

M0032/984

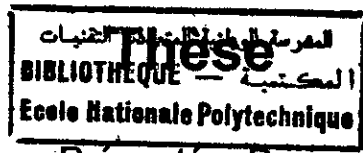
Ministère de l'enseignement Supérieur et de la Recherche Scientifique

Ecole Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات  
Ecole Nationale Polytechnique

DER- Génie Electrique & Informatique  
Département d'Electronique



Présentée Par

**M<sup>r</sup> Chaâbane BOUBAKIR**, Ingénieur d'Etat en Electronique  
En vue d'obtenir le Grade  
de **Magister** en Electronique

Option: Télécommunications

**THEME**

**REHAUSSEMENT  
DE LA  
PAROLE HYPERBARE**

*Soutenue le 25. 10. 1998... devant le Jury composé de :*

- Président: Mr A.FARAH .....Professeur à L'ENP
- Rapporteur: Mr D. BERKANI.....Maître de Conférences à L'ENP
- Examineurs: Mr R. AKSAS.....Maître de Conférences à L'ENP
- Mr A. BELOUHRANI .....DR, Enseignant à L'ENP
- Mme. L. HAMAMI .....Chargée de Cours à L'ENP

ملخص  
في هذا العمل قمنا بتصميم عدة خوارزميات لتحسين نوعية الإشارات الصوتية الخاضعة الناتجة تحت ضغوط عالية و في اعماق كبيرة .  
بالإضافة الى عدة تقنيات لتخفيض نسبة الاضطرابات. النظام المقترح يعتمد على خوارزمية التنبؤ الخطي لتحليل و تركيب الإشارات  
الصوتية. أما عملية التصويب فتتم في فضاءي الأزمنة و النواترات.  
كلمات جوهرية: الإشارات الصوتية في الهليوم او تحت ضغوط عالية. تحسين النوعية. الضوضاء. تخفيض الاضطرابات. الاتصالات تحت  
الماء.

## Abstract

In this work we studied several algorithms for helium speech enhancement. The system to be proposed is based on the use of a Linear prediction for speech analysis and synthesis, corrections are made directly in the time domain or in the frequency domain. Obtained results show that the designed algorithms restore the intelligibility and improve the speech diver's quality. Several noise reduction techniques are incorporated to improve the quality of noisy helium speech.

**Keys words: Helium Speech, Enhancement, Intelligibility, Noise Reduction, Diver's Communication**

## Résumé

Dans ce travail nous avons étudié plusieurs algorithmes de restauration de l'intelligibilité de la parole hyperbare produite à de grandes profondeurs sous haute pression. Le système proposé est basé sur la prédiction linéaire pour l'analyse et la synthèse de la parole hyperbare, les corrections ont été réalisées directement dans le domaine temporel ou dans le domaine fréquentiel. Les résultats obtenus montrent que les algorithmes proposés restaurent efficacement l'intelligibilité et améliorent la qualité de la parole en plongée sous-marine. Plusieurs techniques de réduction de bruit sont incorporées aux systèmes de correction afin d'améliorer la qualité du signal dans un milieu très hostile.

**Mots clés : Parole Hyperbare, Rehaussement, Intelligibilité, Réduction de bruit, Communication sous-marine**

# *Dédicace*

المدرسة الوطنية المتعددة التقنيات  
BIBLIOTHEQUE — المكتبة  
Ecole Nationale Polytechnique

*A ma mère, A ma mère, A ma mère.*

*A mon père.*

*A mes frères et sœurs.*

*A mes amis et collègues.*

*Chaâbane*

**L**E travail présenté dans ce mémoire a été effectué au "Signal & Communications Laboratory" du Département d'Electronique, DER-GEI de l'Ecole Nationale Polytechnique (ENP), sous la direction de Monsieur D.BERKANI, Maître de conférences à l'ENP, qu'il trouve ici l'expression de toute ma gratitude pour ses conseils avisés et ses critiques qui m'ont été d'une aide précieuse, ses encouragements et ses suggestions qui ont beaucoup contribué à susciter, à enrichir et à mener à bien ce travail.

Il m'est particulièrement agréable de remercier, tous les Enseignants qui m'ont aidé, conseillé et qui, par l'environnement crée, m'ont facilité grandement la tâche.

Je remercie Monsieur A.FARAH, Professeur à l'ENP, pour l'honneur qu'il m'a fait d'accepter la présidence de mon jury de thèse.

Je voudrais exprimer mes remerciements les plus sincères à Monsieur A. BELOUHRANI, Docteur, et Enseignant à l'ENP; et R.AKSAS, Maître de conférences à l'ENP, pour l'intérêt qu'ils ont manifesté pour notre travail, malgré ses nombreuses tâches, et pour avoir voulu participer au jury.

Je suis très heureux de pouvoir remercier ici, Madame L.HAMAMI, Chargée de cours à l'ENP; pour ses aides et encouragements continues, et de m'avoir fait le grand honneur d'assister au jury, malgré ses multiples occupations.

*Mes remerciements vont aussi :*

- ❖ *A tous les membres du laboratoire qui, en toute circonstance, nous ont donné l'exemple de l'esprit d'équipe, en particulier : B.Zekrini, K.Laldi, F.Terrant, F.Fliti, M.Damou, N.Bouali, S.Fiala,*
- ❖ *A tous mes amis et collègues qui ont montré un grand intérêt à mon travail et qui m'ont encouragé à toujours aller de l'avant, en particulier : M.Boufenar, N.Boussendel, M.Boutaghane, H.Bouteldj, S.Doulkbrachef, F.Sayoud, S.Labiod, M.Bourboune, O.Bouball, ....*
- ❖ *Aux personnels du centre de documentation de l'ENP, et l'ensemble du corps administratif de la D.E.R Génie Electrique et Informatique, pour l'aide, le soutien et l'encouragement qu'ils m'ont offert, en particuliers : Krimou, Saleh, Mahmoud, Foutal, Natma.Zabia....*

# SOMMAIRE



Dédicace	ii
Avant Propos	iii
Liste des Notations et Abréviations Utilisées	vi
Liste des Figures et Tableaux	viii

## **Chapitre I** **1**

---

### **Introduction**

I.1	Problématique	1
I.2	Contribution de cette Thèse	3
I.3	Organisation de la Thèse	5

## **Chapitre II** **5**

---

### **Le Signal Vocal**

II.1	Introduction	5
II.2	Production du Signal Vocal	6
II.3	La Prédiction Linéaire (LPC)	12
II.3.1	La Méthode d'Autocorrélation	14
II.3.2	Les Considérations de Choix des Conditions d'Analyse LPC	17
II.3.3	La Préaccentuation (Preamphasis)	18
II.4	Conclusion	19

## **Chapitre III** **20**

---

### **Le Signal Hyperbare**

III.1	Introduction	20
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	27
III.6	Les Caractéristiques du Signal Hyperbare	28
III.7	Conclusion	32

## Chapitre IV

33

### Les Systèmes de Traitement

IV.1	Introduction.....	33
IV.2	Synthèse des Systèmes Existants.....	34
IV.2.1	Les Systèmes à Approche Temporelle.....	34
IV.2.1.a	Enregistrement / Reproduction ou "Tape Recorder Playback".....	34
IV.2.1.b	Codage Numérique des Signaux "Digital Coding".....	34
IV.2.1.c	Traitement par signal analytique.....	34
V.2.1.d	Traitement par Convolution.....	35
V.2.2	Les Systèmes à Approche fréquentielle.....	35
IV.3	Considération de Conception du Système de Traitement.....	36
IV.4	Correction des Corrélations.....	37
IV.5	Correction Spectrale.....	40
IV.6	Résultats.....	43
IV.6.1	Description des Fichiers de Données & Conditions d'Expérimentation.....	43
IV.6.2	Commentaires et Interprétations des Résultats.....	46
IV.7	Conclusion.....	54

## Chapitre V

55

### Réduction de Bruit d'un Signal Hyperbare

V.1	Introduction.....	55
V.2	Caractéristiques du Bruit dans un Milieu Hyperbare.....	56
V.3	Réduction de Bruit dans le Domaine Temporel.....	57
V.3.1	Pondération du résidu.....	57
V.3.2	La Représentation SMC de la Parole Hyperbare Bruitée.....	61
V.4	Réduction de Bruit dans le Domaine Spectral.....	63
V.4.1	La Soustraction Spectrale.....	63
V.4.2	Estimation du Spectre du Bruit.....	66
V.4.3	Algorithme d'Amélioration Basé sur la Soustraction Spectrale.....	66
V.5	Conclusion.....	69

## Conclusion Générale

70

## Annexes

72

## Références Bibliographiques

75

WLD	Wiener- Levinson – Durbin
L	Longueur du Conduit Vocal
$\mu$	Facteur de Préaccentuation
$H_e - O_2$	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
** <sub>a</sub> , ** <sub>h</sub>	Valeur de " ** " " dans l'air ou de " ** " " dans l'Hélium
C <sub>a</sub> , C <sub>h</sub>	Célérité du Son
CP <sub>a</sub> , CP <sub>h</sub>	Chaleur Spécifique à Pression Constante
CV <sub>a</sub> , CV <sub>h</sub>	Chaleur Spécifique à Volume Constant
$\mu_a$ , $\mu_h$	Coefficient de Viscosité
$\lambda_a$ , $\lambda_h$	Coefficient de Conductibilité de la Chaleur
$\gamma_a$ , $\gamma_h$	Constante Adiabatique
$\rho_a$ , $\rho_h$	Densité du Gaz
Spec-a(w)	Enveloppe Spectrale du Conduit Vocal
CSpec-a(w)	Enveloppe Corrigée
SMC	Short Time Modified Coherence Representation
VAD	Détecteur d'Activité Vocal (Voice Activity Detection)
RSB (SNR)	Rapport Signal sur Bruit

## Liste des Figures et Tableaux

Figure		Page
Fig. 2.1 :	L'appareil phonatoire et la formation du signal vocal .....	6
Fig. 2.2 :	Analogie électrique d'un tube uniforme sans perte et approximation du conduit vocal par un tube acoustique .....	7
Fig. 2.3 :	la forme d'onde et le spectre d'un son voisé .....	9
Fig. 2.4 :	la forme d'onde et le spectre d'un son non voisé .....	10
Fig. 2.5 :	Modèle de production de la parole dans le domaine spectrale .....	11
Fig. 2.6 :	Modèle de production de la parole dans le domaine temporel .....	12
Fig. 2.7 :	Organigramme de la prédiction linéaire .....	18
Fig. 3.1 :	La proportion d'O <sub>2</sub> en fonction de la profondeur .....	24
Fig. 3.2 :	Les rapports hélium – air .....	24
Fig. 3.3 :	Translation des fréquences des formants pour 90%He-10%O <sub>2</sub> à 10 ATA .....	27
Fig. 3.4 :	Forme d'onde et Spectrogramme du mot "The Rainbow" .....	30
Fig. 3.5 :	Forme d'onde et Spectrogramme du mot "loi" .....	31
Fig. 4.1 :	Approche temporelle de traitement de la parole hyperbare .....	35
Fig. 4.2 :	Approche fréquentielle de traitement de la parole hyperbare .....	36
Fig. 4.3 :	Principe du système de correction .....	37
Fig. 4.4 :	L'étalement temporel des autocorrélations .....	39
Fig. 4.5 :	Algorithme de correction des autocorrélations .....	41
Fig. 4.6 :	Algorithme de correction spectrale .....	41
Fig. 4.7 :	Correction linéaire du spectre .....	43
Fig. 4.8 :	Les valeurs de K pour une compression non linéaire .....	43
Fig. 4.9.a :	Exemples d'une tranche du signal original et son spectre .....	48
Fig. 4.9.b :	Exemples des résultats obtenus avec correction des autocorrélations .....	49
Fig. 4.9.c :	Exemples des résultats obtenus avec correction spectrale .....	50
Fig. 4.10 :	Spectrogramme du (a) signal hyperbare à 200m et (b)-(c) la parole synthétisée (correction des autocorrélations et spectrale) .....	51
Fig. 4.11 :	Spectrogramme du (a) signal hyperbare à 300m et (b)-(c)-(d) La parole synthétisée (correction des autocorrélations et Correction linéaire et non linéaire du spectre) .....	52
Fig. 5.1.a :	Spectrogramme de SOF0.IRS .....	56
Fig. 5.1.b :	Spectrogramme de SNF300P.IRS .....	57
Fig. 5.2 :	Spectre de F(z)- ( $\gamma = 0.5, 0.8, 1$ ) .....	58



Fig. 5.3 :	La correction des corrélations avec filtre de pondération.....	
Fig. 5.4 :	La valeur optimale de $\gamma$ en fonction du RSB.....	59
Fig. 5.5 a :	Spectrogramme de SYN300P (correction des corrélations).....	60
Fig. 5.5.b :	Spectrogramme de SYN300P (pondération du résidu $\gamma = 0.8$ ).....	60
Fig. 5.6 :	Spectrogramme de SYN300P (SMC).....	62
Fig. 5.7 :	Le principe de base de la soustraction spectrale.....	64
Fig. 5.8 :	Correction spectrale avec réduction de bruit.....	66
Fig. 5.9 :	Spectrogramme de SYN300P (Soustraction spectrale).....	68

## Tableau

## Page

Tab. 3.1 :	Rapports des propriétés dans l'He-O2 et l'air.....	23
Tab. 4.1 :	Les Fichiers de test utilisés.....	44
Tab. 4.2 :	Conditions d'expérimentation.....	45

# CHAPITRE I

## Introduction

### I.1 Problématique

**L**ES recherches et les travaux effectués sous les océans à des profondeurs de plus en plus grandes, se heurtent à de nombreuses difficultés, notamment en ce qui concerne les difficultés rencontrées par les plongeurs à communiquer avec les équipes de surface et entre eux, avec un pourcentage de mots reconnus correctement élevé.

Ainsi, les plongeurs professionnels qui réalisent des plongées à des profondeurs très supérieures à 60m, sont amenés à vivre dans une ambiance de haute pression, et on est obligé de remplacer l'azote diluant l'oxygène 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,.... Cela permet de telle plongée dite de saturation et évite les effets très graves de l'azote en particulier la narcose, qui entraîne une sorte de dégradation mentale.

Cependant, les propriétés acoustiques de ces mélanges sous pression diffèrent radicalement de celles dans l'air. Par conséquent, La parole produite par un plongeur qui respire de tel mélange dans de tel environnement, dite la parole sous haute pression ou la parole hyperbare ou encore la parole dans l'hélium est inintelligible et nasillarde, surtout quand le pourcentage d'hélium est important. C'est l'effet «Donald Duck» (Walt Disney faisait parler la doublure du célèbre canard dans une atmosphère d'hélium).

Ce problème d'intelligibilité de la parole en plongée sous-marine, est dû principalement aux effets combinés de deux phénomènes physiques incontournables :

- ❖ La pression, tout d'abord : à 100 mètres de profondeur, elle est 11 fois plus importante qu'à la surface, (Dix mètres d'eau exerce une pression d'une atmosphère). Elle s'exerce sur le corps et dans

le gaz respiratoire des plongeurs, dont la densité augmente en proportion, et on a l'impression de parler dans un tuyau dont le diamètre se rétrécit. Il en résulte une modification du conduit vocal (de la glotte aux lèvres). Les fréquences des formants de la parole augmentent, et la voix devient plus aiguë. Les trois premiers formants situés dans les basses fréquences (BF) sont les plus affectés, et les consonnes perdent plus d'énergie par rapport aux voyelles, ce qui dégrade l'intelligibilité. Par exemple, à 60m de profondeur, "pile" ressemble beaucoup à "bile".

❖ La seconde déformation est due au mélange gazeux à haute pression. La présence d'hélium élève les fréquences des résonances (formants) de façon pratiquement linéaire. Cette variation est due principalement à un changement de la vitesse  $C$  du son. A 100m de profondeur le son dans le mélange se propage trois fois plus vite que dans l'air.

Les effets combinés des deux phénomènes, résultent en un décalage non linéaire des fréquences des formants.

Au début, plusieurs travaux de recherches ont été consacrés 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 des 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 à cet inconvénient.

Le problème de rehaussement de la parole hyperbare intéresse plusieurs organismes, en particulier : les centres qui exploitent les océans; la marine, pour l'amélioration des communications entre les plongeurs et avec les équipes de surface, les études et les recherches sous-marines spécifiques aux militaires; les compagnies d'exploitation pétrolières;...

## I.2 Objet & Contribution de Cette Thèse

L'objet de cette thèse est l'étude et l'évaluation de plusieurs 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 traitements, avec des variantes plus ou moins complexes, capables de travailler en temps réel et les plus adaptés à une éventuelle implémentation sur un DSP. Le travail présenté dans cette thèse est le premier dans notre laboratoire, qui soit consacré à la parole hyperbare et qui nous a permis de mettre en œuvre un environnement riche et varié des signaux de la parole, dans différentes conditions normales et hyperbares avec les outils de traitement et de correction appropriés. De plus, la réalisation de cette thèse a nécessité l'élaboration d'autres travaux, qui peuvent être considérés comme des contributions supplémentaires de cette thèse.

## 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.

Des notions générales sur la production du signal vocal, ses caractéristiques temporelles et spectrales sont présentées au chapitre deux. Nous présenterons également une étude détaillée sur la prédiction linéaire (LPC), qui sera utilisée dans tous les algorithmes de rehaussement que nous développerons par la suite.

Le troisième chapitre est consacré au signal hyperbare. Une étude comparative entre les propriétés de la parole dans les conditions normales et hyperbares a été présentée. De nombreuses simulations numériques nous ont permis de confirmer les études théoriques développées dans une large bibliographie. 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 quatrième chapitre, ensuite une description détaillée des algorithmes de traitement que nous avons proposé. Enfin, des résultats de simulation de chaque variante sont présentés.

Les systèmes de réduction de bruit sont exposés en détails dans le chapitre cinq. On y étudie la réduction de bruit dans le domaine temporel par la pondération du résidu et par une nouvelle représentation de la fonction d'autocorrélation. On y expose notamment la soustraction spectrale et le détecteur d'activité vocal (VAD); tâche à la-quelle plusieurs travaux de recherches sont effectués dans le monde entier, cette dernière section de ce chapitre n'est qu'une introduction à la soustraction spectrale qui sera incorporée au système de réduction de 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. Toutes les simulations et les tests effectués dans le cadre de cette thèse ont été réalisés sur un micro-ordinateur Pentium 166 Mhz, 16 M de RAM; en utilisant le Langage C et les logiciels "VISU" (Laboratoire); "MATLAB"; "GOLDWAVE".

# CHAPITRE II

## Le Signal Vocal

### II.1 Introduction

D'une manière générale, les méthodes d'analyse et de rehaussement de la parole hyperbare, qui seront détaillées par la suite, dans cette thèse fondent leur justification sur une simple extension des propriétés du signal vocal ou "AIR".

Ainsi, dans la première partie de ce chapitre, une étude détaillée de la modélisation du signal vocal, ses caractéristiques temporelles et spectrales basées sur les hypothèses de la quasi-stationnarité du filtre vocal et la quasi-périodicité de l'excitation glottique s'avère nécessaire.

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.

Enfin, il convient de préciser que la prédiction linéaire est utilisée pour l'estimation des paramètres essentiels du signal vocal, une discussion sur ces paramètres et conditions d'analyse fera l'objet de la dernière partie de ce chapitre ; à savoir la méthode d'analyse, l'ordre de prédiction, la longueur des trames et les fenêtres utilisées.

De nombreuses figures et simulations numériques, ont permis de mieux présenter les aspects théoriques développés dans ce chapitre.

## II.2 Production du Signal Vocal

Le signal vocal est produit par l'appareil phonatoire (fig. 2.1), qui se compose essentiellement de trois étages [1,2] :

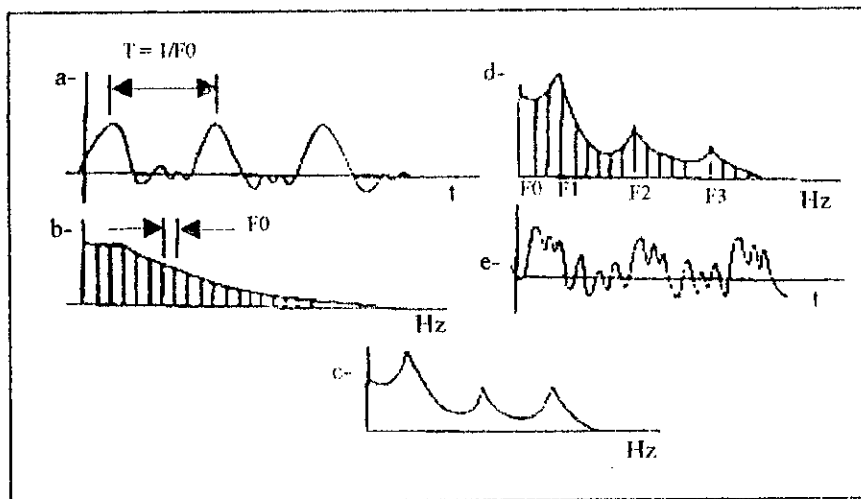
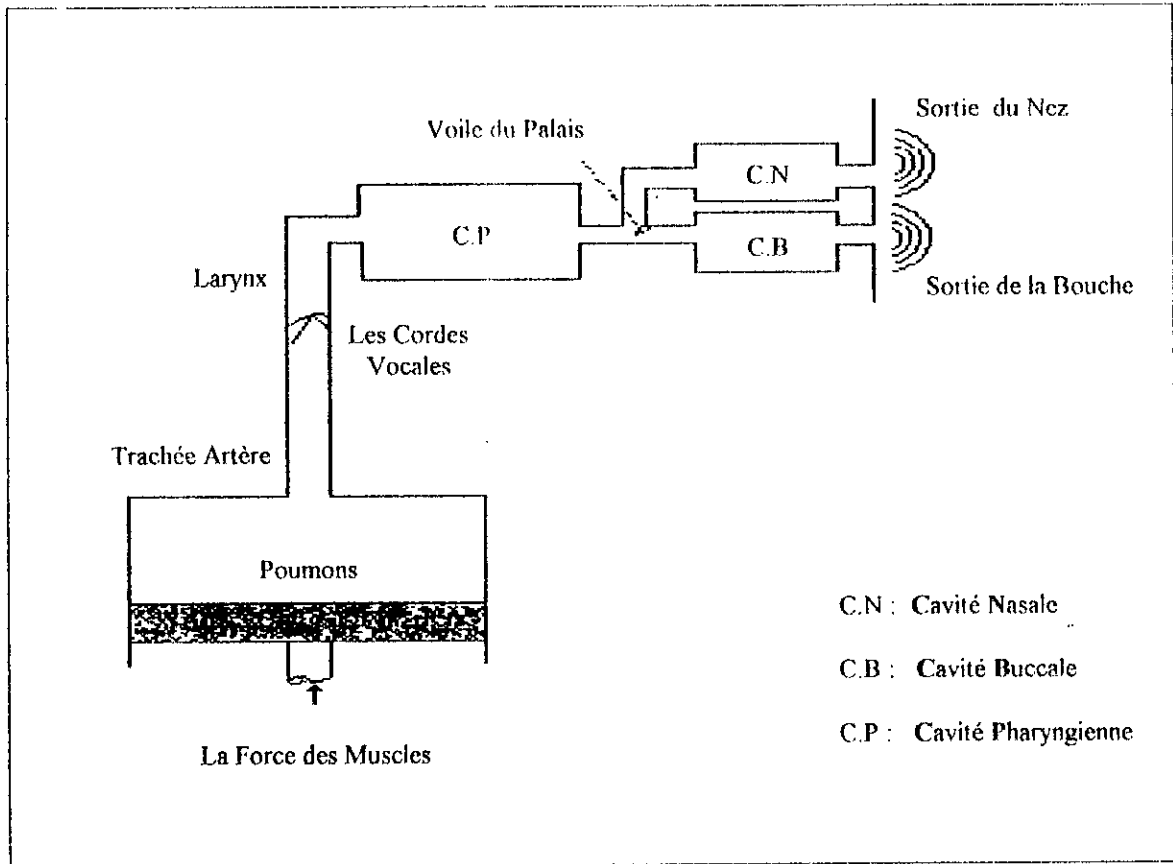
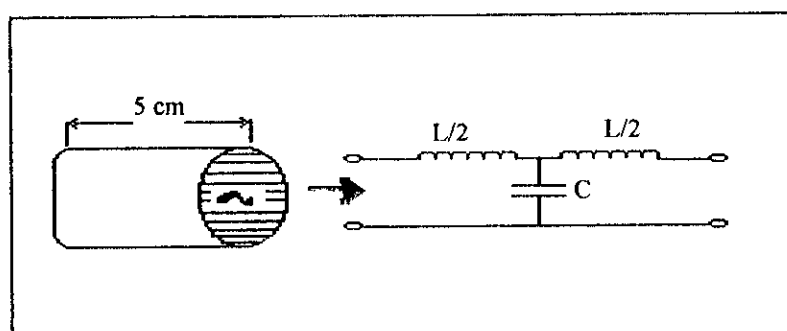
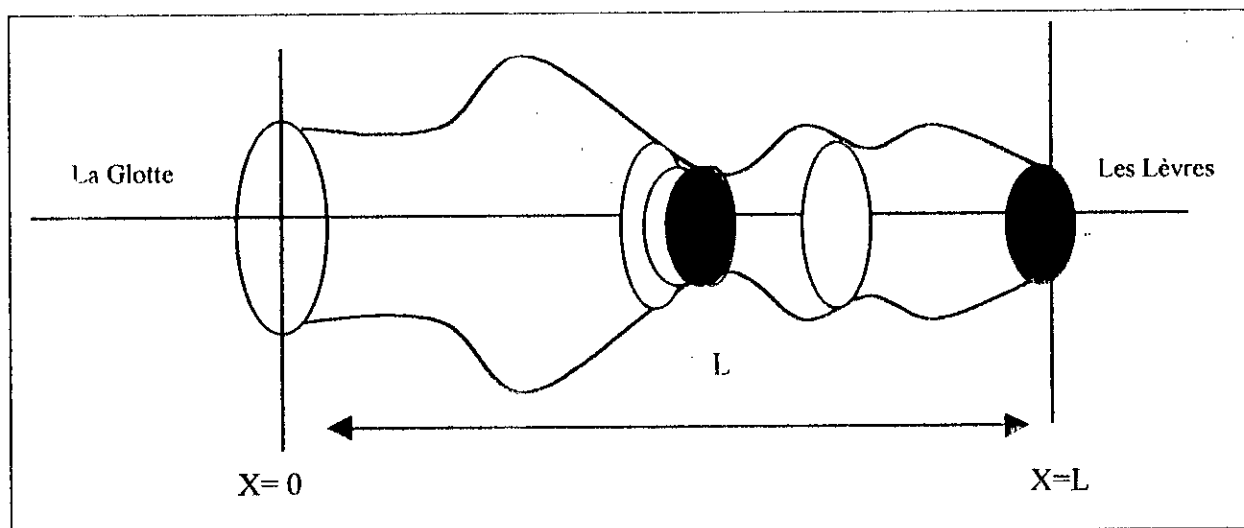


Fig. 2.1 L'appareil phonatoire et la formation du signal vocal: a- le signal émis par les cordes Vocales; b- son spectre; c- la fonction de transfert du conduit vocal; d- spectre de la parole au niveau des lèvres; e- sa représentation temporelle.

- L'étage de respiration : Les poumons, qui insufflent de l'air. L'ensemble poumons trachée artère se comporte comme un générateur d'air qui alimente le larynx.
- Le larynx et les cordes vocales, qui servent de vibreurs. Ainsi le débit d'air fourni par les poumons est modulé par les cordes vocales ; celles-ci vibrent à des fréquences données appelées PITCH. Elles fournissent ainsi le signal glottal.
- Le conduit vocal de longueur 17 Cm environ est composé d'un ensemble de cavités et éléments à géométrie variable. Dans le cas statique, le conduit pourrait être modélisé par un tube à section variable (fig. 2.2). L'analogie électrique serait un filtre à plusieurs résonances et antirésonances.



**Fig. 2.2** Analogie électrique d'un tube uniforme sans perte et approximation du conduit vocal par un tube acoustique.



En ce qui concerne le modèle de simulation, 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.[29]

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 'f' 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 [a] dans le mots "save") et celui du signal non voisé ( la consonne [s] dans le mots "save" ) présentés dans les figures 2.3 et 2.4 ont été limités à 10 KHz avant un échantillonnage à 20 KHz.

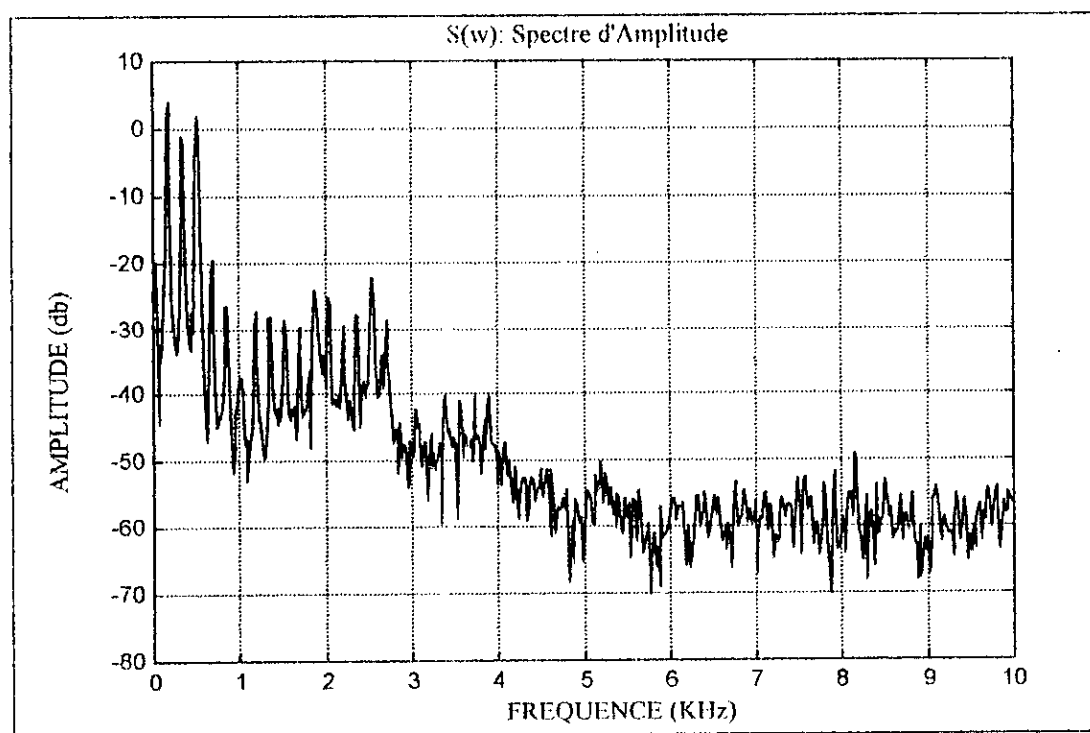
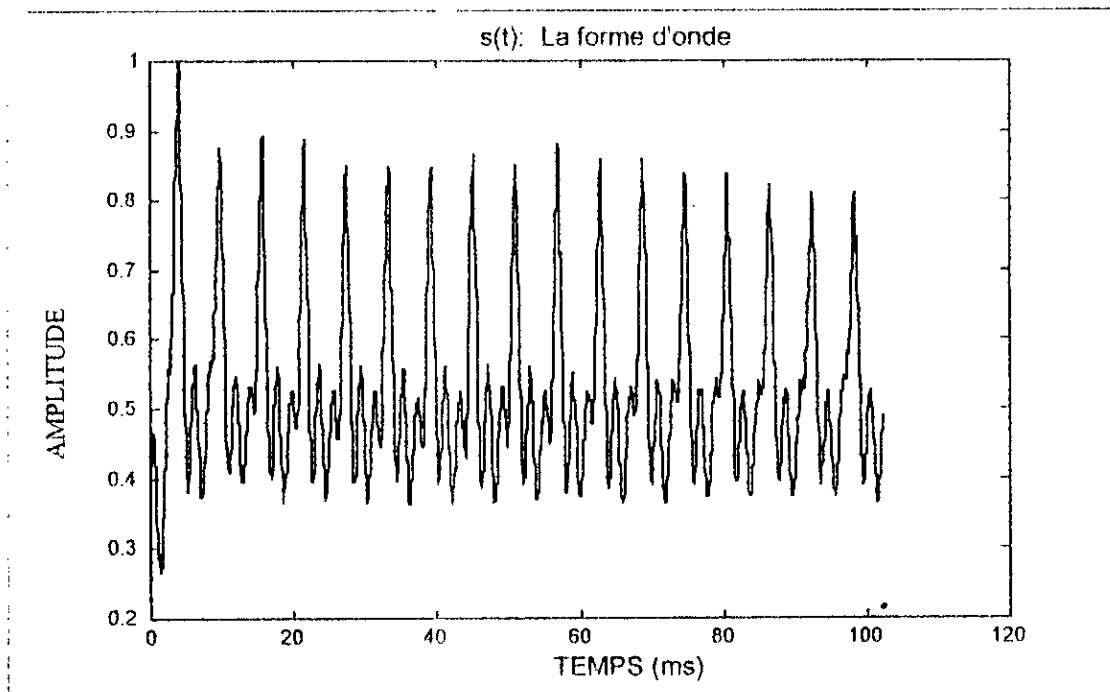


Fig. 2.3 la forme d'onde et le spectre d'un son voisé

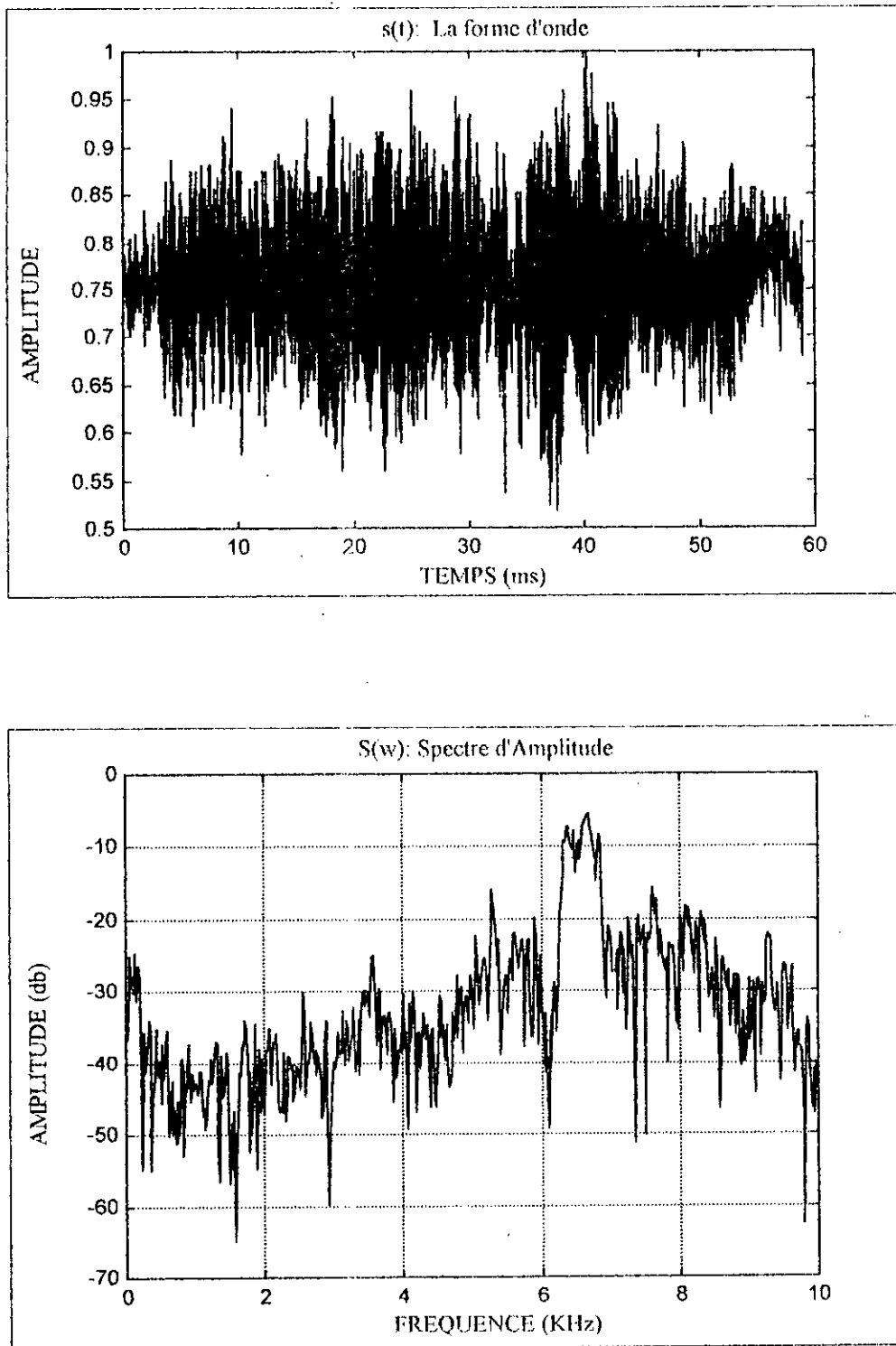


Fig. 2.4 la forme d'onde et le spectre d'un son non voisé

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

$$H(z) = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^p a_k \cdot z^{-k}} \quad (2.1)$$

Ce modèle de production du signal vocal est appelé autoregressif (AR) ; Il est clair que le modèle AR est une approximation plus ou 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 :

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

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.

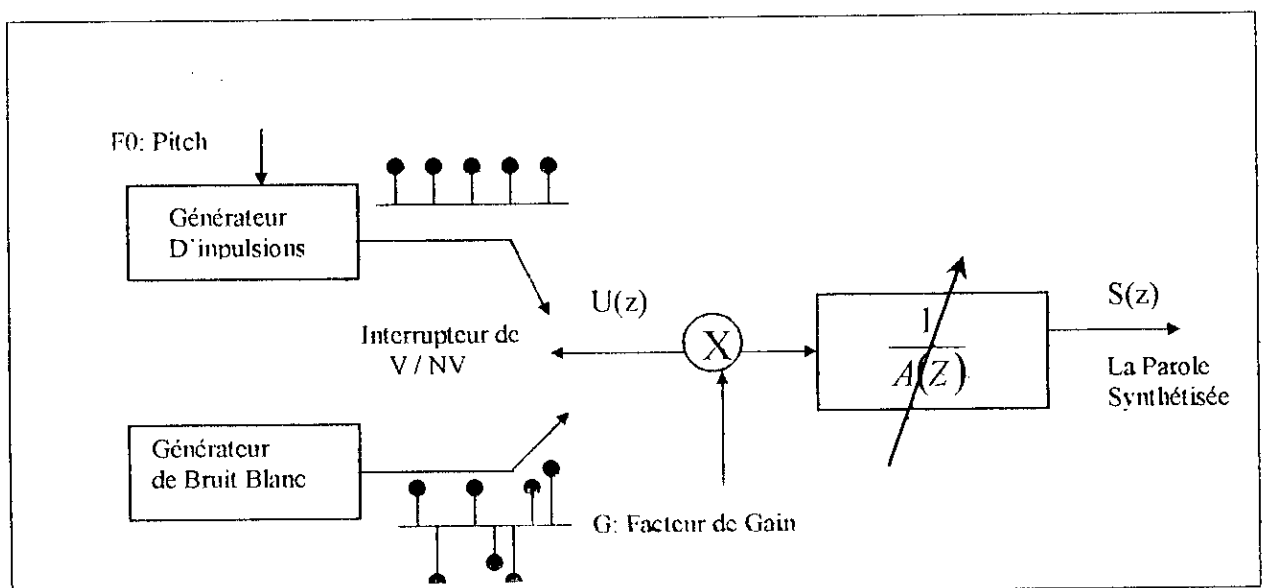


Fig. 2.5 Modèle de production de la parole dans le domaine spectrale

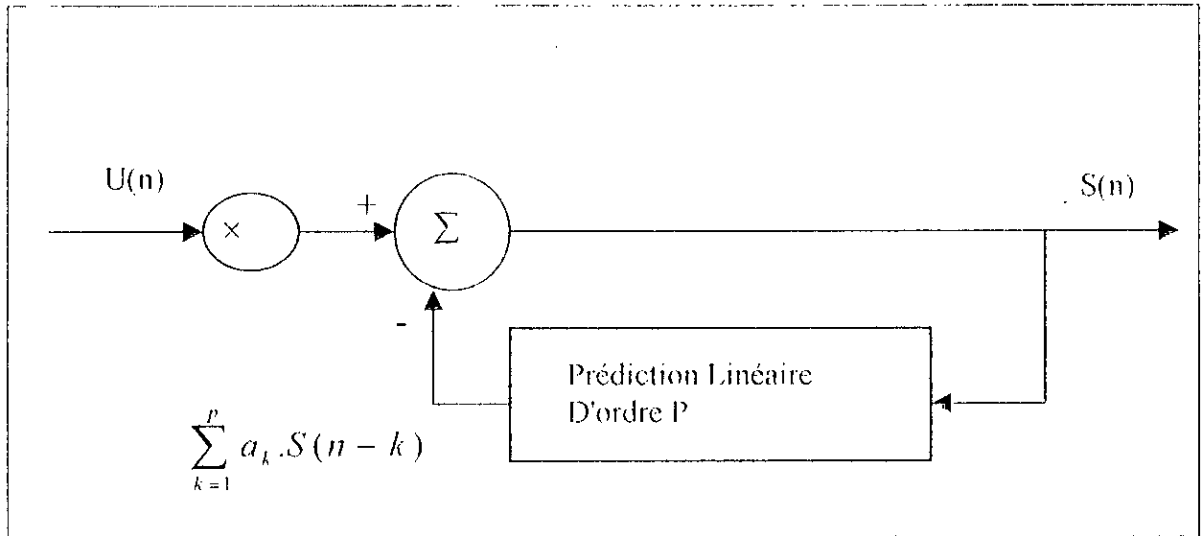


Fig. 2.6 Modèle de production de la parole dans le domaine temporel

### 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 production 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 parole peut être prédit par une combinaison linéaire d'un nombre fini des échantillons qui le précède. Ainsi, un échantillon  $s(n)$  peut être approximé par une combinaison linéaire de  $P$  échantillons précédents ( $p=8, \dots, 16$  échantillons) :

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

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 la sortie du prédicteur linéaire à l'instant d'échantillonnage  $n$ , qui a la fonction de transfert suivante :

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

Cette prédiction induit une erreur de prédiction  $e(n)$  entre l'échantillon original  $s(n)$  et l'échantillon prédit  $\tilde{s}(n)$  et elle est définie par :

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

Si on compare (2.2) et (2.5), on constate que lorsqu'on trouve un ensemble de coefficients  $\tilde{a}_k$  plus proche de l'ensemble  $a_k$  en minimisant l'erreur de prédiction ( $\tilde{a}_k = a_k$ ,  $(k = 1, 2, \dots, p)$ ). On obtient alors l'égalité  $e(n) = G.U(n)$ . Ceci montre que l'erreur de prédiction  $e(n)$  contient beaucoup d'informations sur l'excitation du signal de parole. La transformée en Z de l'équation (2.5) donne :

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

$A(z)$  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 un intervalle donné. 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 :

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

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

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

Ce qui conduit au système :

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

Où :

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

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

Par conséquent l'équation (2.10) devient :

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

Le problème de la technique LP se réduit donc à la résolution d'un ensemble de  $P$  équations avec  $P$  inconnus, afin d'obtenir les coefficients  $a_k$  :  $k = 1, 2, \dots, p$ . 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 LP, la méthode d'autocorrélation et la méthode de covariance. En plus, d'une troisième méthode qui est basée sur le calcul des coefficients de réflexion, appelée aussi la méthode de covariance en treillis ou l'algorithme de Burg. Bien qu'elle ait plusieurs applications, comme la détection du signal sinusoïdal dans un bruit additif, Gray et al [24] ont montré que l'algorithme de Burg ne présente aucun avantage par rapport aux autres techniques dans les applications d'analyse de la parole. 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 intervalle infini, cependant pour respecter les considérations pratiques on suppose que  $s(n) = 0$  en dehors de l'intervalle:  $0 \leq n \leq N-1$ , où  $N$  est la durée de la fenêtre d'analyse LP. Cette hypothèse sur la nature de  $s(n)$  simplifiée le calcul du signal  $\tilde{s}(n)$ ; donc :

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

On voit 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 :

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

Donc.  $\phi(i, k)$  est l'autocorrelations à court terme de  $s(n)$  évaluée à  $(i - k)$ . Par conséquent :

$$\phi(i, k) = R(i - k) \tag{2.15}$$

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

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

$$\sum_{k=1}^p a_k \cdot R(|i - k|) = -R(i), \quad i = 1, 2, \dots, p \tag{2.17}$$

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

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

On constate que la matrice des valeurs de l'autocorrélation 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) développé dans [5] qui suit représente la solution la plus efficace.



**La procédure WLD**

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

for  $i=1$  to  $p$  do

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

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

for  $j=1$  to  $i-1$  do

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

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

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

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 : 
$$-1 \leq K \leq 1 \quad (2.23)$$

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

$$w(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi \cdot n}{(N-1)}\right) \quad n = 0, 1, \dots, N-1 \quad (2.24)$$

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 du 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 Atal [5].

- Le nombre  $P$  des coefficients de prédiction :

Comme dans le cas du codage des paramètres LP, on doit utiliser un nombre minimal de paramètres nécessaires à une modélisation exacte de l'enveloppe spectrale court - terme de la parole. Rabiner et Shafer [2] ont démontré que pour une représentation adéquate du conduit vocal, la durée de mémorisation  $p.T$  nécessaire au calcul de  $\tilde{s}(n)$  [ 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  $(2.L) / C$ , où  $L$  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  $N_a$  de pôles (4 à 5) qui sont nécessaires pour représenter l'influence de la source et du rayonnement, donc :

$$p = F_s (\text{Khz}) + N_a \quad (2.25)$$

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, cet intervalle 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 cet intervalle 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.

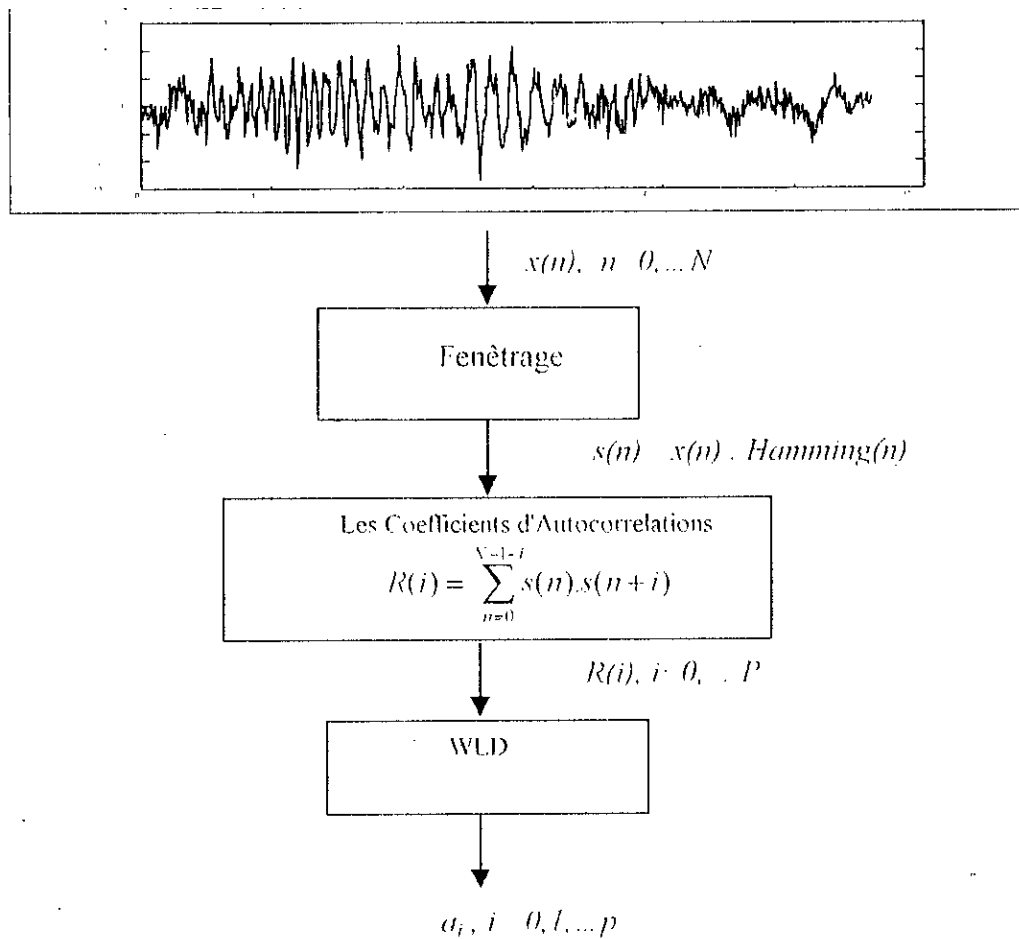


Fig. 2.7 Organigramme de la prédiction linéaire (LP)

### II.3.3 La Préaccentuation (Premphasis)

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 préaccentuation consiste à faire passer, avant l'analyse spectrale par prédiction linéaire, le signal de parole dans un filtre de transmittance :

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

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

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

Ce filtre a pour effet d'accentuer la partie haute fréquence du spectre. Par conséquent, la préaccentuation contribue à compenser 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, développées dans une large bibliographie. Néanmoins, nous avons mis en évidence les principaux 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, dans les différentes conditions hyperbares avec une adaptation appropriée. Notons enfin, qu'on peut donner les conclusions suivantes :

- Que le modèle de simulation considéré par la suite et qui est largement connu, modélise la source glottale par un signal aléatoire pour les sons non voisés et les sons voisés par un train d'impulsions.
- Que le conduit vocal est modélisé par un filtre dont les paramètres varient dans le temps (modèle AR), 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 préaccentuation éventuelle dans le cas d'une analyse spectrale court-terme, qui est basée sur l'analyse LP.

# 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, comme ils l'avaient déjà fait pour le signal vocal dans les conditions normales. Par ailleurs, les travaux présentés par M.R.RICHARDS[7] semblent les plus intéressants. Ainsi, une grande partie de ces travaux 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 mathématiques, afin de mettre en évidence, l'effet de ces variations sur les caractéristiques principales du signal vocal, par rapport à celles du signal hyperbare. Cette étude nous permet de mieux définir les exigences des systèmes de rehaussement de la parole hyperbare. Les résultats théoriques et empiriques sont utilisés ~~vo~~ à cette fin.

Cette analyse soulève de nouvelles questions à propos du 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, d'où les caractéristiques spectrale et temporelle du signal hyperbare.

La deuxième partie décrit sommairement ces notions, 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 à 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 :

$$\mu = \mu_0 \approx 1.87 \times 10^{-4} \text{ dyne} - \text{sec/cm}^2 \quad (3.1)$$

est le même dans l'air et dans tous les mélanges  $\text{N}_2\text{-O}_2$ . 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 :

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

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 toutes les gaz qui existent dans le mélange. De plus,

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

avec  $R$  la constante universelle du gaz :

$$R = 1.987 \text{ cal.mole}^{-1}.\text{deg}^{-1} = 0.082 \text{ Latm/mole.k} \quad (3.4)$$

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 :

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

$$\text{Avec } \lambda_{0\text{air}} = 0.12 \text{ ; et } \lambda_{0\text{H}} = 1.17, \quad (3.7)$$

La densité d'un gaz parfait est donnée par l'équation :  $\rho = P \cdot \rho_0$ ,

$$(3.8)$$

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

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

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

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

$$c = \sqrt{\frac{\gamma \cdot P}{\rho}} = \sqrt{\frac{\gamma}{\rho_0}}, \quad (3.10)$$

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 représentant 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 ( $N_2$ )).

%H <sub>e</sub>	%O <sub>2</sub>	$\frac{C_h}{C_a}$	$\frac{\gamma_h}{\gamma_a}$	$\frac{\lambda_h}{\lambda_a}$	$\frac{\rho_{oh}}{\rho_{oa}}$	$\frac{CP_h}{CP_a}$
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	8.97	0.92	0.94
30	70	1.13	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.

On remarque 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 1.7 ATA (pour des raisons physiologiques) est donnée par :

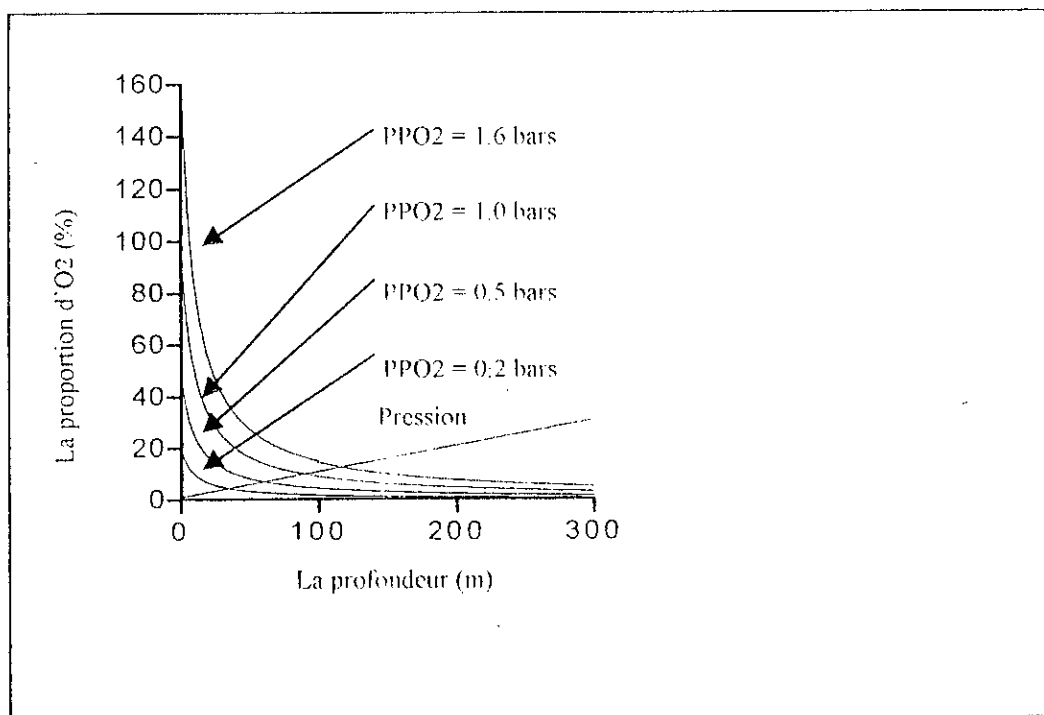
$$PPO_2 = PO_2 \times P \quad (\text{ATA}) \quad (3.11)$$



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 :

$$P = 1 + \frac{D}{10}, \quad (3.12)$$

La figure (3.1) illustre la proportion d'O<sub>2</sub> 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<sub>2</sub> en fonction de la profondeur

La figure 3.2 montre les rapports  $K = C_h/C_a$  et  $\rho_h/\rho_a$  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.

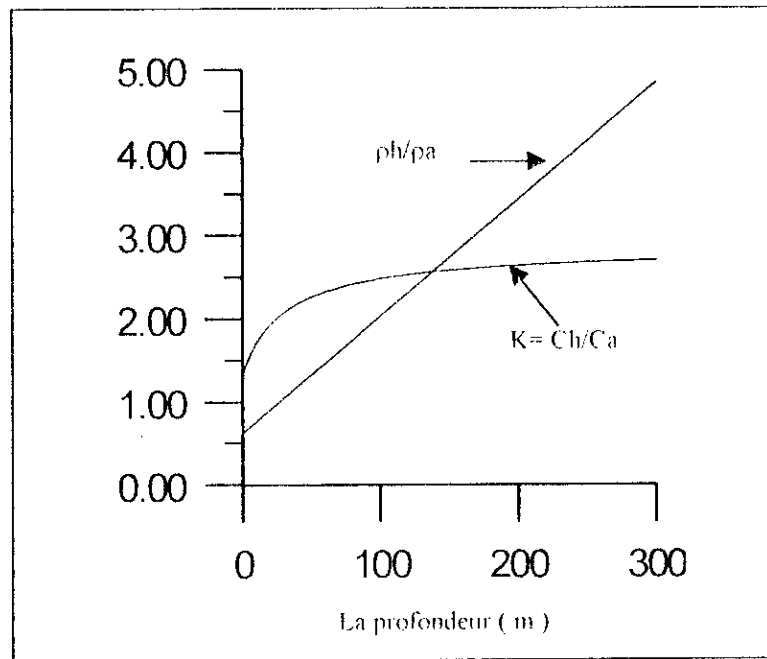


Fig. 3.2 Les rapports hélium – air

### III.3 Variation des Fréquences des Formants

Fant et Sannesson [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

$$F_c^2 = \frac{\gamma \cdot P}{4 \pi^2 L_a}, \tag{3.15}$$

et  $L_a$  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 :

$$F_h^2 = F_{hw}^2 + F_{ch}^2, \tag{3.16}$$

$$\text{et} \quad F_a^2 = F_{mh}^2 + F_{ca}^2, \quad (3.17)$$

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

$$F_{mh}^2 = \left(\frac{c_h}{c_a}\right)^2 F_{mh}^2 = \left(\frac{c_h}{c_a}\right)^2 (F_a^2 - F_{ca}^2), \quad (3.18)$$

à partir de l'équation (3.15),

$$F_{ch}^2 = \frac{\gamma_h \cdot P_h}{\gamma_a \cdot P_a} F_{ca}^2, \quad (3.19)$$

On remplace les valeurs de  $F_{mh}^2$  et  $F_{ch}^2$  de (3.18) et (3.19) dans (3.16) avec  $P_a = 1$  ATA ce qui donne :

$$F_h^2 = \left(\frac{c_h}{c_a}\right)^2 F_a^2 + \left(\frac{\gamma_h \cdot P_h}{\gamma_a} - \left(\frac{c_h}{c_a}\right)^2\right) F_{ca}^2, \quad (3.20)$$

Finalement, vu que

$$C^2 = \frac{\gamma \cdot P}{\rho} = \frac{\gamma}{\rho_0},$$

$$F_h^2 = \left(\frac{c_h}{c_a}\right)^2 F_a^2 + \left(\frac{\gamma_h}{\gamma_a}\right) \left[ \rho_h - \frac{\rho_{0a}}{\rho_{0h}} \right] F_{ca}^2, \quad (3.21)$$

$$\text{Cette equation est sous la forme :} \quad F_h^2 = K^2 F_a^2 + F_0^2, \quad (3.22)$$

$$\text{Où : } K = \frac{c_h}{c_a} \text{ et } F_0^2 = \left(\frac{\gamma_h}{\gamma_a}\right) \left[ \rho_h - \frac{\rho_{0a}}{\rho_{0h}} \right] F_{ca}^2 = K^2 \left( \frac{\rho_h}{\rho_a} - 1 \right) F_{ca}^2, \quad (3.23)$$

La constante  $F_{ca}$  est déterminée à partir de la longueur du conduit vocal, elle est égale à 150Hz 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.

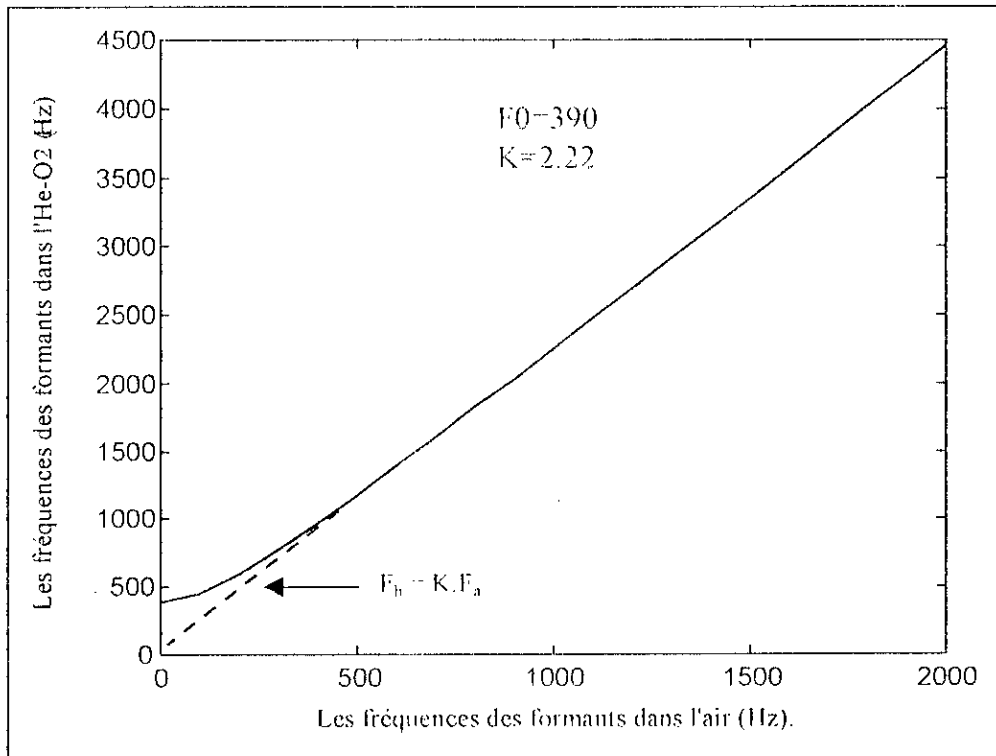


Fig. 3.3 Translation des fréquences des formants pour 90%He-10%O2 à 10-ATA

### 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 dans le modèle avec pertes du conduit vocal, conduisent à des bandes de fréquences limitées des formants et à un décalage des fréquences de ces derniers. De plus, les largeurs des bandes des formants inférieurs sont dominées par la contribution des pertes par friction et par vibration des parois, et celles des formants supérieurs sont dominées 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  $M(f)L(f)$  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 de l'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 3.8 db, des amplitudes des sons non voisés par rapport à celles 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-marines (sections précédentes).

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 :

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

- $F_h$  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 un facteur de propagation, égal au rapport des célérités du son dans l'hélium et l'air.

$$K = \alpha = C_h/C_a \quad (3.25)$$

- $F_{ca}$  est une constante caractérisant la résonance du canal fermé à une pression d'une atmosphère dans l'air, elle varie de 0 à 180hz 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 sans 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 10hz, d'autres donnent une augmentation de 10 jusqu'à 50 %. La variation du pitch ne doit pas 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 et 3.5 illustrent la forme d'onde et le spectrogramme du mot Anglais "The Rainbow" et le mot français "Loi" prononcés par deux plongeurs à 60m, 100m, 200m, 300m 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 conditions d'immersion.

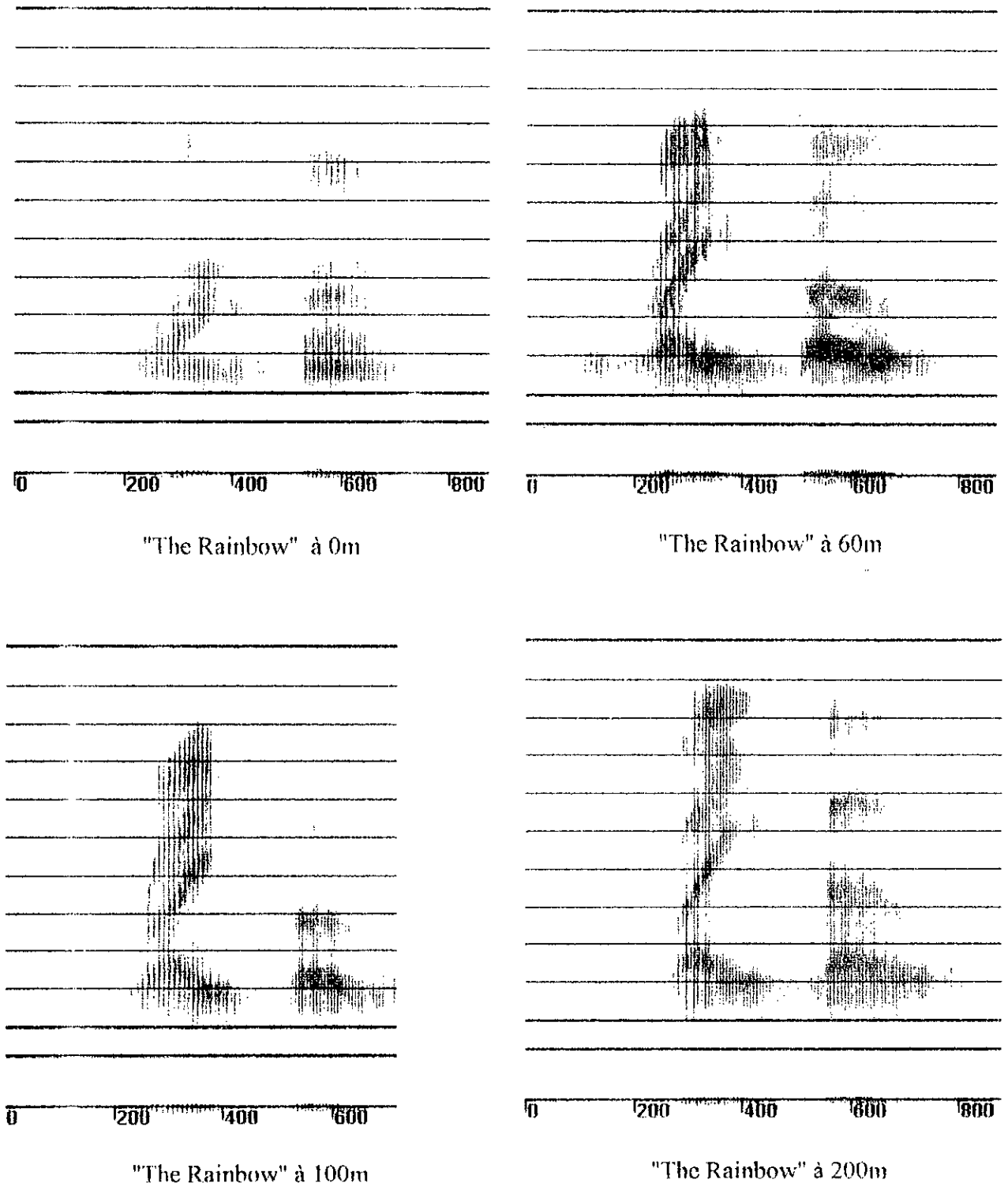


Fig. 3.4 Forme d'onde et Spectrogramme du mot "The Rainbow"

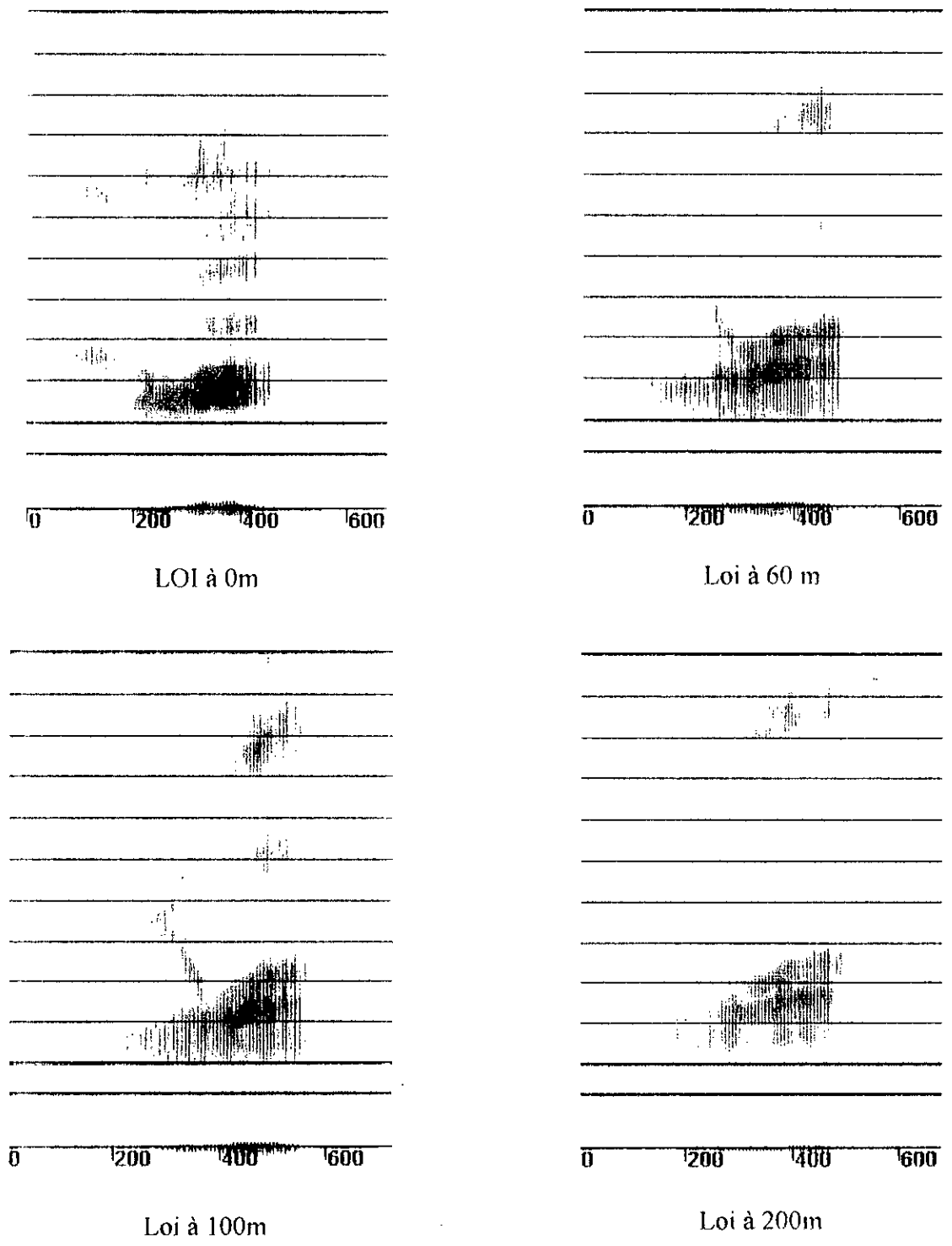


Fig. 3.5 Forme d'onde et Spectrogramme du mot "Loi"



### III.7 Conclusion

L'étude des caractéristiques du signal hyperbare, présentée dans ce chapitre nous a permis de faire le point sur l'étalement non linéaire du spectre du signal hyperbare, qui est 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. Ces deux derniers paramètres confèrent au signal hyperbare une robuste quasi-périodicité et quasi-stationnarité, d'où résulte évidemment la pertinence de l'utilisation de la modélisation du signal hyperbare par un modèle AR. Par conséquent, la prédiction linéaire développée dans le deuxième chapitre, semble la technique de décomposition la plus adaptée au signal hyperbare, qui sera la base de toutes les procédures de correction qui seront explicitées dans le chapitre qui suit.

Ce chapitre présente seulement une vue d'ensemble, qui s'avère être un support privilégié pour comprendre les systèmes de correction. On n'a pas trop insisté sur l'étude du signal hyperbare, car les nombreux travaux de recherches théoriques et expérimentaux, qui ont été consacrés à ce dernier, ont permis d'éclaircir presque toutes ses caractéristiques : en particulier, pour les profondeurs les plus généralement utilisées en plongée sous-marine entre 60m et 300m. Néanmoins, il est d'une importance stratégique de mener une étude sur le signal hyperbare, pour des profondeurs de plus de 500m jusqu'à des milliers de mètres, les plus exploitées par les militaires et les équipes de recherches qui étudient les fonds marins et l'exploitation du pétrole.

# CHAPITRE IV

## Les Systèmes de Traitement

### IV.1 Introduction

L'amélioration de l'intelligibilité de la parole hyperbare est l'objet de recherches depuis plusieurs années (1966). Les travaux présentés dans la littérature ont révélé deux méthodes, qui s'avèrent être les plus performantes. L'une est fondée sur les propriétés de la prédiction linéaire, l'autre utilise la transformée de fourrier à court terme (STFT)[7]. Néanmoins on peut les classer en deux grandes classes, les systèmes à approche temporelle et les systèmes à approche fréquentielle. Une synthèse de ces techniques est brièvement décrite dans la première partie de ce chapitre.

La deuxième partie présente 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 consacré aux résultats de simulation de chaque variante, présentés afin de valider la robustesse des algorithmes et leur puissance vis à vis des exigences de l'intelligibilité, et l'adaptation aux différentes conditions de la plongée sous marine.

## IV.2 Synthèse des Systèmes Existants

Une synthèse des différentes méthodes existant dans la littérature et concernant l'amélioration de l'intelligibilité de la parole hyperbare a révélé deux grandes classes de traitement, l'une est temporelle l'autre fréquentielle.

### IV.2.1 Les Systèmes à Approche Temporelle

Les techniques de traitement du signal hyperbare 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 les reproduire à une vitesse plus lente. Le rapport  $\frac{1}{\alpha}$  entre les vitesses serait le taux de compression spectrale.

Les limites de cette méthode, très simple, sont : l'impossibilité de travailler en temps réel, la qualité médiocre du signal de sortie, et l'altération du pitch. Cette méthode temporelle applique une compression linéaire des fréquences et des largeurs de bandes des formants par un facteur  $\alpha$ . Malheureusement, elle réduit aussi le pitch par le même facteur.

#### IV.2.1.b 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 œuvre la transformée de HILBERT, qui est 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 :

$$\frac{S(t)}{n} = [A(t)]^n \cdot \cos\left[\frac{\phi(t)}{n}\right] \quad (4.2)$$

aucun système connu n'a été conçu selon ce procédé.

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

Il consiste à obtenir une séparation de la source et de la réponse impulsionnelle du conduit vocal (déconvolution), puis de corriger cette réponse et d'exécuter l'opération inverse (convolution). Cette méthode semble la plus intéressante et la plus adaptée à l'analyse LP. L'inconvénient majeur de cette méthode est qu'elle ne peut pas modéliser exactement la non linéarité de l'équation (3.24).

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 vocal normal, qui est suivie par une dilatation temporelle et enfin une reconstitution. La segmentation nécessite la détermination du débit de chaque période, ce qui n'est pas le cas des signaux non voisés. Ce qui provoque des discontinuités du signal reconstitué. La figure 4.1 illustre le principe des systèmes temporels.

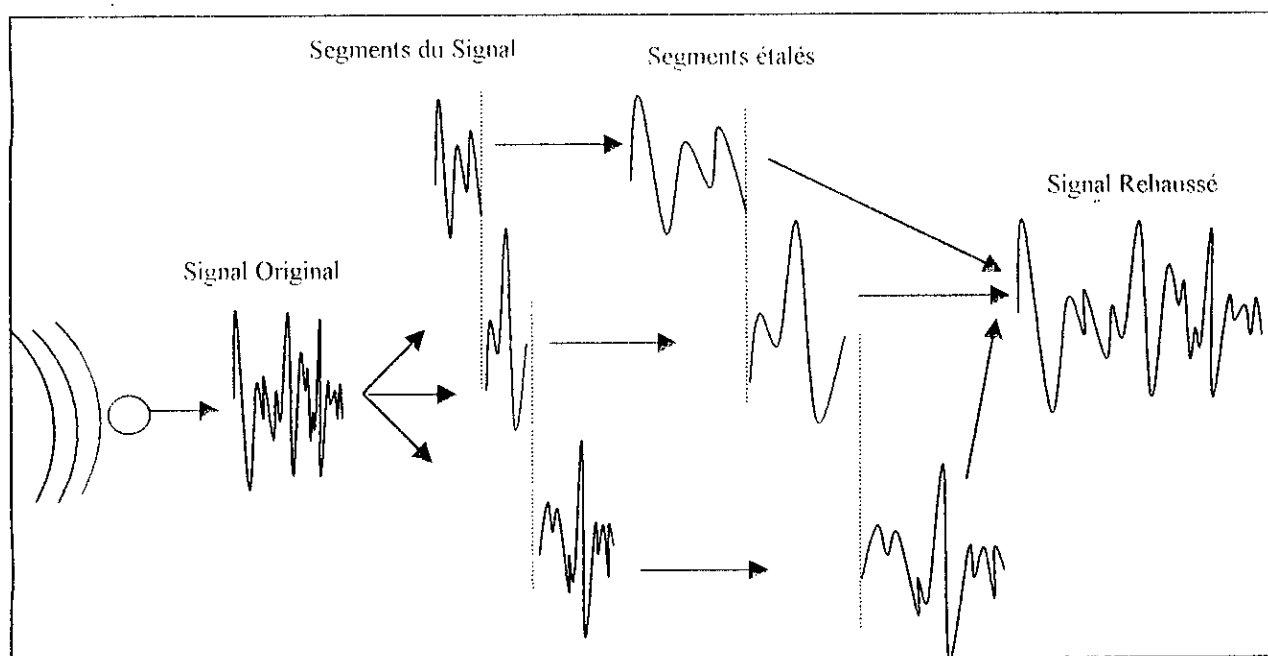


Fig. 4.1 Approche temporelle de traitement de la parole hyperbare [20,22]

#### IV.2.2 Les Systèmes à Approche Fréquentielle

Plusieurs techniques de correction à approche fréquentielle ont été décrites dans plusieurs articles de recherches. Elles reposent en fait 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 :

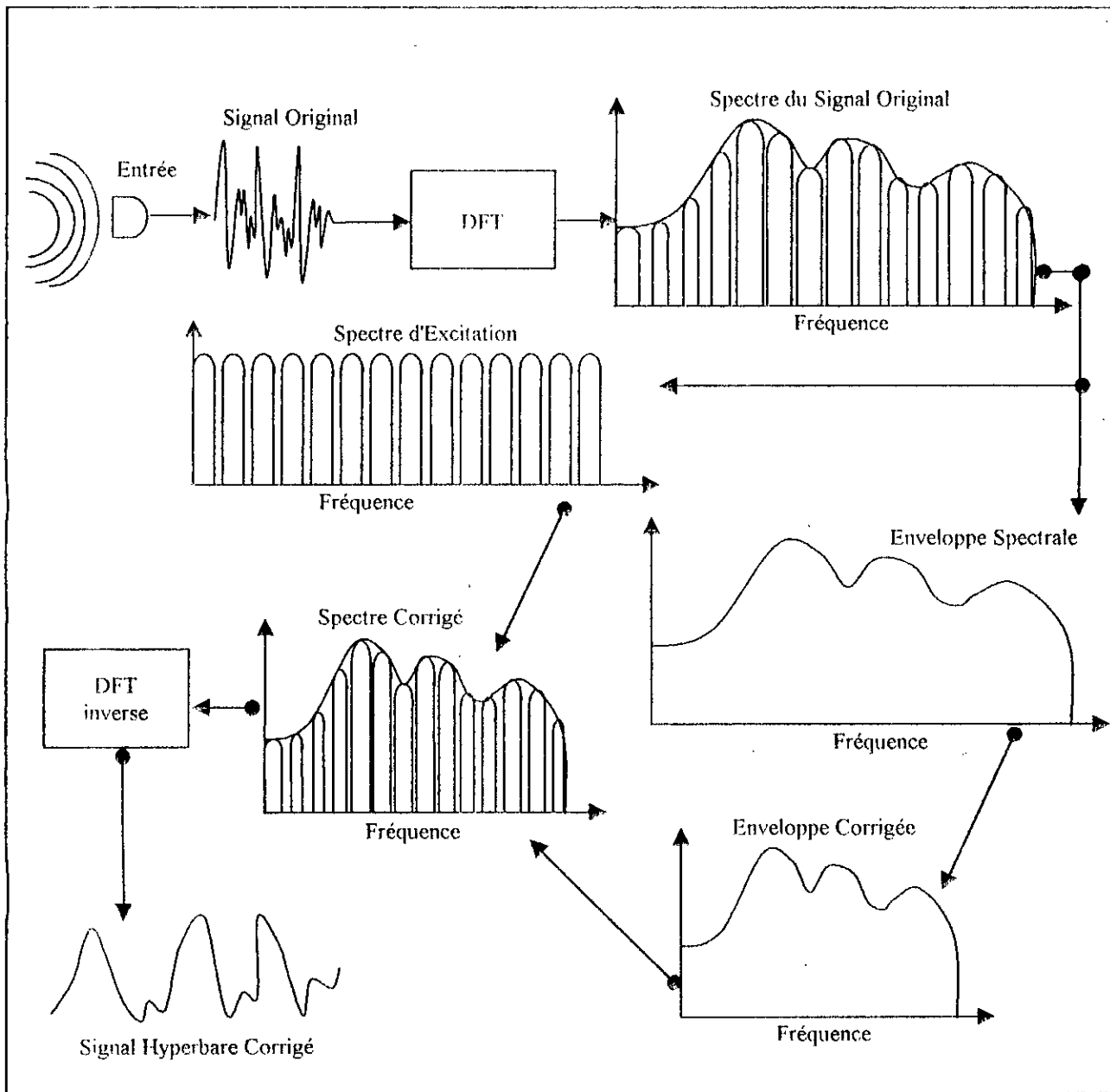


Fig. 4.2 Approche fréquentielle de traitement de la parole hyperbare [39]

### 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 :

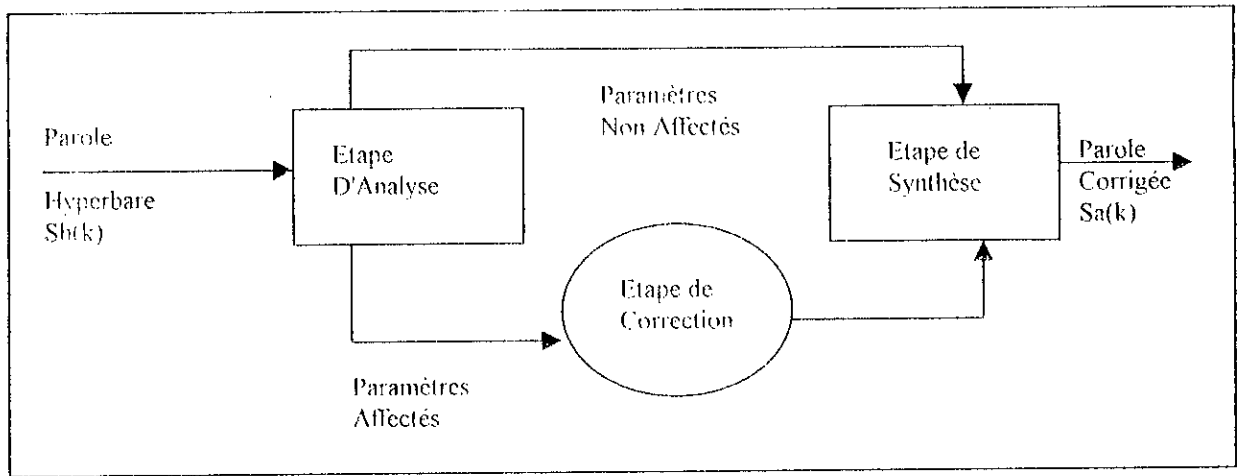


Fig. 4.3 Principe du système de correction

#### IV.4 Correction des Corrélations

Soit  $x_h(n)$  ( $n = 0, 1, \dots, N$ ) la fenêtre en cours d'analyse du signal hyperbare original à 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 :

$$s_h(n) = x_h(n)w(n) \quad \text{pour } n = 0, 1, \dots, NH, \quad (4.3)$$

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

$dH$  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 corrélations 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 :

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

- 3- Détermination des coefficients  $a(j) : j = 1, \dots, p$ , du filtre d'analyse **LPC** du signal hyperbare, à partir des autocorrélations en utilisant l'algorithme de **WLD**.
- 4- Calcul du 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 :

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

- 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 :

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

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

$$R_c(j) = (Fact).R(T_j + 1) + (1 - Fact).R(T_j)$$

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  $M_h = K \times p = 46$ .

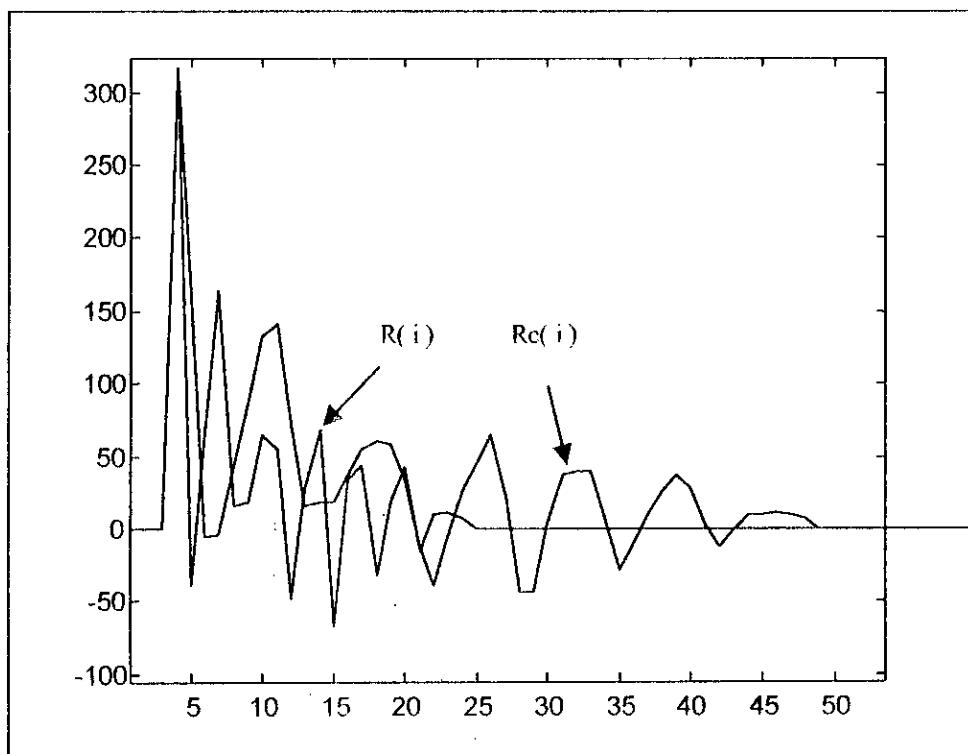


Fig. 4.4 L'étalement temporel des autocorrélations

- 6- Obtenir de nouveaux coefficients **LPC**  $a_c(j)$  ( $j = 0, 1, \dots, K \times p$ ), 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 coefficients **LPC** obtenus, vont servir à 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 :

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



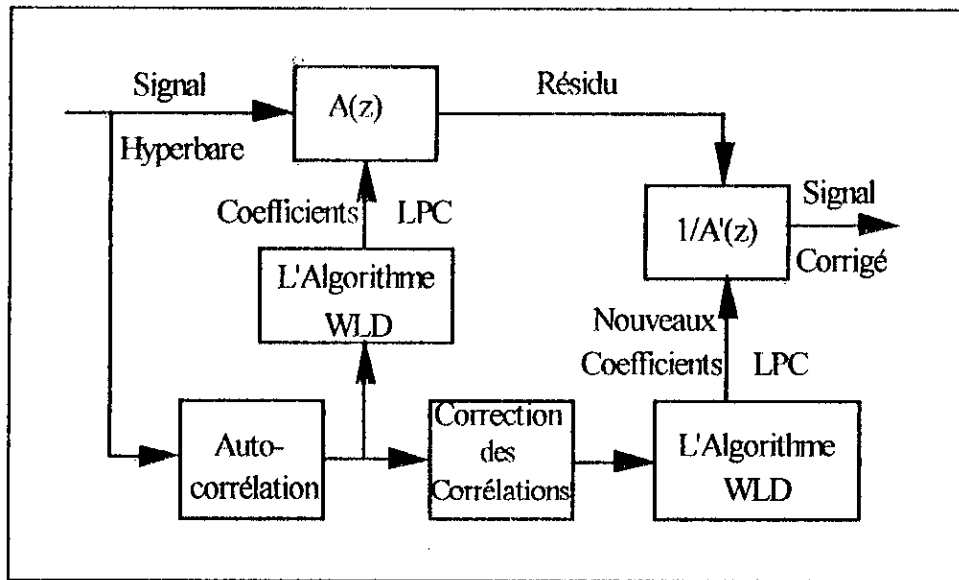


Fig. 4.5 Algorithme de correction des autocorrelations

### 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 autocorrelations, l'algorithme de correction spectrale (fig. 4.6) peut se présenter comme suit :

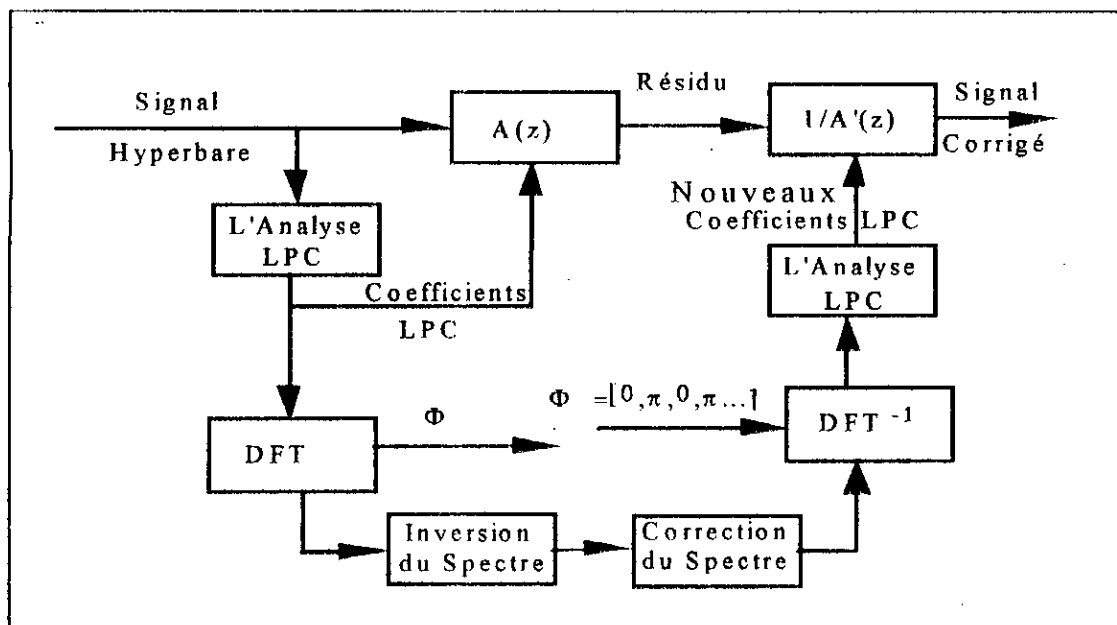


Fig. 4.6 Algorithme de correction spectrale

- 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 la 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) :

$$\{1, a(1), a(2), \dots, a(p), 0, \dots, 0\} \xleftrightarrow{\text{DFT}} A(n) = 1 + \sum_{k=1}^p a(k) e^{-j \frac{2\pi n k}{N}} \quad (4.9)$$

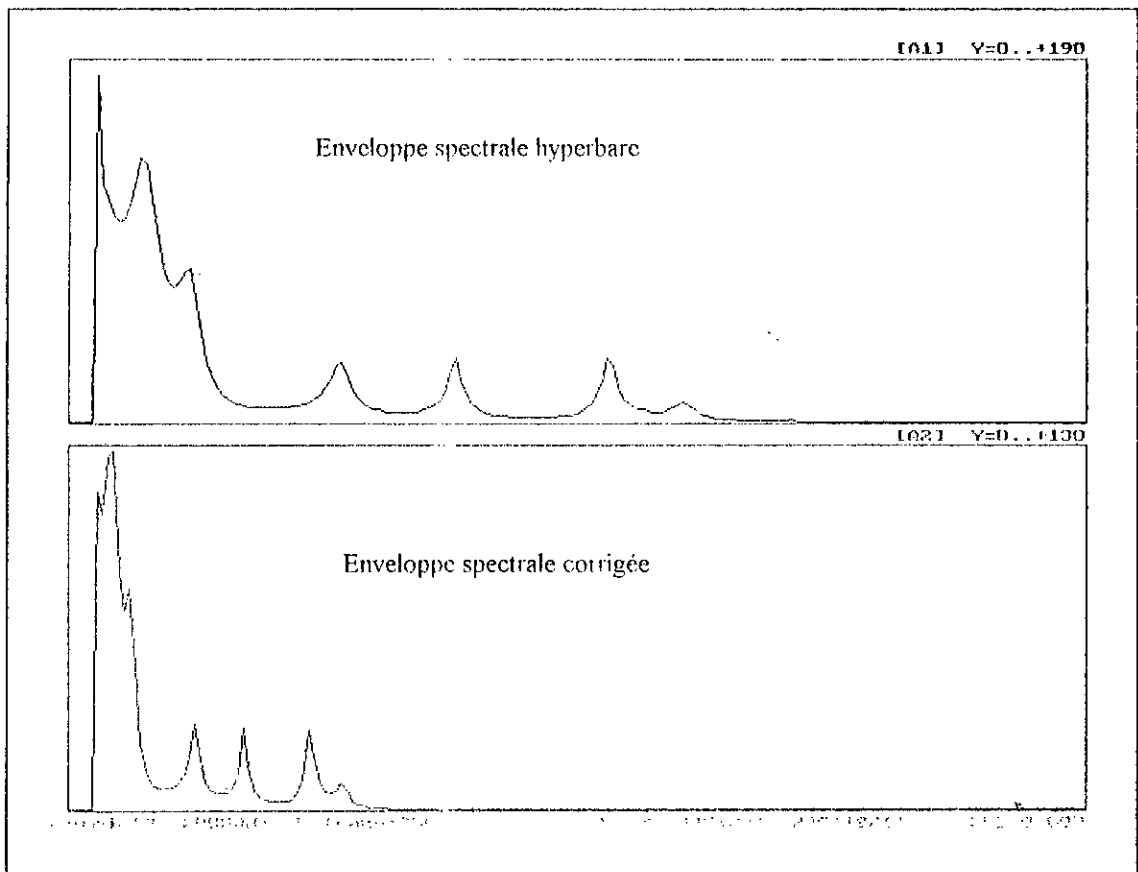
$$n = 0, 1, \dots, \text{NDFT}-1$$

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 (fig. 4.7) est assurée par un filtre interpolateur, capable d'aller rechercher la composante spectrale 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 jouit de la relation de récurrence :

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



**Fig. 4.7** Correction linéaire du spectre

- 6- Revenir à l'espace temps, à 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 échantillon 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'autres effets pour la partie dite "basses fréquences". Une fonction tabulée est utilisée à cette fin inspirée de la formule de G.Fant & J.Lindquist (3.24) ; La figure 4.8 illustre les valeurs de k utilisées en fonctions de la bande de fréquence.

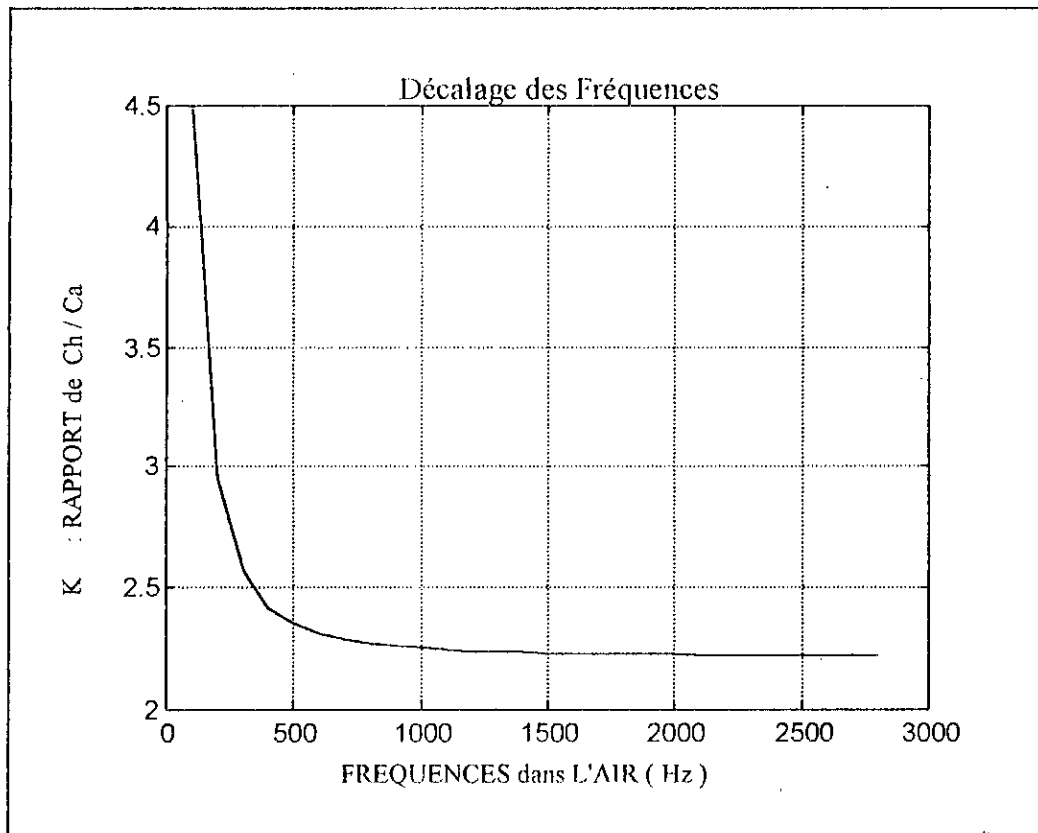


Fig. 4.8 Les valeurs de K pour une compression non linéaire

## 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, constituée 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 (INPP). Elle répond à la demande des professionnels de la plongée et des industriels qui souhaitent disposer aussi bien d'une base de sons pour le développement de nouveaux procédés de "décodage" de la parole hyperbare que d'un outil pour l'évaluation des systèmes de communication vocale.

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 disque CDROM. Tous les fichiers sont conformes au format standard SAM Europec (Projet CEE-ESPRIT "SAM" n°2589).

La fréquence d'échantillonnage des fichiers des signaux originaux est de 40 khz, et le codage est fait sur 16 bits. Les fichiers, en nombre de cinq, de la parole hyperbare qu'on a sélectionné ont été transformés en fichiers dont la fréquence d'échantillonnage est de 20 kHz. En plus, de deux autres fichiers de la parole dans des conditions normales (SOE0.IRS – SOF0.IRS).

- La phrase choisie pour le français (" tous les ans, on envoie mille tonnes de riz au Mali ") est enregistrée à (0m) sans bruit, à (300m) en caisson à l'hélium et en piscine, toutes les deux bruitées.
  - La phrase utilisée pour l'anglais (" The rainbow is a division of white light into many beautiful colours ") est enregistrée à (0m), à (60m), à (100m), à (200m) sans bruit.
- Les conditions d'enregistrement des fichiers sélectionnés sont illustrées au tableau 4.1 :

Fichier Hyperbare	Profondeur (m)	TGZ (C°)	ppO2	$\rho_{O_2}$	$\rho_{He}$	Nombre de Trames Tr=256 échantillons
SOE60.IRS	60	29.2	592mb	8.5	91	361
SOE100.IRS	100	25.3	600mb	5.5	93	361
SOE200.IRS	200	30.5	399mb	5.09	95	371
SNF300c.IRS	300	31.5	400mb	1.31	96.4	332
SNF300p.IRS	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  $\rho_{O_2}$  ;  $\rho_{O_2}$  : 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 faire 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	$M = 16 \dots 20$
Ordre du filtre de synthèse	$M_h = K \times M$
Fenêtre de pondération	Fenêtre de Hamming
Longueur des trames d'analyse	256 (12.8 ms) – 512 (25.6 ms)
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 fait 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. Premièrement, la qualité du signal reconstitué est un concept difficile à expliciter. Elle implique normalement l'intelligibilité (compréhension de ce qui est transmis). C'est un concept objectif auquel il faut joindre le concept subjectif de la parole. L'oreille humaine est en effet souvent le juge final de l'efficacité de tels algorithmes. De plus, la complexité du système de traitement est une caractéristique très importante.

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.

Le paramètre le plus important, qui doit être estimé et incorporé à tous les algorithmes de corrections est le coefficient de compression  $K$  (ou  $\alpha$ ). Ce paramètre a été choisi constant, la valeur de 2.22 a donné de très bons résultats pour des profondeurs de moins de 200m. En plus il facilite beaucoup la mise en œuvre des algorithmes. Par ailleurs, ses meilleures valeurs pour des 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. Ainsi, une procédure d'estimation de  $K$  peut être rajoutée à tous les algorithmes, qui les rendent eux mêmes dynamiques. Cette procédure est basée sur les équations des gaz parfaits développées dans le chapitre trois. Les informations concernant la profondeur d'immersion et le mélange gazeux, qui se trouvent dans un fichier de données, permettent d'estimer le coefficient  $K$  en se basant sur les équations (3.2 - 3.10). Néanmoins, comme cette estimation n'est pas exacte, on se contente 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 simplicité de conception et qualité du signal synthétisé, surtout pour les faibles profondeurs, tout en gardant un coefficient de dilatation constant égal à sa valeur intégrale.

Par ailleurs, cet algorithme de traitement a donné naissance à un bruit essentiellement situé dans la partie haute du spectre du signal de synthèse ou du signal corrigé. Les tentatives de filtrage de ce bruit en agissant sur les autocorrélations donnent parfois, lieu à des instabilités de l'algorithme Wiener-Levinson-Durbin (WLD). Pour réduire les effets de ce bruit, un filtre passe bas a été introduit après la synthèse. Auparavant, un filtre d'accentuation est utilisé avant les corrections afin de réduire les effets d'atténuations du filtre précédent.

Le test de la variante avec des corrections linéaires du spectre du signal hyperbare, qui est une variante des systèmes à approche fréquentielle nécessite des transformées de Fourier de 256 échantillons, afin d'estimer l'enveloppe spectrale du conduit vocal à partir des  $p$  coefficients LPC prolongés par des zéros jusqu'à la longueur de la DFT. Cette variante ne nécessite pas de filtrage ; la visualisation des spectres corrigés confirme cette suggestion et les tests subjectifs montrent bien que le filtre n'apporte aucune amélioration. De plus, elle garde toujours un coefficient de compression constant, pour toute la gamme des fréquences. L'intelligibilité des signaux corrigés par cette méthode est meilleur que celle de la première variante. Malheureusement, le passage au domaine fréquentiel imposé par la DFT rend l'algorithme plus long par rapport à celui des corrections des corrélations, en augmentant les contraintes d'une éventuelle implémentation.

La dernière variante testée est celle qui applique une compression non linéaire du spectre. Elle est la plus adaptée aux grandes variations non linéaires des formants, présentées par les signaux hyperbares prononcés à des grandes profondeurs (plus de 200m et 300m). Celle-ci utilise un coefficient de compression constant pour les composantes spectrales de hautes fréquences et une valeur dynamique pour celles de basses fréquences (fig. 4.8). De plus, le temps de calcul mis par cette variante est très proche de celui de la compression linéaire du spectre.

Les différents algorithmes de traitement ont été tout d'abord optimisés en choisissant les paramètres qui leurs permettent d'obtenir les meilleurs résultats possibles. C'est ainsi que de nombreux essais ont été effectués sur les différents signaux avant d'obtenir les valeurs adéquates des paramètres cités plus haut.

Bien que l'évaluation de la robustesse de nos algorithmes repose sur des tests subjectifs, la figure 4.9 illustre un exemple d'une trame du fichier SOE200.IRS. Une représentation des résultats sous forme de spectrogrammes semble la plus explicite ; Elle illustre mieux les variations et les corrections apportées. Ainsi, la figure 4.10 montre le spectrogramme du signal original hyperbare à 200m et la version du signal synthétisé avec correction des corrélations et correction spectrale. La figure 4.11, quant à elle, illustre les spectrogrammes pour une profondeur de 300m en plus d'un spectrogramme avec compression non linéaire du spectre.



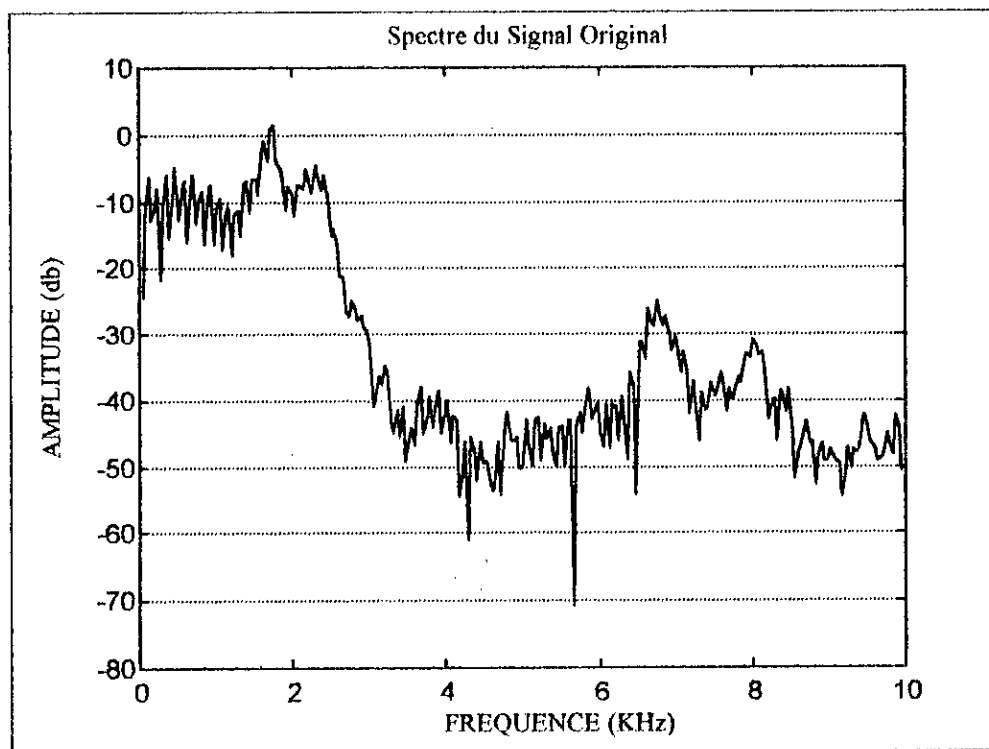
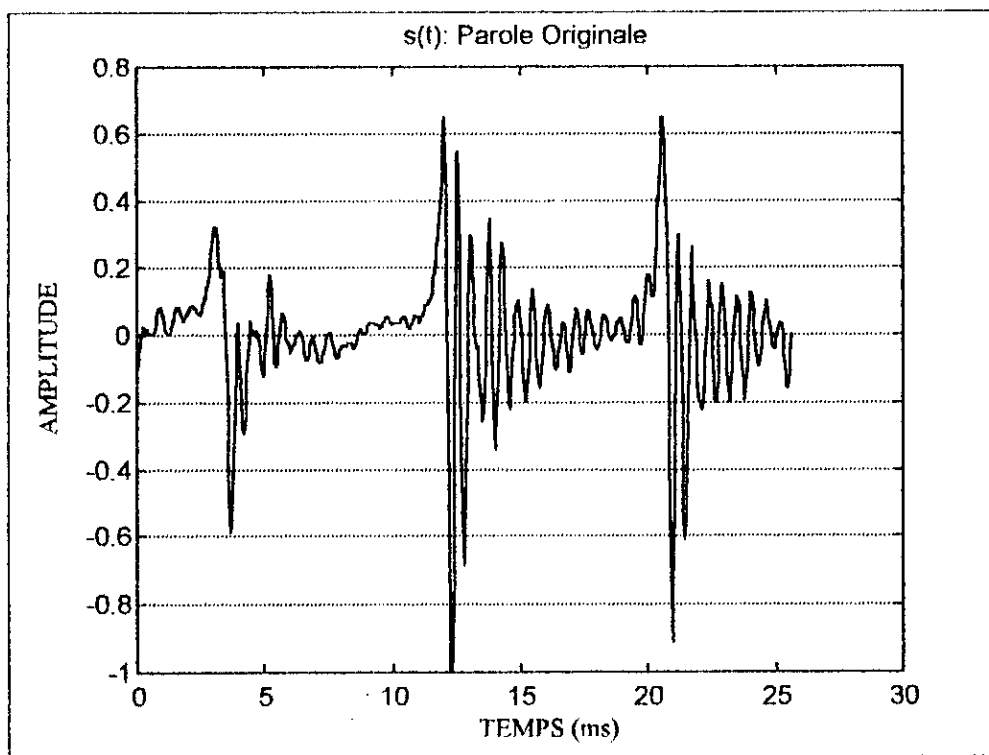


Fig. 4.9.a Exemples d'une tranche du signal original et son spectre

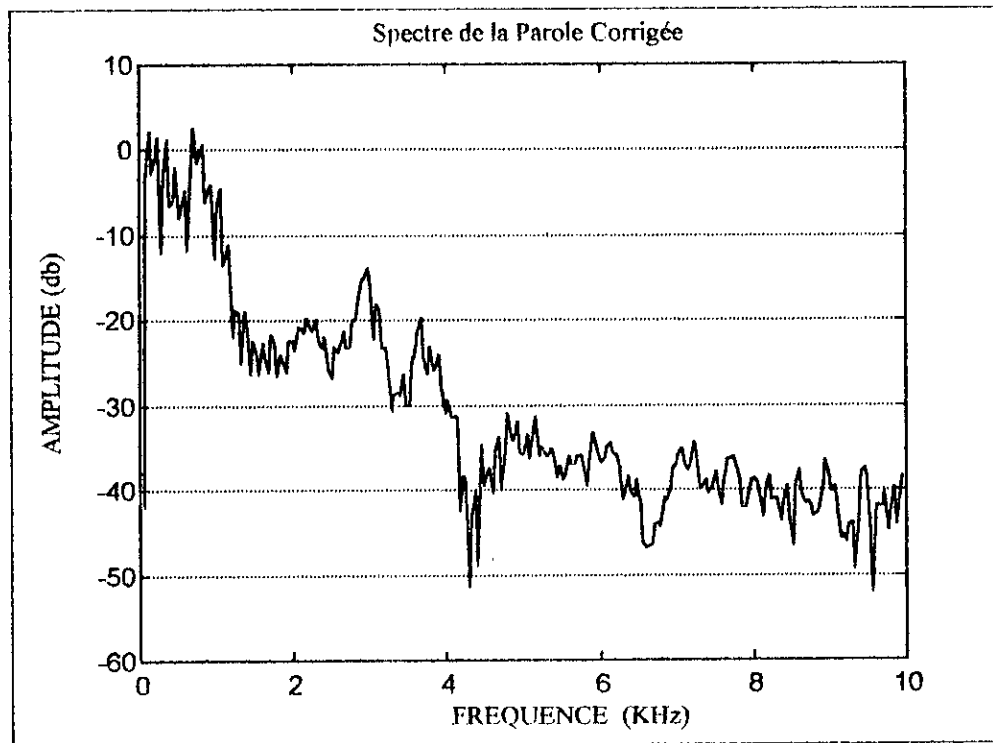
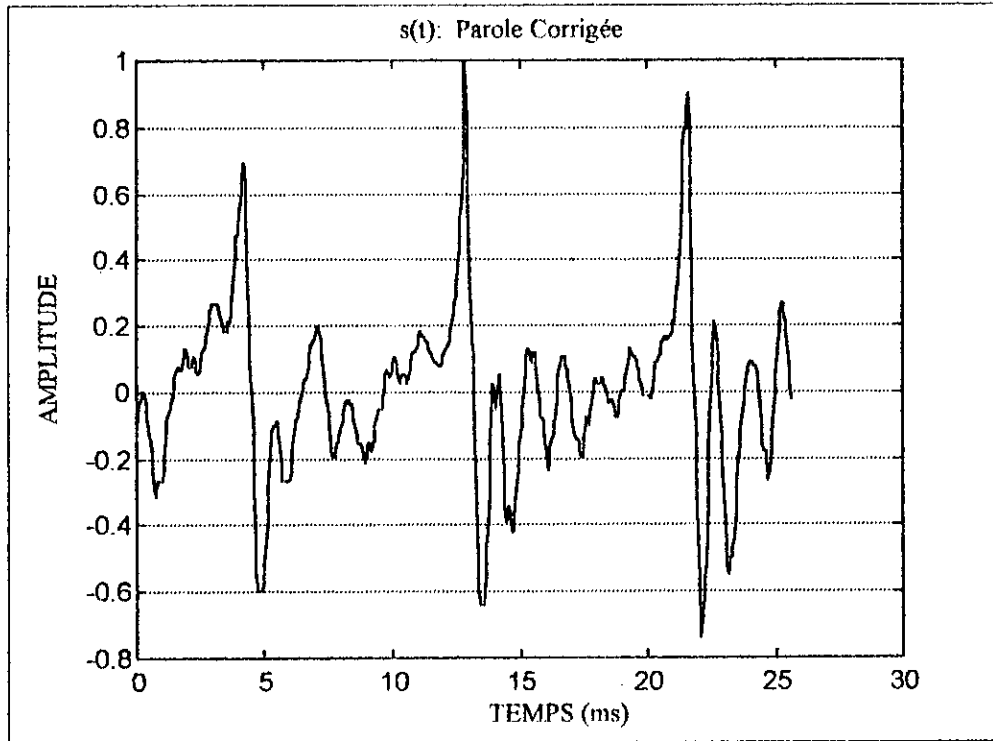


Fig. 4.9.b Exemples des résultats obtenus avec correction des autocorrélations

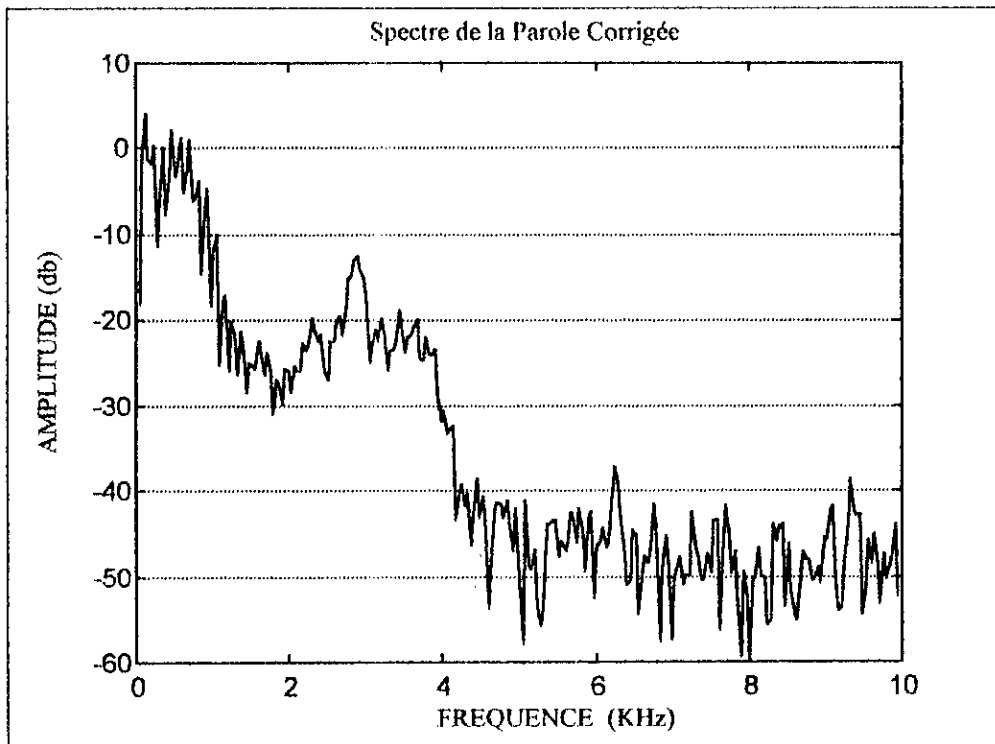
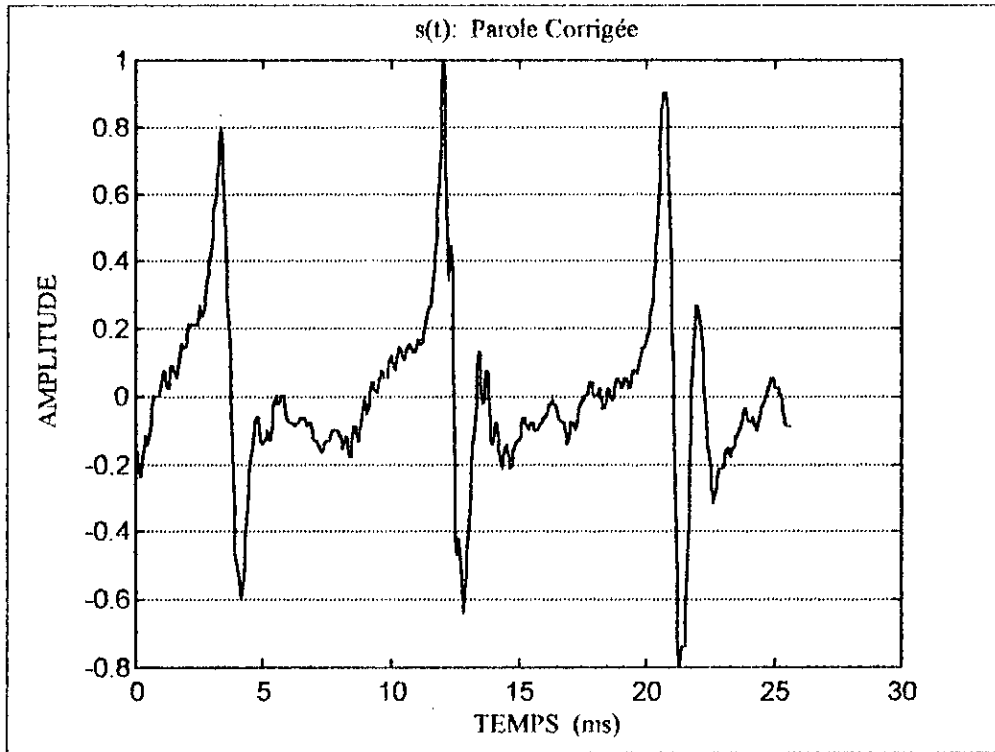


Fig. 4.9.c Exemples des résultats obtenus avec correction spectrale

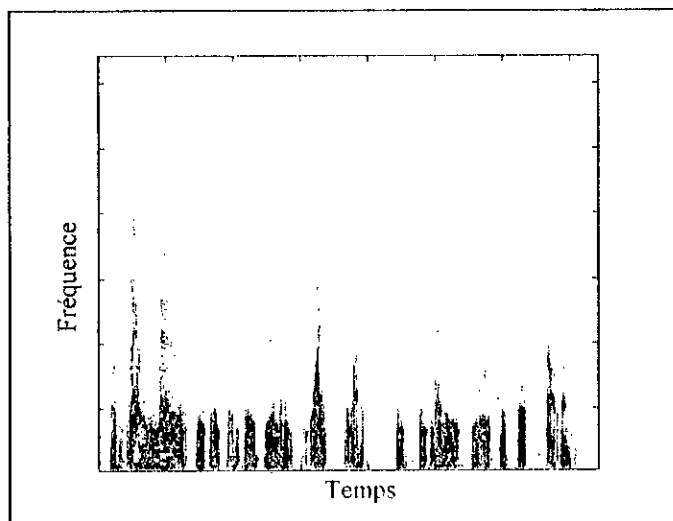


Fig. 4.10.a Spectrogramme du signal hyperbare original à 200m

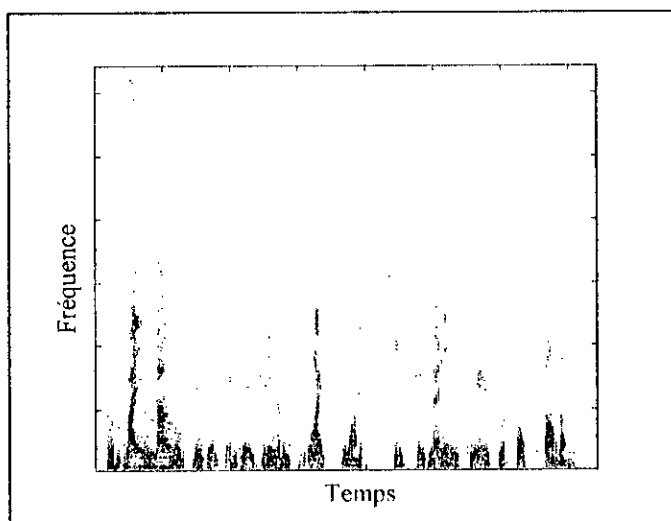


Fig. 4.10.b Spectrogramme du signal synthétisé avec correction des corrélations

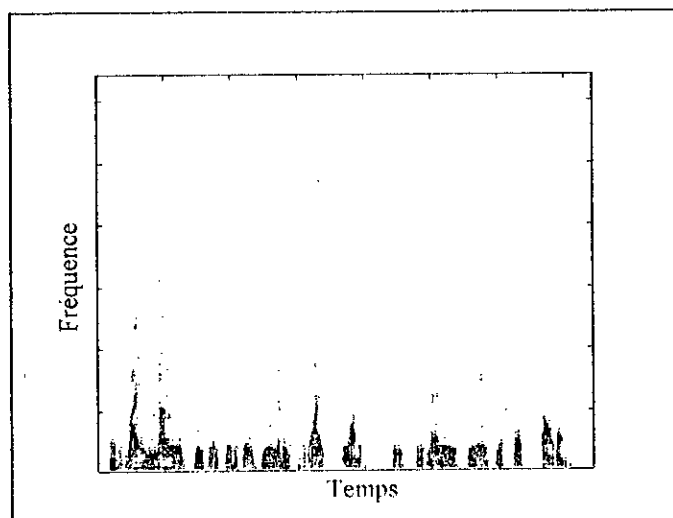


Fig. 4.10.c Spectrogramme du signal synthétisée avec correction spectrale

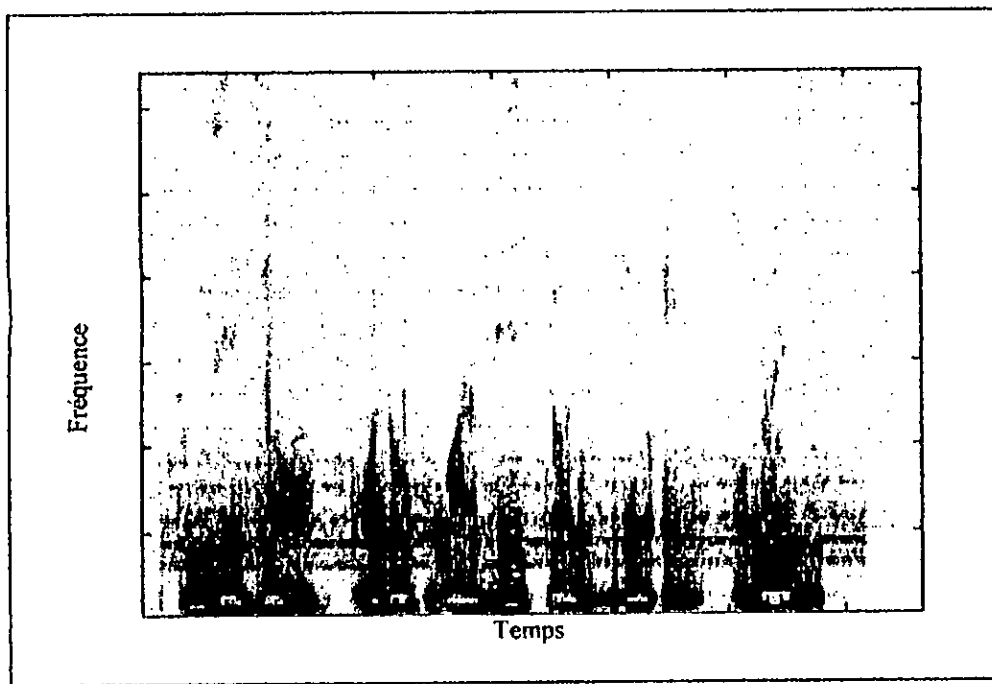


Fig. 4.11.a Spectrogramme du signal hyperbare original à 300m

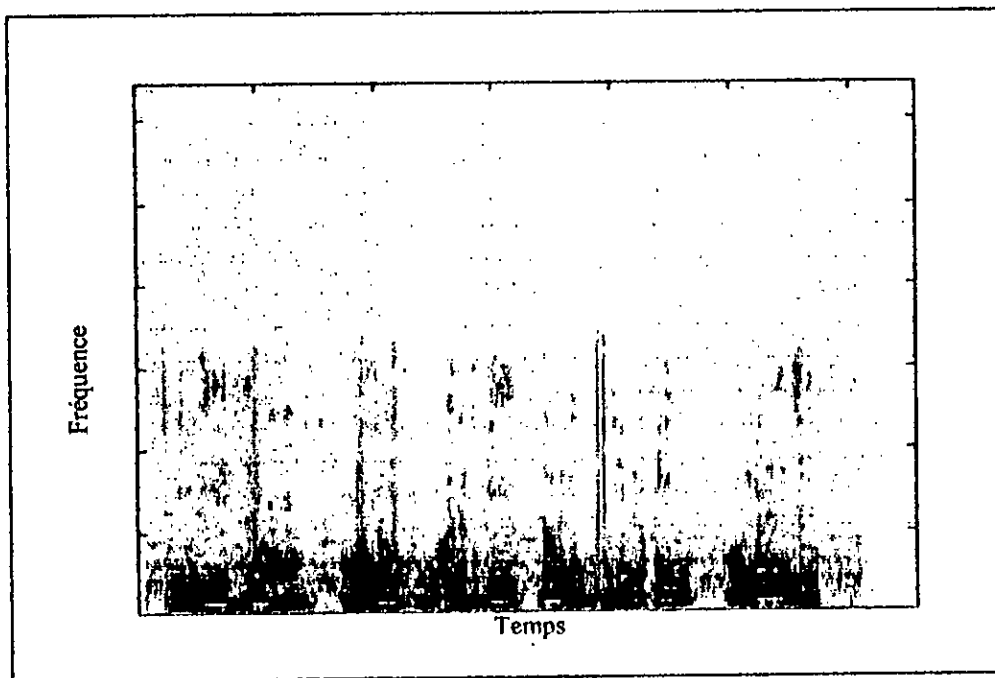


Fig. 4.11.b Spectrogramme du signal synthétisé avec correction des corrélations

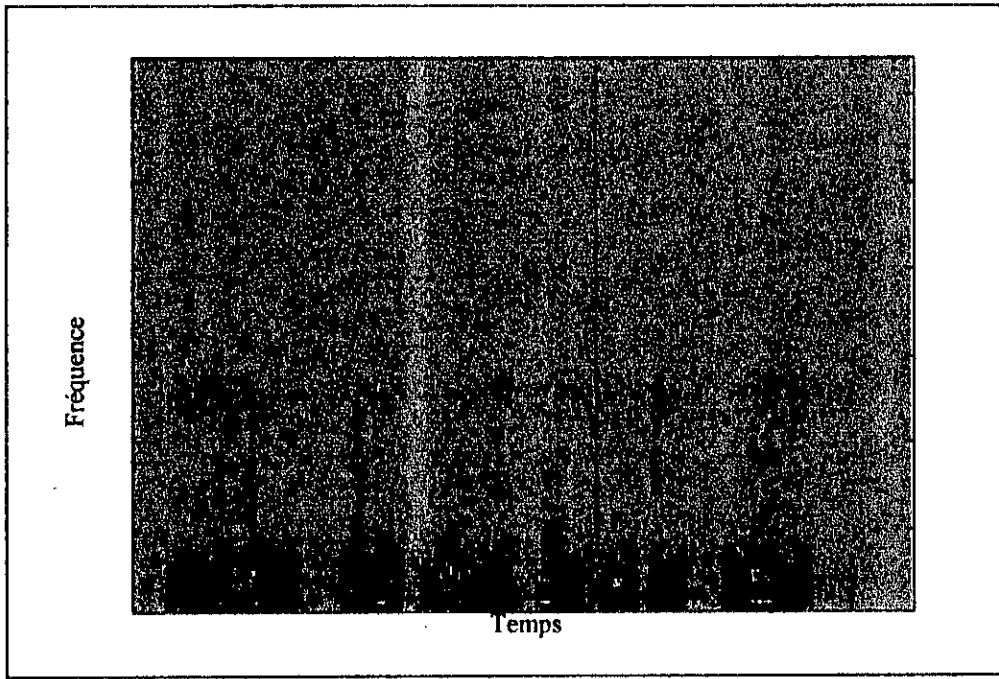


Fig. 4.11.c Spectrogramme du signal synthétisé avec correction linéaire du spectre

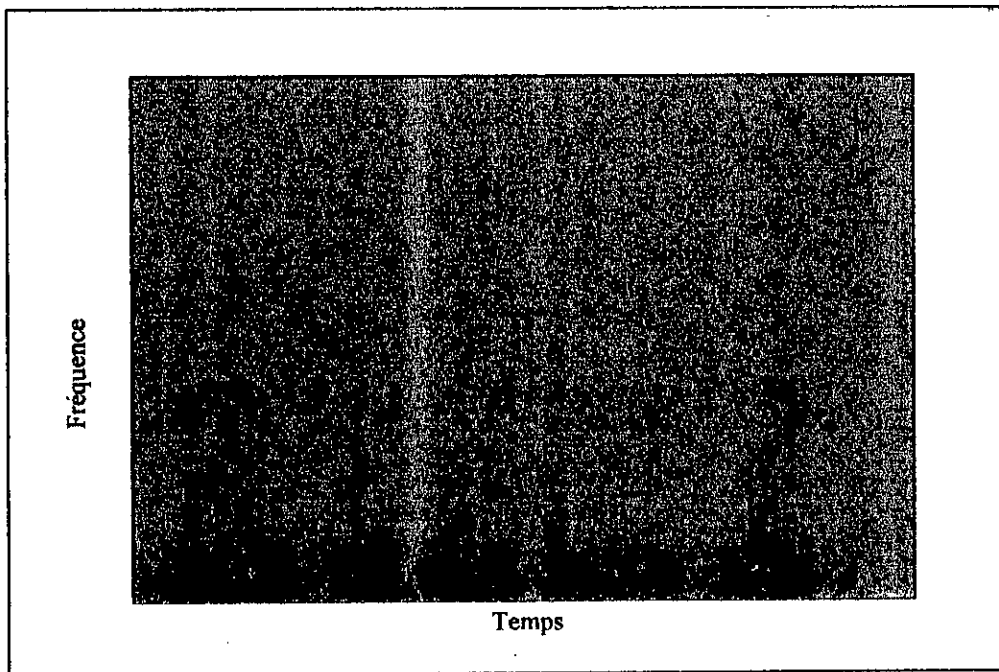


Fig. 4.11.d Spectrogramme du signal synthétisé avec correction non linéaire du spectre

## IV.7 CONCLUSION

En guise de 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 vue 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 la quelle 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 et qui seront développés dans le chapitre suivant.

# CHAPITRE V

## Réduction de Bruit d'un Signal Hyperbare

### V.1 Introduction

Le problème de la suppression de bruits, qui contamine le signal utile est devenu très stratégique ces derniers temps tant pour les applications militaires que pour des applications civiles. Les différents milieux où les signaux utiles sont prononcés se trouvent très hostiles, à savoir le milieu ambiant, l'hélicoptère, la voiture(téléphone main – libres),.... Quant 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 de bruit, qui seront incorporés aux techniques de correction proposées dans le chapitre précédent. Ainsi, deux versions de réduction de bruits seront exposées dans la première et la deuxième partie 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 inintelligible résultat du 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. En 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'où les pertes d'information du signal hyperbare. La figure 5.1 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.

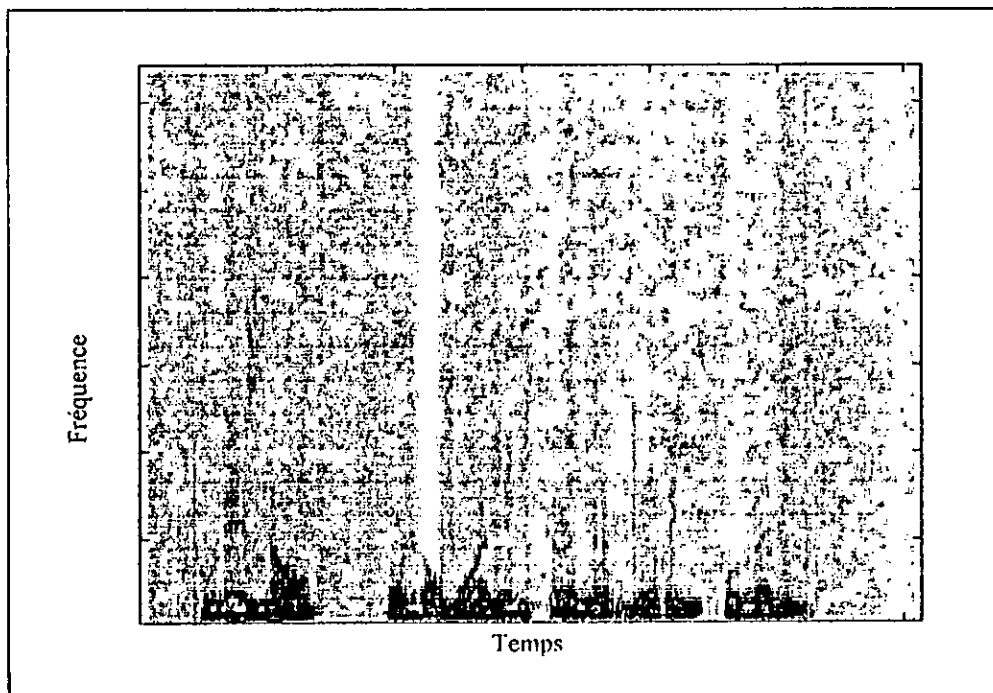


Fig. 5.1.a Spectrogramme de SOF0.IRS

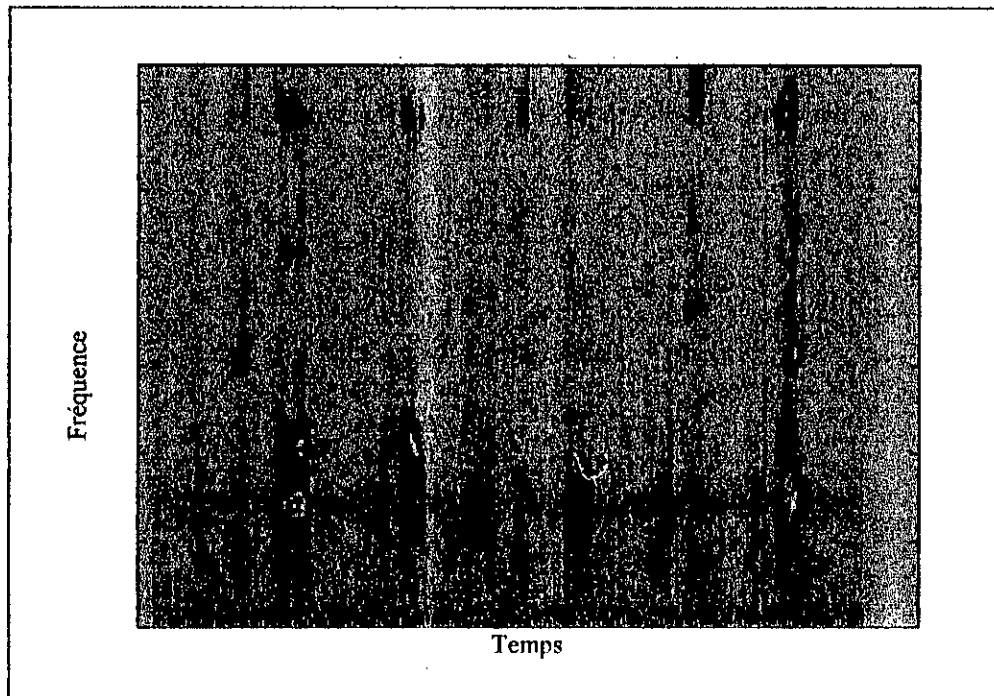


Fig. 5.1.b Spectrogramme de SNF300p.IRS

## 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 plupart des zones de fréquence 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.

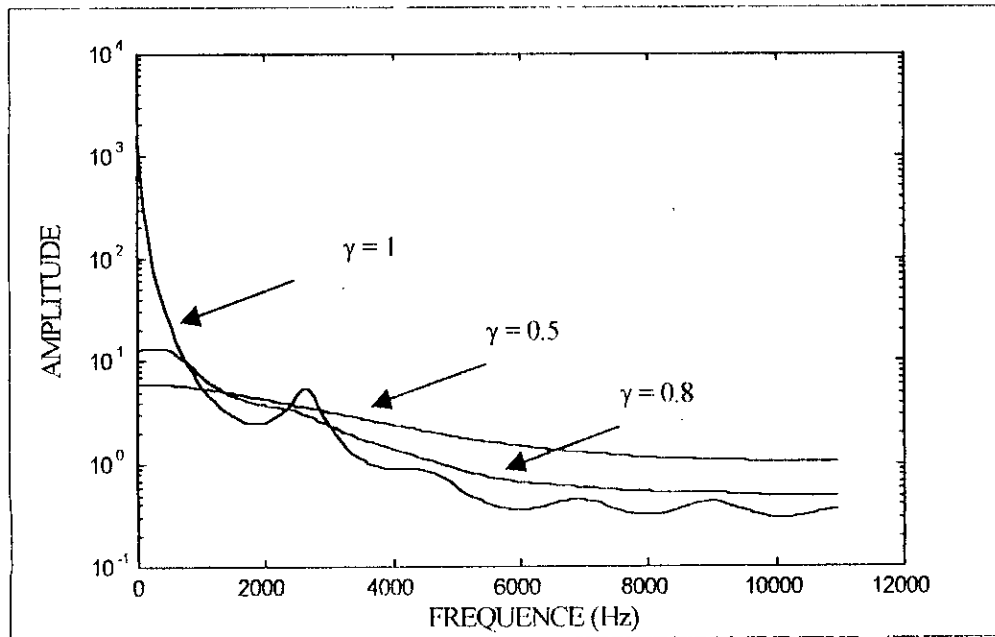


Fig. 5.2 Spectre de  $F(z)$ - ( $\gamma = 0.5, 0.8, 1$ )

La figure 5.3 schématise l'algorithme de correction des corrélations modifié :

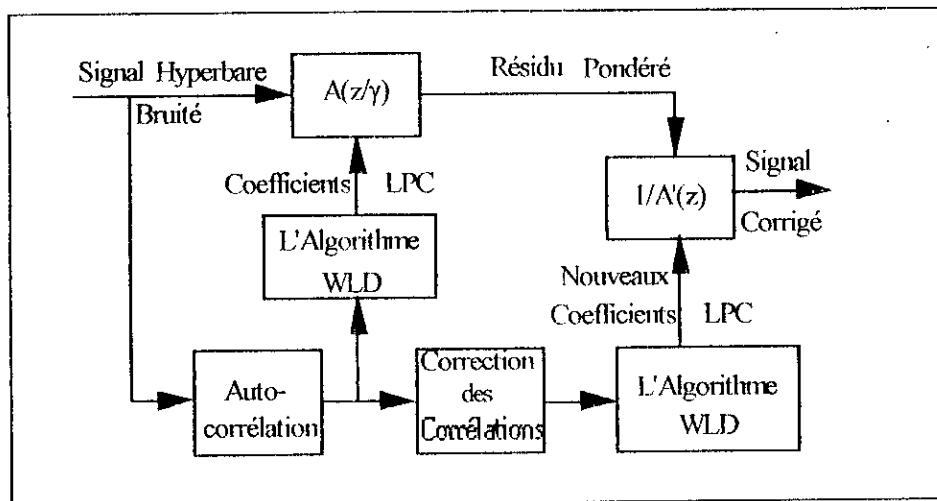


Fig. 5.3 La correction des corrélations avec filtre de pondération

Le nouveau résidu est donné par :

$$residu\_w(n) = X_h(n) + \sum_{j=1}^p a(j) \cdot \gamma^j \cdot X_h(n-j) \quad \text{pour } n = 0, 1, \dots, N \quad (5.1)$$

L'objectif est de déterminer la valeur optimale de  $\gamma$  en fonction de la RSB. Soong et Sondhi[31] par exemple ont utilisés les valeurs de  $\gamma$  qui minimisent la distorsion pondérée d'Itakura entre le modèle AR sans bruit et l'estimé AR du signal bruité(fig. 5.4) pour chaque trame dans le cas de la reconnaissance de la parole bruitée.

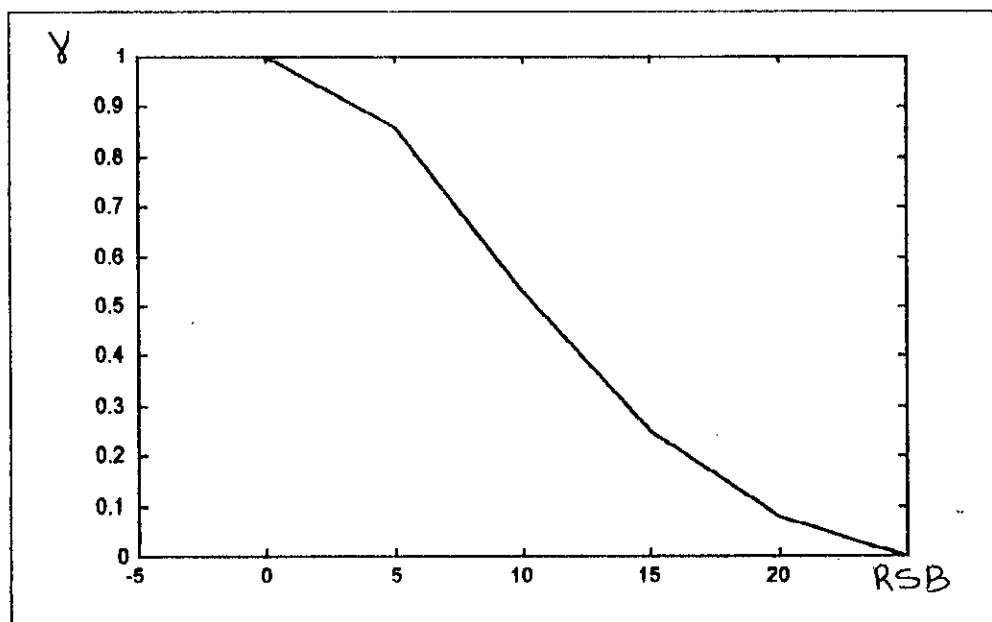


Fig. 5.4 La valeur optimale de  $\gamma$  en fonction du RSB

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 satisfaisantes(fig. 5.5). 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 le rôle d'un filtre passe bas, ce qui assourdit la parole synthétisée résultante, d'où l'utilisation d'un filtre de préaccentuation. 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.

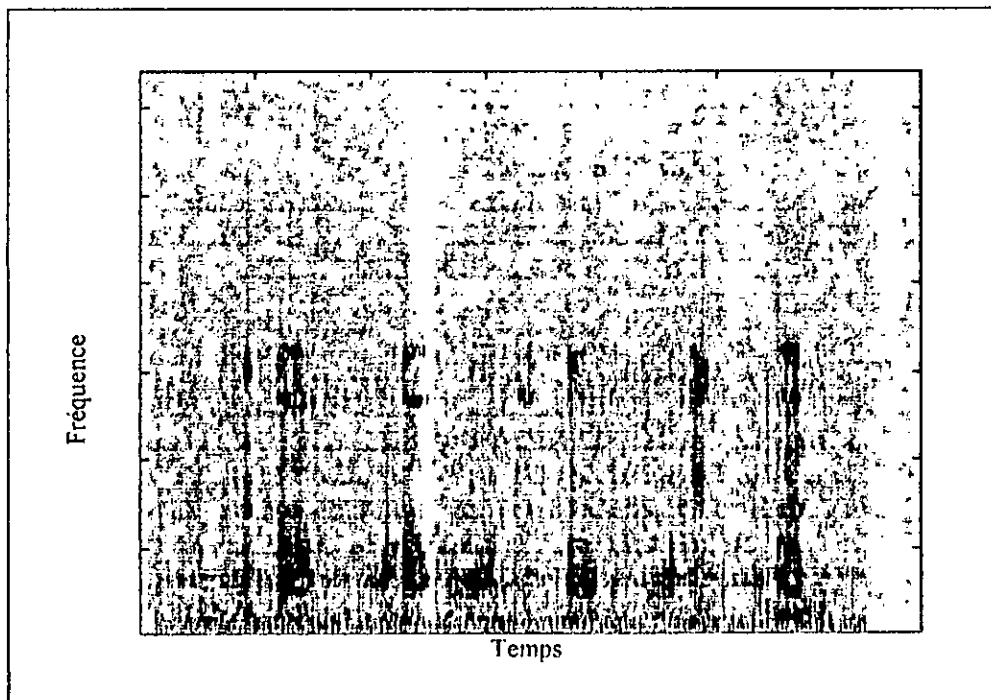


Fig. 5.5.a Spectrogramme du SYN300P(corréction des corrélations)

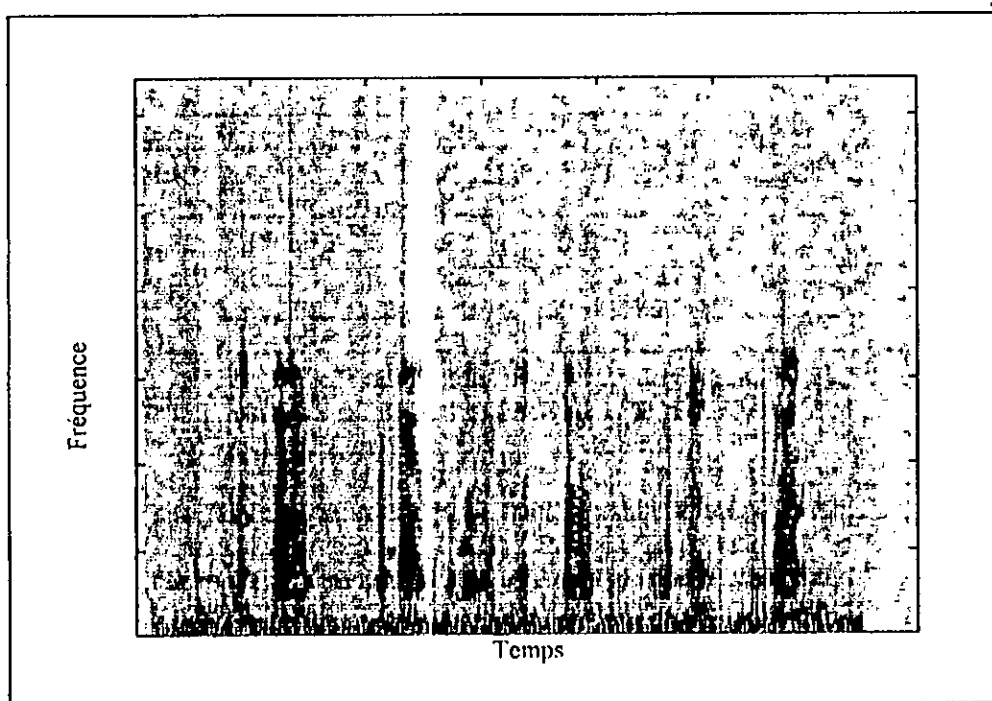


Fig. 5.5.b Spectrogramme du SYN300P(pondération du résidu  $\gamma=0.8$ )

### V.3.2 La Représentation SMC de la Parole Hyperbare Bruitée

Cette méthode est une version de la correction des corrélations 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'autocorrélation (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  $2.N$  échantillon, on calcule  $N + 1$  coefficients de corrélation selon la relation suivante :

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

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

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

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

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

où  $NDFT$  est la longueur de la DFT et  $\rho^h(0)$  n'est pas incluse dans la somme, particulièrement dans ce cas où 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  $R(k)$  :

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

- 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  $2.N. \ln N$  multiplications-

additionset (  $N + 1$  ) opérations de valeurs absolues complexes pour la 3<sup>ième</sup> et la 4<sup>ième</sup> 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 (fig. 5.6).

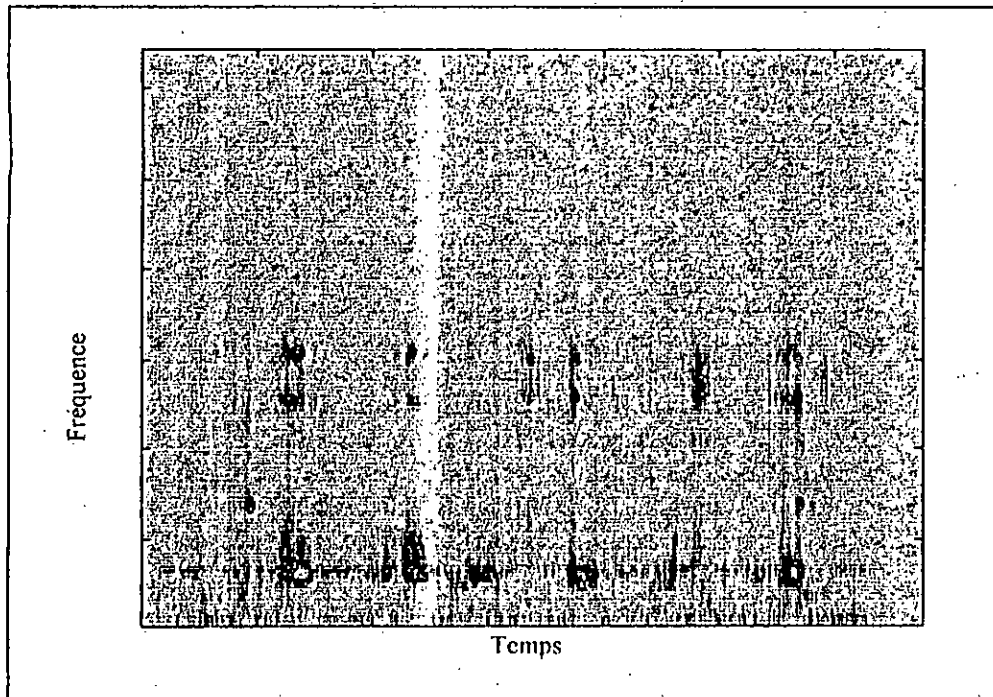


Fig. 5.6 Spectrogramme du SYN300P (SMC)

## 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 de 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 décorréler, et  $S(k)$  de moyenne nulle, avec  $X(w)$ ,  $S(w)$ ,  $N(w)$  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  $|N(w)|^2$  est connu, donc  $S(k)$  peut être facilement estimé par la relation suivante :

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

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 5.7 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.



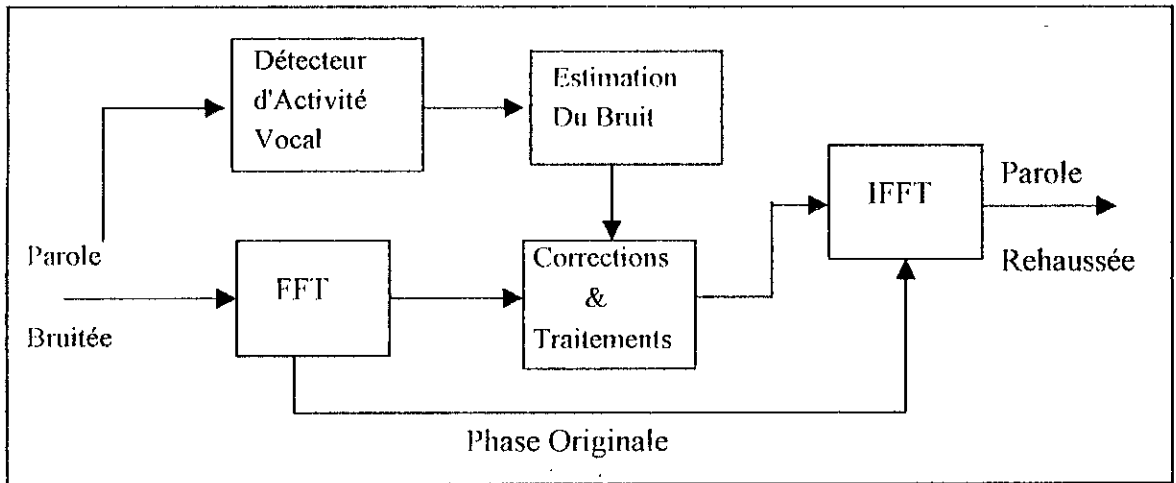


Fig. 5.7 Le Principe de Base de la Soustraction Spectrale

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 :

$$|S(n)|^\mu = |X(n)|^\mu - \lambda |N(n)|^\mu, \quad n = 0, 1, 2, \dots, \text{NDFT}/2 - 1 \quad (5.9)$$

où  $|N(n)|$  est la  $n^{\text{ième}}$  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 :

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

La méthode de soustraction spectrale que nous avons intégré au système complet de réduction et de correction (fig. 5.8) est celle de Berouti, qui peut se formuler de la façon suivante pour chaque trame  $i$  :

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

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 distorsions 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 afin de minimiser la distorsion.

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

$$RSB_i = 10 \cdot \text{Log}_{10} \left[ \frac{EX_i - EN_i}{EN_i} \right] \tag{5.12}$$

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

$$\begin{aligned} \bar{\lambda}_i &= \lambda_0 - \left[ \frac{\lambda_0 - \lambda_{\min}}{RSB_0} \right] RSB_i \\ \lambda_i &= \begin{cases} \lambda_{\min} , & \bar{\lambda}_i > \lambda_{\min} \\ \bar{\lambda}_i , & \lambda_{\min} \leq \bar{\lambda}_i \leq 5 \\ 5 , & \bar{\lambda}_i > 5 \end{cases} \end{aligned} \tag{5.13}$$

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  $RSB_0$ , où  $\lambda_i$  prend sa valeur minimale  $\lambda_{\min}$ . De même, lorsque le RSB décroît, le degré de suppression du bruit augmente jusqu'à le maximum ( $\lambda_i = 5$ ).

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  $RSB_0 = 20\text{dB}$ . 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 un estimé initial du spectre du bruit en moyennant  $|X_i(n)|^2$  sur ces trames :

$$|N_i(n)|^2 = \frac{1}{N_{trames}} \sum_{k=0}^{N_{trames}-1} |X_k(n)|^2, \quad n = 0, 1, \dots, NDFT/2 \quad (5.14)$$

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

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

### 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 de bruit et de correction spectrale, proposé est illustré par la figure (5.8).

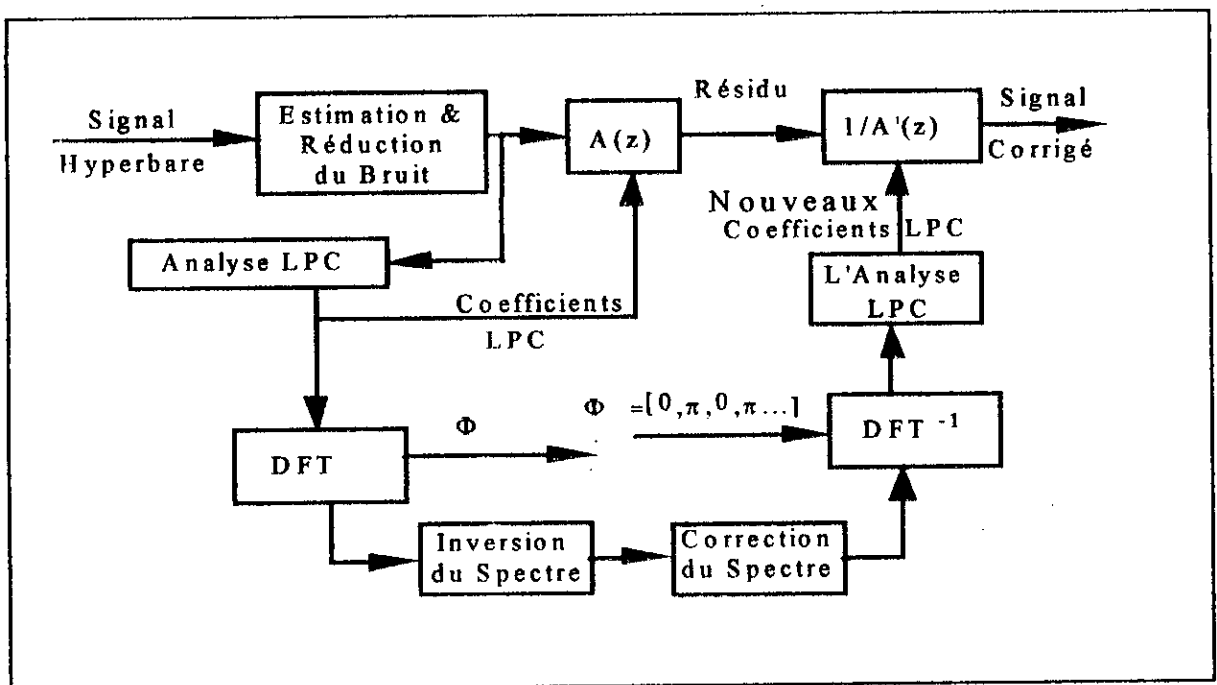


Fig. 5.3 Correction spectrale avec réduction de bruit

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

1. Trame  $i = 1$  ; Ntrames = 10;

2. Lecture de N échantillons de la trame  $i$  :  $X_i(n) : n = 0, 1, \dots, N$  ;

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 (Ntrames), GO TO 9 ;

### L'estimation du spectre du bruit

6. La mise à jour du spectre du bruit :

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

$$\text{où } |N_0(n)|^2 = 0 , \forall n$$

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

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

8. Aller à l'étape 12 :

### La suppression du bruit par soustraction spectrale

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

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

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

$$\bar{\lambda}_i = \lambda_0 - \left[ \frac{\lambda_0 - \lambda_{\min}}{RSB_0} \right] RSB_i$$

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

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

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

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

13. (Trame)  $i \leftarrow (trame) i + 1$  : aller à 2.

La figure 5.9 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.

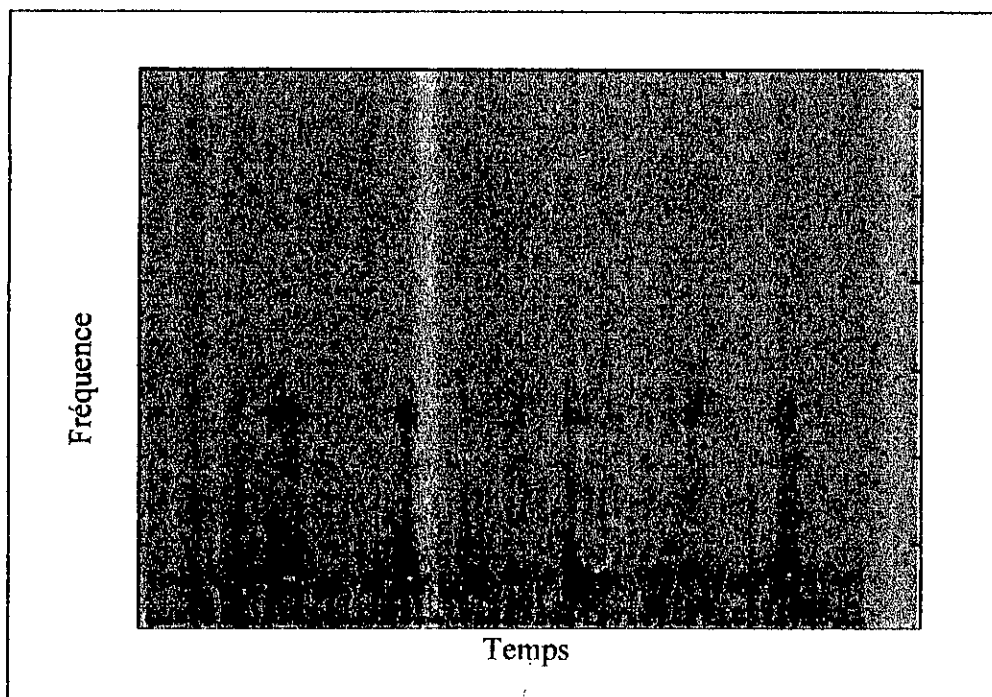


Fig. 5.9 Spectrogramme de SYN300P (Soustraction spectrale)

## **V.5 Conclusion**

C'est à travers les tests d'écoutes que le résultat final peut être 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 de bruit très efficaces à savoir la VAD(annexe C).

# Conclusion Générale

Le travail de recherche accompli a été consacré à 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  $a_i$  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  $F_s = 20$  KHz ) ; un ordre  $P$  de prédiction entre 16 et 20. Enfin, une préaccentuation est envisagée pour compenser la pente globale négative du spectre de la 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. La dernière variante, qui assure une compression non linéaire du spectre semble la plus adaptée aux variations non linéaires des formants.

Les systèmes de réduction de bruit incorporés aux systèmes de corrections améliorent encore la qualité de cette parole. Néanmoins, cette qualité peut être améliorée d'avantage par une parfaite connaissance des statistiques du bruit dans le milieu très hostile où évolue le signal hyperbare, et une meilleure adaptation des algorithmes de réduction de bruit à notre cas. Ainsi, de nombreuses simulations numériques ont permis de confirmer les études théoriques.

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 correspondants, les algorithmes proposés peuvent montrer quelques insuffisances qui peuvent être comblées par une meilleure adaptation.
- ❖ Des systèmes de réduction de bruit, avec plusieurs microphones, utilisant un filtrage adaptatif (filtre de Wiener) ou d'autres procédés.
- ❖ Les modifications causées par le milieu hyperbare réverbérant dû à l'enceinte mécanique utilisée par les plongeurs, en plus des différents types de bruits qui se rajoutent au signal hyperbare original, sont captés par le microphone. Des solutions à la prise de son peuvent être proposées afin de combattre le bruit à sa source, par un choix judicieux du microphone à large bande et de toute la chaîne de communication sous-marine, cela permet d'assurer la robustesse et d'améliorer les performances du système complet.
- ❖ 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 statistiquement 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)

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

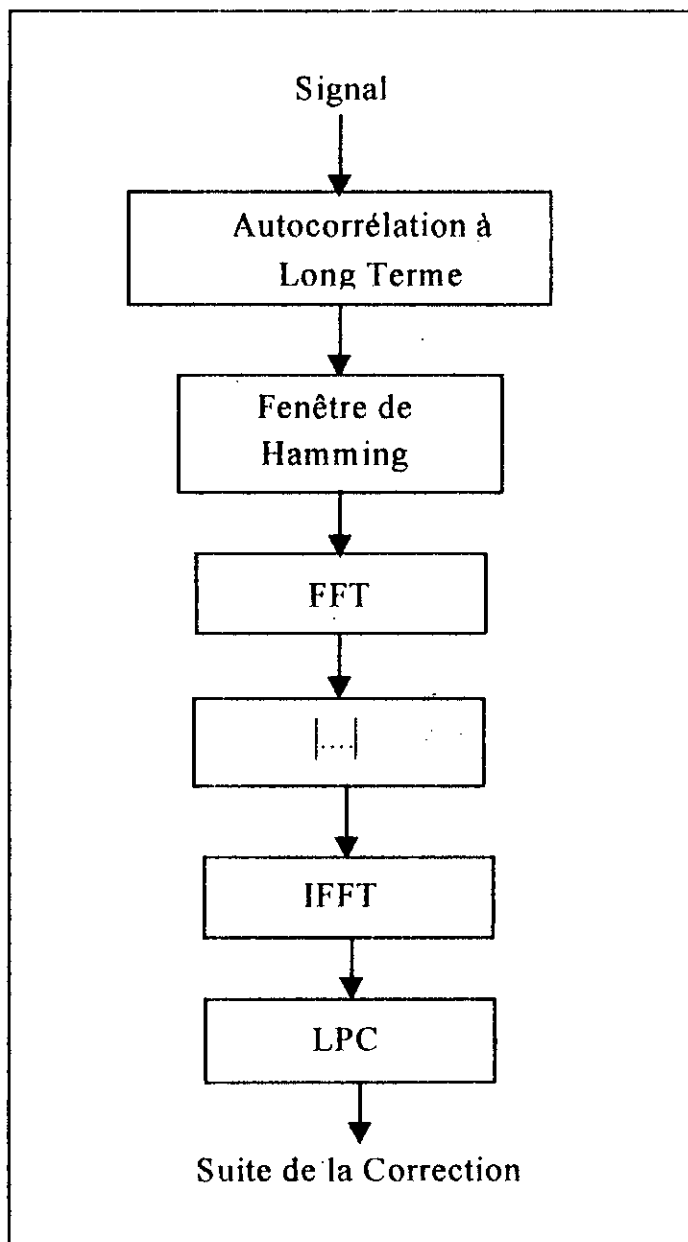
- La résolution du système (2.18) par l'algorithme de WLD,
- On obtient donc les coefficients de prédiction  $a_k$  ( $k = 1, 2, \dots, p$ ),

## 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.

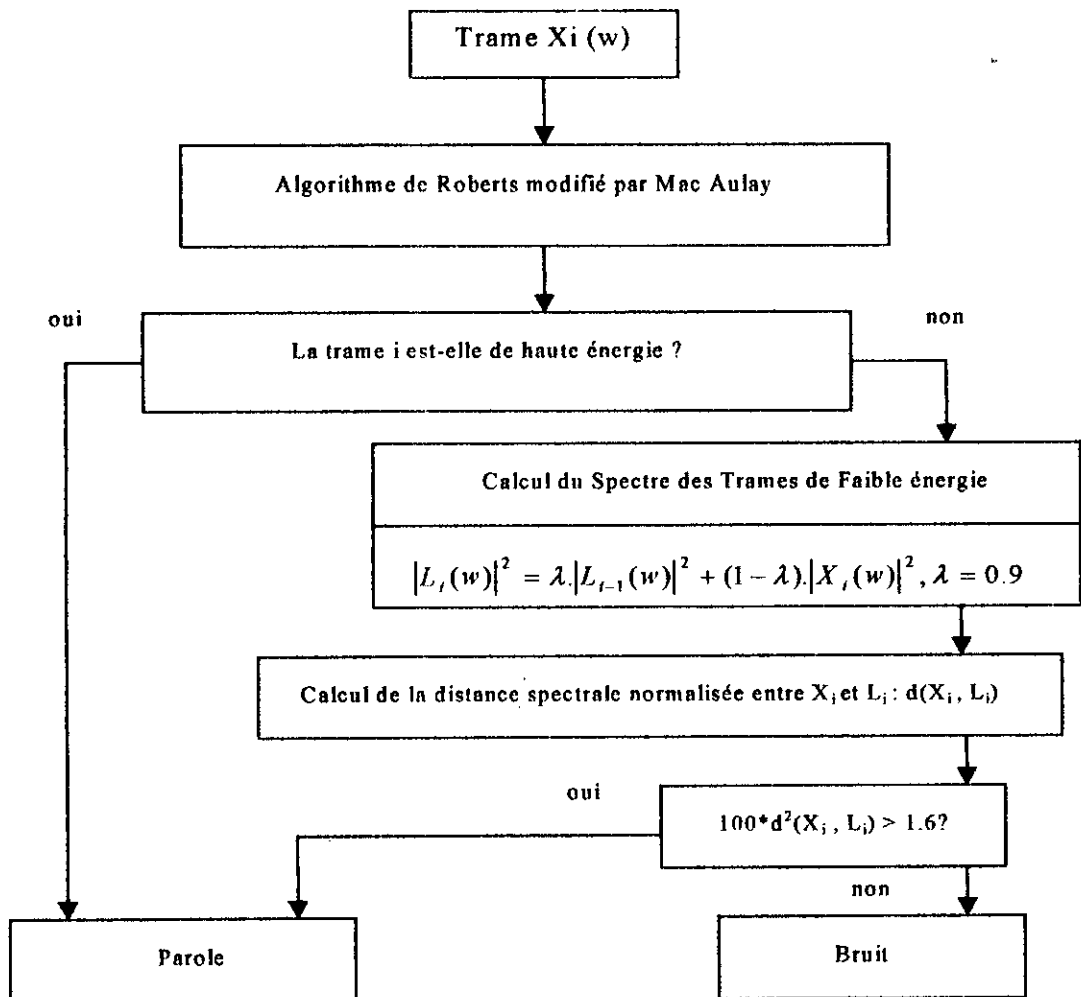


Organigramme de la Représentation SMC

## Annexe C

### La Détection d'Activité Vocale (VAD)

La détection d'activité vocale (VAD) se fait en deux étapes. Premièrement l'algorithme de Rogers modifié par Mac Aulay[30] permet la séparation des trames de haute énergie (parole voisée), des autres trames de faible énergie (Parole non voisée ou bruit), par la détermination d'un seuil d'énergie à partir de la densité de probabilité de l'énergie des trames. Après cette première phase de séparation, les trames classées dans la catégorie "parole non voisée ou bruit" peuvent être classées soit comme des trames de parole non voisée ou des trames de bruit seul. La deuxième étape consiste à assurer cette séparation par le calcul d'un spectre moyen des spectres de ces trames de faible énergie. Enfin, si la distance spectrale normalisée entre le spectre instantané de la trame courante, et le spectre moyen est supérieur à un seuil, la trame en cours est une trame de parole non voisée, sinon c'est une trame de bruit seul.



# Références Bibliographiques

- [1] J.L.Flanagan, "Speech Analysis, Synthesis, and Perception," Second Edition. New York Springer-Verlag, 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.63, 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," Ph.D. Thesis, School of Electrical Engineering, Georgia Institute of Technology, March, 1982.
- [8] E. O. Belcher,S. Hatstad, "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. Nakatsui, T. Takasugi, J. Susuki, "Source characteristics of speech produced under high ambient pressures," J. Radio Res. Lab. 21, May 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. R. R.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, N 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. On ASSP, vol. ASSP-29, N3, 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, NØ 3, pp. 113-120. 1979
- [18] E.J. Cupples and J.L. Foelker, "Air Force speech enhancement program," Military speech tech., Rome Air Development Center, Washington DC. 1987.
- [19] M.Berouti, R.Schwartz, and J.Makhoul, "Enhancement of speech corrupted by acoustic noise," Proc. IEEE ICASSP, pp. 208-211, Washinton, 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 hight 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. Wong. "The Burg algorithm for LPC speech analysis/synthesis," IEEE Trans. On ASSP, vol.28, no6, PP.609-615, Dec 1980.
- [25] G.Fant and B.Sonneson, "Speech at High Ambient Air Pressure," Speech Transmission Laboratory, Royal Inst. Of Tech., Stockholm, Sweden, Quart. Prog. And Status Rep. STL-QPSR-2/1964, pp. 9-21, 1964.

- [26] R.Tanaka, M.Nakatsui, T.Takasugi, J.Susuki, "Source Characteristics of Speech Produced under High Ambient Pressures," J.Radio Res.Lab.21, May 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 Juang, "The Short-Time Modified Coherence Representation and Noisy Speech Recognition," IEEE Trans., Speech, signal Processing, vol.37, NO. 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 Filter," IEEE Trans., Speech, signal Processing, vol.ASSP-28, NO.2. APRIL 1980.
- [31] F.K.Soong and M.M.Sondhi, "A Frequency-Weighted Itakura Spectral Distorsion Measure and Its Application to Speech Recognition in Noise," IEEE Trans., Speech, signal Processing, vol.36, NO.1. JANUARY 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," Acceptée pour le prochain COMAIEI'98, Bejaia, DECEMBRE 1998.