République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche scientifique
Université de Biskra

FACULTE DES SCIENCES ET DE LA TECHNOLOGIE

DEPARTEMENT GENIE ELECTRIQUE

Mémoire de magister
En Electronique

Option : ***Communications***

Thème

Techniques de détection de la période du pitch par les méthodes temps fréquence et temps échelle.

Présenté par :

AJGOU Riad

Soutenu le : 11 /03 /2010 devant le Jury composé de :

O.KAZAR	Maitre de Conférences	Université Biskra	Président
A.BENAKCHA	Maitre de Conférences	Université Biskra	Examinateur
Z.BAARIR	Maitre de Conférences	Université Biskra	Examinateur
S.SBAA	Maitre de Conférences	Université Biskra	Rapporteur

Remerciements

C'est avec émotion que je tiens à remercier tous ceux qui m'a aidé à élaborer ce travail.

Mes premiers remerciements à Mr OKBA KASAR, maitre de conférences à l'Université de Biskra, qui m'a fait l'honneur de présider la commission d'examen.

Je remercie ensuite sincèrement toux qui m'ont fait l'honneur de participer à mon jury de mémoire de magistère, tout d'abord, je tiens à témoigner ma gratitude à Mr A.BENAKCHA, maitre de conférences au d'épatement de génie électrique, d'avoir examiné, corriger et recorriger ce travail et d'avoir me comblé avec ces conseils précieuses au cours de mes recherches pour la réalisation de ce travail, je tiens à remercier également Mr Z.BAARIR, maitre de conférences au sein de département de génie électrique, qui trouve mes profondes gratitude pour ses conseils, d'avoir accepté juger, corriger et recorriger ce travail.

Toute ma gratitude à mon encadreur Dr S.SBAA pour son soutient, ses conseils, et son assistance et patience et encouragement tout au long de cette période de travail ainsi et m'avoir permis d'améliorer mes capacités scientifiques et m'a introduit au domaine du traitement de la parole, et d'avoir me donné un bon état d'esprit pour la recherche.

Dédicace

Je dédie ce travail à toute

ma famille,

mes parents, mers frères, mes sœurs.

Sommaire:

À

Liste des Tableaux

Liste des Symboles et abréviations

ZCR : Zéro crossing rate (passage par zéro).

$\overline{\mathbf{E_0}}$: L'énergie moyenne.

SFM : Spectral flatness mesure (mesure de l'égalité spectrale).

$\mathbf{S_{xx}}$: Densité spectrale de puissance.

F0 : Fréquence fondamentale.

T0 : La période fondamentale.

R(k) : La fonction d'autocorrélation.

W(n) : Fenêtre de pondération.

hk(n) : Réponse impulsionelle.

AMDF: Averaged Magnitude Difference Function (fonction de la différence de la magnitude moyennée).

ASDF: Average Square Difference Function.

SIFT: Simplified Inverse Filtering Tracking (filtrage inverse simplifiée pour la detection de pitch).

LPC : Prédiction linéaire.

Fs : Fréquence d'échantillonnage.

S(f) : Le spectre d'amplitude du signal.

SNR : Rapport signal sur bruit.

V/NV : Voisé/non voisé.

ACF_LPC : Fonction d'autocorrélation basée sur une analyse LPC.

ACF : La fonction d'autocorrélation.

HPS: Harmonic Product Spectrum (produit de spectre des harmoniques).

SHS : Sub-harmonic Summation (sommation des harmoniques).

RTF : Représentation temps-fréquence.

WV : Wigner-Ville.

DWV : Distribution de Wigner-Ville.

DPWV : Distribution pseudo Wigner-Ville.

PWVL : Pseudo Wigner Ville Lissé

DPWVL : Distribution de pseudo Wigner-Ville lissé.

CW : Choi-Williams.

DCW: Distribution de Choi-Williams

AF : La fonction d'Ambigüité.

STFT : Transformée de Fourrier à courts terme.

TO : Transformé en ondelette.

TOC : Transformée en ondelette continue.

TDO : Transformée dyadique d'ondelette.

TH : Transformée d'Hilbert.

Introduction générale

La connaissance de *la fréquence fondamentale* (pitch en Anglais) du signal parole est utile dans nombreux domaines de la reconnaissance vocale, identification du speaker, le codage de la parole et sa synthèse, …etc.

La détermination de la fréquence fondamentale des signaux parole a été étudiée assez largement, les signaux de la parole posent des problèmes spécifiques. D'une part, ils ne peuvent pas être considérés comme stationnaires; en effet, l'évolution de l'enveloppe spectrale, est parfois très rapide. D'autre part les signaux de parole sont constitués schématiquement d'une alternance des zones voisées et non voisées où la décision est difficile.

Le problème est donc de déterminer la période du signal d'excitation du signal voisé. Le travail présenté ici est celui de l'étude du mécanisme de production de la parole en s'appuyant sur la prédiction linéaire et les algorithmes de détermination de la fréquence fondamentale. Bien que le signal vocal fait partie des signaux non stationnaires c'est alors nécessaire de rechercher d'autres formes de représentations qui contribuent mieux à ce type de signaux où les représentations temps-fréquence et temps-échelle sont les plus commodes à l'analyse du signal parole.

Notre travail se divise en sept chapitres :

- le premier chapitre présente une étude et analyse de la parole (le phénomène de la production de la parole, spectre, harmoniques…..).

- le deuxième chapitre explore les différents algorithmes pour la détection du pitch

- le troisième chapitre : dans ce chapitre on explore plusieurs algorithmes appartenant aux domaines temporels. Nous essayerons de mettre en évidence le comportement de chaque algorithme vis-à-vis des problèmes liés à la détermination de la fréquence fondamentale.

- le quatrième chapitre explore plusieurs algorithmes appartenant aux domaines fréquentiels (Cepstre, Hps, SHS…etc.).

- le cinquième chapitre explore quelques outils d'analyse des signaux non stationnaires appartenant au plan temps-fréquence, où on s'intéresse au signal vocal du point de vue représentation et estimation de la fréquence fondamentale.

- le sixième chapitre présente une approche de détection de la fréquence fondamentale fondée sur l'utilisation de la décomposition on ondelettes du signal parole.

- le septième chapitre procède à une étude comparative des méthodes classiques (temporelles, fréquentielles), temps- fréquence, et temps-échelle.

En fin et avec une conclusion générale et perspective on termine notre mémoire.

I.1 Introduction :

L'information portée par le signal parole peut être considéré de plusieurs façons. On distingue généralement plusieurs niveaux de description non exclusifs : *acoustique, phonétique* et *phonologique*, au niveau acoustique, on s'intéresse essentiellement au signal que l'on tentera de caractériser par son intensité, sa fréquence, son timbre et ses propriétés statistiques, au plan phonétique, on considère la génération des sons, les phonèmes qui composent un mot et les classes auxquels ils se rattachent, et enfin la phonologie s'attache à décrire le rythme, la prosodie, la mélodie d'une phrase.

Avant de vouloir analyser le signal parole, il est important de commencer par comprendre ce qu'est la parole, quel est son contenu spectral, quelles sont les parties qui la composent.

I.2 le signal parole :

On définit la parole et on explore les caractéristiques et la modélisation du mécanisme de production de la parole.

I.2.1 Qu'est ce que c'est la parole ?

La parole est le principal moyen de communication dans toute société humaine. Son abstraction par rapport à un support physique en fait un moyen de communication très simple à utiliser : il est plus facile de parler à quelqu'un que de lui écrire ou de lui faire un schéma. L'ère industrielles a par ailleurs permis, mettant en place des moyens d'enregistrement, et donc de sauvegarde, de se hisser au rang de l'écrit pour conservation de la connaissance [2].
L'information portée par le signal de parole peut être analysé de différentes façons.

I.2.2 Fonctionnement de l'appareil vocal :

L'ensemble du système vocal se compose des poumons et du conduit trachéobronchique, du larynx, et du conduit vocal, formé par le pharynx et les cavités nasales et orales [2]. La figure I.1 représente l'appareil phonatoire et modèle mécanique de production de la parole.

- L'ensemble poumons et conduit trachéobronchique se comporte comme un générateur d'air qui alimente le larynx.
- le larynx est l'ensemble de cartilages articulés, ligaments muscles et muqueuses, qui grâce à son action sur les cordes vocales (muscles élastiques), permet de déterminer la nature du flux d'air qui va exciter le conduit vocal.

- Le conduit vocal par des organes articulateur imprime au son émis les caractéristiques spécifiques permettant de distinguer les différents phonèmes et ceci en tant que :

  - résonateur de l'onde glottique pour la production des voyelles.
  - générateur de bruit pour la production des consonnes.

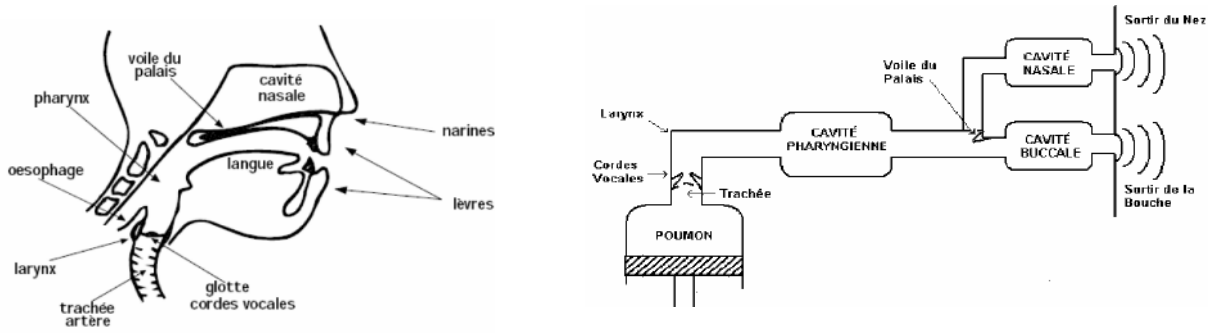


Figure I.1. Appareil phonatoire et Modèle mécanique de production de la parole [3].

I.2.3 Caractéristiques articulatoires et acoustiques de la parole :

Physiquement la parole est un phénomène vibratoire résultant de deux composantes :

- Le passage de l'air à travers les cordes vocales (source d'excitation) produit un signal périodique dont la fréquence caractérise la hauteur de la voix.
- Un système résonant (figure I.2) est composé de quatre cavités : pharyngale, buccale, nasale et labiale. La Figure I.2 représente les quatre résonateurs de l'appareil phonatoire

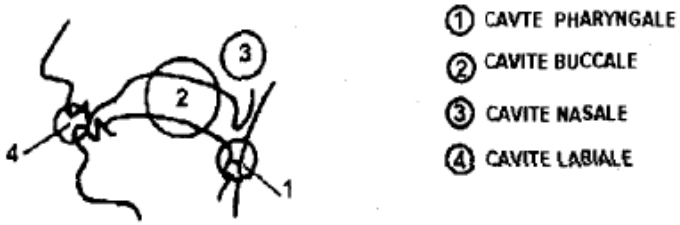


Figure I.2 Les quatre résonateurs de l'appareil phonatoire [4]

Avant d'être rayonné au niveau des lèvres, le signal acoustique se propage à travers le conduit vocal dans lequel il est filtré. En effet les cavités supraglotiques possèdent des fréquences de résonance qui renforcent certaines régions du spectre des sources excitatrices (source sonore et source bruité). Les maximas de la courbe de la réponse en fréquence du conduit vocal sont appelés "Formants". La figure I.3 représente le spectre d'un signal parole présenant trois resonances formantiques.

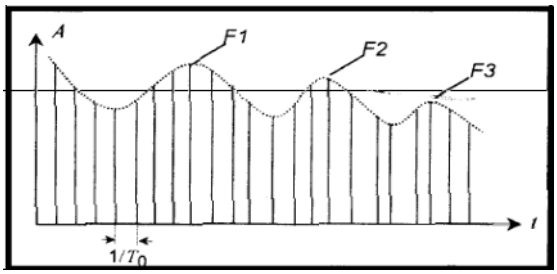


Figure I.3 Spectre présentent trois resonances formantiques [2].

La fréquence du premier formant peut varier de 200 Hz à 800 Hz, celle du second formant de 900 Hz à2400 Hz. Il existe des formants d'ordres supérieurs pouvant aller jusqu'à 5 KHz ; l'ensemble des formants contribue en particulier à caractériser le « timbre » de la voix.

I.2.3.1 Opposition voisée - non voisée :

On peut établir deux catégories de sons, selon qu'ils sont dus :

- à une vibration périodique : "son voisé ou sonore" ;
- à une génération de bruit à travers une constriction du conduit vocal : "son non voisé ou sourd".

Dans la première catégorie de sons, la forme de signal glottique est sensiblement triangulaire, son spectre est riche en harmoniques. La figure I.4 représente la production d'un son voisé.
Dans la seconde catégorie, la figure I.5 exprime que le signal est apériodique son spectre est relativement uniforme sur une large bande de fréquence

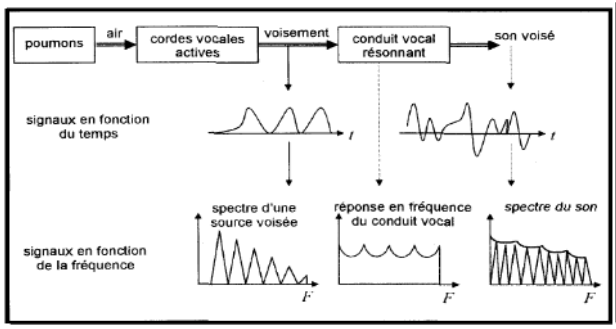


Figure I.4 production d'un son voisé [2].

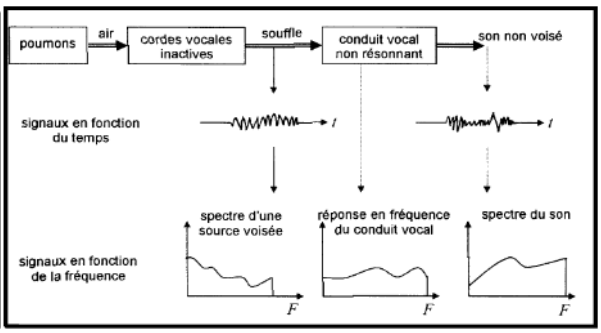


Figure I.5 production d'un son non voisé [2].

I.2.4 Propriétés statistiques du signal parole :

Le signal de parole est une réalisation particulière d'un processus aléatoire non stationnaire c'est-à-dire que ces propriétés statiques changent au cours du temps. Nous faisant l'hypothèse de quasi-stationnarité sur des périodes allants de 10 à 35 ms [10].

I.2.4.1 Détection du voisement :

L'étude du signal de parole a montré que l'analyse spectrale constitue une indication importante pour la détection de voisement, les sons voisé contrairement aux sons non voisé présentant d'avantage d'énergie vers les basses fréquences, ainsi que le calcul de nombre d'échantillons successifs signe opposés constitue un deuxième indicateur de voisement ou non-voisement du signal parole. Pour le taux de passage par zéro pour un segment de signal de « N » échantillons, la formule est représentée comme suit [6,7]:

$$ZCR = \sum_{N=0}^{N-1} \left| sgn[x(n)] - sgn[x(n-1)] \right| \qquad I.1$$

L'énergie pour le même segment de signal est [6]:

$$\overline{E} = \sum_{n=0}^{N-1} x_m^2(n) \qquad I.2$$

Il est important d'éclairer qu'il y a plusieurs méthode de détection de voisement non voisement, exemple une méthode efficace basé sur la géométrie du signal (SFM), le son est considéré voise si la valeur de SFM est comprise entre 0.4 est 0.1 [8] [9]:

$$SFM = \frac{(\prod_{k=0}^{N-1} X_j(k))^{1/N}}{\frac{1}{N}\sum_{n=0}^{N-1} X_j(k)} \qquad I.3$$

La figure I.6 montre un segment voisé(e) et un segment nom voisé, on remarque que les valeurs de taux de passage par zéro sont plus élevées pour les sons non voisés que pour les sons voisés.

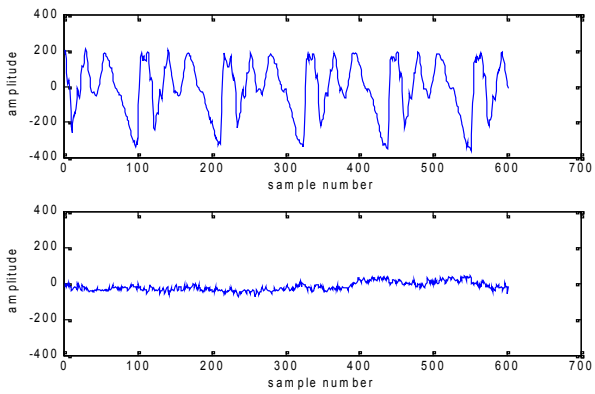


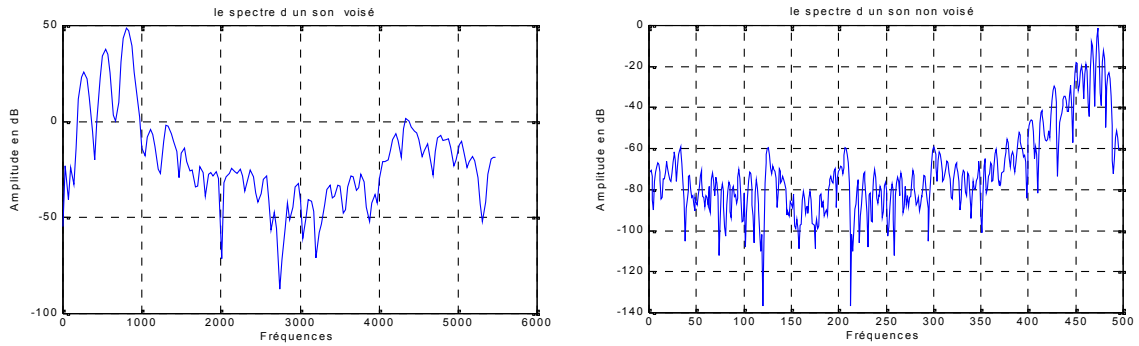
Figure I.6 son voisé (en haut), son non voisé (en bas).

I.2.4.2 Densité spectrale de puissance à court terme :

La densité spectrale de puissance est la transformée de Fourier de la fonction d'autocorrélation, telle que S_{xx} (DSP) peut être calculée après application d'une fonction de pondération

$$S_{xx}(\theta) = \sum_{k=-k}^{k} R_{xx}(k)\, h(k) \exp(-jk\theta) \qquad ;\theta = 2\pi T_e \qquad \text{I.4}$$

Nous remarquons à partir de la figure I.7 une structure périodique (tranche voisée), fine. Elle correspond aux harmoniques d'excitation glottique. Les maximums de l'enveloppe de son spectre correspondent aux formants. Par contre le spectre d'un signal non voisé ne présente aucune structure particulière sauf une accentuation vers les hautes fréquences. La Figure I.7 représente les densités spectrale de puissance (a : un son voisé, b : un son non voisée).



(a) (b)

Figure I.7 Les densités spectrale de puissance (a : un son voisé, b : un son non voisée)

I.2.5 Spectrogramme :

Il est souvent intéressant de représenter l'évolution temporelle du spectre à court terme d'un signal, sous la forme d'un ***spectrogramme***. L'amplitude du spectre y apparaît sous la forme de niveaux de gris dans un diagramme en deux axes : temps et fréquence. Il mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants.

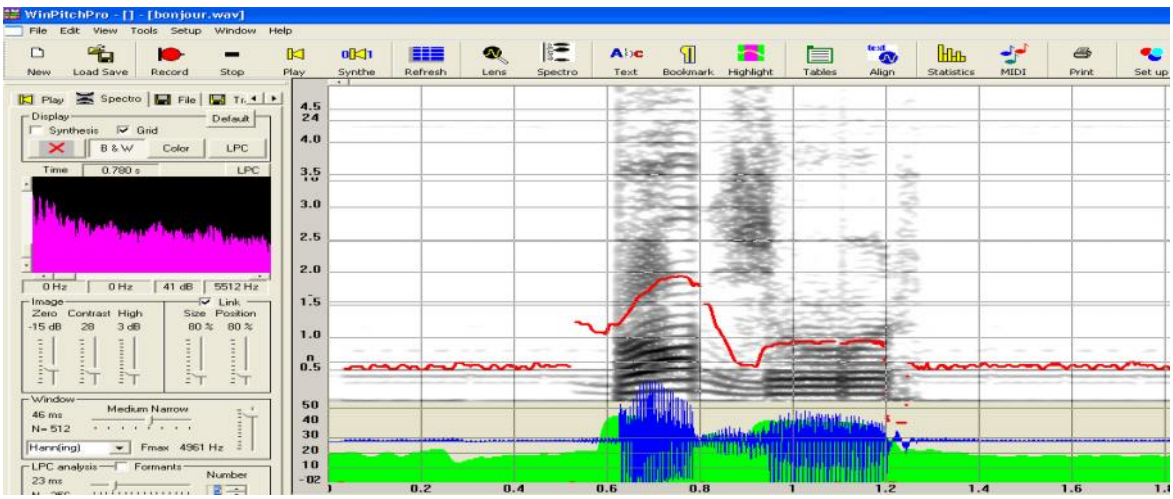


Figure I.8 Spectre de mot « Bonjour », visualisation de la parole réalisé avec le logiciel WinPitchPro [11]

Le spectrogramme est un outil de visualisation utilisant la technique de la transformée de Fourier Le spectrogramme permet de mettre en évidence les différentes composantes fréquentielles du signal à un instant donné.

I.2.5.1 Principe :

Le signal doit être tout d'abord pré accentué par un filtre du premier ordre égaliser les hautes fréquences. Cette phase de préaccentuation du signal est suivie par une phase de fenêtrage nécessaire du fait de la théorie de la transformée de Fourier. Dans cette méthode d'analyse, le signal est supposé stationnaire. Il faut convoluer le signal avec une fenêtre temporelle glissante puisque chaque calcul de spectre nécessite de convoluer le signal avec la fenêtre temporelle à un instant particulier.

I.2.5.2 Type de spectrogramme :

Le choix de la taille de la fenêtre, en nombre de points de convolution, est également important vis-à-vis de la qualité de l'analyse fréquentielle obtenue.

a. Les spectrogrammes à bande large :

Ils sont obtenus avec des fenêtres de pondération de faible durée (typiquement 10ms) ; ils mettent en évidence l'enveloppe spectrale du signal, et permettent par conséquent de visualiser l'évolution temporelle des formants.

b. Les spectrogrammes à bande étroite :

Ils sont moins utilisés. Ils mettent plutôt la structure fine du spectre en évidence : les harmoniques du signal dans les zones voisées y apparaissent sous la forme de bande horizontale.

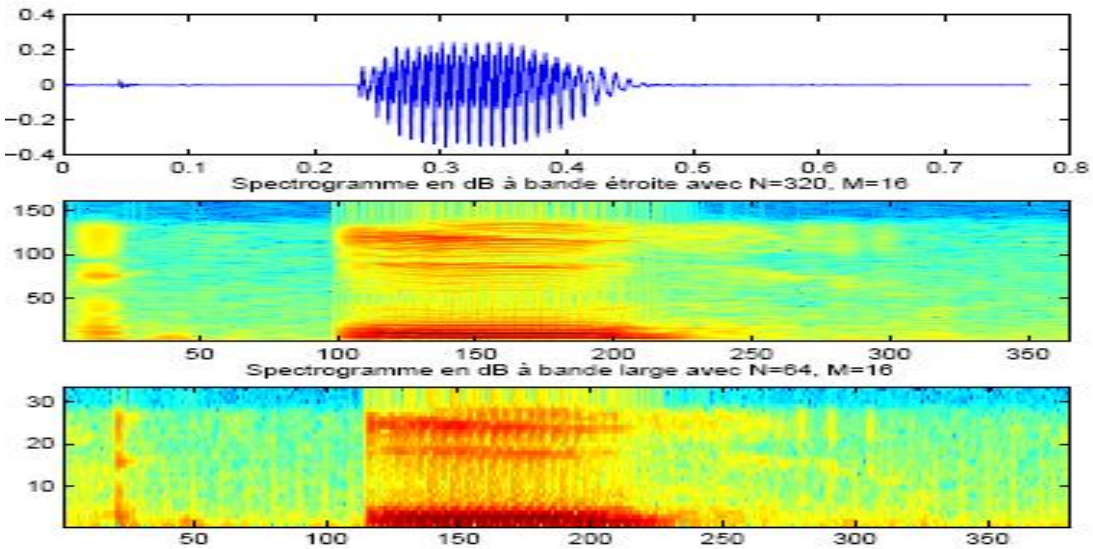


Figure I.9 Un son voisé (phonème "i") et son Spectrogrammes à bande étroite et bande large.

I.2.6 Caractéristiques phonétiques :

Les caractéristiques phonétiques sont : les phonèmes, voyelles, consonnes.

I.2.6.1 Phonème [2]:

Un phonème est la plus petite unité présenté dans la parole [2]. Le nombre de phonèmes est toujours très limité (normalement inférieur à cinquante) et ça dépend de chaque langue.

I.2.6.2 Voyelles [2]:

Les voyelles sont des sons voisés qui résultent de l'excitation du conduit vocal par des impulsions périodiques de pression liées aux oscillations des cordes vocales. Il y a deux types de voyelle : les ***voyelles orales*** (i, e, u, ...) qui sont émises sans intervention de la cavité nasale et les ***voyelles nasales*** (ã, e~ , …) qui font intervenir la cavité nasale. Chaque voyelle se caractérise par les résonances du conduit vocal qu'on appelle "***les formants***". En général, les trois premiers formants sont suffisants pour caractériser toutes les voyelles.

I.2.6.3 Consonnes [2]:

Les consonnes sont des sons qui sont produits par une turbulence créée par le passage de l'air dans une constriction du conduit où une source périodique liée à la vibration des cordes vocales s'ajoute à la source de bruit (les consonnes voisées).

I.2.7 Les paramètres acoustiques de la parole :

La parole est constituée d'une succession de phonèmes, c'est la plus petite unité phonatoire susceptible de changer un mot en un autre, la langue française comporte *37 phonèmes* .

I.2.7.1 Formant :

Un formant est une fréquence résonnante du système acoustique. Le formant se caractérise par la présence de maxima spectraux, c'est-à-dire des zones où les harmoniques sont intenses. Il est utilisé généralement dans la phonétique ou l'acoustique pour décrire les vibrations des tractus vocaux ou des instruments musicaux.

I.2.7.2 Modélisation de la parole :

Les traits ou indices acoustiques d'un signal de parole sont sa fréquence fondamentale (ou pitch), son énergie et son timbre.

Le pitch (le thème de notre sujet) représente la fréquence de vibration des cordes vocales. Cet élément est différent pour la voix d'un homme (entre 60 Hz et 150Hz), la voix d'une femme

(aux alentours de 250Hz) ou celle d'un enfant (entre 300Hz et 400Hz). L'énergie d'un son est liée à la pression de l'air en amont du larynx et caractérise son intensité. Le timbre est la caractéristique d'un son permettant de le différencier d'un autre son.

I.3 Conclusion :

Nous avons pu voir au cours de ce chapitre, le phénomène de la production de la parole, les différentes sources permettant la génération des sons d'une langue donnée.

Nous avons aussi remarqué que le signal vocal est très complexe, du fait de sa grande variabilité, ce qui rend toute tentative de le modéliser ou de reconnaître très délicate.

Un signal de parole est une séquence de sons correspondant à une suite d'états de l'appareil phonatoire. Le signal de parole est un processus aléatoire non stationnaire à long terme.

Bibliographie :

[1] R. Boite et all, Traitement de la parole, PPUR, 2000
[2] Damien VINCEN. (thèse) « Analyse et contrôle du signal glottique en synthèse de la parole» l'École Nationale Supérieure des Télécommunications de Bretagne 2007.
[3] LE Manh Tuan.''ANALYSE DES VOYELLES SPÉCIALES DU VIETNAMIEN'', Institut de la Francophonie pour l'Informatique En collaboration avec le Centre de Recherche MICA, Hanoi.2005
[4] S.ADDAD .mémoire magister « Décodage Acoustico Phonétique en vue de la reconnaissance des voyelles de l'arabe standard » .Ecole militaire polytechnique, Algerie.2001.
[5] Barış Eker, TURKISH TEXT TO SPEECH SYSTEM. BILKENT UNIVERSITY.Turky (Thèse).2002
[6] Dr. Joseph Picone ,"FUNDAMETALS OF SPEECH RECOGNITION, INSTITUTE FOR SIGNAL AND INFORMATION PROCESSING". Mississippi State University. 1998.
[7] Bojan Kotnik1, Harald Höge, Zdravko Kacic1 "Evaluation of Pitch Detection Algorithms in Adverse Conditions", University of Maribor, Slovenia , Siemens AG, Corporate Technology, Germany 2006.
[8] MARK D. SKOWRONSKI, "BIOLOGICALLY INSPIRED NOISE-ROBUST SPEECH RECOGNITION FOR BOTH MAN AND MACHINE " .UNIVERSITY de FLORIDA ,USA . 2004 .
[9] Robert E. The Spectral Autocorrelation Peak Valley Ratio (SAPVR) – A Usable speech Measure Employed as a Co-channel Detection System, (Article) Temple University USA IEEE_WISP_2001_V5.
[10] Codage et décodage LPC de la parole, avril 2003.
[11] www.Winpitch.com

II.1 Introduction :

On explore les différents algorithmes et ses classes pour la détection du pitch. Où certain nombre de problèmes se posent, notamment parce que les signaux réels ne sont pas à proprement parlé périodiques, et aussi parce que le paramètre à estimer et variable au cours du temps.

II.2 nécessité de la fréquence fondamentale :

La fréquence fondamentale est parmi les paramètres acoustique d'un signal parole qui son nécessaire à extraire dans plusieurs domaine comme la synthèse, reconnaissance automatique de la parole, de codage nécessaire au transmission (codage LPC), la fréquence fondamentale est par définition l'inverse de la période de vibration des cordes vocales, elle est appelée aussi le pitch. La figure II.1 représente un sont voisé avec son fréquence fondamentale F0.

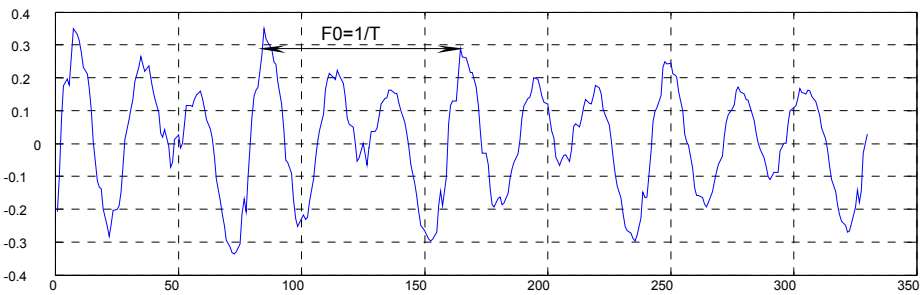


Figure II.1 Un son voisé avec son fréquence fondamentale 'F0' (amplitude Vs échantillon).

L'extraction du pitch n'est pas une tache facile pour les trois raisons suivantes :

- La vibration des cordes vocales n'à pas nécessairement une périodicité complète ;
- Séparation source/conduit vocale.
- La plage de dynamique de la fréquence fondamentale est très grande.

Une analyse d'un signal parole n'est pas complète tant qu'on n'a pas mesuré l'évolution temporelle de la fréquence fondamentale ou *pitch*.

La figure III.2 donne l'évolution temporelle de la fréquence fondamentale de la phrase *"les techniques de traitement de la parole"*. On constate qu'à l'intérieur des zones voisées la fréquence fondamentale évolue lentement dans le temps.

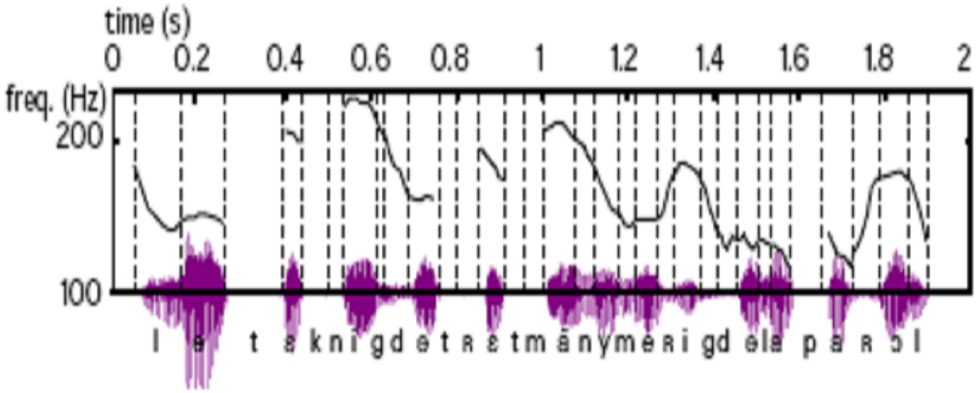


Figure II.2 l'évolution temporelle de la fréquence fondamentale.

Dans la figure précédente la fréquence est donnée sur une échelle logarithmique et les sons non - voisés sont associés à une fréquence nulle

II.2.1 Problèmes liés à la détermination de la période :

a. Ambiguïtés d'octave :

L'une des caractéristiques essentielles d'un signal strictement périodique de période T0, est qu'il vérifie la périodicité non seulement pour T0 mais aussi pour 2T0, 3T0,…. etc. La période d'un tel signal est alors définie mathématiquement par la plus petite valeur vérifiant la propriété de périodicité. Il en résulte qu'un signal réel "presque périodique" à T0, sera aussi "presque périodique" pour tous les multiples de T0 ; les irrégularités, même petites, par rapport à la périodicité stricte, pourront alors entraîner des erreurs de détermination au multiple de la période attendue.

De même, si la forme d'onde du signal est suffisamment régulière à la période *T0/k*, cela peut entraîner une erreur de détermination au sous-multiple k de la période attendue. Les erreurs aux multiple et au sous-multiple de la période attendue sont appelées "erreurs d'octave".

b. Variations temporels de la période :

L'évolution de la période fondamental des signaux sonores oblige d'une part à estimer régulièrement ce paramètre au cours du temps et d'autre part à effectuer des analyses locale du signal.

En effet, pour pouvoir suivre l'évolution de la période, il faut en fournir une estimation en fonction de temps. Mais le fait même que la période évolue au cours du temps implique que le signal ne plus périodique et qu'une analyse globale risque de mélanger des parties

de signaux ayant des périodes différentes. Il faut alors utiliser une hypothèse de *"pseudo périodicité locale"*. Une analyse locale du signal consiste à effectuer l'étude sur des durées finies appelées « trames», alors un calcul d'une valeur de fréquence fondamentale par trame.

II.3 Classification de l'algorithme :

On classe les algorithmes de représentation de signal selon quatre méthodes : temporelles, fréquentielles, temps-fréquence, temps échelle.

II.3.1 Algorithme temporels à court terme :

La représentation à cours terme dans le domaine temporel du signal consiste à extraire une représentation trame par trame du signal, L'idée de base est de travailler sur chaque trame. Dans certains cas, le signal est supposé stationnaire sur la trame donnée. La taille de la trame est un paramètre important de signal parole, et dans le cas ou cette taille est fixe, elle est choisie pour contenir dans tous les cas au moins 2 à 3 période de signal, pour les signaux de parole continue, il est raisonnable de considérer que les voix d'homme descendent jusqu'à 50 Hz (période égale à 20 ms) et les voix des femmes jusqu'à 100Hz (période égale 10ms), ce qui donne des tailles de fenêtre d'environ 30 ms pour les hommes et 20 ms pour les femmes.

II.3.2. Algorithmes fréquentiels à court terme :

D'un point de vue théorique, la théorie des série de Fourier indique qu'un signal parfaitement périodique peut être décomposé en une somme d'exponentielles dont les fréquences sont toutes multiple d'une fréquence particulière; cette fréquence est appelée fréquence fondamentale, et correspondons à l'inverse de la plus petit période du signal ; les différentes exponentielles sont appelées harmoniques du signal, et chacune d'entre elles est associée à une amplitude. En fait, les signaux étudiés n'étant pas périodique, la décomposition en série de Fourier doit être utilisée sur des signaux non périodiques et qui fournit quand même les harmoniques dans le cas d'un signal périodique.

D'un point de vue pratique, la seule analyse réalisable est une analyse à court terme (et à support temporel borné) dans le domaine fréquentiel; cette analyse nécessite l'utilisation d'une fenêtre sur le signal, ce qui modifie la décomposition (de séparation) en harmoniques. Néanmoins, si la taille de la fenêtre est suffisante, alors l'analyse

fréquentielle sera en générale capable de séparer les harmoniques de signal et donc d'en estimer les caractéristiques.

II.3.3 Algorithmes temps -fréquences :

L'évolution assez rapide de la caractéristique de périodicité des signaux étudiés impose de considéré la notion de façon locale; ceci peut être effectué en analysant le signal grâce au technique d'analyse temps-fréquence.

II.3.4 Algorithmes temps -échelles :

Ces algorithmes prends en considération les caractéristiques fréquentielles du signal parole du fait que le pitch se trouvant dans les basses fréquences ce qui conduit à poncer d'utiliser autre outils « appelés Ondelettes » caractérisés par une longueur variable suivant le temps.

II.4 Algorithmes d'estimation du pitch :

Il y a plusieurs méthodes proposées dans la détection de la fréquence fondamentale, chaque méthode possède ses avantages et ses propres inconvénients. On cite quelque uns.

II.4.1 Domaine temporel :

a. Fonction d'autocorrélation :

Dans le domaine de traitement du signal digital, la fonction d'autocorrélation du signal

x(n) est définis par l'équation suivante :

$$R(k) = \sum_{m=-\infty}^{\infty} x(m).x(m + k) \qquad \text{II.1}$$

On sait que la fonction d'autocorrélation d'un signal périodique avec la période P est périodique avec la même période :

$$R(k) = R(k + P) \qquad \text{II.2}$$

La fonction d'autocorrélation possède les caractéristiques importantes suivantes:

- C'est la fonction paire : R(k) = R(-k)
- R(k) est maximal à zéro : $|R(k)| \leq R(0)$ avec n'import quel k
- R(0) est égal à l'énergie du signal :

$$R(0) = \sum_{m=-\infty}^{\infty} x^2(m) \qquad \text{II.3}$$

En raison de ces caractéristiques, la fonction d'autocorrélation est alors maximale aux échantillons 0, ±P, ±2P,... où les valeurs de cette fonction sont égales à la valeur d'énergie du signal. Ces maximas sont appelés les sommets. La détermination de la période fondamentale P du signal de la parole peut être remplacée par la détermination de la période de la fonction d'autocorrélation.

La fonction d'autocorrélation à court temps peut être appliquée pour un segment de signal. D'abord, c'est la multiplication avec une fenêtre convenable w(n), ainsi :

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m).x(n - m).x(m + k).w(n - k - m)$$

clairement $R_n(-k) = R(k)$

et $$R_n(-k) = R_n(k) = \sum_{m=-\infty}^{\infty} [x(m).w(m - k)].[w(n - m).w(+k - m)]$$

Si $h_k(n) = w(n).w(n + k)$ alors :

$$R_n(k) = \sum_{m=-\infty}^{\infty} [x(m).x(m - k)].h_k(n - m) \qquad \text{II.4}$$

Cela veut dire que $R_n(k)$ peut être obtenu en filtrant x(m).x(m-k) par un filtre avec la réponse impulsionnelle $h_k(n)$.

De plus, la fonction d'autocorrélation peut être calculée en utilisant l'équation de définition:

$$R_n(k) = \sum_{m=-\infty}^{\infty} [x(n + m).w'(m)].[x(n + m + k).w(k + m)] \qquad \text{II.5}$$

avec : $w'(n) = w(-n)$

Si w(n) est la fenêtre rectangulaire ou de Hanning ,on aura :

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(n + m).w'(m)].[x(n + m + k).w(k + m)] \qquad \text{II.6}$$

Enfin, on devra donc rechercher $R_n(nT) \approx \alpha R(0)$ avec $\alpha \leq 1$ pour déterminer la période de la fonction d'autocorrélation.

Afin d'améliorer les performances, on peut opérer un prétraitement qui ne conserve que les données les plus typiques de la périodicité, à savoir les extrema du signal:

On réalise alors un évidement central comme le montre la figure II.3.

Le seuil de l'évidement central doit être choisi en fonction de l'amplitude du signal et sera donc adaptatif. Une valeur habituellement proposée et également utilisée avec l'autocorrélation de 30% de Amax (amplitude maximale).

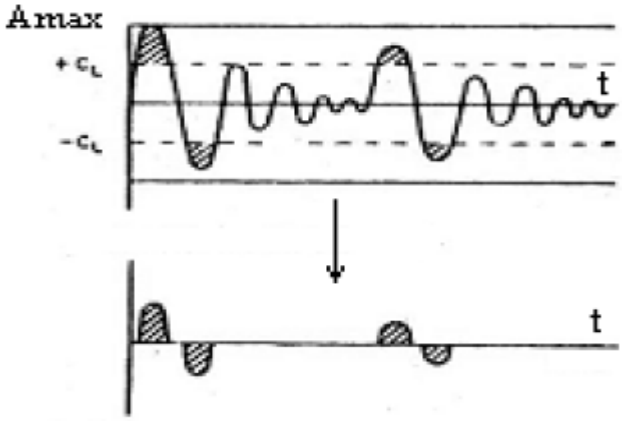


Figure II.3 Evidement central avec le seuil C_L

b. AMDF: Averaged Magnitude Difference Function :

Le calcul de l'autocorrélation est assez coûteux en volume d'opérations. On lui préfère souvent un calcul plus simple basé sur la différence de l'amplitude du signal à différents instants. La fonction de la différence de la magnitude définit par [12,13,14]:

$$AMDF\ (m) = \frac{1}{L}\sum_{t=1}^{L} |s(t) - s(t+m)| \qquad 0 \le m \le t-1 \qquad \text{II.7}$$

L : est la longueur de la fenêtre choisie.

m : est le coefficient glissant de la fenêtre.

Clairement, la fonction AMDF (m) ≥ 0.

AMDF(m) = 0 si et seulement si s(t)=s(t+k) avec tout t = 0,1,..., L.

Cette condition est satisfaite quand m = i*P (avec P est la période du signal x(n)) parce qu'avec k=i*P, la différence d'amplitude entre s(n) et s(t+m) est égale à 0.

Cela veut dire que si nous appliquons cette équation pour calculer nous allons obtenir des chaînes de AMDF(m) : AMDF(iP) = 0 avec i = 0,1,...

À partir de cette remarque, on peut calculer F0 en déterminant les minima de la fonction AMDF(k). La distance entre deux minima successifs constitue la période fondamentale.

c. Méthode de SIFT :

La méthode de SIFT(Simplified Inverse Filtering Tracking) se fonde sur le filtrage inverse du signal et analyse de la périodicité de la source estimée. Le signal est tout d'abord filtré dans la bande (0,800 Hz) puis traité par une analyse LPC pour obtenir les coefficients de prédictions du filtre 1/A(Z), on le passe ensuite dans le filtre inverse A(Z).

Pour obtenir une oscillation en* (signal résidu), on calcul la fonction d'autocorrélation $R_e(k)$ de en * sur laquelle on cherche l'amplitude la plus élevée et supérieur à un seuil $\mu 2$ donné, soit [12]:

$$R_e(k0) = \max \quad R_e(k) = \max \sum e_n . e_{n-k} \quad \text{et} \quad Re(k0) > \mu_2 \qquad \text{II.8}$$

La valeur du fondamental estimé est alors : F0*=Fs/k0 avec : Fs est la fréquence d'échantillonnage, K0 est l'échantillon qui correspond à l'amplitude la plus élevée.

II.4.2 Domaine fréquentiel :

Il y a plusieurs méthodes appartiennes au domaine fréquentiel, on cite :

a. Méthode de Peigne :

Le principe de Peigne est de détecter une structure harmonique dans le spectre du signal par intercorrélation avec une fonction peigne. Physiquement, cela revient à la recherche simultanée des maxima du spectre situés à des fréquences harmoniques les unes des autres. Pour tenir compte de la pente du spectre de la parole, les dents du peigne doivent être d'amplitude décroissante avec la fréquence, soit [12]:

$$AP(f, f0) = \sum_{n=1}^{N} d(f, f0).e^{-bn} \quad \text{avec} \quad N = \frac{Fs}{2f0} \quad \text{et } b = 0,1; \qquad \text{II.9}$$

La fonction d'intercorrélation s'exprime par :

$$I(f0) = \int_0^{Fe/2} P(f, f0).df \qquad \text{II.10}$$

Avec S(f), le spectre d'amplitude du signal. La valeur du fondamental estimé F0 maximise cette fonction, soit [12]:

$$I(f0) = max I(f0) \quad \text{et} \quad I(f0) > \mu_1$$

Où μ_1 est un seuil fixé a priori.

II.5 Conclusion :

L'information prosodique est dominée par la variation de F0, la fréquence fondamentale est un paramètre important dans divers domaines. L'estimation de F0 et la décision de voisement/non-voisement sont des problèmes délicats qui tiennent principalement aux raisons suivantes :

- L'excitation glottale n'est pas rigoureusement périodique.
- Il y a une interaction entre l'excitation et le conduit vocal
- La segmentation des débuts et de fins de voisement est difficile.

- L’étude de la variation de F0 est importante.

Il est plus commode de faire suivi la dynamique de F0, ce qui permet de corriger en particulier les sauts d’harmoniques dans la trajectoire.

Bibliographie :

[12] J P.Haton, J. M.Pierrel , GPereou,J Galelen, J. L.Gauvain, « Reconnaissance Automatique de la parole » France. 1991.

[13] *Li Tan and Montri Karnjanadecha* **«** PITCH DETECTION ALGORITHM: AUTOCORRELATION METHOD AND AMDF » Department of Computer Engineering Faculty of Engineering Prince of Songkhla University Hat Yai, Songkhla Thailand, 90112. 2003

[14] « Overview of Homophonic Pitch Detection algorithms » Alexandre Savard Schulich School of Music - McGill University 555 Sherbrooke St. West Montreal, QC Canada H3A 1E3 . 2003

III.1 Introduction:

On explore plusieurs algorithmes appartenant aux domaines temporels. Nous essayerons de mettre en évidence le comportement de chaque algorithme vis-à-vis des problèmes liés à la détermination de la fréquence fondamentale et la décision voisé non voisé nécessaire pour la détection de pitch.

Si on ne possède pas un bon algorithme de voisement non voisement le résultat peut entaché d'erreur à cause de la possibilité d'estimation d'une fréquence fondamentale qui n'exprime en aucun sens le pitch, pour cette raison on a développé un algorithme de décision V/NV basé sur le ZCR(taux de passage par zéro) et l'énergie moyenne.

III.2 Méthode d'autocorrélation basée sur analyse LPC :

Cette méthode d'autocorrélation est basée sur l'analyse LPC son objectif est la séparation de l'effet de conduit vocale sur la source de signal.

La recherche de pitch est procédé par l'Autocorélation (voire figure III-1) du signal résidu au lieu de signal lui-même, basant sur le fait que le signal résidu est résulte de la déférence entre le signal source et le signal estimé (équation III.11). Le signal résidu peut être estimé par la convolution de signal source et d'*un filtre non récursif* représenté par la fonction de transfert $A(z)$ (équation III.12).

Le codage LPC traite le signal de parole en faisant une distinction entre les parties voisées et non voisées du son. Les parties voisées présentent une certaine périodicité, ce qui permet de trouver une fréquence fondamentale. La partie non voisée n'est aucunement périodique et il n'y a donc pas de fréquence fondamentale. Les paramètres renvoyés par cette analyse sont la *fréquence fondamentale* (uniquement pour les parties voisées), *un gain* et les coefficients d'un *filtre tout pôle* (filtre autorégressif AR) [15].

En effet, l'analyse LPC est une analyse prédictive qui utilise donc les valeurs entourant un échantillon pour le prédire.

Comme les échantillons du début et de la fin n'ont soit pas de prédécesseurs, ne soit pas de successeurs, leur prédiction sera entachée d'erreur, ce phénomène s'appelle *l'effet de bord,* on utilise donc une fenêtre de Hamming, nulle en ses deux extrémités, pour diminuer l'influence de ces échantillons extrêmes. Après cette pondération, chaque tranche va être passée dans un algorithme afin de calculer les coefficients **ai**. Ces coefficients sont ceux du dénominateur du filtre tous pôles dont la transmittance est l'enveloppe spectrale du signal.

Le calcul des coefficients se base sur la résolution des équations de **Yule-Walker** qui peuvent notamment être résolues par l'algorithme de ***Le Vinson*** et l'algorithme de ***Schur*** **[**16].
Le premier coefficient est toujours égal à un, les suivants sont les « ai » qui représentent le dénominateur du filtre (H(z)) autorégressif. (p : l'ordre de paramètre) :

$$H(Z) = \frac{\text{gain}}{1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}} \qquad \text{III.1}$$

III.2.1 Prédiction linéaire :

III.2.1.1 Mesure de l'erreur de prédiction :

La figure III.1 représente le codage LPC qui consiste à estimer la valeur de l'échantillon à venir sur la base de quelques valeurs mesurées précédemment $s[n-k]$ [10].

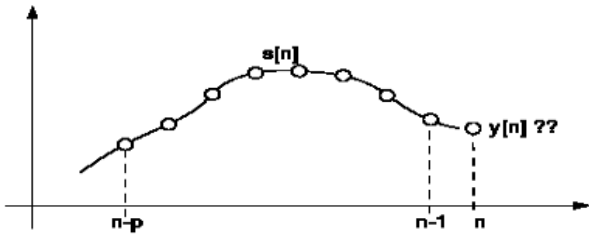


Figure III.1 Les échantillons $s[n-p]$ à $s[n-1]$sont utilisés pour estimer la valeur à venir

La valeur estimée $y[n]$ est calculée à partir des échantillons précédents pondérés par des coefficients $\boldsymbol{a_k}$ qui sont généralement un nombre de 8 à 12 [12].

$$y[n] = -(a_1 s[n-1] + a_2 s[n-2] + ... + a_p s[n-p]) = -\sum_{k=1}^{p} a_k s[n-k] \qquad \text{III.2}$$

Les valeurs des coefficients de prédiction a_k s'obtiennent par minimisation de la variance σ_e^2 de l'écart $e[n]$. où $s[n]$ est la valeur réelle [10]:

$$e[n] = s[n] - y[n] = s[n] + \sum_{k=1}^{p} a_k s[n-k] \qquad \text{III.3}$$

La puissance ou variance de l'écart de l'ensemble des N échantillons ''$e[n]$'' à disposition avec $0 \leq n \leq N-1$ est alors la suivante [10] :

$$\sigma_e^2 = \frac{1}{N}\sum_{n=0}^{N-1} e^2[n] = \frac{1}{N}\sum_{n=0}^{N-1}\left(s[n] + \sum_{k-1}^{p} a_k s[n-k] \right)^2 \qquad \text{III.4}$$

III.2.1.2 Calcul des coefficients de prédiction linéaire :

La procédure pour obtenir les coefficients a_k consiste à trouver les valeurs de ces coefficients pour que la puissance de l'erreur commise lors de la prédiction soit minimum. Un schéma fonctionnel traduisant cette démarche est présenté dans la figure III.2**:**

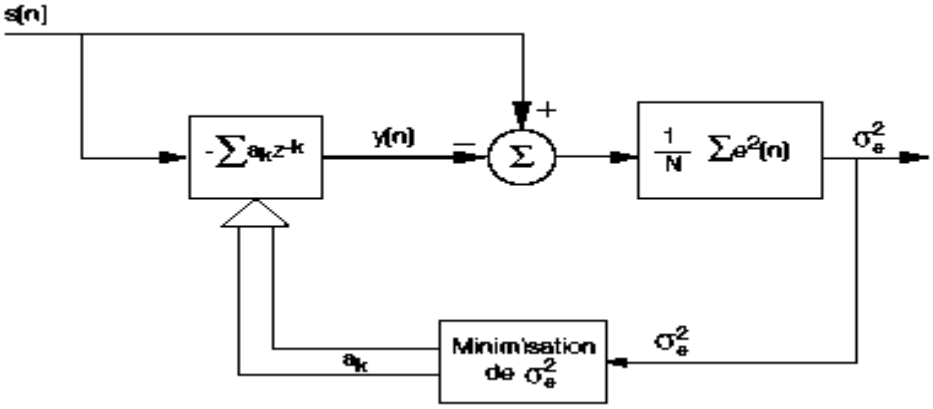


Figure III.2: Schéma fonctionnel de la prédiction linéaire

Mathématiquement, la variance est une fonction des paramètres de prédiction a_k **[16]** :

$$\sigma_e^2 = \sigma_e^2(a_1, a_2, ..., a_p) = \sigma_e^2(a_k) \qquad \text{III.5}$$

Sa valeur minimum s'obtient donc lorsque l'ensemble des dérivées partielles de σ_e^2 par rapport aux paramètres *ak* sont nulles **[10]** :

$$\sigma_{e,\min}^2 = \frac{\delta\sigma_e^2(a_k)}{\delta a_k} = 0, k = 1, ..., p \qquad \text{III.6}$$

Le calcul de ces *p* dérivées partielles conduit à *p* équations :

$$\begin{aligned} &a_1 r_{ss}[0] + a_2 r_{ss}[-1] + ... + a_p r_{ss}[1-p] = -r_{ss}[1] \\ &a_1 r_{ss}[1] + a_2 r_{ss}[0] + ... + a_p r_{ss}[2-p] = -r_{ss}[2] \\ &\qquad\qquad . \qquad\quad . \\ &\qquad\qquad . \quad = \quad . \\ &\qquad\qquad . \qquad\quad . \\ &a_1 r_{ss}[p-1] + a_2 r_{ss}[p-2] + ... + a_p r_{ss}[0] = -r_{ss}[p] \end{aligned} \qquad \text{III.7}$$

Avec **[10]** :

$$r_{ss}[m] = \sum_{n=0}^{N-1} s[n]s[-m], m = 1, ..., p \qquad \text{III.8}$$

Les coefficients des paramètres a_k sont les *p* premières valeurs de la fonction d'autocorrélation $r_{ss}[m]$ (est une forme matricielle) du signal *s*[*n*] comportant *N* échantillons :

$$\mathrm{R_{ss}a} = -\mathrm{r_{ss}} \qquad \text{III.9}$$

Où $\mathbf{R}_{ss}$ est la matrice *p*× *p* d'autocorrélation, $\mathbf{r}_{ss}$ le vecteur *p*×1 d'autocorrélation et **a** le vecteur *p*×1 des paramètres de prédiction. On notera que, la fonction d'autocorrélation étant paire, la matrice $\mathbf{R}_{ss}$ est symétrique [10]. Avec :

$$\mathrm{a} = -R_{ss}^{-1}\mathrm{r_{ss}} \qquad \text{III.10}$$

L'estimation de la valeur à venir sera d'autant meilleure que le nombre de points *N* utilisés pour calculer la fonction d'autocorrélation $r_{ss}[m]$ sera élevé. L'évaluation des paramètres a_k se fait donc après analyse d'une tranche t_k suffisamment longue alors que le calcul de la valeur à venir $y[n]$ n'utilise que les p dernières valeurs échantillonnées. La figure III.3 représente le découpage en tranches tk du signal s[n]. Il est important de relever que si les points *s* [n] ne sont pas corrélés entre eux, aucune prévision n'est possible (cas du bruit blanc).

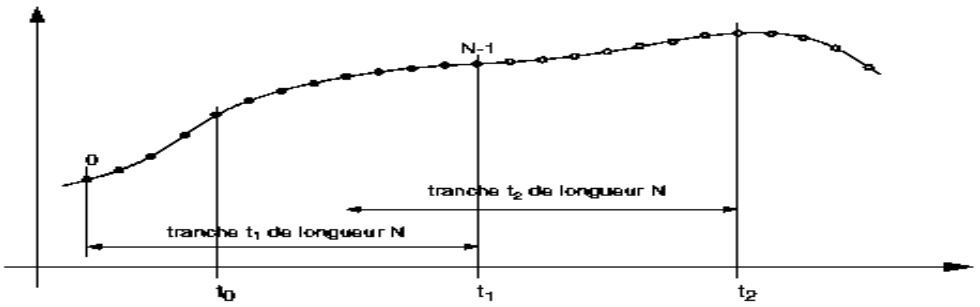


Figure III.3 Découpage en tranches *tk* du signal *s*[*n*].

III.2.1.3 Interprétation de la prédiction linéaire :

On calcul la variance pour calculer les coefficients de prédiction linéaire *ak* [10] :

$$e[n] = s[n] - y[n] = s[n] + \sum_{k=1}^{p} a_k s[n-k] \qquad \text{III.11}$$

Appliquant la transformation en *z* à cette équation, on en tire les 2 relations suivantes :

$$E[z] = s(z)(1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}) = s(z)A(z) \qquad \text{III.12}$$

$$S(z) = E(z)\frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + ... + a_p z^{-p}} = E(z)\frac{1}{A(z)}$$

On notera que le résidu *e*[*n*] peut être considéré comme un signal d'excitation servant à créer le signal *s*[*n*] avec l'aide d'un filtre récursif tous pôles : ***H*(*z*) = 1/*A*(*z*).**

Dans le cas de la parole, ce signal d'excitation peut être périodique (sons voisés) ou aléatoire les tranches successives sont décalées d'une valeur inférieure à leur durée c'est le « *recouvrement* ». Généralement, ce décalage est de 10 msec [10].

III.2.2 Modèle du conduit vocal :

La Figure III.4 représente le modèle de la production de la parole généralement adopté pour créer artificiellement des sons comporte :

– un générateur périodique d'impulsions unité ;

– un générateur de nombres aléatoires à valeur moyenne nulle et variance unité

– un commutateur servant à choisir les sons voisés ou non

– un gain proportionnel à la valeur efficace du signal $s[n]$

– un filtre tous pôles $H(z) = 1/A(z)$.

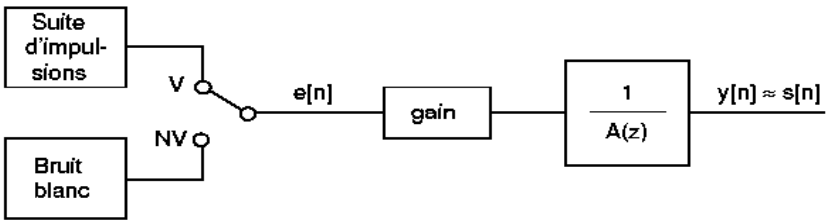


Figure III.4 Modèle du conduit vocal.

L'extraction de générateur et du filtre est faite pour chaque trame (la trame est de 20 à 30 ms)

III.2.3 Spectre de signal :

L'analyse spectrale se fait à l'aide de la transformée de Fourier rapide **FFT.** Idéalement, le nombre de points de la tranche analysée devrait être une puissance de 2 Il est nécessaire d'effectuer préalablement un fenêtrage.

III.2.4 Fonction de transfert du conduit vocal et stabilité :

Après avoir choisir une tranche de 30 ms on procède de trouver les coefficients de filtre tous pôle en faisant l'analyse LPC. *Il est important de rappeler que dans notre travaille on a choisit le nombre de paramètres du filtre d'après l'analyse LPC de* ***12 paramètres***. Les coefficients LPC représentent le dénominateur de la fonction de transfert du conduit vocal. Avant de procéder les autres étapes on doit s'assurer que le conduit vocal est *stable.*

III.2.5 Réponse fréquentielle du conduit vocal :

La réponse fréquentielle nous donne une idée concernant les formants générée par le conduit vocal et le pitch, les fréquences se trouvant au de la de 50 Hz et 500 Hz.

III.2.6 Recherche du pitch :

La recherche de pitch passe par les étapes suivantes :

III.2.6.1 Filtrage du signal :

Comme on l'a vu plus haut, la période du pitch est comprise entre **2** et **20** msec. Il est préférable avant de poursuivre l'analyse, de commencer par éliminer les fréquences supérieures à 500 Hz à l'aide d'un filtre passe-bas de *Butterworth*, généralement d'ordre 8.

III.2.6.2 Fenêtrage de la tranche (trame) de signal :

Après filtrage on passe à effectuer la pondération par une fenêtre de Hamming.

III.2.6.3 Calcul des coefficients :

On calcul des coefficients d'ordre 12 par analyse LPC pour la tranche fenêtrée.

III.2.6.4 Recherche du signal d'excitation *e*[*n*]:

Nous avons vu que le résidu *e*[*n*] de la prédiction linéaire peut être considéré comme le signal d'excitation servant à créer le signal *s*[*n*] en passant à travers le filtre récursif (Equ III.1). Puisque, dans notre cas, le signal *s*[*n*] est connu, par filtrage inverse on obtient le résidu

$$E(z) = A(z)S(z) \qquad \text{III.13}$$

Ce qui revient à convoluer les coefficients $a_k \equiv a[n]$ avec le signal *s*[*n*] :

$$e[n] = a[n] \otimes s[n] \qquad \text{III.14}$$

Le signal d'excitation *e*[*n*] est *la convolution du filtre tous pôles avec le signal s[n] connus.*

III.2.6.5 Autocorrélation de *e*[*n*] :

On a vu que le signal d'excitation est périodique si le son est voisé et aléatoire dans le cas contraire. Comme le signal est passablement bruité, la recherche de la période est grandement facilitée si on l'effectue sur la fonction d'auto -corrélation de *e*[*n*] plutôt que sur le signal lui-même. Le résultat de l'auto -corrélation est un vecteur symétrique de longueur **2*N*** avec un maximum en son milieu. Alors on limite le calcul du maximum dans une seule partie du signal d'auto corrélation (Ex : partie droite) en suite on recherche du premier pic latéral compris entre **0.0020** (1/500hz) sec et **0.0200** sec (1/50hz), d'autres pics distants de la valeur du pitch seront présents. Pour trouver ce dernier, il suffit donc de mesurer cette distance.

III.2.6.6 Estimation de la fréquence fondamentale :

La fréquence fondamentale s'obtient par le maximum de la fonction d'autocorrelation du signal vocal

La valeur du fondamental estimée est alors [12]:

$$\boxed{\mathbf{F0 = Fs/m}} \qquad \text{III.15}$$

Où Fs est la fréquence d'échantillonnage

m : l'échantillons qui correspond au maximum.

III.2.6.7 Critère de décision voisé / non voisé :

Le maximum qui correspond au éventuel pitch que l'on vient de trouver n'est pas nécessairement significative d'un son voisé.

Si le son n'est pas voisé, ce maximum est peu marquer. Par contre, dans le cas des sons voisés, le premier pic latéral vaut généralement plus du tiers du maximum central [12]. C'est finalement ce critère qui est utilisé pour déterminer si le son est voisé ou non.

On note que Cette méthode de décision V/NV est applicable seulement avec la méthode d'atocorrélation, mais n'est pas applicable avec toutes les méthodes.

III.2.7 Méthodes de décision (voisé /non voisé) :

La détection de pitch est dépend essentiellement d'un efface décision voisement non voisement (V/NV). On présente quelque méthode à utiliser.

III.2.7.1 Plus du tiers du maximum central :

Cette méthode efficace est utilisée avec la méthode d'autocorrélation. L'inconvénient essentiel, est qu'elle s'applique au bout d'algorithme ce qui résulte un temps de calcul en plus.

III.2.7.2 Le taux de passage par zéros (PPZ) et énergie de signal :

En rappel (voire chapitre I), les sons voisés contrairement aux sons non voisés présentant d'avantage d'énergie vers les basses fréquences, ainsi le taux de passage par zéro pour un segment de signal constitue un deuxième indicateur de V/NV.

III.2.7.3 Méthode SFM (Spectral Flatness Measure):

L'une des méthodes pour détecter les sections voisé / non voisé de la parole est la mesure de l'égalité spectrale (Spectral flatness mesure). Est une méthode efficace de V/NV inspiré de la géométrie de signal. L'égalité spectrale fait usage de la propriété que le spectre du bruit pur est supposé être plat. En d'autres termes, le spectre de section non voisé est plat et le spectre de section voisé est moins plat. La mesure de l'égalité spectrale « SFM » (voire aussi l'équation I.3) est donnée par [13]:

$$SFM = \frac{G_m}{A_m} \qquad \text{III.16}$$

Il est important d'éclairer qu'il y a plusieurs méthode de détection de pitch ont utilisé cette méthode de décision V/NV, mais toutes ces méthodes s'appuient sur un seuil définit

primitivement. Il y a beaucoup de générations de SFM mais souvent s'appuie sur un seuil définit primitivement (Voir références [17], [18]).

III.2.8 Nouvelle méthode de décision (V /NV) applicable avec toutes les méthodes:

Bien que le taux de passage par zéros pour un son voisé est peut marquer par rapport au son non voisé, par contre contient une énergie grande .Donc le rapport *énergie/ZCR* est grands par rapport au son non voisé. Alors on peut exploiter cette notion énergie/ZCR pour la détection d'un son voisé ou non. Il est important d'éclairer qui' il y a des articles qui adoptent cette méthode mais le problème est le seuil (*EZR* : *energy-to-zero crossing rate*) [7]:

$$EZR\ [m] = \frac{\overline{E_0}[m]}{ZCR\ [m]} \qquad \text{III.17}$$

Où ZCR [m] présents le taux de passage par zéro pour une trame [m], et $\overline{E}_0[m]$ est l'énergie moyenne de la trame [n, m] (n : échantillon, m : trame). Le paramètre présenté EZR [m] est appliqué comme critère de décision voisé / non voisé. L'EZR [m] sera par conséquent relativement bas dans les régions non voisées du signal de la parole et inversement pour les régions voisées. La Figure III.5 qui exprime le spectre d'un signal avec un bruit de fond, ainsi illustre l'énergie de ce signal illustre l'énergie pour un signal.

La décision voisé / non voisé (VUV [m]) pour une trame 'm' est estimé par comparaison d'EZR [m] avec un seuil *v* de la base de données de signal concerné [7] :

$$\begin{aligned} &If \quad EZR\ [m] > v\ Then \\ &VUV\ [m]=1 \qquad (son\ voisé) \\ &Else\ VUV[m]=0 \quad (son\ non\ voié) \end{aligned} \qquad \text{III.18}$$

Donc la méthode suppose un seuil, si pour une trame le rapport EZR[m] atteint ou dépasse ce seuil alors cette trame est considérée voisé.

Problème :

Quel est le seuil optimal à utiliser ?

Solution : Pour notre travaille on propose une méthode de décision voise/non voisée met au début d'algorithmes avec un seuil ***adaptatif*** se change au cours de signal parole sans intervention par l'utilisateur se qui signifie la possibilité d'implantation **en temps réel.**

Dans notre nouvelle méthode, le seuil est calculé automatiquement et placé au début d'algorithme de détection de pitch ce qui résulte un temps réduit.

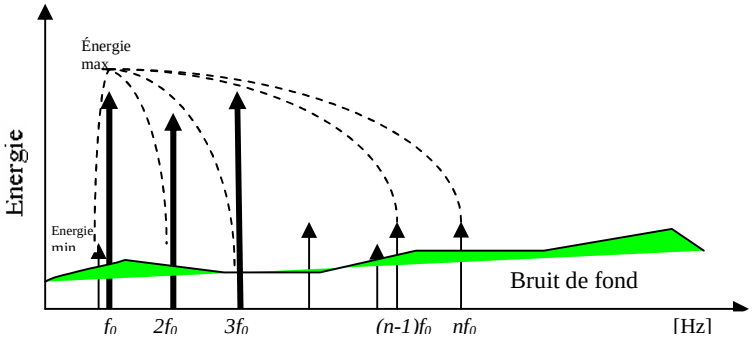


Figure III.5 Illustration d'énergie pour un signal

III.2.8.1 Procédure de la méthode :

La procédure est comme suit :

1. acquisition du signal

2. extraction de la ''i'' èmé trame et décision V/NV basée sur un seuil calculé par l'algorithme (on choisit seuil=1, seulement pour la première trame).

3. calcul d'énergie et le passage par zéro et par suite l'EZR

4. si EZR >seuil alors : l'estimation de pitch sinon on passe à l'tape 5.

5. calcul de maximum et de minimum d'EZR des trames.

6. calcul de la différence entre le maximum *EZR* et le minimum d'EZR, on le nommer « Delta » :

```
%valeur initiales
Rapport_EZR =1;
seuil=1;

(10) Détection V/NVpour la « i »eme trame :
        Calcul de l'énergie 'E'
        Calcul de 'ZCR'
        Calcul d'EZR

Si EZR>seuil
        Un son voisé
                Estimation de pitch
        Sinon
        Un son non voisé
Fin si
% Création des vecteurs de sorties
        max_rapport=max(rapport_EZR);
        min_rapport=min(rapport_EZR);
        delta=max_rapport-min_rapport;
        seuil=min_rapport+0.2*delta;
        rapport_EZR = [EZR];
% Terminaison
Aller à (10)
```

Algo.1 Algorithme principal basé sur le EZR

$$\mathbf{Delta = Max(EZR) - Min(EZR)} \qquad \text{III.19}$$

7- le seuil est égale à :

$$\mathbf{Seuil = Minimum\ (EZR) + 0.2*Delta} \qquad \text{III.20}$$

Ce seuil est estimé d'une façon à prendre la somme de minimum d'EZR et 20% de la différence entre le maximum et le minimum d'*EZR*.

8. le seuil estimé est appliqué comme seuil de décision V/NV pour la trame qui suit ((i+1) ème trame).

Si pour une trame [m] vérifie $EZR[m] > \mathbf{Seuil}$ ors la trame est considérée voisée.

On résume l'algorithme qui exprime la détection V/NV par EZR dans l'algorithme **Algo. 1**.

On note qu'on peut utiliser la méthode d'EZR d'une autre façon où l'exécution en temps réel n'est pas demandé par :

- Tout d'abord l'acquisition du signal complet.
- Le segmenter en plusieurs trames et calcul d'EZR[m] pour chaque trame.
- Estimation de seuil en considérant tout les trames, ce seuil est adopté par toutes les trames pour la décision V/NV.

III.2.9 Rapports signal sur bruit :

Le bruit est présent partout au niveau des signaux utiles, alors dans notre cas la détection de pitch ainsi l'estimation nous oblige d'étudier la méthode d'autocorrelation en présence de bruit, dans le but de cette étude, nous avons essayé de définir un rapport signal sur bruit, comme la division de la puissance moyenne de signal utile *Px* (sans bruit) par celles de bruit *Pb*, on obtient donc la formule suivante **[21]** :

$$SNR = 10\log 10\left(\frac{\sum_{n=1}^{N} Px}{\sum_{n=1}^{N} Pb}\right) \qquad \text{III.21}$$

III.2.10 Le contour du pitch et le choix de la valeur exact du pitch :

Il est clair que pour chaque tranche voisée d'un signal parole on a une fréquence fondamentale (qui peut correspondre au pitch),

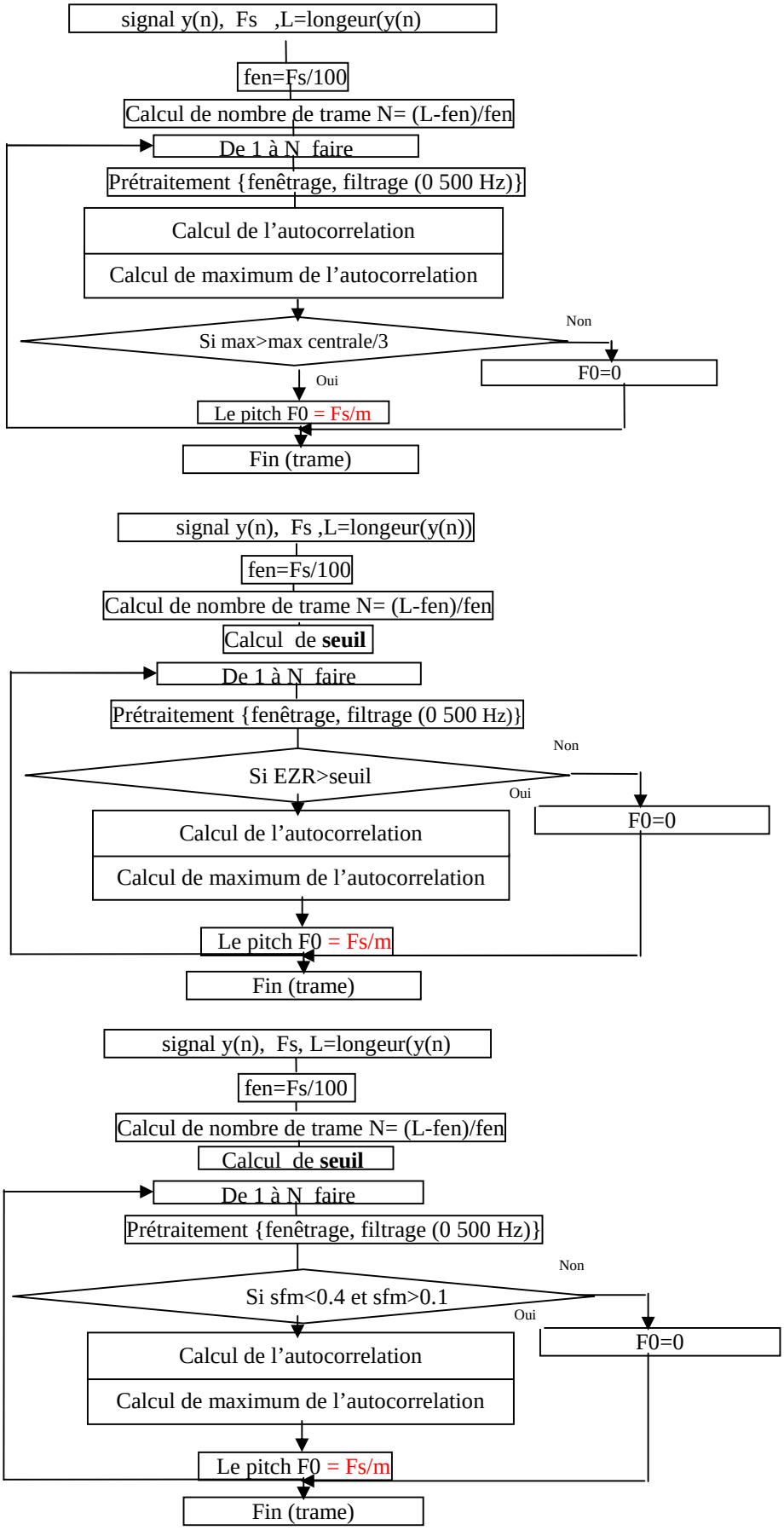
III.2.10.1 Problème de choix de pitch parmi les fréquences fondamentales:

Alors le problème c'est le choix de la valeur de pitch parmi les fréquences fondamentales estimées, on s'appui sur la théorie que le pitch exact est la fréquence fondamentale qui correspond à la *valeur maximale* des valeurs efficace ou valeur maximale d'énergie pour une tel trame. Pour chaque tranche on calcul la valeur efficace et l'éventuel pitch pour le son voisé, par suite on effectue le choix.

III.2.10.2 Choix de pitch pour ACF_LPC et décision V/NV par trois méthodes:

Les algorithmes suivants expriment la détection et l'estimation des fréquences fondamentales de chaque trame (si aura lieu) par la méthode d'ACF_LPC ainsi la décision V/NV par trois méthodes. Algo 2 présente les algorithmes d'estimation de pitch par ACF_LPC et V/NV par trois méthodes :

- La première est la méthode de tiers de maximum centrale.
- La deuxième est le rapport *EZR* (nouvelle méthode)
- La troisième c'est la méthode SFM.



Algo.2 Algorithmes d'estimation de pitch par ACF et V/NV de haut en bas par : 1/3 maximum central, EZR, SFM

III.2.11 Résultats expérimentaux pour la méthode ACF_LPC :

a. signal et fenêtrage :

Prenons le phonème « a » représenté par la figure III.6 prononcé par une voix masculine enregistré par le logiciel «WinPitchPro [11]» et procédons à son fenêtrage par une fenêtre de « Hamming », le résultat de fenêtrage est représente par la Figure III.7.

La taille de ces fenêtres est de 30 ms .Choisissons une de ces fenêtres dans la région du phonème **'a'** où **sa puissance** ou **énergie** est maximale, la durée du temps de la tranche du signal de parole à analyser de phonème 'a' est de **200ms** avec une fréquence d'échantillonnage **de Fs=11025 Hz.** le signal parole est non stationnaire à long terme, d'autre part il est supposé stable pour une durée de 20 à 30 ms [10], ce qui signifie qu'on doit effectuer notre étude sur cet intervalle, on aura :

Nombres d'échantillons=Fs*durée (sec) (pour chaque trame).

AN :

Nombre_ d'échantillons =11025*0.02=220,5 échantillons (pour 20ms)

Nombre_ d'échantillons =11025*0.03=330,75 échantillons (pour 30ms)

Dans notre étude on ne doit pas passer 330 échantillons.

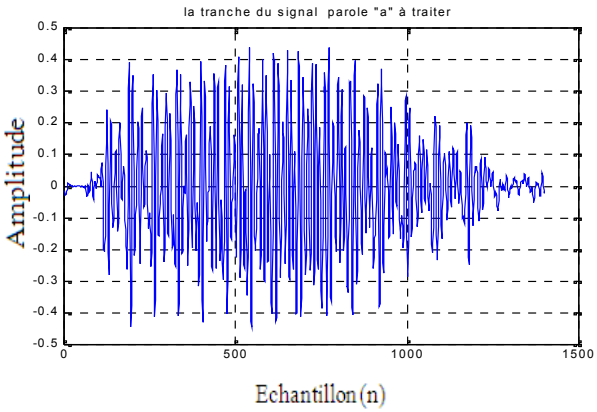


Figure III.6 Le signal parole "a"**.**

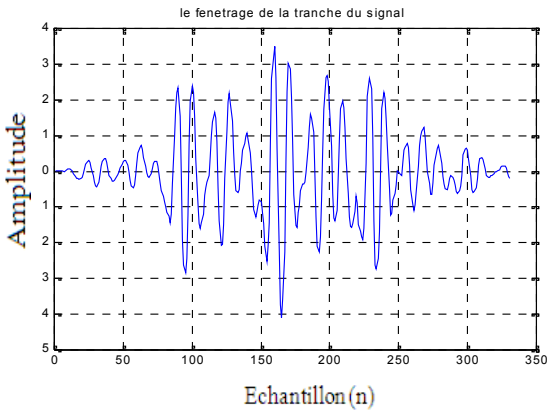


Figure III.7 Une tranche de signal soumis à un fenêtrage de 330 échantillons

b. Spectre de signal

La figure III.8 représente le spectre du signal

c. Fonction de transfert du conduit vocal et stabilité :

On a procédé un LPC de *12 paramètres* pour un éventuel bon résultat. Les coefficients LPC représentent un polynôme en $z-1$ qui n'est autre que le dénominateur de la fonction de

transfert du conduit vocal. Les coefficients LPC sont : 1.0000 -1.7294 0.9686 -0.2029 0.2139 -0.2059 0.5490 -0.7080 0.4909 -0.6391 0.8200 -0.4827 0.1255.

Ces coefficients doivent se trouver à l'intérieur du cercle de rayon unité pour assurer la stabilité. La figure III.9 montre que les coefficients se trouvent à l'intérieure de rayon unité.

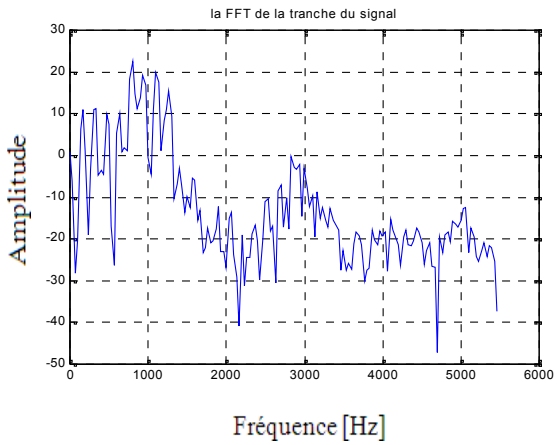


Figure III.8 La FFT de la tranche du signal

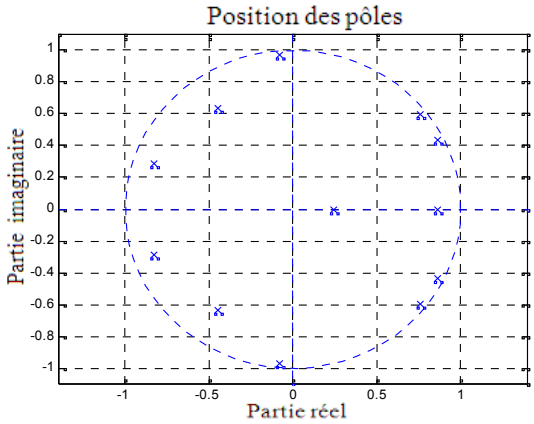


Figure III.9 Les coefficients à l'intérieur de rayon unité.

d. Réponse fréquentielle du conduit vocal :

La réponse fréquentielle nous donne une idée concernant les formants générés par le conduit vocale. La figure III.10 représente la réponse fréquentielle de signal vocale (en bleu) superposé au spectre de la tranche de signal (fenêtré).

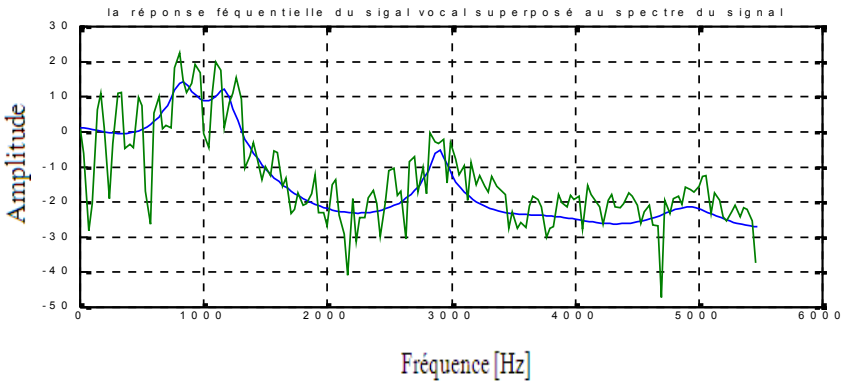


Figure III.10 La réponse fréquentielle du signal vocal superposé au spectre du signal

e. Recherche du pitch :

Il est important d'éclairer qu'avant procéder la recherche de pitch de procéder un filtrage, on a choisit un filtre passe-bas (0 à 500 Hz) de *Butterworth*, généralement d'ordre 8.

e.1 Recherche du signal résidu $e[n]$:

On applique l'équation III.14.

e.2 Autocorrélation de $e[n]$:

On calcule l'autocorrélation de signal résidu. La Figure III.11 représente le résultat l'autocorrélation de signal résidu e[n]. La figure III.12 représente la partie droite de l'autocorrélation où on doit éffectuer la recherche de maximum qui crrespond au pitch.

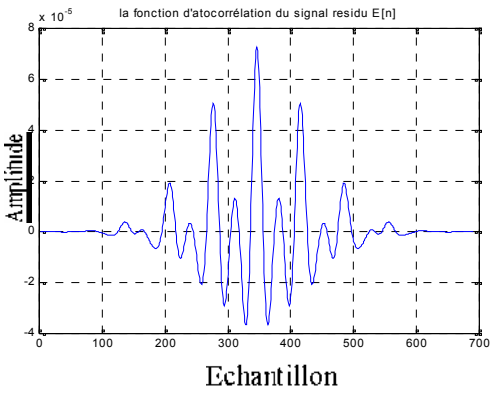


Figure III.11 L'autocorrélation de signal résidu e[n]

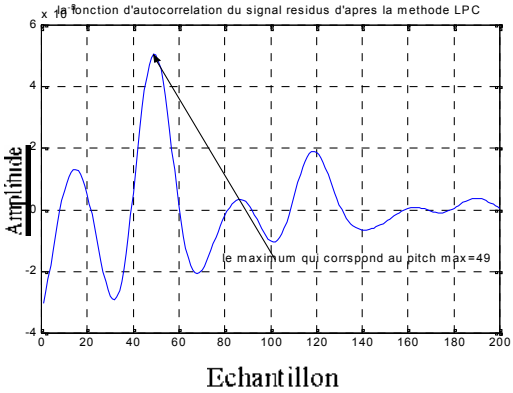


Figure III.12 La partie droite de l'autocorrélation de signal résidu e[n]

On se limite au partie droite de l'autocorrélation de e[n] .Puisque on limite le calcul entre 0.002sec et 0.02sec de la tranche du signal, alors le calcul par rapport aux échantillons correspond à **22** (22 ≈11025*0.002) et **221** (221≈11025*0.02)

Pour notre cas expérimental le maximum qui correspond aux 49 échantillons revient à 48+22=70 échantillons. La Figure III.13 représente l'autocorrélation de signal résidu e[n] (en fonction de temps [sec]).

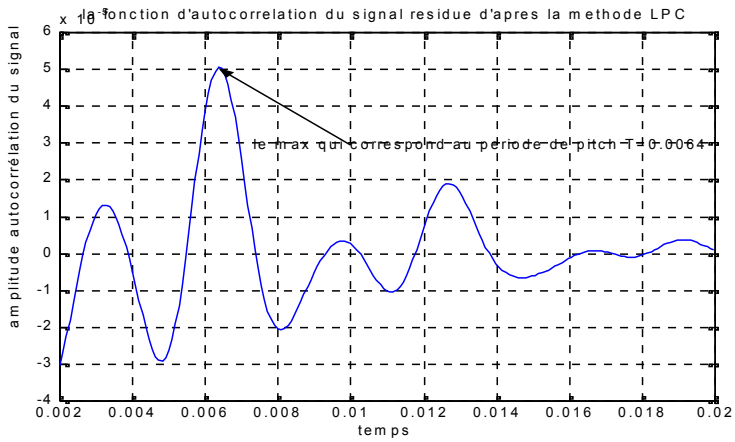


Figure III.13 L'autocorrélation de signal résidu e[n] (en fonction de temps [sec]).

e.4 Calcul de pitch :

F0= Fs/m. alors An : F0=11025/70 = 157.500 Hz (le pitch), T0=1/F0,T0 = 0.0063 sec

f. L'influence de bruit sur l'estimation de pitch :

On prend une tranche d'un signal voisé de durée de 30 ms, sans bruit exprime un phonème « a » Figure III.14. Par suite on effectue l'estimation de pitch par ACF_LPC en présence de différents puissance de bruit afin d'évaluer la méthode concernant l'estimation de pitch.

Le Tableau III.1 représente les résultats de l'estimation de pitch en fonction de bruit.

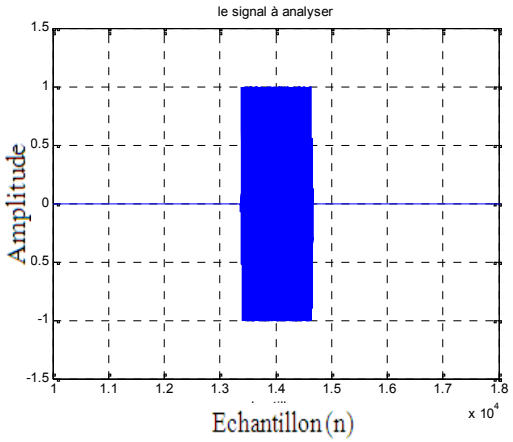


Figure III.14 Le signal « a » sans bruit

SNR [dB]	Le pitch EZR [HZ]
Sans bruit	129.7059
69	129.7059
30	129.7059
29	129.7059
19	129.7059
0	129.7059
-07	130.5000
-11	135.2817

Tableau III.1 Le pitch vs SNR

f1. Choix de pitch parmi plusieurs fréquences fondamentales :

- ✓ On a pris un phonème « a » et rechercher le pitch par ACF_LPC avec V/NV par toits méthodes : Les fréquences fondamentales trouvées par 1/3 maximum central :

0 0 0 157.5000 **157. 5000** 153.1250 131.2500

Les valeurs efficaces : 0.0067 0.0985 0.2091 0.2745 **0.3275** 0.3069 0.2922

- ✓ Les fréquences fondamentales trouvées par EZR :

0 0 0 157.5000 **157. 5000** 153.1250 131.2500

Les valeurs efficaces : 0.0067 0.0985 0.2091 0.2745 **0.3275** 0.3069 0.2922

- ✓ Les fréquences fondamentales trouvées par SFM : 0 0 162.1324 157.5000 **157.5** 153.1250 131.2500 132.8313 128.1977 0 380.1724

Les valeurs efficaces correspondantes : 0.0067 0.0985 0.2091 0.2745 **0.3275** 0.3069 0.2922 0.2540 0.2271 0.1718 0.0939

- On a achevé aux mêmes résultats, la fréquence fondamentale qui exprime le pitch est **157.5Hz** qui correspond à une valeur maximale des valeurs efficace de **0.3275**

Important : Pour avoir un contour de pitch plus claire et supprimant les erreurs dû au saut des harmoniques. On peut et après calcul de pitch exact d'accepter les fréquences fondamentales qui sont supérieur ou inferieur par rapport pitch exact de 20Hz.

g. Influence de bruit sur le contour de pitch :

On prend un signal dents de scie avec des fréquences fondamentales de 120Hz ,160Hz, 90Hz,130 Hz et d'amplitude de 0.5,1.25,0.7,0.95 respectivement pour analyser la méthode ACF_LPC par la méthode EZR et SFM en présence de différents puissance de bruit. Le pitch exact est égal à=**160Hz**. Figure III.15 représente le signal sans bruit avec les deux contours de pitch basé sur V/NV de SFM et EZR où on remarque *une zone entachée d'erreur*. La Figure III.16 représente le résultat pour SNR=30dB. La figure III.17 Le résultat pour SNR=0dB. La figure III.18 représente le résultat pour SNR=-8dB.

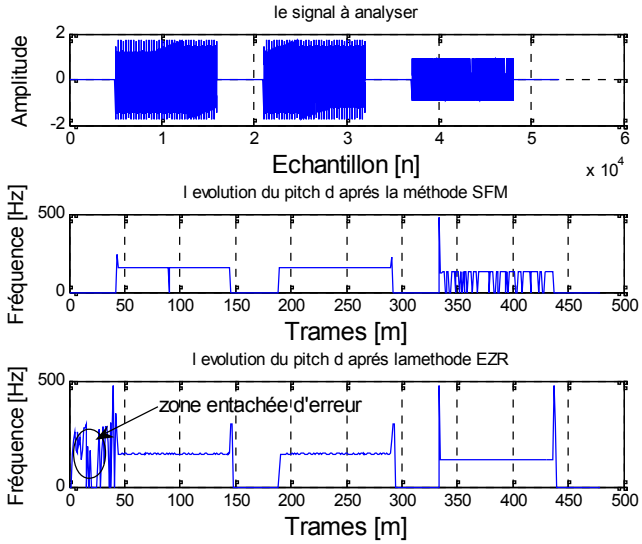


Figure III.15 Le signal sans bruit avec les deux contours de pitch basés sur V/NV par SFM ,EZR :F_0EZR=157.5 Hz ,F_0SFM=155.2717 Hz.

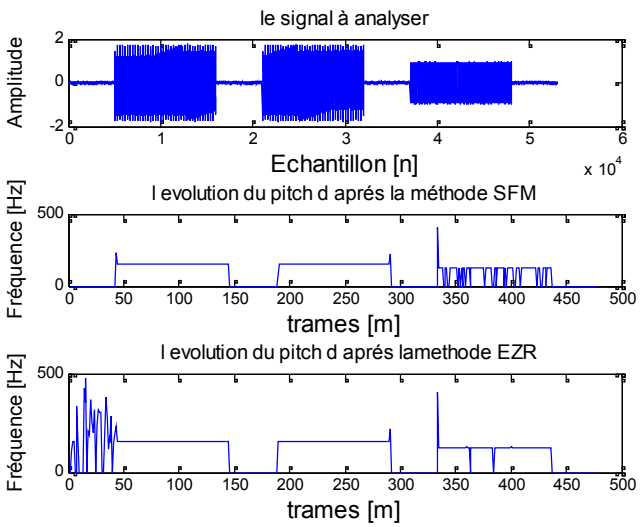


Figure III.16 Le signal avec les deux contours de pitch basés sur V/NV de SFM, EZR pour SNR = 30dB : F_0EZR = 157.5 Hz , F_0SFM = 155.2717 Hz

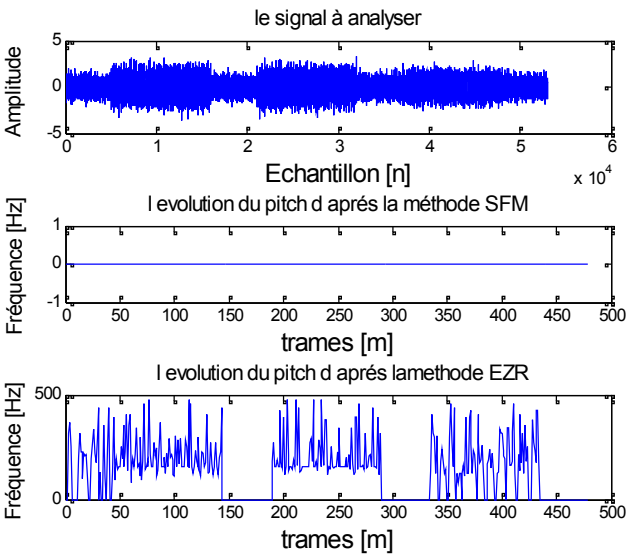


Figure III.17 Le signal avec les deux contours de pitch basé sur V/NV de SFM, EZR pour SNR=0dB : F_0EZR=157.5Hz, F_0SFM=0 Hz

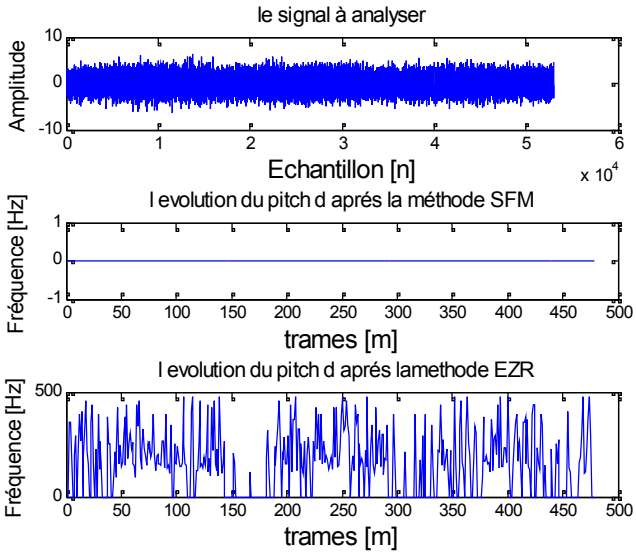


Figure III.18 Le signal avec les deux contours de pitch basés sur V/NV de SFM, EZR pour SNR=-8dB : F_0EZR=159.25Hz, F_0SFM=0 HZ

k. L'influence d'énergie sur la décision voisée non voisée et le contour de pitch :

L'énergie est un paramètre essentiel pour le traitement de signal parole. Pour satisfaire à étudier l'influence de la valeur d'énergie sur le V/NV par SFM et EZR, on a à procéder à enregistrer une phrase de différents maximum d'énergie. La phrase est « how are you » prononcée par un masculin par différents façons de telle manière à changer l'énergie en utilisant un logicielle *WinPitchPro,* pour cela on développe la formule suivante :

$$\overline{E}_0 = 10\log 10(\max(\overline{E}(m)) \qquad \text{III.22}$$

$\overline{E}_0$: l'énergie moyenne.
m : trame

Les résultats sont résumés dans le tableau III.2. On peut constater que la méthode EZR nous donne de bons résultats de V/NV que la méthode de SFM malgré l'énergie faible de signal. Les Figures III.20, III.21présentent le signal parole « how are You » et le contour de pitch par EZR et SFM pour $\overline{E}_0$ = 0.2095dB et $\overline{E}_0$ = -0.2901dB respectivement.

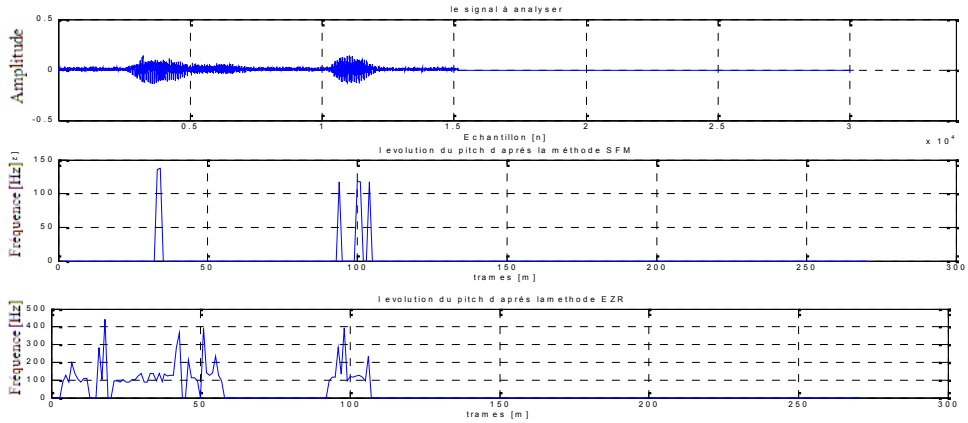


Figure III.19. Contours de pitch basés sur V/NV de SFM, EZR par $\overline{E}_0$ = 0.2095dB, : F_0EZR= 118.5484 Hz, F_0SFM= 118.5484Hz

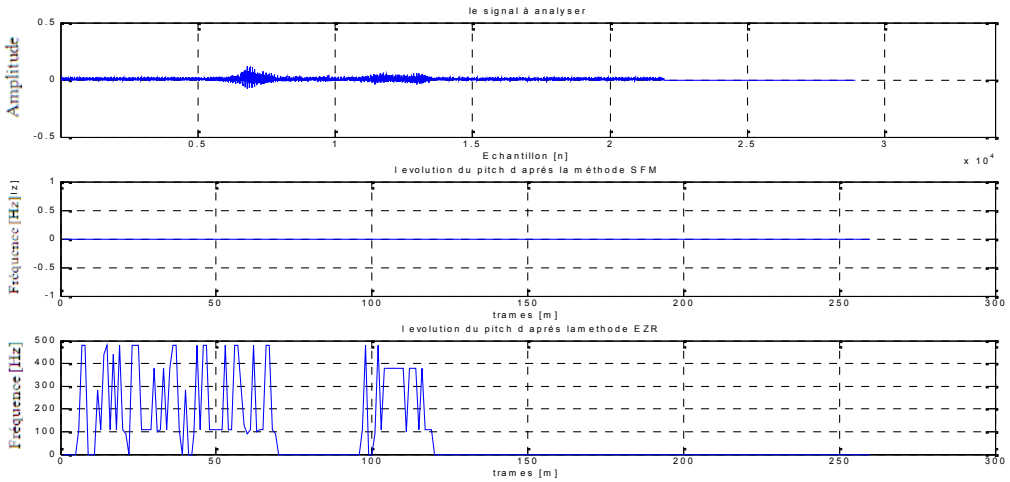


Figure III.20 Contours de pitch basés sur V/NV de SFM, EZR par $\overline{E}_0$ =-0.2901dB, : F_0EZR= 109.1584 Hz, F_0SFM= 0 Hz.

Le max $\overline{E}_0$ [dB]	Le pitch EZR [HZ]	Le pitch SFM [HZ]
2.3089	196.8750	196.8750
0.2095	118.5484	118.5484
-0.2901	109.1584	NV
-0.4165	100.2273	NV
- 0.4351	324.2647	NV

Tableau III.2 Le pitch par EZR, SFM en fonction de $\overline{E}_0$.

k.1 Discussion :

L'algorithme adaptatif (EZR) adapte son seuil au cours du temps en fonction de l'évolution des caractéristiques du signal non stationnaires.

Cet algorithme (EZR) présente une méthode de décision V/NV basé sur l'énergie et le passage par zéro. Cette nouvelle méthode se caractérise par la facilité de mise en ouvre et à implémenter avec n'importe quelle méthode de détection de pitch au début d'algorithme ce qui permet de gagner de temps. Le seuil est estimé automatiquement par l'algorithme (est un seuil adaptatif) au cours de signal même si ce dernier de faible amplitude et énergie.

La méthode EZR est performante, la décision de V/NV est possible malgré un SNR important (SNR= -8dB) grâce au seuil adaptatif et qui dépend du signal lui même.

Malgré l'erreur de la décision au début de signal (Figure III.15, la zone cerclée) cette méthode est performante. Si les premières trames sont voisés vraiment , alors les résultats fournit seront juste et par suite le seuil sera optimal.

III.2.12 Limitations de la méthode d'autocorrélation :

L'estimation de la fréquence fondamentale du signal peut donc être faite par le calcul de l'autocorrelation basé sur une analyse LPC suivie d'une recherche de maximum.Un problème fréquemment rencontre est celui du doublement de période ou le pic situe à 2T0 possède une amplitude supérieure à celui situe à T0, ce qui conduit à estimer une fréquence fondamentale moitié de la fréquence fondamentale réelle (erreur d'octave). La fonction d'autocorrelation est sensible à l'influence des formants.

III.2.13 Conclusion :

- ✓ La fréquence fondamentale s'obtient par le maximum de la fonction d'autocorrelation du signal vocal
- ✓ La méthode EZR est plus performante que celle de SFM ; le contour de pitch pour la méthode EZR est plus étendu que celle de SFM, la décision de V/NV est possible malgré un SNR important (SNR= - 8dB) par la méthode EZR.

- ✓ L'algorithme de l'auto-corrélation est relativement imperméable au bruit.

*L'algorithme de l'auto-corrélation est relativement imperméable au bruit, mais sensible à la fréquence d'échantillonnage Fs***,** parce qu'il calcule directement la fréquence fondamentale d'un changement dans les échantillons, il suit que si nous avons un taux de l'échantillonnage inférieur, nous avons la résolution inférieure dans le pitch.

- ❖ Donc il est aisément de conclure la limitation de la méthode d'autocorrelation.

III.3 Fonction d'Autocorrelation [22,23]:

On applique l'analyse sur le signal lui même, la fonction d'autocorrelation à court terme est donnée par la relation II.1. En général, la fonction d'autocorrelation est une fonction paire qui s'étend de plus l'infini à moins l'infini. La figure III.21 exprime la fonction d'autocorrelation qui ne génère que la partie positive.

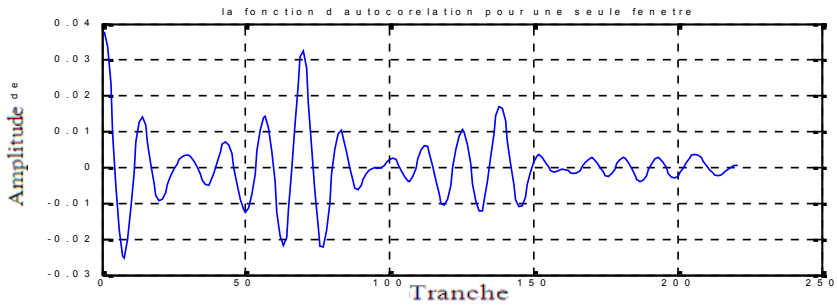


Figure III.21 La fonction d'autocorrélation

Il est clair que la fonction d'autocorrelation appliquée à une seule fenêtre contient des maximas qui se répètent à des intervalles réguliers. En effet, la fonction d'autocorrelation appliquée à un signal périodique présente des maximas aux multiples de la période du fondamentale. Le premier maximum correspond donc à la période du fondamental.

Ce maximum est déterminé après un filtrage ***passe bas*** de la tranche à analyser de 0 à 500 Hz et un fenêtrage d'une fenêtre ***Hamming.*** L'estimation de La fréquence fondamentale s'obtient par équation III.16 de la même façon qu'ACF_LPC.

III.3.1 Résultats expérimentaux :

On présente les résultats de simulation concernant cette méthode.

III.3.1.1 Contour de pitch :

La figure III.22 contient l'évolution de pitch et l'énergie en fonction du temps pour le phonème 'a'. La figure III.23 exprime le contour de pitch superposé sur la forme d'énergie, et

le pitch exact qui correspond à la valeur maximale d'énergie. Le pitch exact est de 155.1217 Hz.

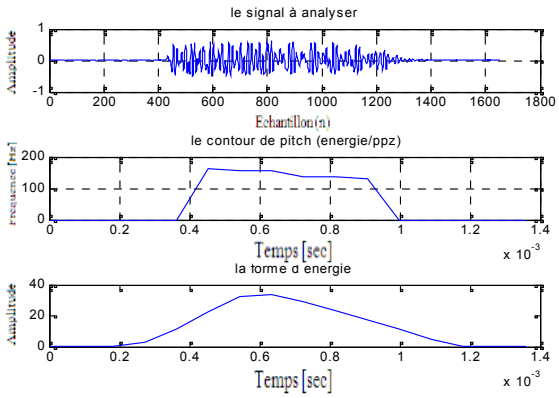


Figure III.22 De haut en bas : le phonème 'a ', Le contour de pitch, la forme d'énergie

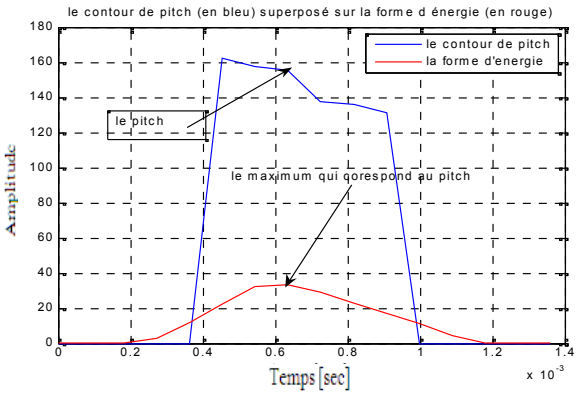


Figure III.23 Le contour de pitch superposé sur la forme d'énergie

III.3.1.2. Influence de bruit sur les résultats obtenus par l'autocorrelation :

On prend un signal parole synthétisé sans bruit exprime un phonème « a » (figure III.14) pour analyser la méthode ACF, avec une présence de différents puissance de bruit. Le tableau III.3 représente l'évolution de pitch estimé par ACF en fonction de SNR (rapport signal sur bruit).

SNR [dB]	Le pitch ACF [HZ]
Sans bruit	129.7059
30	129.7059
19	129.7059
0	129.7059
-07	130.5000
-11	135.2817

Tableau III.3 Le pitch d'ACF en fonction de SNR

III.3.2 Conclusion :

Il est clair qu'on a arrivé aux mêmes résultats que l'ACF_LPC. mais il est aisément de conclure que le temps de calcul d'ACF est plus petit que celle d'ACF_LPC.

L'ACF_ LPC est plus utile dans les systèmes de codage et de transmission.

III.4 AMDF (Average Magnitude Difference Function):

C'est la fonction de la différence de la magnitude moyennée définit par la relation II.7. La figure III.24 représente le boc diagramme de la méthode AMDF.

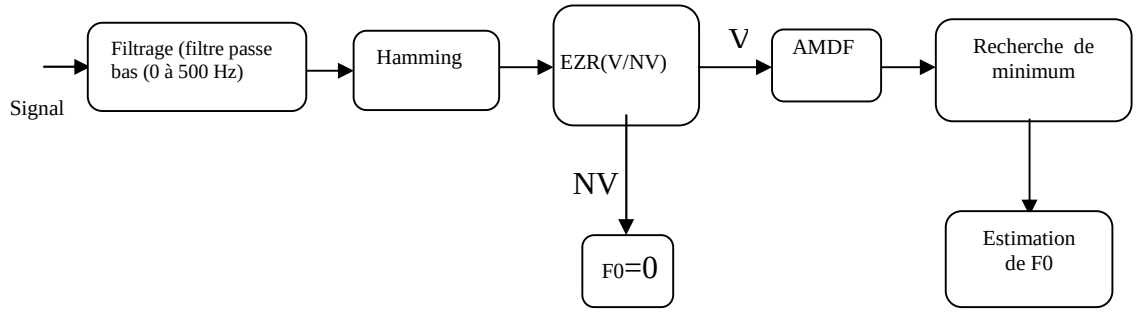


Figure III.24 Bloc diagramme d'AMDF

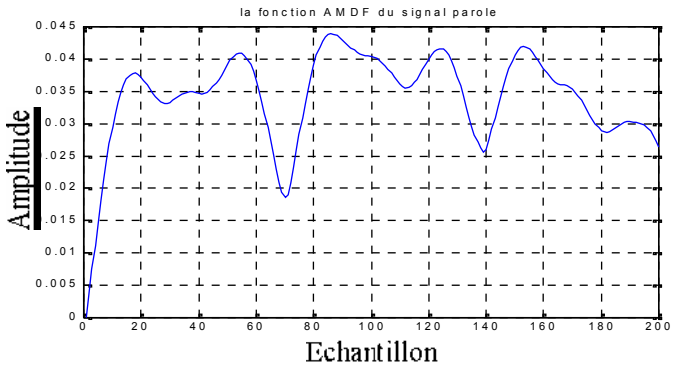


Figure III.25 la fonction AMDF d'une tranche de 30ms d'un phonème « a »

La fenêtre utilisée est la fenêtre Hamming. L'AMDF fournit une alternative plus rapide au calcul que la fonction de l'autocorrelation comme il n'exige pas de multiplications. La Figure III.25 représente le résultat de calcule de la fonction AMDF sur une tranche de 30ms qui recouvre plusieurs périodes fondamentale.

III.4.1 Estimation de la période du fondamental :

Le raisonnement pour le calcul de la période du fondamental est analogue au raisonnement pour l'auto corrélation. On constate que la fonction AMDF appliquée à une seule fenêtre contient des **minimas** qui se répètent à des intervalles réguliers. En effet, a la fonction AMDF appliquée à un signal périodique présente des minima aux multiples de la période du fondamentale. Le premier minimum correspond donc à la période du fondamental. La recherche de minimum de cette fonction permet de connaître la période de fondamentale. La valeur du fondamental estimée est alors :

$$F0 = Fs/m$$

m : c'est l'échantillon qui correspond au minimum.

III.4.2 Résultats expérimentaux :

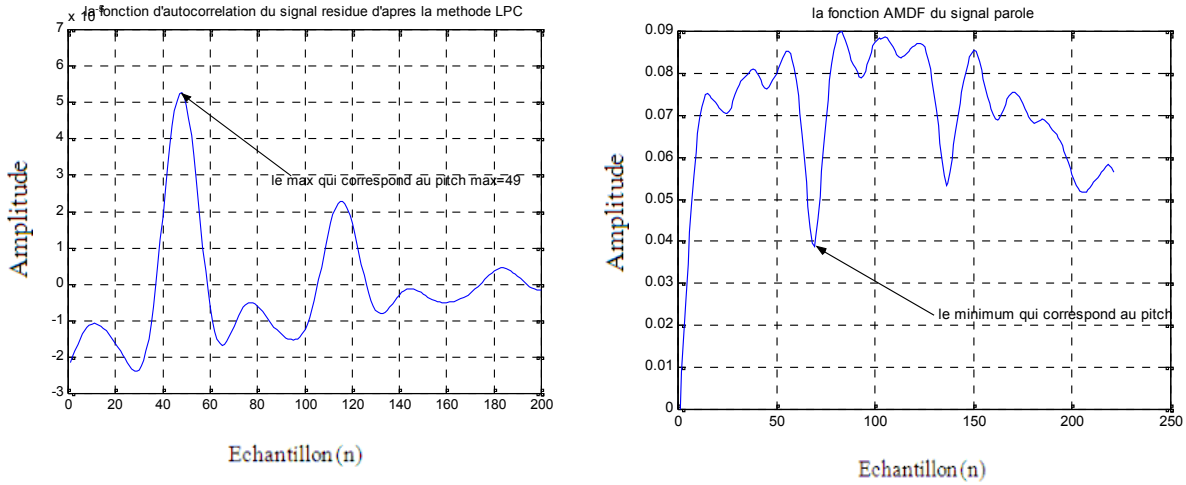
On présente les résultats de simulation par la méthode AMDF.

III.4.2.1. Estimation de pitch et comparaison avec celle d'ACF :

On prend une tranche de 30 ms (phonème 'a'). La figure III.26 (a, b) représente le résultat de la fonction ACF, AMDF, le minimum du signal (figure III.26 (b)) correspond à : m=70 (AMDF). Le maximum de signal (figure III.27(a)) correspond à : m=70 (ACF).

AN : F0= 11025/70=157,5 Hz qui exprime le pitch pour ACF et AMDF.

On trouve comme période **T = 0.0063** s F0 = **157,5 Hz**, résultats identique au résultat trouvée pour l'autocorrélation. La figure III.27 contient le résultat de la fonction AMDF (b) appliquée à la même fenêtre que la fonction d'autocorrelation (a).



(a) (b)

Figure III.26 a-ACF, b-AMDF

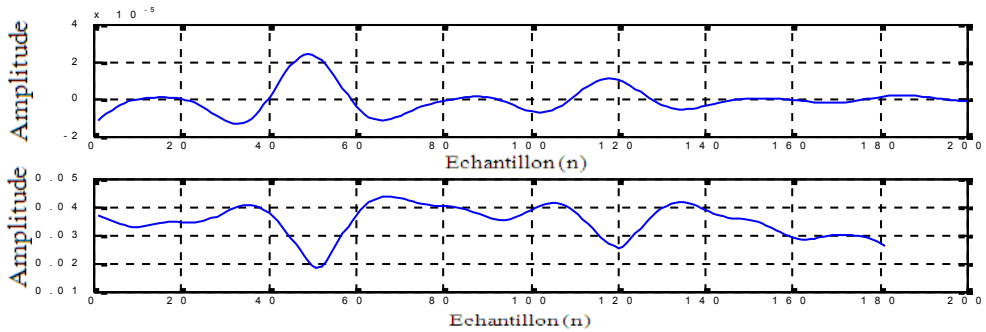


Figure III.27 Autocorrelation (en haut) et AMDF (en bas).

La figure III27 contient le résultat des fonctions AMDF et ACF appliquées à la même fenêtre. La fenêtre d'analyse contient **330** échantillons, et puisque le pitch est comprise entre 50 Hz et 500 Hz alors la recherche de la fréquence fondamentale se fait par la recherche du minimum dans min(c : m), c : le début de l'échantillon, dans notre exemples on a cherché dans l'intervalle **Fm=min (21 :330), F0=Fs /m (m : correspond à Fm). m ϵ [21,221]** qui correspond au **Fmax=Fs / 21= 525 Hz** et **Fmin=Fs/221 ≈ 50 Hz**

III.4.2.2. L'influence de puissance de bruit sur l'estimation de pitch et le contour :

On prend une tranche d'un signal « a » sans bruit et on effectue la recherche de pitch pour différentes valeurs de SNR, les résultats sont résumés dan le tableau III .4.

III.4.2.3. Analyse des phonèmes voisés (comparaison entre ACF et AMDF) :

Calculons la fréquence du fondamental pour quelques autres phonèmes voisés dans le cas d'une voix masculine, et chaque fois à l'aide de l'ACF et de la fonction AMDF. Les résultats sont résumes dans le tableau III.5.

Le pitch [Hz]	SNR [dB]
129.7059 Hz	38 dB
130.5010 Hz	30 dB
132.5013 Hz	26 dB

Tableau III.4 Le pitch d'AMDF en fonction de SNR.

Le phonème	ACF [Hz]	AMDF [Hz]
« a »	157.500	157.500
« b »	155.281	155.280
« I »	153.255	153.255
« O »	157.500	157.500

Tableau III.5 L'estimation de pitch pour plusieurs Phonèmes par ACF et AMDF

III.4.3 Conclusion :

Cette méthode a surtout été utilisée pour sa simplicité numérique (pas de multiplication) lorsque les processeurs de traitement du signal savaient surtout faire des additions, Mais elle se révèle très sensible au bruit. C'est l'une des plus anciennes méthodes de détection de la fréquence fondamentale

- On peut constater que l'autocorrelation et la fonction AMDF donnent toujours les mêmes résultats (d'après tableau III.6).
- La fonction AMDF sert à remplacer la fonction d'autocorrelation pour la détermination de la période du fondamental, parce qu'elle est plus facile `a calculer.
- La fonction AMDF est sensible au Bruit. Un peut de bruit peut provoquer une modification importante de pitch, donc c'est l'inconvénient essentiel de cette méthode,

Donc il est aisément de conclure la limitation de la méthode.

III.5 Average Square Difference Function (ASDF) [25,23]:

Est une l'un des méthodes d'estimation de la période à l'aide *des Fonctions de Différences Moyennées* [25 ,23] :

$$\mathrm{ASDF[m]} = \frac{1}{N-m}\sum_{n=0}^{N-1-m}(x[n]-x[n+m])^2 \qquad \text{III.23}$$

N : nombre d'echantillons
m : est la longueur de la fenêtre d'analyse.

Par définition, le signal x[n] (n : signifie échantillon) est rigoureusement périodique de période P si et seulement si :

$$x[n]-x[n+P] = 0 \quad ,\forall n \qquad \text{III.24}$$

On peut donc estimer la période **P** en recherchant le minimum par rapport à m [m minimal, m maximal] de l'écart quadratique entre les signaux x[n] et x[n + m] **[25]**:

$$E\,[(x[n]-x[n+m])\,2] = 2(\,rx[0]-rx[m]\,) \qquad \text{III.25}$$

Cet écart quadratique est nul en m = 0, et il est toujours supérieur ou égal à $2\,\sigma^2$ pour $m \neq 0$ Il atteint cette seconde valeur minimale pour tout 'm' multiple de P. Pour l'estimer, on définit l'Average Square Difference Function (ASDF).

III.5.1Résultats expérimentaux :

La figure III.28 représente la fonction ASDF pour une tranche de signal (« a ».) On veut comparer cette méthode avec les méthodes ACF (autocorrélation) et AMDF et voire l'influence de bruit vis-à-vis l'estimation de pitch On utilise un signal sans bruit (figure III.14) avec un filtre passe bas de 0 à 500Hz et un fenêtrage de Hamming.

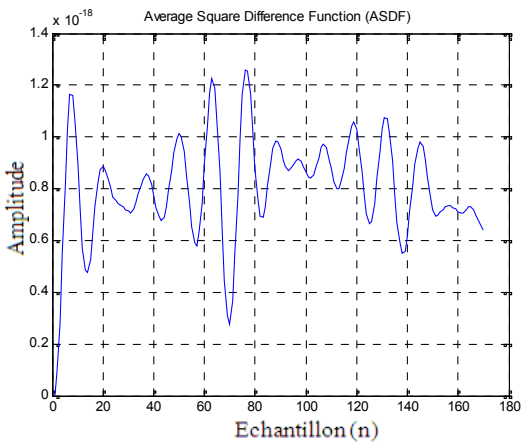


Figure III.28 La fonction d'ASDF

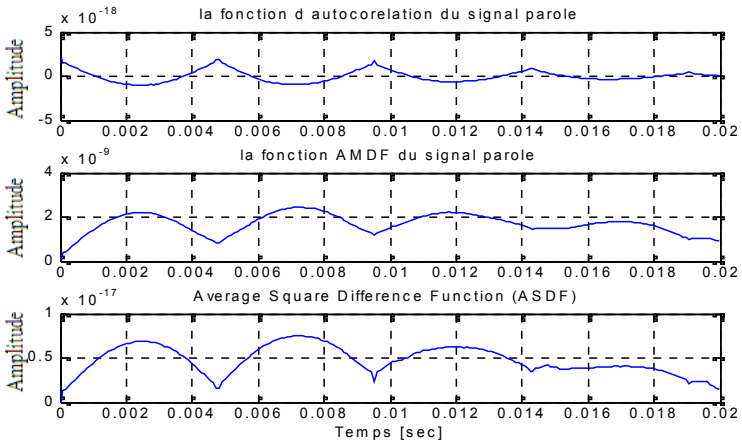


Figure III.29 La fonction ACF, AMDF, ASDF

Bruit [dB]	ACF [Hz]	AMDF [Hz]	ASDF [Hz]
Sans bruit	129.7059	129.7059	129.7059
47	129.7059	129.7059	129.7059
36	129.7059	129.7059	129.7059
27	129.7059	132.5013	110.5013
4	129.7059	***135.0101***	50.5946
3	129.7059	***140.2501***	50.5946

Tableau III.6 Le pitch en fonction de SNR pour ACF, AMDF, ASDF

La figure III.29 représente une comparaison de résultats d'une fenêtre par les méthodes ACF, AMDF, ASDF. Le tableau III.6 présente les résultats d'estimation de pitch en fonction de SNR pour les méthodes : ACF, AMDF, ASDF.

III.5.2 Conclusion :

Les méthodes d'ASDF, ACF, ASDF donnent toujours les mêmes résultats.

- La méthode ASDF est sensible au bruit que la méthode AMDF
- Il s'agit d'un estimateur non biaisé, l'estimateur d'AMDF est souvent préféré qu' ASDF car il nécessite moins de calculs (il n'y a plus de multiplications à effectuer) et il s'avère plus robuste sur des signaux réels.

III.6 Détermination du pitch par la méthode SIFT :

On va donc ici vous présenter une technique assez complète de détermination du pitch : le SIFT. On a ajouté la notion d'EZR pour une décision V/NV au début de traitement qui nous donne une possibilité de réduire le temps de calcul (parmi les inconvénients essentielles de cette méthode est la complexité de calcul pour cela on a introduit la notion EZR pour en l'améliorer), cette méthode prend en compte assez de paramètres pour donner de bons résultats.

III.6.1 Principes de la méthode :

La méthode du SIFT (Simplified Inverse Filtering) se base sur une estimation du maximum de l'autocorrelation d'un signal filtré par le filtre LPC inverse.

Pourquoi c'est deux techniques combinent ?

Premièrement il est logique d'utiliser l'autocorrelation pour trouver la période de répétition dans les échantillons de la fenêtre. Il est donc logique de retrouver '*n*' maximum de l'autocorrelation pour un décalage correspondant à la période de fondamentale du signal. On peut donc ainsi facilement déterminer le nombre d'échantillons correspondant à la période du signal.

En effet, la composition spectrale du signal vocal est composée des formants avec la fréquence fondamentale, ce qui "salit" en quelque sorte le signal duquel on veut retrouver le pitch.

C'est pourquoi la méthode du SIFT procède comme son nom l'indique un filtrage inverse du signal, dont le but est *d'arracher l'influence des formants*. En effet, avec analyse LPC on peut retrouver ces coefficients du filtre correspondant aux formants du signal, si l'on inverse ce filtre on annule donc l'effet des formants sur le signal.

III.6.1.1 Réalisation pratique

La figure III.30 représente le schéma de la réalisation pratique du SIFT (amélioré) :

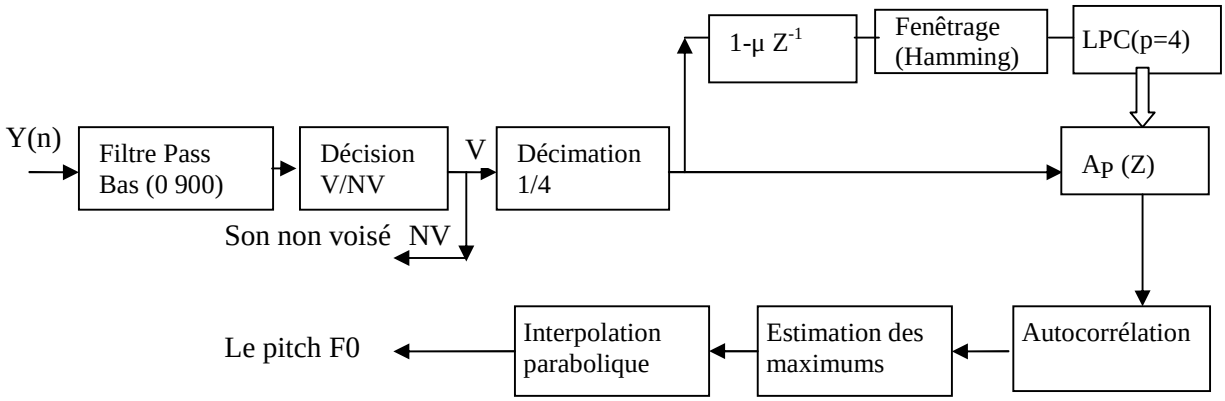


Figure III.30 Blocs diagramme amélioré de la méthode SIFT

On remarque la présence d'un filtre passe bas, dont le but est encore une fois d'arracher l'influence des fréquences autres que la fondamentale. Le filtre coupe 900Hz donne une valeur de pitch comprise entre environ 50 et 500 Hz. En effet le théorème de Shannon nous dit qu'il faut pour préserver la qualité d'un signal par échantillons cette fréquence d'échantillonnage doit être au moins deux fois la fréquence maximale du spectre du signal.

Le but de ce ***sous – échantillonneur*** (décimation ¼) est simplement de diminuer la charge de calcul, en diminuant le nombre de points sur lesquels se porte l'analyse.

Ensuite comme présenté, on effectue une préaccentuation et une pondération par une fenêtre de Hamming.

Une fois ces prétraitements sont effectuées, on effectue l'analyse LPC et on passe les coefficients ***PARCOR*** ainsi obtenus au filtre inverse, Ensuite, on calcule le vecteur d'autocorrelation de la fenêtre afin d'obtenir un maximum, afin d'encore accentuer le maximum, nous allons encore traiter le signal avant de calculer l'autocorrelation. Le traitement appliqué est simple, pour retrouver la période d'un signal on compte le temps entre **2 maximas**.

Le vecteur de corrélation ainsi obtenu présentera des pics pour des valeurs de décalage multiple de la période du signal et aura une valeur nulle pour les autres décalages. Nous pouvons donc maintenant calculer le vecteur d'autocorrelation, vecteur duquel on extrait le maximum.

En effet, afin de gagner en temps de calcul, nous avons un sous-échantillonnage du signal, en l'occurrence nous n'avons gardé qu'un échantillon sur 4, Or, le positionnement du maximum que nous obtenons se fait par valeur entière de la période d'échantillonnage ce qui provoque de perdre tous les efforts consentis pour obtenir une valeur la plus précise possible du pitch. C'est pourquoi on essaie d'interpoler la position du maximum, Cette interpolation nous permet donc de rattraper le manque de précision du positionnement du maximum tout en gardant l'avantage du sou - échantillonnage...

III.6.2 L'analyse de recherche du pitch par la méthode SIFT :

On a pour but d'expliquer le fonctionnement d'un analyseur par la méthode du filtrage inverse. Cette méthode permet de trouver le pitch ainsi que la caractéristique voisée ou non voisée d'une tranche d'un signal parole.

L'idée de cette méthode est de rendre le spectre du signal de parole le plus plat possible par filtrage et ensuite de retrouver le pitch par corrélation de ce signal.

III.6.2.1 le filtre passe bas :

Le but du filtre passe bas est de supprimer l'information spectrale superflue, c'est à dire celle composé par les formants autres que le premier formant.

En effet le signal de parole se compose d'environ **5 formants** espacé de 1KHz chacun. Il est donc naturel de se dire que comme seul nous intéresse le premier formant, il est judicieux de supprimer l'information qui n'est pas li par celui-ci. Ainsi diminuer la charge de calcul.

III.6.2.2 Discriminateur (voisé, non voisé) :

Ce bloc est responsable de la décision voisé ou non voisé de la tranche du signal (30ms un compromis de stationnarité d'un signal vocal)

III.6.2.3 Préaccentuation :

Ce bloc réalise la fonction « $\mathbf{1-\mu z^{-1}}$ », **μ** est le facteur de la préaccentuation.

III.6.2.4 Un fenêtrage de Hamming :

Il s'agit d'un bloc qui applique une fenêtre de Hamming au signal.

III.6.2.5 Un filtre LPC d'ordre 4 :

Ce codeur effectue son codage sur une tranche de 30 ms. On choisit une tranche de 30 ms dont les coefficients d'autocorrelation sont calculés, Le codeur utilise ensuite l'algorithme de Shur pour calculer les coefficients PARCORS de la tranche de signal considérée. Et on s'assure que ces coefficients qui représentent les coefficients de filtre sont à l'état stable.

III.6.2.6 Un filtre inverse :

Le filtre inverse utilise les coefficients comme paramètre pour le filtrage du signal. Il prend directement les coefficients PARCORS comme paramètres.

III.6.2.7 Un corrélateur :

Ce bloc qui est aussi intéressants le codeur LPC calcule les coefficients d'Autocorrélation du signal résultant dans le but de retrouver la composante qui est le pitch du signal.

III.6.2.8 Estimation des maximums :

Ce bloc essaie de trouver les maximums d'autocorrection du signal après effectuation de l'absolu de l'Autocorrélation. Cette recherche pourrait être simple et renvoyer le maximum d'Autocorrélation tel quel mais cette solution n'est pas la plus efficace. C'est pourquoi cette méthode est suivie d'une correction effectuant le choix du maximum parmi les 2 échantillons maximas et cela en fonction des 2 maximas de la tranche précédente.

Le but de cette correction est de suivre le maximum le plus plausible sans être perturbé par des signaux parasites. Cela permet d'atténuer le saut du pitch lors de la décision, ce qui peut provoquer des effets indésirable lors de la reconstruction du signal codé par un décodeur LPC.

III.6.3 Résultats numériques :

On a enregistré un phonème « a » prononcé par un masculin par suite on a procédé l'analyse. La figure III.31 exprime de haut en bas, signal de phonème « a », contour de pitch, zone V/NV (zone voisée=1, zone non voisé= 0), forme d'énergie.

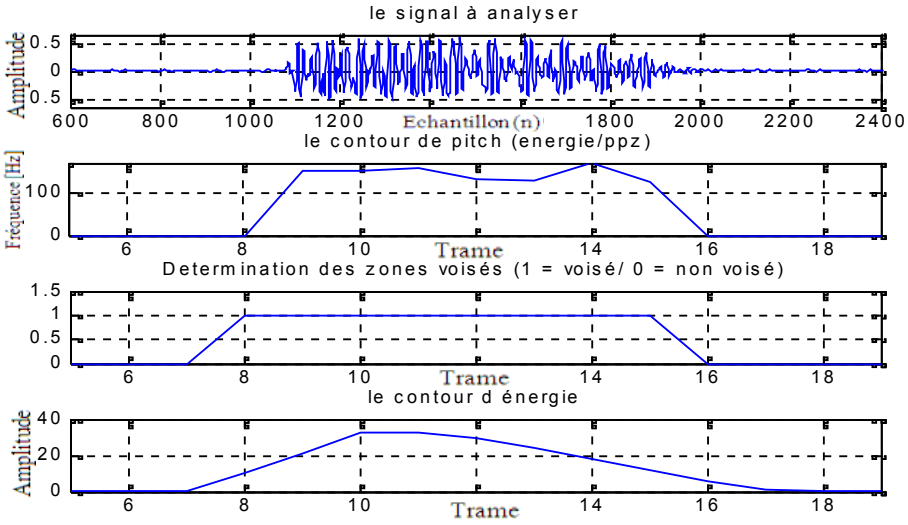


Figure III.31 Signal de phonème « a », contour de pitch, zones V/NV, forme d'Energie.

Les fréquences fondamentales trouvées par EZR (V/NV) qui constitue le contour de pitch:

153.1250 153.1250 **157.5000** 131.2500 128.1977 172.2656 125.2841

Les valeurs efficaces correspondantes :

0.2528 0.3128 **0.3151** 0.2984 0.2691 0.2356 0.1911

- La fréquence fondamentale qui exprime le pitch est **157.5 Hz** qui correspond à une valeur maximale de valeur efficace : **0.3151**

Tant qu'on a analysé le signal complet, alors le signal doit être segmenté en plusieurs trames (de 30ms c à -d 330 échantillons avec Fs=11025 Hz), chaque trame doit être fenêtrée et analysée séparément, dans le cas voisé on procède l'analyse et le calcul des coefficients Parcor, ainsi les fréquences fondamentales. La Figure III.32 présente les résultats graphiques pour une trame voisée.

Les résultats pour la première fenêtre (première trame) :

- Le max après interpolation : 18.0000
- Fréquence fondamentale : 153.1250 Hz
- Coefficients de PARCOR

1.0000 0.3153 0.7264 0.1764 0.5890

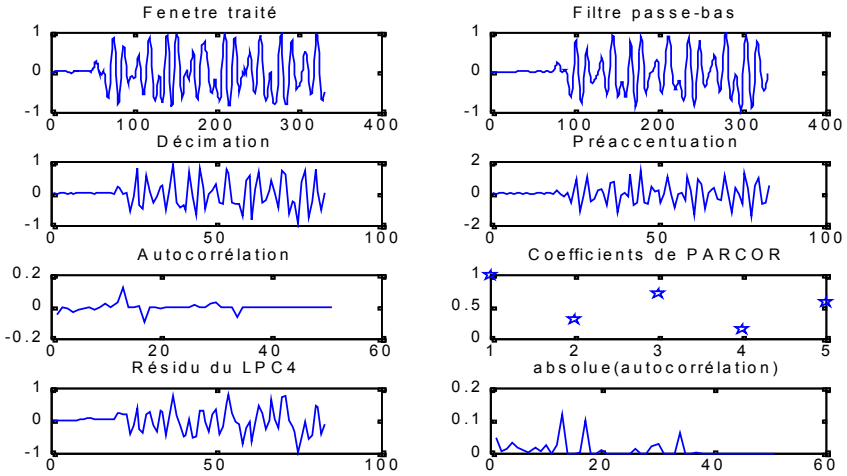


Figure III.32 La première fenêtre (tous ces graphes en fonction de temps sauf les PARCOR)

b. L'influence de bruit sur l'estimation de pitch et la décision V/NV :

On prend un signal parole sans bruit exprime un phonème « a » (figure III.14) pour analyser la méthode SIFT, avec une présence de différents SNR. Le tableau III.7 représente l'évolution de pitch estimé par SIFT en fonction de SNR. La méthode donne toujours les mêmes résultats en dépit de la présence de perturbation (moins de -13 dB) et maintient la même valeur de pitch.

SNR [dB]	Le pitch Energie/PPZ [HZ]
Sans bruit	153.1250
38	153.1250
30	153.1250
2	153.1250
-9	153.1250
-13	Non voisé

Tableau III.7 Le pitch par SIFT en fonction de SNR

III.6.4 Conclusion :

- L'algorithme de SIFT est relativement imperméable au bruit, mais sensible à la fréquence d'échantillonnage Fs.
- On voit bien que, la méthode du SIFT donne de bons résultats.
- Le seul inconvénient c'est la complexité de calcul.
- Je pense que cet algorithme du SIFT en est un bon exemple.

III.7 Conclusion :

- Les méthodes ACf et ACF_LPC, très résistantes au bruit.
- Les deux méthodes AMDF, ASDF donnent presque toujours les mêmes résultats, sont simples concernant la facilitée et la rapidité de calculs, mais sensible au bruit.
- Concernant la méthode SIFT l'inconvénient essentielle est la complexité de l'algorithme, qui peut engendrer des erreurs du calcul.
- Les méthodes temporelles permettent une estimation de la période T0 avec un délai minimal, et calculs très simples. Pour ces deux raisons, ce furent les premières à être utilisées.

Bibliographie :

[15] Alexis Moinet & Maxime Tryhoen, « Implémetation d'un codeur LPC10 complet sous Matlab »**,** Faculté Polytechnique de Mons, Belgique, 9 Rue de Houdain, 7000 Mons, France.

[16] Adeldjalil Ouhabi « Techniques avancées de traitement du signal et Applications »; Alger 1993

[17] *Shlomo Dubnov* ." Non - Gaussian Source - Filter and Independent Components Generalizations of Spectral Flatness Measure". Ben-Gurion University (ocupated Palastine).2003

[18] *Shlomo Dubnov* ."Generalization of Spectral Flatness Measure for Non-Gaussian Linear Processes ".". Ben-Gurion University (ocupated Palastine).2003.

[19] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 2-8, Feb. 1976

[20] *Li Hui, Bei-qian Dai, Lu Wei* « a pitch détection algorithme based on Amdf and Acf » MOE-Microsoft Key Laboratory of Multimedia Computing and Communication, University of Science and Technology of China 2003.

[21] J.MAX, D.Berthier, H.Chevalier, B.Escudie, A.Hellion, M.Martin, M.Trottot «Méthodes et technique de Traitement du signal »Deuxiemme édition MASSON, Paris, Neuw york, Bercelone, Milan. 1977.

[22] J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-timedigital hardware pitch detector," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-24, pp. 2-8, Feb. 1976.

[23] MICHAEL.J.CHENG,STUDENT " A Comparative Performance Study of Several Pitch Detection Algorithms LAWRENCER.RABINER.FELLOW,IEEE, 1970.

[24] Dr. Andrzej Drygajlo .ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE ; EPFL - Faculté STI - ITS SCG Laboratoire de Traitement Numérique de la Parole ; TRAVAUX PRATIQUES B, 2003 ;

[25] Dr. Roland Badeau. http://www.perso.enst.fr , détection de hauteur. Notes de cours.

IV.1 Introduction :

On explore plusieurs algorithmes appartenant aux domaines fréquentiels. Nous essayerons de mettre en évidence le comportement de chaque algorithme vis-à-vis des problèmes liés à la détermination de la fréquence fondamentale et la décision voisée non voisée nécessaire pour la détection du pitch.

IV.2 Méthode de Cepstre :

La méthode de «cepstre » est une des techniques fréquentielle pour déterminer la fréquence du signal, L'idée fondamentale derrière le cepstre est que le signal périodique peut être considère comme la convolution d'un train d'impulsions par un filtre amorti. Dans le domaine des fréquences, les spectres sont multiplies mais en prenant le « log » du résultat, on obtient la somme des résultats indépendants. De cette façon, une convolution dans l'espace des temps correspond à une addition dans le domaine du cepstre: Si les deux spectres ont des caractéristiques différentes, il devient possible de les séparer.

IV.2.1 Modèle source -filtre pour un signal périodique :

Nous modélisons maintenant le signal observé x[n] comme un signal déterministe et rigoureusement périodique de période P (on supprime le terme de bruit additif). Son spectre est :

$$\mathrm{X}(\mathrm{e}^{\mathrm{j}2\pi\mathrm{f}}) = \sum_{-\infty}^{+\infty} x[n] e^{-j2\pi nf} \qquad \text{IV.1}$$

Est alors constitué d'harmoniques à la fréquence fondamentale $f_0 = \frac{1}{P}$.

Si nous supposons connue une enveloppe $H(e^{j2\pi f})$ continue de ces harmoniques, X peut être vu comme le produit de **H** par un peigne de Dirac séparés de la fréquence **f0**, que nous noterons **S**.

Dans le domaine temporel, on obtient alors :

$$x = h \otimes s \qquad \text{IV.2}$$

On peut donc considérer que x est obtenu par filtrage de la source s (qui est un peigne de Diracs séparés de la période P) par le filtre H [25] :

$$H(e^{j2\pi f}) = \frac{\sum_{k=0}^{N} \beta_k e^{-j2k\pi f}}{\sum_{k=0}^{D} \alpha_k e^{-j2\pi kf}} \qquad \text{IV.3}$$

Ce modèle source-filtre apparaît ici comme un artifice mathématique pour représenter le signal, mais il correspond en fait à une réalité physique du mécanisme de production des signaux de parole et de la plupart des signaux de musique.

IV.2.2 Définition du cepstre d'un signal discret

On constate que le logarithme du spectre d'amplitude de X est simplement la somme de la contribution du filtre et de celle de la source [25]:

$$\ln(|X(e^{j2\pi f})|) = \ln(|H(e^{j2\pi f})||S(e^{j2\pi f})|) = \ln(|H(e^{j2\pi f})|) + \ln(|S(e^{j2\pi f})|) \quad \text{IV.4}$$

C'est sur cette propriété fondamentale que repose la définition du cepstre:

Le cepstre d'un signal discret X est la transformée de Fourier inverse du logarithme de son spectre d'amplitude [25]:

$$Cx[n] = \int_{f=-\frac{1}{2}}^{+\frac{1}{2}} \ln(|X(f)|) e^{+j2\pi nf} df \quad \text{IV.5}$$

Il s'agit d'un signal temporel. Si x[n] est un signal réel, Cx[n] est un signal réel et symétrique.

IV.2.3 Propriétés du cepstre :

Le Cepstre d'un signal est la somme des cepstres de h et s:

$$C_x[n] = C_h[n] + C_s[n] \quad \text{IV.6}$$

L'intérêt de cette décomposition cepstrale du signal est que dans la pratique, $C_h[n]$ et $C_s[n]$ ont des supports temporels disjoints, ce qui permet de séparer facilement la source et le filtre.

IV.2.3.1 Cepstre du filtre :

Nous allons commencer par calculer le cepstre du filtre h. On en déduit [25]:

$$|H(e^{j2\pi f})|^2 = A \frac{\prod_{k=1}^{N} (1 - a_k e^{-j2\pi f})(1 - a_k^* e^{+j2\pi f})}{\prod_{k=1}^{D} (1 - c_k e^{-j2\pi f})(1 - c_k^* e^{+j2\pi f})} \quad \text{IV.7}$$

On peut supposer sans perte de généralité que $|a_k| < 1 \quad \forall k = 1...N$ et $|c_k| < 1 \quad \forall k = 1...D$

En prenant le logarithme complexe des deux membres de l'égalité précédente, on obtient:

$$2\ln(|X(e^{j2\pi f})|) = \ln(A) + \sum_{k=1}^{N} \ln(1 - a_k e^{-j2\pi f}) + \sum_{k=1}^{N} \ln(1 - a_k^* e^{+j2\pi f}) - \sum_{k=1}^{D} \ln(1 - cke - j2\pi f) - \sum_{k=1}^{D} \ln(1 - c_k^* e - j2\pi f) \quad \text{IV.8}$$

Comme les a_k et les c_k sont de module < 1, on peut utiliser le développement en série entière du Logarithme complexe [25]:

$$\ln(1-x) = -\sum_{n=1}^{+\infty} \frac{x^n}{n} \qquad \text{IV.9}$$

On obtient alors [25]:

$$2\ln\left(\left|H(e^{j2\pi f})\right|\right) = \ln(A) - \sum_{k=1}^{N}\sum_{n=1} \frac{(a_k)^n e^{-j2\pi nf}}{n} - \sum_{k=1}^{N}\sum_{n=1}^{+\infty} \frac{(a_k^*) + e + j2\pi nf}{n} + \sum_{k=1}^{D}\sum_{n=1}^{+\infty} \frac{(c_k)^n e^{-j2\pi nf}}{n} + \sum\sum \frac{(c_k^*)^n + e^{+j2\pi nf}}{n} \qquad \text{IV.10}$$

Or par définition du cepstre, on en déduit [25] :

$$C_h[n] = \begin{cases} \sum_{k=1}^{N} \frac{(a_k^*)n}{2n} - \sum_{k=1}^{D} \frac{(c_k^*)n}{2n} & Si \quad n < 0 \\ \frac{\ln(A)}{2} & Si \quad n = 0 \\ \sum_{k=1}^{N} \frac{(c_k^*)n}{2n} - \sum_{k=1}^{D} \frac{(a_k^*)n}{2n} & Si \quad n > 0 \end{cases} \qquad \text{IV.11}$$

Par conséquent, si le filtre H et son inverse sont peu résonnants, c'est-à-dire si les a_k et les c_k sont de module suffisamment inférieur à 1, le support temporel du cepstre $C_h[n]$ est localisé au voisinage de zéro.

IV.2.3.2 Cepstre de la source :

On suppose que le spectre de la source est défini de la façon suivante (avec $|c| < 1$) [25]:

$$S\left(e^{j2\pi f}\right) = \frac{1-|c|^2}{(1-(ce^{-j2\pi f})^p)(1-(c^* e + j2\pi f)\, p)} \qquad \text{IV.12}$$

Nous allons vérifier que lorsque c $\longrightarrow$ 1, s[n] converge vers un peigne de Dirac à la période P. En effet,

$$S\left(e^{j2\pi f}\right) = \frac{1}{(1-(ce^{-j2\pi f})^p)} + \frac{1}{1-(c^* e + j2\pi f} \sum_{n=0}^{+\infty} (c)^{nP} e^{-j2\pi nf}) + \sum_{n=0}^{+\infty} (c^*)^{nP} e^{+j2\pi nPf}) - 1 \qquad \text{IV.13}$$

Par identification, on en déduit que [25]:

$$s[n]=\begin{cases}(c^*) & Si\ n<0\ et\ P\ divise\ n\\ 1 & Si\quad n=0\\ (c)n & Si\ n>0\qquad et\ P\ divise\quad n\\ 0 & Sinin\end{cases} \qquad \text{IV.14}$$

Lorsque c $\longrightarrow$ 1, s[n] converge vers $\sum_{k=-\infty}^{+\infty}\delta[n-kP]$

Nous allons maintenant calculer le cepstre de s [25]:

$$\begin{aligned}\ln\left(\left|H(e^{j2\pi f})\right|\right) &= \ln(S(e^{j2\pi f})\\ &= \ln(1-|c|^2)-\ln(1-(e-j2\pi f)P)-\ln(1-(c^*e+j2\pi f)P)\\ &= \ln(1-|c|^2)+\sum_{n=1}^{+\infty}\frac{(c)^{nP}e^{-j2\pi nPf}}{n}+\sum_{n=1}^{+\infty}\frac{(c^*)^{nP}e^{+j2\pi nPf}}{n}\end{aligned} \qquad \text{IV.15}$$

Par identification, on obtient [25]:

$$C_s[n]=\begin{cases}\dfrac{P(c^*)^{-n}}{-n} & Si\ n<0\ et\ P\ divise\ n\\ \ln(1-|c|^2) & Si\quad n=0\\ \dfrac{P(c)^n}{n} & Si\ n>0\qquad et\ P\ divise\quad n\\ 0 & Sinin\end{cases} \qquad \text{IV.16}$$

Lorsque c $\longrightarrow$ 1, Cs[n] diverge en n = 0, mais pour tout n ≠ 0,

$$C_s[n]=P\sum_{k=-\infty}^{+\infty}\frac{1}{|kP|}\delta[n-kP] \qquad \text{IV.17}$$

Le cepstre de la source est donc un peigne d'impulsions à la cadence P, qui décroît lentement lorsque $n \longrightarrow \pm\infty$.

- Le logarithme a pour effet de transformer la multiplication en addition et de séparer les signaux ; il permet également de mettre en évidence les composantes spectrales de faibles amplitudes, contrairement à la fonction d'autocorrélation. Le cepstre du signal vibratoire est donc la somme de deux cepstres

IV.2.4 Estimation de la période à l'aide du cepstre

Ainsi, nous avons vu que le cepstre du filtre **h** a un support temporel localisé au voisinage de 0, alors que celui de la source **s** est un peigne d'impulsions à la période P et à décroissance lente.

On peut donc estimer la période P en déterminant l'intervalle de temps qui sépare deux impulsions successives, ou en recherchant le maximum global du cepstre sur un intervalle n Є $[n_{min}, n_{max}]$, avec nmin > 0.

Dans la pratique, il est nécessaire d'appliquer une *fenêtre de pondération* au signal avant de calculer son cepstre. La figure IV.1 représente le bloc diagramme de la méthode de cepstre :

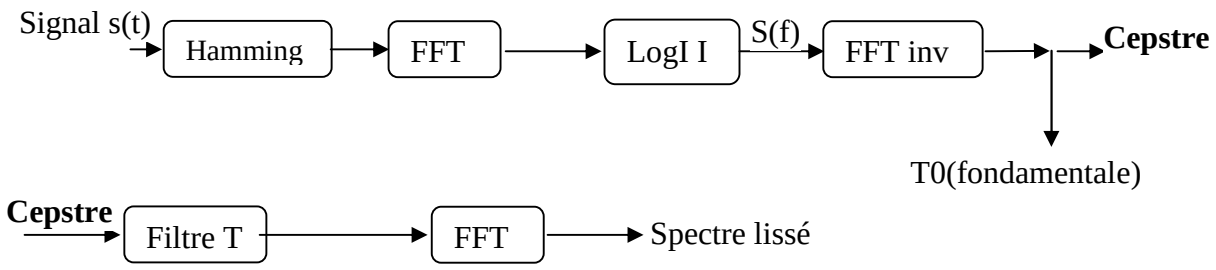


Figure IV.1 Le bloc diagramme de la méthode de cepstre [12]

IV.2.5 Résultats expérimentaux :

On applique la méthode de cepstre sur un signal parole. Dans notre travaille on a procédé à sélectionner une tranche de 330 (correspond au 30ms) échantillons qui correspond au compromis de la stationnarité du signal pour une fréquence d'échantillonnage de 11025Hz pour effectuer l'analyse par suite on prend le signal complet avec un fenêtrage sur le long de signal avec un saut de 10 ms (110 échantillons) pour construire le contour de pitch en s'appuyant sur la décision V/NV par EZR, et enfin on évalue la méthode en fonction de différents puissance de bruit.

a. sur une seule tranche de 30ms :

La tranche exprime un phonème « a » (voire Figure IV.5).

a.1 Filtrage de signal :

A cause que le pitch est comprise entre 50 et 500 Hz on a choisit un filtre qui est un filtre passe bas d'ordre 8 (filtre de butter Worth).

a.2 un fenêtrage de Hamming :

Apres avoir filtré le signal et afin d'éliminer l'effet de bord on a utilisé la fenêtre de Hamming.

a.3 LA FFT du signal :

Dans cette étape on procède la transformé de fourrier rapide (FFT) avec une limitation d'un demi du nombre de points (1/2 du nombre d'échantillons). La figure IV.2 exprime la FFT de signal.

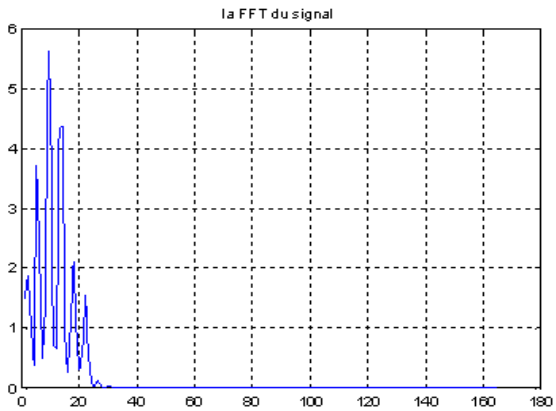


Figure IV.2 La FFT de signal (amplitude Vs fréquence [Hz]).

a.4 logarithme du la FFT du signal :

Dans cette étape on fait le logarithme à la FFT du signal (afin de séparer la source de conduit vocal). La Figure IV.3 représente le logarithme de la FFT.

a.5 la FFT inverse du signal :

La FFT inverse du signal résulte le 'Cepstre', l'estimation de pitch est achevé par la recherche de maximum de cepstre, ce maximum est correspond au pitch recherché. La figure IV.4 représente le cepstre du signal source où le maximum est bien claire.

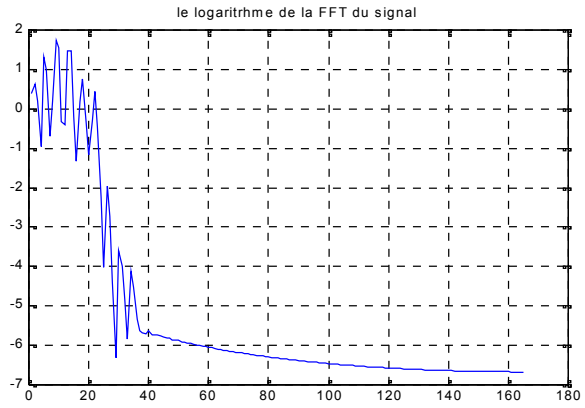


Figure IV.3 Le logarithme de la FFT

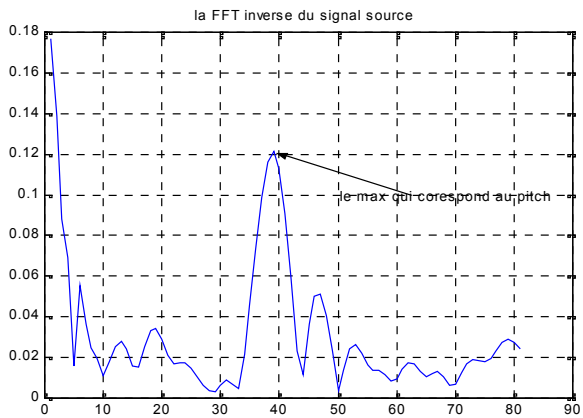


Figure IV.4 Le cepstre de signal

a.6 estimation de pitch :

Dans notre cas et à partir du la figure IV.4 on a max =35, et puisque la FFT est symétrique on a procéder la recherche dans une seule partie (droite) , alors le max devient égale à m= 70, le calcul de pitch est calculé de la même façon que les méthodes temporelle (F0= Fs/m).

AN : Fs= 11025 Hz , m= 70. Alors le pitch est égale à : **F0 = 157.50 Hz.**

b. L'influence de bruit sur la détection de pitch et la décision V/NV :

On prend un signal sans bruit exprime un phonème « a » avec une présence de différents SNR. Le tableau IV.1 représente l'évolution de pitch estimé par la méthode cepstre en fonction de SNR. La Figure IV.5 représente le signal « a » et le contour de pitch par cette méthode.

La Figure IV.6 représente le signal « a » et le contour de pitch(en bas) pour SNR=10 dB

SNR [dB]	Le pitch [HZ]
Sans bruit	129.7059
45	129.7059
10	129.7059
5	122.7059
2	135.4211
0	Non voisé

Tableau IV.1 Le pitch par Cepstre en fonction de SNR

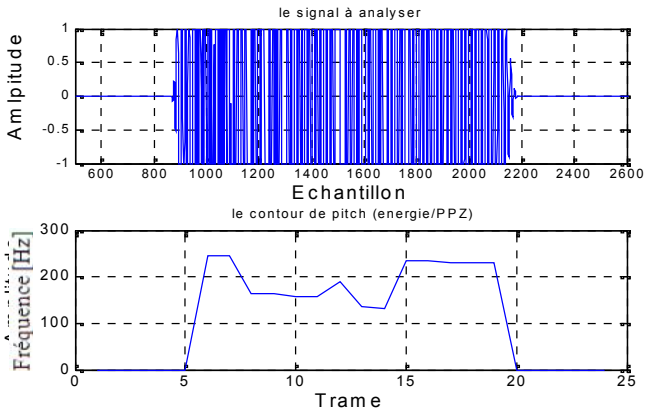


Figure IV.5 : Le signal « a » sans bruit (en haut) et le contour de pitch(en bas).

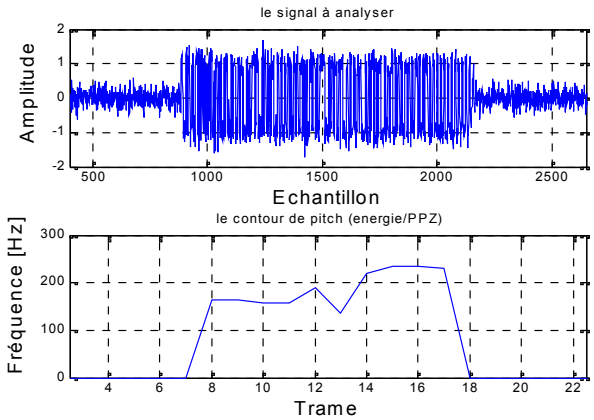


Figure IV.6 : Le signal « a » (en haut) et le contour de pitch(en bas) pour SNR=10 dB.

- On remarque des résultats acceptable jusqu'au SNR= 2 dB.
- Le pitch égal à 129.7059 Hz (sans bruit) et pour SNR =10 dB.

IV.2.6 Problèmes et limitations:

Les méthodes Cepstrales présentent cependant un Certain nombre de problèmes:

- ✓ Il est nécessaire d'appliquer au signal une fenêtre de pondération, ce qui dans le cas de fondamentaux de fréquence basse (faible nombre de périodes dans la fenêtre) atténue fortement les pics cepstraux.
- ✓ Si le signal possède peu d'harmoniques, son cepstre ne présente plus de pic à n = T0 (cas limite d'une sinusoïde).

Le cepstre est utilisé non seulement en traitement de la parole, mais aussi en traitement du son et en traitement d'image (amélioration du contraste).

IV.2.7 Conclusion :

- La fonction cepstre est plus performante à cause qu'elle sépare le spectre de conduit vocal avec le signal source, ainsi le pitch est facile à calculer les résultats obtenus.
- La fonction cepstre n'est pas très sensible au bruit, les résultas obtenus montre que le pitch débute de tomber aux erreurs dans SNR =5 dB mais elle donne encore des résultats acceptables jusqu'au SNR= 2 dB.

IV.3 Méthode HPS (Harmonic Product Spectrum):

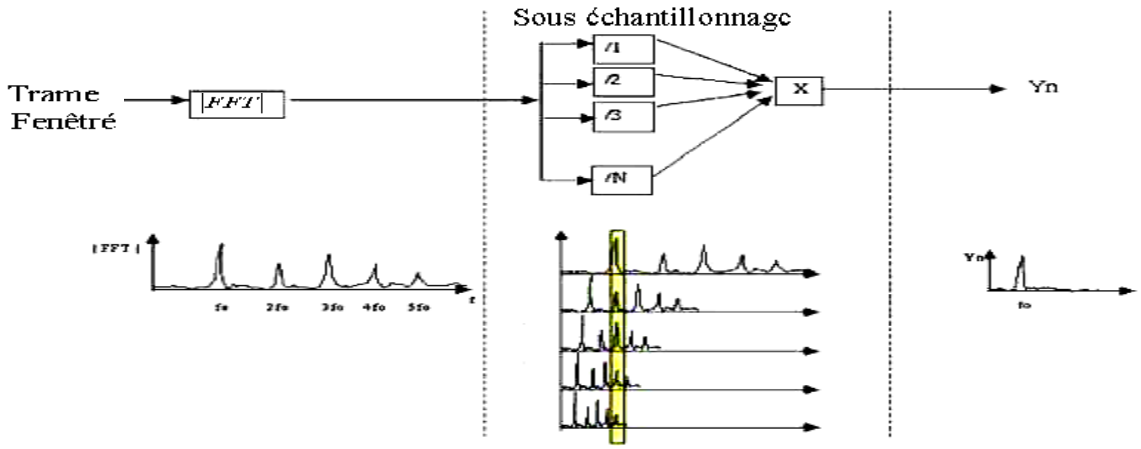


Figure IV.7 llustration d'algorithme d'HPS [10]

Cet algorithme estime le pitch comme la fréquence dont maximise le produit de spectre des harmoniques de cette fréquence. La Figure IV.7 illustre l'algorithme d'HPS (Schroeder, 1968), HPS s'écrit comme suit [10] :

$$F0 = \arg\max_{f} \prod_{k=1}^{n} \left| X(kf) \right| \qquad \text{IV.18}$$

Où : X : c'est le spectre du signal

n : c'est le nombre d'harmonique à utiliser. (De préférence entre 5 et 11[10])

F0 : c'est le pitch à estimer.

VI.3.1 Méthode :

En premier, Dans notre étude on a utilisé un nombre d'harmonique est égale à **3,** ainsi un filtrage passe bas (0 500 Hz) avec une fréquence d'échantillonnage Fs, et appliquant une fenêtre de Hamming où la dimension de la fenêtre est de 30ms, le petit saut est de 10 ms. Pour chaque fenêtre, nous utilisons la transformée de Fourier à Court terme pour convertir le signal du domaine du temps au domaine de la fréquence. Une fois le signal est dans le domaine de la fréquence, nous appliquons la technique du Spectre du Produit Harmonique à chaque fenêtre.

Le HPS implique deux pas: sous échantillonnage et multiplication. Pour sous échantillonner, nous avons compressé deux fois le spectre dans chaque fenêtre par re-échantillonner: la première fois, nous compressons le spectre original par deux et la deuxième fois, par trois. Une fois cela est complété, nous multiplions les trois spectres ensemble et trouvons la fréquence qui correspond au sommet (valeur maximale). Cette fréquence particulière représente la fréquence fondamentale de cette fenêtre particulière.

IV.3.2 Problème de la méthode :

D'après la définition de cette méthode le problème essentiel c'est l'estimation de nombre d'harmonique (n), il n'y a pas une raison logique de limiter le nombre d'harmonique, la seule raison c'est de réduire le calcul en utilisant un nombre d'harmonique faible (exemple n=3).
Mais le problème essentiel de cet algorithme est que si chacun de l'harmoniques manque (son énergie est égale à zéros) le produit est égal à zéro ce qui conduit à un résultas faux.

IV.3.3 Résultats expérimentaux :

a. estimation de pitch :

On a enregistré un son d'un phonème voisé (phonème « a ») et choisissant une fenêtre à analyser de 330 échantillons (Fs=11025) en suite calculant le pitch, le pitch estimé est de **164.1907 Hz.** La Figure IV.8 exprime le résultat d'application de HPS pour une seule fenêtre d'un signal « a ».
Prenant le signal complet de phonème « a » et construisant le contour (la décision V/NV est assurée par EZR) de pitch. La figure IV.9 exprime le contour de pitch d'un signal (« a ») par HPS.

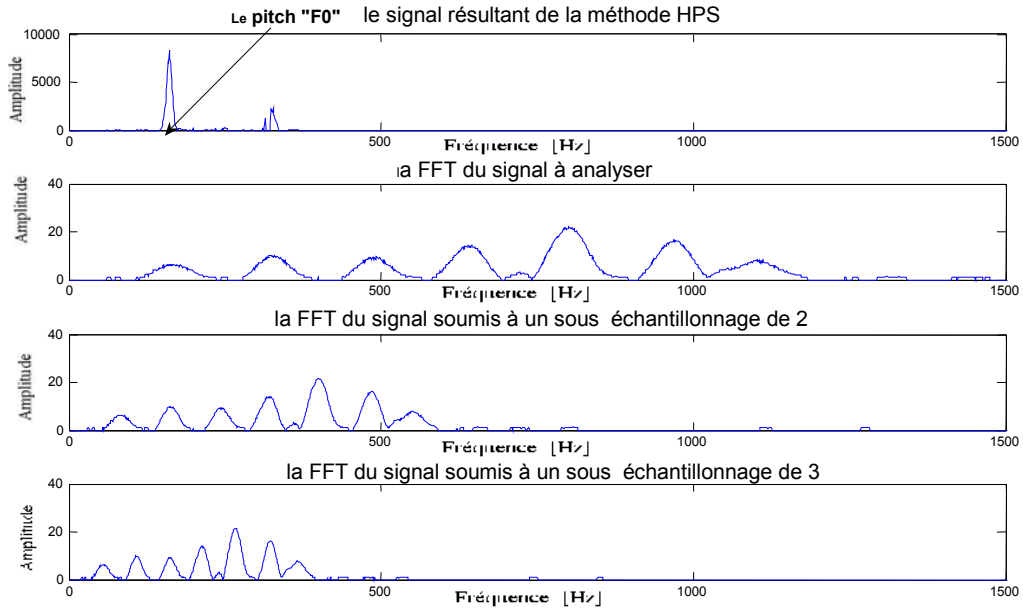


Figure IV.8 HPS pour une seule fenêtre d'un signal du phonème « a » en fonction d'échantillon

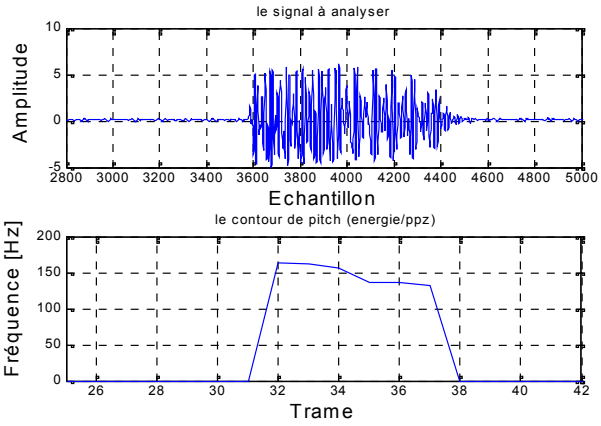


Figure IV.9 Le contour de pitch d'un signal (« a ») par HPS, F0= 164.1907 Hz.

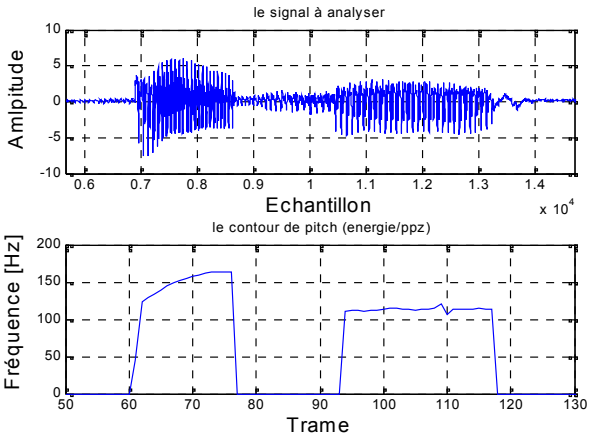


Figure IV.10 Le contour de pitch «Bonjour » de la méthode HPS, F0=139.9658 H.

b. Estimation de pitch en conditions bruitées :

Dans cette étude, on prend un signal sans bruit (phonème « a ») pour analyser la méthode HPS en présence de différentes puissances de bruit (bruit blanc), les résultats sont résumés dans le tableau IV.2.

Le pitch [Hz]	SNR [dB]
129.7059	Sans bruit
129.7059	30 dB
157.1012	17 dB
133.3120	9 dB
136.0160	2 dB
137.1630	1 dB
131.8500	-5 dB

Tableau IV.2 Le pitch par HPS en fonction de SNR

IV.3.4 Amélioration de la méthode :

Pour remédier au problème de la méthode (que si chacun de l'harmonique manque, son énergie est égale à zéros) et depuis que le logarithme est une fonction croissante, une approche équivalente d'estimer le pitch comme la fréquence qui maximise le logarithme du produit du spectre à harmoniques de cette fréquence. Depuis le logarithme d'un produit est égal à la somme des logarithmes des termes, HPS peut être écrit comme suit :

$$F0 = \arg\max_{f} \sum_{k=1}^{n} \log\left|X(kf)\right| \qquad \text{IV.19}$$

Pour éviter le problème de log(0), on ajoute un nombre suffisamment petit $\boldsymbol{\varepsilon}$ au spectre *X(kf)* HPS peut être écrit comme suit :

$$F0 = \arg\max_{f} \sum_{k=1}^{n} \log\left|X(kf) + \varepsilon\right| \qquad \text{IV.20}$$

IV.3.4.1 Evaluation de la méthode :

on prend une trame exprime un lettre « a »,et estimons le pitch. La figure IV.11 représente le HPS logarithmique pour une trame d'un signal « a ».

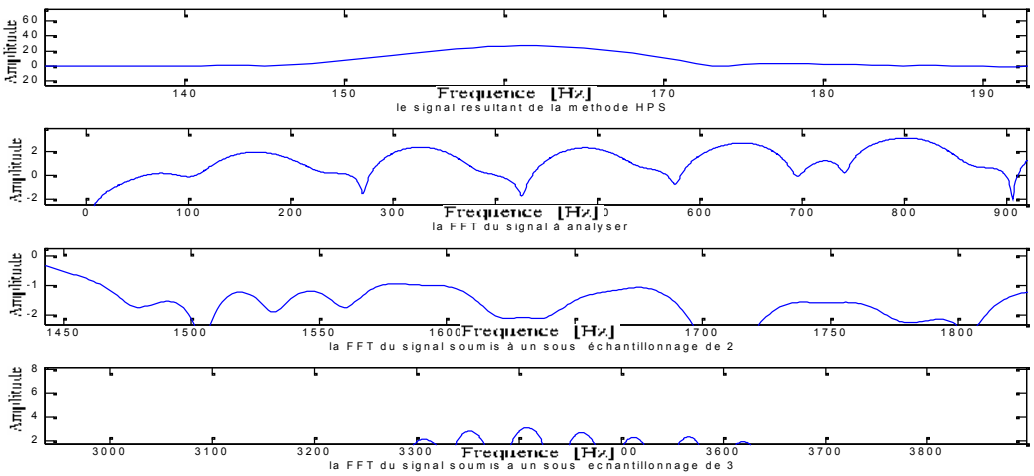


Figure IV.11 HPS logarithmique pour une trame d'un signal du phonème « a » F0=164.1907 Hz

- La première remarque c'est qu'on arrive aux mêmes résultats.

IV.3.5 Conclusion :

Quelques traits agréables de cette méthode incluent: c'est que coté de calcul l'algorithme ne nécessite pas un temps grand, raisonnablement résistant au bruit additif et bruit multiplicatif, et réglable à genre différent d'entrées. Par exemple, nous pourrions changer le

nombre de spectres compressé (sous échantillonnée) pour utiliser et nous pourrions remplacer la multiplication spectrale par une addition spectrale. Cependant, depuis que la perception du pitch humaine est fondamentalement logarithmique, cela signifie que les bas pitch peuvent être détectés correctement que hautes fréquences.

IV.4 Addition sous - harmonique (Sub-harmonic Summation (SHS)):

Un algorithme qui n'à aucun problème avec harmoniques manquant est l'addition Sous harmonique (SHS) (Hermès, 1988), lequel résout le problème (de la méthode HPS) en utilisant l'addition au lieu de multiplication. Par conséquent, si tout harmonique manque, il ne contribuera pas au total. Dans les termes mathématiques, SHS estime le pitch comme suit **[25]** :

$$F0 = \arg\max_{f} \sum_{k=1}^{n} \left| X(kf) \right| \qquad \text{IV.21}$$

IV.4.1 Problème de la méthode :

Une trappe de cet algorithme est que depuis qu'il donne le même poids à tout l'harmonique, *addition sous - harmonique* du pitch peuvent avoir le même score comme le pitch et par conséquent ils sont valides pour être reconnu comme le pitch.

IV.4.2 Résolution du problème de la méthode :

SHS peut remédier à ce problème en pesant l'harmonique avec une progression géométrique comme [25] :

$$F0 = \arg\max_{f} \int_{0}^{\infty} \left| X(f') \right| \sum_{k=1}^{n} r^{k-1}\, \delta(f' - kf)\, df' \qquad \text{IV.22}$$

Où la valeur de 'r' a été mise empiriquement à 0.84 [25] basé sur expériences qui utilisent la parole.

IV.4.3 Résultats expérimentaux :

Dans cette étude, on prend un signal sans bruit (Phonème « a ») en présence de bruit (bruit blanc), les résultats sont résumés dans le tableau IV.3. La Figure IV.12 présente un signal « a » avec son contour de pitch par SHS.

Le pitch [Hz]	SNR [dB]
129.7059Hz	Sans bruit
129.7059Hz	54 dB
129.7059Hz	30 dB
137.1630Hz	1 dB
131.8500Hz	-5 dB

Tableau IV.3 Pitch par SHS en fonction de SNR.

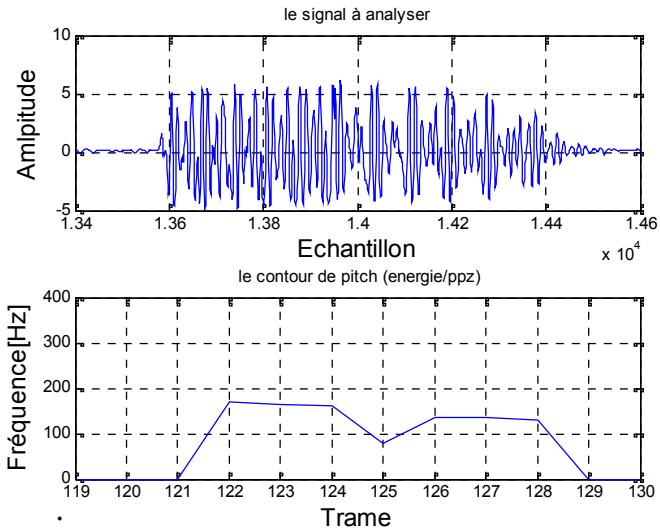


Figure IV.12 Un signal « a » avec le contour de pitch par SHS.

IV.4.4 Conclusion :

Cette méthode remédie le problème de la méthode HPS, mais tomber dans un autre problème (le même poids à tout l'harmonique). Les résultats d'analyse de cette méthode semblable aux résultats obtenus par HPS (résistant au bruit).

IV.5 Conclusion :

- L'objectif essentiel d'utiliser la méthode cepstrale c'est qu'elle permet de séparer facilement la source et le filtre. C'est la propriété fondamentale mais n'est pas résistante au bruit et peut donner des résultats fausses.
- Les deux méthodes HPS et SHS donnent presque toujours les mêmes résultats mais avec une fiabilité de la méthode SHS vis-à-vis le bruit, et un bon contour de pitch par HPS.

V.1 Introduction :

La transformation de Fourier, occupe une place privilégiée dans la théorie et le traitement des signaux. Néanmoins, la nature même de cette transformation ne peut donner d'un signal temporel qu'une information fréquentielle globale de laquelle toute notion de chronologie a disparu : son champ d'application naturel est l'analyse des signaux stationnaires.

Ainsi, dès lors que l'on considère des signaux modulés ou des processus non stationnaires, une analyse spectrale classique fournit une information moyennée sur la durée d'observation. Une solution partielle à ce problème, la plus intuitive, consiste à associer à un signal non stationnaire une suite de *transformées de Fourier à court terme*. Une deuxième solution, plus satisfaisante, consiste à chercher directement un outil adapté à l'étude de phénomènes non stationnaires c'est *la représentation temps - fréquence.*

Dans cet esprit, on explore quelques outils d'analyse des signaux non stationnaires appartenant au plan temps-fréquence, où on s'intéresse au signal vocal point de vue représentation et estimation de la fréquence fondamentale. On cite les plus importantes distributions de *Spectrogramme, Wigner-Ville,Pseudo Wigner-Ville,Distribution de Rihazek-Hill, Pseudo Rihazek-Hill,……ect.*

V.2 La représentation temps -fréquence :

Dans cette représentation, on fait intervenir sur un même graphique la composition en fréquence du signal (ordonnée), sa durée (abscisse), et sa densité d'énergie (code couleur).On obtient ainsi des courbes (des chirps par exemple : morceaux d'hyperboles, indiquant un glissement de fréquence continu au cours du temps comme avec l'effet Doppler) ou des zones en deux dimensions mettant ainsi en relation l'évolution de la composition en fréquence du signal (Fourier) au cours du temps.

Sur la base des limites de la transformation de Fourier, se sont développées les techniques d'analyse temps-fréquence qui peuvent être vues comme une transformée de Fourier locale.

V.2.1 Propriétés d'une représentation temps-fréquence idéale

Avant de parcourir les principales représentations en temps et en fréquence d'un signal monodimensionnel, il faut connaitre les propriétés que doit posséder la représentation temps fréquence : **(a)** orthogonale, **(b)** locale, **(c)** adaptative, **(d)** lisible (interprétation physique).
La propriété **(a)** assure l'unicité de la décomposition. Ainsi, l'analyse est réversible dans le sens où le signal initial peut être reconstruit simplement à partir de la projection.

La propriété **(b)** permet d'observer des états qui sont brefs, voire transitoires. Le fait d'être adaptative **(c)** peut être intéressant dans le cas où le signal possède une grande richesse dynamique et balaye un très large spectre de fréquences au cours du temps. Dans ce cas, il est nécessaire que l'analyse prenne en compte les évolutions du signal en fonction du temps.
Enfin, la propriété **(d)**, qui est plus un critère « de confort », se révèle cependant extrêmement intéressante lors de l'analyse. En effet, l'interprétation et la compréhension des phénomènes physiques sollicites pour élaborer le signal peuvent en être grandement facilitées.
Parmi toutes les méthodes susceptibles de répondre a notre attente, seules les plus classiques vont retenir notre attention. Les représentations que nous allons développer sont les *Représentations temps-fréquence*

V.3 Classification des méthodes temps-fréquences :

On peut classer les méthodes temps–fréquence en deux groupes « paramétrique » et « non paramétrique » : Au-delà des mots, l'idée générale est en fait de disposer de méthodes, soit inspirées de l'analyse de Fourier dans le premier cas, soit reposant sur une approche différente (par exemple à l'aide de modèles utilisant une information apriori quant à la structure possible des signaux analysés) dans le deuxième cas. Il est important d'éclairer que dans notre travaille on s'intéresse aux méthodes *non paramétriques.*

V.3.1 Méthodes non paramétriques :

Toutes les représentations de ce groupe sont des distributions d'énergie du signal analysé. Ce sont également toutes des formes bilinéaires du signal.
Une première formulation unifiée est fournie par la classe de *Cohen* associée aux distributions bilinéaires covariantes par translations temporelles et fréquentielles : dans ce groupe on trouve ; Spectrogramme, Wigner-Ville, pseudo Wigner-Ville lissé, Choï-Williams, Zhao-Atlas-Marks, scalogramme (ondelettes continues), Wigner-Ville lissé, temps-fréquence instantanée…..

V.3.2 Méthodes paramétriques :

Dans ce groupe on trouve ; Lagunas glissant, AR glissant, AR à mémoire variable, méthodes hybrides, AR évolutif, Prony adaptatif, ARMA par Filtrage de Kalman étendu.

V.4 Représentation temporelle et fréquentielle :

Faire introduire la conception de la localisation en temps et en fréquence nous mène à poncer au produit du bonde passante de temps et son associe en fréquence c'est-à-dire introduire l'inégalité de *Heisenberg-Gabor* (équation V.5), par suite introduire la notion de la fréquence instantanée et le retard de groupe, ces deux notions sont la solution de problème de la localisation en temps de spectre.

V.4.1 Localisation et le principe de *Heisenberg -Gabor :*

La moyenne la plus simple de caractériser un signal simultané en temps et fréquence est de considérer la valeur moyenne de la localisation et dispersion pour chacune de ces deux représentations, ce la peut être acheminé en considérant $|x(t)|^2 et |X(f)|^2$ (le module carré de transformée de Fourrier de x(t)) comme des probabilités, et voire les valeurs moyennes et standard de déviation [29]:

$$t_m = \frac{1}{Ex}\int_{-\infty}^{+\infty} t|x(t)|^2 dt \qquad \text{moyenne temporelle} \qquad \text{V.1}$$

$$f_m = \frac{1}{Ex}\int_{-\infty}^{+\infty} f|X(f)|^2 df \qquad \text{moyenne fréquencielle} \qquad \text{V.2}$$

$$T_2 = \frac{4\pi}{Ex}\int_{-\infty}^{+\infty} (t-tm)^2|x(t)|^2 dt \qquad \text{dispersion sur laxe des temps.} \qquad \text{V.3}$$

$$B_2 = \frac{4\pi}{Ex}\int_{-\infty}^{+\infty} (f-fm)^2|X(f)|^2 dt \qquad \text{dispersion sur laxe des fréquences.} \qquad \text{V.4}$$

Où : Ex est l'énergie moyenne de signal x (t).

Le signal est caractérisé dans le plan temps-fréquence par la position moyenne (t_m,f_m) et la localisation d'énergie par [29]:

$$T.B \geq 1 \qquad \text{V.5}$$

Cette dernière relation est connue sous le nom de l'inégalité de *Heisenberg-Gabor*.

- « B » est la largeur de bande fréquentiel d'un signal,
- « T » est la durée du signal

V.4.2 la fréquence instantané et signal analytique :

Une autre méthode (localisation en fréquence d'une composante temporelle) de décrire un signal dans le plan temps-fréquence est de considérer la « *fréquence instantanée* », la figure V.1 représente un exemple d'une estimation de la fréquence instantanée pour un signal « Chirp », tout d'abord on débute par définir un *signal analytique* :

Pour toutes les valeurs réelles d'un signal x(t), nous associons un signal xa(t) de valeur complexe, on définit un signal analytique par [29] :

$$x_a(t) = x(t) + jHT((x(t)) \qquad \text{V.6}$$

Où HT((x(t)) c'est la *transformée d'HILBERT. $X_a(t)$ est la transformée d'Hilbert de x(t).*

$$X(f)=0 \quad si \quad f<0. \qquad \textit{(les fréquences négatives sont annulées)} \qquad \text{V.7}$$

$$X_a(f) = X(0) \quad if \quad f = 0 \qquad \text{V.8}$$

$$X_a(f) = 2X(f) \quad if \quad f > 0 \qquad \text{V.9}$$

C'est possible maintenant de définir le concept de l'amplitude instantanée et la fréque instantanée par [29]:

$$a(t) = |X(t)| \qquad \text{V.10}$$

$$f(t) = \frac{1}{2\pi}\frac{d\arg x_a(t)}{dt} \qquad \text{V.11}$$

(Équation : V.10 amplitude instantanée, équation : V.11 fréquence instantanée)

V.4.3 Le retard de groupe :

Le retard de groupe caractérise la localisation en temps d'une composante fréquentielle, le retard de groupe est définit par :

$$t_x(f) = -\frac{1}{2\pi}\frac{d\arg X_a(f)}{df} \qquad \text{V.12}$$

Figure V.2 représente un exemple d'une estimation de retard de groupe pour un signal 'Chirp'

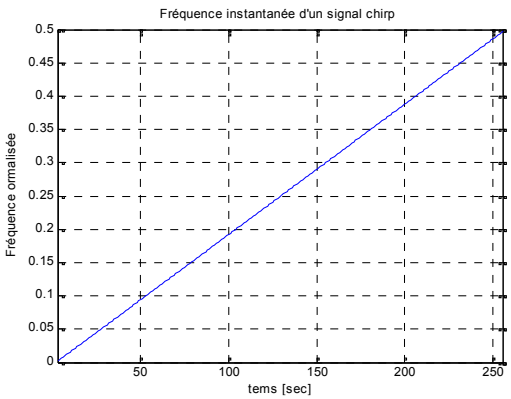


Figure V.1 Exemple d'une estimation de la fréquence instantanée pour un signal 'Chirp'

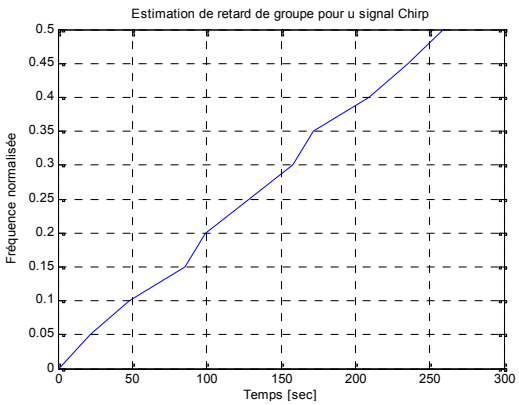


Figure V.2 Exemple d'une estimation de retard de groupe pour un signal 'Chirp'

V.5 Analyse temps-fréquence :

Il est aujourd'hui bien admis que les représentations d'un signal conjointement en temps et en fréquence offrent un réel intérêt : elles permettent une description des signaux non stationnaires, c'est à dire l'analyse des lois de comportement fréquentiel du signal au cours du temps. Les RTF situent l'énergie dans le plan temps-fréquence, parmi les RTF, la classe de

Cohen qui tient une place particulière : *peut être construite de manière objective et* elle *contient l'ensemble des représentations bilinéaires covariantes par translations dans le plan temps-fréquence.*

V.5.1 Spectrogrammes :

L'examen visuel de la distribution d'énergie dans le plan temps-fréquence est un outil fondamental pour l'analyse acoustique descriptive de la parole. Le prototype de la représentation énergétique est le spectrogramme, historiquement c'est la première représentation temps - fréquence utilisée pour l'analyse visuelle de la parole [29]:

$$S_x(t, f) = \left| \int_{-\infty}^{+\infty} x(\tau) w(t-\tau) e^{-2i\pi f\tau} d\tau \right|^2 \qquad \text{V.13}$$

La fenêtre '*w*' fixe la résolution spectrale et temporelle d'analyse. En général les spectrogrammes sont calculés par transformée de Fourier rapide, en utilisant 256 à 1024 coefficients et 1 à 2 ms [26] entre chaque analyse. Il est bien connu que les spectrogrammes ne peuvent pas offrir une grande résolution simultanément en temps et en fréquence. De même que pour les autres distributions énergétiques, des termes d'interférence entre composantes spectrales ont été mis en évidence par Jeong & Williams [27].

V.5.1.1 Exemples pour une Application de spectrogramme avec un signal parole :

On applique l'algorithme de STFT pour représenter un signal parole, on a considéré un signal parole contenant un mot « GABOR » de 338 points et Fs=1KHz. De cette représentation on ne peut rien dire sur la localisation de pique de 140 Hz pour le mot GABOR. Figure V.3 exprime la représentation fréquentielle de mot « GABOR ».

D'autre part on utilise une fenêtre Hamming de 85 points. Dans ce plan temps-fréquence on résulte une partie se situe entre 30sec et 60sec et centré autour de 150Hz, correspond au « GA ». La deuxième partie se situe entre 150sec et 250sec, correspond au « BOR ». Figure V.4 exprime le module carré de module STFT de mot GABOR, nous pouvons voir que la fréquence moyenne se baisse de 140Hz à 110Hz en fonction de temps.

V.5.1.2 Résolution temps fréquence:

Une bonne résolution en temps nécessite un court duré de la fenêtre d'analyse, d'autre part une bonne résolution en fréquence nécessite une bonde étroite c'est-à-dire une longue durée de *w*(t) (la fenêtre d'analyse de STFT). Mais malheureusement on ne peut pas garantir une bonne résolution en temps et en fréquence simultanément. Donc on doit s'assurer d'un

compromis de résolution en temps et fréquence. Cette limitation résulte l'inégalité de *Heisenberg-Gabor*.

a. une parfaite résolution en temps :

Cela implique une fenêtre *w*(t) comme une impulsion de Dirac.

$$w(t) = \delta(t) \Rightarrow F_x(t, f; h) = x(t)\exp[-i2\pi ft] \qquad \text{V.14}$$

Le STFT a une parfaite résolution en temps mais n'a aucune résolution en fréquence.

Prenant un signal ''x'' avec une modulation en fréquence linaire et une modulation d'amplitude gaussienne ainsi une fenêtre d'analyse exprime une impulsion de Dirac. La figure V.5 représente le résultat de la RTF du signal x(t) en utilisant une impulsion de Dirac.

Le signal est parfaitement localisé en temps mais nulle en fréquence

b. une parfaite résolution en fréquence :

Prenant le même signal ''x'' mais avec une fenêtre d'analyse *w*(t) constante, avec cette résolution le STFT se réduit à une transformé de Fourier simple avec aucune résolution en temps mais résolution parfaite en fréquence. La figure V.6 représente la RTF du signal x(t) avec une fenêtre d'analyse constante.

V.5.1.3 Influence de la forme et la longueur de la fenêtre d'analyse:

Pour faire apparaître l'influence de la forme et la nature de la fenêtre d'analyse on propose l'exemple suivant : considérant deux signaux transitoire ont les mêmes amplitudes gaussienne de fréquence constante (voire figure V.7). La figure V.8 exprime les résultats obtenus par une fenêtre Hmming de 65 points.

La première remarque qu'on peut extraire c'est qu'il est difficile de discriminer les deux composantes de signal, la résolution en fréquence est bonne mais une mauvaise résolution en temps. La figure V.9 exprime le résulta obtenue par une fenêtre Hmming de 17 points. On remarque qu'on a une bonne résolution en temps (nous permettre de discriminer les deux composons de signal) mais avec une mauvaise résolution en fréquence

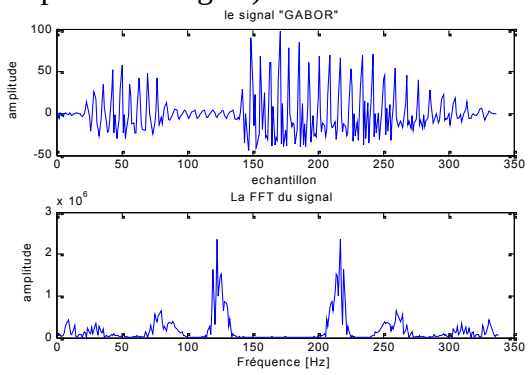


Figure V.3 Représentation fréquentielle de mot « GABOR »

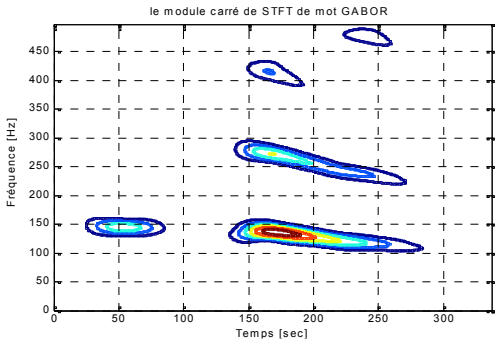


Figure V.4 Le module carré de module STFT de mot GABOR

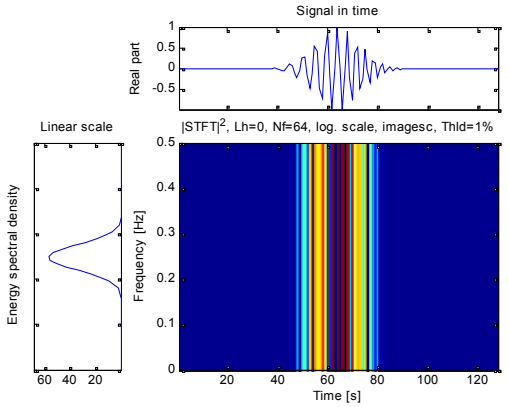


Figure V.5 Une résolution parfaite avec STFT en temps avec impulsion de Dirac.

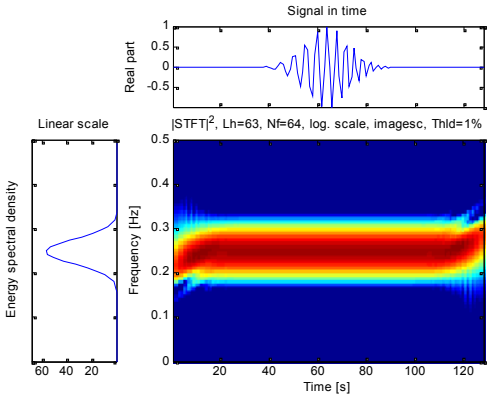


Figure V.6 Une résolution parfaite en fréquence.

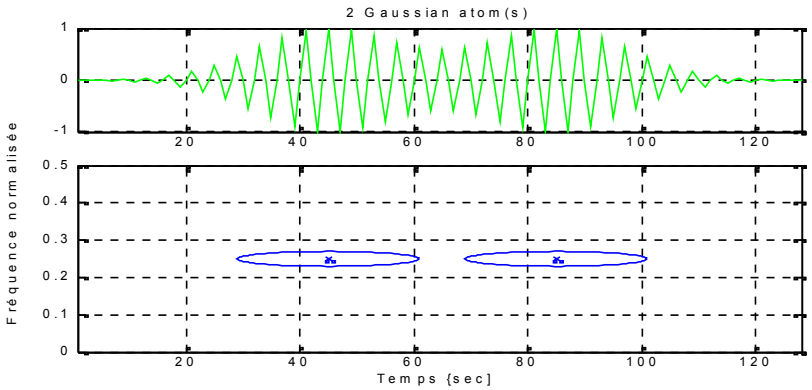


Figure V.7 Le signal comprenant deux atomes avec amplitude gaussienne

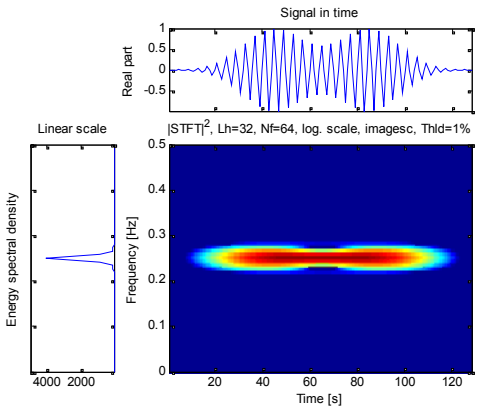


Figure V.8 Les deux atomes gaussienne avec Hamming de 65 points.

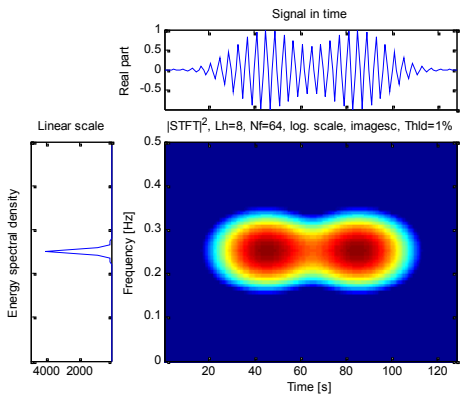


Figure V.9 Les deux atomes gaussienne avec Hamming de 17 points.

V.5.1.4 Conclusion :

À partir des deux expériences on peut conclure des avantages et des inconvénients :

Avantages : - positif,

- extension directe de Fourier, interprétation identique en fréquences
- pas de termes d'interférences

Inconvénients:- principe d'incertitude (compromis en fréquence et en temps).

- la résolution et les lois en fréquence sont en fonction de la fenêtre (Fig V.8,V.9).
- l'optimisation des fenêtres nécessite des informations a priori sur le signal.

V.5.2 Distribution de Wigner-Ville :

La distribution de Wigner-Ville, appartient à la classe de Cohen, présentée à la même période que le spectrogramme par Ville [28], Elle est définie par [26]:

$$W_x(t, f) = \int x^*(t - \lambda/2)\, x(t + \lambda/2)\, e^{-2i\pi f\lambda}\, d\lambda \qquad \text{V.15}$$

L'argument en faveur de cette représentation est sa résolution simultanée en temps et en fréquence, qui ne connaît pas les limitations du spectrogramme. De plus, la distribution de Wigner-Ville possède des propriétés théoriques qui la placent au cœur de la problématique des représentations non paramétriques bilinéaires, puisqu'il suffit qu'une représentation bilinéaire soit invariante par translations temporelles et fréquentielles pour qu'elle s'exprime en fonction de la représentation de Wigner-Ville et d'un noyau spectro-temporel K.

$$C_x(t, f) = \int \int W_x(\lambda, \alpha)\, K(\lambda - t, \alpha - f)\, d\alpha\, d\lambda \qquad \text{V.16}$$

Toutes les distributions citées ici, le spectrogramme en particulier, peuvent s'exprime sous cette forme. On peut la considérer comme une distribution d'énergie dans le plan-temps fréquence, puisque la sommation de la distribution en temps permet de retrouver la densité d'énergie fréquentielle. Cependant, cette interprétation énergétique est discutable à cause de l'existence des *termes d'interférences* importants entre les différentes composantes fréquentielles avec la RTF ainsi la présence de valeurs négatives locales sur la distribution. Ces propriétés peu souhaitables et peuvent être atténuées par l'utilisation du signal analytique à la place du signal réel, ainsi par un lissage.

La DWV est caractérisé par des valeurs réel, conservation temporelle et fréquentielle et satisfaire les propriétés marginale.

L'expression V.15 est la transformée de Fourier d'une forme acceptable de la fonction caractéristique pour la distribution d'énergie.

V.5.2.1 Problèmes d'interférences:

Pour comprendre le problème d'interférences de la DWV on observe les deux exemples suivants :

Exemple1 : Soit le signal synthétique exprimant un signal linéaire « Chirp » représenté par la figure V.10 et faisons la DWV, le résultat est représenté par la figure V.11.

La figure V.12 nous montre la distribution de WV en trois dimensions où les valeurs négatives sont prises en compte, le plan temps -fréquence pour ce signal est assez parfait

Exemple2 : si une voiture de vitesse constate passe devant un observateur, le son de moteur entendu par cet observateur se change au cours de temps : la fréquence diminue d'une valeur à une autre .Ce phénomène s'appel « L'effet DOPPLER », cela exprime la dépendance de la fréquence reçue par l'observateur d'un transmetteur pour la vitesse relative entre l'observateur et le transmetteur. La figure V.13 représente ce signal (effet DOPPLER) [29].

Discussion et problèmes d'interférences:

La Figure V.14 représente La DWV de signal DOPPLER où nous montre beaucoup d'interférences dû au bilinéarité du signal (signal Doppler),dans la distribution temps - fréquence, on note que l'énergie n'est pas distribuée convenablement mais d'une façon générale le signal est bien localisé dans le plan temps-fréquence, beaucoup termes sont présent (termes dû aux interférences) où l'énergie doit être nulle. Dans la suite on voit comment remédier aux problèmes d'interférences.

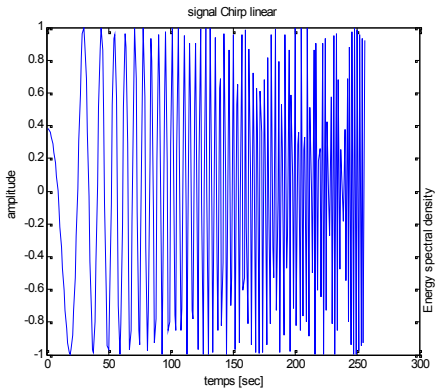


Figure V.10 Signal Chirp.

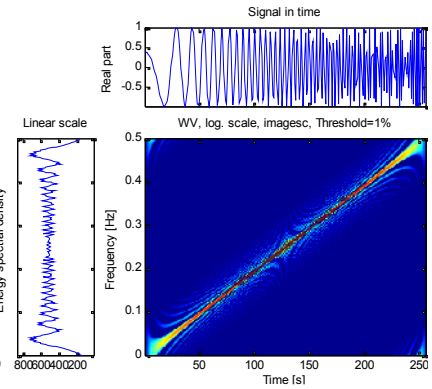


Figure V.11 Signal Chirp et son DWV.

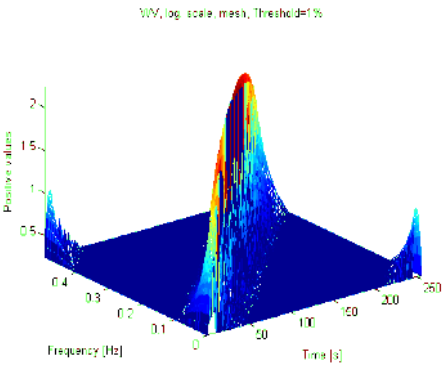


Figure V.12 La DWV de signal Chirp en trois dimensions.

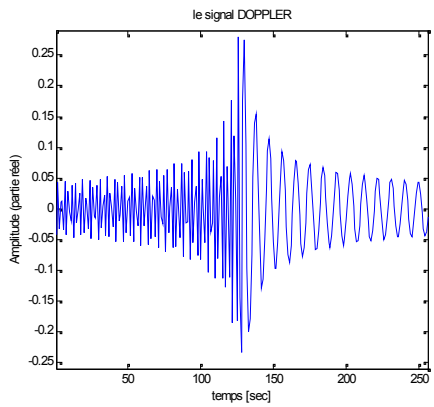


Figure V.13 Le signal DOPPLER

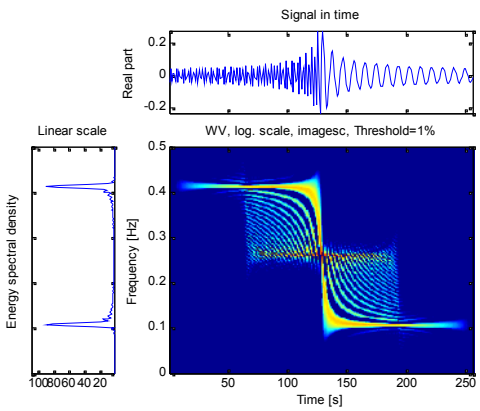


Figure.14 DWV de signal Doppler

Interprétation mathématique des interférences :

Si on a deux signaux x1(t) et x2(t), on construit un signal x(t) qui exprime la sommes des deux signaux, alors le WV est définit par :

$$Wx(t, f) = Wx1(t, f) + Wx2(t, f) + Wx1,x2(t, f) + Wx2,x1(t, f)$$

- ✓ Très gênant pour l'interprétation (explique la non-positivité)
- ✓ *N* composantes donne une représentation temps-fréquence de *N* termes + $N(N-1)/2$ termes d'interférences .
- ✓ Où l'idée est de lisser Wigner-Wille (par convolution) pour réduire les interférences.

V.5.2.2 Propriétés :

La DWV possède un grand nombre des propriétés souhaitables pour une représentation temps-fréquence. En particulier :

1. La conservation d'énergie : la DWV est une fonction réelle répartissant l'énergie d'un signal dans le plan temps-fréquence.

$$E_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t, f)\,dtdf = \int_{-\infty}^{+\infty} |x(t)|^2\,dt \qquad \text{V.17}$$

2. Ses distributions marginales s'identifient à la puissance instantanée et à la densité spectrale du signal : où X (*f*) est la transformée de Fourier de x (t)

$$\int_{-\infty}^{+\infty} W_x(t, f)\,dt = |X(f)|^2 \qquad \text{V.18}$$

$$\int_{-\infty}^{+\infty} W_x(t, f)\,df = |X(t)|^2 \qquad \text{V.19}$$

3. la DWV conserve les supports temporels et fréquentiels du signal :

$$|x(t)| = 0, |t| > T \Rightarrow W_x(t, f) = 0, |t| > T. \qquad \text{V.20}$$

$$|x(f)| = 0, |f| > B \Rightarrow W_x(t, f) = 0, |f| > B. \qquad \text{V.21}$$

4. Translation :

$$y(t) = x(t - t0) \Rightarrow W_x(t, f) = W_x(t - t0, f) \qquad \text{V.22}$$

$$y(t) = x(t)e^{j2\pi 2_0 t} \Rightarrow W_x(t, f) = W_x(t, f - f0) \qquad \text{V.23}$$

5. Dilatation, la DWV conserve la dilatation:

$$y(t) = \sqrt{k}\,x(kt), k > 0 \Rightarrow W_y(t, f) = W_x(kt, \frac{f}{k}) \qquad \text{V.24}$$

6. Compatibilité avec le filtrage linéaire :

$$y(t) = \int_{-\infty}^{+\infty} h(t - s)x(s)\,ds \Rightarrow W_y(t, f) = \int_{-\infty}^{+\infty} W_h(t - s, f)W_x(s, f)\,ds \qquad \text{V25}$$

7. Conservation de produit scalaire, la DWV conserve le produit scalaire de domaine temporel au domaine fréquentiel :

$$\left|\int_{-\infty}^{+\infty} x(t)y^*(t)\,dt\right|^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t, f)W_y^*(t, f)\,dtdf \qquad \text{V.26}$$

Cette dernière est connue sous le nom de la formule de « Moyal ».

8. les caractéristiques de modulation d'un signal réel (enveloppe, fréquence instantanée, retard de groupe): sont définies de manière univoque par l'intermédiaire du signal analytique associé

V.5.2.3 Conclusion :

On peut conclure des avantages et des inconvénients :

Inconvénients: - valeurs négatives locales,

- interférences entre les composantes des signaux ; *ceci dû au produit* $x^*(t-\lambda/2)x(t+\lambda/2)$.

Avantages : -meilleure résolution en temps et fréquence puisque on utilise tout le signal.

- propriétés marginales, du retard de groupe.
- estimation des modulations de fréquences.

V.5.3. Pseudo-DWV :

L'idée générale de cette nouvelle méthode est la réduction des interférences de la distribution de Wigner-Ville par lissage dans le plan temps-fréquence. La Figure V.16 représente la DWV du signal composé de quatre atomes où les interférences révèlent plus claire entre les composants de signal.

V.5.3.1 Pseudo Wigner Ville :

C'est le lissage en fréquence :

$$W_{PWV}(t,f) = \int_{-\infty}^{+\infty} h(\lambda).x^*(t-\lambda/2)x(t+\lambda/2)e^{-2i\pi f\lambda}d\lambda \qquad \text{V.27}$$

La figure V.17 exprime le signal gaussien avec quatre atomes après l'application de PWV.

Le PWV se caractérise par:- moins bonne résolution en fréquence version lissée en fréquence.

- on utilise une portion du signal.

V.5.3.2 Pseudo Wigner Ville Lissé (PWVL) :

C'est le lissage en temps et fréquence : Une forme de lissage particulièrement intéressante met en œuvre un noyau séparable en temps et en fréquence, comme produit d'une fenêtre temporelle *w* et d'une fenêtre spectrale *W* :

$$W_{PWVL}(t,f) = \int_{-\infty}^{+\infty} h(\lambda)\int_{-\infty}^{+\infty} g(u-t).x^*(u-\lambda/2)x(u+\lambda/2)e^{-2i\pi f\lambda}dud\lambda \qquad \text{V.28}$$

La figure V.18 représente la PWVL du signal composé de quatre atomes.

Le PWVL se caractérise par : - convolution bidimensionnelle.

- réduction des interférences
- positivité et perte de résolution temps-fréquence.

Exemple : Soit un signal composé de quatre (4) composants (quatre atomes gaussienne : figure V.15). On représente la distribution énergétique pour WV, PWV et PWVL

- la représentation de WV nous montre les quatre atomes avec présence de six termes d'interférences (deux sont superposés : Figure V.16).
- Si on calcul PWV et faisons la représentation temps-fréquence on constate l'atténuation des interférences clairement au axe de fréquence. Figure (V.17).
- Si on considère la représentation par PWVL on remarque l'atténuation des interférences sur les deux axes temporel et fréquentiel. Figure (V.18).

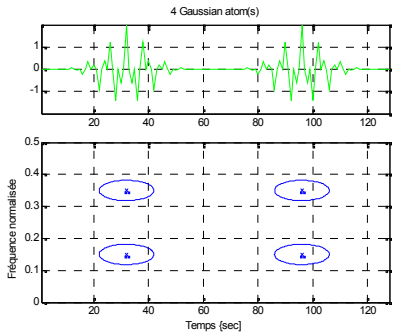


Figure V.15 Le signal gaussien avec ses quatre atomes

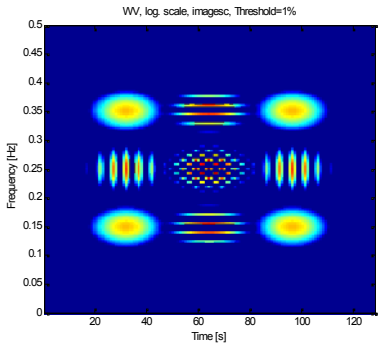


Figure V.16 DWV du signal composé de quatre atomes

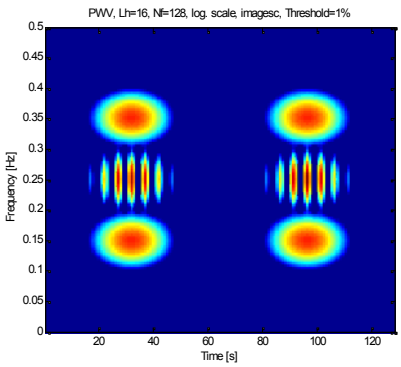


Figure V.17 Le signal gaussien avec ses quatre atomes (PWV)

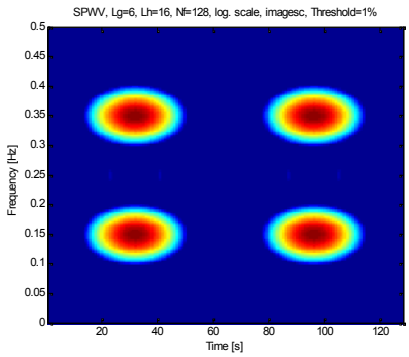


Figure V.18 PWVL du signal composé de quatre atomes.

V.5.4 Distribution de Rihaczek et Margenau-Hill :

Cette distribution correspond au celle de Cohen , si on considère l'interaction d'énergie entre un signal x réduit à un intervalle infiniment petit $\boldsymbol{\delta_T}$ centré à t et un signal x aussi passant par une bande passante d'un filtre $\boldsymbol{\delta_B}$ centré à f , on peut approximer ça par:

$$\delta_T \delta_B [x(t).X^*(f).e^{-j2\pi ft}] \qquad \text{V.29}$$

Alors on résulte la distribution de Margenau-Hill-Rhiaczek appelée «densité d'énergie complexe» par Rhiaczek :

$$R_x(t, f) = x(t).X^*(f).e^{-j2\pi ft} \qquad \text{V.30}$$

Cette distribution n'est pas forcément positive pour toutes les valeurs du temps et de la fréquence. Les formants sont difficiles à extraire, surtout lorsque les fréquences centrales évoluent. Quand il y a plus d'une composante fréquentielle, des termes d'interactions apparaissent (Figure V.17). Cette méthode présente plusieurs propriétés. Les interférences pour les représentations de Rihaczek et Margenau-Hill différentes des interférences de WV.
les interférences qui correspondent au deux points se situent à (t,f1) et (t2,f2) sont positionnées à (t1,f2) et (t2,f1) respectivement . La Figure V.19 représente la distribution *Margenau-Hill des* deux atomes positionnés à (t1,f1) et (t2,f2).

V.5.4.1 Pseudo Rihaczek et Margenau-Hill :

On peut aussi définir la version de Rihaczek et Margenau-Hill lissé comme celle de WV (Figure V.20). Si on considère un signal x à deux atomes, on veut faire le calcul de Margenau-Hill, et pseudo Margenau-Hill. La figure V.20 exprime le résultat de la distribution de psudo Margenau-Hill pour deux atomes où on remarque la disparaissions des interférences vues à la figure V.19.

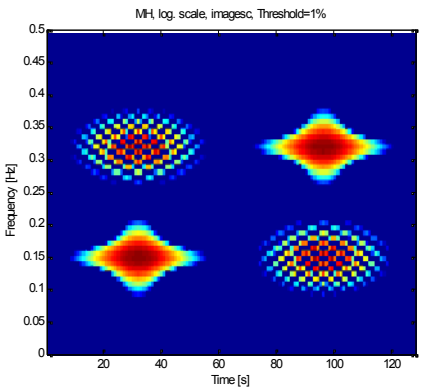


Figure V.19 La distribution de *Margenau-Hill*

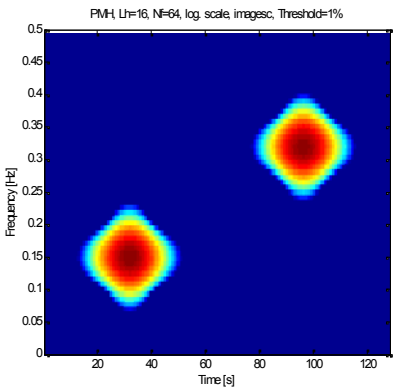


Figure V.20 La distribution de *pseudo Margenau-Hill*

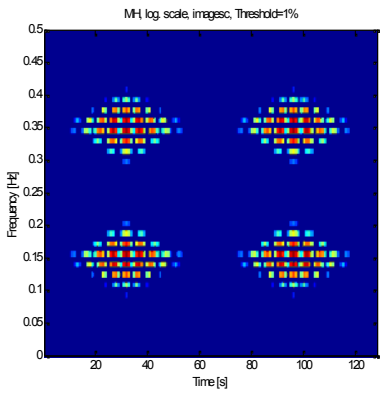


Figure V.21 Distribution de Margenau -Hill pour un signal à quatre atomes

V.5.4.2 Conclusion:

L'utilisation de la distribution de Rihaczek (ou Margenau-Hill) pour un signal multi composants se situent aux mêmes positions en temps ou en fréquence n'est pas recommandé, en raison de la superposition de ces termes sur le signal utile. La distributi Rihacz - Margenau-Hill est semblable à la distribution de WV. La Figure V.21 représente la

distribution de Marg-enau Hill pour un signal à quatre atomes où les interférences superposent sur les atomes.

V.5.5 Distribution de Page :

La distribution de Page, ou « spectre instantané de puissance », est définie par [29]:

$$P_x(t,f) = \frac{d}{dt}\left(\left| \int_{-\infty}^{t} x(u).e^{-j2.\pi.f.u}.du \right|^2 \right). \qquad \text{V.31}$$

$$= 2\,\Re\left(x(t)\left(\int_{-\infty}^{t} x(u).e^{-j2\pi.f.u}\,du \right)^{*} e^{-j2.\pi.f.t} \right)$$

C'est la dérivée de la densité spectrale de puissance de signal considéré avant l'instant 't', cette méthode est similaire aux méthodes précédentes Rihaczek et Margenau-Hill concernant les caractéristiques (interférences, positivité, format …). La figure V.22 nous illustre le résultat de la distribution de Page pour un signal à deux composants.

V.5.5.1 Distribution de pseudo Page :

À cause de présence des interférences (Figure 22), cette distribution (Page) fournit un autre outil de lissage nommé *Pseudo page* (comme celle de WV). La Figure V.23 présente le résultat de la distribution de pseudo page pour un signal à deux composants.

.

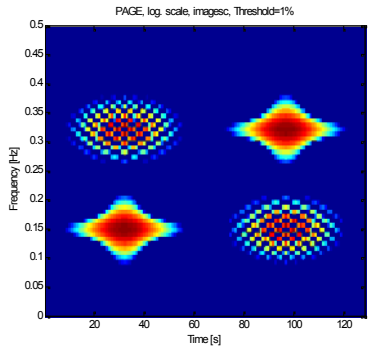


Figure V.22 Distribution de *Page* pour un signal à deux atomes (composants).

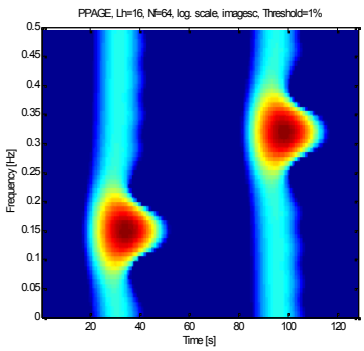


Figure V.23 Distribution de *Pseudo Page* pour un signal de deux atomes

V.5.5.2 conclusion :

La méthode de page c'est la dérivée de la densité spectrale de puissance, cette méthode est similaire au *Margenau-Hill.*

V.5.6 Distribution de Choi-Williams *:*

L'idée générale est la réduction des termes d'interférences de la distribution de Wigner-Ville tout en préservant les distributions marginales temporelles et fréquentielle.

V.5.6.1 Principe de la méthode :

Utiliser une fonction de pondération (dans la classe de Cohen) qui dépende de ses variables via leur produit. Choix effectif : une gaussienne paramétrée par une ‘’variance σ’’, définition à temps continu. La méthode Choi-Williams est définit par [29]:

$$CW_x(t, f) = \sqrt{\frac{2}{\pi}} \int_{-\infty}^{+\infty} \int \frac{\sigma}{|\lambda|} e^{-2\sigma^2 (u-t)^2 / \lambda^2} x\left(u + \frac{\lambda}{2}\right). x\left(u - \frac{\lambda}{2}\right) e^{-j2.\pi . f . \lambda} du d\lambda \quad \text{V.32}$$

V.5.6.2 Quelques caractéristiques de la méthode :

On peut citer les caractéristiques suivantes :

- On obtient WV lorsque le paramètre σ tend vers l’infini.
- Constatation empirique [29] : choisir $1 < \sigma < 80$.
- La méthode de Choi-Williams ne garanti pas la conservation des supports temporels et fréquentiels.
- Réduction de l'amplitude des interférences mais délocalisation.
- Vérifier les caractéristiques de conservation d’énergie, propriétés marginales, valeur réel, translation et dilatation, la fréquence instantanée d’un signal « x » peut être extraite à partir de Choi-Williams comme moment d’ordre 1 en fréquence, le retard de groupe peut être extraite à partir de Choi-Williams comme moment d’ordre 1 en temps.

V.5.6.3 Problèmes d’interférences et la méthode Choi-Williams(CW) :

Pour évaluer la méthode vis-à-vis le problème d’interférence on considère un signal de quatre atomes de différents position l’un par rapport à l’autre. La figure V.24 représente ce signal par la méthode de CW dans le plan temps-fréquence. À partir des résultats obtenus (Figure V.24) il est aisément de conclure que si le signal a différents composants synchronisés en temps ou en fréquence la distribution de Choi-Williams présente des interférences.

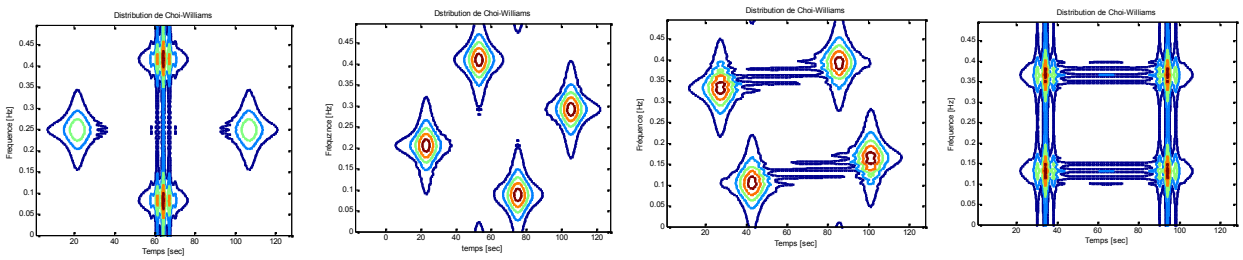


Figure V.24 Distribution de Choi-Williams de quatre atomes à différentes positions (de gauche à droite) où on remarque les interférences dépend de position des atomes.

V.5.6.4 Conclusion :

L’efficacité de cette méthode de distribution dépend de la nature de signal à analyser.

V.5.7 Relation avec la fonction d'Ambigüité (AF):

Une fonction intéressante et particulière spécialement dans le domaine de traitement des signaux radar, est la fonction d'Ambigüité à bande étroite (AF) définit par [31] :

$$A_x(\zeta,\tau) = \int_{-\infty}^{+\infty} x(t+\tau/2)\,x^*(t-\tau/2)\,e^{-2i\pi\zeta t}\,dt \qquad \text{V.33}$$

C'est la mesure de la similarité d'un signal x et son translation dans le plan temps-fréquence, ainsi il y a une relation entre la fonction de WV et AF tel que la TF de WV résulte l'AF [31] :

$$A_x(\zeta,\tau) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W_x(t,f)\,e^{2i\pi(f\tau-\zeta t)}\,dtdf \qquad \text{V.34}$$

V.5.7.1 Propriétés d'AF :

La fonction d'ambigüité vérifie les propriétés de WV, parmi ces propriétés on cite :

1. Propriétés marginale :

L'autocorrélation temporelle et spectrale est l'AF sur l'axe de τ et ξ respectivement [29]:

$$r_x(\tau) = A_x(0,\tau) \;et\; R_x(\zeta) = A_x(\zeta,0). \qquad \text{V.35}$$

L'énergie d'un signal x est la valeur de AF à l'origine ($A_x(0,0)$).

2. Géométrie d'interférence :

La représentation d'AF présente des interférences mais apparaissent d'une distance lointaine de l'origine inversement au DWV qui correspond au composants du signal où se situe à l'origine. La Figure V.25 représente la DWV pour un signal à deux atomes. On note que les interférences pour DWV se situent entre les composants du signal. La Figure V26 représente AF pour un signal à deux atomes.

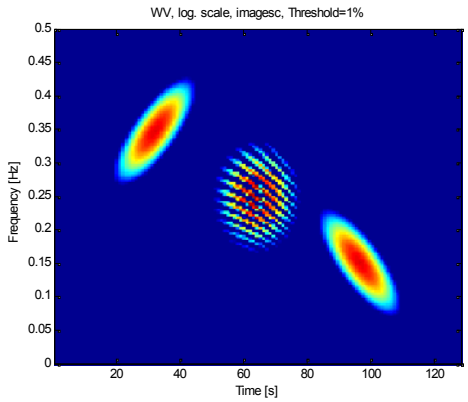


Figure V.25 DWV pour un signal à deux atomes

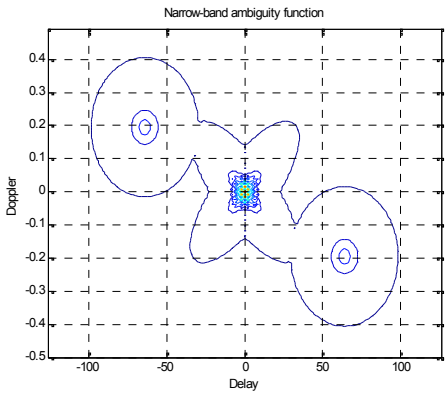


Figure V.26 AF pour un signal à deux atomes

V.5.7.2 Conclusion :

L'AF à bonde étroite présente une nouvelle représentation en temps-fréquence basée sur la mesure de la similarité d'un signal et son translation dans le plan temps-fréquence. Les interférences se situent au loin de l'origine où se trouve l'AF qui représente les composants du

signal, se qui signifie la possibilité d'utiliser un filtre passe bas à deux dimensions pour éliminer carrément les interférences.

V.5.8 Conclusion :

En conséquence, lorsque le signal est multi-composant, les interférences compliquent l'analyse, la réduction de celles-ci est réalisée par l'application d'un lissage dans le plan temps-fréquence.

Une autre solution (n'est pas abordée dans notre étude), plus « géométrique », permettant d'améliorer la lisibilité des distributions de la classe de Cohen est basée sur la méthode de la *réallocation* [32], le principe est de réarranger le plan temps-fréquence pour y améliorer la localisation des Composantes d'un signal.

Alors on n'a pas d'une méthode de représentation exempte d'erreur.

V.6 Estimation du Pitch par les méthodes temps-fréquences :

On explore plusieurs algorithmes d'estimation de pitch appartenant au domaine temps-fréquence.

V.6.1 Spectrogramme :

Parmi les méthodes anciennes et intuitive d'estimation de pitch, la méthode basée sur la transformée de Fourrier à cours terme nommée « spectrogramme » ou (STFT).

V.6.1.1 Comment lire un spectrogramme ? :

Bien que le spectrogramme s'obtient en visualisant le spectre de puissance d'une transformée de Fourier à court terme (STFT) (voire équation V.1), la Figure.27 nous montre les différents composants d'un signal parole via le spectrogramme.

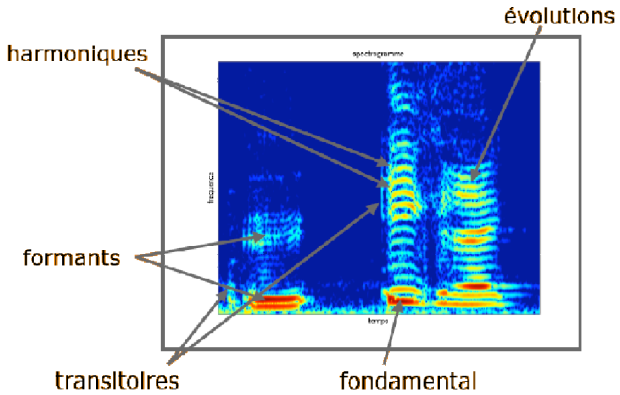


Figure V.27 Différents composants d'un signal parole visualisés par le spectrogramme.

V.6.1.2 Algorithme de détection de pitch :

Dans cet algorithme, une détermination de la fréquence fondamentale est basée sur la découverte de maximum simple de spectre à court terme de signal parole.

Les transformées en ondelette continue (étudier dans le chapitre suivant) (TOC) produit une représentation temps-échelle, sa définition est comme suit [34,36] :

$$TOC \quad (t, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t)\psi \left(\frac{t-\tau}{a}\right) dt \qquad \text{V.36}$$

Où 'a' : est l'échelle, τ : translation sur l'axe des temps. Ψ : l'ondelette mère.

Cette fonction a été utilisé dans notre étude pour atténuer et même éliminer les harmoniques à cause que les ondelettes se comporte comme un *banc de filtre,* afin que le maximum de spectre à cours terme de la transformée en ondelette correspond vraiment à la fréquence fondamentale.

Après avoir calculé les coefficients de TOC, on applique la transformée de Fourier à court terme (STFT) sur les valeurs des coefficients, en suite on utilise un détecteur de maximum de STFT pour trouver la fréquence fondamentale qui correspond à ce maximum.

Le STFT susceptible de fournir une exacte estimation de pitch lorsque le signal de la parole est corrompu par le bruit.

V.6.1.2.1 Atténuation ou élimination des harmoniques (filtrage) :

Le signal parole est naturellement multi-composants où il y a N - 1 harmonique (N : nombre d'échantillons) dans le spectre. Pour utiliser le sommet (pique) de carré STFT afin de détecter la fréquence fondamentale, les harmoniques doivent être enlevé ou atténué afin que le sommet correspond à la fréquence fondamentale.

Dans le cas général, c'est difficile d'utiliser un filtre passe bas fixe des que la fréquence fondamentale change avec le temps, alors un filtre passe bas à temps variable est nécessaire de conserver la fondamentale, pour concevoir un satisfaisant filtre à temps-variable, la connaissance a priori de pitch est exigé, cependant, une telle connaissance n'est pas disponible [33]. Mais, l'usage d'un *banc de filtres* est plus considéré qu'un filtre à temps-variable alors que la TOC semblable à ce banc de filtre.

Les ondelettes sont largement utilisées avec un signal parole [35]. La figure V.28 présente un exemple d'ondelettes mère *Daubetchies d'ordre 2.*

On utilise cette fonction comme un filtre. On note que dans notre cas expérimentale on a choisit une ondelette mère « *Daubetchies* » d'ordre 4 (db4).

La figure V.29 (a) et figure V.29(b) montrent un segment (trame) d'un signal parole pour un masculin et son spectre (sans application de TOC). Dans ce spectre, l'amplitude spectrale maximum ne correspond pas au pitch mais au premier harmonique.

La figure V.30(a) et figure 30(b) montrent la transformé d'ondelette de la trame vue au Figure V.29. On remarque l'atténuation des harmoniques après l'application de la transformée d'ondelette et augmente le maximum qui correspond au pitch recherché ce qui signifie la facilitée d'estimation de pitch.

V.6.1.2.2 Détermination de la fréquence fondamentale :

L'algorithme d'estimation de la fréquence fondamentale suit les étapes suivantes :

1. Sélection d'une tranche de signal soumis à un fenêtrage (Hamming).
2. Détection voisé non voisé par EZR.
3. Elimination ou atténuation des harmoniques basés sur la transformé en ondelettes.
4. Application de calcul de la transformée de Fourrier à courts terme (STFT) aux coefficients de TOC, par suite détection de maximum de STFT qui correspond à la fréquence fondamentale.

V.6.1.2 Evaluation de la méthode:

Pour l'évaluation de la méthode on a enregistré un signal parole « Bonjour » en utilisant le logiciel « winPitchPro » ; on applique l'algorithme d'estimation de pitch où on résulte le contour de pitch vue à la figure V.31. Pour l'évaluation vis-à-vis le bruit on a enregistré un signal sans bruit exprime un phonème 'a' ; par suite le faire noyer dans un bruit avec différentes puissances. Les figures V.32 représentent le signal 'a' et ses contours de pitch par différentes puissances de bruit,le tableau V.1présente l'évolution de pitch en fonction de SNR.

V.6.1.3 Résultats expérimentaux :

Les résultats de simulation sont présentés comme suit.

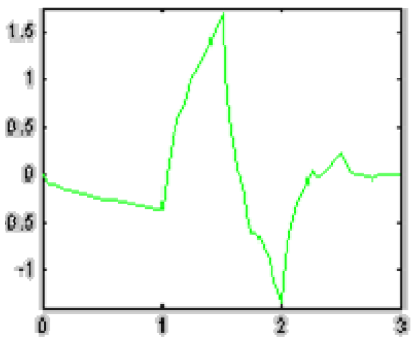


Figure V.28 Exemple d'une fonction d'ondelette.

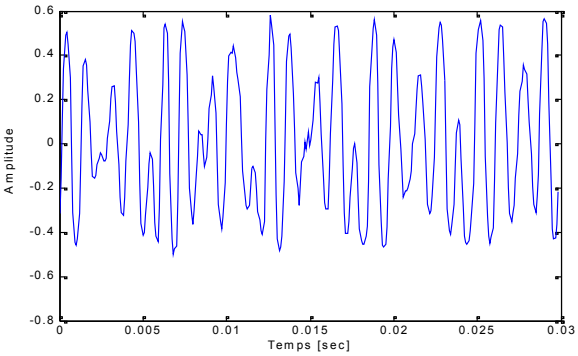


Figure V.29(a) Trame d'un signal masculin

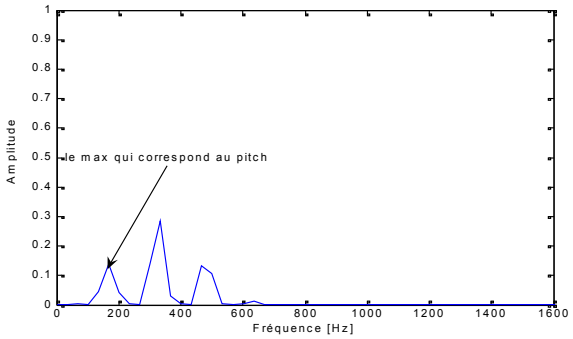


Figure V.29(b) Le spectrogramme de la trame

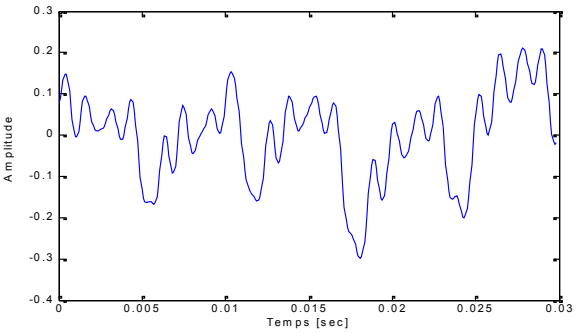


Figure V.30(a)TOC de la trame de fig (V.29(a))

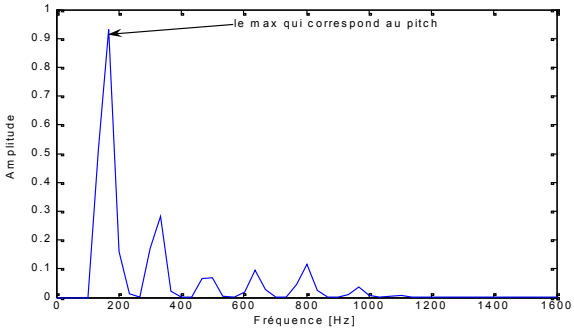


Figure V.30(b) Le spectrogramme de TOC

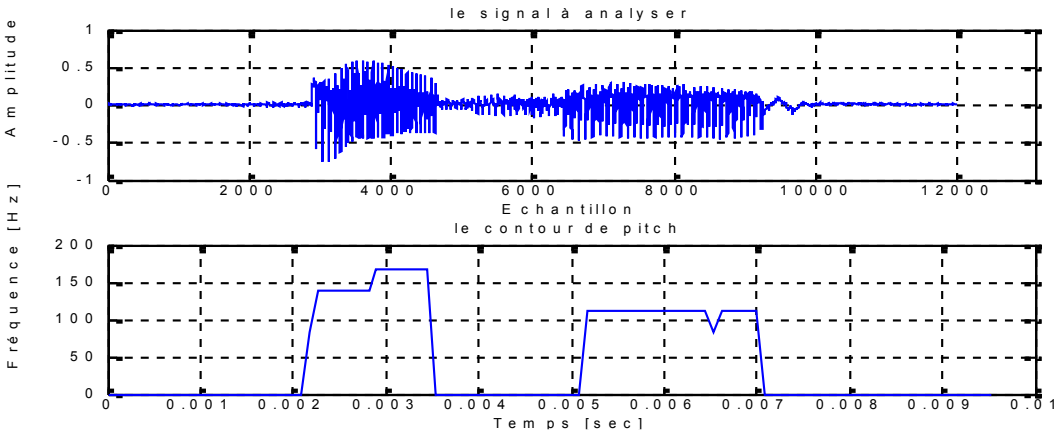


Figure V.31 Le signa parole « Bonjour », avec son le contour de pitch.

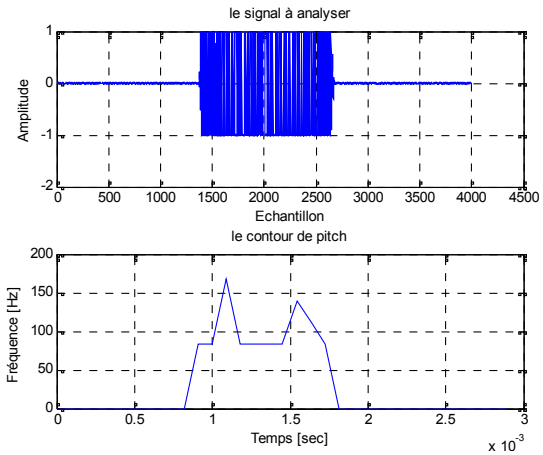


Figure V32(a). Le signal 'a' sans bruit et son contour de pitch : F0= 139.320 Hz

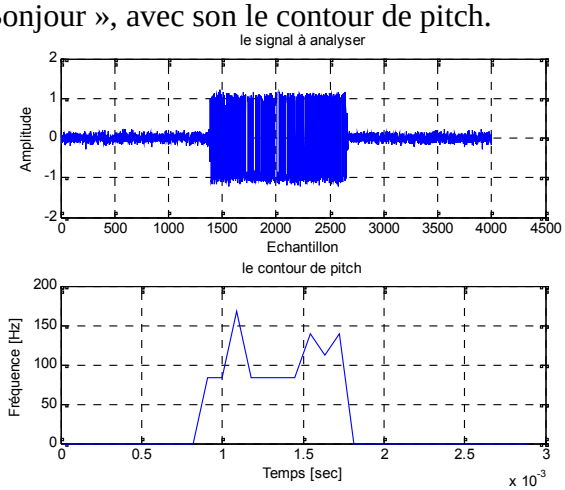


Figure V32(b) Les résultats pour SNR= 10dB : F0= 139.3206Hz

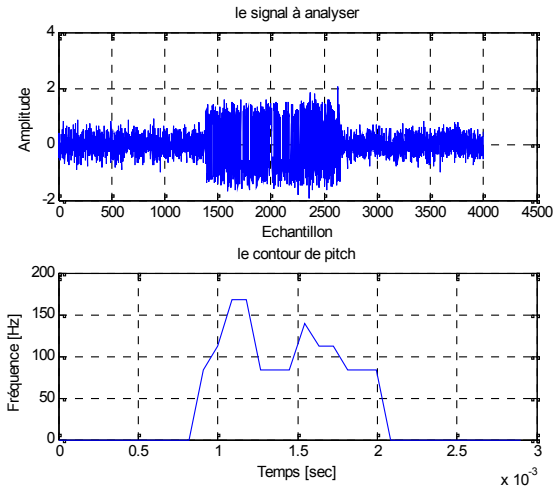


Figure V32(c) Les résultats pour SNR= 0dB. F0= 111.4565 Hz.

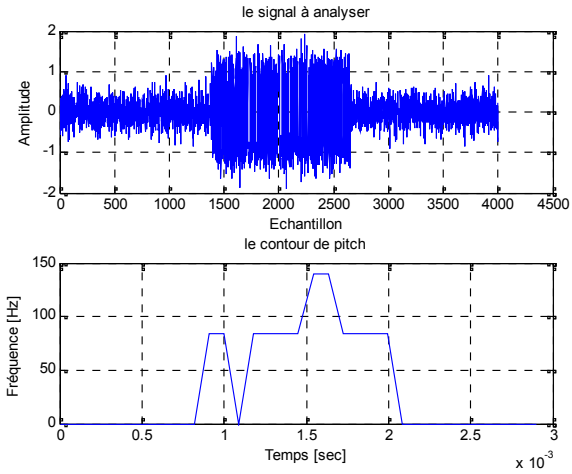


Figure V32(d) Les résultats pour SNR=1dB : F0= 83.5924 Hz

SNR (rapport signal sur Bruit) [dB]	Le pitch [HZ]
Sans bruit	139.3206
50	139.3206
20	139.3206
10	139.3206
0	111.4565
-1	83.5924

Tableau V.1 L'influence de SNR sur l'estimation de pitch.

V.6.1.3 Conclusion :

Nous savon présentés une méthode d'estimation de pitch de STFT basée sur la transformé d'ondelette qui a fournit une robustesse à la méthode vis-à-vis la puissance de bruit (jusqu' à 5 dB) malgré les termes d'interférences, la détection voisée/non voisée basée sur EZR donne une bonne facilité de calculs et contour de pitch bien claire. Mais l'inconvénient traditionnel est le mois résolution dans le plant temps-fréquence.

V.6.2 Wigner-Ville :

Il a été démontré que la DWV possède des meilleures résolutions en temps et fréquence que la méthode STFT, donc c'est raisonnable d'attendre que la DWV exécutera mieux pour le bute d'estimation de pitch. Cependant, il y a un inconvénient pour la DWV, c'est, l'influence des termes d'interférences à la résolution temps-fréquence dû à la bilinéarité de DWV, donc certaines techniques doivent être appliquées pour vaincre cet inconvénient. Les performances d'estimation de pitch basées sur la DWV est d'éliminer l'effet des termes d'interférence, il a été démontré dans [37] que l'effet des interférences peut être réduit grandement en faisant la

moyenne de WVD sur le long de l'axe du temps qui peut être illustré clairement à travers le rapport suivant [35] :

$$\frac{1}{T}\int_{0}^{T} e^{-j2\pi f\tau} \leq \frac{1}{\pi fT} \qquad \text{V.37}$$

Les équations (V.37) et (V.38) impliquent que si le signal est stationnaire dans un intervalle du temps, alors l'influence des termes d'interférences peut être réduite en intégrant la DWV sur cet intervalle du temps. Ce n'est pas difficile de généraliser la conclusion précitée au cas d'un signal multi composants.

Bien que pour un son voisé d'un signal parole, le signal est supposé stationnaire si l'intervalle analytique est assez court de 20ms à 30ms. Par conséquent, si nous définissons la moyenne de DWV comme suit :

$$\overline{W_x(t,f)} = \frac{1}{T}\int_{t-T}^{t} W(t,f)\,dt \qquad \text{V.38}$$

Alors le problème d'interférences peut être négligé pour une période T, ce qui nous fournit une estimation de pitch basée sur la moyenne de DWV [37].

L'équation V.38 est semblable à l'équation marginale (équation V.18) ce qui signifie que le résultat de la moyenne de DWV est une densité spectrale de signal originale pendant une période T supposé stationnaire, c'est la densité spectrale de signal originale [38].

V.6.2.1 Application de transformée d'HILBERT (TH) :

L'application de la transformée d'Hilbert (TH) résulte un signal analytique où les composants négative pour la DWV sont éliminés ce qui élimine par conséquent les interférences dû aux ces composants négatives. L'inconvénient essentielle de la transformé d'Hilbert est d'influencer sur quelque propriétés de DWV, ce qui signifie l'influence sur le résultat d'estimation du pitch [38].

V.6.2.2 Algorithme de détection de pitch :

Dans cet algorithme, une détermination de la fréquence fondamentale est basée sur la découverte des sommets de la moyenne de la DWV (figure V.35) qui représente un spectre, le pitch correspond au premier sommet, mais des que ce dernier est faible dû à l'effet de la réponse vocale. On applique la méthode de *cepstre* (figure V.36) pour la détection et l'estimation de pitch.

V.6.2.2.1 Estimation de la fréquence fondamentale :

Pour la détection et l'estimation de pitch on suit les étapes suivantes :

1. Acquisition de signal, et sélection d'une tranche.
2. Application de la transformée d'Hilbert, et décision voisée /non voisée par EZR.
3. Calcul de DWV et $\overline{W_x(t, f)}$ en utilisant les équations V.15, V.38 respectivement.
4. Calculant le cepstre de la moyenne de DWV.
5. Extraction de maximum de cepstre.
6. Estimation de pitch : Le pitch= Fs /m, m : l'échantillon qui correspond au max de cepstre.

V.6.2.2.2 Evaluation de la méthode:

On prend une trame de 30ms d'un masculin (figure V.29(a)), on applique l'algorithme d'estimation de pitch basé sur la moyenne de DWV. Figures V.29(a), V.33, V.34, V.35 où la figure V.33 exprime la DWV d'une trame d'un signal ''a'' représenté par la figure V.29(a), la figure V.34 représente la DWV moyenne après calcul de DWV, la figure V.35 représente le cepstre de la DWV moyenne pour mieux extraire le sommet qui correspond au pitch.

On enregistre un signal parole« Bonjour», on applique l'algorithme d'estimation de pitch où la figure V.36 exprime le signal parole « bonjour » avec son contour de pitch.

Pour l'évaluation vis-à-vis le bruit, on a enregistré un signal sans bruit (un phonème 'a'); par suite le faire noyer dans un bruit avec différentes puissances. La figure V.37 représente le signal « a » (d'origine sans bruit) et son contour de pitch avec différents rapport signal sur bruit (SNR). Les résultats de calculs de pitch sont résumés dans le tableau V.2.

V.6.2.2.3 Discussion, interprétation, amélioration :

La première remarque qu'on peut extraire est la sensibilité au bruit. Malgré l'application de la DWV et son moyenne en considérant le signal analytique, et en raison des résultats de simulations obtenus, on peut deviner sur la présence encore de problème d'interférences qui persiste. Ce qui nous conduit à penser sur l'interprétation de ces résultats.

Interprétation :

Soit x(t) un signal utile additionné à un bruit additive n(t), on résulte un signal s(t) :

$$s(t) = x(t) + n(t) \qquad \text{V.39}$$

Si nous calculons la DWV :

$$W_s(t, f) = \int s^*(t - \lambda/2)\, s(t + \lambda/2)\, e^{-2i\pi f\lambda}\, d\lambda \qquad \text{V.40}$$

Après développement de la formule précédente on aura :

$$W_s(t,f) = W_x(t,f) + W_n(t,f) + 2\,\mathrm{Re}\left\{ W_{nx}(t,f) \right\} \qquad \text{V.41}$$

Alors d'après l'équation V.41, la DWV d'un signal bruité est la somme de DWV de signal utile plus la DWV de bruit plus les termes d'interférence dû aux deux composants (signal+bruit). Pour un objectif d'une interprétation graphique on prend une tranche (30ms) sans bruit (figure V.29(a)), on prend la DWV de cette tranche où la figure V.38 exprime la DWV de cette tranche, on noyé ce signal dans le bruit (0dB), la figure V.39 représente la DWV de signal bruité (signal utile + bruit) où on constate la présence des interférences justifiée par les équations V.39, V.40, V.41.

Amélioration :

Après plusieurs tentatives de recherches on a achevé qu'il ya une autre méthode de réduire les interférences et le bruit, c'est l'application d'un filtre passe-bas classique pour chaque tranche (30 ms) [38].

V.6.2.2.4 Evaluation de la nouvelle méthode :

Alors on applique un filtre passe bas (dans notre analyse, on a appliqué un filtre passe bas 0-500Hz de Botherworth d'ordre 8) pour une tranche sans bruit, en suite on considère le signal analytique (application de la TH) et la DWV moyenne pour estimer le pitch en suite on additionne un bruit d'un SNR=0dB (pour évaluer la méthode concernant la présence de bruit). La figure V.40 représente la DWV d'une tranche sans bruit soumis au filtrage passe bas en considérant le signal analytique avec résultats de F0=33.33Hz. La figure V.41 représente la DWV d'une tranche par SNR de 0dB soumis au filtrage passe bas en considérant le signal analytique. On aura une bonne représentation (élimination des harmoniques) mais avec moins résolution, ainsi un résultat d'estimation de F0=33.33Hz ce qui est faux que l'on justifié par l'influence de la TH sur quelque propriétés de DWV. Dans la deuxième expérience on élimine la TH et on estime le pitch où les Figures V.42, V.43 expriment la DWV d'une tranche sans bruit soumis au filtrage passe bas et sans application de la TH (le pitch F0=157.5Hz) et la DWV d'une tranche avec SNR de 0dB soumis au filtrage passe bas sans application de la TH respectivement (Le pitch F0=157.5Hz). Les résultats d'estimation de pitch de vis-à-vis la présence de bruit sont présentés dans le tableau V.3.

V.6.2.3 Résultats expérimentaux :

Les résultats de simulation concernant l'analyse et l'amélioration de la méthode DWV sont comme suit :

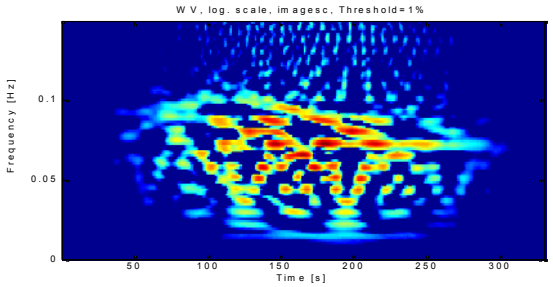


Figure V.33. DWV de la trame de « a »

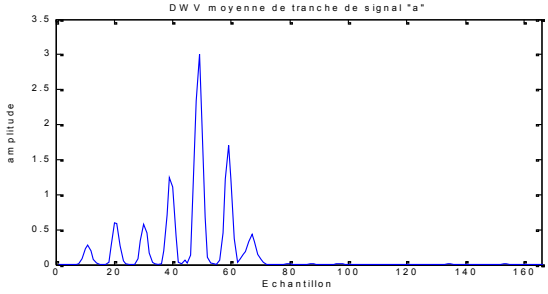


Figure V.34 La DWV moyenne

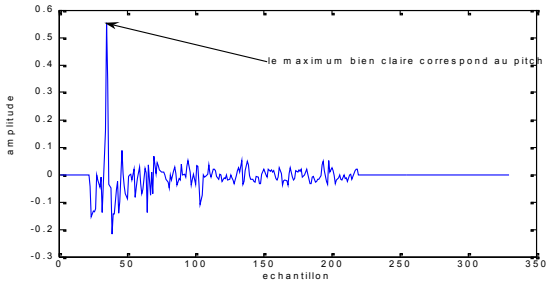


Figure V.35 Le cepstre de la DWV moyenne

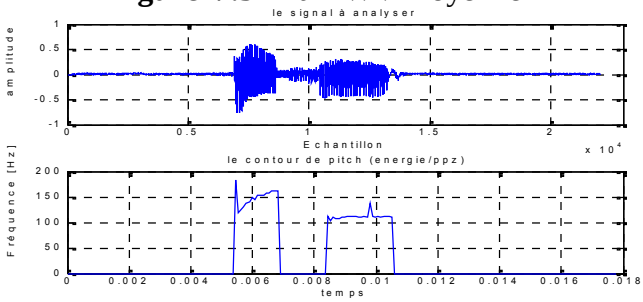
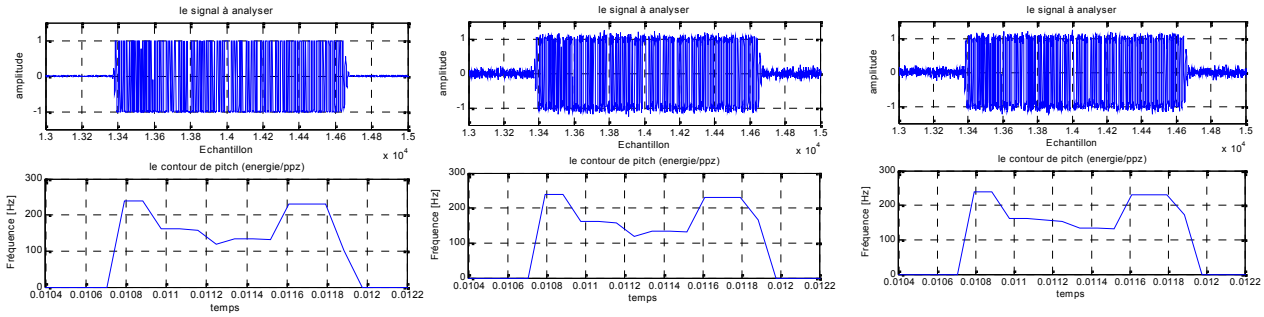


Figure V.36 Signal parole ‘’Bonjour’’ avec son contour de pitch.



(a), Sans bruit, F0=131.25, (b) SNR=30dB. F0=131.25Hz (c)SNR= 20dB. F0=157.5Hz

Figure V.37 Signal « a » avec différents SNR.

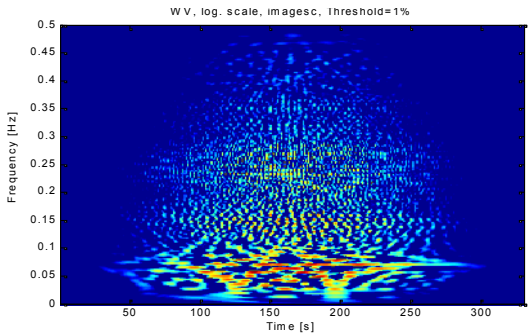


Figure V.38 DWV d’une tranche ‘a’ sans bruit

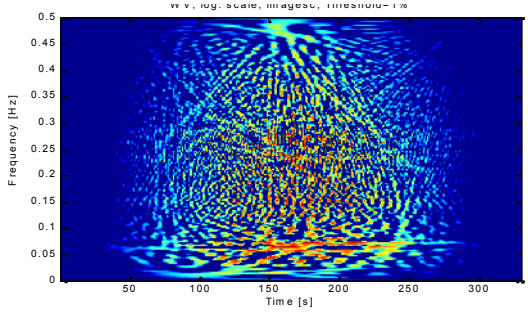


Figure V.39 DWV avec SNR=0dB

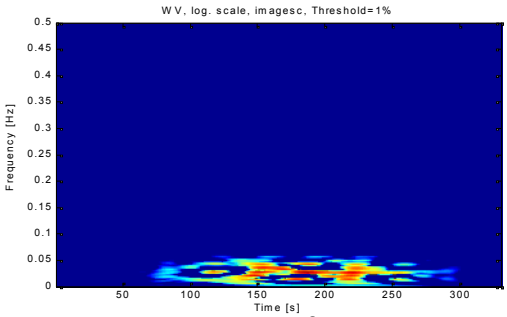


Figure V.40 DWV, filtrage, TH: F0= 33.33Hz

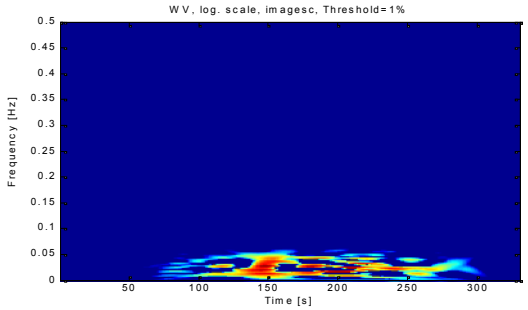


Figure V41 DWV, Filtrage, TH , SNR=0dB

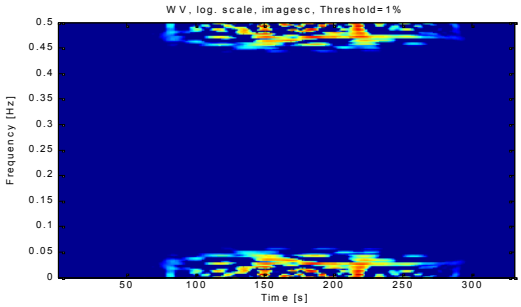


Figure V.42 DWV, filtré, sans TH, : F0=157.5Hz

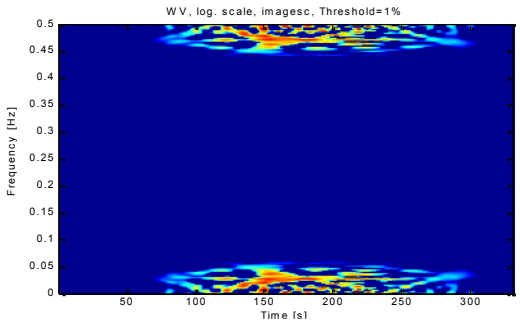


Figure V.43 DWV, filtré, sans TH, SNR=0dB : F0=157.5Hz,

Le pitch [Hz]	SNR [dB]
131.25	Sans bruit
131.25	30
157.5	20
162.1324	0
119.837	-2
134.4512	-8

Tableau V.2 Le pitch par WV vs SNR

Le pitch [Hz]	SNR [dB]
157.50	Sans bruit
157.50	30
157.50	20
157.50	0
157.50	-9
122.50	-15
79.22	-30

Tableau V.3 Pitch par WV vs SNR (amélioration)

V.6.2.4 Conclusion :

Malgré la bonne résolution de DWV, et les bons résultats (estimation de pitch, contour de pitch), et l'utilisation de DWV moyenne, la méthode DWV souffre de problèmes d'interférences à cause que cette distribution est quadratique, ainsi elle soufre de la sensibilité vis-à-vis le bruit (voire tableau V.2). Bien que notre objectif est l'estimation de la fréquence fondamentale on a remédié aux problèmes par application d'un filtre passe bas pour chaque tranche de 30 ms supposée stationnaire, on a aboutit à une bonne immunité aux interférences et au bruit (voire figures V.42, V.43). La méthode de DWV reste donner une bonne résolution dans le plan temps-fréquence.

V.6.3 Méthode de la Distribution de Pseudo Wigner-Ville (DPWV) :

La DPWV (équation V.27) est semblable au STFT, la DPWV est le DWV soit tronqué habituellement par une fenêtre « rectangulaire » glissante de longueur (2L+1) centrée au point « n » [39]:

$$n : h(t) = \begin{cases} 1, |n| \le L \\ 0, |n| > L \end{cases} \qquad \text{V.42}$$

Cette méthode a pour objectif de remédier aux problèmes des interférences par l'ajout d'un lissage sur l'axe des temps. De la même façon que la DWV, on suit les mêmes étapes pour la détection et l'estimation de pitch (application de DPWV moyenne, filtrage passe bas, TH).

V.6.3.1 Evaluation de la méthode et résultats expérimentaux :

On prend une tranche de « a » (sans bruit), on applique l'algorithme d'estimation de pitch en présence de différentes puissances de bruit, ainsi on prend un signal parole « bonjour », la figure V.48 représente le signal parole 'Bonjour' avec son contour de pitch, la fenêtre rectangulaire de lissage est choisit de longueur de 30ms. L'ajout d'un filtrage passe bas pour chaque tranche a pour objectif d'éliminer les harmoniques et de fournir une bonne immunité au bruit. Les figures V.44, V.45, V.46, V.47 expriment la DPWV de signal de la tranche sans bruit, la DPWV avec un filtre passe bas (0-500 Hz), le résultat de la DPWV moyenne, le résultat de la DPWV moyenne avec le filtre passe bas (0-500 Hz) respectivement.

Le tableau V.4 représente les résultats d'estimation de pitch en présence de différentes puissances signal sur bruit (SNR).

D'après les résultats de simulation, on constate presque les mêmes résultats que la méthode DWV.

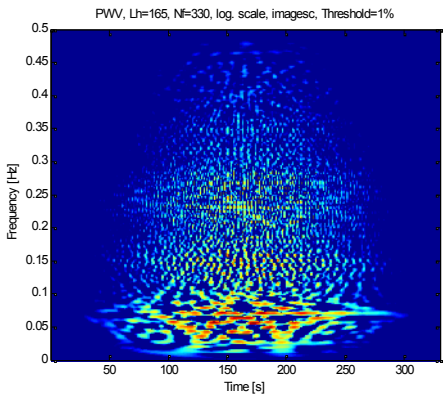


Figure V.44 DPWV

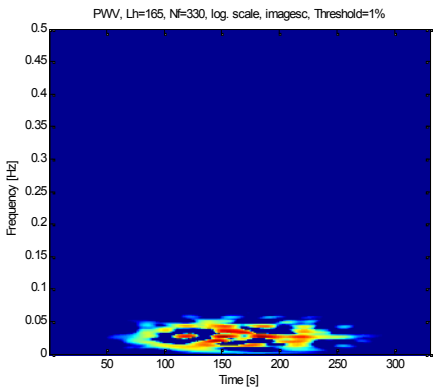


Figure V.45 DPWV avec Filtre Pb

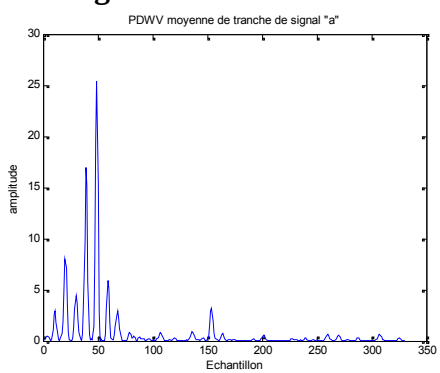


Figure V.46 DPWV moyenne de Fig V.44

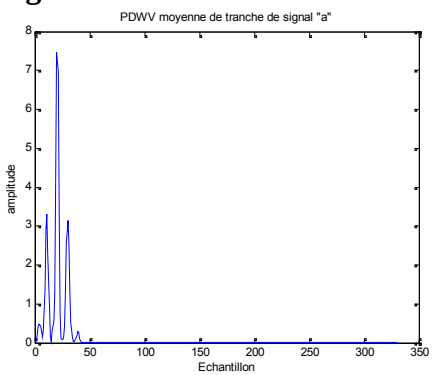


Figure V.47 DPWV moyenne de FigV.45

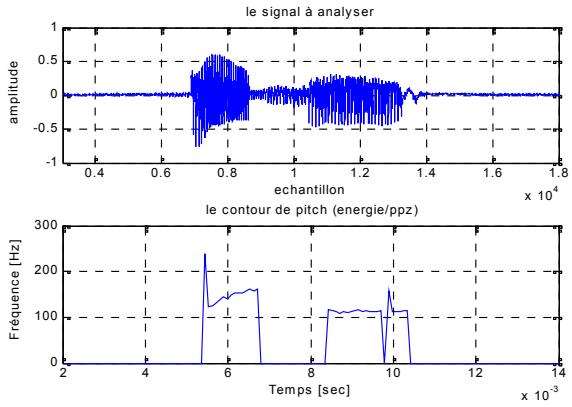


Figure V.48 ‘’Bonjour’’ avec son contour de pitch

Le pitch [Hz]	SNR [dB]
153.25	Sans bruit
153.25	30
153.25	20
153.25	0
153.25	-9
120.50	-15
75.22	-30

Tableau V.4 Pitch par PWV vs SNR.

V.6.3.2 Conclusion :

D’après les résultats de simulation, on conclure des bons résultats, grâce au filtre passe bas à court terme (application de filtre Pb, pour chaque trame de 30ms).la DPWV donne les mêmes résultats que la méthode DWV.

V.6.4. Distribution de Pseudo-Wigner-Ville-Lissé (DPWVL) :

Est une méthode temps-fréquence (voire **V.5.3.2)**, elle se caractérise par son principe d’utiliser une fonction de lissage séparable en temps et fréquence .Définition à temps continu (Equation V.28). Donc Cette méthode à pour objectif de remédier aux problèmes des interférences par l’ajout d’un lissage sur les deux axes (temps et fréquence). Pour la détection et l’estimation de pitch on suit les étapes suivantes :

1- Acquisition de signal.
2- Le choix d’une trame à traiter de longueur (30ms).
3- Application de la transformée d’Hilbert du segment de signal.
4- Le choix de longueur des fenêtres de lissage sur l’axe temps et l’axe de fréquence.
5- Calcul de DPWVL et par suite $\overline{DPWVL}$ (Valeur moyenne).
6- Calculant le cepstre de $\overline{DPWVL}$.
7- Extraction de maximum de cepstre et estimation de pitch de la même façon que DWV.

V.6.4.1 Evaluation de la méthode et résultats expérimentaux :

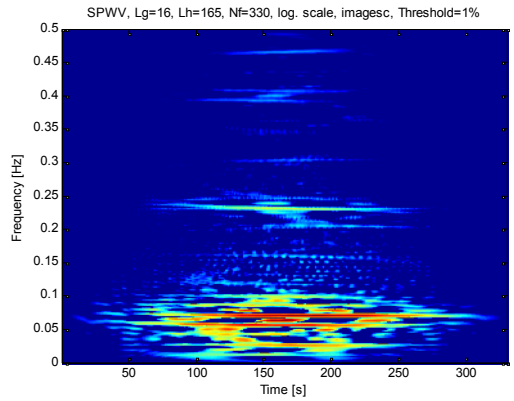
On prend une tranche de « a » (sans bruit), on applique l’algorithme d’estimation de pitch en présence de différentes puissances de bruit, ainsi on prend un signal parole « bonjour » pour construire un contour de pitch.

La figure V.41 représente le signal parole ‘Bonjour’ avec son contour de pitch en appliquant la méthode DPWVL, la fenêtre (rectangulaire) de lissage sur les deux axes temps et fréquence est de longueur 3ms et 30ms respectivement.

On a aboutis un contour de pitch semblable au celle qui résulté par les méthodes DWV et PDWV ainsi une bonne résistivité au bruit.

Les figures V.49. V.50 représentent la DPWVL d’une trame de 30ms sans bruit, le spectre de $\overline{DPWVL}$ pour cette trame où F0=153.25Hz respectivement.

Le tableau V.5 représente les résultats d’estimation de pitch en présence de différentes puissances de bruit (SNR).



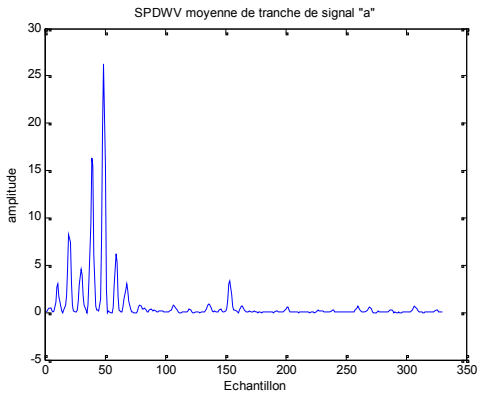
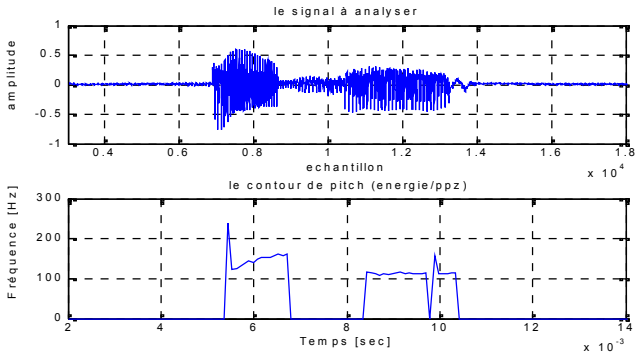


Figure V.49 DPWVL d’une trame « a ». **Figure V.50** $\overline{DPWVL}$ d’une trame « a ».



Le pitch [Hz]	SNR [dB]
153.25	Sans bruit
153.25	30
153.25	20
153.25	0
153.25	-9
120.50	-15
75.22	-30

Figure V.51 ’Bonjour’ et son contour de pitch. par PWVL **Tableau V.5** Pitch par PWVL vs SNR.

V.6.4.2 Conclusion :

D’après les résultats expérimentaux, on conclure des bons résultats. la DPWVL donne les mêmes résultats que la méthode DWV et DPWV. Cette méthode exprime un filtrage sur l’axe de temps et l’axe de fréquence grâce aux fenêtres (rectangulaire) de lissage.

V.6.5 Choi-Williams basé sur la transformée dyadique d’ondelette :

Cette méthode de Choi-Williams est basée sur la transformée dyadique d'ondelette (TDO) [34]:

$$TDO \quad (t, 2^{j}) = \frac{1}{\sqrt{2^{j}}} \int_{-\infty}^{+\infty} s(t)\psi \left(\frac{t-\tau}{2^{j}}\right) dt \qquad \text{V.42}$$

J : utilisé pour représenter les différentes échelles dans l'analyse multi-résolution ainsi les différentes bandes de fréquence. TDO semblable à un *banc de filtres.*

Les ondelettes dyadiques (TO) sont largement utilisées avec un signal parole [35,40]. (Voire figure V.52 pour les ondelettes avec son spectres).

On peut voir (figure V.52) que les ondelettes raisonnablement bien localisé dans le domaine temporel et fréquentiel.

Les ondelettes fonctionnent à échelle j = 4 ont la gamme de fréquence d'approximativement 10 - 300H.z [33] (presque la même plage fréquentiel de pitch) alors peuvent éliminer les harmoniques qui est plus haut que 300Hz et atténue les autres au-dessous de 10H.z, la fonction est semblable à un filtre de bande passante de 10Hz à 400Hz. Aussi, C'est l'équivalent à un filtre passe bas (O – 400 Hz).

On note qu'il y a des expérimentations faites par plusieurs auteurs (voire [33], [40]) ont montré que pour un femelle la bonne transformée dyadique d'ondelette celui avec l'échelle j=3(la gamme de fréquence est entre 100 et 450 Hz) et pour un masculin j=4.

V.6.5.1Estimation de pitch :

Après avoir calculé la TDO, et le calcul de la DCW, on procède la recherche de maximum de DCW moyenne pour trouver la fréquence fondamentale [40].

La méthode de Choi-Williams, (l'équation V.32) dépend d'un variable 'σ' ; c'est le facteur qui a la capacité de supprimer les termes d'interférence. Alors avant de procéder l'estimation de pitch par cette méthode on doit identifier la valeur de paramètre σ.

V.6.5.2 Evaluation de la méthode et résultats expérimentaux :

Tout d'abord on a choisit σ=3 (pour un éventuel bon résultat), la figure V.53 nous montre la représentation de la distribution Choi-Williams pour une tranche d'un voie masculine après application de TDO, cette dernière et après calcul de la DCW moyenne on découverte la simplicité d'estimation de pitch qui correspond au maximum, la figure V.54 présente le spectre qui définit la DCW moyenne et le maximum qui correspond au pitch :

Pitch = Fs /m, (m : échantillon qui correspond au maximum), alors pour Fs=11025Hz et m= 72 on aura : F0=11025/72= 153.25 Hz.

On prend une tranche de « a » (sans bruit), on applique l'algorithme d'estimation de pitch en présence de différentes puissances de bruit, le tableau V.6 présente les résultats d'estimation de pitch en présence de différentes puissances de bruit (SNR).

On prend un signal parole « bonjour », on applique l'algorithme d'estimation de pitch par DCW, on aura la figure V.55 qui représente le signal parole « bonjour » avec son contour de pitch.

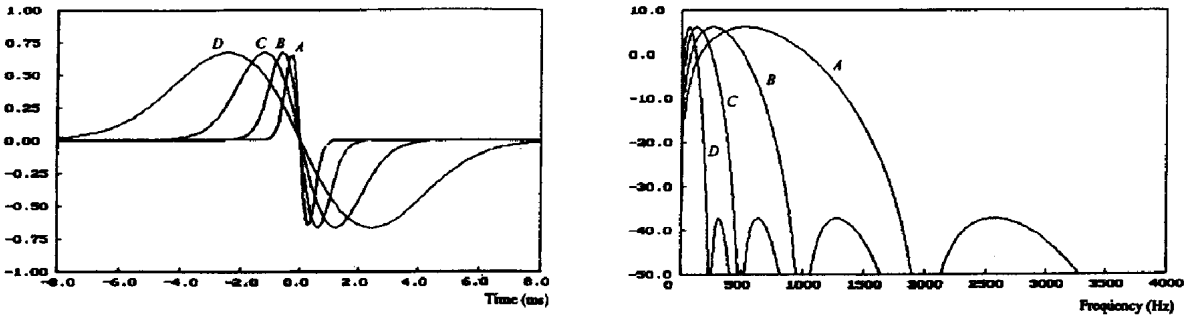


Figure V.52 les ondelettes (a) et ses spectres (A :j=1,B :j=2,C :j=3,D :j=4) (b)

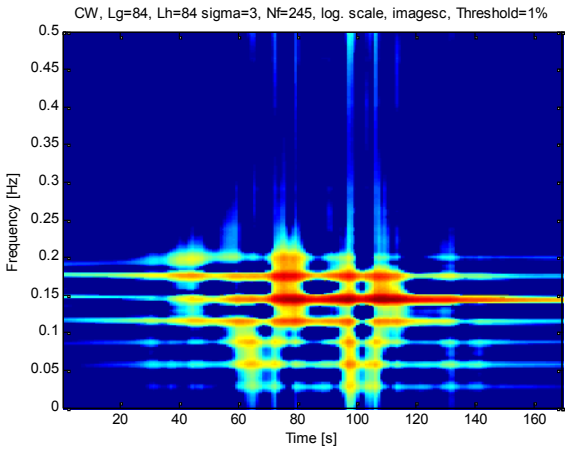


Figure V.53 La DCW de TDO d'une tranche masculine (σ=3).

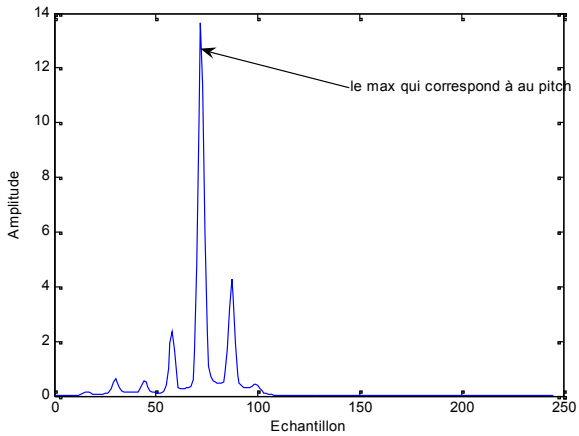


Figure V.54 La DCW moyenne d'une trame et le maximum qui correspond au pitch (prononcé par un masculin) j=4.

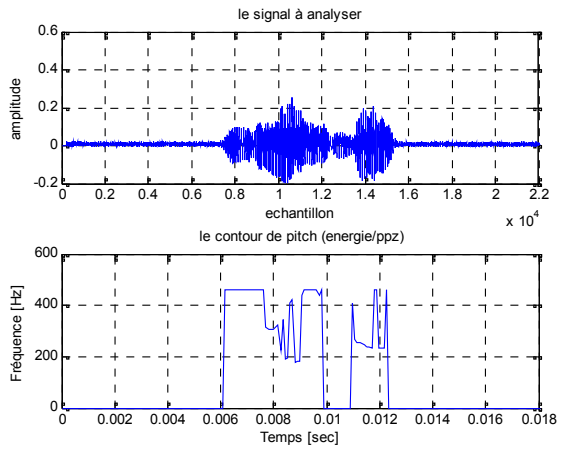


Figure V.55 Le contour de pitch pour un signal Bonjour (masculin), (σ=3).

Le pitch [Hz]	SNR [dB]
134.4512	Sans bruit
116.0526	20
134.4512	10
114.8438	5
134.4512	0
125.2841	-5
112.5000	-9

Tableau V.6 Pitch par DCW vs SNR

V.6.5.3 Conclusion :

Cette dernière méthode garde une bonne performance grâce à la réduction de bruit qui dépend d'un facteur σ ce qui signifie le choix de ce paramètre à chaque fois de l'application d'un signal parole. Cette méthode de Choi-Williams est appliqué à la transformée dyadique d'ondelette qui joue le rôle d'un banc de filtre, la dyadique d'ondelette est appliqué avec un signal parole par certains auteurs (voire : [34], [40]). Dans notre cas on a ajouté la moyenne de CW pour un objectif de réduction des interférences, cette algorithme est sensible au bruit (justifiée par les résultats cités au tableau V.6) mais donne un bon contour de pitch.

V.7 Conclusion :

Dans ce chapitre on a exploré, plusieurs méthodes temps-fréquence, la méthode la plus simple est classique c'est la STFT où elle ne fournit pas une bonne résolution dont on doit toujours définir la largeur de la fenêtre à utiliser.

La méthode DWV avec ses versions (DWV, DPWV, DPWVL) est plus connue que les autres méthodes ainsi donne une bonne résolution, cette méthode est améliorée par certains auteurs par l'utilisation des transformées d'ondelette pour un objectif de filtrage et robustesse à la détection de pique et la fourniture d'un pitch plus exacte.

L'utilisation de la moyenne de la distribution a pour objectif de réduire des termes d'interférences. Dans notre analyse on a entré un filtrage passe bas à court termes pour aider à réduire les harmonique et les termes d'interférences.

Bibliographie :

[26] C.ALESSANDRO,C.DEMARS « Représentations temps-fréquence du signal de parole » LIMSI-CNRS, Université Paris VI,France.1992.
[27] J. JEONG, W . J . WILLIAMS. On the cross-terms in spectrograms. Proceedings of IEEE-ICASSP 1991, pp . 1565-1568. 1990
[28] Patrick FLANDRIN,Bernard ESCUDIÉ. Principe et mise en œuvre de l'analyse temps fréquence par transformation de Wigner-Ville.1990.
[29]*F. Auger,P. Flandrin, P. Gonçalvès,O. Lemoine.* Tutorial"time frequency toolbox for use with Matlab".CNRS(France),Rice university (USA), *26 Octobre 2005.*
[30] Boîte à outils temps-fréquence: **www-isis.enst.fr/TFTB.**
[31] Ronald L. Allen, Duncan W. Mills, «Signal anlysis : time, frequency, scale,and structure » IEE Press 2004.
[32] E. Chassande-Mottin : *Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non-stationnaires*. These de doctorat, Universitede Cergy-Pontoise, 1998.
[33] Lunji Qiu, Haiyun Yang and So0 Ngee Koh « A Fundamental Frequency Detector of Speech Signals Based on Short Time Fourier Transform » Nanyang Technological University Singapore. 1994
[34] **A** Spaargaren, MJ English « Detecting Ventricular Late Potentials using the Continuous Wavelet Transform » University of Sussex**,** Brighton, UK.1999.
[35] Sam Kwong', Wei Gang", and Chan H Lee **«** A Pitch Detection Algorithm Based on Time-Frequency Analysis » Institute of Electronic Engineering and Control,South China University of Technoloiy , Guangzho.1992.
[36] Emmanuel Didiot (These)« Segmentation parole/musique pour la transcription automatique de parole continue » université Henri Poincaré Nancy 1France .NOV 2007.
[37] *S.* Kwong, G. Wei, and J. Z. Ouyang, "Fundamental frequency estimation based on adaptive time averaging Wigner-Ville distribution," Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Victoria, British Columbia,Canada.Oct 1992.
[38] Z. Leonowicz, T. Lobos. «ANALYSIS OF THREE PHASE SIGNAL USING WIGNER SPECTRUM » Chair of the Theory of Electrical Engineering Department of Electrical Engineering Wroclaw University of Technology, Wroclaw, Poland
[39] R. Yu and E. C. Tan « Comparison of Different Time-Frequency Distributions in Pitch Detection » School of Computer Engineering, Nanyang Technological University Nanyang Avenue, Singapore 639798, Singapore,IEE 2003.
[40] Namhoon Kim Heungkyu Lee Hanseok Ko. « Reliable Pitch Period Estimation Based on Wavelet Transform and Choi-William Distribution «Dept. of Electronics Engineering, Korea University ».1995.

VI.1 Introduction :

On explore une approche de détection de la fréquence fondamentale fondée sur l'utilisation de la décomposition en ondelettes du signal parole.

VI.2 Présentation des ondelettes :

Cette section présente rapidement la base de notre approche en parametrisation du signal, à savoir la décomposition de signal en ondelettes.

VI.2.1 Un peu d'histoire :

L'évolution temps – fréquence peut être mise en évidence en décomposant le signal parole en fonctions élémentaires bien concentrée en temps et en fréquence. La transformée de Fourier à fenêtre et la transformée en ondelette sont deux exemples importants de décomposition temps – fréquence. C'est en 1946 que le physicien GABOR [41] propose d'analyser les signaux sonores avec des atomes élémentaires qui sont des fonctions concentrées en temps et en fréquence. En montrant que de telles décompositions étroitement liées à notre perception des sons, et qu'elles isolent les structures importantes des signaux de parole, les travaux de GABOR furent à la base de l'analyse temps – fréquence. GAGOR introduit ainsi en 1946 les atomes de Fourier à fenêtre afin de mesurer « les variations fréquentielles » des sons parole.

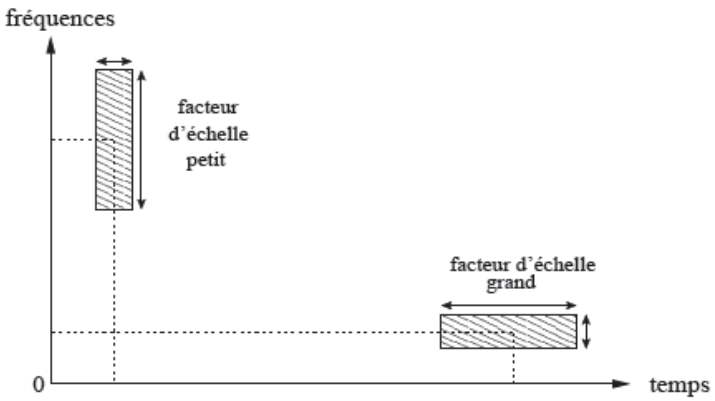


Figure VI.1 Boites d'Heisenberg correspondant au pavage du plant temps-fréquence de la transformée en ondelettes à des échelles différentes. Une échelle plus petite réduit de l'étalement en temps mais augmente la taille du support fréquentiel.

La résolution temps-fréquence de la transformée de Fourier à fenêtre dépend de l'étalement de la fenêtre en temps et en fréquence, cet étalement correspond à la surface de la boite d' HEISENBERG (voire la figure VI.1), en effet les concentrations en temps et en fréquence sont limitées par le principe d'Heisenberg.

Le principe d'incertitude d'Heisenberg indique qu'un signal ne peut pas être simultanément connu avec précision en temps et en fréquence, le produit de ces deux quantités étant bornées inférieurement.

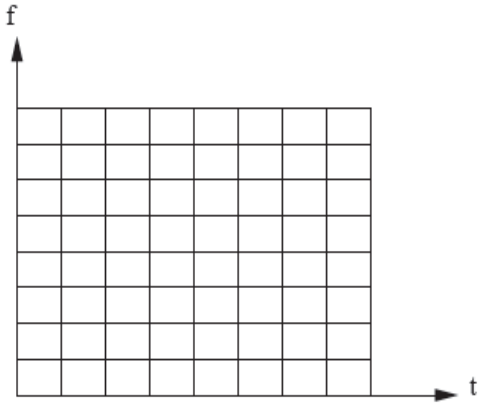


Figure VI.2 Exemple de couverture temps fréquence avec la TF à fenêtre.

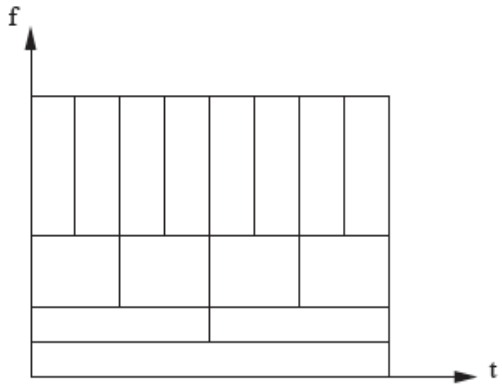


Figure VI.3 Un exemple de couverture temps -fréquence avec la transformée en odelettes

La Figure VI.2 montre que la résolution temporelle et fréquentielle restent inchangée quelque soit le temps et la fréquence. Un inconvénient de la transformée de Fourier à fenêtre est le réglage de la taille de la fenêtre d'analyse. Ce réglage est un compromis entre la résolution temporelle et la résolution fréquentielle. On perd en localisation fréquentielle ce qu'on a gagné en localisation temporelle ceci à cause du principe d'incertitude d'Heisenberg.

. Ainsi une représentation satisfaisante de la structure temporelle fie du signal permettant par exemple de voire les transitions entre phonèmes se fera au détriment de la résolution fréquentielle (analyse large bande). Inversement, une analyse permettant de bien faire apparaître les composants harmoniques du signal se fera au détriment de la résolution temporelle et ne rendra pas compte des événements temporels brefs (analyse bande étroite).

Une fois ce réglage effectué, la taille de la fenêtre sera fixée et la résolution de la transformée de Fourier à fenêtre restera la même sur tout le plant temps-fréquence (Figure VI.2) mais pour analyser des composants transitoires de durées différentes comme c'est souvent le cas de parole, il est nécessaire d'utiliser des atomes dont les supports temporels ont des tailles variables. La transformée en ***ondelettes en est la solution*** (figure VI.3).

VI.2.2 Définitions :

Nous avons vu qu'une alternative, pour dépasser les limitations de la transformée de Fourier à fenêtre, se trouve par l'utilisation de la transformée en ondelettes. Alors nous pouvons à présent définir ce qu'est une ondelette et comment réaliser une transformée en ondelettes du signal.

IV.2.2.1 Les ondelettes :

Une ondelette [41] est une fonction $\Psi \in \geq L^2(R)$ de moyenne nulle [41] :

$$\int_{-\infty}^{+\infty} \Psi(t)dt = 0 \qquad \text{VI.1}$$

Et à énergie finie [41] :

$$\int_{-\infty}^{+\infty} \left|\Psi(t)dt\right|^2 < +\infty \qquad \text{VI.2}$$

Elle est normalisée à $\|\psi\| = 1$[41], et centrée au voisinage de t=0. Une famille d'atomes temps fréquence s'obtient en dilatant l'ondelette par un facteur **a,** et en la translatant par **τ** [41] **:**

$$\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \quad \text{avec} \quad a \in \Re_+^* \qquad \text{VI.3}$$

L'ondelette peut être réelle ou analytique complexe. Selon les applications, on peut choisir l'une ou l'autre. Pour notre part, et pour l'objectif de détection du pitch en utilisant les ondelettes, nous avons opté pour une *odelette réelle.*

IV.2.2.2 La transformée d'ondelettes :

La transformée en ondelettes d'un signal *f(t)* à l'échelles **a** et au temps τ se calcule en corrélant *f(t)* avec l'ondelette $\psi_{\tau,a}$ correspondante. Ceci nous donne la définition de la transformée en odelettes (TO) (équation V.36).

Nous utiliserons par la suite uniquement des transformées en ondelette réelle car elle permette de mesurer la variation de *f(t)* dans un certain voisinage de «*τ*» de taille proportionnelle à « a », Il a été démontré que lorsque a tend vers 0, la décroissance des coefficients d'ondelettes caractérisent la régularité de *f*(t) au voisinage de τ .Cette propriété est très importante car elle permet de détecter des transitoires.

Une transformée en ondelette réelles est complète et préserve l'énergie tant que l'ondelette ψ satisfait une condition admissible donnée par le théorème suivant : $\psi \in L^2(\Re)$

Théoreme 1 [29] soit :

$$C_\psi = \int_0^{+\infty} \frac{\left|\psi(w)\right|^2}{w} dw < +\infty \qquad \text{VI.4}$$

Toute fonction x(t) vérifie [29]:

$$x(t) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_x(\tau, a) \frac{1}{\sqrt{a}} \psi \left(\frac{t-\tau}{a} \right) d\tau \frac{da}{a^2} \qquad \text{VI.5}$$

Et [41]:

$$\int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_0^{+\infty} \int_{-\infty}^{+\infty} |W_x(\tau, a)|^2 d\tau \frac{da}{a^2} \qquad \text{VI.6}$$

L'équation VI.4 du théorème précédent s'appelle la condition *d'admissibilité* de l'ondelette. Pour que l'intégrale soit finie, il faut s'assurer que TF(ψ)=0, ce qui explique pourquoi les ondelettes doivent être de moyenne nulle.

Enfin la dernière équation du théorème démontre la conservation d'énergie entre le domaine temporel et le domaine des ondelettes.

Les signaux de parole sont continus mais nous avons travaillons sur un signal discret (de taille N), la transformée en ondelettes se calcule aux échelles a^j, avec $a= 2^{(1/v)}$.

De plus la transformée en odelette de f ne pourra être calculée que pour les échelles :

$$\frac{1}{N} < a \leq 1 \qquad \text{VI.7}$$

IV.2.2.3 La transformée en ondelettes discrète :

Le traitement du signal basé sur les ondelettes a été utilisé avec succès pour des problèmes très variés, comme la compression d'image, la reconnaissance automatique de la parole, etc....

L'utilisation des ondelettes permet de faire une analyse multi - résolution du signal. Nous verrons l'intérêt de ce type d'analyse dans le cadre de détection de pitch pour un signal parole mais tout d'abord, définissons de la transformée en ondelettes discrète.

Soit un signal *f*(t) échantillonné uniformément sur [0,1] avec un pas d'échantillonnage de 1/N, on obtient un signal discret *f*[n]=*f*(n/N) composé de N échantillons.

Soit ψ(t) une ondelettes en temps continue dont le support est inclus dans |-K/2, K/2], pour $2 \leq a^j \leq \frac{N}{K}$, on définit une ondelettes discrète dilatée par a^j [42] :

$$\psi_j[n] = \frac{1}{\sqrt{a^j}} \psi \left(\frac{n}{\sqrt{a^j}} \right) \qquad \text{VI.8}$$

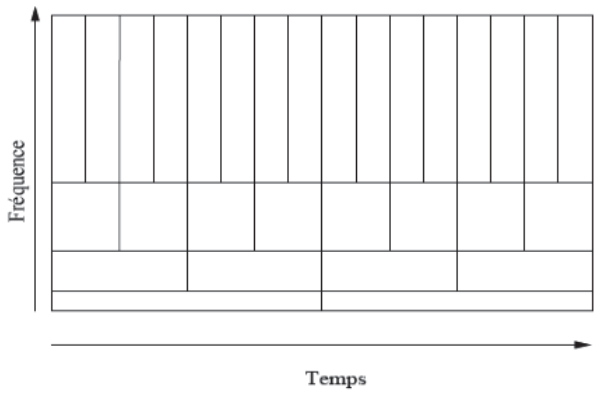


Figure VI.4 décomposition temps fréquence du signal. Une décomposition dyadique appliquée à la fois sure l'axe du temps et l'axe des fréquences.

L'échelle a^j doit être supérieure à 2 pour que le pas d'échantillonnage soit plus petit que le support de l'ondelette.

La transformée en ondelette discrète peut alors s'écrire comme une convolution circulaire avec $\overline{\psi}_j[n] = \psi_j^*[-n]$ [42]:

$$Wf[n,a^j] = \sum_{m=0}^{N-1} f[m]\psi_j^*[m-n] = f \otimes \overline{\psi_j}[n] \qquad \text{VI.9}$$

Où ψ* est le conjugué complexe de ψ et ⊗ l'opérateur de convolution circulaire. Si nous prenons le cas où l'échelle est découpé selon une suite dyadique $\{ 2^j \}_{j \in Z}$, c'est-à-dire lorsque le paramètre d'échelle est $a^j = 2^j$, alors la transformée en odelettes discrète et dyadique s'écrit [42]:

$$Wf[n,2^j] = \sum_{m=0}^{N-1} f[m]\psi_{2^j}^*[m-n] = f \otimes \overline{\psi_{2^j}}[n] \qquad \text{VI.10}$$

Avec :

$$\psi_{2^j}[n] = \frac{1}{\sqrt{2^j}}\psi\left(\frac{n}{2^j}\right) \qquad \text{VI.11}$$

La figure VI.4 montre la décomposition temps – fréquence du signal en utilisant la transformée dyadique d'ondelette. La transformée dyadique de '*f* 'ne peut être calculé que pour des échelles $1 > 2^j > 1/N$. la valeur absolue de **j** utilisé par la suite pour représenter les différentes échelles dans l'analyse multi-résolution ainsi que les différentes bandes de fréquence.

L'utilisation de la transformée dyadique d'ondelette nous permet d'obtenir une partition dyadique du plan temps fréquence de telle sorte que les basses fréquences sont représentées avec une haute résolution fréquentielle et une faible résolution temporelle et vice versa.

VI.2.2.4 Algorithme rapide pour la transformée en ondelettes :

Il a été montré que les coefficients de la décomposition du signal sur une base orthonormée d'ondelettes se calculent par un algorithme rapide (algorithme pyramidal) qui cascade des convolutions discrètes avec des filtres passe bas (G) et passe-haut (H) dont les sorties sont sous-échantillonnées.

Pour un signal parole qui est notre cas, les coefficients de décomposition du signal par la transformée dyadique d'ondelette sont obtenus par filtrage successif passe-haut (H) et passe bas (G) de la sortie du filtre passe bas (G). Les sorties sont sous échantillonnées par un facteur 2, l'algorithme est illustré à la figure VI.5 [41].

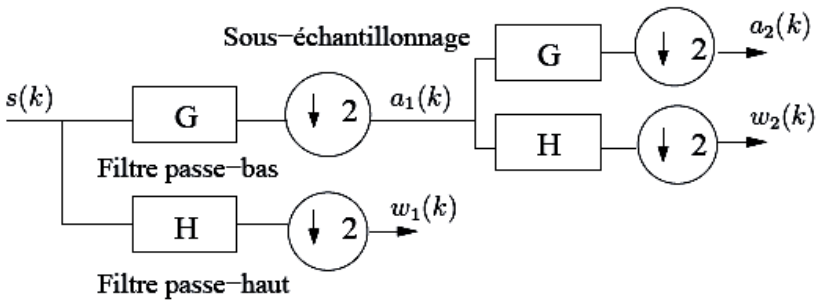


Figure VI.5 Transformée dyadique d'ondelette avec 2 niveaux de décomposition.

Le symbole «2» correspond au sous - échantillonnage par un facteur 2. La figure VI.5 montre qu'à chaque niveau de décomposition j, le signal est décomposé en coefficients *d'approximation* $a^j(k)$ (la sortie du filtre passe bas) et en coefficients de détailles $w_j(k)$ (la sortie du filtre passe-haut H)

Les coefficients d'approximation correspondent à des *moyennes locales* du signal tandis que les coefficients de détaille, aussi appelé *coefficient d'ondelettes,* dépeignent les différences entre deux moyennes locales successives, c'est-à-dire entre deux approximations successives du signal. Les coefficients d'approximation donnent une représentation lissée du signal et les coefficients d'ondelettes de détail nous donnent les détails (le bruit).

On peut reconstruire le signal de départ à partir de ces coefficients d'approximation et de détails [36], la localisation temporelle des fréquences n'et pas perdue et c'est là un autre avantage de la transformée en ondelettes pour une détection de pitch.

VI.2.3 Types d'ondelettes qu'on peut utiliser :

Il existe un nombre très important de types d'ondelettes que l'on appelle aussi famille, parmi lesquels on cite, les Coiflet , les Symlet, les ondelettes Daubechies, les ondelettes bi-orthogoales, l'ondelette de Haar, etc.

Lors de notre étude, nous nous sommes limités à des familles d'ondelettes bien connues en traitement du signal : les ondelettes Daubechies , les Symlet et les coiflets , ces ondelettes sont toutes admissible, selon le théorème 1, car de moyenne nulle et à décroissance rapide . De plus elles ont déjà été étudiée en quelque sujets comme Segmentation parole/musique, reconnaissance de la parole, et donnent de bons résultats [36] ainsi elles ont toutes la propriété d'avoir un *support minimum* pour un nombre de *moment nuls* donné.

- Ces deux caractéristiques généralement sont toujours prises en compte dans le choix d'ondelettes.

a. Les moments nuls :

Le nombre de moments nuls d'une ondelette s'exprime de la manière suivante [36]:

$$\int_{-\infty}^{+\infty} t^k \psi(t) dt = 0 \qquad \text{Pour } 0 \le k < p. \qquad \text{VI.12}$$

Si une ondelette Ψ vérifie cette équation alors on dit que l'ondelette Ψ a *p* moments nuls, cela signifie que Ψ est orthogonale à tout polynôme de degré *p*-1. L'intérêt d'avoir p moments nuls est d'obtenir des coefficients d'ondelettes w_j proche de 0 aux échelles fines 2^j (lorsque 2^j tend vers 0). En effet, si *f* (t) est localement de classe C^k alors *f (t)* est localement bien « approximé » par un polynôme de Taylor de degré k, et si k>p alors les ondelettes seront orthogonales à ce polynôme, la transformée en ondelettes aura donc des valeurs proches de 0, au contraire, quand *f*(t) ne pourra être approximé correctement que par des polynômes de degrés supérieur à p, alors la transformée en ondelettes aura de fortes amplitude. Cette propriété est très utile pour détecter les transitions brutales.

En effet les zones stationnaires d'un signal correspondront à de petits coefficients d'ondelettes et les transitions brutales à de grands coefficients [36].

b. Taille du support :

Si *f*(t) a une singularité isolé en t0, et si t0 est dans le support de l'ondelette Ψ_j, alors la transformée en ondelettes aura des coefficients d'ondelettes de forte amplitude autour de t0.

Si l'ondelette Ψ a un support de taille k, alors à haute fréquence, c'est-à-dire aux fies échelles : lorsque l'échelle 'a' tend vers 0, il y aura K ondelettes Ψ_j dont le support contiendra t0. L'idée est de minimiser la taille du support de Ψ dans le but de diminuer le nombre de coefficients d'ondelettes de grande amplitude. Cela permet ainsi de faire de la détection de singularité.

- Ces deux caractéristiques ne sont pas indépendantes, en effet, la taille de support et le nombre de moment nuls d'une ondelette orthogonale ne sont liés par le fait que si Ψ a *p* moments nuls approprié de choisir une ondelette ayant de nombreux moments nuls afin d'obtenir un grand nombre de coefficients d'ondelettes de petite amplitude. lorsque la densité de singularités augmente, il vaut mieux diminuer la taille du support, quitter à avoir moins de moment nuls.
- En effet les ondelettes dont le support passe par une singularité donnent des coefficients de grande amplitude [36].
- Pour le choix des odelettes, il faut aussi noter qu'en utilisant la transformée en ondelettes discrète, il est préférable de n'utiliser que des ondelettes à filtres. En effet seule les ondelettes à filtre peuvent être utilisé avec la transformée discrète, alors que dans le cas continue n'importe quelle fonction d'intégrale nulle convient [36].

VI.2.3.1 Les ondelettes de Daubechies :

Cette famille d'ondelettes a été créée par Ingrid Daubechies , on note les ondelettes de cette famille ***dbN*** ou *N* est l'ordre de l'odelette ; dans cette famille on trouve l'ondelettes de ***Haar*** qui correspond au **db1** et qui est la plus simple et certainement la plus ancienne des ondelettes. Exemptée db1, les ondelettes de cette famille n'ont pas d'expression explicite, cette famille possède certaines propriétés intéressantes ; le nombre de moment nuls de l'ondelette *dbN* est *N*, les ondelettes de Daubechies ont un support de taille minimale pour un nombre de moments nuls donné ainsi sont très asymétrique en particulier en faible valeur, sauf pour *db1*.Figure IV.6 présente des exemples d'ondelettes Daubechies (db2, db4, db8).

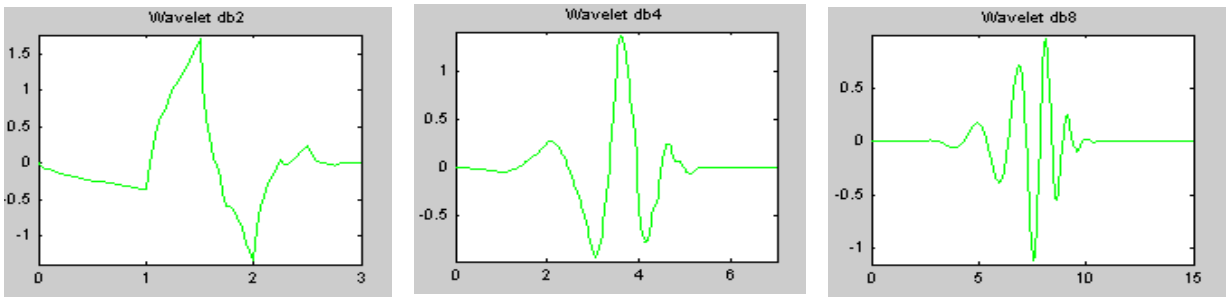


Figure VI.6 Exemples d'ondelettes Daubechies : de gauche à droite 'db2','db4','db8'.

VI.2.3.2 Les odelettes de Symlet :

Les symlets, notées *symN*, ont été proposé par Daubechies en modifiant la construction des ondelettes dbN et constituant une famille d'ondelettes presque symétrie. A part la

symétrie, les propriétés de ces deux familles sont similaires, En regardant des figures des ondelettes de Daubetchies et les symlets, nous pouvons constater que la Symlet ressemble à une odelette de Daubetchies pour un nombre de moment nuls petit, et qu'elle est plus symétrique. Figure VI.7 présente des exemples d'ondelette Symlet (Sym2, Sym4, Sym8).

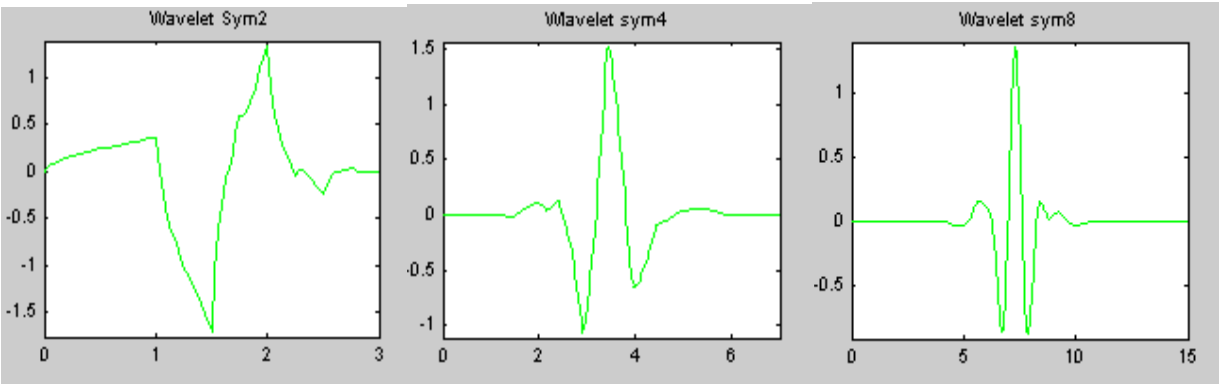


Figure VI.7 Exemples d'ondelettes Symlet : de gauche à droite 'sym2','sym4','sym8'.

VI.2.3.3 Les ondelettes Coiflets :

Les coiflets, comme les symlets, ont été construit par Daubechies, Elles sont crée sur demande de R.coifman [43] pour une application liée à l'analyse numérique, on prend comme notation de cette famille d'ondelettes : *coif N* .

Cette famille d'ondelette est différente des deux précédentes, ici, l'ondelette coif N aura 2N moments nuls, les Coiflets comme on peut le voir sur la figure 8 sont bien plus symétrique que les Symlets ou les ondelette de Daubechies .

L'intérêt principal des Coiflets réside dans le fait que si nous analysons une fonction *f* assez régulière, alors les coefficients d'approximation (pour un nombre de niveau de décomposition assez grand) correspondent à l'échantillonnage de *f*. Figure VI.8 présente des exemples d'ondelette Coiflet (Coif2, Coif4, Coif8).

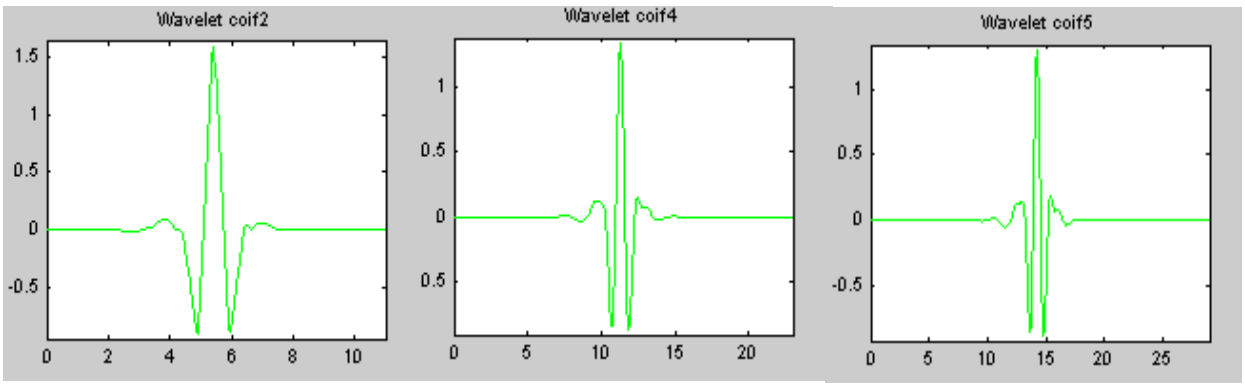


Figure VI.8 Exemples d'ondelettes Coiflet: de gauche à droite 'coif2','coif4','coif5'.

VI.2.4. Les Types d'énergie calculable sur les coefficients d'ondelettes :

L'énergie souvent utilisé en tant que paramètre et donne de bons résultats plus, l'énergie peut être calculé sur les coefficients d'ondelettes obtenus à partir de la transformée en ondelettes Dyadique, comme paramètre à utiliser.

Si w_j^k dénote le coefficient d'ondelettes à la position temporelle k et à la bande de fréquence j ; on rappelle que les décompositions temporelles et en bandes de fréquences suivant une échelle dyadique, c'est-à-dire que la résolution temporelle est divisée par deux alors que la résolution fréquentielle double à chaque niveau de décomposition.

Le nombre de coefficients dans la bande j est noté Nj. Alors et à partir des coefficients d'ondelettes $w^j{}_k$ pour la bande de fréquence j, différents paramètres *fi* pour cette bande de fréquences j en utilisant différents types d'énergie [36] :

- L'énergie instantanée (notée **E**) :

Ce type d'énergie classiquement utilisé dans le domaine de la parole, nous donne la distribution de l'énergie dans chacune des bandes [36] :

$$f_i = \log 10\left(\frac{1}{N_j}\sum_{k=0}^{N_j-1}(w_k^j)^2\right) \qquad \text{VI.13}$$

- L'énergie de Teager (notée **T_E**) :

C'est l'opérateur discret d'énergie de Teager introduit par Kaiser [36].cet opérateur permet de suivre les modulations d'énergie et donne une meilleur représentation de l'information formantique du signal dans le vecteur paramètre, ainsi il permet une réduction de bruit en utilisant sa capacité de suivi de la modulation d'énergie [36].

$$f_i = \log 10\left(\frac{1}{N_j - 1}\sum_{k=0}^{N_j-1}\left|(w_k^j)^2 - wj_{k-1}w_{k+1}^j\right|\right) \qquad \text{VI.14}$$

VI.2.5 Motivation :

Puisque un signal parole est par essence non stationnaire, la nécessité d'une analyse temps –fréquence à été reconnue de longue date alors, dans ce chapitre on a exploré des notions de base des ondelettes et différentes familles qu'on peut utiliser en prenant en compte les caractéristiques de chaque famille d'ondelette.

Nous avons vu que pour analyser des signaux stationnaires, une solution consiste à utiliser une variante de la transformée de Fourier classique : la transformée de Fourier à fenêtre aussi appelée transformée de Fourier à court terme. Cependant, cette résolution a des limites, notamment dans le choix de la taille de la fenêtre d'analyse qui détermine si nous concentrons

sur une analyse fréquentielle du signal ou sur une analyse des événements temporels. Une réponse intéressante à ce problème est la transformée en odelettes qui fournit une **résolution variable.**

On peut effectuer une analyse multi-résolution du signal et ainsi étudier plus finement les détailles du signal en l'observant différentes échelles tout en respectant le principe d'incertitude de Heisenberg.

La décomposition de signal parole à laide de la transformée en odelettes discrète et dyadique est basé sur l'utilisation de filtre passe haut et passe bas en cascade, alors le signal est décomposé en coefficients d'approximation et en coefficients de détails, plus couramment appelés coefficients d'ondelettes.

VI.2.6. Conclusion :

On peut exploiter les caractéristiques de la transformée en ondelettes pour une estimation de la fréquence fondamentale ou le pitch en bénéficiant l'analyse multi - résolution du signal en l'observant différentes échelles tout en respectant le principe d'incertitude de Heisenberg

VI.3 Détection de pitch par ondelettes :

L'analyse temps -fréquence et l'analyse temps-échelle ont été développées pour répondre à un besoin de mise en évidence de phénomènes très localisés en temps et en fréquence.

La Transformée en Ondelettes (TO) possède des propriétés de "*zoom*" qui en fait un outil idéal pour la détection de phénomènes hautes fréquences de courte durée. Rappelons que la TO possède une bonne résolution temporelle aux hautes fréquences et inversement.

La TO utilise une famille de fonctions d'analyses, construite par dilatation/compression et translation d'une fonction appelée *ondelettes mère*.

Dans notre étude, nous avons choisi différents ondelettes appartiens aux différentes familles, nous focalisons nos efforts sur les transformées en ondelettes continues (TOC) et discrète.

VI.3.1 Choix d'ondelettes :

Le but essentiel de notre expérience est de sélectionner les ondelettes les plus adéquates pour les différentes taches de détection du pitch.

Il existe de nombreuses familles d'ondelettes. Mais on s'est limité aux ***ondelettes orthogonales*** (des ondelettes à base de banc de filtres).

On a étudié trois familles d'ondelettes les plus connus et les plus utilisées en traitement du signal : les ondelettes de Daubechies, Les Symlets et les Coiflets que nous avons décrit précédemment.

On doit donc étudié le comportement des ondelettes de ces trois familles aux propriétés de lissage différentes, en faisant varier le nombre de moments nuls de ces ondelettes, En effet nous pouvons voir les ondelettes comme un outil pour lisser un signal.

VI.4 Méthodes de détection de pitch :

On limite les méthodes de détection de pitch basées sur les ondelettes sur deux méthodes en utilisant les différentes familles d'ondelette.

VI.4.1 Détection de pitch basé sur les maximums des coefficients de TOC :

Les coefficients de TOC (transformée en ondelette contenue) sont calculés par la convolution d'un signal $f(t)$ avec une ondelette, d'échelle a et au temps τ. La TOC se reporte souvent par un vecteur de tous les coefficients pour une échelle donnée. Les différentes ondelettes avec différents échelles dont sont illustrées dans les figures VI.6, VI.7, VI.8 fournirent des propriétés différentes dans le signal.

Un grand montant d'information redondante est chiffré dans le TOC parce que les coefficients ne changent pas considérablement sur petits changements dans le petit temps ou échelle, l'information est extraite habituellement uniquement par *des maximums du Coefficients* TOC [43].

Lorsqu'on procède de traiter de grands signaux ou des signaux multidimensionnels, de calculer les données redondantes peut conduire aux difficultés pour le processus de calculs. L'algorithme de la transformée d'ondelette rapide (FWT) [44] est inspiré par l'algorithme de pyramide (voire figure VI.5) élimine la redondance à travers l'orthogonalité [36]. Cela implique que toute information à chiffrer par la première ondelette n'est pas chiffré par la seconde et vice versa.

Le FWT utilise l'échelle dyadique, l'idée générale derrière l'algorithme de FWT est représenté par la figure VI.5, la procédure de décomposition se continue jusqu'au niveau (n) désiré ou s'arrête de la raison qu'on n'a pas des informations à extraie. Pour réduire le bruit on tient compte des coefficients d'approximation(Ai) et omets les coefficients de détaille(Di). Le sous échantillonnage par deux est procédé pour chaque niveau ce qui signifie la réduction de la longueur des coefficients. La reconstruction d'un signal 'S' nécessite les coefficients détaille Di et le dernier coefficient d'approximation An [43] :

$$S=\sum_{i}^{n} D_i + A_n \qquad \text{VI.15}$$

Où : 'A_i'les coefficients d'approximation.

'D_i' les coefficients de détaille.

La figure VI.9 présente un exemple dans Matlab d'un signal du doppler bruité a décomposé en 5 niveaux en utilisant une ondelette symlet4. Le 'S' (gauche) est le signal original. La décomposition du signal est montrée de bas en haut. Le S (droite) est le signal reconstruit, identique à l'original, et le 'cfs' est une visualisation des coefficients, l'échelle sur l'axe vertical et le temps sur l'axe horizontal. Les coefficients correspondent à la contribution des 2^i échelle d'ondelette.

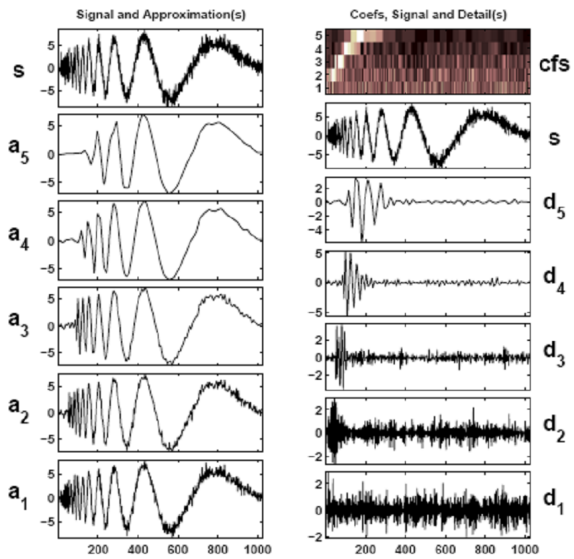


Figure VI.9 un exemple de décomposition d'un signal Doppler en 5 niveaux en présence de bruit.

VI.4.1.1 le choix de la forme d'ondelette :

Il est important d'éclairer que la forme d'ondelette est le facteur qui définit le choix d'ondelette. Certain type d'ondelette sont très utile pour extraire le pitch. Mallât [43] a prouvé que l'analyse avec une ondelette d'ordre1(le premier dérivative) donne des maximums au point de changement de signal. Comme vue au figure IV.10 qui présente une ondelette de Haar, celle ci est une ondelette Daubechies d'ordre1 (le premier dérivé), on remarque qu'elle est positif coté droite de centre et négative coté gauche. La TOC avec cette forme accentue le passage par zéro. Alors on peut utiliser les maximums de TOC pour identifier le passage par zéro dans le signal (chaque maximum correspond à un passage par zéro [43]) originale et

calculer les fréquences correspondent. C'est plus utile à choisir des maximums dans les Coefficients de TOC que le passage par zéro dans le signal original,

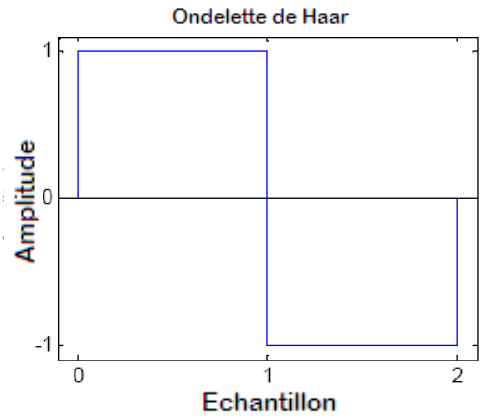


Figure VI.10 ondelette de Haar.

VI.4.1.2 Méthode d'estimation de pitch :

On résume la procédure d'estimation de pitch dans les étapes suivantes :

1- Choisir une trame de 30ms.
2- Exécuter la TOC pour chaque échelle chacun, en suite extraire les coefficients.
3- Extraire les maximums des coefficients pour chaque échelle.
4- Estimer la fréquence entre chaque paire de maximums pour chaque échelle.
5- Calculer la moyenne des fréquences pour chaque échelle qui signifie la fréquence fondamentale.
6- Grouper les fréquences fondamentales en fonction de temps calculées dans un graphe.

En principe, l'analyse devrait être exécutée avec une ondelette sur toutes les échelles dyadiques. Chaque échelle donne différentes fréquences. Alors il ya des hauteurs qui limite le nombre d'échelle à considéré dans l'analyse, *Fitch* et *Shabana* [45] suggèrent que l'analyse peut se faire sur trois échelles adjacent pour le cas d'une guitare, *Valerie Perrier* [46] suggèrent que le nombre d'échelles à considérer pour l'analyse d'un signal parole par ondelettes est se calcule par :

$$a = 2.(\log(n) / \log(2)) \qquad \text{VI.16}$$

Où : n est le nombre d'échantillons pour le signal à analyser.

On note qu'après calcul, on prend le nombre entier de 'a'.

Après avoir utilisé le premier dérivé d'ondelette pour calculer les coefficients, et les

fondamentales pour chaque échelle, la question qu'on pose est qu'elle fréquence qu'on adopte pour cette tranche de signal vocale comme pitch ? (1)

L'énergie d'un signal est la valeur de la somme carrée de ses échantillons. Dans ce cas et intuitivement l'énergie est la somme de carré des coefficients pour chaque échelle, on développe la relation suivante :

$$E(a) = \sum_{i}^{n} \left| c_i \right|^2 \qquad \text{VI.17}$$

Où : a : échelle, a=1,2….

E(a) : l'énergie qui correspond à un échelle a.

C : coefficient d'ondelette.

Une fois les maximums dans les coefficients sont identifiés, la période entre chacun des paires adjacentes est utilisée pour calculer la fréquence à ce point dans le temps. Alors pour chaque échelle on a beaucoup de paires par suite plusieurs fréquences, on doit estimer la fréquence fondamentale, pour une mesure de facilité on estime la fréquence fondamentale comme la fréquence entre le premier paires des deux maximums.

IV.4.1.3.1 Choix de la fréquence fondamentale :

On répond à la question (1) : après calcul des fréquences fondamentale pour chaque échelle, la fréquence fondamentale qu'on adopte comme pitch pour cette tranche de signal est la fréquence pour la quelle l'énergie des coefficients (équation IV.16) soit maximum.

VI.4.1.3.2 Estimation de la fréquence fondamentale :

La fréquence fondamentale pour chaque échelle est définit comme suit :

$$F_0 = \frac{Fs}{m_1 - m_2} \quad [Hz] \qquad \text{VI.18}$$

F_0 : fréquence fondamentale.

m_1 : l'échantillon qui correspond au premier maximum de première paire.

m_2 : l'échantillon qui correspond au deuxième maximum de première paire.

Fs : fréquence d'échantillonnage.

IV.4.1.4 le débruitage :

Le débruitage est la réduction de bruit pour un signal. On éclaire qu'on peut réduire le bruit pour un signal en appliquant la transformée d'ondelette. La figure VI.9 nous montre que les

successives approximations (a1 …..a5) apparait moins bruité. Or progressivement perd les informations concernant les hautes fréquences. Si on remarque l'approximation A5 on estime que 20% [43] de signal est perdue. La figure VI.11 présente le signal originale (en rouge) superposé sur le signal débruité (jaune).

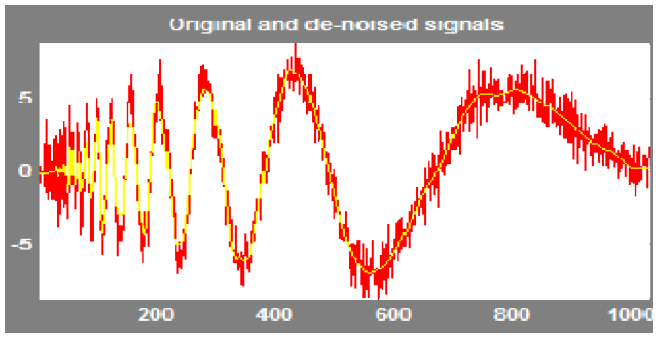


Figure VI.11 le signal bruité en rouge et le signal sans bruit en jaune en appliquant la TOC. (Amplitude vs échantillon)

On a deux méthodes pour appliquer le débruitage, un débruitage *doux* et un débruitage *dur*. Concernant notre étude, et à fin d'éviter de perdre l'information contenue dans le signal on a choisit le débruitage doux.

VI.4.1.5 décision voisé/non voisé :

À cause de l'efficacité et la simplicité de la méthode EZR, on a choisit cette dernière.

VI.4.1.6 Evaluation de la méthode et résultats expérimentaux :

a- Estimation de pitch par ondelette de Daubechies :

On prend une tranche d'un signal vocale « a » de 30ms, on estime la fréquence fondamentale pour chaque échelle, l'ondelette utilisé est l'ondelette Daubechies (db1) ; on a choisit a=1, 2, 3, 4, 5, 6. La figure VI.12 nous montre les coefficients d'ondelette pour l'échelle a=6, où : max1=53, max2=123, Fs =11025 :

F0=Fs/(max2-max1), la fréquence fondamentale est égale à :F0=11025/(123-53)=157,5 Hz.

Le tableau VI.1 présente les résultats obtenus pour chaque échelle de a=1…6, l'énergie maximale dans ce cas est égale à *62.2827* ce qui correspond à F0=157.5Hz, cette fréquence est considérée comme pitch pour cette tranche de signal. La figure VI.13 présente l'évolution de pitch superposé sur la forme d'énergie, où le maximum de la courbe d'énergie correspond au pitch pendant 30ms pour a=1,…30.

La Figure VI.14 présente le contour de pitch d'un son masculin qui représente une phrase « Bonjour » en fonction des trames (chaque trame de 30ms).

b- Résultats par trois ondelettes « Daubechies », « Symlet », « Coifflet » :

On prend une tranche de 30ms d'un son voisé pour extraire et estimer la fréquence fondamentale et voire le contour de pitch par trois ondelettes « Daubechies1, Symlet1, Coifflet 1» pour un objectif de comparaison. Figure VI.15 présente l'évolution de pitch par les trois ondelettes. Le tableau VI.2 présente les valeurs de pitch calculés par les trois ondelettes d'analyse.

c- Influence de bruit sur l'estimation de pitch :

On prend une tranche de « a » (sans bruit), on applique l'algorithme d'estimation de pitch par transformée d'ondelette (Daubechies 1) en présence de différentes puissances de bruit. Le tableau VI.3 présente le pitch en fonction de puissance de bruit (SNR).

VI.4.1.6 Discussion :

Cette méthode nous montre une bonne résolution temps fréquence, pendant une durée de 30ms, on voie une transcription de pitch en fonction d'échelle. La fréquence fondamentale parmi les fréquences calculées pour chaque échelle est choisit d'une façon que cette fréquence correspond à une maximum d'énergie. L'énergie se calcul pour chaque échelle par le carré des coefficients, cette notion d'énergie est inspiré de carrée des échantillons.

On remarque que le pitch estimé par les ondelettes de mères Daubechies1, Coifflet1, Symlet1 donnent des résultats semblable. Les résultats d'estimation de pitch en présence de bruit est généralement bonnes. Le contour de pitch pour un signal en fonction des trames est bien claire (Figure VI.14).

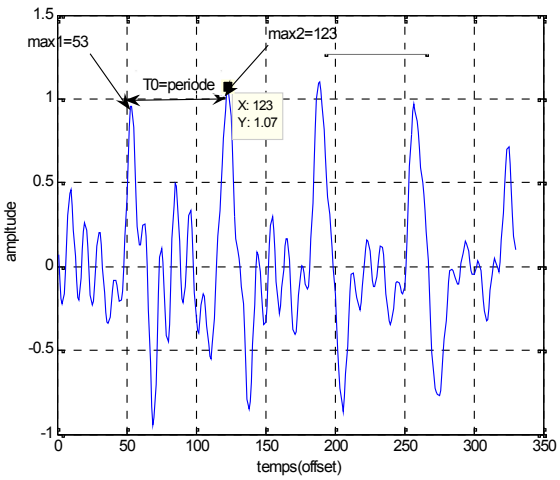


Figure VI.12 Coefficients d'ondelette en fonction des échantillons pour l'échelle a=6, Fo=157.5 Hz

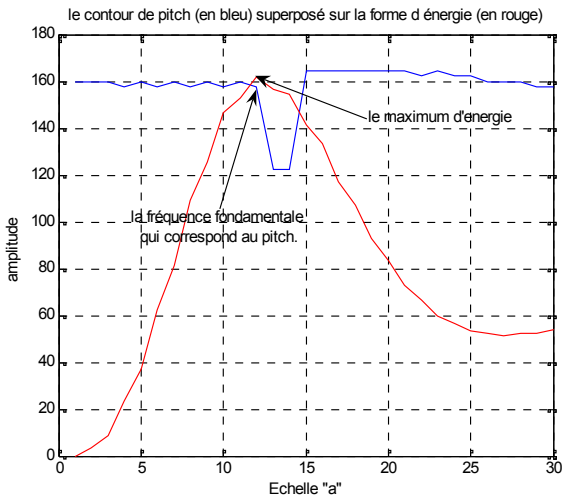


Figure VI.13 l'évolution de pitch en bleu superposé sur la forme d'énergie, a=1,…..30.

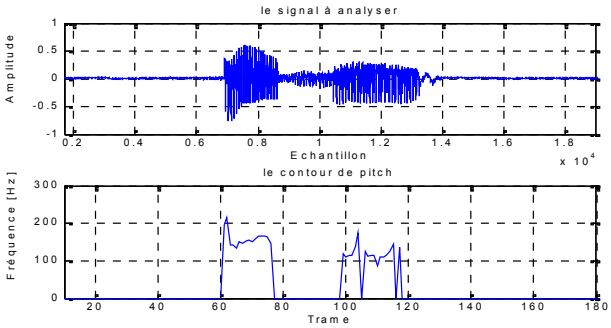


Figure VI.14 Signal « Bonjour » en haus et son contour de pitch en bas par application de TOC de Daubechies 1.

échelle	Max 1	Max 2	Energie	F_0 [Hz]
a=1	73	142	0.00001	159.7826
a=2	74	143	3.5654	159.7826
a=3	74	143	8.4630	159.7826
a=4	74	144	23.3658	157.500
a=5	74	143	37.1512	159.7826
a=6	74	**144**	62.2827	**157.500**

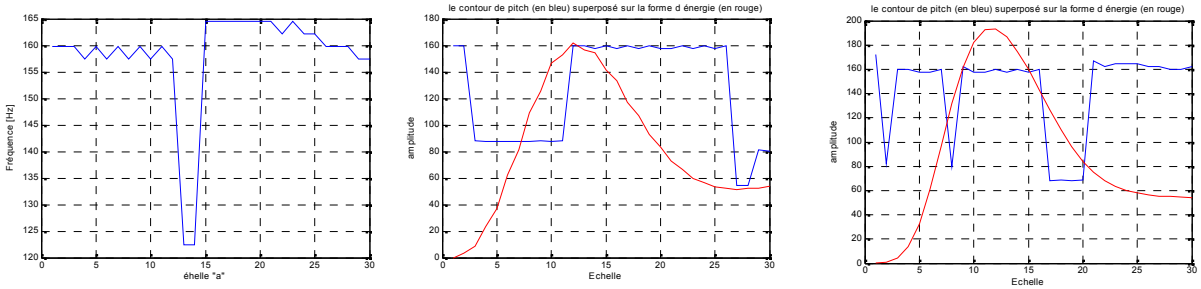
Tableau VI.1 Fréquences fondamentale en fonction d'échelles.

Ondelette d'ordre 1	pitch [Hz]
Daiubetchies	157.5000
Coifflet	159.7826
Symlet	159.7826

Tableau VI.2 Le Pitch par trois ondelettes : Daubechies, Symlet , Cofflet

Le pitch [Hz]	SNR [dB]
177.25	Sans bruit
177.25	30
177.25	20
169.62	4
15325	0
168.54	-3
170.31	-5

Tableau VI.3 Le pitch en fonction de puissance de bruit (SNR)



(a) (b) (c)

Figure VI.15 L' évolution de pitch pour une tranche de signal ''a'' par: (a) Daubetchies 1, (b) symlet 1 avec forme d'énergie en rouge, (c) Coiflet 1 avec forme d'énergie en rouge.

VI.4.1.7 Conclusion :

La première motivation d'utiliser les ondelettes est la bonne résolution temps fréquence, l'idée de base de cet algorithme est que, pour une ondelette convenablement choisit, la transformée d'ondelette expose les maximums locaux aux points de variation du signal ce qui nous permet d'exploiter ça pour une détection et une estimation de pitch, l'estimation de pitch est basé sur l'énergie calculée sur les coefficients d'ondelette. Les résultats d'estimation de pitch par les trois mères d'ondelettes donnent presque les mêmes résultats concernant l'estimation de pitch. On a une bonne résistance à la présence de bruit. On peut deviner que

l'inconvénient essentiel de cette méthode est l'erreur d'octave doublement de pitch à cause que l'estimation de pitch est basé sur deux maximums adjacent.

VI.4.2 Détection de pitch en temps réel basée sur les Ondelettes discrètes :

Cette méthode est développée par *Eric Larson* et *Ross Maddox* [47] [48]. Une méthode de détection de pitch en utilisant les ondelettes basée sur l'algorithme de FLWT (développée sous MATLAB). L'algorithme de FLWT utilise l'ondelette de Haar pour la détection de pitch, la transformée d'ondelette de Haar, comme elle est montré mathématiquement, c'est le sous échantillonnage et la décomposition de signal en coefficients de détail et coefficients d'approximation.

Les approximations sont utilisées en combinaison avec une découverte de maximum qui représente le pitch d'un son voisé.

Cet algorithme a été testé sur des signaux naturels et synthétiques pour savoir les caractéristiques et les performances de la méthode. Les hauteurs dans [47] [48] attendent toujours des erreurs pour la décision voisé / non voisée.

De notre part on redéveloppe cette méthode en intégrant notre algorithme de décision V/NV (EZR) pour un éventuel bon résultat après exécution d'algorithme originale. .

VI.4.2.1. Méthode :

Si on analyse un signal visuellement périodique, c'est facile de voir la périodicité. On peut facilement d'utiliser une simple logique pour extraite la valeur de période. Pour le cas d'uns signal quasiment périodique, on prend une tranche fenêtré pour extraire la valeur de pitch mais c'est difficile d'évaluer ou mesurer la valeur de la période.

A. le FLWT (Fast Lifting Wavelet Transform):

L'algorithme utilise une mise en œuvre du (FLWT). La transformée d'ondelette décompose en composants d'approximation et détail. La réduction de bruit est une étape importante pour détecter la périodicité, alors les coefficients d'approximations peuvent soumettre à des transformations pour réduire le bruit afin d'extraire le pitch qui est normalement plus difficile à extraire du signal original. Bien qu'il y ait plusieurs différentes ondelettes de mère peuvent être utilisées pour exécuter le FLWT. Cet algorithme utilise le FLWT qui utilise l'ondelette de Haar. Le FLWT qui utilise l'ondelette de Haar est mathématiquement équivalent à un filtre passe bas et un sous échantillonnage pour produire les composante d'approximations et exécuter un filtre passe haut et un sous échantillonnage

pour produire les composants de détails. Les équations dérivées précédemment par Debauchies pour le FLWT avec l'ondelette de Haar sont [47] :

$$\begin{aligned} d_0(n) &= x(2n+1). \\ a_0(n) &= x(2n). \\ d_1(n) &= d_0(n) - a_0(n). \\ a_1(n) &= d_0(n) + d_1(n). \end{aligned} \qquad \text{VI.19}$$

Où : x(n) est le signal originale, a_1 est le premier coefficient d'approximation et d_1 est le premier coefficient de détail. L'équation IV.19 équivalent à :

$$\begin{aligned} d_1(n) &= x(2n+1) - x(2n). \\ a_1(n) &= \frac{x(2n+1) + x(2n)}{2}. \end{aligned} \qquad \text{VI.20}$$

De ces équations c'est claire que le composant de l'approximation est simplement une application d'un filtre faisant la moyenne (un filtre passe bas d'ordre 1) avec sous échantillonnage; et le composant du détail est une application d'un premier filtre de la différence avec sous échantillonnage. Dans ce chemin, le FLWT fournit une méthode rapide de décomposer le signal.

La décomposition de signal peut révéler la périodicité sous-jacente. En abandonnant le composant de détail et exécutant une autre fois le FLWT sur le composant d'approximation, les niveaux supplémentaires de transformée d'ondelette sont à produire. L'application répétée (mais a limité) du FLWT dans cette manière idéalement donne une détection de pitch qui est robuste au bruit et une décision voisée/non voisée. Depuis la gamme de la fréquence désirée d'opération est d'approximativement 50Hz et 500Hz l'usage de plusieurs opérations de filtre passe bas et le sous échantillonnage est nécessaire pour détecter le pitch.

A.1 limite de décomposition (niveau de décomposition):

Notez qu'avec chaque transformée d'ondelette, le nombre d'échantillons dans l'approximation est divisé en deux et le signal soit décomposé (approximation, détail); cela nous impose de limiter le nombre de niveau d'ondelette qui peuvent être exécutés sur toute fenêtre donnée d'un signal. Les hauteurs dans [47][48]limitent le niveau de la transformée d'ondelette à *6 niveaux*. Notez la réduction de bruit et simplification (lissage) du signal avec chaque approximation consécutive. La figure VI.16 représente le signal original, et le premier / deuxième / troisièmes approximations d'ondelette.

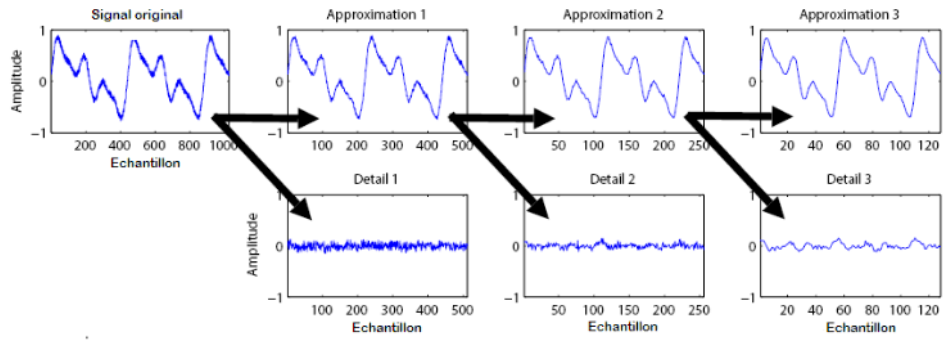


Figure VI.16 le signal original, et le premier / deuxième / troisièmes approximations d'ondelette.

B. Algorithme :

Fenêtrer le signal (fenêtre Hamming) en plusieurs segments de 25ms (N segments):

1. Au niveau 'i' pour la transformée d'ondelette (Haar) des données, on procède:

a) Enlever le composant DC(le composant contenue).

b) Trouver tous le passage par zéro.

c) Trouver en premier le maximum local après chaque passage de zéro.

d) Calculer les distances (en échantillons) entre les maximums locaux.

e) Répétition b), c), d) pour les minimums locaux.

f) Déterminer la moyenne du mode distance.

2. Si la distance du mode au niveau '**i'** égal à la distance du mode au niveau **(i-1)**, on assume que la distance de mode de (i-1) est égale à la période de la fondamentale (pitch); Sinon, aller au prochain niveau de la transformée d'ondelette (jusqu'à i = 6 [47]) et revenir à l'étape 1.

Alors la fréquence fondamentale ou le pitch est égale [47,48]:

$$F_0 = \frac{Fs}{\text{mode}\ (i-1).2^{(i-2)}} \qquad \text{VI.21}$$

A1. Décision voisée/non voisée :

Si la limite de niveau de décomposition (i=6) est dépassée, on suppose que la tranche de signal est non voisée et on passe à la prochaine fenêtre (segment).

A2. Différence des maximums/minimums :

Le processus de détection des maximums implique plusieurs pas. La moyenne d'amplitude est calculée et à soustraire de chaque élément du signal fenêtré, en enlevant le composant DC (le contenue). La valeur de DC est considérée comme un seuil inférieur pour les maximums et un seuil supérieur pour les minimums. Aussi, la distance entre deux

maximums doit être supérieur où égale àδ . La distance minimum δ dépend de la fréquence maximale de pitch (F= 500Hz) 'F' et le niveau de la décomposition d'ondelette en cours ' i', ainsi la fréquence d'échantillonnage Fs [47,48]:

$$\delta = \max\left(\left[\frac{Fs}{2^{i}F}\right],1\right) \qquad \text{VI.22}$$

A.3 Pourquoi la moyenne de mode distance ? :

La moyenne distance a été employé pour améliorer l'estimation de mode distance et par suite améliorer l'exactitude de l'algorithme notamment pour les fréquences supérieures où une petite variation du nombre entier dans la distance provoque une grande variation dans la fréquence. Notez que dans les hauts fréquences, la moyenne améliore l'exactitude, réduire des erreurs par approximativement 100 Hz [48,47]. La figure VI.17 présente une représentation visuelle de cette amélioration, ces expériences faites aussi par *Eric Larson* et *Ross Maddox* [47,48] où un signal sinusoïdal (de fréquence constante dans chaque fenêtre) a été construit, en variant de 200-1400 Hz. Sans faire la moyenne, le détecteur du pitch fait de grosses erreurs sur l'ordre de 100 Hz autour de 1400 Hz.

VI.4.2.2 Résultats et discussion :

On applique notre algorithme sur une tranche d'un phonème « b » prononcée par un masculin, on note que cette tranche est de durée de 30ms avec un Fs=11025Hz, et fréquence maximale de pitch de 500Hz, alors on calcul la distance minimum δ et le DC (la composante contenue de la tranche de signal) :

DC=0.0095 ; $\delta = \max\left(\frac{11025}{2^{1}.500},1\right)$ δ =11.025 ≈ 11 échantillon. Cette distance minimale est valable pour le premier signal d'approximation.

$\delta = \max\left(\frac{11025}{2^{2}.500},1\right)$ δ =5.5125 ≈ 5 échantillon (pour le deuxième signal d'approximation).

Calculs de modes distances pour le premier signal d'approximation (i=1) :

max1=7, max2= 48, max3= 90, max4=132.

min1=2, min2=41, min3=83, min4=124.

1- Pour les maximums :

Mod1= (48-7)=41 échantillons. Mode 2= (90-48)= 42. Mode 3= (132-90)= 42.

2- Pour les minimums :

Mod1=(41-2)=39 échantillons. Mode 2= (83-41)= 42. Mode 3= (124-83)= 41.

Calculs de la moyenne de modes distances :

Moyenne mode = (41+42+42+39+42+41)/6. **Moyenne mode = 41.1667 échantillons.**

Calculs de modes distances pour le deuxième signal d'approximation (i=2):

max1=5, max2= 25, max3= 46, max4=67

min1=2, min2=22, min3=42, min4=61.

1- Pour les maximums :

Mod1=(25-5)=20 échantillons. Mode 2= (46-25)= 21. Mode 3= (67-46)= 21.

2- Pour les minimums :

Mod1=(22-2)=20 échantillons. Mode 2= (42-22)= 20. Mode 3= (61-42)= 19.

Calculs de la moyenne de modes distances **:**

Moyenne mode = (20+21+21+20+20+19)/6. **Moyenne mode = 20.1667 échantillons.**

Comparaison entre modes distances d'approximations 1 et 2 :

Moyenne mode = 41.1667 (approximation 1) pour Fs/2

Moyenne mode = 20.1667 (approximation 2) pour Fs/4 alors : Moyenne mode =20.1667*2= 40.3334 (approximation 2).

À partir de ces deux résultats on constate que les deux modes presque égaux ce qui signifie qu'on arrête la décomposition de signal au niveau 2. Le pitch est égale à :

Pitch = Fs/(41.1667.2) = 11025/(41.1667.2)= 133.9068Hz

Le pitch = 133.9068Hz.

Figure VI.18 représente le signal original bruité.

Figure VI.19 représente le signal original et le DC qui est considéré comme seuil.

Figure VI.20 représente le signal d'approximation 1 (i=1), et le DC.les maximums et les minimums de signal.

Figure VI.21 exprime le signal d'approximation 2 (i=2).

Figure VI.22 exprime le signal d'approximation, le DC, les maximums et les minimums de signal.

Figure VI.23 exprime le signal d'approximation 2, la distance entre les maximums (pour calculer le mode de distance) ainsi la distance entre les minimums.

La première remarque est que l'utilisation de la moyenne a gardé l'analyse corrigée partout dans la gamme (200-1400 [Hz]) justifiée par la figure VI.17 (droite).

La deuxième remarque qu'on peut extraire l'exactitude d'estimation de pitch justifié par les résultats présentés par le tableau VI.4 où on a appliqué l'algorithme sur un signal sinusoïdale (cosinus) avec différentes fréquences. L'erreur maximale (fréquence mesurée-fréquence actuel) était approximativement de 0.6 Hz dans la gamme de 60-500 Hz.

La troisième remarque est la plus importante dans notre analyse est la résolution efficace en temps et en fréquence. Cette méthode de détection de pitch a même bonne résolution dans les basses fréquences (dû à son utilisation de méthodes du domaine temporelle) et dans les haut-fréquences (dû à l'utilisation de la méthode de la moyenne de mode distance). Cette résolution permet la détection du pitch exacte comme présenté dans le tableau IV.4.

La quatrième remarque est que le contour de pitch est moyennement clair (n'est pas mauvaise) à cause de la décision V/NV, les hauteurs dans [47,48] déclarent la limitation de cette méthode de décision. La figure VI.24 présente le contour de pitch pour un signal « égale » prononcé par un masculin.

La cinquième remarque qu'on peut constater l'immunité au bruit. Le tableau IV.5 présente le résultat d'estimation de pitch d'un son voisé « a » sans bruit où on le noyer dans différents puissance de bruit.

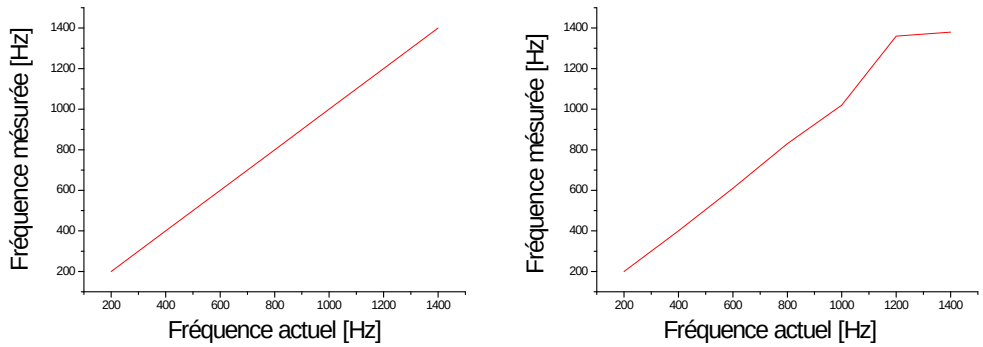


Figure VI.17 Une comparaison de l'exactitude du pitch sur le signale sinusoïdale (cosinus) avec le mode qui fait la moyenne (coté gauche) (adopté dans notre étude), et sans application de la moyenne (coté droite).

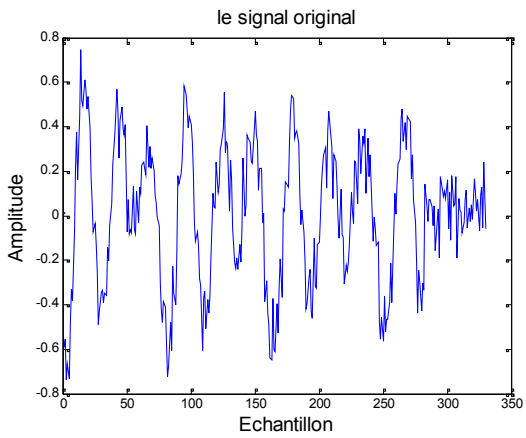


Figure VI.18 Le signal original bruité.

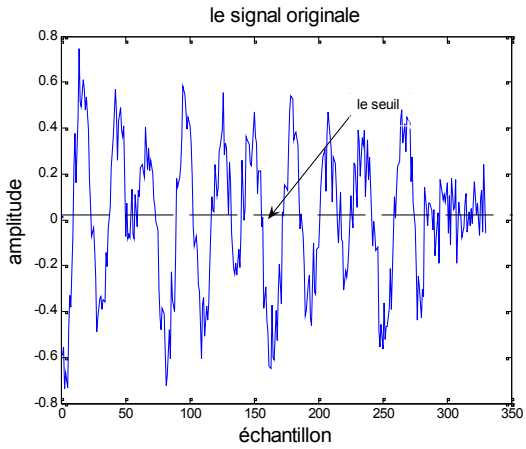


Figure VI.19 Le signal original et DC.

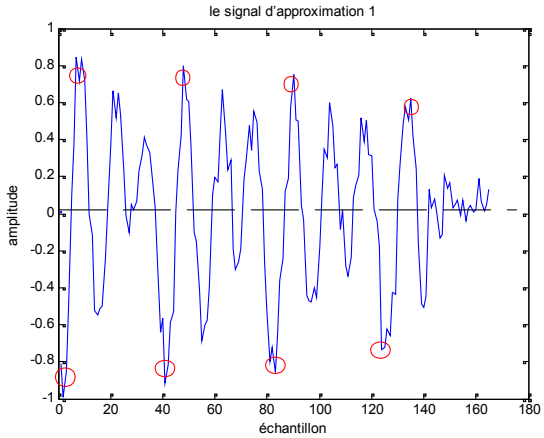


Figure VI.20 Le signal d'approximation , Le DC, les maximas, minimas pour calculer des modes distance.

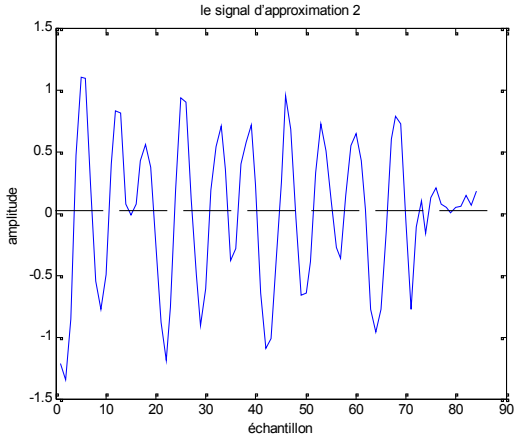


Figure VI.21 Le signal d'approximation2 et le DC.

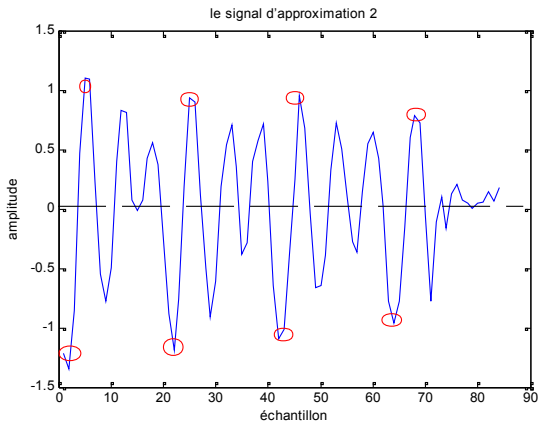


Figure VI.22 Le signal d'approximation 2, le DC , les maximas, minimas, pour calculer les modes distance.

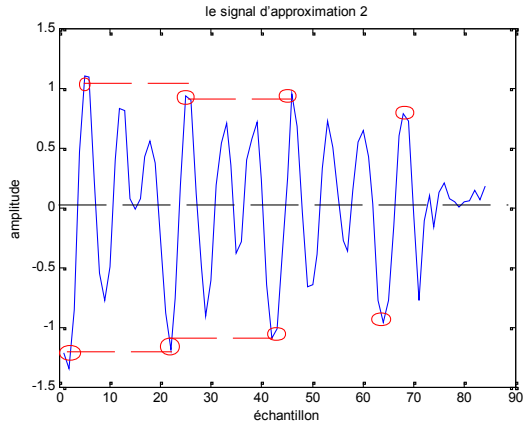


Figure VI.23 Le signal d'approximation2 et comparaison des modes distance avec celui de niveau 1.

Fréquence actuel [Hz]	Fréquence mesurée [Hz]	(Fréq mésurée)-(Fréq actual) [Hz]
60	60.2459	0.2459
80	79.8913	0.1087
110	110.2500	0.2500
150	149.6606	0.3394
200	200.0504	0.0504
250	250.5682	0.5682
300	299.5924	0.4076
350	350.3708	0.3708
460	459.3750	0.6250

Tableau VI.4 Fréquences actuels, fréquences mesurées, et l'erreur d'estimation.

SNR [dB]	Pitch [Hz]
Sans bruit	148.4135
30	148.4135
20	148.4135
10	148.4135
4	150.4532
0	152.6534
-4	170.4325

Tableau VI.5 L'estimation de pitch en fonction de SNR.

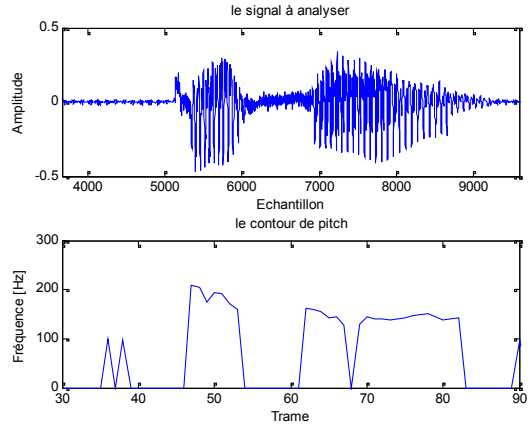


Figure VI.24 Le signal « égale » et son contour de pitch

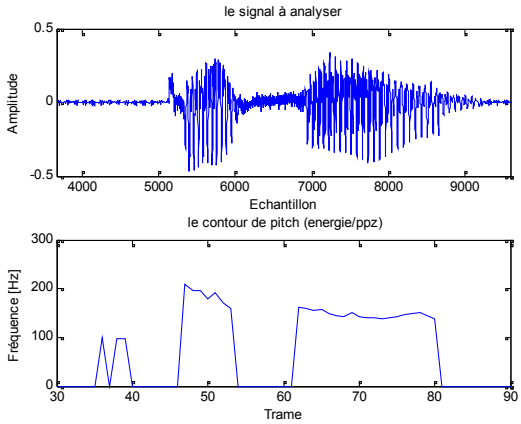


Figure VI.25 Le signal « égale » et son contour de pitch par EZR

VI.4.2.3 Amélioration :

On note que l'estimation de temps en temps réel d'un tel algorithme est de 4ms, où cette estimation est faite par *Eric Larson* et *Ross Maddox*. Alors de notre part on propose d'utiliser l'algorithme d'EZR en temps réel qui conduit à un temps moins de 4ms. La figure VI.25 exprime le résultat d'un signal « égale » avec EZR où on peut aisément conclure l'efficacité en comparaison avec celle de résultat représenté dans la figure VI.24 sans EZR.

VI.4.2.4 Conclusion :

L'objectif essentiel de cette dernière méthode est l'implantation en temps réel d'un algorithme à fin d'extraire le pitch avec son contour en temps réduit. L'approche est basée sur l'utilisation d'une ondelette simple et détection intelligentes des piques pour calculer la fréquence fondamentale ou le pitch.

Après analyse de cette méthode on constate une efficacité concernant :

1- Un pitch plus exact.
2- Capable de donner une bonne résolution en temps et fréquence.
3- Une décision voisé / non voisé exacte sans seuils d'amplitude globaux.
4- Robuste au bruit grâce à la décomposition du signal (approximations, détailles).

VI.5 Conclusion :

Dans ce chapitre on a analysé deux méthodes basées sur les transformées en ondelettes.

La première méthode est basée sur la détection des maximums des coefficients d'ondelette qui correspond physiquement à la variation de signal parole. L'inconvénient essentielle de cette méthodes et de tomber dans les erreurs de doublement ou inversement dans l'estimation de la fréquence fondamentale.

La deuxième méthode se caractérise par la bonne résolution en temps et en fréquence qui reflète sur l'estimation la plus exacte de pitch pour les fréquences inférieurs et supérieurs, l'ajout d'EZR donne une bonne efficacité concernant la décision V/NV et le temps d'exécution réduit.

Bibliographie :

[41] Ronald L. Allen. Duncan W. Mills« signal analysis time frequency, scale, and structure » A John Wiley & Sons, Inc., Publication, IEE press. 2004.
[42] D.Jonathan, B.Michael, F.S´ebastien « les ondelettes ». Universit´e Libre de Bruxelles Facult´e des SciencesD´epartement de Physique.2002.
[43] John McCullough « UsingWavelets for Monophonic Pitch » Computer Science Department ,Harvey Mudd College, USA. HMC-CS-2005-01 September, 2005
[44]*Michel Misiti,Yves Misiti ,Georges Oppenheim,Jean-Michel Poggi. « Wavelet Toolbox »,for use with Matlab*, september 2000 by The MathWorks.
[45] W.Shabana and J.Fitch « a wavelet-based pitch detector for musical signals ». Department of Mathematical Sciences, University of Bath,Bath BA2 7AY, UK.2000.
[46] Valerie Perrier-PROGRAMME DE TRANSFORMEE EN ONDELETTES. Institut nationale polytechnique de Grenoble. LMD novembre 1993.
[47] Eric Larson , Ross Maddox « Real-Time Time-Domain Pitch Tracking Using Wavelets » Departments of Mathematics, Physics and Philosophy, Kalamazoo College, Center for Performing Arts Technology University of Michigan School of Music.USA.2004.
[48] Eric Larson« Real-Time Time-Domain Pitch Tracking Using Wavelets » A paper submitted in partial fulfillment Of the requirements for the degree of Bachelor of Arts at Kalamazoo College. USA. 2005.

VII.1 Introduction :

On procède à une étude comparative des méthodes classiques, temps-fréquence, et temps-échelle. Le choix d'un tel algorithme de détection de la hauteur des sons (pitch) dépend de plusieurs critères. On en cite : la précision d'estimer la période du pitch, la précision de la décision V/NV, le contour du pitch; la robustesse vis-à-vis de la présence du bruit de toute nature, la vitesse d'opération, la complexité de l'algorithme.

On note que toutes les méthodes au cours de notre analyse, on a utilisé l'algorithme d'EZR comme méthode de décision V/NV qui engendre un temps réduit et un contour de pitch bien claire au cours de recherche de la fréquence fondamentale.

VII.2 Etude comparatives des méthodes vis-à-vis la précision de calcul de pitch:

Dés que l'exactitude de pitch est un paramètre important, on exécute tous les algorithmes des méthodes classiques, temps-fréquences, temps-échelle étudiés dans les chapitres précédents par un signal sinusoïdal « cosinus » de fréquence constante (60, 80, 110, 150, 200, 250, 300, 350).

Le tableau VII.1 représente les fréquences fondamentales mesurées par les méthodes classiques en fonction des fréquences fondamentales actuelles.

Le tableau VII.2 représente les fréquences fondamentales mesurées par les méthodes temps fréquence, temps échelle en fonction des fréquences fondamentales actuelles.

VII.3 Etude comparatives des méthodes vis-à-vis la présence de bruit:

Dans cette étape de comparaison on a enregistré un signal sans bruit qui exprime un phonème « a » d'une voix masculine, par suite on effectue la recherche de la fréquence fondamentale en présence de différents puissance de bruit par chaque méthode : classiques, temps- fréquence et temps-échelle.

Le tableau VII.3 représente les fréquences fondamentales mesurées par les méthodes classiques en fonction de rapport signal sur bruit (SNR).

Le tableau VII.4 représente les fréquences fondamentales mesurées par les méthodes temps fréquence, temps échelle en fonction de rapport signal sur bruit.

VII.4 Résultats expérimentaux :

Soit les résultats expérimentaux représentés dans les tableaux VII.1, VII.2, VII.3, VII.4.

Fréq actuel [Hz]	ACF [Hz]	ACF_LPC [Hz]	AMDF [Hz]	ASDF [Hz]	SIFT [Hz]	CEPS [Hz]	HPS [Hz]	SHS [Hz]
60	56.171	56.171	56.171	56.171	55.991	58.723	60.783	60.783
80	75.250	75.250	75.250	75.250	73.501	79.120	80.410	80.410
110	103.250	103.250	103.250	103.250	101.893	108.123	109.004	109.004
150	146.350	146.350	146.350	146.350	145.765	144.543	146.036	146.036
200	193.510	193.510	193.510	193.510	193.123	190.372	191.681	191.681
250	250.381	250.381	250.381	250.381	248.629	230.543	235.310	235.310
300	300.150	300.150	300.150	300.150	300.546	260.543	267.036	267.036
350	350.110	350.110	350.110	350.110	349.324	310.321	312.136	312.136

Tableau VII.1 les fréquences fondamentales mesurées par les méthodes classiques en fonction des fréquences fondamentales actuelles.

Fréquence actuel[Hz]	STFT [Hz]	DWV [Hz]	DPWV [Hz]	DPWVL [Hz]	DCW [Hz]	TOC [Hz]	TDO [Hz]
60	62.8182	59.2742	59.2742	59.2742	56.5611	80.4836	60.2459
80	83.9384	78.7500	78.7500	78.7500	78.2500	90.3689	79.8913
110	113.2512	108.0882	108.0882	108.0882	107.1931	110.2500	110.2500
150	142.5639	145.0658	145.0658	145.0658	144.0011	157.5000	149.6606
200	182.1895	190.0862	190.0862	190.0862	188.1410	200.4545	200.0504
250	226.5023	239.6739	239.6739	239.6739	240.0110	245.0000	250.5682
300	283.1279	266.3100	266.3100	266.3100	270.5430	297.9730	299.5924
350	311.4407	312.1361	312.1361	312.1361	300.321	350.1724	350.3708

Tableau VII.2 les fréquences fondamentales mesurées par les méthodes temps fréquence, temps échelle en fonction des fréquences fondamentales actuelles.

SNR [dB]	ACF [Hz]	ACF_LPC [Hz]	AMDF [Hz]	ASDF [Hz]	SIFT [Hz]	CEPS [Hz]	HPS [Hz]	SHS [Hz]
Sans bruit	131.2500	131.2500	131.2500	131.250	131.250	131.250	131.932	137.274
38	131.2500	131.2500	131.2500	131.250	131.250	131.250	131.932	130.830
27	131.2500	131.2500	131.2500	131.250	131.250	164.552	131.932	130.8300
18	131.2500	131.2500	139.5570	136.111	131.250	139.557	136.710	130.8300
14	131.2500	131.2500	NV	136.111	131.250	355.645	159.862	130.8300
6	131.2500	131.2500	NV	NV	153.125	355.645	135.607	130.8300
2	131.2500	131.2500	NV	NV	153.125	157.500	138.180	130.8300
-1	131.2500	131.2500	NV	NV	NV	136.111	138.180	150.7324

Tableau VII.3 les fréquences fondamentales mesurées par les méthodes classiques en fonction de rapport signal sur bruit (SNR)

SNR [dB]	STFT [Hz]	DWV [Hz]	DPWV [Hz]	DPWVL [Hz]	DCW [Hz]	TOC [Hz]	TDO [Hz]
Sans bruit	139.3206	132.1064	131.2500	131.2500	134.4512	133.1110	132.8313
38	139.3206	132.1064	131.2500	131.2500	134.4512	133.1110	132.8313
27	139.3206	132.1064	131.2500	131.2500	134.4512	13 3.1110	132.8313
18	139.3206	132. 1064	131.2500	131.2500	128.1800	133.1110	132.8313
14	139.3206	132. 1064	131.2500	131.2500	134.4512	133.1110	132.8313
6	120.0100	132. 1064	131.2500	131.2500	112.5000	139.326	134.5522
2	113.1410	132. 1064	131.2500	131.2500	114.8438	140.6012	139.3206
-1	83.5924	132. 1064	131.2500	131.2500	128.1977	213.5237	164.5522

Tableau VII.4 les fréquences fondamentales mesurées par les méthodes temps fréquence, temps échelle en fonction de rapport signal sur bruit

VII.5 Discussion et conclusion :

Concernant l'exactitude d'estimation de la fréquence fondamentale on voit que :

- Les méthodes temporelles généralement sont bien pour détecter les fréquences supérieurs, les méthodes fréquentielles sont bien pour détecter les fréquences inférieurs, ce qui est vérifiée par les résultats enregistrés dans le tableau VII.1.
- D'après le tableau VII.2 la méthode STFT semblable à une méthode fréquentielle ce qui traduit par la moins résolution temps-fréquence.
- Les méthodes DWV, DPWV, DPWVL donnent presque les mêmes résultats d'estimation de la fréquence fondamentale mais avec des propriétés semblable de celle des méthodes fréquentielles (Cepstre), cela nous faire penser aux algorithmes adopter pour l'estimation de pitch où on a utilisé la méthode de Cepstre pour détecter le maximum qui correspond au pitch. La méthode DCW semblable aussi à une méthode Cepsrale, cette dernière méthode n'ai pas recommandée où le résultat dépend du signal lui-même et un facteur de réduction d'interférences « σ » qui doit être changé suivant le signal à analyser (dans notre analyse on a choisit σ =3).
- La méthode TOC (transformée d'ondelette continue) basée sur la détection des maximums des coefficients est semblable aux méthodes temporelles.
- La méthode TDO (transformée d'ondelette discrète) donne une bonne estimation soit pour des fréquences moyennement supérieur ou inférieurs dont les erreurs d'estimation ne dépassent pas 0.23% ce qui se traduit par le bon algorithme adopté et la bonne résolution en temps et fréquence.

Concernant l'influence de bruit on voit que :

- Les méthodes ACf et ACF_LPC, très résistante au bruit, les deux méthodes AMDF, ASDF ces deux méthodes donnent presque toujours les mêmes résultats et simple concernant la facilité et la rapidité de calcul, mais sensible au bruit.
- Concernant la méthode SIFT l'inconvénient essentielle c'est la complexité du calcul, qui peut engendrer des erreurs de calcul.
- L'objectif essentiel d'utiliser la méthode de Cepstre c'est qu'elle permet de séparer facilement la source et le filtre. C'est la propriété fondamentale mais n'est pas résistante au bruit et peut donner des résultats faux.
- D'après ces résultats on peut constater que la méthode SIFT est performante par rapport aux autres méthodes mais l'inconvénient essentiel c'est la complexité du calcul.

- Les deux méthodes HPS et SHS donnent presque toujours les mêmes résultats mais avec une fiabilité de la méthode SHS vis-à-vis le bruit.
- La méthode STFT est généralement sensible au bruit, les méthodes DWV, PDWV, PDWVL ont une bonne résistance au bruit grâce aux améliorations ajoutées. la DCW est généralement nuisible au bruit.
- La méthode TOC donne une bonne résistance au bruit comme celle de la méthode TDO grâce à la décomposition d'ondelettes en coefficients d'approximation et de détaille.

Après cette étude comparative on peut constater que le choix de la méthode dépend de plusieurs critères parmi les, la nature de signal (masculin, féminin), la nécessitée de la bonne précision ou non, temps de calculs si on veut l'implantation en temps réel. Mais on pence que les meilleures méthodes sont les méthodes basées sur la décomposition en ondelettes.

Conclusion générale

Le signal vocal est très complexe, du fait de sa grande variabilité, ce qui rend toute tentative de le modéliser ou de reconnaître très délicate. Le signal de parole est un processus aléatoire non stationnaire à long terme.

Les méthodes temporelles permettent une estimation de la période du pitch avec un délai minimal, et des calculs très simples. Pour ces deux raisons, ce furent les premières à être utilisées. Les méthodes temporelles généralement sont bien pour détecter les fréquences supérieurs, les méthodes fréquentielles sont bien pour détecter les fréquences inférieurs.

La méthode simple classique et intuitive des méthodes temps-fréquence est la STFT où elle ne fournit pas une bonne résolution dont on doit toujours définir la largeur de la fenêtre à utiliser. La méthode DWV avec ses versions (DPWVL, DPWV) est plus connue que les autres méthodes est donne une bonne résolution, cette méthode est amélioré par certains auteurs par l'utilisation des transformées d'ondelette pour un objectif de filtrage et robustesse à la détection de pique responsable de pitch. L'utilisation de la moyenne des distributions DWV, DPWL, DPWV ont pour objectif de réduire les termes d'interférences. Les performances de DCW dépendent de signal lui-même. Au cours de notre analyse on a entré un filtrage passe bas à cours termes pour aider à réduire les harmonique ainsi les termes d'interférences.

Les méthodes temps-échelle en générale sont les plus commodes à l'analyse de signal parole, parmi ces méthodes qu'on a exploré, une méthode basée sur la détection des maximums des coefficients d'ondelette qui correspondent physiquement à la variation de signal parole , l'estimation de pitch est basé sur l'énergie calculée sur les coefficients d'ondelette, nous avons testé cette méthode avec différentes famille d'ondelette (Daubechies, coifelets, Symlets) où on a constaté que ces ondelettes donnent des résultats similaires, le nombre de moments nuls d'ondelette conditionne ça capacité d'extraire la fréquence fondamentale, les résultats avec les ondelettes d'ordre inferieur (on a choisit l'ordre égale à 1) sont bon, l'inconvénient essentielle de cette méthodes et de tomber dans les erreurs de

doublement ou inversement dans l'estimation de la fréquence fondamental, l'autre méthode se caractérise par la bonne résolution en temps et en fréquence qui reflète sur l'estimation la plus exacte de pitch pour les fréquences inférieurs et supérieurs, alors cette dernière méthode est recommandée à utiliser.

Enfin, dans ce travail on a traité le problème de détection de pitch nécessaire au traitement de la parole (reconnaissance automatique de a parole, discrimination parole/musique ….) en explorant plusieurs méthodes, alors comme perspective on traite les problèmes liées à la reconnaissance automatique de locuteur.

Bibliographie :

[1] R. Boite et all, « Traitement de la parole », PPUR, 2000.

[2] Damien Vincent. Thèse« Analyse et contrôle du signal glottique en synthèse de la parole» l'École Nationale Supérieure des Télécommunications de Bretagne 2007.

[3] LE Manh Tuan « Analyse des voyelles spéciale du Vitnamien ». Institut de la Francophonie pour l'Informatique En collaboration avec le Centre de Recherche MICA, Hanoi.2005

[4] S.addad. Mémoire magister « Décodage Acoustico Phonétique en vue de la reconnaissance des voyelles de l'arabe standard » .Ecole militaire polytechnique, Algerie.2001

[5] Barış Eker, These "Turkish text to speech system". Bilkent universIty. Turky 2002.

[6] Dr. Joseph Picone ,"Fundamentals of speech recognition" institute for signal and information processing". Mississippi State University. 1998.

[7] Bojan Kotnik1, Harald Höge, Zdravko Kacic1 "Evaluation of Pitch Detection Algorithms in Adverse Conditions", University of Maribor, Slovenia , Siemens AG, Corporate Technology, Germany 2006.

[8] M. Skowronski, "Biologically inspired noise-robust speech recognition for both man and machine ".Florida university,USA . 2004.

[9] Robert E. The Spectral Autocorrelation Peak Valley Ratio (SAPVR) – A Usable speechMeasure Employed as a Co-channel Detection System, (Article) Temple University USA.IEEE_WISP_2001_V5.

[10] Codage et décodage LPC de la parole, avril 2003.

[11] www.Winpitch.com.(2009).

[12] J P.Haton, J. M.Pierrel , GPereou,J Galelen, J. L.Gauvain, « Reconnaissance Automatique de la parole » France. 1991.

[13] Li Tan and Montri Karnjanadecha **«** Pitch détection algorithme: Autocorrélation méthod and amdf » Department of Computer Engineering Faculty of Engineering Prince of Songkhla University Hat Yai, Songkhla Thailand, 90112. 2003.

[14] « Overview of Homophonic Pitch Detection algorithms » Alexandre Savard Schulich School of Music - McGill University 555 Sherbrooke St. West Montreal, QC Canada H3A 1E3 . 2003

[15] A.Moinet & M.Tryhoen, « Implémetation d'un codeur LPC10 complet sous Matlab »**,** Faculté Polytechnique de Mons, Belgique, 9 Rue de Houdain, 7000 Mons, France.

[16] A. Ouhabi « Techniques avancées de traitement du signal et Applications »; Alger 1993.

[17] Shlomo Dubnov ." Non - Gaussian Source - Filter and Independent Components

Generalizations of Spectral Flatness Measure". Ben-Gurion University (ocupated Palastine). 2003.

[18] Shlomo Dubnov ."Generalization of Spectral Flatness Measure for Non-Gaussian Linear Processes ".". Ben-Gurion University (ocupated Palastine). 2003.

[19] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 2-8, Feb. 1976.

[20] Li Hui, Bei-qian Dai, Lu Wei « a pitch détection algorithme based on Amdf and Acf » MOE-Microsoft Key Laboratory of Multimedia Computing and Communication, University of Science and Technology of China 2003.

[21] J.MAX, D.Berthier, H.Chevalier, B.Escudie, A.Hellion, M.Martin, M.Trottot «Méthodes et technique de Traitement du signal »Deuxiemme édition MASSON, Paris, Neuw york, Bercelone, Milan. 1977.

[22] J. J. Dubnowski, R. W. Schafer, and L. R. Rabiner, "Real-timedigital hardware pitch detector," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, pp. 2-8, Feb. 1976.

[23] Michael.J.CHENG,Student "A Comparative Performance Study of Several Pitch Detection Algorithms". Lawrencer.Rabiner.Fellow, IEEE, 1970.

[24] Dr. Andrzej Drygajlo .École polytechnique fédérale de Lausanne ; EPFL - Faculté STI - ITS SCG Laboratoire de Traitement Numérique de la Parole ; Travaux pratiques B, 2003.

[25] Dr. Roland Badeau. http://www.perso.enst.fr , détection de hauteur. Notes de cours.

[26] C.alessandro, C.demars « Représentations temps-fréquence du signal de parole » LIMSI-CNRS, Université Paris VI,France.1992.

[27] J. JEONG, W . J . Wlliams "On the cross-terms in spectrograms proceedings". IEEE-ICASSP 1991, pp . 1565-1568. 1990.

[28] P.Flzndrin, B.Escudié. Principe et mise en œuvre de l'analyse temps fréquence par transformation de Wigner-Ville.1990.

[29]F. Auger,P. Flandrin, P. Gonçalvès,O. Lemoine. Tutorial"time frequency toolbox for use with Matlab".CNRS(France),Rice university (USA), 26 Octobre 2005.

[30] Boîte à outils temps-fréquence: **www-isis.enst.fr/TFTB.**

[31] Ronald L. Allen, Duncan W. Mills, «Signal anlysis : time, frequency, scale,and structure » IEE Press 2004.

[32] E. Chassande-Mottin : *Méthodes de réallocation dans le plan temps-fréquence pour l'analyse et le traitement de signaux non-stationnaires*. These de doctorat, Universitede Cergy-Pontoise, 1998.

[33] Lunji Qiu, Haiyun Yang and So0 Ngee Koh « A Fundamental Frequency Detector of Speech Signals Based on Short Time Fourier Transform » Nanyang Technological University Singapore. 1994

[34] **A** Spaargaren, MJ English « Detecting Ventricular Late Potentials using the Continuous Wavelet Transform » University of Sussex**,** Brighton, UK.1999.

[35] Sam Kwong', Wei Gang", and Chan H Lee **«** A Pitch Detection Algorithm Based on Time-Frequency Analysis » Institute of Electronic Engineering and Control,South China University of Technoloiy , Guangzho.1992.

[36] Emmanuel Didiot (These)« Segmentation parole/musique pour la transcription automatique de parole continue » université Henri Poincaré Nancy 1France .NOV 2007.

[37] *S.* Kwong, G. Wei, and J. Z. Ouyang, "Fundamental frequency estimation based on adaptive time averaging Wigner-Ville distribution," Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Victoria, British Columbia,Canada.Oct 1992.

[38] Z. Leonowicz, T. Lobos. «Analysis of three phase signal using Wigner spectrum ». Chair of the Theory of Electrical Engineering Department of Electrical Engineering Wroclaw University of Technology, Wroclaw, Poland

[39] R. Yu and E. C. Tan « Comparison of Different Time-Frequency Distributions in Pitch Detection » School of Computer Engineering, Nanyang Technological University Nanyang Avenue, Singapore 639798, Singapore,IEE 2003.

[40] Namhoon Kim Heungkyu Lee Hanseok Ko. « Reliable Pitch Period Estimation Based on Wavelet Transform and Choi-William Distribution «Dept. of Electronics Engineering, Korea University ».1995.

[41] Ronald L. Allen. Duncan W. Mills« signal analysis time frequency, scale, and structure » A John Wiley & Sons, Inc., Publication, IEE press. 2004.

[42] D.Jonathan, B.Michael, F.S´ebastien « les ondelettes ». Universit´e Libre de Bruxelles Facult´e des SciencesD´epartement de Physique.2002.

[43] John McCullough « UsingWavelets for Monophonic Pitch » Computer Science Department ,Harvey Mudd College, USA. HMC-CS-2005-01 September, 2005

[44]*Michel Misiti,Yves Misiti ,Georges Oppenheim,Jean-Michel Poggi. « Wavelet Toolbox »,for use with Matlab,* september 2000 by The MathWorks.

[45] W.Shabana and J.Fitch « a wavelet-based pitch detector for musical signals ». Department of Mathematical Sciences, University of Bath,Bath BA2 7AY, UK.2000.

[46] Valerie Perrier-PROGRAMME DE TRANSFORMEE EN ONDELETTES. Institut nationale polytechnique de Grenoble. LMD novembre 1993.

[47] Eric Larson , Ross Maddox « Real-Time Time-Domain Pitch Tracking Using Wavelets » Departments of Mathematics, Physics and Philosophy, Kalamazoo College, Center for Performing Arts Technology University of Michigan School of Music.USA.2004.

[48] Eric Larson« Real-Time Time-Domain Pitch Tracking Using Wavelets » A paper submitted in partial fulfillment Of the requirements for the degree of Bachelor of Arts at Kalamazoo College. USA. 2005.

ملخص

كشف الدور للكلام يعتبر ذا أهمية كبيرة في عدة مجالات نذكر منها المعرفة اللفظية، معرفة المتحدث، تشفير الكلام، تكوينه لكن تعيين الدور صعب بسبب تعقد إشارة الكلام هذا الدور يميز اللفظ فقط إذن فالمشكل هو كشف هذا الدور.

الكلام يمثل إشارة غير مستقرة من هنا يتضح أن تحويل "فوريي" لا يستطيع تمثيل هذا النوع من الإشارات. كما انه توجد عدة طرق قديمة من اجل كشف الدور منها "الترابط المثلي" والذي يعطي نتائج غير جيدة، من هنا تكون هناك حتمية البحث عن طرق أخري.

من بين الطرق الجديدة والتي تظهر جيدة من اجل هدف تمثيل إشارة الكلام و كشف الدور طريقة التمثيل في معلم "زمن- تردد" و "زمن-سلم"، إذا هدفنا هو كشف الدور و التمثيل بهذين الطريقتين.

المفاتيح: تردد الكلام، الترابط المثلي، التنبؤ الخطي زمن-تردد، زمن-سلم، النوافذ المتموجة.

Abstract

The detection of the period of the pitch from the vocal signal presents a considerable importance however in the vocal recognition, identification of the broadcaster, coding of the speech and its synthesis, the determination of the period of the pitch of the vocal signal is difficult following the complexity of the vocal signal that is considered for the voiced signal, a signal as a result of the exit of a variable system in the time excited by trains of impulses almost-magazine. The problem is therefore to determine the period of the signal of excitation of the voiced signal.

The vocal signal is part of the non stationary signals. it proves to be that the tool of the transformation of Fourier doesn't solve the representation of the non stationary signals. Among the classic methods of detection of the period of the pitch, the method of auto-corrélation that gives less effective results in this detection, from where the necessity of research of other shapes of representations that contributes to this type of signals better.

The analysis time - frequency and the analysis time-scale have been developed to answer a need of bets in evidence of phenomena very localized in time and in frequency.

The representations time-frequency and time-scale are the most convenient to the analysis of signal speech.

Key words: Pitch, auto-corrélation, linear prediction, cepstre, time-frequency, time-scale, wavelet.

Résumé

La détection du période du pitch à partir du signal vocal présente une importance considérable dans la reconnaissance, identification du speaker, codage de la parole et sa synthèse. Cependant la détermination de la période du pitch du signal vocal est difficile suite à la complexité du signal vocal qui est considéré comme pour le signal voisé, un signal issu de la sortie d'un système variant dans le temps excité par des trains d'impulsions quasi-périodique. Le problème est donc de déterminer la période du signal d'excitation du signal voisé.

Le signal vocal fait partie des signaux non stationnaires. Il s'avère que l'outil de la transformation de Fourier ne résout pas la représentation des signaux non stationnaires. Parmi les méthodes classiques de détection de la période du pitch, la méthode d'auto corrélation qui donne des résultats moins performants dans cette détection, d'où la nécessité de recherches d'autres formes de représentations qui contribuent mieux à ce type de signaux.

Notre travail consiste donc à étudier les différents types de représentations temps fréquence et temps échelle, et de les adapter mieux à la détection de la période du pitch du signal vocal.

Les représentations temps-fréquence et temps-échelle s'avèrent les plus commodes à l'analyse d'un signal parole.

Mots clés : Pitch, auto corrélation, prédiction linéaire, cepstre, temps-fréquence, temps-échelle, ondelettes.