma[6]_min > -32.33285
|   |   |   |   |   |   pcm_zcr_sma_de_linregerrQ <= 0.000948: I (30.0)
|   |   |   |   |   |   pcm_zcr_sma_de_linregerrQ > 0.000948: NI (2.0)
|   |   |   |   pcm_RMSenergy_sma_linregc1 > 0.000051: NI (2.0)
|   |   F0_sma_de_kurtosis > 4.507522
|   |   |   mfcc_sma_de[6]_stddev <= 1.867464
|   |   |   |   pcm_zcr_sma_min <= 0.0075: NI (40.0)
|   |   |   |   pcm_zcr_sma_min > 0.0075
|   |   |   |   |   pcm_RMSenergy_sma_min <= 0.000807
|   |   |   |   |   |   mfcc_sma_de[8]_maxPos <= 371
|   |   |   |   |   |   |   mfcc_sma[2]_max <= 19.10769
|   |   |   |   |   |   |   |   mfcc_sma[10]_amean <= -2.776833
|   |   |   |   |   |   |   |   |   voiceProb_sma_stddev <= 0.220992
|   |   |   |   |   |   |   |   |   |   mfcc_sma[8]_amean <= 0.560064
|   |   |   |   |   |   |   |   |   |   |   mfcc_sma[10]_max <= 11.44042: NI (75.0)
|   |   |   |   |   |   |   |   |   |   |   mfcc_sma[10]_max > 11.44042
|   |   |   |   |   |   |   |   |   |   |   |   F0_sma_de_skewness <= 0.139156: NI (18.0)
|   |   |   |   |   |   |   |   |   |   |   |   F0_sma_de_skewness > 0.139156: I (5.0/1.0)
|   |   |   |   |   |   |   |   |   |   mfcc_sma[8]_amean > 0.560064
```

Figure V.9 : L'arbre de décision utilisée par le J48.

V.6 conclusion

Dans ce chapitre nous avons développés et implémentés tous éléments et outils que nous avons vus, le logiciel Matlab et Weka ainsi que les algorithmes de classification SVM, l'arbre de décision et l'ELM, ces derniers nous donnés de bon résultat. En outre, nous avons remarqué que l'implémentation de l'ELM présente un temps d'apprentissage intéressent.

Introduction Générale

La parole est le principal moyen de communication dans toute société humaine, il peut y avoir des maladies qui touchent le système phonatoire, d'où la naissance des signaux vocaux pathologiques, dans ce cas le problème majeur est de montrer l'éligibilité ou la non éligibilité de la voix, cette dernière peut être amélioré en utilisant des systèmes de reconnaissance vocale.

La reconnaissance automatique de la parole est un domaine de lascience ayant toujours eu un grand attrait auprès des chercheurs commeauprès du grand public,qui n'ont jamais rêvé de pouvoir parler avecune machine ou, au moins, piloter un appareil ou un ordinateur par la voix.

Pour ne plus avoir à se lever pour allumer ou éteindre tel ou tel appareil électrique, et ne plus avoir à taper pendant des heures sur un clavier pourrédiger un rapport(par exemple), une telle technologie a toujours suscité chez l'homme une part d'envie et d'intérêt, ce que peu d'autres technologies ont réussi à faire.

La reconnaissance automatique de la parole est l'un des deux domaines du traitement automatique de la parole qui permet à la machine de traiter des informations fournies oralement par un utilisateur humain. Elle consiste à employer des techniques d'appariement afin de comparer une onde sonore à un ensemble d'échantillons, composés généralement de mots mais aussi, plus récemment, de phonèmes (unité sonore minimale), l'autre étant la synthèse vocale qui permet de reproduire d'une manière sonore un texte qui lui est soumis, comme un humain le ferait.

Ces deux domaines et notamment la reconnaissance vocale, fontappel aux connaissances de plusieurs sciences : l'anatomie (les fonctionsde l'appareil phonatoire et de l'oreille), les signaux émis par la parole, laphonétique, le traitement du signal, lalinguistique, l'informatique,l'intelligence artificielle et les statistiques.

Le traitement automatique de la parole ouvre des perspectivesnouvelles en tenant compte de la différence considérable existant entre lacommande manuelle et la commande vocale. L'utilisation du langage naturel dans ledialogue personne/machine met la technologie à la portée de tous etentraîne sa vulgarisation, en réduisant les contraintes de l'usage desclaviers, souris et codes de commandes à maîtriser. En

simplifiant leprotocole de dialogue personne/machine, le traitement automatique de laparole vise donc aussi un gain de productivité puisque c'est la machine quis'adapte à l'homme pour communiquer, et non l'inverse.

En simplifiant leprotocole de dialogue personne/machine, le traitement automatique de laparole vise donc aussi un gain de productivité puisque c'est la machine quis'adapte à l'homme pour communiquer, et non l'inverse. De plus, il rendpossible l'utilisation simultanée des yeux ou des mains à une autre tâche. Ilpermet d'humaniser les systèmes informatiques de gestion de l'information,en axant leur conception sur les utilisateurs.

L'objectif de notre travail est d'extraire plus de paramètres en utilisant Open Smile et les injectés dans l'ELM sous Matlab, afin de comparer ses résultats par quelques algorithmes de classifications sous Weka, tels que SVM et l'arbre de décision, qui ont les mêmes paramètres d'entrée.

Notre projet est organisé de la manière suivante :

-Nous présentons dans le premier chapitre une approche générale sur l'identité du signal de parole

-Le deuxième chapitre étudie les signaux pathologiques et leurs causes

-Le troisième chapitre traite une technique d'extraction des paramètres

-Le quatrième chapitre est consacré aux Algorithmes de classification

-Le cinquième chapitre présente les résultats de l'implémentation de nos algorithmes.

-Et à la fin nous terminerons ce mémoire par une conclusion qui résume les résultats obtenus au cours de notre travail et de proposer quelques perspectives.

Bibliographie

[1] Gilles Léothaud, "*Théorie de la Phonation*"2004-2005;

[2] Thomas HUEBER, "*Reconstitution de la parole par imagerie ultrasonore et vidéo de l'appareil vocal : vers une communication parlée silencieuse*", thèse de doctorat Université Pierre et Marie Curie 2009.

[3] Deller, J.R., Hansen, J.H. L., Proakis J. G ,"*Discrete Time Processing of Speech Signals*" IEEE Press 1999.

[4] Calliope , "*La parole et son traitement automatique*" Paris, Masson 1989.

[5] J. Rouat, R. Pichevar, and S. Loiselle. Perceptive, non-linear speech processing and spikingneural networks. International Summe School on Neural Nets "E.R. Caianiello", IX Course: Nonlinear Speech Processing: Algorithms and Analysis, 2004.

[6] J.B. Allen. How do humans process and recognize speech? IEEE Trans. on Speech and Audio Processing, 2(4):567–577, 1994.

[7] S. Greenberg. Representation of speech in the auditory periphery. Journal of Phonetics, Special Issue, 16(1), January 1998.

[8] H. Hermansky. Auditory modeling in automatic recognition of speech. Proc. Keele Workshop, 1996.

[9] O. Ghitza. Auditory models and human performance in tasks related to speech coding and speech recognition. IEEE Trans. on Speech and Audio Processing, 2(1):115–132, 1994.

[10] J. Mariani. Reconnaissance de la parole- Traitement automatique du langage parlé 2. Hermés Science Pulications, 2002.

[11] H. Hermansky. Should recognizers have ears? Proc. ESCA Tutorial and Research Workshop Pion Robust Speech Recognition for Unknown Communication Channels, pages 1–10, 1997.

[12] J.B. Allen. How do humans process and recognize speech? IEEE Trans. on Speech and Audio Processing, 2(4):567–577, 1994.

[13] H. Hermansky. Should recognizers have ears? Proc. ESCA Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, pages 1–10, 1997.

[14] H. Hermansky. Perceptual linear predictive (plp) analysis of speech. The Journal of the Acoustical Society of America, pages 1738–1752, 1990.

[15] Thèse de projet fin d'étude 2003 BLIDA « Traitement du signal de la parole par le filtre de KALMAN, LPC et LPCC ».

[16] Thèse de projet fin d'étude 2007 BLIDA « Application des MFCCs à la reconnaissance des phonèmes arabes ».

[17]: M.KUNT et CALLIOPE 1987 « traitement de la parole ».

[18] http://www.portaudio.com

[19] http://opensmile.sourceforge.net/

[20] Florian Eyben, Martin Waellmerand Bjoern Schuller. The Munich Open Speech and Music Interpretation by large space extraction toolkit. Institute for Human Machine communication (TUM) may 2010

[21] Florian Eyben, Martin W ollmer, and Bj orn Schuller. openear - introducing the munich open-source emotion and a_ect recognition toolkit. In Proceedings of the 4th International HUMAINE Association Conference on A_ective Computing and Intelligent Interaction 2009 (ACII 2009), volume I, pages 576{581. IEEE, 2009.

[22] Björn Schuller, S. Steidl, and Anton Batliner. The interspeech 2009 emotion challenge. In Interspeech (2009), ISCA, Brighton, UK, 2009. Bibliographie chapitre IV

[23] : Bernhard Schölkopf, Alexander J. Smola, Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond, 2002, MIT Press.

[24] : Vladimir Vapnik et A. Lerner, Pattern Recognition using Generalized Portrait Method, Automation and Remote Control, 1963

[26]: Machine à vecteurs de support, Un article de Wikipédia, l'encyclopédie libre.

[27] : Support Vector Machine / Séparateurs à vaste marge /Arnaud Revel / revel.arnaud@gmail.com

[28]:Vapnik, V. Statistical Learning Theory. Wiley-Interscience, New York, (1998)

[29]: F. Delbos, J.Ch. Gilbert (2005). Global linear convergence of an augmented Lagrangian algorithm for solving convex quadratic optimization problems. Journal of Convex Analysis, 12,

[30] : S.S.COI « A vowel recognition using adjusted fuzzy member ship function ».

[31] : V. Vapnik, *The nature of statistical learning theory*, N-Y, Springer-Verlag, 1995

[32]: PETIBON Stephane : 'Nouvelles architectures distribuees de gestion et de conversion de l'energie pour les applications photovoltaiques', 2009.

[33] Marti A. Hearst, *Support Vector Machines*, IEEE Intelligent Systems, vol. 13, no.4, p. 18-28, Jul/Aug, 1998

[34]: Guang-Bin Huang, Qin-Yu Zhu, Chee-Kheong Siew, Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, International Joint Conference on Neural Networks, Vol. 2,pp: 985-990, 2004.

[35]: Dianhui Wang, Guang-Bin Huang, "Protein Sequence Classification Using Extreme Learning Machine, Proceedings of International Joint Conference on Neural Networks, Vol. 3, pp: 1406- 1411, 2005.

[36] : Corinna Cortes and V. Vapnik, "Support-Vector Networks*, Machine Learning, 20, 1995. http://www.springerlink.com/content/k238jx04hm87j80g/*

[37] : R. Rajesh, J. Siva Prakash Extreme Learning Machines - A Review and State-of-the-art international journal of wisdom based computing, vol. 1(1), 2011

[38]: Guang-Bin Huang, Senior Member Extreme Learning Machine for Regression and Multiclass Classification IEEE, transactions on systems, man, and cybernetics part b: cybernetics, vol. 42, no. 2, april 2012

[39] J. A. K. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300,Jun. 1999.

[40] G.-B. Huang, L. Chen, and C.-K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 879–892,Jul. 2006.

[41] G.-B. Huang and L. Chen, “Convex incremental extreme learning machine,” *Neurocomputing*, vol. 70, no. 16–18, pp. 3056–3062, Oct. 2007.

[42] G.-B. Huang and L. Chen, “Enhanced random search based incremental extreme learning machine,” *Neurocomputing*, vol. 71, no. 16–18, pp. 3460–3468, Oct. 2008.

[43] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagation errors,” *Nature*, vol. 323, pp. 533–536, 1986.

[44] P. L. Bartlett, “The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network,” *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 525–536, Mar. 1998.

[45] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: A new learning scheme of feedforward neural networks,” in *Proc. IJCNN*, Budapest, Hungary, Jul. 25–29, 2004, vol. 2, pp. 985–990.

[46] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: Theory and applications,” *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[47] D. Serre, *Matrices: Theory and Applications*. New York: Springer-Verlag, 2002.

[48] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*. New York: Wiley, 1971.

[49] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” *Technometrics*, vol. 12, no. 1, pp. 55–67,Feb. 1970.

[50] W. Deng, Q. Zheng, and L. Chen, “Regularized extreme learning machine,” in *Proc. IEEE Symp. CIDM*, Mar. 30–Apr. 2, 2009, pp. 389–395.

[51] K.-A. Toh, “Deterministic neural classification,” *Neural Comput.*, vol. 20, no. 6, pp. 1565–1595, Jun. 2008.

[52] G.-B. Huang, Y.-Q. Chen, and H. A. Babri, “Classification ability of single hidden layer feedforward neural networks,” *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 799–801, May 2000.

[53] L. van der Molen, M. van Rossum, A. Ackerstaff, L. Smeele, C. Rasch, andF. Hilgers, “Pretreatment organ function in patients with advanced head and neck cancer: clinical outcome measures and patients’ views,” *BMC Ear Nose Throat Disorders*, vol. 9, no. 10, 2009.

Fichier de configuration de emo_IS09

```
//////////////////////////////////////////////////////////////
///////// > openSMILE configuration file for emotion features
<      /////////////////
/////////   Feature set of the INTERSPEECH 2009 Emotion
Challenge    /////////////////
/////////   384 features, (16 LLD + 16 delta)*12 functionals
//////////////////
/////////
//////////////////
/////////  * written 2009 by Florian Eyben *
//////////////////
/////////
//////////////////
///////// (c) 2009 by Florian Eyben, Martin W◌ollmer, Bj◌orn
Schuller  //////////////////
/////////     see the file COPYING for details
//////////////////
//////////////////////////////////////////////////////////////////

//////////////////////////////////////////////////////////////////
;
; This section is always required in openSMILE configuration
files
;   it configures the componentManager and gives a list of all
components which are to be loaded
; The order in which the components are listed should match
;   the order of the data flow for most efficient processing
;
//////////////////////////////////////////////////////////////////
[componentInstances:cComponentManager]
 ; this line configures the default data memory:
instance[dataMemory].type=cDataMemory
instance[waveIn].type=cWaveSource
instance[fr1].type=cFramer
instance[pe2].type=cVectorPreemphasis
instance[w1].type=cWindower
instance[fft1].type=cTransformFFT
instance[fftmp1].type=cFFTmagphase
instance[mspec].type=cMelspec
instance[mfcc].type=cMfcc
instance[mzcr].type=cMZcr
instance[acf].type=cAcf
instance[cepstrum].type=cAcf
instance[pitchACF].type=cPitchACF
instance[energy].type=cEnergy
instance[lld].type=cContourSmoother
instance[delta1].type=cDeltaRegression
instance[functL1].type=cFunctionals
instance[arffsink].type=cArffSink
printLevelStats=0
nThreads=1
```

```
////////////////////////////////////////////////////////////
//////////////////////////   component configuration //////////
////////////////////////////////////////////////////////////
; the following sections configure the components listed above
; a help on configuration parameters can be obtained with
;  SMILExtract -H
; or
;  SMILExtract -H configTypeName (= componentTypeName)
////////////////////////////////////////////////////////////
[waveIn:cWaveSource]
writer.dmLevel=wave
filename=\cm[inputfile(I){test.wav}:name of input file]
buffersize=4000
monoMixdown=1

[fr1:cFramer]
reader.dmLevel=wave
writer.dmLevel=frames
copyInputName = 1
noPostEOIprocessing = 1
frameSize = 0.0250
frameStep = 0.010
frameMode = fixed
frameCenterSpecial = left
buffersize = 1000

[pe2:cVectorPreemphasis]
reader.dmLevel=frames
writer.dmLevel=framespe
copyInputName = 1
processArrayFields = 1
k=0.97
de = 0

[w1:cWindower]
reader.dmLevel=framespe
writer.dmLevel=winframe
copyInputName = 1
processArrayFields = 1
winFunc = ham
gain = 1.0
offset = 0

  // ---- LLD -----

[fft1:cTransformFFT]
reader.dmLevel=winframe
writer.dmLevel=fftc
copyInputName = 1
processArrayFields = 1
inverse = 0

[fftmp1:cFFTmagphase]
reader.dmLevel=fftc
```

```
writer.dmLevel=fftmag
copyInputName = 1
processArrayFields = 1
inverse = 0
magnitude = 1
phase = 0

[mspec:cMelspec]
reader.dmLevel=fftmag
writer.dmLevel=mspec1
copyInputName = 1
processArrayFields = 1
htkcompatible = 1
nBands = 26
usePower = 0
lofreq = 0
hifreq = 8000
inverse = 0
specScale = mel

[mfcc:cMfcc]
reader.dmLevel=mspec1
writer.dmLevel=mfcc1
copyInputName = 1
processArrayFields = 1
firstMfcc = 1
lastMfcc =  12
cepLifter = 22.0
htkcompatible = 1


[acf:cAcf]
reader.dmLevel=fftmag
writer.dmLevel=acf
nameAppend = acf
copyInputName = 1
processArrayFields = 1
usePower = 1
cepstrum = 0

[cepstrum:cAcf]
reader.dmLevel=fftmag
writer.dmLevel=cepstrum
nameAppend = acf
copyInputName = 1
processArrayFields = 1
usePower = 1
cepstrum = 1

[pitchACF:cPitchACF]
  ; the pitchACF component must ALWAYS read from acf AND
cepstrum in the given order!
reader.dmLevel=acf;cepstrum
writer.dmLevel=pitch
copyInputName = 1
```

```
processArrayFields=0
maxPitch = 500
voiceProb = 1
voiceQual = 0
HNR = 0
F0 = 1
F0raw = 0
F0env = 0
voicingCutoff = 0.550000

[energy:cEnergy]
reader.dmLevel=winframe
writer.dmLevel=energy
nameAppend=energy
rms=1
log=0

[mzcr:cMZcr]
reader.dmLevel=frames
writer.dmLevel=mzcr
copyInputName = 1
processArrayFields = 1
zcr = 1
amax = 0
mcr = 0
maxmin = 0
dc = 0

[lld:cContourSmoother]
reader.dmLevel=energy;mfcc1;mzcr;pitch
writer.dmLevel=lld
writer.levelconf.nT=10
;writer.levelconf.noHang=2
writer.levelconf.isRb=0
writer.levelconf.growDyn=1
nameAppend = sma
copyInputName = 1
noPostEOIprocessing = 0
smaWin = 3

// ---- delta regression of LLD ----
[delta1:cDeltaRegression]
reader.dmLevel=lld
writer.dmLevel=lld_de
writer.levelconf.isRb=0
writer.levelconf.growDyn=1
nameAppend = de
copyInputName = 1
noPostEOIprocessing = 0
deltawin=2
blocksize=1

[functL1:cFunctionals]
reader.dmLevel=lld;lld_de
writer.dmLevel=func
```

```
copyInputName = 1
 ; frameSize and frameStep = 0 => functionals over complete
input
 ; (NOTE: buffersize of lld and lld_de levels must be large
enough!!)
frameSize = 0
frameStep = 0
frameMode = full
frameCenterSpecial = left
functionalsEnabled=Extremes;Regression;Moments
Extremes.max = 1
Extremes.min = 1
Extremes.range = 1
Extremes.maxpos = 1
Extremes.minpos = 1
Extremes.amean = 1
Extremes.maxameandist = 0
Extremes.minameandist = 0
 ; Note: the much better way to normalise the times of maxpos
and minpos
 ; is 'turn', however for compatibility with old files the
default 'frame'
 ; is kept here:
Extremes.norm = frame
Regression.linregc1 = 1
Regression.linregc2 = 1
Regression.linregerrA = 0
Regression.linregerrQ = 1
Regression.qregc1 = 0
Regression.qregc2 = 0
Regression.qregc3 = 0
Regression.qregerrA = 0
Regression.qregerrQ = 0
Regression.centroid = 0
Moments.variance = 0
Moments.stddev = 1
Moments.skewness = 1
Moments.kurtosis = 1
Moments.amean = 0

//////////////////////////////////////////////////////////////
 ////////////////////  data output configuration
//////////////////////////////////////////////////////////////
// ----- you might need to customise the arff output to suit
your needs: ------

[arffsink:cArffSink]
reader.dmLevel=func
filename=\cm[output(O){output.arff}:output arff file for
feature data]
append=1
frameIndex=0
frameTime=0
;relation=IS2012_STC_PSC
; name of @relation in the ARFF file
```

```
relation=\cm[corpus(R){SMILEfeatures}:corpus name, arff
relation]
\{config\arff_targets.conf}

 ; do not print "frameNumber" attribute to ARFF file
;frameIndex = 0
;frameTime = 1
 ; name of output file as commandline option
; filename=\cm[arffout(O){output.arff}:name of WEKA Arff
output file]
 ; name of @relation in the ARFF file
; relation=\cm[corpus{SMILEfeatures}:corpus name, arff
relation]

 ; name of the current instance (usually file name of input
wave file)
;instanceName=\cm[instname(N){noname}:name of arff instance]
 ;; use this line instead of the above to always set the
instance name to the
 ;; name of the input wave file
 ; instanceName=\cm[inputfile]

 ; name of class label
;class[0].name = emotion
 ; list of nominal classes OR "numeric"
;class[0].type = \cm[classes{unknown}:all classes for arff
file attribute]
 ; the class label or value for the current instance
; target[0].all = \cm[classlabel(a){unknown}:instance class
label]
 ; append to an existing file, so multiple calls of
SMILExtract on different
 ; input files append to the same output ARFF file
append=1


//////---------------------- END -------------------------
///////
```

Exemple : extraction de paramètres d'un seul fichier wave avec Open Smile sous Matlab .

```
%%%%%%%path to configuration file
PathToConf='config/emo_IS09.conf';

%%%%%%%path to wave file
PathToWavFile='./bakim.wav';

%%%%%%%name of arff file
NewArff='MasterSISII.arff';

%%%%%%%our class
Class='L'

Str12=['smilExtract -C ',PathToConf ,' -I ' ,PathToWavFile,' –O'
        ,NewArff, ' -classlabel_L ',Class]

%%%%%%%evaluate smilExtract under Dos
eval('dos(Str12)')
```