الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR- ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR - ANNABA

جامعة باجي مختار-عنابة

Faculté : Sciences de l'ingéniorat **Année : 2018/2019**
Département : Electronique

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat 3ème Cycle

Intitulé

Amélioration du signal de la parole par les filtres de Kalman

Option : *Multimédia et Communications Numériques*

Par : MELLAHI Tarek

Directeur de Thèse : HAMDI Rachid Pr. Univ. Badji Mokhtar-Annaba

DEVANT LE JURY

Président :	**DOGHMANE Noureddine**	**Pr. Univ. Badji Mokhtar-Annaba**
Examinateurs :	**BOUKROUCHE Abdelhani**	**Pr. Univ. Guelma**
	BAARIR Zineddine	**Pr. Univ. Biskra**
	BOUSBIA SALAH Mounir	**Pr. Univ. Badji Mokhtar-Annaba**
	MESSADEG Djemil	**Pr. Univ. Badji Mokhtar-Annaba**

Dédicace

Cette thèse est dédiée à mes deux parents. Mr Bachir Mellahi, mon père, ne m'a pas seulement élevé et nourri, mais il s'est également dépensé beaucoup au fil des ans pour mon éducation et mon développement intellectuel. Ma mère, Mme Khamissa Seklouli a été une source de motivation et de force pendant les moments de désespoir et de découragement. Ses soins maternels et son soutien ont été montrés de manière incroyable récemment.

Remerciements

Tout d'abord, je voudrais exprimer ma sincère gratitude à mon directeur de thèse Prof. Rachid Hamdi pour le soutien continu de mon étude de doctorat et de nos recherches, pour sa patience, sa motivation et sa immense connaissance. Ses conseils m'ont aidé tout au long de la recherche et de l'écriture de cette thèse. Je n'aurais pas pu imaginer avoir un meilleur directeur de thèse et un mentor pour mon doctorat.

En plus de mon directeur de thèse, je tiens à remercier le reste de mon comité de thèse : Prof. Noureddine Doghmane, Prof. Mounir Bousbia Salah, Prof. Djemil Messadeg, Prof. Abdelhani Boukrouche et Prof. Zineddine Baarir pour leurs commentaires perspicaces et leurs encouragements, mais aussi pour la question difficile qui m'a poussé à élargir mes recherches de différents points de vue. Je remercie mes camarades de laboratoire pour les discussions stimulantes, pour les nuits sans sommeil, nous travaillions ensemble avant les dates limites, et pour tout le plaisir que nous avons eu au cours des dernières années. En particulier, je suis reconnaissant au Prof. Mr Noureddine Doghmane de m'avoir éclairé sur le premier regard de la recherche.

Dernier point mais non le moindre, je voudrais remercier ma famille : mes parents, mes frères et sœurs de m'avoir soutenu spirituellement tout au long de l'écriture de cette thèse et de ma vie en général.

ملخص

تم توظيف خوارزميات تحسين الكلام بنجاح في العديد من المجالات مثل الهواتف المحمولة والهواتف التي تعمل بدون استخدام اليدين والمؤتمرات عن بعد وفي اتصالات كابينة السيارات وأجهزة السمع والخدمات الصوتية الآلية المعتمدة على التعرف وتركيب على الكلام.

تم تقديم العديد من الطرق في الأدبيات. في هذه الرسالة نركز على تحسين الكلام أحادي القناة المتدهور بالضوضاء البيضاء أو الضوضاء الملونة.

في هذا العمل ركزنا على طريقة تكرارية جديدة للمرشح كالمان حيث يتم تقدير معلمات نموذج التنبؤ الخطي من الكلام الصاخب. ومع ذلك الصوت الوحيد المتاح هو المتلف بالضوضاء، نتائج التحسين لمرشح كالمان تعتمد إلى حد ما على دقة معاملات التنبؤ الخطية (LPCs). ومع ذلك ، فإن تحليل الكلام القائم على التنبؤ الخطي (LPC) معروف بحساسية وجود ضوضاء مضافة.

للتغلب على هذه المشكلة، نقدم تحليلاً وتطبيقًا لطريقة تحسين الصيغ المعتمدة على LPC عن طريق تعديل شدة الطيف لنموذج LPC ومن ثم إعادة تقييم LPCs جديدة ليتم تطبيقها على مرشح كالمان. هذه LPCs المحسنة هي مؤشر مفيد لأداء مرشح كالمان.

في الخوارزميات المذكورة أعلاه يفترض أن يكون نموذج الكلام خطيًا. لمعالجة مشكلة النموذج الغير الخطي نقوم بدراسة بعض الخوارزميات لمرشح كالمان التي بواسطتها نستطيع معالجة المشكل الغير خطي مثل مرشح كالمان الممتد (EKF) وغير المعطر (UKF) وذلك بتدريبه من طرف برسبترون متعدد الطبقات (MLP).

تجاربنا للتحسين تستخدم قاعدة البيانات (NOIZEUS)، حيث تحقق الطريقة المقترحة نتائج أعلى مقارنة مع طرق التحسين الأخرى.

كلمات مفتاحية: مرشح كالمان (KF) ، مرشح كالمان الممتد (EKF) ، مرشح كالمان غير معطر (UKF) ، تحسين الكلام (SE) ، معاملات التنبؤ الخطي (LPCs) ، طريقة تحسين الصفة (FEM)، برسبترون متعدد الطبقات (MLP).

Résumé

Les algorithmes d'amélioration de la parole ont été utilisés avec succès dans nombreux domaines tels que les téléphones portables, les téléphones mains libres, la téléconférence, la communication dans la cabine de voiture, les prothèses auditives et les services vocaux automatisés basés sur la reconnaissance et la synthèse de la parole.

Des nombreuses approches sont présentées dans la littérature. Cette thèse se concentre sur l'amélioration de la parole dégradée par un bruit blanc ou un bruit coloré d'un seul canal de réception (réception par mono-capteur).

Dans ce travail, nous sommes concernés par une nouvelle conception de filtrage de Kalman itératif où les paramètres d'un modèle prédictif linéaire sont estimés à partir d'un signal de parole bruité. Cependant, lorsque le seul signal disponible est corrompu par un bruit, les performances d'amélioration du filtre de Kalman dépendent de la précision des estimations des coefficients de prédiction linéaire (LPCs) et de la variance d'excitation. Néanmoins, on sait que l'analyse de la parole par la prédiction linéaire (LPC) est sensible à la présence de bruit additif.

Pour surmonter ce problème, nous présentons une analyse et une application de la méthode d'amélioration des formants (FEM) basée sur les coefficients LPC en modifiant le spectre d'amplitude logarithmique du modèle LPC puis en réévaluant les nouveaux LPCs pour les implémenter au filtre de Kalman. Ces derniers LPCs sont des indicateurs utiles de la performance du filtre de Kalman.

Dans les algorithmes ci-dessus, le modèle vocal est supposé linéaire. Les modèles du signal de parole non linéaires sont également considérés dans cette thèse. Pour résoudre le problème du modèle non linéaire, nous étudions deux algorithmes d'amélioration (entraînement au filtre Kalman étendu et au filtre de Kalman non parfumé (UKF) d'un perceptron multicouche (MLP)).

Nos expériences d'amélioration utilisent le corpus vocal NOIZEUS où les méthodes proposées atteignent des bons résultats (objectifs et subjectifs) en comparant par des autres méthodes d'amélioration.

Mots-clés: Filtre de Kalman (KF), Filtre de Kalman étendu (EKF), Filtre de Kalman non parfumé (UKF), Amélioration de la parole (SE), Coefficients de prédiction linéaire (LPCs), Méthode d'amélioration du formant (FEM), Perceptron multicouche (MLP).

Abstract

Speech enhancement algorithms have been employed successfully in many areas such as mobile phones, hands-free phones, teleconferencing, in-car cabin communication, hearing aids, automated voice services based on speech recognition and synthesis.

Many approaches are presented in the literature. This thesis focuses on enhancing single channel speech degraded by white noise or colored noise.

In this work, we are concerned by a new iterative Kalman filtering scheme where a linear predictor model parameters are estimated from noisy speech. However, when only noise-corrupted speech is available, the enhancement performance of the Kalman filter is somewhat dependent on the accuracy of the linear prediction coefficients (LPCs) and excitation variance estimates. Nevertheless, linear prediction based speech (LPC) analysis is known to be sensitive to the presence of additive noise.

To overcome this problem we present an analysis and application of the LPC-based formant enhancement method by modifying the log magnitude spectrum of the LPC model and then re-evaluating new LPCs to be applied on the Kalman filter. These enhanced LPCs are useful indicator of Kalman filter performance.

In the above algorithms, the speech model is assumed to be linear. Nonlinear speech models are also considered in this thesis. To address the nonlinear model problem, extended Kalman filter (EKF) and unscented Kalman Filter (UKF) training of multi-layer perceptron (MLP) algorithms are used in the speech enhancement.

Our enhancement experiments use a NOIZEUS speech corpus where the proposed methods achieve higher objective and subjective results compared with other enhancement methods.

Keywords: Kalman Filter (KF), Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), Speech Enhancement (SE), Linear Prediction Coefficients (LPCs), Formant Enhancement Method (FEM), Multi-Layer Perceptron (MLP).

Liste des tableaux

6.1 Comparaison des performances entre la méthode proposée (K-proposed) et les méthodes concurrentes (K-Iter et K-D-C) en termes de PESQ, SNR et SegSNR en utilisant un signal de parole bruité par un bruit blanc gaussien de SNR=5 dB. . . . 86
6.2 Comparaison des performances entre les méthodes proposées (K-proposed, MLP-EKF et MLP-UKF) en termes de PESQ, SNR et SegSNR en utilisant un signal de parole bruité par un bruit blanc gaussien de SNR=5 dB. 86

Table des figures

2.1 Le filtre de Wiener généralisé pour l'amélioration du signal de la parole 9
2.2 Diagramme de la méthode d'amélioration de la parole basée sur la phase 15
2.3 Le diagramme de bloc d'un système [Rezayee 2001] 18
2.4 Schéma du filtre de Kalman standard 19

3.1 Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase bruitée par un bruit blanc gaussien à SNR =5 dB ; (C) la phrase traitée par l'EKF ; (D) la phrase traitée par l'UKF. 47

4.1 Montrant la méthode d'amélioration des formants (FEM). 55
4.2 Enveloppes spectrales d'une trame voisée de parole bruitée par un bruit blanc de SNR = 5 dB et du spectre de parole améliorée par FEM. . . . 56
4.3 Les enveloppes d'une section entre deux formants. 60

6.1 Spectrogramme de la phrase propre (sp01.wav) 'The birch canoe slid on the smooth planks' a été prise de la base de données NOIZEUS. 77
6.2 Spectrogrammes : **A** à bande étroite et **B** à large bande de la phrase propre (sp11.wav) ' He wrote down a long list of items' a été prise de la base de données NOIZEUS. 78
6.3 Comparaison des performances de K-proposed avec d'autres algorithmes d'amélioration dans le cas de bruit blanc gaussien de différents niveaux de SNR en termes de : (a) PESQ, (b) SNR et (c) SegSNR. 87
6.4 Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase bruitée par un bruit blanc gaussien à SNR =5 dB ; (C) la phrase traitée par K-Iter ; (D) la phrase traitée par MMSE ; (E) la phrase traitée par PSC ; (F) la phrase traitée par MAP ; (G) la phrase traitée par SMPO ; (H) la phrase traitée par le K-Clean et (I) la phrase traitée par **K-proposed**. 89

6.5 Comparaison des performances de K-proposed avec d'autres algorithmes d'amélioration de parole dans le cas de bruit de voiture coloré de différents niveaux de SNR en termes de : (a) PESQ, (b) SNR et (c) SegSNR. 90
6.6 Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase corrompue par un bruit coloré de voiture à SNR =5 dB ; (C) la phrase traitée par K-Iter ; (D) la phrase traitée par MMSE ; (E) la phrase traitée par PSC ; (F) la phrase traitée par MAP ; (G) la phrase traitée par SMPO ; (H) la phrase traitée par le K-Clean et (I) la phrase traitée par le textbf K- proposed. 91
6.7 Les résultats d'évaluation subjectifs (SIG = signal, BAK = bruit de fond, et OVRL = global) de la méthode proposée (K-proposed) par rapport les différentes méthodes d'amélioration de SNR=5 dB dans le cas de bruit blanc gaussien. 93
6.8 Les résultats d'évaluation subjectifs (SIG = Signal, BAK = bruit de fond, et OVRL = global) de l'amélioration de la parole proposée (K-proposed) par rapport aux différentes méthodes d'amélioration de la parole de SNR=5 dB dans le cas de bruit coloré de voiture. 94

Liste des abréviations

ADF : Gaussian Assumed Density Filter où Le filtre de densité supposé gaussien

AMS : Analysis Modification Synthesis où Synthèse de modification d'analyse

AR : Autoregressive où auto-régressif

BAK : Background noise où Bruit de fond

CDKF : Central Differences Kalman Filter où Le filtre de Kalman des différences centrales

CKF : Cubature Kalman filters où Les filtres de Kalman cubature

DEKF : Dual Extended Kalman Filter où Filtre de Kalman étendu double

DFT : Discrete Fourier Transform où Transformation de Fourier discrète

DSP : Densité Spectrale de Puissance où Power Spectral Density

DSTFT : Discrete Short-Time Fourier Transform où Transformée de Fourier discrète de courte durée

DUKF : Dual Unscented Kalman Filter où Filtre de Kalman non parfumé double

ED : Eigenvalue Decomposition où Valeur propre de décomposition

EKF : Extended Kalman Filter où Filtre de Kalman étendu

EM : Expectation Maximization où Espérance maximisation

FEM : Formant Enhancement Method où Méthode d'amélioration du formant

FFT : Fast Fourier Transform où Transformation de Fourier rapide

FHKF : Fourier-Hermite Kalman filters où Les filtres de Kalman de Fourier-Hermite

FIR : Finite Impulse Response où Réponse impulsionnelle finie

GHKF : Gauss-Hermite Kalman filters où Les filtres de Kalman de Gauss-Hermite

HOS : Higher Order Statistics où Statistiques d'ordre supérieur

IDFT : Inverse Discrete Fourier Transform où Transformation de Fourier discrète inverse

IDSTFT : Inverse Discrete Short-Time Fourier Transform où Transformée de Fourier discrète de courte durée inverse

IFFT : Inverse Fast Fourier Transform où Transformation de Fourier rapide inverse

KF : Kalman Filter où Filtre de Kalman

KLT : Karhunen-Loeve Transformation où Transformation de Karhunen-Loeve

Kalman LPC-FEM : LPC-based Formant Enhancement Method in Kalman filtering où Méthode d'amélioration des formants basée sur LPC dans le filtre de Kalman

LMMSE : Linear Minimum Mean Square Error où Erreur quadratique moyenne minimale linéaire

LP : Linear Predictive où Prédictif linéaire

LPC : Linear Prediction Coding où Codage de prédiction linéaire

LPCs : Linear Prediction Coefficients où Coefficients de prédiction linéaire

MAP : maximum a posterior estimator of magnitude-squared spectrum où l'estimation de maximum à posteriori du spectre d'amplitude au carré

MCKF : Monte Carlo Kalman Filter où Le filtre de Kalman de Monte Carlo

MLP : Multi-Layer Perceptron où Perceptron à multicouche

MMSE : Minimum Mean Square Error où Erreur quadratique moyenne minimale

MOS : Mean Opinion Score où Score moyen d'opinion

OVRL : overall où Qualité globale

PESQ : Perceptual Evaluation of Speech Quality où Evaluation perceptuelle de la qualité de la parole

PR-bruit : Pure Residual Noise où Bruit résiduel pur

PSC : Phase Spectrum Compensation où Compensation de spectre de phase

SDC : Spectral Domain Constraint où Contrainte du domaine spectral

SDD : Speech Distortion Degree où Degré de distorsion vocale

SE : Speech Enhancement où Amélioration de la parole

SIG : Signal distortion où Distorsion de la parole

SLF : Statistically Linearized Filters où Les filtres statistiques linéarisés

SMPO : Soft Masking using a Posterior SNR uncertainty où masquage doux en utilisant l'incertitude à posteriori de SNR

SNR : Signal to Noise Ratio où Rapport signal-sur-bruit

STFT : Short-Time Fourier Transform où Transformée de Fourier de courte durée

STSA : Short Time Spectral Amplitude où Estimateur spectrale d'amplitude à courte durée

SegSNR : Segmental SNR où SNR segmental

TDC : Time Domain Constraint où Contrainte du domaine temporel

UKF : Unscented Kalman Filter où Filtre de Kalman non parfumé

UT : Unscented Transform où La transformée non parfumée

VAD : Voice Activity Detector où Détecteur d'activité vocale

Table des matières

Dédicace i
Remerciements iii
Résumé v
Abstract vii
Liste des tableaux ix
Table des figures xii
Liste des abréviations xiii
Table des matières xix

1 Introduction générale **1**
1.1 Motivation de la thèse 1
1.2 Objectif de la thèse 3
1.3 Contribution de la thèse 4
1.4 Organisation de la thèse 4

2 Etat de l'art sur l'amélioration de la parole **7**
2.1 Introduction 7
2.2 Les techniques d'amélioration de la parole 8
2.2.1 Les méthodes basées sur le filtrage de Wiener 8
2.2.2 Les méthodes basées sur l'estimation spectrales MMSE 11
2.2.3 Méthode de compensation du spectre de phase 13
2.2.4 Méthode basée sur la transformation de Karhunen-Loeve (KLT) 16
2.2.5 Méthodes basées sur le filtre de Kalman 18
2.3 Conclusion 22

3 Les filtres de Kalman **23**
3.1 Les équations de filtrage formel et les solutions exactes 23
3.1.1 Les modèles d'espace d'état probabilistes 23
3.1.2 Equations de filtrage optimales 26
3.1.3 Filtre de Kalman 27
3.2 Filtrage de Kalman étendu et non parfumé 29
3.2.1 Les expansions de la série de Taylor 30
3.2.2 Filtre de Kalman étendu (EKF) 34
3.2.3 La transformée non parfumée 38
3.2.4 Filtre de Kalman non parfumé (UKF) 42
3.3 Conclusion 46

4 Amélioration du signal de parole par le filtre de Kalman en utilisant la technique d'amélioration des formants 49
4.1 Prétraitement de la parole 51
4.2 L'analyse LPC 51
4.3 Détermination du spectre d'amplitude 53
4.4 L'identification des formants et des vallées 54
4.4.1 Détection des segments non voisés 55
4.4.2 Détermination des amplitudes et des emplacements des formants 56
4.4.3 Trouver les vallées 57
4.5 Modification du spectre d'amplitude 58
4.6 Calcul les coefficients du post-filtre 60
4.7 Réévaluer les nouveaux LPCs 61
4.8 Conclusion 61

5 Amélioration du signal de parole par les filtres de Kalman non linéaire 63
5.1 Introduction 63
5.2 Amélioration du signal de parole à l'aide de l'entraînement au EKF d'un MLP 64
5.2.1 Filtre de Kalman étendu - Estimation d'Etat 65
5.2.2 Filtre de Kalman étendu - Estimation de Poids 66
5.3 Amélioration du signal de parole à l'aide de l'entraînement au UKF d'un MLP 67
5.4 Conclusion 71

6 Résultats et comparaison avec les techniques compétitive et conventionnels 73
6.1 La Qualité et l'intelligibilité de la parole 74
6.2 Les mesures objectives 74
6.2.1 Evaluation perceptuelle de la qualité de la parole (PESQ) 75
6.2.2 Rapport signal sur bruit segmental (SegSNR) 75
6.2.3 Analyse spectrographique des signaux de parole 76
6.3 Les mesures subjectives 77
6.3.1 Calcul du score SIG : degré de distorsion de parole (SDD) 79
6.3.2 Calcul du score BAK : rapport signal/PR-bruit (SPR) 81
6.3.3 Estimation des scores SIG, BAK et OVRL 83
6.4 Expérience d'amélioration de la parole 84
6.4.1 Dispositif expérimental 84
6.4.2 Résultats et discussion 85

7 Conclusion générale **95**
7.1 Résumé du travail 95
7.2 Suggestions des futurs travaux 96

A Matériels supplémentaires **97**
A.1 Propriétés de la distribution gaussienne 97

Bibliographie **99**

Chapitre 1

Introduction générale

Sommaire

1.1 Motivation de la thèse 1
1.2 Objectif de la thèse 3
1.3 Contribution de la thèse 4
1.4 Organisation de la thèse 4

1.1 Motivation de la thèse

Avec le développement des techniques de communication et les exigences accrues sur les systèmes de reconnaissance et de la synthèse de la parole, les techniques d'amélioration de la parole sont devenues un domaine de recherche très intéressant dans les trois dernières décennies. Les dispositifs de communication mobile et les systèmes mains-libres utilisés avec les conditions de bruit très influent exigent un algorithme d'amélioration efficace. Dans le monde réel, les signaux de la parole sont souvent déformés par les bruits (par exemple, le bruit de fond émis par les moyens de transports dans un environnement urbain). Cette dégradation peut réduire l'intelligibilité et la qualité de la parole. En conséquence, les performances du système lié au traitement de la parole seront dégradées. Par exemple, si un système de téléphone mobile encode un signal de parole bruité sans prétraitement, alors les performances du système seront encore dégradés, puisque la plupart des algorithmes de codage dépendent des signaux propres et non pas des signaux bruités.

Les méthodes d'amélioration de la parole peuvent être appliquer à ces systèmes pour augmenter l'intelligibilité et / ou la qualité. La qualité de la parole est une mesure subjective qui reflète la façon de percevoir le signal par les auditeurs, tandis que l'intelligibilité est une mesure objective de la quantité d'information qui peut être extraite par les auditeurs. Les objectifs des algorithmes d'amélioration de la parole varient en fonction de l'application. Dans un système de reconnaissance vocale automatique, le principal objectif d'amélioration de la parole est d'augmenté son intelligibilité, alors que dans les systèmes de communication, la cible dépend du rapport signal-sur-bruit

(signal-to-noise ratio) (SNR) de la parole déformée. Dans les environnements de moyenne à haute SNR (par exemple> 5 dB), l'intérêt est de réduire le niveau de bruit pour produire un signal de parole naturelle. Les faibles SNR peuvent contribuer à la diminution du niveau de bruit, tout en conservant/ augmentant l'intelligibilité et réduire la fatigue auditive causée par le bruit. Les techniques de réduction de bruit sont soumises à un compromis entre le niveau d'efficacité de réduction de l'effet du bruit et de la distorsion sur le signal de la parole.

L'amélioration de la qualité et l'inéligibilité d'un signal capturé par un nombre particulier de microphones varie selon la nature d'application. Certaines d'entre elles utilisent un seul microphone ; c'est le cas d'un système de téléphone cellulaire. Par contre un système de vidéo conférence utilise un réseau de microphones. Cette thèse aborde le 1er problème qui concerne un système de microphone unique. Les algorithmes de rehaussement de la parole dans un système de microphone unique peuvent être classés en deux catégories :

1. La première catégorie comprend les méthodes du domaine fréquentiels tel que les algorithmes de soustraction spectrale [Boll 1979] et [Ephraim 1984], qui est couramment utilisée comme référence standard. Dans ces méthodes on estime la densité spectrale de puissance (DSP) d'un signal de parole propre par la soustraction de la DSP du bruit de la DSP du signal de parole bruité [Boll 1979]. Chaque estimation de la DSP est effectuée dans un segment à court duré. L'avantage des approches fondées sur la soustraction spectrale : ils sont simples et faciles à mettre en œuvre. Cependant, ces méthodes comptent aussi des limitations tel que :
   - L'effet négatif du bruit musical est toujours présent. Ce bruit est souvent perçu dans les trames de silence ou de faible parole.
   - Les performances de ces méthodes dépendent du rapport signal/bruit (SNR).
2. L'autre catégorie des algorithmes d'amélioration de parole comprend les méthodes du domaine temporel, tels que les méthodes basées sur le filtre de Kalman ([Paliwal 1987], [Gibson 1991], [Gannot 1998], [Gabrea 2001], [Popescu 1998], [So 2010], [Mellahi 2015]). L'algorithme de filtrage de Kalman a été proposé pour l'amélioration de la parole par Paliwal et Basu [Paliwal 1987], où les paramètres vocaux sont obtenus à partir d'un signal de parole 'propre' (qui ne contient pas de bruit) et les caractéristiques de bruit sont obtenues à partir des trames de silence. [Popescu 1998] propose une approche basée sur le filtre de Kalman dans le cas de bruit coloré. Ces méthodes ont pour but la détection des trames de silence pour l'estimation de la covariance de bruit. Le filtre de Kal-

man peut fournir une estimation de l'erreur quadratique moyenne minimale (MMSE) du signal propre dans le cas où le bruit est un processus gaussien. Il peut donner une estimation de l'erreur quadratique moyenne minimale linéaire (LMMSE) si le bruit est non gaussien. Avec l'approche de [Gabrea 2001], l'estimation du processus de bruit dans un modèle de parole a été effectué dans les étapes d'estimation durant le calcul des paramètres de filtrage de Kalman, sans nécessité de détecter les trames de parole / silence dans la procédure.

Bien que le coût de calcul est plus élevé que la méthode de soustraction spectrale, la performance du filtre de Kalman est meilleure du point de vue des scores de PESQ (Perceptual Evaluation of Speech Quality [Rix 2001], [Recommendation 2001]). Le PESQ est un modèle d'évaluation objective de la qualité vocale. Le bruit de fond et le bruit résiduel peuvent être évalués en présentant le PESQ avec le signal propre et dégradé. Cette thèse utilise le score de PESQ comme paramètre de performance, car les méthodes conventionnelles comportent un bruit musical dans les résultats, alors les valeurs des PESQ sont typiquement plus faibles que celle du filtre de Kalman. Certaines méthodes ont été proposées pour supprimer le bruit musical comme dans [Wójcicki 2008], mais les résultats ne sont pas entièrement satisfaisante, en particulier avec un très faible SNR d'un signal de parole bruité ($<$5 dB). Le filtre de Kalman a une meilleure performance dans de tels cas. Cependant, il ya encore de la place pour des améliorations significative.

1.2 Objectif de la thèse

Il existe trois principaux objectifs dans l'amélioration de la parole. Pour cette recherche l'objectif principal sera d'améliorer la qualité de la parole et de préserver tout au moins l'intelligibilité du signal, qui est un problème souvent rencontré dans la littérature. Les objectifs de cette thèse sont comme suit :

1. Trois algorithmes sont proposés pour l'amélioration de la parole.
2. Trois nouvelles structures de filtres de Kalman ont été proposées pour la réduction de l'effet du bruit, ainsi que la nature tonale du bruit résiduel.
3. Une évaluation qualitative et quantitative des résultats obtenus, ainsi qu'une comparaison de ces derniers (résultats) avec des travaux récents a été effectués.

On s'intéresse dans cette thèse aux systèmes qui utilisent un seul microphone (systèmes à mono-entrée). Ce type de système est connu comme étant fastidieux et complexe par rapport aux systèmes à multiple microphones. Dans un tel cas de figure, les propriétés statistiques de la parole sont exploitées.

Ainsi, elles ne requièrent aucune information additionnelle. Le système d'amélioration de la parole à mono-entrée est plus général, dans ce cas il peut être appliqué à des phrases vocales enregistrées.

1.3 Contribution de la thèse

La contribution majeure de la thèse est de développer des algorithmes d'amélioration à un seul canal basés sur les filtres de Kalman dans le but de réduire les bruits des environnements. Dans cette thèse et selon les techniques d'estimation des paramètres, plusieurs approches d'amélioration sont proposées :

1. Un algorithme itératif basé sur le filtre de Kalman, où les paramètres du modèle d'espace d'état ont été estimés à partir un signal de parole bruité par la méthode d'amélioration des formants (Kalman LPC-FEM).
2. Deux algorithmes basés sur le filtre de Kalman dans le domaine non-linéaire, tel que le filtre de Kalman étendu et non-parfumé (Extended and Unscented Kalman Filter) entrainés par le perceptron multicouche (multilayer perceptron) (EKF-MLP et UKF-MLP).

1.4 Organisation de la thèse

Le premier chapitre tient lieu d'une introduction générale pour le reste du document. Ce chapitre permet notamment de situer les problèmes abordés par rapport au cadre plus général et de présenter l'organisation des chapitres de ce document.

Dans le chapitre 2, un aperçu sur les différentes techniques principales du domaine temporel et du domaine fréquentiel qui sont utilisés dans l'amélioration du signal de la parole. Diverses transformations couramment utilisées et des algorithmes de filtrage sont également abordés dans ce chapitre.

Dans le chapitre 3, nous présentons d'abord la formule classique du filtrage optimal à temps discret comme interférence bayésienne récursive. Ensuite nous présentons le filtre de Kalman et le filtre de Kalman étendu. En plus des algorithmes classiques, on présente une étude du filtre de Kalman non parfumé.

Dans le chapitre 4, une nouvelle approche du problème de l'amélioration de la parole est présentée. Cette méthode constitue l'amélioration du signal de parole par le filtre de Kalman basique en utilisant la technique d'amélioration des formants, cette approche tente de minimiser la distorsion sans compromettre l'efficacité du processus de réduction du bruit.

Le chapitre 5 est dédié à l'amélioration du signal de parole par les filtres de Kalman non-linéaire tel que le filtre de Kalman étendu et le filtre de Kalman non parfumé entrainés par un perceptron multicouche.

Le chapitre 6 fournit une description détaillée des mesures objectives et des différents tests d'écoute subjective utilisés pour évaluer les algorithmes d'amélioration de la parole en termes de qualité et d'intelligibilité. Il fournit également la corrélation de ces mesures avec les jugements de qualité de l'auditeur humain (mesures de qualité) et les scores d'intelligibilité de l'auditeur (mesures d'intelligibilité).On présente aussi les comparaisons des différentes techniques en utilisant des évaluations objectives et des évaluations subjectives. La conclusion et les recommandations seront données dans le chapitre 7.

Chapitre 2

Etat de l'art sur l'amélioration de la parole

Sommaire

2.1 Introduction 7
2.2 Les techniques d'amélioration de la parole 8
2.2.1 Les méthodes basées sur le filtrage de Wiener 8
2.2.2 Les méthodes basées sur l'estimation spectrales MMSE 11
2.2.3 Méthode de compensation du spectre de phase 13
2.2.4 Méthode basée sur la transformation de Karhunen-Loeve (KLT) 16
2.2.5 Méthodes basées sur le filtre de Kalman 18
2.3 Conclusion 22

2.1 Introduction

Au cours des quatre dernières décennies, beaucoup de recherches ont été effectuées dans le domaine de l'amélioration de la parole, par exemple ([Ephraim 1984], [Lim 1979] et [O'Shaughnessy 1989]). Avec le développement des communications, l'étude des méthodes d'amélioration a été très intensive. L'objectif principal de ces méthodes est d'améliorer les aspects de perception (par exemple, la qualité et l'intelligibilité) d'un signal dégradé.

Dans la plupart des algorithmes d'amélioration du signal de parole, il est supposé que ce dernier est dégradé par un bruit additif qui ne dépend pas de la parole non-bruitée. Certains autres bruits pratiques peuvent être transformés en un bruit additif. Par exemple, une dégradation du bruit de convolution multiplicative ou sera converti en une dégradation de bruit additif par une transformation homomorphique [Oppenheim 1999].

La qualité des signaux de parole est une mesure subjective qui reflète la manière dont le signal est perçu par les auditeurs. D'autre part, l'intelligibilité est une mesure objective de la quantité d'information qui peut être extraite par les auditeurs du signal donné, si le signal est propre ou bruité. Jusqu'à présent,

la plupart des méthodes ne peuvent seulement améliorer qu'un aspect. Dans ce chapitre, certaines approches d'amélioration de la parole seront introduites.

Certains articles ont proposé des approches pour améliorer l'intelligibilité ([Niederjohn 1996], [Witzke 1994]). La philosophie principale est que les éléments pertinents à l'intelligibilité du signal de parole peuvent être extraits à partir du signal bruité, et utilisés pour traiter ce dernier pour augmenter son intelligibilité.

Mais la plupart des recherches menaient dans l'amélioration de la parole ont mis l'accent sur la suppression du bruit, pour améliorer la qualité globale du signal de parole. Les approches peuvent être classées en deux grandes catégories : les méthodes fréquentiel et les méthodes temporel. Pour les méthodes fréquentielles, l'approche de soustraction spectrale est utilisé presque comme standard. Elle estime la densité spectrale de puissance (DSP) du signal propre en soustrayant la DSP du bruit de la DSP du signal bruité [Boll 1979]. Chaque estimation de cette approche est effectuée dans un segment de courte durée.

La thèse porte sur le problème d'amélioration de la parole d'un seule canal. Les méthodes introduites dans les sections sont dérivées et modifiés à partir de l'approche de soustraction spectrale basique. Les autres sections décrivent les méthodes dans le domaine temporel.

2.2 Les techniques d'amélioration de la parole

2.2.1 Les méthodes basées sur le filtrage de Wiener

Les filtres de Wiener sont des filtres optimaux linéaires à temps discrets [Haykin 2000]. Etant donné un certain nombre d'observations, $y(n)$, il est souhaitable de trouver une estimation optimale linéaire $\hat{s}(n)$ du signal $s(n)$. Les observations sont la somme du signal désiré $s(n)$ et le bruit $d(n)$ comme suit :

$$y(n) = s(n) + d(n) \tag{2.1}$$

L'estimation $\hat{s}(n)$ est obtenue en appliquant un filtre linéaire aux observations :

$$\hat{s}(n) = \sum_{k=0}^{M-1} w_k y(n-k) \tag{2.2}$$

où le filtre supposé est un filtre FIR et $w_k (k = 0, \ldots, M-1)$ sont des coefficients du filtre. Selon la théorie de Wiener, le but de ce filtre est de minimiser l'erreur quadratique moyenne :

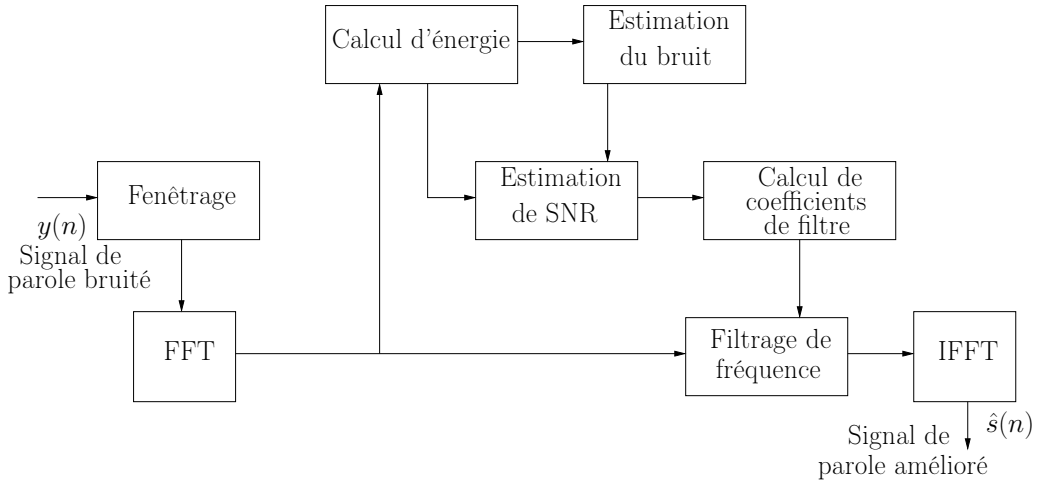

FIGURE 2.1 – Le filtre de Wiener généralisé pour l'amélioration du signal de la parole

$$E\{e(n)^2\} = E\{[\hat{s}(n) - s(n)]^2\} \tag{2.3}$$

Pour minimiser $E\{e(n)^2\}$, on utilise l'erreur orthogonale à toutes les observations :

$$E\{y(n-k)e^{\star}(n)\} = 0, \quad k = 0, \ldots, M-1 \tag{2.4}$$

où la notation $\star$ désigne le conjugué opérateur. On peut montrer [Haykin 2000] que si $y(n)$ et $s(n)$ sont conjointement de moyenne nulle de processus stationnaires de large sens, les coefficients de filtrage doivent satisfaire les équations suivantes :

$$\sum_{i=0}^{M-1} w_i r(i-k) = p(-k) \quad k = 0, 1, \ldots, M-1 \tag{2.5}$$

où $r(i-k) = E\{y(n-k)y^{\star}(n-i)\}$, $p(-k) = E\{y(n-k)s^{\star}(n)\}$.

En résolvant l'équation (2.5) pour trouver les coefficients du filtre de Wiener. Dans le domaine fréquentiel, la réponse en fréquence du filtre peut être représentée par :

$$H(w) = \frac{P_s(w)}{P_s(w) + P_d(w)} = \frac{SNR(w)}{SNR(w) + 1} \tag{2.6}$$

Dans le domaine de l'amélioration de la parole, $s(n)$ est le signal propre et $y(n)$ est le signal bruité. Par conséquent, si les spectres de puissance du signal et du bruit peuvent être obtenu ou estimé, alors un système d'amélioration

de la parole peut être construit. La figure 2.1 est un diagramme de bloque d'un système général d'amélioration à base de filtre de Wiener [Nemer 1999]. Cependant, à partir de la réponse en fréquence , on obtient les méthodes de soustraction de puissance et d'amplitude :

— Les méthodes de soustraction de puissance

A partir d'un filtre de Wiener, on peut obtenir le filtre de soustraction de puissance. La réponse de ce filtre est la racine carrée d'un filtre de Wiener. De ce fait,

$$|H(w)|_p = \sqrt{\frac{P_s(w)}{P_s(w) + P_d(w)}} \tag{2.7}$$

Le filtre peut supprimer le bruit dans le sens de spectre de puissance.

— Les méthodes de soustraction d'amplitude

Un autre type de filtre basé sur la théorie du filtre de Wiener est les filtres de soustraction d'amplitude. La réponse d'un tel type de filtres est la suivante :

$$|H(w)|_M = 1 - \sqrt{\frac{P_d(w)}{P_s(w) + P_d(w)}} \tag{2.8}$$

Le spectre d'amplitude de la sortie de ce filtre est le suivant :

$$M_{out} = \sqrt{P_s(w) + P_d(w)} - |D(w)| \tag{2.9}$$

Ainsi, le filtre peut supprimer le bruit dans le sens de spectre d'amplitude.

— L'implémentation de ces méthodes

La méthode de soustraction de puissance peut être représentée par les équations suivantes :

$$laisser \quad G(w) = P_y(w) - P_d(w) \tag{2.10}$$

$$P_{\hat{s}}(w) = \begin{cases} G(w), & if \quad G(w) > 0 \\ 0 & autrement \end{cases} \tag{2.11}$$

où $P_{\hat{s}}(w)$ est le spectre du signal de parole amélioré $\hat{s}(n)$, $P_y(w)$ est le spectre du signal bruité $y(n)$, et $P_d(w)$ est l'estimation lissée du spectre de bruit $d(n)$. Le signal amélioré est obtenu par une transformée de Fourier inverse :

$$\hat{s}(n) = F^{-1}(\sqrt{P_{\hat{s}}(w)}.e^{j\theta_y(w)}) \tag{2.12}$$

où $\theta_y(w)$ est le spectre de phase du signal de parole bruité. Le problème majeur de la méthode ci-dessus est la présence du "bruit musical" indésirable dans le signal restauré.

Pour réduire l'effet de "bruit musical", dans [Berouti 1979], une méthode de soustraction spectrale généralisée est proposée :

$$laisser \quad G(w) = \tilde{G}[P_y^{\gamma}(w) - \alpha P_d^{\gamma}(w)] \tag{2.13}$$

$$P_{\hat{s}}(w) = \begin{cases} G^{1/\gamma}(w), & si \quad G^{1/\gamma}(w) > \beta P_d(w) \\ \beta P_d(w) & autrement \end{cases} \tag{2.14}$$

avec $\alpha \geq 1$, $0 < \beta << 1$

où α est le facteur du bruit de sur-soustraction utilisé pour compenser l'estimation d'un bruit imparfait, ce qui conduit pour une réduction des pics de bruit résiduels, mais aussi à une augmentation de la distorsion audible. β est le facteur de plancher spectral (spectral floor factor), ce qui conduit à une réduction du bruit résiduel, mais une augmentation du niveau de bruit de fond reste dans la parole améliorée. L'exposant γ détermine la netteté du filtre, et $\tilde{G}$ est le facteur de normalisation. Le choix de la valeur de γ n'est pas aussi critique que celle de α et β. Si $\gamma = 1$, il s'agit un filtre de soustraction de puissance. Si $\gamma = 0.5$, c'est une méthode de soustraction d'amplitude.

L'avantage de l'approche basée sur le filtre de Wiener est qu'elle est simple et facile à mettre en œuvre. Toutefois, les limites sont :

• Le bruit musical est un signal résiduel indésirable souvent perçu dans les trames de parole faibles.

• Les performances de cette méthode dépendent d'une bonne estimation de bruit et de SNR.

2.2.2 Les méthodes basées sur l'estimation spectrales MMSE

En 1984, Ephraïm et Malah ont proposé une méthode d'amélioration de la parole sur la base de l'estimation spectrale de l'erreur quadratique moyenne minimum (MMSE) [Ephraim 1984]. Un estimateur spectrale d'amplitude à courte durée (STSA) de MMSE est dérivé en supposant une distribution de probabilité a priori des coefficients de dilatation de Fourier (Fourier expansion coefficients) de la parole et du bruit. Les coefficients de dilatation de Fourier de chaque processus sont modélisés comme des variables aléatoires gaussiennes statistiquement indépendantes de moyenne nulle.

Si le signal de parole bruité est représenté par l'équation (2.1), la $k^{ième}$ composante spectrale des signaux $s(n)$ et $y(n)$ peut être exprimée comme suit :

$$S_k = A_k e^{j\alpha_k}, \tag{2.15}$$

$$Y_k = R_k e^{jv_k}, \quad k = 0, \pm 1, \pm 2, \cdots \tag{2.16}$$

D_k, est la $k^{ième}$ composante spectrale du bruit $d(n)$. Cependant, l'estimation d'amplitude de MMSE $\hat{A}_k$, est la suivante :

$$\hat{A}_k = E\{A_k|Y_0, Y_1, \cdots\} = E\{A_k|Y_k\} \tag{2.17}$$

où $\{Y_0, Y_1, \cdots\}$ est l'ensemble des observations spectrales sur quelques trames. L'estimateur peut être exprimé comme un gain spectral avec l'équation suivante :

$$G_{MMSE}(\xi_k, \gamma_k) \equiv \frac{\hat{A}_k}{R_k} = \frac{\sqrt{\pi}}{2}\frac{\sqrt{v_k}}{\gamma_k}\exp\Big(-\frac{v_k}{2}\Big)\Big[(1+v_k)I_0\Big(\frac{v_k}{2}\Big) + v_k I_1\Big(\frac{v_k}{2}\Big)\Big]R_k \tag{2.18}$$

avec $I_0(.)$ et $I_1(.)$ sont des fonctions de Bessel modifiés respectivement à zéro et en premier ordre.

$$v_k \equiv \frac{\xi_k}{1+\xi_k}\gamma_k \tag{2.19}$$

où ξ_k et γ_k, sont définies comme des rapports signal sur bruit (SNR) a priori et a posteriori, respectivement.

Si l'incertitude de présence de parole dans les observations bruitées est prise en compte, une fonction de gain modifié est obtenue dans [Ephraim 1984] :

$$G^D_{MMSE}(\eta_k, \gamma_k, q_k) \equiv \frac{\Lambda(\xi_k, \gamma_k, q_k)}{1+\Lambda(\xi_k, \gamma_k, q_k)} G_{MMSE}(\xi_k, \gamma_k)\big|_{\xi_k = \eta_k/(1-q_k)} \tag{2.20}$$

avec :

$\Lambda(Y_k, q_k) = \mu_k \frac{\exp(v_k)}{1+\xi_k}$, qui est dérivée à partir la définition du rapport de vraisemblance généralisée suivant :

$$\Lambda(Y_k, q_k) = \mu_k \frac{p(Y_k|H_k^1)}{p(Y_k|H_k^0)} \tag{2.21}$$

avec $\mu_k \equiv (1-q_k)/q_k$, et q_k est la probabilité du signal absent dans la $k^{ième}$ composante spectrale. H_k^0 et H_k^1 désigne respectivement les deux hypothèses

d'absence et de présence du signal. L'équation (2.20) peut être obtenue quand on suppose que les composantes spectrales soient des variables aléatoires gaussiennes. Dans (2,20), v_k, est défini comme dans (2.19), ξ_k est redéfinie par :

$$\xi_k \equiv \frac{E\{A_k^2|H_k^1\}}{\lambda_d(k)} \tag{2.22}$$

avec $\lambda_d = E\{|D_k|^2\}$.

La performance de la méthode basée sur MMSE dépend de la bonne estimation d'a priori et a posteriori de SNR, qui est encore une question ouverte dans la recherche actuelle.

2.2.3 Méthode de compensation du spectre de phase

Le principe de la méthode de compensation du spectre de phase a été introduit dans [Wójcicki 2008] comme moyen d'amélioration de la parole. Dans ce cas, nous considérons une relation de bruit de fond additive. Cela peut être représenté comme

$$y(n) = x(n) + d(n) \tag{2.23}$$

où $y(n)$, $x(n)$ et $d(n)$ sont les signaux en temps discret de la parole bruitée, de la parole propre et du bruit, respectivement. Puisque la parole est généralement supposée être quasi-stationnaire sur de courtes trames (4-32 ms), elle est analysée fenêtre par fenêtre dans le cadre de synthèse de modification d'analyse (AMS) par l'analyse discrète de Fourier à court terme. La transformée de Fourier discrète de courte durée (DSTFT) du signal de parole bruité y(n) est donnée par

$$Y(m,k) = \sum_{n=-\infty}^{\infty} y(n)w(mS-n)\exp(-j2\pi kn/K), \tag{2.24}$$

où k désigne la $k^{ième}$ fréquence discrète de K fréquences uniformément espacées, $w(n)$ est une fonction de fenêtre d'analyse, m est l'indice de trame d'analyse et S est le décalage de trame dans l'échantillon. Dans le domaine spectral, la relation de bruit est donnée comme

$$Y(m,k) = X(m,k) + D(m,k), \tag{2.25}$$

où $Y(m,k)$, $X(m,k)$ et $D(m,k)$ sont les DSTFTs de la parole bruitée, de la parole propre et du bruit, respectivement. Chacun des segments spectraux de courte durée ci-dessus peut être exprimé en termes de spectre d'amplitude et de spectre de phase. Par exemple, la DSTFT du signal de parole bruité peut être écrite sous forme polaire comme :

$$Y(m,k) = |Y(m,k)|e^{j\angle Y(m,k)} \tag{2.26}$$

où $|Y(m,k)|$ désigne le spectre d'amplitude de la parole bruitée et $\angle Y(m,k)$ désigne le spectre de la phase de la parole bruitée. Etant donné l'amplitude de la parole bruitée et le spectre de phase, notre objectif devient l'estimation du spectre DSTFT de la parole propre $X(m,k)$.

Les méthodes traditionnelles d'amélioration de la parole basées sur AMS améliorent seulement le spectre d'amplitude $|Y(m,k)|$ tout en laissant le spectre de phase bruité $\angle Y(m,k)$ [Allen 1977]. Dans ce travail, nous prenons l'approche inverse en modifiant le spectre de la phase bruité et en laissant le spectre d'amplitude bruité inchangé. La suppression du bruit est obtenue en altérant le spectre de la phase de DSTFT de manière à induire une grande annulation de synthèse parmi les composantes de bruit pendant l'opération DSTFT inverse. Une étude préliminaire de cette technique de suppression du bruit a été rapportée dans [Wójcicki 2008]. Donc, nous étendons ce mécanisme de suppression de bruit de base, ce qui lui permet de fonctionner dans un cadre d'amélioration de la parole en ligne. Ici, nous formulons une procédure qui utilise une heuristique axée sur le bruit pour contrôler le degré de modification du spectre de phase.

La suppression de bruit basée sur [Wójcicki 2008] utilise le cadre de travail AMS communément employé dans le traitement de la parole. Le cadre de travail AMS consiste les étapes suivantes :

1. Une étape d'analyse, où le signal de parole subit l'analyse de DSTFT.
2. Une étape de modification, là où le signal bruité subit une certaine forme de modification.
3. Une étape de synthèse, où l'opération de transformation de Fourier discrète inverse (IDSTFT) effectuée, suivie de la synthèse de chevauchement-ajouter.

Un schéma fonctionnel de la méthode d'amélioration de la parole basée sur la phase est montré à la Figure 2.2 Dans l'approche ci-dessus, l'atténuation est obtenue en modifiant la relation entre les paires DSTFT conjuguées. Ces paires résultent naturellement de la prise de la DSTFT d'un signal à valeur réelle, c'est-à-dire $Y(m,k) = Y^*(m,K-k)$. Dans le travail actuel, nous fournissons un mécanisme pour lier la fonction de modification de phase avec les estimations de bruit. Contrairement à la méthode décrite dans [Wójcicki 2008], la méthode détaillée ici facilite le traitement des conditions de bruit variables en fonction du temps et / ou de la fréquence.

Notre spectre de phase à court terme modifié est calculé comme suit. Premièrement, nous obtenons la fonction de modification du spectre de phase donnée par

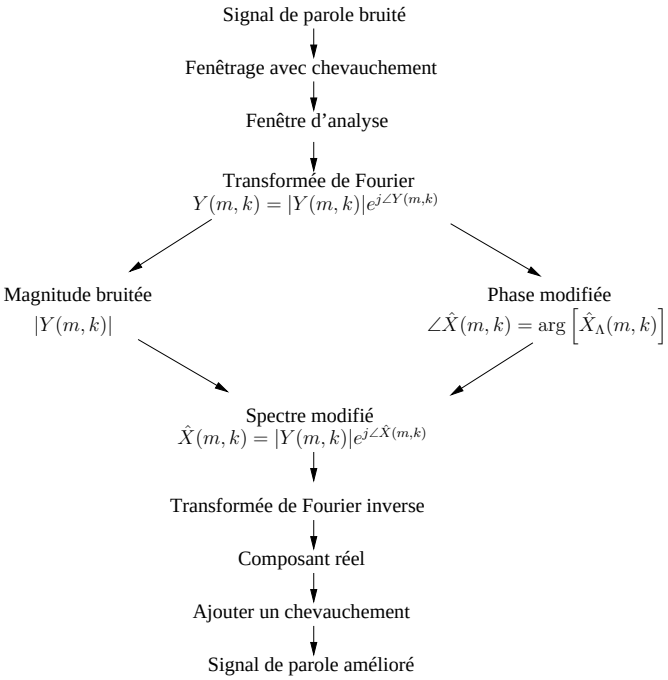

FIGURE 2.2 – Diagramme de la méthode d'amélioration de la parole basée sur la phase

$$\Lambda(m,k) = \lambda\Psi(k)|\hat{D}(m,k)|, \tag{2.27}$$

où λ est une constante déterminée empiriquement à valeur réelle, $\Psi(k)$ est une fonction antisymétrique et $|\hat{D}(m,k)|$ est une estimation du spectre d'amplitude du bruit.[1] La fonction antisymétrie invariante dans le temps est donnée par

$$\Psi(k) = \begin{cases} 1, & si \quad 0 < k/N < 0.5 \\ -1, & si \quad 0.5 < k/N < 1 \\ 0, & autrement, \end{cases} \tag{2.28}$$

Lorsqu'une pondération nulle est donnée aux valeurs correspondant aux vecteurs non conjugués de la DSTFT (c'est-à-dire la valeur $k = 0$ et le singleton possible à $k = K/2$ pour $K = pair$). Puisque l'estimation de l'amplitude du bruit $|\hat{D}(m,k)|$ est symétrique, la multiplication par $\Psi(k)$ produit une fonction $\Lambda(m,k)$ antisymétrique. C'est cette antisymétrie qui constitue la base primaire de l'annulation du bruit pendant la synthèse. L'étape suivante

1. Notez que la définition de l'estimation de bruit $|\hat{D}(m,k)|$ dans (2.27) à l'unité pour tout m,k réduit l'algorithme proposé à l'approche étudiée dans [Wójcicki 2008].

dans le calcul du spectre de phase modifié consiste à décaler le spectre complexe du signal de parole bruité par la fonction de modification additive de la valeur réelle $\Lambda(m,k)$ dépendante de la fréquence.

$$\hat{X}_\Lambda(m,k) = Y(m,k) + \Lambda(m,k). \tag{2.29}$$

le spectre de phase modifié est ensuite obtenu par

$$\angle\hat{X}(m,k) = \arg\left[\hat{X}_\Lambda(m,k)\right], \tag{2.30}$$

Où arg est la fonction d'angle complexe. Le spectre de phase modifié est recombiné avec le spectre d'amplitude bruité pour produire un spectre complexe modifié

$$\hat{X}(m,k) = |Y(m,k)|e^{j\angle\hat{X}(m,k)} \tag{2.31}$$

Dans la phase de synthèse, l'IDSTFT est utilisé pour convertir les trames du domaine spectral $\hat{X}(m,k)$ au domaine temporel. En raison du décalage additif introduit dans (2.29), les trames de domaine temporel résultantes peuvent être complexes. Cela nécessite la suppression explicite de tout composant imaginaire dans le domaine temporel. Le signal de domaine temporel amélioré $\hat{x}(n)$ est ensuite produit en employant la procédure d'addition de chevauchement. Nous nous référons à la méthode d'amélioration de la parole proposée en tant que procédure de compensation de spectre de phase (PSC) à court terme conduit par le bruit.

2.2.4 Méthode basée sur la transformation de Karhunen-Loeve (KLT)

Dans [Ephraim 1995], l'auteur a proposé une approche sous-espace de signal pour l'amélioration du signal de la parole. Le principe fondamental est de décomposer l'espace vectoriel du signal bruité en un sous-espace d'un signal plus bruité et un sous-espace d'un bruit. La décomposition est effectuée par la transformée de Karhunen-Loeve (KLT). L'amélioration est réalisée par l'estimation linéaire du signal propre à partir du signal bruité. Le bruit est supposé être additif et blanc. L'estimation linéaire est obtenue en modifiant les composants KLT qui représentent le sous-espace du signal par une fonction de gain déterminé par le critère d'estimation. Les composants KLT restants sont nulles. Le signal amélioré est obtenu à partir de l'inverse de KLT des composants altérés. Le critère d'estimation est de maintenir le niveau du bruit résiduel en dessous d'un certain seuil, tout en minimisant la distorsion du signal. Deux estimateurs sont étudiés dans [Ephraim 1995]. La première

consiste à maintenir l'énergie du bruit résiduel dans l'ensemble de la trame au-dessous d'un seuil donné. Une autre est de garder chaque composante spectrale du bruit résiduel en dessous d'un seuil donné. La première permet une contrainte du domaine temporel (TDC) sur le bruit résiduel, tandis que la seconde est conçue pour une mise en forme de bruit en utilisant des contraintes de domaine spectral (SDC). En mettant en œuvre les estimateurs de signal linéaire, il faut avoir de bonnes estimations des matrices de covariance, des vecteurs du signal bruité et le processus de bruit. En outre, une bonne estimation de la dimension du sous-espace de signal est nécessaire. Les matrices de covariance des vecteurs du processus de bruit sont estimées à partir des vecteurs de signal bruité au cours du laquelle la parole est absente. Si le bruit est stationnaire, l'estimation peut être effectué à partir d'un segment initial du signal bruité qui a été enregistré avant que le signal de parole est présent. Lorsque le bruit est non stationnaire, un détecteur parole / bruit doit être utilisé, et la covariance de bruit est estimé et mis à jour à partir des trames non-parole du signal bruité.

[Mittal 2000] a proposé une approche basée sur KLT pour améliorer la parole dégradée par un bruit coloré. Le bruit additif est supposé stationnaire. Les trames vocales sont classées en des trames d'activité vocale et des trames silencieuses. Les trames actives peuvent être divisées en des trames voisées et des trames non-voisées. L'énergie de la parole dans les trames voisées est typiquement supérieure à celui des trames non voisées et silencieuses. Les trames de parole bruitée sont classées en deux catégories, nommément celles de parole dominée et de bruit dominé. Les trames de bruit dominé sont soit des trames de l'activité non-vocaux ou des trames de parole non-voisée. Le classement se fait par une équation empirique. La matrice KLT du signal est utilisée pour les trames de la parole dominée et une matrice KLT de bruit est utilisée pour les trames de bruit dominé. L'algorithme dépend de la connaissance a priori de SNR. Ce SNR peut être évalué par la moyenne de long terme.

[Rezayee 2001] ont proposé une approche KLT adaptative pour le cas de bruit coloré. Le nouveau type de l'algorithme de KLT a été nommé le suivi sous-espace de l'approximation de projection. Dans cette méthode, la valeur propre de décomposition (Eigenvalue Decomposition) (ED) est considérée comme un problème d'optimisation sous contrainte et un algorithme adaptatif est utilisé pour trouver une approximation très proche des vecteurs propres de la matrice de covariance d'un signal de parole propre en utilisant des échantillons d'un signal de parole bruité. Cet algorithme est rapide et a une construction simple. Il est supposé que les caractéristiques statistiques du bruit ne varient pas trop jusqu'à ce que le prochain intervalle de bruit est arrivé et les échantillons de bruit sont renouvelées. Un détecteur d'activité

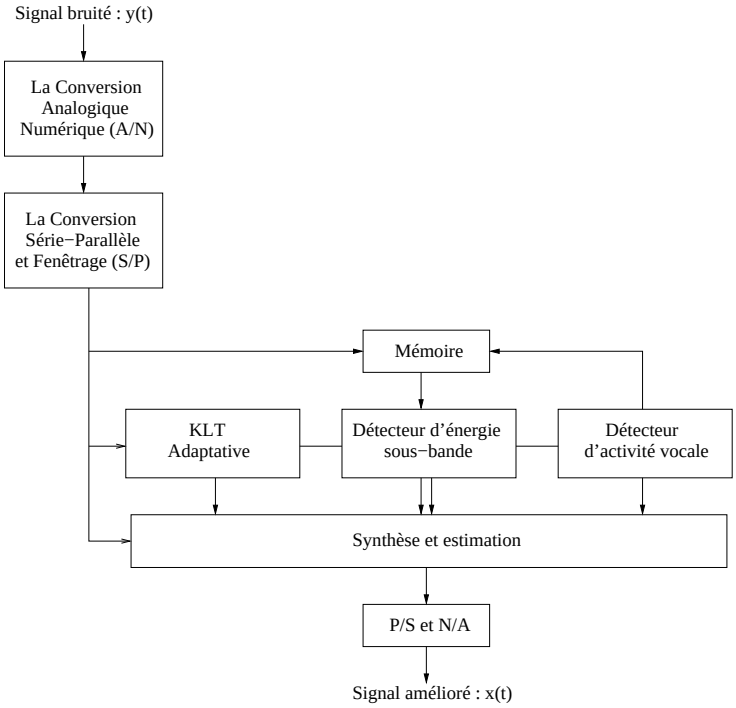

FIGURE 2.3 – Le diagramme de bloc d'un système [Rezayee 2001]

vocale efficace doit être utilisé. La Figure 2.3 représente le schéma synoptique du système proposé.

Depuis le KLT et la DFT sont reliés, la méthode populaire de soustraction spectrale peut être considérée comme une approche sous-espace d'un signal approximative.

2.2.5 Méthodes basées sur le filtre de Kalman

Les approches basées sur le filtre de Kalman sont des méthodes dans le domaine temporel. Lorsque le processus de bruit de mesure et le signal sont conjointement gaussiens, le filtre de Kalman peut donner une estimation MMSE du signal propre. Autrement, le filtre de Kalman peut produire une estimation LMMSE (Linear MMSE) du signal. L'algorithme de filtrage de Kalman a été proposé par Paliwal et Basu [Paliwal 1987] pour l'appliquer à l'amélioration de la parole. Un résumé de l'algorithme standard de filtrage de Kalman est illustré ci-dessous [Chui 1989].

Un système linéaire avec la description d'espace d'état est écrit comme suit :

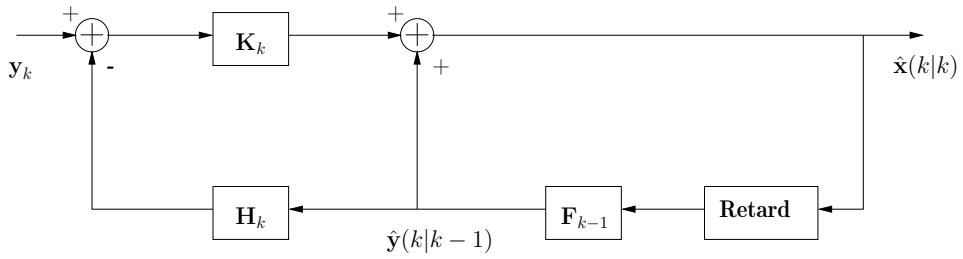

FIGURE 2.4 – Schéma du filtre de Kalman standard

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{B}_k\xi_k + \mathbf{G}_k\mathbf{u}_k \\ \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{D}_k\xi_k + \mathbf{v}_k \end{cases} \tag{2.32}$$

où $\mathbf{F}_k, \mathbf{B}_k, \mathbf{G}_k, \mathbf{H}_k, \mathbf{D}_k$ sont des matrices constantes (connu) $n \times n$, $n \times m$, $n \times p$, $q \times n$, $q \times m$, respectivement, avec $1 \leq m, p, q \leq n$, $\{\xi_k\}$ est une séquence connue de m-vecteurs (appelé une séquence d'entrée déterministe), et $\{\mathbf{u}_k\}$ et $\{\mathbf{v}_k\}$ sont respectivement, des séquences de bruit de système et d'observation (inconnu), avec des informations statistiques connues telles que la moyenne, la variance et le covariance. Quand un filtre de Kalman est appliqué au rehaussement de la parole, l'entrée déterministe est égale à zéro. Ainsi, le modèle d'espace-état pour l'amélioration de la parole est un système linéaire (purement) stochastique :

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{F}_k\mathbf{x}_k + \mathbf{G}_k\mathbf{u}_k \\ \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \end{cases} \tag{2.33}$$

L'estimation du vecteur d'état dépend de l'information statistique des séquences de bruit.

La dérivation de l'algorithme peut être trouvée dans [Chui 1989]. Supposons que $\{\mathbf{u}_k\}$ et $\{\mathbf{v}_k\}$ sont des séquences de bruit blanc gaussien de moyenne nulle de telle sorte que $var(\mathbf{u}_k) = \mathbf{Q}(k)$ et $var(\mathbf{v}_k) = \mathbf{R}(k)$ sont des matrices définies positives et $E(\mathbf{u}_k\mathbf{v}_l^T) = 0$ pour tout k et l. L'état initial $\mathbf{x}_{-1}$ est également supposé être indépendant de $\{\mathbf{u}_k\}$ et $\{\mathbf{v}_k\}$ dans le sens où $E(\mathbf{x}_{-1}\mathbf{u}_k^T) = 0$ et $E(\mathbf{x}_{-1}\mathbf{v}_k^T) = 0$ pour tout k. L'algorithme de filtrage de Kalman pour le système stochastique linéaire (2.33) est donnée par :

$$\begin{cases} \mathbf{P}(-1|-1) = var(\mathbf{x}_{-1}) \\ \mathbf{P}(k|k-1) = \mathbf{F}_k \mathbf{P}(k|k-1)\mathbf{F}_{k-1}^T + \mathbf{G}_{k-1}\mathbf{Q}_{k-1}\mathbf{G}_{k-1}^T \\ \mathbf{K}_k = \mathbf{P}(k|k-1)\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}(k|k-1)\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \mathbf{P}(k|k-1) = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}(k|k-1) \\ \hat{\mathbf{x}}(-1|-1) = E(\mathbf{x}_{-1}) \\ \hat{\mathbf{x}}(k|k-1) = \mathbf{F}_k\hat{\mathbf{x}}(k-1|k-1) \\ \hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}_k[\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}(k|k-1)] \\ k = 0, 1, \cdots \end{cases} \tag{2.34}$$

où $\mathbf{P}(k|k-1) = E\{[\mathbf{x}_k - \hat{\mathbf{x}}(k|k-1)][\mathbf{x}_k - \hat{\mathbf{x}}(k|k-1)]^T\} = var[\mathbf{x}_k - \hat{\mathbf{x}}(k|k-1)]$ est la matrice de covariance du vecteur de l'erreur d'état prédit, $\mathbf{P}(k|k) = E\{[\mathbf{x}_k - \hat{\mathbf{x}}(k|k)][\mathbf{x}_k - \hat{\mathbf{x}}(k|k)]^T\} = var[\mathbf{x}_k - \hat{\mathbf{x}}(k|k)]$ est la matrice de covariance du vecteur de l'erreur d'état estimé. $\mathbf{K}_k$ est appelé le gain de Kalman. L'algorithme est représenté sur la Figure 2.4.

Dans l'amélioration de la parole, le signal propre $x(n)$ est modélisé comme un processus auto-régressif (AR) d'ordre p comme suit :

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + w(n) \tag{2.35}$$

Le signal de parole mesuré $y(n)$ est

$$y(n) = x(n) + v(n) \tag{2.36}$$

où a_k est la $k^{ième}$ coefficients d'un modèle AR, $w(n)$ est le processus de bruit du modèle et $v(n)$ est le bruit de fond. $w(n)$ et $v(n)$ sont des processus de bruit blanc gaussien non corrélés, avec des moyens $\bar{w}$ et $\bar{v}$; et des variances σ_w^2 et σ_v^2.

Le modèle d'espace d'état est écrit comme suit, peut alors être obtenue.

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1) + \mathbf{d}w(n) \tag{2.37}$$

$$y(n) = \mathbf{c}^T\mathbf{x}(n) + v(n) \tag{2.38}$$

où

$$\mathbf{x}(n) = [x(n-p+1)\cdots x(n)]^T \tag{2.39}$$

$$\mathbf{A} = \begin{bmatrix} -a_1 & -a_2 & \cdots & -a_{p-1} & -a_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{2.40}$$

$$\mathbf{d} = \mathbf{c} = [1, 0, \ldots, 0]^T \tag{2.41}$$

où p est l'ordre du processus AR, $\mathbf{A}$ est la matrice de transition $p \times p$, $\mathbf{d}$ et $\mathbf{c}$ sont respectivement le vecteur d'entrée $p \times 1$ et le vecteur ligne d'observation $1 \times p$. L'algorithme de filtrage de Kalman pour l'amélioration de la parole est résumé dans [Gabrea 2001] comme suit :

$$e(n) = y(n) - \mathbf{c}^T \hat{\mathbf{x}}(n|n-1) \tag{2.42}$$

$$\hat{\mathbf{x}}(n|n-1) = \mathbf{A}\hat{\mathbf{x}}(n-1|n-1) \tag{2.43}$$

$$\mathbf{P}(n|n-1) = \mathbf{A}\mathbf{P}(n-1|n-1)\mathbf{A}^T + \sigma_w^2 \mathbf{d}\mathbf{d}^T \tag{2.44}$$

$$\mathbf{K}(n) = \mathbf{P}(n|n-1)\mathbf{c}[\sigma_v^2 + \mathbf{c}^T\mathbf{P}(n|n-1)\mathbf{c}]^{-1} \tag{2.45}$$

$$\hat{\mathbf{x}}(n|n) = \hat{\mathbf{x}}(n|n-1) + \mathbf{K}(n)e(n) \tag{2.46}$$

$$\mathbf{P}(n|n) = [\mathbf{I} - \mathbf{K}(n)\mathbf{c}^T]\mathbf{P}(n|n-1) \tag{2.47}$$

Dans [Gabrea 2001], les estimateurs séquentiels sont dérivés pour une estimation adaptative sous-optimale d'un processus d'apprentissage à priori inconnu et des statistiques de bruit additifs simultanément avec l'état du système en reformulant et en adaptant l'approche classique utilisée pour les applications de contrôle. L'algorithme fournit des estimations de l'état améliorées avec un taux de calcul réduit. Un avantage distinct de l'algorithme proposé est que le détecteur de parole / bruit n'est pas nécessaire.

Dans [Gannot 1998], un algorithme de filtre de Kalman basé sur les méthodes itérative et séquentielle est proposé. L'avantage distinct de l'algorithme proposé par rapport aux autres algorithmes est qu'il améliore la qualité et le SNR de la parole, tout en préservant son intelligibilité et son naturel. L'auteur de cet article a proposé d'incorporer les statistiques d'ordre supérieur pour améliorer la performance de l'algorithme.

Le signal de bruit est exprimé par :

$$z(t) = x(t) + v(t) \tag{2.48}$$

avec :

$x(t)$: le signal de parole propre, qui est modélisée comme un processus AR stochastique :

$$x(t) = -\sum_{k=1}^{p} a_k x(t-k) + \sqrt{g_x} w(t) \tag{2.49}$$

où $w(t)$ est un bruit blanc centré, et g_x est le niveau spectrale.

Le bruit additif $v(t)$ est également supposé pour être une réalisation d'une moyenne nulle, éventuellement le processus AR stochastique coloré :

$$v(t) = -\sum_{k=1}^{q} \beta_k v(t-k) + \sqrt{g_v} u(t) \tag{2.50}$$

Dans cet article, la méthode de l'espérance-maximisation (EM) est appliquée au problème d'amélioration des LPCs et de la variance. L'algorithme opère entre l'estimation de l'état et l'estimation des paramètres. L'estimation d'état utilise une récursion avant de filtrage de Kalman, suivie d'une récursion arrière de lissage de Kalman. Il a été utilisé un processus de prédiction linéaire (LPC) pour obtenir les paramètres initiaux de la parole et un détecteur d'activité vocale (VAD) pour estimer les paramètres initiaux du bruit.

L'algorithme EM itératif nécessite l'utilisation d'une courte fenêtre au cours de laquelle les statistiques de signal et de bruit sont supposées stationnaires. Un algorithme séquentiel d'amélioration de la parole est proposé dans [Gannot 1998] pour éviter la supposition précédente.

2.3 Conclusion

Dans ce chapitre, nous avons abordé les systèmes d'amélioration de la parole à canal unique tels que le filtrage de Wiener, l'estimation spectrales MMSE, la compensation du spectre de phase, la transformation de Karhunen-Loeve (KLT) et le filtre de Kalman. Ce chapitre présente également la nécessité de débruitage aux signaux de parole dans le domaine temporel, et spécialement par l'algorithme de Kalman.

Chapitre 3

Les filtres de Kalman

Sommaire

3.1 Les équations de filtrage formel et les solutions exactes 23
3.1.1 Les modèles d'espace d'état probabilistes 23
3.1.2 Equations de filtrage optimales 26
3.1.3 Filtre de Kalman 27
3.2 Filtrage de Kalman étendu et non parfumé 29
3.2.1 Les expansions de la série de Taylor 30
3.2.2 Filtre de Kalman étendu (EKF) 34
3.2.3 La transformée non parfumée 38
3.2.4 Filtre de Kalman non parfumé (UKF) 42
3.3 Conclusion 46

Dans ce chapitre, nous présentons en premier lieu la formule classique du filtrage optimal à temps discret sous la forme d'une interférence bayésienne récursive. Ensuite, le filtre de Kalman basique (KF) et les filtres de Kalman étendus (EKF) sont présentés. En plus des algorithmes classiques le filtre de Kalman non parfumé (UKF) est également présenté en vue de les appliquer sur la parole.

3.1 Les équations de filtrage formel et les solutions exactes

3.1.1 Les modèles d'espace d'état probabilistes

Avant d'aborder les algorithmes de filtrage non-linéaire pratiques, nous présentons dans les sections suivantes la théorie du filtrage probabiliste (bayésienne). Les équations de filtrage de Kalman sont les solutions de forme fermée au problème du filtrage optimal à temps discret gaussien linéaire.

Définition 1 (Modèle d'espace d'état)**.** . *Un modèle d'espace d'états à temps discret ou un modèle de filtrage probabiliste non linéaire consiste en une séquence de distributions de probabilités conditionnelles :*

$$\begin{aligned} x_k &\sim p(x_k|x_{k-1}) \\ y_k &\sim p(y_k|x_k), \end{aligned} \tag{3.1}$$

pour $k = 1, 2, \ldots,$ *où*

- $x_k \in \mathbb{R}^n$ *est l'état du système au pas de temps* k*.*
- $y_k \in \mathbb{R}^m$ *est la mesure au pas de temps* k*.*
- $p(x_k|x_{k-1})$ *est le modèle dynamique qui décrit la dynamique stochastique du système. Le modèle dynamique peut être une densité de probabilité, une mesure de comptage ou une combinaison de ceux-ci selon que l'état* x_k *soit continu, discret ou hybride.*
- $p(y_k|x_k)$ *est le modèle de mesure, qui est la distribution des mesures en fonction de l'état.*

Le modèle est supposé Markovien, ce qui signifie qu'il a les deux propriétés suivantes :

Propriété 3.1 (propriété de Markov des états)**.** .
Les états $x_k : k = 0, 1, 2, \ldots$ *forment une séquence de Markov (ou chaîne de Markov dans le cas discret). Cette propriété signifie que* x_k *étant donné* x_{k-1} *est indépendant de tout ce qui s'est passé avant le pas de temps* $k - 1$ *:*

$$p(x_k|x_{1:k-1}, y_{1:k-1}) = p(x_k|x_{k-1}). \tag{3.2}$$

le passé est aussi indépendant du futur vu le présent :

$$p(x_{k-1}|x_{k:T}, y_{k:T}) = p(x_{k-1}|x_k). \tag{3.3}$$

Propriété 3.2 (Indépendance conditionnelle des mesures)**.** .
La mesure actuelle y_k *étant donné l'état actuel* x_k *est conditionnellement indépendant des historiques de mesure et d'état :*

$$p(y_{k-1}|x_{1:k}, y_{1:k-1}) = p(y_k|x_k). \tag{3.4}$$

La marche aléatoire gaussienne est un simple exemple d'une séquence markovienne, lorsqu'elle est combinée avec des mesures bruitées, nous obtenons l'exemple d'un modèle d'espace d'état probabiliste.

Exemple 3.1.1 (marche aléatoire gaussienne)**.** *Un modèle de marche aléatoire gaussien peut être écrit comme suit*

$$\begin{aligned} x_k &= x_{k-1} + w_{k-1}, \quad w_{k-1} \sim N(0, q) \\ y_k &= x_k + e_k, \qquad\quad e_k \sim N(0, r), \end{aligned} \tag{3.5}$$

où x_k est l'état caché et y_k est la mesure. En termes de densités de probabilité, le modèle peut être défini comme

$$\begin{aligned} p(x_k|x_{k-1}) &= N(x_k|x_{k-1}, q) \\ &= \frac{1}{\sqrt{2\pi q}} exp\left(-\frac{1}{2q}(x_k - x_{k-1})^2\right) \\ p(y_k|x_k) &= N(y_k|x_k, r) \\ &= \frac{1}{\sqrt{2\pi r}} exp\left(-\frac{1}{2r}(y_k - x_k)^2\right), \end{aligned} \tag{3.6}$$

Cette équation représente un modèle d'espace d'état à temps discret.

La distribution conjointe à priori des états $(\mathbf{x}_0, \ldots, \mathbf{x}_T)$ et la vraisemblance conjointe des mesures $(\mathbf{y}_0, \ldots, \mathbf{y}_T)$ sont représentées avec l'hypothèse markovienne et le modèle de filtrage (3.1) respectivement comme suit

$$p(\mathbf{x}_0, \ldots, \mathbf{x}_T) = p(\mathbf{x}_0) \prod_{k=1}^{T} p(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{3.7}$$

$$p(\mathbf{y}_1, \ldots, \mathbf{y}_T|\mathbf{x}_0, \ldots, \mathbf{x}_T) = \prod_{k=1}^{T} p(\mathbf{y}_k|\mathbf{x}_k). \tag{3.8}$$

Nous pouvons simplement calculer la distribution à posteriori des états selon la règle de Bayes pour un T donné :

$$\begin{aligned} p(\mathbf{x}_0, \ldots, \mathbf{x}_T|\mathbf{y}_1, \ldots, \mathbf{y}_T) &= \frac{p(\mathbf{y}_1, \ldots, \mathbf{y}_T|\mathbf{x}_0, \ldots, \mathbf{x}_T)p(\mathbf{x}_0, \ldots, \mathbf{x}_T)}{p(\mathbf{y}_1, \ldots, \mathbf{y}_T)} \\ &\propto p(\mathbf{y}_1, \ldots, \mathbf{y}_T|\mathbf{x}_0, \ldots, \mathbf{x}_T)p(\mathbf{x}_0, \ldots, \mathbf{x}_T). \end{aligned} \tag{3.9}$$

Cependant, la règle complète de Bayes n'est pas réalisable dans les applications en temps réel, car la complexité de calculs s'augmente lorsque des nouvelles observations arrivent. Ainsi, de cette façon, nous pouvons uniquement travailler avec des petits ensembles, car si la quantité de données est illimitée (comme dans les applications de détection en temps réel), alors les calculs deviendraient intraitables à un certain moment. Pour faire face aux données en temps réel, nous avons besoin d'un algorithme qui effectue une quantité constante de calculs par pas de temps.

Dans ce chapitre, on s'intéresse principalement par le calcul du filtrage et les distributions de prédiction.

3.1.2 Equations de filtrage optimales

Le but du filtrage optimal est de calculer la distribution à postériori marginale de l'état x_k pour chaque instant k par rapport à l'historique des mesures jusqu'au le temps k :

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \tag{3.10}$$

Les équations fondamentales de la théorie de filtrage bayésienne sont données par le théorème suivant :

Theorème 3.1 (Les équations de filtrage optimales bayésiennes)**.** *Les équations récursives pour calculer la distribution prédite $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ et la distribution de filtrage $p(\mathbf{x}_k|\mathbf{y}_{1:k})$ à l'instant k sont données par les étapes suivantes :*

- Initialisation : *La récursivité commence à partir de la distribution a priori $p(\mathbf{x}_0)$.*
- Prédiction : *La distribution prédictive de l'état $\mathbf{x}_k$ sur le temps k donné au modèle dynamique peut être calculée par l'équation de Chapman-Kolmogorov*

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_k. \tag{3.11}$$

- Mise à jour : *Compte tenu de la mesure $\mathbf{y}_k$ au temps k, la loi à posteriori de l'état $\mathbf{x}_k$ peut être calculée par la règle de Bayes*

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = \frac{1}{Z_k}p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1}), \tag{3.12}$$

où la constante de normalisation Z_k est donnée comme

$$Z_k = \int p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})d\mathbf{x}_k. \tag{3.13}$$

Si certaines des composantes de l'état sont discrètes, les intégrales correspondantes sont remplacées par des sommations.

Démonstration. La distribution conjointe de $\mathbf{x}_k$ et $\mathbf{x}_{k-1}$ donnée $\mathbf{y}_{1:k-1}$ peut être calculée comme suit :

$$\begin{aligned} p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) &= p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_{1:k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \\ &= p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}), \end{aligned} \tag{3.14}$$

Où la disparition de l'historique de mesure $\mathbf{y}_{1:k-1}$ est due à la propriété de Markov de la suite $\{\mathbf{x}_k, k = 1, 2, \ldots\}$. La distribution marginale de $\mathbf{x}_k$ donnée

$\mathbf{y}_{1:k-1}$ peut être obtenue en intégrant la distribution (3.14) sur $\mathbf{x}_{k-1}$, ce qui donne l'équation de Chapman-Kolmogorov

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1}. \tag{3.15}$$

Si $\mathbf{x}_{k-1}$ est discret, alors l'intégrale ci-dessus est remplacée par la sommation sur $\mathbf{x}_{k-1}$. La distribution de $\mathbf{x}_k$ donnée $\mathbf{y}_k$ et $\mathbf{y}_{1:k-1}$, c'est-à-dire, donnée $\mathbf{y}_{1:k}$ peut être calculée par la règle de Bayes

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= \frac{1}{Z_k} p(\mathbf{y}_k|\mathbf{x}_k, \mathbf{y}_{1:k-1}) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) \\ &= \frac{1}{Z_k} p(\mathbf{y}_k|\mathbf{x}_k) p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) \end{aligned} \tag{3.16}$$

Où la constante de normalisation est donnée par l'équation (3.13). La disparition de l'historique des mesures $\mathbf{y}_{1:k-1}$ dans l'équation (3.16) est due à l'indépendance conditionnelle de $\mathbf{y}_k$ de l'historique des mesures, donné $\mathbf{x}_k$.

□

3.1.3 Filtre de Kalman

Le filtre de Kalman [Kalman 1960] est une solution aux équations de filtrage optimales à temps discret, où les modèles dynamiques et des mesures sont linéaires gaussiens :

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{H}_k\mathbf{x}_k + \mathbf{r}_k, \end{aligned} \tag{3.17}$$

$\mathbf{x}_k \in \mathbb{R}^n$ représente l'état, $\mathbf{y}_k \in \mathbb{R}^m$ est la mesure, $\mathbf{q}_{k-1} \sim N(0, \mathbf{Q}_{k-1})$ est le bruit de processus, $\mathbf{r}_k \sim N(0, \mathbf{R}_k)$ est le bruit de mesure et $\mathbf{x}_0 \sim N(\mathbf{m}_0, \mathbf{P}_0)$ est la distribution a priori gaussienne. $\mathbf{A}_{k-1}$ représente la matrice de transition du modèle dynamique et $\mathbf{H}_k$ est la matrice du modèle de mesure. En termes probabilistes, le modèle devient

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{x}_{k-1}) &= N(\mathbf{x}_k|\mathbf{A}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1}) \\ p(\mathbf{y}_k|\mathbf{x}_k) &= N(\mathbf{y}_k|\mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k). \end{aligned} \tag{3.18}$$

Algorithme 3.1 (filtre de Kalman)**.** *Les équations de filtrage optimales pour le modèle de filtrage linéaire* (3.17) *peuvent être évaluées sous une forme fermée d'où les distributions résultantes sont gaussiennes :*

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) &= N(\mathbf{x}_k|\mathbf{m}_k^-, \mathbf{P}_k^-) \\ p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= N(\mathbf{x}_k|\mathbf{m}_k, \mathbf{P}_k) \\ p(\mathbf{y}_k|\mathbf{y}_{1:k-1}) &= N(\mathbf{y}_k|\mathbf{H}_k\mathbf{m}_k^-, \mathbf{S}_k). \end{aligned} \tag{3.19}$$

Les paramètres des distributions ci-dessus peuvent être calculés avec les étapes de prédiction *et de* mise à jour *de filtre de Kalman :*

- L'étape de prédiction :

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{A}_{k-1}\mathbf{m}_{k-1} \\ \mathbf{P}_k^- &= \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}. \end{aligned} \tag{3.20}$$

- L'étape de mise à jour :

$$\begin{aligned} \mathbf{v}_k &= \mathbf{y}_k - \mathbf{H}_k\mathbf{m}_k^- \\ \mathbf{S}_k &= \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k \\ \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_k^T\mathbf{S}_k^{-1} \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{k}_k\mathbf{v}_k \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T. \end{aligned} \tag{3.21}$$

L'état initial a une distribution a priori gaussienne donnée $\mathbf{x}_0 \sim N(\mathbf{m}_0, \mathbf{P}_0)$*, qui définit également la moyenne et la covariance initiales.*

Les équations du filtre de Kalman de cet algorithme peuvent être dérivées comme suit :

1. D'après le lemme A.1 (voir l'annexe A), la distribution conjointe de $\mathbf{x}_k$ et $\mathbf{x}_{k-1}$ étant donnée par $\mathbf{y}_{1:k-1}$ est

$$\begin{aligned} p(\mathbf{x}_{k-1}, \mathbf{x}_k|\mathbf{y}_{1:k-1}) &= p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) \\ &= N(\mathbf{x}_k|\mathbf{A}_{k-1}\mathbf{x}_{k-1}, \mathbf{Q}_{k-1})N(\mathbf{x}_{k-1}|\mathbf{m}_{k-1}, \mathbf{P}_{k-1}) \\ &= N\left(\begin{bmatrix}\mathbf{x}_{k-1} \\ \mathbf{x}_k\end{bmatrix}\middle|\mathbf{m}', \mathbf{P}'\right). \end{aligned} \tag{3.22}$$

où

$$\mathbf{m}' = \begin{pmatrix}\mathbf{m}_{k-1} \\ \mathbf{A}_{k-1}\mathbf{m}_{k-1}\end{pmatrix}, \quad \mathbf{P}' = \begin{pmatrix}\mathbf{P}_{k-1} & \mathbf{P}_{k-1}\mathbf{A}_{k-1}^T \\ \mathbf{A}_{k-1}\mathbf{P}_{k-1} & \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1}\end{pmatrix}. \tag{3.23}$$

et la distribution marginale de $\mathbf{x}_k$ est représenté par le lemme A.2 (voir l'annexe A)

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = N(\mathbf{x}_k|\mathbf{m}_k^-\mathbf{P}_k^-), \tag{3.24}$$

où

$$\mathbf{m}_k^- = \mathbf{A}_{k-1}\mathbf{m}_{k-1}, \quad \mathbf{P}_k^- = \mathbf{A}_{k-1}\mathbf{P}_{k-1}\mathbf{A}_{k-1}^T + \mathbf{Q}_{k-1} \tag{3.25}$$

2. D'après le lemme A.1 (voir l'annexe A), la distribution conjointe de $\mathbf{y}_k$ et $\mathbf{x}_k$ est

$$\begin{aligned} p(\mathbf{x}_k, \mathbf{y}_k|\mathbf{y}_{1:k-1}) &= p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) \\ &= N(\mathbf{y}_k|\mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k)N(\mathbf{x}_k|\mathbf{m}_k^-, \mathbf{P}_k^-) \\ &= N\left(\begin{bmatrix}\mathbf{x}_k \\ \mathbf{y}_k\end{bmatrix} \middle| \mathbf{m}'', \mathbf{P}''\right). \end{aligned} \tag{3.26}$$

où

$$\mathbf{m}'' = \begin{pmatrix}\mathbf{m}_k^- \\ \mathbf{H}_k\mathbf{m}_k^-\end{pmatrix}, \quad \mathbf{P}'' = \begin{pmatrix}\mathbf{P}_k^- & \mathbf{P}_k^-\mathbf{H}_k^T \\ \mathbf{H}_k\mathbf{P}_k^- & \mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k\end{pmatrix}. \tag{3.27}$$

3. D'après le lemme A.2 (voir l'annexe A), la distribution conditionnelle de $\mathbf{x}_k$ est

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{y}_{1:k-1}) &= p(\mathbf{x}_k|\mathbf{y}_{1:k}) \\ &= N(\mathbf{x}_k|\mathbf{m}_k, \mathbf{P}_k), \end{aligned} \tag{3.28}$$

où

$$\begin{aligned} \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1}[\mathbf{y}_k - \mathbf{H}_k\mathbf{m}_k^-] \\ \mathbf{P}_k &= \mathbf{P}_k^- + \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1}\mathbf{H}_k\mathbf{P}_k^- \end{aligned} \tag{3.29}$$

et aussi écrit sous la forme (3.21).

La forme fonctionnelle des équations du filtre de Kalman décrite ici n'est pas la seule. D'un point de vue de la stabilité numérique, il serait préférable de travailler avec des racines carrées de la matrice des covariances au lieu des matrices de covariance simples. La théorie et les détails de la mise en œuvre de ce type de méthodes sont bien couverts, par exemple, dans le livre de Grewal et Andrews [Grewal 2001].

3.2 Filtrage de Kalman étendu et non parfumé

Souvent, les applications pratiques des processus dynamiques et de mesure ne sont pas linéaires comme par exemple dans le traitement de la parole, dans ce cas le filtre de Kalman basique ne peut pas être appliqué. Cependant, les distributions de filtrage de ce type de processus peuvent souvent être approchées avec des distributions gaussiennes. Dans cette section, deux types de méthodes pour former les approximations gaussiennes sont considérées : les filtres de Kalman étendus basés sur la série de Taylor (EKF) et les filtres de Kalman non parfumés (UKF) à base de la transformée non parfumée (UT).

3.2.1 Les expansions de la série de Taylor

Nous considérons la transformation d'une variable aléatoire gaussienne $\mathbf{x}$ par rapport la variable aléatoire $\mathbf{y}$:

$$\begin{aligned} \mathbf{x} &= \mathbf{N}(\mathbf{m}, \mathbf{P}) \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}) \end{aligned} \tag{3.30}$$

où $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ et $\mathbf{g} : \mathbb{R}^n \mapsto \mathbb{R}^m$ est une fonction non linéaire générale. Formellement, la densité de probabilité de la variable aléatoire $\mathbf{y}$ [1] est représentée par (voir, par exemple [Gelman 1995])

$$p(\mathbf{y}) = |\mathbf{J}(\mathbf{y})| N(\mathbf{g}^{-1}(\mathbf{y})|\mathbf{m}, \mathbf{P}), \tag{3.31}$$

$|\mathbf{J}(\mathbf{y})|$ représente le déterminant de la matrice jacobienne de la transformée inverse $\mathbf{g}^{-1}(\mathbf{y})$. Cependant, il n'est pas possible de gérer directement cette distribution vue qu'elle est non gaussienne pour tous sauf $\mathbf{g}$.

Les séries de Taylor de premier ordre basées sur l'approximation gaussienne à la distribution de $\mathbf{y}$ peuvent être formées comme suit. Si nous laissons $\mathbf{x} = \mathbf{m} + \delta\mathbf{x}$, où $\delta\mathbf{x} \sim N(\mathbf{0}, \mathbf{P})$, nous pouvons former l'expansion des séries de Taylor de la fonction $\mathbf{g}(.)$ comme suit :

$$\mathbf{g}(\mathbf{x}) = \mathbf{g}(\mathbf{m}, \delta\mathbf{x}) = \mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x} + \sum_i \frac{1}{2}\delta\mathbf{x}^T \mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\delta\mathbf{x}\mathbf{e}_i + \dots \tag{3.32}$$

où $\mathbf{G}_{\mathbf{x}}(\mathbf{m})$ est la matrice jacobienne de $\mathbf{g}$ avec les éléments

$$\left[\mathbf{G}_{\mathbf{x}}(\mathbf{m})\right]_{j,j'} = \left.\frac{\partial g_j(\mathbf{x})}{\partial x_{j'}}\right|_{\mathbf{x}=\mathbf{m}}. \tag{3.33}$$

et $\mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})$ est la matrice de hessienne de $g_i(.)$ évaluée à $\mathbf{m}$:

$$\left[\mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\right]_{j,j'} = \left.\frac{\partial^2 g_i(\mathbf{x})}{\partial x_j \partial x_{j'}},\right|_{\mathbf{x}=\mathbf{m}}. \tag{3.34}$$

Aussi, $\mathbf{e}_i = (0 \cdots 0\, 1\, 0 \cdots 0)^T$ est un vecteur avec 1 à la position i et les autres éléments sont nuls, c'est-à-dire, il est le vecteur unitaire dans la direction de l'axe des coordonnées i. L'approximation linéaire peut être obtenue en approximant la fonction par les deux premiers termes de la série de Taylor :

$$\mathbf{g}(\mathbf{x}) \approx \mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x}. \tag{3.35}$$

1. Cela ne s'applique en réalité qu'à $\mathbf{g}(.)$ inversible, mais il peut facilement être généralisé dans le cas non-inversible.

Le calcul de la valeur espérée par rapport à $\mathbf{x}$ donne :

$$\begin{aligned}\mathbf{E}[\mathbf{g}(\mathbf{x})] &\approx \mathbf{E}[\mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x}] \\ &= \mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{E}[\delta\mathbf{x}] \\ &= \mathbf{g}(\mathbf{m}).\end{aligned} \tag{3.36}$$

La covariance peut alors être approximée comme suit

$$\begin{aligned}&\mathbf{E}\left[(\mathbf{g}(\mathbf{x}) - \mathbf{E}[\mathbf{g}(\mathbf{x})])(\mathbf{g}(\mathbf{x}) - \mathbf{E}[\mathbf{g}(\mathbf{x})])^T\right] \\ &\quad\approx \mathbf{E}\left[(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m}))(\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{m}))^T\right] \\ &\quad\approx \mathbf{E}\left[(\mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x} - \mathbf{g}(\mathbf{m}))(\mathbf{g}(\mathbf{m}) + \mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x} - \mathbf{g}(\mathbf{m}))^T\right] \\ &\quad= \mathbf{E}\left[(\mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x})(\mathbf{G}_{\mathbf{x}}(\mathbf{m})\delta\mathbf{x})^T\right] \\ &\quad= \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{E}[\delta\mathbf{x}\delta\mathbf{x}^T]\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) \\ &\quad= \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}).\end{aligned} \tag{3.37}$$

Nous sommes également souvent intéressés par la covariance conjointe entre les variables $\mathbf{x}$ et $\mathbf{y}$. L'approximation de la covariance conjointe peut être réalisée en considérant la transformation augmentée

$$\tilde{\mathbf{g}}(\mathbf{x}) = \begin{pmatrix}\mathbf{x} \\ \mathbf{g}(\mathbf{x})\end{pmatrix}. \tag{3.38}$$

La moyenne et la covariance résultantes sont :

$$\begin{aligned}\mathbf{E}[\tilde{\mathbf{g}}(\mathbf{x})] &\approx \begin{pmatrix}\mathbf{m} \\ \mathbf{g}(\mathbf{m})\end{pmatrix} \\ \mathbf{Cov}[\tilde{\mathbf{g}}(\mathbf{x})] &\approx \begin{pmatrix}\mathbf{I} \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m})\end{pmatrix}\mathbf{P}\begin{pmatrix}\mathbf{I} \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m})\end{pmatrix}^T \\ &= \begin{pmatrix}\mathbf{P} & \mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P} & \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m})\end{pmatrix}.\end{aligned} \tag{3.39}$$

Dans la dérivation des équations du filtre de Kalman étendu, nous avons besoin d'une transformation plus générale de la forme suivante

$$\begin{aligned}\mathbf{x} &\sim \mathbf{N}(\mathbf{m}, \mathbf{P}) \\ \mathbf{q} &\sim \mathbf{N}(\mathbf{0}, \mathbf{Q}) \\ \mathbf{y} &= \mathbf{g}(\mathbf{x}) + \mathbf{q},\end{aligned} \tag{3.40}$$

où $\mathbf{q}$ est indépendante de $\mathbf{x}$. La distribution conjointe de $\mathbf{x}$ et $\mathbf{y}$ (comme définie ci-dessus) est maintenant la même que dans les équations (3.41), sauf que la covariance $\mathbf{Q}$ est ajoutée au bloc inférieur droit de la matrice de covariance de $\tilde{\mathbf{g}}(\cdot)$. Nous obtenons ainsi l'algorithme suivant :

Algorithme 3.2 (Approximation linéaire d'une transformée additive)**.** *L'approximation linéaire basée sur l'approximation gaussienne de la distribution conjointe de* $\mathbf{x}$ *et de la variable aléatoire transformée* $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{q}$*, où* $\mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{P})$ *et* $\mathbf{q} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q})$ *est donnée comme*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mu_{\mathbf{L}} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C}_{\mathbf{L}} \\ \mathbf{C}_{\mathbf{L}}^T & \mathbf{S}_{\mathbf{L}} \end{pmatrix} \right), \tag{3.41}$$

où

$$\begin{aligned} \mu_{\mathbf{L}} &= \mathbf{g}(\mathbf{m}) \\ \mathbf{S}_{\mathbf{L}} &= \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) + \mathbf{Q} \\ \mathbf{C}_{\mathbf{L}} &= \mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}), \end{aligned} \tag{3.42}$$

où $\mathbf{G}_{\mathbf{x}}(\mathbf{m})$ *est la matrice jacobienne de* $\mathbf{g}$ *en fonction de* $\mathbf{x}$*, évaluée à* $\mathbf{x} = \mathbf{m}$ *avec les éléments suivants*

$$[\mathbf{G}_{\mathbf{x}}(\mathbf{m})]_{jj'} = \left. \frac{\partial g_j(\mathbf{x})}{\partial x_{j'}} \right|_{\mathbf{x}=\mathbf{m}} \tag{3.43}$$

En outre, dans les modèles de filtrage où le bruit de processus n'est pas additif, nous devons approximer les transformations de la forme suivante

$$\begin{aligned} &\mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{P}) \\ &\mathbf{q} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q}) \\ &\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q}), \end{aligned} \tag{3.44}$$

Où $\mathbf{x}$ et $\mathbf{q}$ sont des variables aléatoires non corrélées. La moyenne et la covariance peuvent être calculées en substituant le vecteur augmenté $(\mathbf{x}, \mathbf{q})$ au vecteur $\mathbf{x}$ dans l'équation (3.41). La matrice jacobienne conjointe peut être écrite comme $\mathbf{G}_{\mathbf{x},\mathbf{q}} = (\mathbf{G}_{\mathbf{x}}\mathbf{G}_{\mathbf{q}})$. Ici $\mathbf{G}_{\mathbf{q}}$ est la matrice jacobienne de $\mathbf{g}(\cdot)$ en fonction de $\mathbf{q}$ et les deux matrices jacobiennes sont évaluées à $\mathbf{x} = \mathbf{m}, \mathbf{q} = \mathbf{0}$. Les approximations de la moyenne et de la covariance de la transformée augmentée comme dans l'équation (3.41) sont données comme

$$\begin{aligned} \mathbf{E}[\tilde{\mathbf{g}}(\mathbf{x}, \mathbf{q})] &\approx \mathbf{g}(\mathbf{m}, \mathbf{0}) \\ \mathbf{Cov}[\tilde{\mathbf{g}}(\mathbf{x}, \mathbf{q})] &\approx \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m}) & \mathbf{G}_{\mathbf{q}}(\mathbf{m}) \end{pmatrix} \begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m}) & \mathbf{G}_{\mathbf{q}}() \end{pmatrix}^T \\ &= \begin{pmatrix} \mathbf{P} & \mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) \\ \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P} & \mathbf{G}_{\mathbf{x}}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) + \mathbf{G}_{\mathbf{q}}(\mathbf{m})\mathbf{Q}\mathbf{G}_{\mathbf{q}}^T(\mathbf{m}) \end{pmatrix}. \end{aligned} \tag{3.45}$$

L'approximation ci-dessus peut être formulée comme l'algorithme suivant :

Algorithme 3.3 (Approximation linéaire d'une transformée non additive)**.** *L'approximation linéaire basée sur l'approximation gaussienne de la distribution conjointe de* $\mathbf{x}$ *et de la variable aléatoire transformée* $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q})$ *où* $\mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{P})$ *et* $\mathbf{q} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q})$ *est donnée comme suite*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mu_{\mathbf{L}} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C_L} \\ \mathbf{C}_{\mathbf{L}}^T & \mathbf{S_L} \end{pmatrix} \right), \tag{3.46}$$

où

$$\begin{aligned} \mu_{\mathbf{L}} &= \mathbf{g}(\mathbf{m}) \\ \mathbf{S_L} &= \mathbf{G_x}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) + \mathbf{G_q}(\mathbf{m})\mathbf{Q}\mathbf{G}_{\mathbf{q}}^T(\mathbf{m}) \\ \mathbf{C_L} &= \mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}), \end{aligned} \tag{3.47}$$

Et $\mathbf{G_x}(\mathbf{m})$ *est la matrice jacobienne de* $\mathbf{g}$ *en fonction de* $\mathbf{x}$*, évaluée à* $\mathbf{x} = \mathbf{m}, \mathbf{q} = \mathbf{0}$ *avec les éléments suivants*

$$[\mathbf{G_x}(\mathbf{m})]_{jj'} = \left. \frac{\partial g_j(\mathbf{x}, \mathbf{q})}{\partial x_{j'}} \right|_{\mathbf{x}=\mathbf{m}, \mathbf{q}=\mathbf{0}} \tag{3.48}$$

et $\mathbf{G_q}(\mathbf{m})$ *est la matrice jacobienne correspondante en fonction de* $\mathbf{q}$ *:*

$$[\mathbf{G_q}(\mathbf{m})]_{jj'} = \left. \frac{\partial g_j(\mathbf{x}, \mathbf{q})}{\partial q_{j'}} \right|_{\mathbf{x}=\mathbf{m}, \mathbf{q}=\mathbf{0}} \tag{3.49}$$

Dans les approximations quadratiques, les termes du second ordre dans l'expansion des séries de Taylor de la fonction non-linéaire sont étudiés :

Algorithme 3.4 (L'approximation quadratique d'une transformée non linéaire additive)**.** *L'approximation du second représente par la forme suivante*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mu_{\mathbf{Q}} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C_Q} \\ \mathbf{C}_{\mathbf{Q}}^T & \mathbf{S_Q} \end{pmatrix} \right), \tag{3.50}$$

où les paramètres sont

$$\begin{aligned} \mu_{\mathbf{Q}} &= \mathbf{g}(\mathbf{m}) + \frac{1}{2}\sum_i \mathbf{e}_i \mathrm{tr}\left\{ \mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\mathbf{P} \right\} \\ \mathbf{S_Q} &= \mathbf{G_x}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}) + \frac{1}{2}\sum_{i,i'} \mathbf{e}_i \mathbf{e}_{i'}^T \mathrm{tr}\left\{ \mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\mathbf{P}\mathbf{G}_{\mathbf{xx}}^{(i')}(\mathbf{m})\mathbf{P} \right\} \\ \mathbf{C_Q} &= \mathbf{P}\mathbf{G}_{\mathbf{x}}^T(\mathbf{m}), \end{aligned} \tag{3.51}$$

où $\mathbf{G_x}(\mathbf{m})$ *est la matrice jacobienne* (3.45) *et* $\mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})$ *est la matrice hessienne de* $g_i(\cdot)$ *évaluée à* $\mathbf{m}$ *:*

$$\left[\mathbf{G}_{\mathbf{xx}}^{(i)}(\mathbf{m})\right]_{jj'} = \left.\frac{\partial^2 g_i(\mathbf{x})}{\partial x_j \partial x_{j'}}\right|_{\mathbf{x}=\mathbf{m}}, \tag{3.52}$$

où $\mathbf{e}_i = (0 \cdots 010 \cdots 0)^T$ *est un vecteur avec* 1 *en position* i *et les autres éléments sont nuls, c'est-à-dire, il est le vecteur unitaire dans la direction de l'axe des coordonnées* i.

3.2.2 Filtre de Kalman étendu (EKF)

Le filtre de Kalman étendu (EKF) (voir [Jaswinski], [Maybeck 982a], [Bar 2001], [Grewal 2001]) est une extension du filtre de Kalman basique pour résoudre le problème du filtrage optimal non linéaire. Dans cela, les bruits de processus et de mesure peuvent être considérés additifs, donc le modèle d'espace d'état non-linaire écrit comme suit :

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1} \\ \mathbf{y}_k &= \mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k, \end{aligned} \tag{3.53}$$

où $\mathbf{x}_k \in \mathbb{R}^n$ représente l'état, $\mathbf{y}_k \in \mathbb{R}^m$ est la mesure, $\mathbf{q}_{k-1} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ est le bruit de processus gaussien, $\mathbf{r}_k \sim \mathbf{N}(\mathbf{0}, \mathbf{R}_k)$ est le bruit de mesure gaussien, $\mathbf{f}(\cdot)$ est la fonction du modèle dynamique et $\mathbf{h}(\cdot)$ est la fonction du modèle de mesure. Les fonctions $\mathbf{f}$ et $\mathbf{h}$ peuvent aussi dépendre du nombre de pas k.

L'idée du filtre de Kalman étendu est de former des approximations gaussiennes.

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) \approx \mathbf{N}(\mathbf{x}_k|\mathbf{m}_k, \mathbf{P}_k) \tag{3.54}$$

L'EKF utilise des approximations linéaires aux modèles non-linéaires, le résultat est l'algorithme suivant :

Algorithme 3.5 (filtre de Kalman étendu I)**.** *Les étapes de prédiction et de mise à jour du filtre de Kalman étendu du premier ordre (EKF) sont :*

- *Prédiction :*

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}) \\ \mathbf{P}_k^- &= \mathbf{F}_{\mathbf{x}}(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\mathbf{F}_{\mathbf{x}}^T(\mathbf{m}_{k-1}) + \mathbf{Q}_{k-1}. \end{aligned} \tag{3.55}$$

- *Mise à jour :*

$$\begin{aligned} \mathbf{v}_k &= \mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-) \\ \mathbf{S}_k &= \mathbf{H}_{\mathbf{x}}(\mathbf{m}_k^-)\mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-) + \mathbf{R}_k \\ \mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-)\mathbf{S}_k^{-1} \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\mathbf{v}_k \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T. \end{aligned} \tag{3.56}$$

Ces équations peuvent être résolues en répétant les mêmes étapes de la dérivation du filtre de Kalman dans la section 3.1.3 et en appliquant les approximations de la série de Taylor aux étapes appropriées :

1. La distribution conjointe de $\mathbf{x}_k$ et $\mathbf{x}_{k-1}$ est non gaussienne, mais nous pouvons former une approximation gaussienne en appliquant l'algorithme 3.2 à la fonction suivante :

$$\mathbf{f}(\mathbf{x}_{k-1}) + \mathbf{q}_{k-1}, \tag{3.57}$$

ce qui résulte l'approximation gaussienne suivante

$$p(\mathbf{x}_{k-1}, \mathbf{x}_k, |\mathbf{y}_{1:k-1}) \approx \left(\begin{bmatrix} \mathbf{x}_{k-1} \\ \mathbf{x}_k \end{bmatrix} \middle| \mathbf{m}', \mathbf{P}' \right), \tag{3.58}$$

où

$$\begin{aligned} \mathbf{m}' &= \begin{pmatrix} \mathbf{m}_{k-1} \\ \mathbf{f}(\mathbf{m}_{k-1}) \end{pmatrix} \\ \mathbf{P}' &= \begin{pmatrix} \mathbf{P}_{k-1} & \mathbf{P}_{k-1}\mathbf{F}_{\mathbf{x}}^T \\ \mathbf{F}_{\mathbf{x}}\mathbf{P}_{k-1} & \mathbf{F}_{\mathbf{x}}\mathbf{P}_{k-1}\mathbf{F}_{\mathbf{x}}^T + \mathbf{Q}_{k-1} \end{pmatrix}, \end{aligned} \tag{3.59}$$

et la matrice jacobienne $\mathbf{F}_{\mathbf{x}}$ de $\mathbf{f}(\mathbf{x})$ est évaluée en $\mathbf{x} = \mathbf{m}_{k-1}$. La moyenne et la covariance de $\mathbf{x}_k$ sont

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}) \\ \mathbf{P}_k^- &= \mathbf{F}_{\mathbf{x}}\mathbf{P}_{k-1}\mathbf{F}_{\mathbf{x}}^T + \mathbf{Q}_{k-1}. \end{aligned} \tag{3.60}$$

2. La distribution conjointe de $\mathbf{y}_k$ et $\mathbf{x}_k$ est également non-gaussienne, mais nous pouvons encore la former en utilisant l'algorithme 3.2 à la fonction

$$\mathbf{h}(\mathbf{x}_k) + \mathbf{r}_k. \tag{3.61}$$

Nous obtenons l'approximation suivante :

$$p(\mathbf{x}_k, \mathbf{y}_k, |\mathbf{y}_{1:k-1}) \approx \left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{y}_k \end{bmatrix} \middle| \mathbf{m}'', \mathbf{P}'' \right), \tag{3.62}$$

où

$$\begin{aligned} \mathbf{m}'' &= \begin{pmatrix} \mathbf{m}_k^- \\ \mathbf{h}(\mathbf{m}_k^-) \end{pmatrix} \\ \mathbf{P}'' &= \begin{pmatrix} \mathbf{P}_k^- & \mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T \\ \mathbf{H}_{\mathbf{x}}\mathbf{P}_k^- & \mathbf{H}_{\mathbf{x}}\mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T + \mathbf{R}_k \end{pmatrix}, \end{aligned} \tag{3.63}$$

la matrice jacobienne $\mathbf{H}_{\mathbf{x}}$ de $\mathbf{h}(\mathbf{x})$ est évaluée en $\mathbf{x} = \mathbf{m}_k^-$.

3. D'après le lemme A.2 (voir l'annexe A), la distribution conditionnelle de $\mathbf{x}_k$ est approximativement

$$p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{y}_{1:k-1}) \approx \mathbf{N}(\mathbf{x}_k|\mathbf{m}_k, \mathbf{P}_k) \tag{3.64}$$

où

$$\begin{aligned} \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T (\mathbf{H}_\mathbf{x} \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T + \mathbf{R}_k)^{-1} [\mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-)] \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T (\mathbf{H}_\mathbf{x} \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T + \mathbf{R}_k)^{-1} \mathbf{H}_\mathbf{x} \mathbf{P}_k^-. \end{aligned} \tag{3.65}$$

Un modèle d'espace d'état non-linéaire plus général avec un bruit non additif peut être écrit comme

$$\begin{aligned} \mathbf{x}_k &= \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{q}_{k-1}) \\ \mathbf{x}_k &= \mathbf{h}(\mathbf{x}_k, \mathbf{r}_k), \end{aligned} \tag{3.66}$$

où $\mathbf{q}_{k-1} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ et $\mathbf{r}_k \sim \mathbf{N}(\mathbf{0}, \mathbf{R}_k)$ sont respectivement le bruit gaussien de processus et de mesure. Encore une fois, les fonctions $\mathbf{f}$ et $\mathbf{h}$ peuvent aussi dépendre du nombre des pas k.

Algorithme 3.6 (filtre de Kalman étendu II)**.** *Les étapes de prédiction et de mise à jour du filtre de Kalman étendu (EKF) (premier ordre) dans le cas de bruit non additif sont :*

- *Prédiction :*

$$\begin{aligned} \mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}, \mathbf{0}) \\ \mathbf{P}_k^- &= \mathbf{F}_\mathbf{x}(\mathbf{m}_{k-1}) \mathbf{P}_{k-1} \mathbf{F}_\mathbf{x}^T(\mathbf{m}_{k-1}) + \mathbf{F}_\mathbf{q}(\mathbf{m}_{k-1}) \mathbf{Q}_{k-1} \mathbf{F}_\mathbf{q}^T(\mathbf{m}_{k-1}). \end{aligned} \tag{3.67}$$

- *Mise à jour :*

$$\begin{aligned} \mathbf{v}_k &= \mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-, \mathbf{0}) \\ \mathbf{S}_k &= \mathbf{H}_\mathbf{x}(\mathbf{m}_k^-) \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T(\mathbf{m}_k^-) + \mathbf{H}_\mathbf{r}(\mathbf{m}_k^-) \mathbf{R}_k \mathbf{H}_\mathbf{r}^T(\mathbf{m}_k^-) \\ \mathbf{K}_k &= \mathbf{P}_k^- \mathbf{H}_\mathbf{x}^T(\mathbf{m}_k^-) \mathbf{S}_k^{-1} \\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k \mathbf{v}_k \\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T. \end{aligned} \tag{3.68}$$

$\mathbf{F}_\mathbf{x}(\mathbf{m})$, $\mathbf{F}_\mathbf{q}(\mathbf{m})$, $\mathbf{H}_\mathbf{x}(\mathbf{m})$ *et* $\mathbf{H}_\mathbf{r}(\mathbf{m})$*, sont les matrices jacobiennes de* $\mathbf{f}$ *et* $\mathbf{h}$ *qui dépendent de l'état et le bruit, avec*

$$[\mathbf{F}_\mathbf{x}(\mathbf{m})]_{jj'} = \left. \frac{\partial f_j(\mathbf{x}, \mathbf{q})}{\partial x_{j'}} \right|_{\mathbf{x}=\mathbf{m}, \mathbf{q}=\mathbf{0}}, \tag{3.69}$$

$$[\mathbf{F}_{\mathbf{q}}(\mathbf{m})]_{jj'} = \left.\frac{\partial f_j(\mathbf{x},\mathbf{q})}{\partial q_{j'}}\right|_{\mathbf{x}=\mathbf{m},\mathbf{q}=\mathbf{0}}, \tag{3.70}$$

$$[\mathbf{H}_{\mathbf{x}}(\mathbf{m})]_{jj'} = \left.\frac{\partial h_j(\mathbf{x},\mathbf{r})}{\partial x_{j'}}\right|_{\mathbf{x}=\mathbf{m},\mathbf{r}=\mathbf{0}}, \tag{3.71}$$

$$[\mathbf{H}_{\mathbf{r}}(\mathbf{m})]_{jj'} = \left.\frac{\partial h_j(\mathbf{x},\mathbf{r})}{\partial r_{j'}}\right|_{\mathbf{x}=\mathbf{m},\mathbf{r}=\mathbf{0}}, \tag{3.72}$$

Ces équations de filtrage peuvent être résolues en répétant les mêmes étapes de la dérivation du filtre de Kalman étendu ci-dessus, mais au lieu d'utiliser l'algorithme 3.2, nous utilisons l'algorithme 3.3 pour calculer les approximations.

L'avantage d'EKF par rapport aux autres méthodes de filtrage non linéaires est sa simplicité relative par rapport à ses performances. La linéarisation est une méthode d'ingénierie très courante pour construire des approximations aux systèmes non linéaires. Cependant, elle ne fonctionne pas dans les problèmes des non-linéarités considérables.

L'EKF exige également que le modèle de mesure et le modèle dynamique soient différentiables, mais dans certains cas, on ne peut pas simplement calculer les matrices jacobiennes, ce qui rend l'utilisation d'EKF impossible.

Dans le second ordre d'EKF, la non-linéarité est approximée en utilisant les termes du second ordre de l'expansion des séries de Taylor comme dans l'algorithme 3.4 :

Algorithme 3.7 (filtre de Kalman étendu III)**.** *Les étapes de prédiction et de mise à jour du filtre de Kalman étendu du second ordre (dans le cas du bruit additif) sont :*

- *Prédiction :*

$$\begin{aligned}
\mathbf{m}_k^- &= \mathbf{f}(\mathbf{m}_{k-1}) + \frac{1}{2}\sum_i \mathbf{e}_i \mathrm{tr}\left\{\mathbf{F}_{\mathbf{xx}}^{(i)}(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\right\} \\
\mathbf{P}_k^- &= \mathbf{F}_{\mathbf{x}}(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\mathbf{F}_{\mathbf{x}}^T(\mathbf{m}_{k-1}) \\
&\quad + \frac{1}{2}\sum_{i,i'} \mathbf{e}_i \mathbf{e}_{i'}^T \mathrm{tr}\left\{\mathbf{F}_{\mathbf{xx}}^{(i)}(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\mathbf{F}_{\mathbf{xx}}^{(i')}(\mathbf{m}_{k-1})\mathbf{P}_{k-1}\right\} \\
&\quad + \mathbf{Q}_{k-1}.
\end{aligned} \tag{3.73}$$

- *Mise à jour :*

$$\begin{aligned}
\mathbf{v}_k &= \mathbf{y}_k - \mathbf{h}(\mathbf{m}_k^-) - \frac{1}{2}\sum_i \mathbf{e}_i \mathrm{tr}\left\{\mathbf{H}_{\mathbf{xx}}^{(i)}(\mathbf{m}_k^-)\mathbf{P}_k^-\right\} \\
\mathbf{S}_k &= \mathbf{H}_{\mathbf{x}}(\mathbf{m}_k^-)\mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-) \\
&\quad + \frac{1}{2}\sum_{i,i'} \mathbf{e}_i\mathbf{e}_{i'}^T \mathrm{tr}\left\{\mathbf{H}_{\mathbf{xx}}^{(i)}(\mathbf{m}_k^-)\mathbf{P}_k^-\mathbf{H}_{\mathbf{xx}}^{(i')}(\mathbf{m}_k^-)\mathbf{P}_k^-\right\} + \mathbf{R}_k. \\
\mathbf{K}_k &= \mathbf{P}_k^-\mathbf{H}_{\mathbf{x}}^T(\mathbf{m}_k^-)\mathbf{S}_k^{-1} \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k\mathbf{v}_k \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T.
\end{aligned} \tag{3.74}$$

où $\mathbf{F}_{\mathbf{x}}(\mathbf{m})$ *et* $\mathbf{H}_{\mathbf{x}}(\mathbf{m})$ *sont données par les équations* (3.71) *et* (3.73). $\mathbf{F}_{\mathbf{xx}}^{(i)}(\mathbf{m})$ *et* $\mathbf{H}_{\mathbf{xx}}^{(i)}(\mathbf{m})$ *sont les matrices hessiennes de* f_i *et* h_i *respectivement :*

$$\left[\mathbf{F}_{\mathbf{xx}}^{(i)}(\mathbf{m})\right]_{jj'} = \left.\frac{\partial^2 f_i(\mathbf{x})}{\partial x_j \partial x_{j'}}\right|_{\mathbf{x}=\mathbf{m}}, \tag{3.75}$$

$$\left[\mathbf{H}_{\mathbf{xx}}^{(i)}(\mathbf{m})\right]_{jj'} = \left.\frac{\partial^2 h_i(\mathbf{x})}{\partial x_j \partial x_{j'}}\right|_{\mathbf{x}=\mathbf{m}}, \tag{3.76}$$

3.2.3 La transformée non parfumée

La transformée non parfumée (UT) ([Julier 1996] ; [Juler 2000]) est une méthode numérique relativement récente qui peut être utilisée pour approximer la distribution conjointe des variables aléatoires $\mathbf{x}$ et $\mathbf{y}$ définies comme :

$$\begin{aligned}
\mathbf{x} &\sim \mathbf{N}(\mathbf{m}, \mathbf{P}) \\
\mathbf{y} &= \mathbf{g}(\mathbf{x}).
\end{aligned}$$

Cependant, la philosophie d'UT diffère de la linéarisation, dans ce sens elle tente d'approcher directement la moyenne et la covariance de la distribution cible au lieu d'essayer d'approximer la fonction non-linéaire [Julier 1996].

L'idée d'UT est de choisir de façon déterministe un nombre fixe des points sigma qui capturent exactement la moyenne et la covariance de la distribution originale $\mathbf{x}$. Ces points sigma sont ensuite propagés à travers la non-linéarité donc la moyenne et la covariance de la variable transformée sont estimées à partir de celles-ci [Julier 2004].

La transformée non-parfumée forme une approximation gaussienne[2] par la procédure suivante :

2. Notez que la transformé non parfumée peut également être appliquée sans l'hypothèse de gaussienne. Cependant, comme cette hypothèse rend l'interprétation bayésienne de l'UT beaucoup plus facile, nous l'utiliserons ici.

1. Former un ensemble $2n+1$ de points sigma comme suit :

$$\begin{aligned} \mathcal{X}^{(0)} &= \mathbf{m} \\ \mathcal{X}^{(i)} &= \mathbf{m} + \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}}\right]_i \\ \mathcal{X}^{(i+n)} &= \mathbf{m} - \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}}\right]_i, \quad i = 1, \ldots, n, \end{aligned} \tag{3.77}$$

où $[]_i$ désigne la $i$ème colonne de la matrice, et λ est un paramètre d'échelle qui est défini par les paramètres d'algorithme α et κ comme suit :

$$\lambda = \alpha^2(n+\kappa) - n. \tag{3.78}$$

Les paramètres α et κ déterminent la propagation des points sigma autour de la moyenne [Haykin 2001]. La racine carrée de la matrice désigne une matrice telle que $\sqrt{\mathbf{P}}\sqrt{\mathbf{P}}^T = \mathbf{P}$. Les points sigma sont les colonnes de la matrice.

2. Propager les points sigma à travers la fonction non-linéaire $g(\cdot)$:

$$\mathcal{Y}^{(i)} = \mathbf{g}\left(\mathcal{X}^{(i)}\right), \quad i = 0, \ldots, 2n,$$

ce qui résulte dans des points sigma transformés $\mathcal{Y}^{(i)}$.

3. Les estimations de la moyenne et de la covariance de la variable transformée peuvent être calculées à partir des points sigma comme suit :

$$\begin{aligned} \mathbf{E}[\mathbf{g}(\mathbf{x})] &\approx \sum_{i=0}^{2n} W_i^{(m)} \mathcal{Y}^{(i)} \\ \mathbf{Cov}[\mathbf{g}(\mathbf{x})] &\approx \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}^{(i)} - \mu)(\mathcal{Y}^{(i)} - \mu)^T, \end{aligned} \tag{3.79}$$

où les poids constants $W_i^{(m)}$ et $W_i^{(c)}$ sont donnés comme suit [Haykin 2001] :

$$\begin{aligned} W_0^{(m)} &= \lambda/(n+\lambda) \\ W_0^{(c)} &= \lambda/(n+\lambda) + (1 - \alpha^2 + \beta) \\ W_i^{(m)} &= 1/2\{(n+\lambda)\}, \quad i = 1, \ldots, 2n \\ W_i^{(c)} &= 1/2\{(n+\lambda)\}, \quad i = 1, \ldots, 2n, \end{aligned} \tag{3.80}$$

et β est un paramètre d'algorithme supplémentaire qui peut être utilisé pour incorporer des informations à priori sur la distribution (non gaussienne) de $\mathbf{x}$ [Haykin 2001].

Si nous appliquons la transformation non parfumée à la fonction augmentée $\tilde{\mathbf{g}}(\mathbf{x}) = (\mathbf{x}, \mathbf{g}(\mathbf{x}))$, nous obtenons simplement l'ensemble des points sigma, où les points sigma $\mathcal{X}^{(i)}$ et $\mathcal{Y}^{(i)}$ ont été concaténés au même vecteur. Ainsi, une simple approximation de la distribution conjointe $\mathbf{x}$ et $\mathbf{g}(\mathbf{x}) + \mathbf{q}$ est également formée et le résultat est l'algorithme suivant :

Algorithme 3.8 (Approximation non-parfumée d'une transformée additive)**.** *L'approximation gaussienne basée sur la transformée non-parfumée à la distribution conjointe de* $\mathbf{x}$ *et de la variable aléatoire transformée* $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{q}$ *où* $\mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{P})$ *et* $\mathbf{q} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q})$ *est donnée comme*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mu_U \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C}_U \\ \mathbf{C}_U^T & \mathbf{S}_U \end{pmatrix} \right), \tag{3.81}$$

où les sous-matrices peuvent être calculées comme suit :

1. *Former l'ensemble des points sigma* $2n + 1$ *comme suit :*

$$\begin{aligned} \mathcal{X}^{(0)} &= \mathbf{m} \\ \mathcal{X}^{(i)} &= \mathbf{m} + \sqrt{n + \lambda} \left[\sqrt{\mathbf{P}} \right]_i \\ \mathcal{X}^{(i+n)} &= \mathbf{m} - \sqrt{n + \lambda} \left[\sqrt{\mathbf{P}} \right]_i, \quad i = 1, \ldots, n, \end{aligned} \tag{3.82}$$

   où le paramètre λ *est défini dans l'équation* (3.78)*.*

2. *Propager les points sigma à travers la fonction non-linéaire* $g(\cdot)$ *:*

$$\mathcal{Y}^{(i)} = \mathbf{g}\left(\mathcal{X}^{(i)} \right), \quad i = 0, \ldots, 2n.$$

3. *Les sous-matrices sont alors données comme suit :*

$$\begin{aligned} \mu_U &= \sum_{i=0}^{2n} W_i^{(m)} \mathcal{Y}^{(i)} \\ \mathbf{S}_U &= \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}^{(i)} - \mu_U)(\mathcal{Y}^{(i)} - \mu_U)^T + \mathbf{Q} \\ \mathbf{C}_U &= \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}^{(i)} - \mathbf{m})(\mathcal{Y}^{(i)} - \mu_U)^T, \end{aligned} \tag{3.83}$$

   les poids constants $W_i^{(m)}$ *et* $W_i^{(c)}$ *ont été définis dans l'équation* (3.80)*.*

L'approximation de la transformée non parfumée de la forme $\mathbf{y} = \mathbf{g}(\mathbf{x}, \mathbf{q})$ peut être dérivée en considérant la variable augmentée $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{q})$ comme variable aléatoire dans une transformée. Le résultat est l'algorithme suivant :

Algorithme 3.9 (Approximation non-parfumée d'une transformée non-additive)**.** *L'approximation gaussienne basée sur la transformée non parfumée à la distribution conjointe de* $\mathbf{x}$ *et de la variable aléatoire transformée* $\mathbf{y} = \mathbf{g}(\mathbf{x}) + \mathbf{q}$ *où* $\mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{P})$ *et* $\mathbf{q} \sim \mathbf{N}(\mathbf{0}, \mathbf{Q})$ *est donnée comme*

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} \mathbf{m} \\ \mu_U \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{C}_U \\ \mathbf{C}_U^T & \mathbf{S}_U \end{pmatrix} \right), \tag{3.84}$$

où n *et* n_q *représentent les dimensions de* $\mathbf{x}$ *et* $\mathbf{q}$*, respectivement, et* $n' = n + n_q$*, les sous-matrices peuvent être calculées comme suit :*

1. *Former les points sigma pour la variable aléatoire augmentée* $\tilde{\mathbf{x}} = (\mathbf{x}, \mathbf{q})$

$$\begin{aligned} \tilde{\mathcal{X}}^{(0)} &= \tilde{\mathbf{m}} \\ \tilde{\mathcal{X}}^{(i)} &= \tilde{\mathbf{m}} + \sqrt{n' + \lambda'} \left[\sqrt{\tilde{\mathbf{P}}} \right]_i \\ \tilde{\mathcal{X}}^{(i+n')} &= \tilde{\mathbf{m}} - \sqrt{n' + \lambda'} \left[\sqrt{\tilde{\mathbf{P}}} \right]_i, \quad i = 1, \ldots, n', \end{aligned} \tag{3.85}$$

   où le paramètre λ' *est défini comme dans l'équation* (3.78)*, en remplaçant* n *par* n'*, ainsi la moyenne et la covariance augmentées sont définies par*

$$\tilde{\mathbf{m}} = \begin{pmatrix} \mathbf{m} \\ \mathbf{0} \end{pmatrix} \quad \tilde{\mathbf{P}} = \begin{pmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}.$$

2. *Propager les points sigma à travers la fonction :*

$$\tilde{\mathcal{Y}}^{(i)} = \mathbf{g}\left(\tilde{\mathcal{X}}^{(i),x}, \tilde{\mathcal{X}}^{(i),q} \right), \quad i = 0, \ldots, 2n',$$

   où $\tilde{\mathcal{X}}^{(i),x}$ *et* $\tilde{\mathcal{X}}^{(i),q}$ *désignent les parties du point sigma augmenté* i*, qui correspondent respectivement à* $\mathbf{x}$ *et* $\mathbf{q}$*.*

3. *Calculer la moyenne prédite* μ_U*, la covariance prédite* $\mathbf{S}_U$ *et la covariance croisée* $\mathbf{C}_U$ *:*

$$\begin{aligned} \mu_U &= \sum_{i=0}^{2n'} W_i^{(m)'} \tilde{\mathcal{Y}}^{(i)} \\ \mathbf{S}_U &= \sum_{i=0}^{2n'} W_i^{(c)'} (\tilde{\mathcal{Y}}^{(i)} - \mu_U)(\tilde{\mathcal{Y}}^{(i)} - \mu_U)^T \\ \mathbf{C}_U &= \sum_{i=0}^{2n'} W_i^{(c)'} (\tilde{\mathcal{X}}^{(i),x} - \mathbf{m})(\tilde{\mathcal{Y}}^{(i)} - \mu_U)^T, \end{aligned}$$

où les poids $W_i^{(m)'}$ *et* $W_i^{(c)'}$ *sont définis comme dans l'équation* (3.80), *en remplaçant* n *par* n' *et* λ *par* λ'.

3.2.4 Filtre de Kalman non parfumé (UKF)

Le filtre de Kalman non parfumé (UKF) ([Julier 1995] ; [Julier 2004] et [Haykin 2001]) est un algorithme de filtrage optimal qui utilise la transformée non parfumée pour approcher les modèles d'espace d'état non-linéaires de la forme (3.53) ou (3.66). La forme de l'approximation gaussienne est comme suit :

$$p(\mathbf{x}_k|\mathbf{y}_1,\ldots,\mathbf{y}_k) \approx \mathbf{N}(\mathbf{x}_k|\mathbf{m}_k,\mathbf{P}_k), \tag{3.86}$$

où $\mathbf{m}_k$ et $\mathbf{P}_k$ sont la moyenne et la covariance calculées par l'algorithme suivant :

Algorithme 3.10 (filtre de Kalman non parfumé I)**.** *L'algorithme de filtre de Kalman non parfumé (UKF) peut être appliqué aux modèles de la forme* (3.53), *les opérations suivantes sont effectuées à chaque étape de mesure* $k = 1, 2, 3, \ldots$ *:*

1. Etape de prédiction :

   (a) *Former les points sigma :*

$$\begin{aligned}
\mathcal{X}_{k-1}^{(0)} &= \mathbf{m}_{k-1}, \\
\mathcal{X}_{k-1}^{(i)} &= \mathbf{m}_{k-1} + \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}_{k-1}}\right]_i \\
\mathcal{X}_{k-1}^{(i+n)} &= \mathbf{m}_{k-1} - \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}_{k-1}}\right]_i, \quad i = 1,\ldots,n
\end{aligned} \tag{3.87}$$

   où le paramètre λ *est défini dans l'équation* (3.78).

   (b) *Propager les points sigma à travers le modèle dynamique :*

$$\hat{\mathcal{X}}_k^{(i)} = \mathbf{f}(\mathcal{X}_{k-1}^{(i)}), \quad i = 0,\ldots,2n. \tag{3.88}$$

   (c) *Calculer la moyenne prédite* $\mathbf{m}_k^-$ *et la covariance prédite* $\mathbf{P}_k^-$ *:*

$$\begin{aligned}
\mathbf{m}_k^- &= \sum_{i=0}^{2n} W_i^{(m)} \hat{\mathcal{X}}_k^{(i)}, \\
\mathbf{P}_k^- &= \sum_{i=0}^{2n} W_i^{(c)} (\hat{\mathcal{X}}_k^{(i)} - \mathbf{m}_k^-)(\hat{\mathcal{X}}_k^{(i)} - \mathbf{m}_k^-)^T + \mathbf{Q}_{k-1},
\end{aligned} \tag{3.89}$$

   où les poids $W_i^{(m)}$ *et* $W_i^{(c)}$ *ont été définis dans l'équation* (3.80).

2. Etape de mis à jour :

(a) Former les points sigma :

$$\begin{aligned}\mathcal{X}_k^{-(0)} &= \mathbf{m}_k^-,\\ \mathcal{X}_k^{-(i)} &= \mathbf{m}_k^- + \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}_k^-}\right]_i\\ \mathcal{X}_k^{-(i+n)} &= \mathbf{m}_k^- - \sqrt{n+\lambda}\left[\sqrt{\mathbf{P}_k^-}\right]_i, \quad i = 1,\ldots,n.\end{aligned} \tag{3.90}$$

(b) Propager les points sigma à travers le modèle de mesure :

$$\hat{\mathcal{Y}}_k^{(i)} = \mathbf{h}(\mathcal{X}_k^{(i)}), \quad i = 0,\ldots,2n. \tag{3.91}$$

(c) Calculer la moyenne prédite μ_k, la covariance prédite de la mesure $\mathbf{S}_k$, et la covariance croisée de l'état et la mesure $\mathbf{C}_k$:

$$\begin{aligned}\mu_k &= \sum_{i=0}^{2n} W_i^{(m)} \hat{\mathcal{Y}}_k^{(i)}\\ \mathbf{S}_k &= \sum_{i=0}^{2n} W_i^{(c)} (\hat{\mathcal{Y}}_k^{(i)} - \mu_k)(\hat{\mathcal{Y}}_k^{(i)} - \mu_k)^T + \mathbf{R}_k\\ \mathbf{C}_k &= \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_k^{-(i)} - \mathbf{m}_k^-)(\hat{\mathcal{Y}}_k^{(i)} - \mu_k)^T.\end{aligned} \tag{3.92}$$

(d) Calculer le gain du filtre $\mathbf{K}_k$, la moyenne $\mathbf{m}_k$ et la covariance $\mathbf{P}_k$ de l'état filtrée :

$$\begin{aligned}\mathbf{K}_k &= \mathbf{C}_k \mathbf{S}_k^{-1}\\ \mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k[\mathbf{y}_k - \mu_k]\\ \mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T.\end{aligned} \tag{3.93}$$

Les équations du filtre ci-dessus peuvent être dérivées de manière similaire aux équations d'EKF, mais on utilise la transformée non-parfumée à la place des approximations linéaires.

La forme non additive d'UKF [Julier 2004] peut être dérivée en augmentant le bruit de processus ou de mesure avec le vecteur d'état et en appliquant une approximation UT à celle-ci. Alternativement, on peut d'abord augmenter le vecteur d'état avec le bruit de processus, puis l'approximer dans l'étape de prédiction, ensuite, on fait la même chose avec le bruit de mesure dans l'étape de mise à jour. Les différents algorithmes sont décrits dans l'article [Wu 2005]. Cependant, on applique séparément l'UT non additif aux étapes de prédiction et de mise à jour dans l'algorithme 3.9, nous obtenons l'algorithme suivant :

Algorithme 3.11 (filtre de Kalman non parfumé II)**.** *Dans la forme augmentée de l'algorithme du filtre de Kalman non parfumé (UKF), qui peut être appliqué aux modèles non additifs de la forme* (3.66)*, les opérations suivantes sont effectuées à chaque étape de mesure* $k = 1, 2, 3, \ldots$ *:*

1. Etape de prédiction :

(a) Former les points sigma pour la variable aléatoire augmentée $(\mathbf{x}_{k-1}, \mathbf{q}_{k-1})$ *:*

$$\begin{aligned} \tilde{\mathcal{X}}_{k-1}^{(0)} &= \tilde{\mathbf{m}}_{k-1}, \\ \tilde{\mathcal{X}}_{k-1}^{(i)} &= \tilde{\mathbf{m}}_{k-1} + \sqrt{n' + \lambda'} \left[\sqrt{\tilde{\mathbf{P}}_{k-1}} \right]_i \\ \tilde{\mathcal{X}}_{k-1}^{(i+n')} &= \tilde{\mathbf{m}}_{k-1} - \sqrt{n' + \lambda'} \left[\sqrt{\tilde{\mathbf{P}}_{k-1}} \right]_i, \quad i = 1, \ldots, n', \end{aligned} \tag{3.94}$$

où

$$\tilde{\mathbf{m}}_{k-1} = \begin{pmatrix} \mathbf{m}_{k-1} \\ \mathbf{0} \end{pmatrix} \quad \tilde{\mathbf{P}}_{k-1} = \begin{pmatrix} \mathbf{P}_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{k-1} \end{pmatrix}.$$

Ici $n' = n + n_q$*, où* n *est la dimensionnalité de l'état* $\mathbf{x}_{k-1}$ *et* n_q *est la dimensionnalité du bruit* $\mathbf{q}_{k-1}$*. Le paramètre* λ' *est défini comme dans l'équation* (3.78)*, mais avec* n *remplacé par* n'*.*

(b) Propager les points sigma à travers le modèle dynamique :

$$\hat{\mathcal{X}}_k^{(i)} = \mathbf{f}(\tilde{\mathcal{X}}_{k-1}^{(i),x}, \tilde{\mathcal{X}}_{k-1}^{(i),q}), \quad i = 0, \ldots, 2n'. \tag{3.95}$$

où $\tilde{\mathcal{X}}_{k-1}^{(i),x}$ *désigne les premiers* n *composants dans* $\tilde{\mathcal{X}}_{k-1}^{(i)}$ *et* $\tilde{\mathcal{X}}_{k-1}^{(i),q}$ *désigne les derniers* n_q *composants.*

(c) Calculer la moyenne prédite $\mathbf{m}_k^-$ *et la covariance prédite* $\mathbf{P}_k^-$ *:*

$$\begin{aligned} \mathbf{m}_k^- &= \sum_{i=0}^{2n} W_i^{(m)'} \hat{\mathcal{X}}_k^{(i)} \\ \mathbf{P}_k^- &= \sum_{i=0}^{2n} W_i^{(c)'} (\hat{\mathcal{X}}_k^{(i)} - \mathbf{m}_k^-)(\hat{\mathcal{X}}_k^{(i)} - \mathbf{m}_k^-)^T. \end{aligned} \tag{3.96}$$

où les poids $W_i^{(m)'}$ *et* $W_i^{(c)'}$ *sont définis comme dans l'équation* (3.80)*, en remplaçant* n *par* n' *et* λ *par* λ'*.*

2. Etape de mis à jour :

(a) Former les points sigma pour la variable aléatoire augmentée $(\mathbf{x}_k, \mathbf{r}_k)$ *:*

$$\begin{aligned}
\tilde{\mathcal{X}}_k^{-(0)} &= \tilde{\mathbf{m}}_k^-, \\
\tilde{\mathcal{X}}_k^{-(i)} &= \tilde{\mathbf{m}}_k^- + \sqrt{n'' + \lambda''}\left[\sqrt{\tilde{\mathbf{P}}_k^-}\right]_i \\
\tilde{\mathcal{X}}_k^{-(i+n'')} &= \mathbf{m}_k^- - \sqrt{n'' + \lambda''}\left[\sqrt{\tilde{\mathbf{P}}_k^-}\right]_i, \quad i = 1, \ldots, n'',
\end{aligned} \tag{3.97}$$

où

$$\tilde{\mathbf{m}}_k^- = \begin{pmatrix} \mathbf{m}_k^- \\ \mathbf{0} \end{pmatrix} \quad \tilde{\mathbf{P}}_k^- = \begin{pmatrix} \mathbf{P}_k^- & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_k \end{pmatrix}.$$

Ici $n'' = n + n_r$*, où* n *est la dimensionnalité de l'état* $\mathbf{x}_k$ *et* n_r *est la dimensionnalité du bruit* $\mathbf{r}_k$*. Le paramètre* λ'' *est défini comme dans l'équation* (3.78)*, en remplaçant* n *par* n''*.*

(b) Propager des points sigma à travers le modèle de mesure :

$$\hat{\mathcal{Y}}_k^{(i)} = \mathbf{h}(\tilde{\mathcal{X}}_k^{-(i),x}, \tilde{\mathcal{X}}_k^{-(i),r}), \quad i = 0, \ldots, 2n'', \tag{3.98}$$

où $\tilde{\mathcal{X}}_k^{-(i),x}$ *désigne les premiers* n *composants dans* $\tilde{\mathcal{X}}_k^{-(i)}$ *et* $\tilde{\mathcal{X}}_k^{-(i),r}$ *désigne les derniers* n_r *composants.*

(c) Calculer la moyenne prédite μ_k*, la covariance prédite de la mesure* $\mathbf{S}_k$*, et la covariance croisée de l'état et la mesure* $\mathbf{C}_k$ *:*

$$\begin{aligned}
\mu_k &= \sum_{i=0}^{2n''} W_i^{(m)''} \hat{\mathcal{Y}}_k^{(i)} \\
\mathbf{S}_k &= \sum_{i=0}^{2n''} W_{i-1}^{(c)''} (\hat{\mathcal{Y}}_k^{(i)} - \mu_k)(\hat{\mathcal{Y}}_k^{(i)} - \mu_k)^T \\
\mathbf{C}_k &= \sum_{i=0}^{2n''} W_i^{(c)''} (\mathcal{X}_k^{-(i),x} - \mathbf{m}_k^-)(\hat{\mathcal{Y}}_k^{(i)} - \mu_k)^T,
\end{aligned} \tag{3.99}$$

où les poids $W_i^{(m)''}$ *et* $W_i^{(c)''}$ *sont définis comme dans l'équation* (3.80)*, en remplaçant* n *par* n'' *et* λ *par* λ''*.*

(d) Calculer le gain du filtre $\mathbf{K}_k$*, la moyenne de l'état filtrée* $\mathbf{m}_k$ *et la covariance* $\mathbf{P}_k$ *:*

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{C}_k \mathbf{S}_k^{-1} \\
\mathbf{m}_k &= \mathbf{m}_k^- + \mathbf{K}_k[\mathbf{y}_k - \mu_k] \\
\mathbf{P}_k &= \mathbf{P}_k^- - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^T.
\end{aligned} \tag{3.100}$$

3.3 Conclusion

L'avantage de l'UKF par rapport à l'EKF est que l'UKF n'est pas basé sur l'approximation linéaire locale, il utilise aussi un nombre important de points pour l'approximation de la non-linéarité. La transformée non-parfumée est capable de capturer les moments d'ordre supérieur causés par la transformée non-linéaire mieux que les approximations de la série de Taylor [Julier 2004]. Cependant, c'est que l'estimation moyenne d'UT soit exacte pour les polynômes jusqu'à l'ordre 3, le calcul de covariance n'est pas exact que pour les polynômes du premier ordre. Dans UT, les fonctions dynamique et modélisée ne doivent pas formellement différentiables et les matrices jacobiennes ne doivent pas calculées. L'inconvénient de l'UKF par rapport l'EKF est que l'UKF nécessite souvent un coût de calcul important que l'EKF.

L'UKF appartient à une plus large classe des filtres, c'est les filtres des points sigma [[Van Der Merwe 2004], qui inclut également d'autres types de filtres tels que le filtre de Kalman des différences centrales (CDKF), le filtre de Kalman de Gauss-Hermite (GHKF) et quelques autres ([Ito 2000], [Wu 2006], [NøRgaard 2000], [Arasaratnam 2009]).

La figure 3.1 représente des spectrogrammes pour un exemple d'application du filtre de Kalman étendue et non parfumé sur un signal de parole bruité. Dans le chapitre 6, on va voir une nouvelle extension pour régler le problème de modélisation de l'UKF et l'EKF sur l'amélioration du signal de la parole en utilisant le perceptron multicouche.

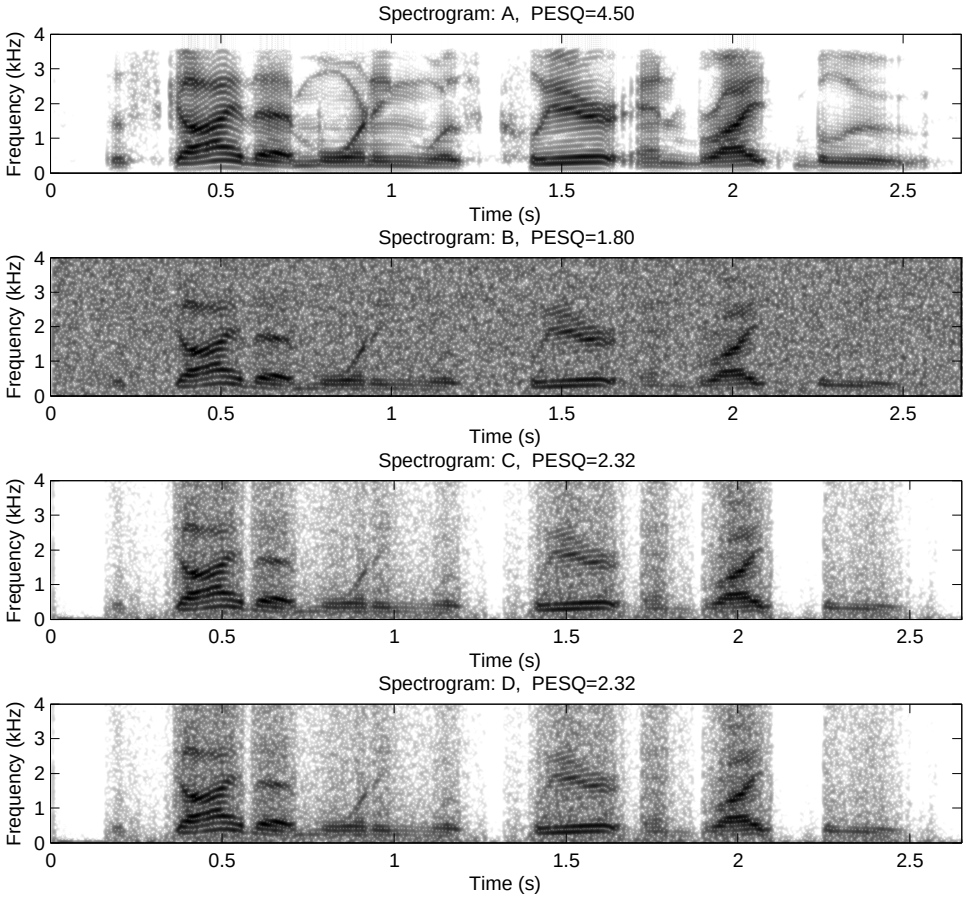

FIGURE 3.1 – Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase bruitée par un bruit blanc gaussien à SNR =5 dB ; (C) la phrase traitée par l'EKF ; (D) la phrase traitée par l'UKF.

Chapitre 4

Amélioration du signal de parole par le filtre de Kalman en utilisant la technique d'amélioration des formants

Sommaire

4.1 Prétraitement de la parole 51
4.2 L'analyse LPC 51
4.3 Détermination du spectre d'amplitude 53
4.4 L'identification des formants et des vallées 54
4.4.1 Détection des segments non voisés 55
4.4.2 Détermination des amplitudes et des emplacements des formants 56
4.4.3 Trouver les vallées 57
4.5 Modification du spectre d'amplitude 58
4.6 Calcul les coefficients du post-filtre 60
4.7 Réévaluer les nouveaux LPCs 61
4.8 Conclusion 61

Le filtre de Kalman est un estimateur non biaisé de l'erreur quadratique moyenne minimal (MMSE) linéaire dans le domaine temporel et qui provient de la théorie des systèmes de contrôle [6]. Son rôle est d'estimer les états inconnus d'un système dynamique, en utilisant une combinaison linéaire d'un bruit corrompu par des observations et des états prédits. Le filtre de Kalman présente un intérêt particulier pour l'amélioration de la parole en raison de plusieurs avantages par rapport aux méthodes d'amélioration du domaine spectral :

— Le modèle de la production de la parole se situe dans les équations de filtre Kalman en utilisant un prédicteur linéaire comme un modèle dynamique ;

— Lorsque des LPCs précis sont disponibles, le signal de parole amélioré par le filtre de Kalman ne contient aucun bruit de musique ;

— Le filtre de Kalman ne fait aucune hypothèse stationnaire comme le filtre de Wiener ;

— Le filtre de Kalman peut être activé au premier échantillon n = 0, où les paramètres de récursivité sont initialisés avec leurs valeurs espérées ;

— Le filtre de Kalman est considéré comme un estimateur commun de l'amplitude et le spectre de phase à la fois [7].

Le type de filtre-source le plus utilisé est le modèle de prédiction linéaire (LP). Qui est utilisé pour le codage [8], la reconnaissance [9] et l'amélioration [10] de la parole. Les LPCs sont trouvés par l'estimation du modèle LPC, qui décrit la fonction de transfert inverse du conduit vocal humain.

La performance d'amélioration du filtre de Kalman est un dépendant de la précision des LPCs et de la variance d'excitation. Idéalement, ces coefficients doivent être obtenus à partir d'un signal de parole propre, comme cela a été fait par [5]. Cependant, dans la pratique, les LPCs et la variance ne sont pas connues a priori ; elles doivent donc être estimées à partir d'un signal bruité. En fonction des caractéristiques de bruit et du rapport signal sur bruit (SNR), les LPCs et la variance d'excitation d'un signal de parole bruité sont médiocres. Pour cette raison, la plupart des méthodes proposées se concentrent sur les méthodes d'estimation des LPCs et de la variance d'excitation. Ainsi, dans [11], les auteurs développent une méthode alternative d'amélioration de la parole itérative sous-optimale en utilisant l'algorithme Expectation-Maximization (EM). Dans [12], l'auteur propose une méthode d'estimation des LPCs en utilisant l'algorithme LS récursif robuste. En outre, les LPCs sont estimés en utilisant la sortie améliorée du filtre de Kalman de l'itération précédente [13,14]. Cependant, ces derniers permettent d'obtenir des SNR inférieurs. Dans cette thèse, nous proposons une nouvelle technique présentant la méthode d'amélioration des formants basée sur LPC (LPC-FEM) pour améliorer la structure du spectre de la parole bruitée dans le filtre de Kalman itératif.

La méthode proposée repose sur la modification du spectre d'amplitude logarithmique du modèle LPC, puis sur la réévaluation de nouveaux LPCs pour réduire la présence du bruit de fond dans le filtre de Kalman itératif. Les améliorations apportées par cette méthode sont vérifiées dans le cadre d'expériences d'amélioration de la parole objectives et subjectives en utilisant le corpus NOIZEUS. Nous montrons que la méthode d'amélioration proposée (Kalman LPC-FEM) donne de meilleures performances que les techniques de filtrage de Kalman itératifs conventionnels [13,14].

Dans ce chapitre, nous présentons notre méthode (où notre contribution) qui concerne la réévaluation des LPCs en utilisant la méthode d'amélioration des formants (FEM). L'objectif est d'implémenter ces nouveaux LPCs dans le filtre de Kalman itératif pour améliorer le signal de parole bruité.

Les avantages de cette méthode par rapport aux autres sont les suivants :

1. Résoudre le problème des mauvaises estimations des LPCs en raison du bruit additif.
2. Le signal de parole amélioré par notre méthode ne contient pas des bruits résiduels (musiques) parce que les performances d'optimisation du filtre de Kalman sont étroitement liées à la précision des LPCs.

L'algorithme de la méthode d'amélioration des formants basée sur LPC dans le filtre de Kalman (Kalman LPC-FEM) est implémenté à travers les étapes suivantes :

4.1 Prétraitement de la parole

Il est effectué par le filtre suivant

$$H_p(z) = 1 - \beta z^{-1} \tag{4.1}$$

Avec β choisi empiriquement, le rôle du filtre de prétraitement est la suppression de l'inclinaison spectrale dans le spectre de la parole et met l'accent sur les plus élevés formants de fréquence pour voir une analyse LPC plus précise.

4.2 L'analyse LPC

Les trames de parole de 20 ms sont extraites par la fenêtre de Hanning [Blackman 1958]. Nous avons arbitrairement choisi cette fenêtre qui est définie par

$$\omega_n = \begin{cases} 0.5 - 0.5\cos\frac{2\pi n}{N}, & 0 \leq n \leq N-1 \\ 0 & autrement \end{cases} \tag{4.2}$$

Le modèle de filtre source dans lequel le filtre est contraint à être un filtre linéaire tout-pôles. Les quantités pour effectuer une prédiction linéaire de l'échantillon suivant en tant que somme pondérée des échantillons passés sont :

$$\hat{y}(n) = -\sum_{k=1}^{p} a_k y(n-k) \tag{4.3}$$

où la variable entière n est l'indice temporel discret, $\hat{y}$ est la prédiction de la parole bruitée y, et a_k sont les LPCs.

L'erreur de prédiction $u(n)$, définie comme étant la différence entre la valeur de l'échantillon actuel y et sa valeur prédite $\hat{y}$ est donnée par

$$\begin{aligned} u(n) &= y(n) - \hat{y}(n) \\ &= y(n) + \sum_{k=1}^{p} a_k y(n-k) \end{aligned} \tag{4.4}$$

A partir de l'équation (4.4) le signal généré ou modélisé par la prédiction linéaire peut être décrit par l'équation de réaction suivante

$$y(n) = -\sum_{k=1}^{p} a_k y(n-k) + u(n) \tag{4.5}$$

Le problème est la détermination des LPCs a_k du signal de parole bruité $y(n)$.

La solution optimale pour obtenir les LPCs en minimisant l'erreur quadratique moyenne de la prédiction, en calculant le gradient de l'erreur quadratique moyenne de la prédiction par rapport au vecteur des LPCs $\mathbf{a}$ et en calculant l'erreur moyenne des moindres carrés :

$$\mathbf{a} = -\mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{\mathbf{yy}} \tag{4.6}$$

où :

$$\mathbf{a}^T = [a_1, a_2, \ldots, a_p]$$

$$\mathbf{r}_{\mathbf{yy}} = [r_{yy}(1), r_{yy}(2), \ldots, r_{yy}(p)]^T$$

$$\mathbf{R}_{\mathbf{yy}} = \begin{pmatrix} r_{yy}(0) & r_{yy}(1) & r_{yy}(2) & \cdots & r_{yy}(p-1) \\ r_{yy}(1) & r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(p-2) \\ r_{yy}(2) & r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{yy}(p-1) & r_{yy}(p-2) & r_{yy}(p-3) & \cdots & r_{yy}(0) \end{pmatrix}$$

où $\mathbf{R}_{\mathbf{yy}} = \mathrm{E}(\mathbf{y}\mathbf{y}^T)$ est la matrice d'autocorrélation du vecteur d'entrée $\mathbf{y}^T = [y(n-1), y(n-2), \ldots, y(n-p)]$, $\mathbf{r}_{\mathbf{yy}} = \mathrm{E}(y(m)\mathbf{y})$ est le vecteur d'autocorrélation et $\mathbf{a}^T = [a_1, a_2, \ldots, a_p]$ est le vecteur des LPCs.

Notez que le matrice d'autocorrélation $p \times p$ est symétrique et les éléments de toute la diagonale sont identiques (c.-à-d., une matrice de Toeplitz). La

méthode efficace pour résoudre l'équation (4.6) est l'algorithme de Levinson-Durbin [Durbin 1960]. La procédure récursive de Durbin est spécifiée comme suit :

pour $i = 1, \ldots, p$

$$\mathrm{E}^{(0)} = r_{yy}(0) \tag{4.7}$$

$$k_i = -[r_{yy}(i) + \sum_{k=1}^{i-1} a_k^{(i-1)} r_{yy}(i-k)]/E^{i-1} \tag{4.8}$$

$$a_i^{(i)} = -k_i \tag{4.9}$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \qquad 1 \leq j \leq i-1 \tag{4.10}$$

$$\mathrm{E}^{(i)} = (1 - k_i^2)\mathrm{E}^{(i-1)} \tag{4.11}$$

4.3 Détermination du spectre d'amplitude

La transformée en z de l'équation (4.5) montre que le modèle LP est un filtre numérique tous-pôles avec la fonction de transfert suivante

$$K(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k z^{-k}} \tag{4.12}$$

L'approximation du spectre de parole est obtenue en calculant le logarithme du spectre d'amplitude de $K(z)$. La première étape est la détermination des LPCs. Dans chaque séquence de parole, on utilise la méthode d'autocorrélation. Cette méthode garantit que $K(z)$ est une phase minimale [Rabiner 1978]. En outre, une fenêtre de Hanning est préférable à une fenêtre rectangulaire, ce qui conduit à une meilleure approximation du spectre de parole dans lequel les pics des formants sont plus évidents ([Rabiner 1978] ; [O'shaughnessy 1987]). Etant donné que l'objectif est l'estimation des emplacements et des amplitudes des formants. Les expériences ont montré qu'une analyse d'ordre supérieur aboutit à trop de pics dans le spectre, ce qui rend difficile l'identification des emplacements des formants, tandis qu'une analyse d'ordre inférieur ne donne pas assez de pics pour résoudre 4 formants.

$K(f)$, $f = 0$ à 255 est la FFT de $N_{pt} = 256$ points de la séquence $[1, a_1, a_2, \ldots, a_p]$. La réponse en fréquence d'un modèle LP est donnée par

$$K(f) = \frac{1}{1 + \sum_{k=1}^{p} a_k e^{-j2\pi kf}} \tag{4.13}$$

où f est l'indice discrète dans le domaine fréquentiel. L'approximation du spectre de parole est obtenue en calculant le logarithme du spectre d'amplitude de $K(f)$:

$$R(f) = 20 \log |K(f)| \tag{4.14}$$

Ceci est une représentation discrète d'une approximation du spectre de parole qui est utilisée ensuite pour identifier les formants.

4.4 L'identification des formants et des vallées

La détection de l'amplitude et l'emplacement des formants est une étape importante dans la détermination des LPCs améliorés. L'extraction de formant pour un signal de parole propre est plus simple que pour notre signal bruité ([Markel 1972], [McCandless 1974]). Néanmoins, dans la pratique, le signal propre n'est pas disponibles, où les paramètres LPC sont obtenus à partir d'un signal de parole bruitée $y(n)$, qui est le seul signal observable dans la pratique. Puisque les autres détecteurs des formants ne sont pas conçus pour utiliser sur la parole bruitée, nous étions incapables de les appliquer sur notre signal et faire une comparaison formelle. Plus récemment, les difficultés de l'estimation des formants à partir d'un signal de parole bruité ont commencé à être abordées. Plusieurs méthodes tentent ceci en divisant le spectre en plusieurs segments, chacun contenant un formant. Les limites de segment peuvent être obtenues grâce à une programmation dynamique [Welling 1996] ou filtrage de Wiener [Chen 2004]. Par exemple, [Bruce 2002] focalise sur l'estimation de chaque formant séparément par le filtrage adaptatif. Le filtrage de Kalman peut être utilisé pour augmenter la précision de la prédiction des formants en utilisant le suivi dans le temps [Yan 2005].

Dans le travail présenté ici, nous nous sommes basés sur la stratégie de cueillette de pic de Kabal [Kabal 1991] pour détecter les formants ensuite déterminer les amplitudes et les emplacements des vallées dans la parole bruitée. Cela est nécessaire pour modifier le logarithme du spectre d'amplitude $R(f)$ pour devenir $S(f)$ de telle sorte les pics des formants sont aiguisés et les vallées spectrales sont approfondies (voir la Figure 4.1). La Figue 4.2 montre l'effet de la méthode d'amélioration des formants (FEM) proposée. A partir le calcul de $R(f)$, on a calculé de façon séquentielle les maximums locaux et choisir uniquement les pics correspondant aux formants. Dans chaque séquence de parole

il y a au maximum quatre formants. Deux problèmes majeurs se produisent avec l'approche de cueillette de pic là où certains pics peuvent être faux et deux formants peuvent apparaître comme un seul pic [Loizou 2007]. Le premier problème doit être résolu en choisissant un pic correspondant un formant afin d'éviter de classer les pics parasites comme des formants. Le deuxième problème n'est pas important puisque les pics fusionnés sont aiguisés par le post-filtre.

Pour chaque séquence de parole, nous déterminons le niveau d'énergie maximale (A_{max}) et son emplacement (L_{max}) dans la gamme de $f = 0$ à $N_{pt}/2$, où $A_{max} = \max(R(f))$ et N_{pt} est le nombre de points $R(f)$ dans la séquence. Ensuite, nous calculons le niveau de bruit (N_{AV}), ainsi que ses approximations de la moyenne L_{AV}. Dans notre base de donnée, les cinq premières séquences du signal vocal correspondent un silence (c.-à-d. bruit pur). Pour chacune de ces séquences on calcule $A(m)$ comme la somme des amplitudes des pics $R(f)$ divisé par le nombre de pics. Puis, N_{AV} est la moyenne des cinq valeurs de $A(M)$.

Nous allons maintenant analyser chaque séquence du signal de parole bruité et dans ce cas nous effectuons trois tâches. Tout d'abord, chaque séquence est classé comme étant non voisée ($L_U = 1$) ou voisée ($L_U = 0$). Deuxièmement, on détermine les amplitudes des formants et leurs positions. En troisième lieu, on trouve les amplitudes et les positions des vallées spectrales. Ces trois tâches sont décrites ci-dessous.

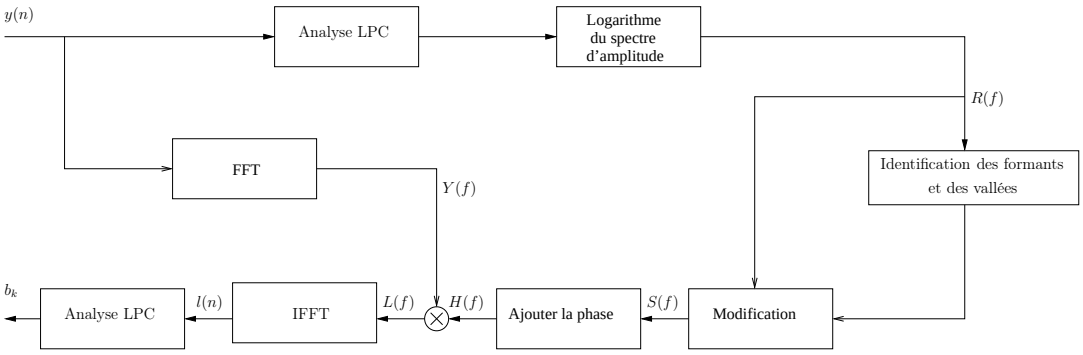

FIGURE 4.1 – Montrant la méthode d'amélioration des formants (FEM).

4.4.1 Détection des segments non voisés

La méthode d'identification des séquences non voisées est donnée par l'algorithme suivant :

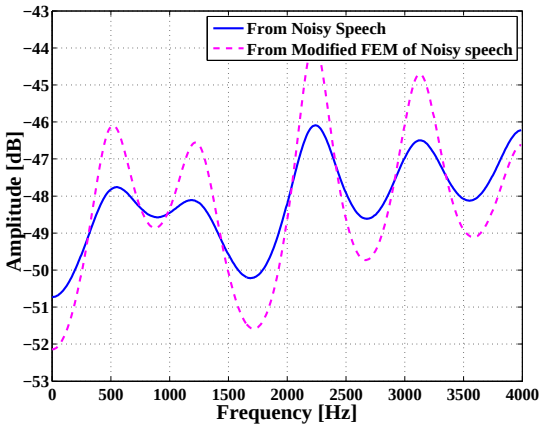

FIGURE 4.2 – Enveloppes spectrales d'une trame voisée de parole bruitée par un bruit blanc de SNR = 5 dB et du spectre de parole améliorée par FEM.

1. Définir la séquence non voisée $L_U = 1$.
2. Si le niveau d'énergie maximale $A_{max} < 2_{AV}$, puis la séquence voisée $L_U = 0$; Arrêtez.
3. Si la position $L_{max} < M_{pt}/4$, puis la séquence voisée $L_U = 0$; Arrêtez.
4. Calculer A_{MX} (cette quantité est discuté plus tard).
5. Si $A_{MX} < N_{AV}/2$, puis $L_U = 0$; Arrêtez.

La quantité A_{MX} est calculé en divisant les premiers échantillons de $R(f)$ en quatre portions. Pour chaque partie, nous calculons S_{R1}, S_{R2}, S_{R3} et S_{R4}, puis, $A_{MX} = \max(S_{R1}, S_{R2}, S_{R3}, S_{R4})$, où les S_{Rn} sont les valeurs moyennes de $R(f)$ dans chaque portion.

$$\begin{aligned} S_{R1} &= \sum_{k=0}^{\frac{N_{pt}}{8}-1} R(f), \quad S_{R1} = \sum_{k=\frac{N_{pt}}{8}}^{\frac{N_{pt}}{4}-1} R(f), \\ S_{R1} &= \sum_{k=\frac{N_{pt}}{4}}^{\frac{3N_{pt}}{8}-1} R(f), \quad S_{R1} = \sum_{k=\frac{3N_{pt}}{8}}^{\frac{N_{pt}}{2}-1} R(f), \end{aligned} \tag{4.15}$$

4.4.2 Détermination des amplitudes et des emplacements des formants

Le problème est séquentiellement pour : examiner chaque valeur de $R(f)$, localiser un pic, et décider si il est un formant. La détection de Formant est

invoquée lorsqu'un pic local $R(f)$ se trouve sous la contrainte $R(f) > R_{min}$. Deux solutions émergent :

1. Si plus d'un formant a été trouvé, la partie 1 est invoquée,
2. Si au plus un formant a été détectée ; seule la partie 2 de l'algorithme est invoqué.

Dans chaque trame de parole, on obtient N_F formants. Les amplitudes de ces formants sont placées dans la matrice $A_p(J)$. Les indices des positions de $R(f)$ à laquelle ces formants se produisent sont stockées dans la rangée $N_p(J)$.

— Algorithme de détection du Formant

Partie 1 :

1. Si $A_{max} \geq C_2 N_{AV}$ et $A_{max}/A_p(2) \leq C_1$, un formant est trouvé. Sinon, passez à la partie 2

Partie 2 :

1. Si les deux conditions ci-dessous sont remplies, un formant est détecté. Sinon, passez à l'étape (2).
   $A_{max} > C_3 N_{AV}$ et $A_{max}/R(f) \leq C_5$
2. Si $A_{max} \leq C_3 N_{AV}$, puis un formant se trouve si les deux conditions ci-dessous sont remplies. Sinon, passez à l'étape (3).
   $A_{max} > C_1 N_{AV}$ et $A_{max}/R(f) \leq C_4$
3. Si $A_{max} \leq C_3 N_{AV}$ et $A_{max} \leq C_1 N_{AV}$, puis un formant est trouvée si $T_U = 1$ (une trame non voisée). Dans le cas contraire, le pic local dans $R(f)$ ne correspond pas un formant.

Les seuils C_1, C_2 et C_3 utilisés pour comparer A_{max} et N_{AV}, ces seuils sont empiriquement choisi en fonction du SNR. Comme le SNR augmente, les valeurs de C_1, C_2 et C_3 diminues. Les valeurs de C_4 et C_5 sont principalement choisis pour détecter le deuxième formant et parfois utilisés pour établir les formants d'ordre supérieur.

4.4.3 Trouver les vallées

Etant donné les amplitudes des formants $A_p(1), \ldots, A_p(N_F)$ et leurs positions correspondants $N_p(1), \ldots, N_p(N_F)$, nous procédons à déterminer les amplitudes et les positions des vallées. Cela est nécessaire pour aiguiser les formants et approfondir les vallées avec le post-filtre. Entre deux positions du formant $N_p(J)$ et $N_p(J+1)$, le minimum local $R(f)$ correspond à une vallée, comme représenté sur la Figure 4.3. L'amplitude de cette vallée est $A_v(J)$ et sa position est $N_v(J)$. De cette manière, $N_F - 1$ vallées se trouvent.

4.5 Modification du spectre d'amplitude

Le spectre d'amplitude $R(f)$ est modifié pour devenir $S(f)$ de telle sorte que dans le signal de parole post-filtré les pics de formants sont aiguisés, les vallées spectrales sont approfondies et aucun signal non désiré est présent. La première étape consiste à diviser $R(f)$ en sous-séquences de $f = 0$ à $N_p(1)$, $N_p(1)$ à $N_p(2)$, ..., et enfin $N_p(Nf)$ à $N_{pt}/2$. Chaque sous-séquence est modifiée individuellement. Cette liberté de modifier indépendamment les différentes séquences de $R(f)$ est l'avantage de cette approche (dans le domaine fréquentiel) par rapport la méthode dans le domaine temporel.

Tout d'abord nous allons concentrer sur les sections de $R(f)$ qui correspondent à des formants actuels, c'est à dire, à partir de $f = N_p(1)$ à $N_p(Nf)$. La Figure 4.3 représente les formes générales des enveloppes $R(f)$ et $S(f)$ dans une de ces sections (de $f = N_p(J)$ à $N_p(J+1)$). Cette section est divisée en deux sous-sections de $f = N_p(J)$ à $N_v(J)$ et de $f = N_v(J)$ à $N_p(J+1)$ (marqué par (A) et (B) sur la Figure 4.3). Dans la sous-section (A), les deux points terminaux de $S(f)$ sont affectés comme suit :

$$\begin{aligned} S(N_p(J)) &= A_{max}, \\ S(N_v(J)) &= R(N_v(J)) + \tau A_{max} \\ &= A_v(J) + \tau A_{max} \end{aligned} \tag{4.16}$$

La réponse en fréquence du post-filtre aura des pics d'amplitude égale A_{max} aux fréquences des formants. Le facteur $\tau < 0$ reflète à quel point nous voulons approfondir les vallées. En outre, τ dépend du SNR en ce que pour un SNR élevé, une grande valeur de τ est utilisée. Pour un SNR de 10dB, on utilise $\tau = -0.05$. Soit $\Delta(N_p(J))$ le changement de $R(f)$ à $f = N_p(J)$. Alors, $\Delta(N_p(J)) = S(N_p(J)) - R(N_p(J)) = A_{max} - A_p(J)$. De même, $\Delta(N_v(J))$ est le changement de $R(f)$ à $f = N_v(J)$. Alors, $\Delta(N_v(J)) = S(N_v(J)) - R(N_v(J)) = \tau A_{max}$. Les changements de $R(f)$ aux points intermédiaires $N_p(J) < f < N_v(J)$ sont calculés par l'interpolation linéaire entre les valeurs $\Delta(N_p(J))$ et $\Delta(N_v(J))$. Plus précisément, $\Delta(f)$, qui est la variation de $R(f)$ pour $N_p(J) < f < N_v(J)$, est donnée par

$$\Delta(f) = \frac{(\Delta(N_v(J)) - \Delta(N_p(J)))(f - N_p(J))}{N_v(J) - N_p(J)} + \Delta(N_p(J)). \tag{4.17}$$

Ainsi, $S(f) = R(f) + \Delta(f)$. Une procédure similaire pour fixer les points d'extrémité de $S(f)$ et interpoler linéairement les changements $\Delta(f)$ entre les deux points terminaux est adoptée pour la sous-section (B) de la figure 4.3. De cette manière, toutes les sections de $R(f)$ correspondant aux formants réels sont modifiées pour donner $S(f)$.

Nous devons maintenant considérer les régions de $0 \leq f \leq N_p(1)$ et $N_p(N_F) \leq f \leq N_{pt}/2$. Tout d'abord, considérons le cas $0 \leq f \leq Np(1)$. Si on a un minimum local de $R(f)$ à $N_v(0)$, on divise cette région en deux sous-sections, à savoir (A) de $f = 0$ à $N_v(0)$ et (B) de $f = N_v(0)$ à $N_p(1)$. Pour la sous-section (A), $R(f)$ est modifié de telle sorte que la vallée à $N_v(0)$ soit approfondie et telle qu'il n'y ait pas de changement à $f = 0$ ($S(0) = R(0) = A_p(0)$) . C'est parce que le pic à $f = 0$ n'est pas un vrai formant. Les points intermédiaires sont modifiés en calculant $\Delta(f)$ comme décrit ci-dessus. La modification de $R(f)$ pour la sous-section (B) suit la méthode générale décrite ci-dessus. Ainsi, pour le premier formant à $f = N_p(1), S(N_p(1)) = A_{max}$. Si $R(f)$ augmente de façon monotone dans $f = 0$ à $N_p(1)$, alors cette entière section est modifiée par la méthode générale décrite ci-dessus de sorte que la vallée à $f = 0$ est approfondie et l'amplitude du formant atteint à $f = N_p(1)$ est égal A_{max}.

Maintenant, nous examinons la région $N_p(N_F) \leq f \leq N_{pt}/2$. Si $R(f)$ diminue de façon monotone dans cette région, nous suivons la méthode générale décrite ci-dessus pour obtenir $S(f)$. Dans le cas contraire, deux sous-sections sont formées : (A) de $f = N_p(N_F)$ à $N_v(N_F)$ et (B) $f = N_v(N_F)$ à $N_{pt}/2$. Pour la sous-section (A), nous suivons la méthode générale pour obtenir $S(f)$ telle que la vallée à $N_v(N_F)$ soit approfondie. La sous-section (B) correspond à des fréquences relativement élevées. Le spectre d'amplitude à ces fréquences devrait être désaccentué pour atténuer les effets du bruit hautes fréquences. Par conséquent, $S(f) = R(f) - 10$ pour la sous-section (B).

La procédure pour obtenir $S(f)$ de $R(f)$ a été donnée. Cependant, il y a deux cas particuliers dans lesquels nous nous écartons de la procédure générale. Tout d'abord, supposons que l'on rencontre une trame dans laquelle le signal de parole est très fort par rapport au niveau de bruit, le premier pic de formant ayant la plus grande amplitude ($A_p(1) = A_{max}$). Ensuite, certaines des vallées peuvent également avoir une amplitude élevée par rapport au niveau de bruit. Dans un tel cas, nous n'avons pas à approfondir ces vallées. Si $A_v(0) >$ le pic de formant de l'amplitude la plus faible, alors $S(f) = R(f)$ pour $f = 0$ à $N_p(1)$ (aucun changement dans le spectre). De même, si $A_v(1) >$ le pic de formant la plus faible amplitude, alors $S(f) = R(f)$ pour $f = N_p(1)$ à $N_v(1)$. Cependant, la section de $f = N_v(1)$ à $N_p(2)$ est modifiée pour affiner le second formant et conserver l'amplitude à $f = N_v(1)$. De cette manière, les amplitudes de chaque vallée sont vérifiées pour décider si ces vallées doivent ou non être approfondies. Cette flexibilité de pouvoir conserver certaines parties du spectre du signal où le faible niveau de bruit est un avantage de l'approche du domaine fréquentiel.

Un second cas particulier se produit lorsqu'une trame est non voisée ($L_v = 1$) et $A_p(N_F) = A_{max}$. Disposant du plus grand pic à une fréquence élevée est

assez courant pour les trames non voisées. En particulier, certaines trames non voisées ont seulement un large formant de résonance à une fréquence relativement élevée. Si l'algorithme de détection des formants ne donne qu'un seul formant, on procède simplement à obtenir $S(f)$ de $R(f)$. Si $N_F > 1$, les amplitudes des autres formants doivent être testées avant le post-filtrage. Si les amplitudes de ces formants sont très faibles, il y a beaucoup de bruit aux basses fréquences qui seraient améliorées par le post-filtre. Pour éviter cela, nous faisons ce qui suit. Supposons $N_F = 2$ et $A_p(1) < 2N_{AV}$. Le premier pic du formant a une faible amplitude est rejeté, ce qui amène N_F à 1. Si $N_F > 2$, on teste les deux premiers pics ayant des amplitudes $A_p(1)$ et $A_p(2)$. Si $A_p(1) < 2N_{AV}$, le pic à $N_p(1)$ est rejeté. De même, si $A_p(2) < 2N_{AV}$, le pic à $N_p(2)$ est rejeté.

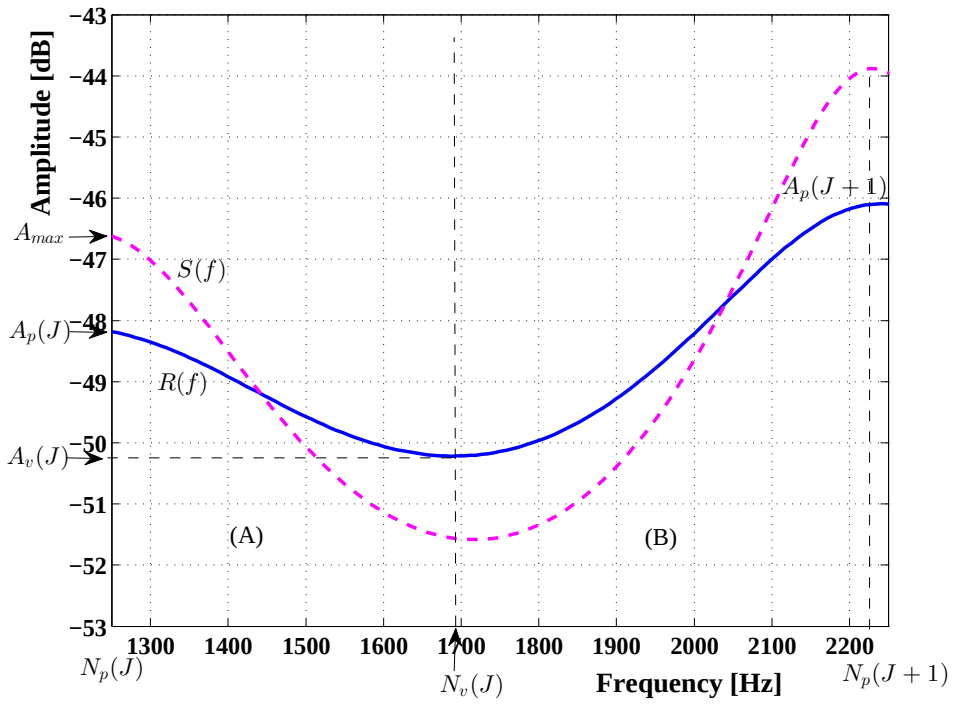

FIGURE 4.3 – Les enveloppes d'une section entre deux formants.

4.6 Calcul les coefficients du post-filtre

Maintenant, les coefficients du post-filtre $H(f)$ doivent être déterminés à partir le spectre d'amplitude logarithmique modifiée $S(f)$. Notez que $S(f)$

est une représentation de $|H(f)|$, à savoir, $S(f) = 20\log|H(f)|$, par conséquent, $|H(f)| = 10^{S(f)/20}$. La phase de $H(f)$ est exactement la même que celle la phase de $K(f)$. Etant donné que la phase de $k(f)$ est $\theta(f)$, $H(f) = |H(f)|e^{j\theta(f)}$. Les coefficients de post-filtre sont obtenus en ne modifiant que l'amplitude du spectre LPC. Le composant de la phase reste inaltéré.

4.7 Réévaluer les nouveaux LPCs

Dans ces cas, les coefficients de FFT $L(f)$ du signal de sortie $l(n)$ sont déterminées comme $L(f) = H(f)Y(f)$, où $Y(f)$ est la FFT de $y(n)$ (voir la Figure 4.1), ainsi nous avons utilisé la méthode d'analyse LPC du signal de parole $l(n)$ pour déterminer les nouveaux LPCs améliorés b_k :

$$l(n) = -\sum_{k=1}^{p} b_k l(n-k) + \mathbf{u}(n) \tag{4.18}$$

En outre, nous avons utilisé ces nouveaux coefficients (b_k) au lieu de a_k dans la matrice de transition de filtre de Kalman.

4.8 Conclusion

Nous avons proposé une nouvelle technique de filtre de Kalman basé sur la méthode d'amélioration des formants en utilisant dans le rehaussement de la parole corrompue par des bruits blancs gaussiens additifs et colorés. Nous avons montré que la méthode d'amélioration des formants (FEM) tend à améliorer la plage dynamique de la structure de la parole bruitée. Les performances du modèle de filtrage de Kalman proposé ont été comparées à d'autres applications. Les résultats de plusieurs expériences ont montré que le modèle proposé (Kalman LPC-FEM) fournit des scores PESQ, SNR et SegSNR plus élevés et un lissage plus précis que les autres méthodes.

Chapitre 5

Amélioration du signal de parole par les filtres de Kalman non linéaire

Sommaire

5.1 Introduction 63

5.2 Amélioration du signal de parole à l'aide de l'entraînement au EKF d'un MLP 64

5.2.1 Filtre de Kalman étendu - Estimation d'Etat 65

5.2.2 Filtre de Kalman étendu - Estimation de Poids 66

5.3 Amélioration du signal de parole à l'aide de l'entraînement au UKF d'un MLP 67

5.4 Conclusion 71

5.1 Introduction

Les approches traditionnelles de suppression du bruit dans la parole impliquent des techniques spectrales, ce qui entraîne souvent une distorsion audible du signal. Des méthodes récentes de filtrage non linéaire dans le domaine temporel utilisent des ensembles de données où le signal de parole propre est disponible en tant que signal cible pour entrainer un réseau de neurone. De telles méthodes sont souvent efficaces dans l'ensemble de l'apprentissage, mais n'ont pas efficace pour les sources réelles avec les différents niveaux de signal et de bruit. En outre, les modèles de réseau dans ces méthodes ne tiennent pas pleinement compte de la nature non-stationnaire de la parole. Dans l'approche présentée ici, nous ne supposons que la disponibilité du signal bruité. En effet, une séquence de perceptron multicouche est entraînée par un signal de parole bruité, résultant un modèle non-stationnaire qui peut être utilisé pour éliminer le bruit du signal.

Un signal de parole bruitée $y(k)$ peut être modélisé avec précision comme une auto-régression non linéaire avec le processus et le bruit d'observation additif :

$$x(k) = f(x(k-1)), \ldots x(k-M), \mathbf{w}) + u(k) \tag{5.1}$$

$$y(k) = x(k) + v(k) \tag{5.2}$$

où $x(k)$ est le signal de parole réelle tirée par le bruit de processus $u(k)$, et $f(\cdot)$ est une fonction non linéaire des valeurs passées de $x(k)$ paramétrées par $\mathbf{w}$. La parole est seulement supposé être stationnaire au-dessus de courts segments, chaque segment ayant un modèle différent. L'observation disponible est $y(k)$, qui contient un bruit additif $v(k)$. L'estimateur optimal compte tenu l'observation bruitée $\mathbf{y}(k) = \{y(k), y(k-l), \cdots y(0)\}$ est $\mathrm{E}[x(k)|(y(k)]$. La façon la plus directe pour estimer cela serait de former un ensemble de données propres dans lequel le vrai $x(k)$ peut être utilisé comme une cible de réseau de neurone. Notre hypothèse est que le signal propre non disponible ; le but est d'estimer $x(k)$ à partir d'un signal de parole bruité $y(k)$.

Pour résoudre ce problème, nous supposons que $f(\cdot, \cdot)$ est dans la classe des modèles de perceptron multicouche à action-directe et calculer l'estimation double des états $\hat{x}$ et des poids $\hat{w}$ basée sur l'approche de filtre de Kalman. Dans ce chapitre nous fournissons une description fondamentale de l'algorithme.

5.2 Amélioration du signal de parole à l'aide de l'entraînement au EKF d'un MLP

En posant le problème d'estimation double dans un cadre d'espace-état, nous pouvons utiliser les méthodes de filtrage de Kalman pour effectuer l'estimation d'une manière récursive et efficace. A chaque point de temps, le filtre de Kalman fournit une estimation optimale en combinant une prédiction à priori avec une nouvelle observation. Connor et al. [Connor 1994] a proposé d'utiliser un filtre de Kalman étendu (EKF) avec un réseau de neurones pour effectuer une seul estimation d'état. [Puskorius 1994] et d'autres ont posé l'estimation du poids dans un cadre d'espace d'état pour un entrainement efficace du filtre de Kalman d'un réseau de neurones. Le travail présenté ici développe ces idées et le perceptron multicouche dans le contexte de traitement de la parole.

La formule d'espace d'état de (5.1) et (5.2) se présente comme suit :

$$\mathbf{x}(k) = F[\mathbf{x}(k-1)] + Gu(k) \tag{5.3}$$

$$y(k) = H\mathbf{x}(k) + v(k) \tag{5.4}$$

où

$$\mathbf{x}(k) = \begin{bmatrix} x(k) \\ x(k-1) \\ \vdots \\ x(k-M+1) \end{bmatrix}, \quad F[\mathbf{x}(k)] = \begin{bmatrix} f(x(k), \ldots, x(k-M+1), \mathbf{w}) \\ x(k) \\ \vdots \\ x(k-M+2) \end{bmatrix}$$
$$C = [1, 0, \ldots, 0], \qquad G = C^T \tag{5.5}$$

Si le modèle est linéaire, donc $f(\mathbf{x}(k))$ prend la forme $\mathbf{w}^T\mathbf{x}(k)$, et $F[\mathbf{x}(k)]$ peut être écrit comme $A\mathbf{x}(k)$, où A est en forme canonique contrôlable. Nous supposons au départ que les termes de bruit $u(k)$ et $v(k)$ sont blancs avec des variances connues σ_u^2 et σ_v^2, respectivement. Les méthodes pour estimer les variances de bruit directement à partir des données bruitées sont décrites plus tard dans ce chapitre.

5.2.1 Filtre de Kalman étendu - Estimation d'Etat

Pour un modèle linéaire avec des paramètres connus, l'algorithme de filtre de Kalman (KF) peut être facilement utilisé pour estimer les états [Lewis 1986]. A chaque pas de temps, le filtre calcule l'estimation linéaire des moindres carrés $\hat{x}(k)$ et la prédiction $\hat{x}^-(k)$, ainsi que leurs covariances d'erreur, $P_{\mathbf{x}}(k)$ et $P_{\mathbf{x}}^-(k)$. Dans le cas linéaire avec les statistiques gaussiennes, les estimations sont quadratiques moyennes minimales. Sans informations à priori sur x, elles sont réduites les estimations du maximum de vraisemblance.

Lorsque le modèle est non linéaire, le filtre de Kalman (KF) ne peut pas être appliqué directement mais nécessite une linéarisation du modèle non linéaire à chaque pas de temps. L'algorithme s'appelle le filtre de Kalman étendu (EKF), qui est rapproche effectivement la fonction non linéaire avec une fonction linéaire variante dans le temps. L'algorithme d'EKF est le suivant :

$$\begin{aligned}
\hat{\mathbf{x}}^-(k) &= F[\hat{\mathbf{x}}(k-1), \hat{\mathbf{w}}(k-1)] \\
\mathbf{P}_{\hat{\mathbf{x}}}^-(k) &= \mathbf{A}\mathbf{P}_{\hat{\mathbf{x}}}(k-1)\mathbf{A}^T + G\sigma_u^2 G^T, \quad o\grave{u} \quad \mathbf{A} = \left.\frac{\partial F[\hat{\mathbf{x}}, \hat{\mathbf{w}}]}{\partial \hat{\mathbf{x}}}\right|_{\hat{\mathbf{x}}(k-1)} \\
\mathbf{K}(k) &= \mathbf{P}_{\hat{\mathbf{x}}}^-(k)C^T(C\mathbf{P}_{\hat{\mathbf{x}}}^-(k)C^T + \sigma_v^2)^{-1} \\
\mathbf{P}_{\hat{\mathbf{x}}}(k) &= (I - \mathbf{K}(k)C)\mathbf{P}_{\hat{\mathbf{x}}}^-(k) \\
\hat{\mathbf{x}}(k) &= \hat{\mathbf{x}}^-(k) + \mathbf{K}(k)(y(k) - C\hat{\mathbf{x}}^-(k)).
\end{aligned} \tag{5.6}$$

5.2.2 Filtre de Kalman étendu - Estimation de Poids

Lorsque le modèle de la parole n'est pas connu, l'algorithme d'EKF standard ne peut pas être appliqué directement. Nous abordons ce problème en construisant une formule d'espace d'état séparée pour les poids sous-jacents comme suit :

$$\mathbf{w}(k) = \mathbf{w}(k-1) \tag{5.7}$$

$$y(k) = f(\mathbf{x}(k-1)), \mathbf{w}(k)) + u(k) + v(k) \tag{5.8}$$

où la transition d'état est simplement une matrice identité et le perceptron multicouche $f(\mathbf{x}(k-1)), \mathbf{w}(k))$ joue le rôle d'une observation non linéaire variée dans le temps sur $\mathbf{w}$. Ces équations d'état pour les poids nous permettent de les estimer avec un seconde EKF.

$$\begin{aligned}
\hat{\mathbf{w}}^{-}(k) &= \hat{\mathbf{w}}(k-1) \\
\mathbf{P}^{-}_{\hat{\mathbf{w}}}(k) &= \mathbf{P}_{\hat{\mathbf{w}}}(k-1) \\
\mathbf{K}_{\hat{\mathbf{w}}}(k) &= \mathbf{P}^{-}_{\hat{\mathbf{w}}}(k) H(k)^T (H(k)\mathbf{P}^{-}_{\hat{\mathbf{w}}}(k) H(k)^T + \sigma_v^2 + \sigma_u^2)^{-1} \\
\mathbf{P}_{\hat{\mathbf{w}}}(k) &= (I - \mathbf{K}_{\hat{\mathbf{w}}}(k) H(k)) \mathbf{P}^{-}_{\hat{\mathbf{w}}}(k) \\
\hat{\mathbf{w}}(k) &= \hat{\mathbf{w}}^{-}(k) + \mathbf{K}_{\hat{\mathbf{w}}}(k)(y(k) - CF(\hat{\mathbf{x}}(k-1), \hat{\mathbf{w}}^{-}(k))).
\end{aligned} \tag{5.9}$$

$$où \quad H(k) = \left. \frac{C \partial F[\hat{\mathbf{x}}, \hat{\mathbf{w}}]}{\partial \hat{\mathbf{w}}} \right|_{\hat{\mathbf{w}}(k-1)} \tag{5.10}$$

La linéarisation de (5.10) peut être calculée en tant que dérivée dynamique [Werbos 1990] pour tenir compte de la nature récurrente du filtre d'estimation d'état, comprenant la dépendance du gain de Kalman $K(k)$ sur les poids. Le calcul de ces dérivées est coûteux et peut être évité en ignorant la dépendance de $\hat{\mathbf{x}}(k)$ sur $\hat{\mathbf{w}}$ [1]. Cette approximation a été utilisée pour produire les résultats.

Nous avons maintenant deux EKF exécutés en parallèle pour estimer à la fois les états $\mathbf{x}$ et les poids $\mathbf{w}$. A chaque pas de temps, l'estimation actuelle de $\mathbf{x}$ est utilisée par le filtre de poids et l'estimation actuelle de $\mathbf{w}$ est utilisée par le filtre d'état. Pour les ensembles des données finis, l'algorithme est exécuté itérativement jusqu'à ce que les poids convergent.

Cette approche de double estimation est liée aux travaux effectués par Nelson [Nelson 1976] dans le cas linéaire, et l'approche de neurone de Matthews [Matthews 1994] pour l'algorithme d'erreur de prédiction récursive [Goodwin 2014] [2].

1. Cela équivalent à une seule étape de propagation inverse à travers le temps [Werbos 1990].

2. Une approche alternative consiste à concaténer $\mathbf{w}$ et $\mathbf{x}$ dans un vecteur d'état commun, et appliquer l'EKF aux équations d'état non linéaires (voir [Goodwin 2014] pour le cas

Dans la littérature, la méthode est étroitement liée à l'approche de Lim et Oppenheim pour adapter les modèles LPC à la parole dégradée [Lim 1978]. Elle se rapporte également à l'approche basée sur le modèle d'Ephraim [Ephraim 1992], mais utilise une estimation non linéaire pour adapter les données au lieu d'utiliser un nombre fixe de modèles linéaires pré-spécifiés.

5.3 Amélioration du signal de parole à l'aide de l'entraînement au UKF d'un MLP

Comme mentionné précédemment, l'UKE peut être utilisé pour remplacer l'approche EKF ci-dessus. L'UKF est une extension directe de la transformation non parfumée (UT) à l'estimation récursive. L'UT est une méthode de calcul des statistiques d'une variable aléatoire qui subit une transformation non linéaire. Il est décrit ci-dessous en résumé.

Etant donné une variable aléatoire $\mathbf{x}$ de L dimension avec la moyenne $\bar{\mathbf{x}}$ et la matrice de covariance $\mathbf{P_x}$, les statistiques d'une variable aléatoire $\mathbf{y} = g(\mathbf{x})$ peuvent être calculées par la procédure suivante :

D'abord, former les $2L+1$ vecteurs sigma $\mathcal{X}_i$, avec les poids correspondants W_i, selon les règles suivantes :

$$\begin{aligned}
\mathcal{X}_0 &= \bar{\mathbf{x}} \\
\mathcal{X}_i &= \bar{\mathbf{x}} + \left(\sqrt{(L+\lambda)+\mathbf{P_x}}\right)_i \quad i = 1, \cdots, L \\
\mathcal{X}_i &= \bar{\mathbf{x}} - \left(\sqrt{(L+\lambda)+\mathbf{P_x}}\right)_{i=L} \quad i = L+1, \cdots, 2L \\
W_0^{(m)} &= \lambda/(L+\lambda) \\
W_0^{(m)} &= \lambda/(L+\lambda) + (1-\alpha^2+\beta^2) \\
W_i^{(m)} &= W_i^{(c)} = 1/2(L+\lambda) \quad i = 1, \cdots, 2L
\end{aligned} \tag{5.11}$$

où $\lambda = \alpha^2(L+\gamma) - L$ est un paramètre d'échelle. α détermine la propagation des points sigma autour de $\bar{\mathbf{x}}$ et il est généralement défini sur une petite valeur positive. γ est un paramètre d'échelle secondaire qui est habituellement mis à 0, et β est utilisé pour incorporer la connaissance à priori de la distribution de $\mathbf{x}$ (pour les distributions gaussiennes, $\beta = 2$ est optimal).

$\left(\sqrt{(L+\gamma)\mathbf{P}_x}\right)_i$ est la $i^{ème}$ ligne de la racine carrée de la matrice. Ces vecteurs sigma se propagent à travers la fonction non linéaire.

$$\mathcal{Y}_i = \mathbf{g}(\mathcal{X}_i) \quad i = 0, \cdots, 2L \tag{5.12}$$

linéaire, et [Williams 1992] pour une application aux réseaux neurones récurrents). Cet algorithme est connu des problèmes de convergence.

la moyenne et la covariance de $\mathbf{y}$ sont approximées par :

$$\bar{\mathbf{y}} \approx \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_i \tag{5.13}$$

$$\mathbf{P}_{\mathbf{y}} \approx \sum_{i=0}^{2L} W_i^{(c)} \{\mathcal{Y}_i - \bar{\mathbf{y}}\}\{\mathcal{Y}_i - \bar{\mathbf{y}}\}^T \tag{5.14}$$

Pour les entrées gaussiennes, les résultats d'UT peuvent atteindre la précision du troisième ordre. Pour les entrées non gaussiennes, les approximations sont précises au moins au second ordre. La précision des moments d'ordre supérieur et inférieur est déterminée par le choix de α et de β [Julier 1997].

Par conséquent, l'UKF est exécuté en appliquant le schéma de sélection de points sigma d'UT au vecteur d'état pour calculer la matrice sigma correspondante, comme indiqué dans l'équation (5.11).

Si les propriétés de masquage des systèmes auditifs humains sont prises en compte, le problème d'optimisation suivant doit être résolu pour obtenir un nouveau gain de Kalman :

$$\begin{aligned}
&\text{minimiser } \Delta_0 \\
&\text{sujet à} \\
&0 \leq \Delta_0 + 2\sum_{i=1}^{M-1} \Delta_i \cos(\frac{2\pi}{256}.i.m) \leq T(m), \quad m = 0, 1, \cdots, 128 \\
&|\Delta_0| \geq |\Delta_j| \\
&\Delta_0 > 0
\end{aligned} \tag{5.15}$$

avec :

$$\Delta_0 = \frac{1}{M}\left[\mathbf{T}_r\mathbf{P} - 2\mathbf{T}_r\mathbf{K}\mathbf{H}\mathbf{P} + \mathbf{T}_r\mathbf{K}(\mathbf{H}\mathbf{P}\mathbf{H}^T)\mathbf{K}^T + \mathbf{T}_r\mathbf{K}\mathbf{R}\mathbf{K}^T\right] \tag{5.16}$$

$$\Delta_j = \frac{1}{M-j}\left[\sum_{i=1}^{p-j} \gamma_{i,i+j} - (\mathbf{K}\mathbf{H}\mathbf{P})_{i,i+j} + -(\mathbf{P}\mathbf{H}^T\mathbf{K}^T)_{i,i+j} + (\mathbf{K}\mathbf{H}\mathbf{P}\mathbf{H}^T\mathbf{K}^T)_{i,i+j} + (\mathbf{K}\mathbf{R}\mathbf{K}^T)_{i,i+j}\right] \tag{5.17}$$

où $T(m)$ est le seuil de masquage total calculé par la méthode décrite dans [Ma 2004]. $\mathbf{T}_r$ est la trace d'une matrice, $\mathbf{P}$ et $\mathbf{K}$ représentent respectivement $P_{\hat{\mathbf{x}}}(k|k-1)$ et $\mathbf{K}(k)$, $\gamma_{i,i+j}$ est l'élément de $\mathbf{P}$ qui se trouve sur la $i^{ème}$ ligne et la $(i+j)^{ème}$ colonne. $(\mathbf{K}\mathbf{H}\mathbf{P})_{i,i+j}$, $(\mathbf{P}\mathbf{H}^T\mathbf{K}^T)_{i,i+j}$, $(\mathbf{K}\mathbf{H}\mathbf{P}\mathbf{H}^T\mathbf{K}^T)_{i,i+j}$

Algorithme 1 Les équations d'UKF à contrainte perceptuelle pour (5.6)

Initialiser avec :
$\hat{\mathbf{x}}(0|0) = \mathrm{E}[\mathbf{x}(0)]$
$\mathbf{P}_{\hat{\mathbf{x}}}(0|0) = \mathrm{E}[\mathbf{x}(0) - \hat{\mathbf{x}}(0|0)][\mathbf{x}(0) - \hat{\mathbf{x}}(0|0)]^T$
en commençant par $k = 1$. Calculer la matrice sigma :
$\mathcal{X}_0(k-1|k-1) = \hat{\mathbf{x}}(k-1|k-1)$
$\mathcal{X}_i(k-1|k-1) = \hat{\mathbf{x}}(k-1|k-1) + \left(\sqrt{(L+\lambda)\mathbf{P}_{\hat{\mathbf{x}}}(k-1|k-1)}\right)_i$
$\mathcal{X}_{i+L}(k-1|k-1) = \hat{\mathbf{x}}(k-1|k-1) - \left(\sqrt{(L+\lambda)\mathbf{P}_{\hat{\mathbf{x}}}(k-1|k-1)}\right)_i \quad i = 1, \cdots, L$ où L est égal à M.
Mise à jour du temps :
$\mathcal{X}_i(k|k-1) = \mathbf{F}\left[\mathcal{X}_i(k-1|k-1), \hat{\mathbf{w}}(k-1|k-1)\right] \quad i = 0, \cdots, 2L$
Vecteur d'état prédit :
$\hat{\mathbf{x}}(k|k-1) = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{X}_i(k|k-1)$
Matrice de covariance d'erreur d'état prédite :
$\mathbf{P}_{\hat{\mathbf{x}}}(k|k-1) = \sum_{i=0}^{2L} W_i^{(c)} \left[\mathcal{X}_i(k|k-1) - \hat{\mathbf{x}}(k|k-1)\right]\left[\mathcal{X}_i(k|k-1) - \hat{\mathbf{x}}(k|k-1)\right]^T + G\sigma_u^2 G^T$
Le gain de Kalman :
$\mathbf{K}(k)$: obtenu en résolvant l'équation (5.15)
Estimation filtrée du vecteur d'état :
$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)(\mathbf{y}(k) - \hat{\mathbf{x}}(k|k-1))$
Matrice de covariance d'erreur d'état filtrée :
$\mathbf{P}_{\hat{\mathbf{x}}}(k|k) = [\mathbf{I} - \mathbf{K}(k)]\mathbf{P}_{\hat{\mathbf{x}}}(k|k-1)[\mathbf{I} - \mathbf{K}(k)]^T + \mathbf{K}(k)\mathbf{R}\mathbf{K}^T(k)$

et $(\mathbf{KRK}^T)_{i,i+j}$ sont les éléments de la matrice correspondante qui sont sur la $i^{ème}$ ligne et la $(i+j)^{ème}$ colonne respectivement, et $\mathbf{H} = \mathbf{I}$.

Les équations d'UKF sont présentées dans l'Algorithme 1 :

Alors que les équations d'UKF pour (5.9) sont montrées dans l'Algorithme 2 :

Algorithme 2 Les équations d'UKF pour (5.9)

Initialiser avec :

$\hat{\mathbf{w}}(0|0) = \mathrm{E}[\mathbf{w}(0)]$

$\mathbf{P}_{\hat{\mathbf{w}}}(0|0) = \mathrm{E}[\mathbf{w}(0) - \hat{\mathbf{w}}(0|0)][\mathbf{w}(0) - \hat{\mathbf{w}}(0|0)]^T$

Vecteur de poids prédit :

$\hat{\mathbf{w}}(k|k-1) = \hat{\mathbf{x}}(k-1|k-1)$

Matrice de covariance d'erreur de poids prédite :

$\mathbf{P}_{\hat{\mathbf{w}}}(k|k-1) = \mathbf{P}_{\hat{\mathbf{w}}}(k-1|k-1)$

Calculer la matrice sigma :

$\omega_0(k-1|k-1) = \hat{\mathbf{w}}(k-1|k-1)$

$\omega_i(k-1|k-1) = \hat{\mathbf{w}}(k-1|k-1) + \left(\sqrt{(L+\lambda)\mathbf{P}_{\hat{\mathbf{w}}}(k-1|k-1)}\right)_i$

$\omega_{i+L}(k-1|k-1) = \hat{\mathbf{w}}(k-1|k-1) - \left(\sqrt{(L+\lambda)\mathbf{P}_{\hat{\mathbf{w}}}(k-1|k-1)}\right)_i \quad i = 1, \cdots, L$ où L est égal le nombre de poids.

Mise à jour du temps :

$\omega_i(k|k-1) = \omega_i(k-1|k-1)$

$\mathcal{Y}_i(k|k-1) = \mathbf{F}\left[\hat{\mathbf{x}}(k-1|k-1), \omega_i(k-1|k-1)\right], \quad i = 0, \cdots, 2L$

$\hat{\mathbf{y}}(k|k-1) = \sum_{i=0}^{2L} W_i^{(m)} \mathcal{Y}_i(k|k-1) + \bar{v}$

Equations de mise à jour de mesure :

$\mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k) = \sum_{i=0}^{2L} W_i^{(m)} \left[\mathcal{Y}_i(k|k-1) - \hat{\mathbf{y}}(k|k-1)\right]^T \left[\mathcal{Y}_i(k|k-1) - \hat{\mathbf{y}}(k|k-1)\right]$

$\mathbf{P}_{\hat{\mathbf{w}}\hat{\mathbf{y}}}(k) = \sum_{i=0}^{2L} W_i^{(c)} \left[\omega_i(k|k-1) - \hat{\mathbf{w}}(k|k-1)\right]^T \left[\omega_i(k|k-1) - \hat{\mathbf{w}}(k|k-1)\right]$

$\mathbf{T}(k) = \mathbf{P}_{\hat{\mathbf{w}}\hat{\mathbf{y}}}(k)\mathbf{P}^T_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k)$

Vecteur de poids filtré :

$\hat{\mathbf{w}}(k|k) = \hat{\mathbf{w}}(k|k-1) + \mathbf{T}(k)(\mathbf{y}(k) - \hat{\mathbf{y}}(k|k-1))$

Matrice de covariance d'erreur de poids filtré :

$\mathbf{P}_{\hat{\mathbf{w}}}(k|k) = \mathbf{P}_{\hat{\mathbf{w}}}(k|k-1) - \mathbf{T}(k)\mathbf{P}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}(k)\mathbf{T}^T(k)$

5.4 Conclusion

L'amélioration de la parole basée sur l'entraînement EKF et UKF d'un MLP sont proposées dans ce chapitre. Nous utilisons le perceptron multicouche pour modéliser le processus non linéaire du signal de parole. Par la suite, nous adoptons l'EKF et l'UKF pour estimer les états de la parole. Enfin, l'efficacité de la méthode développée est démontrée à travers des expériences simples.

Chapitre 6

Résultats et comparaison avec les techniques compétitive et conventionnels

Sommaire

6.1 La Qualité et l'intelligibilité de la parole 74
6.2 Les mesures objectives 74
6.2.1 Evaluation perceptuelle de la qualité de la parole (PESQ) 75
6.2.2 Rapport signal sur bruit segmental (SegSNR) 75
6.2.3 Analyse spectrographique des signaux de parole 76
6.3 Les mesures subjectives 77
6.3.1 Calcul du score SIG : degré de distorsion de parole (SDD) 79
6.3.2 Calcul du score BAK : rapport signal/PR-bruit (SPR) 81
6.3.3 Estimation des scores SIG, BAK et OVRL 83
6.4 Expérience d'amélioration de la parole 84
6.4.1 Dispositif expérimental 84
6.4.2 Résultats et discussion 85

Nous avons décrit de nombreux algorithmes d'amélioration de parole, mais on n'a pas discuté comment évaluer correctement leur performance. La performance peut être évaluée en utilisant soit des tests subjectifs d'écoute ou des tests de mesures objectives. L'évaluation de la qualité subjective, par exemple, implique des comparaisons de signaux de parole originaux et traités par un groupe d'auditeurs qui sont invités pour évaluer la qualité de la parole sur une échelle prédéterminée. L'évaluation objective implique une comparaison mathématique entre des signaux de parole originaux et traités. Les mesures objectives quantifient la qualité en mesurant la «distance» numérique entre les signaux de parole originaux et traités. De toute évidence, pour que la mesure objective soit valide, elle doit être bien corrélée avec les tests d'écoute subjectifs, pour cette raison, beaucoup de recherches ont été focalisées sur le développement des mesures objectives qui modélisent les différents aspects du système auditif.

Ce chapitre se concentre sur les mesures objectives et les mesures subjectives d'écoute qui ont été utilisées pour évaluer les algorithmes d'amélioration de la parole en termes de qualité et d'intelligibilité. Bien que traditionnellement certaines de ces procédures ont été utilisées pour évaluer les algorithmes de codage de la parole, ces méthodes ont également été utilisées pour quantifier les performances des algorithmes d'amélioration de la parole [Quackenbush 1988].

6.1 La Qualité et l'intelligibilité de la parole

Les différentes méthodes d'évaluation sont utilisées pour évaluer la qualité et l'intelligibilité de la parole traitée. La qualité est très subjective dans la nature et elle est difficile à évaluer de façon fiable. Ceci est dû à ce que les auditeurs individuels ont différentes normes auditives de ce qui constitue une «bonne» ou une «mauvaise» qualité. Les mesures de la qualité évaluent «comment» un locuteur produit un énoncé et comment comprendre des attributs tels que «Naturel», «enroué», «rugueux», et etc. La qualité possède de nombreuses dimensions à énumérer. Pour des raisons pratiques, nous nous limitons habituellement à seulement quelques dimensions de la qualité de la parole en fonction de l'application. Les mesures d'intelligibilité évaluent ce que le locuteur a dit, qui est, la signification ou le contenu des paroles prononcées. Contrairement à la qualité, l'intelligibilité n'est pas subjective et elle peut être facilement mesurée par la prononciation d'une phrase ou des mots d'un groupe d'auditeurs en leurs demandant d'identifier les mots prononcés. L'intelligibilité est quantifiée en comptant le nombre des mots ou des phonèmes identifiés correctement.

La relation entre l'intelligibilité et la qualité de la parole n'est pas entièrement compris, et cela est en partie dû au fait que les corrélats acoustiques de la qualité et de l'intelligibilité ne sont pas encore correctement identifié [Voiers 1980]. Le signal de parole peut être très intelligible, tout en étant de mauvaise qualité. Inversement, la parole peut avoir une bonne qualité, mais n'est pas complètement intelligible. Différentes méthodes doivent être utilisées pour évaluer la qualité et l'intelligibilité de la parole améliorée.

6.2 Les mesures objectives

Un nombre de mesures objectives sont examinées dans la présente étude pour prédire l'intelligibilité de la parole dans le bruit. Certaines des mesures objectives (par exemple, PESQ) ont été utilisés avec succès pour l'évaluation

de la qualité de la parole (par exemple [Quackenbush 1986] et [Rix 2001]), tandis que d'autres sont plus appropriées pour l'évaluation de l'intelligibilité.

6.2.1 Evaluation perceptuelle de la qualité de la parole (PESQ)

Parmi toutes les mesures objectives considérées, la mesure PESQ est la plus compliquée, elle est recommandée par [Recommendation 2001] pour l'évaluation de la qualité du signal de la parole de 3.2 kHz ([Rix 2001] ; [Recommendation 2001]). La mesure de PESQ est calculée comme suit :

Les signaux propres et dégradés sont mis au même niveau d'écoute standard, ainsi ils sont filtrés par un filtre à réponse similaire à celle d'un téléphone standard. Les signaux sont alignés dans le temps pour corriger les retards, puis ils sont traitées par une transformée pour obtenir les spectres de l'intensité sonore. La différence de volume entre les signaux originaux et dégradées est calculée dans le domaine temporel et dans le domaine fréquentiel pour trouver la prédiction subjective de la qualité. Le PESQ produit un score entre 1,0 et 4,5, avec les valeurs élevées indiquent une meilleure qualité. Les corrélations élevées ($r > 0,92$) avec des tests d'écoute subjectifs ont été rapportés par [Rix 2001] en utilisant la mesure de PESQ pour un grand nombre de conditions de test à partir l'application de protocole voix sur Internet. Forte corrélation ($r \approx 0,9$) a également été signalée dans [Hu 2008] avec les jugements de la qualité subjective du signal bruité par des algorithmes d'amélioration.

Le score final de PESQ est calculé comme une combinaison linéaire de la valeur de perturbation moyenne symétrique d_{sym} et la valeur de perturbation moyenne asymétrique d_{asym} comme suit :

$$PESQ = 4.5 - 0.1.d_{sym} - 0.0309.d_{asym} \tag{6.1}$$

Pour le matériel de test subjective normal, les valeurs se situent entre 1.0 (mauvais) et 4.5 (pas de distorsion).

6.2.2 Rapport signal sur bruit segmental (SegSNR)

Le rapport signal sur bruit segmentaire peut être évalué soit dans le domaine temporel ou bien dans le domaine fréquentiel. La mesure dans le domaine temporel est peut-être l'un des plus simples mesure objective utilisée pour évaluer les algorithmes d'amélioration de la parole. Pour que cette mesure soit efficace, il est important que les signaux propres et traités soient alignés dans le temps et que les erreurs de phase soient corrigées. Le rapport signal sur bruit segmentaire (SegSNR) est défini comme suit :

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1} log_{10}\frac{\sum_{n=mN}^{mN+N-1} x^2(n)}{\sum_{n=mN}^{mN+N-1}(x(n)-\hat{x}(n))^2} \tag{6.2}$$

où
$x(n)$ est le signal d'origine (propre)
$\hat{x}(n)$ est le signal amélioré
N est la longueur de trame (généralement choisie pour être de 15 à 20 ms)
M est le nombre de trames dans le signal

A noter que la mesure SegSNR est basée sur la moyenne géométrique des rapports signal sur bruit dans l'ensemble des trames du signal de parole.

Un problème potentiel avec l'estimation de SegSNR est que l'énergie du signal pendant les intervalles de silence dans le signal de parole (qui sont abondants dans la parole conversationnelle) soit très faible résultat dans des grandes valeurs de SegSNR négatifs, qui va polariser la mesure globale. Une façon de remédier à cela est d'exclure les cadres silencieuses de la somme dans l'équation 6.2 en comparant les mesures d'énergie de courte durée à un seuil. Dans [Hansen 1998], les valeurs SegSNR ont été limitées dans la gamme de [-10, 35 dB] ce qui évite la nécessité d'un détecteur de parole / silence.

6.2.3 Analyse spectrographique des signaux de parole

La transformée de Fourier de courte durée (STFT) est souvent utilisée pour l'analyse spectrographique des signaux de parole. Le spectrogramme est un affichage graphique du spectre de puissance de la parole en fonction du temps qui est donné par

$$S(n,\omega) = |X(n,\omega)|^2 \tag{6.3}$$

où $X(n,\omega)$ désigne la STFT du signal de parole $x(n)$. La quantité $S(n,\omega)$ peut être considérée comme une «densité spectrale de puissance» bidimensionnelle, la seconde dimension étant le dans temps. Le spectrogramme décrit la concentration d'énergie relative du signal vocal en fréquence en fonction du temps qu'il reflète les propriétés variant dans le temps de la forme d'onde du signal de parole. Les spectres sont généralement affichés en échelle de gris (voir l'exemple de la figure 6.1).

En fonction de la longueur de la fenêtre utilisée dans le calcul de $S(n,\omega)$, deux types de spectrogrammes peuvent être produits (bande étroite et large bande). Une fenêtre de longue durée est typiquement utilisée dans le calcul du spectrogramme à bande étroite et une fenêtre de courte durée est utilisée dans le calcul du spectrogramme à large bande. Le spectrogramme à

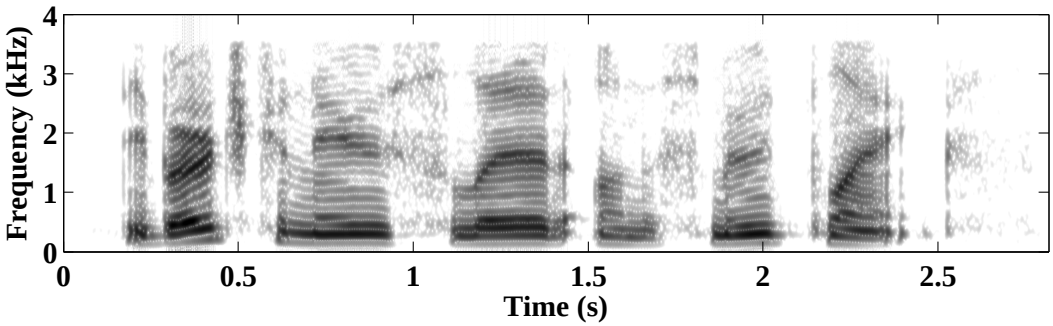

FIGURE 6.1 – Spectrogramme de la phrase propre (sp01.wav) 'The birch canoe slid on the smooth planks' a été prise de la base de données NOIZEUS.

bande étroite donne une bonne résolution en fréquence mais une mauvaise résolution temporelle. La résolution en fréquence permet de résoudre les harmoniques individuelles de la parole. Ces harmoniques apparaissent comme des stries horizontales dans le spectrogramme (voir l'exemple de la figure 6.2a). L'inconvénient principal de l'utilisation de longues fenêtres est la possibilité d'étaler temporellement des segments de parole de courte durée, tels que les consonnes. Le spectrogramme à large bande utilise des fenêtres de courte durée (inférieures de la période de la fréquence fondamental) qui donne une bonne résolution temporelle mais une mauvaise résolution fréquentielle. La principale conséquence cette mauvaise résolution est le frottement des harmoniques individuelles dans le spectre de la parole, ne donnant que l'enveloppe spectrale du spectre (voir l'exemple de la figure 6.2b). Le spectrogramme de large bande offre une bonne résolution temporelle, ce qui le rend approprié pour l'analyse de tous les sons de la langue anglaise. Le spectrogramme à large bande est un outil précieux pour l'analyse de la parole, et nous l'utiliserons dans tout le texte.

6.3 Les mesures subjectives

L'évaluation subjective comprend un test d'écoute informelle qui est conçu pour suivre la procédure proposée dans [Rix 2001] et [Recommendation 2001]. En exigeant cette méthode pour évaluer la parole de test en termes de la distorsion de la parole (SIG), l'intrusion du bruit de fond (BAK) et la qualité globale (OVRL), elle réduit le problème d'incertitude de l'auditeur de se qualifier la parole résultante en raison de la suppression du bruit au détriment de la distorsion de la parole dans les algorithmes d'amélioration de la parole. Une échelle de cinq points de la notation est utilisée. Un score moyen de dix

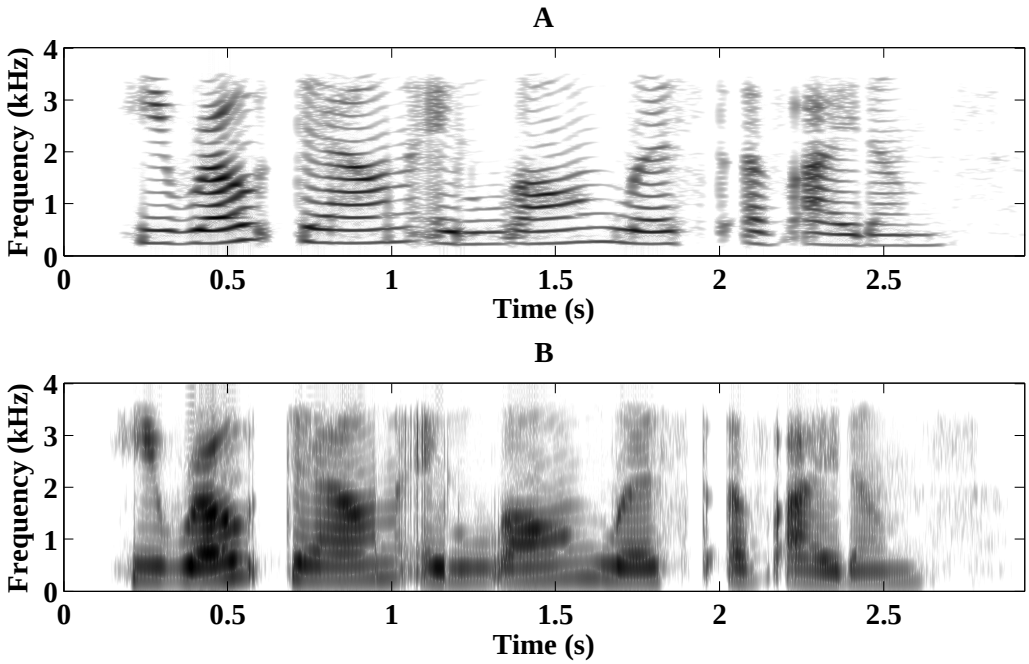

FIGURE 6.2 – Spectrogrammes : **A** à bande étroite et **B** à large bande de la phrase propre (sp11.wav) ' He wrote down a long list of items' a été prise de la base de données NOIZEUS.

phrases vocales corrompus avec bruit de voiture (5 par sexe masculin et 5 par sexe féminin) sont extraits au hasard à partir la base des données NOIZEUS de SNR = 5 dB. Puis un score moyen de dix phrases vocales corrompus par bruit blanc gaussien additif (5 par sexe masculin et 5 par sexe féminin) de SNR=5 dB sont extraits au hasard.

Le signal résiduel est la différence entre la parole traitée $\hat{s}(m)$ et le discours propre $s(m)$ où m est l'indice temporel, qui est utilisé pour l'analyse de corrélation avec les jugements humains. Il est défini comme suit :

$$e(m) = \hat{s}(m) - s(m) \tag{6.4}$$

Le signal traité $\hat{s}(m)$ est supposé être obtenu par des algorithmes de filtrage, qui prend beaucoup de temps pour aligner avec la parole propre $s(m)$. Le signal résiduel et la parole propre sont tronqués par des fenêtres. Ce paramétrage est le même que celui de [Hu 2008]. Pour chaque trame, les analyses suivantes sont appliquées.

6.3.1 Calcul du score SIG : degré de distorsion de parole (SDD)

Il est idéal que le filtre d'amélioration de la parole n'utilise pas de tous les composants de la parole, ce qui signifie que le signal résiduel ne contient pas de composantes de parole pure. Toutefois, cela est irréalisable en pratique. Ainsi, le signal résiduel contient deux parties, la composante de la parole la plus supprimée $e_s(m)$ et le bruit résiduel pur (PR-bruit) $e_n(m)$, ce signal est représenté par :

$$e(m) = e_s(m) + e_n(m) \tag{6.5}$$

La distorsion de la parole est considérée comme le degré de filtrage du signal par des algorithmes d'amélioration de la parole. Une valeur faible indique une faible distorsion de parole, à savoir haute parole naturalité est préservée. Par conséquent, ce paramètre est en mesure de prédire l'état SIG et qui est désigné comme le degré de distorsion vocale (SDD). Il est raisonnable de limiter le SDD dans la plage de [0, 1] où SDD = 0 signifie pas de distorsion de parole a été introduite par l'algorithme d'amélioration alors SDD= 1 signifie la distorsion de la parole est maximale. Afin d'estimer SDD, la valeur λ de la mesure de similarité entre la parole propre et signal résiduel est calculé par :

$$\lambda(j) = \frac{\sum_{m-1}^{M} s(j,m)e(j,m)}{\sum_{m-1}^{M} s^2(j,m)} \tag{6.6}$$

où j est l'indice de trame et M est la longueur de la trame. En effet, la mesure de similarité est la valeur maximale de la corrélation croisée entre la parole propre $s(j,m)$ et le signal résiduel $e(j,m)$. $\hat{s}$ est supposé être aligné temporellement avec s, la valeur maximale qui est égale à la valeur centrale de la corrélation croisée peut être exprimée par l'équation (6.6).

Etant donné que le bruit de fond additif est supposé non corrélé avec la parole propre, la mesure de similarité de e_n et s est égal à zéro. Ainsi λ reflète certaine quantité des composantes de parole dans le signal résiduel e de chaque trame de parole. Dans l'équation. (6.6), la mesure de similarité est normalisée par l'auto-corrélation de la parole propre, qui diffère de la définition de la mesure de similarité standard où le dénominateur doit contenir à la fois la parole propre et le signal résiduel. Il est conçu pour rendre la formule proposée résistant à l'estimation inexacte de $\lambda(j)$ causée par la faible énergie du signal résiduel qui rend le dénominateur tend vers zéro. Bien que la faible énergie de la parole propre ait aussi le même effet, le signal résiduel à basse énergie est plus fréquent, alors que la parole propre avec une faible énergie ne se trouve que dans les périodes de silence du signal de parole qui sont relativement rares dans la parole continue.

Incorporant les équations (6.4) et (6.6), $\lambda(j)$ peut être réécrit comme :

$$\lambda(j) = \frac{\sum_{m-1}^{M} s(j,m)\hat{s}(j,m)}{\sum_{m-1}^{M} s^2(j,m)} - 1 \tag{6.7}$$

Si $s(j,m)$ et $\hat{s}(j,m)$ ont le même signe, la valeur de $\lambda(j)$ résulte dans trois situations :

- $\lambda(j) > 0 \Rightarrow |\hat{s}(j,m)| > |s(j,m)|$ à peu près, l'amplification du signal
- $\lambda(j) < 0 \Rightarrow |\hat{s}(j,m)| < |s(j,m)|$ à peu près, l'atténuation du signal
- $\lambda(j) = 0 \Rightarrow \hat{s}(j,m) \cong s(j,m)$, pas de distorsion du signal

Ici, "approximativement" signifie qu'il peut être intuitivement considérée comme une comparaison entre $|\hat{s}(j,m)|$ et $|s(j,m)|$. Plus précisément, il existe un contraste d'amplitude de $\sum_{m-1}^{M} s(j,m)\hat{s}(j,m)$ et $\sum_{m-1}^{M} s^2(j,m)$. Ici, nous utilisons toujours la manière intuitive afin d'introduire deux termes "l'amplification de signal" et "l'atténuation du signal" pour indiquer les situations de $\lambda(j) > 0$ et $\lambda(j) < 0$, respectivement. Pour plus d'explications, un cas particulier où $|\hat{s}(j,m)| = \alpha.|s(j,m)|$ et α est une constante positive qui est illustrée comme suit :

- $\alpha > 1 \Rightarrow \lambda(j) > 0$, amplification du signal
- $0 < \alpha < 1 \Rightarrow \lambda(j) < 0$, atténuation du signal

D'un autre côté, $\lambda(j)$ tend à être nulle indiquant que la parole traitée soit très proche de la parole propre de telle sorte qu'une bonne débruitage de parole est atteinte. Pendant ce temps, le signal de parole traité a une grande variance par rapport à la parole propre quand la valeur absolue de $\lambda(j)$ est grande. En limitant $\lambda(j)$ à la plage de -1 à 1, $|\lambda(j)|$ est capable d'indiquer la valeur SDD dans l'intervalle [0, 1]. Le processus de limitation est affiché comme :

$$SDD(j) = |\lambda(j)| \leq 1 \tag{6.8}$$

Dans ce cas, la valeur maximale, $\lambda(j) = 1$, signifie $|\hat{s}(j,m)| \cong 2|s(j,m)|$. La valeur minimale $\lambda(j) = -1$, signifie $|\hat{s}(j,m)| \cong 0$. Les situations sévères, où la valeur absolue de $\lambda(j)$ est hors de cette plage, qui sera traitée de la même façon que ces deux cas extrêmes. La raison de l'application de cette limitation est que $\lambda(j)$ soit utilisée pour estimer SDD qui est considérée comme étant le pourcentage de la parole propre intégrée dans le signal résiduel. En outre, sur la base de l'analyse de 240 paires de signaux $s(m)$ et $e(m)$ calculées par l'équation (6.6), seulement 13,57% des valeurs de λ sont hors de la plage de [-1, 1]. Ainsi, le processus de limitation non seulement rend λ apte pour l'estimation de SDD d'une manière raisonnable, mais réduit également les distorsions introduites par les situations sévères qui se produisent rarement. Le SDD globale peut être obtenu en prenant la moyenne de l'opération $SDD(j)$ pour les trames.

Toutes les analyses ci-dessus sont basées sur les hypothèses suivantes : le signal de parole traité $\hat{s}$ est similaire au signal de parole propre s et les deux signaux de parole ont le même signe. Cependant, si $\hat{s}(j,m)$ et $s(j,m)$ ont un signe différent, une mauvaise estimation de la parole propre et la valeur absolue du signal résiduel e est grand. Dans cette situation, $\lambda(j) < -1$ peut être obtenu par l'équation (6.7). Il est inapproprié de considérer cette situation comme "atténuation du signal", mais il appartient à le cas extrême $|\lambda(j)| = 1$, une grande variance se produit entre la parole propre et le signal traité. Par conséquent, l'indicateur SDD est encore réalisable pour cette situation de "différent signe".

6.3.2 Calcul du score BAK : rapport signal/PR-bruit (SPR)

Un autre point qui nous intéresse, à savoir l'aspect de la réduction du bruit. Sur la base de l'équation (6.4), PR-bruit peuvent être obtenues par $e(m) - e_s(m)$ pour calculer le score BAK. Cependant, la composante de parole sur-supprimé, $e_s(m)$ est indisponible. Le SDD peut être estimé par la valeur de $\lambda(j)$, ce qui reflète le pourcentage de la composante de parole propre

incorporés dans le signal résiduel. Ainsi $e_s(j,m)$ peut être quantitativement estimée $\lambda(j)s(j,m)$. Par conséquent, le PR-bruit e_n peut être obtenu par :

$$\hat{e}_n(j,m) = e(j,m) - \lambda(j)s(j,m) \tag{6.9}$$

L'estimation de la quantité de $e_n(j,m)$ ci-dessus est en accord avec l'explication physique. Par exemple, dans le cas particulier $|\hat{s}(j,m)| = \alpha.|s(j,m)|$ où α est une constante positive, $\hat{e}_n(j,(m) = 0$ est obtenue sur la base des équations (6.4), (6.7) et (6.9), à savoir l'absence de la distorsion de bruit, conduisant $\lambda(j)$ à un signal indicateur plain d'amplification / atténuation. Cependant, $\hat{e}_n(j,m)$ est pas le vrai PR-bruit puisque $\lambda(j)s(j,m)$ ne constitue pas une estimation précise de $e_s(j,m)$. Afin d'avoir une meilleure compréhension du niveau de PR-bruit, le rapport signal/PR-bruit (SPR) est choisi pour calculer le score de BAK et la définition est donnée ci-dessous :

$$SPR(j) = 10\log_{10}\frac{\sum_{m=1}^{M} s^2(j,m)}{\sum_{m=1}^{M} \hat{e}_n^2(j,m)} \tag{6.10}$$

Les valeurs $SPR(j)$ sont limités dans la plage de [-10 dB, 35 dB] dans chaque trame pour éviter le biais de mesure globale dans le cas de très petite $\hat{e}_n$ ou s. Cette gamme est également utilisé dans le rapport signal sur bruit segmentaire (SegSNR) dans [Hansen 1998], [Hu 2008]. Le SPR globale peut être obtenue en prenant la moyenne de l'opération $SPR(j)$ à travers les trames.

Comme indiqué dans [Hu 2006a], la mesure SegSNR est fortement corrélée avec les scores BAK. Un score élevé de SegSNR indique une bonne élimination du bruit par l'algorithme d'amélioration. Le SPR proposé est également en mesure d'évaluer la performance de l'algorithme d'amélioration sur la réduction du bruit. La raison est représentée dans la discussion suivante. Incorporant les équations (6.6), (6.9) et (6.10), le SPR dans la trame j est donnée par :

$$\begin{aligned} SPR(j) &= 10\log_{10}\frac{\sum_{m=1}^{M} s^2(j,m)}{\sum_{m=1}^{M} e^2(j,m) - \lambda^2(j)\sum_{m=1}^{M} s^2(j,m)} \\ &= 10\log_{10}\frac{1}{\frac{\sum_{m=1}^{M} e^2(j,m)}{\sum_{m=1}^{M} s^2(j,m)} - \lambda^2(j)} \end{aligned} \tag{6.11}$$

où les valeurs $\frac{\sum_{m=1}^{M} e^2(j,m)}{\sum_{m=1}^{M} s^2(j,m)} - \lambda^2(j) \leq 0$ sont tous à une petite constante positive. Si $\lambda^2 = 0$, ce qui ne signifie pas une distorsion sur la parole, l'équation (6.11) est comme la mesure de SegSNR standard. Dans le cas contraire, le résultat SPR sera affecté par la valeur de λ^2 lorsque $\frac{\sum_{m=1}^{M} e^2(j,m)}{\sum_{m=1}^{M} s^2(j,m)}$ est considéré comme une constante :

- $\lambda^2(j) \uparrow$ signifie la distorsion de parole est sévère => le $SPR(j) \uparrow$ signifie le bruit de fond est supprimée complètement.
- $\lambda^2(j) \downarrow$ signifie la parole distorsion est faible => le $SPR(j) \downarrow$ signifie le bruit de fond reste très élevé.

Cette relation entre $\lambda^2(j)$ et $SPR(j)$ reflète une situation conflictuelle entre la distorsion de la parole et la réduction du bruit introduit par les algorithmes d'amélioration, c'est un problème réel examiné par un certain nombre de travaux ([Ephraim 1984] ; [Ding 2010]). La plupart des algorithmes poursuivent un bon compromis pour obtenir le signal de parole débruité qui est le plus adapté à la perception auditive humaine.

6.3.3 Estimation des scores SIG, BAK et OVRL

SDD et SPR peuvent numériquement évaluer la performance des algorithmes d'amélioration en termes de SIG et BAK. Cependant, leurs valeurs sont dans les plages de [0, 1] et [-10 dB, 35 dB], respectivement. Pour aligner avec la gamme [1, 5] des évaluations subjectives R_{SIG} et R_{BAK}, ainsi que pour la comparaison intuitive et l'estimation de régression linéaire suivantes, nous utilisons deux fonctions de projection linéaire appliquées sur SDD et SPR respectivement pour atteindre notre mesure d'estimer les évaluations SIG et BAK, $\hat{R}_{SIG}$ et $\hat{R}_{BAK}$, comme suit :

$$\hat{R}_{SIG} = 5 - 4.SDD \tag{6.12}$$

$$\hat{R}_{BAK} = \frac{4}{45}SPR + \frac{85}{45} \tag{6.13}$$

où SDD peut être obtenue par l'équation (6,8) et SPR par l'équation (6.10). Pour les cas extrêmes, $SDD = 0$ et $SDD = 1$ correspondent à $\hat{R}_{SIG} = 5$ et $\hat{R}_{SIG} = 1$, respectivement, tandis que $SPR = 35dB$ et $SPR = -10dB$ correspondent à $\hat{R}_{BAK} = 5$ et $\hat{R}_{BAK} = 1$, respectivement.

Dans [Hu 2006a], la régression linéaire est appliquée pour analyser les évaluations subjectives obtenues pour la qualité globale de la parole et le bruit envahissant. La relation entre les trois évaluations est la suivante :

$$R_{OVRL} = -0.0783 + 0.571.R_{SIG} + 0.366.R_{BAK} \tag{6.14}$$

Sur la base de nos données subjectives, une relation linéaire peut être obtenue aussi, qui est présentée comme suit :

$$R_{OVRL} = -0.0533 + 0.521.R_{SIG} + 0.403.R_{BAK} \tag{6.15}$$

Les poids entre les équations (6.14) et (6.15) sont assez similaires. Il est donc raisonnable d'utiliser ces poids directement pour d'autres situations similaires. Selon les tests subjectifs dans [Hu 2006a], nous utilisons la relation linéaire illustrée dans l'équation (6.14) calculer le $\hat{R}_{OVRL}$, qui est la valeur estimée de la qualité globale basée sur nos mesures objectives. Avec les équations (6.12) et (6.13), $\hat{R}_{OVRL}$ est calculé comme suit :

$$\hat{R}_{OVRL} = -0.0783 + 0.571.\hat{R}_{SIG} + 0.366.\hat{R}_{BAK} \tag{6.16}$$

6.4 Expérience d'amélioration de la parole

6.4.1 Dispositif expérimental

Dans nos expériences, nous utilisons la base des données de parole NOIZEUS, qui est développée pour faciliter la comparaison entre les groupes de recherche de l'amélioration de la parole. Cette base contient 30 phrases IEEE (produites par trois hommes et trois femmes) corrompus par différents bruits du monde réel avec différents SNRs [Rix 2001]. Les phrases ont été initialement échantillonnées à 25 kHz et sous-échantillonnés à 8 kHz. Les phrases propres ont été prises à partir de la base de données IEEE, nous générons un ensemble de phrase qui a été corrompu par un bruit blanc additif gaussien, ainsi que le bruit coloré de voiture est généré par les créateurs de la base des données pour quatre niveaux de SNR (0, 5, 10 et 15 dB). Nous évaluons la performance par plusieurs méthodes en utilisant l'évaluation perceptuelle de la qualité de la parole (PESQ), SNR, segmentaire-SNR (SegSNR), les spectrogrammes et l'évaluation subjective. Les définitions des acronymes de traitement qui seront utilisés dans les évaluations sont les suivantes :

- **Noisy :** Le signal de parole bruité ;
- **K-Clean :** Le filtre de Kalman avec des LPCs estimés à partir d'un signal propre, w = 20, s = 0 et $p = 10$;
- **MMSE :** La méthode de l'erreur quadratique moyenne minimum d'amplitude spectrale courte durée ;
- **PSC :** La procédure de la compensation spectral de la phase [Wójcicki 2008] ;
- **K-Iter :** Le filtre de Kalman itératif de Gibson [Gibson 1991] ; trois itérations , w = 20, s = 0 et $p = 10$;
- **K-D-C :** Le filtre de Kalman en utilisant la fenêtre d'analyse de Dolph-Chebyshev [So 2010] ; (atténuation de lobe latéral=-200 dB), la longueur de trame w=80, deux itérations et $p = 10$;

— **MAP :** La technique basée sur l'estimation de maximum à posteriori du spectre d'amplitude au carré [Lu 2011] ;

— **SMPO :** Le masquage doux en utilisant l'incertitude à posteriori de SNR sur le spectre d'amplitude au carré [Lu 2011] ;

— **MLP-EKF** Entraînement au EKF d'un MLP ; une itérations, w = 20, s = 0 et $p = 10$;

— **MLP-UKF** Entraînement au UKF d'un MLP ; une itérations, w = 20, s = 0 et $p = 10$;

— **K-proposed :** Kalman LPC-FEM [Mellahi 2015] ; deux itérations, w = 20, s = 0 et $p = 10$;

(w : la durée de trame d'analyse (ms) ; s : le chevauchement de trame d'analyse (ms) ; p : l'ordre du modèle de prédiction linéaire.)

6.4.2 Résultats et discussion

6.4.2.1 Evaluation objectif

En termes de PESQ, SNR et SegSNR, les valeurs les plus élevées indiquent la meilleure performance, c.-à-d., la meilleure qualité de la parole.

Le Tableau 6.1 montre les comparaisons de performance basées sur les mesures PESQ, SNR et SegSNR entre le filtre de Kalman proposé (K-proposed) et les filtres de Kalman itératives concurrents (K-Iter, K-D-C) pour un bruit blanc gaussien de 5 dB. La meilleure performance en termes de PESQ et SegSNR est obtenue par la méthode proposée (K-proposed) tout en utilisant moins d'itération par rapport à K-Iter. En termes de SNR, Kalman LPC-FEM surpasse K-D-C. Malgré le nombre d'itérations de notre méthode (K-proposed) et la méthode K-D-C est similaire ; le FEM de la méthode proposée (K-proposed) nécessite moins de calcul que la fenêtre chevauchée de la méthode K-D-C.

Le Tableau 6.2 montre les comparaisons de performance basées sur les mesures PESQ, SNR et SegSNR entre les filtres de Kalman proposés tel que Kalman LPC-FEM, MLP-EKF et MLP-UKF pour un bruit blanc gaussien de 5 dB. La meilleure performance en termes de PESQ, SNR et SegSNR est obtenue par la méthode Kalman LPC-FEM (K-proposed).

La Figure 6.3 montre les comparaisons de performance entre les déférents méthodes dans le cas de bruit blanc gaussien (basé sur les scores PESQ, SNR et SegSNR). En termes de scores PESQ, la meilleure performance est obtenue par la méthode proposée (K-proposed) (Figure 6.3 (a)). Dans la Figure 6.3 (b) et la Figure 6.3 (c), nous pouvons voir que Kalman LPC-FEM est compétitif

Méthode	Évaluation objective		
	PESQ	SNR (dB)	SegSNR (dB)
Noisy	1.80	5.00	−1.41
K−Iter (3 Iterations)	2.43	11.80	4.59
K−D−C (2 Iterations)	2.50	11.32	5.64
K−proposed (2 Iterations)	2.57	11.30	5.60

TABLEAU 6.1 – Comparaison des performances entre la méthode proposée (K-proposed) et les méthodes concurrentes (K-Iter et K-D-C) en termes de PESQ, SNR et SegSNR en utilisant un signal de parole bruité par un bruit blanc gaussien de SNR=5 dB.

Méthode	Évaluation objective		
	PESQ	SNR (dB)	SegSNR (dB)
Noisy	1.80	5.00	−1.41
MLP−EKF	2.32	10.73	5.39
MLP−UKF	2.40	10.86	5.42
K−proposed	2.57	11.30	5.60

TABLEAU 6.2 – Comparaison des performances entre les méthodes proposées (K-proposed, MLP-EKF et MLP-UKF) en termes de PESQ, SNR et SegSNR en utilisant un signal de parole bruité par un bruit blanc gaussien de SNR=5 dB.

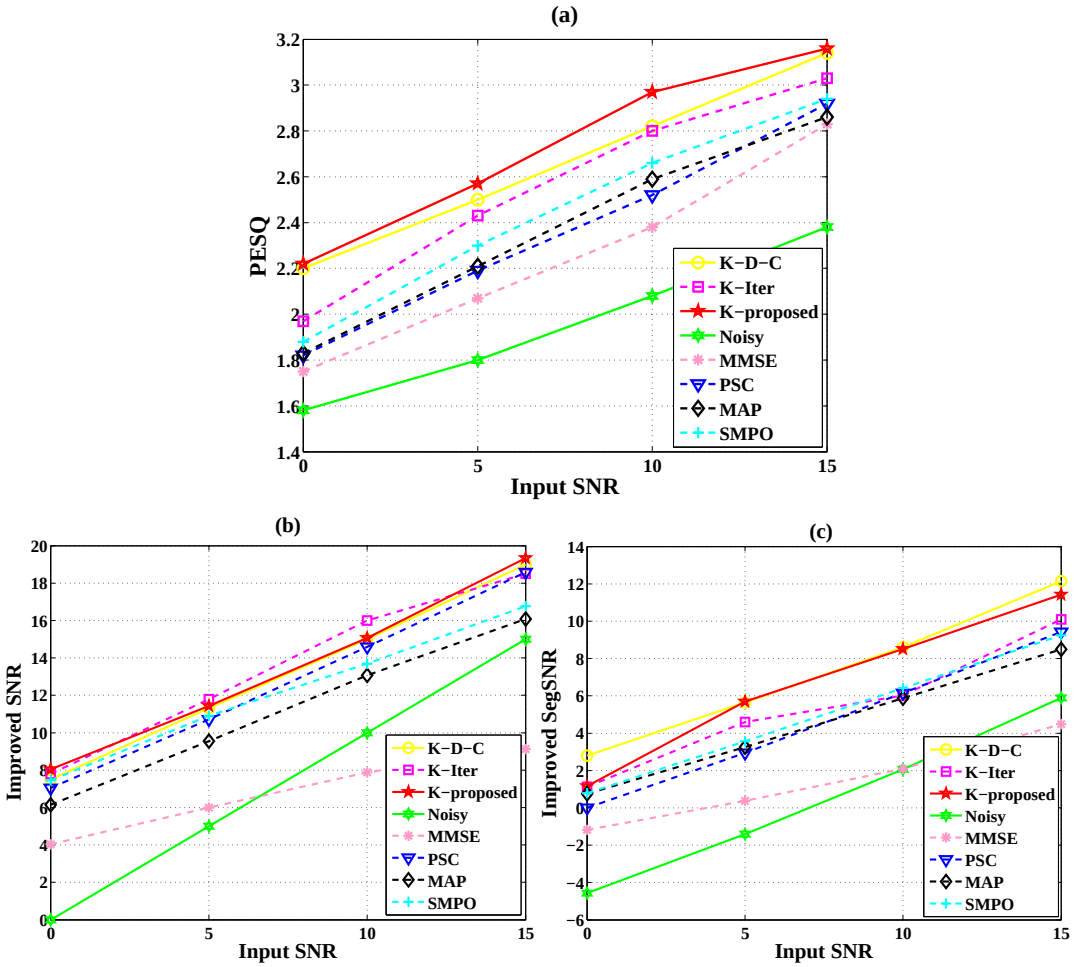

FIGURE 6.3 – Comparaison des performances de K-proposed avec d'autres algorithmes d'amélioration dans le cas de bruit blanc gaussien de différents niveaux de SNR en termes de : (a) PESQ, (b) SNR et (c) SegSNR.

avec K-Iter et K-D-C en termes de SNR et SegSNR respectivement, et elle offre fortement la meilleure performance par rapport aux autres méthodes.

La Figure 6.4 montre les spectrogrammes d'échantillon de parole traités par différentes méthodes. La phrase suivante : 'The sky that morning was clear and bright blue' a été corrompue par un bruit blanc gaussien (Noisy) de SNR=5dB (B). La sortie de K-Clean (H) ressemble clairement le signal propre (A). Le bruit résiduel est évident dans les spectrogrammes de la sortie de MMSE (D), PSC (E), MAP (F) et SMPO (G). Ce bruit résiduel est considérablement réduit dans la sortie de K-Iter (C). Le signal de la parole de la sortie de Kalman LPC-FEM (I) contient moins de bruit résiduel que la parole de la sortie de K-Iter (C). Finalement, la sortie de Kalman LPC-FEM (I) contient moins de distorsion de la parole et une faible distorsion de bruit. Les tests d'écoute informels ont confirmé que l'approche proposée (K-proposed) donne la plus haute qualité, en accord avec les scores de PESQ représentés sur la Figure 6.3 (a).

La Figure 6.5 montre les comparaisons de performance (sur la base des mesures PESQ, SNR et SegSNR) dans le cas du bruit de voiture coloré. En termes de scores PESQ, une meilleure performance est obtenue par la méthode Kalman LPC-FEM (Figure. 6.5 (a)). Dans la Figure. 6.5 (b), nous pouvons voir que la méthode proposée (K-proposed) surpasse nettement K-Iter, K-D-C, MMSE et MAP en termes de SNR. Cependant, notre méthode donne une courbe plus proche à SMPO dans 5 et 10 dB. Le même phénomène peut être observé pour PSC, pour tous les SNRs sauf 0 dB. En termes de SegSNR (Figure. 6.5 (c)), la méthode Kalman LPC-FEM est compétitif avec K-D-C, et elle offre beaucoup de meilleures performances par rapport aux autres méthodes d'amélioration.

Figure 6.6 montre les spectrogrammes d'échantillon de parole traités par des différentes méthodes. La phrase suivante : 'The sky that morning was clear and bright blue' a été corrompue par un bruit coloré de voiture de SNR=5dB (B). La sortie de K-Clean (H) ressemble clairement le signal propre (A). Le bruit résiduel est évident dans les spectrogrammes de la sortie de MMSE (D), PSC (E), MAP (F) et SMPO (G). Ce bruit résiduel est considérablement réduit dans la sortie de K-Iter (C). Le signal de la parole de la sortie de Kalman LPC-FEM (I) contient moins de bruit résiduel que la parole de la sortie de K-Iter (C). Finalement, la sortie de Kalman LPC-FEM (I) contient moins de distorsion de la parole et une faible distorsion de bruit. Les tests d'écoute informels ont confirmé que l'approche proposée (K-proposed) donne la plus haute qualité, en accord avec les scores PESQ représentés sur la Figure 6.5 (a).

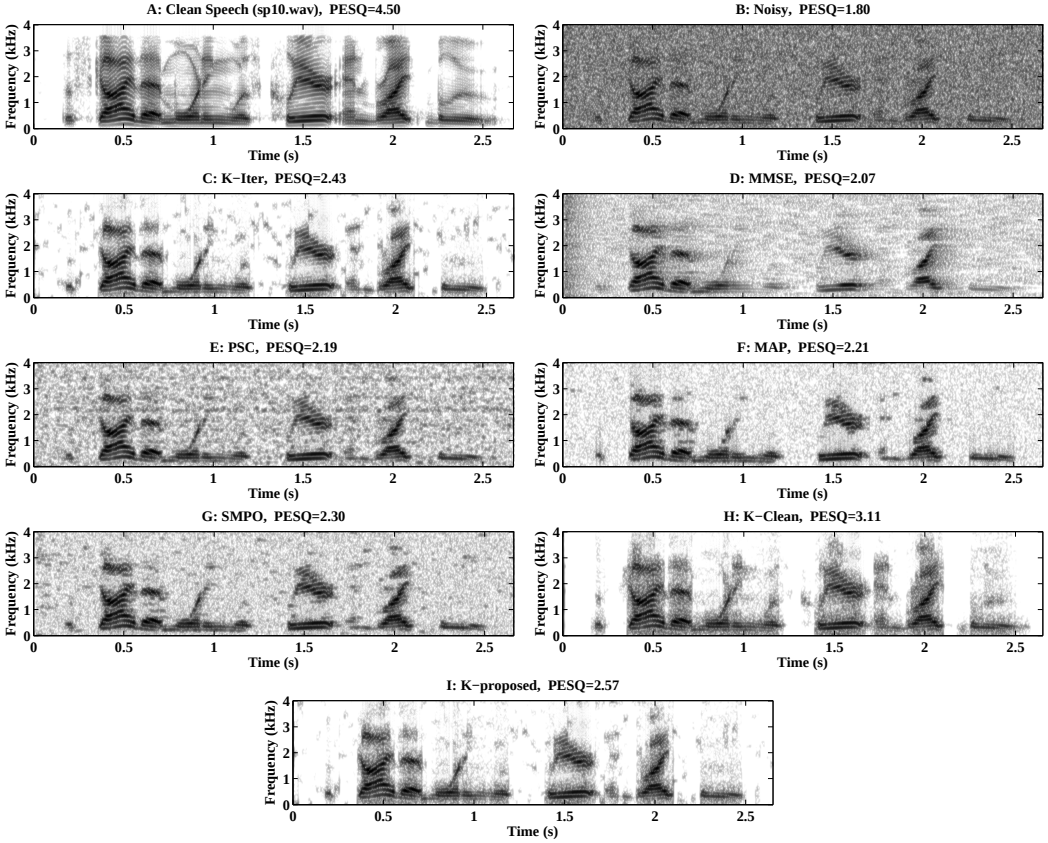

FIGURE 6.4 – Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase bruitée par un bruit blanc gaussien à SNR =5 dB ; (C) la phrase traitée par K-Iter ; (D) la phrase traitée par MMSE ; (E) la phrase traitée par PSC ; (F) la phrase traitée par MAP ; (G) la phrase traitée par SMPO ; (H) la phrase traitée par le K-Clean et (I) la phrase traitée par **K-proposed**.

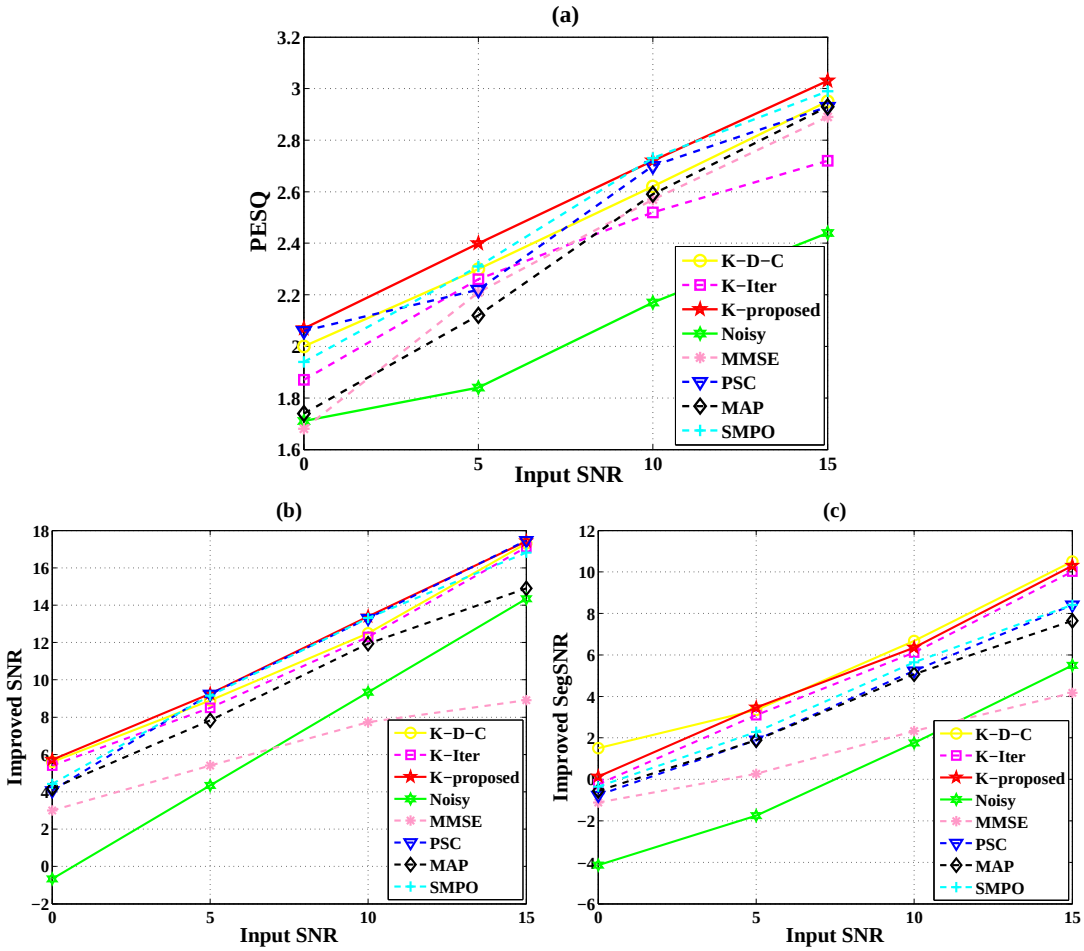

FIGURE 6.5 – Comparaison des performances de K-proposed avec d'autres algorithmes d'amélioration de parole dans le cas de bruit de voiture coloré de différents niveaux de SNR en termes de : (a) PESQ, (b) SNR et (c) SegSNR.

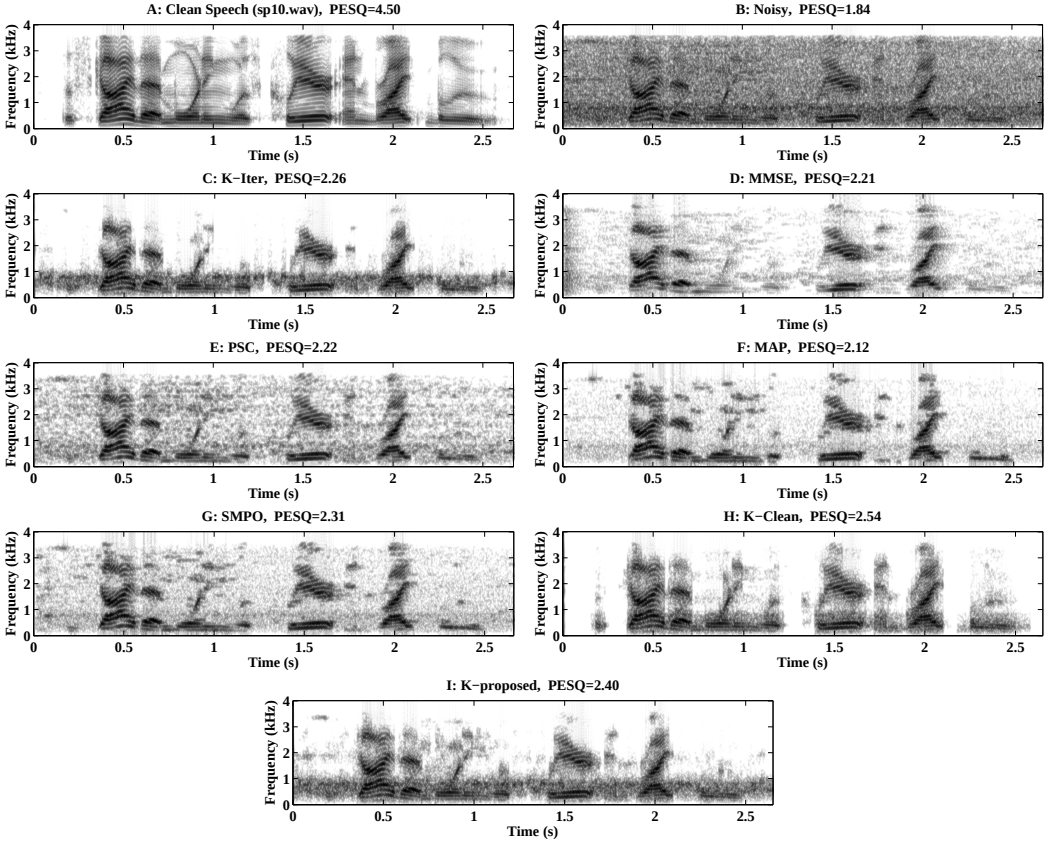

FIGURE 6.6 – Les spectrogrammes de : (A) la phrase propre (sp10.wav) 'The sky that morning was clear and bright blue' a été prise de la base de données NOIZEUS ; (B) la phrase corrompue par un bruit coloré de voiture à SNR =5 dB ; (C) la phrase traitée par K-Iter ; (D) la phrase traitée par MMSE ; (E) la phrase traitée par PSC ; (F) la phrase traitée par MAP ; (G) la phrase traitée par SMPO ; (H) la phrase traitée par le K-Clean et (I) la phrase traitée par le textbf K- proposed.

6.4.2.2 Evaluation subjective

L'évaluation subjective comprend un test d'écoute informelle qui est conçu pour suivre la procédure proposée dans [Loizou 2007] et [Hu 2006b]. L'évaluation est conçue pour un score moyen de dix phrases de parole bruité par un bruit coloré de voiture (5 par homme et 5 par femme), ces dernières sont extraites au hasard de la classe du bruit de voiture de NOIZEUS à SNR=5 dB. Ensuite, les mêmes phrases sont contaminées par un bruit blanc gaussien additif (5 par homme et 5 par femme). Une phrase extraite au hasard avec un SNR de 5dB est envoyée à l'entrée du système conçu. Vingt auditeurs sont chargés d'assister successivement pour évaluer le signal amélioré (aussi le signal bruité) sur :

1. L'échelle de distorsion du signal de parole (SIG) est : [1 = très non naturel, 5 = très naturel] ;
2. L'échelle de bruit de fond (BAK) est : [1 = très visible, très intrusive, 5 = pas perceptible] ;
3. L'échelle du MOS (Mean Opinion Score) de l'effet global (OVRL) est : [1 = mauvais, 5 = excellent].

La Figue 6.7 montre les scores moyens pour l'échelle SIG, BAK et OVRL de l'algorithme Kalman LPC-FEM comparé aux autres algorithmes d'amélioration dans le cas de bruit blanc gaussien de SNR = 5dB. Les scores moyens pour le signal bruité (non traités) sont également indiqués comme référence.

Parmi les algorithmes d'amélioration, la méthode Kalman LPC-FEM réalise un taux d'échelle OVRL élevé par rapport aux autres méthodes (sauf pour K-Clean).

Une distorsion inférieure du signal, c.-à-d. les scores de SIG plus élevés est observés avec les algorithmes de la méthode proposée, K-Clean et K-D-C. Dans la technique Kalman LPC-FEM, un grand nombre d'auditeurs ont constaté que le signal de parole est légèrement déformé par rapport aux autres algorithmes d'amélioration de la parole.

Une distorsion inférieure du bruit, c.-à-d. des scores de BAK plus élevés, ont été obtenus à nouveau avec Kalman LPC-FEM et K-Clean. Dans la technique de K proposed, la partie parole est améliorée.

La Figue 6.8 montre les scores moyens pour l'échelle SIG, BAK et OVRL de l'algorithme Kalman LPC-FEM comparé par les autres algorithmes d'amélioration de la parole dans le cas de bruit de voiture coloré de SNR = 5dB. Les scores moyens pour le signal bruité (non traités) sont également indiqués comme référence.

Dans cette expérience, la méthode Kalman LPC-FEM effectue un taux d'échelle OVRL élevé par rapport aux autres méthodes (sauf pour K-Clean,

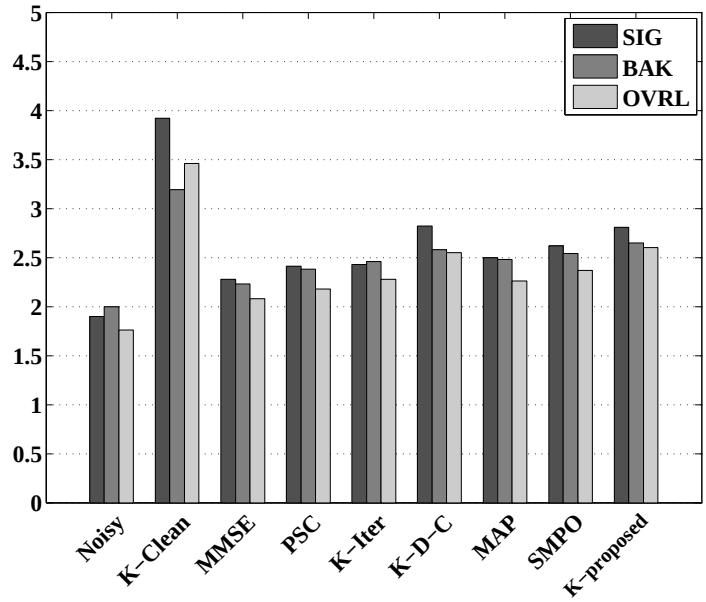

FIGURE 6.7 – Les résultats d'évaluation subjectifs (SIG = signal, BAK = bruit de fond, et OVRL = global) de la méthode proposée (K-proposed) par rapport les différentes méthodes d'amélioration de SNR=5 dB dans le cas de bruit blanc gaussien.

SMPO).

Une distorsion légère du signal (c.-à-d. des scores de SIG plus élevés) a été observée avec les algorithmes de la méthode proposée, K-Clean et PSC. Un grand nombre d'auditeurs ont constaté que le signal de parole est légèrement déformé par rapport les autres algorithmes.

Une distorsion légère du bruit (c.-à-d. des scores de BAK plus élevés) a été obtenue à nouveau avec Kalman LPC-FEM et K-Clean. Donc une nette amélioration de la parole a été obtenue par la méthode proposée.

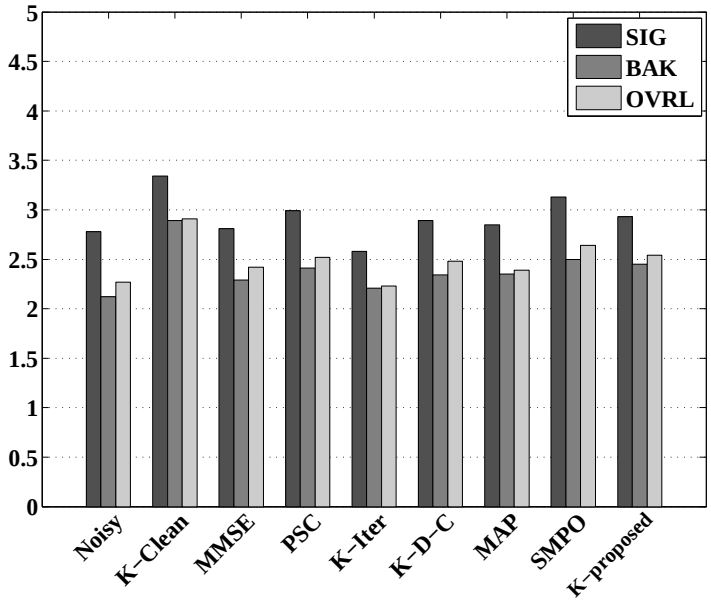

FIGURE 6.8 – Les résultats d'évaluation subjectifs (SIG = Signal, BAK = bruit de fond, et OVRL = global) de l'amélioration de la parole proposée (K-proposed) par rapport aux différentes méthodes d'amélioration de la parole de SNR=5 dB dans le cas de bruit coloré de voiture.

Chapitre 7

Conclusion générale

Sommaire

7.1 Résumé du travail 95
7.2 Suggestions des futurs travaux 96

7.1 Résumé du travail

Dans cette thèse, des algorithmes d'amélioration de parole basés sur les filtres de Kalman qui sont capables de traiter deux types de bruits ont été étudiés. Les algorithmes proposés ont été implémentés dans des cas non idéaux, où les paramètres de modèle d'état du filtre de Kalman, à savoir les LPCs et la variance de bruit sont estimés dans des conditions bruitées sans tenir compte de la connaissance a priori de la parole propre et du bruit additif. Par contre dans la plupart des méthodes existantes basées sur les filtres de Kalman, les informations sur la parole et le bruit sont supposées être disponibles. Ces hypothèses antérieures rendent les algorithmes conventionnels non pratiques dans le sens de l'amélioration réelle pour un signal bruité. Afin de résoudre ces problèmes, de nouvelles méthodes d'estimation des LPCs et de la variance du bruit dans des conditions bruitées ont été proposées. En fonction de ces techniques d'estimation de paramètres, trois méthodes d'amélioration de parole basées sur le filtre de Kalman ont été développées.

Tout d'abord, dans les méthodes non itératives basées sur le filtre de Kalman, les paramètres du modèle d'espace d'état, à savoir les LPCs et la variance du bruit sont estimés dans des conditions bruitées. Une méthode combinée de lissage de la parole et d'autocorrélation a été proposée pour l'estimation des LPCs. Deux autres méthodes de filtre de Kalman non itératives dans le domaine non linéaire ont été proposés, la $1^{ère}$ est basée sur l'expansion de Taylor pour la linéarisation de la parole bruité et la $2^{ème}$ est basée sur la transformée non-parfumée pour l'approximation qui sont respectivement le filtre de Kalman étendu et le filtre de Kalman non parfumé entrainés par un perceptron multicouche.

Bien que le filtre de Kalman non itératif fonctionne relativement bien, il introduit néanmoins des bruits résiduels et de petites distorsions dans la parole améliorée. Afin d'augmenter les performances d'amélioration de la parole et la précision de l'estimation des paramètres dans les conditions bruitées, une méthode itérative basée sur le filtre de Kalman a été présentée comme une deuxième approche.

Du fait que la parole améliorée fournie par le filtre de Kalman itératif soit exempte de bruit résiduel qui apparaît dans la méthode basée sur le filtre de Kalman non itératif, certains artefacts de type musical restent dans la parole filtrée. Pour améliorer encore les résultats de l'amélioration de la parole, un filtre de Kalman itératif basé sur la méthode d'amélioration des formants (Kalman LPC-FEM) a été proposé comme une troisième approche. Cette dernière est proposée pour obtenir un meilleur compromis entre la réduction du bruit, l'intelligibilité de la parole et la complexité des calculs.

Les méthodes proposées ont été testées avec une base de données vocale largement utilisée, à savoir NOIZEUS. Les expériences ont été conduites dans la présence des deux types de bruits pour une large gamme de SNR d'entrée, où les conversations vocales de la vie réelle ont souvent eu lieu. Les performances sont évaluées et comparées avec certaines méthodes d'amélioration. Grâce à des simulations des méthodes proposées, il a été clairement observé que ces méthodes proposées, comparées aux méthodes conventionnelles, sont efficaces pour réduire le bruit, tout en préservant une meilleure qualité. Le temps de calcul pour les méthodes proposées est également raisonnable. En outre, les méthodes proposées peuvent être efficaces dans la présence de différents bruits, tandis que les performances de certaines méthodes existantes sont limitées par des types de bruit spécifiques. Parmi les méthodes proposées, Kalman LPC-FEM est la meilleure, suivi du MLP-EKF et MLP-UKF en termes de métriques d'évaluation rapportées.

7.2 Suggestions des futurs travaux

Les résultats de cette thèse indiquent plusieurs directions intéressantes pour les travaux de futurs, qui devraient être abordées et développées pour obtenir de meilleurs résultats à moindre coût de calcul. Ces points sont les suivants :

— L'amélioration par le filtre de Kalman quadrature (QKF)

— L'amélioration par le filtre de Kalman cubature (CKF)

Annexe A

Matériels supplémentaires

A.1 Propriétés de la distribution gaussienne

Définition 2 (distribution gaussienne)**.** *La variable aléatoire* $\mathbf{x} \in \mathbb{R}^n$ *a une distribution gaussienne avec la moyenne* $\mathbf{m} \in \mathbb{R}^n$ *et la covariance* $\mathbf{P} \in \mathbb{R}^{n \times n}$ *sa densité de probabilité a la forme*

$$p(\mathbf{x}_k|\mathbf{m}, \mathbf{P}) = \frac{1}{(2\pi)^{n/2}|\mathbf{P}|^{1/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{m})\right), \tag{A.1}$$

où $|\mathbf{P}|$*est le déterminant de la matrice* $\mathbf{P}$*.*

Lemme A.1 (Densité conjointe des variables gaussiennes)**.** *Si les variables aléatoires* $\mathbf{x} \in \mathbb{R}^n$ *et* $\mathbf{y} \in \mathbb{R}^n$ *ont les densités de probabilité gaussiennes*

$$\begin{aligned} \mathbf{x} &\sim N(\mathbf{x}|\mathbf{m}, \mathbf{P}) \\ \mathbf{y}|\mathbf{x} &\sim N(\mathbf{y}|\mathbf{H}\mathbf{x}+\mathbf{u}, \mathbf{R}), \end{aligned} \tag{A.2}$$

la densité conjointe de $\mathbf{x}$*,* $\mathbf{y}$ *et la distribution marginale de* $\mathbf{y}$ *sont données comme*

$$\begin{aligned} \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} &\sim N\left(\begin{bmatrix}\mathbf{m}\\ \mathbf{H}\mathbf{m}+\mathbf{u}\end{bmatrix}, \begin{bmatrix}\mathbf{P} & \mathbf{P}\mathbf{H}^T\\ \mathbf{H}\mathbf{P} & \mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{R}\end{bmatrix}\right) \\ \mathbf{y} &\sim N(\mathbf{H}\mathbf{m}+\mathbf{u}, \mathbf{H}\mathbf{P}\mathbf{H}^T+\mathbf{R}). \end{aligned} \tag{A.3}$$

Lemme A.2 (Densité conditionnelle des variables gaussiennes)**.** *Si les variables aléatoires* $\mathbf{x}$ *et* $\mathbf{y}$ *ont la densité de probabilité gaussienne conjointe*

$$\mathbf{x}, \mathbf{y} \sim N\left(\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}, \begin{bmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{C}^T & \mathbf{B}\end{bmatrix}\right), \tag{A.4}$$

alors les densités marginales et conditionnelles de $\mathbf{x}$ *et* $\mathbf{y}$ *sont données comme suit :*

$$\begin{aligned} \mathbf{x} &\sim N(\mathbf{a}, \mathbf{A}) \\ \mathbf{y} &\sim N(\mathbf{b}, \mathbf{B}) \\ \mathbf{x}|\mathbf{y} &\sim N(\mathbf{a}+\mathbf{C}\mathbf{B}^{-1}(\mathbf{y}-\mathbf{b}), \mathbf{A}-\mathbf{C}\mathbf{B}^{-1}\mathbf{c}^T) \\ \mathbf{y}|\mathbf{x} &\sim N(\mathbf{b}+\mathbf{C}^T\mathbf{A}^{-1}(\mathbf{x}-\mathbf{a}), \mathbf{B}-\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}). \end{aligned} \tag{A.5}$$

Bibliographie

[Allen 1977] Jont B Allen et Lawrence R Rabiner. *A unified approach to short-time Fourier analysis and synthesis*. Proceedings of the IEEE, vol. 65, no. 11, pages 1558–1564, 1977. (Cité en page 14.)

[Arasaratnam 2009] Ienkaran Arasaratnam et Simon Haykin. *Cubature kalman filters*. IEEE Transactions on automatic control, vol. 54, no. 6, pages 1254–1269, 2009. (Cité en page 46.)

[Bar 2001] YAAKOV Bar, XR Shalom, T Kirubarajan Li et T Kirubarajan. *Estimation with applications to tracking and navigation*, 2001. (Cité en page 34.)

[Berouti 1979] Michael Berouti, Richard Schwartz et John Makhoul. *Enhancement of speech corrupted by acoustic noise*. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'79., volume 4, pages 208–211. IEEE, 1979. (Cité en page 11.)

[Blackman 1958] Ralph Beebe Blackman et John W Tukey. *The measurement of power spectra*. 1958. (Cité en page 51.)

[Boll 1979] Steven Boll. *Suppression of acoustic noise in speech using spectral subtraction*. IEEE Transactions on acoustics, speech, and signal processing, vol. 27, no. 2, pages 113–120, 1979. (Cité en pages 2 et 8.)

[Bruce 2002] Ian C Bruce, Neel V Karkhanis, Eric D Young et Murray B Sachs. *Robust formant tracking in noise*. In Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on, volume 1, pages I–281. IEEE, 2002. (Cité en page 54.)

[Chen 2004] Bin Chen et Philipos C Loizou. *Formant frequency estimation in noise*. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, volume 1, pages I–581. IEEE, 2004. (Cité en page 54.)

[Chui 1989] CK Chui et G Chen. *Kalman filtering with real time applications*. Applied Optics, vol. 28, page 1841, 1989. (Cité en pages 18 et 19.)

[Connor 1994] Jerome T Connor, R Douglas Martin et Les E Atlas. *Recurrent neural networks and robust time series prediction*. IEEE transactions on neural networks, vol. 5, no. 2, pages 240–254, 1994. (Cité en page 64.)

[Ding 2010] Huijun Ding, Yann Soon et Chai Kiat Yeo. *Over-attenuated components regeneration for speech enhancement*. IEEE transactions on audio, speech, and language processing, vol. 18, no. 8, pages 2004–2014, 2010. (Cité en page 83.)

[Durbin 1960] James Durbin. *The fitting of time-series models.* Revue de l'Institut International de Statistique, pages 233–244, 1960. (Cité en page 53.)

[Ephraim 1984] Yariv Ephraim et David Malah. *Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator.* IEEE Transactions on acoustics, speech, and signal processing, vol. 32, no. 6, pages 1109–1121, 1984. (Cité en pages 2, 7, 11, 12 et 83.)

[Ephraim 1992] Yariv Ephraim. *Statistical-model-based speech enhancement systems.* Proceedings of the IEEE, vol. 80, no. 10, pages 1526–1555, 1992. (Cité en page 67.)

[Ephraim 1995] Yariv Ephraim et Harry L Van Trees. *A signal subspace approach for speech enhancement.* IEEE Transactions on speech and audio processing, vol. 3, no. 4, pages 251–266, 1995. (Cité en page 16.)

[Gabrea 2001] Marcel Gabrea. *Adaptive Kalman filtering-based speech enhancement algorithm.* In Electrical and Computer Engineering, 2001. Canadian Conference on, volume 1, pages 521–526. IEEE, 2001. (Cité en pages 2, 3 et 21.)

[Gannot 1998] Sharon Gannot, David Burshtein et Ehud Weinstein. *Iterative and sequential Kalman filter-based speech enhancement algorithms.* IEEE Transactions on speech and audio processing, vol. 6, no. 4, pages 373–385, 1998. (Cité en pages 2, 21 et 22.)

[Gelman 1995] Andrew Gelman, John B Carlin, Hal S Stern et Donald B Rubin. Bayesian data analysis. Chapman and Hall/CRC, 1995. (Cité en page 30.)

[Gibson 1991] Jerry D Gibson, Boneung Koo et Steven D Gray. *Filtering of colored noise for speech enhancement and coding.* IEEE Transactions on signal processing, vol. 39, no. 8, pages 1732–1742, 1991. (Cité en pages 2 et 84.)

[Goodwin 2014] Graham C Goodwin et Kwai Sang Sin. Adaptive filtering prediction and control. Courier Corporation, 2014. (Cité en page 66.)

[Grewal 2001] MS Grewal et AP Andrews. Kalman filtering : theory and practice using matlab. Wiley New York et al., 2001. (Cité en pages 29 et 34.)

[Hansen 1998] John HL Hansen et Bryan L Pellom. *An effective quality evaluation protocol for speech enhancement algorithms.* In Fifth International Conference on Spoken Language Processing, 1998. (Cité en pages 76 et 82.)

[Haykin 2000] Simon Haykin. *Adaptive filter theory, 1996.* telecommunication systems, radio resource management,(adhoc) multihop relay system,

sensor network and particularly their applicable issues to 4G mobile communication systems and cognitive radio systems. Bath in, pages 12–13, 2000. (Cité en pages 8 et 9.)

[Haykin 2001] Simon S Haykin*et al.* Kalman filtering and neural networks. Wiley Online Library, 2001. (Cité en pages 39 et 42.)

[Hu 2006a] Yi Hu et Philipos C Loizou. *Evaluation of objective measures for speech enhancement*. In Ninth International Conference on Spoken Language Processing, 2006. (Cité en pages 82, 83 et 84.)

[Hu 2006b] Yi Hu et Philipos C Loizou. *Subjective comparison of speech enhancement algorithms*. In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, volume 1, pages I–I. IEEE, 2006. (Cité en page 92.)

[Hu 2008] Yi Hu et Philipos C Loizou. *Evaluation of objective quality measures for speech enhancement*. IEEE Transactions on audio, speech, and language processing, vol. 16, no. 1, pages 229–238, 2008. (Cité en pages 75, 79 et 82.)

[Ito 2000] Kazufumi Ito et Kaiqi Xiong. *Gaussian filters for nonlinear filtering problems*. IEEE transactions on automatic control, vol. 45, no. 5, pages 910–927, 2000. (Cité en page 46.)

[Jaswinski] AH Jaswinski. *Stochastic Processes and Filtering Theory, 1970*. (Cité en page 34.)

[Juler 2000] SJ Juler, JK Uhlmann et HF Durrant-Whyte. *A new method for nonlinear transformation of means and covariances in filters and estimator*. IEEE Trans. Autom. Control, vol. 45, no. 3, pages 477–482, 2000. (Cité en page 38.)

[Julier 1995] Simon J Julier, Jeffrey K Uhlmann et Hugh F Durrant-Whyte. *A new approach for filtering nonlinear systems*. In American Control Conference, Proceedings of the 1995, volume 3, pages 1628–1632. IEEE, 1995. (Cité en page 42.)

[Julier 1996] Simon J Julier et Jeffrey K Uhlmann. *A general method for approximating nonlinear transformations of probability distributions*. Rapport technique, Technical report, Robotics Research Group, Department of Engineering Science, University of Oxford, 1996. (Cité en page 38.)

[Julier 1997] Simon J Julier et Jeffrey K Uhlmann. *New extension of the Kalman filter to nonlinear systems*. In Signal processing, sensor fusion, and target recognition VI, volume 3068, pages 182–194. International Society for Optics and Photonics, 1997. (Cité en page 68.)

[Julier 2004] Simon J Julier et Jeffrey K Uhlmann. *Unscented filtering and nonlinear estimation*. Proceedings of the IEEE, vol. 92, no. 3, pages 401–422, 2004. (Cité en pages 38, 42, 43 et 46.)

[Kabal 1991] Peter Kabal, F-M Wang, Douglas O'Shaughnessy et Ravi P Ramachandran. *Adaptive postfiltering for enhancement of noisy speech in the frequency domain*. In Circuits and Systems, 1991., IEEE International Sympoisum on, pages 312–315. IEEE, 1991. (Cité en page 54.)

[Kalman 1960] Rudolph Emil Kalman. *A new approach to linear filtering and prediction problems*. Journal of basic Engineering, vol. 82, no. 1, pages 35–45, 1960. (Cité en page 27.)

[Lewis 1986] Frank L Lewis et FL Lewis. Optimal estimation : with an introduction to stochastic control theory. Wiley New York et al., 1986. (Cité en page 65.)

[Lim 1978] Jae Lim et Alan Oppenheim. *All-pole modeling of degraded speech*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 3, pages 197–210, 1978. (Cité en page 67.)

[Lim 1979] Jae S Lim et Alan V Oppenheim. *Enhancement and bandwidth compression of noisy speech*. Proceedings of the IEEE, vol. 67, no. 12, pages 1586–1604, 1979. (Cité en page 7.)

[Loizou 2007] PC Loizou. *Subjective evaluation and comparison of speech enhancement algorithms*. Speech Commun, vol. 49, pages 588–601, 2007. (Cité en pages 55 et 92.)

[Lu 2011] Yang Lu et Philipos C Loizou. *Estimators of the magnitude-squared spectrum and methods for incorporating SNR uncertainty*. IEEE transactions on audio, speech, and language processing, vol. 19, no. 5, pages 1123–1137, 2011. (Cité en page 85.)

[Ma 2004] Ning Ma, Martin Bouchard et Rafik A Goubran. *Perceptual Kalman filtering for speech enhancement in colored noise*. In Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, volume 1, pages I–717. IEEE, 2004. (Cité en page 68.)

[Markel 1972] J Markel. *Digital inverse filtering-a new tool for formant trajectory estimation*. IEEE Transactions on Audio and Electroacoustics, vol. 20, no. 2, pages 129–137, 1972. (Cité en page 54.)

[Matthews 1994] Michael B Matthews et George S Moschytz. *The identification of nonlinear discrete-time fading-memory systems using neural network models*. IEEE Transactions on Circuits and Systems II : Analog and Digital Signal Processing, vol. 41, no. 11, pages 740–751, 1994. (Cité en page 66.)

[Maybeck 982a] Peter S Maybeck. Stochastic models, estimation, and control, volume 2. Academic press, 1982a. (Cité en page 34.)

[McCandless 1974] S McCandless. *An algorithm for automatic formant extraction using linear prediction spectra*. IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 22, no. 2, pages 135–141, 1974. (Cité en page 54.)

[Mellahi 2015] Tarek Mellahi et Rachid Hamdi. *LPC-based formant enhancement method in Kalman filtering for speech enhancement*. AEU-International Journal of Electronics and Communications, vol. 69, no. 2, pages 545–554, 2015. (Cité en pages 2 et 85.)

[Mittal 2000] Udar Mittal et Nam Phamdo. *Signal/noise KLT based approach for enhancing speech degraded by colored noise*. IEEE Transactions on Speech and Audio Processing, vol. 8, no. 2, pages 159–167, 2000. (Cité en page 17.)

[Nelson 1976] Lawrence Nelson et Edwin Stear. *The simultaneous on-line estimation of parameters and states in linear systems*. IEEE Transactions on automatic Control, vol. 21, no. 1, pages 94–98, 1976. (Cité en page 66.)

[Nemer 1999] Elias J Nemer. Speech analysis and quality enhancement using higher order cumulants. Citeseer, 1999. (Cité en page 10.)

[Niederjohn 1996] Russell J Niederjohn. *Understanding speech corrupted by noise*. In Industrial Technology, 1996.(ICIT'96), Proceedings of The IEEE International Conference on, pages P1–P5. IEEE, 1996. (Cité en page 8.)

[NøRgaard 2000] Magnus NøRgaard, Niels K Poulsen et Ole Ravn. *New developments in state estimation for nonlinear systems*. Automatica, vol. 36, no. 11, pages 1627–1638, 2000. (Cité en page 46.)

[Oppenheim 1999] Alan V Oppenheim. Discrete-time signal processing. Pearson Education India, 1999. (Cité en page 7.)

[O'shaughnessy 1987] Douglas O'shaughnessy. Speech communication : human and machine. Universities press, 1987. (Cité en page 53.)

[O'Shaughnessy 1989] Douglas O'Shaughnessy. *Enhancing speech degrated by additive noise or interfering speakers*. IEEE Communications Magazine, vol. 27, no. 2, pages 46–52, 1989. (Cité en page 7.)

[Paliwal 1987] K Paliwal et Anjan Basu. *A speech enhancement method based on Kalman filtering*. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'87., volume 12, pages 177–180. IEEE, 1987. (Cité en pages 2 et 18.)

[Popescu 1998] Dimitrie C Popescu et Ilija Zeljkovic. *Kalman filtering of colored noise for speech enhancement*. In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, volume 2, pages 997–1000. IEEE, 1998. (Cité en page 2.)

[Puskorius 1994] Gintaras V Puskorius et Lee A Feldkamp. *Neurocontrol of nonlinear dynamical systems with Kalman filter trained recurrent networks*. IEEE Transactions on neural networks, vol. 5, no. 2, pages 279–297, 1994. (Cité en page 64.)

[Quackenbush 1986] Schuyler R Quackenbush. *OBJECTIVE MEASURES OF SPEECH QUALITY (SUBJECTIVE)*. 1986. (Cité en page 75.)

[Quackenbush 1988] S Quackenbush, T Barnwell et M Clements. *Objective Measures of Speech Quality. 1988*. Ramrez, J., JC Segura, C. Bentez, L. Garca, and A. Rubio." Statistical voice activity, 1988. (Cité en page 74.)

[Rabiner 1978] Lawrence R Rabiner et Ronald W Schafer. Digital processing of speech signals, volume 100. Prentice-hall Englewood Cliffs, NJ, 1978. (Cité en page 53.)

[Recommendation 2001] ITU-T Recommendation. *Perceptual evaluation of speech quality (PESQ) : An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*. Rec. ITU-T P. 862, 2001. (Cité en pages 3, 75 et 77.)

[Rezayee 2001] Afshin Rezayee et Saeed Gazor. *An adaptive KLT approach for speech enhancement*. IEEE Transactions on Speech and Audio Processing, vol. 9, no. 2, pages 87–95, 2001. (Cité en pages xi, 17 et 18.)

[Rix 2001] Antony W Rix, John G Beerends, Michael P Hollier et Andries P Hekstra. *Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs*. In Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, volume 2, pages 749–752. IEEE, 2001. (Cité en pages 3, 75, 77 et 84.)

[So 2010] Stephen So et Kuldip K Paliwal. *Fast converging iterative Kalman filtering for speech enhancement using long and overlapped tapered windows with large side lobe attenuation*. In Eleventh Annual Conference of the International Speech Communication Association, 2010. (Cité en pages 2 et 84.)

[Van Der Merwe 2004] Rudolph Van Der Merwe. *Sigma-point Kalman filters for probabilistic inference in dynamic state-space models*. 2004. (Cité en page 46.)

[Voiers 1980] WD Voiers. *Interdependencies among measures of speech intelligility and speech" Quality"*. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'80., volume 5, pages 703–705. IEEE, 1980. (Cité en page 74.)

[Welling 1996] Lutz Welling et Hermann Ney. *A model for efficient formant estimation*. In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, volume 2, pages 797–800. IEEE, 1996. (Cité en page 54.)

[Werbos 1990] Paul J Werbos. *Backpropagation through time : what it does and how to do it*. Proceedings of the IEEE, vol. 78, no. 10, pages 1550–1560, 1990. (Cité en page 66.)

[Williams 1992] Ronald J Williams. *Training recurrent networks using the extended Kalman filter*. In Neural Networks, 1992. IJCNN., International Joint Conference on, volume 4, pages 241–246. IEEE, 1992. (Cité en page 67.)

[Witzke 1994] Lillian I Witzke, RJ Niederjohn, JA Heinen et MG Sommerville. *Speech synthesis based on feature extraction to enhance noise-corrupted speech*. In Industrial Electronics, Control and Instrumentation, 1994. IECON'94., 20th International Conference on, volume 3, pages 1946–1951. IEEE, 1994. (Cité en page 8.)

[Wójcicki 2008] Kamil Wójcicki, Mitar Milacic, Anthony Stark, James Lyons et Kuldip Paliwal. *Exploiting conjugate symmetry of the short-time Fourier spectrum for speech enhancement*. IEEE Signal processing letters, vol. 15, pages 461–464, 2008. (Cité en pages 3, 13, 14, 15 et 84.)

[Wu 2005] Yuanxin Wu, Dewen Hu, Meiping Wu et Xiaoping Hu. *Unscented Kalman filtering for additive noise case : augmented vs. non-augmented*. In American Control Conference, 2005. Proceedings of the 2005, pages 4051–4055. IEEE, 2005. (Cité en page 43.)

[Wu 2006] Yuanxin Wu, Dewen Hu, Meiping Wu et Xiaoping Hu. *A numerical-integration perspective on Gaussian filters*. IEEE Transactions on Signal Processing, vol. 54, no. 8, pages 2910–2921, 2006. (Cité en page 46.)

[Yan 2005] Qin Yan, Saeed Vaseghi, Esfandiar Zavarehei et Ben Milner. *Formant-tracking linear prediction models for speech processing in noisy environments*. In Ninth European Conference on Speech Communication and Technology, 2005. (Cité en page 54.)