

3/04

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de La
Recherche Scientifique



المدرسة الوطنية المتعددة التقنيات
BIBLIOTHEQUE — المكتبة
Ecole Nationale Polytechnique

Ecole Nationale Polytechnique
Département d'Electronique
Laboratoire Signal et
Communication

Mémoire de Fin d'Etudes
En vue de l'obtention du Diplôme d'Ingénieur d'Etat en
Electronique

Thème

Amélioration des
Performances du Codec
G.723.1

Fait par :
Amine BOUABDALLAH

Proposé par :
M^{elle}: F.MERAZKA

Promotion 2003/2004



Remerciements

Mes remerciements vont, spécialement, vers mes parents pour leurs soutient indéfectible sans oublier mon encadrement et tout le corps enseignant et administratif et à tous ceux qui m'ont offert leur aide, consciemment ou inconsciemment.

Résumé

Dans le codec standard de l'ITU, le G.723.1, nous avons observé que, lors de l'occurrence de pertes de trames, les erreurs engendrées ne se limitaient pas aux trames perdues mais se propageaient aux trames suivantes. Comme l'origine de cette propagation est la quantification Inter-trames (quantification prédictive) des coefficients de paires de raies spectrales LSP, nous l'avons remplacé par une quantification Intra-trame qui respecte le débit et n'introduit pas de complexité supplémentaire significative par rapport à la méthode originale.

Les résultats obtenus montrent, dans des conditions de pertes de trames, une nette amélioration des performances du codec G.723.1.

Mots clés:

Codage de la parole, Voix sur IP, Quantification Inter-trame, quantification Intra-trame, Masquage de perte.

ملخص

لوحظ في الرّامزة التّمونجية ج.723.1 (G.723.1) المقّمة عند اتحاد الدّولي للمواصلاّت البعيدة (I.T.U) أنّ عند ضياع قطع من الكلام، الأخطاء المتولّدة لا تتحصّر في القطع المفقودة فقط بل تنتشر إلى القطع الموالية و بما أنّ الأصل في هذا الانتشار هو التكميم بين القطع "التكميم بالمتنبأ" (quantification Inter-trame) لمعاملات أزواج الخطوط الطيفية، (LSP) ، فغيرنا هذا التكميم بتكميم داخل القطع (quantification Intra-trame) حيث أنه يراعي المعدل الثنائي و لا يدخل تعقيدا إضافيا بالغا مقارنة مع الطريقة الأصلية. نهاية، النتائج المحصل عليها تظهر، في حالة ضياع قطع من الكلام، تحسن واضح في فعالية الرامزة ج.723.1

مفاتيح الكلمات: ترميز الكلام، الكلام عبر شبكة انترنت، تكميم بين القطع، تكميم داخل القطع، إخفاء فقدان القطع،

Abstract

In the codec G.723.1, error propagation can be observed with the occurrence of loss frames. As a matter of fact, the errors propagate to the next frames following those which are lost and don't stop within the lost frames. Since that propagation originates from the Inter-frame quantization (predictive quantization) of Line Spectral Pairs (LSP), we have replaced this method by an Intra-frame quantization which is likely to be respectful of the rate and does not introduce any significant additional complexity.

The results have shown a clear improvement of the performances of the G.723.1 in lossy conditions.

Key words:

Speech coding, Voice Over Internet protocol, Inter-frame quantization, Intra-frame quantization, packet loss concealment.

Sommaire

| | |
|---|-----------|
| LISTE DES FIGURES | 1 |
| LISTE DES TABLES | 2 |
| LEXIQUE | 3 |
| INTRODUCTION | 4 |
| CHAPITRE.I LE CODAGE DE LA PAROLE..... | 6 |
| I.1 Signal vocal..... | 6 |
| I.2 Redondance du signal vocal..... | 6 |
| I.3 Mécanisme de la phonation | 7 |
| I.4 Modèle de production de la parole | 10 |
| I.5 Prédiction linéaire | 12 |
| I.5.1 Méthode d'autocorrélation | 13 |
| I.5.2 Méthode de covariance | 15 |
| I.5.3 Considérations pratiques..... | 16 |
| I.5.4 Représentation des paramètres de prédiction | 17 |
| I.6 Mesure de distorsion | 18 |
| I.6.1 Mesure de distorsion subjective | 19 |
| I.6.2 Mesure de distorsion objective..... | 19 |
| CHAPITRE.II LA VOIX SUR LE PROTOCOLE INTERNET (VOIP)..... | 24 |
| II.1 Le Protocole Internet IP..... | 25 |
| II.1.1 Format des en-têtes IP | 27 |
| II.1.2 La qualité de service..... | 28 |
| II.2 La Voix sur IP..... | 29 |
| II.3 Les composants VoIP..... | 29 |
| II.3.1 Les Codec | 30 |
| II.3.2 Les protocoles TCP/IP | 31 |
| II.3.3 Les protocoles VoIP | 32 |
| II.4 Techniques de recouvrements de paquets, issus de flux sonores et perdus sur les réseaux IP. | 34 |
| II.4.1 Réparation basée au niveau de l'émetteur (Sender-Based repair) | 34 |
| II.4.2 Réparation basée au niveau du récepteur "Masquage d'erreur" | 36 |
| II.5 Avantages de la technologie VOIP..... | 38 |

| | | |
|---|---|-----------|
| II.6 | Activités de normalisation de la téléphonie IP..... | 39 |
| II.7 | Conclusion | 40 |
| CHAPITRE.III LE CODEC STANDARD G.723.1..... | | 41 |
| III.1 | Caractéristiques générales..... | 41 |
| III.1.1 | Domaine d'application..... | 41 |
| III.1.2 | Débits binaires..... | 42 |
| III.1.3 | Signaux d'entrée possibles..... | 42 |
| III.1.4 | Délai..... | 42 |
| III.2 | Principes de codage..... | 43 |
| III.2.1 | Description générale..... | 43 |
| III.2.2 | Quantificateur des coefficients de paires de raies spectrales (LSP)..... | 45 |
| III.3 | Principes de décodage..... | 47 |
| III.3.1 | Description générale..... | 47 |
| III.3.2 | Décodeur des coefficients LSP..... | 48 |
| III.3.3 | Interpolation des coefficients LSP..... | 48 |
| III.3.4 | Traitement par interpolation des trames..... | 49 |
| CHAPITRE.IV ETUDES, SIMULATIONS ET RESULTATS | | 51 |
| IV.1 | Propagation de l'erreur lors d'effacements de trames..... | 51 |
| IV.2 | Quantification Intra-trame des LSP..... | 53 |
| IV.2.1 | Base de données et étude statistique..... | 53 |
| IV.2.2 | Quantification scalaire des différences (DSQ)..... | 58 |
| IV.2.3 | Quantification vectorielle..... | 61 |
| IV.2.4 | Quantification vectorielle des différences en sous-vecteur (DSVQ)..... | 62 |
| IV.2.5 | Quantification vectorielle des différences en sous-vecteurs améliorée (EDSVQ)..... | 64 |
| IV.2.6 | Quantification vectorielle en sous-vecteurs des différences améliorée et modifiée (MEDSVQ)..... | 66 |
| IV.3 | Masquage de trames perdues par interpolation..... | 72 |
| IV.3.1 | Modélisation des pertes de paquets de données dans les réseaux de paquets..... | 72 |
| IV.3.2 | Masquage..... | 75 |
| IV.3.3 | Simulations et résultats..... | 76 |
| CONCLUSION GENERALE..... | | 81 |
| ANNEXE A | | 83 |
| ANNEXE B | | 86 |
| ANNEXE C | | 87 |
| BIBLIOGRAPHIE | | 90 |

Liste des figures

| | |
|--|----|
| Figure 1. Appareil phonatoire | 8 |
| Figure 2. Un son voisé et son spectre | 9 |
| Figure 3. Un son non voisé et son spectre..... | 10 |
| Figure 4. Modèle simplifié de production de la parole..... | 11 |
| Figure 5. Spectre LPC avec LSF superposés[6] | 18 |
| Figure 6. Paquet IP et son format d'en-tête..... | 27 |
| Figure 7. Classification des techniques de réparations basées à l'émetteur | 34 |
| Figure 8. Forward error correction | 35 |
| Figure 9. Interleaving | 36 |
| Figure 10. Classification des techniques de masquage d'erreur..... | 37 |
| Figure 11. G.723.1/ Organigramme du codeur | 45 |
| Figure 12. G.723.1/Organigramme du décodeur. | 47 |
| Figure 13. Propagation de l'erreur dans le codage inter-trame (PSVQ). | 52 |
| Figure 14. Histogramme normalisé représentant les variations des coefficients <i>LSP</i> | 55 |
| Figure 15. Histogramme normalisé des coefficients ΔLSP | 57 |
| Figure 16. Distribution des ΔLSP avec leurs dictionnaires..... | 59 |
| Figure 17. Propagation de l'erreur dans le codage intra-trame (DSQ)..... | 60 |
| Figure 18. Modèle de perte de paquets de Gilbert..... | 73 |
| Figure 19. Variation des EMBSD en fonction des taux de pertes. | 79 |
| Figure 20. Variation de la distorsion spectrale en fonction du taux de perte..... | 80 |
| Figure 21. Autres histogramme des LSP (cartésiennes)..... | 86 |
| Figure 22. Histogramme LSP (polaires)..... | 86 |
| Figure 23. Histogramme des distorsions spectrales (R = 0%)..... | 87 |
| Figure 24. Histogramme des distorsions spectrales (R = 10%)..... | 87 |
| Figure 25. Histogramme des distorsions spectrales (R = 20%)..... | 88 |
| Figure 26. Histogramme des distorsions spectrales (R = 30%)..... | 88 |
| Figure 27. Histogramme des distorsions spectrales (R = 40%)..... | 89 |



Liste des tables

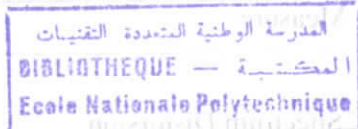
Liste des figures

| | |
|--|----|
| Table 1. Abréviations..... | 3 |
| Table 2. Les codec et leurs performances..... | 31 |
| Table 3. Répartition des orateurs de la base de donnée TIMIT..... | 53 |
| Table 4. Longueurs de la base de données..... | 54 |
| Table 5. Ecart type et dynamique des coefficient <i>LSP</i> | 56 |
| Table 6. Coefficients de corrélation linéaire des coefficients <i>LSP</i> | 56 |
| Table 7. Ecart type et dynamique des coefficients ΔLSP | 57 |
| Table 8 : Allocation optimale de bits pour la DSQ..... | 58 |
| Table 9. Performance de la méthode DSQ..... | 61 |
| Table 10. Découpage et allocation de bits pour la DSVQ..... | 62 |
| Table 11. Découpage et allocation de bits pour la DSVQ..... | 65 |
| Table 12. Performances de la DSVQ..... | 65 |
| Table 13. Performance de la MEDSVQ..... | 68 |
| Table 14. Performance de la MEDSVQ pour différentes allocations de bits..... | 69 |
| Table 15. Performances de la MEDSVQ pour une allocation de bits optimisée..... | 69 |
| Table 16. Performances de la MEDSVQ optimisée par l'allocation de bits et le métrique de FarVardin..... | 70 |
| Table 17. Comparaison entre PSVQ (G. 723.1) et MEDSVQ..... | 71 |
| Table 18. Récapitulatif des performances des différentes méthodes utilisées..... | 71 |
| Table 19. Statistique des pertes en réseaux..... | 74 |
| Table 20. Comparaison des EMBSD des différentes méthodes de quantification sous l'effet de pertes..... | 77 |
| Table 21. . Comparaison des EMBSD des différentes méthodes de quantification sous l'effet de pertes..... | 78 |
| Table 22. Performances complètes de la MEDSVQ..... | 83 |
| Table 23. Résultats complets de la DSQ..... | 84 |
| Table 24. Résultats complets de la PSVQ (standard)..... | 85 |

Lexique

| | |
|--------|--|
| ACELP | Excitation par Séquences Codées à Structure Algébrique |
| AR | Auto Régressif |
| DAM | Diagnostic Acceptability Measure |
| DS | Distorsion Spectrale |
| EMBSD | Enhanced Modified Bark Spectrum Distorsion |
| IETF | Internet Engineering T Forum |
| ITU | International telecommunication Union |
| LPC | Codage par Prédiction Linéaire |
| LSP | Line Spectral Frequencies |
| LSP | Coefficient de Paires de raies Spectrales |
| MGCP | Media Gateway Control protocol |
| MIC | Modulation d'Impulsion Codée |
| MOS | Mean Opinion Score |
| MP-MLQ | Quantification d'Impulsions Multiples selon le critère du Maximum de Vraisemblance |
| PESQ | Perceptual evaluation of speech quality |
| PL | Prédiction Linéaire |
| PSVQ | Predictive Split Vector Quantizer |
| RSB | Rapport Signal sur Bruit |
| RSBseg | RSB segmenté |
| RTCP | Real Time Control Protocol |
| RTP | Real Time Protocol |
| SIP | Session Initiation Protocol |
| SNR | Signal to Noise Ratio |
| TFD | Transformé de Fourier Discrète |

Table 1. Abréviations.



Introduction

Le transport, fiable et rapide, de l'information, et plus généralement, la communication, a été à travers les âges et jusqu'au jour d'aujourd'hui, un domaine de recherche inépuisable. A l'aube du 21^{ème} siècle, l'utilisation de réseaux informatiques, tel le réseaux IP(Internet), pour le transport de communication vocale, est l'objet de recherches soutenues.

L'intérêt majeur à faire passer de la parole sur des réseaux IP est d'ordre économique. En effet l'utilisation d'Internet permettra de contourner les réseaux téléphoniques traditionnels, et ainsi, réduire les coûts d'une communication.

Dans les application Voix sur IP (VoIP), la parole est numérisée puis codée, trame par trame (à intervalle de temps régulier), par le biais de codec (codeur/décodeur) tel le G.723.1 ou le G.729, ensuite, les flux de trames recueillis sont empaquetés dans des paquets de données réseaux (IP Datagramme). Lesquels seront envoyés, sur le réseau, vers le récepteur qui fera la conversion inverse et entamera la lecture sur les sorties du terminal.

Cependant, l'intégration de deux technologies distinctes à savoir: les réseaux informatiques à commutations de paquet et le codage de la parole, engendre plusieurs problèmes qui peuvent être affiliés à deux causes majeures :

Introduction générale

les réseaux IP n'ont pas été, initialement, conçus afin de transporter des flux en temps réels.

En effet, les réseaux IP étant du type "Best Effort", il n'y a, simplement, aucune garantie quant à la réception des paquets envoyés; tout au moins dans des délais de temps convenables.

Les pertes de paquets n'ont pas été la préoccupation principale (relativement au débit et à la complexité) des concepteurs de codec de parole. En effet, dans le codec G.723.1, lequel nous a servi de plateforme pour mener nos travaux, la corrélation inter-trame (entre trames successives) est mise à profit pour quantifier les paramètres obtenus après analyse des échantillons de la parole par prédiction linéaire (PL). De ce fait, un quantificateur prédictif (PSVQ) a été adopté assurant, de la sorte, un débit minimum et une complexité moyenne pour une qualité similaire à celle fournie par la téléphonie traditionnelle (qualité "Toll"). Néanmoins, puisque chaque trame envoyée contient une information commune à elle et à la trame suivante, lors de perte de trames, l'erreur conséquente ne se limitera pas à la trame perdue mais se propagera aux trames suivantes reçues, et ceci se poursuivra jusqu'à la ré-synchronisation du codeur et du décodeur.

L'objectif de notre travail est d'implémenter, sur le codec G.723.1, de nouvelles techniques de quantification qui, en enrayant la propagation des erreurs, améliorent la robustesse du codec aux pertes, et sans pour autant détériorer ses autres performances (débit et complexité).

La thèse est organisée en quatre chapitres:

Le chapitre 1 aborde les points essentiels du codage de la parole : la prédiction linéaire, la représentation paramétrique et la mesure de distorsion. Le chapitre 2 est consacré à la Voix sur IP (VoIP), le réseau IP (Internet) et les différentes méthodes de recouvrement de pertes. Le chapitre 3 détaille le codec G.723.1, principes de codage et de décodage. Le chapitre 4 regroupe toutes les études, simulations et résultats entrepris. Enfin une conclusion générale résume et clos le travail.

CHAPITRE.I Le codage de la parole

1.1 Signal vocal

La parole est la faculté de communiquer la pensée par un système de sons articulés. L'information d'un message parlé réside dans les fluctuations de la pression de l'air, engendrées, puis émises, par l'appareil phonatoire. Ces fluctuations constituent le signal vocal ; elles sont détectées par l'oreille, laquelle procède à une analyse [1].

1.2 Redondance du signal vocal

Le signal vocal est caractérisé par une très grande redondance, condition nécessaire pour résister aux perturbations du milieu ambiant. Cette redondance sera mise à profit par les techniques de codage de la parole, pour réduire le débit binaire nécessaire au stockage ou à la transmission de la parole, sans, pour autant nuire à son intelligibilité. Pour la langue française, qui comporte 36 phonèmes¹, on a calculé l'information moyenne ou "entropie" de la source par l'Equation (1.1) et trouvé $H = 4.73$, valeur à comparer avec 5.17 ($2^{5.17} = 36$) cas où tous les phonème serait équiprobable. [1]

¹ Phonème : plus petite unité présente dans la parole et susceptible par sa présence de changer la signification d'un mot [1].

$$H(X) = -\sum_i p(x_i) \log_2(p(x_i)) \quad (1.1)$$

Dans la conversation courante, environ dix phonèmes sont prononcés chaque seconde; l'information moyenne est donc inférieure à 50 bits/s [1]. Or, par exemple, pour un canal téléphonique, supposé continu et sans erreur, de bande passante $B = 3000\text{Hz}$ et avec un rapport signal sur bruit $S/N = 30\text{ dB}$, on trouve que le débit maximum d'information C calculé par l'Equation (1.2) vaut $C = 30000\text{ bits/s}$ [1].

$$C = B \log_2(1 + S/N) \quad (1.2)$$

Il y a donc apparemment une redondance énorme dans le signal vocal. La suppression partielle des redondances permet une représentation plus efficace des données.

La compression des données peut se faire sans pertes d'information, ou avec pertes en exploitant dans ce cas la tolérance de l'organe récepteur (l'oreille). La compression du signal consistera à réduire les redondances du signal de la parole.

1.3 Mécanisme de la phonation

Les principaux organes composant l'appareil phonatoire sont : les poumons, la trachée-artère, le pharynx, les cavités buccales et nasales (Figure 1). L'appareil respiratoire fournit l'énergie nécessaire à la production de sons, en poussant de l'air à travers la trachée-artère. Au sommet de celle-ci se trouve le *larynx* ou la pression de l'air est modulée avant d'être appliquée au conduit vocal. Le larynx est un ensemble de muscles et de cartilages mobiles qui entourent une cavité située à la partie supérieure de la trachée. Les *cordes vocales* sont en fait deux lèvres symétriques placées en travers du larynx. Ces lèvres peuvent fermer complètement le larynx et, en s'écartant progressivement, déterminer une ouverture triangulaire appelée *glotte*. L'air y passe librement pendant la respiration et la voix chuchotée, ainsi que pendant la phonation des sons non voisés.

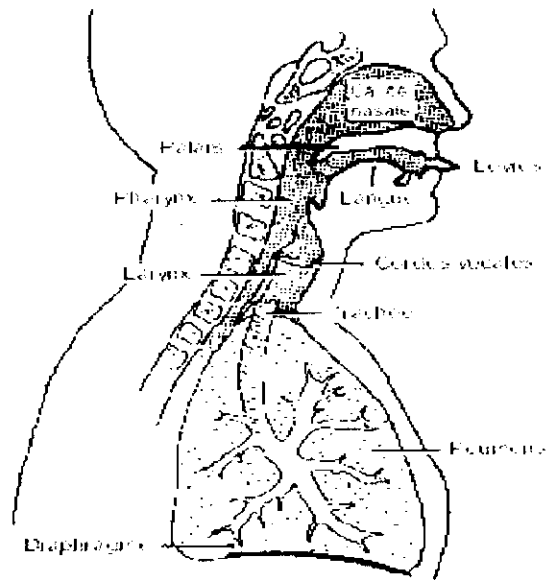


Figure 1. Appareil phonatoire

Les sons voisés résultent au contraire d'une vibration périodique des cordes vocales. Le larynx est d'abord complètement fermé, ce qui accroît la pression en amont des cordes vocales, et les force à s'ouvrir, ce qui fait tomber la pression, et permet aux cordes vocales de se refermer; des impulsions périodiques de pression sont ainsi appliquées au conduit vocal, composé des cavités pharyngienne et buccale pour la plupart des sons. Lorsque la *lucette* est en position basse, la cavité nasale vient s'y ajouter en dérivation. Notons pour terminer le rôle prépondérant de la langue dans le processus phonatoire. Sa hauteur détermine la hauteur du pharynx : plus la langue est basse, plus le pharynx est court. Elle détermine aussi le *lieu d'articulation*, région de rétrécissement maximal du canal buccal, ainsi que l'aperture, écartement des organes au point d'articulation.

L'intensité du son émis est liée à la pression de l'air en amont du larynx ; sa hauteur est fixée par la fréquence de vibration des cordes vocales, appelée fréquence du fondamentale ou *pitch*.

La fréquence du fondamentale peut varier [1] :

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 2. On y observe les raies qui correspondent aux harmoniques du fondamentale F_0 (structure de *pitch*), l'enveloppe de ces raies présente des maximums appelés *formants* et qui correspondent aux fréquences propres F_i ($i=1, 2, 3, \dots$) du conduit vocal (structure formantique).

Les trois premiers formants sont essentiels pour caractériser le spectre vocal, les formants d'ordre supérieur ont une influence plus limitée.

Un son non voisé ne présente pas de structure périodique, il peut être considéré comme un bruit blanc filtré par la transmittance de la partie du conduit vocal situé entre la constriction et les lèvres (Figure 3), son spectre ne présente donc pas de structure de *pitch*.

La classification qui vient d'être exposée est forcément un peu sommaire et surtout elle concerne la production normale de la parole. Ainsi une voyelle peut être chuchotée, c'est-à-dire produite avec la glotte largement ouverte, dans ce cas le spectre du signal résulte de l'excitation du conduit vocal par une source aléatoire : c'est un spectre continu qui présente une structure formantique semblable à celle d'une voyelle voisée. Par contre, il ne possède pas de structure de *pitch* (raies dues aux harmoniques du fondamentale).



Figure 2. Un son voisé et son spectre

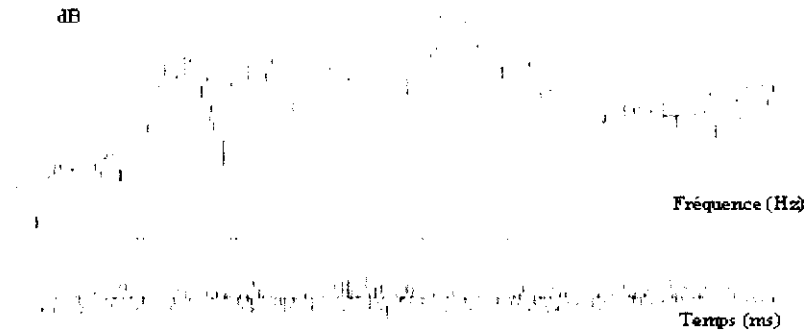


Figure 3. Un son non voisé et son spectre.

1.4 Modèle de production de la parole

Fant [6] a proposé en 1960 un modèle de production dont nous résumons ici la version numérique. Un signal voisé peut être modélisé par le passage d'un train d'impulsions $u(n)$ à travers un filtre numérique récursif de type tout-pôles (AR^2). On montre que cette modélisation reste valable dans le cas de sons non voisés, à condition que $u(n)$ soit cette fois un bruit blanc. Le modèle final est illustré à la Figure 4 Il est souvent appelé modèle auto régressif, parce qu'il correspond dans le domaine temporel à une régression linéaire de la forme :

$$X(n) = G \cdot u(n) + \sum_{i=1}^p -a_i X(n-i) \quad (1.3)$$

(Où $u(n)$ est le signal d'excitation), ce qui exprime que chaque échantillon est obtenu en ajoutant un terme d'excitation à une prédiction obtenue par combinaison linéaire de p échantillons précédents.

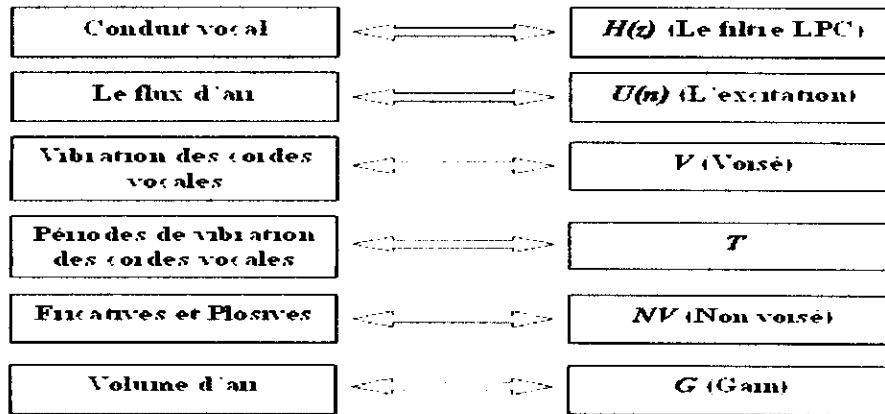
Les coefficients du filtre sont d'ailleurs appelés coefficients de prédiction et le modèle AR est souvent appelé modèle de prédiction linéaire.

Les paramètres du modèle AR sont : la période du train d'impulsions (sons voisés uniquement), la décision Voisé / Non Voisé (V/NV), le gain σ , et les coefficients du filtre $1/A(z)$, appelé filtre de synthèse.

² Abréviation de "Auto Régressif"

Chapitre I: Le codage de la parole

Les relations d'équivalence entre le modèle physique et le modèle mathématique sont :



Le problème de l'estimation d'un modèle AR, souvent appelée analyse LPC^3 revient à déterminer les coefficients d'un filtre tout pôles dont on connaît le signal de sortie, mais pas l'entrée. Il est par conséquent nécessaire d'adopter un critère, afin de faire un choix parmi l'infinité de solutions possibles. Le critère classiquement utilisé est celui de la minimisation de l'énergie de l'erreur de prédiction.

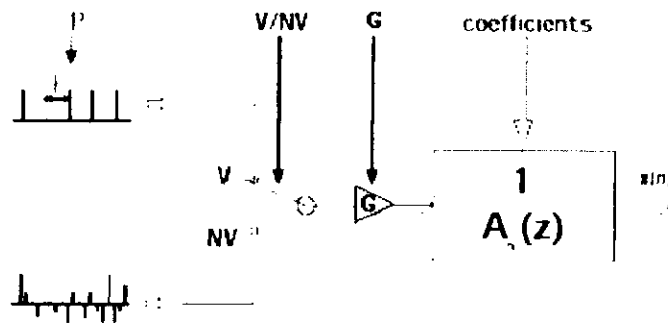


Figure 4. Modèle simplifié de production de la parole

³ Dans ce mémoire "Linear Prediction Coding" est abrégé LPC

1.5 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 $u(n)$ [3] et [5] :

$$s(n) = \sum_{k=1}^p a_k s(n-k) + G \sum_{l=0}^q b_l u(n-l), \quad b_0 = 1, \quad (1.4)$$

Où $\{a_k\}$, $\{b_l\}$, et le gain G sont les paramètres du système. L'Equation (1.4) prédit la sortie courante en utilisant une combinaison linéaire des sorties antérieures et des entrées courantes et antérieures.

Dans le domaine fréquentiel, la fonction de transfert du modèle de prédiction linéaire de la parole est de la forme :

$$H(z) = \frac{B(z)}{A(z)} = \frac{G[1 + \sum_{l=1}^q b_l z^{-l}]}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (1.5)$$

Les racines du dénominateur et numérateur sont, respectivement, les pôles et les zéros du système ou modèle pôle-zéro $H(z)$.

Si $a_k = 0$ pour $1 \leq k \leq p$, $H(z)$ devient un modèle tout zéro ou modèle à moyenne ajustée (MA). Si $\{b_l = 0\}$ pour $1 \leq l \leq q$, $H(z)$ devient un modèle tout pôle ou modèle Auto Régressive (AR) :

$$H(z) = \frac{1}{A(z)} \quad 6.$$

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 $H(z)$. Par contre, les voyelles contiennent des résonances qui peuvent être modélisées par le modèle tout-pôle. Pour des raisons de simplicité, ce modèle est préféré pour l'analyse par prédiction linéaire de la parole.

Ainsi, le signal prédit est égal à :

$$s(n) = \sum_{k=1}^p a_k s(n-k) \quad (1.7)$$

Et l'erreur de prédiction ou résiduel du signal est la sortie $e(n)$:

$$e(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (1.8)$$

Pour avoir les coefficients du filtre court-terme $\{a_i\}$ du processus AR, la méthode classique des moindres carrés peut être utilisée. La variance ou l'énergie, du signal erreur $e(n)$ est minimisée sur une trame de parole. Deux grandes approches sont utilisées pour l'analyse LPC court-terme : la méthode d'*autocorrélation* et la méthode de *covariance*.

1.5.1 Méthode d'autocorrélation

Dans la méthode de l'autocorrélation le signal est défini pour toutes les valeurs de temps ; et il est identiquement nul en dehors d'une séquence de N échantillons, où N est un entier; ceci équivaut à multiplier le signal de parole par une fenêtre de longueur finie correspondant à N échantillons.

$$\begin{cases} S_f(n) = W(n) \cdot S(n) & \text{pour } 0 \leq n \leq N-1 \\ S_f = 0 & \text{ailleurs} \end{cases} \quad (1.9)$$

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

$$\begin{cases} W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1} & \text{pour } 0 \leq n \leq N-1 \\ W(n) = 0 & \text{ailleurs} \end{cases} \quad (1.10)$$

Chaque échantillon peut être prédit approximativement à partir de p échantillons précédents. Ceci est valable pour toutes les valeurs du temps : $-\infty < n < +\infty$.

Chapitre I: Le codage de la parole

L'erreur quadratique totale entre le signal fenêtré et le modèle (signal prédit) est minimisée sur l'ensemble des échantillons.

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

$$E = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} (s_f(n) - \sum_{k=1}^p a_k s_f(n-k))^2 \quad (1.11)$$

L'erreur de prédiction peut être minimisée en annulant les dérivées partielles par rapport aux coefficients du filtre :

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

On obtient p équation linéaire avec " p " coefficient inconnus a_k :

$$\sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n-k) = \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n) \quad 1 \leq i \leq p \quad (1.12)$$

$$\sum_{k=1}^p a_k R(i, k) = R(0, i) \quad (1.13)$$

Où,

$$R(i, j) = \sum_{n=-\infty}^{\infty} s_f(n-i) s_f(n-j)$$

Et puisque le signale fenêtré $S_f(n)$ est durée limitée,

$$R(i, j) = R(|i - j|) \quad (1.14)$$

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 : $R(i) = R(-i)$

Sous la forme matricielle, l'ensemble des équations linéaires représenté par l'Equation (1.12) peut être réécrit comme suit:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(2) & \dots & R(p-2) \\ R(2) & R(0) & \dots & \\ \vdots & \dots & \dots & \\ \vdots & \dots & \dots & \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ \vdots \\ R(p) \end{bmatrix} \quad (1.15)$$

Et regroupées selon $R \cdot a = r$, où la matrice $p \times p$ R est une matrice Toeplitz nommée la matrice d'autocorrélation. L'algorithme de Levinson-Durbin est utilisé pour résoudre le système et trouver les coefficients de prédiction minimisant la moyenne quadratique de l'erreur de prédiction.

1.5.2 Méthode de covariance

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

$$E = \sum_{n=-\infty}^{\infty} e_f^2(n) = \sum_{n=-\infty}^{\infty} e^2(n)w^2(n) \quad (1.16)$$

En annulant les dérivations partielles par rapport aux coefficients du filtre $\frac{\delta E}{\delta a_k} = 0$

Pour $1 \leq k \leq p$, on a "p" équations linéaires.

$$\sum_{k=1}^p \Phi(i, k) = \Phi(i, 0) \quad 1 \leq i \leq p \quad (1.17)$$

Où la fonction de covariance $\Phi(i, k)$ est définie par l'Equation (1.18)

$$\Phi(i, k) = \sum_{n=-\infty}^{\infty} w(n)s(n-1)s(n-k) \quad (1.18)$$

Sous la forme matricielle, les p équations deviennent $\Phi a = \Psi$

$$\begin{bmatrix} \Phi(1,1) & \Phi(1,2) & \dots & \Phi(1,p) \\ \Phi(2,1) & \Phi(2,2) & \dots & \Phi(2,p) \\ \dots & \dots & \dots & \dots \\ \Phi(p,1) & \Phi(p,2) & \dots & \Phi(p,p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} \Psi(1) \\ \Psi(2) \\ \dots \\ \Psi(p) \end{bmatrix} \quad (1.19)$$

Où : $\Psi(i) = \Phi(i,0)$ pour $1 \leq i \leq p$.

La matrice Φ n'est pas une matrice Toeplitz et ne garantit pas la stabilité du filtre LPC. Elle est symétrique et définie positive ce qui permet de résoudre le système par la décomposition de Cholesky.

1.5.3 Considérations pratiques

Pour mener à bien une analyse LPC, il faut pouvoir choisir :

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

Le choix de la fréquence d'échantillonnage est fonction de l'application visée et de la qualité du signal à analyser. On choisira plutôt 8 kHz pour les signaux téléphoniques, 10 kHz pour les applications de reconnaissance, et 16 kHz pour les applications de synthèse.

L'ordre de prédiction P , est choisi de façon à ce qu'il permette de bien représenter toute séquence de signal de parole.

L'ordre p est fonction de la fréquence d'échantillonnage, on estime en général qu'une paire de pôles est nécessaire par 1 KHz de bande passante. De plus, il a été établi expérimentalement que l'énergie de l'erreur de prédiction diminue rapidement lorsqu'on

augmente p , pour tendre vers une asymptote horizontale: il devient inutile, alors, d'augmenter l'ordre, puisque l'erreur ne diminue plus.

La durée des trames d'analyse et leur décalage sont souvent fixés à 30 et 10 ms respectivement. Ces valeurs ont été choisies empiriquement; elles sont liées au caractère quasi-stationnaire du signal parole.

Enfin, comme vu précédemment dans la méthode d'autocorrélation, pour compenser les effets de bord, on multiplie en général préalablement chaque tranche d'analyse par une fenêtre de pondération $w(n)$, la plus souvent utilisée est celle de Hamming (Equation (1.10))

1.5.4 Représentation des paramètres de prédiction

Les coefficients de prédiction linéaire PL ne sont pas appropriés à la transmission de paroles codées. En effet, du à leurs grandes dynamiques, ces coefficient se quantifient mal et la stabilité du filtre de synthèse reste difficile à vérifier. Pour palier ce problème, avant quantification, les coefficients PL sont transformés vers un ensemble de paramètres ayant des propriétés convenable à la quantification et à la restitution. Plusieurs représentations ont été développées, nous citerons les paires de raies spectrales (LSP), les coefficients de réflexion, *log area ratio* (LAR), les coefficients cepstraux, ...etc. Cependant la représentation la plus répandue et la plus prisée pour ses performances reste la représentation en paires de raies spectrales LSP.

1.5.4.1 Paires de raies spectrales

Connue aussi sous le nom de fréquence de raies spectrales, la représentation LSP a été introduite par Itakura [2].

Soient les polynômes $P(z)$ et $Q(z)$ définis par:

$$\begin{cases} P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \\ Q(z) = A(z) - z^{-(p+1)} A(z^{-1}) \end{cases} \quad (1.20)$$

Donc:
$$A(z) = \frac{1}{2} [P(z) + Q(z)] \quad (1.21)$$

Chapitre I: Le codage de la parole

Les zéros des polynômes $P(z)$ et $Q(z)$ sont appelés les LSP. Ces polynômes ont les propriétés suivantes:

1. tous les zéros de $P(z)$ et $Q(z)$ se trouvent sur la cercle unité.
2. 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 [3] 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[4] [5] [6]:

- Comme illustré à la Figure 5, 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

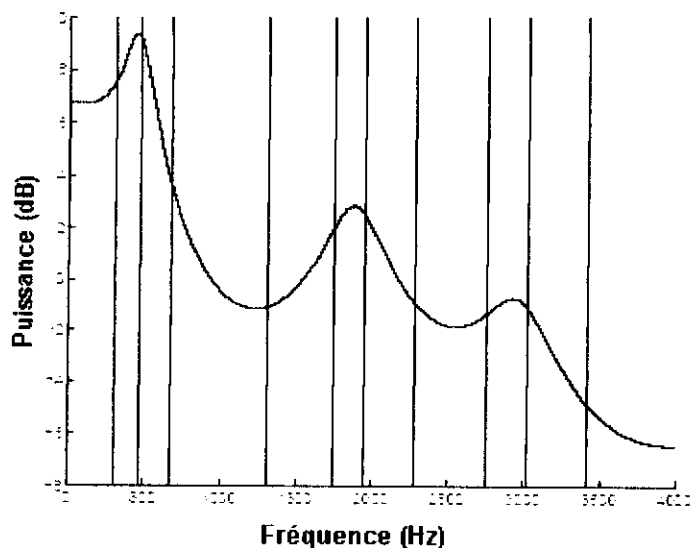


Figure 5. Spectre LPC avec LSF superposés[6]

1.6 Mesure de distorsion

Les mesures de distorsion sont, généralement, utilisées pour comparer les performances de différents systèmes ou de différentes configurations. Aussi, pour qu'une

mesure de distorsion soit utile, il faut qu'elle reflète le plus fidèlement possible la qualité subjective (perçue) de la parole. Les nombreuses mesures de distorsion actuellement disponible peuvent être regroupées dans deux catégories principales : la mesure de distorsion subjective et la mesure de distorsion objective.

1.6.1 Mesure de distorsion subjective

Ce genre de distorsion se base sur l'appréciation ou l'opinion, qu'ont un groupe d'auditeurs, sur la qualité et l'intelligibilité de la parole. Puisque le système auditore humain reste l'ultime juge de la qualité et de l'intelligibilité d'un signal parole, la mesure subjective est l'évaluation la plus précise des performances d'un codeur de parole. Néanmoins, cette technique revient chère en terme de temps et requière tout un ensemble d'auditeurs avérés et distincts. De plus, la distorsion perçue peut varier en fonction de facteurs tels que l'haut-parleur utilisé (écouteur, combiné téléphonique...) et le volume de lecture, ce qui requière en plus un environnement d'écoute homogène[5] [7].

Le MOS (Mean Opinion Score) et le DAM (Diagnostic acceptability Measure) sont les tests de qualité subjective les plus utilisés.

1.6.2 Mesure de distorsion objective

Les mesures de ce genre peuvent être calculées directement du signal parole, de son spectre ou de n'importe quels paramètres dérivés. Puisque elles ne requièrent pas de tests d'écoute ces mesures peuvent donner une estimation immédiate de la qualité d'un algorithme de codage. De plus, elles peuvent servir, pendant les étapes de quantification, comme un souple critère mathématique à minimiser. Les deux principaux éléments à vérifier lors du choix d'une distorsion objective sont ses performances et sa complexité. Les performances d'une mesure objective résident dans sa corrélation avec la mesure subjective correspondante (qualité ou intelligibilité). Les mesures de distorsion sont classifiées en trois domaines: domaine temporelle (RSB et RSBseg), domaine fréquentielle (Distorsion spectrale) et domaine perceptuel (EMBSD).

1.6.2.1 Rapport signal sur bruit (RSB)

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 alors le RSB a la forme suivante:

$$RSB = 10 \text{ Log}_{10} \frac{\sum_{n=0}^{N_T-1} S(n)^2}{\sum_{n=0}^{N_T-1} [S(n) - \tilde{S}(n)]^2} \quad (dB) \quad (1.22)$$

Le RSB donne une valeur après avoir traité tout le fichier, donc il n'y a pas moyen de retrouver les instants où les divergences ont été enregistrées. De plus, Le RSB est dominé par les portions de forte énergie (tranches voisées), alors que le bruit a un effet perceptuel plus important sur les portions de faible énergie [5].

1.6.2.2 Rapport signal sur bruit segmenté (RSBseg)

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 :

$$RSB_{seg} = \frac{1}{N_F} \sum_{i=0}^{N_F-1} 10 \text{ Log}_{10} \frac{\sum_{j=0}^{N_S-1} S(N_S i + j)^2}{\sum_{j=0}^{N_S-1} [S(N_S i + j) - \tilde{S}(N_S i + j)]^2} \quad (dB) \quad (1.23)$$

Le RSBseg est meilleur que le RSB. Cependant, les tranches de silences renvoient de grandes valeurs 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.

1.6.2.3 Distorsion Spectrale

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

$$DS_i = \sqrt{\frac{1}{F_e} \int_0^{F_e} \left[10 \text{Log}_{10} \frac{S_i(f)}{\tilde{S}_i(f)} \right]^2 df} \quad (dB) \quad (1.24)$$

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

$$S_i(f) = \frac{1}{A_i(e^{j2\pi f / F_e})} \quad (1.25)$$

$$\tilde{S}_i(f) = \frac{1}{\tilde{A}_i(e^{j2\pi f / F_e})} \quad (1.26)$$

Où, $A_i(z)$ et $\tilde{A}_i(z)$ sont, respectivement, les polynômes PL Original et quantifié vu plus haut, pour la trame i . Au lieu de l'intégration, une sommation des coefficients obtenus après application de la TFD (transformer de Fourier Discrète) aux coefficients LPC, peut être utilisée pour calculer DS_i . La distorsion devient donc:

$$DS_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{k=n_0}^{n_1-1} \left[10 \text{Log}_{10} \frac{S_i(e^{j2\pi k / N})}{\tilde{S}_i(e^{j2\pi k / N})} \right]^2} \quad (dB) \quad (1.27)$$

Dans notre travail, les signaux d'entrés sont échantillonnés à $F_e = 8$ KHz et nous avons calculé la distorsion sur une bande allant de 0 Hz à 3 KHz avec une TFD sur $N=256$ points. Ce qui donne $n_0 = 0$ et $n_1 = 95$. La résolution fréquentielle est de 31.25 Hz (8000/256).

Une distorsion spectrale moyenne (la moyenne des distorsions spectrales calculées pour toutes les trames) de 1 dB est habituellement acceptée. Cependant, selon [4] et [3] les conditions de "transparence spectrale" (pas de distorsion audible) établies expérimentalement sont les suivantes:

- La moyenne DS est inférieure à 1 dB.
- Le nombre de trames ayant DS_i dans l'intervalle 2-4 dB est inférieur à 2%.
- Pas de trames ayant DS_i supérieure à 4 dB.

La distorsion spectrale est très utilisée dans l'évaluation des performances d'un quantificateur de paramètres PL.

I.6.2.4 Mesure de distance euclidienne LSP pondérée.

Cette distance a été développée dans le but d'optimiser la quantification des paramètres PL[4], elle a la forme suivante:

$$d_{LSP} = \sum_{i=1}^p [c_i w_i (\omega_i - \tilde{\omega}_i)]^2 \quad (1.28)$$

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

$$c_i = \begin{cases} 1.0, & \text{pour } 1 \leq i \leq 8, \\ 0.8, & \text{pour } i = 9, \\ 0.4, & \text{pour } i = 10. \end{cases} \quad (1.29)$$

Ces poids sont utilisés pour donner plus d'importance aux basses fréquences par rapport aux hautes fréquences. Ceci est justifié par le fait que l'oreille humaine est plus sensible aux basses fréquences qu'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 donnée par:

$$w_i = [S(e^{j\omega})]^r \quad (1.30)$$

Où, r est une constante empirique qui contrôle le degré de la pondération, empiriquement $r = 0.15$ [4].

Une pondération plus simple a été proposé par [8]. Elle a la forme suivante:

$$w_i = \frac{1}{\omega_i - \omega_{i-1}} + \frac{1}{\omega_{i+1} - \omega_i} \quad (1.31)$$

Où $\omega_0 = 0$ et $\omega_{p+1} = \pi$

Chapitre I: Le codage de la parole

Les mesures dans le domaine perceptuel sont basées sur les modèles d'audition humaine. Le signal est transformé vers un domaine perceptuel adéquat de telle manière qu'on puisse exploiter effets de masquage psycho-acoustique [5]. Parmi les mesures perceptuelles les plus utilisées nous pouvons citer: Perceptual evaluation of speech quality (PESQ) et Enhanced Modified Bark Spectrum Distorsion (EMBSD).

L'EMBSD estime la distorsion perceptuel d'un signal déformé en le comparant au signal original dans le domaine des sons forts (loudness domain) tout en tenant compte du seuil de masquage de bruit modifié et du modèle cognitif basé sur le post-masquage [9][10][11].

CHAPITRE.II La voix sur le protocole Internet (VoIP)

L'utilisation croissante des réseaux fondés sur le protocole Internet (IP) pour les services de communication, y compris pour les applications telles que la téléphonie, est devenue une question cruciale pour l'industrie des télécommunications dans le monde entier. L'acheminement du trafic vocal sur des réseaux IP ouvre la porte à une utilisation plus efficace des ressources du réseau et, en générale, à une multitude de nouveaux services. Cependant, sa mise en œuvre engendre, tout aussi bien, toutes une série de nouveaux problèmes (fiabilité du réseau, perte de paquets). Ces problèmes trouvent leur origine dans les différences de conception des deux technologies qui les supportes, à savoir :

- Le réseau téléphonique public commuté, RTCP ou (Public Switched telephone Network, PSTN), fondé sur la technologie de la commutation de circuits et qui, jusqu'à ces dernières années, était fortement réglementé dans la plupart des pays;

L'Internet, fondé sur la technologie de la commutation par paquets, et qui s'est transformé en réseau de données est peu, voire pas du tout, réglementé.

II.1 Le Protocole Internet IP

Pour pouvoir comprendre les difficultés qui existent quant à l'implémentation de transmissions sonores en temps réel, il faudrait, d'abord, bien comprendre les motivations et les buts recherchés, par les concepteurs du réseau qui est, aujourd'hui, le support de ces transmissions, puis, analyser la nature et les principes selon lesquels fonctionne.

Afin d'éclaircir ces points, on a résumé les grandes lignes du protocoles IP dans les paragraphes suivants:

Un réseau « idiot » mais robuste : Le réseau IP ne possède pas « d'intelligence ».les routeurs et les nœuds qui supportent un flux de paquets n'échangent entre eux aucun accusés de réception, ne maintiennent pas d'états sur les paramètres d'une connexion, le chemin emprunter n'est pas fixé à l'avance et puisque chaque paquet arrive à « l'improviste », il n'y a bien sur pas de réservation de ressource. Le rôle du routeur se contente d'examiner individuellement les paquets et de calculer pour chacun d'eux, en fonction de sa destination et de la topologie du réseau, le meilleur chemin à emprunter. Le nœud ne fait donc que router aveuglément les paquets un par un, dans l'ignorance totale du contexte, de la connexion, de leur appartenance à un flot et de la redondance des calculs de routages. Cet apparent aveuglement est voulu. Il offre, dans le cas de défaillance d'un des nœuds, une souplesse totale, et confère une robustesse exemplaire au réseau. Lors de congestion ou de rupture d'une liaison, les paquets sont aiguillés vers une autre route de manière transparente, sans que l'application soit alertée, ni que la connexion soit interrompue. Moins le réseau offre de service, plus il est universel.

Une gestion de bout en bout (end to end) par des terminaux « intelligents » : Corollaire du principe précédent, ce sont les terminaux d'extrémité qui prennent en charge la totalité de la gestion et du contrôle de la connexion. C'est l'application chargée dans le terminal qui gère la mise en paquet, l'expédition, la réception et la restitution de l'information. Ce sont les terminaux de réceptions qui repèrent les paquets manquants et

déclenchent la procédure de remplacement du paquet manquant (exemple : réémission par le terminal source).

Le « best effort » ou le service « pour le mieux » : Le réseau créé en suivant les principes précédents ne fournit qu'un service dit « pour le mieux ». Les paquets arrivent à destination si :

Les routeurs surchargés ne les détruisent pas faute de place dans les buffers.

Leur temps de vie n'est pas dépassé à force de transiter par de trop nombreuses escales.

Le délai est lui aussi très long, variable et imprévisible : il dépend de l'encombrement du réseau, ou de la taille des files d'attente aux portes des routeurs. Dans le cas d'une transmission de données peu sensible aux délais et à leurs variations, le service « pour le mieux » est tout à fait acceptable. Néanmoins, dans le cas de service en temps réel, la variation de délai, appelée gigue, rend très délicate la synchronisation du signal reçu avec le signal d'origine.

Le partage des ressources et l'attente indifférenciée : Réseau de type data, l'Internet, sans jamais rejeter la moindre demande de connexion, partage équitablement la bande passante disponible entre tous les usagers. Sur IP, si un paquet ne peut être traité au moment même où il se présente dans la mémoire du routeur, il est mis en attente dans une file. Enfin, si les besoins en bande passante et en puissance de traitement dépassent les capacités offertes par le réseau et les routeurs, alors il y a congestion et la mémoire des buffers est libérée.

L'indépendance par rapport aux sous réseaux : Le protocole IP a été conçu pour assurer le routage des données entre sous réseaux Il a été étudié pour être indépendant des technologie en vigueur sur ces sous réseaux, indépendant des machines hôtes ou de leurs systèmes d'exploitation. C'est ainsi que l'IP n'est attaché à aucune architecture particulière, et s'adapte aussi bien à des micro-ordinateurs qu'à des stations de travail ou à des mainframes.

L'évolutivité : Le protocole IP offre toutes les garanties d'évolutivité et de capacité de croissance requises. Pour rattacher un nouveau sous réseau, il suffit de tirer une nouvelle liaison, et l'ensemble fonctionne sans la moindre modification.

Le multicast : Le multicast, c'est-à-dire l'envoi d'un même message vers une liste de destinataire, est un fondement de l'Internet et imprègne la plupart de ses protocoles. La diffusion multicast permet d'économiser près de 50% des ressources. Un même paquet émis à l'intention de plusieurs destinataires n'est présent qu'en un seul exemplaire sur une même liaison ; ce sont les nœuds intermédiaires qui se chargent de sa duplication à chaque croisement de chemins.

II.1.1 Format des en-têtes IP

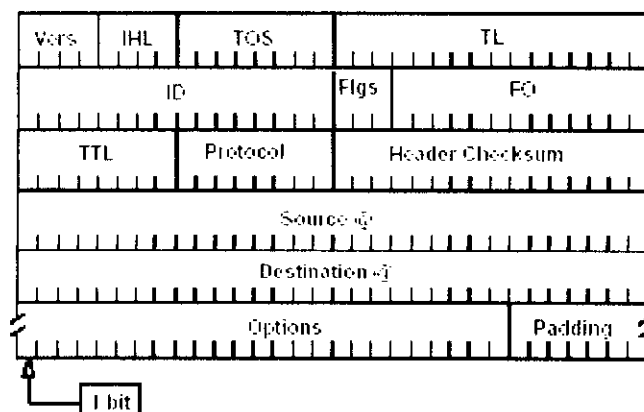


Figure 6. Paquet IP et son format d'en-tête.

Signification des champs se trouvant dans l'en-tête

Version sur 4 bits : définit le numéro de la version du protocole IP

Longueur d'en-tête sur 4 bits : code la longueur de l'en-tête.

Type de service sur 8 bits : donne une indication sur la qualité de service souhaité. Ce champ est, malheureusement pour VOIP, peu pris en compte par les routeurs actuels.

Longueur totale sur 16 bits : contient la longueur du datagramme entier y compris en-tête et données (en octets).

Identification sur 16 bits : identifie les fragments d'un même datagramme.

Flags sur 4 bits : divers commutateurs de contrôle :

0 , 0 , AF , DF

Bit 0-1 : réservés

Bits 2 : (AF) 0 = fragmentation possible, 1=non fractionnable.

Bits 3 : (DF) 0 = dernier fragment, 1 = fragments intermédiaires

Fragment offset sur 12 bits : indique le décalage du premier octet du fragment par rapport au datagramme complet.

Durée de vie sur 8 bits : limite le temps pendant lequel un datagramme reste dans le réseau. Si ce champ prend la valeur zéro, le datagramme doit être détruit. Chaque module IP (routeur) retire au moins une unité de temps à ce champ. Ce mécanisme est motivé par la nécessité de détruire les datagrammes qui n'ont pu être acheminés, en limitant la durée de vie même du datagramme et ainsi éviter tout problème de surcharge réseau.

Protocole sur 8 bits : indique quel protocole de niveau supérieur est utilisé dans la section « données » du datagramme IP (UDP, TCP, ...).

Checksum d'en-tête sur 16 bits : donne la valeur de checksum calculé sur l'en-tête uniquement. Comme certains champs de l'en-tête sont modifiés (ex : durée de vie) pendant leur transit à travers le réseau, ce checksum est recalculé et vérifié à chaque point du réseau ou l'en-tête est réinterprétée.

Adresse source sur 32 bits : l'adresse IP de la source

Adresse destination sur 32 bits : l'adresse IP de la destination.

Option variable : les datagrammes peuvent contenir des options

II.1.2 La qualité de service

Le champ " Type de Service " est utilisé pour guider le choix de services, lorsqu'un datagramme transite dans un réseau. Certains réseaux offrent un mécanisme de priorité, traitant préférentiellement un tel trafic par rapport à un autre. En général, en acceptant seulement de véhiculer des paquets d'un haut niveau de priorité lors d'une surcharge momentanée. Cependant, ce champ est encore peu utilisé par les modules du réseau.

Principalement, le choix offert est une négociation entre les trois contraintes suivantes : faible retard, faible taux d'erreur, et haut débit

TOS = | Priorité | D | T | R | 0 | 0 |

- Bits 0-2 : Priorité.
- Bit 3 : 0 = Retard standard, 1 = Retard faible.
- Bits 4 : 0 = Débit standard, 1 = Haut débit.
- Bits 5 : 0 = Taux d'erreur standard, 1 = Taux d'erreur faible.
- Bit 6-7 : Réservé.

Le "Type de Service" sert à préciser le traitement effectué sur le datagramme pendant sa transmission à travers un réseau. Les bits de priorité servent à hiérarchiser les paquets, les bits DTR permettent de personnaliser le routage en fonction du service désiré.

II.2 La Voix sur IP

La voix sur IP, ou VoIP, est le transfert de conversations vocale sous forme de données sur un réseau IP. Contrairement aux réseaux traditionnels à commutation de circuit (RTCP) ; dans les appels VoIP, la connexion téléphonique est à commutation de paquets.

Avec un appel VoIP, la partie de l'établissement de l'appel doit être simulée : la tonalité, les signaux de sonneries et les signaux occupés. La partie audio de l'appel elle-même (la conversation) a besoin d'être converti de son format analogique à un format numérique, découpée en paquets, envoyée à travers le réseau dans un format paquet, ré-assemblée, et à présent, reconverti du format numérique au format analogique . Les codec à chaque bout font la conversion de l'analogique au numérique et vice vers ça.

II.3 Les composants VoIP

Pour transférer de la voix sous forme de données sur le même réseau transportant les emails et les pages Web, un nouvel ensemble de composant est rajouté à ceux déjà existants,. Parmi ces composant on peut citer les suivants :

Les codec

Les protocoles TCP/IP

Les protocoles VoIP

Les serveurs de téléphonie IP et les PBXs

Les routeurs et les gateways VOIP

Les gatekeepers

Les téléphones IP et les softphones

II.3.1 Les Codec

Un codec (acronyme de « codeur/décodeur ») est le hardware ou software qui échantillonne les sons analogiques et les convertis aux formats numériques avec un débit binaire prédéterminé. Les codec font aussi de la compression afin d'économiser la bande passante.

Il existe une multitude de codec, chacun avec ses propres caractéristiques. Par exemple, les codec nommés **G.711u** et **G.711a** font la conversion de l'analogique au numérique et du numérique à l'analogique en assurant, relativement, une bonne qualité. Comme avec la plupart des appareils numériques, plus on requiert de qualité et plus il y a de bits.

La Table 2 rassemble quelques codec et leurs performances.

Chapitre II : la Voix Sur IP

| Standard (N) | Algorithm (N) | Complexity (MIPS) | Frame Size (bytes/frame) | Compression | Bit rate (kb/s) | MOS (N) |
|----------------|---------------|-------------------|--------------------------|-------------|-----------------|---------|
| G.711 | PCM | 0 | 0.125:0 | 1 | 64 | 4.10 |
| G.726 G.727 | ADPCM | 1 | 0.125:0 | 1/2 7/2 1/6 | 16/24/32/40 | 3.85 |
| G.722 | SB-ADPCM | 10 | 0.125:1.5 | 1.3:1.1:1 | 48/56/64 | 3.3 |
| G.728 | LD-CELP | 30 | 0.125:0 | 1 | 16 | 3.61 |
| G.729 | CS-ACELP | 20 | 10:5 | 8 | 8 | 3.92 |
| G.729A | CS-ACELP | 14 | 10:5 | 8 | 8 | 3.7 |
| G.723.1 | MPC-MBQ | 16 | 30:7.5 | 10:2:12:1 | 6.3/5.3 | 3.9 |
| GSM 06.10 | RPE-LTP | 10 | 20:0 | 10 | 13 | 3.5 |
| IS-54 | VSELP | 24 | 20:5 | 8 | 8 | 3.51 |
| IS-96 | QCELP | 20 | 20:5 | 7.5:16/32 | 8.5/1/2 | |
| IS-106 | CELP | 30 | | 13:3 | 18 | 3.0 |
| IS-105 | LPC10E | 15 | | 26:7 | 24 | 2.1 |

Table 2. Les codec et leurs performances.

II.3.2 Les protocoles TCP/IP

La famille des protocoles TCP/IP forme la base de l'Internet et des réseaux d'entreprises les plus récentes. Les programmes informatiques envoient et reçoivent les données sur un réseau IP en faisant appel au logiciel TCP/IP, connu comme la pile de protocole, dans leur ordinateur local.

Les piles TCP/IP dans l'ordinateur local échange de l'information avec la pile TCP/IP dans l'ordinateur cible pour accomplir le transfert de données d'un bout à un autre. L'information qu'ils échangent consiste en la dimension des paquets de données qu'ils s'échangent (la dimension du datagramme), l'identificateur associé avec chaque datagramme (l'en-tête du datagramme) et qu'est ce qui devrait se produire si un datagramme est perdu ou a été endommagé pendant son transit.

II.3.2.1 Le protocole TCP

Le rôle du protocole "Transmission Control protocol (TCP)" est d'éviter que des données soient perdues, dupliquées, ou désordonnées. TCP est, aussi, connu comme étant un protocole en mode connecté parce qu'il a besoin d'une requête et d'un acquittement avant de commencer le transfert de données.

II.3.2.2 Le protocole UDP

Le "User Datagramme Protocol (UDP)" est, aussi, connu comme étant un protocole en mode non connecté, ce qui veut dire qu'en utilisant ce protocole, les deux côtés d'une transmission ne s'échangent aucune requête ou acquittement avant d'entamer un transfert de données. Ainsi l'émetteur n'a aucune garantie sur la livraison des paquets émis, et il a à gérer cela.

II.3.3 Les protocoles VoIP

Ils sont de deux sortes:

II.3.3.1 Les protocoles de signalisation

Les protocoles de signalisation d'appel utilisent les protocoles TCP et UDP pour encapsuler les phases d'établissement et de libération des ressources d'un appel. Pour exemple dans la téléphonie sur IP, ils prennent en charge différentes fonctions comme la correspondance entre les numéros téléphoniques et les adresses IP, génération de la tonalité et des signaux occupés. Le protocole H.323 est le protocole de signalisation d'appel le plus largement déployé⁴. Le protocole MGCP (Media Gateway Control protocol) est moins flexible pour une utilisation avec un équipement traditionnel comme les téléphones de maison.

La famille des protocoles H.323 est une famille de protocoles robustes et flexibles ceci a été la conséquence de plusieurs années de raffinage. Mais le coût de cette robustesse est qu'il faut toute une série d'acquittements et de données échangées pour chaque fonction exécutée pendant une session d'appel.

⁴ Source : Insight Research janvier 2001

Le SIP (Session Initiation Protocol) et MGCP sont des protocoles légers développés par l'IETF dans les réseaux de données. Le SIP en particulier représente typiquement la logique des réseaux de données qui demande pourquoi utiliser de lourds protocoles (tel que H.323) quand un protocole léger (tel que le SIP) accomplira, la plupart du temps, le même travail.

II.3.3.2 Les protocoles temps réels (steaming protocols)

II.3.3.2.1 RTP: Real Time Transport Protocol

RTP, standardisé en 1996, est un protocole adapté aux applications présentant des propriétés temps réel. Il permet ainsi de :

Reconstituer la base de temps des flux (horodatage des paquets : possibilité de re-synchronisation des flux par le récepteur)

Détecter les pertes de paquets et en informer la source

Identifier le contenu des données pour leurs associer un transport sécurisé.

En revanche, ce n'est pas "la solution" qui permettrait d'obtenir des transmissions temps réel sur IP. En effet, il ne procure pas de :

Réservation de ressources sur le réseau (pas d'action sur le réseau),

Fiabilisation des échanges (pas de retransmission automatique, pas de régulation automatique du débit).

Garantie dans le délai de livraison (seules les couches de niveau inférieur le peuvent) et dans la continuité du flux temps réel.

II.3.3.2.2 RTCP: Real Time Transport Control Protocol (RFC 1889)

Ce protocole a pour but de transmettre périodiquement des paquets de contrôle à tous les participants d'une session. Pour expliquer les différents paquets de contrôle fournis par RTCP, il faut voir dans quel contexte RTP est utilisé. Par exemple, une application de visioconférence pourra l'utiliser pour transmettre les caractéristiques de chacun des participants.

Ce protocole définit quatre principaux paquets de contrôle :

SR (Sender Report) : ce rapport regroupe des statistiques concernant la transmission (pourcentage de perte, nombre cumulé de paquets perdus, variation de délai (gigue), ... Ces rapports sont issus de l'émetteurs actifs d'une session.

RR (Receiver Report) : ensemble de statistiques portant sur la communication entre les participants. Ces rapports sont issus des récepteurs d'une session.

SDES (Source Description) : carte de visite de la source (nom, e-mail, localisation).

BYE : message de fin de participation à une session.

II.4 Techniques de recouvrements de paquets, issus de flux sonores et perdus sur les réseaux IP.

Vu l'impact très néfaste qu'ont les pertes de paquets sur la qualité des transmissions des flux sonores, plusieurs techniques de recouvrement de paquets perdus ont été mises en œuvre. Ces techniques résumées en [12] se divisent en deux parties complémentaires:

1. réparation basée au niveau de l'émetteur (Sender-Based repair)
2. réparation basée au niveau du récepteur (Receiver-Based repair)

II.4.1 Réparation basée au niveau de l'émetteur (Sender-Based repair)

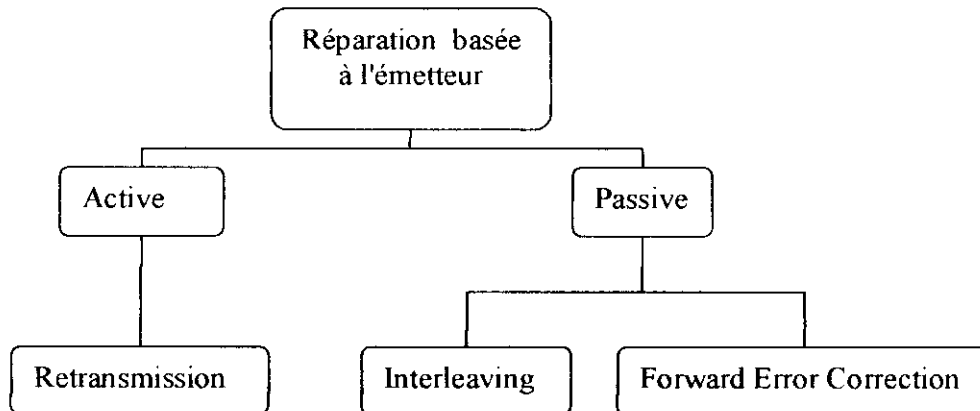


Figure 7. Classification des techniques de réparations basées à l'émetteur

Plusieurs techniques requérant la participation de l'émetteur du flux sonore afin de réparer les pertes de paquets ont été réalisées. Comme dans [12], la taxonomie de toutes ces techniques a été rassemblée dans la Figure 7 redessinée de [12].

Pour éviter les malentendus dans ce qui suit, nous avons distingué une trame de donnée d'un paquet de donnée. Une trame représente un intervalle du flux sonore. Un paquet peut contenir une ou plusieurs trames encapsulées afin d'être envoyées sur le réseau

Forward error correction

Ce schéma de recouvrement repose sur l'addition de donnée de réparation au flux sortant. De ces données, les paquets manquant peuvent être réparés le cas échéant.

Le principe est illustré dans la Figure 8 [12]

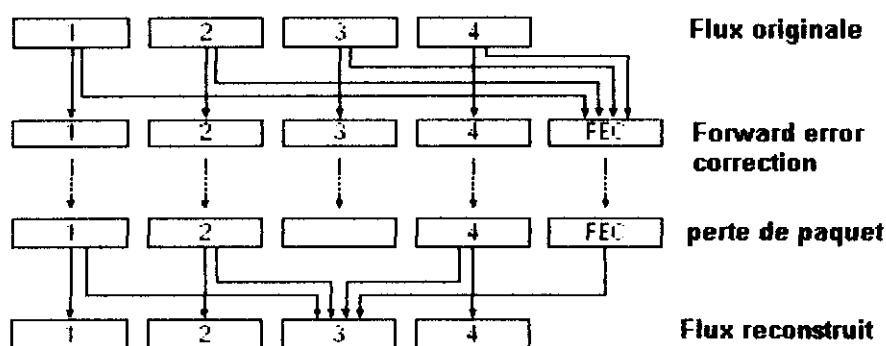


Figure 8. Forward error correction

Plusieurs avantages découlent de cette méthode, nous pouvons citer la faible demande en ressource de calcul et la simplicité de l'implémentation. En contre partie, cette technique impose un retard supplémentaire, une augmentation de la bande passante et une difficile implémentation au niveau du décodeur

Interleaving

Cette technique est très utile lorsque, les paquets contiennent plusieurs trames et le délai de bout-en-bout n'est pas important[13]. Avant transmission du flux, les trames sont ré-arrangées de telle manière que celles, initialement, adjacentes se retrouvent séparées dans le flux transmis, puis remises dans leur ordre original au niveau du récepteur.

En conséquence, les effets d'effacement de paquets, sont dispersés. La Figure 9 illustre un exemple où chaque paquet contient 4 trames [12].

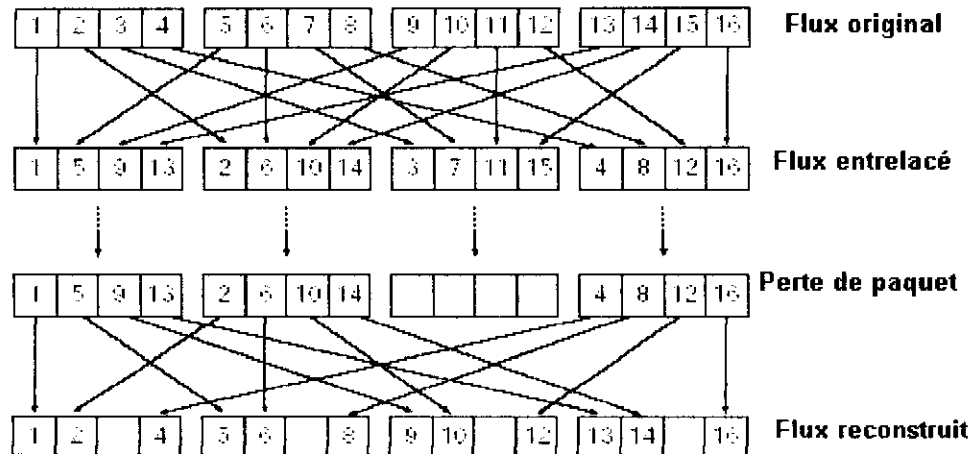


Figure 9. Interleaving

L'augmentation de latence constitue un sérieux inconvénient à l'utilisation de l'*interleaving* dans des applications interactives [12]. Alors que le maintien d'une bande passant stable avant et après son implémentation représente son avantage majeur.

Retransmission

Cette technique retransmet, simplement les paquets perdus, elle est difficilement applicable pour les applications interactives et pour lesquelles les délais de bout-en-bout sont réduits. Cependant pour des conditions de délai plus souple, cette méthode peut être implémentée.

II.4.2 Réparation basée au niveau du récepteur "Masquage d'erreur"

Comme pour la réparation basée à l'émetteur, plusieurs techniques, de masquage d'erreur, initiées par le récepteur d'un flux sonore, ont été réalisées. Ces techniques peuvent travailler soit en tandems avec celles entreprises au niveau de l'émetteur, soient seules.

Le masquage d'erreur repose sur le principe de remplacer les paquets perdus par des paquets similaires aux originaux. Ceci reste possible du fait de la similarité à court-terme du flux[12]. Une taxonomie des différentes techniques est illustrée dans la Figure 10.

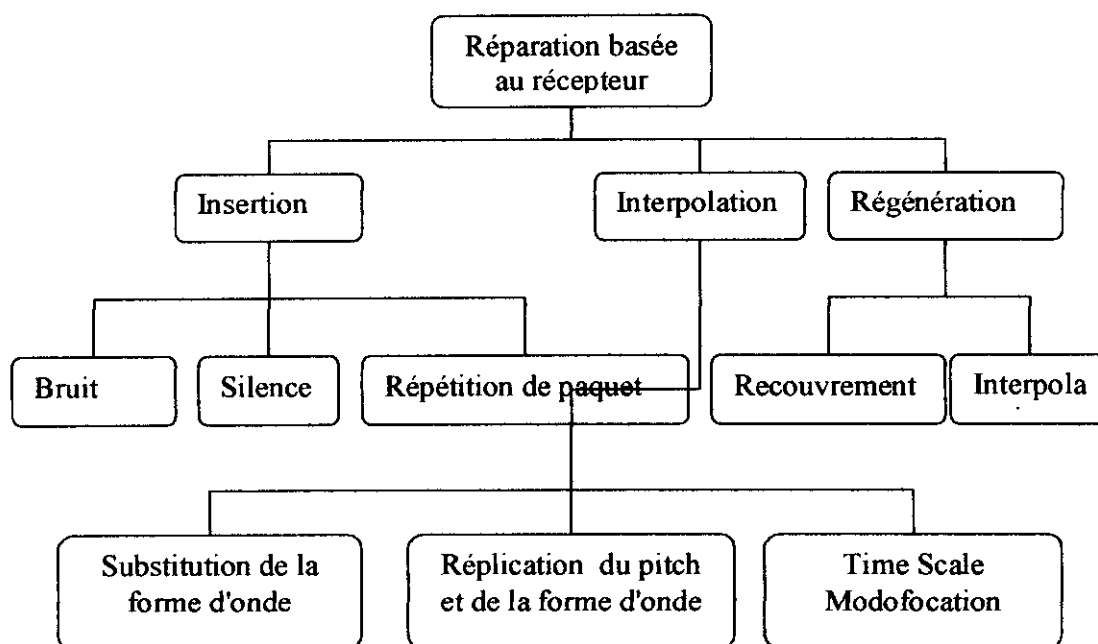


Figure 10. Classification des techniques de masquage d'erreur.

L'insertion

Cette technique répare les erreurs par l'insertion de paquets-remplaçants. Ces paquets peuvent être soit des silences, du bruit ou carrément la répétition du paquet précédent la perte. Ces techniques sont faciles à implémenter mais, avec l'exception de la répétition, offrent de faibles performances.

L'interpolation

Cette technique utilise un genre d'identification de paramètre et l'interpolation pour remplacer les paquets perdus. Elle est plus difficile à implémenter et requière plus de ressource de calcul que la méthode d'insertion, cependant, parce que cette technique prend en compte les changements des caractéristiques du signal, ses performances sont meilleures. Plusieurs techniques d'interpolation existent on en site:

II.4.2.1.1 Substitution de la forme d'onde

Cette technique met à profit le signal d'avant et, optionnellement, d'après perte pour trouver un signal convenable pour combler le pertes [12] [13].

II.4.2.1.2 Réplication du pitch et de la forme d'onde

Cette méthode est une amélioration de la méthode précédente II.4.2.1.1 et semble donner de meilleurs résultats, elle utilise, en plus, un algorithme de détection du pitch des deux cotés d'une perte de paquet [12] [5].

II.4.2.1.3 Time scale modification

Cette technique permet au signal audio, des deux cotés d'une perte, d'être étiré sur toute la longueur de la perte [12] [3]. Malgré une demande, en ressource de calcul, importante, cette méthode semble travailler mieux que les deux méthodes précédente.

La régénération

Cette technique exploite les connaissances sur les algorithmes de compression audio afin de retirer les paramètres audio manquants, et ainsi synthétiser les paquets perdus. En conséquence, cette méthode est forcément utilisée dans les codec. De plus, elle est, généralement, d'un point de vue complexité, coûteuse. Cependant, et grâce à une grande quantité d'information, cette méthode a de bonnes performances.

II.4.2.1.4 Interpolation de l'état transmis

Pour les codec fondés sur le codage par transformation ou par prédiction linéaire, il est possible que le décodeur interpole entre plusieurs états [12]. Cette méthode est utilisée, comme nous le verrons dans le CHAPITRE.III, dans le codec G.723.1.

II.4.2.1.5 Modèle

Le signal parole, d'un coté ou des deux cotés d'une perte, est ajusté à un modèle, qui est utilisé pour régénérer le signal de parole qui comblera la période de pertes.

II.5 Avantages de la technologie VOIP

Intégration de la voix et des données

L'intégration des trafics voix et données sera demandée par des logiciels multi-application. L'évolution inévitable sera vers des serveurs Web capable de gérer de la voix, des données et des images.

La simplification

Une infrastructure intégrée qui supporte toutes les formes de communication permet plus de standardisation et une gestion des équipements facilitée.

L'efficacité du réseau

Une bonne intégration de la voix et des données sur la même infrastructure charge le canal de communication des données efficacement, en découle une bonne utilisation de la bande passante. L'idée est de s'éloigner du plan où l'utilisateur se voit attribué une bande passante qu'il parle ou qu'il ne parle pas. Les réseaux de données n'opèrent pas de la même façon. C'est une grande économie quand on considère que 50% d'une conversation est silence. En outre, l'efficacité du réseau peut être améliorée en enlevant la redondance dans la parole.

Réduction du coût

Le réseau téléphonique traditionnel peut être mis hors circuit en utilisant la transmission via Internet qui va réduire les prix des appels à longues distances.

II.6 Activités de normalisation de la téléphonie IP5

La normalisation des technique de la téléphonie IP est en cours dans le cadre de nombreuses entités industrielles et régionales ainsi que d'organismes de normalisation tels que le Secteur de la normalisation des télécommunications de l'UIT (UIT-T), le Secteur des radiocommunications de l'UIT (UIT-R), l'Institut européen des normes de télécommunication (ETSI) et le Groupe d'étude sur l'ingénierie Internet (IETF).

Un exemple de normalisation dans le cadre de l'UIT est la série de Recommandations H.323 rédigées par la Commission d'études 16 de l'UIT-T. Le champ d'application des recommandations de cette série est très vaste: audioconférence, visioconférence multimédia, établissement et commande d'appel, gestion de la largeur de bande, interfaces entre différentes architectures de réseau. Il faut également signaler le Protocole d'initiation de session (SIP) défini par l'IETF pour la conférence, la téléphonie, la détection de présence, la notification d'événements et la messagerie instantanée. Dans certains cas, l'IETF et l'UIT-T ont collaboré directement à la normalisation de la

⁵ source ITU

téléphonie IP, c'est ainsi qu'ils ont mis au point le protocole commun H.248 (nom donné par l'UIT-T) ou MGCP (nom donné par l'IETF).

II.7 Conclusion

Les réseaux IP se montrent robuste et flexible aux pertes se produisant sur des données de type fax, images, bande son stockée. En effet, les données perdues peuvent être ré-émises à partir de la source et ré-introduites à la destination sans que l'intégralité ou que la qualité des fichiers reçus soit altérée. Cependant, dans des applications en temps réel, tel la VOIP, la procédure de ré-émission risque d'être impraticable tant les exigences temporelles sont strictes. Ainsi, lors d'envoi de séquences de paroles, c'est le protocole UDP/IP qui est utilisé. En effet, comme vu plus haut et contrairement au TCP/IP, ce protocole ne requière pas d'interactions initiales avec le destinataire, ni de ré-émission de paquets lors de pertes.

Les paquets peuvent être perdus de deux façons :

Remise à zéro, délibérée, des buffers des routeurs afin de gérer les liens encombrés.

Temps de vie dépassé à force de transiter par de nombreux nœuds.

Enfin, et puisque les conversations humaines ne tolèrent pas de long retard, les paquets de voix ayant été trop retardés sont considérés comme perdus.

Les pertes de paquets varient entre 0% et 20%. Ce chiffre peut augmenter selon le type de réseau, le nombre de saut et les quantités de congestion.

CHAPITRE.III Le codec standard G.723.16

Le présent chapitre est une vue globale sur le codec G.723.1, il aborde les caractéristiques générales du codec, ses principes de codage et de décodage avec et sans perte de trames.

III.1 Caractéristiques générales

III.1.1 Domaine d'application

La Recommandation G.723.1 spécifie une représentation codée qui peut être utilisée pour compresser la composante vocale ou un autre signal audio de services multimédias à très faible débit binaire. Lors de la conception de ce codeur, la principale application étudiée a été la visiophonie à très faible débit binaire, dans le cadre de la série des normes du type H.324

⁶ Tous les détails de ce chapitre peuvent être trouvés dans [14]

III.1.2 Débits binaires

Ce codeur possède deux débits binaires associés: 5,3 kbit/s et 6,3 kbit/s. Le débit supérieur correspond à une meilleure qualité. Le débit inférieur assure un niveau de qualité satisfaisant et offre aux concepteurs de systèmes un degré supplémentaire de flexibilité.

Ces deux débits sont des caractéristiques obligatoires du codeur et du décodeur. Il est possible de passer de l'un à l'autre à chaque limite de trame (toutes les 30 ms). Il est également possible de choisir le fonctionnement en débit variable au moyen d'une émission discontinue avec remplissage de bruit pendant les intervalles sans paroles.

III.1.3 Signaux d'entrée possibles

Ce codeur a été optimisé afin de représenter les paroles avec une qualité élevée aux débits indiqués ci-dessus, et au prix d'un faible degré de complexité. Les signaux acheminant des fréquences musicales et autres signaux audio ne sont pas restitués aussi fidèlement que les signaux vocaux mais ils peuvent être comprimés et décomprimés au moyen de ce codeur.

III.1.4 Délai

Ce codeur transforme les signaux de parole ou d'autres signaux audio en trames de 30 ms plus une période d'exploration de 7,5 ms, soit un délai algorithmique total de 37,5 ms. Tous les délais additionnels apportés par la réalisation et par le fonctionnement de ce codeur se composent des durées suivantes:

- i) délais de traitement des données dans le codeur et dans le décodeur;
- ii) délais de transmission dans la liaison de communication;
- iii) délais additionnels dus au stockage temporaire (tampons) pour le protocole de multiplexage.

III.2 Principes de codage

III.2.1 Description générale

Ce codeur est conçu de façon à fonctionner avec un signal numérique que l'on obtient en filtrant d'abord la bande téléphonique (Recommandation G.712) de l'entrée analogique, puis en l'échantillonnant à 8000 Hz puis en le convertissant linéairement en mots MIC de 16 éléments binaires pour l'entrée du codeur. Il convient d'effectuer la conversion inverse vers le mode analogique par des moyens similaires à la sortie du décodeur. D'autres caractéristiques d'entrée/sortie, comme celles qui sont spécifiées par la Recommandation G.711 pour les données MIC à 64 kbit/s, seront converties linéairement en mots MIC de 16 bits avant codage, ou seront reconverties dans leur format approprié après décodage linéaire des mots MIC de 16 bits. Le flux binaire allant du codeur au décodeur est défini dans le cadre de la présente Recommandation.

Le codeur est fondé sur les principes du codage prédictif linéaire (LPC) par analyse et synthèse, en vue de minimiser un signal d'erreur pondéré par une courbe de perception. Le codeur travaille sur des blocs (trames) contenant chacun 240 échantillons, ce qui correspond à une longueur de trame de 30 ms à 8 kHz. Chaque bloc passe d'abord par un filtre passe-haut afin d'éliminer la composante apériodique. Il est ensuite subdivisé en quatre sous-trames de 60 échantillons chacune.

Au moyen du signal d'entrée non traité, on calcule pour chaque sous-trame un filtre de codage prédictif linéaire (LPC) (*linear prediction coder*) du 10ème ordre. Le filtre LPC pour la dernière sous-trame est calculé au moyen d'un quantificateur des vecteurs prédictifs dédoublés (PSVQ) (*predictive split vector quantizer*). Les coefficients de codage LPC non quantifiés servent à construire le filtre de pondération perceptive à court terme qui est utilisé pour filtrer la trame complète afin de donner le signal vocal pondéré par la perception. Le signal vocal ainsi pondéré sert à calculer, toutes les deux sous-trames (120 échantillons), la période fondamentale en boucle ouverte, L_{OL} . Cette estimation de la hauteur tonale est effectuée sur des blocs de 120 échantillons. La période

fondamentale (inverse de la fréquence fondamentale) est recherchée dans l'étendue de 18 à 142 échantillons.

A partir de ce point, le signal vocal est traité à raison de 60 échantillons par sous-trame. Au moyen de la période fondamentale estimée, on construit un filtre de mise en forme du bruit harmonique. La combinaison du filtre de synthèse à codage LPC, du filtre de pondération perceptive des formants, et du filtre de mise en forme du bruit harmonique sert à créer une réponse impulsionnelle qui est utilisée pour d'autres calculs. La réponse impulsionnelle et l'estimation de la période fondamentale, L_{OL} , permettent de calculer un prédicteur de tonie en boucle ouverte. On fait appel à un prédicteur de tonie du cinquième ordre. La période fondamentale est calculée sous la forme d'une petite valeur différentielle de part et d'autre de l'estimation L_{OL} . La contribution du prédicteur de tonie est ensuite soustraite du vecteur cible initial. La période fondamentale et la valeur différentielle sont toutes deux transmises au décodeur.

Finalement, la composante apériodique de l'excitation est approchée. Pour le débit supérieur, on fait appel à l'excitation par quantification d'impulsions multiples selon le critère du maximum de vraisemblance (MP-MLQ). Pour le débit inférieur, on fait appel à l'excitation par séquences codées à structure algébrique (ACELP).

Le schéma fonctionnel du codeur est représenté sur la Figure 11.

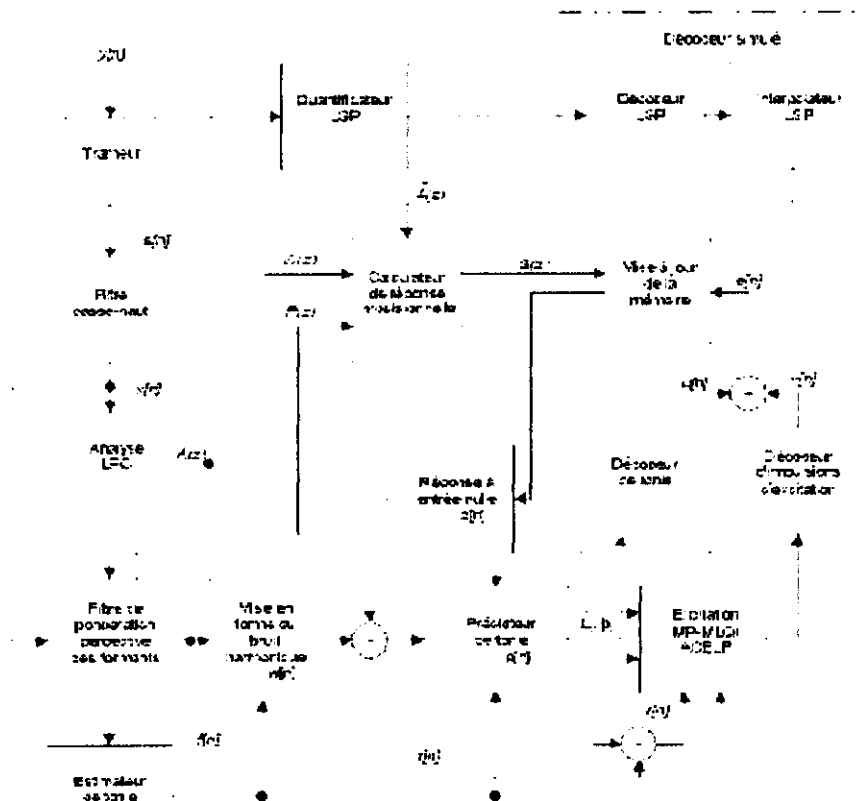


Figure 11. G.723.1/ Organigramme du codec

III.2.2 Quantificateur des coefficients de paires de raies spectrales (LSP)

On effectue d'abord une petite expansion de bande additionnelle (7,5 Hz). Puis le filtre prédictif linéaire résultant $A(z)$ est quantifié au moyen d'un quantificateur de vecteurs prédictifs dédoublés (PSVQ) comme suit.

Les coefficients LP, $\{a_j\}_{j=1..10}$, sont convertis en coefficients LSP, $\{p'_j\}_{j=1..10}$ par recherche sur le cercle unité et interpolation des passages à zéro.

La composante apériodique à long terme, p_{DC} , est soustraite des coefficients LSP, p'_j , et un nouveau vecteur LSP sans composante apériodique, p , est obtenu.

Chapitre III : Le codec G.723.1

Un prédicteur fixe (non adaptatif) du premier ordre, $b = (12/32)$, est appliqué au vecteur LSP déjà décodé \tilde{p}_{n-1} , ce qui donne le vecteur LSP prédit sans composante apériodique, \bar{p}_n , ainsi que le vecteur d'erreur LSP résiduelle, e_n , au temps (à la trame) n .

$$p_n^T = [p_{1,n} \ p_{2,n} \dots \ p_{10,n}] \quad (3.1)$$

$$\bar{p}_n^T = [\bar{p}_{1,n} \cdot \bar{p}_{2,n} \cdot \Lambda \ \bar{p}_{10,n}] \quad (3.2)$$

$$\tilde{p}_n^T = [\tilde{p}_{1,n} \cdot \tilde{p}_{2,n} \cdot \Lambda \ \tilde{p}_{10,n}] \quad (3.3)$$

$$e_n = p_n - \bar{p}_n \quad (3.4)$$

Le vecteur des coefficients LSP non quantifiés, p'_n , le vecteur des coefficients LSP quantifiés, \tilde{p}_n , et le vecteur des erreurs LSP résiduelles, e_n , sont subdivisés en 3 sous-vecteurs de dimensions 3, 3 et 4 respectivement. Chaque $m^{\text{ième}}$ sous-vecteur est quantifié vectoriellement au moyen d'un répertoire de séquences codées sur 8 éléments binaires. L'index, l , de l'entrée appropriée dans ce répertoire pour sous-vecteurs est choisi de façon à minimiser le critère d'erreur $E_{l,m}$.

$$p_{i,m}^T = [P_{1,i,m} \ P_{2,i,m} \ \dots \ P_{K_i,i,m}], \quad K_i = \begin{cases} 3, & m=0 \\ 3, & m=1 \\ 4, & m=2 \end{cases} \quad (3.5)$$

$$\tilde{p}_{i,m}^T = [\tilde{P}_{1,i,m} \ \tilde{P}_{2,i,m} \ \dots \ \tilde{P}_{K_i,i,m}], \quad \begin{matrix} 0 \leq m \leq 2 \\ 1 \leq i \leq 256 \end{matrix} \quad (3.6)$$

$$p'_i = p_i + P_{LPC} \quad (3.7)$$

$$\tilde{p}_{i,m} = \bar{p}_{i,m} - P_{LPC} e_{i,m} \quad \begin{matrix} 0 \leq m \leq 2 \\ 1 \leq i \leq 256 \end{matrix} \quad (3.8)$$

$$E_{l,m} = (p'_{i,m} - \tilde{p}_{i,m}^T W_m (p'_{i,m} - \tilde{p}_{i,m}^T)) \quad \begin{matrix} 0 \leq m \leq 2 \\ 1 \leq i \leq 256 \end{matrix} \quad (3.9)$$

où $e_{l,m}$ est la $l^{\text{ième}}$ entrée du $m^{\text{ième}}$ répertoire de codes LSP résiduels dédoublés et où W_m est une matrice de pondération par filtrage diagonal, déterminée d'après le vecteur p' des coefficients LSP non quantifiés, les pondérations étant définies comme suit:

$$\begin{aligned}
 W_{jk} &= \frac{1}{\min\{p_j^i - p_{j-1}^i, p_{j-1}^i - p_j^i\}} \quad 0 \leq j \leq 10 \\
 W_{1,1} &= \frac{1}{p_1^i - p_0^i} \\
 W_{10,10} &= \frac{1}{p_{10}^i - p_0^i}
 \end{aligned}
 \tag{3.10}$$

Les index ainsi sélectionnés sont transmis au canal.

III.3 Principes de décodage

III.3.1 Description générale

Comme le codeur, le décodeur fonctionne trame par trame. Il décode d'abord les index de codage LPC quantifiés, puis construit le filtre de synthèse des coefficients LPC. Pour chaque sous-trame, on décode aussi bien les séquences d'excitation par répertoire adaptatif que les séquences d'excitation par répertoire fixe, avant de les injecter dans le filtre de synthèse. Le postfiltre adaptatif se compose d'un postfiltre de formants et d'un postfiltre fondamental anticipée/différée. Le signal d'excitation envoyé dans le postfiltre fondamental, qui à son tour attaque le filtre de synthèse dont la sortie va vers le postfiltre de formants. Un module de normalisation du gain maintient l'énergie du signal au niveau d'entrée du postfiltre de formants.

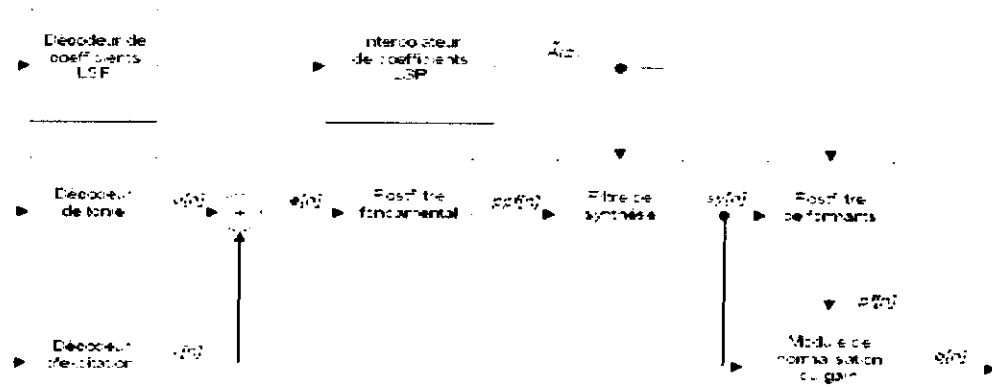


Figure 12. G.723.1/Organigramme du décodeur.

III.3.2 Décodeur des coefficients LSP

Le décodage des coefficients LSP est effectué comme suit:

Tout d'abord, les trois sous-vecteurs, $\{e_{m,n}\}_{m=0,2}$, sont décodés afin de former un vecteur du dixième ordre, \tilde{e}_n ;

Le vecteur prédit \tilde{p}_n , est ajouté au vecteur décodé, \tilde{e}_n , et au vecteur des composantes a périodiques, p_{DC} , afin de former le vecteur des coefficients LSP décodés, \tilde{p}_n ;

- 1) une vérification de stabilité est effectuée sur le vecteur des coefficients LSP décodés, \tilde{p}_n , afin de s'assurer que ce vecteur respecte la condition d'ordre suivante:

$$\tilde{F}_{j+1} - \tilde{F}_j \geq \Delta_{min} \quad 1 \leq j \leq 9 \quad (3.11)$$

Δ_{min} est égal à 31,25 Hz. Si cette condition de stabilité (Equation (3.11)) n'est pas vérifiée pour \tilde{p}_i et pour \tilde{p}_{i+1} , on modifie comme suit \tilde{p}_i et \tilde{p}_{i+1} :

$$F_{avg} = (\tilde{F}_i + \tilde{F}_{i+1}) / 2$$

$$\tilde{F}_i = F_{avg} - \Delta_{min} / 2$$

$$\tilde{F}_{i+1} = F_{avg} + \Delta_{min} / 2 \quad (3.11)$$

Cette modification est effectuée jusqu'à ce que la condition (Equation (3.11)) soit vérifiée. Si, à l'issue de 10 itérations, cette condition de stabilité n'est toujours pas vérifiée, on fait appel au vecteur LSP précédent.

III.3.3 Interpolation des coefficients LSP

Une interpolation linéaire est effectuée entre le vecteur LSP décodé, \tilde{p}_n , et les vecteur LSP précédent, \tilde{p}_{n-1} , pour chaque sous-trame. Quatre vecteurs LSP interpolés, $\{\tilde{p}_i\}_{i=0,3}$, sont convertis en vecteurs LPC, $\{\tilde{a}_i\}_{i=0,3}$.

$$p_{ii} = \begin{cases} 0.75\tilde{p}_{i-1} + 0.25\tilde{p}_i, & i = 0 \\ 0.5\tilde{p}_{i-1} + 0.5\tilde{p}_i, & i = 1 \\ 0.25\tilde{p}_{i-1} + 0.75\tilde{p}_i, & i = 2 \\ \tilde{p}_i, & i = 3 \end{cases} \quad (3.13)$$

$$a_i^T = [a_{i1} a_{i2}]^T, \quad 0 \leq i \leq 3 \quad (3.14)$$

Le filtre de synthèse des coefficients LPC quantifiés est utilisé pour restituer le signal vocal décodé. Il est défini comme suit:

$$\tilde{A}_{i,t} = \frac{1}{1 - \sum_{j=1}^3 a_{ij} z^{-j}}, \quad 0 \leq i \leq 3 \quad (3.15)$$

III.3.4 Traitement par interpolation des trames

Ce codeur a été conçu pour résister aux effacements de trame indiqués. Une stratégie de masquage des erreurs par effacement de trames a été intégrée dans le décodeur. Il est toutefois nécessaire que cette stratégie soit déclenchée par une indication externe du fait que le flux binaire a été effacé pour la trame actuelle. Etant donné que le codeur a été conçu pour résister à des erreurs sporadiques, aucun mécanisme de correction n'est prévu pour les erreurs aléatoires sur les bits. Si un effacement de trame s'est produit, le décodeur passe du mode de décodage normal au mode de masquage d'effacement de trame. La procédure d'interpolation de trame est lancée indépendamment pour les coefficients LSP et pour le signal résiduel.

III.3.4.1 Interpolation des coefficients LSP

Le décodage des coefficients LSP en mode d'interpolation de trame se déroule comme suit:

le vecteur \tilde{e}_n est mis à zéro;

le vecteur prédit, \tilde{p}_n , est ajouté au vecteur, \tilde{e}_n , et au vecteur a périodique, p_{DC} , afin de former le vecteur LSP décodé, \tilde{p}_n . Un prédicteur non adaptatif différent est utilisé pour prédire le vecteur \tilde{p}_n : $b_e = 23/32$.

A partir de là, le décodage des coefficients LSP se poursuit comme indiqué au III.3.2, sauf que $\Delta_{\min} = 62,5$ Hz au lieu de 31,25 Hz.

III.3.4.2 Interpolation des résidus

On effectue l'interpolation des résidus du signal de deux façons différentes, selon la dernière bonne trame trouvée avant la trame effacée. Cette trame est vérifiée par un classificateur d'éléments voisés et non voisés. Ce classificateur est fondé sur une fonction de maximisation d'une intercorrélacion. Les 120 derniers échantillons de la trame sont intercorrélés avec le délai $L_2 \pm 3$. L'index qui atteint la valeur maximale d'intercorrélacion est retenu comme éventuel index d'interpolation. Puis on teste le gain prédictif du meilleur vecteur. Si ce gain est supérieur à 0,58 dB, la trame est déclarée de type voisé, sinon de type non voisé. Le classificateur revient à 0 dans le cas d'une trame non voisée et à la valeur tonale estimée dans le cas d'une trame voisée. Si la trame a été déclarée non voisée, on sauvegarde la moyenne des index de gain pour les sous-frames 2 et 3.

Si la trame actuelle était marquée comme effacée et que la trame précédente ait été classifiée comme non voisée, l'excitation de la trame actuelle est produite au moyen d'un générateur de nombres aléatoires uniformes, dont la sortie est normalisée au moyen de la valeur de gain précédemment calculée. Dans le cas d'une trame voisée, la trame actuelle est régénérée par une excitation périodique dont la fondamentale est égale à la valeur fournie par le classificateur. Si l'état d'effacement de trame se maintient pour les deux trames suivantes, le vecteur régénéré est affaibli de 2,5 dB supplémentaires dans chaque trame. Après 3 trames interpolées, la sortie est totalement supprimée.

CHAPITRE.IV Etudes, simulations et résultats

Ce chapitre rassemble et synthétise, comme son titre l'indique, les études, les simulations et les résultats de tous les travaux accomplis lors de ce projet. Son ordonnancement ne suit aucune règle précise si ce n'est la progression temporelle des travaux accomplis. A travers ce chapitre nous avons, tout d'abord, posé la problématique, cherché, ensuite, son origine et enfin proposé des solutions respectueuses des normes exigées.

IV.1 Propagation de l'erreur lors d'effacements de trames

Comme nous avons vu au chapitre III, dans le codec G.732.1, la quantification du vecteur LSP, p' , se fait par la méthode PSVQ, cette dernière met à profit la bonne corrélation qui existe entre les trames successives (corrélation inter-trames) afin d'obtenir une bonne quantification.

Cependant, lors de l'effacement d'une trame, et à cause de la méthode de quantification elle-même (inter-trame), l'erreur commise dans le masquage ne va pas se limiter à la trame elle-même mais se propagera, au moins, aux 3 trames suivantes.

Ce phénomène est illustré dans la, Figure 13, où est tracé, pour comparaison, les distorsions spectrales d'une séquence de parole codée avec et sans perte. Aussi, nous observons qu'après chaque effacement de trame signalé par des impulsions en vert, la

Chapitre IV : Etudes, simulations et résultats.

courbe en points tillés, représentant la distorsion avec perte, est surélevée par rapport à la courbe en continu (séquence sans pertes), signifiant ainsi, une plus grande distorsion pour les 3 à 4 trames suivantes.

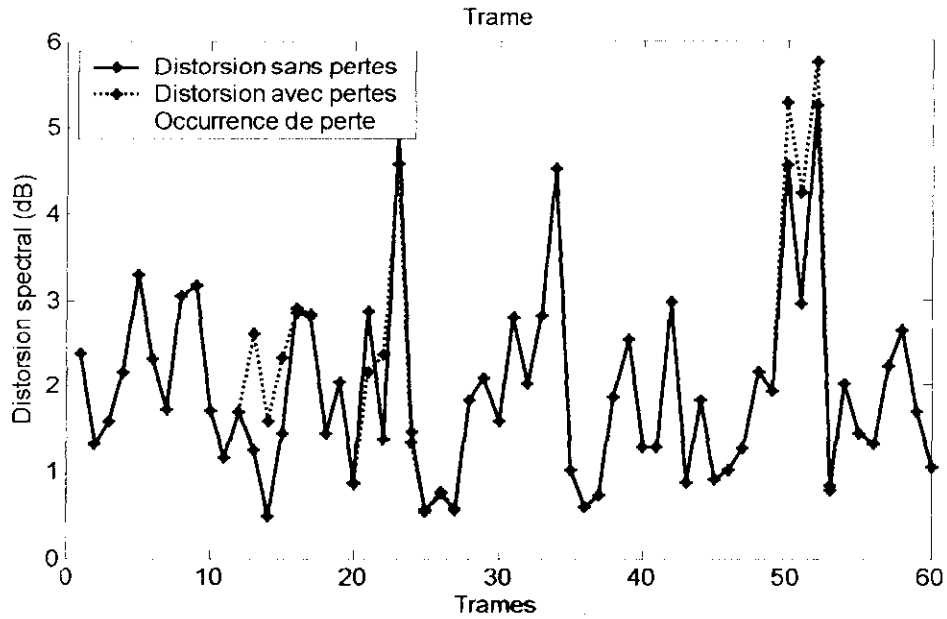


Figure 13. Propagation de l'erreur dans le codage inter-trame (PSVQ).

Puisque cette propagation d'erreur est due essentiellement à la façon dont les *LSP* sont quantifiés (quantification inter-trame), une nouvelle méthode de quantification agissant sur chaque trame séparément et exploitant la corrélation existante entre chaque élément d'une même trame (quantification intra-trame) pourrait empêcher la propagation des erreurs et donnerait, ainsi, de bons résultats.

Cependant cette méthode devra respecter trois impératifs :

La qualité.

Le débit.

La complexité.

IV.2 Quantification Intra-trame des LSP

Avant d'entamer la conception du quantificateur, une petite étude statistique nous donnera de précieuses informations sur les données à quantifier (i.e : coefficients *LSP*).

IV.2.1 Base de données et étude statistique

IV.2.1.1 Base de données

Une « bonne » base de donnée reste la condition sine qua non pour la validation d'un quelconque résultat trouvé. Dans nos travaux, nous avons utilisé une base de donnée mondialement reconnue, à savoir, « *The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT), Training and Test Data* ».

Le corpus TIMIT, formé de paroles lues, a été conçu afin de fournir des données de parole, pour l'acquisition de connaissance acoustique-phonétique et pour le développement et l'évaluation de systèmes automatiques de reconnaissance de parole.

IV.2.1.1.1 Répartition des orateurs

630 orateurs, provenant de régions pratiquant les 8 principaux dialectes des États-Unis, ont participé à l'élaboration du corpus TIMIT. Chacun d'eux lit 10 phrases distinctes, totalisant ainsi, 6300 phrases.

| (région) | Male | Femelle | Total |
|----------|-----------|-----------|------------|
| 1 | 31 (63%) | 18 (27%) | 49 (8%) |
| 2 | 71 (70%) | 31 (30%) | 102 (16%) |
| 3 | 79 (67%) | 23 (23%) | 102 (16%) |
| 4 | 69 (69%) | 31 (31%) | 100 (16%) |
| 5 | 62 (63%) | 36 (37%) | 98 (16%) |
| 6 | 30 (65%) | 16 (35%) | 46 (7%) |
| 7 | 74 (74%) | 26 (26%) | 100 (16%) |
| 8 | 22 (67%) | 11 (33%) | 33 (5%) |
| Total | 438 (70%) | 192 (30%) | 630 (100%) |

Table 3. Répartition des orateurs de la base de donnée TIMIT.

Chapitre IV : Etudes, simulations et résultats.

IV.2.1.1.2 Test et Training

La base de donnée a été divisée en 2 parties, une pour l'entraînement (Training) et l'autre pour le test.

Les données de test ont un noyau composé de 24 orateurs, 2 males et 1 femelle pour chaque région, récitant 192 phrases.

Il est à noter que les séquences de parole de la base de données sont échantillonnées à 16KHz et qu'il comporte un entête. Donc, pour adapter ces signaux de base aux signaux d'entrés du codec G.723.1, il a fallu effacer les entêtes, ré-échantillonner à 8KHz, et regrouper les phrases pour faciliter le travail.

La Table 4 rassemble les caractéristiques :

| | Longueurs (Temps) | Nbr Trame de 30ms |
|----------------------|------------------------|-------------------|
| Entraînement | 3 ^h 08'42'' | 377403 |
| Test complet | 1 ^h 08'56'' | 13787 |
| Noyau du test | 0 ^h 09'43'' | 19433 |

Table 4. Longueurs de la base de données.

Les données d'entraînement ont été utilisées dans les études statistiques et la conception de tous les dictionnaires. Alors que ces mêmes dictionnaires ont été testés par le noyau du test.

IV.2.1.2 Etude statistique

Cette étude statistique comprend des tracés d'histogramme normalisé ⁷ (Figure 14), des calculs de l'écart type et de la dynamique⁸ (Table 5) pour chaque élément du vecteur p' . Enfin une matrice regroupant les coefficients de corrélations linéaires $\rho(i,j)$ entre chaque paire des éléments du vecteur LSP suivant la formule de l'Equation (4.1) .

⁷ Fréquences d'apparitions sont normalisées par max.

⁸ Dynamique de $s = \max(s) - \min(s)$,

$$\rho[i, j] = \frac{\text{cov}(p'[i], p'[j])}{\sqrt{\text{cov}(p'[i], p'[i])} * \sqrt{\text{cov}(p'[j], p'[j])}} \quad (4.1)$$

Où $\text{cov}(p'[i], p'[j])$

Est la covariance entre les élément i et j du vecteur LSP, p' , à travers toutes les observations faites (377403 trames).

A noter que les LSP utilisés dans cette étude ont été calculé par le standard ITU G.732.1 en utilisant les fichiers d'entraînements (Train) de la base de données TIMIT, puis, ont été normalisés par π .

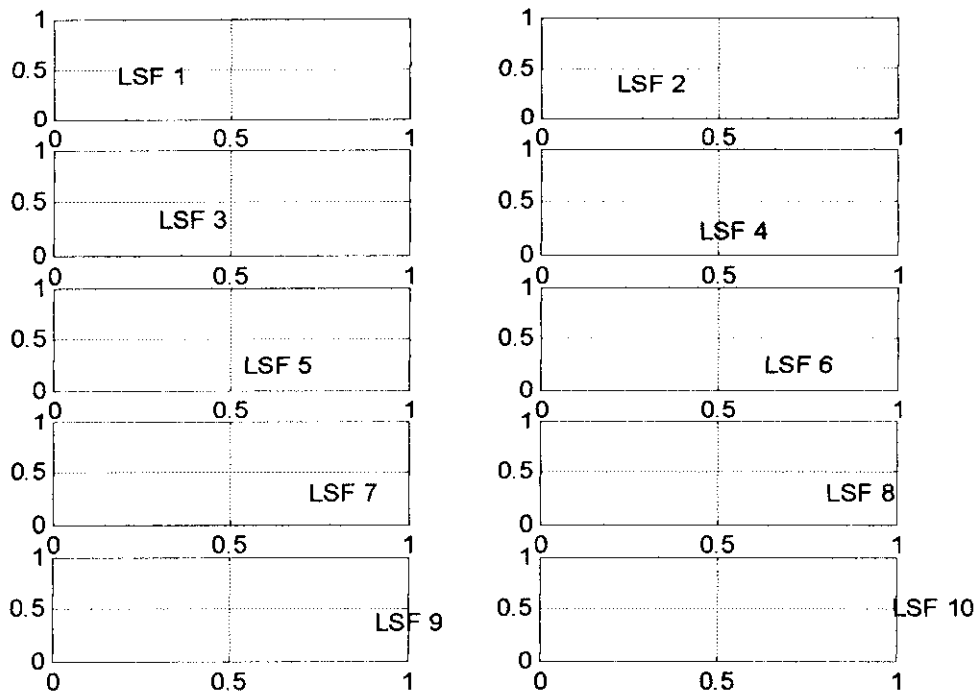


Figure 14. Histogramme normalisé représentant les variations des coefficients LSP .

| LSP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|
| Ecart type* 10^{-4} | 267 | 353 | 465 | 547 | 645 | 603 | 526 | 469 | 386 | 285 |
| Dynque* 10^{-4} | 2001 | 2852 | 3458 | 3753 | 3890 | 4022 | 3961 | 3707 | 2982 | 2277 |

Chapitre IV : Etudes, simulations et résultats.

Table 5. Ecart type et dynamique des coefficient LSP.

| LSP | 1 | 2 | 3 | 4 | 8 | 6 | 7 | 8 | 9 | 10 |
|-----|--------|-------|-------|-------|--------|--------|--------|-------|-------|--------|
| 1 | 1 | 0.585 | 0.276 | 0.088 | -0.065 | -0.118 | -0.018 | 0.009 | 0.030 | -0.044 |
| 2 | 0.585 | 1 | 0.751 | 0.395 | 0.237 | 0.196 | 0.267 | 0.342 | 0.244 | 0.178 |
| 3 | 0.276 | 0.751 | 1 | 0.651 | 0.401 | 0.439 | 0.382 | 0.496 | 0.371 | 0.256 |
| 4 | 0.088 | 0.395 | 0.651 | 1 | 0.697 | 0.511 | 0.414 | 0.361 | 0.404 | 0.186 |
| 5 | -0.065 | 0.237 | 0.401 | 0.697 | 1 | 0.722 | 0.508 | 0.411 | 0.256 | 0.204 |
| 6 | -0.118 | 0.196 | 0.439 | 0.511 | 0.722 | 1 | 0.709 | 0.582 | 0.402 | 0.225 |
| 7 | -0.018 | 0.267 | 0.382 | 0.414 | 0.508 | 0.709 | 1 | 0.687 | 0.466 | 0.345 |
| 8 | 0.009 | 0.342 | 0.496 | 0.361 | 0.411 | 0.582 | 0.687 | 1 | 0.553 | 0.335 |
| 9 | 0.030 | 0.243 | 0.371 | 0.404 | 0.256 | 0.401 | 0.466 | 0.553 | 1 | 0.487 |
| 10 | -0.044 | 0.178 | 0.256 | 0.186 | 0.204 | 0.225 | 0.345 | 0.335 | 0.487 | 1 |

Table 6. Coefficients de corrélation linéaire des coefficients LSP.

De la Figure 14 nous observons que :

Les tracés des éléments 1, 2,9 et 10 sont très proches du tracé d'une distribution Gaussienne.

Les éléments intermédiaires 3,4 jusqu'a 8 ont une grande dynamique

De plus, la matrice de corrélation renvoie d'assez grandes valeurs sur la diagonale au-dessus de la diagonale centrale (surlignée en jaune) indiquant une grande corrélation entre chaque paire d'élément successive.

Afin de mieux illustrer cette dernière propriété, nous avons calculé les différences entre chaque paire d'éléments du vecteur LSP, p' , selon les Equation (4.2), et refait la même étude statistique

$$\begin{cases} \Delta p_1 = p'_1 \\ \Delta p_i = p'_i - p'_{i-1}, K \text{ pour } i = 2,10 \end{cases} \quad (4.2)$$

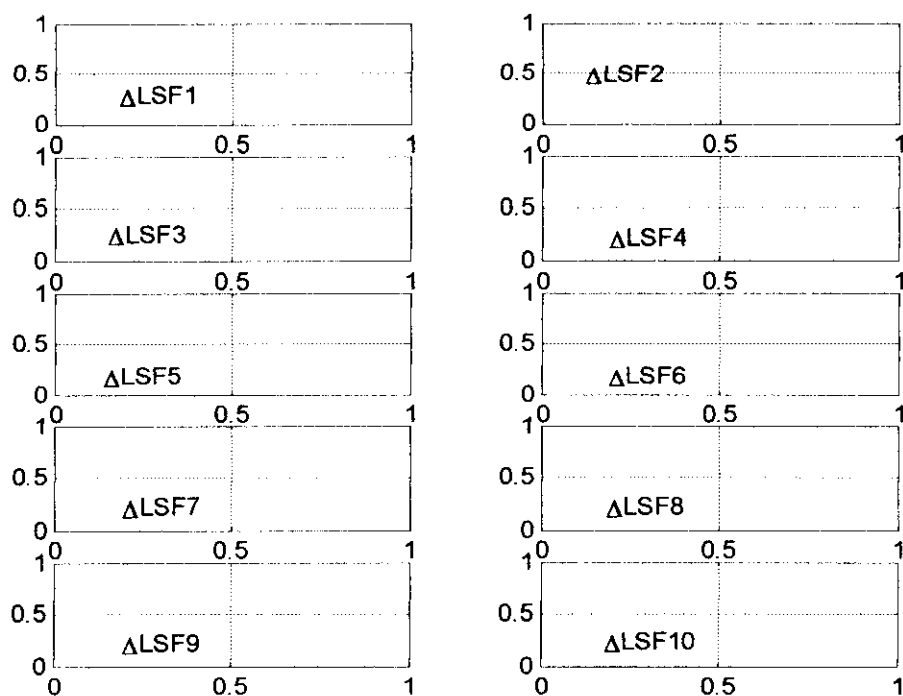


Figure 15. Histogramme normalisé des coefficients ΔLSP

| ΔLSP | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|------|------|------|------|------|------|------|------|------|------|
| Ecart type* 10^{-4} | 267 | 293 | 307 | 429 | 472 | 466 | 436 | 397 | 410 | 351 |
| Dynque* 10^{-4} | 2001 | 2468 | 2655 | 3257 | 3256 | 3291 | 3102 | 3129 | 2981 | 2624 |

Table 7. Ecart type et dynamique des coefficients ΔLSP .

Comme la Figure 15 et la Table 7 le montrent les ΔLSP ont une dynamique plus petite que celle des LSP et se prêteraient, donc, mieux à la quantification.

La première méthode de quantification essayée est la quantification scalaire des ΔLSP nommés DSQ.

IV.2.2 Quantification scalaire des différences (DSQ)

Cette méthode a été décrite en [15], elle utilise la technique utilisée dans la DPCM,

Cette méthode requière l'utilisation d'un décodeur intégré dans le codeur afin d'empêcher la propagation des erreurs au fur et à mesure de la quantification.

IV.2.2.1 Procédure de quantification

1. quantifier $\Delta p_1 \xrightarrow{Q_1} \tilde{\Delta p}_1$
2. calculer $\Delta p_i = p'_i - \tilde{p}'_{i-1} K$ pour $i = 2, 10$
3. quantifier $\Delta p_i \xrightarrow{Q_i} \tilde{\Delta p}_i$ pour $i = 2, 10$
4. si $i > 10$ stop, Sinon reconstruire

$$\tilde{p}'_i = \tilde{\Delta p}_i + \tilde{p}'_{i-1} \dots \text{pour } i = 2, 10 \tag{4.3}$$

Puis aller à 2.

IV.2.2.2 Allocation de bits et conception des dictionnaires

Concernant l'allocation des bits nous reprenons l'allocation faite par [15] et illustrée dans la Table 8

| Index | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Total |
|-------|---|---|---|---|---|---|---|---|---|----|-------|
| Bits | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 27 |

Table 8 : Allocation optimale de bits pour la DSQ.

Les dix dictionnaires ont été conçus par la méthode de Lloyd qui prend en compte la distribution statistique du signal. Ils sont illustrés dans la Figure 16

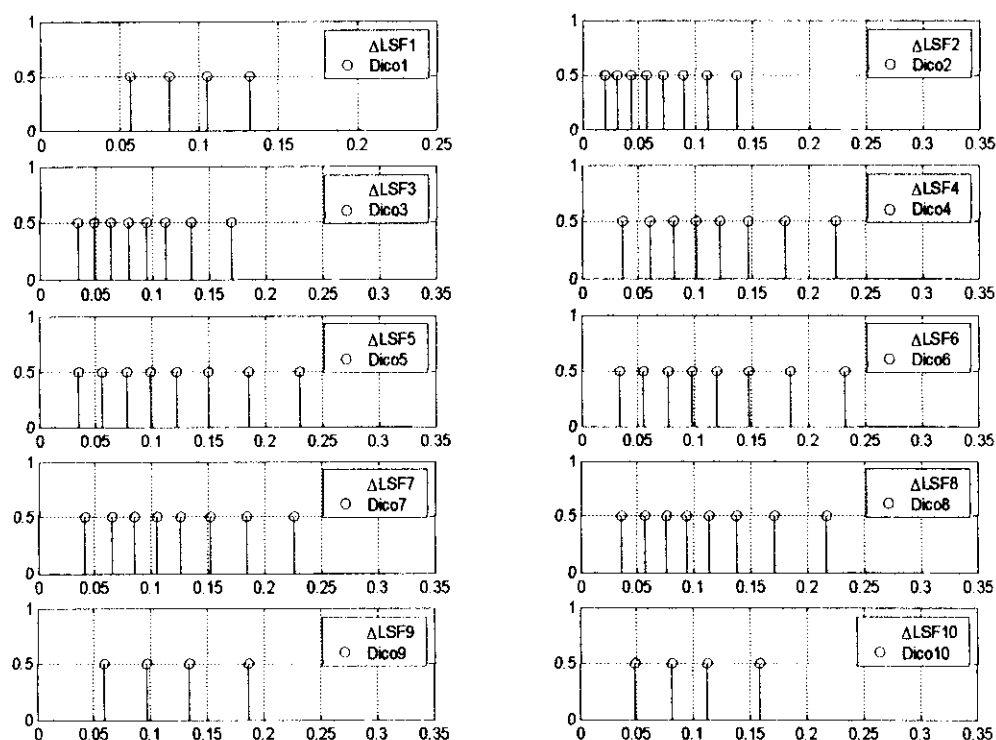


Figure 16. Distribution des ΔLSP avec leurs dictionnaires.

Une fois le quantificateur implémenté, nous avons refait le test sur la propagation de l'erreur utilisant un masquage, de trames perdues, par interpolation⁹ linéaire entre la dernière trame reçue avant perte et la première reçue après perte.

IV.2.2.3 Propagation de l'erreur entre trames successives

Comme nous le remarquons sur la Figure 17. (Trame), et contrairement à ce qu'on pouvait attendre, l'erreur se propage, comme même, à une seule trame après la trame perdue.

⁹ Plus de détail sur cette méthode de masquage sont donnés plus loin.

Chapitre IV : Etudes, simulations et résultats.

Cette synchronisation d'une trame s'explique par le fait que le codec, comme détaillé plus haut, ne quantifie que la dernière sous-trame de la trame et que les 3 autres sous-trames sont interpolées entre la dernière sous-trame de la trame précédente (perdue) et la dernière sous-trame de la trame présente (reçue). donc, une erreur dans une trame se propagera forcément vers les 3 premières sous-trames de la trame suivante, provoquant ainsi une erreur sur toute la trame. Toute fois dans Figure 17 (sous-trame 4) on remarque bien qu'il n'y a aucune propagation d'erreur de sous-trame 4 vers sous-trame 4 ce qui induit que l'erreur ne peut se propager au-delà d'une seule trame après la trame perdue.

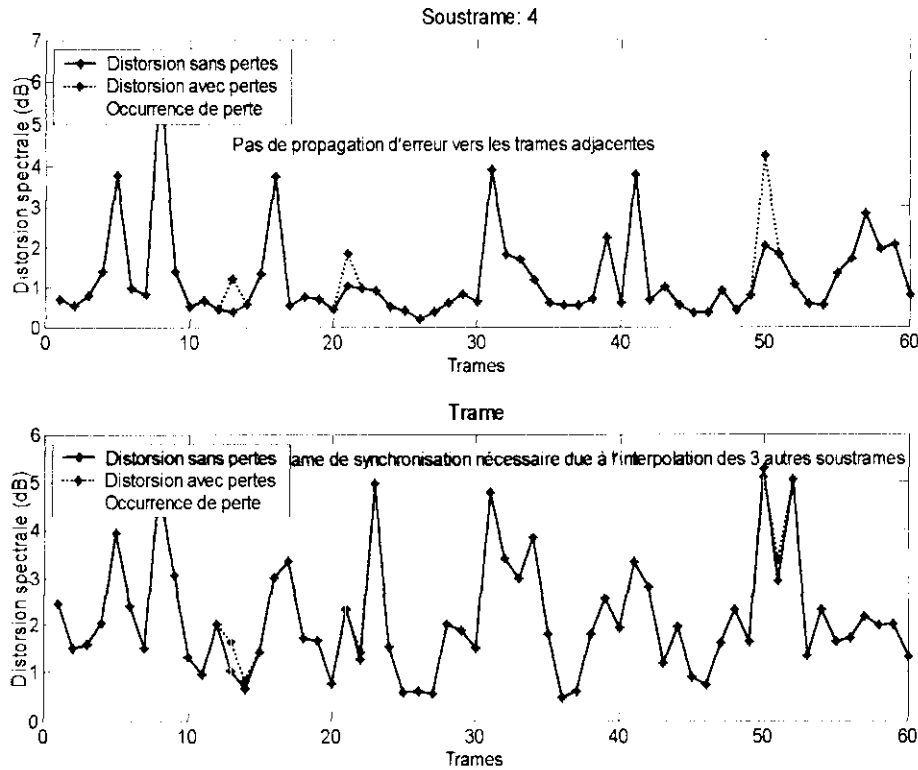


Figure 17. Propagation de l'erreur dans le codage intra-trame (DSQ)

Malgré la propagation d'erreur, qui se fait au maximum, sur une seule trame, dans les cas de transmissions avec pertes, ce genre de codage (Intra-trame) reste meilleur que le codage du standard ITU G.732.1 (Inter-trame) où la propagation se fait, au minimum, sur 3 trames comme vu plus haut.

Chapitre IV : Etudes, simulations et résultats.

Maintenant que nous sommes sûrs de l'efficacité du codage intra-trame, nous nous focalisons sur les performances du quantificateur.

IV.2.2.4 Performance du quantificateur

Cette méthode, comme les autres, sera testé par le noyau de test de la base de données TIMIT. Les résultats de cette méthode DSQ et du standard ITU G.723.1 sont regroupés dans la Table 9

| | DS.Moy (dB) | 2-4 dB | > 4dB | EMBSD |
|------|-------------|--------|-------|-------|
| PSVQ | 1.84 | 32.99 | 4.67 | 1.551 |
| DSQ | 1.81 | 32.99 | 4.44 | 1.498 |

Table 9. Performance de la méthode DSQ

Les résultats de la DSQ sont très proches voire légèrement meilleurs que ceux du standard ITU, néanmoins cette méthode requière 27 bits pour la quantification des coefficients *LSP*, soit 3 bits de plus par trame par rapport au standard ITU G.732.1 (24 bits), soit une augmentation de débit de 100 bits/s.

La DSQ revoie de bons résultats mais introduit un excédent de débit de 100bits/s.

Une bonne méthode pour réduire le débit sans altérer la qualité de la quantification est le passage à la quantification vectorielle.

IV.2.3 Quantification vectorielle

Pour des raisons évidentes de complexités, le vecteur de 10 éléments à quantifier, a été divisé en plusieurs sous-vecteurs.

Tous les dictionnaires utilisés dans les différentes méthodes de quantifications ont été conçus par approche statistique et calculés grâce à l'algorithme LBG.

Pour, au moins, égaler les performances du standard la nombre total de bit réservés est de 24 bits.

La première méthode utilisée à été nommée DSVQ

IV.2.4 Quantification vectorielle des différences en sous-vecteur (DSVQ).

Cette méthode est la plus simple, elle quantifie chaque sous-vecteur séparément et n'intègre pas de décodeur au niveau du codeur.

IV.2.4.1 Procédure de quantification

La procédure de quantification est la suivante :

Au codeur

1. on forme les différences selon les Equation (4.2)
2. on quantifie chaque sous-vecteur séparément par le dictionnaire qui lui est propre :

$$\Delta p^j \xrightarrow{Q^j} \tilde{\Delta p}^j \dots \text{pour } j = 1, j_{\max}$$

Au décodeur

1. à partir de l'index on récupère les valeurs quantifiées pour chaque sous-vecteur.
2. on forme les *LSP*

$$\tilde{p}'_i = \tilde{\Delta p}_i + \tilde{p}'_{i-1} \dots \text{pour } i = 2, 10 \tag{4.4}$$

IV.2.4.2 Découpage et allocation de bits

Pour tester les performances de cette méthode nous avons réalisé qu'un seul découpage (3, 3, 4) en octroyant pour chaque sous vecteur 8 bits.

IV.2.4.3 Performance et analyse du quantificateur

| | DS.Moy (dB) | 2 – 4 dB | > 4dB | EMBSD |
|-------------|--------------------|-----------------|-----------------|--------------|
| PSVQ | 1.84 | 32.99 | 4.67 | 1.551 |
| DSVQ | 2.23 | 40.82 | 16.76 | Non Calculé |

Table 10. Découpage et allocation de bits pour la DSVQ.

Comme on peut la voir dans la Table 10, la DSVQ engendre presque 0.4 dB de distorsion de plus que la méthode standard. Cette détérioration du signal n'est pas acceptable.

Néanmoins, une analyse plus fine de cette méthode nous renseignera sur les erreurs à éviter.

Soit
$$d_i = \|\Delta p_i - \tilde{\Delta p}_i\| \quad (4.5)$$

. 1

L'erreur commise à la quantification de l'élément Δp_i de la différence ΔLSP

Et soit
$$e_i = \|p'_i - \tilde{p}'_i\| \quad (4.6)$$

L'erreur commise à la quantification de l'élément p_i des coefficients LSP.

De l'Equation (4.2) et l'Equation (4.4) nous obtenons

$$e_i = \|\left(\Delta p_i - \tilde{\Delta p}_i\right) + \left(p'_{i-1} - \tilde{p}'_{i-1}\right)\| \quad (4.7)$$

Donc l'erreur $e(i)$, commise sur p'_i est la somme de 2 erreurs, à savoir : $d(i)$, l'erreur sur Δp_i et $e(i-1)$, l'erreur sur p'_{i-1} .

De la même manière, nous obtenons que l'erreur sur p'_{i-1} est la somme de 2 erreurs : $d(i-1)$ et $e(i-2)$. Ainsi de suite nous formons une récurrence qui démontre que l'erreur commise lors de la quantification du premier élément du vecteur LSP se propagera en s'additionnant jusqu'au dernier élément du vecteur.

Dans la DSVQ, une bonne quantification des ΔLSP n'implique pas forcément une bonne quantification des LSP .

Afin de palier, partiellement, à ce problème, nous avons essayé de limiter la propagation d'erreurs à l'intérieur des frontières de chaque sous-vecteur. Résultant, ainsi, en une nouvelle méthode que nous avons appelés DSVQ améliorée.

IV.2.5 Quantification vectorielle des différences en sous-vecteurs améliorée (EDSVQ)

Cette méthode est très proche de la méthode précédente, mise à part qu'elle introduit un mécanisme pour arrêter la propagation des erreurs à la frontière de chaque sous-vecteur.

IV.2.5.1 Procédure de quantification

Au codeur

1. on forme le premier sous vecteur de différence

$$\begin{cases} \Delta p_1^1 = p_1^1 \\ \Delta p_i^1 = p_i^1 - p_{i-1}^1 \text{ K pour } i = 2, i_{\max} \end{cases}$$

2. on quantifie le premier sous vecteur

$$\Delta p_i^1 \xrightarrow{Q^1} \tilde{\Delta p}_i^1 \dots \text{ pour } i = 1, i_{\max}$$

3. on reforme le dernier élément *LSP* du premier sous-vecteur.

$$\tilde{p}_{i_{\max}}^1 = \tilde{\Delta p}_{i_{\max}}^1 + \tilde{p}_{i_{\max}-1}^1$$

4. on forme les ΔLSP du sous-vecteur *j*

$$\begin{cases} \Delta p_1^j = p_1^j - \tilde{p}_{i_{\max}}^{j-1} \\ \Delta p_i^j = p_i^j - p_{i-1}^j \text{ K pour } i = 2, i_{\max} \end{cases}$$

5. on quantifie le sous-vecteur *j*

$$\Delta p_i^j \xrightarrow{Q^j} \tilde{\Delta p}_i^j \dots \text{ pour } i = 1, i_{\max}$$

6. on reforme le dernier élément *LSP* du sous-vecteur *j*.

$$\tilde{p}_{i_{\max}}^j = \tilde{\Delta p}_{i_{\max}}^j + \tilde{p}_{i_{\max}-1}^j$$

Si *j* > nombre de sous-vecteurs, stop

Sinon, *j* = *j*+1 et aller à 4

Au décodeur

À partir de l'index on récupère les différences quantifiées pour chaque sous-vecteur.

On forme : $\tilde{p}_i^j = \tilde{\Delta p}_i^j + \tilde{p}_{i-1}^j \dots \text{ pour } i = 2, 10$

IV.2.5.2 Découpage et allocation de bits

Plusieurs découpages ont été essayés et pour chaque découpage plusieurs allocations de bits ont été attribuées.

Dans la Table 11, les allocations de bits qui ont données le meilleur résultat pour chaque découpage ont été retenues.

| Découpage | Allocation de bits | | |
|--------------|--------------------|----------------|----------------|
| | Sous-vecteur 1 | Sous-vecteur 2 | Sous-vecteur 3 |
| 3 3 4 | 7 | 8 | 9 |
| 3 4 3 | 8 | 9 | 7 |
| 4 3 3 | 9 | 8 | 7 |

Table 11. Découpage et allocation de bits pour la DSVQ

IV.2.5.3 Performance et analyse du quantificateur

La Table 12 rassemble les performances obtenues par la EDSVQ pour chaque découpage et les comparent aux performances de la DSVQ et de la PSVQ (standard)

| | | DS.Moy (dB) | 2 – 4 dB | > 4db | EMBSD |
|--------------|-------------|-------------|----------|-------|-------|
| | PSVQ | 1.84 | 32.99 | 4.67 | 1.551 |
| | DSVQ | 2.23 | 40.82 | 16.76 | |
| EDSVQ | 334 | 2.06 | 40.35 | 5.81 | 1.552 |
| | 343 | 2.14 | 41.90 | 7.18 | 1.535 |
| | 433 | 2.04 | 38.00 | 6.47 | 1.604 |

Table 12. Performances de la DSVQ

Les valeurs de la distorsion spectrale confirment l'amélioration théorique, de telle sorte qu'on remarque, un gain proche de 0.2 dB, pour tout découpage, par rapport aux performances de la DSVQ. Cependant, la EDSVQ reste relégué à près de 0.2db de la méthode standard PSVQ.

Limiter la propagation de l'erreur aux frontières d'un sous-vecteur améliore les performances du codeur mais, en même temps, les garde au dessous de celle du standard.

Notre objectif, maintenant, est de carrément effacer cette propagation d'erreur afin que les erreurs commises sur les coefficients *LSP* soient seulement celles dues à la quantification. Et pour ce, nous avons modifié la façon de faire résultant en une nouvelle méthode que nous avons nommé EDSVQ modifiée.

IV.2.6 Quantification vectorielle en sous-vecteurs des différences améliorée et modifiée (MEDSVQ)

Avec comme idée dans l'esprit d'empêcher la propagation des erreurs entre les sous-vecteurs successifs et à l'intérieur de chaque sous-vecteur, nous sommes arrivés à la méthode suivante :

IV.2.6.1 Procédure de quantification

Procédure de quantification :

Au codeur

1. on forme le premier sous vecteur de différence

$$\Delta p_i^1 = p_i^1 - K \text{ pour } i = 1, i_{\max}$$

2. on quantifie le premier sous vecteur

$$\Delta p_i^1 \xrightarrow{Q^1} \tilde{\Delta p}_i^1 \dots \text{pour } i = 1, i_{\max}$$

3. on reforme le dernier élément *LSP* du premier sous-vecteur.

$$\tilde{p}_{i_{\max}}^1 = \tilde{\Delta p}_{i_{\max}}^1$$

4. on forme les ΔLSP du sous-vecteur *j*

$$\Delta p_i^j = p_i^j - \tilde{p}_{i_{\max}}^{j-1} \dots \text{pour } i = 1, i_{\max} \tag{4.8}$$

5. on quantifie le sous-vecteur *j*

$$\Delta p_i^j \xrightarrow{Q^j} \tilde{\Delta p}_i^j \dots \text{pour } i = 1, i_{\max}$$

6. on reforme le dernier élément *LSP* du sous-vecteur *j*.

$$\tilde{p}_{i_{\max}}^j = \tilde{\Delta p}_{i_{\max}}^j + \tilde{p}_{i_{\max}}^{j-1} \dots \text{pour } i = 1, i_{\max}$$

Si *j* > nombre de sous-vecteurs, stop

Sinon, *j* = *j*+1 et aller à 4

Au décodeur

À partir de l'index on récupère les différences quantifiées pour chaque sous-vecteur.

On forme :
$$\tilde{p}'_i = \tilde{\Delta p}'_i + \tilde{p}'_{i_{\max}} \dots \text{pour } i = 1, i_{\max} \quad (4.9)$$

Et l'erreur de l'Equation (4.6) devient par les Equation (4.8) et Equation (4.9)

$$e'_i = \left\| (\Delta p'_i - \tilde{\Delta p}'_i) + (\tilde{p}'_{i-1} - \tilde{p}'_{i-1}) \right\|$$

Et donc

$$e'_i = d'_i = \left\| \Delta p'_i - \tilde{\Delta p}'_i \right\| \quad (4.10)$$

Et de ce fait, nous avons accomplis notre objectif, à savoir, empêcher la propagation des erreurs.

IV.2.6.2 Découpage et allocation de bits

Puisque cette méthode engendre un nouveau genre de données, une nouvelle étude statistique a été menée. Suivant la dynamique de chaque élément du vecteur *LSP* et les corrélations qui existent entre chacun d'eux. Le découpage suivant a été sélectionné :

- 3 éléments pour le premier sous-vecteur
- 3 éléments pour le deuxième sous-vecteur
- 4 éléments pour le troisième sous-vecteur

De ce qui est de l'allocation de bits, nous avons commencé par partager 24 bits équitablement sur les trois sous-vecteurs octroyant ainsi 8 bits à chacun d'eux, puis nous leurs avons rajouté un bit, totalisant de ce fait 9 bits, de telle sorte que nous ayons, maintenant, 2 dictionnaires pour chaque sous-vecteur.

IV.2.6.3 Performance et analyse du quantificateur

| | | DS.Moy (dB) | 2 – 4 dB | > 4db | EMBSD |
|---------------|------------|--------------------|-----------------|-----------------|--------------|
| PSVQ | | 1.84 | 32.99 | 4.67 | 1.551 |
| DSVQ | | 2.23 | 40.82 | 16.76 | |
| EDSVQ | 334 | 2.06 | 40.35 | 5.81 | 1.552 |
| | 343 | 2.14 | 41.90 | 7.18 | 1.535 |
| | 433 | 2.04 | 38.00 | 6.47 | 1.604 |
| MEDSVQ | 334 | 1.88 | 35.05 | 4.11 | 1.514 |

Table 13. Performance de la MEDSVQ

Les valeurs de la distorsion spectrale confirment l'amélioration théorique, de telle sorte qu'on remarque, un gain proche de 0.15 dB, pour tout découpage, par rapport aux performances de la EDSVQ. De plus, la MEDSVQ se rapproche de près des performances de la méthode standard PSVQ. En effet, la distorsion spectrale de la MEDSVQ n'est supérieure de la PSVQ que de 0.04 dB.

Enfin et sachant que les performances de la MEDSVQ ne sont pas très éloignées de celles de la PSVQ (standard ITU), une optimisation de l'allocation de bits pourrait, encore, donner de meilleurs résultats.

IV.2.6.4 Optimisation

L'optimisation a été faite en deux étapes :

1. optimisation de l'allocation des bits pour chaque sous-vecteur
2. utilisation du métrique de Farvardin au niveau du quantificateur.

IV.2.6.4.1 L'allocation de bits

La combinaison des différents dictionnaires conçus à partir des différentes allocations de bits résulte dans 8 cas possibles regroupés dans la Table 14.

Des résultats illustrés dans la Table 14 nous pouvons remarqué que :

Chapitre IV : Etudes, simulations et résultats.

| Méthode | Débits MEDSVQ (bits/Trame) | Bit Allocation (bits) | | | EMBSD | DS.Moy (dB) | 2<SD<4 (%) | SD>4 (%) |
|---------|-------------------------------|-----------------------|----|----|-------|----------------|---------------|-------------|
| | | V1 | V2 | V3 | | | | |
| 1 | 24 | 8 | 8 | 8 | 1.514 | 1.88 | 35.05 | 4.11 |
| 2 | 25 | 9 | 8 | 8 | 1.528 | 1.87 | 35.11 | 4.02 |
| 3 | 25 | 8 | 9 | 8 | 1.516 | 1.84 | 34.17 | 4.05 |
| 4 | 25 | 8 | 8 | 9 | 1.503 | 1.83 | 33.83 | 4.00 |
| 5 | 26 | 9 | 9 | 8 | 1.534 | 1.84 | 34.17 | 3.76 |
| 6 | 26 | 9 | 8 | 9 | 1.505 | 1.82 | 33.41 | 3.84 |
| 7 | 26 | 8 | 9 | 9 | 1.499 | 1.79 | 32.32 | 3.72 |
| 8 | 27 | 9 | 9 | 9 | 1.462 | 1.79 | 32.26 | 3.83 |

Table 14. Performance de la MEDSVQ pour différentes allocations de bits.

L'ajout d'un bits à la première bande n'a pas vraiment influencé les résultats et que par contre, le même ajout aux autre bandes donne des résultats plus significatifs.

Les méthodes (3) et (4) ont des performances proches voire meilleures que la méthode de quantification PSVQ Utilisée dans le standard.

Par conséquent et à fin de maintenir le même débit que celui du standard, nous avons retranché au premier sous-vecteur 1 bit que nous avons rajouté, à tour de rôle, au deuxième puis au troisième sous-vecteur. Les résultats sont dans la Table 15.

| Méthode | Rate MEDSVQ (bits/Trame) | Bit Allocation (bits) | | | EMBSD | DS.Moy (dB) | 2<SD<4 (%) | SD>4 (%) |
|---------|-----------------------------|-----------------------|----|----|-------|----------------|---------------|-------------|
| | | V1 | V2 | V3 | | | | |
| 1 | 24 | 7 | 8 | 9 | 1.509 | 1.84 | 33.63 | 4.43 |
| 2 | 24 | 7 | 9 | 8 | 1.501 | 1.86 | 34.19 | 4.33 |

Table 15. Performances de la MEDSVQ pour une allocation de bits optimisée

Chapitre IV : Etudes, simulations et résultats.

Comme on pouvait s'y attendre les deux nouvelles allocations donnent de meilleurs résultats, ramenant, de la sorte, la distorsion spectrale de la MEDSVQ au niveau de celle du standard.

IV.2.6.4.2 Métrique de Farvardin

Dans cette étape, au lieu d'utiliser, dans la procédure de quantification, la distance euclidienne comme métrique de recherche du vecteur le plus proche, on utilise, à sa place, le métrique de Farvardin comme défini dans les Equation (1.28), (1.29) et (1.31)

Comme on peut le voir dans la Table 16, les performances de la MEDSVQ sous ses trois variantes, sont aussi bonnes que celles du standard G.723.1, en terme de distorsion spectrale, et meilleures en terme D'EMBSD.

| Méthode | Débit | Bit Allocation (bits) | | | EMBSD | DS.Moy | 2<SD<4 | SD>4 |
|---------------------|-------|-----------------------|----|----|-------|--------|--------|------|
| | | V1 | V2 | V3 | | | | |
| MEDSVQ (bits/Trame) | | | | | (dB) | (%) | (%) | |
| 1 | 24 | 8 | 8 | 8 | 1.482 | 1.86 | 34.21 | 4.17 |
| 2 | 24 | 7 | 8 | 9 | 1.529 | 1.83 | 33.07 | 4.38 |
| 3 | 24 | 7 | 9 | 8 | 1.445 | 1.84 | 33.55 | 4.19 |

Table 16. Performances de la MEDSVQ optimisée par l'allocation de bits et le métrique de FarVardin.

Des trois variantes MEDSVQ, nous avons opté pour l'allocation accordant 7 bits au premier sous-vecteur, 9 bits au deuxième sous-vecteur, et 8bits au dernier sous-vecteur. Ce choix reste difficile tant leurs performances sont proches. En effet, les trois variantes offrent des distorsions spectrales très voisines, cependant, celle choisie offre une EMBSD significativement plus basse et donc meilleure.

Donc, au terme de cette méthode, et dans des condition de perte de trame nulles, nous sommes arrivés, par une quantification intra-trame et tout en respectant les impératifs exigés au début du chapitre, à savoir : qualité, débits et complexité. (Voir Table 17), à légèrement améliorer, du moins, à égaler les performances de la quantification inter-trame (PSVQ) utilisée dans le standard ITU G.723.1.

Chapitre IV : Etudes, simulations et résultats.

| Volet de comparaison | | PSVQ | MEDSVQ |
|----------------------|------------------|-------------------|---------|
| Qualité | Dist.Spectr (dB) | 1.84 | 1.84 |
| | EMBSD | 1.551 | 1.445 |
| Débit (bits/trame) | | 24 | 24 |
| Complexité | | Celle du standard | La même |

Table 17. Comparaison entre PSVQ (G.723.1) et MEDSVQ

| Méthode de quantification | | Std PSVQ | DSQ | DSV Q | EDVSQ | | | MEDVSQ10 | | |
|--|-------|----------|-------|-------|--------------------|--------------------|--------------------|----------------------------|-------|-------|
| Allocation de bits (bits) et découpage | | 24 | 27 | 24 | 3, 3, 4 7, 8, 9 | 3, 4, 3 8, 9, 7 | 4, 3, 3 9, 8, 7 | 3 3 4 8 8 8 7 8 9 7 9 8 | | |
| DS dB | Moy | 1.84 | 1.81 | 2.23 | 2.06 | 2.14 | 2.04 | 1.86 | 1.83 | 1.84 |
| | SsTr4 | 1.10 | 1.05 | NnCa | 1.45 | 1.59 | 1.43 | 1.12 | 1.08 | 1.11 |
| 2<ds<4(%) | Moy | 32.99 | 32.99 | 40.82 | 40.53 | 41.9 | 38 | 34.21 | 33.07 | 33.55 |
| | SsTr4 | 10.16 | 9.66 | NnCa | NnCa | NnCa | NnCa | 10.63 | 10.18 | 10.46 |
| ds>4(%) | Moy | 4.67 | 4.44 | 16.76 | 5.81 | 7.18 | 6.47 | 4.17 | 4.38 | 4.19 |
| | SsTr4 | 1.34 | 0.69 | NnCa | NnCa | NnCa | NnCa | 0.72 | 0.64 | 0.75 |
| EMBSD | | 1.551 | 1.498 | NnCa | 1.552 | 1.535 | 1.604 | 1.482 | 1.529 | 1.445 |

Table 18. Récapitulatif des performances des différentes méthodes utilisées

Enfin pour pouvoir apprécier l'avancement du travail, la Table 18 rassemble les performances des différentes méthodes proposées.

Une fois le quantificateur conçu et testé dans des conditions de perte nulles, nous nous intéressons, maintenant, à ses performances lors d'effacements de trames.

¹⁰ MEDSVQ optimisée par le métrique de Farvardin

IV.3 Masquage de trames perdues par interpolation

IV.3.1 Modélisation des pertes de paquets de données dans les réseaux de paquets

Plusieurs modèles de perte de paquets ont été établis. Cependant, pour correctement évaluer les qualités sonores d'un codec, l'utilisation de modèle simulant des pertes en rafales est particulièrement recommandée. Ceci est d'autant plus vrai que, l'efficacité de la méthode de masquage de paquets perdus est directement corrélée à la nature statistique des pertes.

IV.3.1.1 Modèle de perte de paquets sans mémoire (modèle de Bernoulli)

Le modèle de Bernoulli est un modèle très simple, il est caractérisé par un seul paramètre, le taux de perte « R ». Ce modèle peut être implémenté très simplement en utilisant un générateur de nombre aléatoire et un processus de décision. Néanmoins ce genre de modèle modélise faiblement les pertes en rafales.

IV.3.1.2 Modèle de pertes en rafales (modèle de Gilbert)

Une chaîne de Markov à 2 états modélise bien les pertes en rafales. Le réseau avec pertes peut être modélisé comme ayant 2 états, un état « 0 » indiquant un paquet reçu ou bon paquet, et un état « 1 » indiquant un paquet perdu ou mauvais paquet. A la différence du modèle de Bernoulli le modèle de Gilbert est caractérisé, comme montré dans la figure 7, par 2 paramètres, à savoir :

- P : probabilité de passer de l'état « 0 » vers l'état « 1 ».
- Q : probabilité de passer de l'état « 1 » vers l'état « 0 ».

Le taux de perte moyen est donné par $R = \frac{P}{P+Q}$. (4.11)

La probabilité Q est reliée à la longueur des rafales tel que, la probabilité d'avoir une perte en rafale de longueur n est : $P(n) = Q * (1 - Q)^{n-1}$ (4.12)

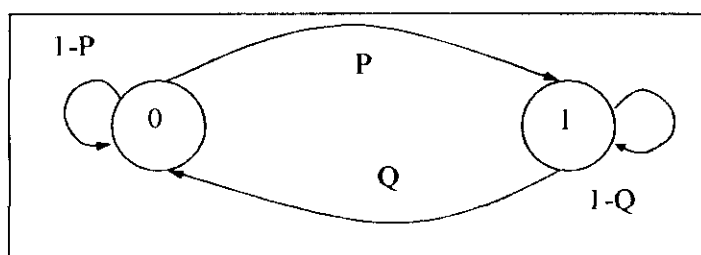


Figure 18. Modèle de perte de paquets de Gilbert

IV.3.1.3 Implémentation

Pour bien évaluer et comparer les performances de la nouvelle méthode de quantification par rapport à celle du standard, et afin de bien mesurer les effets des pertes en rafales sur la qualité sonore du signal reconstruit, différentes conditions de perte ont été simulées, de telle sorte que pour chaque taux de perte R , sept valeurs différentes de Q ont été expérimentées.

La simulation s'est faite sur un nombre total de 19433 paquets, ce qui correspond à la longueur de notre base de données « TEST ». De plus, nous avons supposé que chaque paquet de données du réseau ne contenait qu'une seule trame de voix.

La distribution, la valeur moyenne et l'écart type de la longueur des rafales ; plus les valeurs théorique (recherchées) et pratiques (trouvées après simulation) de R et Q sont rassemblées dans la Table 19 .

Chapitre IV : Etudes, simulations et résultats.

| Taux de perte R (%) | | Prob Q (%) | | Nbre de paquets perdus | Distribution des pertes en rafale de longueurs | | | | | | | | | | | | Moy. long. Rafs | Ecart type Long. rafs |
|---------------------|-------|------------|-------|------------------------|--|------|-----|-----------------|----|----|----|---|---|----|----|----|-----------------|-----------------------|
| Thé | Pra | Thé | Pra | | 1 | 2 | 3 | ¹¹ 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | | |
| 10 | 10.11 | 100 | 100.0 | 1965 | 1965 | | | | | | | | | | | | 1.0 | 0.0 |
| | 09.86 | 95 | 95.15 | 1917 | 1736 | 83 | 5 | | | | | | | | | | 1.05 | 1.25 |
| | 10.23 | 90 | 88.78 | 1988 | 1572 | 168 | 21 | 3 | 1 | | | | | | | | 1.13 | 2.35 |
| | 10.14 | 85 | 83.86 | 1970 | 1393 | 209 | 41 | 9 | | | | | | | | | 1.19 | 1.72 |
| | 9.80 | 80 | 80.78 | 1904 | 1241 | 241 | 45 | 9 | 2 | | | | | | | | 1.24 | 2.26 |
| | 10.30 | 75 | 73.63 | 2002 | 1098 | 26 | 83 | 21 | 6 | 1 | 1 | | | | | | 1.36 | 3.31 |
| | 9.91 | 60 | 60.16 | 1925 | 708 | 266 | 111 | 39 | 18 | 10 | 3 | 2 | 1 | | | | 1.66 | 4.22 |
| 20 | 20.06 | 100 | 100.0 | 3899 | 3899 | | | | | | | | | | | | 1.00 | 0.00 |
| | 19.62 | 95 | 94.81 | 3813 | 3426 | 180 | 9 | | | | | | | | | | 1.05 | 1.25 |
| | 20.39 | 90 | 88.80 | 3963 | 3125 | 351 | 37 | 5 | 1 | | | | | | | | 1.13 | 2.35 |
| | 19.98 | 85 | 84.72 | 3882 | 2805 | 391 | 78 | 14 | 1 | | | | | | | | 1.18 | 2.30 |
| | 19.87 | 80 | 80.61 | 3862 | 2513 | 481 | 94 | 21 | 3 | 1 | | | | | | | 1.24 | 2.83 |
| | 20.20 | 75 | 74.81 | 2224 | 520 | 131 | 45 | 14 | 2 | 1 | | | | | | | 1.34 | 3.33 |
| | 20.04 | 60 | 59.99 | 1411 | 544 | 225 | 101 | 32 | 13 | 5 | 3 | 1 | 1 | | | | 1.67 | 4.79 |
| 30 | 29.96 | 100 | 99.98 | 5822 | 5819 | 1 | | | | | | | | | | | 1.00 | 0.71 |
| | 29.49 | 95 | 95.31 | 5730 | 5203 | 245 | 12 | | | | | | | | | | 1.05 | 1.25 |
| | 30.17 | 90 | 89.34 | 5863 | 4679 | 504 | 46 | 7 | 2 | | | | | | | | 1.12 | 2.35 |
| | 30.15 | 85 | 85.15 | 5860 | 4262 | 606 | 104 | 16 | 2 | | | | | | | | 1.17 | 2.31 |
| | 29.76 | 80 | 80.75 | 5783 | 3778 | 711 | 144 | 35 | 1 | 1 | | | | | | | 1.24 | 2.83 |
| | 30.31 | 75 | 74.92 | 5890 | 3338 | 787 | 203 | 61 | 20 | 3 | 1 | | | | | | 1.33 | 3.33 |
| | 29.87 | 60 | 60.07 | 5805 | 2114 | 802 | 344 | 142 | 49 | 19 | 12 | 2 | 2 | 1 | | | 1.66 | 4.79 |
| 40 | 39.98 | 100 | 99.99 | 7770 | 7767 | 1 | | | | | | | | | | | 1.00 | 0.71 |
| | 39.84 | 95 | 95.18 | 7743 | 7014 | 338 | 16 | 1 | | | | | | | | | 1.05 | 1.83 |
| | 39.94 | 90 | 89.94 | 7762 | 6274 | 644 | 54 | 7 | 2 | | | | | | | | 1.11 | 2.36 |
| | 40.34 | 85 | 85.06 | 7839 | 5690 | 811 | 143 | 22 | 2 | | | | | | | | 1.18 | 2.31 |
| | 39.60 | 80 | 81.06 | 7696 | 5075 | 921 | 192 | 48 | 1 | 1 | | | | | | | 1.23 | 2.84 |
| | 40.21 | 75 | 75.29 | 7814 | 4449 | 1064 | 270 | 77 | 19 | 4 | | | | | | | 1.33 | 2.76 |
| | 39.95 | 60 | 59.80 | 7764 | 2777 | 1100 | 469 | 188 | 60 | 27 | 13 | 3 | 2 | 2 | 0 | 1 | 1.67 | 5.93 |

Table 19. Statistique des pertes en réseaux.

¹¹ Excitation nulle pour plus de 3 pertes successives.

De toutes les valeurs développées dans la table 15, nous choisirons 3 valeurs de Q pour chaque taux de perte R. ces valeurs, surlignées en jaune dans la Table 19, ont été choisies afin d'évaluer les méthode de masquages de trames perdus du codec dans tous ses modes de fonctionnement. En effet, pour

1. $Q = 100$, les pertes sont de longueurs 1
2. $Q = 85$, les pertes de longueurs 2 et 3 sont très présentes.

$Q = 60$, les longueurs de rafales dépassent largement la limite de masquage du codeur c'est-à-dire 3 pertes de suite.

IV.3.2 Masquage

Comme vu dans III.3.4 le masquage consiste à régénérer, pour chaque trame perdue, l'excitation et les coefficients LSP. Afin de bien évaluer l'impacte de la quantification des coefficients LSP sur les performances générales du codec, nous avons, d'une part, gardé le même schéma de masquage des excitations du standard vu dans III.3.4.2 et d'autre part, nous avons mis en place une procédure d'interpolation des LSP différente.

IV.3.2.1 Interpolation des LSP

Contrairement à ce qui est réalisé dans le standard ITU G.723.1 où le masquage de trames perdues consiste à remplacer la trame perdue par une version "dérivée" de la trame précédente, le masquage par interpolation consiste à remplacer les trames perdues par des trames issues de l'interpolation linéaire entre la dernière trame reçue avant perte et la première trame reçues après perte. Intuitivement, on peut dire que l'apport d'information postérieure aux pertes ne peut qu'améliorer les performances du codec. Néanmoins le revers de la médaille de l'interpolation est l'introduction d'un retard du à l'attente de la prochaine bonne trame pour pouvoir poursuivre l'exécution du décodage. Cet inconvénient ne peut être éliminé, mais peut, par contre, être limité. En effet, connaissant la nature très incertaine des réseaux IP, les concepteurs de terminaux VOIP, où les codec G.723.1 sont beaucoup employés, les ont initialement munis d'un tampon mémoire (buffer). Aussi l'implémentation du masquage par interpolation n'introduira aucun coût supplémentaire.

IV.3.3 Simulations et résultats

Dans cette section nous avons simulé l'occurrence de pertes à différents taux sur le codec G.723.1, standard (PSVQ) et les trois variantes de la MEDSVQ.

Les conclusion que nous pouvons tiré de la Table 20 (EMBSD) et de la Table 21 (distorsion spectrale) sont :

Tant que la transmission se passe sans perte, les performances des différentes méthodes sont, au centième près, quasi égales, en terme de distorsion spectrale (PSVQ = 1.84 dB, MEDSVQ = 1.86-1.83-1.84 dB) et restent très proches en terme d'EMBSD (PSVQ = 1.551, MEDSVQ = 1.482 -- 1.529 -- 1.445).

Dés que les pertes surviennent, et même à faible taux ($R = 10\%$), des écarts commencent à apparaître, de telle sorte que, les méthodes exploitant la quantification intra-trame donnent déjà des résultat sensiblement meilleurs : EMBSD : PSVQ = 2.567, MEDSVQ = 1.884 -- 2.114 -- 2.086
Distorsion spectrale : PSVQ = 2.07 dB, MEDSVQ = 1.97 – 1.94 – 1.96 dB

L'impacte, néfaste, des pertes en rafales se confirme, ainsi nous remarquons que pour un même taux de perte R (même nombre de trames perdues), les valeurs de la distorsion spectrale DS croît quand Q diminue, c'est-à-dire, plus les pertes se produisent en séries, moins la procédure de masquage donne de bons résultats. La remarque est aussi valable pour l'EMBSD pour $Q \geq 85\%$, en dessous nous remarquons l'EMBSD varie inversement. Ceci "pourrait" être expliqué par le fait qu'à $Q = 60\%$ le nombre de rafales de plus de trois pertes successives est important (Table 19), ce qui induit de nombreuses portions de silence, dans le signal reconstruit, qui ne sont pas prises en comptes lors du calcul de l'EMBSD.

nous remarquons, Comme illustré dans la Figure 19 (EMBSD) et dans la Figure 20 (Distorsion spectrale), qu' au fur et à mesure que les pertes se multiplient, les performances de la quantification intra-trame (MEDSVQ) se montrent bien au dessus de celles de la quantification inter-trame (PSVQ), de telle sorte qu'à $R = 40\%$ la distorsion spectrale de la MEDSVQ est de plus de 0.4 dB inférieure à la méthode standard PSVQ, alors que l'EMBSD est meilleur de plus de 1.6.

Chapitre IV : Etudes, simulations et résultats.

Il est maintenant clair que, dans des conditions de pertes de trames, les quantifications MEDSVQ donnent de meilleurs résultats que la PSVQ, cependant le choix entre les différentes variantes de la MEDSVQ semble être plus délicat tant leurs performances sont proches.

| EMBSD | | Quantification Inter-trame | Quantification Intra-trame | | |
|----------------------------|-------|-------------------------------|----------------------------|-------------------|-------------------|
| | | | MEDSVQ 8, 8, 8 | MEDSVQ 7, 8, 9 | MEDSVQ 7, 9, 8 |
| Taux de pertes R (%) | Q (%) | Standard (24 Bits) | | | |
| 0 | 0 | 1.551 | 1.482 | 1.529 | 1.445 |
| | 100 | 2.567 | 1.884 | 2.114 | 2.086 |
| | 85 | 2.862 | 2.509 | 2.339 | 2.409 |
| 10 | 60 | 2.330 | 1.987 | 2.289 | 1.936 |
| | 100 | 3.688 | 2.488 | 2.783 | 2.771 |
| | 85 | 3.785 | 3.031 | 2.834 | 2.965 |
| 20 | 60 | 2.863 | 2.410 | 2.648 | 2.230 |
| | 100 | 4.610 | 3.024 | 3.217 | 3.169 |
| | 85 | 4.701 | 3.533 | 3.309 | 3.533 |
| 30 | 60 | 3.414 | 2.606 | 3.070 | 2.711 |
| | 100 | 4.944 | 3.682 | 3.500 | 3.390 |
| | 85 | 5.230 | 3.610 | 3.632 | 3.607 |
| 40 | 60 | 3.948 | 2.782 | 3.203 | 2.962 |

Table 20. Comparaison des EMBSD des différentes méthodes de quantification sous l'effet de pertes.

Chapitre IV : Etudes, simulations et résultats.

| Distorsion spectrale | | Quanti. Inter-frame | Quantification Intra-frame | | |
|----------------------|------------|------------------------|----------------------------|----------------|----------------|
| Taux de pertes R (%) | Prob Q (%) | Standard (24 Bits) | MEDSVQ 8, 8, 8 | MEDSVQ 7, 8, 9 | MEDSVQ 7, 9, 8 |
| 0 | 0 | 1.84 | 1.86 | 1.83 | 1.84 |
| 10 | 100 | 2.07 | 1.97 | 1.94 | 1.96 |
| | 85 | 2.08 | 1.99 | 1.96 | 1.97 |
| | 60 | 2.08 | 2.02 | 1.99 | 2.00 |
| 20 | 100 | 2.31 | 2.09 | 2.06 | 2.07 |
| | 85 | 2.31 | 2.12 | 2.09 | 2.11 |
| | 60 | 2.32 | 2.17 | 2.15 | 2.15 |
| 30 | 100 | 2.54 | 2.20 | 2.18 | 2.19 |
| | 85 | 2.56 | 2.25 | 2.22 | 2.23 |
| | 60 | 2.55 | 2.32 | 2.30 | 2.30 |
| 40 | 100 | 2.77 | 2.33 | 2.30 | 2.31 |
| | 85 | 2.80 | 2.38 | 2.35 | 2.37 |
| | 60 | 2.80 | 2.48 | 2.45 | 2.47 |

Table 21. . Comparaison des EMBSD des différentes méthodes de quantification sous l'effet de pertes.

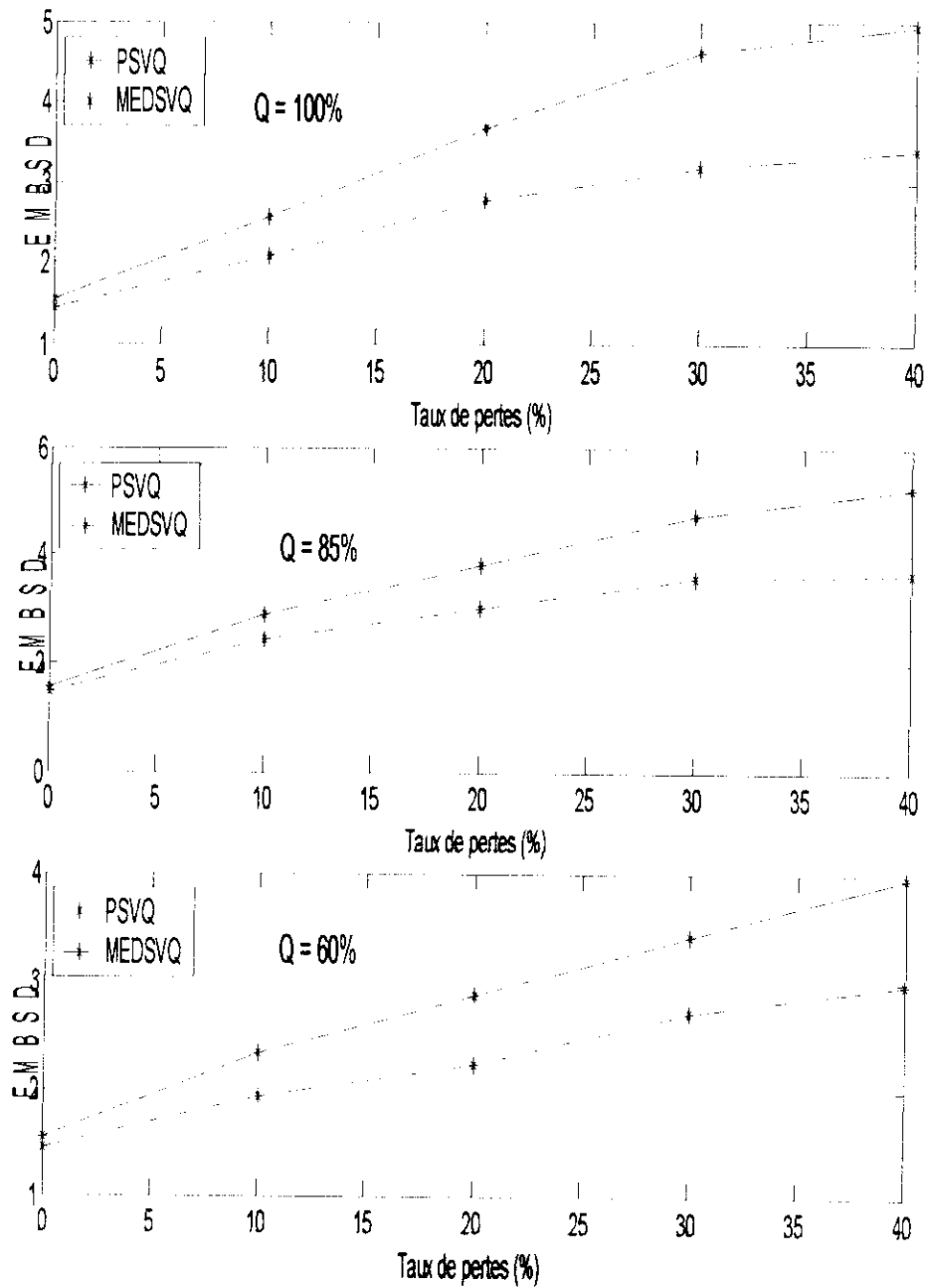


Figure 19. Variation des EMBS D en fonction des taux de pertes.

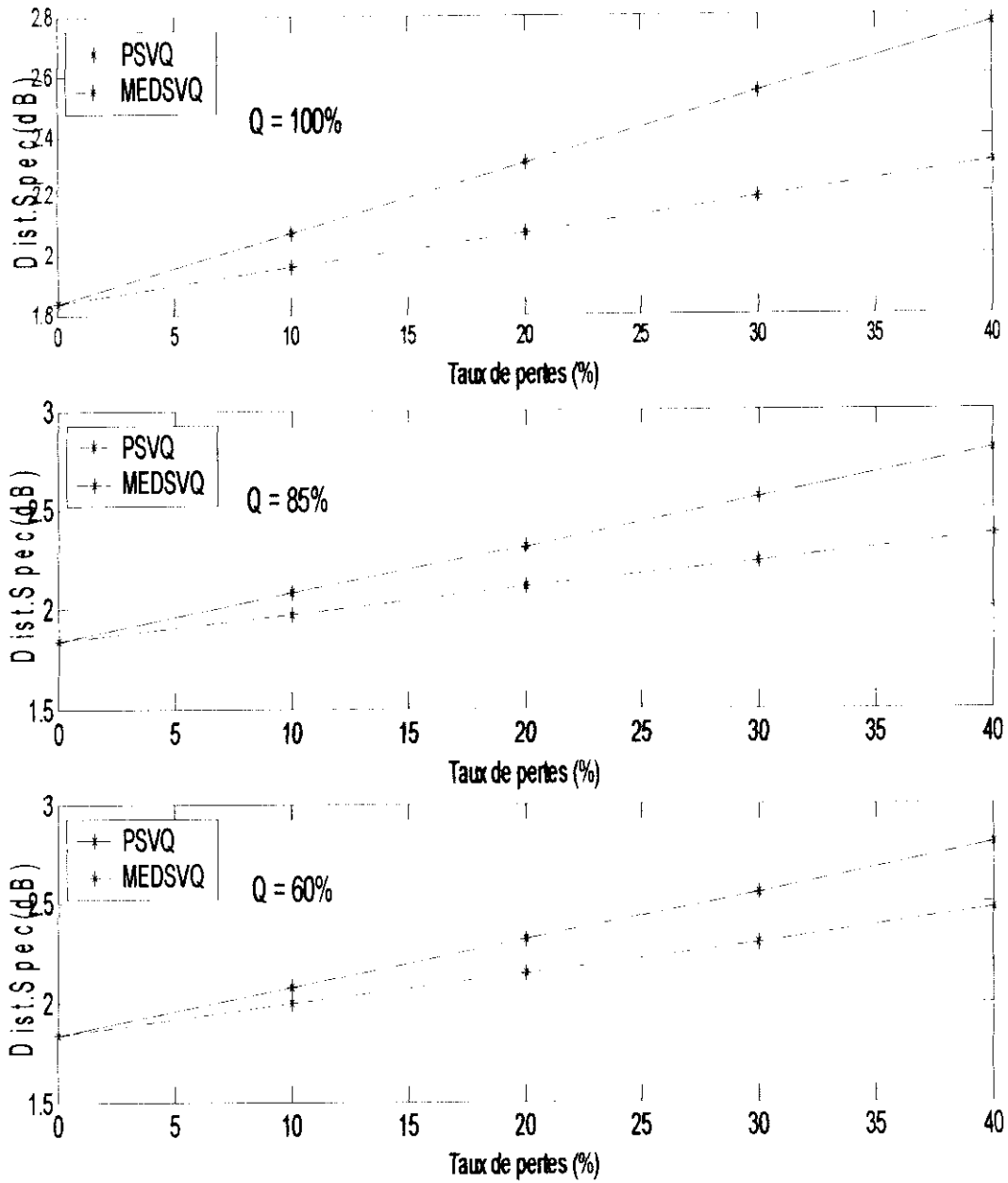


Figure 20. Variation de la distorsion spectrale en fonction du taux de perte.

Conclusion générale

A

près avoir constaté que, lors de pertes de trames, la quantification prédictive des coefficients LSP (PSVQ, Inter-trame) utilisée par le codec G.723.1, engendre une erreur qui ne se limite pas à la trame perdue mais qui se propage systématiquement aux trames suivantes, affectant ainsi, les performances globales du codec; nous avons essayé d'implémenter de nouvelles méthodes de quantification qui traitent chaque trame séparément et qui exploitent les corrélations Intra-trame, c'est-à-dire la corrélation entre chaque élément d'une même trame, espérant de la sorte, enrayer la propagation d'erreur sans détériorer les autres performances du codec (débit complexité).

La première étape fut la réalisation d'une petite étude statistique sur les données à quantifier, c'est-à-dire, les coefficients LSP (IV.2.1.2). Cette étude a montré la large variabilité des LSP et les difficultés qui en découlent pour les quantifier. Néanmoins, elle a mis à jour la forte corrélation qui existe entre chaque couple d'éléments successifs LSP.

Afin d'exploiter la corrélation Intra-trame, la deuxième étape consistait à quantifier les différences Δ LSP par le biais d'un quantificateur scalaire non uniforme[15]. Cette méthode a renvoyé de bon résultat en égalant la PSVQ dans des conditions de perte nulles et, surtout, en confirmant la non propagation des erreurs lors de pertes de trames.

Conclusion générale

L'augmentation du débit de 3bits/trame soit 100bits/s constitue le revers de la médaille de cette méthode.

Afin de réduire le débit sans affecter la qualité, le passage à la quantification vectorielle s'est imposé. De plus, pour des considérations de complexité et de simplicité de calcul, la quantification vectorielle s'est faite sur plusieurs bandes (SVQ).

La troisième étape consistait à quantifier les Δ LSP vectoriellement bande par bande. Bien que la reconstruction des LSP à partir des Δ LSP engendre une propagation des erreurs d'élément en élément LSP, de telle sorte, que les résultats de cette méthode soit très mauvais, les enseignements appris ont été très utiles pour la suite du travail.

La quatrième étape proposait une solution pour stopper la propagation des erreurs aux frontières de chaque bande, diminuant ainsi, les distorsions, mais pas suffisamment pour approcher les performances du standard (PSVQ).

La cinquième et dernière étape proposait une solution qui élimine complètement la propagation des erreurs entre éléments LSP mais qui abandonnait aussi la quantification des Δ LSP successifs et leurs propriétés au profit d'une autre forme. Cette dernière méthode optimisée par une allocation de bits adaptée et l'utilisation du métrique de Farvardin dans la procédure de quantification, et sans introduire d'augmentation de débit ni de complexité, a donné, dans des conditions de pertes nulles, des résultats semblables à ceux du standard PSVQ. L'essai de cette méthode sous des conditions de pertes de trames a montré un gain significatif de 0.4dB en terme de distorsion spectrale et de 1.6 en terme d'EMBSD

Ainsi, nous sommes arrivé à trouver une méthode de quantification qui enrayer la propagation d'erreurs et donc qui améliore la qualité du codec G.723.1 sans détériorer ses autres performances (débits et complexité).

Perspectives futures:

Confirmer les résultats mathématiques par des tests d'écoute.

Améliorer les performances de l'interpolation des sous-trames.

Annexe A

| Taux de perte R (%) | Prob. Q(%) | DS.Moy (dB) | | 2<DS<4 (%) | | DS>4 (%) | | EMBSD |
|---------------------|------------|-------------|--------|------------|--------|----------|--------|-------|
| | | Tr | S.Tr.4 | Tr | S.Tr.4 | Tr. | S.Tr.4 | |
| 0 | 0 | 1.84 | 1.11 | 33.55 | 10.46 | 4.19 | 0.75 | 1.445 |
| | 100 | 1.96 | 1.29 | 35.43 | 13.40 | 5.83 | 3.05 | 2.086 |
| 10 | 95 | 1.96 | 1.29 | 35.63 | 13.42 | 5.92 | 3.10 | 2.098 |
| | 90 | 1.97 | 1.30 | 35.11 | 13.62 | 6.21 | 3.30 | 2.015 |
| | 85 | 1.97 | 1.31 | 35.28 | 13.63 | 6.31 | 3.41 | 2.409 |
| | 80 | 1.96 | 1.29 | 35.41 | 13.50 | 6.05 | 3.20 | 1.996 |
| | 75 | 1.99 | 1.32 | 35.15 | 13.42 | 6.54 | 3.69 | 2.165 |
| | 60 | 2.00 | 1.32 | 34.98 | 13.48 | 6.89 | 3.77 | 1.936 |
| | 100 | 2.07 | 1.47 | 37.54 | 16.30 | 7.47 | 5.45 | 2.771 |
| 20 | 95 | 2.08 | 1.47 | 37.19 | 16.18 | 7.73 | 5.53 | 2.766 |
| | 90 | 2.09 | 1.50 | 37.22 | 16.34 | 8.09 | 5.97 | 2.637 |
| | 85 | 2.11 | 1.50 | 36.93 | 16.51 | 8.44 | 6.04 | 2.965 |
| | 80 | 2.09 | 1.50 | 37.10 | 16.64 | 8.13 | 5.79 | 2.389 |
| | 75 | 2.12 | 1.52 | 36.78 | 16.29 | 8.62 | 6.46 | 2.693 |
| | 60 | 2.15 | 1.55 | 36.40 | 16.60 | 9.60 | 6.95 | 2.230 |
| | 100 | 2.19 | 1.66 | 38.99 | 19.12 | 9.29 | 7.90 | 3.169 |
| 30 | 95 | 2.20 | 1.66 | 38.97 | 19.19 | 9.55 | 7.92 | 3.116 |
| | 90 | 2.21 | 1.68 | 39.23 | 19.26 | 9.75 | 8.24 | 3.251 |
| | 85 | 2.23 | 1.69 | 38.83 | 19.52 | 10.36 | 8.49 | 3.533 |
| | 80 | 2.22 | 1.69 | 38.66 | 19.19 | 10.29 | 8.53 | 2.843 |
| | 75 | 2.26 | 1.73 | 38.60 | 19.57 | 10.91 | 9.08 | 3.356 |
| | 60 | 2.30 | 1.77 | 37.88 | 19.46 | 12.18 | 9.92 | 2.711 |
| | 100 | 2.31 | 1.85 | 41.05 | 22.03 | 11.24 | 10.47 | 3.390 |
| 40 | 95 | 2.31 | 1.85 | 40.37 | 22.11 | 11.40 | 10.32 | 3.441 |
| | 90 | 2.34 | 1.87 | 40.94 | 22.16 | 11.63 | 10.73 | 3.585 |
| | 85 | 2.37 | 1.90 | 40.76 | 22.14 | 12.34 | 11.37 | 3.607 |
| | 80 | 2.36 | 1.89 | 39.94 | 22.21 | 12.69 | 11.25 | 3.564 |
| | 75 | 2.40 | 1.94 | 39.98 | 22.03 | 13.30 | 12.20 | 3.730 |
| | 60 | 2.47 | 2.00 | 39.25 | 22.51 | 14.98 | 12.97 | 2.962 |

Table 22. Performances complètes de la MEDSVQ.

Annexe A

| Taux de perte R (%) | Prob. Q(%) | DS.Moy (dB) | | 2<DS<4 (%) | | DS>4 (%) | | EMBSD |
|---------------------|------------|-------------|--------|------------|--------|----------|-------|-------|
| | | Tr | S.Tr.4 | Tr | S.Tr.4 | Av | s.t.4 | |
| 0 | 0 | 1.81 | 1.05 | 32.99 | 9.66 | 4.44 | 0.69 | 1.498 |
| 10 | 100 | 1.93 | 1.24 | 34.85 | 12.53 | 6.23 | 3.11 | 1.933 |
| | 95 | 1.94 | 1.24 | 34.79 | 12.66 | 6.29 | 3.15 | 2.238 |
| | 90 | 1.94 | 1.25 | 34.65 | 12.83 | 6.37 | 3.32 | 2.070 |
| | 85 | 1.95 | 1.25 | 34.49 | 12.43 | 6.61 | 3.45 | 2.344 |
| | 80 | 1.94 | 1.25 | 34.83 | 12.37 | 6.31 | 3.24 | 2.108 |
| | 75 | 1.96 | 1.27 | 34.66 | 12.65 | 6.80 | 3.62 | 2.116 |
| | 60 | 1.97 | 1.27 | 34.47 | 12.54 | 7.00 | 3.77 | 2.092 |
| 20 | 100 | 2.05 | 1.42 | 36.47 | 15.67 | 8.14 | 5.43 | 2.676 |
| | 95 | 2.05 | 1.42 | 36.38 | 15.59 | 8.09 | 5.48 | 2.823 |
| | 90 | 2.07 | 1.45 | 36.65 | 15.76 | 8.33 | 5.95 | 2.912 |
| | 85 | 2.08 | 1.45 | 36.09 | 15.42 | 8.75 | 6.12 | 2.927 |
| | 80 | 2.08 | 1.45 | 36.57 | 15.84 | 8.28 | 5.85 | 2.581 |
| | 75 | 2.10 | 1.48 | 36.36 | 15.61 | 8.84 | 6.38 | 2.720 |
| | 60 | 2.13 | 1.51 | 36.02 | 15.61 | 9.74 | 7.04 | 2.580 |
| 30 | 100 | 2.17 | 1.61 | 38.11 | 18.47 | 9.85 | 7.97 | 3.227 |
| | 95 | 2.18 | 1.61 | 38.00 | 18.39 | 10.00 | 7.95 | 3.179 |
| | 90 | 2.20 | 1.64 | 38.18 | 18.64 | 10.20 | 8.27 | 3.436 |
| | 85 | 2.21 | 1.66 | 38.03 | 18.50 | 10.80 | 8.73 | 3.366 |
| | 80 | 2.21 | 1.65 | 38.26 | 18.45 | 10.41 | 8.58 | 3.215 |
| | 75 | 2.24 | 1.69 | 38.13 | 18.99 | 11.20 | 9.06 | 3.476 |
| | 60 | 2.29 | 1.73 | 37.37 | 18.55 | 12.44 | 10.14 | 2.749 |
| 40 | 100 | 2.30 | 1.81 | 40.08 | 21.36 | 11.73 | 10.39 | 3.417 |
| | 95 | 2.30 | 1.81 | 39.55 | 21.57 | 11.89 | 10.42 | 3.561 |
| | 90 | 2.33 | 1.83 | 39.60 | 21.45 | 12.24 | 10.85 | 3.871 |
| | 85 | 2.35 | 1.86 | 39.99 | 21.43 | 12.72 | 11.38 | 3.347 |
| | 80 | 2.35 | 1.85 | 39.40 | 21.38 | 12.81 | 11.37 | 3.738 |
| | 75 | 2.39 | 1.91 | 39.38 | 21.39 | 13.76 | 12.35 | 3.675 |
| | 60 | 2.46 | 1.97 | 38.66 | 21.90 | 15.22 | 13.09 | 2.854 |

Table 23. Résultats complets de la DSQ

Annexe A

| Taux de perte R (%) | Prob. Q(%) | DS.Moy (dB) | | 2<DS<4 (%) | | DS>4 (%) | | EMBSD |
|---------------------|-------------|-------------|-------------|------------|--------|----------|--------------|--------------|
| | | Tr | S.Tr.4 | Tr | S.Tr.4 | Av | s.t.4 | |
| 0 | 0 | 1.84 | 1.10 | 32.99 | 10.16 | 4.67 | 1.34 | 1.551 |
| 10 | 100 | 2.07 | 1.43 | 36.95 | 15.96 | 7.71 | 5.06 | 2.567 |
| | 95 | 2.07 | 1.42 | 36.81 | 15.76 | 7.80 | 4.96 | 2.580 |
| | 90 | 2.07 | 1.44 | 36.41 | 15.73 | 8.07 | 5.31 | 2.608 |
| | 85 | 2.08 | 1.44 | 36.45 | 15.84 | 8.13 | 5.38 | 2.862 |
| | 80 | 2.07 | 1.44 | 36.47 | 15.60 | 7.93 | 5.01 | 2.425 |
| | 75 | 2.08 | 1.44 | 36.80 | 15.41 | 7.71 | 5.06 | 2.673 |
| | 60 | 2.08 | 1.44 | 36.52 | 15.60 | 8.16 | 5.22 | 2.330 |
| 20 | 100 | 2.31 | 1.77 | 41.28 | 22.13 | 10.98 | 8.55 | 3.688 |
| | 95 | 2.30 | 1.75 | 40.42 | 21.35 | 10.84 | 8.55 | 3.179 |
| | 90 | 2.31 | 1.78 | 40.53 | 21.11 | 11.38 | 9.28 | 3.054 |
| | 85 | 2.31 | 1.77 | 39.94 | 21.25 | 11.36 | 9.07 | 3.785 |
| | 80 | 2.29 | 1.74 | 40.50 | 21.23 | 10.92 | 8.60 | 3.234 |
| | 75 | 2.31 | 1.76 | 40.59 | 20.58 | 11.35 | 9.44 | 3.561 |
| 30 | 60 | 2.32 | 1.77 | 39.62 | 20.72 | 12.00 | 9.49 | 2.863 |
| | 100 | 2.54 | 2.09 | 45.76 | 27.41 | 13.93 | 12.28 | 4.610 |
| | 95 | 2.53 | 2.08 | 44.34 | 26.65 | 14.21 | 12.57 | 4.069 |
| | 90 | 2.54 | 2.10 | 44.86 | 26.91 | 14.30 | 12.86 | 4.709 |
| | 85 | 2.56 | 2.11 | 44.42 | 27.17 | 14.67 | 12.96 | 4.701 |
| | 80 | 2.53 | 2.07 | 43.79 | 26.27 | 14.50 | 12.73 | 4.186 |
| | 75 | 2.55 | 2.10 | 44.09 | 26.23 | 15.02 | 13.33 | 4.766 |
| 40 | 60 | 2.55 | 2.09 | 42.58 | 25.15 | 15.71 | 13.87 | 3.414 |
| | 100 | 2.77 | 2.42 | 48.97 | 33.46 | 17.44 | 16.00 | 4.944 |
| | 95 | 2.77 | 2.41 | 48.94 | 32.49 | 17.31 | 16.44 | 5.430 |
| | 90 | 2.79 | 2.43 | 48.97 | 32.59 | 17.80 | 16.56 | 5.209 |
| | 85 | 2.80 | 2.44 | 48.57 | 32.70 | 18.18 | 16.79 | 5.230 |
| | 80 | 2.76 | 2.40 | 47.48 | 31.96 | 18.06 | 16.65 | 4.969 |
| | 75 | 2.79 | 2.43 | 47.68 | 31.36 | 18.68 | 17.66 | 4.602 |
| 60 | 2.80 | 2.42 | 46.22 | 30.48 | 19.38 | 18.09 | 3.948 | |

Table 24. Résultats complets de la PSVQ (standard)

Annexe B

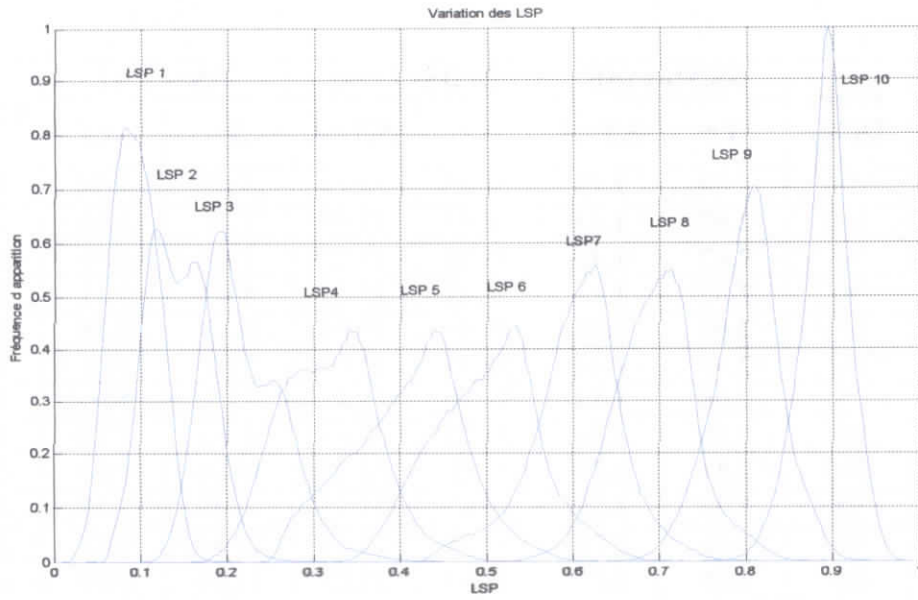


Figure 21. Autres histogramme des LSP (cartésiennes)

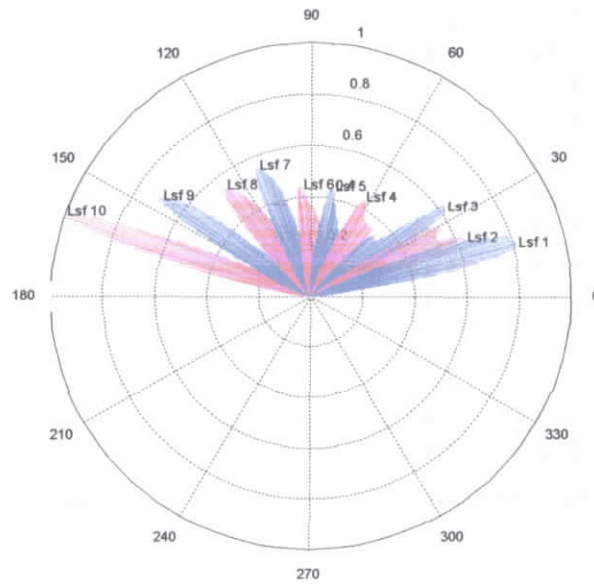


Figure 22. Histogramme LSP (polaires)

Annexe C

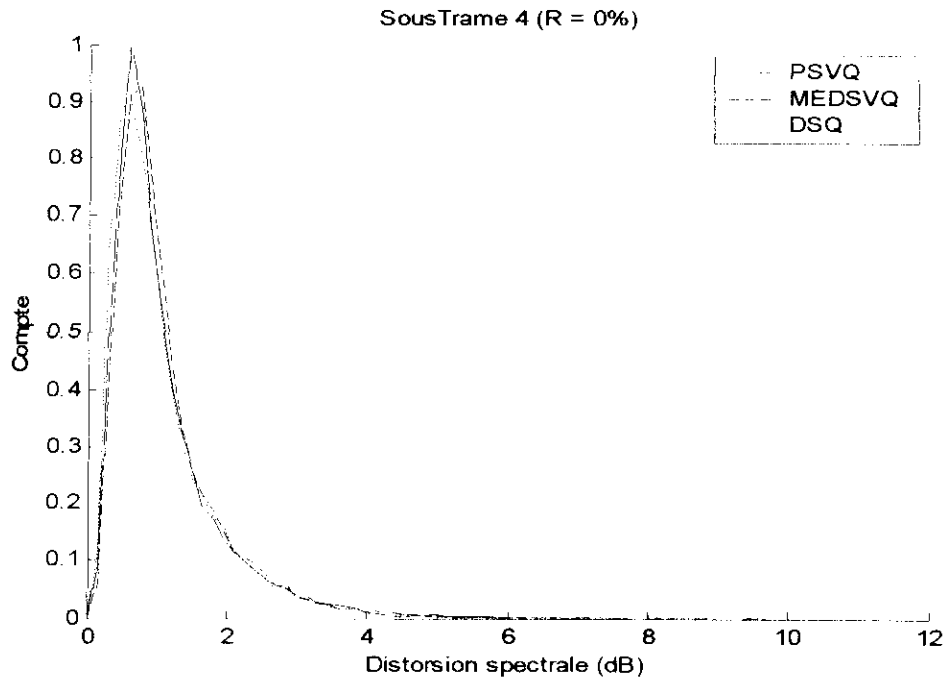


Figure 23. Histogramme des distorsions spectrales (R = 0%)

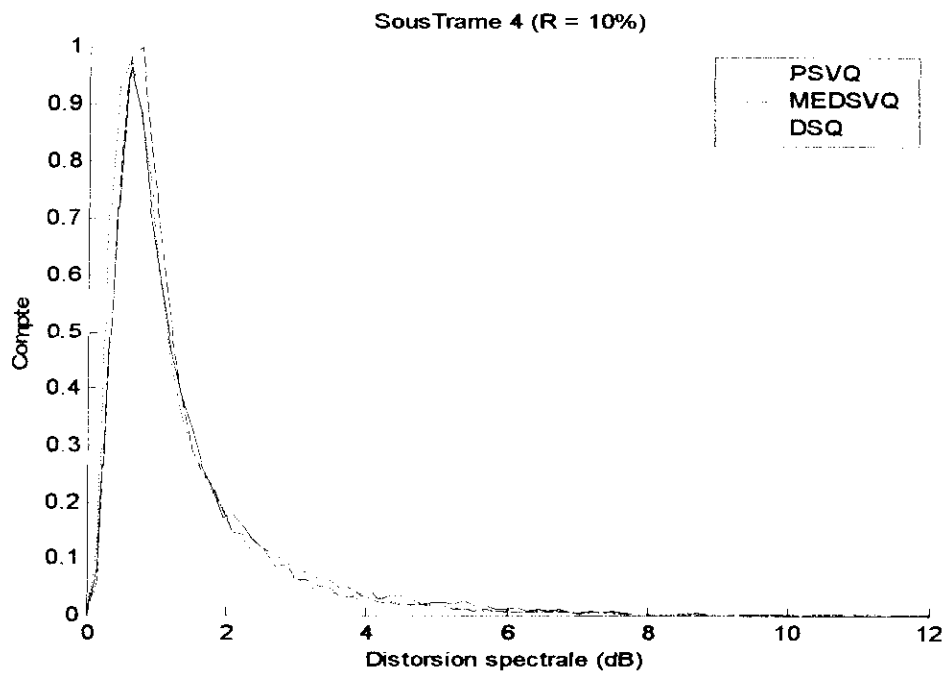


Figure 24. Histogramme des distorsions spectrales (R = 10%).

Annexe c

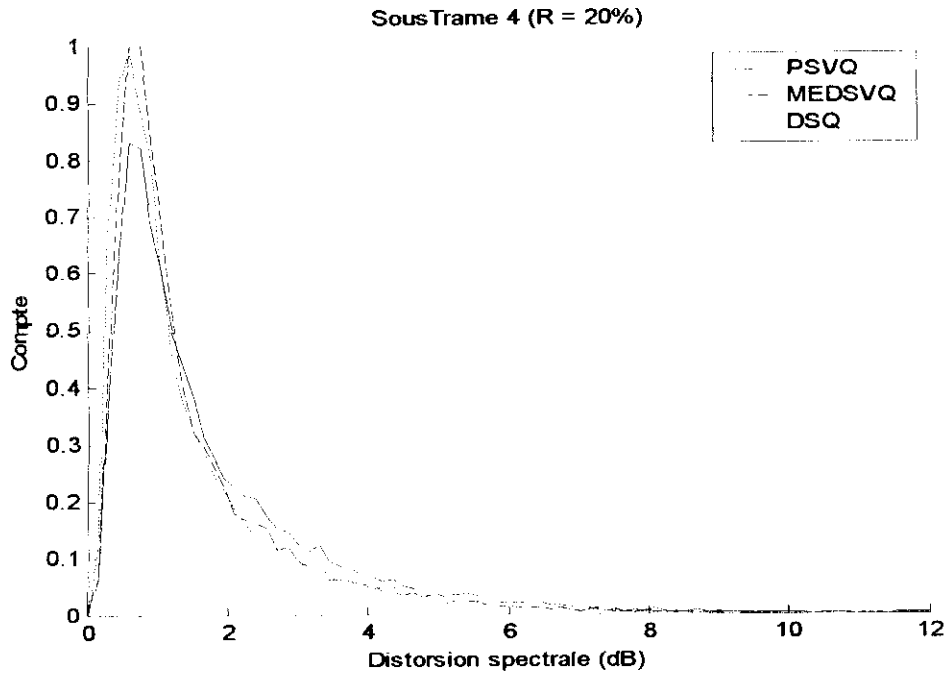


Figure 25. Histogramme des distorsions spectrales (R = 20%).

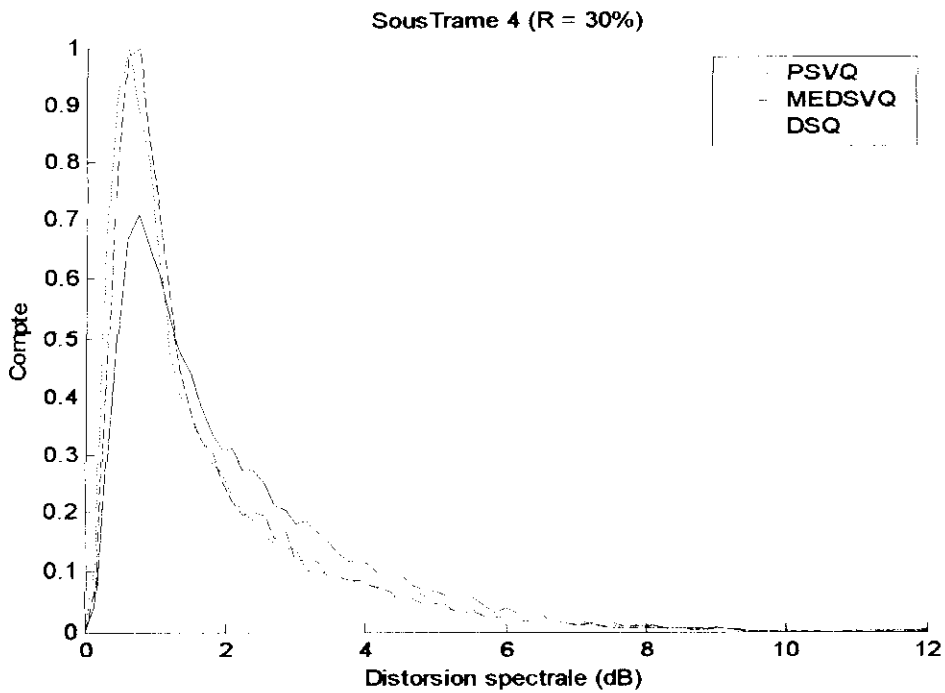


Figure 26. Histogramme des distorsions spectrales (R = 30%).

Annexe c

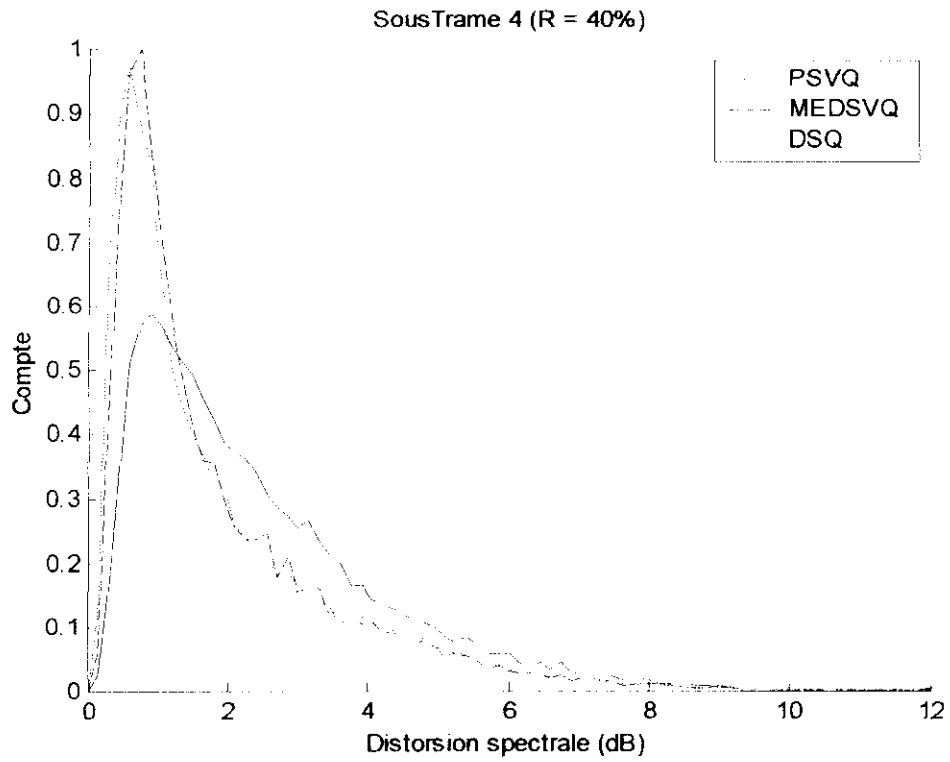


Figure 27. Histogramme des distorsions spectrales (R = 40%).

Bibliographie

- [1] R. Boite et M. Kunt, "Traitement de la parole," 1987, Presses Polytechniques romandes, première édition
- [2] F. Itakura, "Line spectrum representation of linear prediction coefficients of speech signals," *Journal Acoustical Society America*, vol. 57, p. 535, 1975. (Abstract).
- [3] 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.
- [4] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 Bits/Frame," *IEEE Trans. Acoustic, Speech, and Sig. Processing*, Vol. ASSP_1, Jan. 1993, pp. 3-14.
- [5] W. Pariera, "Modifying LPC parameter dynamics to improve speech coder efficiency," *Thesis, Departement of Electrical & Computer Engineering, McGill University, Montreal, Canada, Sep 2001*
- [6] J. H. Y. Loo, "Intraframe and interframe coding of speech spectral parameters", " *Thesis, Departement of Electrical & Computer Engineering, McGill University, Montreal, Canada, Sep 1996*
- [7] S. Dimolitsas, F. L. Corcoran, and C. Ravishankar, "Dependence of opinion scores on listening set Used in degradation category rating assesment," *IEEE Trans. Acoustic, Speech, and Sig. Processing*, Vol. ASSP 3, Sep. 1995, pp. 421-424.
- [8] R. Laroia, N. Phamdo, and N. Farvardin, "Robust and 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.
- [9] W. Yang, K. Krishnamachari and R. Yantom, "Improvement of the MBSD objective speech quality measure using TDMA data," *Submitted to IEEE Speech Coding Workshop, 1999*

- [10] W. Yang, M. Benbouchta and R. Yantom, "Performance of the modified bark spectral distortion as an objective speech quality measure," ICASSP, vol. 1, pp. 541-544, Seattle, 1998
- [11] W. Yang, M. Dixon and R. Yantom, "A modified bark spectral distortion measure which uses noise masking threshold," IEEE Speech Coding Workshop, pp. 55-56, Pocono Manor, 1997
- [12] C. Perkins, O. Hodson and V. Hardman " A survey of Packet Loss Recovery Techniques for Streaming Audio". 1998 IEEE Network.
- [13] D. J. Goodman et al., "Waveform Substitution techniques for recovering missing speech segments in packets voice communications", IEEE Trans. Acoustic, Speech, and Sig. Processing, Vol. ASSP-34, no. 6, Dec. 1986, pp. 1440-48.
- [14] ITU, *ITU-T G.723.1 : Codeur de Signaux Vocaux à Double Débit pour Communications Multimédia acheminées à 5.3 kbit/s et à 6.3kbit/s*, ITU 1996.
- [15] J.Wang and J.D.Gibson, « Performance comparaison of intraframe and interframe LSP quantization in packet network ». Proc. 2000 IEEE Workshop on speech Coding, Delevan, WI, USA, Septembre 2000.
- [16] J. L. Ramsey, "Realisation of optimum interleavers," IEEE trans. Info. Theory, vol IT-16 May1970, pp. 338-45.
- [17] O. J. Wassem et al., "the effects of waveform substitution on the quality of PCM packet communication", IEEE Trans. Acoustic, Speech, and Sig. Processing, Vol. 36, no. 3, Mar. 1988, pp. 342-48.
- [18] H. Sanneck et al., "a new technique for audio packet loss concealment," IEEE Global Internet 1996, IEEE, Dec. 1996, pp. 48-52.
- [19] F. K. Soong and B. H. Juang, "Line Spectrum pair (LSP) and speech data compression," in Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Processing, (San Diego, Ca, USA), Mar. 1984, pp. 1.10.1-1.10.4.
- [20] P. Kabal and R. P. Ramachandran, "The Computation of Line Spectral Frequencies Using Chebyshev Polynomials," ", IEEE Trans. Acoustic, Speech, and Sig. Processing, Vol. ASSP-34, no. 6, Dec. 1986, pp. 342-48.

- [21] G. Fant, "Acoustiic theory of speech production, s-Gravenhage, Mouton, 1960.
- [22] J. Q. Walker and J. T. Hicks from NetIQ Corporation." The Essential Guide toVoIP Implementation and Managemet"
- [23] J; F; Susbielle, "Internet multimédia et temps réel", Eyrolles, 2000
- [24] G.Pujolle, "Les Réseaux; 3ème édition", 2001, 1069 p
- [25] D.Comer , "TCP/IP Architecture Protocoles Applications", 4ème édition, 2001, 821p
- [26] R. Jagadeesan, Cisco Systems, Inc," packet Loss Model", Feb 2000.