

UNIVERSITE DE SAAD DAHLEB DE BLIDA

Faculté des Sciences de l'Ingénieur

Département d'Electronique

MEMOIRE DE MAGISTER

Spécialité : Image et Parole

CODAGE DE LA PAROLE PAR LES FONCTIONS DE SIGNATURE

Par

BOUNIF Aouda

Devant le jury composé de

| | | |
|----------------------------|------------------------------------|---------------|
| Dr H.SALHI | Maître de conférence, U. de Blida | Président |
| Dr M ^{me} .GUERTI | Maître de conférence, E.N.P, Alger | Examineur |
| Dr M.BENSEBTI | Maître de conférence, U. de Blida | Examineur |
| Pr A.GUESSOUM | Professeur, U. de Blida | Rapporteur |
| M ^{me} A. CHENTIR | Chargé de cours, U. de Blida | Co-rapporteur |

Blida, 28 février 2007

ملخص

تشهد الإتصالات الرقمية تطورا هائلا، و أصبح كلا، من ترميز و تقليص المعلومات، مسموعة أو مرئية كانت، أمرا ضروريا. من أجل هذا وجب البحث عن طرق جديدة و فعالة لهذا الغرض.

يتناول هذا العمل دراسة تقنية لتقليص الإشارة الصوتية، تعتمد على الترميز بالتحويل. يستعمل تحويل خطي يعرف بـ "تحويل كرهونن- لويف" للحصول على قاعدة أشعة متعامدة و متجانسة، مقلصة الأبعاد، تسمى بـ: « د وال الإمضاء »

يتم الحصول على هذه الد وال بإستعمال الخصائص الإحصائية للصوت. نستطيع تركيب أي صوت بواسطة هذه الد وال حيث يمكن تمثيل كل قطعة صوتية بدالة إمضاء واحدة فقط، مضروبة في معامل. إذن نحصل على نسبة تقليص حقيقية للإشارة. من خلال تحليل الإشارة الصوتية وجدنا أن الإرتياب التريبيعي المتوسط بين القطعة الأصلية و المركبة يكون معتبرا بالنسبة للأصوات المتحركة. لتقليص هذا الإرتياب إستعملنا نظام هجين يسمح بتركيب القطع المهموسة بمعامل و دالة إمضاء واحدة و القطع المتحركة بمعاملين و دالتين. في الحالة الأخيرة نحتاج إلى مجموعتين من د وال الإمضاء. تم تقييم مجموعة الخوارزميات على قاعدة معلومات من الأصوات الطبيعية.

الكلمات مفاتيح : تحويل كرهونن-لويف، ترميز و تقليص المعلومات، الإشارة الصوتية، د وال الإمضاء، الإرتياب التريبيعي المتوسط.

RESUME

Les communications numériques sont en plein développement, le codage et la compression des informations audio-visuelles sont devenus nécessaires. Donc, il faut chercher des méthodes nouvelles et efficaces.

Notre travail consiste à étudier une technique de compression de la parole basée sur le codage par transformée. On utilise une transformation linéaire appelée " transformée de Karhunen-Loève : KLT " pour obtenir une base orthonormale de vecteurs, de dimension réduite, appelée " fonctions de signature ". Les fonctions de signature sont obtenues en utilisant les propriétés statistiques du signal de parole. Celui-ci peut être reconstruit par les fonctions de signature.

Chaque trame du signal peut être représentée par une seule fonction de signature multipliée par un coefficient. D'ici un taux de compression réel du signal est atteint. En analysant le signal de parole, on a trouvé que l'erreur quadratique moyenne entre signal original et reconstruit d'une trame voisée est importante. Pour minimiser cette erreur, on a utilisé un système hybride qui consiste à modéliser les trames non voisées avec un seul coefficient et une seule fonction de signature et les trames voisées avec deux coefficients. Dans ce dernier cas, on a utilisé deux ensembles de fonctions de signature. L'ensemble des algorithmes étudiés a été évalué sur des bases de données en parole naturelle.

Mots clés : transformation de Karhunen-Loeve (TKL), codage et compression de données, signal de la parole, fonctions de signature, l'erreur quadratique moyenne (EQM).

ABSTRACT

Because of the great development in digital communications, data coding and compression become a necessary task. Moreover, we must look for new and efficient methods. In this work, we study a speech compression technic, based on transform coding. We use a linear transform which is known as " Karhunen-Loeve Transform KLT" to generate orthonormal basis vectors with reduced dimension, which are known as " signature functions ". The signature functions are generated using the statistical proprieties of speech signals. Each frame of the speech signal is represented with only one signature function multiplied by a coefficient. Hence, substantial signal compression rate is achieved. Then speech signal can be synthesized with a few basic signature functions. By analyzing speech signals, we have found that the least mean square error (LMSE) between original and synthesized voiced frame is important. To reduce this error, we have used a hybrid system which allows as modeling the unvoiced frame with only one signature function multiplied by a coefficient and voiced frame with two coefficients. In this case we need two sets of signature functions. The set of algorithms is evaluated on a natural speech data basis.

Keywords: Karhunen-Loeve Transform (KLT), data coding and, speech signal, signature functions, speech compression, least mean square error (LMSE).

DEDICACES

Je dédie ce travail à:

La mémoire de mon père qui avait souhaité me voir arriver au terme de mes études.

Je voudrais exprimer tout mon amour à la plus adorable mère, grâce à qui j'ai pu réussir.

Je souhaiterais aussi dédier ce modeste travail à mes chers frères et soeurs.

A toute ma famille.

A tous mes amies et collègues de travail.

REMERCIEMENTS

Sans la contribution directe ou indirecte de certaines personnes, ce travail n'aurait pu être réalisé.

En premier lieu, je tiens à remercier le P^r GUESSOUM Abderrezak, directeur du Laboratoire de Recherche en Traitement de Signal et Imagerie (LATSI) au département d'électronique à l'USD de Blida, pour m'avoir accordé l'honneur de diriger cette thèse qui s'est déroulée au sein du laboratoire LATSI. Merci pour m'avoir indiqué souvent le bon chemin.

Je remercie aussi M^{me} A.CHENTIR chargée de cours au département d'électronique à l'USD de Blida, ma co-promotrice, pour ses nombreux conseils.

Je veux également remercier M^{er} Kaddai chargé de cours à l'université Hassiba Benbouali de chlef, pour l'aide qu'il m'a apportée.

Merci à ceux qui ont consacré leur temps pour corriger les nombreuses fautes commises dans la rédaction de ce mémoire.

Mes sincères remerciements vont également à tous les autres membres de jury pour avoir apprécié mon travail :

- D^r SALHI Hassen, Maître de conférence au département d'électronique à l'USD de Blida, pour m'avoir fait l'honneur d'être président de mon jury.
- D^r GUERTI M'hania, Maître de conférence au département d'électronique à l'école Nationale Polytechnique ENP Alger, et D^r Bensebti messaoud, Maître de conférence au département d'électronique à l'USD de Blida, pour avoir accepté d'examiner mon travail.

TABLE DES MATIERES

| | |
|--|-----|
| RESUME..... | 1 |
| REMERCIEMENTS..... | 5 |
| TABLE DES MATIERES..... | 6 |
| LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX..... | 7 |
| INTRODUCTION GENERALE..... | 9 |
| | |
| 1. LA TRANSFORMATION DE KARHUNEN-LOEVE | 13 |
| 1.1. Introduction | 13 |
| 1.2. Compression de données..... | 13 |
| 1.3. Compression par transformée orthogonale | 17 |
| 1.4. La transformée de Karhunen-Loève | 21 |
| 1.5. Application de la transformée de Karhunen-Loève à la parole | 28 |
| 1.6. Conclusion | 32 |
| | |
| 2. LES FONCTIONS DE SIGNATURES | 33 |
| 2.1. Introduction | 33 |
| 2.2. Une méthode statistique pour générer les FS | 34 |
| 2.3. Algorithme de génération de la FS | 38 |
| 2.4. Qualité de la parole et évaluation | 42 |
| 2.5. Sélection de la longueur de tranche LF par le « MOS » | 43 |
| 2.6. Capacité de compression du signal de la parole | 44 |
| 2.7. Conclusion | 45 |
| | |
| 3. SIMULATIONS ET RESULTATS | 46 |
| 3.1. Introduction | 46 |
| 3.2. Les sons ciblés | 46 |
| 3.3. Compression par la KLT traditionnelle (algorithme KLT1) | 47 |
| 3.4. Algorithme FS (pour une seule phrase) | 51 |
| 3.5. Algorithme FS (pour plusieurs phrases) | 59 |
| 3.6. Résultat comparatif | 77 |
| 3.7. Conclusion..... | 78 |
| | |
| CONCLUSION GENERALE ET PERSPECTIVES | 79 |
| | |
| APPENDICES | 82 |
| | |
| REFERENCES BIBLIOGRAPHIQUES | 105 |

LISTE DES ILLUSTRATIONS, GRAPHIQUES ET TABLEAUX

| | | |
|--------------|--|----|
| Figure 1 | Evolution des styles de communications et des contenus de messages. | 10 |
| Figure 2 | Localisation de la FS à l'intersection de la compression vidéo, de la vision par ordinateur et de la reconnaissance de formes. | 11 |
| Figure 1.1 | Classification des techniques de compression de données | 15 |
| Figure 1.2 | Le schéma d'un système de compression par transformée orthogonale. | 19 |
| Figure 1.3 | Application de la transformée de Fourier discrète a un signal $x(n)$ (En haut). $y(n)$ (en bas) est le signal transformé. | 21 |
| Figure 1.4 | Génération d'un processus $x(n)$ fortement corrélé. | 29 |
| Figure 1.5 | a) Un processus gaussien fortement corrélé, b) le processus après la transformation de décorrélation | 29 |
| Figure 1.6 | Décomposition du signal $x(n)$ en fenêtres de longueur m | 30 |
| Figure 1.7 | Représentation temporelle des deux premiers vecteurs propres | 31 |
| Figure 1.8 | Représentation de trois signaux reconstruits avec les deux premières composantes principales, comparés à leurs vecteurs originaux | 31 |
| Figure 2.1 | Segmentation d'un signal discret | 34 |
| Figure 2.2 | Séquences de signature prédéfinies pour $LF=16$, $S=[S_1, S_2, \dots, S_{15}]$ | 40 |
| Figure 3.1 | Variation de l'eqm en fonction du nombre de coefficients retenus. | 48 |
| Figure 3.2 | Signal original (en haut) et signal reconstruit (en bas) de la phrase : « الدرس الثاني العرب قبل » a) en utilisant 3 coefficients par trame. b) en utilisant 5 coefficients par trame. | 49 |
| Figure 3.3.a | Détection de voisement d'un signal de parole arabe. | 54 |
| Figure 3.3.b | Détection de voisement d'un signal de parole français | 55 |
| Figure 3.4 | Les fonctions de signatures : a) pour $LF=40$, b) pour $LF=16$ | 58 |
| Figure 3.5.a | Parole originale et synthétisée des phrases 1, 2,10 pour $LF=16$. | 61 |

| | | |
|---------------|---|----------|
| Figure 3.5.b | Parole originale et synthétisée des phrases 1,3 pour LF=16. | 63 |
| Figure 3.6 | L'erreur de reconstruction de la phrase 1, prononcée par les différents Locuteurs. a) LF=16, b) LF=24. | 70 |
| Figure 3.7 | L'erreur de reconstruction des dix phrases(LF=16) prononcées par a) Locuteur femme1 b) Locuteur homme1 c) Locuteur homme10 | 71 |
| Figure 3.8 | Erreur de reconstruction de la phrase 1 avec prétraitement, LF=16. | 75 |
| Figure 3.9 | Evaluation de l'EQM de reconstruction des dix phrases, LF=16 a) Locuteur F1, b) Locuteur H1 c)) Locuteur H10 | 76 |
| Tableau 3.1 | LMSE et MOS pour différentes longueurs de trame LF. | 50 |
| Tableau 3.2 | Erreur quadratique moyenne des trames de parole (en retenant un seul coefficient par trame) a) Parole en arabe b) Parole en français | 52 55 |
| Tableau 3.2.c | Erreur quadratique moyenne des trames de parole en français (en retenant deux coefficients pour les trames erronées). | 56 |
| Tableau 3.3 | Evaluation de la qualité de parole (trames sans recouvrement) en fonction de LF et de N_C . | 57 |
| Tableau 3.4 | Nombre de fonctions de signatures retenues. a) Trames sans recouvrement. b) Trames avec recouvrement de 50% | 58 |
| Tableau 3.5 | Evaluation de la qualité de synthèse par des trames avec 50 % de recouvrement . | 59 |
| Tableau 3.6.a | Résultats de synthèse de 3 phrases prononcées par homme1 | 60 |
| Tableau 3.6.b | Résultats de synthèse de 3 phrases prononcées par femme1 | 62 |
| Tableau 3.7 | LMSE (obtenue avec un seul et deux coefficients) pour LF=16 a) Phrases (1, 2,3,4,5) b) Phrases (6,7,8,9,10) | 65 66 |
| Tableau 3.8 | LMSE (obtenue avec un seul et deux coefficients) pour LF=24 a) Phrases (1, 2,3,4,5) b) Phrases (6,7,8,9,10) | 67 68 |
| Tableau 3.9 | LMSE (obtenue avec un seul et deux coefficients) avec prétraitement LF=16. a) Phrases (1,2,3,4,5) b) Phrases (6,7,8,9,10) | 73 74 |
| Tableau 3.10 | Comparaison entre la technique proposée et la LPC-10 E (FS 1015) | 77 |

INTRODUCTION GENERALE

La vitesse des communications ne peut pas être augmentée sans l'aide des techniques de compression. Tous les types des signaux intéressants pour les transmissions : les logiciels, les textes, la parole, la musique, les images, doivent être comprimés. La compression ne doit pas conduire à des distorsions saisissables et doit être réalisée avec des taux de compression les plus élevés possibles. Le nombre d'opérations et le volume de mémoire requis par les algorithmes de compression ne doivent pas être très grands.

Notre but est d'analyser l'apport de la théorie de la transformation de Karhunen-Loève aux techniques de compression à pertes contrôlées [7], [10], [9], [12].

L'objectif de ce travail est de faire l'étude et l'essai d'un algorithme de compression et de restitution de signaux de parole arabe, à l'aide d'une base de fonctions prédéfinies dites « Fonctions de Signature : FS ». Il s'agit d'une méthode qui utilise les propriétés statistiques du signal de parole pour trouver les fonctions de signature qui modélisent les signaux audio [1], [2], [3], [4].

Les travaux sur le traitement de la parole ont une histoire de plus de cinquante ans : c'est en 1939 que, pour la première fois, un chercheur des laboratoires Bell aux États-Unis, M. Dudley, proposa un appareil nommé Vocoder visant à coder électriquement le signal vocal selon des paramètres limités, puis à le transmettre avec un débit d'informations réduit et à le reproduire enfin dans un système de synthèse effectuant l'opération réciproque du Vocoder. Plus tard, le même chercheur réalisa un autre système de synthèse électrique de la parole, actionné par un clavier. Enfin apparut, en 1947, le sonagraphe, premier analyseur de la composition du signe vocal en fréquence et en amplitude.

L'évolution des vitesses des processeurs actuels permet d'essayer quelques nouveaux traitements de complexité supérieure. Trouver de nouveaux algorithmes ou ajuster les paramètres des algorithmes traditionnels, avec le compromis (économie-qualité), par une analyse de haut niveau mais qui puisse être implantée en temps réel, est l'objectif de nombreux travaux. L'exécution des algorithmes de compression dépend de l'application et du modèle de données utilisées [5]. Le domaine d'application choisi ici est la communication audio.

Depuis sa création en 1990 avec la définition du http, le WWW a révolutionné presque chaque aspect de communication et du calcul personnel et professionnel. Le GSM est devenu le protocole le plus important pour le téléphone mobile. Cet environnement technologique crée une situation dans laquelle il est nécessaire de transmettre une quantité de données par les lignes de communications avec une largeur de bande très limitée, qu'elles soient par radio ou par fil. Ceci mène aux demandes techniques suivantes [5] :

- Représentation de données efficace, élimination d'information redondante.
- Protection de la sécurité et de l'espace privé pendant la communication et la transmission ou l'accès aux données.
- Accès omniprésent et instantané / permanent à l'information.

La figure 1 illustre les changements des habitudes de communications (mobile) pour un tel genre de systèmes.

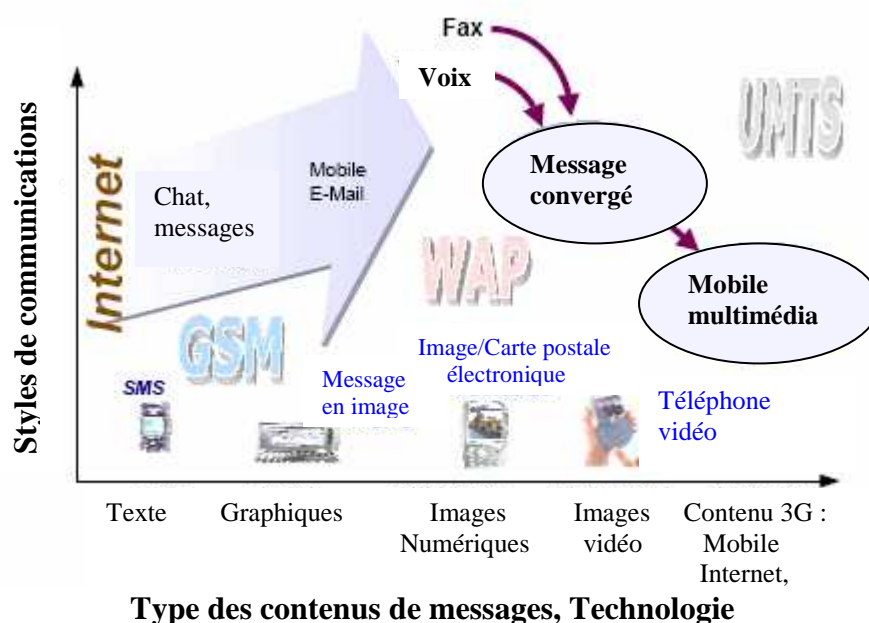


Figure1: Evolution des styles de communications et des contenus de messages [5].

Nous notons de gauche à droite le développement du texte, d'après graphiques, images numériques, et vidéo numérique à l'intégration de contenu mobile multimédia. Verticalement nous voyons l'intégration de services fax, voix, et Internet (WWW, E-Mail, Chat) séparés dans un seul service.

Les applications sont nombreuses : réseau téléphonique public, communication avec les mobiles, radiodiffusion numérique, télévision numérique,

Les nouveaux périphériques et techniques telles que les caméras pilotables avec suivi intégrées , les écrans à contact, les capteurs- intelligents, et les logiciels de reconnaissance de la parole ouvrent de nouvelles voies d'interaction homme- machine. La FS peut être appliquée dans le courrier électronique vidéo (vidéo E-mail), où une séquence est enregistrée pour être transmise comme E-mail.

La figure 2 localise la FS à l'intersection de la compression (vidéo, parole), de la vision par ordinateur et de la reconnaissance de forme [5].

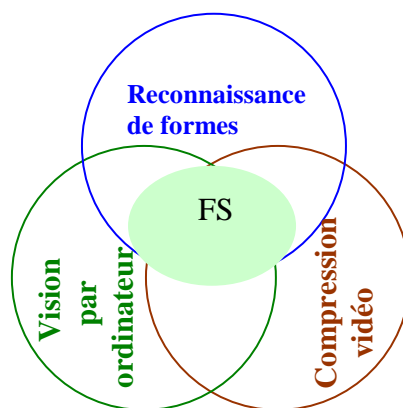


Figure 2: Localisation de la FS à l'intersection de la compression vidéo, de la vision par ordinateur et de la reconnaissance de formes [5].

Le chapitre 1 est une description de la transformation de Karhunen-Loève (KLT). Les propriétés de la transformée et la décomposition en valeurs propres, ainsi que la compaction d'énergie sont présentées. La KLT consiste à projeter le signal d'entrée sur la base orthonormée constituée par les vecteurs propres. On introduit une méthode de compression à pertes contrôlées basée sur l'utilisation d'une transformée orthogonale. Le rôle de cette transformée est la décorrélation du signal à compresser. On présente la transformée idéale, c'est la transformée de Karhunen-Loève. On propose un schéma de compression basé sur une transformée orthogonale. On prouve la convergence asymptotique de quelques transformées vers la transformée de Karhunen-Loève [7].

Le chapitre 2 développe l'algorithme du codage avec les fonctions de signatures

« FS ». Puisque la façon dont les données sont perçues, détermine l'approche pour leur traitement, nous étudions la compression de parole par le modèle de données utilisé. En conclusion, en raison de leur importance et leur fréquente utilisation, les mesures de qualité de reconstruction les plus communes telles que MSE, SNR et MOS sont discutées.

Le chapitre 3 présente les résultats qui montrent que la FS est une alternative pour le standard et d'autres algorithmes de compression audio. Des résultats de codage sont évalués, pour plusieurs morceaux de paroles, prononcés par différents locuteurs (hommes et femmes). On montre que, la FS produit des résultats comparables à un encodeur représentant les arrangements répandus de compression, tel que le FS 1015 LPC10E. Dans certains cas, la FS surpasse de manière significative le codec de référence en termes de qualité et/ou de compression. En effet le logiciel encodant de « FS » a été optimisé en ce qui concerne la vitesse de calcul. Des résultats comparatifs, avec la KLT traditionnelle, sont donnés en utilisant la durée de calcul comme mesure. Le codec « FS » présenté dans cette dissertation peut être vu comme une première étude de faisabilité d'un codec de parole basé sur la signature dans une séquence de son en remplacement de la corrélation statistique entre les valeurs des échantillons.

Nous terminerons par une conclusion générale qui résumera le travail effectué et fournira quelques perspectives pour de futurs travaux de recherche dans ce domaine.

CHAPITRE 1

LA TRANSFORMATION DE KARHUNEN-LOEVE (KLT)

(ANALYSE EN COMPOSANTES PRINCIPALES : PCA)

1.1. Introduction

Dans ce chapitre, nous donnerons un aperçu sur la transformation de Karhunen-Loève qui représente le fondement théorique sur lequel repose la méthode de compression FS. Les standards actuels d'ITU-T et d'ISO, pour le codage d'information, se fondent principalement sur la relation statistique entre les échantillons [5]. Les méthodes pour l'extraction de contenu de parole, basées sur la signature, sont plus flexibles, face à un contenu de parole changeant. Elles sont plus faciles à utiliser que des modèles géométriques [5].

L'objectif de l'analyse de données consiste, à obtenir un résumé descriptif, d'un ensemble de n observations de p variables numériques. Les valeurs de n et de p étant suffisamment grandes pour qu'il soit nécessaire de trouver une méthode permettant de « synthétiser » toute l'information. Parmi les exemples d'applications, on peut noter:

- L'analyse de données classiques: par exemple, le dépouillement d'enquêtes sociologiques (par exemple 1000 personnes interrogées sur la base d'un questionnaire de 60 questions).
- Le traitement d'images :
Image 512 x 512 pixels - 3 bandes spectrales (exemple d'une sous scène SPOT XS). Et on cherche à réduire toute l'information à une seule image 512 x 512 en y concentrant le maximum d'information.
- Le traitement de parole : compression d'un signal de parole par élimination de l'information redondante.

1.2. Compression de données

Pour la compression de la parole, il y a déjà une variété de techniques. Plusieurs normes ont été recommandées par l'UIT-T pour le réseau téléphonique public. Depuis 1972, la norme internationale G.711 précise un codage MIC (PCM) correspondant à un débit de 64 kbit/s: l'amplitude des échantillons est simplement quantifiée sur 8 bits après une compression

de type non-linéaire. Depuis 1984, la norme G.721 définit le codage MICDA correspondant à un débit de 32 kbit/s: on ne quantifie plus directement l'amplitude de l'échantillon mais la différence entre l'amplitude et une valeur prédite déterminée par un filtrage de type adaptatif. Un codeur à 16 kbit/s basé sur des techniques de modélisation et de quantification vectorielle, a été sélectionné par l'UIT-T en 1991. Cette norme G.728 est également appelée LD-CELP mettant en évidence que c'est un codeur de type CELP et qu'il présente un faible délai de reconstruction, propriété particulièrement importante pour un échange téléphonique.

Les communications avec les mobiles sont en plein développement actuellement. La nature du canal de transmission, une liaison radio, réclame d'économiser au maximum la largeur de bande du signal transmis pour permettre un grand nombre d'utilisateurs. Depuis 1989, il existe une norme européenne dite GSM. L'ouverture du service Itineris par France Télécom a eu lieu le 1er juillet 1992. Cette première génération est basée sur les techniques TDMA et sur le codeur de source RPELTP à 13 kbit/s. Cette première génération ne permet qu'une multiplication par trois environ des capacités de ce réseau par rapport au réseau analogique (Radiocom 2000 en France). Pour permettre une multiplication par dix et même plus, le ETSI a choisi les standards de la deuxième génération $\frac{1}{2}$ GSM.

Tout en étant très efficaces, ces standards ont leurs limites de rentabilité imposant des compensations entre la qualité de reconstruction et le taux de compression. Ceci a naturellement mené, grâce à des efforts de recherche à repousser ces limites plus loin.

Les techniques de compression diffèrent l'une de l'autre et on peut les diviser en groupes d'approches fondamentales comme le montre la figure 1.1. Cependant, il n'y a aucun algorithme universel de compression qui comprime efficacement toutes les sortes de données.

1.2.1. Réduction de la redondance

Plusieurs travaux de recherche ont été développés pour le traitement du signal et la perception auditive.

Dans les années (1959, 1961, 1989) Horace Barlow développa des recherches significatives sur la perception [6]. Partant des recherches récentes, Helmholtz et Match ont observé que l'environnement dans lequel on vit, est un grand facteur de développement de notre perception. Malheureusement, à cette époque il n'y avait ni fondations mathématiques ni moyens pour développer de telles observations.

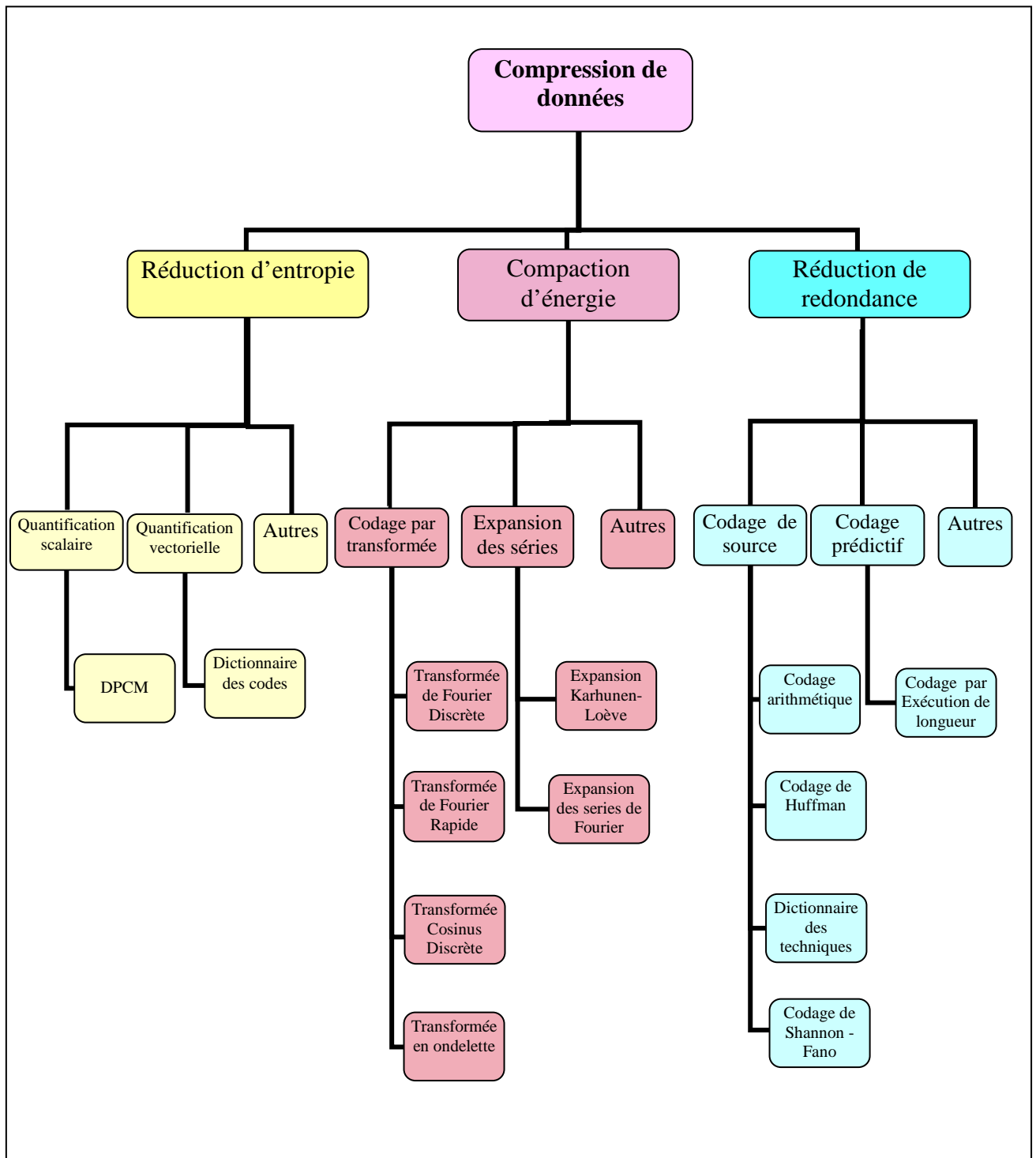


Figure 1.1 : Classification des techniques de compression de données [5]

En s'appuyant sur la théorie de l'information créée par Shannon en 1940, un nouveau langage a été utilisé pour exprimer les idées de recherches sur la perception. C'est en 1954 qu'Attneave utilisa, pour la première fois, l'information principale, la capacité du canal et la redondance pour exprimer les processus de perception (son observation fut que l'information pertinente des images naturelles est les bords de l'image alors que le reste n'est que de l'information redondante) [6].

Barlow (1959, 1961, 1989) exprima les mêmes idées et explora la structure possible de notre système nerveux pour exécuter de telles optimisations. Il donna plus de détails sur l'observation d'Attneave (la redondance sensorielle d'entrée est cruciale pour la perception). Il a suggéré que le mécanisme sensoriel peut donner « l'économie de pensée » de Mach et finit par exécuter le codage factoriel. Plus tard, Barlow laissa l'idée de « l'économie principale » tout en considérant que l'exploitation de la redondance est très importante [6].

En 1960, Watanabe introduisit l'idée que la déduction inductive a été reliée à la MLD [6]. Ceci montre qu'il y a une variété de modèles possibles pour décrire l'ensemble de données. Le modèle préféré est celui qui produit la plus courte et compacte description.

Toutes ces idées furent des thèmes importants à la fin de 1980 où les recherches menées par Linsker (1986a, 1986b, 1988) et Field (1987), utilisèrent les statistiques pour comprendre le système visuel au niveau neuronal. En partant de l'opinion que le système perceptuel est défini pour s'adapter à son environnement, une approche a été détaillée par Atick (1991) en utilisant la théorie de l'information et l'adaptation écologique pour expliquer le traitement sensoriel.

Des expériences ont été conduites par Atick et Redlich (1990) et Redlich (1993), qui explorèrent expérimentalement les possibilités de notre système nerveux à exploiter la redondance. Leurs simulations ont été réussies, en montrant une cartographie de traits, qui sont en accord avec les fonctions de transfert de notre traitement neuronal. Bien que convaincantes, ces expériences étaient compliquées à cause du développement limité des techniques de traitement à cette époque.

La majorité des développements récents vinrent de Olshausen et Field (1996), Bell et Sejnowski (1997), et Hyvärinen et Hoyer (2000), qui ont utilisé des algorithmes d'optimisation de la théorie moderne de l'information pour analyser abondamment les images naturelles. En essayant d'obtenir différentes interprétations de la réduction de redondance, leur analyse donna des décompositions très semblables au champ de réception de notre système visuel. Ils

ont fait un argument très convaincant que le codage abondant et factoriel sont des fonctions très liées à la perception et à son développement [6].

La majorité des recherches sur la perception auditive est incarnée par le travail de Bregman (1990). Son livre explique le fonctionnement d'analyse d'une scène audio à l'aide de soigneuses expériences. Une importante masse de recherche, pour analyser une scène audio, a été influencée par ce travail et a appliqué la translation directe de ses observations sur des programmes d'ordinateur. De telles implémentations aspirent à construire des systèmes pouvant exécuter l'audio analyse, des jugements sur les objets existants et le traitement en un temps mémoire court.

Le travail majeur dans la perception auditive inclut les systèmes construits par Vercoe et Cumming(1988), Duda et Al(1990), Mellinger(1991), Cooke(1991), Brown(1992) et Ellis (1992,1996). Ses systèmes incluent une étape de décomposition inspirée de notre mécanisme d'audition et passèrent ensuite, sélectivement, par des groupes de composantes de cette décomposition pour extraire les objets individuels. A l'écart de toutes les recherches mentionnées, il avait des travaux, sur le traitement et l'analyse audio, rattachés au processus d'audition humaine. La majorité de ces derniers était spécifique aux applications de reconnaissance de configuration et de traitement du signal (Kars et Brandenburg (1998) et Rhoads(1996)).

En pratique, on n'a pas besoin de toutes les informations possibles d'un contenu de parole. On devrait se concentrer sur ce qui est important dans le signal. Cette distinction est importante parce que la façon dont on regarde les données, détermine l'approche pour leur traitement. C'est chercher donc, la représentation optimale de la donnée, sans perdre de l'information. Il s'agit de la compression d'information. Cette méthode de compression permet le calcul d'un espace de base orthonormale, à partir d'un ensemble de signaux de paroles. Si on réussit à concentrer le maximum de contraste sur un minimum d'axes, on dira qu'on a une compression du signal : c'est la procédure qui va être détaillée dans la section suivante.

1.3. Compression par transformée orthogonale

Il s'agit d'une compression de données à perte contrôlée. Le rôle de la transformée orthogonale est de décorréler le signal qui doit être traité. Ainsi, chaque échantillon du

signal obtenu, contient seulement sa propre information et aucune information spécifique pour les échantillons voisins. Ainsi, par l'élimination de certains échantillons, on ne perd pas que leur information spécifique (l'information contenue dans les échantillons restés n'est pas affectée).

Une méthode de compression à perte d'information contrôlée a les pas suivants:

1. On calcule la transformée orthogonale directe $y(n)$, du signal qui doit subir la compression, $x(n)$.
2. On réalise la compression en rejetant les échantillons du signal $y(n)$ inférieurs à un seuil imposé. On obtient le signal $y_s(n)$. On fait la quantification et on obtient le signal $\hat{z}(n)$. Celui-ci est le résultat de la compression.
3. On calcule la transformée inverse du signal $\hat{z}(n)$ et on obtient le signal $\hat{x}(n)$. Celui-ci représente le signal reconstruit par l'opération inverse (décompression).

La valeur du seuil est choisie telle que l'erreur quadratique moyenne d'approximation du signal $x(n)$ par le signal $\hat{x}(n)$ serait inférieure à un certain pour-cent de l'énergie du signal $x(n)$.

Comme chaque transformée orthogonale conserve l'énergie, les relations suivantes sont valables:

$$\begin{aligned} E_x &= \sum_{k=0}^{N-1} x^2(k) = \sum_{k=0}^{N-1} y^2(k) \\ E_{\hat{x}} &= \sum_{k=0}^{M-1} \hat{x}^2(k) = \sum_{k=0}^{M-1} \hat{z}^2(k) \end{aligned} \quad (1.1)$$

Où :

- N représente le nombre d'échantillons du signal $x(n)$.
- M indique le nombre d'échantillons du signal dans l'espace transformé après compression

On peut écrire aussi:

$$y_s(n) = \begin{cases} y(n), & \text{si } |y(n)| > p \\ 0, & \text{ailleurs} \end{cases} \quad (1.2)$$

Où p représente le seuil d'élimination des échantillons faibles.

Soit $y_{s0}(n)$, la séquence obtenue par l'organisation des échantillons du signal $y_s(n)$ dans l'ordre décroissant.

L'erreur quadratique moyenne d'approximation du signal $x(n)$ par le signal $\hat{x}(n)$ est donnée par:

$$\varepsilon = \sum_{k=M}^{N-1} y_{s0}^2(k) \quad (1.3)$$

La valeur du nombre M peut être obtenue comme solution de l'équation:

$$\max_{M \in \mathbb{Z}} \varepsilon = \frac{E_x}{100} \quad (1.4)$$

La transformée orthogonale qui réalise la meilleure décorrélation est la transformée de Karhunen-Loève. Le codage par transformée consiste à envoyer un vecteur $X = (X_1, X_2, \dots, X_n)$ de n échantillons dans une transformée (inversible) T . On obtient ainsi, un vecteur $Y = T(X)$ dans le domaine transformé.

Le schéma d'un système basé sur l'utilisation d'une transformée orthogonale est celui présenté par la figure suivante:

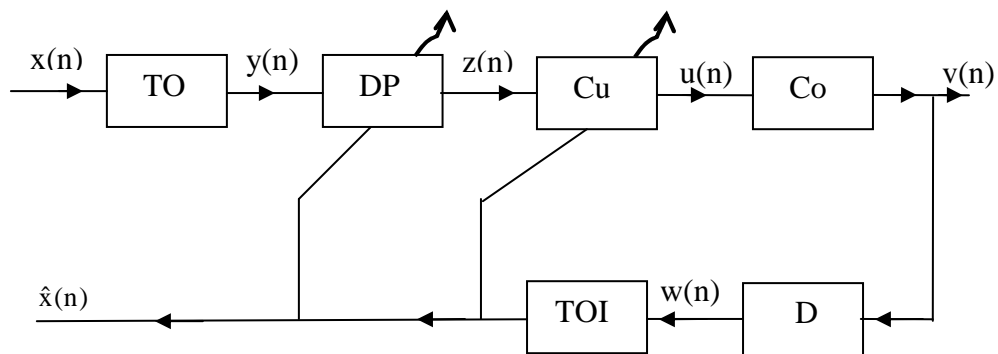


Figure 1.2 : Le schéma d'un système de compression par transformée orthogonale [7].

On a utilisé les notations suivantes :

TO - système de calcul de la transformée orthogonale ;

DP - détecteur de seuil (c'est le système qui élimine les échantillons de valeur petite, tous les échantillons inférieurs à un certain seuil sont éliminés) ;

Cu - système de quantification : Chaque échantillon y_i , en sortie de la transformée, est ensuite quantifié par un quantificateur Q_i sur M_i niveaux de quantification. On a ainsi un taux de quantification de $R_i = \log_2 M_i$ bits par échantillon [9] ;

Co - système de codage ;

D - système de décodage, l'inverse du système Co. Le dé quantificateur $Q^{-1}=D$ reconstruit les échantillons w_i ;

TOI - système de calcul de la transformée inverse à celle calculée en utilisant le bloc TO.

La transformée inverse T^{-1} est finalement appliquée au vecteur w pour fournir le signal global reconstruit $\hat{X} = T^{-1}.w$.

Le taux de quantification global R est le nombre moyen de bits codés par échantillon de source

X , c'est-à-dire $R = \frac{1}{n} \sum_i R_i$ [9].

Ici, la quantification scalaire porte sur des coefficients transformés d'un vecteur de source.

Ce système, muni avec une transformée discrète en KL, sera analysé, dans ce qui suit. Cette analyse montrera l'apport de la théorie de la KLT à la compression. Les principaux signaux de cette figure sont:

- Le signal à traiter $x(n)$.
- Le signal obtenu après la compression, $v(n)$.
- Et le signal reconstruit, obtenu après l'opération de décompression, $\hat{x}(n)$.

Les échantillons du signal $x(n)$ sont corrélés. Cela signifie que l'information contenue dans l'échantillon courant est aussi présente dans les échantillons voisins. C'est le motif pour lequel, par l'élimination de l'échantillon courant n'est pas affectée seulement l'information, contenue dans cet échantillon, mais aussi l'information contenue dans les échantillons voisins. Le rôle de la transformée orthogonale est de décorrélérer le signal. Après l'application de la transformée orthogonale, on obtient un nouveau signal.

La dépendance du contenu informationnel de l'échantillon courant de ce nouveau signal, $y(n)$, de l'information contenue dans les échantillons voisins est plus faible. Donc, par l'élimination de l'échantillon courant, l'information contenue dans les échantillons voisins est moins affectée que dans le cas du signal $x(n)$. L'élimination de l'échantillon courant conduit à une perte d'information moins importante si la valeur de l'échantillon était moins importante.

La transformée orthogonale utilisée, dans l'exemple de la figure 1.3, est la valeur absolue de la transformée de Fourier discrète. En analysant cette figure, on constate que par l'application de la transformée de Fourier discrète, l'entière information s'est groupée en six

échantillons de valeur significative. En éliminant n'importe quel groupe d'échantillons parmi ceux à l'indice compris entre 60 et 190, le contenu informationnel du signal $y(n)$ ne se modifie pas. Evidemment, on ne peut pas dire la même chose pour le signal $x(n)$, tous les échantillons à l'indice entre 60 et 190 ayant une contribution importante à la forme de ce signal.

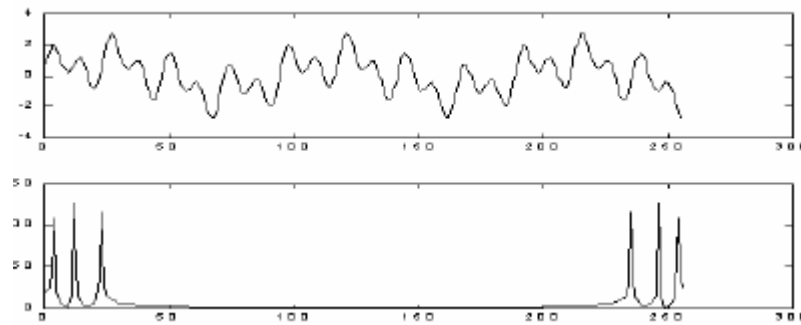


Figure 1.3 : Application de la transformée de Fourier discrète à un signal $x(n)$ (En haut). $y(n)$ (en bas) est le signal transformé[7].

La transformée qui réalise la décorrélation la plus importante d'un signal est la transformée de Karhunen-Loève. Malheureusement, il n'y a pas d'algorithmes rapides pour le calcul de cette transformée. Son calcul exige l'inversion d'une matrice. En plus, il y a des signaux pour lesquels cette matrice est singulière. En pratique, on utilise des transformées sous optimales, comme par exemple la transformée en cosinus discrète ou la transformée en ondelettes discrète. Ces deux transformées convergent asymptotiquement vers la transformée de Karhunen-Loève [7], [9].

1.4. La transformée de Karhunen-Loève

1.4.1. Définition

Soit le signal aléatoire, $x(\tau)$, de durée D et sa décomposition en série:

$$x(\tau) = \lim_{N \rightarrow \infty} \sum_{k=1}^N c_k V_k(\tau) \quad 0 \leq \tau \leq D \quad (1.5)$$

Où la limite est prise en moyenne quadratique et les fonctions $V_k(\tau)$ sont orthogonales.

On s'intéresse de l'approximation du signal $x(\tau)$ par le signal $x_N(\tau)$ avec:

$$x_N(\tau) = \sum_{k=1}^N c_k V_k(\tau) \quad 0 \leq \tau \leq D \quad (1.6)$$

On désire la construction de cette approximation tel que le signal à temps discret c_k n'ait pas des éléments corrélés.

La relation (1.5) devient :

$$\lim_{N \rightarrow \infty} E \left\{ (x_N(\tau) - x(\tau))^2 \right\} = 0 \quad (1.7)$$

Où on a noté avec E l'opérateur d'espérance statistique.

Les coefficients c_k peuvent être calculés avec la relation :

$$c_k = \int_0^D x(\tau) V_k^*(\tau) d\tau \quad (1.8)$$

Si l'espérance statistique du signal $x(\tau)$ est nulle alors:

$$E \{ c_k \} = \int_0^D E \{ x(\tau) \} V_k^*(\tau) d\tau = 0 \quad \forall k \in Z \quad (1.9)$$

Si les coefficients c_k ne sont pas corrélés alors:

$$E \{ c_k c_l^* \} = \lambda_k \delta(k-l) \quad (1.10)$$

Ou:

$$E \left\{ \int_0^D x(\tau) V_k^*(\tau) d\tau \int_0^D x^*(u) V_l(u) du \right\} = \lambda_k \delta(k-l)$$

Or:

$$\int_0^D V_l(u) \left(\int_0^D E \{ x(\tau) x^*(u) \} V_k^*(\tau) d\tau \right) du = \lambda_k \delta(k-l)$$

Mais, $E \{ x(\tau) x^*(u) \}$ représente l'autocorrélation $R_x(\tau, u)$ du signal aléatoire $x(\tau)$. La dernière relation devient:

$$\int_0^D V_l(u) \left(\int_0^D R_x(\tau, u) V_k^*(\tau) d\tau \right) du = \lambda_k \delta(k-l) \quad (1.11)$$

Grâce à l'orthogonalité des fonctions $V_k(\tau)$, la dernière condition est satisfaite aussi si:

$$\int_0^D R_x(\tau, u) V_k^*(\tau) d\tau = \lambda_k V_k^*(u)$$

Ou, parce que la fonction d'autocorrelation est réelle:

$$\int_0^D R_x(\tau, u) V_k(\tau) d\tau = \lambda_k V_k(u) \quad (1.12)$$

On a obtenu ainsi, une équation intégrale Fredholm. $R_x(\tau, u)$, représente le noyau de l'équation, $V_k(\tau)$ sont les fonctions propres et λ_k les valeurs propres.

Sachant l'autocorrélation $R_x(\tau, u)$ du signal à décomposer $x(\tau)$ on peut résoudre l'équation (1.12), et trouver les fonctions $V_k(\tau)$ et les valeurs λ_k .

On peut montrer que la transformée Karhunen-Loève, définie dans la relation (1.6) rend à sa plus petite valeur l'erreur quadratique moyenne d'approximation:

$$E(e^2) = E\{[x(\tau) - x_N(\tau)]^2\} \quad (1.13)$$

Il n'existe pas une autre transformée (décomposition) décrite par une somme à N termes qui conduit à une erreur moyenne quadratique d'approximation du signal original, inférieure à l'erreur introduite par la transformée de Karhunen-Loève.

On présente dans la suite deux exemples.

Soit, pour le commencement, le signal à traiter est un bruit blanc.

Dans ce cas:

$$R_x(\tau, u) = \delta(\tau, u) \quad (1.14)$$

La relation (1.11) devient:

$$\int_0^D V_1(u) \left(\int_0^D \delta(\tau, u) V_k^*(\tau) d\tau \right) du = \lambda_k \delta(k-1) \quad (1.15)$$

$$\text{Où: } V_1(0) \cdot V_k^*(0) = \lambda_k \delta(k-1) \quad (1.16)$$

Donc:

$$\lambda_1 = |V_1(0)|^2 = E\{|c_1|^2\} = \sigma_1^2 \quad (1.17)$$

Tenant compte des relations (1.9) et (1.16) on peut écrire:

$$E\{c_k c_1^*\} = E\{c_k\} \cdot E\{c_1^*\} = 0, \quad k \neq 1 \quad (1.18)$$

Donc, les échantillons du signal c_k ne sont pas seulement décorrélés, ils sont aussi indépendants. Voilà pourquoi on utilise la transformée de Karhunen-Loève pour la compression des données.

Une classe large de signaux aléatoires a la fonction de corrélation de la forme:

$$R_x(\tau, u) = e^{-|\tau-u|} = R_x(\tau - u) = R_x(v) \quad (1.19)$$

Dans ce cas, l'équation Fredholm associée (relation (1.12)) devient:

$$\int_0^D e^{-|\tau-u|} V_k(u) du = \lambda_k V_k(\tau) \quad (1.20)$$

Les solutions de cette équation sont:

$$V_k(\tau) = e^{jk\omega_0\tau} ; \quad \lambda_k = \frac{2}{1+k^2\omega_0^2} \quad (1.21)$$

Voilà la source d'intérêt pour les développements en série de Fourier.

Parce qu'il s'agit de données, la transformée de Karhunen-Loève des signaux à temps discret est très intéressante.

On se donne une source continue X, modélisée par des échantillons aléatoires.

Le signal à transformer, $x(n)$, de durée N peut être exprimé dans la forme:

$$x^T = [x(0) \dots x(N-1)] \quad (1.22)$$

Soit T, une transformée orthogonale. Si on applique cette transformée au signal $x(n)$, on obtient:

$$X = T \cdot x \quad (1.23)$$

Où:

$$X^T = [X(0) \dots X(N-1)] \quad (1.24)$$

La transformée T peut être décrite à l'aide de la matrice T, définie par:

$$T^T = [V_0 \ V_1 \dots \ V_{N-1}] \quad (1.25)$$

Où les V_k , $k=0, \dots, N-1$, sont des vecteurs orthogonaux:

$$V_k^T \cdot V_l = \delta(k - l) \quad (1.26)$$

La transformée inverse est décrite à l'aide de la matrice transposée T^T .

On peut écrire:

$$x = T^T \cdot X = [V_0 \ V_1 \ \dots \ V_{N-1}] X = \sum_{k=0}^{N-1} X(k) V_k \quad (1.27)$$

Une compression de données peut être obtenue si on réduit la dimension de l'espace transformé de N à M ($M < N$).

L'expression du signal obtenu après la compression est:

$$x_M = \sum_{k=0}^{M-1} X(k) V_k \quad (1.28)$$

L'erreur d'approximation de x par x_M est:

$$\Delta x = x - x_M = \sum_{k=M}^{N-1} X(k) V_k \quad (1.29)$$

L'erreur quadratique moyenne d'approximation est:

$$\varepsilon = E \left\{ \|\Delta X\|^2 \right\} = E \left\{ \sum_{k=M}^{N-1} \sum_{l=M}^{N-1} X(k) X(l) V_k^T V_l \right\} \quad (1.30)$$

Tenant compte de l'orthogonalité des vecteurs V_k , la dernière relation, dévient:

$$\varepsilon = \sum_{k=M}^{N-1} E \{ X^2(k) \} \quad (1.31)$$

Parce que le signal aléatoire à l'entrée, $x(n)$ est caractérisé statistiquement, par sa matrice de covariance, il est utile d'exprimer l'erreur quadratique moyenne à l'aide de cette matrice:

$$\text{cov}_x = E \{ x x^T \} \quad (1.32)$$

Tenant compte de la relation (1.27) et de l'orthogonalité des vecteurs V_k on peut écrire:

$$X(k) = \varphi_k^T x \quad (1.33)$$

Où $\varphi_k = [V_0 \ V_1 \ \dots \ V_k]$

Voilà pourquoi la relation (1.31) devient :

$$\varepsilon = \sum_{k=M}^{N-1} E \{ \varphi_k^T \cdot x \cdot (\varphi_k^T \cdot x)^T \} = \sum_{k=M}^{N-1} \varphi_k^T E(x \cdot x^T) \varphi_k \quad (1.34)$$

Où:

$$\varepsilon = \sum_{k=M}^{N-1} \varphi_k^T \text{cov}_x \cdot \varphi_k \quad (1.35)$$

Comme la matrice cov_x est connue, il faut déterminer les vecteurs orthogonaux φ_k et la transformée T qui minimise l'erreur quadratique moyenne ε . On peut trouver la solution de ce problème dans la littérature. Cette solution a été trouvée, en utilisant la méthode des multiplicateurs de Lagrange [10].

On obtient le système d'équations:

$$\text{cov}_x V_k = \lambda_k V_k \quad (1.36)$$

Donc les vecteurs V_k sont des vecteurs propres de la matrice de covariance du signal x .

La plus petite valeur de l'erreur quadratique moyenne est :

$$\varepsilon_{\min} = \sum_{k=M}^{N-1} \lambda_k \quad (1.37)$$

La transformée définie, dans la relation (1.36), est appelée la transformée de Karhunen-Loève du signal à temps discret $x(n)$.

On présente, dans la suite, une liaison entre les transformées de Karhunen-Loève à temps continu et à temps discret.

La relation (1.12) peut être écrite sous la forme:

$$\langle R_x(\tau, u), V_k(\tau) \rangle_{L^2} = \lambda_k V_k(u) \quad (1.38)$$

Avec la notation:

$$R_x(m, n) = \text{cov}_x(m, n) \quad (1.39)$$

La relation (1.38) devient:

$$\langle \text{cov}_x[m, n], V_k(n) \rangle_{L^2} = \lambda_k V_k(m), m = \overline{0, N-1} \quad (1.40)$$

Où:

$$\begin{bmatrix} \text{cov}_x[0,0] & \text{cov}_x[0,1] & \dots\dots\dots\text{cov}_x[0,N-1] \\ \text{cov}_x[1,0] & \text{cov}_x[1,1] & \dots\dots\dots\text{cov}_x[1,N-1] \\ \vdots & \vdots & \vdots \\ \text{cov}_x[N-1,0] & \text{cov}_x[N-1,1] & \dots\dots\dots\text{cov}_x[N-1,N-1] \end{bmatrix} \begin{bmatrix} V_k(0) \\ V_k(1) \\ \vdots \\ V_k(N-1) \end{bmatrix} = \lambda_k \begin{bmatrix} V_k(0) \\ V_k(1) \\ \vdots \\ V_k(N-1) \end{bmatrix} \quad (1.41)$$

Donc:

$$\text{Cov}_x V_k = \lambda_k V_k \quad (1.42)$$

La dernière relation est identique à (1.36).

Mais les coefficients C_k ne sont pas corrélés. A l'aide de la dernière relation, on peut affirmer que les coefficients $X(k)$, ne sont pas corrélés eux aussi. Voilà pourquoi, on peut affirmer que la matrice Cov_x , est diagonale.

Grâce à la relation (1.42) on peut écrire:

$$\text{cov}_x = \begin{bmatrix} \lambda_0 & 0\dots\dots\dots 0 \\ 0 & \lambda_1\dots\dots\dots 0 \\ \vdots & \vdots & \vdots \\ 0 & 0\dots\dots\dots \lambda_{N-1} \end{bmatrix} \quad (1.43)$$

La compression du signal $X(k)$, $k=0, M-1$, peut être réalisée en retenant les premières M valeurs contenues dans ce signal. On obtient le signal:

$$X_c(k) = \begin{cases} X(k), & k=0, \overline{M-1} \\ 0, & \text{ailleurs} \end{cases} \quad (1.44)$$

Le nombre $N-M$, d'échantillons $X(k)$ qui peuvent être négligés correspond au nombre de coefficients λ_k de la matrice de covariance C_x qui peuvent être négligés. Malheureusement, il n'existe pas d'algorithmes rapides pour le calcul de la transformée de Karhunen-Loève d'un signal.

Voilà pourquoi, il est très difficile de déterminer les vecteurs propres de la matrice de covariance. En pratique on utilise des transformées sous-optimales. Celles-ci convergent

asymptotiquement vers la transformée de Karhunen-Loève mais peuvent être calculées à l'aide des algorithmes rapides. On peut mentionner parmi ces transformées les suivantes:

- la transformée de Welch-Hadamard,
- la transformée en cosinus discrète, DCT,
- la transformée en ondelettes discrète.

1.4.2. Propriétés de la transformée de Karhunen-Loève [12]

La KLT crée une représentation optimale des données d'entrée en ce qui concerne le compactage d'énergie et fournit une représentation permettant la manipulation du contenu avec d'autres méthodes basées sur la perception auditive. Les propriétés qui suivent sont énoncées sans démonstration. Elles s'obtiennent pour la plupart, sans difficulté.

1. La transformée de Karhunen-Loève a été montrée comme optimale sous plusieurs aspects, ce qui fait qu'elle est prise comme base de comparaison, comme limite de performance pour d'autres transformées dites sous optimales. C'est une transformation orthogonale.
2. Elle décorrèle les coefficients de la transformée.
3. Elle concentre le plus d'énergie en peu de coefficients, ceux d'indices inférieurs de telle façon que l'erreur due à la troncature est plus petite qu'avec les autres transformations.
4. Elle minimise l'erreur des moindres carrés entre le signal original et le signal reconstruit.
5. Elle minimise l'entropie totale comparée aux autres transformées.
6. Le principal inconvénient est que le calcul des valeurs propres de la matrice de covariance et des vecteurs propres associés doit être recommencé pour chaque tranche du signal.
7. Il n'y a pas d'algorithme rapide dans le cas général.
8. Dans le cas de signaux AR du 1^{er} ordre une approximation possédant un algorithme rapide peut être trouvée, la transformée en sinus discrète (TSD)
9. D'autres transformées sous optimales sont la TCD, la DFT, la DWHT.
10. Approximation du développement de Karhunen-Loève en vue d'un algorithme plus rapide

1.5. Application de la transformée de Karhunen-Loève à la parole [16]

Dans les application de paroles et de musique [11] par exemple, moins d'un dixième (1/10) des coefficients sont nécessaires pour la reconstruction avec une dégradation imperceptible.

Voici un exemple qui illustre le codage par cette transformée :

Un processus $x(n)$ a été généré par passage, d'une séquence $w(n)$ de bruit blanc (de moyenne nulle et de variance l'unité), à travers un filtre numérique passe-bas de Butterworth (d'ordre 3 et de fréquence de coupure 0.1Hz), tel qu'il est montré par la figure 1.4. Les vecteurs des échantillons x_i sont extraits de la séquence $x(n)$ comme le montre la figure 1.6. Le filtre élimine les composantes hautes fréquences de l'entrée. Donc, le processus obtenu en sortie doit varier lentement dans le temps.

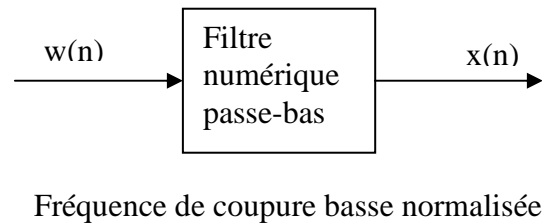


Figure 1.4 : Génération d'un processus $x(n)$ fortement corrélé [16].

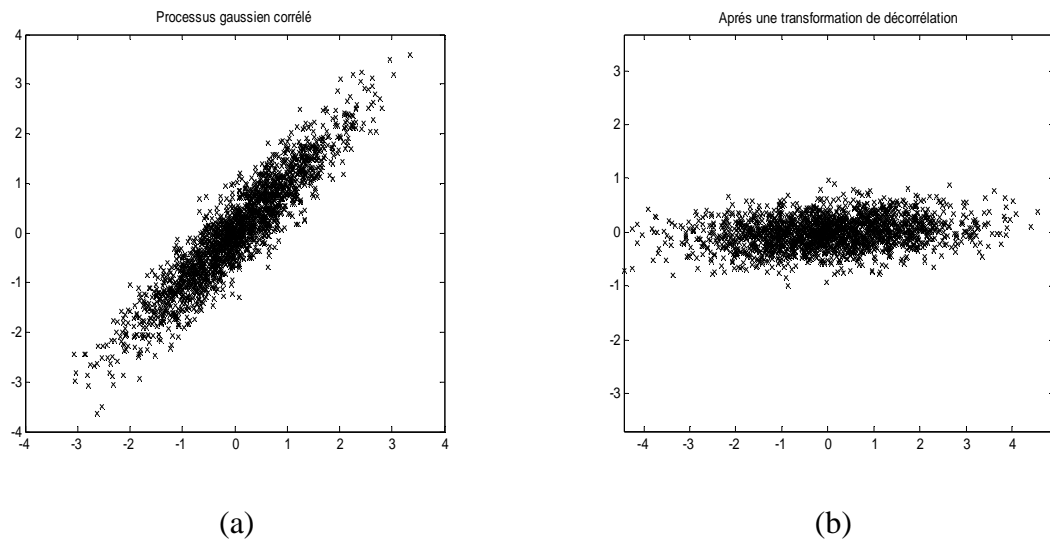


Figure 1.5 : a) Un processus gaussien fortement corrélé,
b) Le processus après la transformation de décorrélation

La figure 1.5 permet de donner une vue statistique du codage par transformée. On remarque qu'après l'application de la transformation de décorrélation, le processus a subi une rotation. Cette rotation permet de :

- 1- maximiser la variance des projections.

2- Minimiser l'erreur de reconstruction.

3- Trouver le système de coordonnées le plus « naturel » qui représente mieux la donnée.

Pour plus de détails sur la représentation de données et la recherche de la projection maximale de variance voir appendice B.

Une variable présentant une très grande variation aurait contribué beaucoup plus à la définition des vecteurs propres qu'une variable montrant peu de variations.

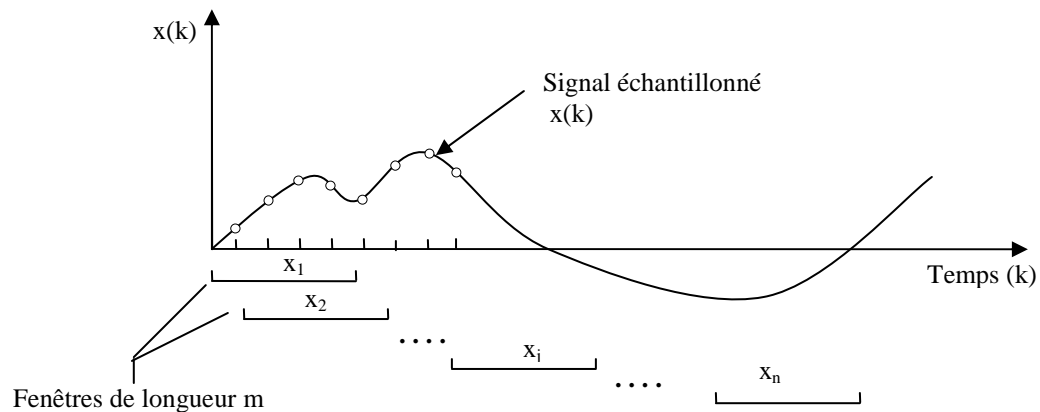


Figure 1.6 : Décomposition du signal $x(n)$ en fenêtres de longueur m [16].

Les échantillons de la $i^{\text{ème}}$ fenêtre forment le vecteur x_i , $i=1,2,\dots,N$. L'expansion KL utilise, quelques composantes principales (vecteurs propres principaux), pour obtenir des gains de compression significatifs.

On peut illustrer ce que nous venons d'énoncer précédemment par l'exemple suivant: Les dix valeurs propres de la matrice de covariance \hat{R}_{xx} (générée à partir des données en sortie du filtre) de x , sont :

$$\begin{array}{ccccc} 0.5468 & 0.1975 & 0.1243 \times 10^{-1} & 0.5112 \times 10^{-3} & 0.2617 \times 10^{-4} \\ 0.1017 \times 10^{-5} & 0.6437 \times 10^{-7} & 0.3895 \times 10^{-8} & 0.2069 \times 10^{-9} & 0.5761 \times 10^{-11} \end{array}$$

Cet exemple est pris pour $m=10$. L'erreur ϵ_j^* , évaluée à partir des valeurs précédentes (pour $j=2$), vaut 0.0130. Cette dernière est ensuite comparée à la valeur 0.7573 qui représente la somme de toutes les valeurs propres. L'erreur normalisée est $\frac{0.0130}{0.7573} = 0.0171$. Puisque cette

erreur peut être considérée comme suffisamment faible, seulement les deux premières composantes KL ($j=2$), sont considérées comme significatives. Dans ce cas, on a un gain de compression égal à $10/2=5$. L'expansion KL exige seulement $1/5$ de bits relatifs pour représenter directement le signal.

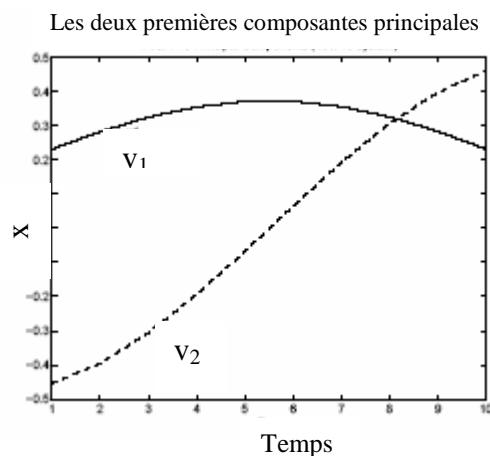


Figure 1.7 : Représentation temporelle des deux premiers vecteurs propres [16].

La figure 1.7 montre la variation du $k^{\text{ème}}$ élément V_k du vecteur propre en fonction de son indice k ($k=1,2,\dots,m$). Dans ce cas, on s'attend à ce que toute observation x_i , puisse être exprimée exactement, par une combinaison linéaire des deux premières formes d'ondes de vecteurs propres tel qu'il est montré dans la figure 1.7.

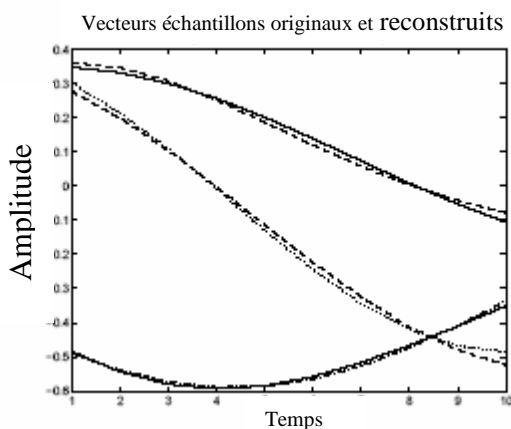


Figure 1.8: Représentation de trois signaux reconstruits avec les deux premières composantes principales (en pointillé), comparés à leurs vecteurs originaux (trait continu) [16].

Dans la figure 1.8, on montre la vraie observation x , dont les échantillons sont représentés par une forme d'onde temporelle, comparée à la reconstruction \hat{x} obtenue avec deux vecteurs propres seulement. Il est clair que, la différence entre les vrais échantillons et les reconstruits, est faible comme il a été attendu.

1.6. Conclusion

Les points positifs de la transformation KL sont les suivants :

- Est une transformation linéaire d'un espace de données à dimension élevée vers un espace de faible dimension.
- Dé corrélacion totale des nouveaux coefficients KL.
- Compression importante des données.

Le point négatif important est que la nouvelle base de KL dépend de la statistique de chaque signal. Cependant, des signaux de même source donnent des bases quasiment identiques.

Plus de détails sur la transformation KL (PCA : analyse en composantes principales) peuvent être trouvés dans l'appendice B .

CHAPITRE 2

LES FONCTIONS DE SIGNATURES

2.1. Introduction

Après avoir présenté, au chapitre précédent, la transformée de Karhunen-Loève, nous développons dans ce chapitre, la technique proposée pour la compression de la parole, basée sur cette transformation.

Les fonctions de signatures ont été créées, expérimentalement, sur une base ad-hoc. En considérant, le comportement quasi-stationnaire des signaux de la parole, une méthode statistique est utilisée pour générer "les fonctions de signatures". A cet égard, on exécute plusieurs milliers d'expériences pour analyser le signal de la parole.

Dans les expériences menées, chaque morceau du signal est décomposé en petites trames (Figure 2.1). Pour chaque trame, la matrice de corrélation est construite, et ses vecteurs et valeurs propres sont calculés. Le vecteur propre associé à la plus grande valeur propre, est sélectionné et ensuite mémorisé pour une évaluation antérieure. Une grande zone de stockage, comme entrepôt de données, est éventuellement obtenue comme résultats de ces expériences.

En employant un algorithme de comparaison efficace, les vecteurs propres de formes semblables ont été éliminés. En conclusion, pour la famille des signaux de la parole, on a fini avec un ensemble de données qui contient seulement 15 ou 16 différentes formes de vecteurs propres. Dans cette approche, chaque vecteur propre est considéré comme une séquence temporelle. Sa forme continue est appelée "Fonction de Signature : FS".

Dans le processus de modélisation des signaux, par les séquences ou fonctions de signature, chaque trame du signal de la parole est représentée avec seulement une fonction de signature multipliée par un coefficient. Chaque trame est alors désignée par :

- Un nombre d'indice qui est associé avec la fonction ou la séquence de signature.
- Un coefficient (multiplicateur).

D'ici, un taux de compression du signal réel est atteint. Dans la section suivante, la génération des fonctions de signature ou de séquences prédéfinies est illustrée. En se basant sur les résultats des expériences quelques fonctions de signatures sélectionnées, pour la modélisation des paroles, sont tracées.

Des exemples sont donnés pour montrer l'implémentation pratique de cette méthode. On s'attend à ce que le modèle proposé, pour modéliser les signaux de la parole, peut être utilisé dans :

- Le codage de la parole.
- L'enregistrement efficient avec un taux de compression élevé.
- La transmission.

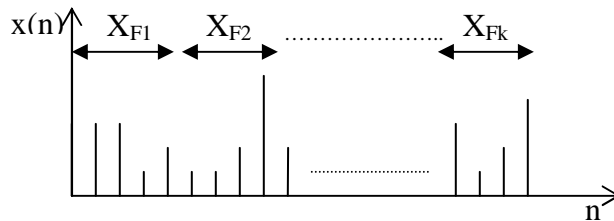


Figure 2.1 : Segmentation d'un signal discret

2.2. Une méthode statistique pour générer les FS:

2.2.1 Description

Dans cette méthode, un signal quasi-stationnaire donné sur une longue période, est divisé en "trames" comme le montre la figure 2.1

Soit N le nombre d'échantillons sur la période considérée. La décomposition du signal en série est donc:

$$x(n) = \sum_{i=1}^N x_i \delta(n - i) \quad (2.1)$$

Où :

- δ est l'impulsion unité.
- x_i est l'amplitude du i^{eme} échantillon.

Cette suite provient de l'échantillonnage d'un signal à temps continu. Le signal $x(n)$, est décomposé en trames, de longueurs égales de LF échantillons. Les séquences correspondantes peuvent être représentées par un vecteur X_{FK} tel que :

$$\mathbf{X}_{Fk} = \begin{bmatrix} \mathbf{X}_{(k-1).LF+1} \\ \mathbf{X}_{(k-1).LF+2} \\ \dots\dots\dots \\ \mathbf{X}_{k.LF} \end{bmatrix} \quad k = 1, 2, \dots\dots N_{FT} \quad (2.2a)$$

Dans (2.2a), $N_{FT}=N/LF$ désigne le nombre total de trames.

Bien qu'on puisse faire les calculs pour N et LF quelconques, il est conseillé, pour des raisons de rapidité et de simplification, de prendre N et LF égaux à une puissance de 2. Donc, N_{FT} est égal à une puissance de 2.

Il est clair que, $N \gg LF$ et $N \rightarrow \infty$, pour les signaux de parole continus.

En utilisant la notation vectorielle, le signal complet \mathbf{X}_F peut être considéré comme la collection des trames \mathbf{X}_{Fk} tel que:

$$\mathbf{X}_F = [\mathbf{X}_{F1} \mathbf{X}_{F2} \dots\dots\dots \mathbf{X}_{FN_{FT}}] \quad (2.2b)$$

Pour chaque trame, la matrice de corrélation peut être définie par :

$$r_k(l+1) = \frac{1}{LF-l} \sum_{j=(k-1)LF+1}^{k.LF-l-1} \mathbf{X}_{j+1} \mathbf{X}_{j+1+l} \quad l=0, 1, 2, \dots\dots\dots (LF-1) \quad (2.3)$$

$$\mathbf{R}_{Fk} = \begin{bmatrix} r_k(1) & r_k(2) & \dots\dots & r_k(LF) \\ r_k(2) & r_k(1) & \dots\dots & r_k(LF-1) \\ \vdots & & & \\ r_k(LF) & r_k(LF-1) & \dots\dots & r_k(1) \end{bmatrix} \quad (2.4)$$

Où \mathbf{R} est une matrice $L_F \times L_F$.

Cette matrice possède des propriétés remarquables :

La fonction de corrélation étant paire, on voit évidemment que la matrice \mathbf{R} est symétrique. Elle possède en plus la propriété d'être " Toeplitz ", ce qui signifie que ses éléments ij , ne dépendent que de la différence $i-j$. Ceci entraîne que tous les éléments sur une parallèle à la

première diagonale sont égaux. En particulier, les éléments diagonaux sont tous égaux à la variance du signal $x(n)$.

Il est clair qu'une matrice de Toeplitz est entièrement définie par sa première ligne ou sa première colonne. On admet qu'elle est définie positive, c'est-à-dire qu'il n'y a aucune valeur propre nulle.

Ses valeurs propres et vecteurs propres, λ_{ik} et V_{ik} , sont calculés à partir de l'équation classique suivante:

$$R_{Fk} \cdot V_{ik} = \lambda_{ik} V_{ik} \quad (2.5)$$

Notons que les vecteurs propres d'une matrice réelle symétrique sont orthogonaux et peuvent être normalisés à la norme unitaire $\|V_{ik} = 1\|$. En plus, les valeurs propres de cette matrice, sont positives et s'ordonnent habituellement dans l'ordre descendant. Dans ce cas, chaque trame peut être exprimée, en fonction des vecteurs propres de sa matrice de corrélation :

$$X_{Fk} = \sum_{i=1}^{LF} C_i V_{ik} \quad (2.6)$$

On s'intéresse au premier terme de cette série :

V_{1k} représente le vecteur propre ayant la plus grande énergie puisqu'il est associé à la plus grande valeur propre λ_{1k} .

Les coefficients C_i sont déterminés par la condition d'orthonormalité des vecteurs propres :

$$V_{ik}^T V_{jk} = \begin{cases} 1 & \text{pour } i = j \\ 0 & \text{pour } i \neq j \end{cases} \quad (2.7a)$$

Donc :

$$C_i = [X_{Fk}]^T \cdot V_{ik} \quad (2.7b)$$

Où :

- X_{Fk} désigne le vecteur de la $k^{\text{ème}}$ trame.
- V_{ik} désigne le $i^{\text{ème}}$ vecteur propre avec LF composantes.
- L'exposant T désigne la transposée d'un vecteur ou d'une matrice.

Jusque là, on désigne le premier vecteur propre comme « vecteur de signature S_k » de la trame X_{Fk} car il contient la plus grande quantité d'énergie par rapport aux autres vecteurs. Si on considère les composantes de V_{1k} comme une séquence temporelle ordonnée alors cette dernière est appelée « séquence de signature ».

Le plus intéressant, dans la méthode proposée est que pour tout signal de parole, la longueur LF peut être choisie de telle sorte que la trame considérée soit approximée par le premier terme de (2.6) :

$$X_{Fk} \cong C_{1k} V_{1k} \quad (2.8)$$

Dans ce cas, pour les signaux aléatoires de parole, il en vaut l'effort de générer les premiers vecteurs propres de (2.6) pour différentes longueurs LF, afin d'étudier la qualité de cette approximation.

Au cours des expériences, des milliers de trames de parole ont été examinées, avec des différentes longueurs LF, obtenues de différentes sources (hommes et femmes).

En considérant les premiers vecteurs propres V_{ik} comme des séquences temporelles, leurs formes ont été examinées. Il résulte de cette étude que beaucoup de similarités existent entre ces formes. Ceci a été attendu à cause du comportement quasi-stationnaire des signaux de parole. A cet égard, on a éliminé les formes qui se ressemblent.

Par exemple, pour LF=24, on a trouvé approximativement $N_S = 16$ différentes formes de V_{ik} , suffisantes à décrire toute trame de parole. Chaque forme distinguée est ensuite désignée par S_r pour former l'ensemble $S = \{S_r ; r=1, 2, \dots, N_S\}$.

Où S est appelé l'ensemble de la « séquence de signature prédéfinie »

Autrement dit, toute trame arbitraire ou aléatoire de parole peut être décrite par un vecteur « S_r » sélectionné de l'ensemble S et un coefficient constant C_{kr} .

Cette propriété facilite significativement la représentation des signaux de parole avec les termes des séquences ou fonctions de signature prédéfinies.

L'ensemble S est construit expérimentalement. Toute trame de parole X_{Fk} est approximée par un élément de S_r , convenablement sélectionné de cet ensemble, en utilisant l'erreur quadratique moyenne :

$$X_{Fk} \cong C_{kr} S_r \quad (2.9)$$

Dans ce qui suit, un algorithme qui permet la construction de l'ensemble S de la séquence de signature prédéfinie, sera présenté.

2.3. Algorithme de génération de la FS

Dans cette section, on consacre quelques pages à l'algorithme de calcul de la fonction de signature désigné par FS. La méthode proposée, est implémentée par deux algorithmes :

2.3.1. Algorithme 1 : Génération des fonctions de signatures

Cet algorithme, permet de trouver l'ensemble S , dont les éléments forment une base orthonormale de dimension réduite (15 ou 16). Les éléments de cette base, représentent les vecteurs propres principaux calculés à partir des matrices de corrélations des trames de paroles. Le paragraphe qui suit tente d'énumérer les principales étapes de cet algorithme.

Etape 1 :

Enregistrer plusieurs signaux de parole (obtenus de différentes sources), avec une fréquence d'échantillonnage régie par le théorème de Shannon. Il est clair que le signal ne doit pas être échantillonné à une fréquence inférieure au double de sa fréquence de battement (On peut reconstruire de façon exacte, pour tout t , un signal réel de bande B à l'aide de ses échantillons pourvu que la fréquence d'échantillonnage soit telle que $f_e \geq 2B$). La fréquence $2B$ porte le nom de fréquence de Nyquist). Pour les signaux de parole f_e ne doit pas être inférieure à 8 kHz.

Etape 2 :

Pour chaque morceau du signal, sélectionner la longueur convenable de trame, LF , à l'aide des essais. De telle sorte que la trame considérée soit approximée par un seul terme selon la relation (2.8). Dans cette étape, LF peut être déterminée par évaluation de la qualité de parole reconstruite. On peut utiliser des mesures de qualité subjective ou objective telle que :

- La moyenne des notes de plusieurs écouteurs (Mean Opinion Score : MOS). C'est une mesure subjective.

- L'erreur quadratique moyenne (LMSE : Least Mean Square Error). C'est une mesure objective.

Dans ce travail, MOS et LMSE ont été employées toutes les deux pour avoir un LF raisonnable.

En se basant sur les tests MOS, on a trouvé que :

- LF=24 donne une bonne qualité du signal reconstruit.
- LF=40 produit un son intelligible seulement.

Les détails des expériences seront donnés dans le chapitre 3.

Etape 3 :

Construire la matrice de corrélation de chaque trame. Ensuite, calculer les vecteurs propres et les valeurs propres de cette matrice selon les relations (2.3) et (2.5).

Etape 4:

Sélectionner les premiers vecteurs propres, qui correspondent aux plus grandes valeurs propres. Sauvegarder les vecteurs sélectionnés dans un ensemble appelé « ensemble de séquence de signature augmenté » désigné par S_{EG} . Il est évident de noter que S_{EG} peut contenir plusieurs centaines de milliers de vecteurs. C'est le résultat, des diverses prononciations de phrases, par un locuteur donné ou plus généralement par un très grand nombre de locuteurs.

Etape 5 :

Comparer tous les éléments de S_{EG} et éliminer les formes similaires. A cet égard, les normes des vecteurs doivent être comparées. On élimine un des deux vecteurs comparés, si la distance entre eux, est inférieure à un certain seuil défini au préalable.

Par exemple, en choisissant LF=24, on peut éliminer l'un des deux vecteurs V_{1k} ou V_{1j} si :

$$\|V_{1k} - V_{1j}\| < \varepsilon \quad (2.10)$$

Où ε est le seuil de comparaison. En choisissant convenablement la valeur de ε , on obtient 15 ou 16 différents vecteurs « séquences de signature », qui constituent l'ensemble S :

$$S = [S_1, S_2, \dots, S_{15}] \quad (2.11)$$

Où les vecteurs S_1, S_2, \dots, S_{15} , sont considérés comme des séquences temporelles. La forme continue de ces dernières, nommée fonction de signature $S_k(t)$, peut être exprimée par une technique de régression mathématique ou tout simplement par la sommation des fonctions d'échantillonnages :

$$S_k(t) = \sum_{l=1}^{LF} S_{lk} \cdot S_a[\omega_m(t - lT_p)] \quad (2.12)$$

Avec :

- $S_a[\omega_m(t - lT_p)] = \frac{\sin[\omega_m(t - lT_p)]}{\omega_m(t - lT_p)}$ représente la fonction d'échantillonnage.

- $\omega_m = 2\pi f_{max}$, désigne la pulsation (fréquence angulaire), elle est fonction de la fréquence maximale de la bande passante.

- T_p désigne la période d'échantillonnage.

- S_{lk} désigne la $l^{ème}$ composante de la séquence de signature :

$$S_k^T = [S_{1k} \ S_{2k} \ S_{lk} \ \dots \ S_{LFk}] \tag{2.13}$$

Si $T_p = (1/2) \cdot f_{max}$, donc :

$$S_k(t) = \sum_{l=1}^{LF} S_{lk} S_a(\omega_m t - l\pi) \tag{2.14}$$

Pour les signaux de parole, il est pratique de prendre $f_{max} = 4$ kHz.

Sur la figure 2.2, on peut observer les 15 différentes fonctions de signature, obtenues pour $LF=16$.

Une fois, l'ensemble de la séquence de signature défini, alors tout signal de parole peut être synthétisé. Le signal est reconstruit, trame par trame, avec les éléments de l'ensemble S et une constante réelle C_{kr} . Cette dernière est déterminée à l'aide de l'erreur quadratique moyenne.

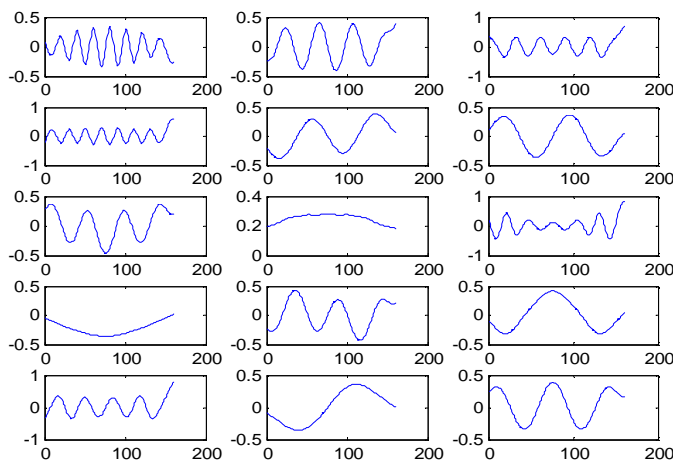


Figure 2.2 : Séquences de signature prédéfinies pour $LF=16$, $S=[S_1, S_2, \dots, S_{15}]$

2.3.2. Algorithme 2 : Reconstruction du signal de parole avec les fonctions de signatures

Le plus important de cette approche réside dans le fait qu'on peut modéliser tout signal de parole par les séquences de signature prédéfinies $S=S_1, S_2, \dots, S_{N_s}$. Avec $N_s=15$ (ou 16) formes différentes obtenues avec $LF=16$ (ou 24) échantillons. Pour chaque trame, on doit trouver le vecteur S_r optimal et le coefficient C_{kr} qui l'approxime mieux. Autrement dit, on sélectionne S_r et C_{kr} qui minimise l'erreur quadratique moyenne entre la trame originale et la trame synthétisée. La reconstruction se fait selon la procédure suivante :

Toute trame aléatoire de parole peut être exprimée par :

$$X_{Fk} \cong C_{kr} S_r \text{ tel que } C_{kr} = X_{Fk}^T \cdot S_r \quad (2.15)$$

Où

$$S_r = [S_{1r}, S_{2r}, \dots, S_{LFr}]^T \quad (2.16)$$

S_r est obtenu de l'ensemble des séquences de signature et correspond à une erreur quadratique moyenne minimale définie par :

$$\delta_{kr} = \|X_{Fk} - C_{kr} S_r\|^2 = \sum_{i=1}^{LF} (X_{iFk} - C_{kr} S_{ir})^2 \quad (2.17)$$

A cet égard, la valeur minimale $\delta_{\min} = \delta_{kr}$ est donnée par la distance

$$\delta_{kr} = \|X_{Fk} - C_{kl} S_l\|^2 \quad (2.18)$$

; entre la trame originale X_{Fk} et les séquences $C_{kl} \cdot S_l$.

Où

- $S = \{S_l ; l=1, 2, \dots, N_s\}$ est l'ensemble de signature

- $C_{kl} = X_{Fk} S_l$ désigne les coefficients des trames.

En prenant le minimum de tous les δ_{kl} , on sélectionne l'indice « r » de la séquence S_r et le coefficient de la trame C_{kr} , nécessaires à la synthèse des signaux de parole, sur la longueur de trame donnée.

Donc :

$$\delta_{kr} = \min\{\delta_{kl}, l=1, 2, \dots, N_s\} \quad (2.19)$$

; donne l'indice « r » et le coefficient $C_{kr} = X_{Fk}^T \cdot S_r$ de la trame en considération.

On en déduit alors que la trame d'un signal peut être décrite par deux quantités :

- l'indice « r » indiquant le nom de la séquence de signature prédéfinie.
- Le coefficient de la trame C_{kr} .

L'expérience montre que la procédure qui vient d'être décrite assure une synthèse convenable des sons.

2.4. Qualité de la parole et évaluation [18]

L'évaluation d'un algorithme de codage de la parole dépend des attributions suivantes :

- 1- Le taux de bits du signal compressé.
- 2- La qualité de la parole codée.
- 3- La complexité de l'algorithme.
- 4- Le retard introduit et
- 5- la robustesse de l'algorithme à enchaîner les erreurs et l'interférence acoustique.

De façon générale, le codage de la parole à haute qualité et à faible taux, est atteint grâce à des algorithmes de haute complexité. Sur un autre plan, le temps réel d'implémentation impose des restrictions sur la complexité et le retard du codeur. Dans les systèmes de la transmission des messages par exemple, le retard du codeur peut ne pas être une solution et le système central de la sauvegarde peut ne pas exiger un codeur à faible complexité. Alors que dans de nombreuses applications, le but principal est d'assurer la similarité entre signal reconstitué et l'original. Dans certains cas (par exemple les systèmes dans lesquels la sécurité est l'intérêt principal) il est suffisant que le parlé reconstitué soit intelligible et naturel.

La qualité de la parole peut être classée en plusieurs catégories (voir appendice C.1).

Evaluer la qualité de la parole est important mais c'est aussi une tâche très difficile. Il existe deux manières pour mesurer la qualité de la parole :

2.4.1. Mesures subjectives de la qualité de la parole

Habituellement, la procédure d'évaluation subjective est faite avec des tests d'écoutes formels. Les différentes mesures subjectives, largement utilisées, sont les suivantes:

- Diagnostic Rhyme Test (DRT) : La DRT est une mesure de l'intelligibilité de la parole.
- Diagnostic Acceptability Measure (DAM) : Le score DAM évalue la qualité d'un système de communication. Cette mesure est basée sur l'acceptabilité de la parole telle qu'elle est perçue par un écouteur (normal et entraîné).
- Mean Opinion Score (MOS) : Cette mesure est largement utilisée pour quantifier la qualité de la parole.

Il est noté que la cotation MOS peut varier significativement d'un test à l'autre et donc elle n'est pas une mesure absolue pour comparer entre différents codeurs.

2.4.2. Mesures objectives de la qualité de la parole

Le système auditif humain est l'évaluateur ultime de la qualité et de la performance d'un codeur de la parole, en percevant l'intelligibilité et le naturalisme du signal audio. Alors que, les tests d'écoute extensifs fournissent l'évaluation la plus exacte des codeurs de la parole, ils nécessitent beaucoup de temps et sont inconsistants. Les mesures objectives peuvent donner une estimation immédiate et sûre de la qualité perceptuelle d'un algorithme de codage. Les principales mesures de distorsion dans le domaine temporel sont les suivantes :

- Signal-to-Noise Ratio (SNR) : Le rapport signal sur bruit, est l'une des plus importantes mesures objectives utilisée pour évaluer la performance d'un algorithme de compression.
- Segmental SNR (SEGSNR) : Les variations temporelles de la performance peuvent être mieux détectées et évaluées en utilisant des SNR de courte durée (trame par trame). Le SEGSNR pénalise les codeurs dont la performance est variable.

Autres détails sur les mesures de qualité se trouvent dans l'appendice C.

2.5. Sélection de la longueur de trame LF par les tests de la qualité d'audition « MOS »

Ce paragraphe, consacré à la détermination du LF optimale, a deux objectifs :

- Montrer l'influence de LF sur la qualité du signal reconstitué.
- Associer cette influence à l'utilisation de l'algorithme FS.

La détermination du LF optimale nécessite les étapes suivantes :

- Etablir le corpus de données comme illustré dans l'appendice D
- Fixer la valeur de LF et décomposer le signal de parole selon cette valeur.
- Appliquer l'algorithme FS au signal décomposé.
- Utiliser les fonctions de signatures trouvées, dans l'étape précédente, pour synthétiser le signal de la parole.
- Evaluer la qualité du signal reconstitué, en utilisant le MOS et la LMSE comme mesure d'évaluation de qualité de la parole.

En pratique, le mode opératoire décrit ci-dessus, ne produit pas toujours le signal reconstitué désiré en une seule passe. Autrement dit, le signal synthétisé à l'issue d'une passe, peut être mauvais. Si c'est le cas, la longueur LF doit être réajustée de façon à obtenir le signal souhaité.

Parmi les valeurs de LF, à tester, on peut citer :

$$LF = 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48.$$

Durant le processus de génération de l'ensemble S, On peut utiliser un flot de parole qui inclut la majorité des phonèmes de la langue arabe (un vocabulaire limité : 500 mots). Ce flot de voix contient une variété de sons masculins et féminins avec 24.000 échantillons. Si on utilise LF=24 échantillons par trame, il en résulte $24.000/24 \approx 1000$ trames ou « premiers vecteurs propres ». Pour LF=40, on aura $24.000/40 = 600$ trames. Autrement dit, on commence avec 1000 ou 600 « premiers vecteurs » dans l'ensemble S_{EG} . On obtient alors, des séquences de signature de 15 ou 16 configurations différentes comme résultat de l'application de l'algorithme FS.

2.6. Capacité de compression du signal de la parole

Dans les sections précédentes nous avons développé les outils de base indispensables pour générer les « fonctions de signature : FS ». La FS a été présentée comme étant un outil de codage par transformée du signal de la parole. Il est temps à présent d'utiliser la FS dans la compression du signal de la parole. Autrement dit, il nous reste à examiner comment la FS réalise la compression. Enfin, l'opération de compression est illustrée comme suit :

- * Si chaque trame de parole X_{Fk} , de longueur LF=24 échantillons, est codée avec 8 bits par échantillon, qui est typique au standard PCM de communication de voix, donc le nombre total de bits nécessaire est :

$$N_b = 24 * 8 = 192 \text{ bits à transmettre ou à mémoriser pour chaque trame.}$$

En utilisant la technique proposée, on a besoin de 8 bits pour coder C_{kr} et 4 bits pour coder l'indice de trame « r ». Le nombre total de bits nécessaire pour décrire la trame considérée est égal à 12 (au lieu de 192). Cela veut dire qu'un taux de compression $CR=192/12=16$ est atteint.

*Si on choisit 40 échantillons pour chaque trame on obtient :

- $N_b = 40 \times 8 = 320$ bits à transmettre (standard PCM).
- $N_b = 4 + 8 = 12$ bits à transmettre (le modèle proposé).

Dans ce cas le taux de compression est : $CR = 320/12 = 26.66$

2.7. Conclusion :

Pour cette application il convient, de choisir d'une façon judicieuse les unités de base, pour la synthèse. Les paramètres FS associés à chaque trame sont mémorisés pour constituer une banque de données. Le choix des fonctions de base dont la sélection doit en principe permettre la synthèse d'un signal de la parole quelconque, doit répondre aux critères suivants :

- Le nombre de fonctions de base doit être limité.
- Leur extraction doit être facile et fiable et éventuellement pouvoir être automatisée.
- L'utilisation de ces fonctions ne doit pas faire intervenir de règles trop complexes.

La FS crée une représentation optimale des données d'entrée en ce qui concerne le compactage d'énergie et fournit une représentation permettant la manipulation du contenu de tous les signaux [5].

Quant à la partie réalisation des algorithmes, dans la méthode proposée, on a développé des modifications des versions traditionnelles. On peut dire que la FS, représente un nouvel algorithme de compression dont la complexité a été évaluée avant la mise en œuvre.

Les deux mesures utilisées pour évaluer la qualité de compression et de reconstruction de la parole sont: l'erreur quadratique moyenne LMSE et la cotation MOS.

Dans ce chapitre, on a décrit l'application de la FS au traitement du signal de la parole, en se basant sur la description de la technique et tout argument mathématique. Selon notre point de vue, la FS est très importante en tant qu'une approche de transformation dans le traitement du signal du son. Le deuxième point qui a été traité concerne l'extraction d'une base pratique pour le codage et la compression de la parole. Un algorithme de sélection de bases a été élaboré afin de trouver la meilleure base.

CHAPITRE 3

SIMULATIONS ET RESULTATS

3.1. Introduction

Dans ce chapitre, on décrit les spécificités de notre travail. Afin de mettre en expérience les concepts décrits et de décrire par la suite la nature et les résultats de notre expérimentation, en référence avec les chapitres précédents, nous avons trois points dans notre étude. Premièrement, nous avons étudié l'importance de la KLT lorsqu'elle est appliquée dans l'analyse, la transformation et la synthèse des sons pour des fins de compression. Le deuxième point concerne l'extraction d'une base optimale en utilisant une approximation de la KLT. Un algorithme de sélection d'une base optimale a été testé afin de nous aider à trouver une meilleure séquence de signature. On cherche une fonction de signature FS de forme indépendante du signal transformé.

Troisièmement, en plus de la représentation et de la transformation, notre premier objectif est la compression du signal audio. Plusieurs transformations et approches de sélection de bases qui existent sont originellement motivées pour des fins de compression. Alors, l'approche FS doit être testée dans ce sens (capacité de réduction de données).

La performance de chaque test a été mesurée en utilisant deux critères d'évaluation de la qualité du signal reconstruit telles que :

- L'erreur quadratique moyenne LMSE (Least Mean Square Error).
- La notation MOS (Mean Opinion Score).

Dans la dernière section, on compare la qualité de l'algorithme FS à celle de « FS1015 LPC10E ».

3.2. Les sons ciblés

Avant de décrire la modification et la sélection des fonctions de signatures, on décrit le type des sons qui nous intéresse et pourquoi la FS semble efficace pour ces sons. On étudie des sons sur lesquels se base cette approche tels que les sons environnementaux. C'est-à-dire les sons qui ont lieu dans l'environnement dans lequel on se retrouve, soit de façon naturelle

ou bien artificiellement produits (sons non musicaux). Pour ces sons la précision temporelle doit être maintenue dans n'importe laquelle des applications de transformation si nous devons percevoir (distinguer) le caractère de base du son. Dans ce cas, l'utilisation d'une transformation à résolution temporelle fine, va aider à percevoir ce type de sons avec précision. Les sons sur lesquels on travaille sont courts. On se base principalement sur ce genre de sons à qualité courte parce que les sons sont généralement complexes et changent tout le temps (et ne gardent jamais la même forme).

Les sons sources utilisés dans nos tests sont énumérés dans le tableau D.1 (appendice D). Ces sons seront modifiés de différentes manières et les fichiers résultats prennent les noms des sons sources avec un préfixe. Chaque entrée du tableau D.1 inclut le nom du son, une brève description, la fréquence d'échantillonnage et sa longueur. Tous les sons sont monovoies et obtenus par lecture des passages écrits. Pourtant, il est intéressant de voir comment cette technique s'applique bien aux autres classes de sons.

Les sons traités ont été enregistrés par nous-mêmes ou pris des CD-ROM data book. Dans le dernier cas, les sons ont été re-échantillonnés à 8 KHz et pris sur de courtes périodes. Puisque tous les fichiers contiennent des sons non compressés, ils sont enregistrés dans le format wave sous l'extension « .wav ». Ils ont été aussi normalisés en amplitudes $-1 \leq x(n) \leq 1$.

3.3. Compression par la KLT classique (Algorithme KLT1)

On utilise comme signal de parole la phrase en arabe « addarso thani » prononcée par une femme. La compression est réalisée au moyen de la KLT, appliqué sur les vecteurs obtenus en partitionnant le signal en blocs (LF=16, 24,40) et regroupés dans une matrice qui subira la décomposition en vecteurs propres et valeurs propres.

La figure 3.1 illustre la variation de l'erreur quadratique moyenne (normalisée par rapport au nombre d'échantillons) en fonction du nombre de coefficients retenus et de la longueur de la trame LF. On peut constater, la redondance dans le signal original, car toute l'information utile est comprise dans les quelques premiers coefficients pour chaque bloc.

Sur le premier graphe (LF=16 échantillons par trame), on peut observer que l'erreur quadratique moyenne devient négligeable lorsque le nombre de coefficients retenus est ≥ 3 . Donc, le signal peut être reconstruit, avec erreur minimale, en retenant 3 coefficients pour

chaque trame du signal de la parole. Dans ce cas, le taux de compression atteint est de $16/3 = 5.33$.

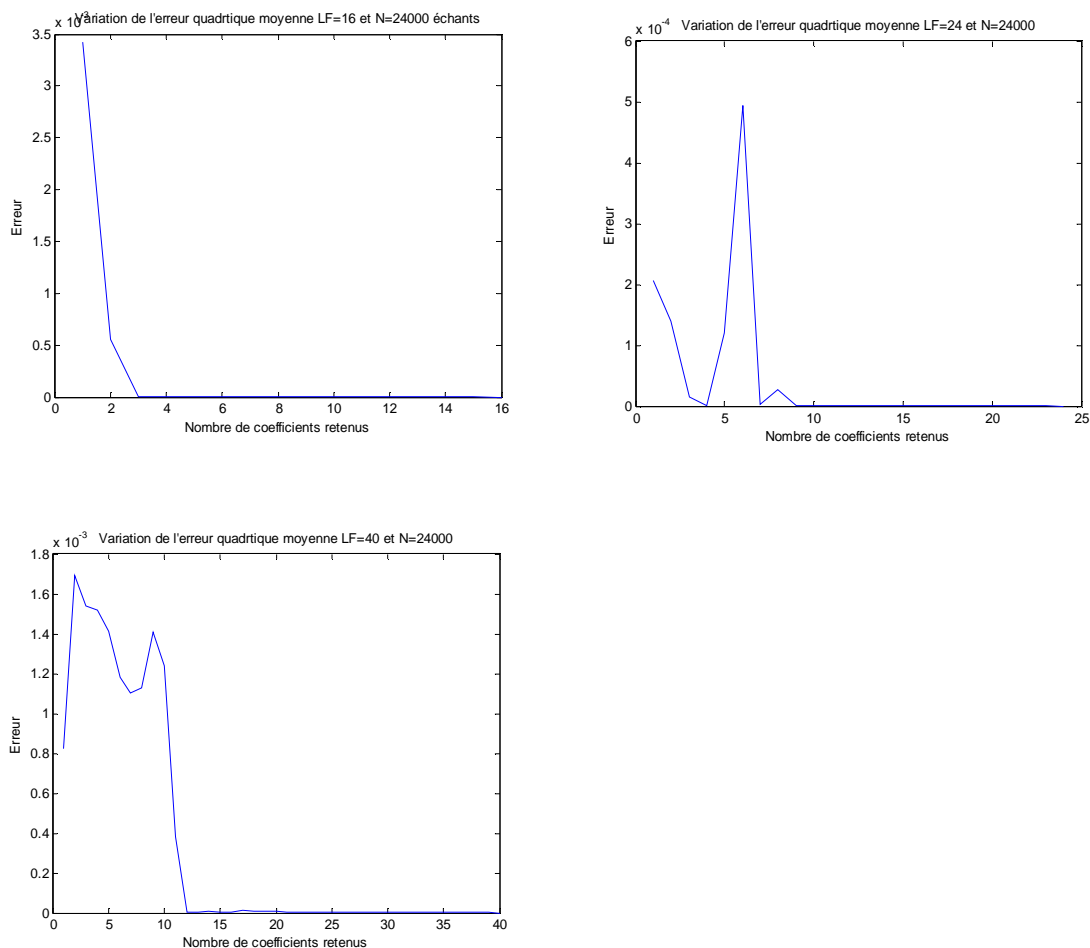


Figure 3.1 : Variation de l'erreur quadratique moyenne (eqm) en fonction du nombre de coefficients retenus.

De la même façon, pour le deuxième graphe (LF=24 échantillons par trame), on doit retenir 9 coefficients afin de reconstruire le signal avec une erreur minimale. Le taux de compression sera $24/9 = 3.66$.

Pour LF= 40 échantillons par trame, ce qui correspond au troisième graphe, on a besoin de 12 coefficients par trame pour pouvoir synthétiser le signal avec la distorsion minimale. Ainsi, on obtient un taux de compression de $40/12 = 3.33$.

On remarque qu'on a un compromis entre la compression et la distorsion du signal.

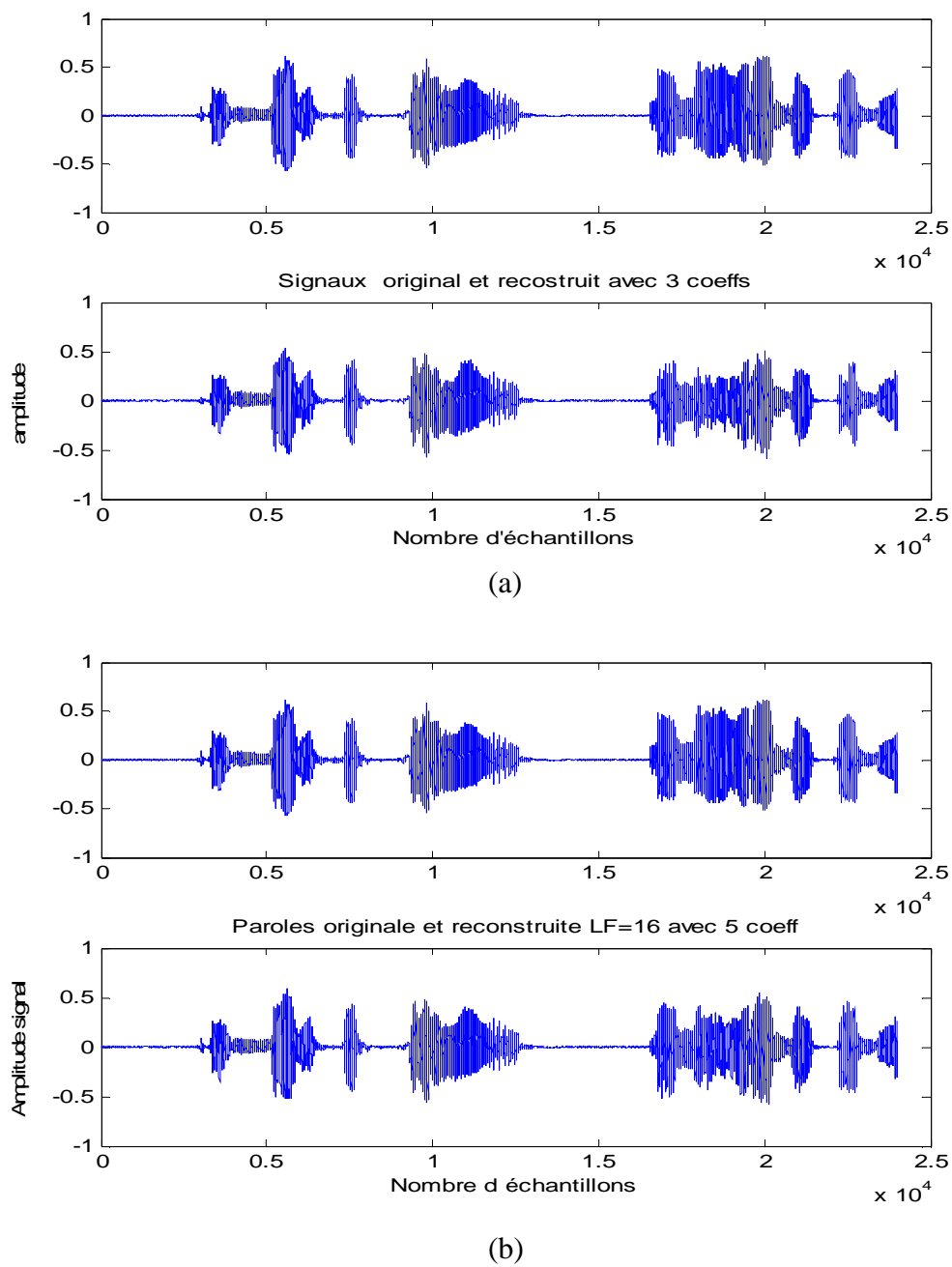


Figure 3.2: Signal original (en haut) et signal reconstruit (en bas) de la phrase :

« الدرس الثاني العرب قبل » a) en utilisant 3 coefficients par trame.

b) en utilisant 5 coefficients par trame.

La figure 3.2.a représente le signal de la parole originale « addarso thani el arabo kabla » et le reconstruit par trois coefficients, pour LF=16 échantillons par trame. On observe une légère différence entre ces deux signaux. En évaluant la qualité de la parole reconstruite par l'écoute, on peut dire que la dégradation est perceptible mais non gênante.

En comparant les figures 3.2.a et 3.2.b, qui correspondent à LF=16, on constate qu'il n'y a pas de grandes différences entre ces deux figures. On peut conclure que, l'augmentation du nombre de coefficients en dessus de 3 n'a pas beaucoup d'influence sur la qualité de reconstruction. Ce résultat est en accord avec le premier graphe de la figure 3.1, sur lequel la LMSE s'annule à partir du 3^{ème} coefficient.

La figure 3.2.b montre la légère amélioration du signal synthétisé en fonction de l'augmentation du nombre de coefficients utilisés.

Plusieurs expériences ont été menées, pour évaluer la qualité du signal reconstruit avec les différentes mesures de la qualité de la parole (mesures subjectives et objectives). Les résultats de ces expériences sont résumés dans le tableau 3.1

| Nombre de coefficients retenus | LF=16 | | LF=24 | | LF=40 | |
|--------------------------------|--------|-----------------|--------|-----------------|--------|---------------------|
| | LMSE | MOS | LMSE | MOS | LMSE | MOS |
| 3 | 0.0644 | Il y a du bruit | 0.1461 | Il y a du bruit | 0.3101 | Mauvais signal |
| 4 | 0.0283 | Bonne qualité | 0.0937 | Il y a du bruit | 0.2445 | Qualité moins bonne |
| 5 | 0.0149 | Bonne qualité | 0.0633 | Bonne qualité | 0.1988 | Parole intelligible |

Tableau 3.1 : LMSE et MOS pour différentes longueurs de trame LF.

Le tableau ci-dessus montre les résultats d'évaluation de la qualité du signal reconstruit en fonction de la longueur de la trame, à l'aide des deux critères les plus utilisés (LMSE et MOS). On constate que les résultats semblent moins bons lorsque LF s'éloigne d'une certaine valeur critique.

On constate qu'en augmentant la longueur de la trame, on améliore le taux de compression sans pour autant en améliorer la qualité du signal synthétisé. Malheureusement, on n'améliore jamais un paramètre sans en dégrader d'autres. L'utilisation de cet algorithme

pose des problèmes pratiques. On remarque dans ce sens le temps de calcul important. C'est la raison pour laquelle on cherche à trouver d'autres algorithmes plus performants.

3.4 .Algorithme FS (pour une seule phrase)

Comme il a été dit précédemment, l'algorithme de la KLT classique (l'algorithme précédent) demande un temps de calcul important. C'est la raison pour laquelle, la KLT est utilisée uniquement dans les laboratoires, à des fins de comparaison. Pour résoudre ce problème on utilise une approximation de cette transformée. Cette tâche est accomplie par l'algorithme FS (la base orthonormée des fonctions de signatures).

Dans ce cas, une seule phrase de paroles arabe, est traitée, puis reconstruite trame par trame séparément. Pour chaque trame on conserve uniquement les plus grandes composantes des vecteurs propres. Au cours des expériences d'analyse et de synthèse des différentes trames du signal de la parole, on a remarqué que :

- La majorité des trames reconstruites sont très proches des originales. C'est-à-dire l'erreur de distorsion est négligeable. Cela est vérifié pour les trames non voisées.
- Les trames voisées sont accompagnées d'erreurs importantes. Donc, elles sont erronées.

Les résultats de ces analyses sont donnés dans les tableaux et les figures qui suivent :

- 1-Le tableau 3.2.a donne l'erreur quadratique moyenne de reconstruction des différentes trames d'un segment de parole arabe, constitué de 8000 échantillons, segmentés en trames de 16 échantillons. Donc, le nombre de trames traitées est 500.
- 2- Les trames sont représentées en groupes de 50 trames par colonne. Les erreurs des trames voisées sont données en rouge.
- 3- La figure 3.3.a localise le voisement dans le signal de parole. On observe sur ce graphe quatre zones de voisement ce qui correspond à quatre plages d'erreurs importantes sur le tableau 3.2.

| | | "الدرس الثاني" | | | | | | | | | | |
|-----------------------------------|----|----------------|----------------------------|--------|--------|--------|--------|--------|--------|--------|--------|-----|
| | | N° de trame | Trames du signal de parole | | | | | | | | | |
| L'erreur quadratique moyenne LMSE | 1 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0040 | 0.0158 | 0.2989 | 0.0026 | 0.0009 | 451 |
| | 2 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0035 | 0.0029 | 0.4353 | 0.0137 | 0.0013 | 452 |
| | 3 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0008 | 0.0194 | 0.0018 | 0.3596 | 0.0022 | 0.0013 | 453 |
| | 4 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0020 | 0.0100 | 0.0091 | 0.2565 | 0.0045 | 0.0019 | 454 |
| | 5 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0005 | 0.0010 | 0.0066 | 0.2460 | 0.0013 | 0.0008 | 455 |
| | 6 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0013 | 0.0061 | 0.0012 | 0.3249 | 0.0041 | 0.0009 | 456 |
| | 7 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0031 | 0.0243 | 0.0037 | 0.2502 | 0.0033 | 0.0022 | 457 |
| | 8 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0097 | 0.0006 | 0.0164 | 0.1781 | 0.0031 | 0.0366 | 458 |
| | 9 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0346 | 0.0065 | 0.0004 | 0.3021 | 0.0043 | 0.0103 | 459 |
| | 10 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0343 | 0.0028 | 0.0034 | 0.2408 | 0.0075 | 0.0194 | 460 |
| | 11 | 0.0001 | 0.0000 | 0.0000 | 0.0002 | 0.0479 | 0.0004 | 0.0091 | 0.2966 | 0.0005 | 0.0212 | 461 |
| | 12 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.1235 | 0.0005 | 0.0004 | 0.1793 | 0.0034 | 0.0531 | 462 |
| | 13 | 0.0000 | 0.0001 | 0.0000 | 0.0001 | 0.0307 | 0.0038 | 0.0094 | 0.1651 | 0.0015 | 0.0539 | 463 |
| | 14 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.1096 | 0.0031 | 0.0048 | 0.0543 | 0.0043 | 0.0824 | 464 |
| | 15 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0668 | 0.0036 | 0.0008 | 0.1308 | 0.0022 | 0.0607 | 465 |
| | 16 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0265 | 0.0251 | 0.0040 | 0.0935 | 0.0013 | 0.1129 | 466 |
| | 17 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0167 | 0.0043 | 0.0008 | 0.1593 | 0.0024 | 0.1425 | 467 |
| | 18 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0927 | 0.0035 | 0.0017 | 0.0808 | 0.0009 | 0.0882 | 468 |
| | 19 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0385 | 0.0140 | 0.0004 | 0.0097 | 0.0020 | 0.0776 | 469 |
| | 20 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0171 | 0.0091 | 0.0102 | 0.0506 | 0.0015 | 0.0980 | 470 |
| | 21 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0477 | 0.0019 | 0.0041 | 0.0264 | 0.0013 | 0.0771 | 471 |
| | 22 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0208 | 0.0054 | 0.0054 | 0.0038 | 0.0008 | 0.0903 | 472 |
| | 23 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0456 | 0.0170 | 0.0055 | 0.0041 | 0.0006 | 0.0761 | 473 |
| | 24 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0199 | 0.0006 | 0.0180 | 0.0024 | 0.0004 | 0.0690 | 474 |
| | 25 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0369 | 0.0089 | 0.0392 | 0.0028 | 0.0005 | 0.0638 | 475 |
| | 26 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0875 | 0.0042 | 0.0461 | 0.0092 | 0.0004 | 0.0451 | 476 |
| | 27 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0375 | 0.0003 | 0.0779 | 0.0107 | 0.0007 | 0.0662 | 477 |
| | 28 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0047 | 0.0004 | 0.1926 | 0.0365 | 0.0006 | 0.0288 | 478 |
| | 29 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0485 | 0.0016 | 0.0690 | 0.0204 | 0.0005 | 0.1192 | 479 |
| | 30 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0268 | 0.0023 | 0.2394 | 0.0141 | 0.0004 | 0.0392 | 480 |
| | 31 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0061 | 0.0029 | 0.0703 | 0.0246 | 0.0002 | 0.0340 | 481 |
| | 32 | 0.0000 | 0.0000 | 0.0000 | 0.0016 | 0.0269 | 0.0167 | 0.1636 | 0.0566 | 0.0003 | 0.0317 | 482 |
| | 33 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0342 | 0.0030 | 0.1812 | 0.0514 | 0.0002 | 0.0117 | 483 |
| | 34 | 0.0000 | 0.0000 | 0.0000 | 0.0006 | 0.0062 | 0.0046 | 0.1004 | 0.0574 | 0.0004 | 0.0266 | 484 |
| | 35 | 0.0000 | 0.0000 | 0.0000 | 0.0005 | 0.0049 | 0.0089 | 0.2183 | 0.0178 | 0.0003 | 0.0039 | 485 |
| | 36 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0010 | 0.0067 | 0.1976 | 0.0067 | 0.0006 | 0.0020 | 486 |
| | 37 | 0.0000 | 0.0000 | 0.0000 | 0.0013 | 0.0486 | 0.0023 | 0.2251 | 0.0214 | 0.0011 | 0.0032 | 487 |
| | 38 | 0.0000 | 0.0000 | 0.0000 | 0.0011 | 0.0015 | 0.0044 | 0.1904 | 0.0100 | 0.0005 | 0.0013 | 488 |
| | 39 | 0.0000 | 0.0000 | 0.0000 | 0.0019 | 0.0048 | 0.0167 | 0.3141 | 0.0136 | 0.0022 | 0.0077 | 489 |
| | 40 | 0.0000 | 0.0000 | 0.0000 | 0.0358 | 0.0114 | 0.0004 | 0.2513 | 0.0149 | 0.0025 | 0.0011 | 490 |
| | 41 | 0.0000 | 0.0000 | 0.0000 | 0.0143 | 0.0034 | 0.0097 | 0.1312 | 0.0155 | 0.0033 | 0.0100 | 491 |
| | 42 | 0.0002 | 0.0000 | 0.0000 | 0.0083 | 0.0003 | 0.0047 | 0.2420 | 0.0137 | 0.0023 | 0.0008 | 492 |
| | 43 | 0.0000 | 0.0000 | 0.0000 | 0.0051 | 0.0120 | 0.0004 | 0.0460 | 0.0189 | 0.0040 | 0.0049 | 493 |
| | 44 | 0.0000 | 0.0000 | 0.0000 | 0.0030 | 0.0260 | 0.0008 | 0.2259 | 0.1146 | 0.0015 | 0.0022 | 494 |
| | 45 | 0.0000 | 0.0000 | 0.0000 | 0.0015 | 0.0008 | 0.0008 | 0.3896 | 0.0103 | 0.0036 | 0.0005 | 495 |
| | 46 | 0.0000 | 0.0000 | 0.0000 | 0.0007 | 0.0013 | 0.0019 | 0.0914 | 0.0689 | 0.0013 | 0.0013 | 496 |
| | 47 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0010 | 0.3742 | 0.0099 | 0.0016 | 0.0003 | 497 |
| | 48 | 0.0000 | 0.0000 | 0.0000 | 0.0002 | 0.0018 | 0.0078 | 0.2238 | 0.0553 | 0.0007 | 0.0004 | 498 |
| | 49 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0056 | 0.0028 | 0.2537 | 0.0024 | 0.0013 | 0.0008 | 499 |
| | 50 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 0.0185 | 0.0019 | 0.3517 | 0.0036 | 0.0021 | 0.0015 | 500 |

Tableau 3.2.a : LMSE des trames de parole (en retenant un seul coefficient par trame)

On remarque que, les trames des 3000 premiers échantillons (les 200 premières trames) ont une erreur quadratique moyenne nulle (cette plage du signal correspond à un silence).

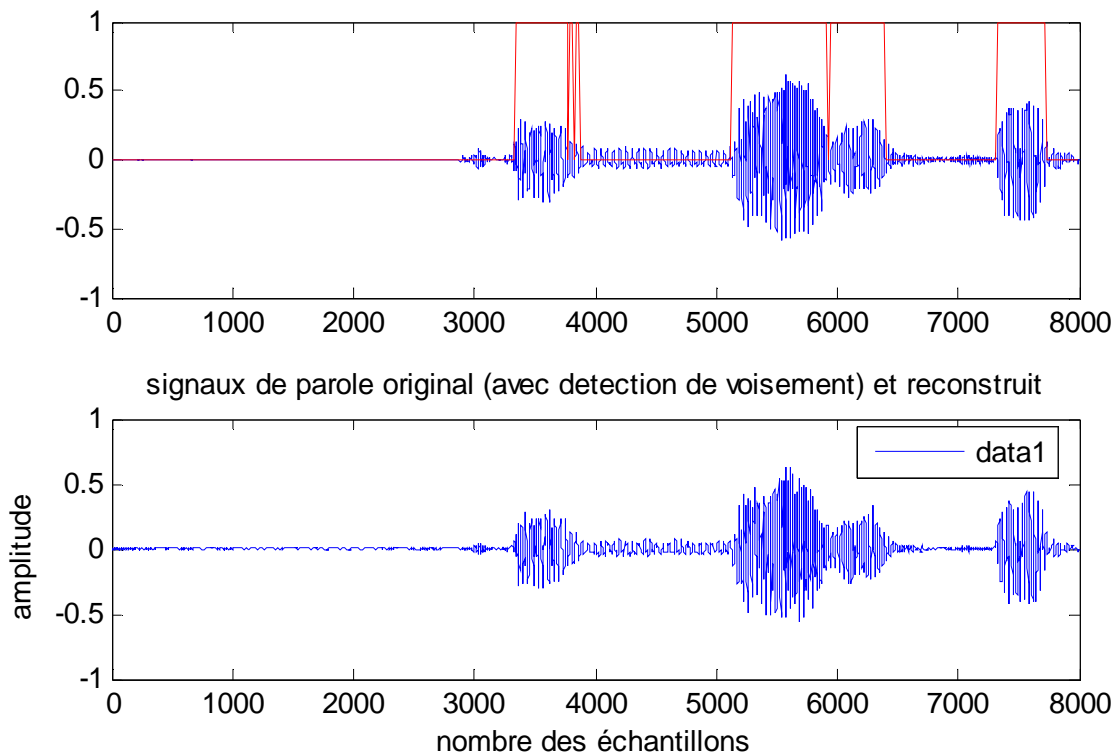


Figure 3.3.a : Détection de voisement (signal en rouge) d'un signal de parole arabe.

On vérifie ces résultats pour un signal de parole en français. La figure 3.3.b permet d'observer les zones de voisement de ce signal, ainsi que sa reconstruction avec un certain nombre de coefficients retenus pour les trames voisées.

Le signal traité est « la brise et le soleil ». Le tableau 3.2.b donne l'erreur quadratique moyenne des trames reconstruites avec un seul coefficient. On a remarqué que les erreurs sont plus grandes dans les parties voisées (en rouge). On a proposé d'utiliser un système hybride SR1/SR2 dans lequel on utilise SR1 dans les parties où l'erreur est inférieure à un certain seuil et SR2 dans les parties voisées (déterminées à partir de l'erreur qui est supérieure au seuil utilisé. Le seuil utilisé ici est de 0.05). Donc, on a essayé de minimiser l'erreur, en approximant les trames erronées avec plus de coefficients.

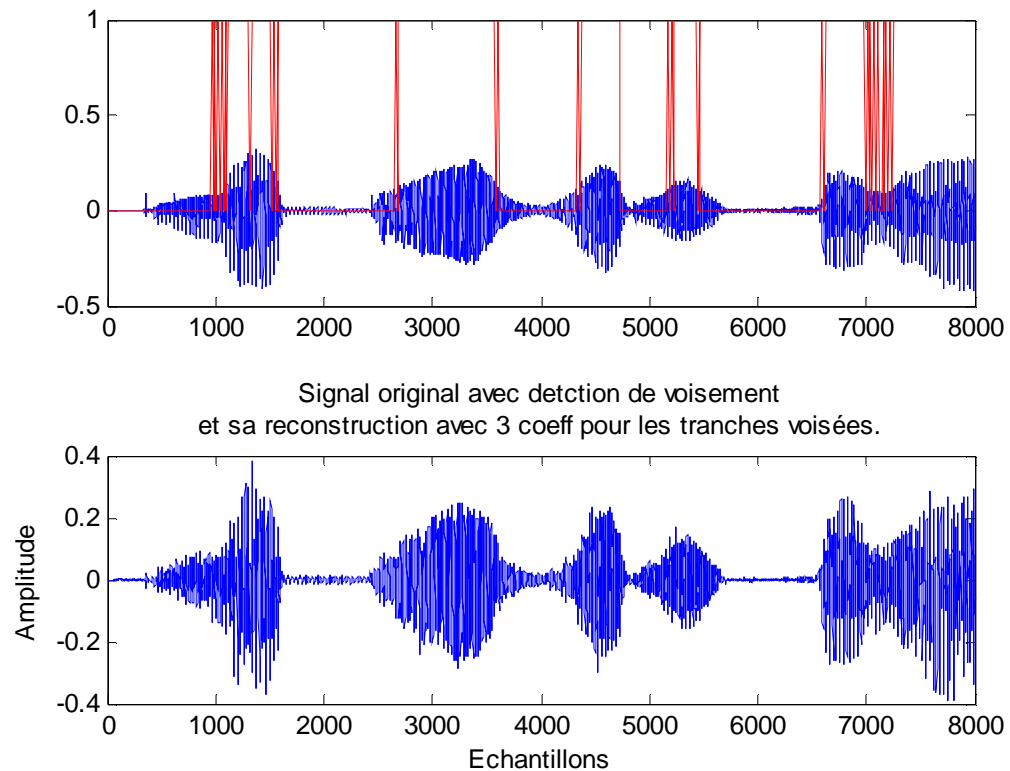


Figure 3.3.b : Détection de voisement (signal en rouge) d'un signal de parole français (en bleu).

Voici les résultats obtenus :

- Le tableau 3.2.c donne l'erreur quadratique moyenne des trames reconstruites avec deux coefficients pour les trames voisées. Sur ce tableau, on constate la réduction de l'erreur pour les zones voisées.
- En comparant les tableaux 3.2.b et 3.2.c on trouve que le nombre de trames ayant une erreur supérieure à 0.05 est 136 pour le premier alors qu'il est de 46 seulement pour le deuxième.

| | N° de trame | Trames du signal de parole | | | | “ la brise et le soleil” | | | | | | |
|-----------------------------------|-------------|----------------------------|--------|--------|--------|--------------------------|--------|--------|--------|--------|--------|-----|
| L'erreur quadratique moyenne LMSE | 1 | 0 | 0.0066 | 0.0095 | 0.0000 | 0.1208 | 0.0013 | 0.0004 | 0.0050 | 0.0011 | 0.0050 | 451 |
| | 2 | 0 | 0.0286 | 0.0003 | 0.0003 | 0.1333 | 0.0009 | 0.0014 | 0.0017 | 0.0004 | 0.0156 | 452 |
| | 3 | 0.0000 | 0.0040 | 0.0017 | 0.0097 | 0.1410 | 0.0005 | 0.0003 | 0.0036 | 0.0007 | 0.0122 | 453 |
| | 4 | 0 | 0.0095 | 0.0020 | 0.0037 | 0.0278 | 0.0036 | 0.0046 | 0.0011 | 0.0010 | 0.0472 | 454 |
| | 5 | 0.0000 | 0.0187 | 0.0001 | 0.0007 | 0.1159 | 0.0019 | 0.0063 | 0.0012 | 0.0017 | 0.0510 | 455 |
| | 6 | 0.0000 | 0.0317 | 0.0009 | 0.0117 | 0.1247 | 0.0006 | 0.0019 | 0.0011 | 0.0007 | 0.0927 | 456 |
| | 7 | 0.0000 | 0.0130 | 0.0006 | 0.0091 | 0.1553 | 0.0029 | 0.0062 | 0.0004 | 0.0009 | 0.0362 | 457 |
| | 8 | 0.0000 | 0.0101 | 0.0014 | 0.0404 | 0.1390 | 0.0038 | 0.0043 | 0.0003 | 0.0006 | 0.0589 | 458 |
| | 9 | 0.0000 | 0.0268 | 0.0004 | 0.0303 | 0.0354 | 0.0021 | 0.0137 | 0.0002 | 0.0010 | 0.0297 | 459 |
| | 10 | 0.0000 | 0.0112 | 0.0007 | 0.0244 | 0.0999 | 0.0029 | 0.0044 | 0.0005 | 0.0006 | 0.0469 | 460 |
| | 11 | 0.0000 | 0.0124 | 0.0009 | 0.0295 | 0.0610 | 0.0063 | 0.0188 | 0.0004 | 0.0062 | 0.0188 | 461 |
| | 12 | 0.0000 | 0.0082 | 0.0002 | 0.0291 | 0.2059 | 0.0077 | 0.0050 | 0.0003 | 0.0057 | 0.0944 | 462 |
| | 13 | 0.0000 | 0.0600 | 0.0002 | 0.0341 | 0.1646 | 0.0069 | 0.0036 | 0.0002 | 0.0456 | 0.0183 | 463 |
| | 14 | 0.0000 | 0.0100 | 0.0005 | 0.0157 | 0.1290 | 0.0042 | 0.0114 | 0.0001 | 0.0281 | 0.1596 | 464 |
| | 15 | 0.0000 | 0.0206 | 0.0004 | 0.0304 | 0.1718 | 0.0019 | 0.0121 | 0.0003 | 0.1838 | 0.0315 | 465 |
| | 16 | 0.0000 | 0.0040 | 0.0002 | 0.0180 | 0.0417 | 0.0126 | 0.0049 | 0.0001 | 0.0597 | 0.0894 | 466 |
| | 17 | 0.0000 | 0.0525 | 0.0007 | 0.0604 | 0.0938 | 0.0255 | 0.0044 | 0.0003 | 0.1538 | 0.0758 | 467 |
| | 18 | 0.0000 | 0.0102 | 0.0012 | 0.0499 | 0.0124 | 0.0272 | 0.0251 | 0.0004 | 0.0243 | 0.0935 | 468 |
| | 19 | 0.0000 | 0.0924 | 0.0002 | 0.0177 | 0.0829 | 0.0175 | 0.0178 | 0.0002 | 0.0866 | 0.1062 | 469 |
| | 20 | 0.0000 | 0.0313 | 0.0003 | 0.0398 | 0.0256 | 0.0241 | 0.0061 | 0.0001 | 0.0103 | 0.1229 | 470 |
| | 21 | 0.0000 | 0.1113 | 0.0004 | 0.0629 | 0.0968 | 0.0316 | 0.0034 | 0.0004 | 0.0532 | 0.0838 | 471 |
| | 22 | 0.0004 | 0.0529 | 0.0003 | 0.0418 | 0.0224 | 0.0704 | 0.0074 | 0.0002 | 0.0103 | 0.1043 | 472 |
| | 23 | 0.0109 | 0.1168 | 0.0003 | 0.0357 | 0.0719 | 0.0297 | 0.0237 | 0.0001 | 0.0449 | 0.1388 | 473 |
| | 24 | 0.0005 | 0.0597 | 0.0004 | 0.0831 | 0.0214 | 0.0479 | 0.0384 | 0.0001 | 0.0172 | 0.0747 | 474 |
| | 25 | 0.0001 | 0.1368 | 0.0012 | 0.0420 | 0.0430 | 0.0419 | 0.0219 | 0.0001 | 0.0607 | 0.1959 | 475 |
| | 26 | 0.0002 | 0.2173 | 0.0003 | 0.0438 | 0.0146 | 0.0898 | 0.0668 | 0.0002 | 0.0273 | 0.0576 | 476 |
| | 27 | 0.0007 | 0.0328 | 0.0004 | 0.0527 | 0.0351 | 0.0182 | 0.0030 | 0.0002 | 0.0721 | 0.2321 | 477 |
| | 28 | 0.0029 | 0.2158 | 0.0003 | 0.0336 | 0.0137 | 0.1135 | 0.0405 | 0.0001 | 0.0284 | 0.0576 | 478 |
| | 29 | 0.0019 | 0.0489 | 0.0002 | 0.0803 | 0.0257 | 0.0660 | 0.0060 | 0.0001 | 0.0593 | 0.2435 | 479 |
| | 30 | 0.0048 | 0.1720 | 0.0003 | 0.0428 | 0.0131 | 0.0254 | 0.0239 | 0.0000 | 0.0158 | 0.0576 | 470 |
| | 31 | 0.0015 | 0.0799 | 0.0006 | 0.0892 | 0.0151 | 0.0185 | 0.0241 | 0.0002 | 0.0573 | 0.2656 | 481 |
| | 32 | 0.0035 | 0.3435 | 0.0011 | 0.0453 | 0.0171 | 0.0651 | 0.0308 | 0.0001 | 0.0050 | 0.0585 | 482 |
| | 33 | 0.0024 | 0.0549 | 0.0002 | 0.0902 | 0.0051 | 0.0976 | 0.0489 | 0.0001 | 0.0933 | 0.2845 | 483 |
| | 34 | 0.0028 | 0.3936 | 0.0002 | 0.0422 | 0.0059 | 0.1988 | 0.0496 | 0.0003 | 0.0078 | 0.0818 | 484 |
| | 35 | 0.0082 | 0.0404 | 0.0003 | 0.0906 | 0.0009 | 0.1277 | 0.0658 | 0.0002 | 0.1727 | 0.1922 | 485 |
| | 36 | 0.0025 | 0.2677 | 0.0004 | 0.0616 | 0.0067 | 0.1714 | 0.0605 | 0.0003 | 0.0558 | 0.1640 | 486 |
| | 37 | 0.0072 | 0.0630 | 0.0002 | 0.0541 | 0.0069 | 0.0736 | 0.0466 | 0.0002 | 0.0416 | 0.1235 | 487 |
| | 38 | 0.0023 | 0.2828 | 0.0007 | 0.1030 | 0.0018 | 0.1050 | 0.0709 | 0.0002 | 0.0087 | 0.1996 | 488 |
| | 39 | 0.0307 | 0.1377 | 0.0006 | 0.0439 | 0.0020 | 0.0502 | 0.0290 | 0.0003 | 0.0694 | 0.1788 | 489 |
| | 40 | 0.0089 | 0.2732 | 0.0000 | 0.0696 | 0.0037 | 0.0818 | 0.0620 | 0.0004 | 0.0125 | 0.1760 | 490 |
| | 41 | 0.0034 | 0.0838 | 0.0001 | 0.0842 | 0.0008 | 0.0450 | 0.0172 | 0.0005 | 0.0090 | 0.1180 | 491 |
| | 42 | 0.0097 | 0.2259 | 0.0002 | 0.0463 | 0.0019 | 0.0959 | 0.0364 | 0.0003 | 0.0296 | 0.1982 | 492 |
| | 43 | 0.0248 | 0.1460 | 0.0006 | 0.0682 | 0.0024 | 0.0528 | 0.0067 | 0.0003 | 0.0611 | 0.0646 | 493 |
| | 44 | 0.0054 | 0.0820 | 0.0004 | 0.1091 | 0.0012 | 0.1396 | 0.0159 | 0.0004 | 0.0127 | 0.2673 | 494 |
| | 45 | 0.0076 | 0.3019 | 0.0002 | 0.0685 | 0.0009 | 0.0762 | 0.0054 | 0.0005 | 0.0203 | 0.1066 | 495 |
| | 46 | 0.0082 | 0.0416 | 0.0001 | 0.0427 | 0.0017 | 0.0625 | 0.0039 | 0.0006 | 0.0207 | 0.2069 | 496 |
| | 47 | 0.0049 | 0.0733 | 0.0003 | 0.1860 | 0.0010 | 0.0194 | 0.0156 | 0.0007 | 0.0158 | 0.0607 | 497 |
| | 48 | 0.0316 | 0.0352 | 0.0004 | 0.1408 | 0.0020 | 0.0085 | 0.0081 | 0.0008 | 0.0434 | 0.2563 | 498 |
| | 49 | 0.0115 | 0.0696 | 0.0002 | 0.0369 | 0.0005 | 0.0014 | 0.0112 | 0.0012 | 0.0149 | 0.0803 | 499 |
| | 50 | 0.0197 | 0.0088 | 0.0001 | 0.1028 | 0.0011 | 0.0028 | 0.0109 | 0.0012 | 0.0140 | 0.3174 | 500 |

Tableau 3.2.b : LMSE des trames de parole en français (en retenant un seul coefficient pour chaque trame)

| | N° de trame | Trames du signal de parole “ la brise et le soleil” | | | | | | | | | | |
|-----------------------------------|-------------|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----|
| L'erreur quadratique moyenne LMSE | 1 | 0 | 0.0066 | 0.0095 | 0.0000 | 0.0306 | 0.0013 | 0.0004 | 0.0050 | 0.0011 | 0.0050 | 451 |
| | 2 | 0 | 0.0286 | 0.0003 | 0.0003 | 0.0364 | 0.0009 | 0.0014 | 0.0017 | 0.0004 | 0.0156 | 452 |
| | 3 | 0.0000 | 0.0040 | 0.0017 | 0.0097 | 0.0824 | 0.0005 | 0.0003 | 0.0036 | 0.0007 | 0.0122 | 453 |
| | 4 | 0 | 0.0095 | 0.0020 | 0.0037 | 0.0278 | 0.0036 | 0.0046 | 0.0011 | 0.0010 | 0.0472 | 454 |
| | 5 | 0.0000 | 0.0187 | 0.0001 | 0.0007 | 0.0580 | 0.0019 | 0.0063 | 0.0012 | 0.0017 | 0.0054 | 455 |
| | 6 | 0.0000 | 0.0317 | 0.0009 | 0.0117 | 0.0346 | 0.0006 | 0.0019 | 0.0011 | 0.0007 | 0.0228 | 456 |
| | 7 | 0.0000 | 0.0130 | 0.0006 | 0.0091 | 0.0363 | 0.0029 | 0.0062 | 0.0004 | 0.0009 | 0.0362 | 457 |
| | 8 | 0.0000 | 0.0101 | 0.0014 | 0.0404 | 0.0771 | 0.0038 | 0.0043 | 0.0003 | 0.0006 | 0.0336 | 458 |
| | 9 | 0.0000 | 0.0268 | 0.0004 | 0.0303 | 0.0354 | 0.0021 | 0.0137 | 0.0002 | 0.0010 | 0.0297 | 459 |
| | 10 | 0.0000 | 0.0112 | 0.0007 | 0.0244 | 0.0326 | 0.0029 | 0.0044 | 0.0005 | 0.0006 | 0.0469 | 460 |
| | 11 | 0.0000 | 0.0124 | 0.0009 | 0.0295 | 0.0330 | 0.0063 | 0.0188 | 0.0004 | 0.0062 | 0.0188 | 461 |
| | 12 | 0.0000 | 0.0082 | 0.0002 | 0.0291 | 0.0981 | 0.0077 | 0.0050 | 0.0003 | 0.0057 | 0.0383 | 462 |
| | 13 | 0.0000 | 0.0205 | 0.0002 | 0.0341 | 0.0296 | 0.0069 | 0.0036 | 0.0002 | 0.0456 | 0.0183 | 463 |
| | 14 | 0.0000 | 0.0100 | 0.0005 | 0.0157 | 0.0287 | 0.0042 | 0.0114 | 0.0001 | 0.0281 | 0.0559 | 464 |
| | 15 | 0.0000 | 0.0206 | 0.0004 | 0.0304 | 0.0826 | 0.0019 | 0.0121 | 0.0003 | 0.0267 | 0.0315 | 465 |
| | 16 | 0.0000 | 0.0040 | 0.0002 | 0.0180 | 0.0417 | 0.0126 | 0.0049 | 0.0001 | 0.0124 | 0.0306 | 466 |
| | 17 | 0.0000 | 0.0247 | 0.0007 | 0.0324 | 0.0278 | 0.0255 | 0.0044 | 0.0003 | 0.0256 | 0.0249 | 467 |
| | 18 | 0.0000 | 0.0102 | 0.0012 | 0.0499 | 0.0124 | 0.0272 | 0.0251 | 0.0004 | 0.0243 | 0.0497 | 468 |
| | 19 | 0.0000 | 0.0421 | 0.0002 | 0.0177 | 0.0259 | 0.0175 | 0.0178 | 0.0002 | 0.0241 | 0.0493 | 469 |
| | 20 | 0.0000 | 0.0313 | 0.0003 | 0.0398 | 0.0256 | 0.0241 | 0.0061 | 0.0001 | 0.0103 | 0.0664 | 470 |
| | 21 | 0.0000 | 0.0829 | 0.0004 | 0.0307 | 0.0499 | 0.0316 | 0.0034 | 0.0004 | 0.0341 | 0.0494 | 471 |
| | 22 | 0.0004 | 0.0356 | 0.0003 | 0.0418 | 0.0224 | 0.0283 | 0.0074 | 0.0002 | 0.0103 | 0.0499 | 472 |
| | 23 | 0.0109 | 0.0667 | 0.0003 | 0.0357 | 0.0461 | 0.0297 | 0.0237 | 0.0001 | 0.0449 | 0.0662 | 473 |
| | 24 | 0.0005 | 0.0295 | 0.0004 | 0.0462 | 0.0214 | 0.0479 | 0.0384 | 0.0001 | 0.0172 | 0.0446 | 474 |
| | 25 | 0.0001 | 0.0876 | 0.0012 | 0.0420 | 0.0430 | 0.0419 | 0.0219 | 0.0001 | 0.0376 | 0.1203 | 475 |
| | 26 | 0.0002 | 0.0641 | 0.0003 | 0.0438 | 0.0146 | 0.0518 | 0.0262 | 0.0002 | 0.0273 | 0.0371 | 476 |
| | 27 | 0.0007 | 0.0328 | 0.0004 | 0.0316 | 0.0351 | 0.0182 | 0.0030 | 0.0002 | 0.0482 | 0.0992 | 477 |
| | 28 | 0.0029 | 0.1335 | 0.0003 | 0.0336 | 0.0137 | 0.0283 | 0.0405 | 0.0001 | 0.0284 | 0.0427 | 478 |
| | 29 | 0.0019 | 0.0489 | 0.0002 | 0.0476 | 0.0257 | 0.0237 | 0.0060 | 0.0001 | 0.0333 | 0.0934 | 479 |
| | 30 | 0.0048 | 0.0966 | 0.0003 | 0.0428 | 0.0131 | 0.0254 | 0.0239 | 0.0000 | 0.0158 | 0.0335 | 480 |
| | 31 | 0.0015 | 0.0569 | 0.0006 | 0.0308 | 0.0151 | 0.0185 | 0.0241 | 0.0002 | 0.0332 | 0.0956 | 481 |
| | 32 | 0.0035 | 0.1605 | 0.0011 | 0.0453 | 0.0171 | 0.0172 | 0.0308 | 0.0001 | 0.0050 | 0.0463 | 482 |
| | 33 | 0.0024 | 0.0450 | 0.0002 | 0.0521 | 0.0051 | 0.0299 | 0.0489 | 0.0001 | 0.0386 | 0.1190 | 483 |
| | 34 | 0.0028 | 0.1944 | 0.0002 | 0.0422 | 0.0059 | 0.0192 | 0.0496 | 0.0003 | 0.0078 | 0.0582 | 484 |
| | 35 | 0.0082 | 0.0404 | 0.0003 | 0.0370 | 0.0009 | 0.0266 | 0.0069 | 0.0002 | 0.0410 | 0.1158 | 485 |
| | 36 | 0.0025 | 0.1560 | 0.0004 | 0.0260 | 0.0067 | 0.0338 | 0.0213 | 0.0003 | 0.0058 | 0.0661 | 486 |
| | 37 | 0.0072 | 0.0244 | 0.0002 | 0.0399 | 0.0069 | 0.0295 | 0.0466 | 0.0002 | 0.0416 | 0.0599 | 487 |
| | 38 | 0.0023 | 0.0991 | 0.0007 | 0.0437 | 0.0018 | 0.0284 | 0.0196 | 0.0002 | 0.0087 | 0.1090 | 488 |
| | 39 | 0.0307 | 0.0906 | 0.0006 | 0.0439 | 0.0020 | 0.0250 | 0.0290 | 0.0003 | 0.0105 | 0.1262 | 489 |
| | 40 | 0.0089 | 0.1925 | 0.0000 | 0.0372 | 0.0037 | 0.0277 | 0.0059 | 0.0004 | 0.0125 | 0.0899 | 490 |
| | 41 | 0.0034 | 0.0526 | 0.0001 | 0.0334 | 0.0008 | 0.0450 | 0.0172 | 0.0005 | 0.0090 | 0.0571 | 491 |
| | 42 | 0.0097 | 0.0702 | 0.0002 | 0.0463 | 0.0019 | 0.0299 | 0.0364 | 0.0003 | 0.0296 | 0.1439 | 492 |
| | 43 | 0.0248 | 0.0812 | 0.0006 | 0.0332 | 0.0024 | 0.0153 | 0.0067 | 0.0003 | 0.0111 | 0.0484 | 493 |
| | 44 | 0.0054 | 0.0370 | 0.0004 | 0.0336 | 0.0012 | 0.0372 | 0.0159 | 0.0004 | 0.0127 | 0.1434 | 494 |
| | 45 | 0.0076 | 0.1073 | 0.0002 | 0.0291 | 0.0009 | 0.0062 | 0.0054 | 0.0005 | 0.0203 | 0.0353 | 495 |
| | 46 | 0.0082 | 0.0416 | 0.0001 | 0.0427 | 0.0017 | 0.0036 | 0.0039 | 0.0006 | 0.0207 | 0.1537 | 496 |
| | 47 | 0.0049 | 0.0509 | 0.0003 | 0.0457 | 0.0010 | 0.0194 | 0.0156 | 0.0007 | 0.0158 | 0.0448 | 497 |
| | 48 | 0.0316 | 0.0352 | 0.0004 | 0.0539 | 0.0020 | 0.0085 | 0.0081 | 0.0008 | 0.0434 | 0.2105 | 498 |
| | 49 | 0.0115 | 0.0406 | 0.0002 | 0.0369 | 0.0005 | 0.0014 | 0.0112 | 0.0012 | 0.0149 | 0.0335 | 499 |
| | 50 | 0.0197 | 0.0088 | 0.0001 | 0.0482 | 0.0011 | 0.0028 | 0.0109 | 0.0012 | 0.0140 | 0.2239 | 500 |

Tableau 3.2.c: LMSE du signal précédent (en retenant 2 coefficients pour les trames erronées).

3.4.1 Traitement du signal décomposé en trames sans recouvrement

| Nombre de coefficients des trames erronées N_C | LF=16 T =0.3233 | | LF=24 T=0.2776 | | LF=40 T=0.2179 | |
|---|-----------------|-------------------|----------------|-------------------|----------------|-------------------|
| | LMSE | MOS (dégradation) | LMSE | MOS (dégradation) | LMSE | MOS (dégradation) |
| 1 | 0.1031 | Peu gênante | 0.1861 | Gênante | 0.4116 | très Gênante |
| 2 | 0.0513 | Moins gênante | 0.1055 | Peu gênante | 0.2568 | Gênante |
| 3 | 0.0314 | Non gênante | 0.0717 | Peu gênante | 0.1969 | Peu gênante |
| 4 | 0.0244 | Non gênante | 0.0549 | Moins gênante | 0.1482 | Peu gênante |
| 5 | 0.0000 | Imperceptible | 0.0002 | Non gênante | 0.0002 | Non gênante |

Tableau 3.3 : Evaluation de la qualité de parole (trames sans recouvrement) en fonction de LF et de N_C . T désigne le seuil de comparaison utilisé pour éliminer les vecteurs semblables.

Les valeurs du tableau 3.3 sont obtenues de la manière suivante :

- les valeurs de la première ligne : le signal est reconstruit, en utilisant un seul coefficient et un seul vecteur des fonctions de signatures, pour toute les trames.
- les valeurs de la deuxième ligne : le signal est reconstruit, par un système hybride. Les trames non voisées sont synthétisées avec un seul coefficient et un seul vecteur. Les trames voisées sont reconstruites avec deux coefficients et deux vecteurs SR1/SR2.
- les valeurs des lignes 3,4 et 5 correspondent à une synthèse avec 3,4 et 5 coefficients respectivement, pour les trames voisées.

On constate que la qualité du son s'améliore quand la taille des trames diminue et le nombre de coefficients retenus augmente. Par exemple, pour le cas de 5 coefficients l'erreur est très faible (=0.0002). Ce qui veut dire que, le signal original et le reconstruit sont confondus. Alors que, pour le cas d'un seul coefficient par trame, la dégradation du signal reconstruit est entre perceptible et gênante. En plus, cette expérience met en évidence l'influence de la longueur de la trame LF sur la qualité du signal reconstitué.

La figure 3.4 montre qu'on peut sélectionner le nombre désiré de vecteurs propres de façon à limiter la reconstruction du signal aux seuls vecteurs qui nous semblent vraiment signifiants.

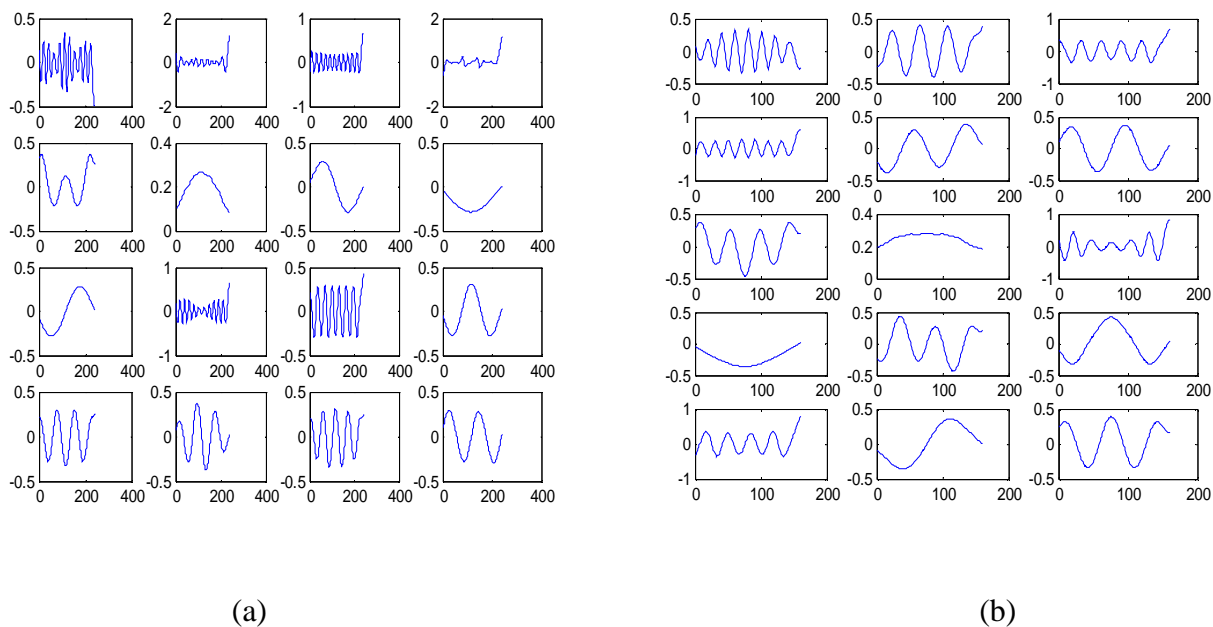


Figure 3.4 : Les fonctions de signatures a) pour LF=40, SR= 16 vecteurs, b) pour LF=16, SR= 15 vecteurs.

On peut ainsi filtrer de la matrice de corrélation toute information qui nous semble non pertinente. Les tableaux ci-dessous, donnent les séquences de signatures pour chaque longueur de trame.

| Nombre de vecteurs propres retenus | LF=16 | LF=24 | LF=40 |
|------------------------------------|-------|-------|-------|
| 1 | 15 | 16 | 16 |
| 2 | 19 | 15 | 18 |
| 3 | 20 | 17 | 17 |
| 4 | 19 | 18 | 18 |
| 5 | 18 | 20 | 22 |

(a)

| Nombre de vecteurs propres retenus | LF=16 | LF=24 | LF=40 |
|------------------------------------|-------|-------|-------|
| 1 | 15 | 16 | 16 |
| 2 | 14 | 24 | 17 |
| 3 | 12 | 25 | 16 |
| 4 | 13 | 25 | 17 |
| 5 | 12 | 25 | 19 |

(b)

Tableau 3.4 : Nombre de fonctions de signatures retenues. a) Trames sans recouvrement. b) Trames avec recouvrement de 50%

On peut effectuer des tests de signification sur les valeurs propres pour introduire les coefficients qui ont une contribution de représentation (les séquences SR1, SR2, SR3, SR4 et

SR5 correspondant respectivement aux premier, deuxième, troisième, quatrième et cinquième coefficients retenus pour l'approximation du signal).

3.4.2 Traitement du signal décomposé en trames avec recouvrement de 50%

| Nombre de coefficients des trames erronées | LF=16 T=0.3459 | | LF=24 T=0.2689 | | LF=40 T=0.2204 | |
|--|----------------|--------------------|----------------|--------------------|----------------|---------------------|
| | LMSE | MOS | LMSE | MOS | LMSE | MOS |
| 1 | 0.0448 | Il y a du bruit | 0.0828 | Il y a du bruit | 0.1603 | Dégradation gênante |
| 2 | 0.0202 | Bonne qualité | 0.0381 | Bonne qualité | 0.0894 | Parole intelligible |
| 3 | 0.0127 | Bonne qualité | 0.0229 | Bonne qualité | 0.0599 | Il y a du bruit |
| 4 | 0.0099 | Très bonne qualité | 0.0154 | Très bonne qualité | 0.0437 | Bonne qualité |
| 5 | 0.0080 | Excellente qualité | 0.0119 | Très bonne qualité | 0.0334 | Bonne qualité |

Tableau 3.5 : Evaluation de la qualité de synthèse par des trames avec recouvrement de 50 %

Dans cette section, on refait les mêmes expériences qui ont été faites au paragraphe précédent, mais appliquées à un signal de paroles décomposé avec un recouvrement de 50%.

Les résultats obtenus après traitement de la phrase : « addarssso thani el arabo kabla », pour le cas d'un recouvrement de 50% entre trames, sont résumés dans le tableau 3.5.

En comparant les tableaux 3.3 et 3.5, on constate une amélioration des résultats. Dans chaque cas LMSE diminue ce qui donne une meilleure qualité.

3.5. Algorithme FS pour plusieurs phrases

Afin d'étudier la robustesse des théories et des modèles précédemment présentés, nous allons appliquer l'algorithme mis au point sur un corpus de parole constitué de plusieurs phrases. Pour les deux premières expériences, concernant un seul locuteur (homme, puis femme), les dix phrases arabes traitées ont été extraites du corpus de parole (MSA_DARS01_NAS) et (MSA_DARS02_NAS) qui a été enregistré dans des conditions convenables. Les passages de l'arabe traités (lus par les deux locuteurs) sont donnés dans l'appendice D tableau D.1.

La dernière expérience, consiste à appliquer l'algorithme FS sur un corpus de parole plus important. Il s'agit de parole échantillonnée à 8KHz, enregistrée dans des conditions ordinaires (bruitées). Cette base de données est intéressante puisqu'elle permet de se

rapprocher des conditions réelles d'utilisation d'un système de compression et de synthèse. Les données correspondent au jeu de 20 locuteurs (dix hommes et dix femmes) sur les passages lus de dix phrases. Les passages lus par les 20 locuteurs sont regroupés dans l'annexe D tableau D.2. Au total 200 phrases, de 24000 échantillons chacune, sont lues ce qui représente 4800000 échantillons de parole.

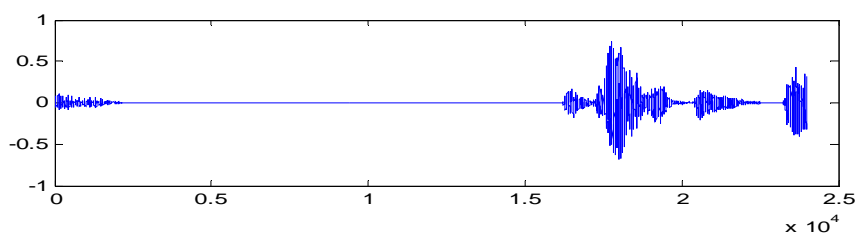
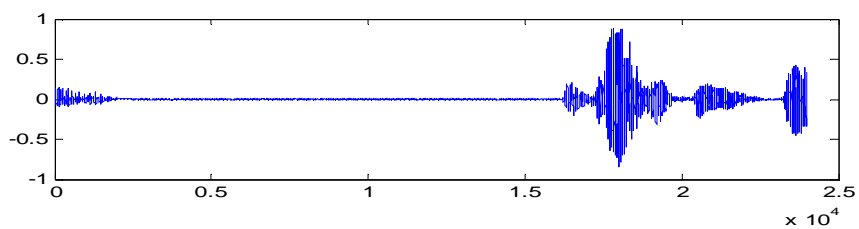
3.5.1. Système mono locuteur

3.5.1.1. Locuteur homme1

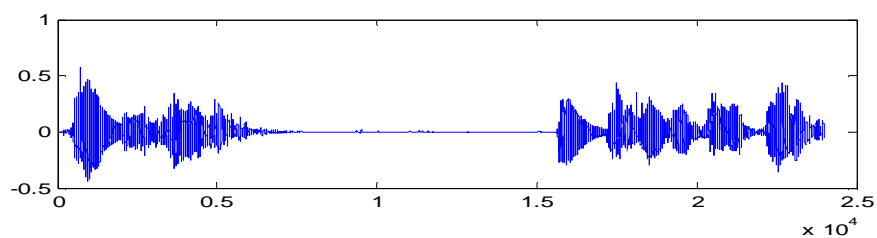
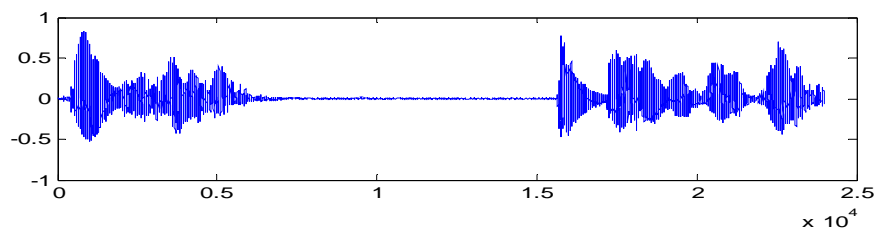
Dans ce cas, une seule et même personne prononce une dizaine de phrases (en arabe) de 24000 échantillons, de façon à créer une séquence de signature (références acoustiques). Les résultats présentés dans cette section concernent la reconstruction de quelques phrases. Trois phrases, parmi le corpus d'entrée, ont été synthétisées. On constate, que la qualité du son reconstruit, s'est améliorée en la comparant aux résultats précédents.

| | N _c | LF= 16 | | LF= 24 | | LF= 40 | |
|----------|----------------|--------|---|--------|---|--------|---|
| | | LMSE | MOS | LMSE | MOS | LMSE | MOS |
| Phrase 1 | 1 | 0.0227 | Bonne qualité en croissance avec N _c (excellente) | 0.0553 | Qualité acceptable et parfois très bonne | 0.1189 | Dégradation gênante surtout pour le cas d'un seul coefficient |
| | 2 | 0.0111 | | 0.0259 | | 0.0749 | |
| | 3 | 0.0080 | | 0.0173 | | 0.0542 | |
| | 4 | 0.0069 | | 0.0118 | | 0.0355 | |
| | 5 | 0.0064 | | 0.0099 | | 0.0279 | |
| Phrase 2 | 1 | 0.0150 | Très bonne qualité | 0.0321 | Bonne qualité en croissance avec N _c | 0.0656 | Qualité moyenne en retenant un seul coeff et bonne ailleurs |
| | 2 | 0.0072 | | 0.0161 | | 0.0423 | |
| | 3 | 0.0046 | | 0.0112 | | 0.0314 | |
| | 4 | 0.0037 | | 0.0076 | | 0.0219 | |
| | 5 | 0.0031 | | 0.0062 | | 0.0185 | |
| Phrase10 | 1 | 0.0620 | Qualité moyenne en retenant un seul coeff et très bonne ailleurs. | 0.1354 | Dégradation gênante surtout pour le cas d'un seul coefficient | 0.2832 | Dégradation gênante surtout pour le cas d'un seul coefficient |
| | 2 | 0.0259 | | 0.0608 | | 0.1695 | |
| | 3 | 0.0163 | | 0.0392 | | 0.1188 | |
| | 4 | 0.0128 | | 0.0244 | | 0.0801 | |
| | 5 | 0.0109 | | 0.0191 | | 0.0633 | |

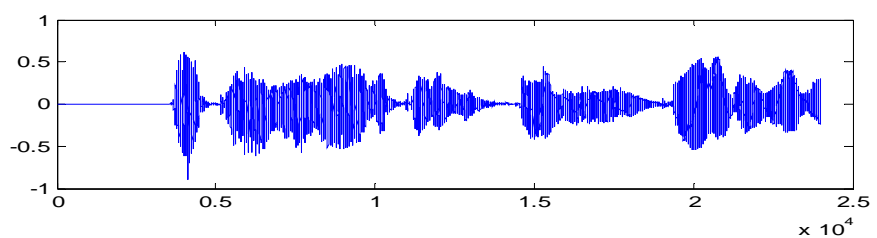
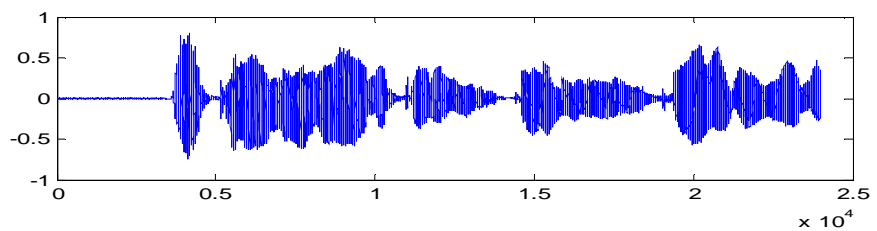
Tableau 3.6.a : Résultats de synthèse de 3 phrases prononcées par homme1



(a)



(b)



(c)

Figure 3.5.a : Parole originale et synthétisée des phrases 1, 2,10 pour LF=16.

La figure 3.5.a montre le signal original et le reconstruit de trois phrases. La synthèse a été faite à l'aide des fonctions de signature, obtenues avec un corpus de données de dix phrases, extraites du fichier son MSA_DARS01_NAS.wav d'un locuteur homme.

On observe, sur les différents graphes que les signaux originaux et les reconstitués sont très proches. Les résultats donnés par, le tableau 3.6.a et la figure 3.5.a, sont en accord.

3.5.1.2. Locuteur femme1

Dans cette partie, on refait les mêmes expériences qui ont été faites au paragraphe précédent, mais appliquées à des signaux de paroles extraits du fichier son MSA_DARS02_NAS.wav, d'un locuteur femme.

Les résultats obtenus, après traitement des dix phrases, sont résumés dans le tableau 3.6.b

| | N _c | T=0.3309 LF= 16 | | T= 0.2759 LF= 24 | | T= 0.2214 LF= 40 | |
|----------|----------------|--------------------|---|---------------------|---|---------------------|---|
| | | LMSE | MOS | LMSE | MOS | LMSE | MOS |
| Phrase 1 | 1 | 0.0412 | Bonne qualité. | 0.0852 | Qualité acceptable et parfois très bonne. | 0.1660 | Dégradation gênante surtout pour le cas d'un seul coefficient |
| | 2 | 0.0187 | | 0.0409 | | 0.1080 | |
| | 3 | 0.0117 | | 0.0273 | | 0.0872 | |
| | 4 | 0.0089 | | 0.0164 | | 0.0664 | |
| | 5 | 0.0073 | | 0.0125 | | 0.0477 | |
| Phrase 3 | 1 | 0.0233 | Très bonne qualité. | 0.0467 | Bonne qualité. | 0.0908 | Dégradation gênante surtout pour le cas d'un seul coefficient |
| | 2 | 0.0122 | | 0.0245 | | 0.0596 | |
| | 3 | 0.0086 | | 0.0175 | | 0.0458 | |
| | 4 | 0.0073 | | 0.0123 | | 0.0356 | |
| | 5 | 0.0066 | | 0.0104 | | 0.0288 | |
| Phrase6 | 1 | 0.0341 | Bonne qualité. en croissance avec N _c (devient excellente) | 0.0672 | Qualité moyenne. | 0.1243 | Dégradation gênante surtout pour le cas d'un seul coefficient |
| | 2 | 0.0143 | | 0.0288 | | 0.0721 | |
| | 3 | 0.0097 | | 0.0194 | | 0.0508 | |
| | 4 | 0.0081 | | 0.0128 | | 0.0351 | |
| | 5 | 0.0074 | | 0.0105 | | 0.0278 | |

Tableau 3.6.b : Résultats de synthèse de 3 phrases prononcées par femme1

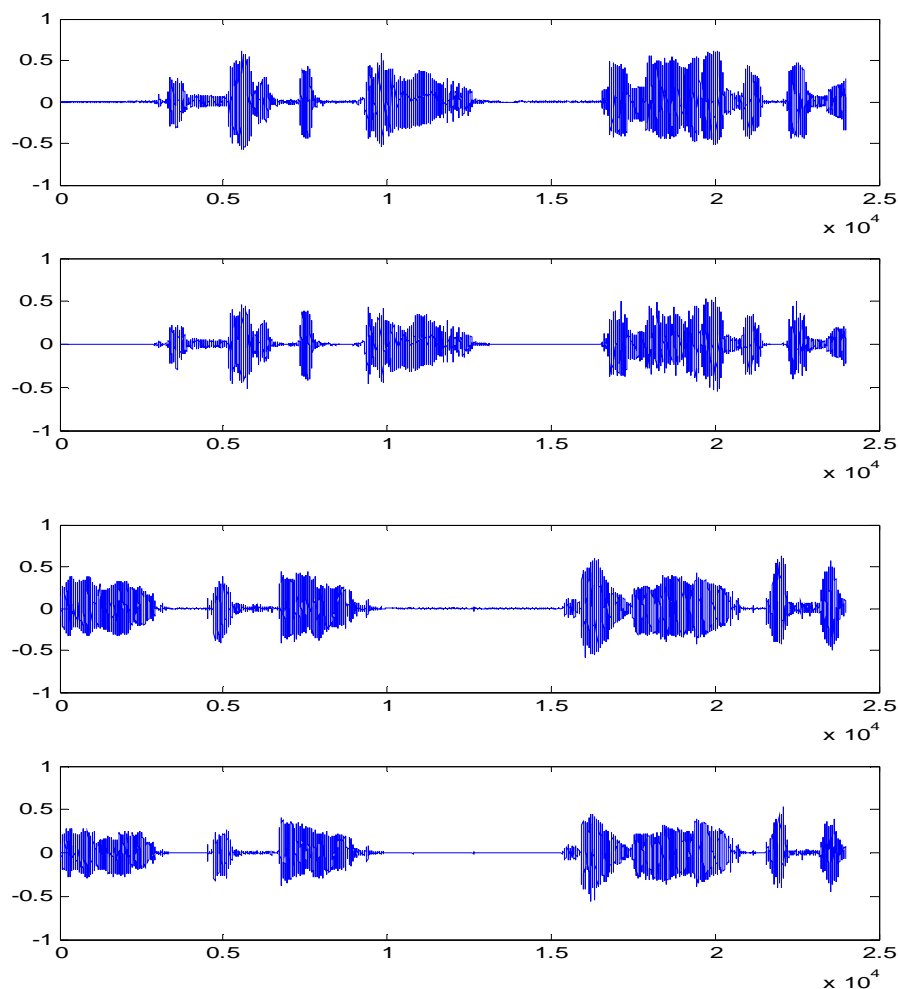


Figure 3.5.b : Parole originale et synthétisée des phrases 1,3 pour LF=16.

On peut faire les mêmes observations et remarques que précédemment, pour le tableau 3.6.b. Les résultats restent presque inchangés.

Les figures 3.5 sont obtenues en retenant deux coefficients pour les trames ayant une LMSE supérieure à 0.05.

On remarque que le meilleur score de qualité s'obtient avec LF=16. La deuxième phrase est mieux reconstruite que les autres (tableau 3.6.a), peut être parce qu'elle est courte (ne contient pas beaucoup de mots). La dégradation est perceptible et même gênante dans le cas où LF=40 et surtout pour les phrases longues.

En comparant, les tableaux 3.6.a et 3.6.b, on constate que LMSE varie en fonction de la « quantité de parole » et du type de locuteur.

Les phrases reconstruites pour les deux locuteurs sont données dans l'appendice D.3.

Il n'y a pas de recette magique pour interpréter les résultats de cette analyse. Il s'agit d'un travail de synthèse qui demande de considérer plusieurs paramètres (longueur de trame, nombre d'échantillons considérés, taux de recouvrement, types de locuteurs et leur nombre).

3.5.2. Système multilocuteur

Pour être pratique, on a besoin de trouver des fonctions de signature optimales. Cela veut dire, appliquer l'algorithme de sélection de base (l'algorithme FS), à un corpus de données plus important.

On enregistre dix phrases, prononcées par 20 locuteurs (10 femmes et 10 hommes), échantillonnées à une fréquence de 8 kHz. Les détails sur le corpus utilisé dans ces expériences se trouvent dans l'appendice D.

Les résultats des différentes expériences ont été regroupés dans les tableaux 3.7 (a,b) et 3.8 (a,b) qui peuvent être interprétés comme suit :

- Les tableaux 3.7.a et b montrent que la qualité du son reproduit est très bonne et même excellente dans l'ensemble des cas et surtout lorsque deux coefficients sont retenus pour les trames erronées. Durant les essais d'écoute, on a remarqué que la dégradation du signal devient perceptible lorsque la voix du locuteur est masculine et la phrase prononcée est longue (contient plusieurs mots).
- La plus petite LMSE générée est 0.0075 (avec un seul coefficient), elle est produite par le locuteur femme 2 (phrase 1).
- En prenant deux coefficients pour les trames erronées, les deuxièmes valeurs (en rouge) des tableaux 3.7.a et 3.7.b montrent que la LMSE diminue d'une façon considérable dans la plupart des cas. Donc la qualité de la parole reconstruite s'améliore. Dans ce cas on a :
- $LMSE_{\min} = 0.0050$ (femme 8 phrase 1) ce qui donne une bonne qualité.

| Phrases | | P 1 | | P 2 | | P 3 | | P4 | | P5 | |
|------------------------|---|------------|--------|------------|--------|------------|--------|-------------|--------|-------------|--------|
| Nombre de coefficients | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Locuteurs Femme1 | E | 0.0369 | 0.0157 | 0.0626 | 0.0265 | 0.0539 | 0.0223 | 0.0637 | 0.0240 | 0.1196 | 0.0470 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Femme2 | E | 0.0075 | 0.0054 | 0.0262 | 0.0122 | 0.0409 | 0.0168 | 0.0434 | 0.0210 | 0.0232 | 0.0113 |
| | Q | Très bonne | | Très bonne | | Bonne | | Bonne | | Très bonne | |
| Femme3 | E | 0.0264 | 0.0129 | 0.0856 | 0.0358 | 0.0412 | 0.0172 | 0.0591 | 0.0228 | 0.0863 | 0.0333 |
| | Q | Très bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Femme4 | E | 0.0171 | 0.0088 | 0.0167 | 0.0092 | 0.0204 | 0.0095 | 0.0245 | 0.0104 | 0.0314 | 0.0142 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Bonne | | Bonne | |
| Femme5 | E | 0.0221 | 0.0091 | 0.0214 | 0.0098 | 0.0140 | 0.0063 | 0.0233 | 0.0105 | 0.0293 | 0.0124 |
| | Q | Bonne | | Bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme6 | E | 0.0111 | 0.0060 | 0.0264 | 0.0101 | 0.0459 | 0.0178 | 0.0663 | 0.0224 | 0.0624 | 0.0200 |
| | Q | Très bonne | | Bonne | | Bonne | | Très bonne | | Très bonne | |
| Femme7 | E | 0.0175 | 0.0055 | 0.0122 | 0.0075 | 0.0149 | 0.0073 | 0.0292 | 0.0127 | 0.0172 | 0.0081 |
| | Q | Très bonne | | Bonne | | Très bonne | | Bonne | | Excellente | |
| Femme8 | E | 0.0076 | 0.0050 | 0.0266 | 0.0122 | 0.0394 | 0.0165 | 0.0697 | 0.0273 | 0.0629 | 0.0255 |
| | Q | Bonne | | Très bonne | | moyenne | | Bonne | | Moins bonne | |
| Femme9 | E | 0.0222 | 0.0097 | 0.0315 | 0.0133 | 0.0466 | 0.0196 | 0.0496 | 0.0199 | 0.0396 | 0.0153 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme1 | E | 0.0129 | 0.0049 | 0.2485 | 0.1093 | 0.3168 | 0.1359 | 0.4598 | 0.1809 | 0.3934 | 0.1667 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme2 | E | 0.0854 | 0.0356 | 0.1257 | 0.0483 | 0.1566 | 0.0603 | 0.1189 | 0.0414 | 0.1601 | 0.0579 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme3 | E | 0.0632 | 0.0293 | 0.1159 | 0.0520 | 0.2862 | 0.1160 | 0.5638 | 0.2314 | 0.5035 | 0.2121 |
| | Q | Bonne | | Bonne | | Bonne | | Moins bonne | | Moins bonne | |
| Homme4 | E | 0.1496 | 0.0696 | 0.3647 | 0.1623 | 0.4834 | 0.2029 | 0.4946 | 0.2070 | 0.7163 | 0.3131 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Moins bonne | |
| Homme5 | E | 0.2162 | 0.0971 | 0.4213 | 0.1846 | 0.3930 | 0.1681 | 0.3656 | 0.1541 | 0.4185 | 0.1764 |
| | Q | Bonne | | Très bonne | | Bonne | | Bonne | | Moins bonne | |
| Homme6 | E | 0.0988 | 0.0430 | 0.3116 | 0.1331 | 0.2239 | 0.0883 | 0.2950 | 0.1194 | 0.2805 | 0.1153 |
| | Q | Bonne | | Bonne | | Très bonne | | Bonne | | Bonne | |
| Homme7 | E | 0.0824 | 0.0388 | 0.0962 | 0.0430 | 0.1370 | 0.0594 | 0.1731 | 0.0724 | 0.1213 | 0.0537 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Bonne | |
| Homme8 | E | 0.0274 | 0.0124 | 0.0592 | 0.0245 | 0.0971 | 0.0352 | 0.0812 | 0.0294 | 0.0525 | 0.0203 |
| | Q | Très bonne | | Très bonne | | Bonne | | Bonne | | Bonne | |
| Homme9 | E | 0.0205 | 0.0100 | 0.0210 | 0.0111 | 0.0297 | 0.0127 | 0.0364 | 0.0152 | 0.0296 | 0.0129 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Homme10 | E | 0.2279 | 0.1040 | 0.0783 | 0.0360 | 0.1586 | 0.0686 | 0.1408 | 0.0587 | 0.1574 | 0.0593 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |

(a)

| Phrases | P 6 | | P 7 | | P 8 | | P9 | | P10 | | |
|------------------------|-----|-------------|--------|-------------|--------|-------------|--------|-------------|--------|------------|--------|
| Nombre de coefficients | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | |
| Locuteurs Femme1 | E | 0.1196 | 0.0494 | 0.0432 | 0.0168 | 0.0362 | 0.0140 | 0.0638 | 0.0248 | 0.0606 | 0.0244 |
| | Q | Bonne | | Bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme2 | E | 0.0624 | 0.0182 | 0.0377 | 0.0160 | 0.0384 | 0.0180 | 0.0538 | 0.0148 | 0.0525 | 0.0240 |
| | Q | Bonne | | Très bonne | | Très bonne | | Bonne | | Bonne | |
| Femme3 | E | 0.0829 | 0.0333 | 0.0836 | 0.0324 | 0.0691 | 0.0313 | 0.0809 | 0.0330 | 0.0559 | 0.0243 |
| | Q | Bonne | | Bonne | | Bonne | | Très bonne | | Excellente | |
| Femme4 | E | 0.0337 | 0.0156 | 0.0376 | 0.0183 | 0.0384 | 0.0180 | 0.0256 | 0.0122 | 0.0216 | 0.0113 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Femme5 | E | 0.0178 | 0.0085 | 0.0224 | 0.0101 | 0.0211 | 0.0098 | 0.0181 | 0.0089 | 0.0174 | 0.0083 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme6 | E | 0.0724 | 0.0282 | 0.1009 | 0.0394 | 0.1176 | 0.0402 | 0.1061 | 0.0380 | 0.0700 | 0.0255 |
| | Q | Bonne | | Bonne | | Bonne | | Moins bonne | | Très bonne | |
| Femme7 | E | 0.0187 | 0.0078 | 0.0173 | 0.0090 | 0.0146 | 0.0074 | 0.0223 | 0.0093 | 0.0128 | 0.0067 |
| | Q | Excellente | | Excellente | | Excellente | | Très bonne | | Excellente | |
| Femme8 | E | 0.0672 | 0.0291 | 0.0342 | 0.0154 | 0.0811 | 0.0356 | 0.0340 | 0.0156 | 0.0508 | 0.0231 |
| | Q | Moins bonne | | Bonne | | Bonne | | Très bonne | | Très bonne | |
| Femme9 | E | 0.0665 | 0.0287 | 0.0645 | 0.0255 | 0.0330 | 0.0141 | 0.0815 | 0.0335 | 0.0317 | 0.0158 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme1 | E | 0.3987 | 0.1790 | 0.4043 | 0.1810 | 0.2858 | 0.1217 | 0.3767 | 0.1646 | 0.1854 | 0.0875 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme2 | E | 0.1786 | 0.0659 | 0.2045 | 0.0738 | 0.1178 | 0.0440 | 0.1358 | 0.0501 | 0.0435 | 0.0188 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme3 | E | 0.5303 | 0.2242 | 0.4864 | 0.1952 | 0.3274 | 0.1344 | 0.3688 | 0.1544 | 0.3042 | 0.1302 |
| | Q | Moins bonne | | Moins bonne | | Bonne | | Bonne | | Bonne | |
| Homme4 | E | 0.6485 | 0.2968 | 0.5721 | 0.2504 | 0.4808 | 0.2008 | 0.4889 | 0.2082 | 0.3024 | 0.1406 |
| | Q | Moins bonne | | Moins bonne | | Bonne | | Bonne | | Bonne | |
| Homme5 | E | 0.3887 | 0.1788 | 0.4382 | 0.1894 | 0.3627 | 0.1624 | 0.3109 | 0.1384 | 0.2882 | 0.1341 |
| | Q | Bonne | | Moins bonne | | Moins bonne | | Bonne | | Bonne | |
| Homme6 | E | 0.3248 | 0.1310 | 0.4702 | 0.1785 | 0.2538 | 0.1035 | 0.4075 | 0.1600 | 0.2169 | 0.0923 |
| | Q | Bonne | | Moins bonne | | Bonne | | Moins bonne | | Bonne | |
| Homme7 | E | 0.1022 | 0.0459 | 0.1619 | 0.0692 | 0.0654 | 0.0313 | 0.0909 | 0.0413 | 0.0922 | 0.0427 |
| | Q | Bonne | | Très bonne | | Très bonne | | Très bonne | | Excellente | |
| Homme8 | E | 0.0661 | 0.0254 | 0.0439 | 0.0177 | 0.0418 | 0.0179 | 0.0441 | 0.0189 | 0.0319 | 0.0135 |
| | Q | Bonne | | Bonne | | Très bonne | | Bonne | | Excellente | |
| Homme9 | E | 0.0293 | 0.0136 | 0.0347 | 0.0141 | 0.0223 | 0.0115 | 0.0315 | 0.0138 | 0.0130 | 0.0066 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Excellente | |
| Homme10 | E | 0.2475 | 0.1006 | 0.2473 | 0.1075 | 0.2480 | 0.0975 | 0.2807 | 0.1202 | 0.2567 | 0.1041 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Très bonne | |

(b)

Tableau 3.7 : LMSE (obtenue avec un seul et deux coefficients) pour LF=16.

a) Phrases (1, 2,3,4,5) b) Phrases (6,7,8,9,10)

| Phrases | | P 1 | | P 2 | | P 3 | | P4 | | P5 | |
|------------------------|---|------------|--------|-------------|--------|-------------|--------|-------------|--------|-------------|--------|
| Nombre de coefficients | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Locuteurs Femme1 | E | 0.0643 | 0.0292 | 0.1228 | 0.0530 | 0.1064 | 0.0506 | 0.1161 | 0.0526 | 0.2370 | 0.1095 |
| | Q | Bonne | | Moyenne | | Moyenne | | Bonne | | Moins bonne | |
| Femme2 | E | 0.0937 | 0.0447 | 0.3362 | 0.1475 | 0.4257 | 0.2010 | 0.3567 | 0.1601 | 0.4351 | 0.1921 |
| | Q | Bonne | | moins bonne | | moins bonne | | moins bonne | | moins bonne | |
| Femme3 | E | 0.0459 | 0.0223 | 0.1759 | 0.0762 | 0.0783 | 0.0332 | 0.1183 | 0.0492 | 0.1639 | 0.0700 |
| | Q | Très bonne | | moins bonne | | Bonne | | Bonne | | Moyenne | |
| Femme4 | E | 0.0322 | 0.0161 | 0.0330 | 0.0183 | 0.0416 | 0.0212 | 0.0498 | 0.0248 | 0.0654 | 0.0319 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Moyenne | |
| Femme5 | E | 0.0478 | 0.0210 | 0.0427 | 0.0194 | 0.0270 | 0.0132 | 0.0454 | 0.0203 | 0.0570 | 0.0249 |
| | Q | Bonne | | Bonne | | Très bonne | | Moyenne | | Moyenne | |
| Femme6 | E | 0.0239 | 0.0100 | 0.0540 | 0.0238 | 0.0926 | 0.0391 | 0.1327 | 0.0586 | 0.1464 | 0.0656 |
| | Q | Très bonne | | Bonne | | Bonne | | Moins bonne | | Moins bonne | |
| Femme7 | E | 0.0262 | 0.0122 | 0.0253 | 0.0128 | 0.0279 | 0.0137 | 0.0576 | 0.0276 | 0.0345 | 0.0175 |
| | Q | Bonne | | Bonne | | Bonne | | Moyenne | | Bonne | |
| Femme8 | E | 0.0143 | 0.0080 | 0.0522 | 0.0226 | 0.0769 | 0.0350 | 0.1321 | 0.0563 | 0.1295 | 0.0578 |
| | Q | Très bonne | | Bonne | | Moyenne | | Moyenne | | Moins bonne | |
| Femme9 | E | 0.0453 | 0.0220 | 0.0652 | 0.0289 | 0.0965 | 0.0470 | 0.0970 | 0.0460 | 0.0800 | 0.0373 |
| | Q | Bonne | | Moyenne | | Moyenne | | Moyenne | | Bonne | |
| Homme1 | E | 0.0232 | 0.0113 | 0.5534 | 0.2519 | 0.6415 | 0.3174 | 0.8965 | 0.4138 | 0.7910 | 0.3784 |
| | Q | Bonne | | Moins bonne | | Moins bonne | | Moins bonne | | Moins bonne | |
| Homme2 | E | 0.1781 | 0.0906 | 0.2665 | 0.1165 | 0.3168 | 0.1409 | 0.2231 | 0.0934 | 0.3177 | 0.1381 |
| | Q | Bonne | | Bonne | | Moyenne | | Moyenne | | Moyenne | |
| Homme3 | E | 0.1304 | 0.0681 | 0.2424 | 0.1238 | 0.5550 | 0.2848 | 1.0496 | 0.4758 | 0.9586 | 0.4728 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme4 | E | 0.2956 | 0.1653 | 0.6630 | 0.3301 | 0.9357 | 0.4785 | 1.0046 | 0.4682 | 1.4147 | 0.6816 |
| | Q | Bonne | | Bonne | | Moins bonne | | Moins bonne | | Moins bonne | |
| Homme5 | E | 0.4386 | 0.2264 | 0.8503 | 0.3963 | 0.7736 | 0.3978 | 0.6239 | 0.3041 | 0.7105 | 0.3516 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme6 | E | 0.1971 | 0.0935 | 0.6183 | 0.2799 | 0.4276 | 0.2059 | 0.5588 | 0.2355 | 0.5518 | 0.2591 |
| | Q | Bonne | | Moyenne | | Moyenne | | Moyenne | | Moyenne | |
| Homme7 | E | 0.1698 | 0.0954 | 0.1831 | 0.0888 | 0.2754 | 0.1461 | 0.3524 | 0.1680 | 0.2756 | 0.1432 |
| | Q | Bonne | | Bonne | | Bonne | | Moyenne | | Bonne | |
| Homme8 | E | 0.0537 | 0.0278 | 0.1170 | 0.0592 | 0.1950 | 0.0966 | 0.1468 | 0.0666 | 0.1069 | 0.0513 |
| | Q | Bonne | | Bonne | | Moyenne | | Moyenne | | Moyenne | |
| Homme9 | E | 0.0419 | 0.0224 | 0.0444 | 0.0222 | 0.0620 | 0.0307 | 0.0725 | 0.0349 | 0.0570 | 0.0286 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme10 | E | 0.4401 | 0.2203 | 0.1589 | 0.0800 | 0.2910 | 0.1547 | 0.2837 | 0.1358 | 0.2790 | 0.1318 |
| | Q | Moyenne | | Moyenne | | Moyenne | | Moyenne | | Moyenne | |

(a)

| Phrases | P 6 | | P 7 | | P 8 | | P9 | | P10 | | |
|------------------------|-----|-------------|--------|-------------|--------|-------------|--------|-------------|--------|------------|--------|
| Nombre de coefficients | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | |
| Locuteurs Femme1 | E | 0.2517 | 0.1145 | 0.0845 | 0.0385 | 0.0650 | 0.0260 | 0.1142 | 0.0499 | 0.1137 | 0.0511 |
| | Q | Moyenne | | Moyenne | | Moyenne | | Moyenne | | Moyenne | |
| Femme2 | E | 0.3667 | 0.1639 | 0.2812 | 0.1284 | 0.2681 | 0.1168 | 0.1941 | 0.0922 | 0.2127 | 0.0987 |
| | Q | Bonne | | Bonne | | Moyenne | | Moyenne | | Bonne | |
| Femme3 | E | 0.1637 | 0.0736 | 0.1599 | 0.0643 | 0.1276 | 0.0581 | 0.1628 | 0.0745 | 0.1005 | 0.0481 |
| | Q | Bonne | | Moyenne | | Moyenne | | Moyenne | | Moyenne | |
| Femme4 | E | 0.0664 | 0.0340 | 0.0779 | 0.0400 | 0.0752 | 0.0375 | 0.0501 | 0.0250 | 0.0413 | 0.0239 |
| | Q | Bonne | | Moyenne | | Moyenne | | Bonne | | Bonne | |
| Femme5 | E | 0.0355 | 0.0169 | 0.0476 | 0.0226 | 0.0449 | 0.0198 | 0.0368 | 0.0174 | 0.0344 | 0.0155 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Femme6 | E | 0.1585 | 0.0716 | 0.2083 | 0.0977 | 0.2564 | 0.1158 | 0.2138 | 0.0985 | 0.1506 | 0.0662 |
| | Q | Moyenne | | Moins bonne | | Moins bonne | | Moyenne | | Moyenne | |
| Femme7 | E | 0.0377 | 0.0159 | 0.0367 | 0.0174 | 0.0311 | 0.0151 | 0.0437 | 0.0185 | 0.0251 | 0.0125 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Très bonne | |
| Femme8 | E | 0.1548 | 0.0656 | 0.0752 | 0.0334 | 0.1524 | 0.0701 | 0.0669 | 0.0313 | 0.0997 | 0.0456 |
| | Q | Moyenne | | Bonne | | Moins bonne | | Moyenne | | Moyenne | |
| Femme9 | E | 0.1424 | 0.0684 | 0.1338 | 0.0622 | 0.0660 | 0.0293 | 0.1619 | 0.0782 | 0.0673 | 0.0328 |
| | Q | Moins bonne | | Moins bonne | | Moyenne | | Moins bonne | | Moyenne | |
| Homme1 | E | 0.8273 | 0.3712 | 0.8948 | 0.4165 | 0.5440 | 0.2783 | 0.7569 | 0.3509 | 0.3955 | 0.1854 |
| | Q | Moins bonne | | Moins bonne | | Bonne | | Moins bonne | | Bonne | |
| Homme2 | E | 0.4007 | 0.1669 | 0.4254 | 0.1888 | 0.2333 | 0.1065 | 0.2682 | 0.1075 | 0.0809 | 0.0364 |
| | Q | Moyenne | | Moyenne | | Moyenne | | Moyenne | | Bonne | |
| Homme3 | E | 1.0257 | 0.5295 | 0.9025 | 0.4364 | 0.6309 | 0.2892 | 0.7104 | 0.3425 | 0.5819 | 0.2898 |
| | Q | Moins bonne | | Moyenne | | Bonne | | Bonne | | Bonne | |
| Homme4 | E | 1.2320 | 0.6186 | 1.1427 | 0.5394 | 0.9222 | 0.4336 | 0.9288 | 0.4367 | 0.5522 | 0.2866 |
| | Q | Moins bonne | | Moins bonne | | Bonne | | Bonne | | Bonne | |
| Homme5 | E | 0.8070 | 0.4077 | 0.8387 | 0.4193 | 0.8565 | 0.4158 | 0.6904 | 0.3599 | 0.5273 | 0.2884 |
| | Q | Bonne | | Moyenne | | Moyenne | | Bonne | | Bonne | |
| Homme6 | E | 0.6384 | 0.3052 | 0.8668 | 0.3908 | 0.5320 | 0.2316 | 0.8083 | 0.3540 | 0.4416 | 0.2039 |
| | Q | Moyenne | | Moyenne | | Bonne | | Moyenne | | Bonne | |
| Homme7 | E | 0.2271 | 0.1233 | 0.3166 | 0.1626 | 0.1228 | 0.0652 | 0.1835 | 0.0977 | 0.1818 | 0.0944 |
| | Q | Moyenne | | Moyenne | | Bonne | | Bonne | | Bonne | |
| Homme8 | E | 0.1294 | 0.0655 | 0.0868 | 0.0412 | 0.0782 | 0.0396 | 0.0866 | 0.0457 | 0.0608 | 0.0323 |
| | Q | Moyenne | | Moyenne | | Bonne | | Bonne | | Bonne | |
| Homme9 | E | 0.0576 | 0.0288 | 0.0676 | 0.0346 | 0.0458 | 0.0232 | 0.0588 | 0.0299 | 0.0246 | 0.0126 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme10 | E | 0.5075 | 0.2382 | 0.5203 | 0.2505 | 0.5132 | 0.2335 | 0.5870 | 0.2686 | 0.4859 | 0.2312 |
| | Q | Moyenne | | Moyenne | | Moyenne | | Moyenne | | Moyenne | |

(b)

Tableau 3.8 : LMSE (obtenue avec un seul et deux coefficients) pour LF=24.

a) Phrases (1,2,3,4,5) b) Phrases (6,7,8,9,10)

- Les tableaux 3.8.a et 3.8.b résument les résultats obtenus pour $LF=24$. On observe que LMSE a augmenté par rapport aux valeurs précédentes ($LF=16$).

Bien que, LMSE ait changé, la dégradation du signal est perceptible pour les premières valeurs des tableaux (avec un seul coefficient) et non gênante pour les deuxièmes valeurs, pour la plupart des signaux reconstruits.

Il en résulte que, la qualité de reconstruction du signal est commandée par de nombreux facteurs tels que : La longueur de la trame, la longueur de la phrase et le locuteur.

Pour faciliter l'interprétation des résultats donnés par les tableaux précédents, on peut représenter quelques cas sous forme graphiques par les figures suivantes :

1) La figure 3.6 montre la variation de l'erreur de reconstruction de la phrase1 en fonction de :

- * Type de locuteurs (de F1-H20).
- * Longueur de la trame.
- * Nombre de coefficients retenus pour chaque trame.

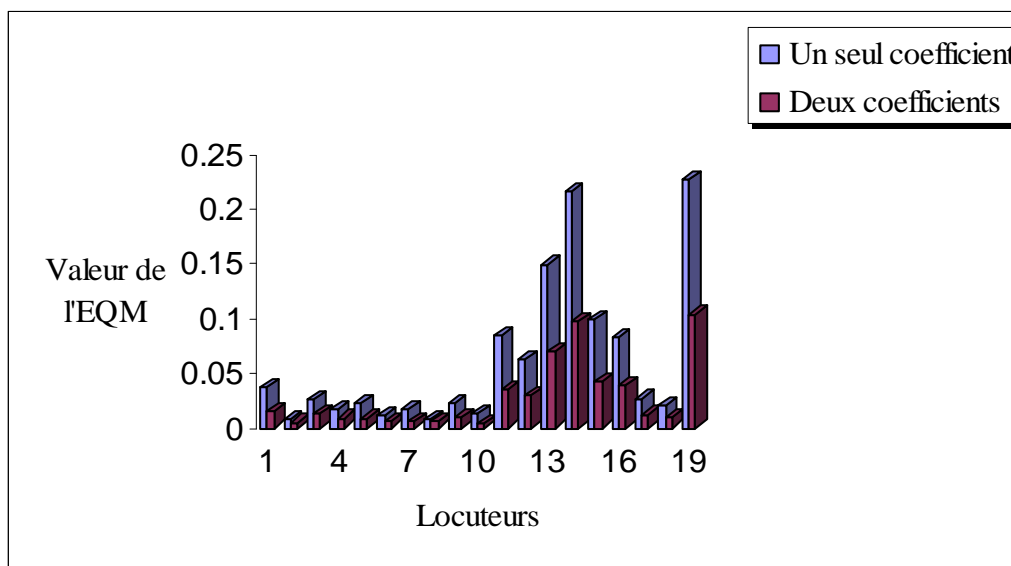
On constate que :

- L'erreur de reconstruction diminue lorsqu'on utilise deux coefficients pour les trames voisées.
- L'erreur varie d'un locuteur à un autre. Elle est importante pour les voix masculines.
- L'erreur augmente avec la taille de la trame.

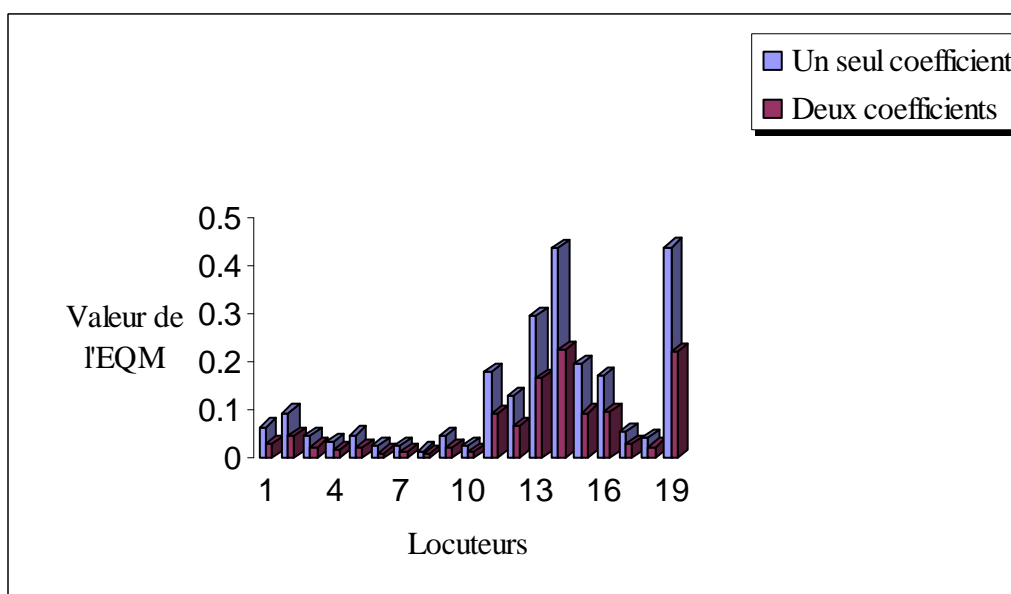
2) La figure 3.7 permet d'illustrer la variation de l'erreur de reconstruction des dix phrases prononcées par un seul locuteur (femme1, homme1 et homme10 respectivement).

On constate que :

- Pour un même locuteur, l'erreur de reconstruction varie d'une phrase à une autre. Par exemple, cette dernière atteint une valeur maximale de 0.2807 pour le locuteur homme 10 (phrase 9). Pour le locuteur femme1 cette erreur est maximale pour les phrases 5 et 6 (vaut 0.1196).
- Dans l'ensemble des cas, l'erreur de reconstruction du locuteur femme1 est faible par rapport aux deux autres locuteurs hommes 1 et 10.
- En utilisant deux coefficients pour approximer les trames voisées, on réduit l'erreur quadratique moyenne entre signal original et reconstruit.



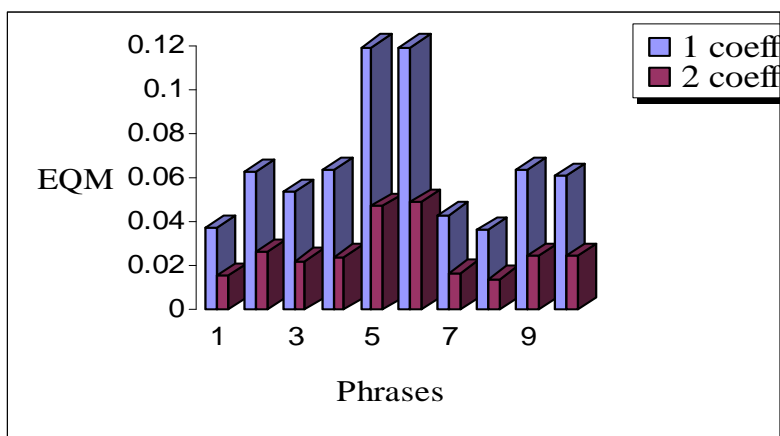
(a)



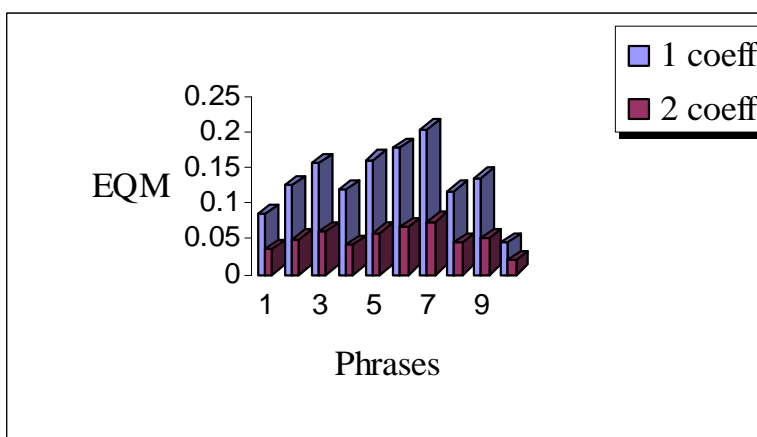
(b)

Figure 3.6 : L'erreur de reconstruction de la phrase 1, pour différents locuteurs avec un coefficient puis deux coefficients pour les trames voisées.

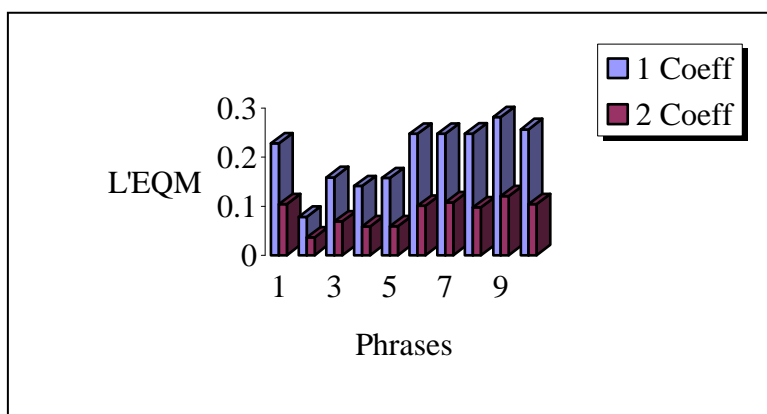
a) LF=16, b) LF=24.



(a)



(b)



(c)

Figure 3.7 : L'erreur de reconstruction des dix phrases avec un seul et deux coefficients respectivement pour les trames voisées . LF=16
 a) Locuteur femme1 b) Locuteur homme1 c) Locuteur homme10

Comme la prise des sons est effectuée dans un local bruyant, le bruit de fond additionné au signal vocal peut dégrader les performances du système de traitement, notamment, en perturbant le fonctionnement de l'algorithme chargé de trouver les fonctions de signatures optimales.

Lors de l'écoute des signaux reconstruits, on a perçu des sons parasites. Pour séparer le bruit du signal reçu par le microphone afin d'en extraire le signal vocal émis par le locuteur, on a appliqué un filtrage de préaccentuation. Cette technique permet de réduire l'amplitude du bruit et atténuer son effet sur le signal de la parole. Ce type de filtre est peu sensible aux petites variations du signal d'entrée mais réagit fortement lorsque celui-ci présente des changements brusques. Ceci équivaut à dire que les signaux basses fréquences (variations lentes) sont peu atténués alors que les signaux hautes fréquences (variations rapides) le sont fortement. Donc, ce filtre est un passe-bas.

Les tableaux 3.9 montrent les résultats obtenus par l'algorithme FS appliqué aux données pré-traitées (enregistrements avec pré-traitement) et les histogrammes qui en résultent sont illustrés sur les figures 3.9.

On remarque que l'opération de préaccentuation a amélioré les résultats en général. On remarque également que cette amélioration est beaucoup plus marquée pour les hommes.

| Phrases | | P1 | | P2 | | P3 | | P4 | | P5 | |
|------------------------|---|------------|--------|------------|--------|------------|--------|------------|--------|------------|--------|
| Nombre de coefficients | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Locuteur Femme 1 | E | 0.0280 | 0.0165 | 0.0285 | 0.0174 | 0.0329 | 0.0172 | 0.0212 | 0.0127 | 0.0257 | 0.0136 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme2 | E | 0.0303 | 0.0188 | 0.0176 | 0.0110 | 0.0370 | 0.0197 | 0.0373 | 0.0210 | 0.0494 | 0.0269 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Bonne | | Bonne | |
| Femme3 | E | 0.0404 | 0.0237 | 0.0140 | 0.0085 | 0.0168 | 0.0087 | 0.0124 | 0.0069 | 0.0313 | 0.0162 |
| | Q | Bonne | | Excellente | | Excellente | | Excellente | | Très bonne | |
| Femme4 | E | 0.0126 | 0.0082 | 0.0091 | 0.0055 | 0.0120 | 0.0077 | 0.0282 | 0.0163 | 0.0316 | 0.0180 |
| | Q | Excellente | | Excellente | | Excellente | | Très bonne | | Très bonne | |
| Femme5 | E | 0.0204 | 0.0107 | 0.0237 | 0.0121 | 0.0118 | 0.0060 | 0.0286 | 0.0149 | 0.0246 | 0.0128 |
| | Q | Très bonne | | Très bonne | | Excellente | | Très bonne | | Très bonne | |
| Femme6 | E | 0.0697 | 0.0338 | 0.0242 | 0.0120 | 0.0294 | 0.0155 | 0.0430 | 0.0192 | 0.0457 | 0.0233 |
| | Q | Bonne | | Très bonne | | Très bonne | | Très bonne | | Bonne | |
| Femme7 | E | 0.0210 | 0.0118 | 0.0369 | 0.0186 | 0.0126 | 0.0068 | 0.0796 | 0.0374 | 0.0621 | 0.0284 |
| | Q | Très bonne | | Bonne | | Excellente | | Bonne | | Bonne | |
| Femme8 | E | 0.0241 | 0.0138 | 0.0145 | 0.0095 | 0.0144 | 0.0087 | 0.0258 | 0.0142 | 0.0260 | 0.0142 |
| | Q | Bonne | | Excellente | | Excellente | | Très bonne | | Très bonne | |
| Femme9 | E | 0.0500 | 0.0273 | 0.0150 | 0.0084 | 0.0626 | 0.0324 | 0.0368 | 0.0202 | 0.0486 | 0.0256 |
| | Q | Bonne | | Excellente | | Bonne | | Bonne | | Bonne | |
| Homme1 | E | 0.0448 | 0.0246 | 0.0422 | 0.0263 | 0.0502 | 0.0278 | 0.0568 | 0.0330 | 0.0600 | 0.0358 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme2 | E | 0.0209 | 0.0130 | 0.0395 | 0.0176 | 0.0572 | 0.0250 | 0.0443 | 0.0218 | 0.0528 | 0.0240 |
| | Q | Très bonne | | Très bonne | | Bonne | | Bonne | | Bonne | |
| Homme3 | E | 0.0099 | 0.0068 | 0.0134 | 0.0085 | 0.0575 | 0.0293 | 0.0493 | 0.0265 | 0.0529 | 0.0298 |
| | Q | Excellente | | Excellente | | Bonne | | Bonne | | Bonne | |
| Homme4 | E | 0.0239 | 0.0152 | 0.0368 | 0.0222 | 0.0488 | 0.0284 | 0.0361 | 0.0217 | 0.0766 | 0.0461 |
| | Q | Très bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme5 | E | 0.0260 | 0.0165 | 0.0434 | 0.0262 | 0.0376 | 0.0218 | 0.0343 | 0.0201 | 0.0498 | 0.0291 |
| | Q | Très bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme6 | E | 0.0355 | 0.0202 | 0.0318 | 0.0186 | 0.0333 | 0.0171 | 0.0258 | 0.0143 | 0.0252 | 0.0142 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Homme7 | E | 0.0107 | 0.0071 | 0.0114 | 0.0070 | 0.0336 | 0.0174 | 0.0231 | 0.0128 | 0.0293 | 0.0159 |
| | Q | Excellente | | Excellente | | Très bonne | | Très bonne | | Très Bonne | |
| Homme8 | E | 0.0165 | 0.0094 | 0.0284 | 0.0162 | 0.0354 | 0.0173 | 0.0479 | 0.0234 | 0.0364 | 0.0178 |
| | Q | Excellente | | Très bonne | | Très Bonne | | Très bonne | | Très bonne | |
| Homme9 | E | 0.0158 | 0.0097 | 0.0242 | 0.0140 | 0.0319 | 0.0142 | 0.0551 | 0.0244 | 0.0594 | 0.0274 |
| | Q | Excellente | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Homme10 | E | 0.0363 | 0.0217 | 0.0156 | 0.0090 | 0.0193 | 0.0105 | 0.0213 | 0.0113 | 0.0212 | 0.0112 |
| | Q | Très bonne | | Excellente | | Très bonne | | Très bonne | | Très bonne | |

(a)

| Phrases | | P 6 | | P 7 | | P 8 | | P9 | | P10 | |
|------------------------|---|------------|--------|------------|--------|------------|--------|------------|--------|------------|--------|
| Nombre de coefficients | | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 |
| Locuteurs Femme1 | E | 0.0483 | 0.0261 | 0.0382 | 0.0199 | 0.0376 | 0.0192 | 0.0191 | 0.0114 | 0.0277 | 0.0153 |
| | Q | Bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme2 | E | 0.0403 | 0.0228 | 0.0552 | 0.0305 | 0.0195 | 0.0118 | 0.0141 | 0.0091 | 0.0183 | 0.0111 |
| | Q | Bonne | | Bonne | | Très bonne | | Excellente | | Très bonne | |
| Femme3 | E | 0.0531 | 0.0275 | 0.0279 | 0.0144 | 0.0149 | 0.0088 | 0.0154 | 0.0088 | 0.0082 | 0.0047 |
| | Q | Bonne | | Très bonne | | Excellente | | Excellente | | Excellente | |
| Femme4 | E | 0.0421 | 0.0242 | 0.0314 | 0.0185 | 0.0120 | 0.0079 | 0.0111 | 0.0071 | 0.0157 | 0.0092 |
| | Q | Bonne | | Très bonne | | Excellente | | Excellente | | Excellente | |
| Femme5 | E | 0.0317 | 0.0165 | 0.0203 | 0.0119 | 0.0078 | 0.0050 | 0.0144 | 0.0082 | 0.0137 | 0.0077 |
| | Q | Très bonne | | Très bonne | | Excellente | | Excellente | | Excellente | |
| Femme6 | E | 0.0572 | 0.0280 | 0.0405 | 0.0223 | 0.0295 | 0.0166 | 0.0742 | 0.0373 | 0.0383 | 0.0178 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Très bonne | |
| Femme7 | E | 0.0593 | 0.0245 | 0.0357 | 0.0191 | 0.0335 | 0.0171 | 0.0699 | 0.0250 | 0.0271 | 0.0127 |
| | Q | Bonne | | Bonne | | Très bonne | | Bonne | | Très bonne | |
| Femme8 | E | 0.0348 | 0.0192 | 0.0261 | 0.0138 | 0.0303 | 0.0177 | 0.0283 | 0.0159 | 0.0285 | 0.0153 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Très bonne | | Très bonne | |
| Femme9 | E | 0.0587 | 0.0318 | 0.0618 | 0.0315 | 0.0303 | 0.0174 | 0.0469 | 0.0256 | 0.0229 | 0.0136 |
| | Q | Bonne | | Bonne | | Très bonne | | Bonne | | Très bonne | |
| Homme1 | E | 0.0671 | 0.0405 | 0.0637 | 0.0390 | 0.0403 | 0.0241 | 0.0614 | 0.0361 | 0.0288 | 0.0176 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Très bonne | |
| Homme2 | E | 0.0325 | 0.0168 | 0.0404 | 0.0209 | 0.0141 | 0.0083 | 0.0419 | 0.0203 | 0.0250 | 0.0129 |
| | Q | Très bonne | | Bonne | | Excellente | | Bonne | | Très bonne | |
| Homme3 | E | 0.0498 | 0.0296 | 0.0549 | 0.0309 | 0.0243 | 0.0149 | 0.0333 | 0.0188 | 0.0253 | 0.0152 |
| | Q | Bonne | | Bonne | | Très bonne | | Très bonne | | Très bonne | |
| Homme4 | E | 0.0882 | 0.0536 | 0.0661 | 0.0397 | 0.0504 | 0.0300 | 0.0461 | 0.0284 | 0.0321 | 0.0179 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Très bonne | |
| Homme5 | E | 0.0568 | 0.0340 | 0.0544 | 0.0321 | 0.0408 | 0.0246 | 0.0379 | 0.0234 | 0.0353 | 0.0211 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme6 | E | 0.0568 | 0.0340 | 0.0544 | 0.0321 | 0.0408 | 0.0246 | 0.0379 | 0.0234 | 0.0353 | 0.0211 |
| | Q | Bonne | | Bonne | | Bonne | | Bonne | | Bonne | |
| Homme7 | E | 0.0248 | 0.0135 | 0.0233 | 0.0140 | 0.0103 | 0.0062 | 0.0229 | 0.0133 | 0.0117 | 0.0067 |
| | Q | Très bonne | | Très bonne | | Excellente | | Très bonne | | Excellente | |
| Homme8 | E | 0.0268 | 0.0138 | 0.0289 | 0.0152 | 0.0293 | 0.0160 | 0.0439 | 0.0218 | 0.0139 | 0.0072 |
| | Q | Très bonne | | Très bonne | | Très bonne | | Bonne | | Excellente | |
| Homme9 | E | 0.0622 | 0.0305 | 0.0391 | 0.0186 | 0.0403 | 0.0199 | 0.0466 | 0.0199 | 0.0180 | 0.0089 |
| | Q | Bonne | | Très bonne | | Très bonne | | Très bonne | | Excellente | |
| Homme10 | E | 0.0309 | 0.0177 | 0.0529 | 0.0288 | 0.0353 | 0.0178 | 0.0425 | 0.0242 | 0.0249 | 0.0138 |
| | Q | Très bonne | | Bonne | | Très bonne | | Bonne | | Très bonne | |

(b)

Tableau 3.9 : LMSE (obtenue avec un seul et deux coefficients) avec prétraitement
 LF=16. a) Phrases (1,2,3,4,5) b) Phrases (6,7,8,9,10)

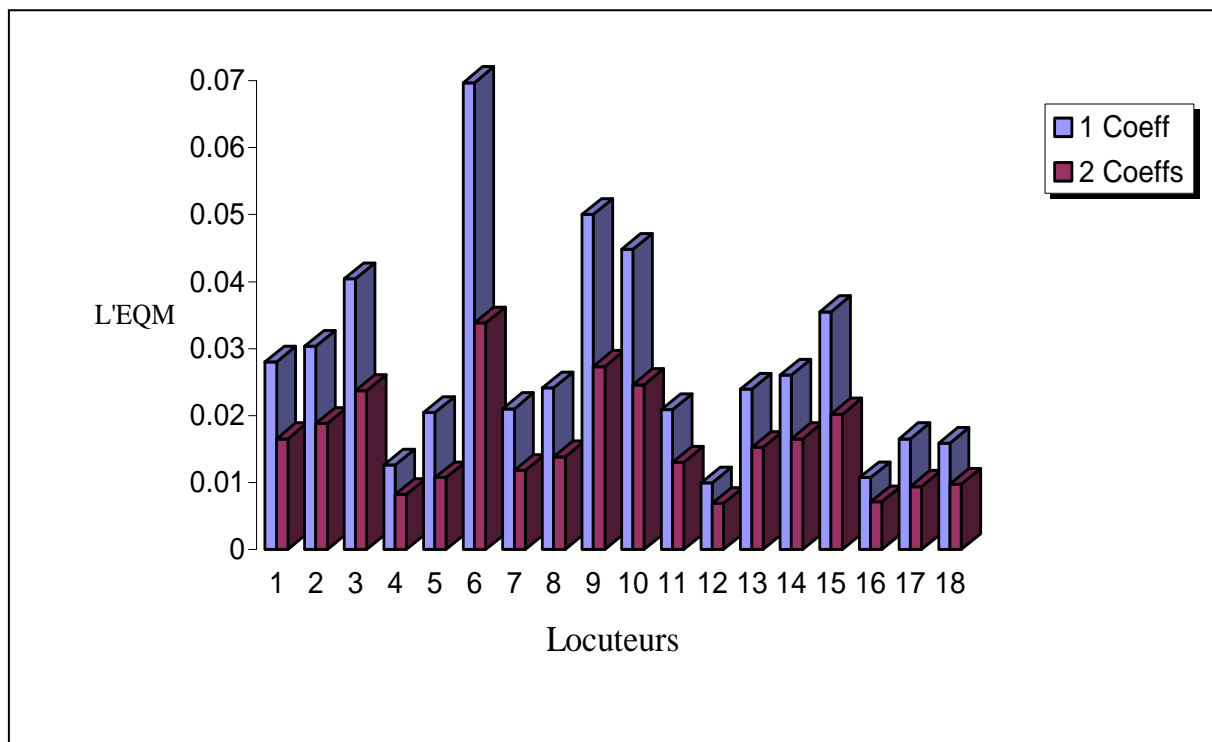
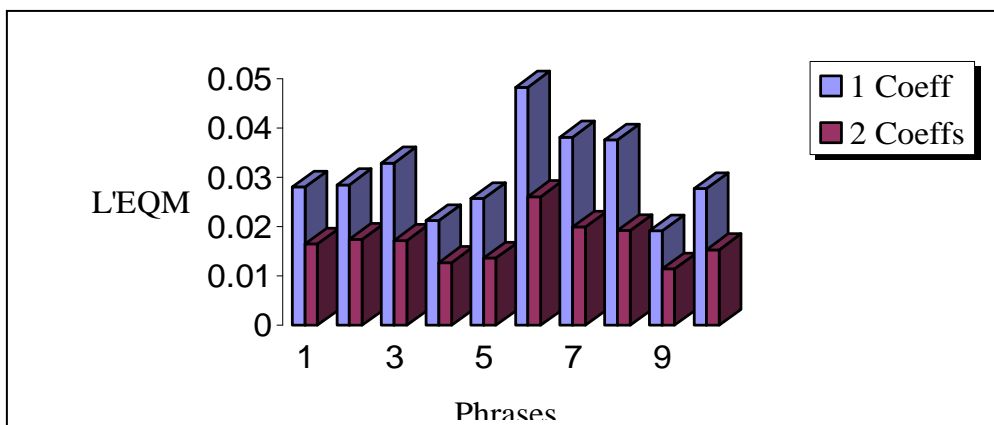
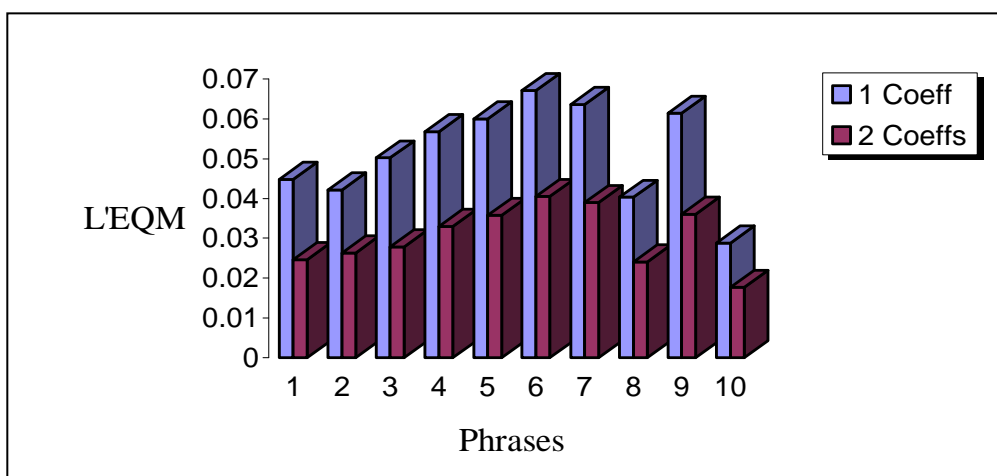


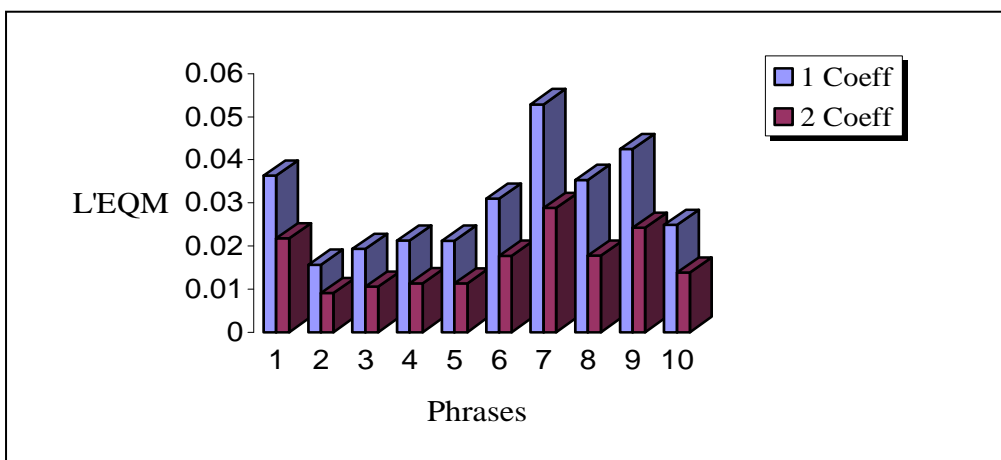
Figure 3.8 : L'erreur de reconstruction de la phrase 1, avec prétraitement. LF=16. pour différents locuteurs avec un coefficient puis deux coefficients pour les trames voisées.



(a)



(b)



(c)

Figure 3.9 : Evaluation de l'EQM de reconstruction des dix phrases, LF=16

a) Locuteur F1, b) Locuteur H1, c) Locuteur H10

3.6. Résultat comparatif.

Ceci constitue une deuxième étape de présentation pratique de la fonction de signature comme algorithme de compression et de synthèse de la parole.

Pour donner plus de signification à ce que nous venons d'énoncer sur la technique proposée, il convient de la comparer à une autre technique dans la même application. Par exemple, on la compare à la LPC-10 E(FS 1015).

3.6.1. Technique proposée :

Pour $LF=16$, le débit de transmission est $64 \text{ Kb}/10.66 = 6 \text{ Kb/s}$ et la moyenne MOS est de 3.20 (tableau 3.7).

Pour $LF=24$, le débit de transmission est $64 \text{ Kb}/16=4 \text{ Kb/s}$ et la moyenne du MOS calculée à partir du tableau 3.8 est 2.76. De la même manière, pour $LF=40$, le débit de transmission est $64 \text{ Kb}/26.6=2.4 \text{ Kb/s}$ avec une moyenne du MOS de 2.5 [1]

3.6.2. Technique LPC-10 E (FS 1015) :

Le débit de cette technique est 2.4 Kb/s et sa cotation MOS est de 2.3 [1,17].

Ces résultats sont d'une grande importance car la qualité d'audition, donnée par la méthode étudiée, est meilleure que celle de la LPC-10 E (FS 1015) dans le cas de $LF=24$. Les résultats de la FS et la LPC-10 E (FS 1015) sont cependant similaires dans le cas de $LF=40$, malgré une petite différence entre les deux approches en ce qui concerne l'évaluation MOS. Les résultats de comparaison sont résumés dans le tableau 3.10.

| Technique | | Débit (Kbits/s) | | Moyenne du MOS | |
|-----------------------|-------------------|-------------------|---------------------|----------------|----------|
| Algorithme FS | Longueur de trame | Sans recouvrement | Recouvrement de 50% | SR | R de 50% |
| | LF=16 | 6 | 12 | 3.20 | 3.40 |
| | LF=24 | 4 | 8 | 2.76 | 2.86 |
| | LF=40 | 2.4 | 4.8 | 2.50 | 2.60 |
| LPC-10E (FS 1015)[17] | | 2.4 | | 2.3 | |

Tableau 3.10 : Comparaison entre la technique proposée et la LPC-10 E (FS 1015)

On remarque que, pour le même débit (2.4 Kb/s), on obtient un meilleur MOS avec la technique FS. Dans le cas de recouvrement, on constate que le débit de l'algorithme FS augmente tout en améliorant la qualité d'écoute du signal de parole reconstruite.

3.7. Conclusion

Dans ce chapitre on a décrit, premièrement des simulations avec la KLT classique. Ensuite, on a appliqué l'algorithme FS à des corpus de paroles différents afin d'obtenir les fonctions de signature optimales. Les essais des différents algorithmes ont été réalisés de manière séparée.

Les mesures de la qualité de reconstruction, utilisées dans ce travail considèrent le système auditif humain comme étant l'ultime évaluateur de la qualité et la performance d'un tel codeur. La méthode objective d'évaluation de la qualité qui a été utilisée est l'erreur quadratique moyenne (LMSE).

On a aussi montré que, La FS a une bonne performance comparée aux techniques traditionnelles en termes de qualité de reconstruction et de taux de compression [1] [7]. Par exemple, la FS a été comparée à la (LPC-10E).

On peut noter que, la dégradation de la parole liée à la compression par cette technique peut être essentiellement différente à celle imposée par d'autres méthodes de compression utilisant des bases fixes telles que (DCT, DFT, etc) [5].

L'approche que nous avons exploitée est basée sur l'utilisation de la transformée KL. En pratique seul un nombre fini de termes est réellement utilisé. Autrement dit, la FS est obtenue en tronquant la séquence d'origine. Cette troncation a bien évidemment des effets néfastes sur la reconstruction. Elle introduit notamment une distorsion du signal original et surtout pour les signaux voisés qui sont caractérisés par une grande énergie (variance).

Pour les différentes expériences menées, les fichiers sons à traiter (paroles arabe), sont extraits des fichiers sons (MSA_DARS.._NAS) ou obtenus par des enregistrements personnels. Les enregistrements et les modifications des sons ont été exécutés à l'aide des logiciels GoldWave et Praat.

Tous les calculs ont été exécutés avec Matlab version 7 par un PC 4 sous windows XP.

CONCLUSION GENERALE ET PERSPECTIVES

La nécessité de comprimer des données volumineuses et difficiles à transmettre ou à stocker est accrue avec le développement des techniques de communications. Les performances d'un système de compression sont évaluées par sa complexité calculatoire, son coût de stockage, le débit (ou taux de compression) qu'il permet d'atteindre et la dégradation qu'il engendre sur les données compressées. Les techniques de compression ont fait l'objet de nombreuses recherches pour tenter d'optimiser ces différents aspects. Il demeure cependant difficile d'optimiser tous ces critères en même temps.

Il n'a pas été aisé de traiter notre sujet sans passer par la transformée de Karhunen-Loève (KLT) qui représente le fondement théorique sur lequel repose la méthode de compression étudiée. Dans le premier chapitre, on a présenté un aperçu théorique sur la KLT. Les principales définitions, notions et propriétés de cette transformation ont été évoquées, notamment la décomposition en vecteurs et valeurs propres. On a étudié son utilisation dans la compression du signal de parole c'est-à-dire qu'on a montré quelles sont les possibilités offertes pour extraire de l'information pertinente de ce signal. Nous avons vu quelles sont les potentialités offertes par la KLT en vue de la compression de la parole, notamment comment cette transformation permettrait une décorrélation de données. La difficulté principale est de savoir quelles sont les informations les plus importantes et comment les modéliser. Puisqu'il n'y a pas d'algorithme rapide dans le cas général, cette transformée ne peut être appliquée en pratique telle quelle était définie.

Ensuite, on a approximé cette méthode en utilisant la FS « fonctions de signature ». On a présenté un algorithme FS qui permet de trouver une base orthogonale constituée d'un nombre réduit de vecteurs propres décorrélés. On a montré qu'à l'aide de cet ensemble de vecteurs, on peut synthétiser, trame par trame, tout signal de parole.

On a analysé l'influence du nombre de vecteurs propres retenus, pour chaque trame, sur la qualité de parole synthétisée ainsi que l'influence du nombre de fonctions de signatures sur

cette qualité. On a remarqué que dans les parties voisées, on obtient une erreur plus grande. Pour remédier à cela, on a proposé un système hybride dans lequel, dans les parties voisées, on a utilisé les deux plus importants vecteurs propres.

Un premier essai a consisté à appliquer directement la KLT traditionnelle pour comprimer le signal de parole. Cependant, les résultats de compression sont loin d'être satisfaisant avec un seul coefficient par trame. En effet pour $LF=16$, un taux de compression de $16/3=5.33$ est atteint, en retenant 3 coefficients pour chaque trame.

Les expériences menées avec l'algorithme FS tel que les trames voisées ont été modélisées par un modèle hybride, ont donné de meilleurs résultats. Les premiers tests de reconstruction ne sont pas effectués sur tout l'ensemble de données, composé de dix phrases, mais on a reconstruit trois phrases seulement, pour deux locuteurs : un homme et une femme.

Afin d'étudier la robustesse de la méthode proposée, nous avons appliqué l'algorithme mis au point sur un corpus de parole plus important (10 phrases prononcées par 20 locuteurs ce qui fait 200 phrases au total). Le corpus considéré a permis d'évaluer les performances du système FS hybride. Le taux de compression et la qualité de reconstruction ont augmenté fortement. Nous supposons que ces résultats sont dus principalement à l'existence d'information pertinente suffisante pour déterminer la « fonction de signature optimale » qui permet une synthèse efficace de tout signal de la parole.

Afin de tenir compte de l'influence de la longueur de trame, nous avons évalué la qualité de parole reconstruite, par la mesure de LMSE et du MOS, pour différentes valeurs de LF.

Une deuxième étape d'expériences a consisté à comparer l'algorithme FS à une technique connue telle que la LPC-10. Les résultats obtenus montrent qu'on a un signal reconstruit de bonne qualité pour $LF=16, 24$ alors qu'il est intelligible pour $LF=40$ avec des scores MOS de 2.76 et 2.5 respectivement. D'après les résultats obtenus, la représentation du signal de la parole par la méthode proposée donne une meilleure qualité que la LPC pour $LF=16, 24$. Dans le cas de $LF=40$, les deux techniques donnent la même qualité (parole intelligible).

Donc les résultats obtenus en comparaison avec un codeur connu (FS1015 LPC 10E), sont satisfaisants, en donnant un meilleur MOS pour un même débit.

Dans les futurs travaux de recherche, il peut être utile d'étudier l'ordre séquentiel et la fréquence d'utilisation des séquences de signature dans les processus de reconstruction qui peuvent mener à la reconnaissance de parole et à l'identification du locuteur en employant la méthode proposée de modélisation de parole.

Comme autre perspective, on peut mentionner l'utilisation de l'ICA (Independent Component Analysis) qui est considérée comme une généralisation de la PCA (Principal Component Analysis).

APPENDICE A

LISTE DES SYMBOLES ET DES ABREVIATIONS

| | |
|-------------------|---|
| CCITT | : International Consultative Committee for Telephone and Telegraph. |
| Cor , Cov | : Correlation, Covariance. |
| C_{kr} | : Coefficient associé aux fonctions de signature. |
| CR | : Compression Rate (taux de compression). |
| dB | : Décibel. |
| DCT | : Discrete Cosine Transform. |
| DFT | : Discrete Fourier Transform. |
| DST | : Discrete Sine Transform. |
| DWHT | : Discrete Welch Hadamard Transform. |
| δ_{kr} | : Erreur quadratique de la trame k. |
| E , ε | : Energie, Erreur d'approximation. |
| ETSI | : European Telecommunications Standard Institute. |
| FS | : Fonctions de signature. |
| 3 G | : Troisième generation. |
| GSM | : Global System for Mobile. |
| http | : Hyper Text Transport Protocol. |
| ISO | : International Standardization Organisation. |
| Kbps, Kb/s | : Kilo bits par seconde. |
| KLT (TKL) | : Karhunen-Loève Transform. |
| LD-CELP | : Low Delay Code Excited Linear Predictive Coder. |
| LF | : Longueur de la trame. |
| LMSE (EQM) | : Least Mean Square Error (Erreur Quadratique Moyenne). |

| | |
|-----------------|---|
| LPC | : Linear Predictive Coding. |
| \mathcal{L} | : Le lagrangien. |
| MIC (PCM) | : Codage par Modulation d'Impulsion(Pulse Coding Modulation). |
| MICDA | : Codage MIC Différentiel Adaptatif (ADPCM). |
| MLD | : Principale Longueur Minimale de Description |
| MOS | : Mean Opinion Score. |
| N_b | : Nombre de bits |
| N_{FT} | : Nombre de trame. |
| PCA | : Principal Component Analysis. |
| Σ, S | : Matrice de covariance, Matrice de corrélation. |
| R_{FK} | : Matrice de corrélation de la trame X_{FK} . |
| RPELTP | : Regular Pulse Excitation-Long Term Prediction. |
| S_{EG} | : Ensemble « augmented signature ». |
| SNR | : Signal Noise Ratio (rapport signal à bruit). |
| SR | : Ensemble de signature. |
| T, TO et TOI | : Transformations orthogonales directe et inverse. |
| TDMA | : Accès Multiples par Division du Temps. |
| UIT-T | : International Telecommunications Union-Terminal. |
| UMTS | : Universal Mobile Telecommunications System. |
| V, λ | : vecteur propre, valeur propre ,. |
| Var, σ^2 | : Variance, écart-type |
| Vocoder | : Voice Coder (codeur de voix) |
| (.wav) | : Format wave. |
| WAP | : Wirless Application Protocol. |
| WWW | : World Wide Web. |

APPENDICE B [10]

ANALYSE EN COMPOSANTES PRINCIPALES (ACP OU PCA)

B.1.Introduction

La PCA est un outil permettant une meilleure visualisation de nos données. Il s'agit de mettre un peu d'ordre dans le fichier de données, souvent de taille considérable (quelques centaines ou milliers d'analyses et quelques dizaines de variables). Cette méthode repose sur l'analyse de la structure de la matrice de variance-covariance (variabilité et dispersion des données). Si l'une des variables peut s'exprimer comme une fonction des autres, on a besoin des p variables pour prendre en compte toute la variabilité du système.

La PCA permet de décrire à l'aide de q ($< p$) composantes un maximum de cette variabilité. Ce qui permet :

- une réduction des données à q nouveaux descripteurs.
- une visualisation des données à 2 ou 3 dimensions ($q=2$ ou 3).
- une interprétation des données : liaisons inter-variables.

La PCA peut être considérée comme une étape intermédiaire souvent utilisée avant d'autres analyses.

B.2.Recherche des composantes principales

Soient $C_1, C_2, \dots, C_k, \dots, C_q$, les composantes principales à déterminer.

C_k est une nouvelle variable qui s'exprime comme une combinaison linéaire des variables d'origine X_1, X_2, \dots, X_p telle que:

$C_k = a_{1k}X_1 + a_{2k}X_2 + \dots + a_{pk}X_p \Rightarrow$ les coefficients a_{jk} à déterminer tel que les C_k soient:

- 2 à 2 non corrélées.
- de variance maximale.
- d'importance décroissante.

C_1 = première composante principale doit être de variance maximale.

Géométriquement, la PCA consiste à projeter les points sur une droite, un plan, un sous-espace à q dimensions c'est-à-dire chercher le sous-espace donnant la meilleure visualisation possible de notre nuage de points.

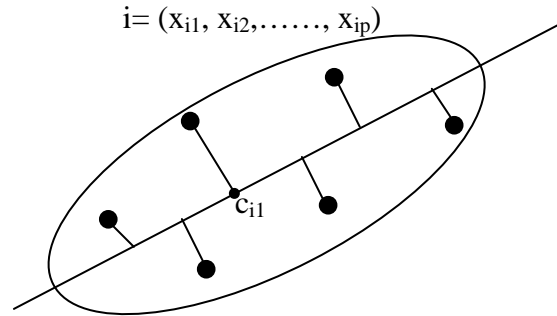


Figure B.1 : Représentation d'un nuage de points à l'aide de la PCA.

c_{i1} = coordonnée du point i sur l'axe $C_1 \Rightarrow$ Projection de x_i sur $C_1 : c_{i1} = \sum_{j=1}^p a_{1j} x_{ij}$

C_1 détermine une nouvelle direction dans le nuage de points qui suit l'axe d'allongement (étirement) maximal du nuage. C_1 de variance maximale \Rightarrow les projections c_{i1} sont les plus dispersées possible. Pour fixer la droite, on impose qu'elle passe par le centre de gravité g . Elle réalise le meilleur ajustement possible du nuage c'est-à-dire qui conserve au mieux la distance entre les points (après projection) \Rightarrow droite de projection assurant une distorsion minimale.

$C_2 = 2^{\text{ème}}$ composante, orthogonale à C_1 et de variance maximale.

C_1 et C_2 déterminent le plan principal : le meilleur plan de projection (de distorsion minimale).

C_3 est la droite orthogonale à C_1 et C_2 passant par g telle que la variance des coordonnées soit maximum.

B.3. Décomposition de la variance

• La variance se décompose de la manière suivante :

$$\begin{aligned} \sigma^2 &= \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^T (x_i - x_j) \\ &= \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n ((x_i - g) + (g - x_j))^T ((x_i - g) + (g - x_j)) \\ &= \frac{1}{2n(n-1)} \sum_{i=1}^n \sum_{j=1}^n ((x_i - g)^T (x_i - g) + (g - x_j)^T (g - x_j) + 2(x_i - g)^T (g - x_j)) \\ &= \frac{1}{(n-1)} \sum_{i=1}^n (x_i - g)^T (x_i - g) \end{aligned}$$

B.4. Projection sur une droite

• L'opérateur de projection orthogonale, π , sur une droite de vecteur directeur unitaire V s'écrit :

$$\pi = VV^T \quad \text{Avec} \quad V^T V = 1$$

• La variance des observations projetées s'écrit alors

$$\sigma_V^2 = \frac{1}{(n-1)} \sum_{i=1}^n (\pi(x_i - g))^T (\pi(x_i - g))$$

B.5. Recherche de la projection de variance maximale

Nous avons donc :

$$\begin{aligned} \sigma_V^2 &= \frac{1}{(n-1)} \sum_{i=1}^n (\pi(x_i - g))^T (\pi(x_i - g)) \\ &= \frac{1}{(n-1)} \sum_{i=1}^n (VV^T(x_i - g))^T (VV^T(x_i - g)) \\ &= \frac{1}{(n-1)} \sum_{i=1}^n (x_i - g)^T (VV^T VV^T)(x_i - g) \\ &= \frac{1}{(n-1)} \sum_{i=1}^n (x_i - g)^T VV^T(x_i - g) \end{aligned}$$

Et donc

$$\begin{aligned} \sigma_V^2 &= \frac{1}{(n-1)} \sum_{i=1}^n ((x_i - g)^T V) (V^T(x_i - g)) \\ &= \frac{1}{(n-1)} \sum_{i=1}^n V^T(x_i - g)(x_i - g)^T V \\ &= \frac{1}{(n-1)} V^T \left[\sum_{i=1}^n (x_i - g)(x_i - g)^T \right] V \\ &= V^T \Sigma V \end{aligned}$$

Observons que la matrice Σ est la matrice variance-covariance. Cette matrice est symétrique définie positive. Nous devons donc maximiser cette variance des observations projetées :

$$\max_V (V^T \Sigma V) \quad \text{subject to} \quad (V^T V = 1)$$

Il s'agit d'un problème d'optimisation sous contrainte. Nous formons donc la fonction de Lagrange suivante :

$$\mathfrak{S} = V^T \Sigma V + \lambda(1 - V^T V)$$

Et nous calculons les conditions nécessaires d'optimalité

$$\partial_v \mathcal{J} = 0$$

Nous obtenons ainsi l'équation aux valeurs propres suivante :

$$\sum V = \lambda V$$

Comme la matrice variance-covariance est symétrique définie positive, les valeurs propres sont réelles positives. Les vecteurs propres peuvent être choisis orthonormés

• La variance des observations projetées s'écrit alors :

$$\begin{aligned} \sigma_v^2 &= V^T \Sigma V \\ &= V^T \lambda V \\ &= \lambda \end{aligned}$$

Et donc la solution est de projeter les données sur le vecteur propre ayant la valeur propre λ la plus élevée.

B.6. Recherche des projections de variance maximale orthogonales au premier axe

Afin de trouver le second axe de variance maximale, nous recherchons :

$$\max_v (V^T \Sigma V) \text{ subject to } (V^T V = 1) \text{ and } (V^T V_1 = 0)$$

Avec V_1 étant le premier vecteur propre à valeur propre maximale. Comme les vecteurs propres de Σ sont naturellement orthonormés, la solution est de choisir le deuxième vecteur propre de Σ (à deuxième valeur propre maximale)

B.7. Matrice variance-covariance

Notons que si X est la matrice de données, qui contient les vecteurs $(x_i - g)^T$ en ligne.

La matrice de variance-covariance est $\Sigma = (n - 1)^{-1} X^T X$.

B.8. Interprétation des valeurs propres

La somme des valeurs propres correspond à la variance totale :

$$\text{tr}(\Sigma) = \sigma^2 = \sum_{i=1}^n \lambda_i$$

Chaque valeur propre mesure la part de variance expliquée par l'axe factoriel correspondant.

B.9. APPROCHE ALTERNATIVE DE LA PCA

Nous présentons une approche alternative de la PCA. En utilisant cette fois-ci la notion de vecteur aléatoire. Et donc en ne partant pas des données empiriques comme précédemment.

Soit $x = [x_1, x_2, \dots, x_n]$ le vecteur aléatoire des n variables aléatoires (caractéristiques) mesurées sur les individus. Nous définissons une nouvelle variable y qui est une combinaison linéaire des variables aléatoires x_i :

$$y = V^T x$$

Nous supposons que V est normalisé $V^T V = 1$. Nous recherchons la projection du vecteur aléatoire x qui maximise la variance projetée:

$$\text{var}(y) = E[(y - E[y])^2]$$

Calculons d'abord la moyenne de y :

$$\begin{aligned} E(y) &= E[V^T x] \\ &= V^T E[x] \\ &= V^T g \end{aligned}$$

Et ensuite la variance:

$$\begin{aligned} \text{var}(y) &= E[(y - E[y])^2] \\ &= E[(V^T x - V^T g)^2] \\ &= E[V^T (x - g) V^T (x - g)] \\ &= V^T E[(x - g)(x - g)^T] V \end{aligned}$$

Il faut donc calculer le maximum de cette variance par rapport à V , ce qui nous ramène au problème d'optimisation suivant (le même que pour l'approche précédente)

$$\max_V (V^T S V) \text{ subject to } (V^T V = 1)$$

Avec S étant la matrice variance-covariance (notons que Σ était la matrice variance-covariance empirique):

$$S = E[(x - g)(x - g)^T].$$

Et S peut être estimé à partir de l'échantillon par :

$$\begin{aligned} S &\cong \frac{1}{(n-1)} \sum_{i=1}^n (x_i - g)(x_i - g)^T \\ &\cong \Sigma \end{aligned}$$

Nous sommes donc amenés, comme précédemment, à calculer les valeurs/vecteurs propres de Σ .

Résultats:

L'ACP remplace les p variables de départ (variances différentes, corrélation inter-variable) en q nouvelles composantes C_k ($q \leq p$).

- Orthogonales 2 à 2 c-à-d $\text{cov}(C_k, C_{k'}) = 0$ (pour tout $k \neq k'$), et de variances maximales

On peut noter que :

- $V(C_1) \geq V(C_2) \dots \geq V(C_q)$ d'importance décroissante
- Le nombre maximum de composantes principales $q \leq p$ avec $q < p$ dès que l'une des variables d'origine est une combinaison linéaire d'autres!

⇒ Mise en évidence de relations linéaires dans les données.

⇒ Les données occupent, en réalité, un sous-espace de dimensions réduites ($q < p$)

Le nombre maximum de composantes principales = dimension intrinsèque des données

- Choix des r premières composantes principales

$r \ll p \Rightarrow$ réduction de la dimension.

Objectif : garder un maximum d'information des données initiales.

Mesure de cette information : le % de variance expliquée est :
$$= \frac{\sum_{k=1}^r V(C_k)}{\text{Inertie totale}}$$

Si les variables originales sont fortement corrélées entre elles, un nombre réduit de composantes permet d'expliquer 80% à 90% de variance !

Géométriquement : revient à projeter les données dans un sous-espace de dimension r , centré sur g , reprenant les r premiers axes principaux d'allongement du nuage ! \Rightarrow Les projections c_{ij} sont les plus dispersées possibles!

En général :

- Le % de variance expliquée par C_1, C_2, \dots, C_r = mesure d'aplatissement du nuage sur le sous-espace des composantes (à r dimensions). Plus ce % est grand, meilleure est la représentation des données dans le sous-espace !
- Les composantes principales sont entièrement déterminées par la matrice V variance-covariance (vecteurs propres).

⇒ Toute modification de $V \Rightarrow$ modification des composantes !

Remarques :

- Si certaines variables initiales sont très dispersées (σ_j^2), elles vont prendre le pas sur les autres.

⇒ Les composantes principales tenteront essentiellement d'expliquer la variance due à ces variables !

⇒ On peut travailler en données réduites (variables normalisées par s_j)

⇒ Toutes les variables auront la même importance: Il se peut qu'on perde de l'information

$$z_{ij} = \frac{(x_{ij} - g_j)}{\sigma_j}$$

⇒ Données centrées réduites

⇒ Matrice variance covariance = R et l'ACP explique la structure de R !

• Autre possibilité : travailler sur les rangs

⇒ ACP non-paramétrique

⇒ Plus robuste : - pour des données très hétérogènes

- aux dissymétries des distributions

- aux valeurs extrêmes ! (Augmente anormalement la variance !)

⇒ Permet d'intégrer des variables qualitatives ordinales !

• Contributions des variables aux composantes

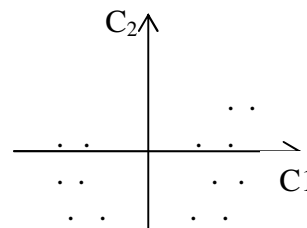
Composante = combinaison linéaire des variables : $C_k = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p$

Coeff. a_{jk} = contribution de la variable X_j à la composante C_k

• Interprétation des résultats

1/ Représentation des individus dans le plan principal

⇒ Peut faire apparaître des groupes d'individus présentant des similitudes. Proximités abusives dues aux projections.



⇒ La représentation n'est valable que si le % de variance expliquée par C_1 et C_2 est suffisamment grand ! (Nuage assez aplati sur le plan)

⇒ Vérifier si les proximités se maintiennent dans d'autres plans de projection:

$$C_1 - C_3, C_2 - C_3, \dots$$

Les individus les mieux représentés: points proches du plan (projection peu importante).

2/ Interprétation des composantes principales

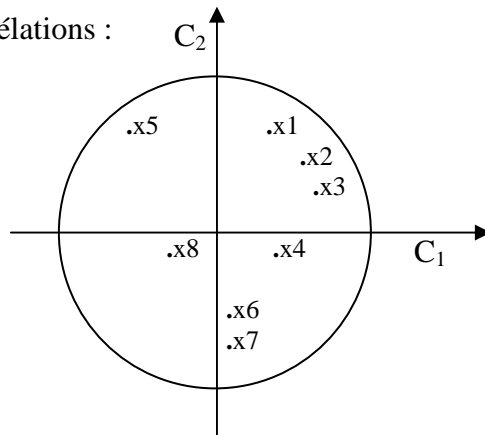
Corrélations avec les variables initiales :

Repérer les variables très corrélées ($r \approx 1$ ou $r \approx -1$)

| | C_1 | C_2 | C_3 |
|-------|----------|----------|----------|
| X_1 | r_{11} | r_{12} | r_{13} |
| X_2 | r_{21} | r_{22} | r_{23} |
| ⋮ | ⋮ | ⋮ | ⋮ |
| X_p | r_{p1} | r_{p2} | r_{p3} |

Interprétation des 2 premières composantes C_1, C_2 :

Cercle des corrélations :



C_1 et C_2 étant non-corrélées, on a $r^2(c_1, x_j) + r^2(c_2, x_j) \leq 1$

\Rightarrow Chaque variable représentée par les coordonnées : $(r(c_1, x_j), r(c_2, x_j))$ est dans un cercle de rayon 1 \Rightarrow groupes de variables liées ou opposées

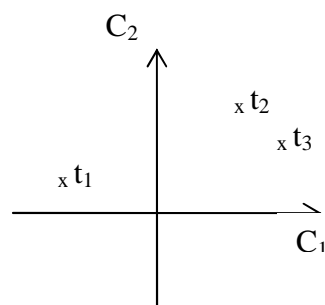
Si proches de la circonférence \Rightarrow Bien représentées par les 2 composantes !

3/ Projection de points supplémentaires sur le plan principal (Après le calcul des composantes)

- individus typiques de groupes d'individus : exemple t_1, t_2, t_3 pour 3 groupes différents.

\Rightarrow Permet de voir si les composantes C_1 et C_2 sont à même de les distinguer

\Rightarrow Idée de ce qui est distinguable (t_1 des autres) ou non (t_2, t_3)



APPENDICE C

MESURE DE PERFORMANCE OBJECTIVE ET SUBJECTIVE

C.1. Gammes de qualité de la parole

Dans les communications numériques, la qualité de paroles est classée selon quatre catégories générales, nommées :

- Qualité de diffusion : qui fait référence à la bande élargie (typiquement 50-7000Hz, mais les 20-20.000Hz pour le compact disque). La parole de haute qualité peut être, généralement, atteinte à des taux d'au moins de 32-64 Kbps.
- Qualité réseau ou ligne : décrit la parole par rapport à la manière dont elle est écoutée, à travers le réseau téléphonique en marche (c'est approximativement la gamme 200-3200Hz). Avec un rapport signal sur bruit plus grand que 30 dB et une distorsion harmonique moins de 2-3%
- Qualité communication : implique la qualité de la parole dégradée qui est plus au moins naturelle et fortement intelligible. La parole de communication peut être atteinte à des taux au dessus de 4 Kb/s.
- Parole synthétique : est habituellement intelligible mais elle peut être non naturelle et associée avec perte de reconnaissance du locuteur.

Le but général dans le codage de la parole est d'atteindre la qualité réseau à 4.0Kb/s. Couramment, les codeurs de la parole qui opèrent bien en dessous de 4.0 Kb/s, tendent à produire de la parole synthétique [14,18].

C.2. SNR

Le SNR ou le SNR global, est la mesure la plus utilisée à cause de la simplicité de calcul.

Soient les signaux suivants :

- $y(n)$: signal de parole bruité.
- $x(n)$: signal de parole propre (sans bruit).
- $d(n)$: signal du bruit.

En notant $\hat{x}(n)$ le signal de parole estimé, le signal erreur $\varepsilon(n)$ s'écrit alors :

$$\varepsilon(n) = x(n) - \hat{x}(n) \tag{C.1}$$

L'énergie du signal erreur peut être calculée par :

$$E_{\epsilon} = \sum_n \epsilon^2(n) = \sum_n [x(n) - \hat{x}(n)]^2 \quad (C.2)$$

Et l'énergie du signal est exprimée par :

$$E_x = \sum_n x^2(n) \quad (C.3)$$

Il en résulte que la valeur du SNR (dB) est donnée par :

$$\text{SNR} = 10 \log_{10} \frac{E_x}{E_{\epsilon}} = 10 \log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - \hat{x}(n)]^2} \quad (C.4)$$

C.3. SNR segmental (SNR_{seg})

Le SNR_{seg} représente la variation du SNR, il est calculé par :

$$\text{SNR}_{\text{seg}} = \frac{1}{M} \sum_{j=0}^{M-1} 10 \log_{10} \frac{\sum_{n=m_j-N+1}^{m_j} x^2(n)}{\sum_{n=m_j-N+1}^{m_j} [x(n) - \hat{x}(n)]^2} \quad (C.5)$$

Où m_0, m_1, \dots, m_{M-1} , sont les instants de fin des trames M , ayant chacune une longueur N . On calcule le SNR pour chaque trame, et la valeur finale est obtenue comme étant la moyenne de toutes les mesures effectuées sur tous les segments du signal. Le problème principal du SNR_{seg} est que pour quelques trames, la valeur du SNR est trop élevée ou trop faible ce qui peut biaiser le SNR_{seg}.

C.4. LMSE (Least Mean Square Error)

L'erreur quadratique moyenne définie par :

$$\text{EQM}(x, \hat{x}) = \frac{1}{N} \|x - \hat{x}\|^2 \quad (C.6)$$

Est considérée comme indice de qualité de la compression.

Où

- x représente le signal original.
- $\hat{x}(n)$ représente le signal compressé.
- N est le nombre d'échantillons du signal.

Notons également que pour le cas de la compression de parole, l'estimé $\hat{x}(n)$ ne représente pas toujours la meilleure solution du point de vue du système auditif humain.

C.5. MOS (Mean Opinion Score)[13,17,18]

Le MOS est une mesure, largement utilisée pour évaluer la qualité de parole. Il nécessite de 12 à 24 évaluateurs (les tests formels CCTTT et TIA incluent typiquement 32-64 évaluateurs) qui ont été entraînés à fixer la valeur phonétique des enregistrements par rapport à une qualité à 5 niveaux. Alors, le niveau de l'évaluation et sa description est présenté dans le tableau C.1.

| Niveau | Qualité de la parole | Dégradation (niveau de distorsion) | Relation entre échelle MOS et qualité de la parole |
|--------|----------------------|------------------------------------|--|
| 1 | Mauvaise | Très gênante et désagréable | 4 → 4.5 qualité réseau. |
| 2 | Faible | Gênante mais pas désagréable | 3.5 → 4 qualité de |
| 3 | Assez bonne | Perceptible mais peu gênante | communication. |
| 4 | Bonne | Perceptible mais non gênante. | 2.5 → 3.5 qualité de |
| 5 | Excellente | Imperceptible | synthèse. |

Tableau C.1 : description de l'échelle MOS et sa relation avec la qualité de la parole

Le résultat final d'un test MOS peut être donné par [1] :

$$(\text{MOS})_{kl} = \frac{1}{N_e} \sum_{j=1}^{N_e} \text{OS}_j \quad (\text{C.7})$$

Où

$(\text{OS})_j$ indique la note donnée par l'écouteur « j ». N_e : le nombre d'écouteurs.

k : indice indiquant le locuteur qui a prononcé la phrase « l ».

C.5.1. Application du MOS à l'évaluation de l'algorithme FS

Pour chaque signal de parole et chaque longueur LF un ensemble S de séquences de signature est construit selon l'algorithme de génération de la FS cité auparavant. On reconstruit les signaux de parole sélectionnés aléatoirement, à l'aide des FS et on examine la qualité des signaux synthétisés, par la méthode subjective MOS :

Pour évaluer la qualité, 10 hommes et 10 femmes ont écouté les sons reconstruits. Chaque individu donna une note. Les évaluateurs suggèrent la classification des signaux de parole reconstruits comme suit :

$\text{MOS} \leq 2.2$ (mauvaise qualité). $2.2 \leq \text{MOS} \leq 2.6$ (intelligible). $2.6 \leq \text{MOS} \leq 3.2$ (bonne qualité). $3.2 \leq \text{MOS} \leq 4.0$ (très bonne qualité). $\text{MOS} > 4.0$ (excellente qualité)

APPENDICE D

CORPUS DE DONNEES TRAITEES

D.1.Introduction

La base des données est réalisée par enregistrements des passages lus par différents locuteurs. Ainsi, cette base de données inclut un grand nombre de locuteurs. Pour construire ce corpus on passe par les étapes suivantes :

- L'enregistrement proprement dit des signaux. Le logiciel utilisé pour l'enregistrement est le « Goldwave ». Les signaux sont échantillonnés à 8KHz ,16 bits.
- L'annotation des signaux : elle inclut le texte qui a été prononcé par le locuteur accompagné de quelques caractéristiques.

D.2.Description des sons tests

| Nom du son | Description | Caractéristiques |
|----------------|--|-------------------------------------|
| MSA_DARS01_NAS | الدرس الأول " نص طويل جدا. | Fe : 16 KHz |
| MSA_DARS02_NAS | الدرس الثاني " نص طويل جدا. | Débit : 256 Kbps |
| Phrase 1 | " بسم الله الرحمن الرحيم " | Durée totale : |
| Phrase 2 | " رأيت كلا من الصلاح و العلم " ليس كمثلته صاحب . | Son1 =3.31min |
| Phrase 3 | " و وجدته مكسبه " إذا وفق الله و أعان يسيرا . | Son2=3.38 min |
| Phrase 4 | " و وجدته يدل على الخير " و يشير بالنصح . | Phrases 1 :10 |
| Phrase 5 | " و وجدته لا ينقص على " إلا نفاق بل يزداد. | Fe : 8KHz |
| Phrase 6 | " و لا يخلق على كثرة الإستعمال " بل يزهو و يكثر. | Débit : 128 Kbps |
| Phrase 7 | " سأعيش رغم الداء و الأعداء " كالنسر فوق القمة السماء. | Durées totales : |
| Phrase 8 | " أرنو إلى الشمس المضيئة " هازنا بالسحب و الأمطار. | 3 - 15s (variable) |
| Phrase 9 | " شعب الجزائر مسلم " و إلى العروبة ينتسب. | Durée utilisée pour |
| Phrase 10 | " و فل ربي زدني علما " | tous les sons : 3s (2400 échant) |

Tableau D.1 : Les noms et la description des sons utilisés dans les expériences.

| Numéro de la phrase | Les locuteurs femmes | | | | | | | | | |
|---------------------------|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
| 1 | lf1s1 | lf2s1 | lf3s1 | lf4s1 | lf5s1 | lf6s1 | lf7s1 | lf8s1 | lf9s1 | lf10s1 |
| 2 | lf1s2 | lf2s2 | lf3s2 | lf4s2 | lf5s2 | lf6s2 | lf7s2 | lf8s2 | lf9s2 | lf10s2 |
| 3 | lf1s3 | lf2s3 | lf3s3 | lf4s3 | lf5s3 | lf6s3 | lf7s3 | lf8s3 | lf9s3 | lf10s3 |
| 4 | lf1s4 | lf2s4 | lf3s4 | lf4s4 | lf5s4 | lf6s4 | lf7s4 | lf8s4 | lf9s4 | lf10s4 |
| 5 | lf1s5 | lf2s5 | lf3s5 | lf4s5 | lf5s5 | lf6s5 | lf7s5 | lf8s5 | lf9s5 | lf10s5 |
| 6 | lf1s6 | lf2s6 | lf3s6 | lf4s6 | lf5s6 | lf6s6 | lf7s6 | lf8s6 | lf9s6 | lf10s6 |
| 7 | lf1s7 | lf2s7 | lf3s7 | lf4s7 | lf5s7 | lf6s7 | lf7s7 | lf8s7 | lf9s7 | lf10s7 |
| 8 | lf1s8 | lf2s8 | lf3s8 | lf4s8 | lf5s8 | lf6s8 | lf7s8 | lf8s8 | lf9s8 | lf10s8 |
| 9 | lf1s9 | lf2s9 | lf3s9 | lf4s9 | lf5s9 | lf6s9 | lf7s9 | lf8s9 | lf9s9 | lf10s9 |
| 10 | lf1s10 | lf2s10 | lf3s10 | lf4s10 | lf5s10 | lf6s10 | lf7s10 | lf8s10 | lf9s10 | lf10s10 |

| Numéro de la phrase | Les locuteurs hommes | | | | | | | | | |
|---------------------------|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|---------|
| | H1 | H2 | H3 | H4 | H5 | H6 | H7 | H8 | H9 | H10 |
| 1 | lh1s1 | lh2s1 | lh3s1 | lh4s1 | lh5s1 | lh6s1 | lh7s1 | lh8s1 | lh9s1 | lh10s1 |
| 2 | lh1s2 | lh2s2 | lh3s2 | lh4s2 | lh5s2 | lh6s2 | lh7s2 | lh8s2 | lh9s2 | lh10s2 |
| 3 | lh1s3 | lh2s3 | lh3s3 | lh4s3 | lh5s3 | lh6s3 | lh7s3 | lh8s3 | lh9s3 | lh10s3 |
| 4 | lh1s4 | lh2s4 | lh3s4 | lh4s4 | lh5s4 | lh6s4 | lh7s4 | lh8s4 | lh9s4 | lh10s4 |
| 5 | lh1s5 | lh2s5 | lh3s5 | lh4s5 | lh5s5 | lh6s5 | lh7s5 | lh8s5 | lh9s5 | lh10s5 |
| 6 | lh1s6 | lh2s6 | lh3s6 | lh4s6 | lh5s6 | lh6s6 | lh7s6 | lh8s6 | lh9s6 | lh10s6 |
| 7 | lh1s7 | lh2s7 | lh3s7 | lh4s7 | lh5s7 | lh6s7 | lh7s7 | lh8s7 | lh9s7 | lh10s7 |
| 8 | lh1s8 | lh2s8 | lh3s8 | lh4s8 | lh5s8 | lh6s8 | lh7s8 | lh8s8 | lh9s8 | lh10s8 |
| 9 | lh1s9 | lh2s9 | lh3s9 | lh4s9 | lh5s9 | lh6s9 | lh7s9 | lh8s9 | lh9s9 | lh10s9 |
| 10 | lh1s10 | lh2s10 | lh3s10 | lh4s10 | lh5s10 | lh6s10 | lh7s10 | lh8s10 | lh9s10 | lh10s10 |

Tableaux D.2 : Codage des fichiers sons. Les 200 fichiers ont été enregistrés sous l'extension (.wav)

Plusieurs passages sont donnés permettant l'enregistrement de différents énoncés. Les dix phrases du tableau D.1 ont été enregistrées par nous même, pour tester l'algorithme FS multi phrases. L'enregistrement a été fait à une fréquence de 8 KHz, à l'aide du logiciel « Goldwave ».

La durée d'enregistrement d'une même phrase varie d'un locuteur à un autre. Donc, pour des signaux de parole de 24000 échantillons d'un même passage lu, on n'aura pas la même quantité de parole.

Les deux premiers fichiers sons ont été utilisés dans l'algorithme FS pour une seule phrase. Ces fichiers ont été décomposés en signaux de 24000 échantillons pour obtenir les dix phrases de test. Les deux sons correspondent respectivement à deux locuteurs homme et femme.

Ces signaux de parole ont été extraits d'un CD-ROM. Ils ont été enregistrés à 16 KHz 16 bits. Pour les utiliser dans nos expériences, on les a re-échantillonnés à 8 KHz à l'aide du logiciel « Goldwave ». Le tableau D.3 montre les phrases reconstruites dans ce cas.

| Phrases reconstruites pour une femme | | Phrases reconstruites pour un homme | |
|--------------------------------------|----------------------------|-------------------------------------|----------------------------|
| P1 | "الدرس الثاني العرب قبل" | P1 | "الدرس الأول أسعد زوجين" |
| P2 | "في الجزيرة العربية و بعض" | P2 | "جلس يستمع" |
| P6 | "قسمين بد و و حضر" | P3 | "و كان ذلك المستمع الكريم" |

Tableau D.3 : Les phrases reconstruites pour l'algorithme FS mono locuteur.

APPENDICE E

COMPRESSION DU SIGNAL DE PAROLE PAR LA TECHNIQUE LPC

E.1. Description

Le programme ci-dessous effectue la compression d'un signal de parole arabe codé par « MSA_DARS01_NAS.wav ». Le principe repose sur la synthèse de parole par modèle AR. Chaque bloc (trame) de N échantillons (240 éch = 30 ms) est remplacé par 12 coefficients du modèle.

- Extraction des paramètres AR : En prenant une fenêtre de 30 ms de signal, et en utilisant une fonction nommée « xtoa », qui calcule les coefficients a_k , on comprime l'enregistrement en prenant 50% de recouvrement entre fenêtres (pour atténuer l'effet des variations brutales des coefficients d'une fenêtre à la suivante).
- Synthèse : On effectue un filtrage inverse $A(z)$ de ce bloc de signal de parole. En admettant que le conduit vocal se comporte comme un filtre linéaire pendant des trames de temps de l'ordre de 10 à 30ms, on peut donc modéliser :
 - Les sons non voisés comme la sortie d'un filtre excité par un bruit blanc,
 - et les sons voisés comme la sortie d'un filtre excité par une source périodique dont la période est appelée, *pitch*. Dans les deux cas, on choisit pour le filtre une structure tout-pôle $1/A(z)$ dont l'ordre va, en pratique de 10 à 50.

E.2. Résultats de simulation

E.2.1. Evaluation de la qualité de la parole en fonction de l'ordre du filtre

Dans ce cas, on a pris une fenêtre d'analyse de 0.03s et un recouvrement de 0.015s (50%). Les résultats de cette expérience sont résumés dans le tableau E.1. On remarque que la qualité du signal est moins bonne et reste presque constante en fonction de l'ordre du filtre.

LMSE1 : Erreur quadratique moyenne normalisée par rapport au nombre d'échantillons.

LMSE2 : Erreur quadratique moyenne normalisée par rapport au nombre de trames.

| Ordre du filtre | SNR (dB) | LMSE1 | LMSE2 | Qualité du son reconstruit |
|-----------------|----------|--------|--------|---|
| 10 | -1.4751 | 0.0196 | 2.3863 | La qualité est moins bonne dans tous les cas. |
| 12 | -1.3661 | 0.0191 | 2.3272 | |
| 16 | -1.0484 | 0.0178 | 2.1630 | |
| 20 | -0.8331 | 0.0169 | 2.0584 | |
| 40 | -0.8694 | 0.0170 | 2.0784 | |
| 50 | -0.8638 | 0.0170 | 2.0730 | |
| 60 | -0.6842 | 0.0163 | 1.9890 | |

Tableau E.1 : Résultats de la synthèse de la parole par la LPC (en variant l'ordre du filtre)

E.2.2. Influence de la fenêtre d'analyse et le taux de recouvrement sur la qualité de la parole

Dans ce cas, on a fixé l'ordre du filtre à 10 et on a varié la largeur de la fenêtre d'analyse. Les résultats obtenus sont regroupés dans le tableau suivant :

| Fenêtre d'analyse (s) | Recouvrement (s) | SNR (dB) | LMSE1 | LMSE2 | Qualité de la parole |
|-----------------------|------------------|----------|--------|--------|----------------------|
| 0.0150 | 0.0075 | -0.6599 | 0.0162 | 0.9815 | Dégradation |
| 0.0200 | 0.0100 | -2.5335 | 0.0250 | 2.0197 | perceptible |
| 0.0300 | 0.0150 | -3.0840 | 0.0284 | 4.6414 | et peut être |
| 0.0400 | 0.0200 | -3.0840 | 0.0284 | 4.6320 | gênante. |

Tableau E.2 : Résultats de la synthèse de la parole par la LPC (en variant la fenêtre d'analyse)

APPENDICE F

ORGANIGRAMES ET DESCRIPTION DE LA FS

F.1.introduction

Cet appendice fournit une description détaillée de l'algorithme FS, à travers une série d'organigrammes et une description des algorithmes qui les accompagnent. La FS a été obtenue à l'aide de deux principaux algorithmes :

- détermination de la FS en analysant une série de sons : c'est l'étape d'analyse.
- Synthèse d'un signal de la parole à l'aide de la FS prédéfinie : c'est l'étape de synthèse.

F.2.Terminologie utilisée dans les organigrammes

| Symbole | | Signification |
|----------|-----------|---|
| Analyse | P | Indice indiquant le numéro du signal traité |
| | SEG | Matrice regroupant les premiers vecteurs propres de chaque matrice de corrélation. |
| | S | Ensemble des fonctions de signature. |
| | V_{1k} | Premier vecteur propre de la trame k (vecteur associé à la plus grande valeur propre) |
| | P_{max} | Nombre total de signaux de parole traités (corpus de 200 phrases). |
| | Seuil | Seuil de comparaison des vecteurs de SEG. |
| | N | Nombre de vecteurs dans SEG. |
| Synthèse | k_{max} | Nombre total de trame d'un morceau de parole. |
| | j_{max} | Nombre de vecteurs constituant l'ensemble des fonctions de signatures. |
| | C_{kr} | Coefficient associé au vecteur S_r de signature permettant la synthèse d'une trame. |
| | d | L'erreur quadratique moyenne (distance euclidienne entre deux vecteurs). |
| | seuil | Pour déterminer l'erreur minimale de reconstruction. |

Tableau F.1 : Terminologie utilisée dans les différents organigrammes.

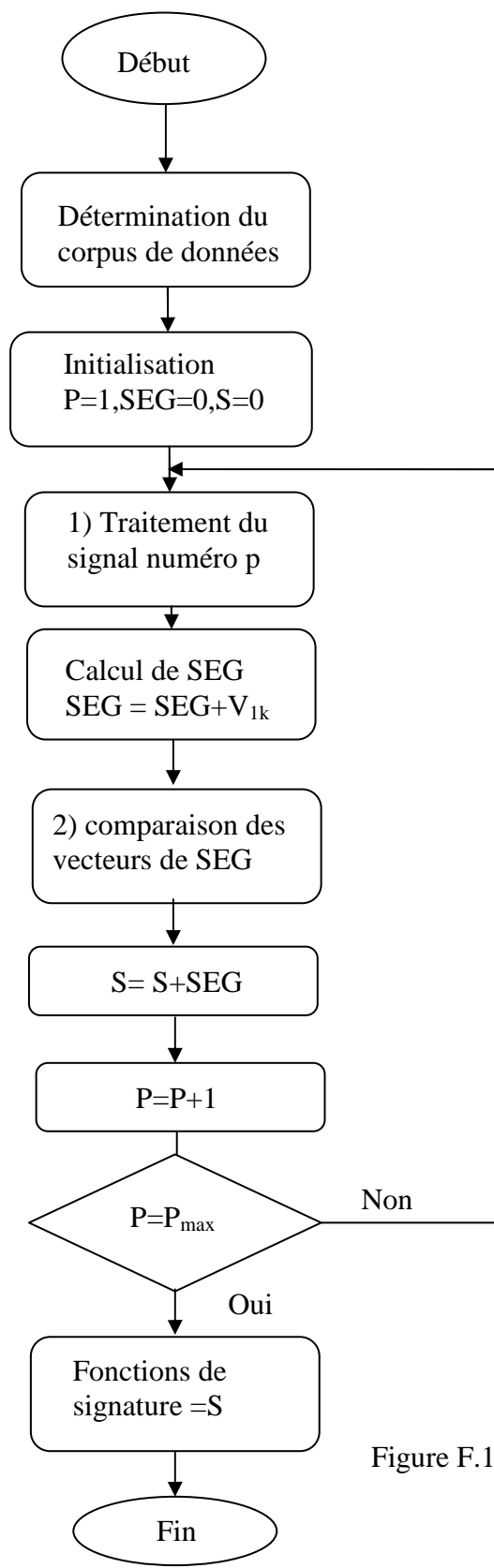
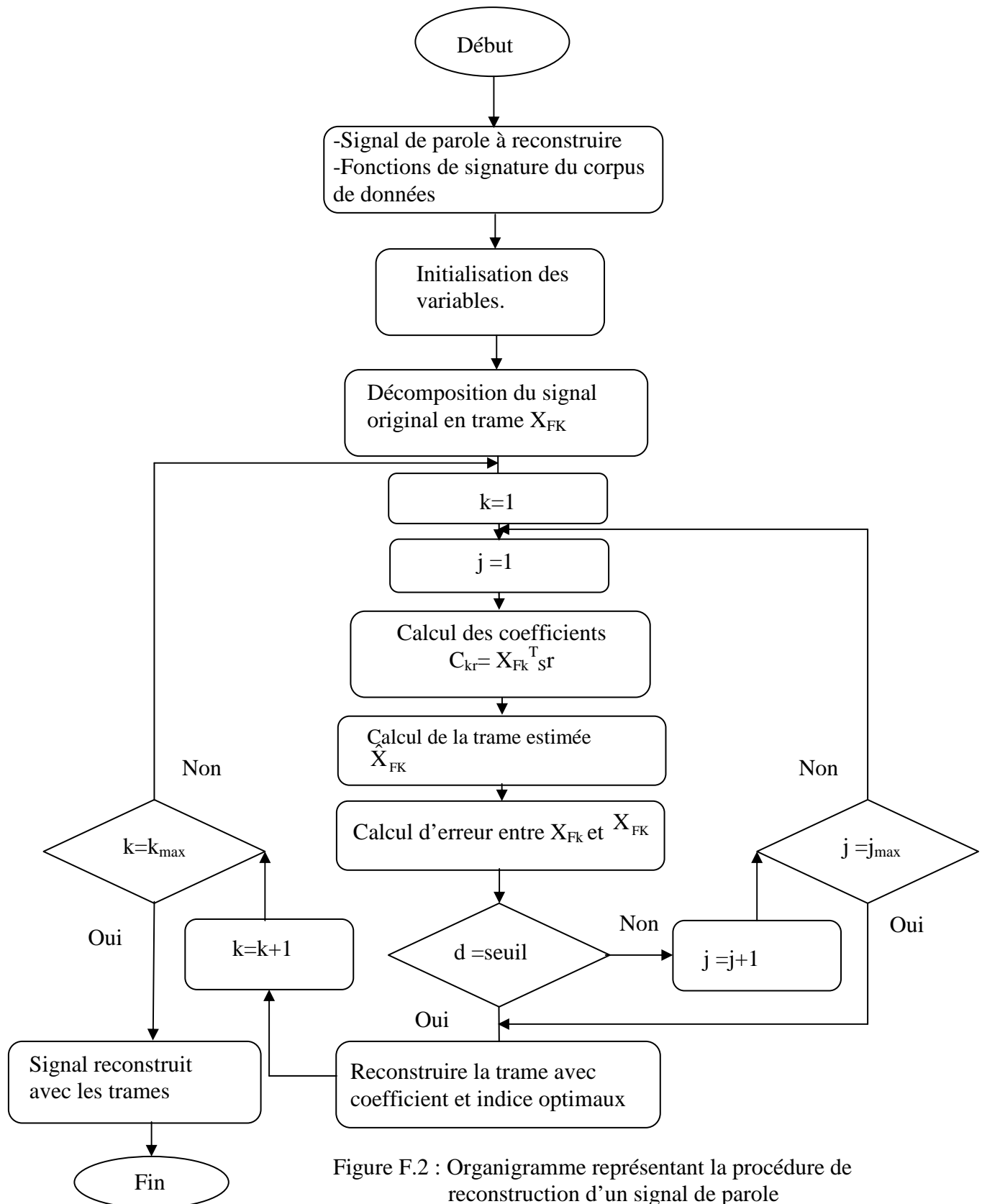


Figure F.1 : Organigramme représentant la procédure d'extraction des fonctions de signature



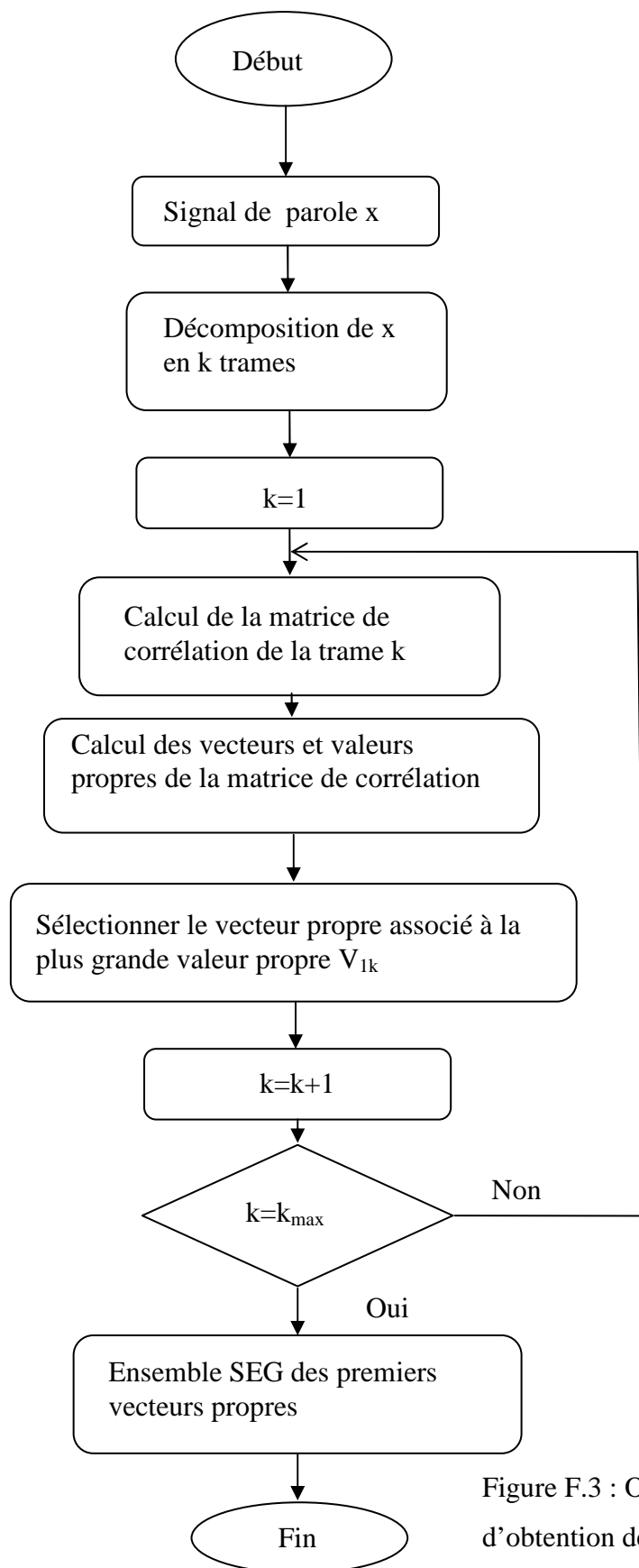


Figure F.3 : Organigramme représentant la procédure d'obtention de l'ensemble SEG : étape 1 de l'algorithme 1

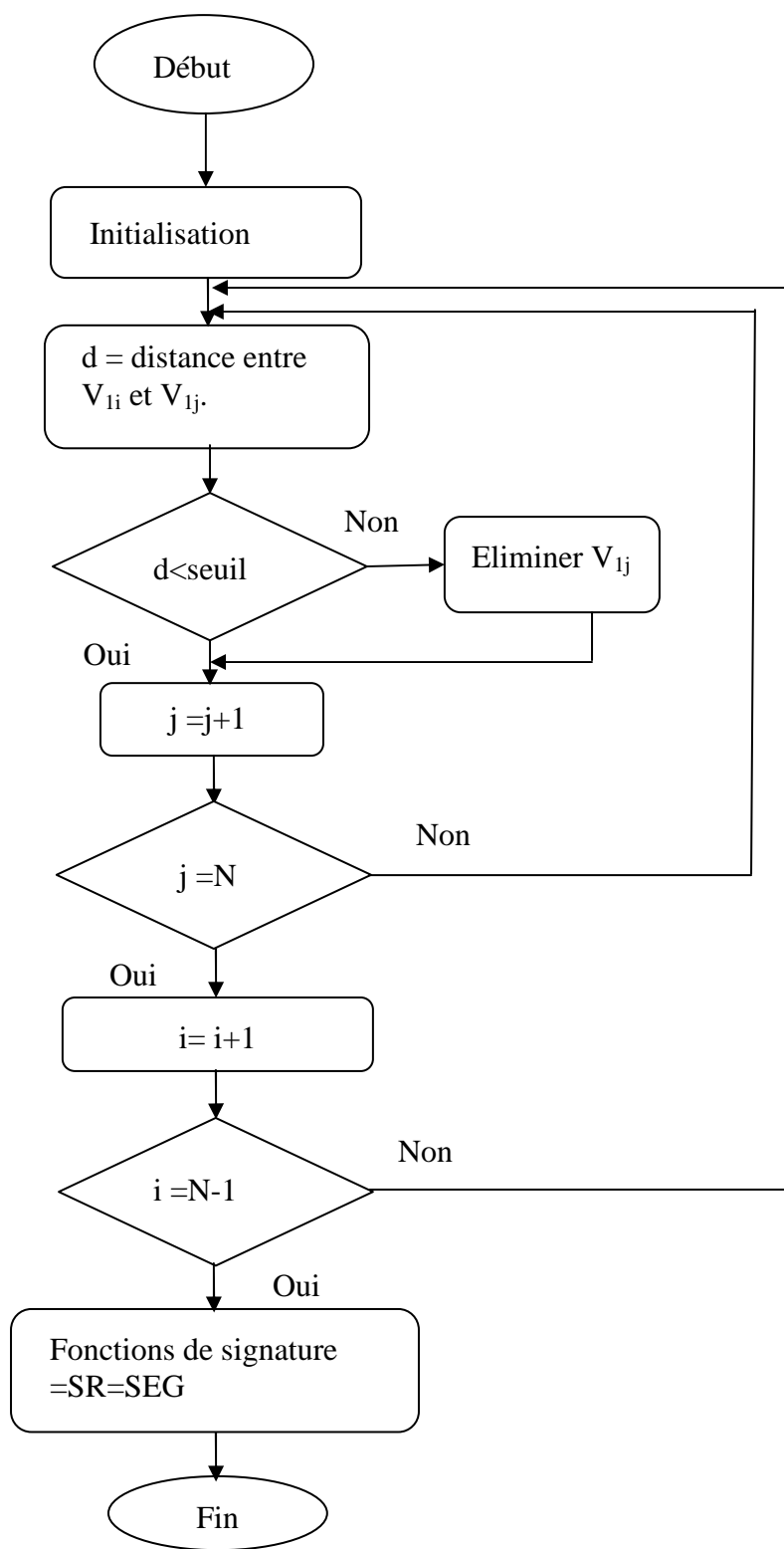


Figure F.4 : Organigramme représentant la procédure de comparaison des vecteurs de SEG.

REFERENCES BIBLIOGRAPHIQUES

1. R.Adeniz , S.Yarman « speech coding signature functions » Corlu Engineering Faculty Isik University Turkey sptember 17,2004.
2. Umit Guz, BS Yarman , Hakam Gurkan « A New Method to represent speech signals via predefined functional Bases » ECC TD 01-European Conference on circuits theory and Design. August 28-31, 2001, Espao,Finland.
3. R.Adeniz , BS.Yarman « Representation of speech signals by single signature base function within optimum frame length » Proc,43rd IEEE Midwest Symp on circuits and systems,lansing MI Aug 8-11-2000.
4. R.Adeniz, BS.Yarman, AM Karas «Speech coding by signature base sequences» Turkey, 1998
5. Karl Schwerdt, « Compression fondée sur l'apparence » version préliminaire. 18, mai 2001, école mathématiques, sciences et technologie de l'information, Grenoble.
6. Paris Smaragdis, « Redundancy reduction for computational Audition,a unifying approach» May 7, 2001. Media Arts and Sciences School of Architecture and planning© Massachusetts institute of technology.
7. Alexandru Isar, Andrei Cubitch, Miranda Nafornta « Algorithmes et techniques de compression ». Iditura orizonturi politehnice, 2002
8. José Hernández « Algorithmes d'acquisition, compression et restitution de la parole à vitesse variable, étude et mise en place », Avril 1995.
9. Olivier Rioul, Pierre DuHamel « Information et codage ». ENSTA voie Sic (systèmes d'information et de communication). 09-2003
10. Christine Decaestecker & Macro Saerens (ULB & UCL). « Analyses en Composantes Principales » LINF 2275 Stat.explor-multidim.
11. John C.Stapleton, Steven Craig Bass « Synthesis of musical Tones Basedon the Karhunen-Loève Transform » IEEE Transactions on Acoustics, Speech and Signal Processing,Vol.36,No.3,March 1988
12. C.Demars « Représentations Bidimensionnelles d'un signal de parole, éléments de monographie », version 2000, revue et augmentée. Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur, BP.133.91403 Orsay Cedex France.

13. D^r Ir Bernard Gosselin, « Introduction au codage de l'information » Faculté Polytechnique de Mons.
14. Slim ESSID, « Normes de codage audio » INT-Février 2006, slim.essid@enst.fr.
15. C.O Etemoglu and V.Cuperman, « Matching pursuit sinusoidal speech coding», IEEE Trans speech Audi. Proc September 11, 2003.
16. James P.Reilly© , « ECE Lecture Notes: Matrix computations for signal processing» Departement of Electrical McMaster University, October 17,2005.
17. ANDREAS S.SPANIAS « Speech coding: A tutorial Review » Proceeding of the IEEE, vol.82, N^o:10, October 1994.
18. MOHAMMAD M.A.Khan «Coding of Excitation Signals In a waveform interpolation Speech Coder », Departement of Electrical & Computer Engineering, Mc Gill University Montreal,Canada, July 2001.

