

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR  
ET DE LA RECHERCHE SCIENTIFIQUE



الدرسة الوطنية المتعددة التقنيات  
المكتبة — BIBLIOTHEQUE  
Ecole Nationale Polytechnique

ECOLE NATIONALE POLYTECHNIQUE

Département d'Électronique

**Mémoire de fin d'études**

En vue de l'obtention du diplôme d'ingénieur d'État en  
Électronique

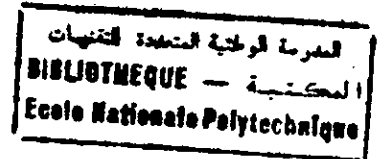
**Rehaussement de l'Intelligibilité de la Parole  
Hyperbare**

Dirigé par :  
M<sup>r</sup> D.BERKANI

Réalisé par :  
M<sup>r</sup> LOUAM Hamim  
M<sup>r</sup> BOUABSA Mehdi

*Promotion : 2002*

## *Avant propos*



*Nos vifs remerciements vont à notre  
promoteur*

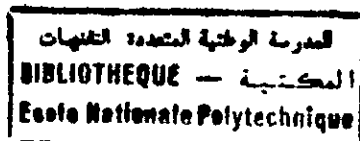
*Mr, Berkani , pour l'aide efficace  
qu'il nous a apporté , sa compréhension  
et sa patience qui ont permis la  
réalisation de ce travail.*

*On remercie également les  
examineurs qui ont accepté de juger  
ce travail.*

*Nos remerciements vont aussi à tous  
ceux qui , de près ou de loin , nous ont  
apporté leur aide et soutien dans la  
réalisation de ce mémoire ;*

*Enfin , on rends hommage à tous les  
enseignants qui ont contribué à notre  
formation depuis l'école primaire*

Dédicace  
Avant propos  
Liste des notations et abréviations utilisées



<b>CHAPITRE I</b>	<b>1</b>
<b>Introduction</b>	<b>1</b>
I.1 Problématique.....	2
I.2 Objet & contribution de cette thèse.....	3
I.3 Organisation de la thèse.....	3
<b>CHAPITRE II</b>	<b>5</b>
<b>le signal vocal</b>	<b>5</b>
II.1 Introduction.....	6
II.2 Production du Signal Vocal.....	6
II.3 La prédiction linéaire (LP).....	12
II.3.1 La Méthode d'Auto-corrélation.....	14
II.3.2 Les considération de choix des conditions d'analyse LP.....	16
II.3.3 La pré accentuation .....	18
II.4 Conclusion .....	19
<b>CHAPITRE III</b>	<b>20</b>
<b>Le signal Hyperbare</b>	<b>20</b>
III.1 Introduction.....	21
III.2 Comparaison entre les propriétés de la parole dans les conditions normales et hyperbares .....	21
III.3 Variation des fréquences des formants .....	25
III.4 Variation des largeurs de bandes des formants .....	27
III.5 Variation des amplitudes des sons voisés et non voisés .....	28
III.6 Les caractéristiques du signal Hyperbare .....	28
III.7 Conclusion .....	31

## Les Systèmes de traitement

IV.1	Introduction.....	33
IV.2	Les systèmes de traitement.....	33
IV.2.1	Les systèmes à approche temporelle.....	33
IV.2.1.a	Enregistrement/Reproduction ou "Tape Recorder Playback".....	33
IV.2.1.b	Le codage numérique des signaux "Digital Coding"....	34
IV.2.1.c	Traitement par signal analytique.....	34
IV.2.1.d	Traitement par Convolution.....	34
IV.2.2	Les systèmes à approche fréquentielle.....	35
IV.3	Considération pour la conception d'un système de traitement .....	37
IV.4	Correction des corrélations .....	37
IV.5	Correction spectrale .....	40
IV.6	Résultats.....	42
IV.6.1	Description des Fichiers de Données et Conditions d'Expérimentation.....	42
IV.6.2	Commentaires et Interprétations des résultats.....	44
IV.7	Conclusion.....	49

## CHAPITRE V

### Réduction de bruit d'un signal hyperbare

V.1	Introduction.....	50
V.2	Caractéristiques du bruit dans un milieu hyperbare.....	51
V.3	Réduction de bruit dans le domaine temporel.....	53
V.3.1	Pondération du résidu.....	53
V.3.2	La représentation SMC de la parole hyperbare bruitée.....	55
V.4	Réduction de bruit dans le domaine spectral.....	56
V.4.1	La soustraction spectrale.....	56
V.4.2	Estimation du spectre du bruit.....	59
V.4.3	Algorithme d'amélioration basé sur la soustraction spectrale...	60
V.5	Conclusion .....	62

**Conclusion générale** 63

**Annexes** 65

**Référence Bibliographique** 67

## Liste des Notations et Abréviations Utilisées

T ou k ou n	Temps , instant directe
Z.	Variables de la transformée en z
F	Fréquence
$\omega=2\pi f$	Fréquence angulaire
$\phi$	La phase
RSB(SNR)	Rapport signal sur bruit
STFT	Transformée de fourier à court de terme
DSP	Digital signal processor
$F_0$	Fréquence fondamentale ou pitch
$F_i (i= 1,2,3,...)$	Fréquences des formants
$F_c$	Fréquence de résonance du conduit vocal fermé aux lèvres
$F_s$	Fréquence d'échantillonnage
S(t), S(k)	L'évolution temporelle , séquence directe
S(w), S(n)	Le DFT de S(t) ou de S(k)
S(k)	Approximation de s(k)
W(k)	Fenêtre de hamming
U(k)	Terme d'excitation
E(k)	L'erreur de prédiction
A(k)	$K^{1\text{eme}}$ paramètre du modèle
a (k)	$K^{1\text{eme}}$ estimé de a(k)
$A_c(k)$	$K^{1\text{eme}}$ paramètre corrigé du modèle
R(i)	L'autocorrélation à court terme évaluée à l'instant i
$R_c(i)$	L'autocorrélation corrigée
K(i)	Coefficients de réflexion
E	Erreur quadratique moyenne
G	Facture du gain
P, m , nh	Ordre du modèle
N	Longueur de la fenêtre
AR	Auto- régressif
LP	Prédiction linéaire
LPC	Le codage prédictif linéaire
V/NV	Voisé ou non voisé
H(z)	Fonction de transfert du modèle ou filtre de synthèse
A(z)	Filtre inverse ou filtre d'analyse
P(z)	Predicteur linéaire
WLD	Wiener -levrinson-durbin
L	Longueur de conduit vocal
$\mu$	Facteur de pre-accentation
$H_e-O_2$	Mélange gazeux d'hélium d'oxygène
ATA	Atmosphère (bar)
D	Profondeur d'immersion en mètre (m)
$\alpha ,k$	Rapport des célérités ou coefficient de compression
P	Pression(ATA)
PPO <sub>2</sub>	Pression partielle d'oxygène
PO <sub>2</sub>	Proportion d'oxygène dans le mélange

$C_a, C_h$	كَلَرِيْتَة دُو سُون
$CP_a, CP_h$	كَلَرَة سَبْعِيْفِيَة اِلى پَرَسْيُون كُونَسْتَانْتَة
$CV_a, CV_h$	كَلَرَة سَبْعِيْفِيَة اِلى وَوَلَم كُونَسْتَانْت
$\mu_a, \mu_h$	كُوْفِيْفِيَنْت دِي وَوِيْصِيْتَة
$\lambda_a, \lambda_h$	كُوْفِيْفِيَنْت دِي كُونْدُوْكْتِيْبِيْلِيْتَة دِي الْكَلَرَة
$\gamma_a, \gamma_h$	كُونَسْتَانْتَة اِدْيَابَاتِيْقَة
$\rho_a, \rho_h$	دَنْسِيْتَة دُو گَاز
Spec-a(w)	اِنْوَلُوْپَة سَبْعْرَالَة دُو كُونْدُوْ وُكَال
Cspec-a(w)	اِنْوَلُوْپَة كُورِيْجِيْتَة
SMC	شُورْت تِيْم مِودِيْفِيْد كُوْهَرَنْس رِيْپْرِيْزَنْتَاشَن
VAD	دِيْتِكْتُوْر د'اِكْتِيْوِيْتَة وُكَال

## CHAPITRE I

# *Introduction*



## 1.1 Problématique :

Les techniques de communications sont arrivées avec l'avancée technologique à un progrès considérable plus précisément en ce qui concerne la transmission sonore, mais cela n'exclut pas les difficultés qui résident en la communication dans et entre milieux complètement différents. Nous donnons comme exemple des travaux effectués à des profondeurs sous-marines importantes en évoquant des transmissions entre plongeurs d'une part, et avec l'équipe qui travaille à la surface d'autre part.

Ainsi, les plongeurs professionnels qui réalisent des plongées à des profondeurs supérieures à 60m, sont amenés à évoluer dans une ambiance de haute pression, ce qui nous oblige à remplacer l'azote diluant de l'air, qui devient impropre à la respiration, par des diluants de masse moléculaire plus faible, le plus souvent de l'hélium ou un mélange d'oxygène d'hélium et d'hydrogène, ce qui forme des mélanges respiratoires synthétiques tels que : l'héliox, l'hydrox, l'hydréliox, ... Permettant ainsi de telle plongée dite de saturation et évite les effets très nocifs de l'azote en particulier la narcose, qui entraîne une sorte de dégradation mentale.

Cependant, les propriétés acoustiques produites par ces mélanges varient de manière radicale sous la pression sous-marine, cela explique la déformation de la parole chez les plongeurs inhalant de tels mélanges dans de tel environnement, cette déformation dite la parole sous haute pression ou la parole hyperbare ou encore la parole dans l'hélium. Cela a comme conséquence un décalage linéaire des fréquences de formant vers des hautes fréquences, c'est l'effet « de Donald le canard » .

Le problème de l'intelligibilité de la parole dans des conditions de plongée de l'eau est généralement dû aux effets combinés de deux phénomènes physiques :

- Les grandes pressions de profondeurs élevées imposent des contraintes au corps et aux mélanges du gaz inhalés par les plongeurs. La densité du gaz augmente et les plongeurs semblent parler dans un tube avec le diamètre décroissant. Ceci a comme conséquence une modification du tractus faisant pour être déplacé vers des hautes fréquences, et pour avoir des fréquences de formant de la parole de plus grandes largeurs de bande. Les premiers formants sont les plus affectés et les consonnes perdent une certaine énergie pendant que l'énergie aux hautes fréquences est réduite en condition hyperbare. L'effet est une voix aiguë.
- Le deuxième facteur de déformation est dû aux mélanges de gaz inhalés par les plongeurs. Le pourcentage élevé du contenu d'azote en air au niveau extérieur devient fortement narcotique une fois inhalé sous des hautes pressions, et a donc pour être remplacé par le gaz inerte d'hélium. Les fréquences de formant sont une fonction de la géométrie de tractus et de la vitesse du bruit. La vitesse du bruit en air est 340 m/s, tandis que dans l'hélium elle atteint 967 m/s.

Au début, plusieurs travaux de recherche ont été consacré à la caractérisation du signal hyperbare, ainsi que la modélisation du conduit vocal dans les deux milieux « AIR » et « HELIUM ».

Il a été constaté que l'amplitude, la largeur de bande des formants, et les fréquences des trois premiers formants sont les paramètres qui ont une variation appréciable dans l'hélium. De plus, l'élocution des plongeurs est plus lente, en plus de l'invariance du pitch confère au signal hyperbare une robuste quasi-périodicité et quasi-stationnarité ; d'où la pertinence de l'utilisation du modèle AR. Ce résultat est d'un intérêt immédiat pour l'élaboration des méthodes de restitution de l'intelligibilité de la parole hyperbare, basées sur la prédiction linéaire et la transformée de Fourier (DFT).

Dans plusieurs cas, le signal hyperbare est contaminé par un bruit de niveau élevé ( $\approx 0$  dB ; il peut masquer même le signal utile) surtout la gamme des faibles fréquences. Ce bruit peut être la combinaison d'un bruit de respiration (breathing noise), d'un bruit de machines (machinery noise), et le bruit des océans (ocean noise). De plus, la chambre ou le masque utilisés dans la plongée sous-marine affectent le milieu qui devient réverbérant. En fin, le microphone utilisé pour la transmission sous-marine, est un facteur essentiel, qui peut être une source importante des pertes d'information du signal hyperbare. Le choix d'un microphone de qualité supérieure, qui possède une large bande passante et une réduction importante du bruit, permet de palier cet inconvénient.

Le problème du rehaussement de la parole hyperbare intéresse plusieurs organismes, en particulier : les centres qui exploitent les océans ; la marine, pour l'amélioration des communications spécifiques aux militaires ; les compagnies d'exploitation pétrolières ;...

## **I.2 Objet & contribution de cette thèse :**

L'objet de cette thèse est l'étude de quelques algorithmes d'amélioration de la qualité de la parole hyperbare, basés sur l'utilisation de la prédiction linéaire. Le premier but du système de rehaussement est d'augmenter l'intelligibilité de la parole hyperbare. Le deuxième but est d'améliorer la qualité de cette dernière, ce but inclut d'autres objectifs, à savoir, la minimisation du bruit de fond et le maintien du naturel de la parole synthétisée. Les algorithmes proposés doivent profiter des caractéristiques du signal hyperbare, de l'analyse LPC, afin de mettre en œuvre des algorithmes de traitement, avec des variantes plus ou moins complexes, capables de travailler en temps réel et les plus adaptés à une éventuelle implémentation sur une carte DSP.

## **I.3 Organisation de la thèse :**

Notre travail consiste en l'étude et la mise au point par simulation numérique des méthodes de rehaussement de l'intelligibilité de la parole hyperbare, ainsi que l'amélioration de la qualité de cette parole dans un milieu très hostile.

Le deuxième chapitre contient une étude du signal vocal émis dans un milieu dont les conditions sont normales, précédé par les définitions des propriétés acoustiques temporelles et spectrales de ce milieu, mais aussi en présentant la modification du système vocal dans sa forme la plus simple en utilisant la méthode de prédiction linéaire (LPC), qui sera utilisée dans tous les algorithmes de rehaussement que nous développerons par la suite.

Nous consacrerons le troisième chapitre au signal hyperbare, une étude comparative entre les propriétés acoustiques de la parole dans les conditions normales et hyperbares a été présentée, en illustrant par plusieurs simulations numériques afin de consolider les études théoriques effectuées auparavant. Enfin, nous avons mis en évidence le comportement des largeurs et les fréquences des formants, l'amplitude des sons voisés et non voisés, le pitch, et l'excitation glottale dans l'hélium par rapport à ceux dans l'air.

L'amélioration de l'intelligibilité de la parole hyperbare est l'objet des recherches depuis plusieurs années (1966). Une synthèse des techniques de rehaussement qui sont présentées dans la littérature est brièvement décrite dans le chapitre, ensuite une description détaillée des algorithmes de traitement que nous avons proposé. Enfin, des résultats de simulation de chaque variante sont présentés.

Dans le chapitre cinq nous exposons les systèmes de réduction du bruit, et ceci par une nouvelle représentation de la fonction d'auto corrélation qui nous permet d'étudier par la pondération du résidu la réduction du bruit dans le domaine temporel. Enfin la dernière partie de ce chapitre introduit la soustraction spectrale qui sera incorporée au système de réduction du bruit dans le domaine spectral.

Une conclusion générale sur le travail de recherche accompli, ainsi que des suggestions pour des travaux à venir sont données à la fin de ce mémoire, quelques algorithmes utilisés sont présentés en annexe.

## CHAPITRE II

# ***Le Signal Vocal***

## II.1 introduction :

Pour comprendre comment l'hélium et la pression déforment la parole, il est préférable de revoir brièvement le procédé de production de la parole.

Nous préciserons dans la première partie de ce chapitre, en plus des caractéristiques temporelles et spectrale du signal vocal, la quasi-stationnarité du filtre vocal et la quasi-périodicité de l'excitation glottique.

De plus, il est montré que l'approximation du conduit vocal, par un modèle AR reste valable dans le cas du signal hyperbare, ces propriétés permettent l'utilisation de la prédiction linéaire qui sera exposée plus en détails dans la deuxième partie.

## II.2 Production du Signal Vocal :

La parole humaine est produite par le système phonatoire, schématisé figure 2.1. On peut l'assimiler à un tuyau acoustique de section variable, terminé à son extrémité par la glotte (c'est la partie du larynx comprise entre les cordes vocales), au sommet de la trachée, et à l'autre par les lèvres. Sa longueur est voisine de 17 cm ; sa section est constamment déformée pendant la production de la parole «phonation» par les mouvements des lèvres, des mâchoires, de la langue et du voile du palais ; elle peut varier de 0 (lors d'une fermeture complète) à 20 cm<sup>2</sup>. Le conduit nasal constitue un trajet auxiliaire pour la transmission de la voix. Il va du voile du palais aux narines. Sa longueur est voisine de 12 cm. Le couplage acoustique entre le conduit buccal et nasal est contrôlé par le voile, lors de l'émission de sons non nasalisés, le voile interdit le passage du son dans la cavité nasale. La figure 2.2 montre une section du larynx, la cavité laryngée est divisée en trois étages par les bandes ventriculaires et les cordes vocales.

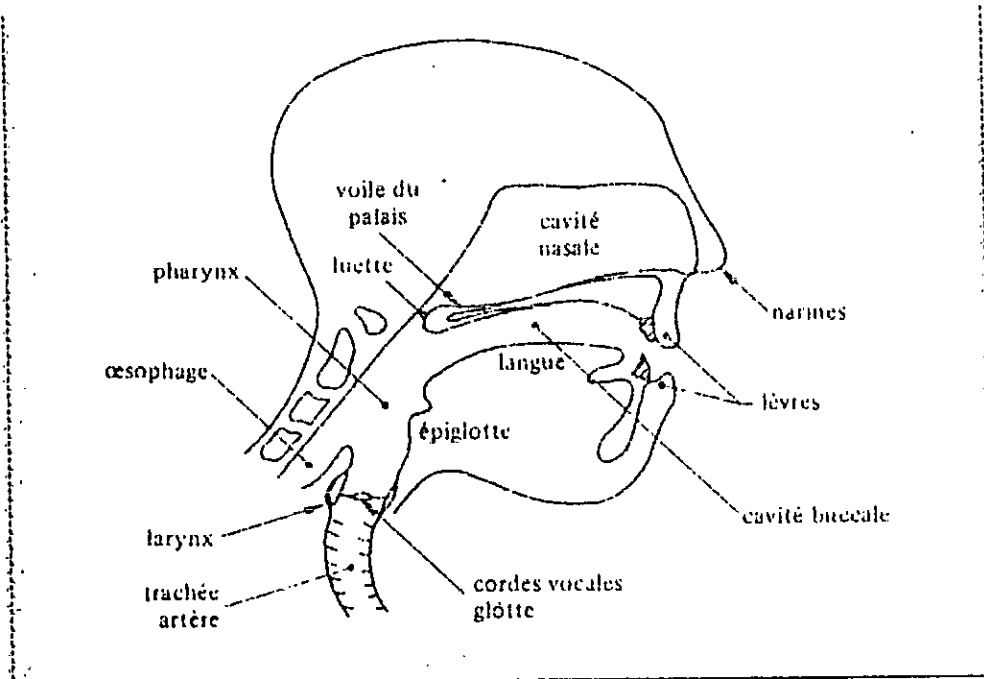


Fig. 1.1 appareil phonatoire

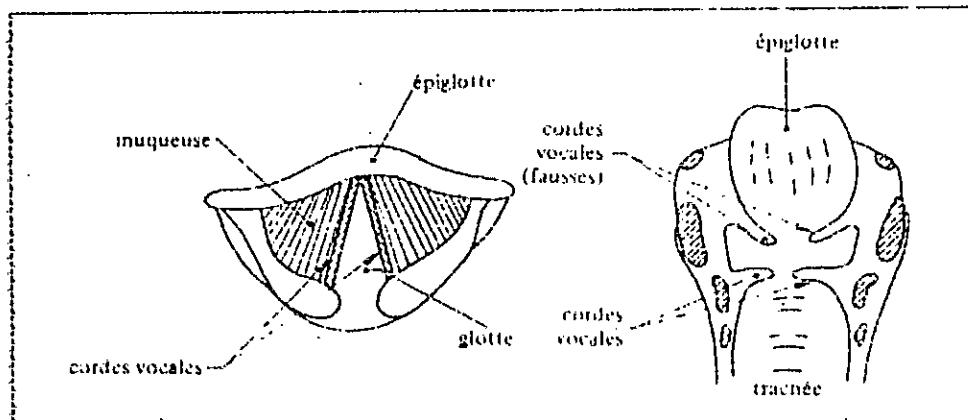


Fig. 1.2 vue de larynx : (a) vue de haut, (b) coupe verticale

Nous pouvons assimiler l'appareil phonatoire à un instrument à vent dans lequel nous trouvons, schématiquement une source d'énergie (semblable à la soufflerie de l'orgue), un système d'excitation servant à mettre le flot d'air en vibration, enfin un système de résonateurs qui vont modifier le spectre du signal d'excitation. Et on considère que la parole résulte de l'excitation du conduit vocal par deux types de sources :

- Les sons voisés comme les voyelles, qui résultent de l'excitation du conduit vocal par des impulsions périodiques, avec une fréquence de récurrence appelée fréquence fondamentale (PITCH), qui peut varier de
  - 80 à 200 Hz pour les hommes.
  - 150 à 450 Hz pour les femmes.
  - 200 à 600 Hz pour les enfants.

Cette différence de fréquence est due à la longueur et la masse des cordes vocales. La figure 2.3 représente la forme d'onde et le spectre d'un son voisé. On y observe les raies qui correspondent aux harmoniques du pitch ( $F_0$ ); l'enveloppe de ces raies présente des maximums appelés formants, qui correspondent aux fréquences propres  $F_i$  ( $i = 1, 2, 3, \dots$ ) dont les trois premiers sont essentiels pour caractériser le spectre vocal.

- Les sons non voisés comme certaines consonnes 's' et 't' sont engendrées par un signal aléatoire (généralement considéré comme un bruit blanc). La figure 2.4 montre la forme d'onde et le spectre d'un son non voisé.

Le spectre du signal voisé (une tranche de la voyelle (o) ) et celui du signal non voisé (la consonne (s)) présentés dans les figures 2.3 et 2.4.

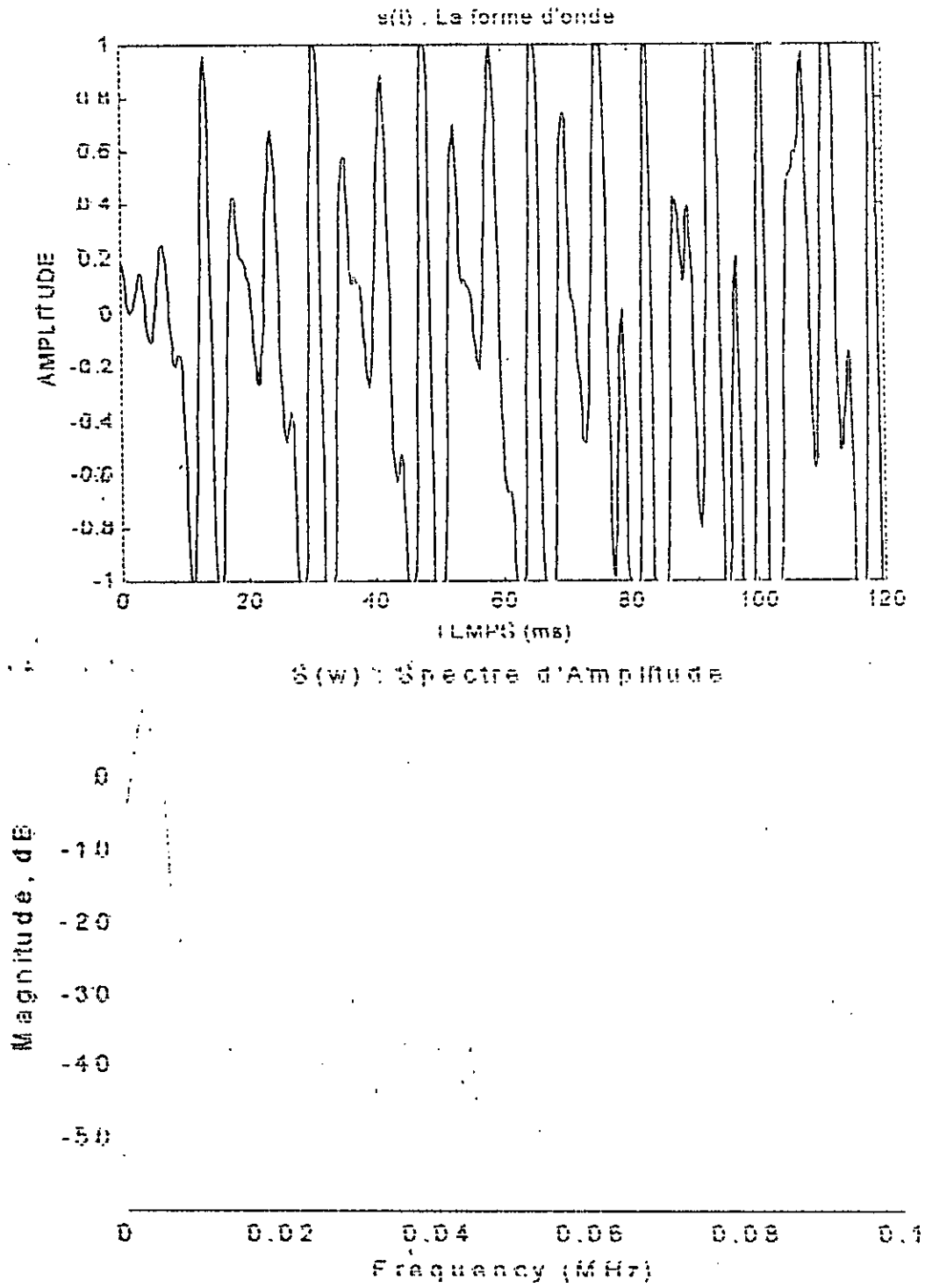


Fig.2.3 la forme d'onde et le spectre d'un son voisé



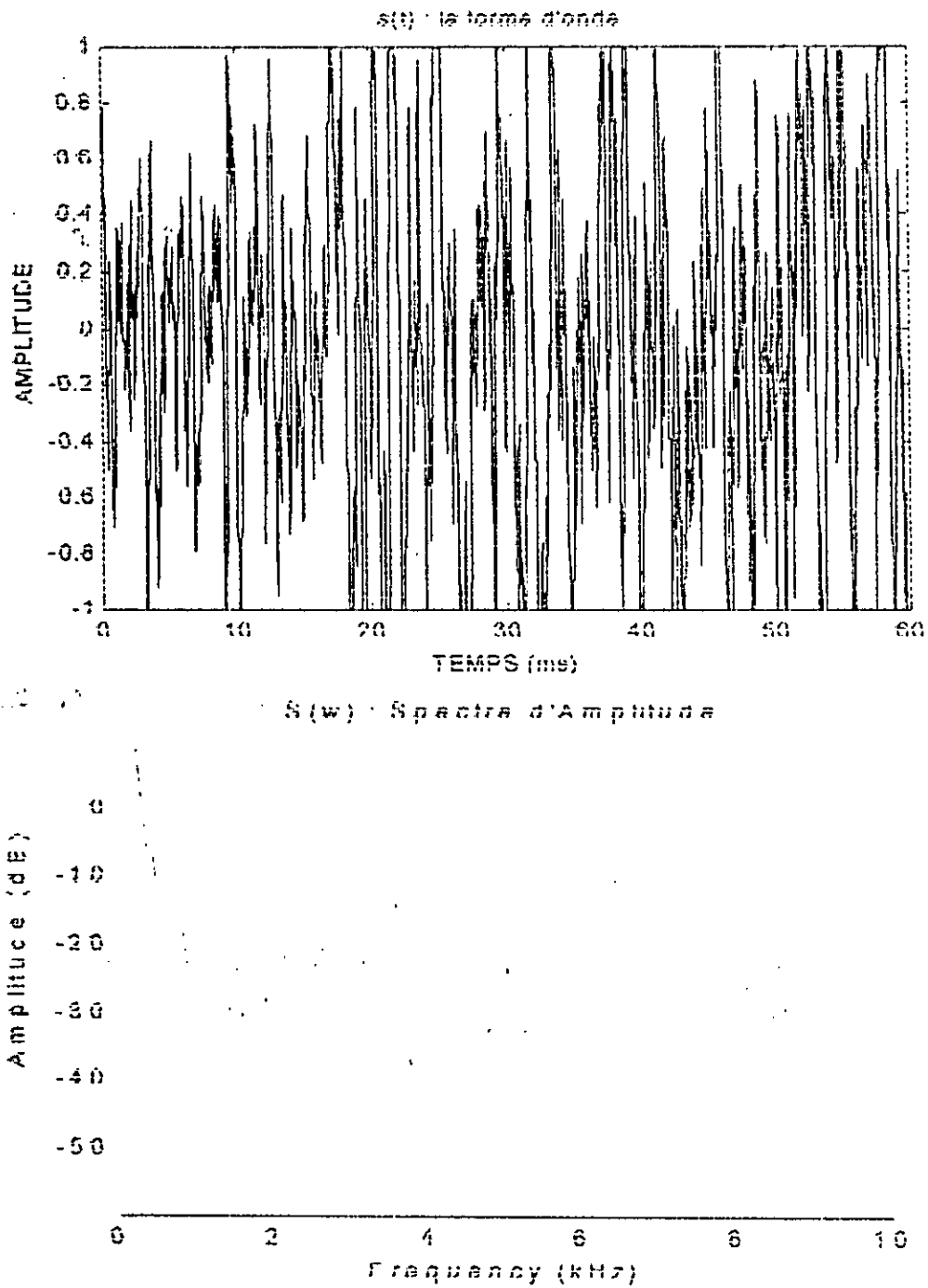


Fig.2.4 la forme d'onde et le spectre d'un son non voisé

Le conduit vocal est modélisé par un filtre tout pôle dont les paramètres varient dans le temps et sa fonction de transfert est :

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^p a_k z^{-k}}$$

Ce modèle de production du signal vocal est appelé auto régressif (AR) ; Il est clair que le modèle AR est approximation plus au moins approchée de la production du signal vocal ; en particulier, pour les sons nasalisés où la transmittance du conduit vocal est plutôt modélisée par un modèle Auto Régressif à Moyenne Ajustée (ARMA) il est dit d'ordre (p, q). Néanmoins, l'estimation d'un modèle ARMA est beaucoup plus délicate que celle d'un modèle AR ; on préfère souvent utiliser un modèle AR dont l'ordre p est un peu surestimé.

En effet à l'expression (2. 1) correspond dans le domaine temporel l'équation suivante :

$$s(n) = -\sum_{k=1}^p a_k s(n-k) + G.U(n) \tag{2.2}$$

qui exprime qu'un échantillon s(n) est une combinaison linéaire des p échantillons qui le précèdent avec un terme d'excitation. Où p est l'ordre du modèle et les coefficients a<sub>k</sub> sont appelés paramètres du modèle, G représente le facteur de gain. La fonction H(Z) est appelée filtre de synthèse et A(Z) le filtre inverse ou filtre d'analyse. Les figures 2.5 et 2.6 montrent le modèle de production de la parole dans le domaine fréquentiel et temporel.

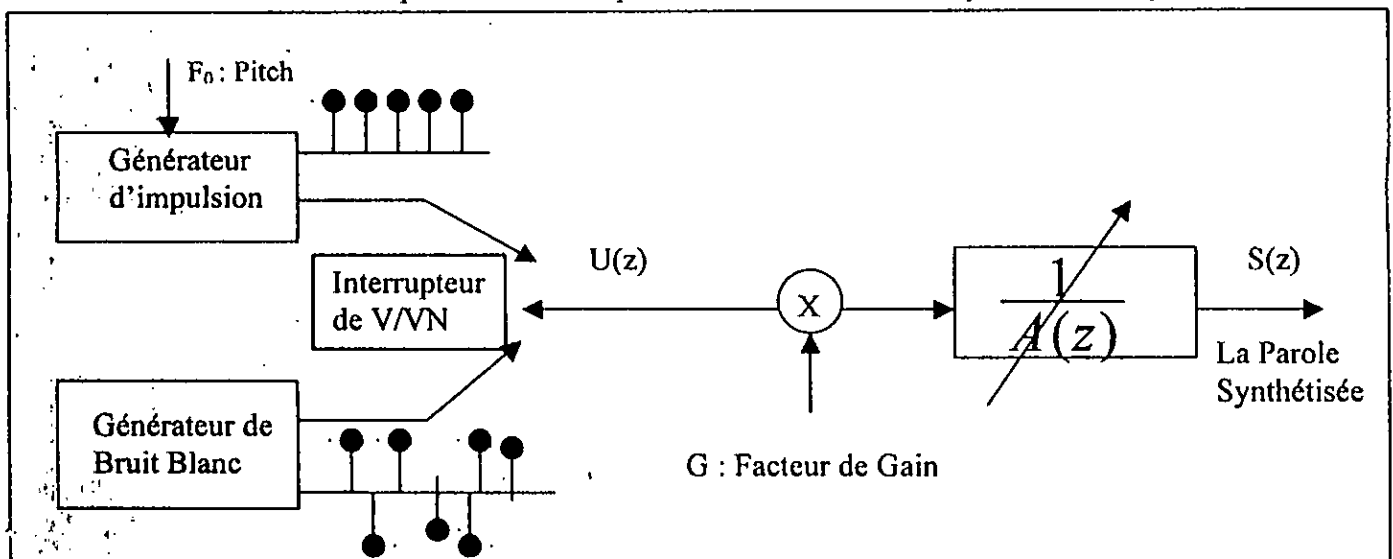


Fig.2.5 Modèle de production de la parole dans le domaine spectral

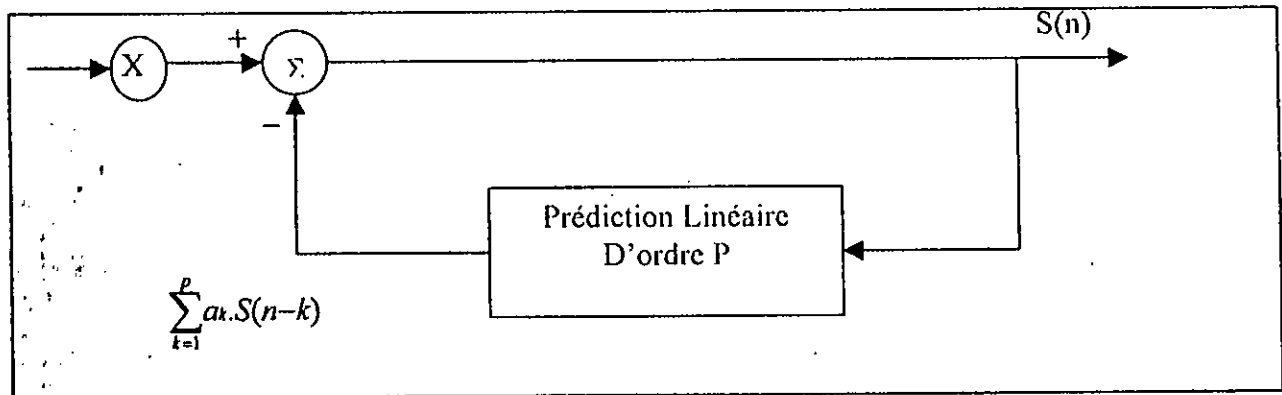


Fig.2.6 : Modèle de production de la parole dans le domaine temporel

### II.3 La prédiction linéaire (LP) :

La prédiction linéaire est l'une des méthodes les plus puissantes du traitement de la parole [3,4,5], utilisée pour l'estimation des paramètres essentiels du signal de la parole, son succès est dû au fait qu'elle représente une solution linéaire au problème de l'estimation des paramètres du modèle de la production de la parole.

La technique de prédiction linéaire est basée sur le modèle de la parole, décrit dans le domaine fréquentiel par l'équation (2.1) et dans le domaine temporel par l'équation (2.2).

L'idée principale de la prédiction linéaire est qu'un échantillon  $s(n)$  de la parole peut être prédit par une combinaison linéaire d'un nombre fini des échantillons qui le précède. Ainsi, un échantillon  $s(n)$  peut être approximatif par une combinaison linéaire de  $P$  échantillons précédents ( $p = 8, \dots, 16$  échantillons) :

$$\tilde{s}(n) = -\sum_{k=1}^p \tilde{a}_k s(n-k) \quad (2.3)$$

Les coefficients  $\tilde{a}_k$ , ( $k = 1, 2, \dots, p$ ) sont appelés paramètres LP ou coefficients de prédiction, et le nombre  $p$  des coefficients s'appelle ordre de prédiction. De plus  $s(n)$  est l'échantillon de la parole et  $\tilde{s}(n)$  est le prédicteur linéaire à l'instant d'échantillonnage  $n$ , qui a la fonction de transfert suivante :

$$P(z) = -\sum_{k=1}^p \tilde{a}_k z^{-k} \quad (2.4)$$

Cette prédiction induit une erreur de prédiction  $e(n)$  entre l'échantillon original  $s(n)$  et

l'échantillon prédit  $\hat{s}(n)$  et elle est définie par :

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p \tilde{a}_k s(n-k) \quad (2.5)$$

Si nous comparons (2.2) et (2.5), nous constatons que lorsque nous trouvons un ensemble de coefficients  $\tilde{a}_k$  plus proche de l'ensemble  $a_k$  en minimisant l'erreur de prédiction ( $\tilde{a}_k = a_k$ ), ( $k = 1, 2, \dots, p$ ), nous obtenons alors l'égalité  $e(n) = G.U(n)$ . Ceci montre que l'erreur de prédiction  $e(n)$  contient beaucoup d'information sur l'excitation du signal de parole. La transformée en Z de l'équation (2.5) donne :

$$E(z) = S(z) \left( 1 - \sum_{k=1}^p \tilde{a}_k z^{-k} \right) = S(z) \left( 1 - \sum_{k=1}^p a_k z^{-k} \right) = S(z) \cdot A(z) \quad (2.6)$$

$A(z)$  est l'inverse de  $H(z)$  dans (2.1),  $A(z)$  est appelé filtre inverse ou filtre d'analyse, car on peut obtenir  $e(n)$  à partir du signal parole en utilisant le filtre  $A(z)$ , suivant l'équation (2.6).

Le problème de l'analyse par prédiction linéaire se réduit donc à trouver un ensemble de coefficients  $\tilde{a}_k$  de façon à minimiser l'erreur de prédiction  $e(n)$  dans une intervalle donnée. L'approche de base est de calculer l'ensemble des coefficients  $a_k$  qui minimisent l'erreur quadratique moyenne de prédiction sur un court segment de la parole, cette erreur est définie par :

$$E = \sum_n e^2(n) = \sum_n \left[ s(n) - \sum_{k=1}^p a_k s(n-k) \right]^2 \quad (2.7)$$

Les valeurs de  $a_k$  qui minimise  $E$  sont obtenues en annulant les dérivées partielles, par rapport à chaque coefficient prédicateur  $a(i) = \frac{\delta E}{\delta a_i} = 0$  pour  $i = 1, 2, \dots, p$  donc :

$$\frac{\delta E}{\delta a_i} = -2 \sum_n \left[ s(n) - \sum_{k=1}^p a_k s(n-k) \right] \cdot s(n-i) = 0 \quad (2.8)$$

Ce qui conduit au système :

$$-\sum_n s(n) \cdot s(n-i) = \sum_n \sum_{k=1}^p a_k s(n-k) \cdot s(n-i) \quad (2.9)$$

où :

$$-\sum_n s(n) \cdot s(n-i) = \sum_{k=1}^p a_k \sum_n s(n-k) \cdot s(n-i) \quad \text{pour } i = 1, 2, \dots, p \quad (2.10)$$

soit :

$$\phi(i, k) = \sum_n s(n-k) \cdot s(n-i) \quad \text{pour } i, k = 1, 2, \dots, p \quad (2.11)$$

Par conséquent l'équation (2.10) devient :

$$\sum_{k=1}^p \alpha_k \phi(i,k) = -\phi(i,0) \quad \text{pour } i = 1, 2, \dots, p \quad (2.12)$$

Le problème de la technique LP se réduit à la résolution d'un ensemble de  $P$  équations avec l'inconnu, afin d'obtenir les coefficients  $\alpha_k : k = 1, 2, \dots, p$ . premièrement, on commence par le calcul des valeurs  $\phi(i,k)$  pour  $i = 1, 2, \dots, p$  et  $k = 0, 1, \dots, p$  en utilisant l'équation (2.11) dont la limite de sommation doit être spécifiée. Deux méthodes généralement utilisées dans l'analyse L.P, la méthode d'auto corrélation et la méthode de covariance. En plus, la méthode de covariance en treillis ou l'algorithme de Burg. On se limitera, dans le cadre de ce travail à l'étude de la méthode d'autocorrélation pour la stabilité qu'elle offre et l'espace mémoire réduit qu'elle nécessite.

### II.3.1 La Méthode d'Autocorrélation :

L'énergie de l'erreur de prédiction (2.7) est calculée sur un interval infini, cependant pour respecter les considérations pratiques nous supposons que  $s(n) = 0$  en dehors de l'intervall ; où  $N$  est la durée de la fenêtre d'analyse LP, cette hypothèse sur la nature de  $s(n)$  simplifiée le calcul du signal  $\hat{s}(n)$ ; donc :

$$\phi(i,k) = \sum_{n=0}^{N+p-1} s(n-i).s(n-k) \quad \text{pour } i = 1, 2, \dots, p \text{ et } k = 0, \dots, p \quad (2.13)$$

Nous observons que seuls les valeurs définies pour  $0 \leq n \leq N+p-1$ , devront être calculés. Par un changement de variable  $m = n-i$ , l'équation (2.13) devient :

$$\phi(i,k) = \sum_{m=0}^{N+p-1} s(m).s(m+i-k) \quad (2.14)$$

Donc  $\phi(i,k)$  est l'auto corrélation à court terme de  $s(m)$  évaluée à  $(i-k)$ . par conséquent :

$$\phi(i,k) = R(i-k) \quad (2.15)$$

où:

$$R(j) = \sum_{n=0}^{N-1-j} s(n).s(n+j) = \sum_{n=j}^{N-1} s(n).s(n-j) \quad \text{pour } j = 0, 1, \dots, p \quad (2.16)$$

L'ensemble de  $p$  équations dans (2.12) devient :

$$\sum_{k=1}^p \alpha_k R(|i-k|) = -R(i) \quad i = 0, 1, \dots, p \quad (2.17)$$

L'équation (2.17) utilisée pour évaluer les coefficients  $\alpha_k$  du modèle ; peut être représentée sous forme matricielle comme suit :

$$\begin{pmatrix} R(0) & R(1) & R(2) & \dots & \dots & \dots & R(p-1) \\ R(1) & R(0) & R(1) & \dots & \dots & \dots & R(p-2) \\ R(2) & R(1) & R(0) & \dots & \dots & \dots & R(p-3) \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & R(0) \\ R(p-1) & R(p-2) & R(p-3) & \dots & \dots & \dots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} R(1) \\ R(2) \\ R(3) \\ \vdots \\ \vdots \\ R(p) \end{pmatrix} \quad (2.18)$$

Nous constatons que la matrice des valeurs de l'auto corrélation est symétrique et que les éléments situés sur la diagonale sont identiques, cette forme matricielle est appelée forme de TOEPLITZ, cette propriété peut être exploitée pour obtenir un algorithme efficace pour résoudre cette équation.

La procédure récursive de Wiener – Levinson – Durbin (WLD) développée dans [5] qui suit représente la solution la plus efficace.

**La procédure WLD**

$$E(0) = R(0)$$

$$K_i = \frac{R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} \cdot R(i-j)}{E(i-1)} \quad (2.19)$$

$$a_i^{(i)} = K_i$$

$$a_j^{(i)} = a_j^{(i-1)} - K_i a_{i-j}^{(i-1)} \quad (2.20)$$

$$E(i) = (1 - K_i^2) \cdot E(i-1) \quad (2.21)$$

La solution finale est :  $a_i = a_i^{(p)} \quad j = 1, 2, \dots, p \quad (2.22)$

La quantité  $E(i)$  dans l'équation (2.21) est l'erreur de prédiction d'ordre  $i$ . les quantités intermédiaires  $K_i$  sont appelées coefficients de réflexion, qui sont les mêmes coefficients qui apparaissent dans le modèle du tube sans pertes du conduit vocal.

La valeur de  $K_i$  jouit de la propriété :  $-1 \leq K_i \leq 1 \quad (2.23)$

La relation (2.23) est une condition nécessaire et suffisante pour que le filtre soit stable. La méthode d'auto corrélation garantit la stabilité, de plus le calcul de  $R(i)$  nécessite un fenêtrage de  $s(n)$  par une fenêtre de Hamming :

$$\omega(n) = 0,54 - 0,46 \cdot \cos\left(\frac{2\pi n}{N-1}\right)$$

$$N = 0, 1, \dots, N-1 \quad (2.24)$$

Ainsi, l'analyse est faite sur le signal :  $x(n) = s(n) \cdot w(n)$

### II.3.2 Les considérations de choix des conditions d'analyse LP

Les variables dans l'analyse LP sont :

- la méthode d'analyse

les deux méthodes d'autocorrélation et de covariance donnent presque les mêmes résultats.

La méthode de covariance ne stipule aucune hypothèse sur la morphologie de signal  $s(n)$ , en dehors des  $N$  points disponibles. ainsi, la minimisation de l'erreur quadratique présentée à l'équation (2.7) ne s'effectue que sur les  $N$  points de  $s(n)$ . Il faut noter que cette approche, contrairement à la technique d'autocorrélation, peut induire une solution instable du modèle. Par ailleurs, la méthode la plus utilisée est celle de l'autocorrélation, malgré l'existence d'une méthode de covariance stabilisée proposée par Atlas.[5]

- le nombre  $P$  des coefficients de prédiction :

comme dans le cas du codage des paramètres LP, on doit utiliser un nombre minimal de paramètre nécessaire à une modélisation exacte de l'enveloppe spectrale court – terme de la parole.

la durée de mémorisation  $p.T$  nécessaire au canal de  $s(n)$  [ l'estimé de  $s(n)$  dans (2.3) ] doit être égale à deux fois le temps mis par l'onde de parole pour se propager depuis la glotte jusqu'aux lèvres. C'est à dire  $(2L) / C$  ou  $L$  est la longueur du conduit vocal et  $C$  la célérité du son dans les conditions normales. Ainsi pour les valeurs de  $C = 34$  (cm/ms) et  $L = 17$  cm, on obtient un temps de 1 ms.

Si la fréquence d'échantillonnage  $F_s = 8$  KHz ; la valeur correspondante de  $p$  est alors :  $8000/1000=8$  au moins , par ailleurs , on ajoute généralement un nombre  $N_p$  de pôles ( 4 à 5) Qui sont nécessaires pour représenter l'influence de la source du rayonnement , donc :

$$P = F_s (\text{KHz}) + N_p$$

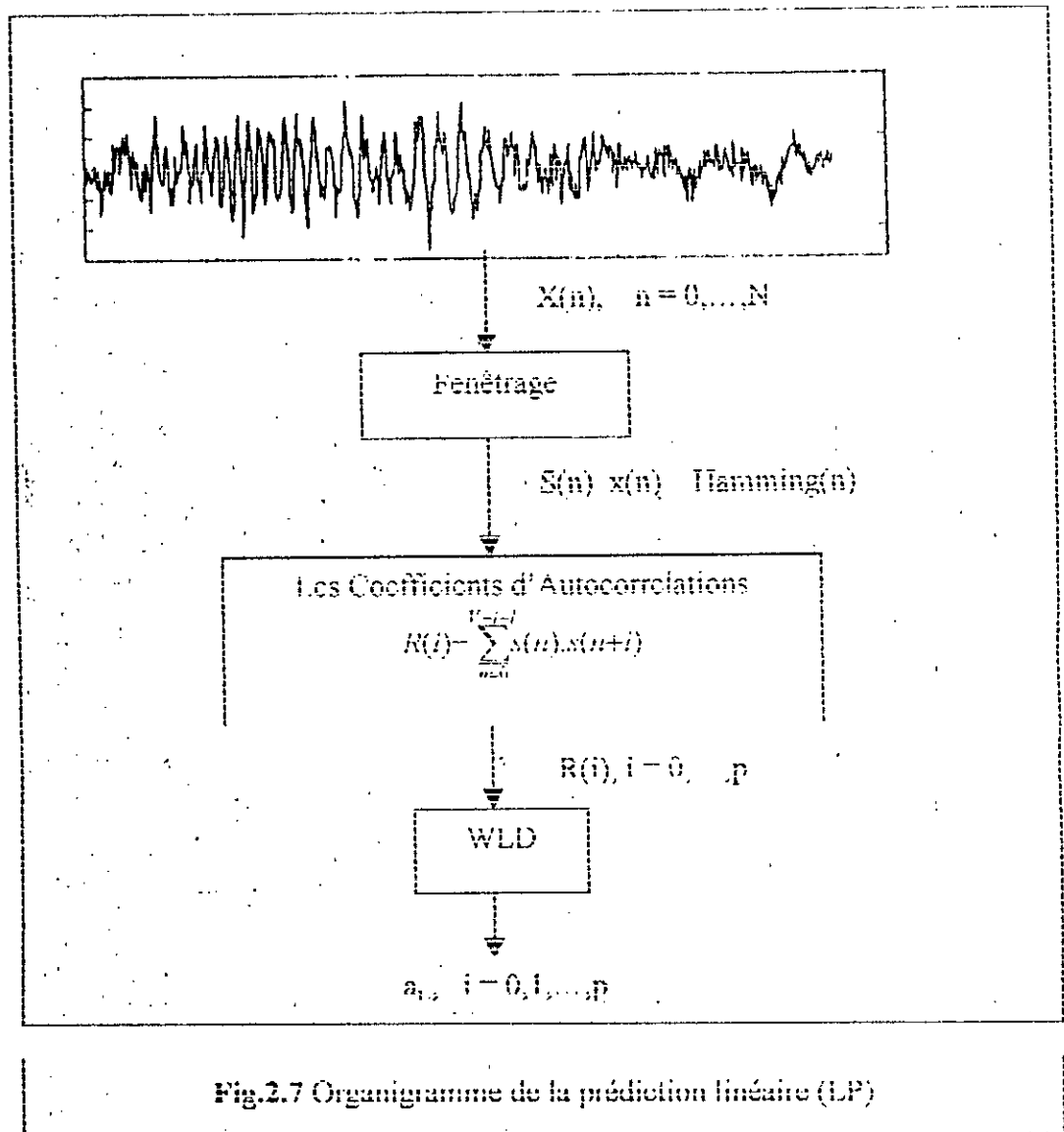
Dans notre cas  $p$  est choisi entre 16 et 20.

- la durée de la trame d'analyse :

pour assurer une bonne résolution spectrale , la durée  $N$  doit être l'intervalle où le mouvement du conduit vocal est négligeable . généralement , cette interval est d'environ 15-20 ms pour la majorité des voyelles , et plus réduit dans le cas des sons non voisés . on peut dire que cette interval est suffisant pour maintenir une bonne qualité de la parole , bien qu'il peut introduire une faible dégradation , surtout pour les sons transitoires , qui ont des changements rapides des caractéristiques spectrales.

La figure 2.7 donne une représentation complète de la prédiction linéaire (annexe A) sous forme d'organigramme.





### 11.3.3 La pré accentuation :

Le spectre de la parole présente une pente globale négative. Par ailleurs, l'oreille humaine est plus sensible à la région du spectre située au-delà de 1 KHz.

La pré accentuation consiste à faire passer, avant l'analyse spectrale par prédiction linéaire, le signal de la parole dans un filtre de transmission :

$$H_p(z) = 1 - \mu z^{-1} \quad \text{avec } 0.9 < \mu < 1. \quad (2.26)$$

Et dans le domaine temporel par la relation de récurrence suivante :

$$y(n) = x(n) - \mu x(n-1) \quad \text{pour } n = 0, 1, \dots, N \quad (2.27)$$

Ce filtre a pour effet d'accentuer la partie haute fréquence du spectre, par conséquent, la pré accentuation contribue à la pente globale négative spectrale. Elle assiste donc, l'algorithme d'analyse spectrale dans la modélisation des aspects du spectre, qui sont plus importants du point de vue de la perception auditive.

#### II.4 Conclusion :

Dans ce chapitre nous avons repris les notions de base du traitement du signal vocal, pour mettre en évidence les principales caractéristiques et outils du signal vocal dans les conditions normales, qui seront appliquées par la suite dans les différents algorithmes proposés de l'analyse et la synthèse de la parole. Enfin, nous pouvons donner les conclusions suivantes :

- Que nous pouvons simuler la source glottale par deux signaux : un signal aléatoire pour les sons voisés et le deuxième un train d'impulsion pour les sons voisés.
- Que le conduit vocal peut être modélisé par un filtre de modèle AR dont les paramètres varient dans le temps, ce qui correspond à la géométrie et aux conditions variables.
- Que le filtre d'analyse et de synthèse LP utilise l'algorithme de WLD, la méthode d'autocorrélation, et un traitement en temps réel basé sur la méthode d'overlap-add avec une fenêtre de Hamming, et une pré accentuation éventuelle dans le cas d'une analyse spectrale court-terme.

## CHAPITRE III

# ***Le Signal Hyperbare***

### III.1 Introduction :

Plusieurs travaux de recherches [7, ..., 14], ont été consacrés à l'étude des propriétés du signal hyperbare. Une grande partie des travaux de M.R.Richards [7] a été reprise dans ce chapitre avec les simulations nécessaires.

La première partie est consacrée aux variations des paramètres physiques des mélanges, qui est basée sur une formulation mathématique, afin de mettre en évidence, l'effet de ces variations sur les caractéristiques principales du signal vocal.

La deuxième partie décrit en général le comportement des largeurs des bandes des formants, les fréquences des formants, de l'amplitude des sons voisés et non voisés, le pitch, et l'excitation glottale dans l'hélium, afin d'avoir une vue d'ensemble sur leurs qualités, et de motiver l'étude des procédures simples et efficaces de rehaussement proposées dans le chapitre suivant.

### III.2 Comparaison entre les propriétés de la parole dans les conditions Normales et hyperbares :

La propagation de l'onde acoustique dans le conduit vocal, dans l'air ou dans un autre milieu hyperbare est due aux vibrations des particules du milieu. Flanagan [1] a montré que les pertes d'énergie dans le conduit vocal, aux conditions normales, dépendent de la célérité du son ( $c$ ), de la chaleur spécifique à pression et volume constants ( $C_p$  ;  $C_v$ ), du coefficient de viscosité ( $\mu$ ), du coefficient de conductibilité de la chaleur ( $\lambda$ ), de la constante adiabatique ( $\gamma$ ) et de la densité du gaz ( $\rho$ ).

Les paramètres cités peuvent être estimés tout simplement, dans l'air et dans l'hélium sous l'hypothèse que les deux milieux se comportent comme un gaz parfait. Quoique, l'hypothèse du gaz parfait n'est pas toujours valable aux pressions extrêmes, mais il est raisonnable, pour plusieurs mélanges respiratoires synthétiques, utilisés sous des pressions modérées rencontrées en plongée sous-marine, d'initier une analyse simplifiée, qui permet d'illustrer les variations des caractéristiques du conduit vocal en fonction du changement de milieu [7].

Le coefficient de viscosité :

$$\mu = \mu_0 \approx 1,87 \times 10^{-4} \text{ dyne-sec/cm}^2 \quad (3.1)$$

Est le même dans l'air et dans tous les mélanges  $H_2-O_2$ . Par conséquent, la viscosité n'est pas un facteur dans l'analyse de la parole hyperbare.

Pour un gaz parfait, la chaleur spécifique à volume constant ( $C_v$ ) est donnée par :

$$C_v = 1 + 2 \times \sum_i n_i p_i \quad (3.2)$$

Où  $n_i$  est le nombre d'atomes par molécule du gaz  $i$ , et  $P_i$  son pourcentage en volume dans le mélange. La sommation est prise sur tous les gaz qui existent dans le mélange. De plus,

$$C_p = C_v + R \quad (3.3)$$

Avec  $R$  la constante universelle du gaz

$$R = 1,987 \text{ cal.mole}^{-1}.\text{deg}^{-1} = 0,082 \text{ l.atm/mole.k} \quad (3.4)$$

La constante adiabatique n'est seulement que :

$$\gamma = \frac{C_p}{C_v} \quad (3.5)$$

Le coefficient de conductibilité de la chaleur ( $\lambda$ ) est donné par rapport à  $C_v$  par :

$$\lambda = \lambda_0 C_v \quad (3.6)$$

$$\text{avec } \lambda_0 = 0,12 ; \text{ et } \lambda_{0H} = 1,17 \quad (3.7)$$

$$\text{La densité d'un gaz parfait est donnée par l'équation : } \rho = P \rho_0 \quad (3.8)$$

Où  $P$  est la pression, et  $\rho_0$  la densité du gaz dans l'air (à 0m) donnée par :

$$\rho_0 = \sum_i n_i P_i A_i \quad (3.9)$$

$A_i$  est la masse atomique du gaz  $i$ .

Finalement, la célérité du son dans un gaz parfait est donnée par la relation :

$$C = \sqrt{\frac{\gamma P}{\rho}} = \sqrt{\frac{\gamma}{\rho_0}} \quad (3.10)$$

Les équations de (3.1 – 3.10) montrent que toutes les constantes du gaz dépendent seulement du mélange gazeux, à l'exception de la densité qui dépend de la pression ambiante. Le tableau suivant résume les paramètres déjà cités, en présentant les rapports entre les quantités  $C$ ,  $\gamma$ ,  $\lambda$ ,  $\rho_0$ ,  $C_p$ , dans plusieurs mélanges gazeux par rapport à ces quantités dans l'air. (l'air est considéré comme étant un mélange de 21% oxygène ( $O_2$ ) et 79% Azote ( $N_2$ )).

%He	%O <sub>2</sub>	$\frac{C_h}{C_a}$	$\frac{\gamma_h}{\gamma_a}$	$\frac{\lambda_h}{\lambda_a}$	$\frac{\rho_{wh}}{\rho_{oa}}$	$\frac{CP_h}{CP_a}$
0	100	0,95	1,00	9,75	1,11	1,00
10	90	1,00	1,01	9,36	1,01	0,97
20	80	1,06	1,02	8,89	0,92	0,94
30	70	1,1	1,04	8,58	0,82	0,91
40	60	1,21	1,05	8,19	0,72	0,89
50	50	1,31	1,07	7,80	0,62	0,86
60	40	1,44	1,09	7,41	0,53	0,83
70	30	1,61	1,11	7,02	0,43	0,80
80	20	1,85	1,13	6,63	0,33	0,77
85	15	2,01	1,15	6,44	0,28	0,76
89	11	2,17	1,16	6,28	0,25	0,75
90	10	2,22	1,16	6,24	0,24	0,74
91	9	2,27	1,16	6,20	0,23	0,74
92	8	2,32	1,17	6,16	0,22	0,74
93	7	2,38	1,17	6,12	0,21	0,73
94	6	2,44	1,17	6,08	0,20	0,73
95	5	2,50	1,17	6,04	0,19	0,73
96	4	2,58	1,18	6,01	0,18	0,73
97	3	2,65	1,18	5,97	0,17	0,72
98	2	2,74	1,18	5,93	0,16	0,72
99	1	2,83	1,19	5,89	0,15	0,72
100	0	2,93	1,19	5,85	0,14	0,71

Tab. 3.1 Rapports des propriétés dans l'He-O<sub>2</sub> et l'air.

Nous remarquons que le coefficient de viscosité et la constante adiabatique sont presque les mêmes dans l'air et dans tous les mélanges He-O<sub>2</sub>, par contre, la conductibilité et la chaleur spécifique augmentent cinq fois ou plus. Cependant, c'est la célérité du son et la densité du mélange gazeux qui restent les plus importantes.

Le mélange respiratoire synthétique utilisé dans la plongée sous marine, doit être modifié pour chaque pression ambiante, correspondante à chaque profondeur d'immersion. Cependant, un des gaz du mélange, l'oxygène, qui est indispensable aux fonctions vitales de l'organisme, sa pression partielle doit être comprise entre 0,2 et 0,7 ATA (pour des physiologiques) est donnée par :

$$PPO_2 = PO_2 \times P \quad (ATA) \quad (3.11)$$

Où  $PPO_2$  est la pression partielle d'oxygène, et  $PO_2$  sa proportion dans le mélange, avec la pression ambiante  $P$  donnée à une profondeur d'immersion  $D(m)$  par la relation suivante :

$$P = 1 + \frac{D}{10} \tag{3.12}$$

La figure (3.1) illustre la proportion d' $O_2$  appropriée pour une profondeur donnée, ainsi que la pression à cette profondeur, pour une  $PPO_2$  constante et égale à 0,2 ; 0,5 ; 1,0 ; 1,6 ATA.

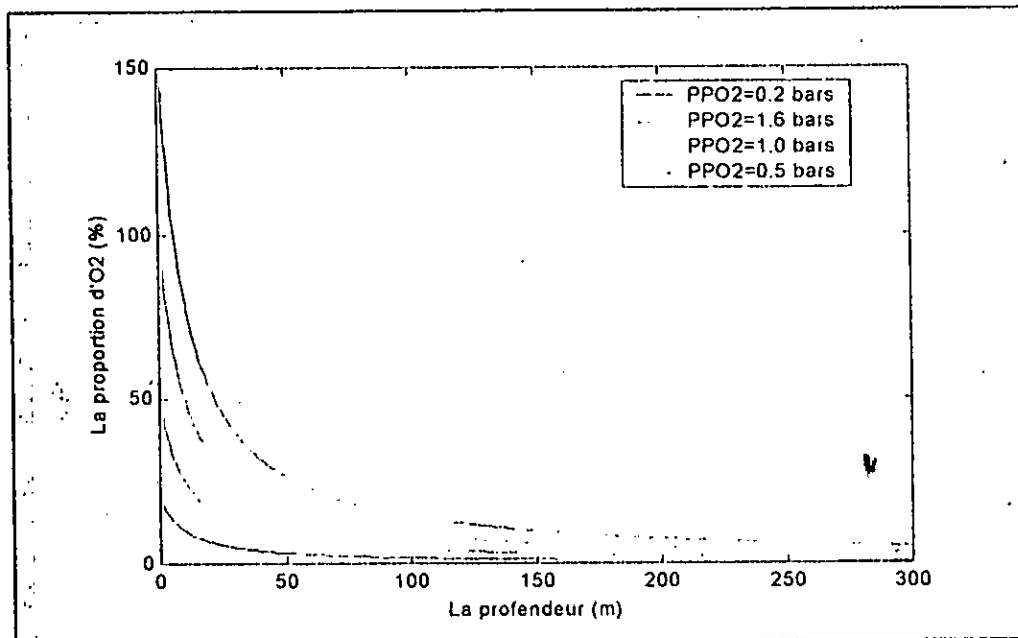


Fig. 3.1 la proportion d' $O_2$  en fonction de la profondeur

La figure 3.2 montre que les rapports  $K = Ch/Ca$  et  $ph/pa$  entre l'héliox et l'air. Ainsi, la célérité est 2 fois jusqu'à 3 fois plus grande dans l'héliox que dans l'air ; et la densité, qui est proportionnelle à la pression, augmente considérablement avec la profondeur.

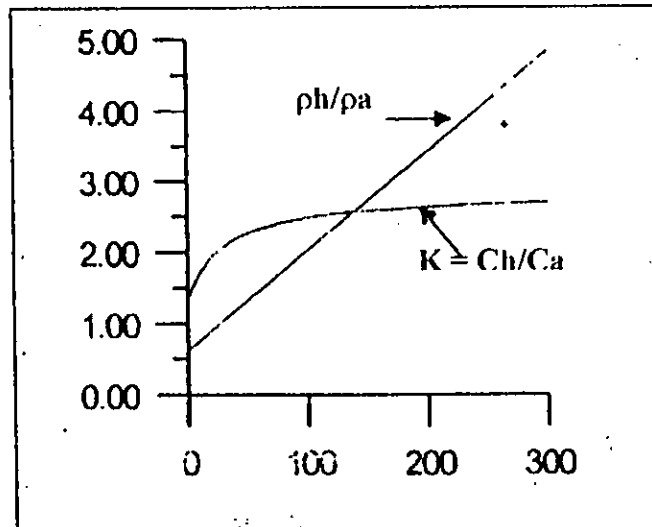


Fig.3.2 Les rapports hélium-air

### III.3 Variation des fréquences des formants :

Fant et Sanneson [25] ont développé une équation pour décrire l'effet de vibration des parois qui affecte essentiellement le premier formant. Il est responsable d'une augmentation de la fréquence de ce formant et de la quasi-totalité de sa bande passante. Soit  $F_{hw}$  la fréquence du premier formant d'un conduit vocal sans pertes (hard-walled), et  $F_c$  la fréquence de résonance du conduit vocal fermé aux lèvres, alors la fréquence  $F$  du premier formant de ce conduit vocal avec pertes (soft-walled) ouvert aux lèvres satisfait la relation suivante :

$$F^2 = F_{hw}^2 + F_c^2 \tag{3.14}$$

où 
$$F_c^2 = \frac{\gamma p}{4\pi^2 L_0} \tag{3.15}$$

Et  $L_0$  est l'inertie des parois de la cavité (cavity wall), qui est indépendante des propriétés atmosphériques. L'équation (3.15) est valable pour tous les gaz ; bien qu'elle soit obtenue initialement pour le premier formant, elle est valable pour tous les formants. Ces deux équations peuvent être utilisées pour faire une comparaison entre les fréquences des formants « air » et dans les mélanges He-O<sub>2</sub> comme suit :

Premièrement, on a : 
$$F_h^2 = F_{hwh}^2 + F_{ch}^2 \tag{3.16}$$

Et 
$$F_a^2 = F_{hwa}^2 + F_{ca}^2 \tag{3.17}$$

Pour le tube sans pertes (hard-walled) seule la vitesse du son affecte le résonance, et on a donc :



$$F_{hwh}^2 = \left( \frac{C_h}{C_a} \right)^2 \cdot F_{hwa}^2 = \left( \frac{C_h}{C_a} \right)^2 (F_a^2 - F_{ca}^2) \quad (3.18)$$

à partir de l'équation (3.15)

$$F_{ch}^2 = \frac{\gamma_h \cdot p_h}{\gamma_a \cdot p_a} \cdot F_{ca}^2 \quad (3.19)$$

on remplace les valeurs de  $F_{hwh}^2$  et  $F_{ch}^2$  de (3.18) et (3.19) dans (3.16) avec  $p_a = 1ATA$  ce qui donne

$$F_h^2 = \left( \frac{C_h}{C_a} \right)^2 \cdot F_a^2 + \left( \frac{\gamma_h \cdot p_h}{\gamma} - \left( \frac{C_h}{C_a} \right)^2 \right) \cdot F_{ca}^2 \quad (3.20)$$

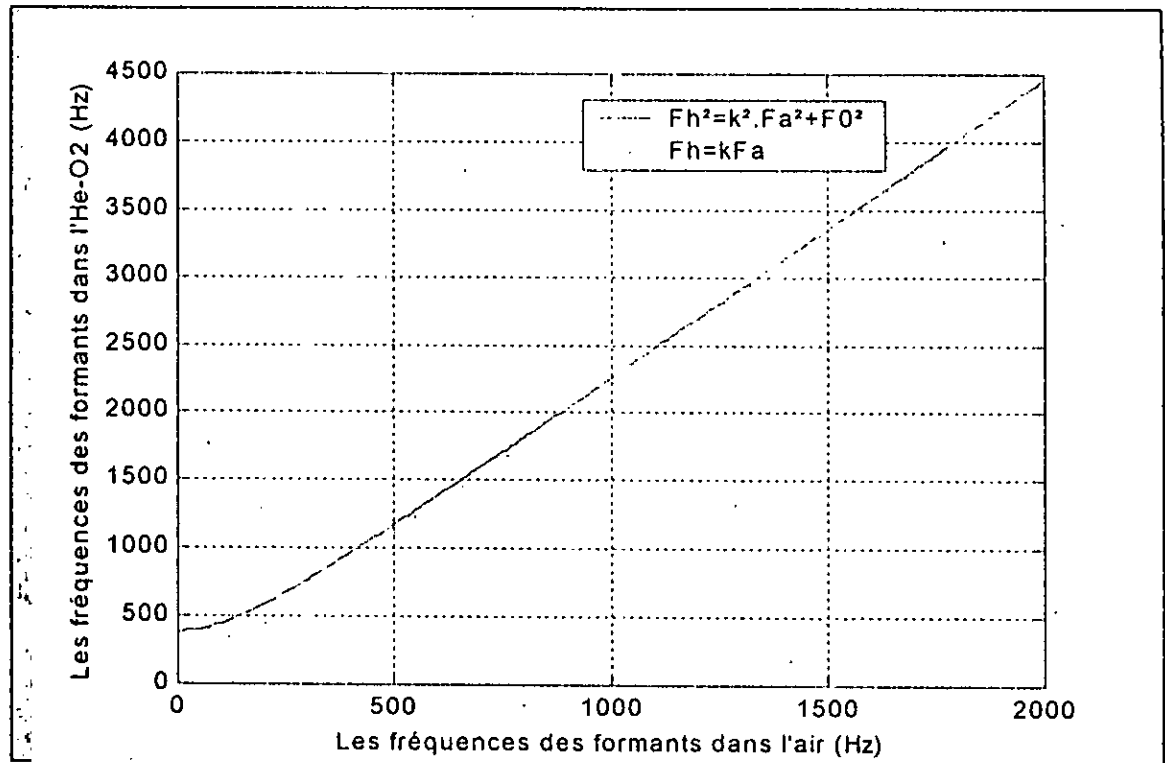
finalement ,vu que  $C^2 = \frac{\gamma \cdot p}{\rho} = \frac{\gamma}{\rho_0}$

$$F_h^2 = \left( \frac{C_h}{C_a} \right)^2 \cdot F_a^2 + \left( \frac{\gamma_h}{\gamma_a} \right) \left[ p_h - \frac{\rho_{0a}}{\rho_{0h}} \right] \cdot F_{ca}^2 \quad (3.21)$$

cette question est sous la forme :  $F_h^2 = K^2 \cdot F_a^2 + F_0^2$  (3.22)

$$\text{ou : } K = \frac{C_h}{C_a} \text{ et } F_0^2 = \left( \frac{\gamma_h}{\gamma_a} \right) \left[ p_h - \frac{\rho_{0a}}{\rho_{0h}} \right] \cdot F_{ca}^2 = K^2 \left( \frac{\rho_h}{\rho_a} - 1 \right) F_{ca}^2 \quad (3.23)$$

La constante  $F_{ca}$  est déterminée à partir de la longueur du conduit vocal, elle est égale à 150 Hz pour un plongeur masculin, ce qui résulte de  $F_0 = 390$  pour un mélange gazeux de 90% - 10% He-O<sub>2</sub> à 10 ATA. La relation (3.21) est illustrée dans la figure suivante sous les conditions citées plus haut. La courbe approche vite l'asymptote  $F_h = K \cdot F_a$  qui décrit un conduit sans pertes.



**Fig.3.3** Translation des fréquences des formants pour 90% He- 10% O2 à 10 ATA

#### III.4 Variation des largeurs de bandes des formants :

La théorie acoustique du signal vocal dans les conditions normales, développée par Flanagan [1] a révélé que les largeurs des bandes des formants dépendent des caractéristiques des gaz et de la pression ambiante. Richard a étendu ce concept au signal hyperbare, il a établi que les pertes inclus des formants et à un décalage des fréquence de ces derniers. De plus, les largeurs des bandes formants inférieurs sont dominées par la contribution des pertes par friction et par vibration des parois, et celles des formants supérieurs sont dominés par la contribution des pertes par rayonnement. Par conséquent, la largeur pouvait augmenter de K fois pour les formants supérieurs, et du carré de K pour les formants inférieurs.

### III.5 Variation des amplitudes des sons voisés et non voisés :

L'approximation du spectre du signal vocal  $s(n)$ , la plus utilisée dans les systèmes de correction proposés dans la littérature, basée sur le modèle de la production de la parole, peut se formuler comme suit :

$$S(f) = G(f) \cdot V(f) \cdot L(f) \cdot M(f) \quad (3.26)$$

Où  $G(f)$  est le spectre d'excitation glottique ;

$L(f)$  caractérise les radiations ;

$M(f)$  est la réponse fréquentielle combinée masque micro ;

$V(f)$  est la fonction de transfert du conduit vocal, qui est le rapport de la vitesse volumique au niveau des lèvres ou au niveau de l'ouverture nasale par rapport à la vitesse volumique glottale.

Fant [9] a remarqué que l'amplitude du spectre de la vitesse volumique aux lèvres est proportionnelle à  $\rho^{-1/2}$  pour les sons voisés. Par contre, pour les sons non voisés, elle est proportionnelle à  $\rho^{-1}$ . De plus, la fonction de transfert  $M(f) \cdot L(f)$  est proportionnelle à  $\rho$ . L'amplitude spectrale de la parole sous pression est donc proportionnelle à  $\rho^{1/2}$  pour les sons voisés, mais elle est indépendante de  $\rho$  pour les sons non voisés. Par conséquent, le changement de la densité atmosphérique conduit à une différence des changements d'amplitude des sons voisés par rapport à celle des sons non voisés. Une augmentation de  $\rho$  résulte en une réduction de l'amplitude des sons non voisés comparativement à celle des sons voisés. Par exemple, pour un mélange 90% He & 10% O<sub>2</sub>, le rapport des densités est de 2,40. Donc un rapport d'atténuation de  $\sqrt{2,4} = 1,55$ , ou 2,8 dB, des amplitudes des sons non voisés par rapport à celle des sons voisés dans un milieu pressurisé.

### III.6 Les caractéristiques du signal Hyperbare :

En premier lieu les recherches sont orientées vers une meilleure détermination des propriétés du signal hyperbare. Ainsi, les caractéristiques principales du signal vocal sont affectées par les conditions sous-marine (section précédente).

On peut retenir les faits suivants :

- Un étalement du spectre du signal hyperbare d'environ trois fois, par rapport au spectre du signal vocal dans les conditions normales, qui est proportionnel à la célérité du son. De plus une modification de la répartition des formants d'autant plus importante que les fréquences sont basses. Il est généralement admis que la loi de Fant [9] caractérise correctement cette modification formulée par :

$$F_h^2 = K^2 \cdot F_a^2 + K^2 \left( \frac{\rho_h}{\rho_a} - 1 \right) F_{ca}^2 \quad (3.24)$$

- $F_h$  et  $\rho_h$  sont respectivement la fréquence centrale du formant et la densité hyperbare.

- $F_a$  et  $\rho_a$  sont la fréquence centrale du formant et la densité dans les conditions normales.
- $K$  est le facteur de propagation, égal au rapport des célérités du son dans l'hélium et l'air.

$$K = \alpha = Ch/Ca \quad (3.25)$$

- $F_{ca}$  est une constante caractérisant la résonance du canal fermé à une pression d'une atmosphère dans l'air, elle varie de 1 à 180 Hz pour 54 à 500 m de profondeur. Elle est choisie pour chaque profondeur et individu donné. De plus, en basses fréquences les glissements de fréquences varient selon que le plongeur est avec ou son masque
- Cette transposition spectrale est une fonction non linéaire, qui provoque des variations sensibles sur les autres paramètres tels que : les largeurs de bandes des formants, surtout les trois premiers formants. Richard [7] a établi que la largeur pouvait augmenter de  $K$  fois pour les formants supérieurs et du carré de  $K$  pour les formants inférieurs. L'erreur sur la fréquence des formants est d'environ 15%.
- La période du pitch augmente sensiblement avec la célérité du son augmente dans l'hélium [26], il n'y a que peu de changement du pitch. L'auteur de l'article [10] donne par exemple une légère variation de moins de 10 Hz, d'autres donnent une augmentation de 10 jusqu'à 50%. La variation du pitch ne doit par normalement dégrader l'intelligibilité, mais elle affecte surtout l'identification du locuteur. Par conséquent, le pitch n'est pas un paramètre dans le rehaussement du signal hyperbare.
- L'atténuation de la partie des hautes fréquences, autrement dit, un affaiblissement des sons non voisés par rapport aux sons voisés.
- La vitesse de l'élocution diminue.
- De plus, il a été constaté un niveau de bruit élevé en plus du phénomène de réverbération du milieu hyperbare.
- Pour tenter de se faire mieux comprendre, le plongeur a tendance à élever sa voix, ce qui provoque parfois une augmentation de la fréquence du pitch et une modification de son articulation surtout lorsqu'il cherche à dominer le bruit ambiant.
- Après un séjour prolongé, le naturel des sons s'améliore.

Les figures 3.4 illustrent la forme d'onde du mot français « loi » prononcés par des plongeurs à 60m, 100m, 200m en caisson à l'hélium, en plus d'une version enregistré dans un studio (i.e. à 0m). On remarque ainsi, les différentes variations des caractéristiques du signal hyperbare suivant les condition d'immersion.

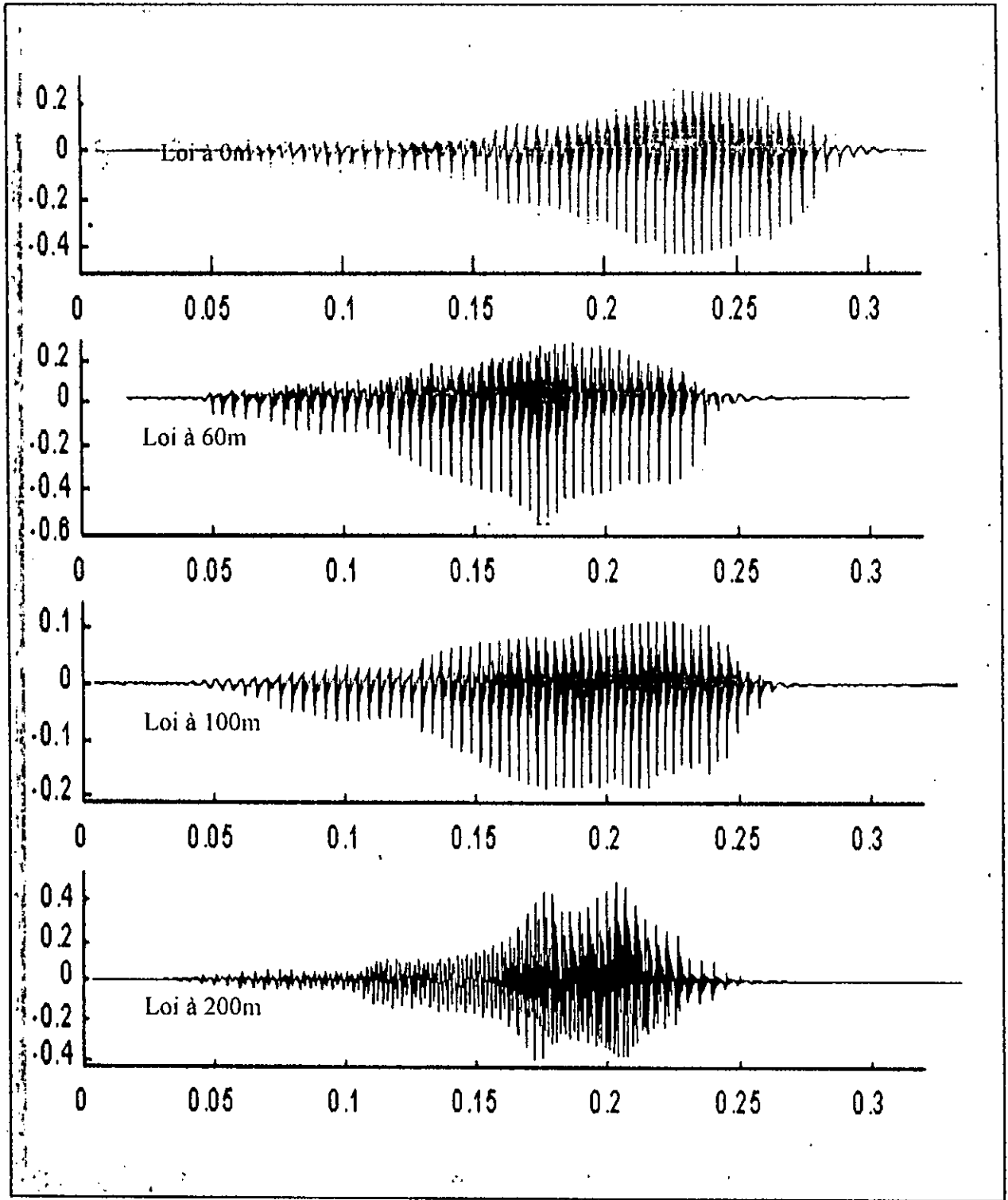


Fig.3.4 Forme d'onde du mot « loi »

### **III.7 Conclusion :**

Nous concluons en disant que la quasi-stationnarité et la quasi-périodicité sont des caractéristiques du signal hyperbare, ce qui a permis de faire le point sur l'étalement non linéaire du spectre du signal hyperbare (*caractérisé par l'équation 3.24*), l'invariance du pitch et l'élocution des plongeurs dans l'hélium qui est plus lente que celle dans l'air. Tout cela nous permet l'utilisation de la modélisation de ce signal par un modèle AR.

Ce chapitre présente seulement une vue d'ensemble, qui s'avère être un support privilégié pour comprendre les systèmes de correction.

Enfin, il est à noter que dès à présent, la disponibilité d'une base de données sur les signaux hyperbares est plus que nécessaire afin de mener ce travail à terme.

## CHAPITRE IV

# **Les Systèmes de traitement**

## **IV.1 Introduction**

L'amélioration de l'intelligibilité de la parole hyperbare est l'objet de recherche depuis 1966. Deux méthodes révélées des travaux présentés auparavant, s'avèrent être les plus performantes. L'une est fondée sur les propriétés de la prédiction linéaire, l'autre utilise la transformée de Fourier à court terme (STFT) [7]. On peut les répertorier en deux grandes classes, les systèmes à approche temporelle et les systèmes à approche fréquentielle. Une synthèse de ces techniques est brièvement présentée dans la première partie de ce chapitre.

La deuxième partie explique le principe général des systèmes de correction, ensuite une description détaillée des algorithmes de traitement que nous avons proposés, fait l'objet de la troisième et quatrième partie de ce chapitre.

La dernière partie est consacrée aux résultats de simulation de chaque variante, présentés afin de valider la robustesse des algorithmes et puissance vis à vis des exigences de l'intelligibilité, et l'adaptation aux différentes conditions de la plongée sous-marine.

## **IV.2 Les systèmes de traitement**

### **IV.2.1 Les systèmes à approche temporelle**

Les techniques de traitement des signaux hyperbares initialement employées ont une approche temporelle. On peut les classer en quatre catégories [20] :

#### **IV.2.1.a Enregistrement/Reproduction ou "Tape Recorder Playback"**

Elle consiste à enregistrer les signaux hyperbares à une vitesse donnée et de les reproduire à une vitesse plus lente; le rapport entre les vitesses serait le taux de compression spectrale [27]. Les limites de cette méthode, très simple, sont fort bien



connues. Elle a de nombreux inconvénients tels que l'altération du pitch, l'impossibilité de travailler en temps réel, la qualité du signal de sortie.

#### IV.2.1.b Le codage numérique des signaux "Digital Coding"

Cette technique permet une expansion dans l'espace temps du signal sans affecter le pitch et permet une utilisation en temps réel. Diverses variantes concernant également la procédure technique telle que la segmentation existent ; même un système de communication entre plongeurs et la surface a été décrit [21, 22].

#### IV.2.1.c Traitement par signal analytique

Cette méthode met en oeuvre la transformée de Hilbert capable de mettre en évidence l'enveloppe du signal afin d'atteindre les objectifs de compression.

Si  $S(t)$  est le signal d'entrée:

$$S(t) = A(t) \cos[\Phi(t)] \quad (4.1)$$

où  $A(t)$  est l'enveloppe et  $\Phi(t)$  la phase.

en utilisant la Transformée de Hilbert, on a :

$$S(t)/n = [a(t)]^{1/n} \cdot \cos[\Phi(t)/n] \quad (4.2)$$

Il apparaît clairement que la fonction réalise une division de l'échelle fréquentielle de  $1/n$ . Aucun système connu n'a été conçu selon ce procédé.

#### IV.2.1.d Traitement par Convolution

Elle consiste à obtenir une séparation de la source et de la réponse impulsionnelle du conduit vocal (déconvolution), puis de corriger cette dernière et d'exécuter l'opération inverse (convolution). Il est évident que la difficulté réside dans cette correction.

Il semble que le principe général des systèmes temporels reste la segmentation du signal temporel hyperbare dont le spectre occupe une plage de fréquences égale pratiquement à trois fois celle du signal normal. Puis une dilatation temporelle et enfin une reconstitution sont réalisées (figure 4.1). Mais la segmentation pour permettre une bonne reconstruction nécessite la détermination du début de chaque période, ce qui n'est pas toujours possible vu qu'il existe des signaux non voisés qui n'ont aucune période. Ce phénomène cause des discontinuités [31] sur le signal temporel reconstitué.

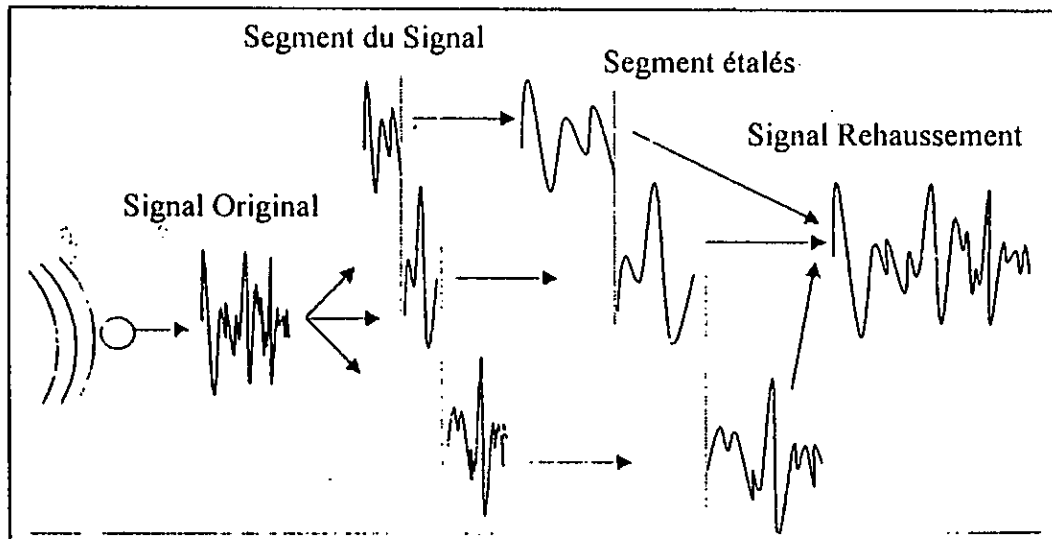


Fig.4.1 Approche temporelle de traitement de la parole hyperbare [20.22]

#### IV.2.2 Les systèmes à approche fréquentielle

Plusieurs méthodes fréquentielles ont été décrites, mais en fait elles sont toutes de même nature à quelques exceptions près. Elles reposent toutes sur la correction du spectre du signal hyperbare par une compression de ce dernier avec un facteur de compression adapté aux conditions du milieu hyperbare. Le même procédé de segmentation – traitement – concaténation est suivi, mais opère sur le spectre [7-23]. Le principe général de la méthode est décrit par la figure 4.2 :

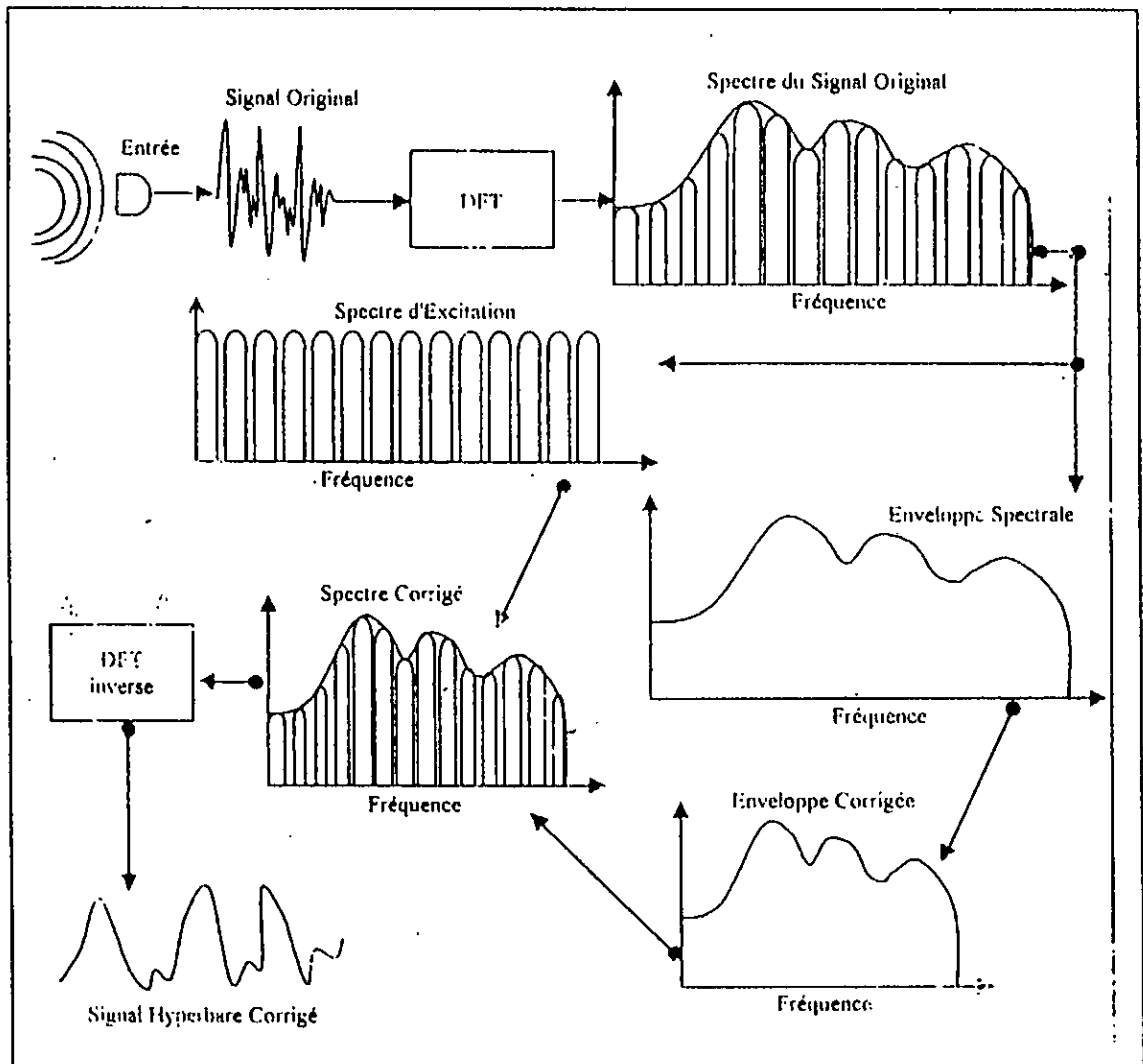


Fig. 4.2 Approche fréquentielle de traitement de la parole hyperbare[32]

### IV.3 Considération pour la conception d'un système de traitement

Pour pouvoir répondre aux exigences d'une future implémentation du système de traitement sur DSP, la recherche d'une procédure à la fois simple et efficace s'avère nécessaire. Cependant, la structure générale du système de traitement reste invariante dans sa globalité et devrait passer par trois étapes nécessaires pour une meilleure restauration de l'intelligibilité du signal vocal :

1. L'estimation de la fonction de transfert du conduit vocal et de l'excitation glottale dans l'étape d'analyse par prédiction linéaire.
2. L'étape de correction du conduit vocal, dans le domaine temporel ou fréquentiel.
3. L'étape de la synthèse à partir d'une réponse corrigée et de l'excitation.

La figure 4.3 schématise le principe général de cette méthode.

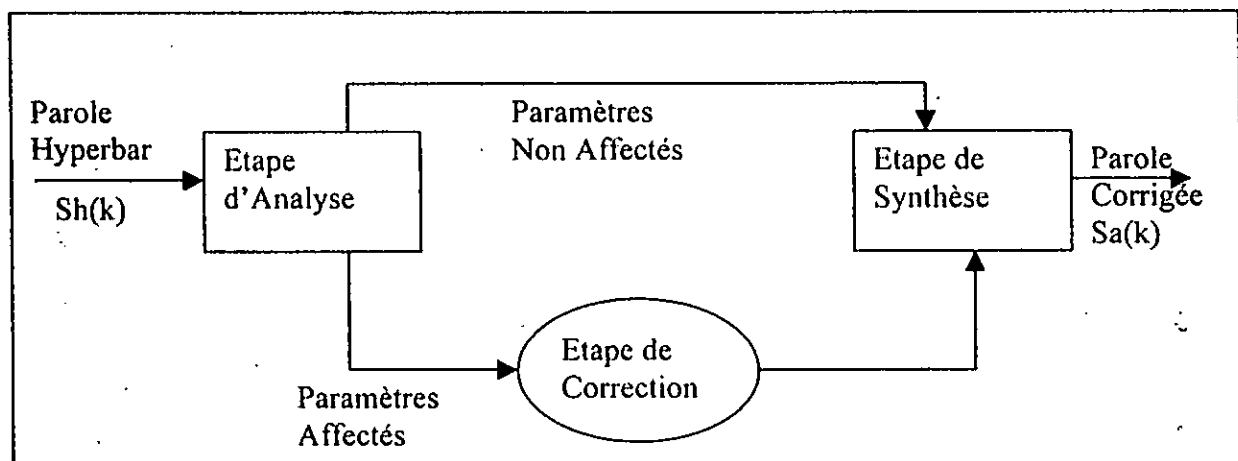


Fig. 4.3 principe du système de correction

### IV.4 Correction des corrélations :

Soit  $x_h(n)$  ( $n = 0, 1, \dots, N$ ) la fenêtre en cours d'analyse du signal hyperbare original une profondeur donnée, et  $x_a(n)$  ( $n = 0, 1, \dots, N$ ) la fenêtre correspondante du signal corrigé

dans la phase de synthèse. Le signal hyperbare est pondéré par une fenêtre de Hamming de longueur (NH), selon la relation suivante :

$$S_h(n) = x_h(n) \cdot \omega(n) \quad \text{pour } n = 0, 1, \dots, \text{NH}, \quad (4.3)$$

$$\text{Où : } \text{NH} = \text{N} + 2 \cdot \text{dH} \quad (4.4)$$

DH est le débordement de la fenêtre de Hamming de chaque côté de la fenêtre d'analyse N,  $s_h(n)$  est le signal pondéré qu'on nomme tout simplement  $x_h(n)$  dans ce qui suit.

Comme on l'a déjà vu dans les sections précédentes, la correction des corrections est basée sur l'analyse LPC qui nous permet de séparer l'excitation glottale et la réponse du conduit vocal, afin de mettre au point les corrections nécessaires à cette réponse, qui est affectée par les conditions du milieu et de la convoluer par la suite avec l'excitation glottique, supposée intacte vis à vis des conditions sous-marines.

L'algorithme général de correction des corrélations (fig. 4.5) se présente comme suit :

1. Lecture des paramètres et données d'entrée.
2. Calcul des autocorrélations du signal d'entrée hyperbare données par la relation (2.16) rappelée ici :

$$R(j) = \sum_{n=0}^{\text{NH}-1-j} s(n) \cdot s(n+j) = \sum_{n=j}^{\text{NH}-1} s(n) \cdot s(n-j) \quad \text{pour } j = 0, 1, \dots, p \quad (4.5)$$

3. Détermination des coefficients  $a(j) : 1, \dots, p$ , du filtre d'analyse LPC du signal hyperbare, à partir des autocorrélations en utilisant l'algorithme de WLD.
4. Calcul de résidu du signal hyperbare, en vue de son utilisation future en synthèse, par un filtrage du signal hyperbare par le filtre d'analyse  $A(z)$  selon la formule suivante :

$$\text{residu}(n) = X_h(n) + \sum_{j=1}^p a(j) X_h(n-j) \quad \text{pour } n = 0, 1, \dots, \text{N} \quad (4.6)$$

5. Correction des autocorrélations du signal hyperbare par dilatation dans le domaine temporel, réalisée par un filtre interpolateur capable de fournir la composante temporelle désirée selon le taux de dilatation exigé, comme indiqué dans l'équation de récurrence suivante :

$$T_j = \text{int} \left[ \frac{j}{K} \right]$$

$$\text{Fact} = \frac{j}{K} - \text{int} \left[ \frac{j}{K} \right] \quad j = 0, 1, \dots, K \cdot p \quad (\text{int} : \text{partie entière}) \quad (4.7)$$

$$R_c(j) = (\text{fact}) \cdot R(T_j+1) + (1 - \text{Fact}) \cdot R(T_j)$$

La figure 4.4 illustre les autocorrélations du signal hyperbare original ( $R(i)$ ) et celles corrigées ( $R_c(i)$ ) qui correspondent à un taux de compression  $K = 2,34$ , un filtre d'analyse d'ordre  $p = 20$  et un filtre de synthèse d'ordre  $Mh = K * p = 46$ .

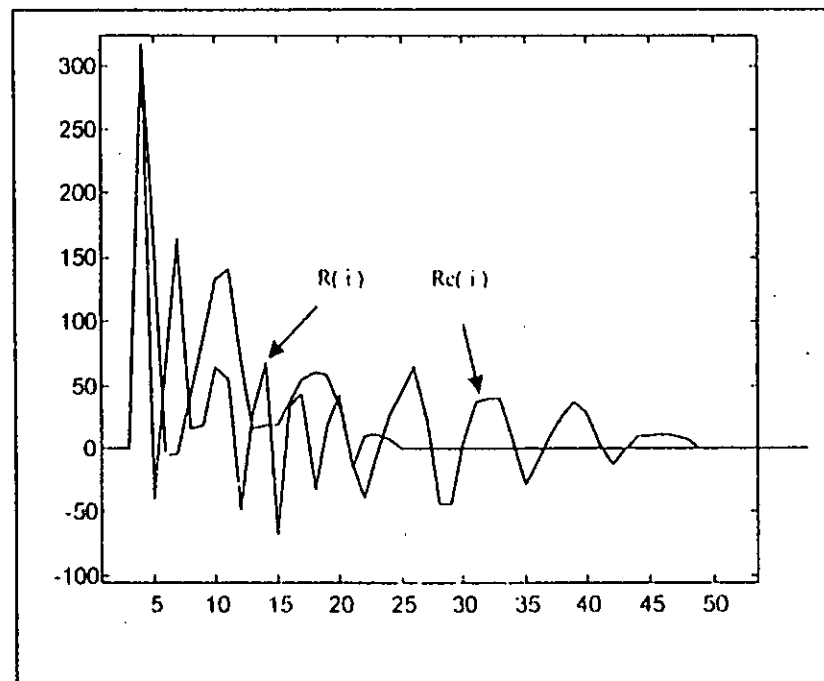


Fig. 4.4 L'étalement temporel des autocorrélations

6. Obtenir de nouveaux coefficients LPC  $a_c(j)$  ( $j = 0, 1, \dots, K * p$ ), avec l'algorithme de WLD. Ces coefficients correspondent à un signal qui n'avait pas existé ou été prononcé auparavant, ou encore à un signal qui aurait pu être émis dans un milieu ambiant normal.
7. les nouveaux LPC obtenus, vont servir la synthèse du signal de sortie. Un filtre LPC de synthèse fait appel au résidu déjà obtenu dans la première étape et aux coefficients LPC obtenus après correction selon la formule suivante :

$$X_a(n) = \text{residu}(n) - \sum_{j=1}^{K,p} a_c(j) \cdot \text{residu}(n-j) \quad n = 0, 1, \dots, N \quad (4.8)$$

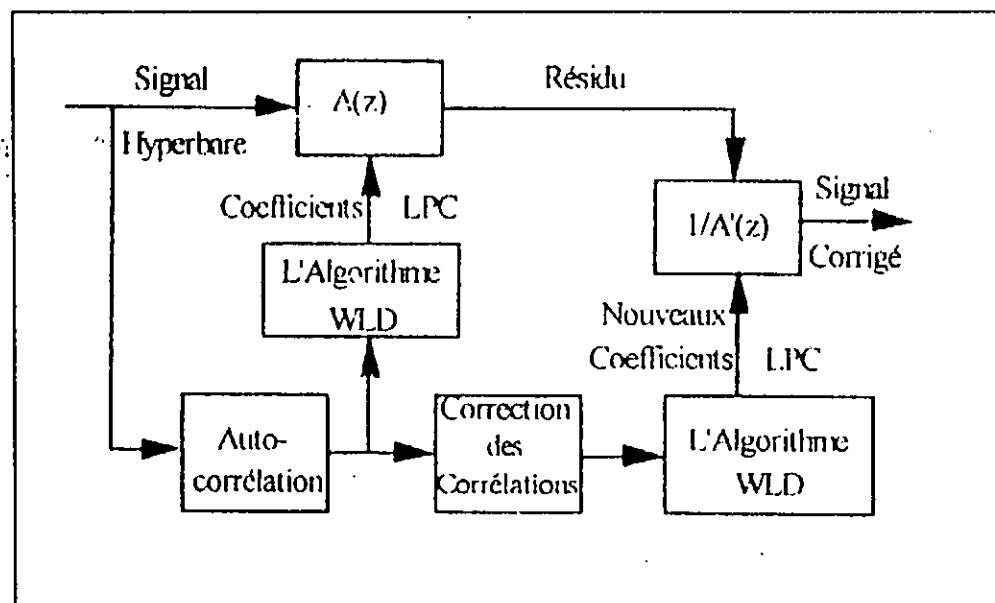


Fig. 4.5 Algorithme de correction des autocorrélations

### IV.5 Correction spectrale

Les deux étapes d'analyse et synthèse restent invariantes dans le cas de correction spectrale, comme décrit précédemment dans les étapes 1,2,3,4,7. Par ailleurs, au lieu de corriger les autocorrélations, l'algorithme de correction spectrale (fig. 4.6) peut présenter comme suit :

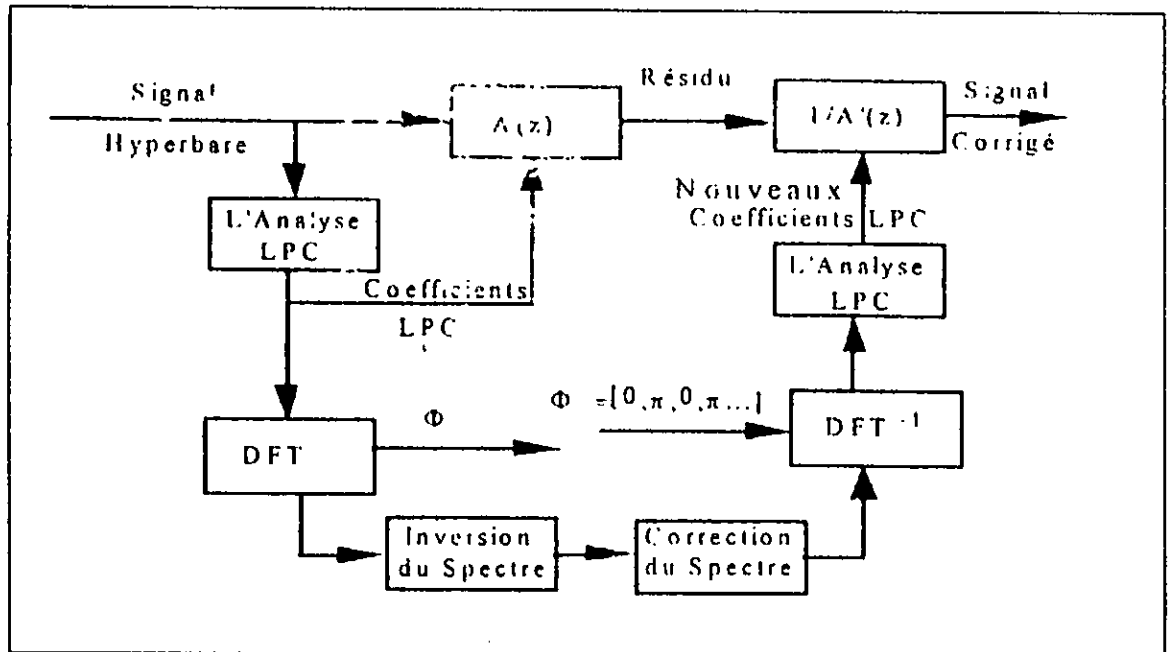


Fig. 4.6 Algorithme de correction spectrale

- les étapes 1 –2 –3 –4 (de l'algorithme précédent)
- 5. obtenir le spectre correspondant au signal hyperbare à partir des coefficients LPC du filtre non corrigé par application de DFT à la suite des coefficients  $a(j)$  ( $j = 1, \dots, p$ ) prolongée par des zéros jusqu'à la longueur de la DFT (NDFT = nombre puissance de 2) :

$$\{1, a(1), a(2), \dots, a(p), 0, \dots, 0\} \xrightarrow{DFT} A(n) = 1 + \sum_{k=1}^p a(k) e^{-j2\pi nk/N} \quad (4.9)$$

$$n = 0, 1, \dots, \text{NDFT} - 1$$

En effet, cette suite représente la réponse impulsionnelle du filtre inverse  $A(z)$  ; il suffit donc de prendre l'inverse des valeurs fournies par (4.9),

$$\text{spec}_a[n] = \frac{1}{A(n)} \quad n = 0, 1, \dots, \text{NDFT} \quad (4.10)$$

la correction de l'amplitude de l'enveloppe spectrale du conduit vocal est assurée par un filtre interpolateur, capable d'aller rechercher la composante désirée selon le taux



de compression exigé. En réalité, nous réalisons une opération de modification de la bande spectrale d'un facteur réel quelconque, le spectre corrigé juit de la relation de récurrence :

$$\begin{aligned} T_j &= \text{int}[(j+1).k] \\ \text{Fact} &= (j+1).k - \text{int} [(j+1).k] & j = 0,1,\dots,\text{NDFT}/2 \\ \text{Cspec}_a(j) &= (\text{Fact}).\text{Spec}_a(T_{j+1}) + (1-\text{Fact}).\text{Spec}_a(T_j) \end{aligned} \quad (4.11)$$

6. Revenir à l'espace temps, partir de l'amplitude du spectre corrigé et les phases artificielles obtenues par inversion des échantillons successifs, ce qui correspond à deux phase 0 et  $\pi$  pour chaque deux échantillons successifs.

7. La synthèse du signal hyperbare corrigé (comme à l'étape 7 du premier algorithme).

Nous avons étudié deux variantes de correction spectrale, une Linéaire avec indice de compression ( $k$ ) constant. La deuxième variante dite non linéaire utilise un facteur de compression ( $k$ ) constant dans la partie dite « hautes fréquences », et garde une valeur dynamique qui permet de corriger d'autre effets pour la partie dite « base fréquence ». Une fonction tabulée est utilisée cette fin inspirée de la formule de G.Fant & J.Lindquist (3.24).

## IV.6 Résultats

### IV.6.1 Description des Fichiers de Données et Conditions d'Expérimentation

Afin de pouvoir mettre en épreuve nos algorithmes de rehaussement, il a fallu disposer d'une base de données riche, constitué de signaux enregistrés dans différentes conditions hyperbares et normales.

La base de données disponible dans notre laboratoire est le PSH/DISPE [27]. Elle est issue d'une collaboration entre le laboratoire CNRS « Parole et Langage » d'Aix-en-Provence et l'Institut de Plongée Professionnelle de Marseille (I NPP). [33]

Les enregistrements qui composent cette base de données PSH ont été réalisés à l'INPP et à la Marine Nationale Française (Gismer, Toulon) en 1990-1991. ces enregistrements sont stockés sur un CDROM. Tous les fichiers sont conformes au format standard SAM Europec (Projet CEE-ESPRIT « SAM » nØ2589).

La fréquence d'échantillonnage des fichiers des signaux originaux est de 40 kHz, et le codage est fait sur 16 bits.

- Les fichiers choisis qui contiennent les phrases françaises sont « dcpf0033.sfs » et « dcpf0010.sfs », sont enregistrés à 300m en caisson à hélium et en piscine, toutes les deux bruitées.

- Les fichiers choisis qui contiennent les phrases Anglaises sont « hsep0112.pes », « hsep0102.pes » et « hsep0092.pes », sont enregistrés à 60m, 100m, 200m sans bruit.
- Les conditions d'enregistrement des fichiers sélectionnés sont illustrées au tableau 4.1 :

Fichiers Hyperbare	Profondeur (m)	TGZ (C°)	PPO2	$\rho O_2$	$\rho He$	Nombre de trames Tr=256 échantillons
Hsep0112.pes	60	29,2	592mb	8,5	91	361
Hsep0102.pes	100	25,3	600mb	5,5	93	361
Hsep0092.pes	200	30,5	399mb	5,09	95	371
Dcpf0033.sfs	300	31,5	400mb	1,31	96,4	332
Dcpf0102.sfs	300	*	737mb	2,38	93,5	338

**Tab. 4.1** Les Fichiers de test utilisés

Où TGZ : est la température du gaz ; PPO2 : la pression partielle d'oxygène ; K : le facteur de compression, et  $\rho O_2$ ,  $\rho He$  : fraction volumique de l'oxygène et de l'hélium dans le mélange.

Le test de chaque variante de nos algorithmes proposés nécessite un ensemble de paramètres et de données que l'on devra varier et adapter. Les paramètres sont contenus dans un fichier, illustré par le tableau 4.2 :

Fichier d'Entrée	Signal Hyperbare (60m-100m-200m-300m)
Fréquence d'échantillonnage	20000 Hz
K : Coefficient de correction	Formule de Fant & Lindquist
Ordre du filtre d'analyse	M = 16 ... 20
Ordre du filtre de synthèse	Mh = K*M
Fenêtre de pondération	Fenêtre de Hamming
Longueur des trames d'analyse	256 (12,8 ms)-512 (25,6ms)
LPC Overlapping	128 échantillons
La longueur de la DFT	256
Fichier de sortie	Parole Synthétique Corrigé

**Tab. 4.2** Conditions d'expérimentation

Les essais ont été faits surtout pour des fichiers de signaux hyperbares correspondant aux profondeurs 200m et 300m

#### IV.6.2 Commentaires et Interprétations des résultats

Les algorithmes de rehaussement de la parole hyperbare proposés dans les sections précédentes, ont pour seul objectif l'amélioration de la qualité de cette parole inintelligible. La qualité du signal reconstitué est devenue un concept difficile à juger et de favoriser une méthode sur une autre, et cela grâce aux techniques du traitement amélioré et la facilité de les implémenter sur des cartes DSP, en plus l'intelligibilité (compréhension de ce qui est transmis) est un concept subjectif et l'oreille humaine reste le seul juge final de l'efficacité de tels algorithmes.

Par conséquent, un système qui présente un tel compromis entre qualité du signal reconstitué et simplicité de conception est un système qu'on peut qualifier de robuste et efficace.

On peut dire que le coefficient de compression  $K$  est le paramètre le plus important, qui doit être estimé et incorporé à tous les algorithmes de corrections. Dans notre cas on a choisi  $k = 2,2$  et nous avons obtenu de bons résultats pour des profondeurs de moins de 200m. par ailleurs, les meilleures valeurs de ce paramètre pour les profondeurs supérieures à 200 m sont comprises entre 2,22 et 2,46.

Ce paramètre  $K$  peut avoir une valeur dynamique, qui s'adapte automatiquement aux conditions de la plongée, en ajoutant une procédure d'estimation de  $k$  à tous les algorithmes de correction, en basant sur les équations (3.2-3.10). Mais comme cette estimation n'est pas exacte, on se contente alors des choix par tâtonnement décrit précédemment.

Le test de la variante avec des corrections sur les autocorrélations du signal hyperbare, qui est une variante des systèmes à approche temporelle présente un très bon compromis entre la simplicité de conception, et la qualité du signal synthétisé, surtout pour de faibles profondeurs.

Par ailleurs, en agissant sur les autocorrélations donnent parfois lieu à des instabilités de l'algorithme Wiener-Levinson-Durbin (WLD), c'est pour cela qu'on remarque une naissance d'un bruit blanc situé dans la partie haute du spectre du signal corrigé (Fig.4.9.b). auparavant, il est préférable d'utiliser un filtre passe bas a fin de réduire ce bruit.

Le test de la variante avec des corrections linéaires de spectre du signal hyperbare, qui est une variante des systèmes à approche fréquentielle nécessite des transformées de Fourier de 256 échantillons, a fin d'estimer l'enveloppe spectrale du conduit vocal à partir des  $p$  coefficients LPC prolongés par des zéros jusqu'à la longueur de DFT. Après les essais qu'on fait, on a remarqué que le changement de coefficient de compression

n'agit pas sur le signal corrigé, et que même sans filtrage cette variante donne presque le même résultat que la variante précédente (fig. 4.9.c). On aurait pu dire que cette méthode est meilleure que la précédente si cet algorithme n'était pas plus long que celui des corrections des corrélations, cela est dû au passage au domaine fréquentiel imposé par la DFT.

Les différents algorithmes de traitement ont été tout d'abord optimisés en choisissant les paramètres qui leur permettent d'obtenir les meilleurs résultats possibles.

Bien que le résultat soit jugé en comparant les signaux corrigés avec le signal original, nous avons voulu comme même consolider ce jugement avec des tests subjectifs, en entendant les signaux corrigés, cela nous a été possible grâce à un logiciel qui lit les fichiers de la base de donnée fig.4.9.

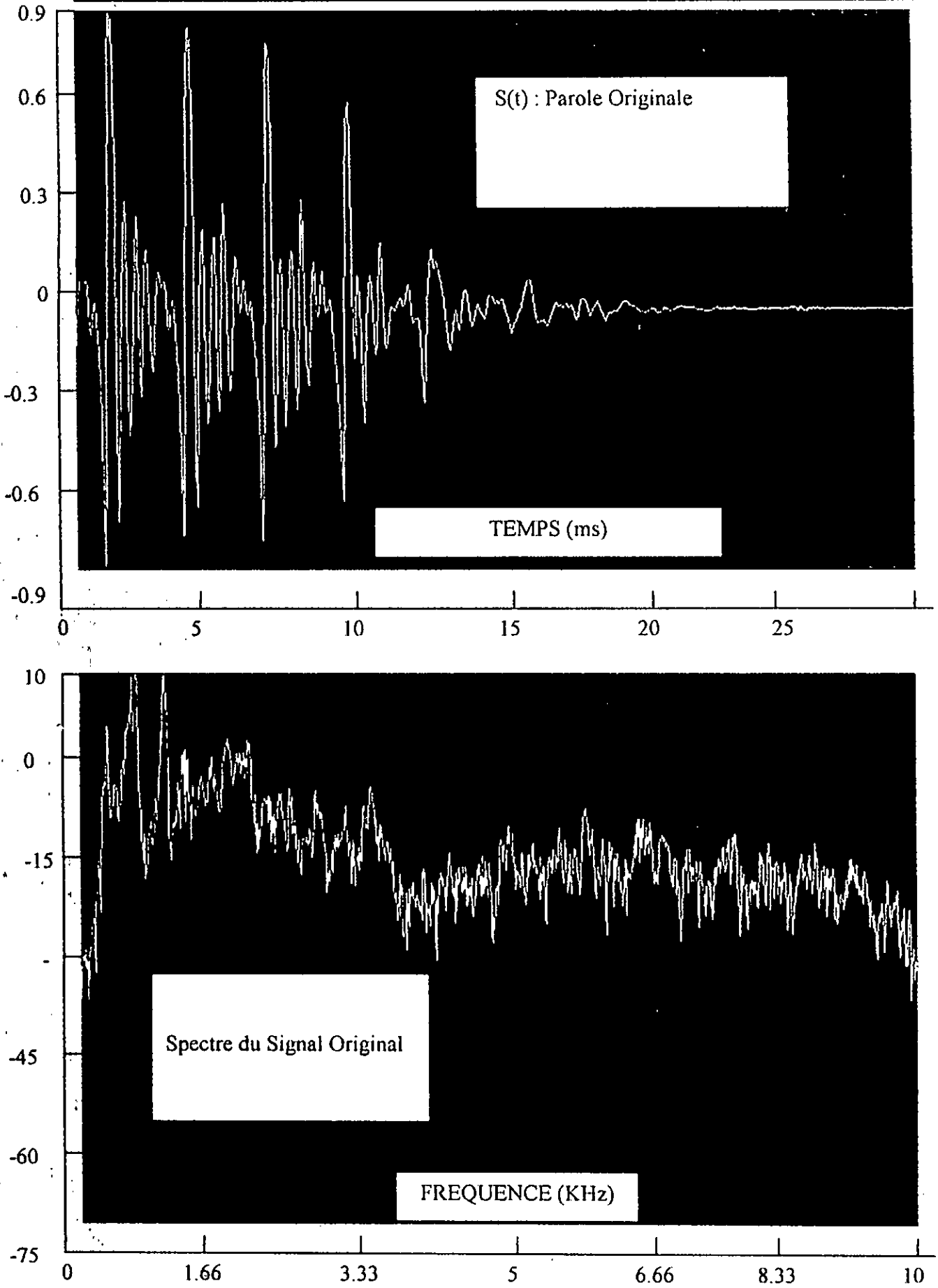


Fig. 4.9.a Exemples d'une tranche du signal original et son spectre

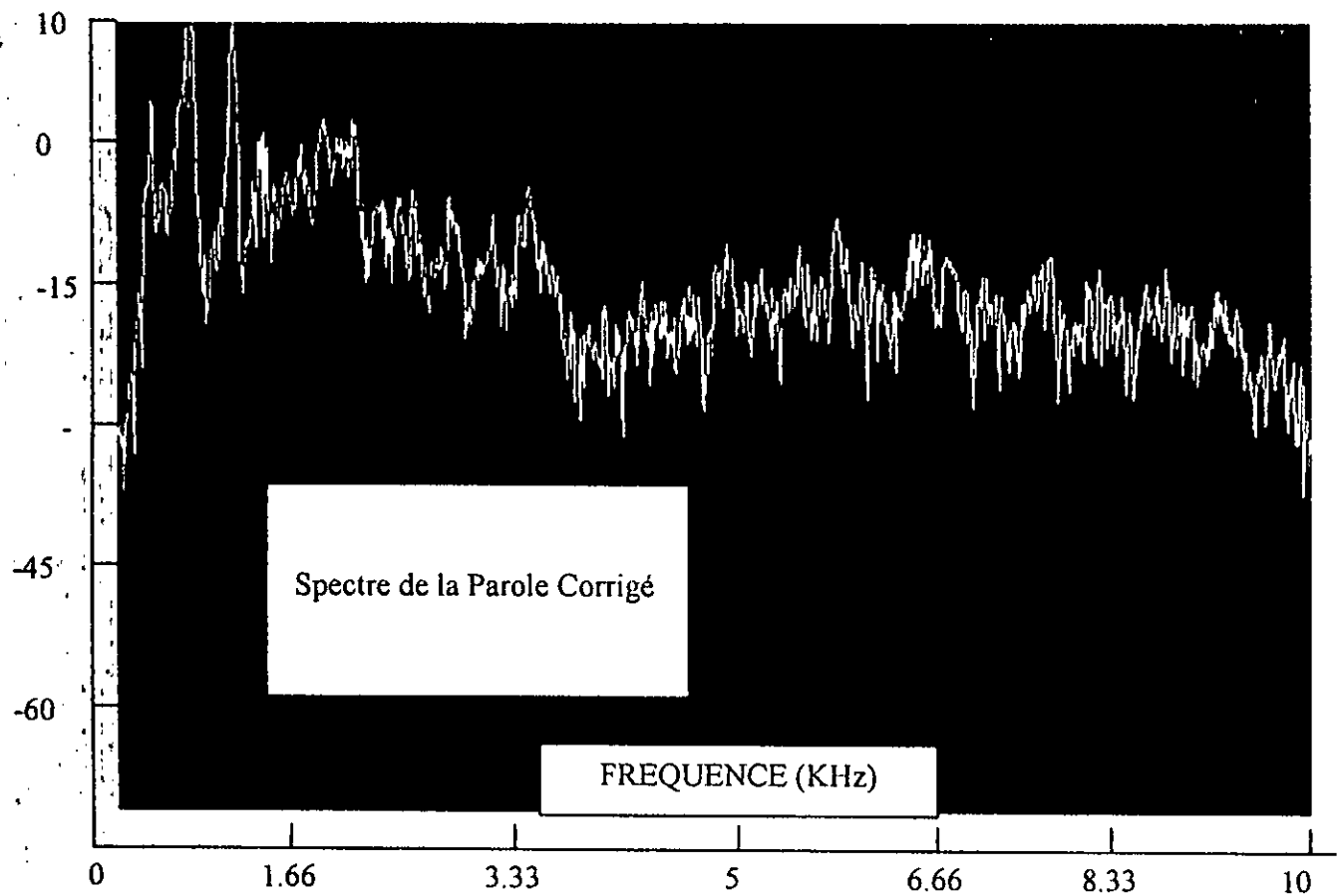
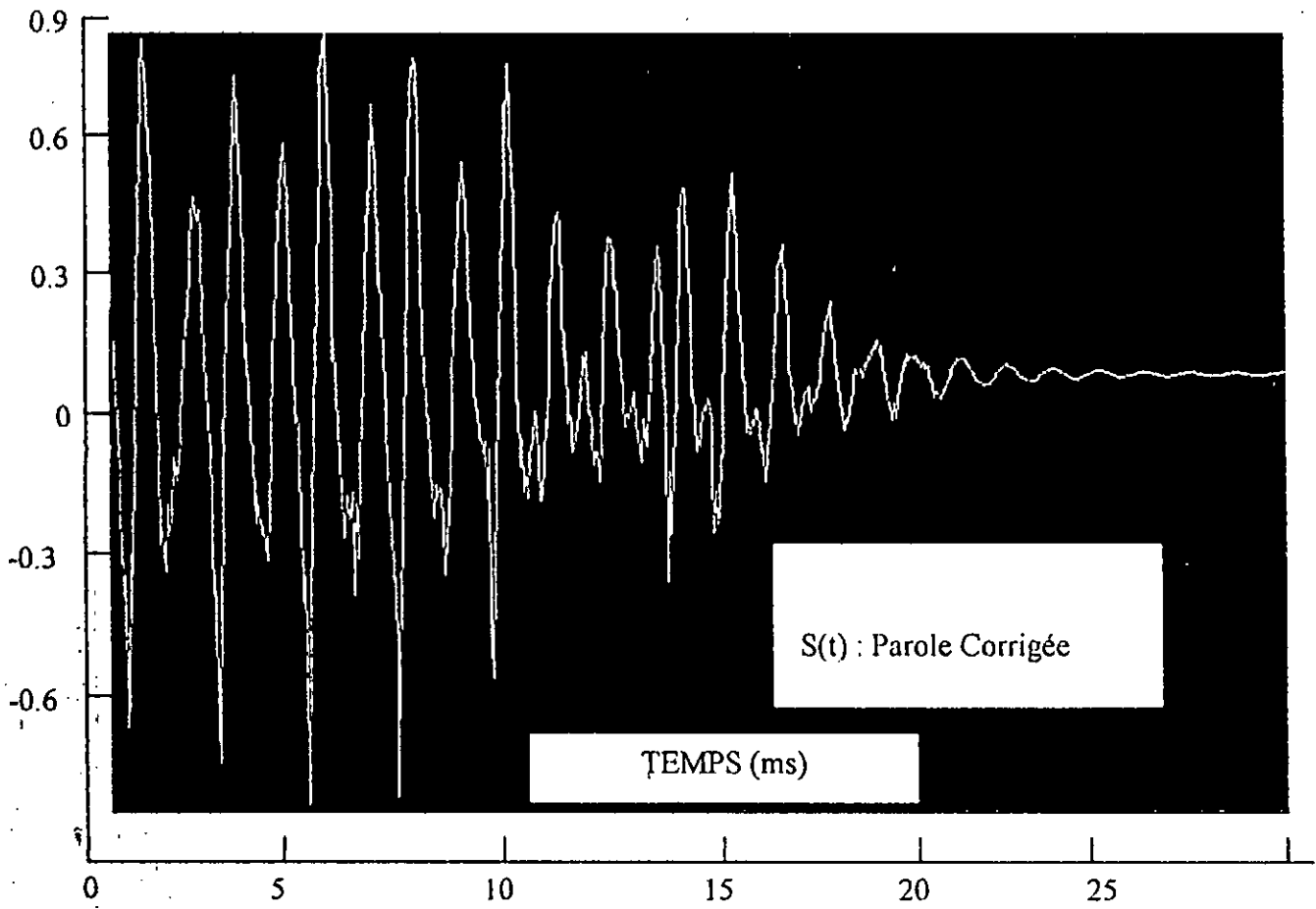


Fig. 4.9.b Exemples des résultats obtenus avec correction des autocorrélations 47

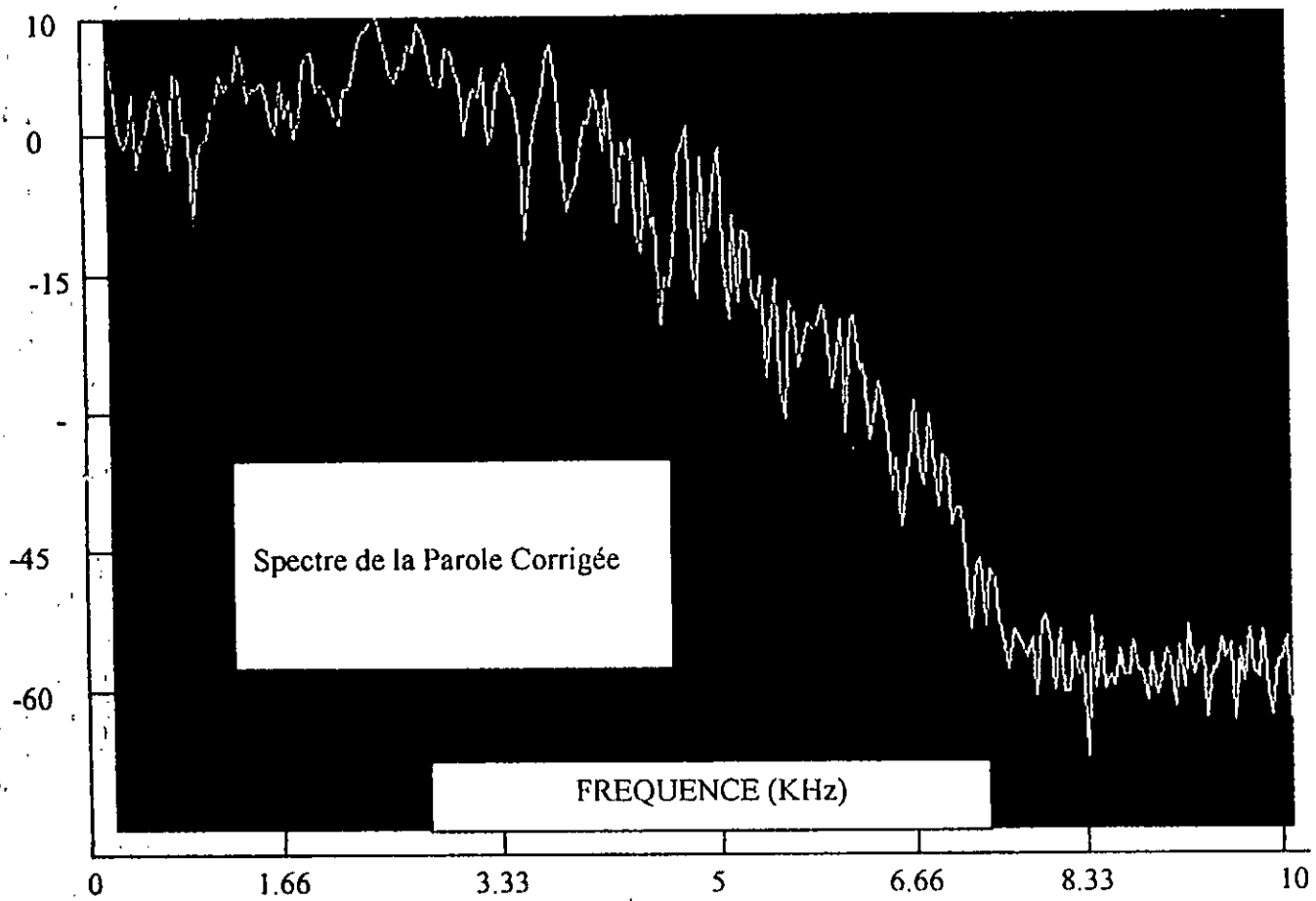
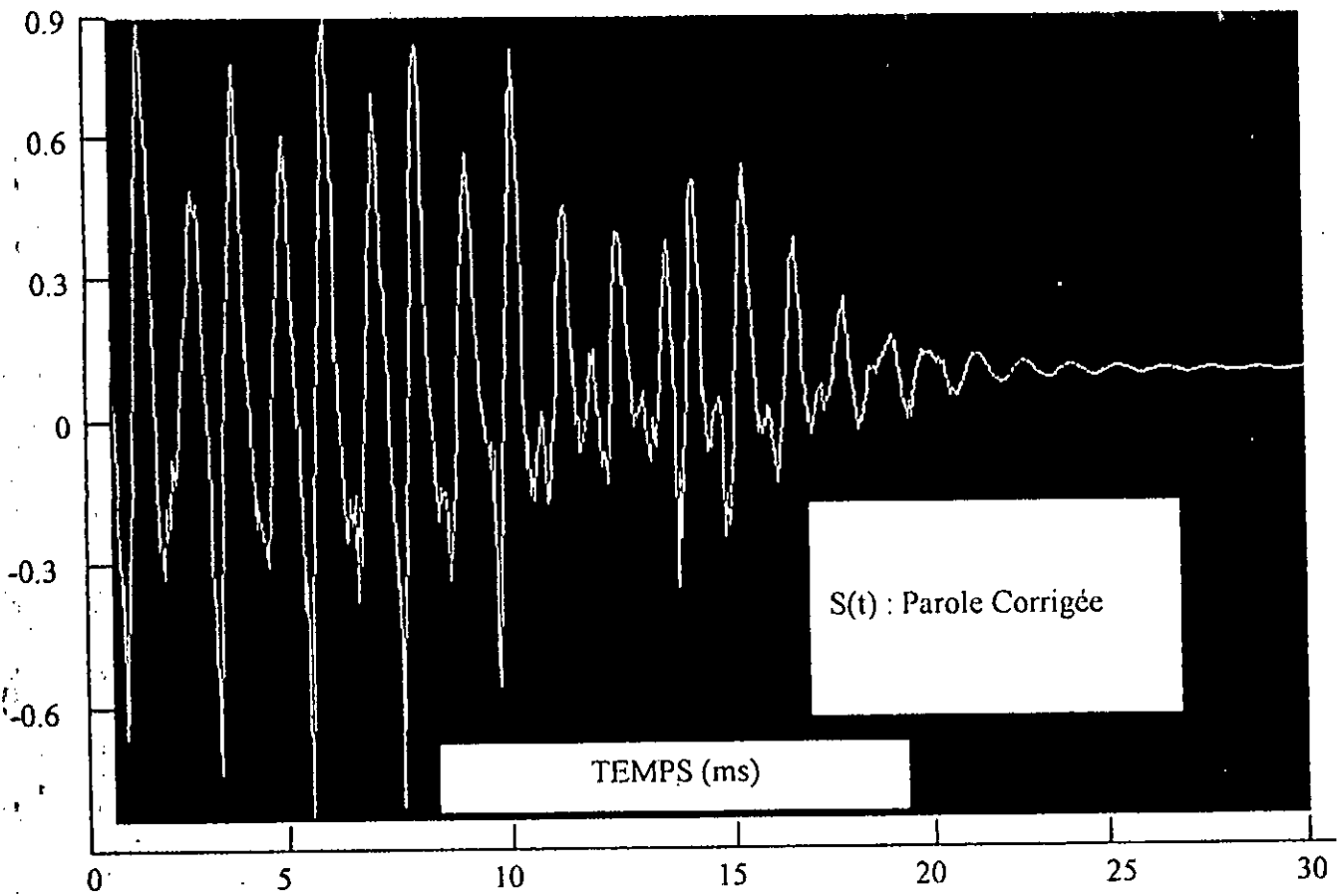


Fig. 4.9.c Exemples des résultats obtenus avec correction spectrale

### **IV.7 Conclusion**

En conclusion à ce chapitre, on a remarqué que l'intelligibilité des signaux corrigés est très bonne pour toutes les variantes. Pour des profondeurs inférieures à 200 m, au vu de la qualité des résultats obtenus, les systèmes utilisant la correction avec compression uniforme (ou linéaire) peuvent remplacer la variante utilisant la correction non linéaire. De plus, Si on se contente d'une certaine qualité du signal synthétisé au prix d'un traitement simple et rapide, la variante de correction des corrélations est la plus sollicitée.

Partant de résultats déjà obtenus avec les systèmes de rehaussement proposés. et vu l'intérêt stratégique que représente la parole hyperbare, la qualité des signaux synthétisés peut être jugée satisfaisante par rapport à la parole originale qui aurait pu être prononcée dans les conditions normales, malgré qu'on ne peut pas identifier correctement le locuteur. Tâche à laquelle on ne s'intéresse pas.

Notons enfin, que les signaux utilisés jusqu'à maintenant ne sont pas contaminés par des bruits du milieu hyperbare, mais il est primordial de tenir compte du contexte dans sa globalité, en intégrant à la démarche de correction, les systèmes de réduction de bruit, qui doivent être efficaces.



## CHAPITRE V

# ***Réduction de bruit d'un signal hyperbare***

## V.1 Introduction

Jusqu'à présent, on a illustré les méthodes du rehaussement du signal hyperbare sans évoquer le problème de la suppression du bruit qui contamine le signal utile. Les différents milieux où les signaux utiles sont prononcés se trouvent très hostiles, à savoir le milieu ambiant, l'hélicoptère, la voiture (téléphone main-libre),... quand au milieu hyperbare, il est réverbérant et présente de plus un bruit acoustique de niveau élevé, ce qui affecte encore la qualité de la parole hyperbare, déformée déjà par les conditions de la plongée.

L'amélioration de la qualité de cette parole hyperbare bruitée résultante, peut être assurée par des systèmes de réduction du bruit, qui seront incorporés aux techniques de correction proposées. Sans trop s'étaler sur le sujet on exposera deux versions de réduction du bruit dans la première et la deuxième parties de ce chapitre. Toutes les deux réalisent des corrections dans le domaine temporel, la première consiste en une pondération du filtre d'analyse LPC, la deuxième par contre, utilise une nouvelle représentation cohérente modifiée de la parole hyperbare. Enfin un système basé sur la soustraction spectrale est utilisé.

## V.2 Caractéristiques du bruit dans un milieu hyperbare

Le bruit est tout phénomène perturbateur gênant l'interprétation ou la perception d'un signal utile. Dans notre cas, le signal hyperbare déformé déjà par les conditions sous marines, devient encore plus inintelligible suite au bruit de niveau élevé qui contamine le signal hyperbare utile, surtout la gamme des faibles fréquences.

Ce bruit peut être la combinaison de plusieurs sources à savoir : le bruit de respiration, le bruit des machines, et le bruit des océans. De plus, l'effet de la chambre et le masque utilisés dans la plongée sous- marine affectent le milieu qui devient réverbérant.

En fin, le microphone utilisé pour la transmission sous- marine, est un facteur essentiel, qui peut être une source importante du bruit d'ou des pertes d'information du signal hyperbare. La figure 5.1.a . montre l'effet du bruit, à l'aide du spectrogramme de la phrase (tous les ans, on envoie mille tonnes de riz au Mali) à 300m en piscine par rapport à celui dans les conditions normales.

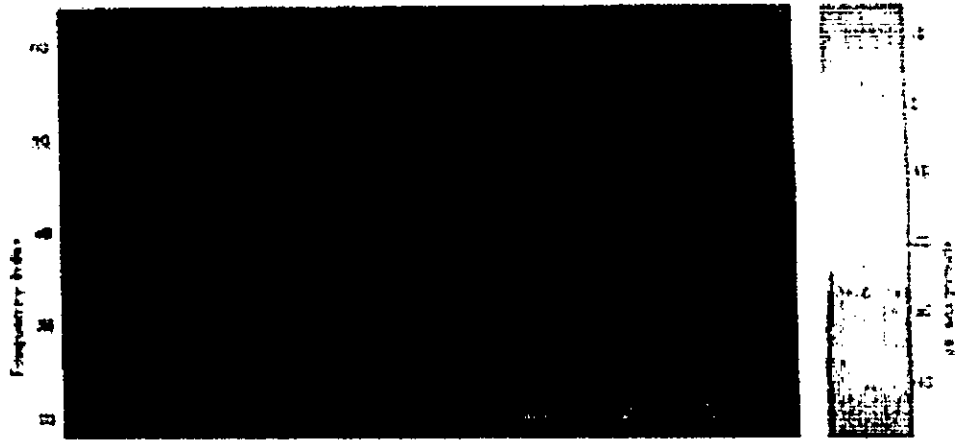


Fig .5.1 .a Spectrogramme de signal bruité

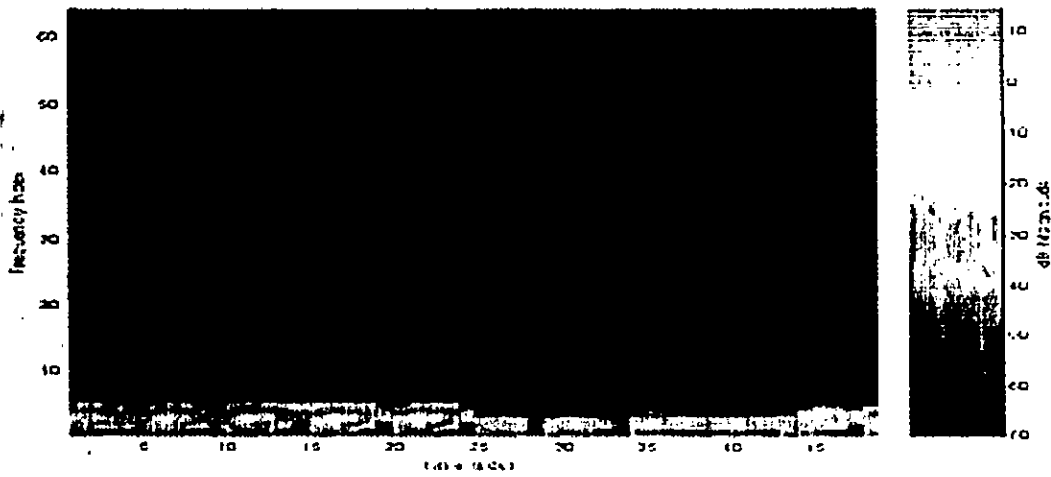


Fig.5.1.b Spectrogramme de signal original dans les conditions normales

### V.3 Réduction de bruit dans le domaine temporel

L'algorithme de correction des autocorrélations développé dans le chapitre précédent, présente en plus de sa simplicité d'implémentation, un bon potentiel pour le rehaussement de la parole hyperbare dans un milieu hostile, en particulier concernant la suppression de bruit, résultat de la non corrélation entre les échantillons du bruit.

Deux variantes ont été testées, l'inconvénient principal de l'algorithme est la haute distorsion introduite par l'opération de l'autocorrélation qui rend la parole rehaussée moins agréable pour l'écoute.

#### V.3.1 Pondération du résidu

L'écoute simultanée de deux sons de fréquences différentes mais assez proches, rend le son de faible amplitude inaudible. Cet effet de masquage fréquentiel peut être total ou partiel et dépend des intensités et des fréquences des sons.

Le principe de masquage peut être utilisé pour réduire les divers bruits qui se rajoutent au signal hyperbare. Le bruit perçu provient pour la plus part des zones de fréquences où le signal est faible. Il est nécessaire donc de tenir compte du spectre du bruit. Le bruit situé dans les zones de formant peut être totalement ou partiellement masqué par le signal de parole.

On utilise un filtre d'analyse de pondération perceptuelle  $F(z)=A(z/\gamma)$  pour redistribuer la puissance du bruit (fig.5.2). On réduit la densité spectrale du bruit dans les zones de fréquence où le niveau du signal est faible, tout en l'augmentant dans les zones des formants où le signal est fort et qui peut masquer efficacement le bruit.

Notons que  $F(z)$  est défini à partir de  $A(z)$  en ramenant ses zéros vers le centre du cercle unité du plan  $Z$  de plus, le facteur  $\gamma$  est compris entre 0 et 1.

La figure 5.2 schématise l'algorithme de correction des corrélations modifié :

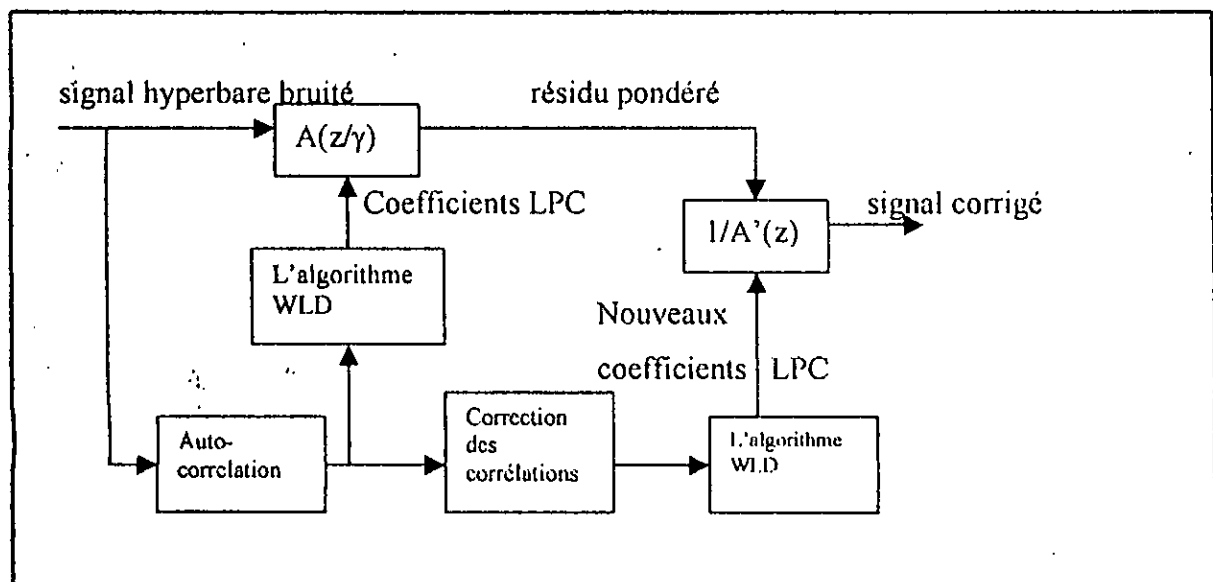


Fig5.2 la correction des corrélations avec filtre de pondération

Le nouveau résidu est donné par :  $\text{residu}_w(n) = X_u(n) + \sum_{j=1}^p a(j) \gamma^j X_u(n-j)$  (5.1)

pour  $n=0,1,\dots,N$

L'objectif est de déterminer la valeur optimale de  $\gamma$  en fonction du rapport signal sur bruit (RSB). Soong et Sondhi [31] par exemple ont utilisé les valeurs de  $\gamma$  qui minimisent la distorsion pondérée d'Itakura entre le modèle AR sans bruit et l'estimé AR du signal bruité.

Dans notre cas, on a opté pour un choix par tâtonnement de  $\gamma$ , en évitant le calcul du RSB pour chaque trame. Les valeurs de  $\gamma$  entre 0.5 et 0.8 donnent des résultats satisfaisants. L'inconvénient majeur de cette méthode est la légère variation des largeurs des bandes des formants. De plus, vu que le spectre du filtre est incliné, il joue un rôle d'un filtre passe bas, ce qui assourdit la parole synthétisée résultante, d'où l'utilisation d'un filtre de Pré-accentuation. Ce post-traitement cause en général une amplification du signal qui sera compensée par la technique du contrôle automatique du gain entre la trame analysée et celle synthétisée, soit échantillon par échantillon soit bloc par bloc.

### V.3.2 La représentation SMC de la parole hyperbare bruitée

Cette méthode est version de la corrélation qu'on a adaptée au signal hyperbare dans un milieu hostile. Elle permet de réduire l'effet du bruit additif de RSB faible, en exploitant les propriétés de la corrélation vis à vis du bruit, mais avec une nouvelle représentation cohérente modifiée de la fonction d'auto corrélation (SMC), du signal bruité au lieu de la corrélation court-terme utilisée jusqu'à maintenant (2.16). Cette représentation (annexe b) a été utilisée pour la première fois en 1989 dans la reconnaissance de la parole bruitée [28], et a été présentée comme suit :

- A partir de  $2.N$  échantillons, on calcule  $N+1$  coefficients de corrélation selon la relation suivante :

$$\rho(i) = \frac{1}{N} \sum_{j=0}^{N-1} S(j)S(j+1) \quad \text{pour } i=0,1,\dots,N \quad (5.2)$$

- On applique une fenêtre de Hamming à la séquence d'autocorrélation :

$$\rho^h(i) = \rho(i) \left( 0.54 - 0.46 \cos\left(\frac{2\pi i}{N+1}\right) \right) \quad \text{pour } i=0,1,\dots,N \quad (5.3)$$

- Le calcul de la DFT de la séquence  $\rho^h(i)$  :

$$R(k) = \sum_{i=1}^{NDFT} \rho^h(i) \omega^{ki} \quad \text{pour } k=0,1,\dots,NDFT \quad (5.4)$$

Ou  $NDFT$  est la longueur de la DFT et  $\rho^h(0)$  n'est pas incluse dans la somme, particulièrement dans ce cas ou le signal est affecté par un bruit blanc additif. Dans les autres cas ceci n'est pas nécessaire.

- Le calcul de la DFT inverse de l'amplitude de  $R(K)$  :

$$\overline{\rho}(i) = \sum_{k=0}^{NDFT} R(k) |\omega^{-ki}| \quad \text{pour } i=0,1,\dots,p \quad (5.5)$$

- A partir de  $\overline{\rho(i)}$  on applique la LPC et la suite de l'algorithme de correction.

Cette représentation exige des calculs plus complexes par rapport à ceux de l'analyse LPC standard, qu'on peut réduire en utilisant la technique d'overlap-save. La complexité de l'algorithme SMC est  $2.N.(Ln N+2)$  pour l'étape 1,  $(N+1)$  multiplication pour l'étape 2, et  $2.N. Ln N$  multiplication additions et  $(N+1)$  opérations de valeurs absolues complexes pour la troisième et la quatrième étape. Si on suppose que l'opération de la valeur absolue prend 4 multiplications- additions, donc la complexité de l'algorithme SMC est  $4.N.(ln N+2)$  par rapport à  $P.N$  de l'analyse LPC. Néanmoins, les tests d'écoutes confirment la robustesse de cette représentation vis à vis du bruit par rapport aux résultats obtenus avec correction des corrélations seulement.

## V.4 Réduction de bruit dans le domaine spectral

### V.4.1 La soustraction spectrale

La soustraction spectrale est une famille des techniques de réduction du bruit dans le domaine fréquentiel. Elle consiste à soustraire le spectre d'amplitude court- terme estimé du bruit à partir du spectre d'amplitude de la parole bruitée. La phase originale du signal bruité est intacte vis à vis du bruit.

Comme le milieu hyperbare est très hostile, la parole hyperbare produite dans de tel environnement se trouve contaminée par un bruit acoustique de niveau élevé. Alors, la technique de la soustraction spectrale semble la plus appropriée pour résoudre ce problème, car elle utilise le spectre court terme. Par conséquent, elle devient facile pour l'intégrer au système de rehaussement basé sur la correction du spectre, en plus de son potentiel d'adaptation aux variations lentes des caractéristiques du bruit.

Le signal hyperbare  $X(k)$  peut être considéré comme étant la somme d'un signal hyperbare  $S(k)$  non bruité et un signal du bruit additif  $N(k)$ . Les deux signaux  $S(k)$  et  $N(k)$  sont supposés décorrélés, et  $S(k)$  de moyenne nulle, avec  $X(w)$  leurs spectres respectifs.

$$X(k) = S(k) + N(k) \quad (5.6)$$

Tout d'abord le spectre de puissance court terme est donné par :

$$|X(w)|^2 = |S(w)|^2 + |N(w)|^2 \quad (5.7)$$

Par ailleurs, si  $N(w)$  est connu, donc  $S(k)$  peut être facilement estimé par la relation suivante :

$$|S(w)|^2 = |X(w)|^2 - |N(w)|^2 \quad \text{et } S(k) = |S(k)| e^{j \arg(X(k))} \quad (5.8)$$

L'équation (5.8) montre que l'amplitude du signal non bruité est atténuée par l'amplitude du bruit estimé, et la phase originale est retenue. La figure 5.3 illustre le schéma de base de tous les algorithmes de soustraction spectrale. Ils diffèrent par la manière d'estimation du spectre du bruit et par le fait que certains utilisent des spectres de puissances, alors que d'autres utilisent des spectres d'amplitude.

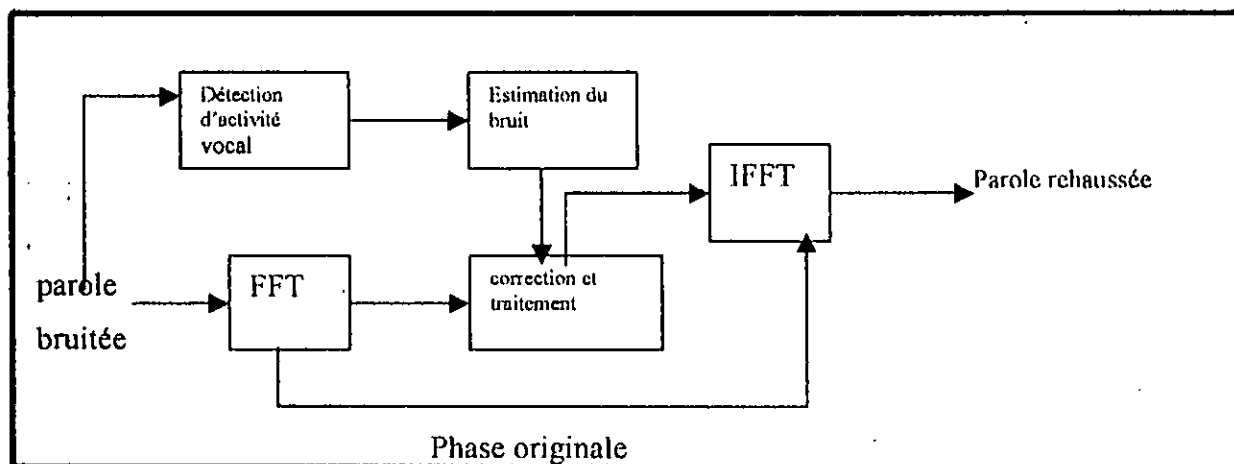


Fig 5.3 le principe de base de la soustraction spectrale

Cette technique a été largement étudiée par les auteurs [17-18-19], et par plusieurs d'autres. La forme générale du spectre d'amplitude estimé est donnée par :

$$|S(n)|^{\lambda} = |X(n)|^{\lambda} - |N(n)|^{\lambda} \quad , \quad n=0,1,2,\dots[(NDF/2) - 1] \quad (5.9)$$



Où  $|N(n)|$  est la  $n^{i\text{ème}}$  composante du spectre d'amplitude du bruit, et  $|X(n)|$  celle du signal d'entrée bruité et  $|S(n)|$  celle du signal de sortie désiré. Les deux paramètres  $\mu$  et  $\lambda$  doivent être  $\geq 1$ . Berouti Schwartz, et Makhoul[19] ont utilisé  $\mu=2$  avec  $\lambda$  variable ; Boll [17] a utilisé  $\mu=1$  et  $\lambda=1$ .

Le principe général de la soustraction spectrale selon Boll est donné par la relation suivante :

$$|S_i(n)| = \begin{cases} |X_i(n)| - |N_i(n)| & \text{si } |X_i(n)| \geq |N_i(n)| \\ \beta |N_i(n)| & \text{ailleurs} \end{cases} \quad n=0,1,\dots [N_{DFT}/2 - 1] \quad (5.10)$$

La méthode de soustraction spectrale que nous avons intégrée au système complet de réduction et de correction est celle de Berouti, qui peut se formuler de la façon suivante pour chaque trame  $i$  :

$$|S_i(n)|^2 = \begin{cases} |X_i(n)|^2 - \lambda_i |N_i(n)|^2 & \text{si } |X_i(n)|^2 - \lambda_i |N_i(n)|^2 \geq |N_i(n)|^2 \\ \beta |N_i(n)|^2 & \text{ailleurs} \end{cases} \quad n=0,1,\dots N_{DFT}/2 \quad (5.11)$$

Où  $\lambda_i > 1$  et  $\beta \ll 1$ . Vu que  $\lambda_i$  est supérieur à l'unité, une surestimation du spectre du bruit est assurée dans l'équation (5.11). Cela induit une augmentation du degré de suppression du bruit et qui peut causer des distorsion supplémentaires de la parole synthétisée. Cet inconvénient est remédié par une adaptation du paramètre  $\lambda_i$ , en se basant sur un compromis entre une suppression maximale du bruit tout en gardant un minimum de distorsion. Le paramètre  $\lambda_i$  est choisi en fonction du rapport signal sur bruit (RSB) de la trame courante  $i$  comme dans [19]. Les trames de RSB faible peuvent être considérées comme des trames de bruit seul, donc  $\lambda_i$  est choisi élevé pour maximiser la suppression du bruit. Les trames de RSB élevé sont considérées comme des trames de parole seule, et  $\lambda_i$  est choisi faible à fin de minimiser la distorsion.

Le RSB de chaque trame est calculé par le rapport suivant :

$$RSB_i = 10 \text{Log}_{10} \left[ \frac{EX_i - EN_i}{EN_i} \right] \quad (5.12)$$

ou  $EX_i = \sum_{n=0}^N X_i^2(n)$  est l'énergie de la trame courante et  $EN_i = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |N_i(w)|^2 dw$ , l'énergie du bruit estimée. Enfin, le paramètre  $\lambda_i$  est donné par la relation suivante :

$$\bar{\lambda} = \lambda_0 \left| \frac{\lambda_0 - \lambda_{\min}}{RSB_0} \right| RSB,$$

$$\lambda_i = \begin{cases} \lambda_{\min} & , \quad \bar{\lambda} > \lambda_{\min} \\ \bar{\lambda}_i & , \quad \lambda_{\min} \leq \bar{\lambda}_i \leq 5 \\ 5 & , \quad \bar{\lambda}_i > 5 \end{cases} \quad (5.13)$$

La constante  $\lambda_0$  donne la valeur de  $\bar{\lambda}_i$  lorsque  $RSB=0$ . lorsque le RSB augmente, le degré de suppression de bruit diminue jusqu'à ce que le RSB atteigne la valeur  $RSB_0$ , ou  $\lambda_i$  prend sa valeur minimale  $\lambda_{\min}$ . De même, lorsque le RSB décroît, le degré de suppression du bruit augmente jusqu'au maximum ( $\lambda=5$ ).

Il apparaît clairement que la mesure du RSB est utilisée comme une décision 'SOFT', qui permet de discriminer les trames de parole et celles qui contiennent du bruit. Cette séparation est efficace dans le cas des sons voisés, mais pour les sons non voisés, il est difficile dans un environnement très bruité de détecter les fricatifs (/s/, /k/, /t/, ...) du bruit. On est amené donc à améliorer le contrôle de la valeur de  $\lambda_i$  par d'autres mesures simples, afin de conserver les sons de faible RSB. De plus, il est important de mettre à jour l'estimé de  $|N_i(n)|^2$  et donc de  $EN_i$ . Par conséquent, la mesure du RSB reste correcte.

Le choix des paramètres utilisés dans l'équation (5.13) est :  $\lambda_0=4$ ,  $\lambda_{\min}=1$  et  $RSB_0=20$ db. Le paramètre  $\beta$  est généralement compris entre 0.005 et 0.05, ceci limite la suppression maximale entre -23db et -13db.

#### V.4.2 Estimation du spectre du bruit

Le dernier chaînon dans l'implantation de la technique de soustraction spectrale est l'estimation du spectre du bruit  $|N_i(n)|^2$  (et son énergie  $EN_i$ ). La méthode la plus efficace exige un certain nombre de trames de silence (généralement=10) au début du fichier sous test, afin de calculer l'estimé initial du spectre du bruit en moyennant  $|X_i(n)|^2$  sur ces trames :

$$|N_i(n)|^2 = \frac{1}{N_{frames}} \sum_{k=0}^{N_{frames}-1} |X_k(n)|^2, \quad n=0,1,\dots,NDFT/2 \quad (4.15)$$

Une adaptation de cet estimé permet de tenir compte des variations lentes des caractéristiques du bruit, la relation suivante assure cette adaptation :

$$|N_i(n)|^2 = \frac{1}{i} \left( (i-1) |N_{i-1}(n)|^2 + |X_i(n)|^2 \right), \quad n=0,1,\dots,NDFT/2 \quad (5.15)$$

### V.4.3 Algorithme d'amélioration basé sur la soustraction spectrale

La soustraction spectrale est incorporée dans l'approche fréquentielle (fig. 4.6). L'algorithme global de réduction du bruit et de correction spectrale, proposé est illustré par la figure (5.4).

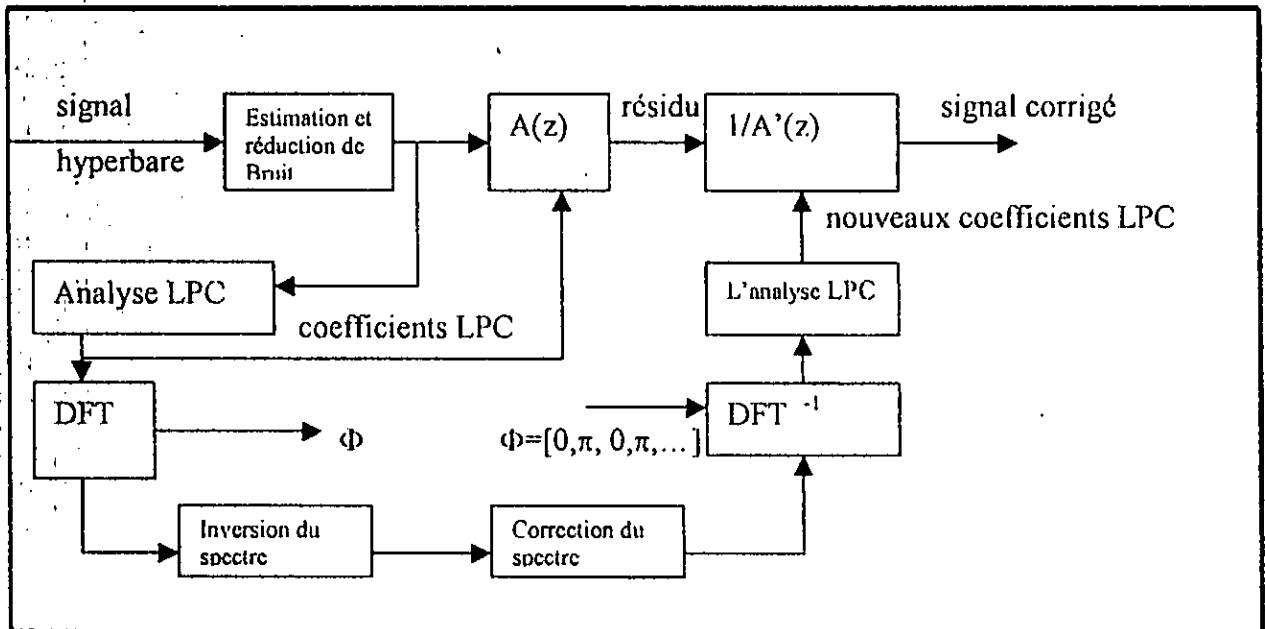


fig 5.4 correction spectrale avec réduction du bruit

L'algorithme utilisé dans le bloc de réduction de bruit est le suivant :

1. trame<sub>i</sub>=1 ; N<sub>trame</sub>=10 ;
2. lecture de N échantillon de la trame i : X<sub>i</sub>(n) : n=0,1...N ;

3. le calcul de l'énergie de la trame :  $EX_i = \sum_{n=0}^N X_i^2(n)$

4. le calcul de la DFT de  $X_i(n)$  :  $X_i(k)$  :  $k=0,1 \dots \text{NDFT}-1$  ;

5. si trame  $i >$  nombre de trame de mise à jour ( $N_{\text{trames}}$ ), GO TO 9 ;

**L'estimation du spectre du bruit**

6. la mise à jour du spectre du bruit :

$$|N_i(n)|^2 = \frac{1}{i} \left( (i-1) |N_{i-1}(n)|^2 + |X_i(n)|^2 \right), \quad n=0,1, \dots \text{NDFT}/2$$

ou  $|N_0(n)|^2 = 0, \forall n$

7. la mise à jour de l'énergie du bruit estimée :

$$EN_i = \frac{1}{i} \left( (i-1) EN_{i-1} + EX_i \right), \quad \text{avec } EN_0 = 0$$

8. aller à l'étape 12 :

**La suppression du bruit par soustraction spectrale**

9. Le calcul du RSB de la trame  $i$  :

$$RSB_i = 10 \text{ Log}_{10} \left[ \frac{EX_i - EN_i}{EN_i} \right]$$

10. Le calcul du facteur  $\lambda$  de soustraction :

$$\bar{\lambda} = \lambda_0 \cdot \left| \frac{\lambda_0 - \lambda_{\min}}{RSB_0} \right| RSB_i$$

$$\lambda_i = \begin{cases} \lambda_{\min} & , \quad \bar{\lambda} > \lambda_{\min} \\ \bar{\lambda} & , \quad \lambda_{\min} \leq \bar{\lambda} \leq s \\ s & , \quad \bar{\lambda} > s \end{cases}$$

11. Le calcul de l'estimé du spectre du signal désiré :

$$|S_i(n)|^2 = \begin{cases} |X_i(n)|^2 - \lambda_i |N_i(n)|^2 & \text{si } |X_i(n)|^2 - \lambda_i |N_i(n)|^2 \geq |N_i(n)|^2 \\ \beta |N_i(n)|^2 & \text{ailleurs} \end{cases} \quad n=0,1,\dots,NDFT/2$$

12. Le retour dans le domaine temporel et la suite de l'algorithme de correction .

13. (trame)<sub>j</sub> ← (trame)<sub>j+1</sub> : aller à 2.

illustre le spectrogramme d'un signal synthétisé avec l'algorithme proposé, à partir de la parole hyperbare bruitée à 300m . La nécessité d'avoir plusieurs fichiers de bruit différents, nous permet d'avoir une meilleure adaptation des paramètres de notre algorithme, et de comparer nos résultats avec ceux qui existe dans les publications.

## V.5 Conclusion :

C'est à partir de tests d'écoutes que le résultat sonore final peut être apprécié et jugé. Ainsi, ces tests montrent l'efficacité des algorithmes de réduction de bruit proposés dans ce chapitre. Une attention particulière doit être apportée à l'approche du filtrage perceptuel du résidu, qui en plus de sa rapidité et sa simplicité d'implantation, est très efficace vis à vis de la suppression du bruit. La représentation cohérente modifiée (SMC) assure une qualité synthétique de la parole comparable à celle obtenue avec la première variante , l'inconvénient principal reste le temps de calcul nécessaire à son implantation . de plus , les résultats obtenus par l'incorporation de la soustraction spectrale sont satisfaisants , et pourraient être améliorés dans les travaux futurs , par l'utilisation d'algorithmes d'estimation et de séparation du bruit très efficaces à savoir la VAD (annexe C).

# Conclusion Générale

Ce travail accompli a été la mise au point des systèmes de rehaussement de la parole hyperbare.

Afin d'aborder cette étude, nous avons commencé par une présentation générale des notions de base du traitement de la parole, dans les conditions normales. On a montré que l'approximation du conduit vocal par un modèle tout pôle (AR) est valable dans le cas du signal vocal et hyperbare. En plus, on a détaillé l'analyse LPC qui est utilisée pour l'estimation des paramètres  $a_1$  et qui nécessite plusieurs variables : à savoir une méthode d'autocorrélation pour l'analyse ; une fenêtre de Hamming pour le fenêtrage ; un nombre N (256 ou 512) d'échantillons par trame, correspondant à 12,8ms et 25,6ms ( pour une  $F_s = 20\text{KHz}$  ) ; un ordre P de prédiction entre 16 et 20. Enfin, une pré accentuation est envisagée pour compenser la pente globale négative du spectre de 'a parole.

Le chapitre suivant a été consacré aux caractéristiques principales du signal hyperbare, par rapport à celles du signal vocal. Cette étude nous a permis de faire certaines constatations à savoir : un étalement du spectre du signal hyperbare d'environ trois fois, par rapport au signal vocal ; que cette transposition spectrale est une fonction non linéaire que le pitch n'est pas un facteur dans le rehaussement et que le milieu hyperbare est contaminé par un bruit de fond très élevé

Les algorithmes de rehaussement développés dans le chapitre IV, nous ont permis d'améliorer l'intelligibilité des signaux synthétisés pour les profondeurs les plus utilisées en plongée sous-marine, entre 60m et 300m. La variante de correction des corrélations est la plus simple et la plus rapide, la deuxième variante qui applique une compression linéaire du spectre, permet d'avoir de bons résultats au prix d'un traitement plus lent dû à la DFT.

Enfin, notre travail ouvre d'autres perspectives pour d'autres travaux de corrections, à savoir :

- ❖ L'utilisation d'une méthode hybride combinant les propriétés du domaine temporel et spectral.
- ❖ Pour des profondeurs de plus de 300m et 500m, par manque de fichiers correspondant, les

algorithmes proposés peuvent montrer quelques insuffisances qui peuvent être comblées par une meilleure adaptation.

- ❖ Dans le cas des milieux hyperbares très bruités, il se peut que les systèmes de débruitage classiques s'avèrent inefficaces. Par ailleurs, la séparation de source, qui est une technique relativement récente dans le domaine du traitement de signal, qui vise à extraire des sources statiquement indépendantes à partir d'un mélange linéaire de ces sources. Et comme le signal hyperbare et certains types de bruits sont indépendants et linéairement mélangés dans le milieu hyperbare, il apparaît donc tout à fait logique de pouvoir appliquer cette méthode dans le cas de la parole hyperbare, comme plusieurs travaux récents consacrés à la téléphonie main - libres en voiture.

## Annexe A

### La Prédiction Linéaire (LP)

L'algorithme général de la prédiction linéaire adopté dans notre étude peut être résumé comme suit :

- Acquisition du signal hyperbare  $xh(n)$ .
- Lecture d'une fenêtre de longueur  $N$ ,
- Préaccentuation éventuelle ( $\mu$ ),
- Application de la pondération par une fenêtre de Hamming de longueur  $L = N + 2.D$  où  $D$  : est le débordement de cette fenêtre de chaque cotée de la fenêtre d'analyse.
- Le calcul de la matrice d'autocorrélation selon (2.16)

$$R(i) = \sum_{n=0}^{N-1-i} s(n).s(n+i) = \sum_{n=i}^{N-1} s(n).s(n-i). \quad \text{pour } i = 0, 1, \dots, p$$

- La résolution du système (2.18) par l'algorithme de WLD,
- On obtient donc les coefficients de prédiction  $a_k$  ( $k = 1, 2, \dots, p$ ),

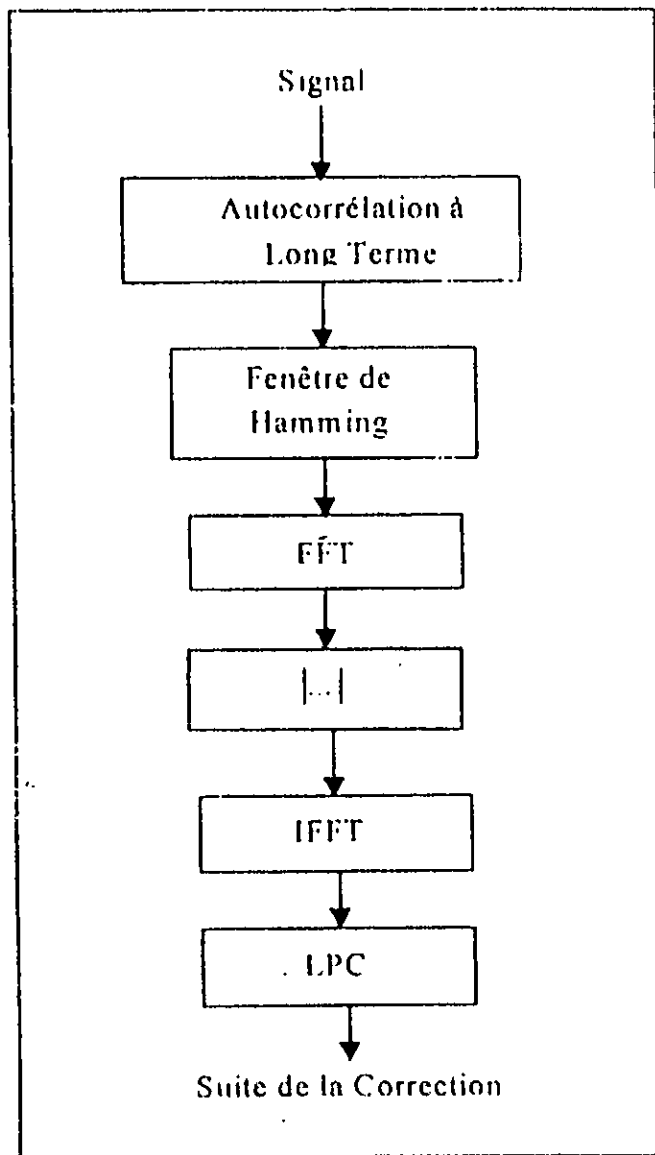
## Annexe B

### La Représentation SMC de la Parole Hyperbare Bruitée

La représentation cohérente modifiée de la parole hyperbare bruitée illustrée par l'organigramme qui suit, utilise la séquence d'autocorrélation pour estimer  $A^2(z)$ , ce qui induit une augmentation artificielle de l'ordre de  $P$  à  $2 \times P$ . Cela ne complique pas seulement la complexité de la modélisation, mais en pratique induit une interaction entre le pitch et les fréquences des formants.



Pour corriger cet effet, on applique la racine carrée dans le domaine spectral, ce qui permet de réduire la dynamique par deux et de rendre l'interaction entre le pitch et les fréquences des formants comme dans l'analyse LP standard. De plus, vu la dynamique élevée, une fenêtre avec des lobes secondaires très atténués doit être utilisée.



Organigramme de la Représentation SMC

## Références Bibliographiques

- [1]- J.L. Flanagan, *Speech Analysis, Synthesis, and Perception*, Springer-Verlag, New York, 1972.
- [2]- L.R. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals", Englewood Cliffs, New Jersey: Prentice-Hall, 1978.
- [3]- J. Makkoul "Linear Prediction: A Tutorial Review", *Proceeding of the IEEE*, VOL.3, NO 4, April 1975.
- [4]- M. Xie, D. Berkani, "Amélioration des Performances des Codeurs de Parole", *AJOT* 97.
- [5]- R. Steele, R.A. Salami, "Mobile Radio Communications : Chap 3 -Speech Coding", Pentech Press. 92
- [6]- D. Berkani, J.P. Adoul, H. Hassanein, "Intelligibility Enhancement For Diver's Speech", *Biennial Symp. On Communications Proceed. IEEE Kingst. SEC., Communication Canada*. Pp 115-120. Kingston. Ont. Canada. May 1994.
- [7]- M.A. Richards, "Helium speech enhancement using the short time fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, N° 6, pp. 841-853, Decembre 1982.
- [8]- E. O. Belcher, S. Hatestad, "Formant frequencies, bandwidths, and Qs in helium speech", *JASA* 74(2), August 1983.
- [9]- G. Fant and J. Lindquist, "Pressure and gas mixture effects on diver's speech, in *Speech Transmission Lab. Prog. and Status Report, STL-QPSR 1/1968*, Stockholm 1968
- [10]- Hisayoshi Suzuki, T. Nakai, K. Shinizu, "Measurement and analysis of speech sound radiated from vocal tract wall", *ICASSP 86*, pp. 1625-1628, Tokyo.
- [11]- J. Crestel, M. Guitton, V. Le Calvé, M. Corazza, "Sur la quasi-stationnarité du filtre vocal en conditions hyperbares". *Colloque GRETSI*, Juan-Les-Pins, Sept. 1991.

- [12]-R Tanaka, M. Nakatani, J. Susuki, "Formant frequency shifts under high ambient pressure", J. Radio Res. Lab. 21, 1974.
- [13]-T. Takasugi, M. Nakatsui, J. Susuki, "Long-term speech spectrum in a He-N<sub>2</sub>-O<sub>2</sub> mixture at depth of 30 m", J. Radio Res. Lab. 18, 1971.
- [14]-D.J. MacLean, "Analysis of speech in a helium-oxygen mixture under pressure," The Journal of Acoustical Society of America, vol. 40, NO 3, pp. 625-627, May 1966.
- [15]-E.O. Belcher, "A new model for unscrambling helium speech," in International Underwater Systems Design, N° 5 USD Pub., pp. 22-27, London, 1982.
- [16]-M.R. Portnoff, "Short-time fourier analysis of sampled speech," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-29, NO 3, pp. 364-373, June 1981.
- [17]- S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoust, Speech, Signal Processing, vol. ASSP-27, no 3, pp. 113-120. 1979.
- [18]- E.J. Cupples and J.L. Foelker, "Air Force speech enhancement program," Military speech tech, Rome Air Development Centre, Washington DC. 1987.
- [19]- M.Bérouti, R Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. IEEE ICASSP, pp. 208-211, Washington, DC, April. 1979.
- [20]-T.A. Giordano, H.B. Rothman, H. Hollien, "Helium speech unscramblers- A critical review of a state of the art," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-21, NO 5, pp.436-444, October 1973.
- [21] -D. Gravitz, "Hydrocom - A high quality solution for three diver communications", Hydro Products Inc./Honeywell 11777 Sorrento Valley Road, San Diego, CA 92121.
- [22]-T. Takasugi, J. Susuki, "Translation of helium speech by segmentation-expansion method", J. Radio Res. Lab. 19, 1972.
- [23]-R.J. Mac Aulay, T.F. Quatieri, "Speech analysis/synthesis based on sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 374-381, August 1986.

- [24]-A.H. Gray, Jr. and D.Y. Wrong, "The Burg algorithm for LPC speech analysis/synthesis," IEEE Trans. On ASSP, vol. 28, no , pp. 609-615, Dec 1980.
- [25]-G. Fant and B. Sonneson, "Speech at High Ambient Air Pressure," Speech Transmission Lab. Royal Inst. Of tech., Prog. Stockholm, Sweden, Quart. Prog. and Status Report, STL-QPSR 1/1964.
- [26]-R Tanaka, M. Nakatoni, J. Susuki, "Formant frequency shifts under high ambient pressure", J. Radio Res. Lab. 21. 1974.
- [27]-A. Marchal, C. Meunier and C. Cavé, "A tool for hyperbaric speech improvement: the PSH/DISPE CDROM". Proc. Speech Processing in Adverse conditions ". Cannes-Mandelieu, 1992.
- [28]-D.Mansour, B.Hwang Jang, "The Short-Time Modified Coherence Representation and Noisy Speech Recongnition,". IEEE Trans., Speech, Signal Processing, vol. 37, N<sup>o</sup>. 6, June 1989
- [29]-R. Boite, M. Kunt, "Traitement de la parole," presses polytechniques romandes. Lausanne. Suisse 1987.
- [30]-R.J.Mc Aulay and M.L.Malpass, "Speech Enhancement using a Soft Decision Noise Suppression Filre," IEEE Trans. Speech, Signal Processing, vol. ASSP-28, N<sup>o</sup>2, April 1980.
- [31]-F.K. Soong and M. M. Sondhi, "A Frequency-Weighted Ifakura Spectral Distortion. Measure and Its Application to Speech Recongnition in noise," IEEE Trans. Speech, Signal Processing; vol. 36, N<sup>o</sup> 1, Januaryl 1988.
- [32]-E.O. Belcher and K. Andersen, "Helium speech enhancement by frequency-domain processing," ICASSP '83, Boston, April 1983.
- [33]-C.Boubakir et D.Berkani, "Rehaussement de l'Intelligibilité de la Parole Hyperbare," COMAIEI'98, Bejaia, December 1998.

## ملخص

تناولنا في هذه الأطروحة دراسة بعض الخوارزميات التي تعمل على تحسين وتجويد الكلام الصادر في ظروف و شروط معينة (ضغط كبير و نسبة عالية من غاز الهليوم) و نجد هذه الظروف عند الغواصين على أعماق تتراوح بين 60 م إلى 300 م. هذه الدراسة مرتكزة على استعمال التنبؤ الخطي LP .  
أول هدف نظام التعليق هو زيادة وضوح الكلام , الهدف الثاني هو تحسين الكلام , هذا الأخير يتضمن أغراض و أهداف أخرى و هي تخفيض و تقليل ضجيج القعر و الحفاظ على طبيعة الكلام المركب.  
يجب على الخوارزميات المقترحة أن تستخلص خصائص الإشارة العالية الضغط و التحليل LPC و هذا لوضع خوارزميات المعالجة.

## Abstract

object of this thesis is study of some algorithm of amelioration of the quality of the word hyperbare, based on utilisation of the linear prediction. The first objective of the system of raising is increase the intelligibility of the word hyperbare. The second objective is ameliorate the quality of the latter, this target includes the others objective, namely, the minimization of the background noise and the maintenance of the naturalness of the synthesized word. The algorithms proposed must benefit from the characteristics of the signal hyperbare, the analysis LPC, in order to bring into play algorithms of treatment, with more or less complex alternatives, able to work in real times and the most adapted to a possible implementation on a chart DSP.

## résumé

L'objet de cette thèse est l'étude de quelques algorithmes d'amélioration de la qualité de la parole hyperbare, basés sur l'utilisation de la prédiction linéaire. Le premier but du système de rehaussement est d'augmenter l'intelligibilité de la parole hyperbare. Le deuxième but est d'améliorer la qualité de cette dernière, ce but inclut d'autres objectifs, à savoir, la minimisation du bruit de fond et le maintien du naturel de la parole synthétisée. Les algorithmes proposés doivent profiter des caractéristiques du signal hyperbare, de l'analyse LPC, afin de mettre en œuvre des algorithmes de traitement, avec des variantes plus ou moins complexes, capables de travailler en temps réel et les plus adaptés à une éventuelle implémentation sur une carte DSP.

## **Les mots clés**

Signal hyperbare ; signal vocal ; milieu hyperbare ; hélium ; intelligibilité de la parole ; la prédiction linéaire ; l'autocorrélation ; le spectre ; la fréquence fondamentale ; les formants ; les fenêtres de hamming ; compression temporelle ; correction spectral ; correction temporelle ; enveloppe spectrale ; réduction de bruit ; pondération du résidu ; soustraction spectrale