REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOHAMED BOUDIAF - M'SILA

FACULTE DE TECHNOLOGIE
DEPARTEMENT D'ELECTRONIQUE
N° d'ordre :2020/ ……. / ….

DOMAINE : SCIENCES ET TECHNOLOGIES
FILIERE : ELECTRONIQUE
OPTION : SYSTEMES TELECOMMUNICATIONS

Mémoire présenté pour l'obtention
Du diplôme de Master Académique

Par :BEDDIAR ANTAR

Intitulé

Etude et élaboration d'algorithme de débruitage de signaux : Application à la reconnaissance de la parole

Soutenudevant le jury composé de :

Dr. …….	Université de ……Président
Dr. BOURAS MOUNIR	Université de M'sila……Rapporteur
Dr. ……	Université de ……Examinateur

Année universitaire : 2019 /2020

Remerciement

JE voudrais très sincèrement remercier les messieurs

Dr. BOURAS MOUNIR *.....RAPPORTEUR*

pour avoir assuré l'encadrement de ce travail. Leurs disponibilités, Leurs expériences, et savoir scientifique et qualités humaines. Ont été déterminants dans l'aboutissement de ce travail.

Dr. BEN HAMIDA MOHAMED**…….Doyen de la faculté**

d ́électroniques

Pour toute les facilités qu'il m′a fournies dans la réalisation de ce mémoire

Je remercie:

Mrs : les membres le jury pour l'intérêt qu'ils ont manifesté pour notre travail.

Je voudrais terminer en saluant la promotion d'électronique 2020

Dédicaces

Je dédie ce modeste travail

A mon père et ma mère innocente

A ma femme qui n'àcesser de m'encourager atout moment.

A mes enfants :

*** RAOUA / MOUNIB / MOUNCIF ***

A toutes ma famille BEDDIAR et la famille MIRA

*Abstract:

The algorithm we have presented for word isolation in noisy environments provides very acceptable results in most cases.

As it uses the cepstral distance notation, it can be implemented very easily, With all the speech recognition algorithms that calculate, for their operation, the cepstral parameters. This has the enormous advantage of not excessively increasing the volume of computation necessary for the actual recognition.

The spectral subtraction method is not limited to recognition applications alone, but can be used in many other applications such as, for example, a processing aimed at improving the intelligibility of a speech signal disturbed by noise added

*Résumé:

L'algorithme que nous avons présenté, pour l'isolation des mots en milieu bruité, fournit des résultats très acceptables dans la plupart des cas.

Comme il utilise la notation de distance cepstrale, il peut être mis en œuvre très facilement, avec tous les algorithmes de reconnaissance de la parole qui calculent, pour leur fonctionnement, les paramètres cepstraux. Ceci possède l'énorme avantage de ne pas augmenter excessivement volume de calcule nécessaire à la reconnaissance proprement dite.

La méthode de soustraction spectrale, n'est pas limitée aux seules applications de reconnaissance, mais peut être utilisée dans de nombreuses autres applications comme, par exemple un traitement visant l'amélioration de l'intelligibilité d'un signal de parole perturbé par du bruit additionne

S O M M A I R E

CHAPITRE III : Méthodes de comparaison

CHAPITRE IV : Algorithme basé sur la mesure de la distance cepstrale

CHAPITRE V : Résultats expérimentaux

*Liste des figures :

Introduction Générale

INTRODUCTION GENERALE

Un problème important dans le traitement de la parole est celui de la détection de la présence ou l'absence de signal de parole dans un environnement bruité. En particulier, le problème de reconnaissance de mots isolés suppose qu'il est possible de localiser l'endroit où le mot débute et où il se termine. Afin d'optimiser l'algorithme de reconnaissance, nous proposons d'y adjoindre un algorithme de débruitage des signaux.

Ceci permettra par la même occasion de réduire le volume de calcul ; une bonne isolation des signaux à traiter, conduira à une optimisation du nombre d'opérations effectuées lors de la reconnaissance. Le problème d'isolation des mots apparaît lors de la reconnaissance sur les lignes téléphoniques, ce qui sera l'objectif principal de cette étude, mais il est clair que d'autres applications peuvent être envisagées. Par exemple, un pilote dans un cockpit d'avion, un locuteur dans une salle d'ordinateur, etc...

DONC POUR CELA

Nous désirons un algorithme à la fois simple et efficace pour des niveaux de bruit raisonnable, c'est à dire comparable à ceux rencontrés dans les lignes téléphoniques.

C'est pour cette raison que nous avons choisi de développer l'algorithme basé sur la mesure de la distance cepstrale.

La plupart des méthodes actuelles de reconnaissance utilisent, en effet, la notion de paramètres cepstraux ; ces derniers, une fois calculés, pourront d'être utilisés aussi bien pour l'isolation des mots que pour la reconnaissance

CHAPITRE I

Généralités sur le signal de parole

Introduction

La parole étant la forme la plus naturelle de communication entre les êtres humains, on s'est beaucoup intéressé à son utilisation pour communiquer avec les machines malgré l'obstacle majeur que constituant l'énorme puissance de traitement nécessaire.

Les progrès de la technologie des microprocesseurs ont permis de réaliser des systèmes de reconnaissance et de synthèse de la parole pour un certain nombre d'application en temps réel.

Dans cechapitre nous présentons tout d'abord le processus de production de la parole chez les humains. Nous exposons le fonctionnement acoustique de l'appareil vocal et les différentes méthodes de modélisation et analyse de signal vocal

I-1.Production de la parole :

I-1-1. Les sons de parole :

La parole se distingue des autres sons par des caractéristiques acoustiques ayant leurs origines dans les mécanismes de production. Les sons de parole sont produits soit par les vibrations des cordes vocales (source de voisement), soit par l'écoulement turbulent de l'air dans le conduit vocal, soit lors du relâchement d'une occlusion de ce conduit (source de bruit).

Dans le processus de communication parlée, pour une langue donnée, les sons permettent de distinguer les différentes unités de signification du langage.L'unité élémentaire d'un son permettant la distinction des différents mots est le phonème.

La notion de phonème ne tient compte que des caractéristiques acoustiques qui permettent une distinction entre des mots. On ne tient pas compte des phénomènes physiques de production du son, tant que la différence d'articulation (fonction du dialecte, de la cadence d'élocution. du contexte) ne permet pas dedistinguer des mots différents. La représentation phonétique d'un texte dépend de la langue dans laquelle il est écrit.

Les phonèmes peuvent être rangés en catégories selon des « traits distinctifs » qui indiquent une similitude au niveau articulatoire, acoustique ou perceptif. Les voyelles peuvent être rangées selon :

- ✓ La nasalité;
- ✓ L'ouverture du conduit vocal;
- ✓ La position de la constriction du conduit vocal;
- ✓ L'arrondissement des lèvres.

Les consonnes sont classées selon :

- ✓ Le voisement;
- ✓ Le mode d'articulation (occlusif, nasal, fricatif);
- ✓ Le lieu d'articulation (labiale, dentale, palatale).

D'autres aspects de la parole permettent de distinguer les différentes significations. Des phénomènes comme la prosodie, la durée ou l'intensité des phonèmes, le timbre de la voix permettent à l'auditeur d'identifier le locuteur ou de se faire une idée sur son attitude.

I-1-2. Production de parole :

Le processus de production de la parole se présente certaines caractéristiques :

- ✓ *Continuité*; lorsqu'on écoute parler une personne, on perçoit une suite de mots que l'analyse du signal vocal sépare difficilement. Le mêmeproblème de segmentation se retrouve à l'intérieur du mot, perçu comme une suite de sons élémentaires, les phonèmes.
- ✓ *Variabilité*; à contenu phonétique égal, le signal vocal est très variable, tant pour différents individus que pour un même locuteur, en raison des différences anatomiques.
- ✓ *Le conduit vocal* est un tuyau tridimensionnel qui est excité par une ou deux sources acoustiques. La source laryngienne peut être considérée comme quasi périodique, avec une fréquence pouvant évoluer très

rapidement. La seconde source génère du bruit de friction ou d'explosion (glotte, lèvres).

- ✓ *Encodages* ; depuis l'idée jusqu'au signal sonore, interviennent plusieurs niveaux successifs de traitement : sémantique (concept), syntaxique (structure du langage), lexical (mots), morphologique, phonétique (phonèmes et leurs interactions).

Le larynx est un lieu important pour les mécanismes phonatoires. Il est situé dans la région moyenne du cou et il est constitué de cartilages, de muscles, de muqueuse et de nerfs. Il contient les « cordes vocales » qui sont un ensemble de muqueuses, de ligaments et de muscles.

La langue joue un rôle dans la phonation, car sa mobilité lui permet d'agir avec précision et rapidité sur la taille du conduit vocalique.

Les lèvres sont situées à l'extrémité du conduit vocal et c'est leur écartement (et les variations de cet écartement) qui est important du point de vue acoustique.

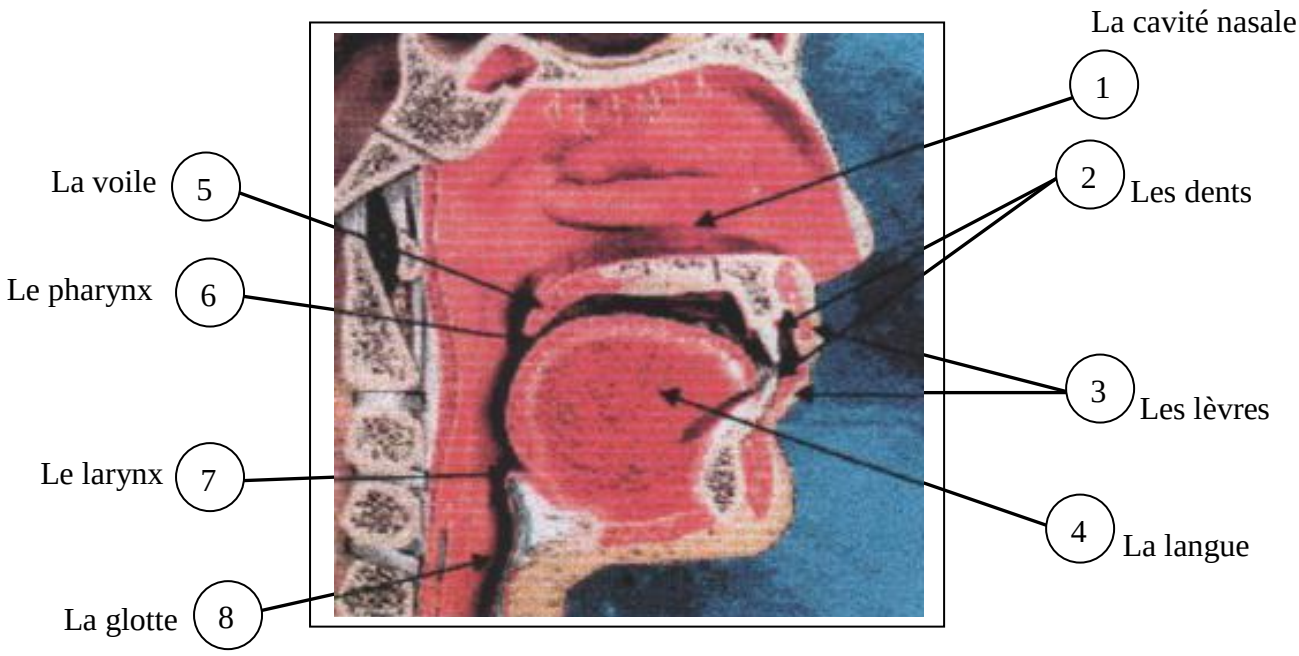


Figure I-1 : L'appareil phonatoire chez l'homme

I-1-3. Modélisation autorégressive du signal vocal :

La représentation fonctionnelle du modèle de production de signal vocal, séparant sources, conduit et rayonnement aux lèvres, est donnée par la figur I-2.

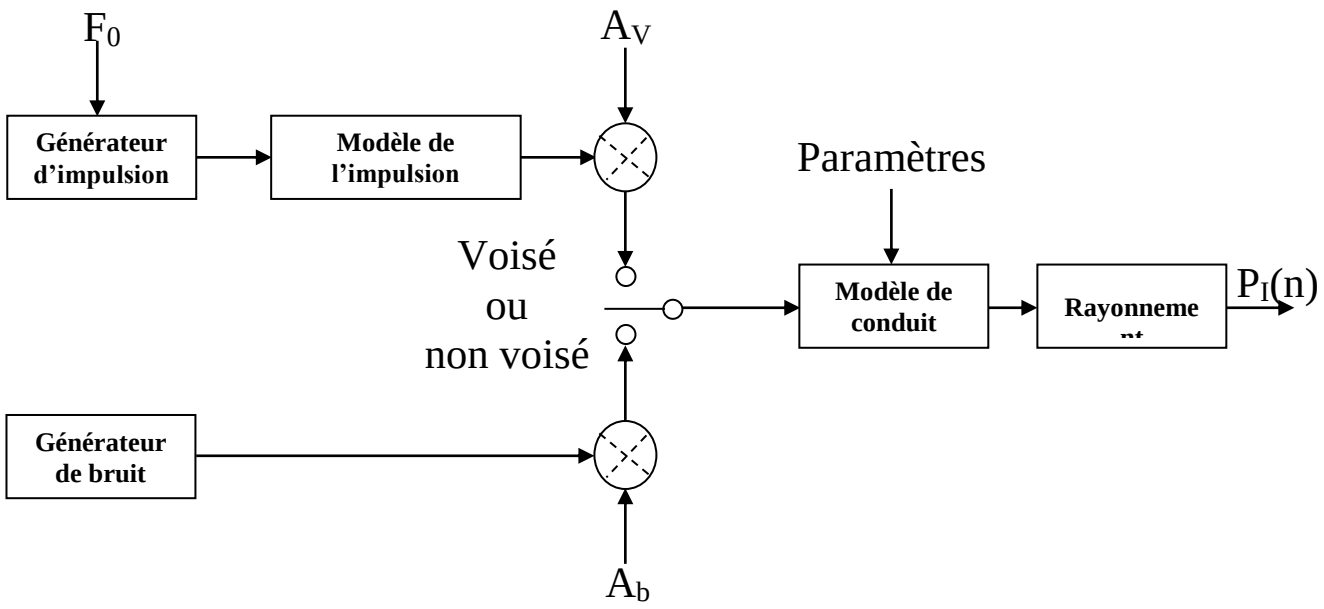


Figure I-2. Modèle fonctionnel de production de la parole

Une phrase est une suite de sons voisés, de sons non voisés et de silences. Pour la générer, il faut connaître pour chaque intervalle de temps dT, intervalle durant lequel le modèle est considéré comme invariant (5 à 25 ms):

*La fréquence fondamentale F_0;

*Les amplitudes A_v et A_b;

*Les coefficients des filtres modélisant le conduit vocal, l'impulsion glottale et le rayonnement aux lèvres.

I-2. Le signal acoustique :

La parole est un signal réel, continu, d'énergie finie et non stationnaire. Sa structure est complexe et variable avec le temps. Sa composition, figure. 1-3, est la suivante :

* Pseudo-périodique (D) sons voisés ;
* Aléatoire (A) sons fricatifs;
* Impulsionel (C) phase explosive des sons occlusifs. (B est du bruit)

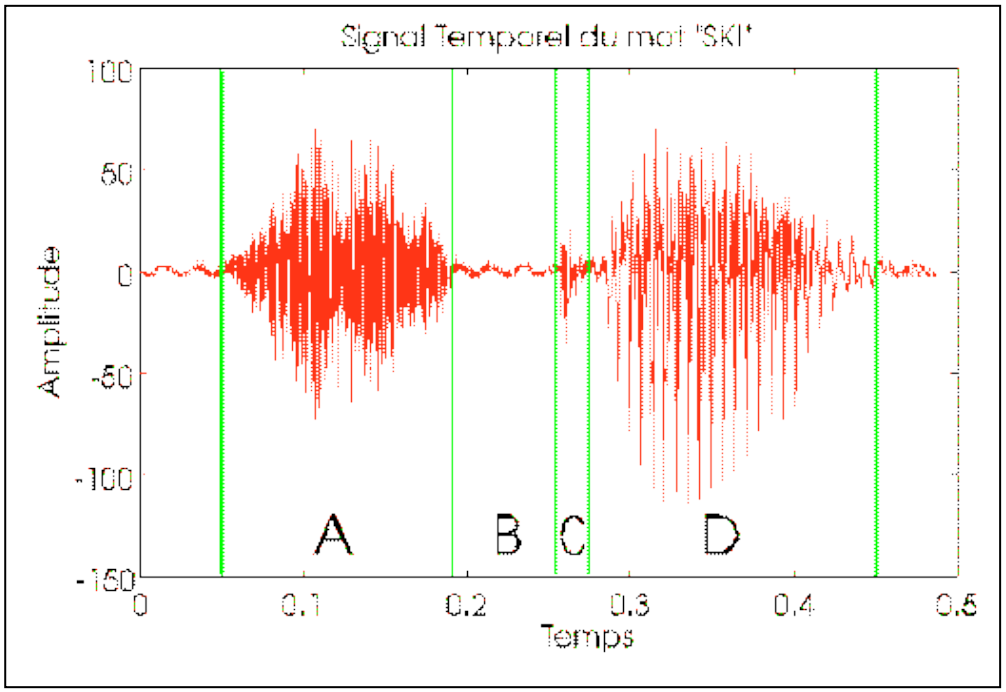


Figure I-3. Représentation temporelle du signal acoustique de la parole

Une manière aisée de décrire le signal acoustique est d'utiliser une représentation sous forme de spectrogramme (les termes couramment employés de *Sonagraph*et *Sonagram* sont des marques déposées), comme dans la figure. 1-4

Le spectrogramme est une représentation tridimensionnelle, où le temps est représenté sur l'axe X, la fréquence sur l'axe Y et le niveau de chaque fréquence, sur l'axe Z, est symbolisé par le niveau de noir. Cette analyse temps-fréquence, d'abord réalisée de manière analogique à l'aide de bancs de filtres, est maintenant réalisée de manière numérique par TFR. Elle sera détaillée dans le chapitre consacré aux outils d'analyse et de traitement du signal.

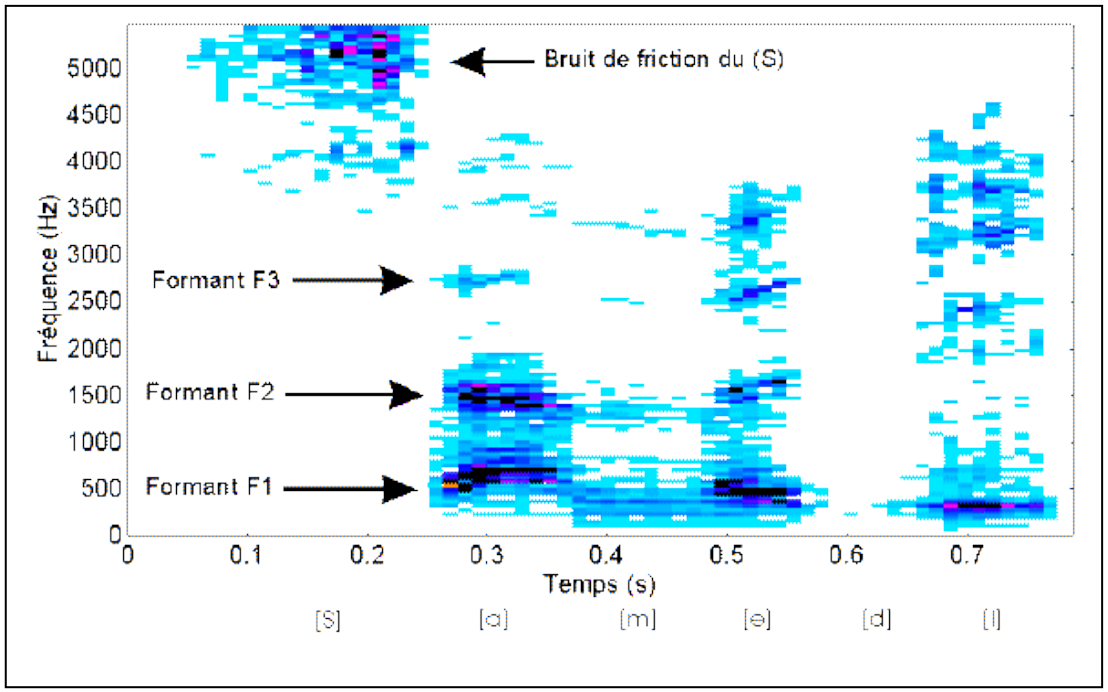


Figure I-4. Spectrogramme de (samedi)

I-3. Système de reconnaissance de la parole :

Le système de reconnaissance de la parole est un système qui reçoit en entrée un signal acoustique, puis essaye d'identifier les informations continues dans celui-ci en vue de la délivrance d'un message ou la réalisation d'une section.

I-3-1. Classification des systèmes de reconnaissance de la parole :

Les systèmes utilisés peuvent être classés selon les hypothèses utilisées:

- *Reconnaissance monolocuteur ou multilocuteur:*

Pour le premier cas, monolocuteur, on suppose que le locuteur est celui qui à donner des exemples de prononciation lors de la phase d'apprentissage, alors que la deuxième approche ne fait aucune contrainte sur l'identité du locuteur.

- ***Reconnaissance de mots isolés et parole continue :***

Selon que les mots énoncés sont séparés par des systèmes de silence ou non, les méthodes d'analyse et de reconnaissance différente. Le problème que se pose pour l'analyse de la parole continue est la segmentation du signal, alors que la reconnaissance est confrontée au problème de la coarticulation.

- ***Reconnaissance globale au analytique :***

Le terme de reconnaissance globale s'applique aux systèmes où l'unité de décision est l'entité lexicale qui peut être un mot ou un groupe de mots. L'idée de base de tels systèmes est de constituer un dictionnaire de référence qui représente les formes acoustiques de chacun des mots qui doivent être identifié par la suite. Au moment de la reconnaissance, l'image acoustique du mot test sera comparée (en utilisant les techniques de programmation dynamique) à celle des mots se trouvant dans le dictionnaire de référence préalablement construit.

Les limités des méthodes globales on conduit les chercheurs vers un second type de méthodes, dites analytiques ou phonétiques dont l'objectif, dans un premier temps, est de déterminer des éléments minimaux ou sons élémentaires (phonèmes, diphonèmes, syllabe...) pour ensuite reconstruire la phase prononcée comme une séquence de ces sons élémentaires. Le processus de traitement nécessite les différentes étapes suivantes :

1. La segmentation du message en unités ;
2. Identification de ces unités ;
3. Reconnaissance du message.

I-3-2. Les principaux problèmes liés à la reconnaissance :

Plusieurs facteurs liés à la complexité du signal vocal redent la tâche de reconnaissance automatique de la parole difficile à implémenter ont cité en général les problèmes majeurs :

- *Bruit et distorsion :*

L'analyse et la reconnaissance de la parole sont plus complexes en présence de bruit, ambiance sonore ou bruits électroniques introduits par le microphone

- *Variation inter et intra locuteurs :*

Un locuteur ne prononce jamais deux fois la même phase de la même façon, autrement le problème de reconnaissance de la parole deviendrait un simple problème de reconnaissance des formes. La voix d'une personne donnée évolue en fonction de l'heure de la journée, l'état physique et moral de la personne...; ces facteurs présentent les problèmes de variation intra-locuteur.

En passant d'un locuteur à l'autre, ces problèmes se compliquent avec les différences de l'intensité de la voix et problèmes des accents régionaux.

Ces problèmes majeurs sont encore mal résolus, et les recherches récentes sont orientées vers la détermination d'autres indices de décodage qui pourront être indépendant du locuteur.

- *La continuité :*

Nous percevons un discours dans une langue connue comme une suite de mots alors même que l'examen du signal acoustique né révélé aucune marque de séparation entre mots. De plus, l'homme perçoit une suite discrète de phonèmes sont floues.

- *La redondance :*

La parole naturelle véhicule de nombreux éléments faisant double emploi les uns avec les autres ; ceci correspond à la présence, dans le signal vocal, de redondance qui explique les performances souvent étonnantes de la perception humaine : même dans de mauvaises conditions de perception où

certainséléments du signal peuvent disparaître avec le bruit, il en reste encore suffisamment d'information pour identifier les mots du discours.

- *Contrainte articulatoire :*

Le problème de la coarticulation, liée à l'influence exercée par les phonèmes les uns sur les autres, constitue l'un des problèmes majeurs de la reconnaissance automatique de la parole en utilisant l'approche analytique.

1-4. Les domaines d'application de la reconnaissance de la parole :

La parole est un moyen de communication très utilisé par les êtres humains dans leurs échanges d'information, et tous les gens attendent le jour où on pourra s'adresser normalement à une machine en utilisant la parole.

Ces affirmations sont dues à des avantages que présente la communication vocale, en voici quelques-uns.

- *Programmation sans clavier :*

On peut s'adresser à une machine et lui dicter un programme par un combiné téléphonique, à travers un micro servant de terminal et ceci grâce à des systèmes de reconnaissance ou encore de compréhension.

- *Communication avec l'ordinateur à distance :*

Relier les instituts, laboratoires, centres de documentation autorisé, banques de donnée, fichiers, etc.…, deviendra très facile en communiquant avec les ordinateurs, à distances par la voie orale.

- *Facilité de renseignement :*

La reconnaissance de la parole peut servir au développement de l'enseignement assisté par ordinateur.

- *L'aide aux handicapés :*

Il est possible de permettre aux sourds grâce à un système de reconnaissance de la parole de communiquer avec une machine par voie orale, au moins pour un vocabulaire limité. Par contre, pour les handicapés moteurs, il suffira qu'ils prononcent une simple commande vocale pour actionner toutes sortes de dispositifs mécaniques.

- *Facilité de la vie quotidienne :*

L'utilisation du clavier vocal permet une interaction très rapide entre l'homme et la machine, cela augmentera la vitesse de transfert des informations dans le sens homme machine.

Conclusion

Dans ce chapitre, nous avons présenté le mécanisme de production de la parole chez les humains ainsi que le fonctionnement acoustique de l'appareil phonatoire d'humain. Par la suite, nous avons décrit les différentes méthodes de système de reconnaissance de la parole et Les principaux problèmes liés à la reconnaissance ….

CHAPITRE II

Analyse et traitement du signal de parole

I. Introduction :

La phase d'analyse du signal de la parole est le module de base de toutes les applications traitant le domaine de la communication parlée, on cite la reconnaissance, le codage et la synthèse de la parole.

Le traitement du signal vocal a pour but de fournir une représentation moins redondante de la parole que celle obtenue par codage direct de l'onde temporelle tout en permettant l'extraction précise des paramètres significatifs.

En acoustiquele son se définit classiquement au moyen de son amplitude, de sa durée et de son timbre. La parole est un son particulièrement complexe, n'échappe pas à cette définition. Le traitement du signal a pour but d'extraire la valeur de ces trois grandeurs pour faire correspondre à Tonde sonore une description multidimensionnelle.

Les notions d'amplitude et de timbre n'ont véritablement de sens que si le son considéré est stationnaire ou pour le moins stable (stationnaire par morceaux), aux contraire le signal de parole est généralement non stationnaire et a une structure très variée, tantôt périodique impulsionnelle et ou bruité. Pour pallier à ces problèmes, l'analyse du signal de parole est réalisée sur des fenêtres d'analyses courtes dans lesquelles le signal est considéré comme stable (analyse à court terme).

II-2. Prétraitement :

Le prétraitement du signal de parole sert d'une part à amplifier le signal capté qui est de faible amplitude, et le scinder en blocs pour permettre l'analyse à court terme et pour préserver la stationnarité du signal.

Le prétraitement consiste à réaliser les trois opérations suivantes :

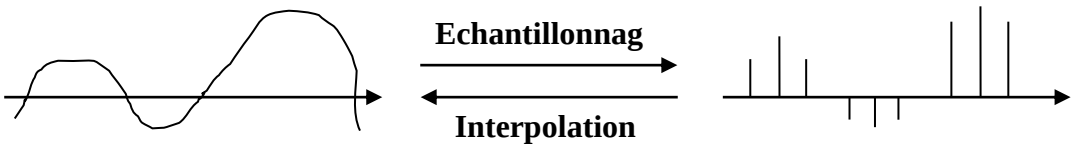
- Acquisition
- Préaccentuation
- Fenêtrage

II-2-1. Acquisition :

L'analyse du signal de parole sur l'ordinateur nécessite numérisation du signal de parole (une discrétisation) réalisée par un convertisseur A/N. Ce convertisseur réalise les opérations suivantes :

II-2-1-1. L'échantillonnage :

Lors de l'acquisition du signal de parole par le microphone, celui-ci enregistre et envoie un signal analogique ; seulement avec le développement des calculateurs et des circuits numériques, le traitement analogique a subi un déclin important vis-à-vis du traitement numérique. En conséquence la numérisation (ou discrétisation) du signal continu sortant du microphone ou d'un appareil d'enregistrement (magnétoscope, magnétophone, etc.…) est devenu nécessaire afin de bénéficier des avantages du traitement numérique de ce signal, cette opération s'appelle échantillonnage du signal ; et l'opération inverse est l'interpolation.



Pour réaliser l'échantillonnage d'un signal on doit appliquer le théorème de SHANNON qui garantit la préservation de toutes les caractéristiques du signal analogique, et la reconstitution de celui-ci.

- *Théorème de SHANNON :*

La perte d'information entre le signal continu et le signal discret correspondant est nulle si, et seulement si, la fréquence d'échantillonnage (fe) est au moins supérieure ou égale au double de la fréquence la plus haute (fm) contenue dans celui-ci :

$$fe \geq 2\ fm \quad \ldots\ldots\ldots\ldots\ 1$$

II-2-1-2. La quantification :

Elle consiste à discrétiser les échantillons analogiques en les rapportant àune unité convenable. Les rapports $S(t_1)/V$ …… $S(t_n)/V$ sont arrondis à la valeurentière la plus proche, ces valeurs sont en nombre fini du fait que la valeur absolus des échantillons est supérieurement bornée. On peut dire, cette fois, que l'information a été mise sous forme numérique, l'ensemble d'un échantillonnage et d'une quantification constitue une conversion analogique/numérique.

II-2-1-3. Le codage :

C'est une opération qui s'impose juste après la quantification. Il s'agit de choisir un code bien déterminé.

L'intérêt du codage apparaît notamment dans la simplification de la traduction numérique.

II-2-2. Préaccentuation :

L'idée de base de cette phase est d'extraire la dérivée du signal enregistré en utilisant un filtre de préaccentuation dont la fonction de transfert est donnée par :

$G(z) = 1 – mz^{-1}$ avec $0,95 < m < 1$ ……..2

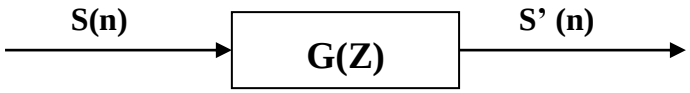


Figure II-1. Le filtre de préaccentuation

D'après la fonction de transfert,S'sera définie par la formule :

$$S'(n) = S(n) – m\, S(n –1)……..3$$

Le rôle de préaccentuation est principalement de rehausser les amplitudes faibles par rapport aux hautes amplitudes afin de tenir compte de l'ensemble du

signal, en d'autres termes elle sert à compresser ladynamique du signal de parole.

II-2-3. Fenêtrage :

Les propriétés du signal de parole sont considérées par une évolution lenteau cours du temps. Ceci a conduit au principe d'analyse à court terme par lequel plusieurs segments du signal sont isolés et traités successivement comme s'ils résultaient d'un son stationnaire avec des propriétés invariables.

Du fait du non stationnarité du signal de la parole, on déplace la fenêtre d'analyse pour le segment suivant avec un mouvement régulier en maintenant une région commune de 5 à 12 millisecondes entre deux segments successifs.

L'idée principale du fenêtrage est de multiplier le signal de parole par une fenêtre de longueur fixe. Il a objectif de subdiviser l'échelle temporelle du signal en segments sur lesquels le signal est considéré comme stationnaire.

Pour l'analyse du signal dans le domaine du traitement automatique de la parole, on utilise trois types de fenêtres :

- *Fenêtre de HAMMING,*
- *Fenêtre de HANNING,*
- *Fenêtre rectangulaire.*

La fenêtre de HAMMING est la plus utilisée dans le domaine de l'analyse du signal de parole, parce qu'elle introduit un minimum de perturbation dans lesignal.

La fonction fenêtre est donnée par la relation suivante :

$$\mathbf{F\,(nT) = } \begin{cases} \boldsymbol{\alpha + (\alpha - 1)\cos [2n\pi T / N]} & \textbf{si} |\mathbf{nT}| < \mathbf{N / 2} \\ \mathbf{0} & \textbf{ailleurs} \end{cases}$$

Pourα = 0,54 : cecicorrespondàlafenêtre de HAMMING

Pourα = 0,50 : cecicorrespondàlafenêtrede HAMMING

Pourα = 1 :cecicorrespondàlafenêtrerectangulaire

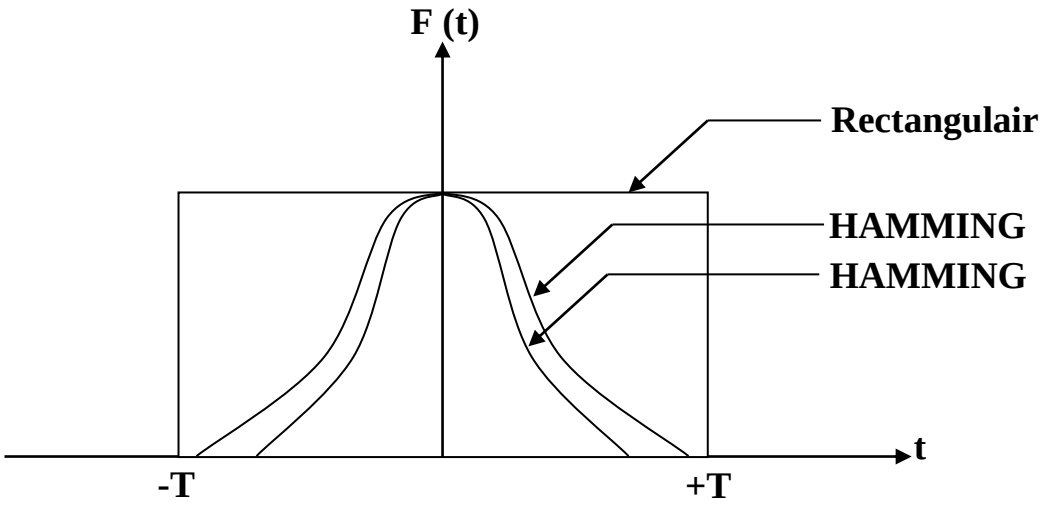


Figure II-2. Représentation temporelle des différentes

II-3. Les descriptions temps-fréquences :

Le signal acoustique de la parole est variable dans le temps. Aussi, les descriptions temps-fréquences sont des formes de représentation couramment utilisées en analyse de la parole.

Les termes description temps et fréquence doivent être pris dans un sens suffisamment large pour inclure diverses formes de représentation et plusieurs notions pour le temps ou la fréquence.

II-3-1. La fréquence :

La notion de fréquence évoque la répétition dans le temps d'un même motif (par exemple la sinusoïde). On peut distinguer :

- La fréquence υ au sens de Fourier. Elle permet de représenter un signal d'énergie finie par une somme d'exponentielles complexes.

$$X(\upsilon) = \int_{-\infty}^{+\infty} x(t) \, . \, e^{-2i\pi\upsilon t} dt \ldots\ldots\ldots\ldots\ldots\ldots 4$$

- La fréquence instantanéeυ_i, déterminée à partir de la phase instantanée de la partie réelle du signal analytique $x_a(t)$ de x(t).

$$\upsilon_i = \frac{1}{2\pi}.\frac{d\Phi}{dt} \text{ avec } x(t) = Re[x_a(t)] = Re\left[A(t).e^{-i\Phi(t)}\right] \ldots\ldots\ldots 5$$

'*A*' est l'enveloppe instantanée.

- Dans le cas des signaux monochromatiques, υ_i correspond à la fréquence au sens de Fourier.
- Il est possible de rechercher une répétition avec d'autres formes que des sinusoïdes, ou une invariance de formes à des échelles de temps et de fréquences différentes. On parle alors « d'échelles ».

II-3-2. Le temps :

La notion de temps peut donner lieu à deux interprétations permettant de distinguer :

- *Les méthodes adaptatives* pour lesquelles le temps est un ensemble de dates, avec hypothèse de stationnarité locale.
- *Les méthodes évolutives* où le temps est une variable de la représentation, sans notion de stationnarité locale.

II-3-3. La description :

Le signal peut être représenté par une distribution de grandeurs physiques (énergie par exemple) dans le domaine temps-fréquence utilisé, par une décomposition sur une famille de fonction, ou par paramétrisation si on fait référence à un modèle.

II-4. Le spectrogramme :

Le spectrogramme est la représentation temps-fréquence la plus courante. C'est une représentation non paramétrique de la distribution énergétique du signal dans le domaine spectro-temporel.

Le *sonagraph* est le plus ancien outil utilisé par les phonéticiens pour caractériser la parole. Appareil analogique, il a été supplanté par les calculateurs

mettant en œuvre des algorithmes de TFR ou de TFD récursive. Il est ainsi possible, en utilisant des processeurs de signaux, d'obtenir des spectres en temps réel.

Avec l'utilisation des calculateurs, et donc des méthodes numériques, il faut échantillonner et numériser le signal. La fréquence d'échantillonnage est généralement comprise entre 8 et 16 kHz tandis que la quantification se fait sur 8 à 16 bits.

Pour obtenir un spectrogramme, on effectue sur le signal une TFR à fenêtre glissante. C'est à dire qu'on analyse une portion limitée du signal, prélevée à l'aide d'une fenêtre de pondération (fenêtre de HAMMING par exemple). Pour ne pas perdre d'information et assurer un meilleur suivi des non-stationnarités, les fenêtres se recouvrent. Elles ont généralement une longueur de 256 ou 512 points et le recouvrement est de 50%, soit 128 ou 256 points.

Afin de compenser le niveau plus faible des aigus, il est généralement utilisé un filtre passe-haut, dit de préaccentuation (avec par exemple $H(z) = 1 - 0{,}9z^{-1}$).

Le fondamental (la fréquence de vibration des cordes vocales) produit de nombreux lobes qui perturbent la lecture du spectrogramme, en particulier la position des formants (fréquences de résonance du conduit vocal).Afin de s'en affranchir, plusieurs types de lissage sont possibles. Un des plus courants est la pondération de chaque trame spectrale par des fenêtres triangulaires. Ce lissage présente aussi l'avantage de réduire le nombre d'informations, en vue d'une éventuelle reconnaissance sur le spectrogramme.

II-5. Les méthodes d'analyses :

Le but de l'analyse du signal est traiter le plus grand nombre d'informations contenues dans le signal, afin de dégager les paramètresnécessaires à la reconnaissance de la parole en vue de réaliser une compression de données, et d'éliminer les informations redondantes.

On peut classer des méthodes d'analyse en trois classes :

- Méthode spectrale ou périodogramme,
- Méthode temporelle,
- Méthode d'identification et de connaissance.

II-5-1. Méthode spectrale :

Elle traite le signal dans le domaine fréquentiel afin de déterminer la densité spectrale du signal ; elles permettent de mettre en évidence certaines propriétés du signal, comme les formants (fréquences de résonance du conduit vocal), qui sont difficiles à observer le domaine temporel.

II-5-1-1. Transforme de FOURIER discrète :

C'est une méthode permettant d'obtenir une estimation de la densité spectrale. Elle consiste en la décomposition du signal en somme de fonction sinusoïdale, telle :

$$X(f) = 1/N \sum_{K=0}^{N-1} S(K) exp\left[-\frac{2jKf\pi}{N}\right] \ldots\ldots\ldots\ldots\ldots 6$$

Ou f = 0, …. N – 1 ; N : Longueur de la fenêtre.

En pratique, le calcul de la TFD est réalisé en utilisant une version améliorée de l'algorithme connue pour sa rapidité de calcul, qui porte le nom de FFT (Fast Fourrier Transforme).

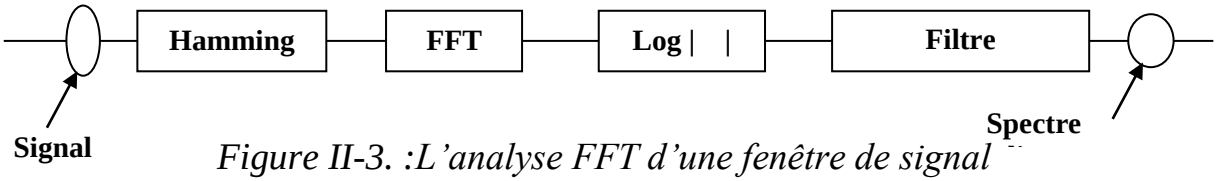


Figure II-3. :L'analyse FFT d'une fenêtre de signal

II-5-1-2. Le banc de filtres : vocodeur à canaux :

Cette technique se propose d'évalu

er l'énergie du signal vocal dans les bandes des fréquences bien choisies, celle justement auxquelles l'oreille humaine est sensible. Le banc de filtre est composé de filtre passe bande, couvrant un spectre étendu de la voix intéressante (généralement de 200 à 600 Hz). Le résultat obtenu représente l'évolution au court d'élocution de la densité spectrale relative à la bande passante du filtre d'analyse considère.

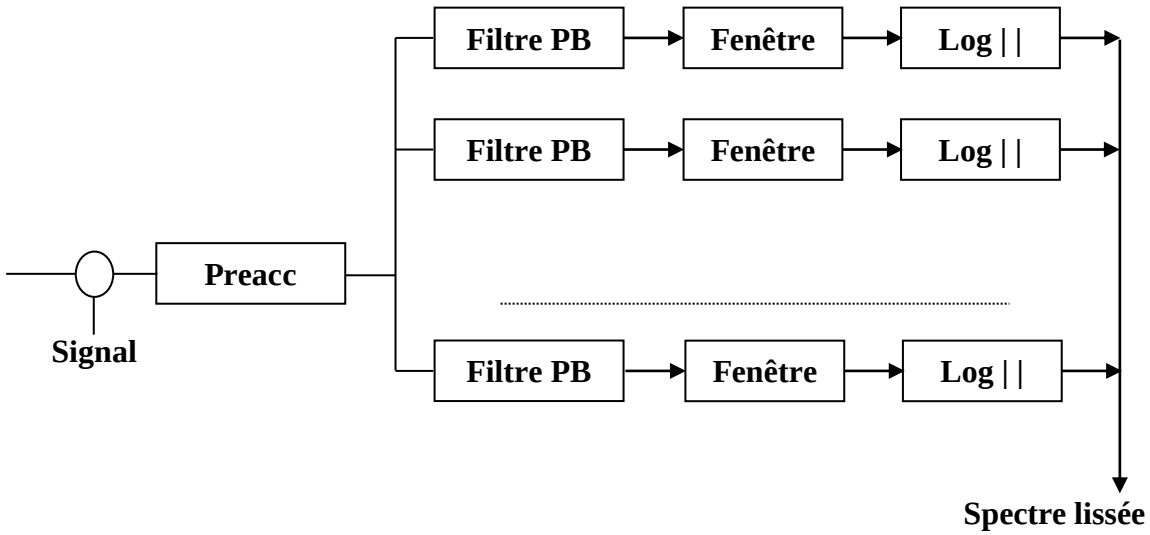


Figure II-4.: L'analyse par vocodeur à canaux

Elles permettent d'extraire des informations du signal acoustique issu directement du microphone (domaine temporel). En général, elles se basent sur la quantification de trois grandeurs :

II-5-2-1. L'énergie totale :

Elle est calculée par la formule suivante :

$E_n = \frac{1}{N}\sum_{k=0}^{N-1} s^2(k)$…….. 7

Où : s(K) : est la valeur du signal à l'échantillon K.

N : est la longueur de la fenêtre.

Ce paramètre nous permet de détecter les segments parole/non parole c'est-à-dire, l'élimination du silence lorsque le signal enregistré est de bonne qualité (absence de bruit), et d'autre part pour localiser les voyelles qui sont caractérisées par une forte énergie comparée avec les autres sons.

II-5-2-2. La densité de passage par zéro :

On a:

$DPZ = 1/2 \sum_{i=0}^{n} \left|Sign\left(s(k+1)\right) - Sign\left(s(k)\right)\right|$…….. 8

Où Sign est une fonction définie par :

$Sign\ (y) = \begin{cases} 1\ Si\ y >= 0 \\ -1\ Si\ y\ < 0 \end{cases}$ …….. 9

La densité de passage par zéro permet de faire une distinction entre le signal de la parole et le bruit, et entre les sons voisés et les sons non voisés.

Remarque :

On note que le taux de passage par zéro est faible pour les sons voisées (environ 1500/s) et élevé pour les sons non voisées (environ 5000/s).

II-5-2-3. L'autocorrélation :

Cette méthode calcule à partir de chaque fenêtre de traitement un vecteur de paramètres R définie par :

$R_{xx}(i) = \frac{1}{N}\sum_{k=0}^{N-1} s(k) * s(k+1)$ i = 0 …….N – 1 …… 10

R(i) sont utilisés pour estimer la fréquence du fondamental (Pitch), utilisée aussi dans l'analyse par prédiction linéaire.

II-5-3. Méthode d'identification et de connaissance :

Elles sont fondées, contrairement aux méthodes précédentes, sur une connaissance des mécanismes de production. Pour la première, le cepstre, cette connaissance est minime, on tente simplement de déconvoluer la «source» et le «conduit». Dans la seconde, le codage prédictif linéaire (LPC: Linear Prédictif Coding) les hypothèses sont de nature structurelle.

II-5-3-1. Analyse cepstrale (spectrale homomorphique) :

Le cepstre est basé sur une connaissance du mécanisme de production de la parole. On part de l'hypothèse qui la suite x_n constituant le signal vocal est le résultat de la convolution du signal de la source par le filtre correspondant au conduit:

$x_n = u_n * h_n$ avec: x_n:Lesignaltemporel

u_n:Lesignal excitateur

h_n: La contribution du conduit.

L'analyse cepstrale permet de séparer sous certaine condition des signaux u et h on dira qu'on effectue une «déconvolution».

Le principe de la méthode consiste à calculer le logarithme de la TZ du signal puis on détermine l'original :

$$\hat{X}(z) = \ln X(z) = \sum_n \hat{x}(n) Z^{-1} \quad \text{..... 11}$$

$$\hat{X}(z) = \ln|U(z)| + \ln|H(z)|$$

Le signal $\hat{x}$(n) obtenue à partir de x(n) par une opération non linéaire noté: «cepstre complexe» associé à x(n):

$$\hat{x}_e(n) = \hat{u}_e + \hat{h}_e(n) \quad \text{..... 12}$$

** Cepstre réel :*

La suite $\hat{x}_e(n)$obtenue à partir Ln |X(exp(jθ))| est appelé «cepstre réel» et on le note c(n):

$$c(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left|X\left(e^{j\theta}\right)\right| e^{jn\theta} d\theta \quad \text{…….. 13}$$

Le calcul est effectué par un algorithme de la TFR, on obtient :

$$X_k = X\left(exp\left(jk\frac{2\pi}{N}\right)\right) = \sum_{n=0}^{N-1} x(n) exp\left(-jnK\frac{2\pi}{N}\right) \quad \text{……14}$$

et : $\tilde{c}(n) = \frac{1}{N}\sum_{k=0}^{N-1} \ln|X_k|\, exp\left(jkn\frac{2\pi}{N}\right)$

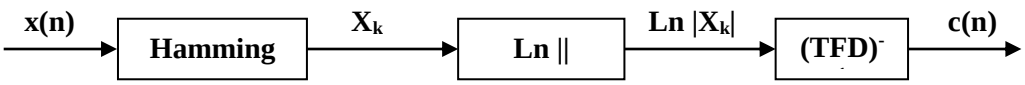


Figure II-5. :Calcul du cepstre réel

et avec la caractéristique de la discrétisation de la fréquence on obtient:

$$\tilde{c}(n) = \sum_{1} c(n + 1N) \quad \text{… … … … … … … … ….15}$$

On doit choisir une valeur de N assez grande pour minimiser le phénomène de recouvrement: N>>1 ⇨$c(n) \cong \tilde{c}(n)$

Si: x(n) est un signal vocal on utilisé de «cepstre à court terme»:

x(n) = [p(n) * g(n) * h_c(n) * r(n)] w(n) …… 16

avec:

- P(n) est un train d'impulsion à la fréquence fondamental :
  $p(n) = \sum_{v} \delta(n - kp)$ ……..17
- g(n) est l'onde glottique
- h_c(n) est la réponse implusionnelle du conduit vocal.
- r(n) réponse impulsionnelle correspondant aux rayonnements des lèvres.

- w(n) est la fonction fenêtre

On pose: $h(n)= g(n) * h_c(n) * r(n)$ donc

$x(n) = [p(n) * h(n)] w(n)$

La procédure de déconvolution exige que le signal observé soit un produit de convolution, c'est pour cela on fait l'hypothèse que la fenêtre w(n) recouvre un nombre suffisant L de période du fondamental où sa variation est faible sur la durée effective de la réponse impulsionnelle, donc on pourra écrire :

$x(n) \cong [p(n) w(n)] * h(n)$

$p_w(n) = p(n) w(n)$

Le cesptre complexe veut : $x(n) = p_w(n) * h(n)$

En pratique la reconstitution de la phase est malaisée c'est pour cela on utilise beaucoup plus cepstre réel calculé par la TFD : on prend des blocs de B échantillons pondérés par une fonction fenêtre, le cepstre c(n) caractérisé le module spectre de x; on aura:

$$\ln|X\ (exp(j\theta))| = \sum_n c(\hat{}n)ex(-jn\theta) \dots\dots\dots\dots..18$$

II-5-3-2. Codage prédictif linéaire (LPC) :

L'analyse par prédiction linéaire utilise un modèle autoregressif (AR) pour modéliser la production du signal de la parole. Elle est très utilisée dans le domaine de la compression de la parole pour la transmission à faible débit.

La méthode LPC pose deux hypothèses:

- Le signal vocal résulte de la convolution d'une source et le conduit.
- Le système de production est linéaire.

Le modèle de production de la parole est un système dont la transmittance de la forme:

$\sigma/A(z)$

une excitation idéalisée: $u(n) = \sum_k \sigma(n - kp)$pour les sons voisés, etpour les sons non voisés l'excitation est un bruit blanc de moyenne nulle.

La transformé en Z du signal donné:

$S(z) = U(z)\frac{\sigma}{A(z)}$ ou $A(z) = \sum_{i=0}^{P} a(i)Z^{-i}$ ……19

$a(0) = 1$…….

Ce modèle est appelé «autoregressif» correspondant dans le domaine temporelle à la récurrence linéaire:

$$S(n) + \sum_{i=1}^{P} a(i)\, s(n-i) = \sigma\, u(n)$$

Avec: a(i) sont les coefficients de prédiction

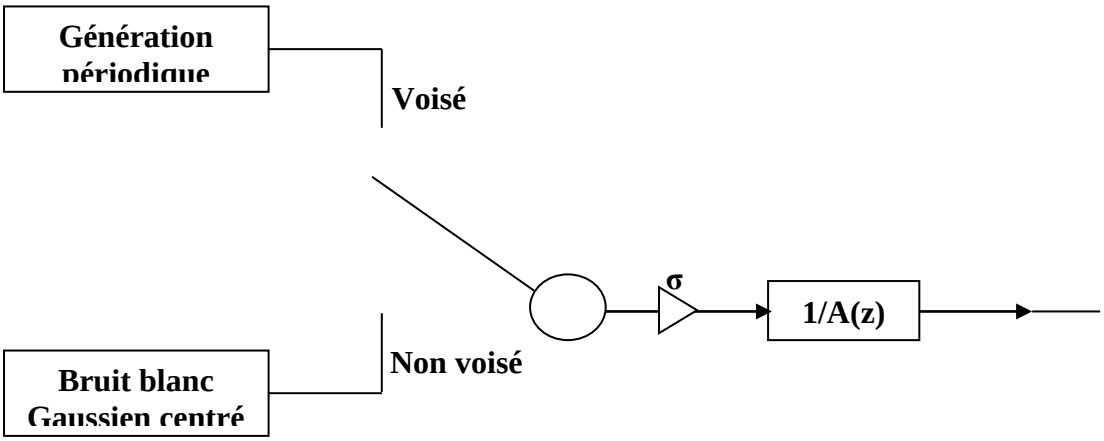


Figure II-6. :Modèle autorégressif de production de la parole

- *Principe de base:*

L'idée principale de l'analyse par prédiction linéaire est de supposer que chaque échantillon du signal de parole s(n) peut être estimé par une combinaison linéaire des P échantillons passés.

$$\hat{s}(n) = -\sum_{i=1}^{P} \hat{a}(i)s(n-i) \text{ …… } 20$$

ou:$\hat{s}(n)$:estimationdel'échantillons(n)

$\hat{a}(i)$: sont les estimées des coefficients a(i) de la récurrence.

P : Ordre de prédiction.

Pour cette estimation, on a commis une erreur e(n) de la valeur définie par:

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i-1}^{P} \hat{a}(i)s(n-1)$$

$$e(n) = \sum_{i=0}^{P} \hat{a}(i)s(n-1)\,; \quad \hat{a}(0) = 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 21$$

Lorsque l'adéquation du modèle est parfaite c'est à dire: â(i)=a(i)

Le problème qui se pose est de déterminer les valeurs des a(i) qui minimisent l'erreur quadratique totale:

Pour cela il existe plusieurs approches parmi lesquelles on cité:

- ✓ Méthode d'autocorrélation
- ✓ Méthode de covariance

- *Approche basée sur la méthode d'autocorrélation:*

Pour cette méthode, le signal est défini pour toute les valeurs du temps, il est supposé nul en dehors d'une fenêtre de longueur finie correspondant à N échantillons (exemple fenêtre de HAMMING)

$$E = \sum_{n=-\infty}^{+\infty} e^2(n) = \sum_{n=-\infty}^{+\infty} \left[\sum_{i=0}^{P} a(i)\, s(n-i)\right]^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 22$$

Pour retrouver les valeurs des a(i) qui minimisent E, il suffit d'annuler les dérivées partielles de E par rapport à chaque a(i).

$$\frac{\partial E}{\partial a(j)} = 0 \Longrightarrow \sum_{n=-n}^{+\infty} s(n)\, s(n-j) = \sum_{i=1}^{P} a(i) \sum_{n=-\infty}^{+\infty} s(n-i)\, s(n-j)\ pour\ j = 1 \ldots\ldots . . P$$

$$\Longrightarrow \sum_{n=-n}^{+\infty} s(n)\, s(n-j) = \sum_{i=1}^{P} a(i) \sum_{m=-\infty}^{+\infty} s(m)\, s(m+1-j) \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

Ces équations (4) sont plus connues sous le nom des équations de « YULLE-WALKER » ou d'équation normale.

On pose :$R(k) = \sum_{m=-\infty}^{+\infty} s(m)\, s(m+k)$ ……23

Avec K = i – j, k = 0 …. P

Le coefficient R est symétrique c'est-à-dire : R(k) = R(-k)

Mais on avait supposé que s(m) était nul en dehors d'une séquence de N échantillons ; en conséquence la formule (5) devient :

$$R(k) = \sum_{m=0}^{N-k-1} s(m)\, s(m+k) \qquad pour\ k = 0 \ldots\ldots p, \ldots\ldots\ldots\ldots\ldots (6)$$

En remplaçant, les valeurs de R(k) dans l'équation (4), on trouve :

$$\sum_{i=1}^{P} a(i)R(i-j) = R(j) \qquad pour\ j = 0 \ldots\ldots p, \ldots\ldots\ldots\ldots\ldots (7)$$

Ces équations ont une forme particulière comme on peut le remarquer en développant la relation (7) sous la forme matricielle :

$$\begin{bmatrix} R(0) & R(1) & \ldots\ldots\ldots & R(p-2) & R(p-1) \\ R(1) & R(0) & \ldots\ldots\ldots & R(p-3) & R(p-2) \\ & & & & \\ R(p-2) & R(p-1) & \ldots\ldots\ldots & R(0) & R(1) \\ R(p-1) & & & & \\ R(p-1) & R(p-2) & \ldots\ldots\ldots & R(1) & R(0) \\ R(p) & & & & \end{bmatrix} \times \begin{bmatrix} a(1) & R(1) \\ a(2) & R(2) \\ & \\ a(p-1) & \\ & \\ a(p) & \end{bmatrix}$$

$$Q(p,p) \qquad A(p) = R(p)$$

La matrice d'autocorrélation Q(p,p) est dites de «TOEPLITZ» possédant les caractéristiques suivantes:

- ✓ Symétrique,
- ✓ Positive,
- ✓ Elément situé sur la diagonal parallèle à la diagonale principale sont identique.

Afinderésoudre ce système, plusieurs algorithmes traitantlastructureparticulière de lamatriceQont été conçus pour accélérerlescalculs:

L'algorithme de GAUSS puis de CHOLESKY arrivent à résoudre rapidement ce système; mais le critère de TOEPLITZ n'a pas été pris en compte.

Cependant, un très grandpas a été réalisé dans la vitesse de calcul des a(I)avecl'arrivéedel'algorithmede LEVINSON qui se base surlastructuresymétrique de la matrice de TOEPLITZ.

Algorithme de LEVISSOS:

Pour résoudre les équations normales, l'algorithme de LEVINSON nécessite seulement $2p^2$ opération de multiplication.

On pose comme l'hypothèse la suivant:

R(i , j) = R(|i – j|) = R(k)

*initialisation: a_i (0) =1 (i=l …… p), E_0=R(0)

*Récursion: pour i=l,2, …… p

$$k_i = \frac{\left[R(i) + \sum_{j=1}^{i=1} a_{i+1}(j)R(i-j)\right]}{E_{i-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 24$$

pour 1 <=j<= i – 1

$a_i(i) = k_i$

$a_i (j)= a_{i–1} (j) + k_i a_{i–1} (i – j)$

$E_i = (l – k_i^2) E_{i–1}$

E_i :énergie résiduelle de prédiction.

K_i: coefficient de réflexion.

La solution du système est donnée par a(j) = a_p(j) pour 1< = j <= p.

- *Avantage de LPC :*

L'analyse LPC présente deux avantages principaux:

- ✓ Une grande part de la redondance de la parole est éliminée par la prédiction. Cette redondance se traduit par l'arrivée de l'élément qui ne fournisse pas d'informations nouvelles. Ainsi, savoir prédire la valeur du signal à un instant (t) en fonction des valeurs

antérieures permet de dispenser l'ordinateur de ces pseudo-informations encombrantes. Néanmoins, il faut préciser que cette redondance n'est que partielle et que les coefficients de prédiction doivent être régulièrement réajustés.

- ✓ Elle est adaptée à l'étude des phénomènes évoluent rapidement dans le temps, ce qui est le cas de parole. Les analyses fréquentielle, par contre, doivent être effectuées sur une durée suffisamment longue si l'on veut que le spectre soit déterminé avec une précision convenable.

- *Relation du cepstre avec les coefficients de prédiction :*

Le calcul du cepstre réel par la méthode de la TFD est assez coûteux en temps de calcul, mais on peut réduire ce temps de calcul on l'estimant avec les coefficients de prédictions:

$$\ln\left[\frac{1}{A_P(z)}\right] = \sum_{n=1}^{\infty} c(n)\, Z^{-n}$$

si on dérive chaque membre par rapport à Z^{-1} on aura:

$$-\frac{A'_P(z)}{A_P(z)} = \sum_{n=1}^{\infty} nc(n) Z^{-(n+1)}$$

$$-\sum_{i=1}^{P} i a_P(i) Z^{-i+1} = \left[\sum_{j=0}^{P} a_P(i) Z^{-j}\right]\left[\sum_{n=1} nc(n) Z^{-n+1}\right]$$

On aura :

$$-i a_P(i) = \sum_{n=1}^{i=1} nc(n) a_p(i-n) + ic(i)$$

$$\Longrightarrow c(i) = -a_P(i) - \sum_{n=1}^{i-1}\left(1 - \frac{n}{i}\right) a_P(n)\, c(i-n)\ , \qquad i > 0$$

C (0) = Ln σ^2 : Correspond au gain du modèle.

II-5-4. Conclusion :

Comme dans n'importe quel système de reconnaissance de formes, le système d'extraction des paramètres est indispensable pour la reconnaissance automatique de la parole. L'information perdue à ce niveau est irrécupérable aux niveaux supérieurs et risque par conséquent de provoquer des erreurs de reconnaissance. Quelle est la meilleure représentation paramétrique pour la parole? On ne sait pas. On essaie plutôt de trouver une solution que de trouver la solution. Actuellement, les représentations basées sur des notions spectres (obtenues par FFT, LPC, cepstre, etc.) sont les plus utilisées. Ces représentations semblent mieux adaptées car, comme on l'a déjà évoqué, certains comportements sus système auditif humain peuvent être expliqués par l'analyse spectrale.

Toutes les techniques de paramétrisation du signal, qu'elles soient dérivées de la FFT comme MFCC, LFCC ou du LPC comme LPCC, Ki ou d'autres que nous n'avons pas exposées ici, s'efforcent d'extraire le maximum d'information et d'en prendre le minimum. Il est difficile de faire un choix parmi ces techniques.

CHAPITRE III

Méthodes de comparaison

III-1. Introduction :

On aborde dans ce chapitre, les méthodes qui compare le mot debruité avec le mot original isolé pour calculer ta distance ou le facteur dissemblance entre les deux mots.

La reconnaissance de la parole est basée sur la comparaison du mot énoncé avec ceux du dictionnaire de référence afin d'extraire la référence la plus proche du mot. Seulement, un locuteur, même entraîné, ne peut par prononcer plusieurs fois une même séquence vocale avec exactement le même rythme et la même durée. Les échelles temporelles de deux occurrences d'un même mot ne coïncident donc pas. et les formes acoustiques issues de l'étage de paramétrisation ne peuvent être simplement comparées point à point. On peut distinguer a priori deux sources de modification de l'échelle temporelle :

- Le changement de la vitesse d'élocution qui est représentable par une transformation linéaire de l'axe des temps.
- Les variations dans le rythme de prononciation qui se traduisent par une transformation non linéaire.

Pour pallier à ce problème, on utilise un algorithme qui réalise un alignement temporel basé sur des algorithmes de programmation dynamique qui va mettre en correspondance optimale les échelles temporelles des deux mots.

III-2. Programmation dynamique :

A cette étape, les deux mots à comparer sont représentés par des vecteurs de paramètres issus de l'étage de paramétrisation du signal; ces deux vecteurs se présentent ainsi :

T = {t(1)……, t(I)} Vecteur de dimension I représentant le mot énoncé.

R = {r(1)……,K(J)} vecteur de dimension J représentant le signalréférence.

On considère un graphe dont l'ensemble de points est défini par le produit cartésien [1...I] x [1...J] (voir figure III-l).

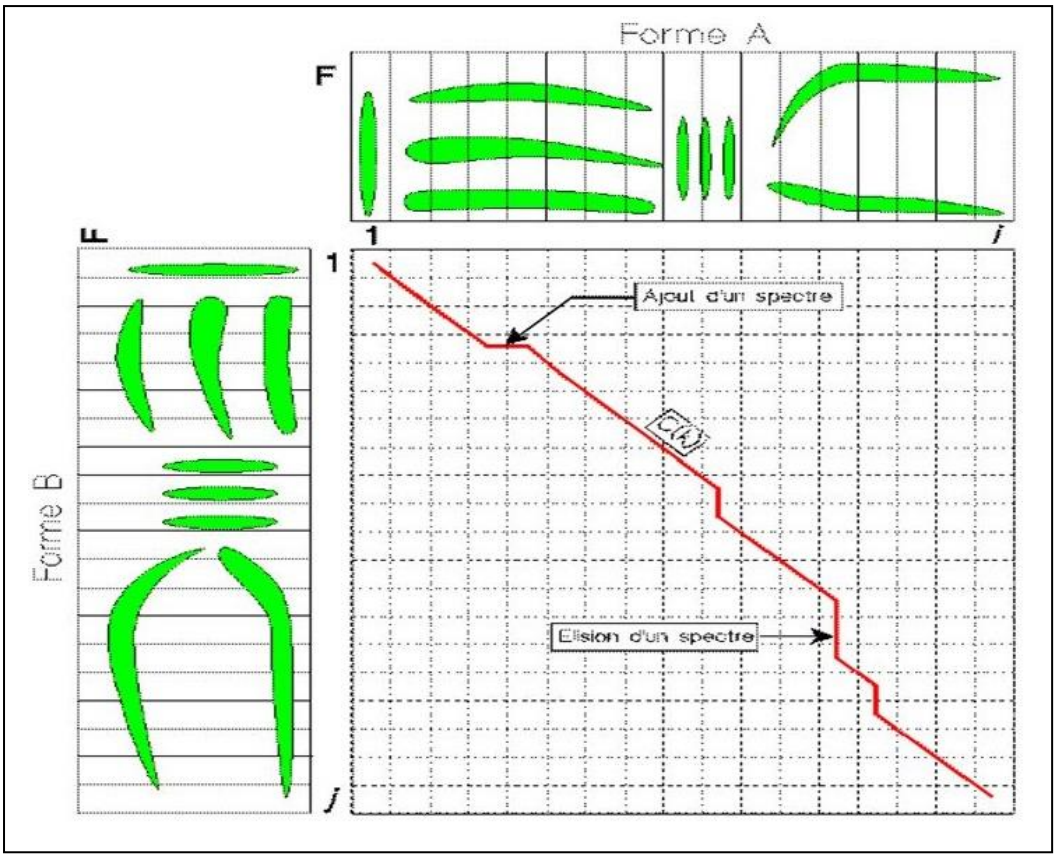


Figure III-1. Recherche du chemin de recalage

La programmation dynamique ou alignement temporel dynamique (DTW : Dynamic Time Warping) consiste à rechercher un chemin C défini par la suite de points {c(i) = (n(i), m(i)), i = 1...K.} où K est la longueur du chemin satisfaisant les contraintes suivantes:

* Les fonctions n(K) et m(K) doivent être croissantes et doivent respectés certaines conditions de continuité, pour cela plusieurs contraintes ont été définies pour déterminer les chemins valables arrivant au point (i, j).

• *Contrainte simple :* (figure III-3-a)

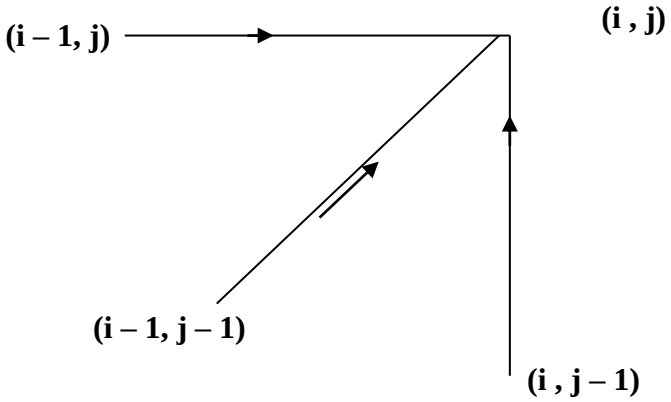
Les seuls chemins valides arrivant au point (i, j) viennent des points (i-1, j), (i-1, j-1)ou (i, j-1).

• *Contrainte de SAKOE et SHIBA :* (figure III-3-b)

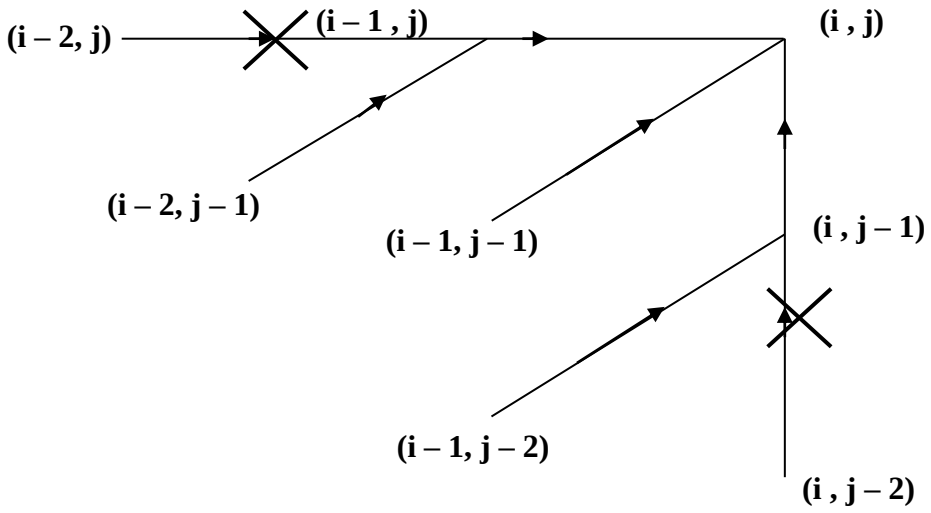
Cette contrainte interdit deux déplacements consécutifs dans la même direction si ceux-ci sont verticaux ou horizontaux, le taux de compression est compris entre 0.5 et 2.

• *Contrainte d'/TAKURA :* (figure III-3-c)

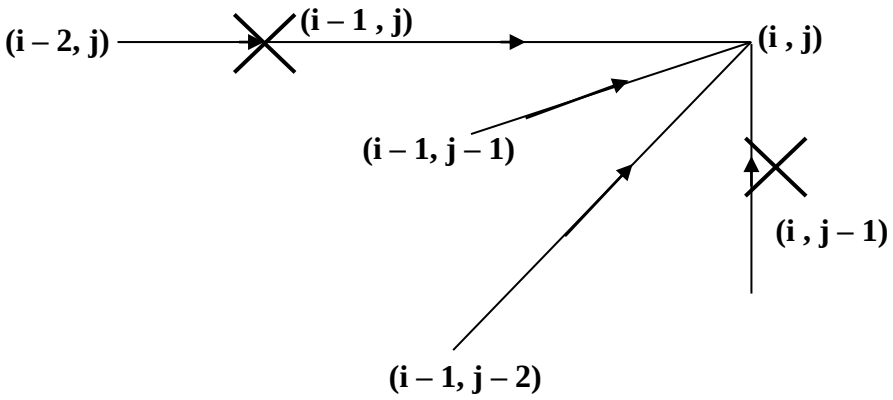
Cette contrainte interdit deux déplacements consécutifs horizontaux ainsi que tout déplacement vertical. Le taux de décompression est compris entre 0.5 et 2.



a – Contrainte simple



b – Contrainte de SAKOE et SHIBA



c – Contrainte d'ITAKURA

Figure III-2 : Les contraintes utilisées

En supposant que les limites du mot sont correctement définies,le chemin C doit vérifier les deux conditions suivantes :

$$\begin{bmatrix} c(1)=(1\ ,\ 1) \\ c(K)=(I,J) \end{bmatrix}$$

L'algorithme consiste à choisir parmi tous les chemins valables (satisfaisant les contraintes), celui qui passe par les distances d(i, j) les plus petites de sorte que la somme des distances le long de ce chemin soit minimal.

La distance finale entre les deux mots T et R est données par:

$$\boldsymbol{D}(T,\boldsymbol{R})\ \min_{\boldsymbol{c}}\left[\frac{\sum_{K=1}^{K}\boldsymbol{d}(\boldsymbol{c}(\boldsymbol{k}))\,\boldsymbol{w}(\boldsymbol{k})}{\boldsymbol{N}(\boldsymbol{w})}\right] \qquad \text{(III-1)}$$

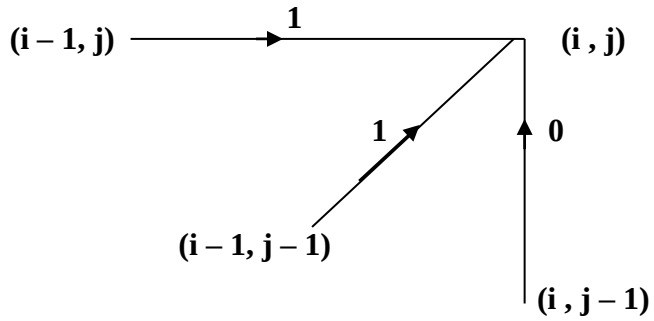
où w(k): est un coefficient de pondération appliqué sur le k^{ième}arc du chemin C.

N(w) : est un coefficient de normalisation qui dépend de la fonctionw.

Deux catégories de coefficients de pondération w(k) peuvent être utilisées, on cite :

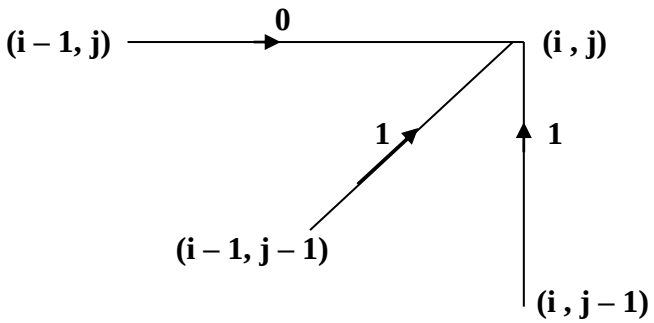
** Pondération asymétrique*

a) w(k) = n(k) – n(k – 1)



b) w(k) = m(k) – m(k – 1)

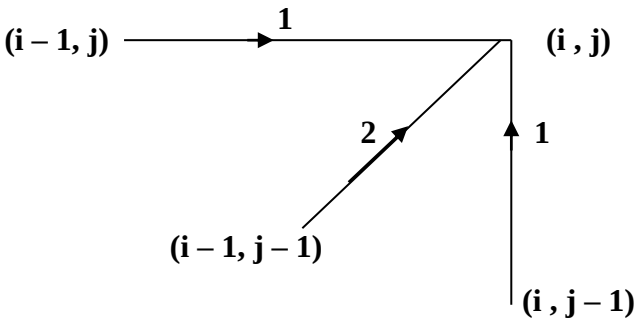
N(w) = J



** Pondération symétrique*

w(k) = m(k) – m(k – 1)

N(w) = 1 + J



On remarque que N(w) ne dépend pas du chemin C, ce qui implique que la recherche de D revient à minimiser seulement le numérateur de l'équation (III-1).

Ce problème peut être facilement résolu par un algorithme de programmation dynamique (FORD BELLMAN) qui consiste à rechercher le plus court chemin dans un graphe.

En utilisant contrainte simple décrite par la figure (III-3-a) et la fonction de pondération symétrique, la quantification de D(T,R) se définie par le calcul récursif des valeurs du paramètre g pour chaque point du graphe [1 ..I] x [I ..J] par la formule suivante :

$$g(i,j)=\min\begin{bmatrix} g(i-1,j)+d(i,j) \\ g(i-1,j-1)+2d(i,j) \\ g(i,j-l)+d(i,j) \end{bmatrix}$$

et g(1,1) = 2d (1, 1)

g (i, j) représente la distance cumulée du chemin optimal allant du point (1, l)au point (i, j).

Les valeurs des g(i, j) peuvent être calculées ligne par ligne ou colonne par colonne.

Le facteur de dissemblance D(A, B) est donné par:

$$\boldsymbol{D(A,B)=\frac{g(I,J)}{I+J}}$$

III-3. Distance spectrale :

A ce stade, les deux mots à comparer sont représentés par deux vecteurs, chacun contenant les indices retenus pour chaque fenêtre du signal correspondant aux deux mots. Pour comparer deux mots, on utilise généralement l'une des deux distances suivantes :

III-3-1. Distance associée au norme :

Elle est généralement utilisée pour les paramètres de l'analyse spectrale ou cepstrale; elle se définie par :

$$\boldsymbol{d_n(a,b) = \left(\sum_{k=1}^{P} |a_k - b_k|^n\right)^{1/n}}$$

Les distances les plus employées sont la distance d_1 vu sa simplicité de calcul et la distance d_2 (distance euclidienne).

III-3-2. Distance d ITAKURA:

Elle est utilisée pour comparer les coefficients issus de la méthode d'analyse par prédiction linéaire, elle se définit par la formule suivante:

$$\boldsymbol{d_1 = \log\left(\frac{\delta_{RT}}{\delta_T}\right)}$$

Oû :

δ_{RT}: représente l'erreur quadratique commise en supposant que le signal est modélisé par le modèle AR de la référence, il peut être calculé par la formule suivante: $\delta_{RT} = aQa^t$

δ_R:représente l'erreur quadratique calculée lors de l'analyse par prédiction linéaire: $\delta_T = bQb^t$

Q: Matrice d'autocorrélation correspondante au mot enregistré.

Donc:

$$\boldsymbol{d_1 = \log\left(\frac{aQa^t}{bQb^t}\right)}$$

La valeur de bQb^t est donnée directement par l'algorithme de LEVINSON, E(p), lors de la phase d'extraction des coefficients LPC.

La valeur de aQa^t peut être évaluée rapidement en utilisant la formule simplifiée suivante:

$$\boldsymbol{aQa^t = r(0)r_a(0) + 2\sum_{k=1}^{P} r(k)r_a(k)}$$

Où :

$r_a(k)$: sont les coefficients d'autocorrélation sur le segment du signal correspondant à b.

$r_b(k)$: sont les coefficients d'autocorrélation calculés sur les coefficients de prédiction a.

$$\boldsymbol{r_a(k) = \sum_{i=0}^{p-k} a(i)a\,(k+1)}$$

On remarque que le calcul du taux de dissemblance entre deux fenêtres du signal utilise les coefficients $r_a(k)$ au lieu des a_i, donc il vaut mieux les calculer dans la phase d'apprentissage et de les stocker dans le dictionnaire afin d'alléger les calculs dans la phase de reconnaissance qui peut prendre un temps considérable si le vocabulaire est grand.

Remarque :

Il faut noter que d_1 n'est pas une distance au vrai sens du terme puisque $d(a,b) \neq d(b,a)$. En pratique, *b* est le vecteur correspondant au mot énoncé et *a* le vecteur du mot référence.

III-3-3. Distance cepstrale:

La distance cepstrale est d'un usage très courant en reconnaissance de la parole, ou les coefficients du cepstre réel sont donnés par :

$$\boldsymbol{\ln f(\theta) = \sum_{n} c(n)\, exp(-jn\theta)}$$

La distance cepstrale est une distance spectrale, elle est donnée par la relation :

$$\boldsymbol{d_P = \left[\frac{1}{2\pi}\int_{-\pi}^{\pi}|V(\theta)|^P d\theta\right]^{1/P}}$$

V(θ) = Ln (f(θ)) – Ln (f'(θ)), ou f(θ) et f'(θ) sont des densités spectrales d'énergie

$$\boldsymbol{d_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\sum_n \big(c(n)-c'(n)\big)\,exp(-jn\theta)\right|^2 d\theta}$$

$$= \sum_{1=-\infty}^{+\infty}\big(c(1)-c'(1)\big)^2 = [c(0)-c'(0)]^2 + 2\sum_{i=1}^{+\infty}[c(1)-c'(1)]^2$$

Ou la distance cepstrale est basée sur un nombre finis L:

$$d_{cep} = [c(0)-c'(0)]^2 + 2\sum_{i=1}^{L}[c(1)-c'(1)]^2$$

Les variations fines du spectre sont d'autant mieux prises en considération que L est élevé.

Les coefficients de prédiction d'ordre p définissant les coefficients Cepstraux c(n) d'ordre L =p et par extension plus ⇨ le choix de L = 2p conduit a une excellente approximation de d_2.

Le terme c(0) correspond au gain du modèle:

$$c(0) = \ln\sigma^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\ln\left|\sigma/A\left(e^{j\theta}\right)\right|^2 d\theta$$

CHAPITRE IV

Algorithme basé sur la mesure de la distance cepstrale

Introduction

Nous avons vu au chapitre III que les méthodes de reconnaissance de la parole étaient généralement basées sur l'utilisation de la mesure de distance cepstrale.

Afin d'améliorer les performances de cette méthode, nous y adjoindrons un algorithme de débruitage basé sur la méthode de soustraction spectrale.

IV-1. Mise en œuvre de l'algorithme :

IV-1-1. Organigramme générale :

L'algorithme réalisé permet de traiter des signaux de parole échantillonnés à 8000 ou 16000 Hz. les fichiers de mots sont constitués d'entiers (16 bits).

Il possède un bruit permettant d'ajouter au signal de type bruit blanc gaussien.

Le signal est d'abord analysé par tranche de 512 échantillons. On applique le principe de recouvrement des tranches suivantes.

Chaque tranche de 512 échantillons contient donc une tranche de signal « nouveau » de 256 échantillons.

Il y a possibilité d'appliquer au signal une pondération par une fenêtre de HAMMING. Ceci permet d'atténuer les effets d'extrémité et de donner un poids moindre aux anciens échantillons.

Nous calculons ensuite la transformée de Fourier rapide de chaque tranche de 256 échantillons ainsi obtenue. Après extraction de module carré des vecteurs spectraux des vecteurs obtenus, donc le spectrogramme du mot bruité. D'un autre côté, on calcule la valeur moyenne du spectre du bruit.

On soustrait cette moyenne du spectre de signal bruité, ce qui nous donne le spectrogramme du mot débruité.

Il ne reste plus que le calcul des coefficients cepstraux (mot débruité. mot isolé, mot bruité), suivis de la distance cepstrale entre le mot débruité et le mot isolé et la distance cepstrale entre le mot bruité et le mot isolé.

Cette distance est calculée dans différents cas pour permettra une comparaison des résultats.

L'organigramme complet est présenté à la figure (IV-1).

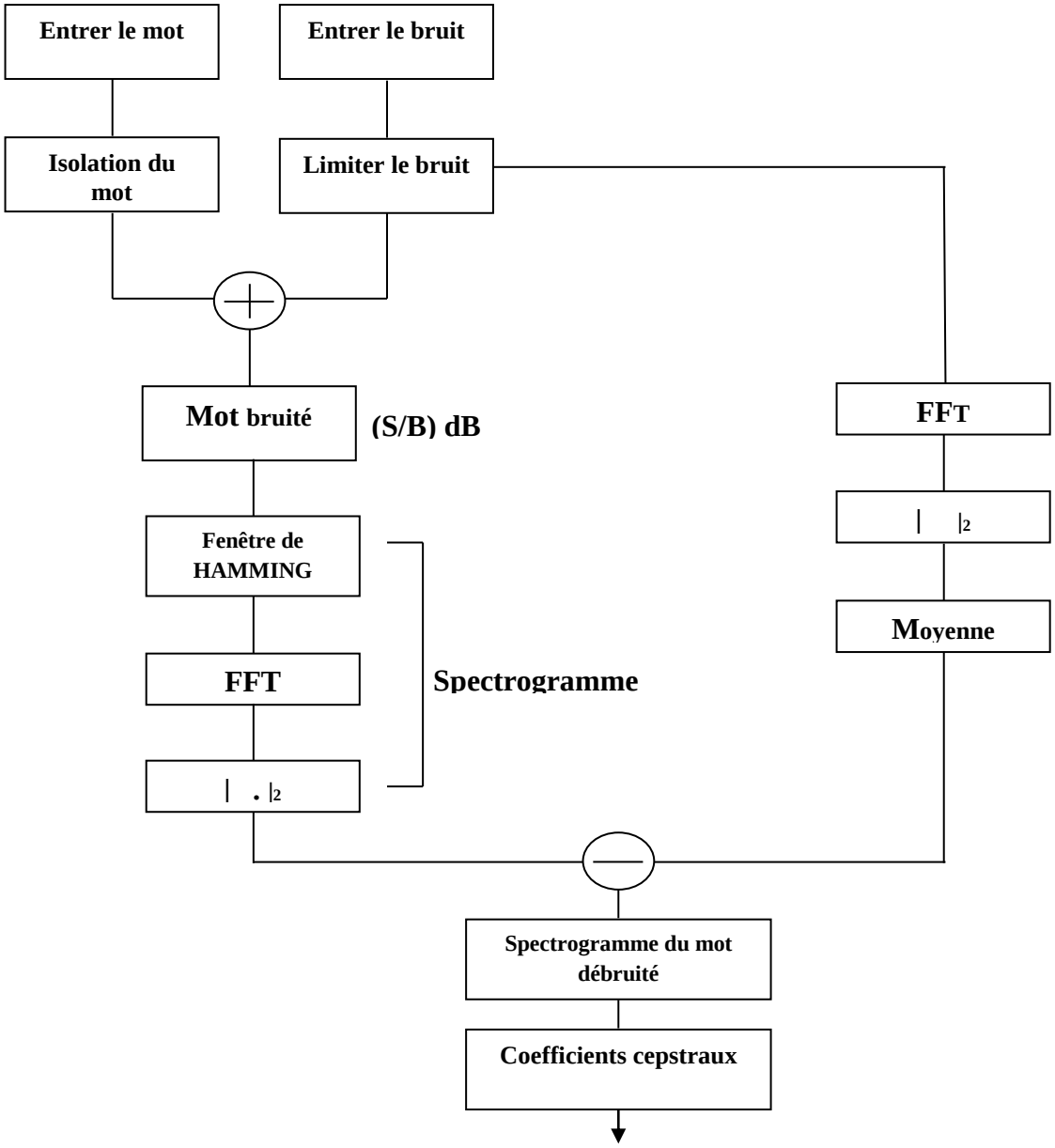


Figure IV-1. : Organigramme de l'algorithme de débruitage

IV-1-2. Détail de certains éléments :

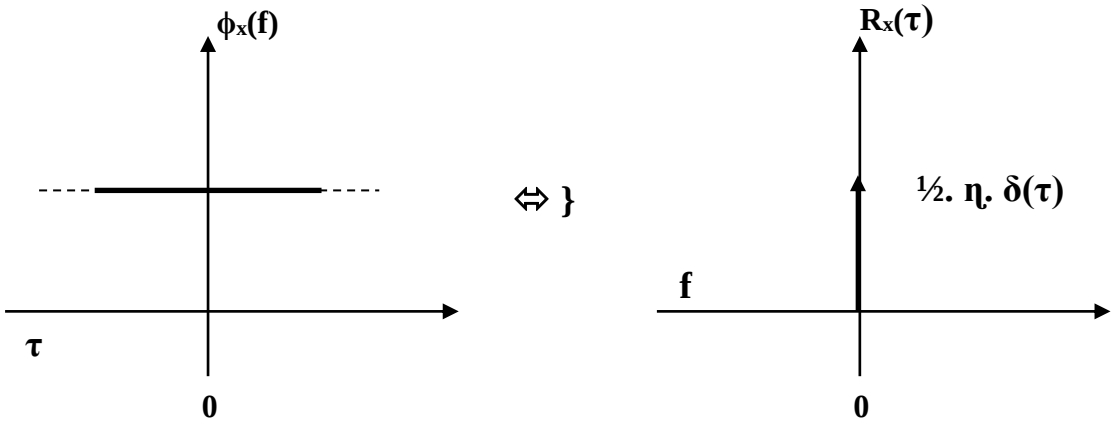
1. Le bruit blanc gaussien :

Le bruit blanc est un processus aléatoire dont la densité spectrale de puissance (D.S.P) est constante pour toutes les fréquences :

- $\phi_x(f) = ½.\eta = cst$ pour $|f| < \infty \Rightarrow$ modèle théorique.
- La D P (distribution) est Gaussienne.

La fonction d'autocorrection du bruit blanc est une impulsion de Dirac de poids ½.η :

Rx (τ) = TF-1(½. η) = ½.η. δ(τ)



Si :τ = $t_2 - d_1 = 0$ alors $R_x(\tau) \neq 0$: le bruit corrélé à τ = 0.

Si :τ = $t_2 - d_1 \neq 0$ alors $R_x(\tau) = 0$: les variables $x(t_1)$ et $x(t_2)$ sont totalement non corrélées.

Si : le processus aléatoire est gaussien : les deux variables sont indépendantes.

•*Bruit blanc à bande limité :*

Le bruit blanc à bande limité est un processus aléatoire dont la (DSP) est constante dans une bande de fréquence finie et nulle ailleurs.

$$\Phi_x(f) = \begin{cases} 1/2.\eta & pour\ f_1 \leq |f| \leq f_2 \\ 0 & ailleurs \end{cases}$$

Si : $f_1 = f_2 - \frac{B}{2}$ et $f_2 = f_0 - \frac{B}{2}$, alors le bruit blanc de type passe-bande.

$$\phi 1(f) = \frac{1}{2}.\eta.rect\left[\left(\frac{f+f_0}{B}\right)\right] + \frac{1}{2}.\eta.rect\left[\left(\frac{f-f_0}{B}\right)\right]$$

et :

$R_1(\tau) = \eta.B.\sin c(B\tau).\cos(2.\pi.f_0.\tau)$

Si : $f_1 = 0$ et $f_2 = B$, alors le bruit blanc de type passe-bas.

$$\phi 1(f) = \frac{1}{2}.\eta.rect\left[\left(\frac{f}{2.B}\right)\right]$$

et :

$R_1(\tau) = \eta.B.\sin c(2.B\tau)$

2. Coefficients cepstraux :

Ils sont calculés comme décrit au chapitre II.

3. La distance cepstrale :

Elle est calculée comme décrit au chapitre III.

4. Représentation graphique :

Le programme réalisé permet de visualiser :

- le mot avant l'isolation.
- le mot après isolation.
- le signal de bruit.
- le mot bruité.
- le rapport signal au bruit.
- les spectrogrammes des mots isolés, mot bruit et mot débruité.
- Le spectre de bruit et sa valeur moyenne.
- les distances cepstrales entre le mot isolé et deux mots (bruit et débruité).

Tous ceci nous a permet de montrer plus clairement les performances de la méthode.

Le programme est présenté en annexe A.

CHAPITRE V

Résultats Expérimentaux

Des essais ont été effectués avec différents mots, différents rapports signal/bruit.

V-1. Mot « Bonjour.wav » :

V-1-1. Expérience N° 1 : S/B= 0.5

1ère étape : **le mot « Bonjour.wav» isolé avec : début = 50000 fin = 150000**

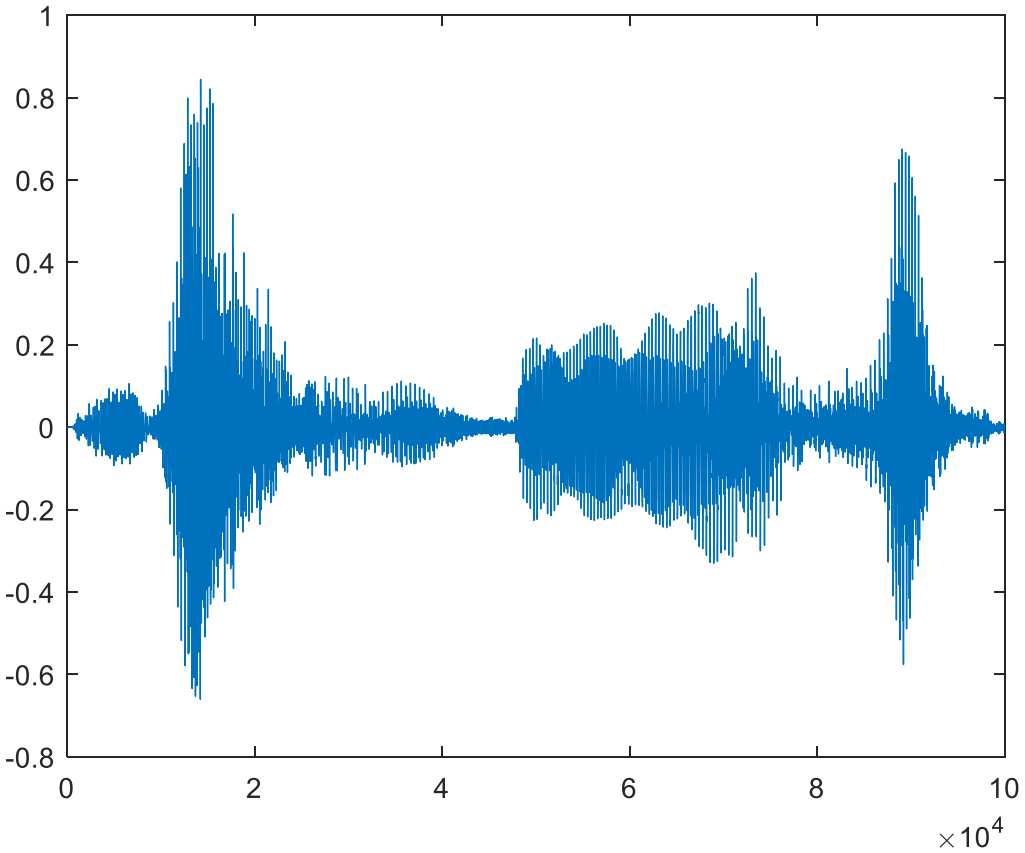


Figure V-1-1. a. : Représentation temporelle du le mot « Bonjour.wav» isolé

2ème étape : **le mot « Bonjour.wav » bruité**

- **Le rapport signal sur bruit est S/R (dB) = 0.5**

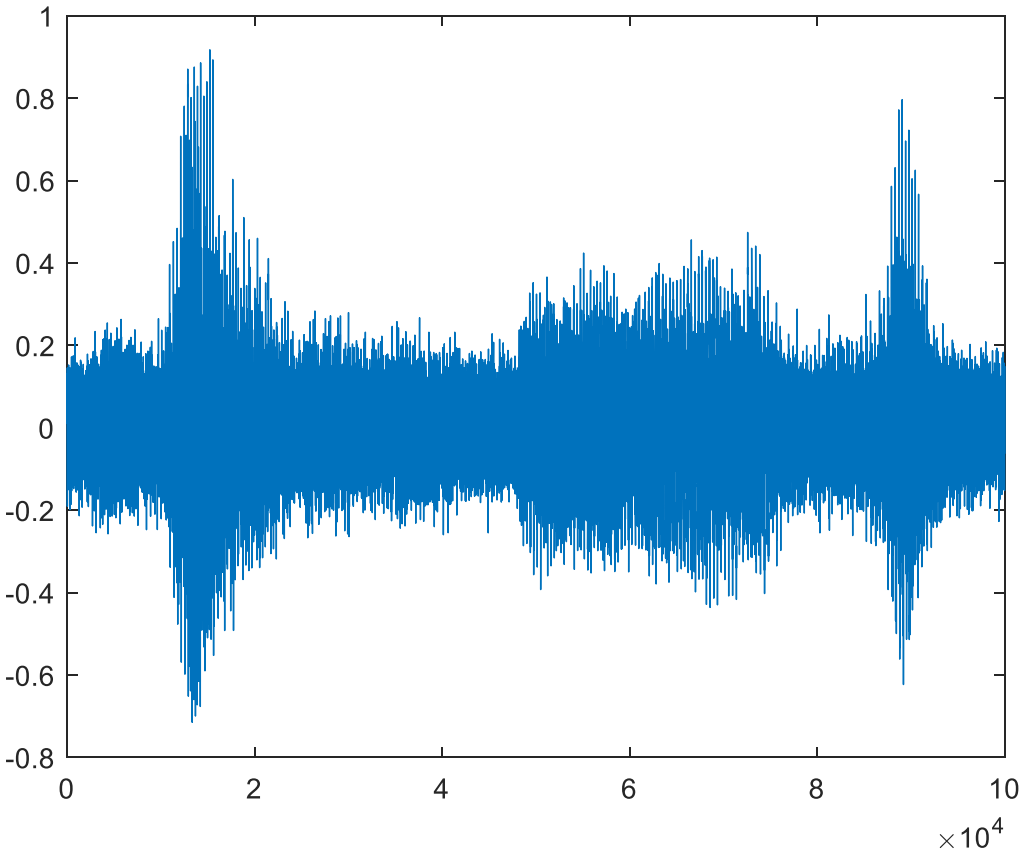


Figure V-1-1. b : Représentation temporelle du mot « Bonjour.wav»bruité

3ème étape : **calcul le spectrogramme du mot bruité avec :**

- **Le type de fenêtre est fenêtre de HAMMING.**
- **Taille de fenêtre est 512 échantillons.**
- **Le chevauchement de la fenêtre est 256 échantillons.**
- **La résolution de FFT est 8192.**

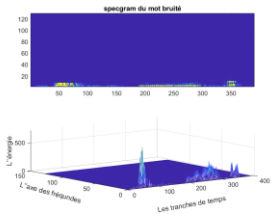
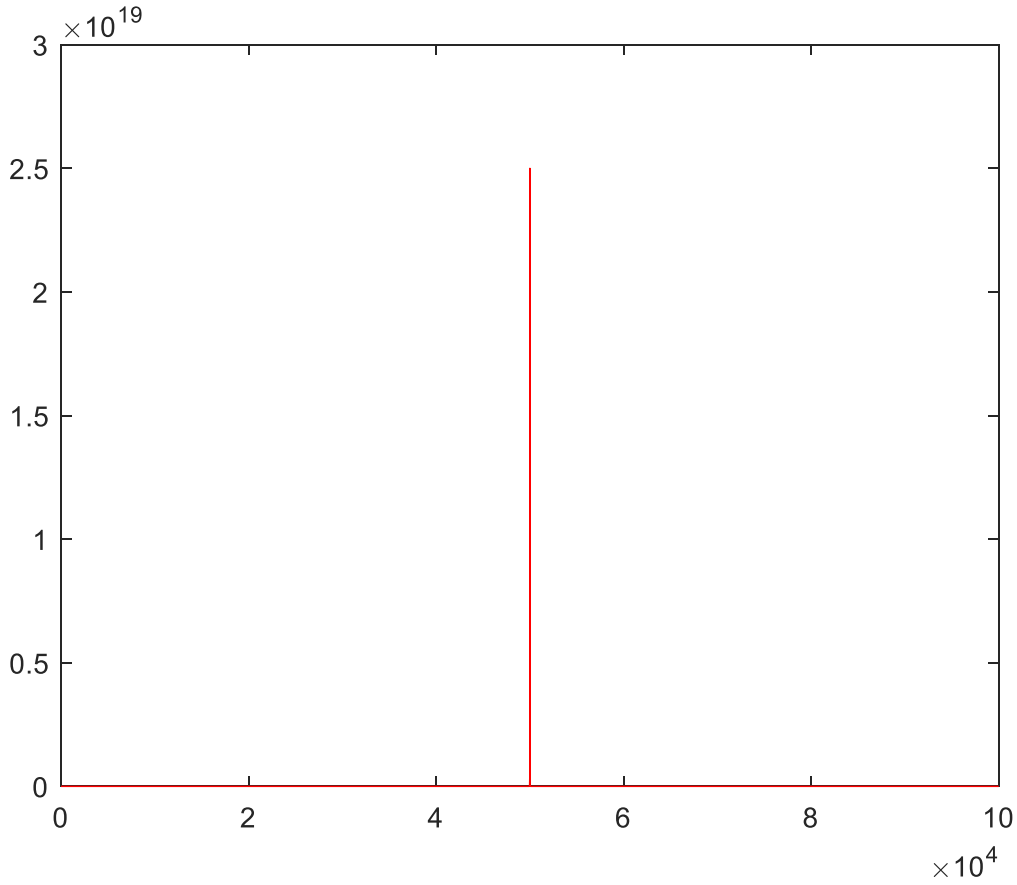


Figure V-1-1.c. : spectrogramme du mot bruité

4ème étape : **calcul la valeur moyenne du spectre de bruit m**


×10^19
3
2.5
2
1.5
1
0.5
0
0
2
4
6
8
10
×10^4

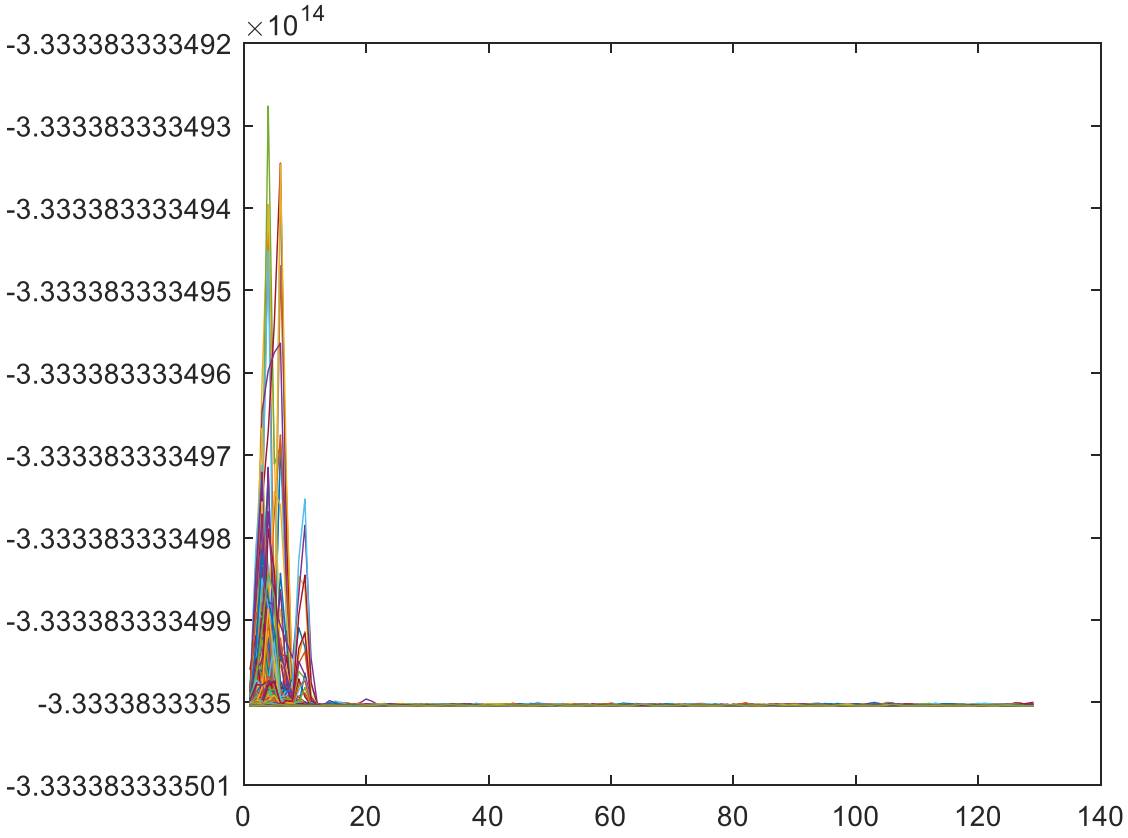
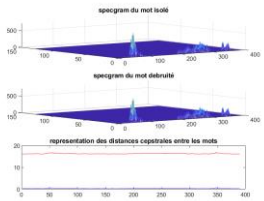

×10^14
-3.333383333492
-3.333383333493
-3.333383333494
-3.333383333495
-3.333383333496
-3.333383333497
-3.333383333498
-3.333383333499
-3.3333833335
-3.333383333501
0
20
40
60
80
100
120
140

Figure V-1-1.d : La valeur moyenne du spectre du bruit

5ème étape : **Le spectrogramme du mot débruité**



Entre le mot isolé et le mot bruité.

Entre le mot isolé et le mot débruité.

Figure V-1-1.e : Spctrogramme du mot isolé, débruité et la distance

*7ème étape :***calcule :**

- **Les coefficients cepstraux du mot débruité.**
- **Les coefficients cepstraux du mot isolé.**
- **Les coefficients cepstraux du mot bruité.**
- **Distance cepstrale entre le mot débruité et le mot isolé :md1=0.2543**
- **La distance cepstrale entre le mot bruité et le mot isolé :md=16.8821**

V-1-2. Expérience N° 2 : S/B= 5

1ère étape : **le mot « Bonjour.wav» isolé avec : début = 50000 fin = 150000**

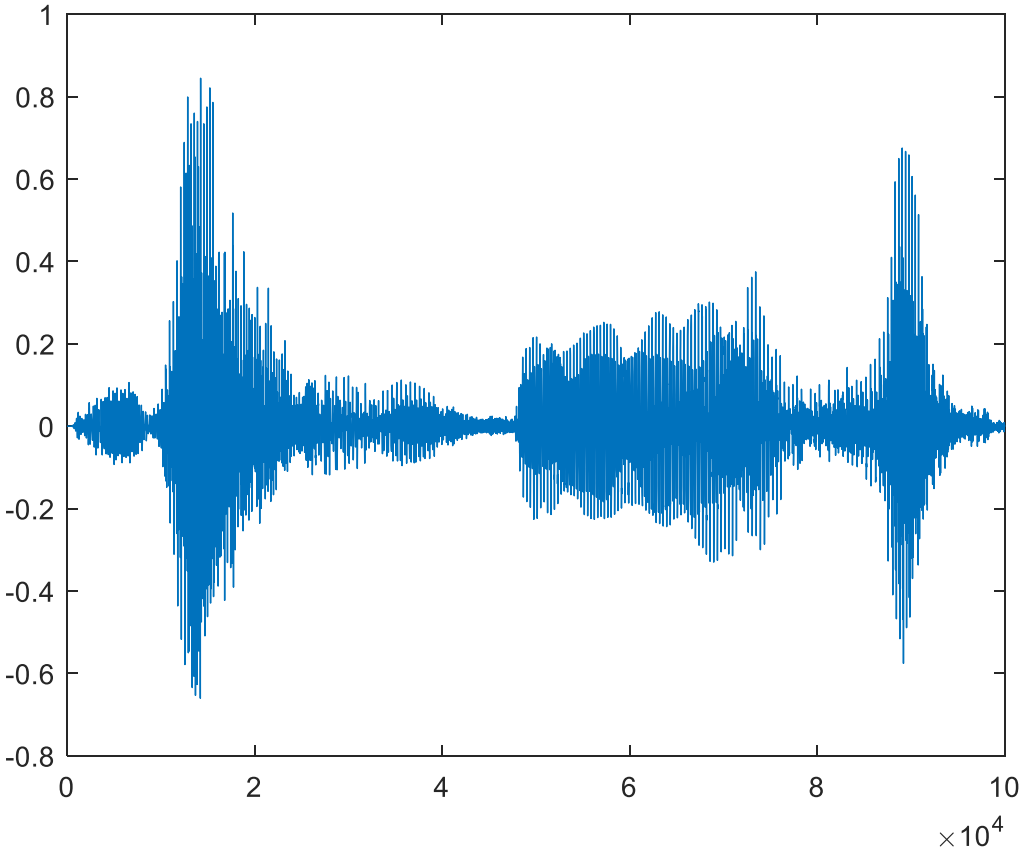


Figure V-1-2. a. : Représentation temporelle du le mot « Bonjour.wav» isolé

2ème étape : **le mot « Bonjour.wav » bruité**

- **Le rapport signal sur bruit est S/R (dB) = 5**

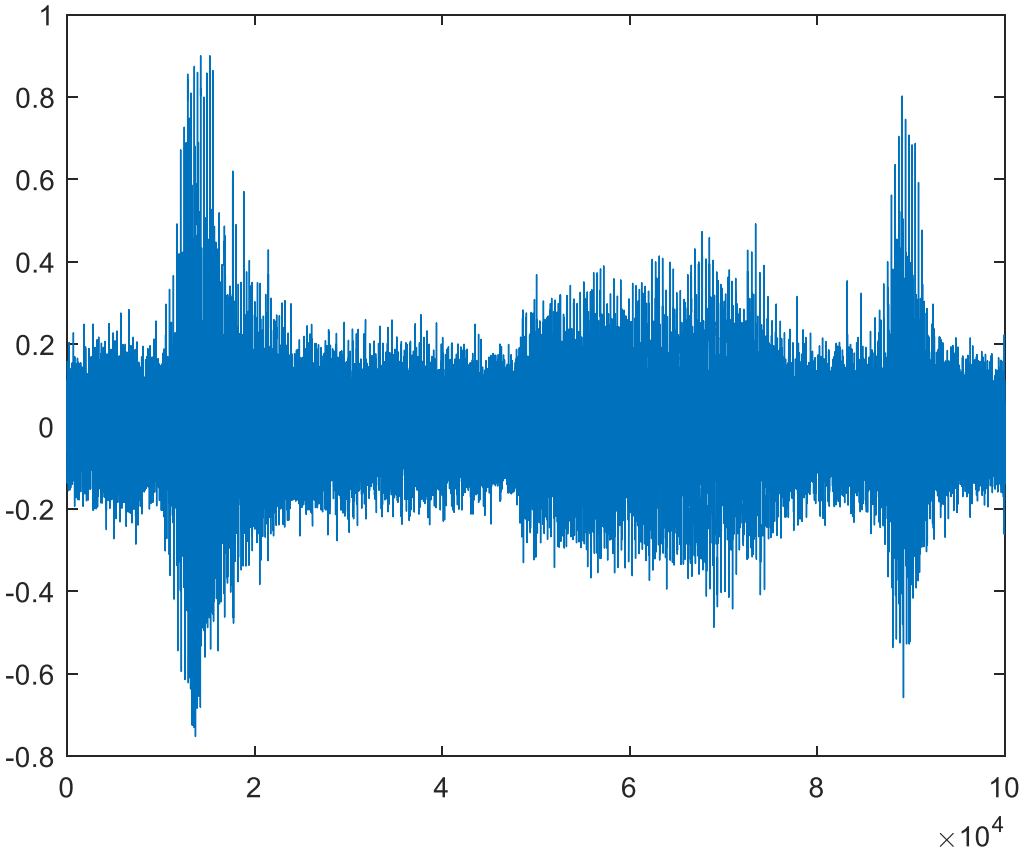


Figure V-1-2. b : Représentation temporelle du mot « Bonjour.wav»bruité

3ème étape : **calcul le spectrogramme du mot bruité avec :**

- **Le type de fenêtre est fenêtre de HAMMING.**
- **Taille de fenêtre est 512 échantillons.**
- **Le chevauchement de la fenêtre est 256 échantillons.**
- **La résolution de FFT est 8192.**

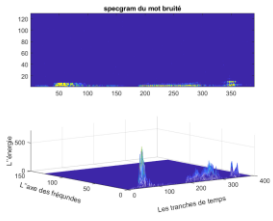
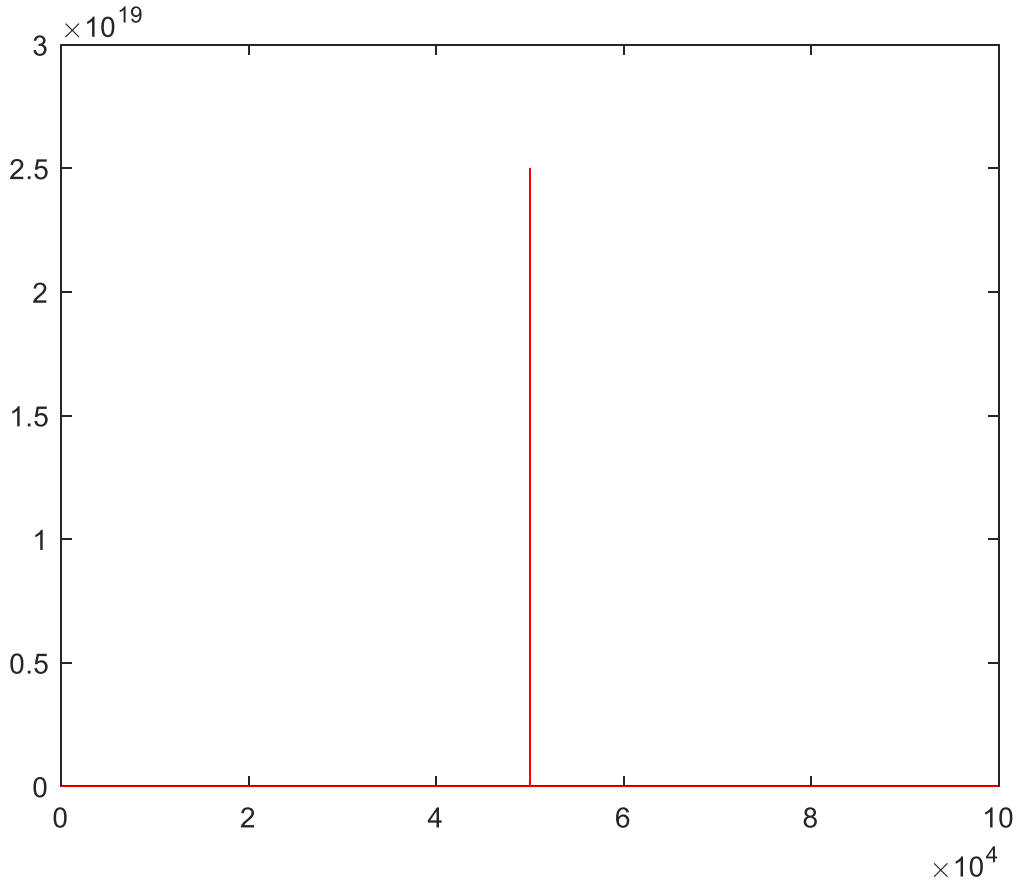


Figure V-1-2.c. : spectrogramme du mot bruité

4ème étape : **calcul la valeur moyenne du spectre de bruit m**


×10^19
3
2.5
2
1.5
1
0.5
0
0
2
4
6
8
10
×10^4

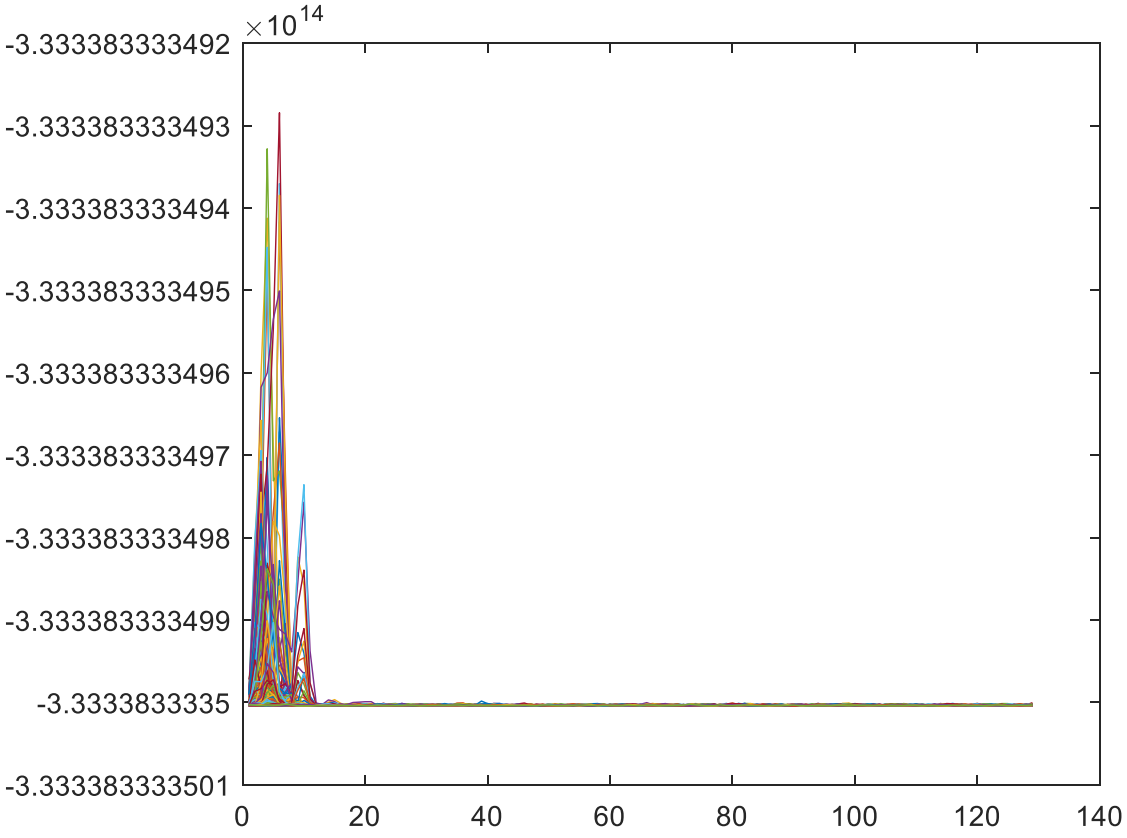
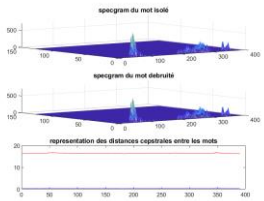

×10^14
-3.333383333492
-3.333383333493
-3.333383333494
-3.333383333495
-3.333383333496
-3.333383333497
-3.333383333498
-3.333383333499
-3.3333833335
-3.333383333501
0
20
40
60
80
100
120
140

Figure V-1-2.d : La valeur moyenne du spectre du bruit

5ème étape : **Le spectrogramme du mot débruité**



Entre le mot isolé et le mot bruité.

Entre le mot isolé et le mot débruité.

Figure V-1-2.e : Spctrogramme du mot isolé, débruité et la distance

*7ème étape :***calcule :**

- **Les coefficients cepstraux du mot débruité.**
- **Les coefficients cepstraux du mot isolé.**
- **Les coefficients cepstraux du mot bruité.**
- **Distance cepstrale entre le mot débruité et le mot isolé :md1=0.2716**
- **La distance cepstrale entre le mot bruité et le mot isolé :md=16.3768**

Tableau de comparaison

	S/B (dB)	*md*	*md1*	*(md1-md)/md1*
Noise 1	**0.5**	**16.8821**	**0.2543**	**98.49%**
Noise 2	**5**	**16.3768**	**0.2716**	**98.34%**

Conclusion Générale

CONCLUSION GENERALE

L'algorithme que nous avons présenté, pour l'isolation des mots en milieu bruité, fournit des résultats très acceptables dans la plupart des cas.

Comme il utilise la notation de distance cepstrale, il peut être mis en œuvre très facilement, avec tous les algorithmes de reconnaissance de la parole qui calculent, pour leur fonctionnement, les paramètres cepstraux. Ceci possède l'énorme avantage de ne pas augmenter excessivement volume de calcule nécessaire à la reconnaissance proprement dite.

La méthode de soustraction spectrale, n'est pas limitée aux seules applications de reconnaissance, mais peut être utilisée dans de nombreuses autres applications comme, par exemple un traitement visant l'amélioration de l'intelligibilité d'un signal de parole perturbé par du bruit additionnel.

ANNEXES

ANNEXE A

PROGRAMMATION MATLAB

```
function [M,c,m,cs_mdb,cs_mi,md,cs_M,md1]=noising(a,SNR)  % SNR=b
% ********************************************************************
% ********************************************************************
% Université de M'sila
% Faculté de Technologie
% Département: Electronique
% Option: systèmes télécommunications
%
% Réalisé par :
% BEDDIAR ANTAR
%
% Etude et élaboration d'algorithme de débruitage de signaux:
% Application à la reconnaissance de la parole
% ********************************************************************
% ********************************************************************
%
% [M,c,m,cs_mdb,cs_mi,md,cs_M,mdl]=noising(a,b)
%
% Les paramètres d'entrée:
%   % a représente le mot original.
%   % b représente le bruit.
%
% Les paramètres de sortie:
%   % M représente le mot bruité.
%   % c représente le spectrogramme du mot bruité.
%   % m représente la moyenne du spectre de bruit.
%   % cd_mdb représente les coefficients cepstraux du mot débruité.
%   % cd_mi représente les coefficients cepstraux du mot isolé.
%   % md1 représente la distance cepstrale entre le mot isolé et le mot
débruité.
%   % cd_M représente les coefficients cepstraux du mot bruité.
%   % md représente la distance cepstrale entre le mot isolé. et le mot
débruité.

[q,Fs] = audioread (a) ; pause; % 249165
%[s,f2] = audioread (b) ; pause; % Noise signal
e = isolation (q);
%end
figure(01); plot(e) ;
soundsc (e,Fs) ;
pause


%% Add gaussian noise
 M = awgn(e,SNR,'measured');
%for i = 1:length (e)
%M(i) = e(i) + s(i) ;
%end
figure(02); plot(M ) ;
soundsc(M,Fs);pause


%calcule le rapport signal au bruit
%Ee = 0 ;
%Es = 0 ;
%for i = 1:length (e)
```

```
%Ee = Ee + abs(e(i))^2 ;
%Es = Es + abs(s(i))^2 ;
%end

disp ('le rapport signal sur bruit est SNR');
%R = 10 * log10 (Ee/Es);

%calcule le specgram du mot bruité et du mot original isolé
%taille = input ('donner le chevauchement entre les fenêtres=\n');
%chevauchement = input ('donner le chevauchement entre les fenêtres=\n');
%resolution = input ('donner la résolution de la FFT=\n');

resolution = min(256,length(M)); %nfft
taille=resolution ;
chevauchement = length(window)/2 ; %numoverlap

x = specgram(M,resolution,Fs,hamming(taille),chevauchement) ;  % mot bruité
spectre_2 = specgram(M,resolution,Fs,hamming(taille),chevauchement) ;  %
mot isolé
c = abs(x).^2;
se = abs (spectre_2).^2;   %spectre_2 not defined
subplot(211) ; image (c) ; set(gca,'ydir','normal') ;
title ('specgram du mot bruité') ;
subplot(212) ; mesh (c) ;
h = ylabel ('L''axe des fréqundes') ;
set(h,'rotation',-15)
h = xlabel('Les tranches de temps') ;
set(h,'rotation',10) ;
zlabel ('L''énergie') ;
%pause
%close

%calcule la valeur moyenne du spectre du bruit
n = (1:length(e)) ;  % error n = s(1:length(e))
f_bruit = abs(fft(n)).^2    ;
f_bruit = fftshift(f_bruit) ;
figure(04); plot(f_bruit,'r');pause
m = 0 ;

for i = 1:length (f_bruit)
m = m + f_bruit(i)/length(f_bruit) ;
end
figure(05); plot(m ) ;

disp('la valeur moyenne du spectre du bruit est m') ;
line([0 length(f_bruit)],[m m] );
spec_db = c - m ; plot(spec_db ) ;
%pause
%close
figure(06);
subplot(311) ; mesh(se)          ; set(gca,'ydir','normal') ;
title ('specgram du mot isolé') ;
subplot(312) ; mesh(c)           ; set(gca,'ydir','normal') ;
title ('specgram du mot debruité');
subplot(313) ; mesh(spec_db)     ; set(gca,'ydir','normal') ;
title('specgram du mot debruité')   ;
%pause
%close

%calcule des coefficientscepstraux du mot débruité
lnspec_db = (1/2) * log(abs(spec_db)) ;
cs_mdb = abs(ifft(lnspec_db)) ;
```

```
%calcule des coefficients cepstraux du mot isolé
ln_se = (1/2) * log(se) ;
cs_mi = abs(ifft(ln_se)) ;

%calcule la distance cepstrale entre le mot débruité et le mot original
isolé
tranch1 = cs_mdb(1:16,:) ;
tranch2 = cs_mi(1:16,:)  ;
[z,y] = size(tranch2);% size© ;
disp('le nombre de colonne est') ; length(z);
disp('le nombre de ligne est') ; length(y);

for i = 1:y
  d1(i)= (tranch1(1,1)-tranch2(1,1)^2) ;
  for j = 2:16
    d1(i)= d1(i) + 2 * ( tranch1(j,1)-tranch2(j,i) )^2 ;
  end
end

md = 0  ;
d1 = d1';

for i = 1:y
 md = md + d1(i) ;
end

md = md / y ;

%calcule des coefficients cepstraux du mot bruité
c1 = (1/2) * log(c) ;
cs_M = abs(ifft(c1)) ;

%calcule la distance cepstrale entre le mot bruité et le mot original isolé
tranch3 = cs_M(1:16,:) ;
tranch2 = cs_mi(1:16,:);

for i = 1:y
d2(i) = (tranch3(1,1)-tranch2(1,1)^2 );
for k = 2:16
d2(i)= d2(i) + 2 * ( tranch3(k,1)-tranch2(k,i) )^2 ;
end
end

md1 = 0 ;
d2 = d2';

for i = 1:y
md1 = md1 + d2(i) ;
end

md1 = md1 / y ;
i=1:y;
plot(i,d1,'r',i,d2,'b');
title('representation des distances cepstrales entre les mots');
```

```
%*************************FONCTION CONVERSION*****************************

%function [y,f]=conversion(chaine)
%[y,f]=audioread(chaine)
%Plot(y);
%Soundsc(y,f);

%*************************FONCTION ISOLATION*****************************



function a=isolation(chaine)
disp('debut de signal est moins que');length(chaine)
debut=input('debut');
disp('fin de signal est moins que');length(chaine)
fin=input('enter fin=');
n=fin-debut;
for i=1:n
 a(i)=chaine(i+debut);
end
%pilot(a);


%************************************************************************
%************************************************************************
```

.

ANNEXE B

LES ALGORITHMES DE PROGRAMMATION DYNAMIQUE

**********Algorithme DTW**********

début

g(0,0) = 0

pour j de1à J

faire

g(0,j) = INFINI

Fin j

pour i de1à I

faire

g(i,0) = INFINI

pour j de1à J

faire

d = dist (i,j)

g(i,j) = min (g(i – 1,j) + d, g(i – 1, j – 1) + 2d, g(i, j – 1) + d)

Fin j

Fin i

D = g(I,J)/(I+J)

Fin

**Remarque :*

L'algorithme analyse chaque point du graphe, alors qu'il existe des points où on est sur que le chemin de recalage ne passera pas, sinon on aura une compression irréaliste, donc d'autres critères sont ajoutés pour améliorer la vitesse de calcul de cet algorithme.

Réduction de l'espace de comparaison

Au lieu de rechercher le chemin de recalage sur tout le graphe, nous allons réduire l'espace de comparaison, afin de minimiser le nombre de calcul de distance lors de la phase de détermination du chemin de recalage.

Pour cela on définit une fenêtre pour laquelle, il n'est pas nécessaire d'inspecter les points situés en dehors de celle-ci.

Méthode 1 :

A chaque itération k, on ne considère que les points qui vérifient contraintes suivantes :

$|i(k) - j(k)| \leq R$ **où R est un entier positif**

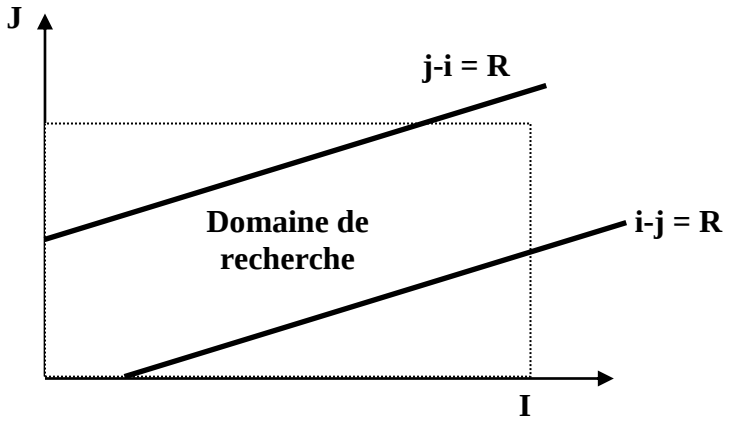


Figure I : Réduction de l'espace de recherche

Le domaine de recherche est représenté par la figure I.

Le paramètre R est en relation avec la taille de l'espace de comparaison, plus il est petit plus l'espace est réduit et le temps de calcul amélioré. Seulement une valeur trop faible de R peut altérer les résultats de l'algorithme et conduire à des résultats erronés, donc ilest recommandé de ne pas prendre une valeur trop faible pour R.

**********Algorithme**********

début

g(0) = 0

***pour** j **de**1**à**J*

faire

g(j) = INFINI

***Fin** j*

***pour** i **de**1**à** I*

faire

vt1 = INFINI

pour j dede max (1, i-R) à min (J, i + R)

faire

d = dist (i, j)

vt2 = min (g(j)) + d, g(j-1) + 2d, vt1 + d)

g(j-1) = vt1

vt1 = vt2

Fin j

Fin i

D = g(mag (J,J + R)) / (I+J)

Fin

Méthode 2 :

Pour cette méthode, on définit deux limites maximales de la pente doit se situer le chemin de recalage. L'espace de comparaison est définie par un parallélogramme dont la surface est inférieure par rapport à la première méthode.

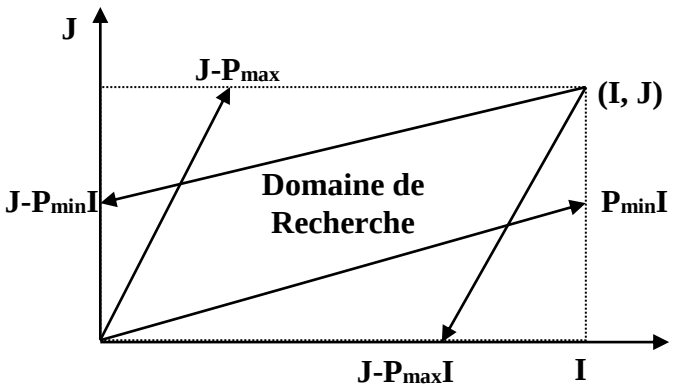


Figure II : Réduction du domaine de recherche

Oû P_{max} : représente la pente maximale.

P_{min} : représente la pente minimale.

Par exemple, si on adopte les contraintes de SAKOE et SHIBA les deux pentes auront les valeurs suivantes :

P_{min} = 0,5 , P_{max} = 2

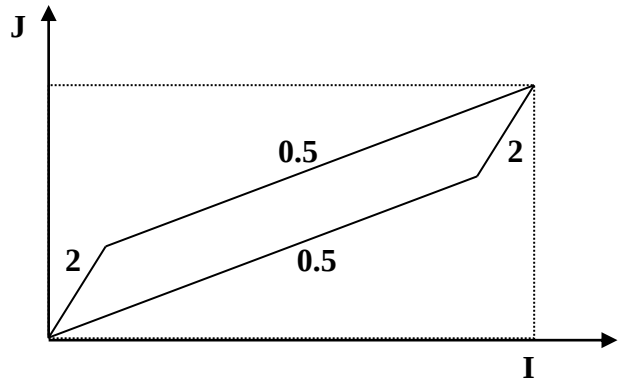


Figure III : Représentation du domaine de recherche pour la contrainte de SAKOE et CHIBA

Dans ce cas, si J = ½ I l'espace de recherche se réduit à une droite (la diagonale) et la recherche du chemin avorté.

** Remarque :*

Nous avons opté pour première méthode que nous avons trouvé simple à implémenter par rapport à la seconde.

Bibliographie

BIBLIOGRAPHIE

1. R. Boité et M. Kunt, « Traitement de la parole », Presses polytechniques romandes, 1987.

   Nous y avons trouvé :

   - **Les fondements de la prédiction linéaire.**
   - **Description de la DTW.**
   - **Description des méthodes statistiques.**

2. Calliope, « La parole et son traitement automatique », Masson, 1989.

   Ce traité collectif fait un résumé de toutes les méthodes utilisées pour le traitement de la parole, il contient une description détaillée des globales et analytiques, ainsi que la DTW.

3. M. Kunt, « Traitement numérique des signaux », Dunod, 1990.

   Nous y avons trouvé toutes les bases nécessaires pour le traitement numérique telle :

   - **La TFD.**
   - **Le fenêtrage.**
   - **Analyse homomorphique.**

4. R. Dehak, « Elaboration d'un système de commande vocale », PFE info, USTO, 1987.

5. Sites internet :**http://perso.aricia.fr/alliun/parole/PRParole.htm**

6. Introduction au traitement automatique de la parole

   * Notes de cours /DEC2 *