

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université Akli Mohand Oulhadj - Bouira -
Tasdawit Akli Muḥend Ulḥağ - Tubirett -



وزارة التعليم العالي والبحث العلمي
جامعة أكلي محمد أولحاج
- البويرة -

Faculté des Sciences et des Sciences Appliquées

كلية العلوم والعلوم التطبيقية

Références :...../MM/2022

المرجع :...../م/2022

Mémoire de Master

Présenté au

Département : Génie Electrique

Domaine : Sciences et Technologies

Filière : Télécommunications

Spécialité : Systèmes des Télécommunication

Réalisé par :

CHAIBI Hamid

Thème

Codage de la parole a large Bande

Soutenu le : 08/10/2023

Devant la commission composée de :

Mr : SAIDI Mohammed

M.A.A Univ. Bouira promoteur

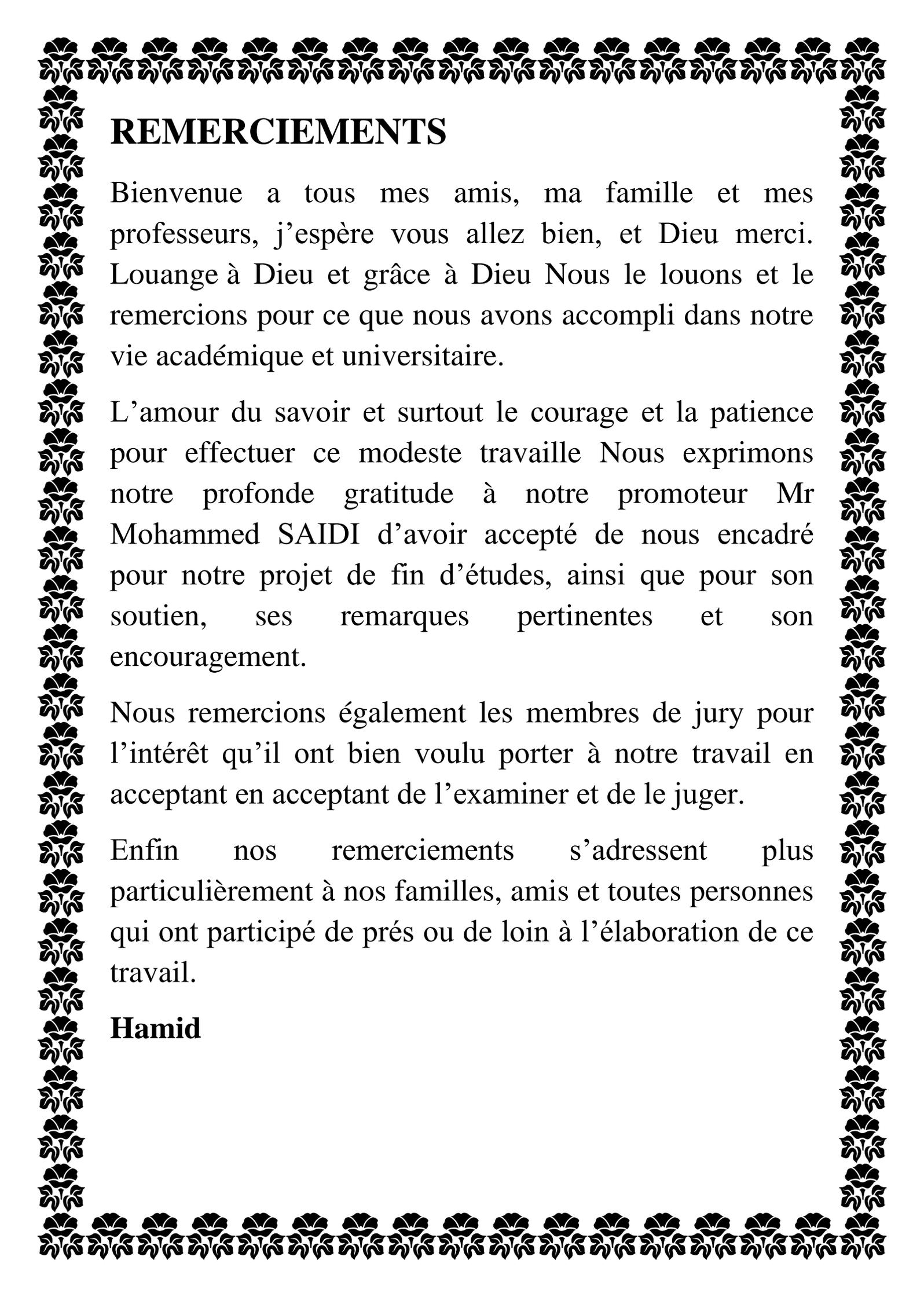
Mr : BOUZIDA Ahcene

M.C.A Univ. Bouira Président

Mr : MEDJEOUB Smail

M.A.A Univ. Bouira Examineur

Année Universitaire 2022/2023

A decorative border of repeating floral motifs surrounds the text. The motifs are stylized, resembling small flowers or leaves, arranged in a continuous line along the top, bottom, and sides of the page.

REMERCIEMENTS

Bienvenue a tous mes amis, ma famille et mes professeurs, j'espère vous allez bien, et Dieu merci. Louange à Dieu et grâce à Dieu Nous le louons et le remercions pour ce que nous avons accompli dans notre vie académique et universitaire.

L'amour du savoir et surtout le courage et la patience pour effectuer ce modeste travail Nous exprimons notre profonde gratitude à notre promoteur Mr Mohammed SAIDI d'avoir accepté de nous encadré pour notre projet de fin d'études, ainsi que pour son soutien, ses remarques pertinentes et son encouragement.

Nous remercions également les membres de jury pour l'intérêt qu'il ont bien voulu porter à notre travail en acceptant en acceptant de l'examiner et de le juger.

Enfin nos remerciements s'adressent plus particulièrement à nos familles, amis et toutes personnes qui ont participé de près ou de loin à l'élaboration de ce travail.

Hamid

Dédicaces

A mes chers parents

*A mes frères et mes sœurs et
Mes neveux qui m'ont soutenu
dans mes études*

*A tous mes amis, en particulier
ceux qui ont étudié avec moi à
l'université en dernière année*

Merci à tous pour tout

Hamid

ملخص

في هذه الدراسة سوف نشرح و نتعمق في موضوع تشفير الكلام واسع النطاق. لهذا قمنا بدراسة جودة الصوت عن طريق برنامج تشفير تكيفي متعدد وواسع النطاق لكي ندرس جودة الصوت حيث لاحظنا وجود علاقة بين التدفق ومتوسط الجودة (PESQ) بحيث كلما زادت عدد التدفقات ارتفع متوسط تباين الجودة وهذا ما استنتجناه في الأخير.

كلمات مفتاحية : تشفير الكلام, برنامج تشفير تكيفي واسع النطاق, متوسط تباين الجودة, التدفق

Résumé

Dans cette étude, nous expliquerons et approfondirons le sujet du codage vocal à grande échelle, Pour cela, nous avons étudié la qualité sonore au moyen d'un programme de codage adaptatif multi-large bande afin d'étudier la qualité sonore, ou nous avons remarqué une relation entre le débit et la qualité moyenne (PESQ) de sorte que plus le nombre de flux est élevé, plus la variance de qualité moyenne, et c'est ce que nous avons finalement conclu.

Mots clefs

Codage de la parole, Programme Adaptatif Multi-débit Large bande, Écart de qualité moyen, le flux

Abstract

In this study, we will explain and deepen the subject of large-scale speech coding, For this, we studied the sound quality by means of a multi-wide band Adaptive coding program to study the sound quality, or we have noticed a relationship between bit rate and average quality (PESQ) such that the higher the number of streams, the greater the average quality variance, and this is what we finally concluded.

Key words

Speech coding, program Adaptive Multi-Rate wide band, Average quality gap, the flo

Sommaire

Sommaire

Liste des figures.....	7
Liste des figures.....	7
Liste des tableaux	8
Liste des tableaux	8
Abréviations	9
Introduction générale.....	11
CHAPITRE I : Codage de la parole	12
Introduction :	13
I.1. Le Signal vocal.....	13
I.1.1. Mécanisme de phonation.....	13
I.1.2. Modèle de production de la parole:.....	15
I.1.3 Prédiction Linéaire	16
I.2 Principe de la quantification.....	22
I.2.1 Quantification scalaire	22
I.2.2 Quantification vectorielle.....	23
I.3 Techniques de codage de la parole.....	23
I.3.1 Le codage de forme d'onde.....	24
I.3.2 le codage par synthèse.....	24
I.4 Qualité des codeurs	24
I.4.1 Mesure de distorsion subjective	24
I.4.2 Mesure de distorsion objective.....	25
I.4.2.1 Domaine temporel.....	25
Conclusion.....	27
CHAPITRE II : Codeur Adaptive Multi-Rate a wide-Bande.....	28
Introduction	29
II.1 Le codec AMR.....	29
II.1.2 Contexte de L'AMR/AMR-WB et des principes de conception	29
II.1.3 le codec vocal adaptive multi-Rate Wide Bande	29
II.2 Pondération perceptive	31
II.2.1 Analyse de Pitch	33
II.2.2 Structure et recherche du livre de codes fixe algébrique	34
II.3 Description Algorithmique de décodeur (AMR-WB)	34

Sommaire

II.3.1 Décodage post-traitement d'excitation et synthèse vocale	34
II.3.2 échantillonnage et génération de hautes fréquences	35
Chiffre de complexité	36
Codeur	36
II.4 Complexité des codecs	36
II.5 Fonctionnement adaptatif sur le canal GSM.....	37
II.6 Détection d'activité vocale et discontinu transmission.....	37
II.7 Qualité de la parole.....	37
II.7.1. Test formel du codec AMR-WB.....	38
II.7.2 Qualité de base.....	39
Conclusion :.....	40
CHAPITRE III : Résultats et Simulation.....	41
III.1.Les etapes des Simulation :	42
III.1.1. étape 1 :	42
III.1.2. étape 2 :	43
III.2.Description et Commentaires :	46
III.3.Remarque :	47
Conclusion général	48
Bibliographie.....	50

Liste des figures

Liste des figures

Figure 1. Appareil phonatoire	13
Figure 2. Un son voisé et son spectre	14
Figure 3. Un son non voisé et son spectre.	Error! Bookmark not defined.
Figure 4. Modèle simplifié de production de la parole	Error! Bookmark not defined.
Figure 5. Spectre LPC avec LSF superposé	22
Figure 6. Quantification Scalaire.	Error! Bookmark not defined.
Figure 7. Compression de la qualité de codage de parole [14]	23
Figure 8. bloc de diagramme a AMR/WB ACELP codeur	30
Figure 9. bloc de diagramme a AMR/WB ACELP décodeur	Error! Bookmark not defined.
Figure 10.Exemple de AMR-WB Mode à adaptation de GSM canal plein débit.	Error! Bookmark not defined.
Figure 11.Qualité de l'AMR-WB avec un discours clair de l'expérience 1a test de Sélection ITU-T réalisé en langage française	37
Figure 12.Qualité de l'AMR-WB en présence de bruit de fond. DE l'expérience 3a du test de Sélection ITU-T en anglais américain	Error! Bookmark not defined.
Figure 13. Performances AMR-WB dans le canal GSM à plein débit sous des erreurs de canal et avec bruit de fond de voiture de 15 dB. De l'expérience 6a de caractérisation 3GPP phase avec La langage anglaise.....	38
Figure 14 Performance AMR-WB avec musique. De l'expérience 3b du test de caractérisation ITU-T.	39
Figure 15.dossier de 10 fichier	41
Figure 16.écran de codeur AMR-WB.	42
Figure 17 écran de codeur AMR-WB et les 9 débits de transissions.	42
Figure 18.écran de décodeur d'AMR-WB	43
Figure 19. écran de standard PESQ.....	43
Figure 20. les variations des valeurs de la moyenne de PESQ (Qualité des signaux).	45

Liste des tableaux

Liste des tableaux

Table 1. Abréviations.	10
Table 2. affectation des bits du mode codec AMR-WB. .. Error! Bookmark not defined.	
Table 3. Positions d'implisison valides dans le sous-Châssis	33
Table 4. Mise en œuvre Complexité de AMR-WB et AMR-NB codeur vocal.	35
Table 5. représenté les valeurs de PESQ1 et PESQ2 avec fonction de débits	44

Abréviations

Abréviations

AR	Auto Régressif
ARMA	Auto Régressif à Moyenne Ajustée
LPC	l'analyse par prédiction linéaire
Fe	Fréquence d'échantillonnage
LSP	la représentation en paires de raies spectrales
LAR	Log area ratio
LSF	Line Spectrale Fréquences
QS	Quantification Scalaire
VQ	Quantification Vectorielle
codec	Codeur et décodeur
Cfd	Codage de forme d'onde
CP	Codage Paramétrique
Ch	Codage hybride
PCM	Pulse Codec Modulation
DPCM	Differential PCM
ADPCM	Adaptive Differential PCM
ADM	Adaptive Delta Modulation
DRT	Diagnostic Rhyme Test
DAM	Diagnostic Acceptability Measure

Abréviations

MOS	Mean Opinion Score
RSB	Rapport Signal sur Bruit
RSBseg	Rapport Signal sur Bruit segmenté
TFD	Transformée de Fourier Discret
PESQ	Perceptuel Evaluation of Speech quality
EMBSD	Amélioré Modifié Aboyer Spectrum Distorsion
AMR	Adaptive Multi-Rate
WB	Wide Bande
ACELP	Excitation par Séquences Codées à Structure Algébrique
VAD	la détection d'activité vocale
NB	Nero Bande
ITU	International Télécommunications Union
ITU-T	ITU-Télécommunication standradizationsector
MOS	Mean Opinion Score
ACBK	Adaptatif Code Book

Table 1. Abréviations.

Tableau I : table des Abréviations

Introduction générale

Ce projet de fin d'étude traite le codage de parole à large bande. La transmission en large bande correspond à l'élargissement de la bande passante utilisée pour la transmission du signal de parole. [1]

La bande passante couramment utilisé en téléphonie est de 200-3400 Hz, les nouvelles technologies permettent d'améliorer la qualité de transmission et le codage de la parole. Donc Le nouvelle gamme de bande passante 50-7000 Hz, résultant en la Qualité du signal vocal transmis. L'étude de la Qualité de la parole est une science psychoacoustique importante pour ses applications dans la structure phonémique, la médecine et d'autres domaines. [2]

Le développement des applications multimédia sur l'Internet ainsi que les systèmes de conférence téléphonique feraient bon usage d'un système adaptatif permettant de régler le niveau de qualité du codage selon le débit disponible.

Nous avons organisé notre travail en trois chapitres :

Le premier chapitre : étude le codage de la parole, la prédiction linéaire, le modèle de production de la parole humaine et sa distorsion.

Le deuxième chapitre : étude le codeur AMR-WB, Bloc de diagramme de codeur AMR-WB, Bloc de diagramme de décodeur AMR-WB.

Le troisième chapitre : étude la simulation et résultats de le testé de la qualité des signaux de la parole à le codeur AMR-WB.

CHAPITRE I : Codage de la parole

Introduction :

La parole est un ensemble de gestes articulatoires rendus audibles et visibles. Le traitement des paroles modernes est une composante fondamentale des sciences de l'ingénieur, ce traitement est divisé à trois disciplines (La reconnaissance, La synthèse, et Le codage de la parole).

Le codage de la parole comprend la physiologie de la phonation, les propriétés de base du signal parole, elle représente source-filtre des codeurs actuels à une modélisation du système phonatoire humaine. Ce chapitre regroupe les généralités fondamentales de la production du signal parole, les propriétés, et des Compréhension de les l'évolution des techniques de codage de la parole. [1,2]

I.1. Le Signal vocal

Les Critères de la parole est un résultats de l'action volontaire et coordonnée d'un certain nombre d'organes. Pour que le système nerveux soit celui qui contrôle cette action et reçoit les informations auditif et tactile permanent [3].

I.1.1. Mécanisme de phonation

Les principaux organes composant l'appareil phonatoire sont [4]: les poumons, la trachée artère, le pharynx, les cavités buccales et nasales qui sont schématisés par la Figure I.1.

L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, pour que l'air soit expulsé la trachée-artère, vers larynx qui module l'air par le biais des cordes vocales. Telle que les cordes vocales qui aident l'air passer librement pendant le processus de parole et de respirer.

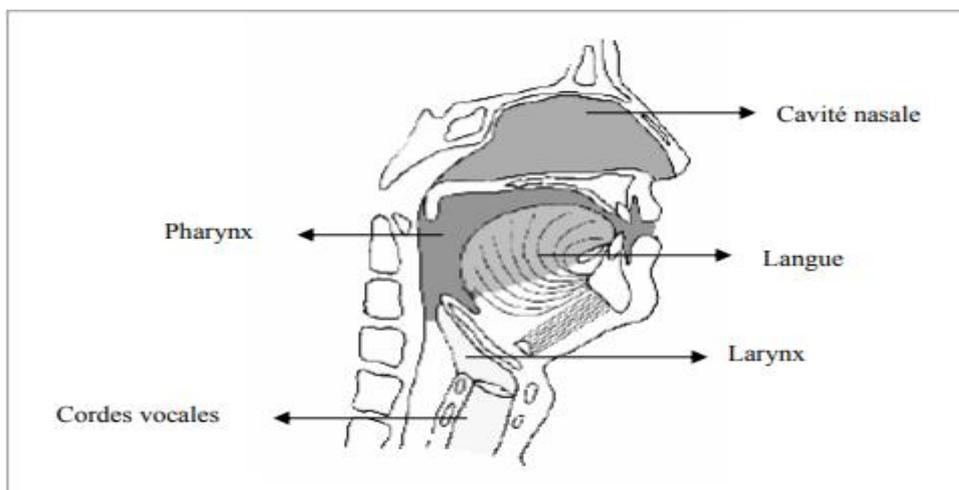


Fig. I.1 Appareil phonatoire

Le son voisé est une vibration périodique des cordes vocales résultant de la pression générée lorsque le larynx est fermé, et se vide progressivement lorsque les cordes vocales sont ouvertes. Ainsi, des impulsions périodiques de pression sont appliquées à la zone à deux

Cavités (cavité nasale et cavité buccales) pour la plupart des sons. De sorte que le son est bas au niveau du nez et monte au niveau de la bouche, qui a son tour est l'endroit qui définit l'expression dans la parole. L'intensité du son est liée à la pression de l'air émis au-dessus du larynx, on l'appelle donc la fréquence fondamentale ou pitch. Telle que la fréquence fondamentale est varier : [3,5]

- De 80 à 200 Hz pour une voix masculine.
- De 150 à 450 Hz pour une voix féminine.
- De 200 à 600 Hz pour une voix d'enfant.

Un son voisé est un signal quasi périodique dont le spectre est tracé à la Figure I.2. On y observe les raies qui correspondent aux harmoniques du fondamentale F_0 (pitch). Si bien qu'au niveau de ces raies, il ya des pics appelés formules qui déterminent les fréquences propres de sorte que les trios premières formules sont nécessaires pour la spectre audio. Telle qui le son non voisé ne présente pas de structure périodique. Telle qui en données comme un bruit blanc filtré par la transmutante de la partie du conduit vocal montre sur la Figure I.2.

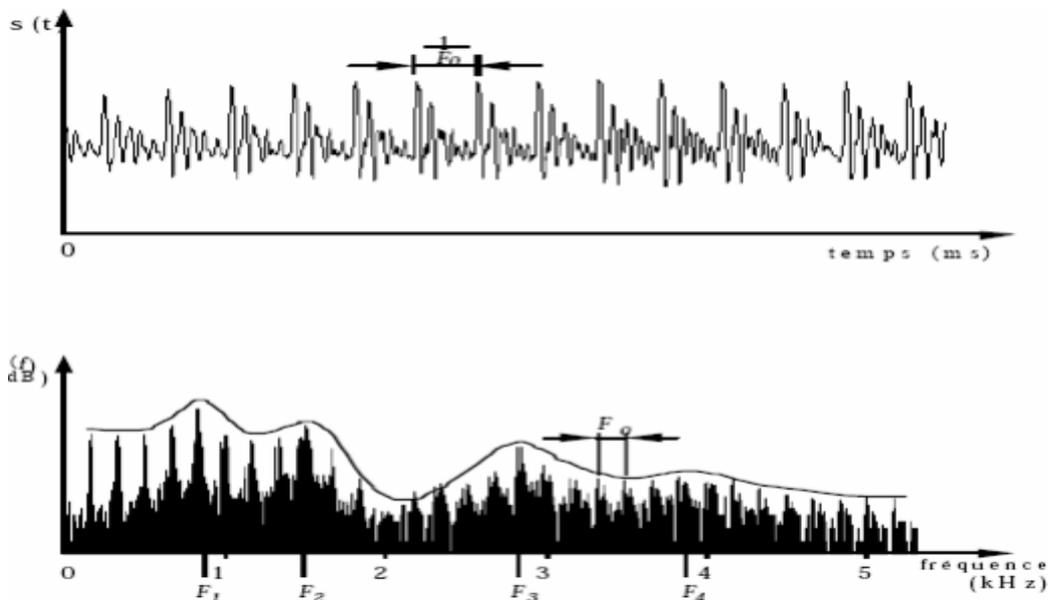


Fig. I.2 Un signal vocal voisé et son spectre [3][5]

Le son non voisé n'a pas de structure périodique donc c'est un bruit blanc qui est filtré par la perméabilité du segment du conduit vocal représenté sur la figure I.3. Ainsi, la taxonomie qu'il révèle est nécessairement un petit résumé portant avant tout sur la production de la parole ordinaire.

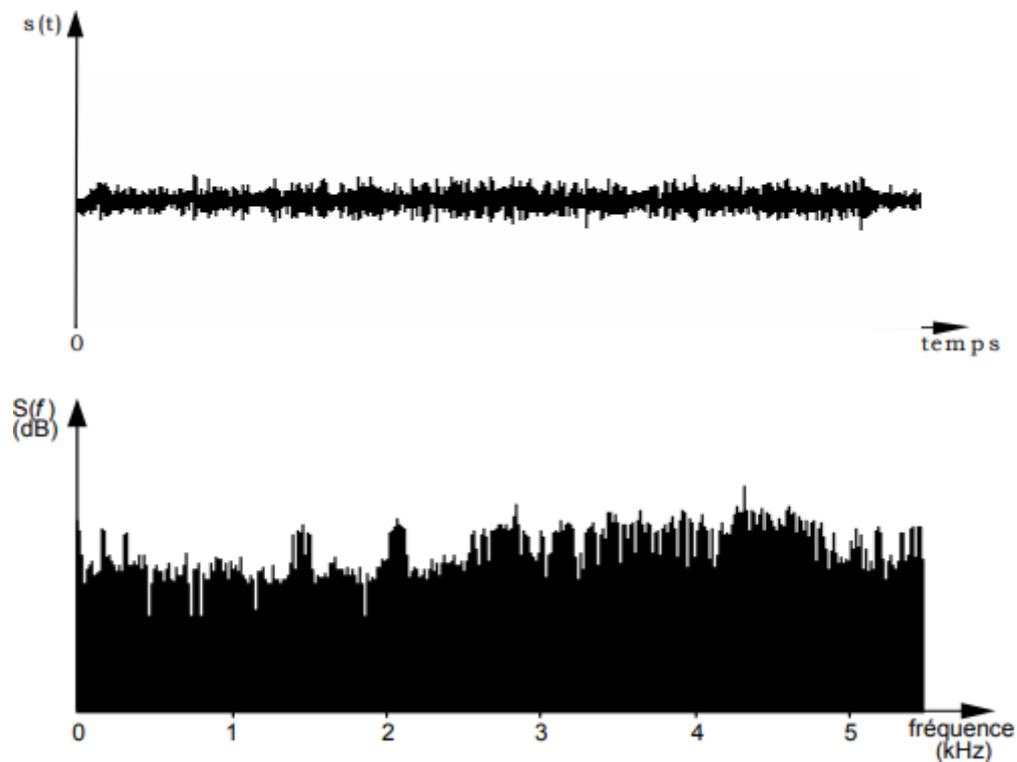


Fig. I.3 Un signal vocal non voisé et son spectre [3][5]

I.1.2. Modèle de production de la parole:

L'analyse de la parole est une étape essentielle dans toute application de synthèse, soit le codage et la reconnaissance. Le modèle électrique linéaire a été proposé par Fant[5] en 1960, qui identifie le signal audio formé en faisant passer des impulsions $u(n)$ multipolaires répétées (Auto Régressif), et l'équation de ce modèle s'écrit sous la forme :

$$X(n) = G \cdot u(n) + \sum_{i=1}^p -a_i X(n-1) \quad (\text{I.1})$$

Où $u(n)$ est le signal d'excitation et p l'ordre du système et Les coefficients du filtre $\{a_i\}$ sont appelés coefficients de prédiction et le modèle AR est souvent appelé modèle de prédiction linéaire. Et les paramètres du modèle AR sont : la période du train d'impulsions (sons voisés uniquement, la décision Voisé/Non Voisé (V/NV), le gain G et les coefficients du filtre $1/A(z)$, appelé filtre de synthèse.

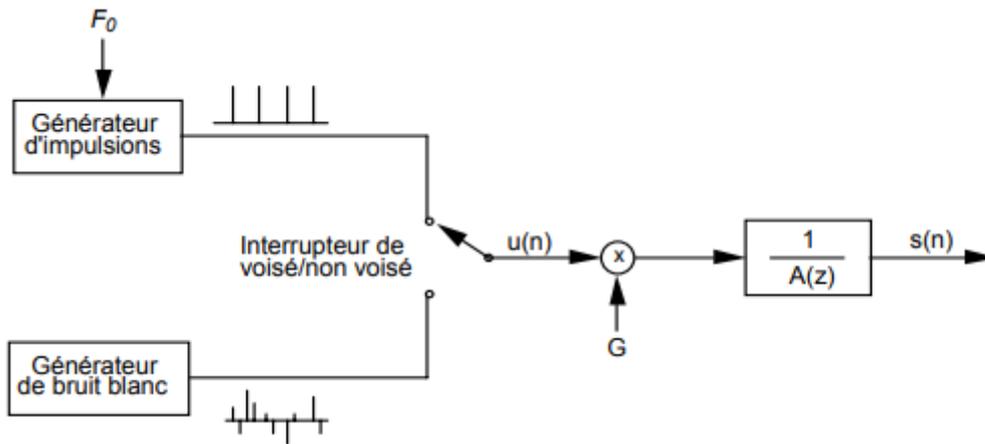


Fig. I.4 Modèle simplifié de production de la parole [3][5]

Les relations d'équivalences entre le modèle physique et le modèle mathématique [5] peuvent être données comme suit :

Conduit vocal	↔	H(z), le filtre LPC
Le flux d'air	↔	Le signal d'excitation u(n)
Vibrations de cordes vocales	↔	Voisé
Fricatives et plosives	↔	Non Voisé/Voisé
Volume d'air	↔	G, Le gain

I.1.3 Prédiction Linéaire

Le principe fondamental de la prédiction linéaire est qu'un échantillon du signal $S(n)$ peut être modélisé comme la sortie d'un système Auto Régressif à Moyenne Ajustée (ARMA) avec une entrée $u(n)$ [6] et [7] dans l'équation suivante :

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \sum_{l=0}^q b_l u(n-l), \quad b_0 = 1, \quad (I.2)$$

Où le gain G , les coefficients $\{a_k\}$ et $\{b_l\}$ sont les paramètres du système, et p et q sont les ordres des polynômes. L'équation (I.3) prédit la sortie courante en utilisant une combinaison linéaire des sorties précédentes et les entrées courantes et précédentes. Dans le domaine fréquentiel, la fonction de transfert du modèle de prédiction linéaire de la parole est de la forme :

$$H(z) = \frac{B(z)}{A(z)} = \frac{G[1 + \sum_{l=1}^q b_l z^{-l}]}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (\text{I.3})$$

$H(z)$ est le modèle pôle-zéro dans lequel les racines du dénominateur et de numérateur sont, respectivement, les pôles et les zéros du système.

Si $a_k = 0$ pour $1 \leq k \leq p$, $H(z)$ devient un modèle tous-zéros ou modèle à moyenne ajustée (MA). Si pour $b_i = 0$, pour $1 \leq i \leq q$, $H(z)$ devient un modèle tous-pôles ou modèle auto régressif (AR), exprimé par :

$$H(z) = \frac{1}{A(z)} \quad (\text{I.4})$$

Dans l'analyse spectrale, les pôles correspondent aux pics du spectre vocal, tandis que les zéros correspondent à la résonance inverse dans l'analyse de la parole, les classes de phonèmes comme les fricatives et les nasales contiennent des vallées spectrales qui correspondent aux zéros dans $H(z)$. Ainsi, le signal prédit est égal à :

$$s(n) = \sum_{k=1}^p a_k s(n-k) \quad (\text{I.5})$$

La différence entre l'échantillon original $s(n)$ et l'échantillon prédit $\hat{s}(n)$ est appelée erreur de prédiction (ou résidu) et elle est définie par:

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (\text{I.6})$$

Par conséquent, Le problème de la prédiction linéaire est réduit en trouvant un ensemble de coefficients a_k qui réduisent l'incidence de l'erreur n dans un certain période de temps. Il existe de nombreuses méthodes d'estimation des coefficients a_k sont nombreuses [8]. Mais il existe deux méthodes principales pour l'analyse par prédiction linéaire LPC court-terme : La méthode d'autocorrélation et la méthode de covariance.

I.1.3.1 Méthode d'Autocorrélation

La méthode d'autocorrélation garantit la stabilité du filtre LP, Ainsi, Le signal est défini pour toutes les valeurs du temps. Donc, le zéro est en dehors de la plage de valeurs de N échantillons, où N est un entier; ceci est équivalent à multiplier le signal de parole $s(n)$ par une fenêtre $w(n)$ de longueur finie correspondant à N échantillons pour obtenir un segment du signal de parole fenêtré $S_w(n)$ [9].

$$\begin{cases} S_f(n) = W(n) \cdot S(n) & \text{pour } 0 \leq n \leq N-1 \\ S_f = 0 & \text{ailleurs} \end{cases} \quad (\text{I.7})$$

La fonction de pondération la plus courante est la fenêtre de Hamming :

$$\begin{cases} W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1} & \text{pour } 0 \leq n \leq N-1 \\ W(n) = 0 & \text{ailleurs} \end{cases} \quad (\text{I.8})$$

Chaque échantillon peut être prédit approximativement à partir de p échantillons précédents.

Ceci est valable pour toutes les valeurs du temps : $-\infty < n < +\infty$.

Cette méthode renvoie les coefficients $\{a_k\}_{k=1..p}$ qui minimiser l'erreur de prédiction E_p :

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} (s_f(n) - \sum_{k=1}^p a_k s_f(n-k))^2 \quad (\text{I.9})$$

La fonction d'autocorrélation est une fonction paire: $R(i) = R(-i)$. Pour trouver les coefficients du filtre LPC, l'énergie du résiduel de prédiction doit être minimisée sur l'intervalle fini : $0 \leq n \leq N-1$

$$\frac{\partial E}{\partial a_k} = 0 \quad 1 \leq k \leq p$$

Cette erreur peut être minimisée en annulant les dérivées partielles par rapport aux coefficients du filtre :

$$\sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n-k) = \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n) \quad 1 \leq i \leq p \quad (\text{I.10})$$

$$\sum_{k=1}^p a_k R(i, k) = R(0, i) \quad (\text{I.11})$$

On obtient p équation linéaire avec p coefficient inconnus :

$$R(i, j) = \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n-j) \quad (\text{I.12})$$

Alors, les équations linéaires peuvent être écrites sous la forme :

$$R(i, j) = R(|i-j|) \quad (\text{I.13})$$

$$\text{Où, } R(i) = \sum_{n=i}^{N-1} s_f(n) s_f(n-i)$$

Est la fonction d'autocorrélation du signal fenêtré $S_f(n)$ vérifiant : $R(i) = R(-i)$

La forme matricielle de l'ensemble des équations linéaires (I.14) est représenté par $R^*a=v$ et peut être réécrite comme suit :

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(2) & \dots & R(p-2) \\ R(2) & R(0) & \dots & \\ \cdot & \dots & & \\ \cdot & \dots & & \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \cdot \\ \cdot \\ \cdot \\ R(p) \end{bmatrix} \tag{I.14}$$

La matrice d'autocorrélation $p \times p$ obtenue est symétrique dont tous les éléments de la diagonale sont égaux, c'est une matrice de Toeplitz. Ce qui nous permet de trouver les coefficients de prédiction minimisant la moyenne quadratique de l'erreur de prédiction par l'algorithme de Levinson – Durbin (Annexa A).

I.1.4.2. Méthode de Covariance

Les méthodes d'autocorrélation et de covariance diffèrent dans l'emplacement de la fenêtre d'analyse. Dans cette méthode c'est le signal erreur qui est fenêtré au lieu du signal parole, de façon à ce que l'énergie à minimiser soit :

$$E = \sum_{n=-\infty}^{\infty} e_f^2(n) = \sum_{n=-\infty}^{\infty} e^2(n)w^2(n) \tag{I.15}$$

En annulant les dérivées partielles en utilisant l'équation (I.13) on obtient p équations linéaires :

$$\frac{\delta E}{\delta a_k} = 0$$

Où la fonction de covariance :

$$\sum_{k=1}^p \Phi(i,k) = \Phi(i,0) \quad 1 \leq i \leq p \tag{I.16}$$

On peut exprimer les p équations, sous la forme : $\Phi^*a=\Psi$

$$\Phi(i,k) = \sum_{n=-\infty}^{\infty} w(n)s(n-1)s(n-k) \tag{I.17}$$

Tel que; $\Psi(i)=\Phi(i,0)$ pour $1 \leq i \leq p$

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & \dots & \Phi(1,p) \\ \Phi(2,1) & \Phi(2,2) & \dots & \Phi(2,p) \\ \cdot & \dots & \dots & \cdot \\ \cdot & \dots & \dots & \cdot \\ \Phi(p,1) & \Phi(p,2) & \dots & \Phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = \begin{bmatrix} \Psi(1) \\ \Psi(2) \\ \cdot \\ \cdot \\ \Psi(p) \end{bmatrix} \quad (\text{I.18})$$

La matrice Φ n'est pas une matrice Toeplitz, et ne garantit pas la stabilité du filtre LPC, elle est symétrique et définie positive. Donc, la matrice de covariance peut être décomposée en deux matrices, l'une triangulaire inférieure L et l'autre triangulaire supérieure

$$U. \Phi = L.U \quad (\text{I.19})$$

Le vecteur a est obtenu en résolvant d'abord l'équation La matrice Φ n'est pas une matrice Toeplitz, et ne garantit pas la stabilité du filtre LPC, elle est symétrique et définie positive. Donc, la matrice de covariance peut être décomposée en deux matrices, l'une triangulaire inférieure L et l'autre triangulaire supérieure

$$U. \Phi = L.U \quad (\text{I.20})$$

La décomposition de Cholesky peut être utilisée pour convertir la matrice de covariance sous la forme :

$\Phi = C.C^T$ telle que ; $C = L$ et $C^T = U$, Le vecteur a est obtenu en résolvant d'abord l'équation (I.22) :

$$L.y = \Psi \quad (\text{I.21})$$

$$\text{Puis : } U.a = y \quad (\text{I.22})$$

I.1.3.3 Considération Pratiques

Lorsque d'une analyse correcte, les quatre fréquences suivantes sont choisies :

- 1) La fréquence d'échantillonnage F_e .
- 2) La méthode d'analyse et l'algorithme correspondant.
- 3) L'ordre p de l'analyse LPC.
- 4) Le nombre d'échantillons par tranche N et le décalage entre tranches successives L.

Le choix de la fréquence d'échantillonnage dépend de l'application visée et de la qualité visée et de la qualité du signal à analyser :

- 8 kHz pour les signaux téléphoniques.
- 10 kHz pour les applications de reconnaissance.
- 16 kHz pour les applications de synthèse.

Telle que L'ordre p est fonction de la fréquence d'échantillonnage de sorte que la prédiction p est choisie pour bien représenter le signal de la séquence vocale.

I.1.4 Représentation des paramètres de prédiction

Les coefficients de prédiction linéaire LP ne conviennent pas pour transmettre des paroles codées. Du fait de sa dynamique, il le rend non quantifiable, d'où une instabilité du filtre de synthèse. Pour résoudre ce problème, avant quantification, les coefficients LP sont convertis en un ensemble de paramètres ayant des propriétés adaptées à la quantification et à la restitution. Plusieurs représentations ont été développées, nous citerons les paires de raies spectrales (LSP) [10], les coefficients de réflexion, *log area ratio* (LAR) [11], Cependant la représentation la plus répandue et la plus prisée pour ses performances reste la représentation en paires de raies spectrales LSP.

I.1.4.1 Paires de raies spectrales

Connus aussi sous le nom de fréquences de raies spectrales. La représentation LSP a été introduite par Itakura [11]. Les LSP sont les solutions des deux équations suivantes :

$$\begin{cases} P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \\ Q(z) = A(z) - z^{-(p+1)} A(z^{-1}) \end{cases} \quad (\text{I.23})$$

Ce qui nous donne :

$$A(z) = \frac{1}{2} [P(z) + Q(z)] \quad (\text{I.24})$$

Soong et Juang [12] ont montrés que si $H(z)$ est stable, où $A(z)$ est à phase minimale, alors les zéros des polynômes $P(z)$ et $Q(z)$ sont appels les LSP. Ces polynômes ont les propriétés suivantes [3]:

- tous les zéros de $P(z)$ et $Q(z)$ se trouvent sur la cercle unité.
- les zéros de $P(z)$ et $Q(z)$ sont entrelacés les uns aux autres, les LSP sont dans un ordre croissant.

Il a été montré **Error! Reference source not found.** que le filtre LPC $A(z)$ est à phase minimum si et seulement si les LSP satisfont les deux propriétés citées plus haut, donc la stabilité du filtre de synthèse est facilement vérifiable. De plus, les caractéristiques suivantes ont été relevées :

- Comme illustré à la **Error! Reference source not found.**, il y a une relation évidente entre les LSP et le spectre du filtre LPC. Une concentration des LSP dans une certaine bande de fréquences correspond approximativement à une résonance dans cette bande.

- Sensibilité spectrale: Un changement d'une LSP cause seulement un changement dans la forme du filtre d'analyse dans une petite gamme de fréquence autour de cette LSP.

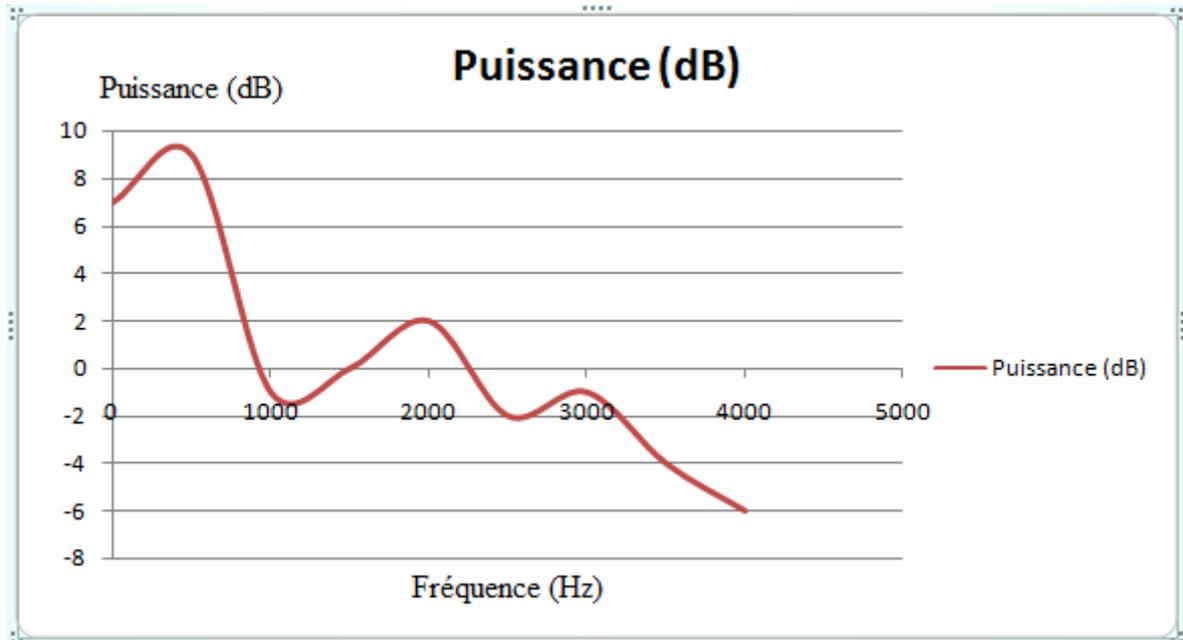


Fig. I.5 Spectre LPC avec LSF superposé

I.2 Principe de la quantification

La quantification est le processus de substitution des échantillons d'un signal analogique par des valeurs arrondies prises parmi un nombre fini de valeurs possibles [3]. La quantification peut être scalaire ou vectorielle selon que les signaux sont plusieurs dimensions. La quantification vectorielle peut être de deux types soit statistique ou algébrique.

I.2.1 Quantification scalaire

Dans la quantification scalaire (QS), chaque échantillon du signal d'entrée est quantifié séparément des autres échantillons. Comme l'illustre la figure I.6, un échantillon x du signal d'entrée est spécifié par l'indice k s'il se trouve dans l'intervalle suivant :

$$\{x_k \leq x \leq x_{k+1}\} \quad k = 1, 2, \dots, N \quad (\text{I.26})$$

Les valeurs x_k et x_{k+1} sont appelées niveaux de décision ou seuils.

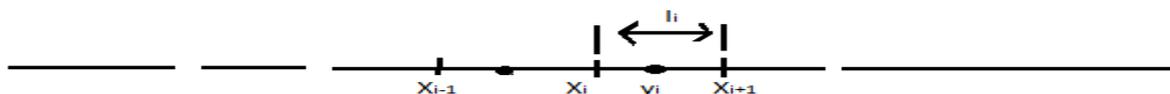


Fig. I.6 Quantification scalaire

Tous les échantillons situés dans l'intervalle I_i seront remplacés par une valeur y_i appelée niveau de reconstruction ou représentant.

I.2.2 Quantification vectorielle

La quantification vectorielle (VQ) est l'extension de la quantification scalaire à un espace multidimensionnel.

Nous appellerons quantifier vectoriel de dimension m à N niveaux une application Q qui, à un vecteur d'entrée $x = \{x_1, x_2, \dots, x_m\}$, fait correspondre une valeur approchée y choisie dans un ensemble fini de N éléments $y = \{y_i, i=0, 1, \dots, N-1\}$.

L'ensemble y est un dictionnaire de N représentants. En posant $R = \log_2(N)$, nous dirons que les vecteurs d'entés sont quantifiés sur N niveaux et codés avec R bits.

Contrairement à la quantification scalaire, un quantificateur vectoriel peut fonctionner avec un débit fractionnaire ($R < 1$) [3].

I.3 Techniques de codage de la parole

Un système de codage de la parole comprend deux parties: le codeur et le décodeur (codec). Le coder analyse le signal pour en extraire un nombre réduit de paramètres pertinents qui sont représentés par un nombre restreint de bits archivée ou transmission. Le décodeur utilise ces paramètres pour reconstruire un signal de parole synthétique.

Les algorithmes de codage de la parole peuvent être divisés en trois catégories [14]

- Codage de forme d'onde (waveform coding).
- Codage paramétrique (parametric coding).
- Codage hybride (hybrid coding)

La figure I.7 montre la différence de qualité de parole qui existe entre les codecs :

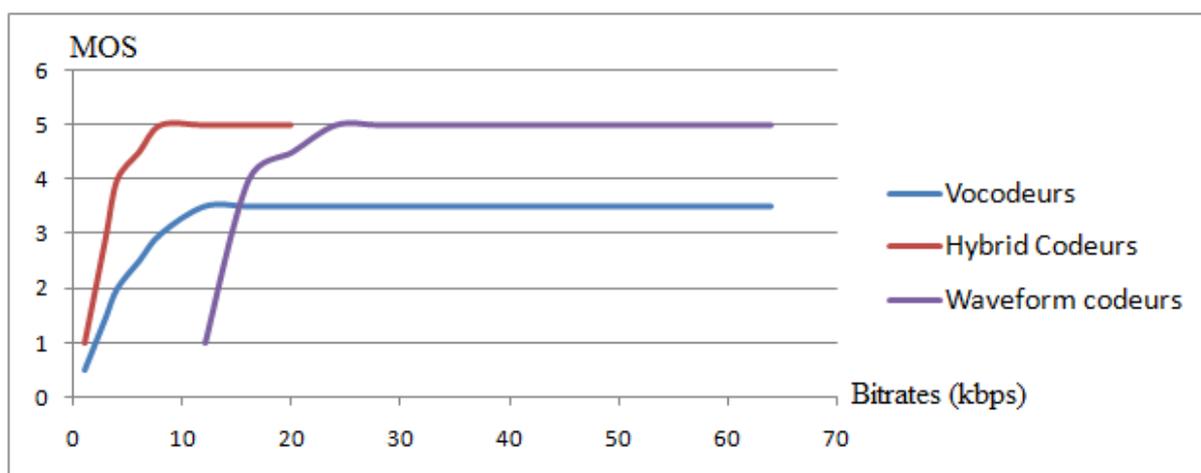


Fig. I.7 Comparaison de la qualité de codage de parole [14]

I.3.1 Le codage de forme d'onde

Les codages de forme d'onde sont relativement faciles à mettre en œuvre, produisant une qualité acceptable jusqu'à 16 Kbit/s. Ainsi, la qualité du signal reconstruit se détériore rapidement. L'algorithme de codage le plus simple est celui qui revient à échantillonner le signal analogique à quantifier les échantillons (convertir des valeurs réelles en valeurs précises finies). Ce codage est appelé PCM (Pulse Codec Modulation).

Le codage PCM est à la base d'une famille de codages différentiels qui est basé sur l'observation que des échantillons successifs d'une source audio sont fortement corrélés. On peut citer :

- Le codage DPCM (Differential PCM).
- Le codage ADPCM (Adaptive Differential PCM).
- Le codage ADM (Adaptive Delta Modulation).

I.3.2 le codage par synthèse

Il connu sous le nom de codeurs source, ils travaillent pour préserver l'intelligibilité de la parole à de faibles débits binaires et s'appuient sur un codage par prédiction linéaire dont les performances dépendent du modèle de sortie de la parole.

I.3.3 Le Codage Hybride

La qualité des codeurs d'onde reste faible à un débit inférieur à 16 Kbit/s, apportant une légère amélioration de la qualité audio à un débit supérieur à 4 Kbit/s, ce qui donne une bonne qualité de parole à des débits moyens. Cependant, ces codeurs nécessitent un plus grand nombre d'opérations, et tous les codeurs hybrides s'appuient sur l'analyse LPC pour obtenir les paramètres du modèle de réglage. Depuis les années 1980, l'intérêt pour les codeurs CELP. Ce codeur est basé sur l'algorithme de codage de la parole les plus utilisés dans les téléphones sans fil [6].

I.4 Qualité des codeurs

L'estimation de la qualité d'un codeur est un problème complexe. Une première approche consiste à utiliser une mesure objective de la ressemblance qui existe entre le signal original et le signal reconstitué. Lorsque l'on cherche une évaluation plus fine des codeurs, il faut faire appel à la dimension subjective de la qualité de la parole.

I.4.1 Mesure de distorsion subjective

L'évaluation subjective est obtenue par des tests d'écoutes ; dans ces tests, la qualité de la parole est mesurée par l'intelligibilité spécifiquement définie par le pourcentage de mots ou phonèmes correctement écoutés et avec une sonorité naturelle.

Il existe trois types de mesures subjectives [3] de la qualité généralement utilisées :

- Le test DRT (Diagnostic Rhyme Test)
- Le test DAM (Diagnostic Acceptability Measure)
- Le test MOS (Mean Opinion Score)

I.4.2 Mesure de distorsion objective

Le système auditif de l'être humaine est l'estimateur le plus adéquat de la qualité et des performances d'un codeur de la parole. Il permet de préciser l'intelligibilité et la sonorité naturelle des sons. Bien que, Les tests d'écoute subjectifs donnent une bonne évaluation pour les codeurs de la parole, ils peuvent exiger beaucoup de temps et sont non conformé. Les mesures objectives peuvent donner une estimation immédiate de la qualité perceptuelle de la parole [15]. Enduite Les mesures de distorsions sont classifiées comme suite [3] [16] :

- Domaine temporel (RSB et RSBseg)
- Domaine fréquentiel (distorsion spectrale)

I.4.2.1 Domaine temporel

- Rapport Signal sur Bruit :

Si $\{S(n)\}_{n=0.N_t}$ sont les N_t échantillons du signal parole original et $\{\tilde{S}(n)\}_{n=0.N_t}$ sont les N_t échantillons du signal parole codé dans le RSB a la forme suivante :

$$RSB = 10 \text{ Log}_{10} \frac{\sum_{n=0}^{N_T-1} S(n)^2}{\sum_{n=0}^{N_T-1} [S(n) - \tilde{S}(n)]^2} \quad (dB) \quad (I.25)$$

Le RSB donne une valeur après avoir traité tout le fichier, donc il n'y a pas moyen de retrouver les instants ou les divergences ont été enregistrées. De plus le RSB est dominé par la portion de forte énergie (tranches voisées).

- Rapport Signal sur Bruit segmenté :

Le RSBseg mesuré en dB, est la moyenne du RSB calculé sur de courts intervalles de temps du signal parole. Le RSBseg calculé sur N_F trames de longueur N_s est donné par :

$$RSB_{seg} = \frac{1}{N_F} \sum_{i=0}^{N_F-1} 10 \text{ Log}_{10} \frac{\sum_{j=0}^{N_s-1} S(N_s i + j)^2}{\sum_{j=0}^{N_s-1} [S(N_s i + j) - \tilde{S}(N_s i + j)]^2} \quad (dB) \quad (I.26)$$

Le RSBseg est meilleur que le RSB. Cependant, les tranches de silences renvoient de grandeurs négatives, biaisant de la sorte le résultat final. Ce problème peut être résolu en éliminant dans le calcul de la distorsion les trames de silence.

I.4.2.2 Domaine fréquentiel

La distorsion spectrale est définie comme étant la racine carrée de la moyenne au carrée des différences entre le logarithme décimale du spectre LPC original et le logarithme décimale du spectre LPC quantifier. La définition mathématique est comme suit :

$$DS_i = \sqrt{\frac{1}{F_e} \int_0^{F_e} \left[10 \text{Log}_{10} \frac{S_i(f)}{\tilde{S}_i(f)} \right]^2 df} \quad (dB) \quad (I.27)$$

Où F_e est la fréquence d'échantillonnage, $S_i(f)$ et $\tilde{S}_i(f)$ sont les spectres de trame i donnés par :

$$S_i(f) = \frac{1}{A_i(e^{j2\pi f / F_e})} \quad (I.28)$$

$$\tilde{S}_i(f) = \frac{1}{\tilde{A}_i(e^{j2\pi f / F_e})} \quad (I.29)$$

Ou, $A_i(z)$ et $\tilde{A}_i(z)$ sont respectivement, au lieu de l'intégration, une sommation des coefficients obtenus après application de la TFD (transformée de Fourier Discret) aux coefficients LPC, peut utilisée pour calculer DSI. La distorsion devient donc :

$$DS_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{k=n_0}^{n_1-1} \left[10 \text{Log}_{10} \frac{S_i(e^{j2\pi k / N})}{\tilde{S}_i(e^{j2\pi k / N})} \right]^2} \quad (dB) \quad (I.30)$$

Une distorsion spectrale moyenne (la moyenne des distorsions spectrales calculées pour toutes les trames) de 1 dB est habituellement acceptée. Cependant, selon Atal et pali wal les conditions de transparence spectrale (pas de distorsion audible) établies expérimentalement sont les suivantes :

- La moyenne DS inférieur à 1dB
- Le nombre de trames ayant DSI dans l'intervalle 2-4 dB est inférieur a 2%
- Pas de trames ayant DSI supérieur à 4 dB

I.4.3 Mesure de distance euclidienne LSP pondérée

Cette distance a été développée le but d'optimiser la quantification des paramètres LP, elle a la forme suivante :

$$d_{LSF} = \sum_{i=1}^p [c_i w_i (\omega_i - \tilde{\omega}_i)]^2 \quad (I.31)$$

Où c_i et w_i sont les poids du 1^{er} coefficient LSP ω_i , et p est l'ordre du filtre LP. pour un filtre d'ordre 10, les poids fixes c_i sont donnés par:

$$c_i = \begin{cases} 1.0, & \text{pour } 1 \leq i \leq 8, \\ 0.8, & \text{pour } i = 9, \\ 0.4, & \text{pour } i = 10. \end{cases} \quad (\text{I.32})$$

Ces poids sont utilisés pour donner plus d'importance aux basses fréquences par rapport aux hautes fréquences. Les poids adaptatifs w_i sont utilisés pour accentuer les régions de l'enveloppe spectrale $S(e^{j\omega})$ à forte énergie (formants). Ces poids sont données par :

$$w_i = [S(e^{j\omega})]^r \quad (\text{I.33})$$

Où r est une constante empirique qui contrôle le degré de la pondération, empiriquement $r=0.15$. Une pondération plus simple a été proposée par [17], elle à la forme suivante :

$$w_i = \frac{1}{\omega_i - \omega_{i-1}} + \frac{1}{\omega_{i+1} - \omega_i} \quad \text{Où } \omega_0 = 0 \text{ et } \omega_{p+1} = \pi \quad (\text{I.34})$$

Les mesures dans le domaine perceptuel sont basées sur les modèles d'audition humaine. Le signal est transformé vers un domaine adéquat telle que la manière qu'on puisse exploiter effets de masquage psycho-acoustique, et les mesures perceptuelles les plus utilisées : Perceptuel Evaluation of Speech quality (PESQ) et Amélioré Modifié Aboyer Spectrum Distorsion (EMBSD).

Conclusion

Le comportement du signal de la parole est déterminé en extrayant les coefficients de la prédiction linéaire de la fréquence du signal de parole, de sorte que sa simplicité et sa précision en font le plus largement utilisées dans le codage de la parole.

CHAPITRE II : Codeur Adaptive Multi-Rate a wide-Bande

Introduction :

Le codage à large bande a connu un bon développement ces dernières années car il est désormais applicable aux communications transmises dans les systèmes 3GPP et WCDMA et GSM. Après ce développement approfondi, l'algorithme Broadband Speech Codec a été créé en décembre 2000 et approuvé en 2001. Le Broadband Speech Codec est un programme adaptatif qui fonctionne avec de nombreux débits de fichiers. Et sont désignés par le symbole AMR-WB (Adaptive Multi-Rate Wide Bande) [18].

II.1 Le codec AMR

L'AMR est un codec vocal adaptatif multi-débit à bande étroite avec huit profils de débit binaire de 4,75 kbps à 12,2 kbps qui comprend un détecteur d'activité sonore, un générateur de bruit pratique et un mécanisme de masquage d'erreur pour améliorer la qualité de la parole sur les supports de transmission [19].

II.1.2 Contexte de L'AMR/AMR-WB et des principes de conception

AMR/AMR-WB a été conçu à l'origine pour la radio à communication de circuits de téléphonie mobile en raison de sa flexibilité et de sa robustesse qui le rendent adapté aux services de communication vocale tels que les réseaux à communication de paquets (internet).

II.1.3 le codec vocal adaptive multi-Rate Wide Bande

Le codec vocal adaptatif Multi-Débit Large-bande (AMR-WB) est un codage vocale multi-mode qui prend en charge 9 modes large bande, chacun débits binaires allant de 6,6 à 23,85 kbps. La fréquence d'échantillonnage utilisée dans AMR-WB est 16000 Hz et la parole est traitée des trames de 20 ms de sorte que toutes les 320 s représentent une trame codée AMR-WB.

II.1.4 Présentation de codec AMR-WB

Le codec AMR-WB est basé sur le codage algébrique (ACELP) [20]. La technologie ACELP est très efficace dans le codage de large bande passant de la téléphonie, où les signaux audio sont traités et diffusés vers une large gamme d'applications numériques et cellulaires et VOIP (P (e.g., 3GPP AMR (TS 26.090) [21], ETSI EFR (TS 06.60) [22], NA-TDMA IS-641, NA-CDMA IS-127 and ITU-T G.729 and G.723.1 codecs).

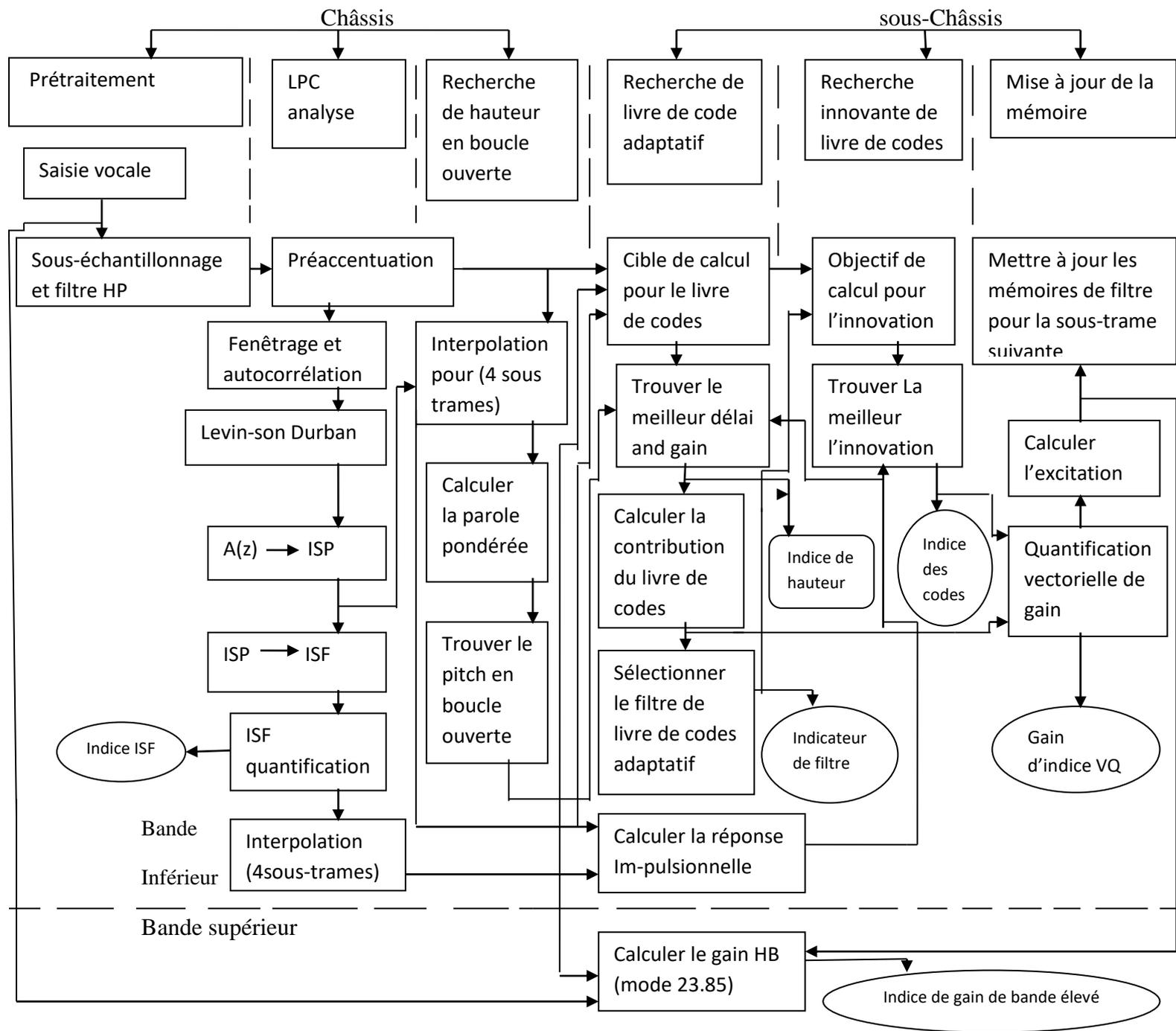


Figure 1 Bloc de diagramme à AMR/WB ACELP codeur

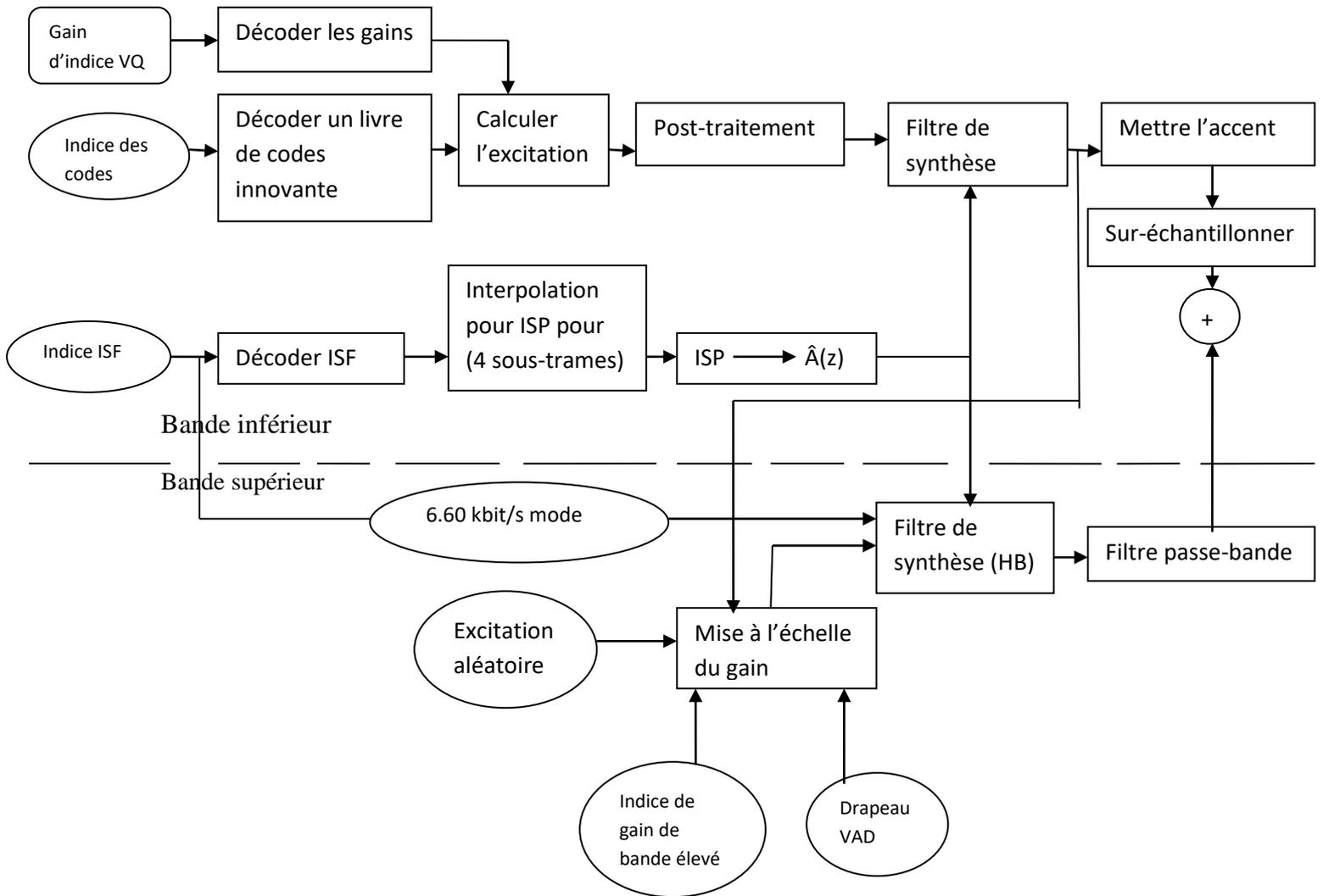


Figure 2 Block de diagramme à AMR/WB ACELP décodeur

II.2 Pondération perceptive

Dans les codeurs à décomposition par synthèse, les paramètres de hauteur sont recherchés et l'innovation optimale se fait en réduisant le carré moyen :

Paramètres	Mode codec [kb/s]								
	6.60	8.85	12.65	14.25	15.85	18.25	19.85	23.05	23.85
Drapeau vad	1	1	1	1	1	1	1	1	1
Indicateur de filtrage LTP	0	0	4	4	4	4	4	4	4
ISP	36	46	46	46	46	46	46	46	46
Retard de hauteur	23	26	30	30	30	30	30	30	30
Code algébrique	48	80	144	176	208	256	288	256	256
Gains	24	24	28	28	28	28	28	28	28
Energie haute bande	0	0	0	0	0	0	0	0	0
Total par image	132	177	253	285	317	365	397	461	477

Tableau 1 : affectation des bits du mode codec AMR-WB

La pondération perceptuelle est mise en œuvre grâce à l'utilisation d'un filtre de pondération et la formule de filtre de pondération :

$$W'(z) = \frac{A'(\frac{z}{y_1})}{A'(\frac{z}{y_2})} \quad 0 < y_2 < y_1 \leq 1 \quad (I.1)$$

OU $A'(z)$ est le filtre de prédication linéaire et y_1 et y_2 sont des facteurs qui contrôlent la quantité de pondération perceptuelle limitée par On peut montrer que le bruit de codage (qui est supposé avoir un spectre blanc) est pondéré par un transfert $1/W'(z)$ fonction, qui est l'inverse de la fonction de transfert du filtre de pondération perceptuelle.

Une nouvelle solution à ce problème consiste à introduire une préaccentuation du filtre de forme en entrée afin que le filtre LP soit basé sur le filtre modificateur de parole et en fixant son dénominateur, par exemple : de sorte que le filtre de pondération perceptuelle sont choisi par cette équation :

$$W = \frac{A(\frac{z}{y_1})}{(1 - y_2 z^{-1})} \quad 0 < y_2 < y_1 \leq 1 \quad (I.2)$$

II.2.1 Analyse de Pitch

Dans le codeur AMR-WB, la recherche de hauteur de compose de trois étapes. Pour le 1^{er} étape On à un retard de hauteur en boucle ouverte est estimé toutes les 10 ms sur la base du filtre passe-bas décimé pondéré signal de parole. Les détails de l'analyse de hauteur en boucle ouverte peuvent être trouvés dans [23]. Dans la deuxième étape une recherche de hauteur en boucle fermée est effectuée pour des retards de tonalité entiers autour du retard de tonalité estimé en boucle ouverte à une gamme de 7 échantillons, ce qui simplifie considérablement la procédure de recherche. Dans la troisième étape, de la recherche passe par la fraction autour de cette valeur entière optimale.

Le recherche de hauteur en Boucle fermée est effectuée en minimisant la erreur quadratique moyenne pondérée entre la parole d'origine et la parole Synthétise

$$C_k = \frac{\sum_{n=0}^{N-1} x(n)y_k(n)}{\sqrt{\sum_{n=0}^{N-1} y_k(n)y_k(n)}} \quad (I.3)$$

Ou N est la taille de la sous-trame, x(n) est le signal cible et y_k(n) est l'excitation passée au décalage k, filtrée à travers le filtre W(z)/A'(z) de synthèse pondéré (notez qu'il s'agit du filtre de pondération) et 1/Â(z) est le filtre de synthèse avec coefficients LP.

$$v(x) = \sum_{i=-1}^1 b_L(i+1)v'(n+i) \quad (I.4)$$

Dans le codec AMR-WB, un bit par sous-trame est alloué pour caractérisant le filtre passe-bas utilisé pour mettre en forme l'adaptatif excitation du livre de codes.

Ou y(n)= v(n) * h(n) est le livre de codes adaptatif filtré compromis pratique entre la complexité et l'atténuation haute fréquence souhaitée. Le gain du livre de codes adaptatif est alors trouvé par :

$$g_p = \frac{\sum_{n=0}^{N-1} x(n)y(n)}{\sum_{n=0}^{N-1} y(n)y(n)} \quad \text{avec} \quad 0 \leq g_p \leq 1.2 \quad (I.5)$$

Piste	Positions d'impulsion valides dans le sous-châssis
T0	0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60
T1	1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61
T2	2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62
T3	3, 7, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63

Tableau 02 : Positions d'impulsion valides dans le sous-châssis

Décodeur en cas d'erreurs de canal, le gain du livre codes adaptatif. Telle que γ est majoré 0,95, si le livre de codes adaptatif gagne les sous-trames précédentes ont été hautes et le filtre LP des sous-trames précédentes ont été proches d'être instables.

II.2.2 Structure et recherche du livre de codes fixe algébrique

Dans le codage de la parole à grande échelle, des livres de codes sont nécessaires pour garantir une qualité intrinsèque élevée. La structure de livre codes algébrique décrit des codes à une taille aussi élevée que 88 bits, Dans notre implémentation, la taille de sous-trames est de 64 échantillons correspondant à 5 ms à une fréquence d'échantillonnage de 12,8 KHZ. Telle que le livre de codes est un livre dynamique dont le code est forcé par un pré-filtre adaptatif qui améliore des composantes spectrales spécifiques pour améliorer la qualité de l'auto-conversation.

II.2.3 Quantification du Gain

Le gain de livre de code adaptatif et le gain de livre de code fixe sont quantifiés par faisceau à l'aide de symboles à 6 bits pour les modes 8.85 et 6.60 kb/s. La quantification du code fixe dans le livre de codes de prédiction de moyenne mobile utilise des coefficients constants du quatrième ordre qui sont appliqués à l'énergie inventive dans le domaine logarithmique. Le livre de codes est similaire à l'article : GSM EFR [20] ou G.729 [24].

II.3 Description Algorithmique de décodeur (AMR-WB)

La fonction de décodeur (AMR-WB) consiste à décoder les paramètres transmis (drapeau VAD), paramètres LP, indices de livre de codes, gains de livre de codes adaptatif, livre de codes fixe. Ensuite reconstruite la parole traitée extraite, le signal de bande haute est généré dans la gamme de fréquences de 6,4 à 7 KHz.

II.3.1 Décodage post-traitement d'excitation et synthèse vocale

Les indices reçus sont utilisés pour quantifier l'ISP afin de reconstruire le vecteur ISP quantifié, il y a quatre vecteurs ISP interpolés correspondant à 4 sous-trames qui sont calculés puis convertis en un champ de paramètre LP α_k , qui est utilisé pour la synthèse vocale de sous-trames afin que chaque trame affiche le fichier d'index de code adaptatif reçu pour trouver le numéro correct.

$$u(x) = \hat{g}_p * v(n) + \hat{g}_c * c(n) \quad (I.6)$$

La technique de lissage de gain non linéaire est appliquée au gain de code fixe pour améliorer l'excitation sur le bruit. Sur la base de la stabilité et de l'articulation, le gain est lissé pour réduire les signaux statiques afin d'améliorer la qualité de la parole par rapport au bruit de fond statique.

$\lambda = \frac{E_c}{(E_v + E_c)}$ Tel que E_c et E_v sont les énergies du livre codes adaptatif et excitation de livre de codes fixe et λ est le facteur de voisement est délimité dans [0,1].

II.3.2 échantillonnage et génération de hautes fréquences

Le signal de synthèse est filtré passe-haut par précaution contre les composants basse fréquences indésirables. Enfin, le signal est sur échantillonné pour obtenir la bande inférieure synthèse à une fréquence d'échantillonnage de 16 KHz.

L'excitation en bande haute est obtenue à partir d'un blanc signal de bruit comme :

$$u_H(n) = \frac{\hat{g}_H * w(n)}{\sqrt{\sum_{k=1}^{63} w^2}} \quad (I.7)$$

\hat{g}_H est le gain haute fréquence en mode 23.85 kb/s

$$\text{L'inclinaison } \tau \text{ du signal de synthèse se trouve comme } \tau = \frac{\sum_{n=0}^{63} \hat{S}(n)\hat{S}(n-1)}{\sum_{n=0}^{63} \hat{S}^2} \quad (I.8)$$

$$\text{On à } 0 \leq \tau \leq 1 - 2\lambda \text{ et l'équation de } \hat{g}_H \text{ est : } \hat{g}_H = w_s(1 - \tau) + 1.25(1 - w_s) \quad (I.9)$$

Chiffre de complexité	Codeur	
	AMR-WB	AMR-NB
Complexité de calcul [WMOPS]		
Codeur vocal	31.1	14.2
Décodeur vocal	7.8	2.6
total	38.9	16.8
Data RAM [k Words, 16-bit]	6.5	5.3
Data ROM [k Words, 16-bit]	9.9	14.6
Program ROM [num. of ESTI basi-cops]	3889	4851

Tableau :03 Mise en œuvre Complexité de AMR-WB et AMR-NB codeur vocal

Le filtre de synthèse LP bande haute est obtenu pour chaque sous-trame d'un filtre de synthèse LP à bande basse utilisant :

$$\hat{A}_H(z) = \hat{A} \left(\frac{12,8z}{16} \right) \quad (I.10)$$

$1/\hat{A}(z)$ est le taux d'échantillonnage été introduit à 12.8 KHz

II.4 Complexité des codecs

Le codec AMR-WB est défini en arithmétique à virgule fixe à l'aide d'un ensemble de facteurs fondamentaux spécifiés par 3GPP/ETSI. Mais la base pour exécuter ce programme est le langage C, l'algorithme du langage C permet le mappage en convertissant le code C en un langage d'assemblage DSP qui est également spécifié comme la complexité approximative de l'algorithme C. Le Complexité théorique du mauvaise cas codec, ou suppose que le chemin le plus complexe pour le codec vocal est AMR-WB, le tableau précédent montre également la complexité du codec vocal et son degré de complexité.

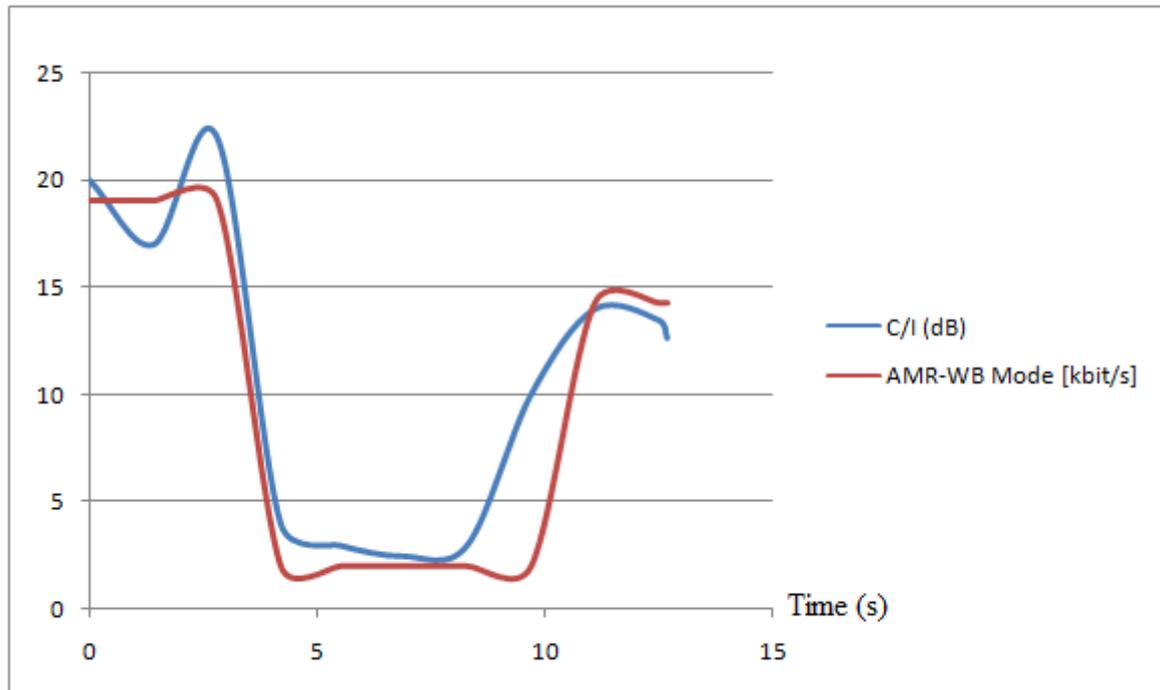


Figure 03 : Exemple de AMR-WB Mode à adaptation de GSM canal plein débit

II.5 Fonctionnement adaptatif sur le canal GSM

AMR-WB a une granularité élevée des débits binaires entre 6,6 et 23,85 Kb/s. Pour les canaux GSM, cela permet de maximiser la qualité de la parole en adaptant le débit binaire du codec pour augmenter robustesse face aux erreurs de transmission. Pour la 3G non adaptation les canaux UTRAN utilisant un contrôle de puissance rapide. [25,26]

II.6 Détection d'activité vocale et discontinu transmission

Le codec AMR-WB inclut la détection d'activité vocale (DAV) qui permet au codec de passer en mode d'encodage à faible bruit de fond. Cette fonction permet d'économiser de l'énergie sur une station en mouvement et réduit le niveau d'interférence global sur l'interface radio. Le DAV calcule la résolution (DAV) logique toutes les 20 ms dans une trame de parole il divise le signal de parole dans la gamme de fréquence [0, 6.4 KHz] en 12 sous-bandes et calcule le niveau du signal dans chaque bande. [27]

II.7 Qualité de la parole

AMR-WB fournit une qualité vocale supérieure sur tous les codes GSM et offre une qualité similaire à AMR-NB. L'AMR-WB fournit également une haute précision pour les débits binaires, ce qui le rend adapté à de nombreuses applications dans les systèmes 2G et 3G ainsi qu'aux applications vocales dans les services câblés pris en charge par l'UIT-U.

II.7.1. Test formel du codec AMR-WB

Lors de la standardisation du programme AMR-WB dans le 3GPP, il a été largement testé en plusieurs phases : la phase de sélection, la phase de validation et la phase de caractérisation il a été testé par six laboratoires d'écoute indépendants et cinq langages. Le test a couvert différents niveaux d'entrée tels que les performances de bruit de fond et les performances DTX/DAV. DE plus, il a été testé sur divers canaux de communication mobile en GSM et 3G WCDMA.

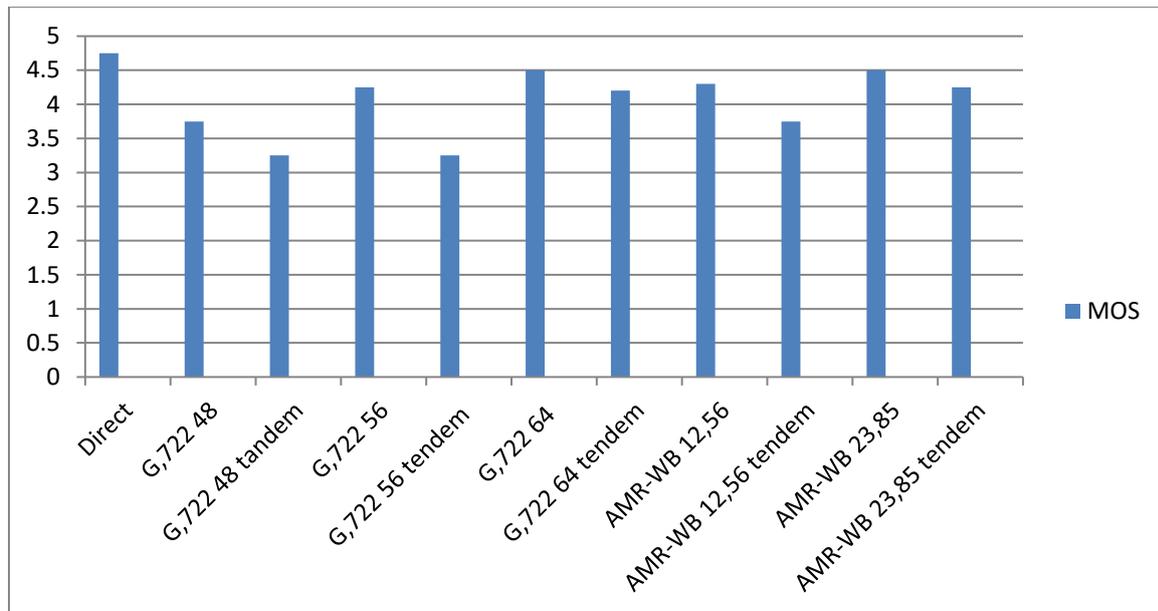


Figure : 05 Qualité de l'AMR-WB avec un discours clair. De l'expérience 1a du test de sélection ITU-T réalisée en langage française

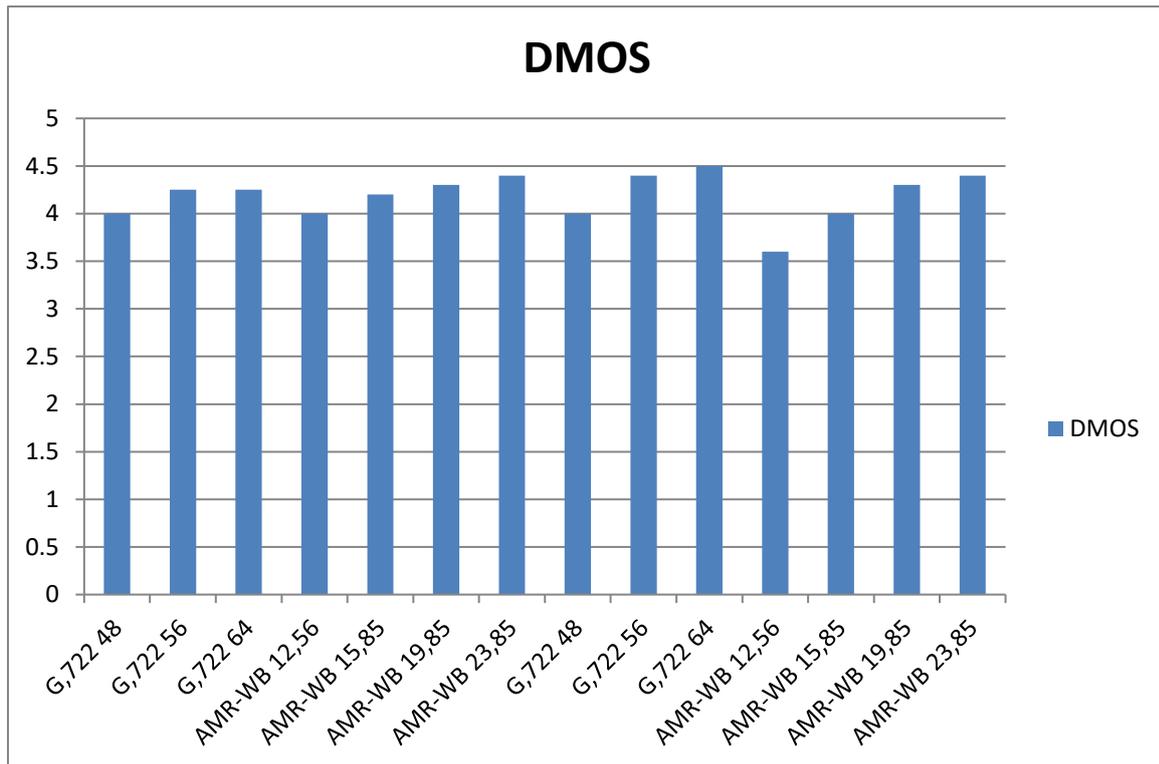


Figure : 06 Qualité de l'AMR-WB en présence de bruit de fond. De l'expérience 3a du test de sélection ITU-T en anglais américain.

II.7.2 Qualité de base

La qualité vocale propre fournie par les six profils les plus élevés (14.25, 15.85, 18.25, 19.85, 23.05, 23.85) AMR-WB est la meilleure qualité de codec à large bande ITU-T G.722 à 64 ko/s. et mettre 12,65 kb/s égal à au moins G.722 à 56 kb/s avec une qualité égale à 48 kb/s.

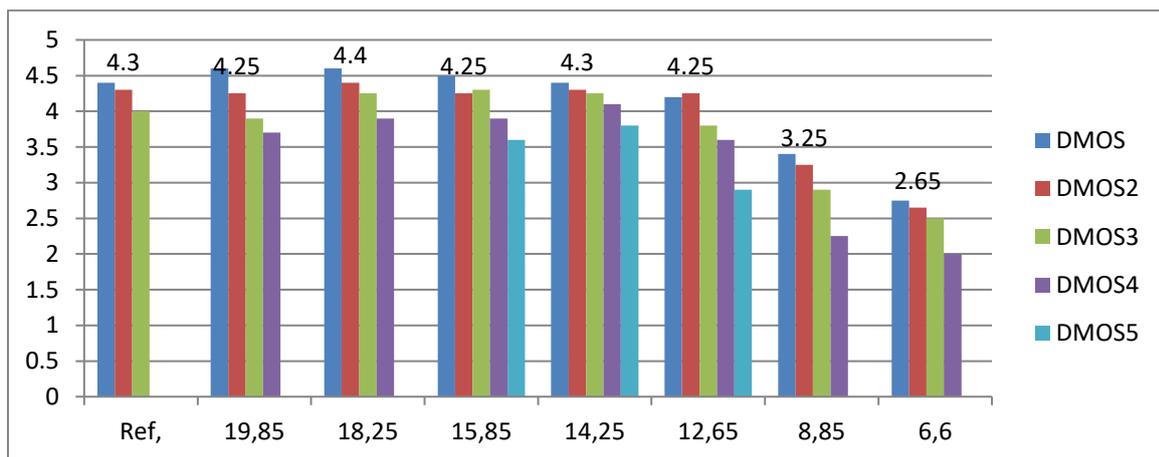


Figure : 07 Performances AMR-WB dans le canal GSM à plein débit sous des erreurs de canal et avec bruit de fond de voiture de 15 dB. De l'expérience 6a de caractérisation 3GPP phase avec La langage anglaise.

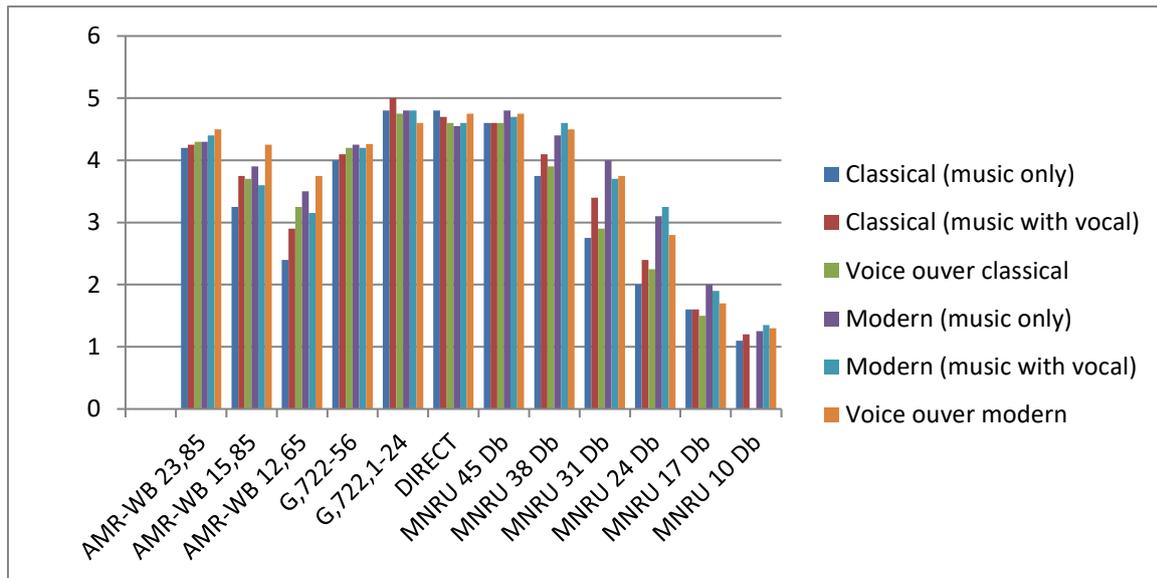


Figure : 08 Performance AMR-WB avec musique. De l'expérience 3b du test de caractérisation ITU-T.

Conclusion :

L'AMR-WB étend la bande passante audio à 7 KHz et offre une qualité vocale par rapport aux programmes actuels dans les réseaux téléphoniques fixes et Le deuxième et troisième génération de systèmes de communications mobile. [28]

Chapitre III : Résultats et simulations

Introduction :

L'objectif de cette travaille on à teste et simulation avec le codeur AMR-WB. Ensuite Nous avons trouvé 10 fichiers guère et testé avec le codeur AMR-WB telle que nous avons transformée la fréquence d'échantillonnage 8 KHZ à 16 KHZ.

Cette travaille la programmation on utilisé le langage C (builder C++ 6.0).

Alors cette travaille contient à trois étapes :

III.1.Lesetapes des Simulation :

III.1.1. étape 1 :

Le 1^{er} étape nous avons testé les 10 fichiers suivants :

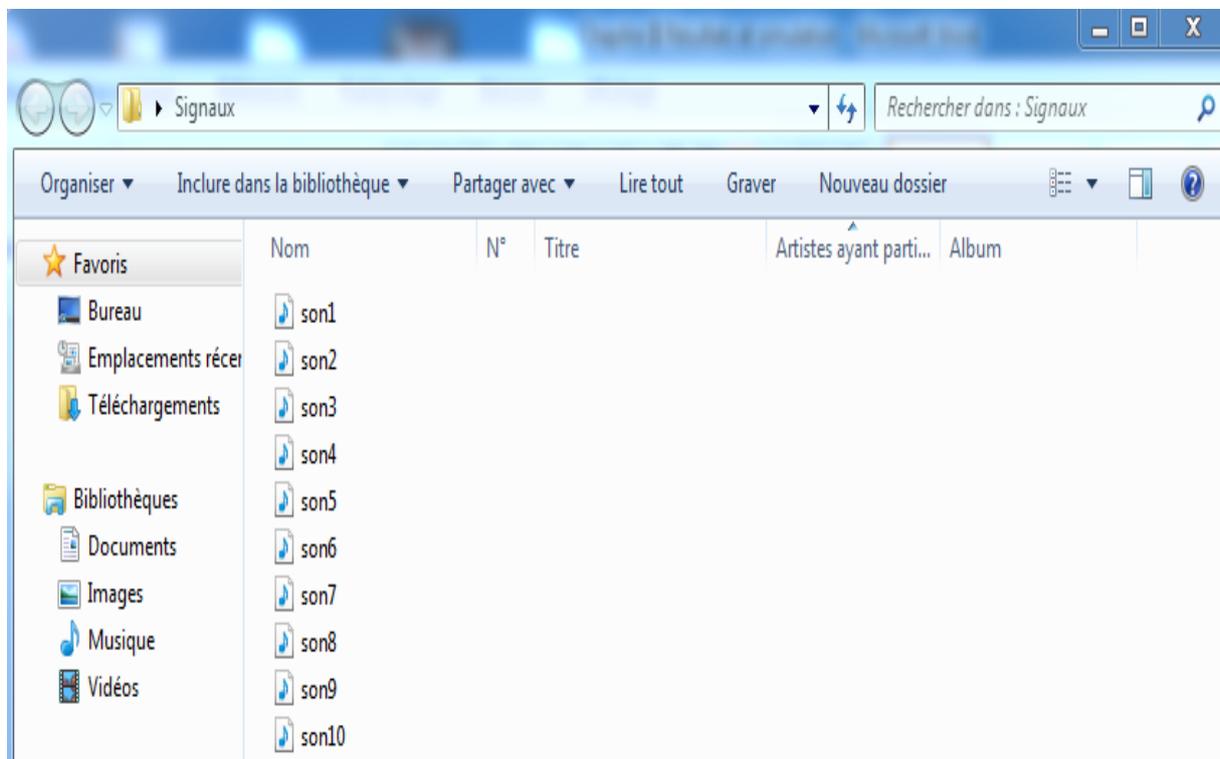


Figure 01 : dossier de 10 fichier

- Ensuite les 10 fichiers on à testé avec le codeur AMR :

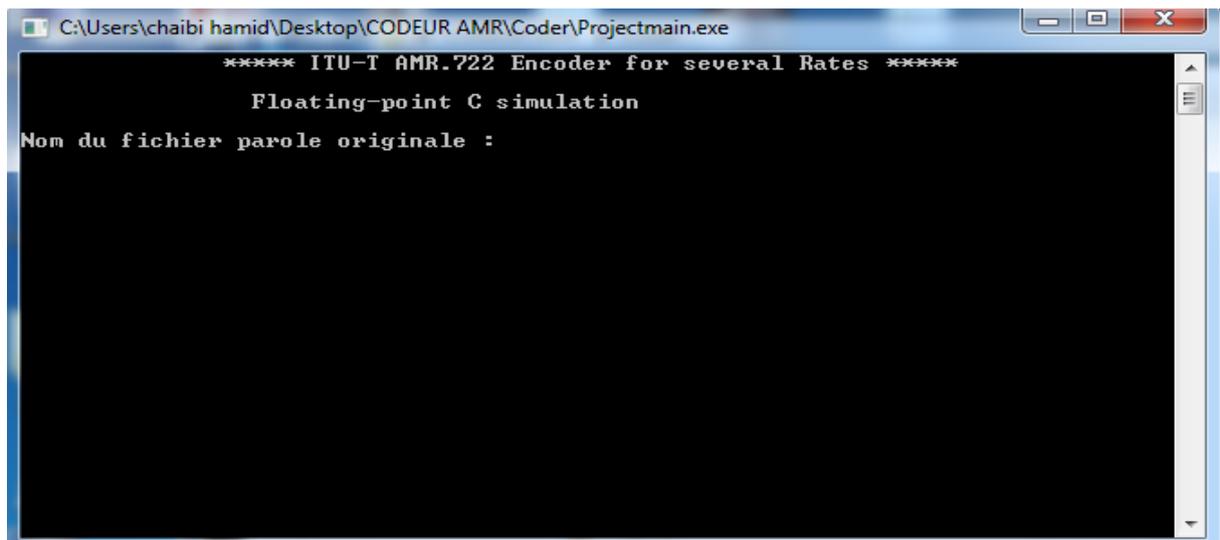


Figure 02 : écran de codeur AMR-WB

- Telle que on à testé 10 fichiers à déférentes débit :

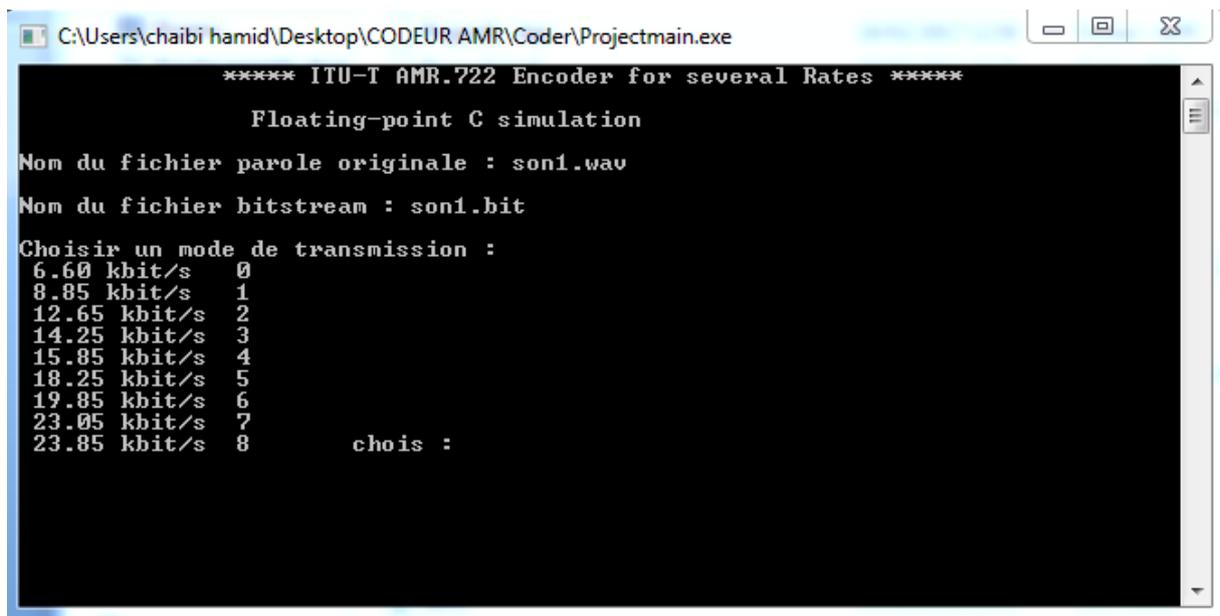


Figure 03 : écran de codeur AMR-WB et les 9 débits de transissions

- Donc on à testé les 10 fichiers avec le débit 6.60 kbit/s (0).
- Pour Le même travaille avec les débits suivants : (8.85 kbit/s (1), 12.65 kbit/s (2), 14.25 kbit/s (3) 15.85 kbit/s (4), 18.25 kbit/s(5) , 19.85 kbit/s (6) , 23.05 kbit/s (7) , 23.85 kbit/s (8)).

III.1.2. étape 2 :

- Les 10 fichiers on à testé avec le débit 6.6 kbit/s (0) à le décoder AMR :

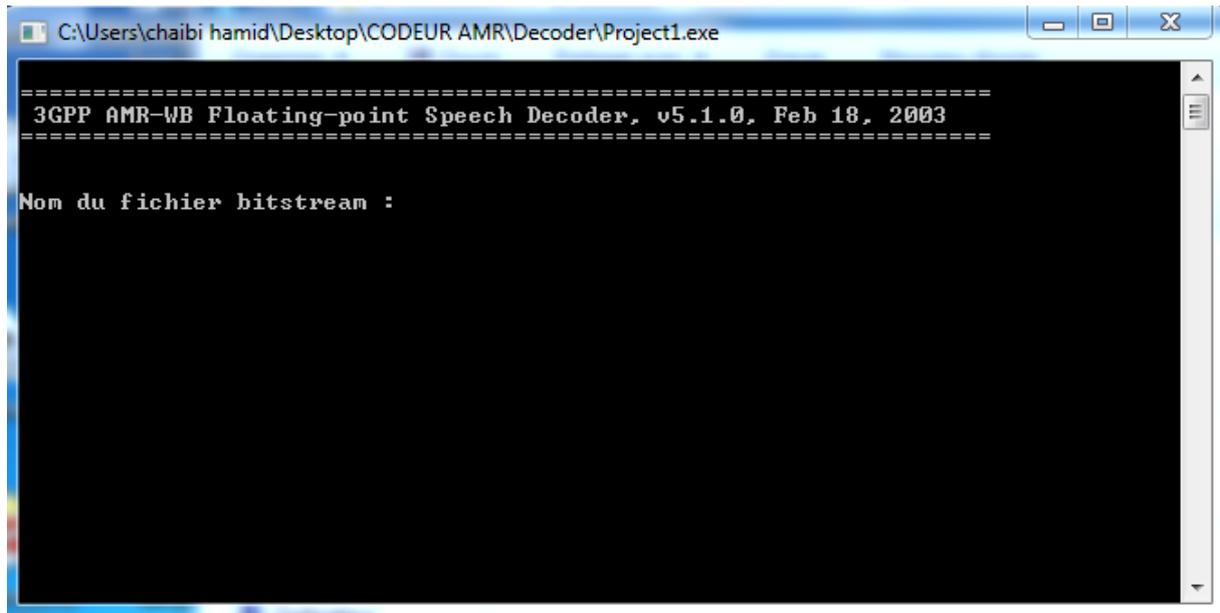


Figure 4 : écran de décodeur d'AMR-WB

- Pour le même travail avec les débits suivants : (8.85 kbit/s (1), 12.65 kbit/s (2) 14.25 kbit/s (3) 15.85 kbit/s (4), 18.25 kbit/s(5), 19.85 kbit/s (6), 23.05 kbit/s (7), 23.85 kbit/s (8)).

III.1.3. étape 3 :

- Enfin on a testé les 10 fichiers pour le débit 6.6 kbit/s (0) avec le PESQ pour calculer les valeurs de PESQ1 et PESQ2 :

```
donner le Nom du fichier de parole originale = son1.wav
donner le Nom du fichier de parole synthétique = son1d0s.wav
donner la fréquence d'échantillonnage +8000/+16000= : +16000
donner la bande passante Narrowband/Wideband +nb/+wb= : +nb

Reading reference file son1.wav...done.
Reading degraded file son1d0s.wav...done.
Level normalization...
IRS filtering...
Variable delay compensation...
Acoustic model processing...

P.862 Prediction (Raw MOS, MOS-LQO): = 3.352 3.344
```

Figure 05 : écran de standard PESQ

- Pour le même travail avec les débits suivants : (8.85 kbit/s (1), 12.65 kbit/s (2) 14.25 kbit/s (3) 15.85 kbit/s (4), 18.25 kbit/s(5), 19.85 kbit/s (6), 23.05 kbit/s (7), 23.85 kbit/s (8)).

- Ensuite en fait calculer le moyenne de PESQ1 et PESQ2, et nous avons fait un tableau pour les calculs des PESQ1 et PESQ2 et la moyenne de PESQ1 et PESQ2 et les débits précédent :

Les débits	La moyenne de PESQ1	La moyenne de PESQ2
0→ 6.60	3.4277	3.426
1→ 8.85	3.6561	3.7151
2→ 12.65	3.7769	3.8829
3→ 14.25	3.8870	3.936
4→ 15.85	3.8501	3.9259
5→ 18.25	3.8905	3.8094
6→ 19.85	3.9008	3.9173
7→ 23.05	3.9314	4.0570
8→ 23.85	3.9205	4.0344

Tableau 1 : représenté les valeurs de PESQ1 et PESQ2 avec fonction de débits

- Après avoir fait le tableau précédent, Ensuite nous avons tracé deux graphes telle que : le 1^{er} graphe représente les valeurs de la moyenne de PESQ1 et le 2^{ème} graphe représente les valeurs de PESQ2, telle que La moyenne de PESQ1 représente la Qualité de (RAW- MOS) et La moyenne de PESQ2 elle représenté la Qualité de (MOS-LQ0) :

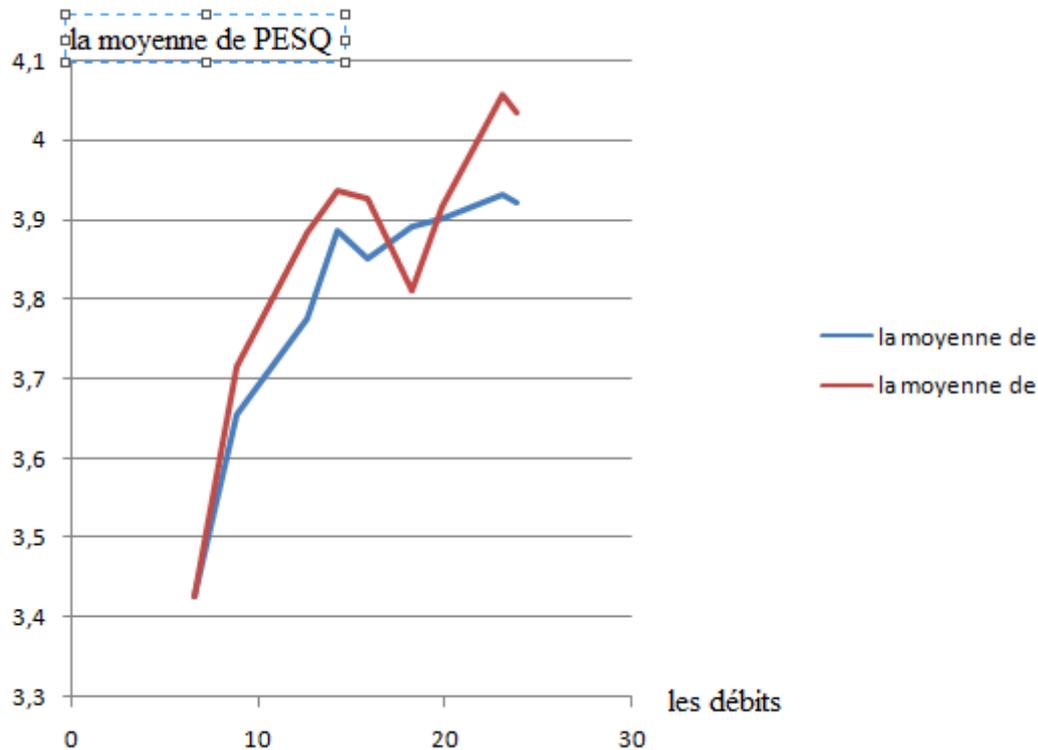


Figure 06 : les variations des valeurs de la moyenne de PESQ (Qualité des signaux)

III.2.Description et Commentaires :

La figure 06 représente les variations des valeurs de la moyenne de PESQ (Qualité des signaux) en fonction des débits telle que la fréquence d'échantillonnage 16 KHZ. Ensuite nous avons regardé deux courbes telle que le 1^{er} graphe représente la variation de la moyenne de PESQ1 (la Qualité de RAW-MOS).

- le 1^{er} courbe commencé à la valeur 6.6 kbit/s tu la rencontre la valeur 3.42 quand la valeur 6.6 kbit/s de débits après la 1^{er} valeur de débit la variation de courbe 1 est augmenté à l'intervalle [6.6 kbit/s : 14.25 kbit/s] après l'intervalle précédent la variation elle diminuer à l'intervalle [14.25 kbit/s : 15.85 kbit/s], donc la variation augmenté à l'intervalle [15.85 kbit/s : 23.05 kbit/s] et la valeur le plus élevée 3.9314. Après diminuer quand la valeur de débit 23.85 kbit/s la moyenne de PESQ1 = 3.9205.
- Le 2^{émé} graphe représenté la courbe de la moyenne de PESQ2 (la Qualité de MOS-LQ0) elle regardé le 2^{émé} courbe commencé à la valeur 3.42 quand la valeur 6.6 kbit/s de débits, Ensuite la courbe augmenté à l'intervalle [6.6 kbit/s : 14.25 kbit/s] la valeur de la moyenne de PESQ2 il ya 3.936 après la courbe diminuer à l'intervalle [14.25 kbit/s : 15.85 kbit/s] l'intervalle suivante la courbe augmenté à l'intervalle [15.85 kbit/s : 23.05 kbit/s], il ya la valeur le plus élevée 4.57. Après diminuer quand la valeur 23.85 kbit/s la moyenne de PESQ2 = 4.0344.

III.3.Remarque :

On a remarqué chaque fois les débits augmentés la variation des les deux courbes des Qualité augmenté, et puis chaque fois nous avons évalué la Qualité avec le PESQ.

Donc Il ya relation pour le débit et la moyenne de PESQ, telle que les débits augmenté la variation augmenté de la moyenne de PESQ.

Conclusion général

La technique standard de codage utilisée dans la téléphonie s'effectue dans une bande de fréquences étroite (200-3400 Hz) utilisant un ordre de prédiction du filtre analyse-synthèse égale à 10, cette dernière ne permet pas la transmission de signaux musicaux.

Nous avons vu le long des chapitres précédents que pour avoir une gamme de fréquences large s'étalant jusqu'à 16000 Hz (la large bande), et donc avoir plus de flexibilité et de performance, il faut augmenter l'ordre de prédiction des filtres.

AMR-WB étend la plage de fréquences audio à 7 KHz et offre également une qualité sonore supérieure à celle des logiciels de réseau téléphonique actuels et des systèmes de communication mobile 2G et 3G.

Le débit correspond à la variance de qualité moyenne des valeurs du PESQ de sorte que plus il y a de flux, plus la variance du PESQ moyen est grande alors il y a relation pour le débit et la moyenne de PESQ.

En premier lieu nous avons extraits les paires de raies spectrales LSP du signal parole échantillonné utilisé dans notre simulation, le codage se fait sur des trames de 80 échantillons chacune.

En deuxième lieu nous avons appliqué l'algorithme de Linde Buso Gray (LBG) afin de constituer un dictionnaire de quantification (code book) pour un nombre de bits par trame allant de 20 à 80 bits/trame.

En dernier lieu nous avons quantifié les LSP en utilisant les code books trouvés, par la suite on a schématisé les enveloppes spectrales des LSP originale et quantifiée et on a calculé les distorsions spectrales entre celles-ci pour des ordres de 10, 16, 18, 20 pôles et le résultat final est 16 pôles.

Annexe A

Algorithme de Levinson-Durban

Les coefficients d'autocorrélation $R(k)$, $k=0,1,\dots,P$, sont utilisés pour obtenir les coefficients du filtre LP après résolution du système linéaire (I.14)

Il s'agit donc d'inverser une matrice d'ordre ' p '. Les méthodes algébriques classiques exigent pour cela un nombre d'opérations (multiplication+ addition) de l'ordre de p^3 , ce que l'on note $O(p^3)$.

L'algorithme qui va être décrit profite de la structure particulière (Toeplitz symétrique) de la matrice d'autocorrélation pour résoudre (I.13) par une récursion sur l'ordre de prédiction : autrement dit, ils fournissent toutes les solutions d'ordre $M=1,2,\dots,p$, le nombre d'opérations est seulement $O(p^2)$.

La variance de l'erreur de prédiction α_p sera obtenue également par une récurrence sur l'ordre m .

Rappelons que la fonction d'autocorrélation est supposée connue et que pour un signal stationnaire, on a :

$$R(i, j) = R(|i - j|) = R(k) \quad (A.1)$$

Initialisation :

$$a_m(0) = 1, \quad (m=1,2,\dots,p) \quad E_0 = R(0) = \sigma_x^2$$

Récursion :

Pour : $m=1,2,\dots,p$.

$$k_m = -\frac{1}{E_{m-1}} [R(m) - \sum_{k=1}^{m-1} a_{m-1}(k)R(m-k)] \quad (A.2)$$

Pour $k = 1,2, \dots, m-1$.

$$a_k(m) = a_k(m-1) - k_m a_{m-k}(m-1) \quad (A.3)$$

$$E_m = E_{m-1}(1 - k_m^2) \quad (A.4)$$

Les coefficients $a_k(m)$ résultant, quand $m = p$ représentent les coefficients de prédiction d'un prédicteur linéaire d'ordre p :

La valeur de k_m joint à la propriété : $-1 \leq k_m \leq 1$

Cette relation est une condition nécessaire et suffisante pour que le filtre soit stable.

La méthode d'autocorrélation garantit la stabilité du filtre, de plus le calcul de $R(i)$ nécessite un fenêtrage de la $S(n)$ par un la fenêtre de Hamming.

Bibliographie

- [1] L. Vandendorpe, "Cours de Télécommunications" Université de Louvain, 2002.
- [2] André DIDIER, "Acoustique audiométrique" Encyclopédie UNIVERSALIS 5, 1999.
- [3] F.Merazka, "Techniques de codage de la parole : applications aux LSPs et aux systèmes VoIP", Thèse de Doctorat d'État, Présenté a l'École National Polytechnique Alger 2004.
- [4] T.Dutoit, "Introduction au Traitement Automatique de la Parole", Faculté Polytechnique de Mons 1989.
- [5] M. Xie et D.Berkani. "Amélioration des performances des codeurs de parole" Août 97
- [6] B; S. Atal, R. V. Cox, and P.Kroon, "Spectral quantization and interpolation for CELP coders," in Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Processing, (Glasgow UK), May. 1989, pp. 69-72.
- [7] W. Pariera, "Modifying LPC parameter dynamics to improve speech coder efficiency," Thesis, Department of Electrical & Computer Engineering, McGill University, Montreal, Canada, Sep 2001
- [8] P. E. Papamichlis, "Practical Approaches to Speech Coding", Prentice-Hall, Englewood Cliffs, N. J. 1987.
- [9] D.O'Shaughnessy,"speech communication, Human and machine. Reading", MA: Addison-Wesley, 1987.
- [10] J.D. Markel and A. H. Gray, Jr., "Linear prediction of speech", New York: Springer-Verlag, 1976.
- [11] F.Itakura, "Line spectrum representation of linear predictive coefficients of speech signals" J. Acoust. Soc. Amer., vol. 57, suppl. 1 p. S35 (A), 1975.
- [12] F. K. Soong and B. H. Juang, "Line spectrum pair (LSP) and speech data compression", in Proc. IEEE Int Conf. Acoust. Speech, Signal Processing, San Diego, CA, pp.1.10.1-1.10.4, Mar.1984.
- [13] B.S Atal, R,V Cox and P.Kroon,"Spectral quantization and interpolation for CELP coders", in Proc. IEEE int. Conf. On Acoustics, speech and signals.
- [14] Alexis Pascal Bernard, "Source-Channel Coding of Speech", Master of Science in Electrical Engineering University of California Los Angeles, 1998.
- [15] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective"

Bibliographie

- [16] R.Boite et M.Kunt,"Traitement de la parole", Presses Polytechniques Romandes, première édition
- [17] R. Laroia, N Phambo, and N,Favardin,"Robust abs=d efficient quantization of speech LSP parameter using structured vector quantizer", in Proc.IEEE Int. Conf on acoustics , speech , and Sig.processing(Toronto, Canada) ,may 1991 pp 641-644.
- [18] "AMR Wideband Speech Codec; General Description," 3GPP TS 26.171.
- [19] 3GPP 1999 TS 26.071, "AMR speech Codec; General description." <http://www.3gpp.org/ftp/Specs/html-info/26071.htm>.
- [20] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "GSM enhanced full rate codec," in IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Munich, Germany, Apr. 20–24, 1997, pp. 771–774.
- [21] 3GPP TS 26.190 "AMR Wideband speech codec; Transco-ding functions", version 5.0.0 (2001-03), 3rd Generation Partnership Project (3GPP).
- [22] K. Järvinen, J. Vainio, P. Kapanen, T. Honkanen, P. Haavisto, R. Salami, C. Laflamme, and J.-P. Adoul, "GSM enhanced full rate codec," in IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Munich, Germany, Apr. 20–24, 1997, pp. 771–774.
- [23] Adaptive Multi-Rate Wideband Speech Trans-coding, 3GPP TS 26.190.
- [24] R.Salami, C, Laflamme, J.P. Adoul, A.Kataoka, S. Hayashi, T.Moriya, C.Lamblin, D.Massaloux, S. Proust, P.Kroon, and Y. Shoham, "Design and description of CS-ACELP : A toll quality 8kb/s speech coder," IEEE Trans. Speech Audio Processing, vol. 6, no. 2, pp. 116-130, 1998.
- [25] Adaptive Multi-Rate Inband Control and Link Adaptation, 3GPP TS 45.009.
- [26] S. Bruhn, P. Blocher, K. Hellwig, and J. Sjöberg, "Concepts and solutions for link adaptation and inband signaling for the GSM AMR speech coding standard," in IEEE Vehicular Technology Conf., Amsterdam, The Netherlands, Sept. 19–22, 1999, pp. 2451–2455.
- [27] AMR Wideband Speech Codec; Voice Activity Detector (VAD), 3GPP TS 26.194.
- [28] AMR Wideband Speech Codec; Comfort Noise Aspects, 3GPP TS 26.192.