

UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

Extension de bande paramétrique pour les signaux audio de parole et de
musique

Mémoire de maîtrise
Spécialité : Génie électrique

Jean-Philippe GOUIN

Jury : Roch LEFEBVRE (directeur)
Philippe GOURNAY (co-directeur)
Martin BOUCHARD
Éric PLOURDE

RÉSUMÉ

Dans le domaine du codage audio, l'utilisation d'une extension de bande paramétrique pour encoder les hautes fréquences permet une économie de débit. Contrairement à la majorité des méthodes d'extension de bande qui effectuent leur traitement dans le domaine temporel, le traitement proposé dans ce mémoire s'effectue dans le domaine fréquentiel pour créer une extension de bande audio de 8 à 16 kHz sur les signaux monophoniques de parole et de musique. Le projet de maîtrise étudie plusieurs éléments : les techniques de régénération des hautes fréquences, l'utilisation du cepstre pour le calcul et la représentation de l'enveloppe spectrale, ainsi que des techniques d'amélioration de la dynamique du spectre. Les résultats de tests subjectifs formels montrent un gain sur certaines catégories de signaux par rapport au codec AMR-WB+, lequel est reconnu comme étant un standard international développé à l'Université de Sherbrooke.

Mots clés : compression, codec, audio, extension, domaine fréquentiel, cepstre

REMERCIEMENTS

Je tiens, tout d'abord, à remercier mes parents qui m'ont soutenu tout au long de mes études. Ce sont eux qui m'ont soutenu et encouragé à terminer cette maîtrise qui aura duré plusieurs années. Mes enfants Charles-Antoine et Lily-Ann qui ont su comprendre que papa devait travailler beaucoup pour écrire son livre de 100 pages. Le plus difficile pour eux aura été la période de rédaction.

Un merci à ma conjointe Catherine qui a accepté le fait que de terminer ma maîtrise était important pour moi et même à l'occasion, plus important que de passer du temps avec elle. Mon ex-conjointe, la mère de mes enfants, Marie-Claude qui a été présente dans ma vie durant une partie de ma maîtrise.

Mon professeur Roch Lefebvre pour avoir cru en moi et mes capacités de mener ce projet à terme. Un merci particulier à mon codirecteur Philippe Gournay, professeur associé, qui m'a soutenu dans ma maîtrise. Je l'ai interpellé à plusieurs reprises et il m'a toujours répondu sans jamais me dire qu'il n'avait pas le temps.

Finalement, mes collègues du groupe de recherche qui m'ont supporté et aidé durant la maîtrise. Je me rappellerai longtemps nos dîners au Pot-au-Feu et au St-Malo : Bruno Bessette, Claude Laflamme, Danielle Poirier, Milan Jelinek, Frédérick Lavoie, Guy Richard, Jimmy Lapierre, Philippe Gournay, Roch Lefebvre, Tommy Vaillancourt, Vaclav Eskler, Vilayphone Vilaysouk et Vladimir Malenovsky. J'oublie sûrement quelques amis et collègues de travail qui m'ont supporté durant ma maîtrise, merci.

TABLE DES MATIÈRES

CHAPITRE 1	INTRODUCTION	1
1.1	Techniques d'extension de bande	1
1.2	Description du projet de maîtrise	2
1.3	Questions de recherche	2
1.4	Objectifs	3
1.5	Contributions	3
1.6	Organisation du mémoire	4
CHAPITRE 2	ÉTAT DE L'ART	5
2.1	Caractéristiques des signaux	5
2.1.1	Fréquence fondamentale	5
2.1.2	Enveloppe temporelle et la structure fine temporelle	5
2.1.3	Enveloppe spectrale et signal blanchi	6
2.1.4	Stationnarité	6
2.2	Codec	7
2.3	Codeur temporel	7
2.4	Codeur par transformée	8
2.5	Codage perceptuel	8
2.6	Extension de bande	10
2.6.1	High Frequency Regeneration (HFR)	11
2.6.2	Bandwidth Extension (AMR-WB+)	12
2.6.3	Non-linéarité pour régénérer les HF	14
2.6.4	Estimation de l'enveloppe spectrale	15
2.6.5	Correction de l'enveloppe estimée	15
2.6.6	Modèle paramétrique sinusoïdale	16
2.6.7	Spectral Band Replication (SBR)	16
2.6.8	Enhanced Voice Service (EVS)	17
2.6.9	Générer les HF	20
2.7	Conclusion	20
CHAPITRE 3	MÉTHODOLOGIE	23
3.1	Choix préliminaires	23
3.1.1	Choix du domaine de traitement	23
3.1.2	Codec de référence	23
3.2	Développement	23
3.3	Test	25
3.3.1	Test A/B	25
3.3.2	Test MUSHRA	28
3.3.3	Résumé des requis et caractéristiques des tests subjectifs	29
3.4	Support technique disponible	30
CHAPITRE 4	MODÈLE PARAMÉTRIQUE PROPOSÉ	33
4.1	Structure du modèle	33
4.2	Trames, fenêtrage et recouvrement	34

4.2.1	Fenêtrage	35
4.2.2	Recouvrement	36
4.2.3	Longueur de trame	38
4.2.4	Choix de la longueur de trame	39
4.3	Générer les HF	40
4.4	Changement de l'enveloppe spectrale	42
4.5	Ajout et retrait de tonalité	44
4.6	Ajustement de la variance	45
4.7	Enveloppe temporelle	47
4.7.1	Modèle sans enveloppe temporelle	47
4.7.2	Modèle avec enveloppe temporelle	47
4.8	Résumé du modèle proposé	48
CHAPITRE 5 TECHNIQUES D'AMÉLIORATION		49
5.1	Calcul de l'enveloppe spectrale	49
5.1.1	Calcul d'enveloppe par filtrage à moyenne mobile	49
5.1.2	Calcul d'enveloppe par sous-bandes	50
5.1.3	Calcul d'enveloppe par filtrage LPC	52
5.1.4	Calcul d'enveloppe par cepstre	54
5.1.5	Sensibilité des coefficients au bruit de quantification	59
5.1.6	Comparaison de la performance	61
5.2	Générer les HF	64
5.2.1	Symétrie et translation	64
5.2.2	Non-linéarité	66
5.2.3	Ajout de bruit blanc gaussien	67
5.2.4	Somme de suréchantillonnages	68
5.2.5	Méthode de sélection	68
5.3	Amélioration du signal blanchi	72
5.3.1	Variance	72
5.3.2	Changement de variance	73
5.3.3	Courbe de variances	74
5.3.4	Comparaison des performances	76
5.4	Résumé des outils proposés	77
CHAPITRE 6 QUANTIFICATION ET COMPRESSION DES DONNÉES		79
6.1	Données à compresser	79
6.2	Quantification de l'enveloppe spectrale	81
6.2.1	Recherche du plus proche voisin dans les domaines cepstral et fréquentiel	82
6.3	Compression de la variance du signal blanchi	83
6.4	Budget de bits	85
6.5	Résumé de la quantification	86
CHAPITRE 7 RÉSULTATS ET ANALYSES		87
7.1	Test A/B	87
7.2	Analyse des résultats du test A/B	90
7.3	Test MUSHRA	91
7.4	Analyse des résultats du test MUSHRA	93
7.5	Résumé des analyses des résultats	95

CHAPITRE 8	DISCUSSION	97
8.1	Discussion du modèle paramétrique	97
8.2	Discussion des résultats	98
8.2.1	Discussion des performances sur les fichiers de musique	98
8.2.2	Discussion des performances sur les fichiers de parole	98
8.2.3	Discussion des performances sur les fichiers de chant sur musique	99
8.2.4	Discussion des performances sur les fichiers de parole sur musique	99
8.2.5	Discussion des performances pour tous les fichiers	100
8.2.6	Résumé des discussions	100
8.3	Améliorations possibles	100
8.3.1	Débit	101
8.3.2	Modèle paramétrique	102
CHAPITRE 9	CONCLUSION	103
9.1	Contributions	105
RÉFÉRENCES	107	
ANNEXE A	109	
ANNEXE B	110	

LISTE DES FIGURES

Figure 1-1 : Diagramme de codec utilisant une extension de bande	1
Figure 2-1 : Seuil d'audition en dB en fonction des fréquences en Hz	9
Figure 2-2 : Masquage fréquentiel avec seuil d'audition	9
Figure 2-3 : Masquage temporel	10
Figure 2-4 : Schéma de l'encodeur du HFR	11
Figure 2-5 : Schéma du décodeur du HFR	12
Figure 2-6 : Schéma de l'encodeur du BWE de l'AMR-WB+	13
Figure 2-7 : Schéma du changement d'enveloppe spectrale HF par filtre LPC	13
Figure 2-8 : Schéma du décodeur du BWE de l'AMR-WB+	14
Figure 2-9 : Schéma du décodeur avec non-linéarité pour créer les HF	15
Figure 2-10 : Diagramme de la quantification vectorielle avec classification, 1 bit par trame	16
Figure 2-11 : Schéma de l'encodeur SBR	17
Figure 2-12 : Schéma du décodeur SBR	17
Figure 2-13 : Schéma du décodeur de l'extension de bande du EVS par réseau de neurones	18
Figure 2-14 : Schéma du décodeur de l'extension de bande du EVS par DCT	19
Figure 2-15 : Schéma du décodeur de l'extension de bande du EVS avec filtre QMF	20
Figure 3-1 : Interface de test A/B pour les votes	27
Figure 3-2 : Interface de test A/B pour les questions	28
Figure 3-3 : Interface de test MUSHRA	29
Figure 4-1 : Structure du codec avec extension en parallèle	33
Figure 4-2 : Structure du codec avec extension en série	33
Figure 4-3 : Schéma de l'encodeur de l'UFR	34
Figure 4-4 : Schéma du décodeur de l'UFR	34
Figure 4-5 : Caractéristiques temporelles des fenêtres	35
Figure 4-6 : Caractéristiques fréquentielles des fenêtres en dB	36
Figure 4-7 : Fenêtre de Hanning avec recouvrement de 50%	37
Figure 4-8 : Fenêtre de Hanning modifiée avec recouvrement de 25%	37
Figure 4-9 : Fenêtres de Hanning courtes modifiées avec recouvrement de 128 échantillons	38
Figure 4-10 : Fenêtres de Hanning courtes modifiées et combinées à recouvrement de 50%	39
Figure 4-11 : Diagramme du choix de la longueur de trame	40
Figure 4-12 : Module du spectre original	41
Figure 4-13 : Module du spectre original et généré HF par la symétrie du spectre	41
Figure 4-14 : Module du spectre original avec son enveloppe HF	42
Figure 4-15 : Module du spectre original et généré par la symétrie avec son enveloppe HF	43
Figure 4-16 : Module du spectre original et du signal blanchi généré par la symétrie	43
Figure 4-17 : Module du spectre original et coloré généré par la symétrie	44
Figure 4-18 : Module du spectre blanchi original et généré par la symétrie	45
Figure 4-19 : Module du spectre blanchi original et généré par la symétrie avec variance	46
Figure 4-20 : Module du spectre original et généré par la symétrie avec variance	46
Figure 5-1 : Module du spectre original avec son enveloppe HF par filtrage à moyenne mobile et le signal blanchi HF	50
Figure 5-2 : Module du spectre du signal avec son enveloppe HF par 5 sous-bandes et le signal blanchi HF	51
Figure 5-3 : Module du spectre avec son enveloppe HF par 5 sous-bandes interpolées linéairement et le signal blanchi HF	51

Figure 5-4 : Module du spectre avec son enveloppe HF par 5 sous-bandes interpolées par spline cubique et le signal blanchi HF	52
Figure 5-5 : Module du spectre avec son enveloppe HF par LPC et le signal blanchi HF	53
Figure 5-6 : Module du spectre avec son enveloppe HF par cepstre et le signal blanchi HF	55
Figure 5-7 : Module du spectre du signal de parole Es02 avec ses enveloppes spectrales avec et sans symétrie	56
Figure 5-8 : Module du spectre du signal de musique Brahms avec ses enveloppes spectrales avec et sans symétrie	56
Figure 5-9 : Module du spectre avec les enveloppes spectrales par cepstre avec 3, 5 et 7 coefficients pour le signal de parole Es02	57
Figure 5-10 : Module du spectre avec les enveloppes spectrales par cepstre avec 3, 5 et 7 coefficients pour le signal de musique Brahms	58
Figure 5-11 : Signal blanchis par cepstre avec 3, 5 et 7 coefficients pour le signal de parole Es02	58
Figure 5-12 : Signal blanchis par cepstre avec 3, 5 et 7 coefficients pour le signal de musique Brahms	59
Figure 5-13 : Module du spectre original avec l'enveloppes spectrales originale et les enveloppes spectrales bruitées	60
Figure 5-14 : Module du spectre original avec les enveloppes spectrales par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de parole Es02.	62
Figure 5-15 : Module du spectre original avec les enveloppes spectrales par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de musique Brahms.	62
Figure 5-16 : Signaux blanchis par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de parole Es02.	63
Figure 5-17 : Signaux blanchis par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de musique Brahms	63
Figure 5-18 : Module du spectre original et généré HF par la symétrie	65
Figure 5-19 : Module du spectre original et généré HF par la translation	65
Figure 5-20 : Module du spectre original et généré HF par la double symétrie	66
Figure 5-21 : Module du spectre original et généré HF par la double translation	66
Figure 5-22 : Module du spectre original et généré HF par la convolution	67
Figure 5-23 : Module du spectre original et généré avec bruit blanc	68
Figure 5-24 : Module du spectre original et généré par suréchantillonnages multiples	68
Figure 5-25 : Statistique de la performance des techniques de régénération de signaux HF	70
Figure 5-26 : Distribution des techniques entre la symétrie et la double translation	71
Figure 5-27 : Distribution des techniques entre la translation et la symétrie	71
Figure 5-28 : Distribution des techniques entre la translation et la double translation	72
Figure 5-29 : Signal blanchi original et généré sans modification de la variance	72
Figure 5-30 : Signal blanchi original et généré avec modification de la variance	74
Figure 5-31 : Signal blanchi original et généré sans modification avec courbes de variances	75
Figure 5-32 : Signal blanchi original et généré modifié par les courbes de variances	75
Figure 5-33 : Signal blanchi original et généré modifié par les courbes de variances interpolées	76
Figure 5-34 : Module du spectre original et final par variances interpolées	76
Figure 6-1 : Distributions des quéfrences de 0 à 4	80
Figure 6-2 : Distributions des variances de 1 à 4	80
Figure 6-3 : Enveloppe spectrale par cepstre avec estimation de l'amplitude	82
Figure 6-4 : Signal blanchi original et généré	84
Figure 6-5 : Signal blanchi original et généré avec courbe de variances estimées	85
Figure 7-1 : Résultats du test A/B par catégorie de fichiers	90

Figure 7-2 : Résultats du test MUSHRA par catégorie par rapport au signal original	92
Figure 7-3 : Différences de points du test MUSHRA entre les signaux UFR(1,2,3) et AMR-WB+ par catégorie par rapport au signal original	92
Figure 7-4 : Résultats du test MUSHRA par catégorie par rapport au signal AMR-ORI	94
Figure 7-5 : Différences de points du test MUSHRA entre les signaux UFR(1,2,3) et AMR-WB+ par catégorie par rapport au signal AMR-ORI	94

LISTE DES TABLEAUX

Tableau 3-1 : Tableau des caractéristiques des tests d'écoute	30
Tableau 3-2 : Tableau des logiciels disponibles	30
Tableau 5-1 : Comparaison des moyennes de la distorsion des enveloppes et coefficients bruités	61
Tableau 6-1 : Allocation du budget de bits pour la compression	86
Tableau 7-1 : Groupes d'auditeurs selon les critères de sélections du test A/B	89
Tableau 7-2 : Moyennes des écarts par catégorie pour le test A/B pour le groupe d'auditeurs ayant réussi à trouver 6/8 fichiers originaux, n'ayant pas de vote aléatoire	90
Tableau a-1 : Séquence des fichiers sonores et questions du test A/B	109
Tableau b-2 : Points par catégorie pour le test MUSHRA par rapport au signal original	110
Tableau b-3 : Points par catégorie pour le test MUSHRA par rapport au signal AMR-ORI	110
Tableau b-4 : Écarts par catégorie pour le test MUSHRA par rapport au signal original	111
Tableau b-5 : Écarts par catégorie pour le test MUSHRA par rapport au signal AMR-ORI	111

LEXIQUE

Noms	Définitions
Cepstre	Transformée de Fourier du logarithme du module d'un spectre
CODEC	COdeur et DÉCodeur
Liffrage	Filtrage par cepstre
Quéfrencence	Élément d'un cepstre ou coefficient cepstral

LISTE DES SYMBOLES

Symboles	Définitions
σ	Écart-type
V	Variance
$V_{HF-estimé}$	Variance estimée
\hat{V}	Erreur d'estimation de la variance
\bar{x}	Moyenne de x
I	Intervalle de confiance
$x[n]$	Signal temporel discret (échantillon)
$X[k]$	Coefficient spectral discret (fréquence)
$c[\tau]$	Coefficient cepstral discret (quéfrency)
$\hat{c}[0]$	Erreur d'estimation du coefficient cepstral zéro
\mathcal{F}	Transformée de Fourier discrète
\mathcal{F}^{-1}	Transformée de Fourier discrète inverse

LISTE DES ACRONYMES

Acronymes	Définitions
3GPP	3rd Generation Partnership Project
ACELP	Algebraic Code-Excited Linear Prediction
ABE	Artificial Bandwidth Extension
AMR-WB+	Adaptative Multi-Rate WideBand Plus
BF	Basse fréquence
BWE	Bandwidth Extension (AMR-WB+)
LPC	Coefficients de prédiction linéaire Linear Predictive Coefficient
EQM	Erreur quadratique moyenne
EVS	Enhanced Voice Services
HE-AAC+	High-Efficiency Advanced Audio Coding plus
HF	Haute fréquence
HFR	High Frequency Regeneration
LPC	Codage à prédiction linéaire Linear predictive coding
LSF	Line Spectral Frequencies
MBE	Multiband excitation
MUSHRA	Stimuli multiples avec référence et ancrage cachés Multiple Stimuli with Hidden Reference and Anchor
MVP	Produit Minimum Viable Minimum Viable Product
QMF	Quadrature Mirror Filter-bank
RIF	Réponse Impulsionnelle Finie
RII	Réponse Impulsionnelle Infinie
RSB	Rapport signal sur bruit
SBR	Spectral Band Replication
TCD	Transformée en cosinus discrète
TFD	Transformée de Fourier discrète
TFDI	Transformée de Fourier discrète inverse
UFR	Upper Frequency Regeneration

CHAPITRE 1 INTRODUCTION

L'oreille humaine n'accorde pas la même importance à toutes les fréquences audibles. La majeure partie du contenu utile se trouve dans les signaux en dessous de 8 kHz. Les codeurs de parole et de musique utilisent cette propriété en limitant la bande passante des signaux qu'ils encodent et compressent afin de minimiser la quantité d'information à transmettre. Cette technique d'élimination des hautes fréquences (HF) n'affecte pas la compréhension du message. Cependant, la qualité du message décodé est beaucoup moindre. Pour remédier à ce problème, il existe des techniques d'extension de bande permettant d'élargir la bande passante du signal reconstitué afin d'augmenter l'impression de qualité du signal transmis. Ces techniques d'extension de bande génèrent un signal artificiel permettant de doubler, tripler ou quadrupler la largeur de bande. En se basant sur la corrélation entre les signaux de basses fréquences (BF) et de HF, les algorithmes d'extension de bande nécessitent un très faible débit en comparaison avec les débits utilisés par les codeurs de base. La figure 1-1 montre le diagramme d'un codec utilisant une extension de bande.

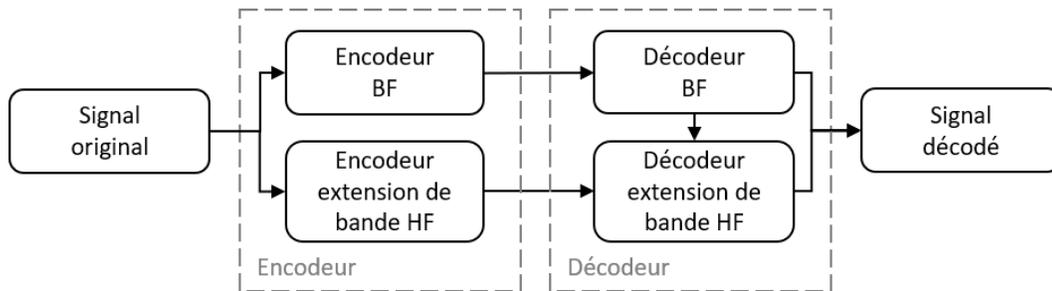


Figure 1-1 : Diagramme de codec utilisant une extension de bande

1.1 Techniques d'extension de bande

Les algorithmes d'extension de bande s'exécutent, pour la plupart, dans le domaine temporel. Ils utilisent des traitements sur les échantillons temporels et des filtres pour modifier le signal temporel BF afin de générer les signaux manquants en HF. Certaines extensions de bande traitent les données dans le domaine fréquentiel, mais elles sont plutôt l'exception. Ces dernières utilisent aussi des filtres et effectuent des traitements sur les données fréquentielles du signal.

Les extensions de bande paramétriques extraient et transmettent des paramètres du signal HF original afin que le décodeur puisse reconstruire le signal HF. Ces paramètres sont généralement un sous-ensemble de ceux-ci : l'enveloppe temporelle, l'enveloppe spectrale, les bandes de fréquences à

1.2 Description du projet de maîtrise

répliquer, des tonalités à ajouter, des niveaux de bruit et des niveaux d'harmonicité. Par la suite, les coefficients de ces paramètres sont quantifiés et compressés pour la transmission.

1.2 Description du projet de maîtrise

Le but du projet de maîtrise est de concevoir un nouvel algorithme d'extension de bande pour le codec AMR-WB+ dans le domaine fréquentiel dont le débit est paramétrable selon la qualité souhaitée. Cette extension de bande est créée spécifiquement pour ce codec dans le but d'améliorer la qualité du signal reconstruit. Le domaine fréquentiel offre un degré de liberté plus grand comparativement au domaine temporel. Le détail de cette affirmation est fait dans les prochains chapitres.

Tout d'abord, la synthèse du signal HF peut se faire par de multiples méthodes : symétrie, translation, double symétrie, double translation, addition de multiples interpolations, ajout de bruit blanc gaussien paramétrable. Le projet de maîtrise a pour but d'explorer toutes ces méthodes et soit d'en déterminer la meilleure ou de transmettre une méthode de synthèse parmi celles développées selon le type de signal.

Il existe plusieurs techniques de calcul d'enveloppe spectrale pour le changement d'enveloppe spectrale. La technique de calcul d'enveloppe choisie dans le projet de maîtrise est par l'analyse du cepstre. Le projet de maîtrise a comme mandat d'évaluer la performance de cette technique en comparaison avec les autres méthodes de la littérature.

De plus, l'ajustement de la dynamique du signal par l'ajout et l'atténuation de fréquences est un degré de liberté supplémentaire du domaine fréquentiel.

En conclusion, le projet de maîtrise a comme objectif d'évaluer la performance d'un algorithme d'extension de bande dans le domaine fréquentiel dont le débit est paramétrable selon la qualité désirée.

1.3 Questions de recherche

Les questions de recherche portent sur le domaine du traitement des signaux, le codage paramétrique et l'évaluation des performances de l'extension de bande développée par des tests audio subjectifs.

Dans le cadre du codage paramétrique des signaux :

- Est-ce que les paramètres de modélisation utilisés dans les modèles d'extension de bande sont les mêmes pour les signaux de parole et de musique?

- Est-il possible de créer un modèle d’extension de bande compatible avec plusieurs codeurs de base?
- Dans le domaine fréquentiel, est-ce que la représentation des signaux offre plus de degrés de liberté lors de la conception d’une extension de bande?
- À débit égal, est-il possible d’atteindre ou de dépasser la qualité de l’extension de bande de référence (AMR-WB+) en utilisant une représentation fréquentielle des signaux ?

Dans le contexte de la pandémie de la Covid-19, il est impossible de respecter les protocoles de tests audio subjectifs conventionnels. Il a été proposé de concevoir un nouveau protocole de test. Dans le cadre d’un test audio subjectif large public sans contrôle du matériel et de l’environnement de test :

- Quels sont les moyens de discriminer les résultats des auditeurs selon leur compétences et selon la validité de leur résultats?
- Les données obtenues permettent-elles d’obtenir les mêmes conclusions qu’un test dont le matériel et l’environnement sont contrôlés?

1.4 Objectifs

L’objectif principal de la maîtrise est de concevoir une extension de bande permettant de corriger des lacunes dans l’AMR-WB+. Les sous-objectifs ci-dessous guident le projet de maîtrise afin d’en améliorer les performances du standard.

- Créer une extension de bande pour l’AMR-WB;
- Atteindre les performances de l’AMR-WB+ avec un débit inférieur ou égal;
- Avoir de bonnes performances subjectives pour la parole et la musique;
- Évaluer les techniques de régénération des HF;
- Avoir un débit ajustable selon la qualité désirée;
- Concevoir un protocole de test audio subjectif large public sans contrôle du matériel et de l’environnement de test.

1.5 Contributions

Les contributions de cette maîtrise, au domaine scientifique, sont multiples. Les contributions scientifiques se trouvent principalement dans les outils et les techniques développées pour mettre en œuvre le codec. Cette maîtrise a permis d’explorer une avenue qui n’avait pas encore été documentée. Les innovations sont multiples :

1.6 Organisation du mémoire

- Développer et comparer des techniques de représentation paramétrique des signaux;
- Développer et comparer des méthodes de régénération des HF : Symétrie, translation, utilisation de non-linéarité, suréchantillonnage multiple et ajout de bruit;
- Développer et comparer des techniques de calcul et de représentation de l'enveloppe spectrale;
- Développer et comparer des techniques d'amélioration des signaux HF tels : filtrage, variance statistique, ajout de tonalités manquantes, modification de l'enveloppe;
- Développer un protocole de tests audio subjectifs A/B large public sans contrôle du matériel et de l'environnement.

1.6 Organisation du mémoire

Après cette introduction, le chapitre 2 présente l'état de l'art sur les sujets des codecs audio et des extensions de bandes. Le chapitre 3 présente la définition, les objectifs et la méthodologie du projet de maîtrise. Au cœur même du document, les chapitres 4 à 6 présentent la conception en partant des caractéristiques des signaux HF, le modèle paramétrique de l'extension de bande, les techniques d'amélioration de la qualité du signal, la quantification et la compression. Pour finir, les chapitres 7 à 9 présentent les résultats, la discussion et la conclusion.

CHAPITRE 2 ÉTAT DE L'ART

Ce chapitre présente l'état de l'art des techniques d'extension de bande audio. Il présente les différentes techniques de codage et d'algorithmes utilisés depuis les 50 dernières années pour élargir la bande des signaux audio par des représentations paramétriques et peu coûteuses en débit numérique.

2.1 Caractéristiques des signaux

Les signaux audio sont catégorisés selon la nature du son, ce qui leur donne des caractéristiques temporelles et fréquentielles spécifiques. Les signaux de musique sont subdivisés en deux catégories, soit : harmoniques ou non harmoniques. Les signaux et les plosives de parole sont catégorisés en trois types : les signaux voisés, les signaux fricatifs et les signaux voisés fricatifs. Les signaux voisés et les signaux de musique harmonique sont composés d'une fréquence fondamentale et de ses harmoniques. Les signaux fricatifs sont en général constitués de bruit blanc gaussien paramétrable et de gains [1], [2], [3]. La majorité des signaux ne peuvent pas être catégorisés aussi simplement et sont de manière générale un mélange de plusieurs catégories.

2.1.1 Fréquence fondamentale

Chaque signal audio harmonique a une fréquence fondamentale. La parole et la musique sont majoritairement harmoniques ou poly-harmoniques. La fréquence porteuse est définie par un pic de forte intensité et ses harmoniques par la répétition des pics aux fréquences de multiples entiers de la fondamentale dans le domaine fréquentiel. Dans le domaine temporel, l'autocorrélation offre les mêmes caractéristiques; cependant, la répétition est présente à toutes les périodes de la fréquence fondamentale.

2.1.2 Enveloppe temporelle et la structure fine temporelle

L'amplitude des signaux audio varie en fonction du temps. L'enveloppe temporelle correspond au volume du signal que nos oreilles perçoivent. L'enveloppe peut être retirée du signal pour en garder la structure fine temporelle composée de la porteuse modulée par les fréquences typiques au timbre des instruments ou du locuteur [4].

L'enveloppe temporelle joue un rôle important dans la perception des sons de parole et de musique. Les variations rapides de l'enveloppe temporelle contiennent les informations des attaques et des coupures des sons. Tandis que les variations lentes sont reliées aux tempos et à la rythmique. Dans les

2.1 Caractéristiques des signaux

signaux de parole, les variations lentes (16 Hz et moins) sont aussi associées aux changements de syllabes [5].

La structure fine joue plusieurs rôles dans la perception de la hauteur des sons, du timbre, ainsi que dans l'identification du locuteur et la prosodie (ensemble des traits oraux d'une expression verbale pour transmettre les émotions dont : l'inflexion, le ton, la tonalité, l'intonation, l'accent et la modulation) [4].

2.1.3 Enveloppe spectrale et signal blanchi

Les ondes sonores sont représentées dans le temps selon la pression qu'elles appliquent sur notre système auditif. La transformée de Fourier discrète (TFD) donnée par l'équation (2-1) permet de transposer l'information temporelle dans le domaine des fréquences et l'inverse de la transformée de Fourier discrète inverse (TFDI) est donnée par l'équation (2-2). Ces transformations sont des projections sur les cosinus et les sinus complexes selon le théorème d'Euler (2-3).

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-\frac{j2\pi kn}{N}} \quad \text{pour } 0 \leq k \leq N-1 \quad (2-1)$$

$$x[n] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} X[k] \cdot e^{\frac{j2\pi kn}{N}} \quad \text{pour } 0 \leq n \leq N-1 \quad (2-2)$$

$$e^{j\theta} = \cos \theta + j \sin \theta \quad \text{et} \quad e^{-j\theta} = \cos \theta - j \sin \theta \quad (2-3)$$

Les données du spectre fréquentiel sont des nombres complexes qui peuvent être représentés selon une représentation polaire (module et phase). Le domaine des fréquences permet de représenter l'amplitude et la phase des fréquences présentes dans le signal.

L'enveloppe spectrale est calculée sur le module du signal en dB. Elle correspond à l'amplitude moyenne locale du signal qui représente l'allure générale du spectre. Le blanchissage est l'opération qui consiste à diviser le module du spectre par son enveloppe spectrale, ou à soustraire l'enveloppe spectrale en dB du module du spectre en dB. Le signal blanchi résultant permet plusieurs analyses dont le calcul de la fréquence fondamentale et des formants. Plusieurs manipulations sur le signal blanchi sont utilisées dans les algorithmes de compression des signaux [2].

2.1.4 Stationnarité

La stationnarité d'un signal est définie comme suit : un signal est dit stationnaire, si ses propriétés statistiques caractérisées par des espérances mathématiques ne varient pas dans le temps. Les signaux

de parole peuvent être considérés stationnaires pour des périodes d'environ 10 à 30 millisecondes. Quant aux signaux de musique, la période est plus variable, elle varie entre 5 et 80 millisecondes [1]. Ces caractéristiques statistiques considérées stationnaires sont : la fréquence fondamentale, l'enveloppe temporelle, l'enveloppe spectrale, la structure fine temporelle, le signal blanchi, la moyenne ainsi que la variance de certains de ces paramètres. Les algorithmes de compression paramétriques utilisent les caractéristiques stationnaires pour compresser les signaux.

2.2 Codec

Le mot codec provient de la contraction des mots codeur et décodeur. Dans le domaine de l'audio et de l'image, les codecs sont utilisés afin de diminuer la quantité d'informations nécessaires pour transmettre des signaux.

Les codecs peuvent transmettre les données avec ou sans perte. Un codeur sans perte signifie que mathématiquement, il n'y a aucune différence entre le signal original et le signal décodé. Un codeur avec perte introduit des déformations dans le signal décodé soit par des pertes d'informations ou par la création d'artefacts. Ces pertes et artefacts sont évalués par rapport au signal original en calculant par exemple l'erreur quadratique moyenne (EQM).

Puisque le système auditif et le système visuel de l'humain ne sont pas parfaits, il est possible d'avoir un codec, avec pertes, sans que celles-ci ne soient perçues par l'humain. Ce niveau de reproduction du signal se nomme la transparence. Elle ne peut pas être évaluée, de façon fiable, par des opérations mathématiques tant pour la quantité de perte que pour la distorsion des signaux. Le niveau de transparence par rapport à l'original doit être évalué par des humains en effectuant des tests subjectifs.

Les types de codage sont nombreux et puisque chaque signal est unique, il existe une multitude de techniques permettant d'extraire les paramètres des signaux et de les compresser.

2.3 Codeur temporel

Par définition les signaux acoustiques sont des oscillations dans le temps. Il paraît donc intuitif de les encoder dans le domaine temporel. Cette technique fonctionne bien sur les signaux ayant des changements brusques et qui sont moins harmoniques. La parole est un bon exemple de l'utilisation de codeur temporel. Le codec *algebraic code-excited linear prediction* (ACELP), un des codecs les plus répandus sur la planète, est le standard utilisé dans la majorité des réseaux cellulaires pour transmettre les signaux de la parole.

2.4 Codeur par transformée

Le signal de parole comporte des caractéristiques stationnaires sur de courtes périodes [1]. Pour tirer parti de cette caractéristique, le signal est subdivisé, analysé, puis transmis par trames de 20 à 30 millisecondes. La technique utilisée pour compresser l'information est inspirée du système vocal de l'humain. Dans les modèles simples, le conduit vocal est modélisé par un filtre et les cordes vocales ainsi que le diaphragme lesquels sont représentés par une série d'impulsions [1]. Ainsi, pour chaque trame, les informations transmises au décodeur sont : les coefficients du filtre, la valeur de la fréquence fondamentale, les gains et le vecteur d'impulsions. Cette technique est à la base de plusieurs technologies utilisées en téléphonie.

2.4 Codeur par transformée

Contrairement aux codeurs temporels, inspirés de la source pour la parole, les codeurs par transformées pour les signaux audio sont inspirés du récepteur, soit l'appareil auditif. Dû au fait que l'oreille analyse les sons selon leur fréquence en hertz, les codeurs par transformées effectuent une transformation permettant de passer du domaine temporel au domaine fréquentiel. Les transformées utilisées sont multiples : transformée de Fourier, cosinus, sinus, ondelettes et quelques autres. Le signal est encodé selon les fréquences présentes dans le spectre fréquentiel. Comme dans les codeurs temporels, l'algorithme évalue la pertinence des informations à transmettre selon un filtre perceptuel et en fonction de leur amplitude. Le changement de domaine permet de représenter l'information de la même manière que l'oreille humaine analyse les sons reçus. Cette représentation ne permet pas de diminuer la quantité d'information, mais plutôt de faciliter le choix des informations à transmettre. Les concepts de codage perceptuel et les algorithmes de compression permettent d'obtenir un codec ayant le débit souhaité.

2.5 Codage perceptuel

L'oreille humaine n'étant pas parfaite, elle perçoit les sons de 20 Hz à près de 20 kHz. Elle est plus sensible à certaines fréquences, soit de 100 Hz à 8000 Hz, ce qui correspond aux signaux utiles pour bien comprendre une conversation. Le seuil d'audition en fonction des fréquences varie selon les individus et dont la largeur de bande devient plus étroite avec l'âge, voir la figure 2-1.

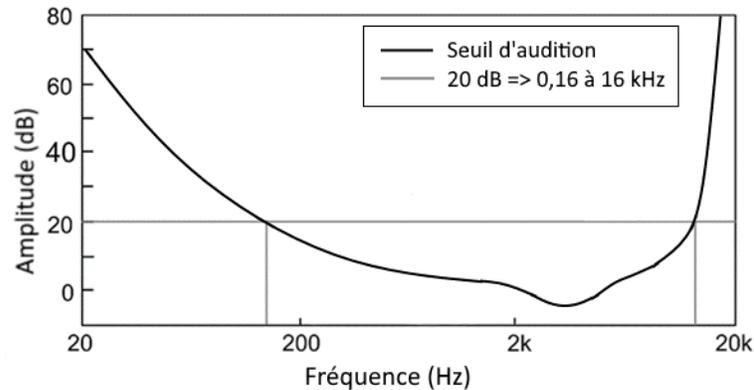


Figure 2-1 : Seuil d'audition en dB en fonction des fréquences en Hz

Le système sensoriel de l'oreille n'est pas seulement de moins en moins sensible à la puissance des sons au-delà de 2 kHz, mais il est aussi moins sensible à la phase et à la précision des fréquences reçues [6]. Puisque l'oreille est peu sensible aux sons aigus comparativement aux sons graves, les codecs utilisent cette caractéristique pour maximiser l'information à compresser.

Le codage perceptuel est inspiré de l'oreille humaine. Il revient à enlever ou cacher tous les signaux non audibles par l'oreille humaine. Cette opération permet de minimiser la quantité d'informations dans le signal audio. Il existe deux principaux types de masquages soit : le masquage fréquentiel et le masquage temporel.

Le masquage fréquentiel se produit lorsqu'une fréquence d'amplitude plus forte qu'une autre vient rendre inaudibles les fréquences voisines [6], [7], voir la figure 2-2.

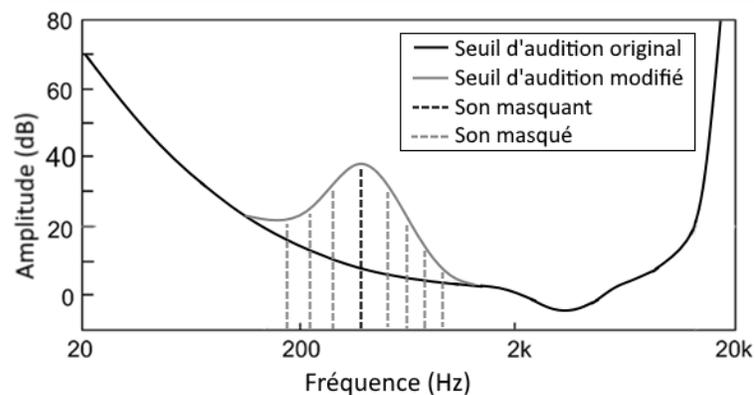


Figure 2-2 : Masquage fréquentiel avec seuil d'audition

Le masquage temporel survient lorsqu'un son est caché par un événement sonore plus fort. Il peut s'agir de deux événements sonores simultanés ou consécutifs. Le masquage survient lorsqu'un son fort réduit la sensibilité de l'oreille en fonction du temps. Deux sons courts sont alors perçus comme un seul son [7]. Un événement masquant affecte les sons passés sur un temps d'environ 20 millisecondes ainsi

2.6 Extension de bande

que les sons futurs de 100 à 200 millisecondes, voir la figure 2-3. Le masquage temporel utilisant les principes du masquage fréquentiel est optimal pour les sons de mêmes fréquences. Un son ayant une large distribution en fréquence masquera presque tous les autres sons [7]. Ces effets temporels sont dus au fait que le cerveau prend plus de temps pour traiter les sons faibles que pour les sons forts [6].

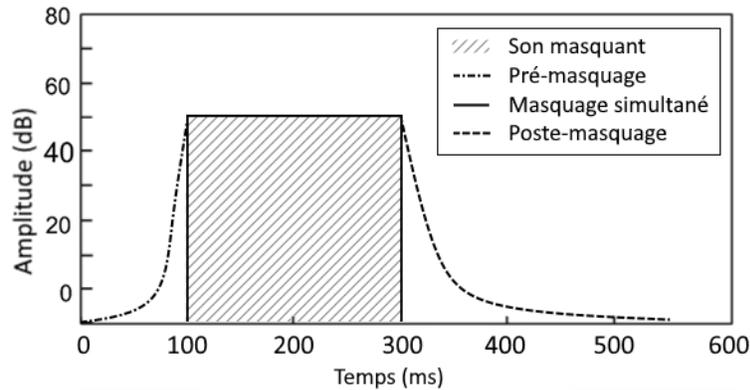


Figure 2-3 : Masquage temporel

En conclusion, les algorithmes d'extension de bande sont inspirés du codage perceptuel afin de facilement tromper l'oreille. Ils analysent le contenu fréquentiel des HF afin d'en extraire les caractéristiques principales et de régénérer le signal selon ses caractéristiques principales, soit : l'enveloppe temporelle et l'enveloppe fréquentielle.

2.6 Extension de bande

Les premiers travaux sur les extensions de bande pour la parole sont apparus dans les années 70. Une des premières problématiques était d'élargir la largeur de bande des systèmes téléphoniques. Ces systèmes fonctionnent en bande étroite (300 à 3300 Hz). Cette caractéristique perçue limite tant la qualité du signal que sa compréhension. Afin de résoudre ce problème, le principe d'extension de bande a été inventé pour doubler la bande passante des signaux et augmenter la qualité des signaux de parole. Le défi est de trouver un moyen de régénérer le signal manquant sans avoir à transmettre d'information. Plusieurs techniques permettent de régénérer les signaux manquants : ajout de bruit, symétrie ou translation fréquentielle du spectre avec ajustement de l'enveloppe spectrale [3], [8]. Le choix d'enveloppe spectrale se fait avec un dictionnaire associatif selon des paramètres extraits des BF [9], [10]. À la suite de ces recherches, les extensions de bande ont été appliquées à des codeurs de musique afin d'en diminuer le débit alloué aux HF.

2.6.1 High Frequency Regeneration (HFR)

Dans les années 70, les extensions de bande pour la parole étaient formées de bruit blanc modulé par une enveloppe spectrale. Du côté de l'encodeur, voir la figure 2-4, le signal temporel $x[t]$ est filtré passe-haut, puis transféré en BF par une symétrie du spectre en appliquant l'équation (2-4) au signal $x_{HF}[t]$. Cette étape permet d'éliminer les BF. Afin d'extraire des paramètres de modélisation, le signal $x'_{HF}[t]$ est subdivisé en trame de 15 à 30 millisecondes, puis fenêtré. Pour chaque trame, l'enveloppe spectrale du signal est modélisée par le filtre de codage à prédiction linéaire (LPC). Le calcul des coefficients du filtre permet une représentation compacte de l'enveloppe spectrale. Finalement, les coefficients du filtre sont compressés et transmis au récepteur.

$$x'[n] = x[n] \cdot e^{j\pi n} = x[n] \cdot (-1)^n \quad (2-4)$$

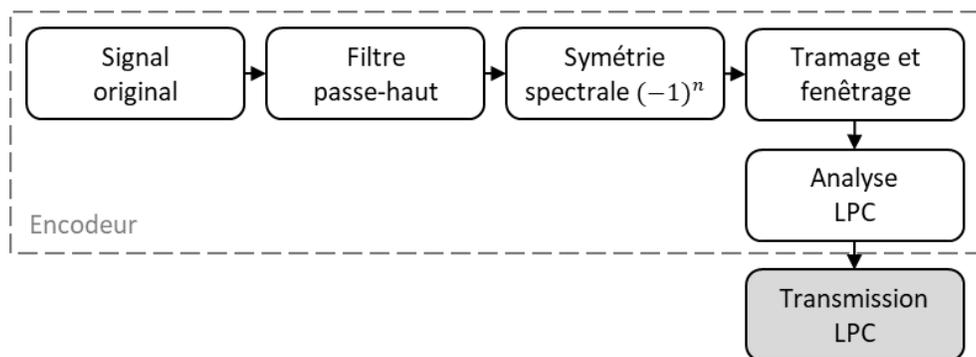


Figure 2-4 : Schéma de l'encodeur du HFR

Au récepteur, voir la figure 2-5, le décodeur décompresse et suréchantillonne le signal BF reçu. Pour créer la partie HF manquante, un bruit blanc gaussien est généré, puis subdivisé en trame et fenêtré selon la même durée qu'à l'encodeur. Pour chaque trame, le bruit est filtré par l'inverse du filtre LPC afin de changer l'enveloppe spectrale. Par la suite, le signal est transféré en HF, par une symétrie du spectre, en appliquant (2-4) au signal. Finalement, le signal BF est additionné au signal HF conditionné.

Plus tard, deux nouvelles techniques permettant de recréer le signal résiduel HF manquant sont présentées. Il est proposé d'utiliser la corrélation entre les BF et les HF en créant une copie du spectre BF par symétrie ou translation [8]. Cette technique requiert un suréchantillonnage du signal temporel sans filtrage passe-bas. L'insertion de zéros entre chaque échantillon temporel permet de créer une symétrie spectrale, tandis que le changement de signe d'un échantillon sur deux combinés à l'ajout de zéros permet de créer une translation spectrale [8].

2.6 Extension de bande

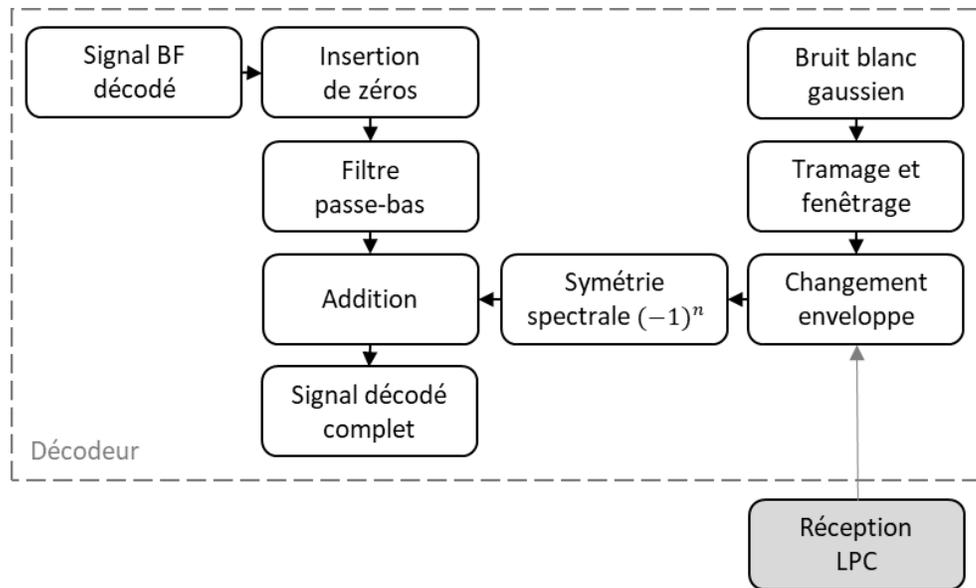


Figure 2-5 : Schéma du décodeur du HFR

Ces techniques sont simples, mais permettent d'obtenir de bien meilleurs résultats en comparaison avec l'ajout de bruit blanc. Cependant, elles ajoutent des artefacts en HF dues au fait que la dynamique des BF est répliquée dans les hautes fréquences de la partie HF, pour la symétrie par exemple. Il faut donc diminuer la dynamique du signal entre 0 et 2 kHz avant de faire la symétrie. De plus, il peut y avoir des pertes d'harmonicité puisque les deux moitiés du spectre ne sont pas synchronisées après le repliement spectral. Il y a une discontinuité dans la série d'harmoniques. La synchronisation de la fréquence de coupure du filtre passe-bas et la modulation avec un multiple entier de la fréquence fondamentale permettent de remédier à ce problème d'harmonicité. Cependant, ce traitement est très onéreux en complexité dans le domaine temporel, mais, facilement réalisable dans le domaine fréquentiel, et ce, par une simple translation du spectre pour assurer la bonne position des harmoniques.

2.6.2 Bandwidth Extension (AMR-WB+)

L'extension de bande du codec AMR-WB+ se nomme Bandwidth Extension (BWE). Elle est une implémentation de la méthode de symétrie décrite à la section 2.6.1 [8].

L'encodeur, présenté à la figure 2-6, est semblable à ce qui est fait dans le HFR par symétrie spectrale et avec l'ajout d'une enveloppe temporelle. Pour faire l'analyse, le signal est subdivisé en trames de 20, 40 et 80 millisecondes. À chaque trame, les coefficients des filtres LPC $A(z)$ sont calculés avec des fenêtres d'analyse permettant le recouvrement. Par la suite, l'énergie du signal est calculée à toutes les 5 millisecondes par l'équation (2-5) afin de créer l'enveloppe temporelle permettant de modéliser

l'amplitude du signal en fonction du temps. Pour chaque trame, les valeurs d'énergie et les coefficients LPC sont compressés et transmis [11].

$$\text{Énergie} = \sum_{n=1}^N x[n]^2 \quad (2-5)$$

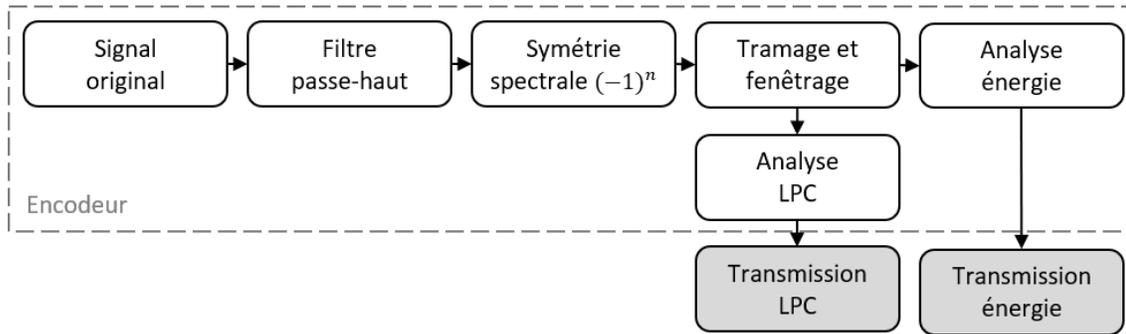


Figure 2-6 : Schéma de l'encodeur du BWE de l'AMR-WB+

Au récepteur, voir la figure 2-8, le décodeur régénère les signaux BF. L'extension de bande crée les HF, il y insère des zéros entre chaque échantillon du signal BF créant ainsi le suréchantillonnage et la symétrie spectrale. Par la suite, le signal est filtré passe-haut et transféré en HF en appliquant (2-4). Le changement d'enveloppe, voir la figure 2-7, est obtenu en filtrant le signal HF généré par le filtre inverse LPC, $A_{Déc}(z)$, puis le signal résiduel obtenu est filtré avec les coefficients LPC reçus $1 / A_{Enc}(z)$.

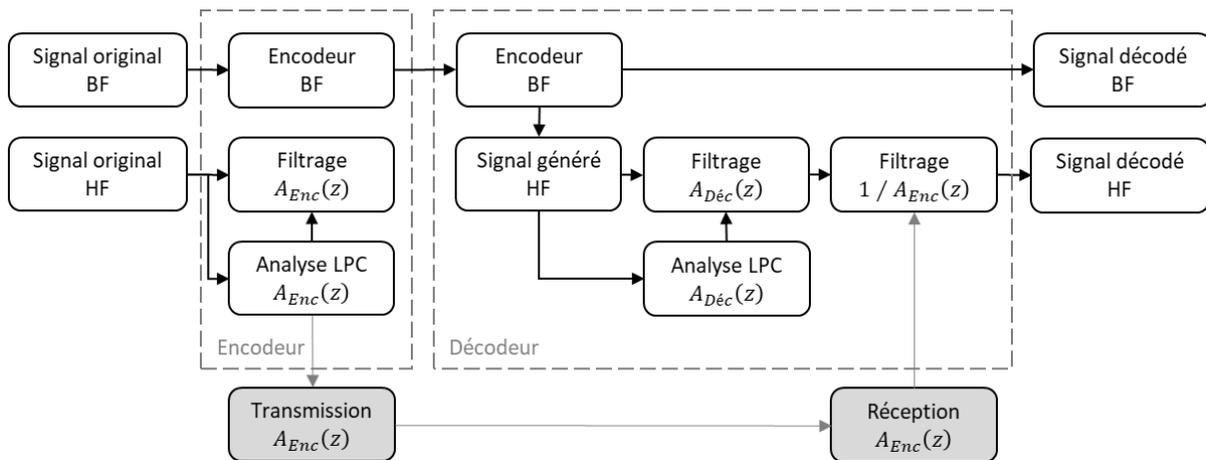


Figure 2-7 : Schéma du changement d'enveloppe spectrale HF par filtre LPC

Pour procéder à l'ajustement de l'enveloppe temporelle, l'énergie du signal généré HF est calculée et les valeurs reçues sont interpolées linéairement. L'ajustement de l'enveloppe temporelle se fait en deux étapes : le signal généré est d'abord divisé par sa valeur d'énergie interpolée, puis il est multiplié

2.6 Extension de bande

par sa valeur d'énergie interpolée. Finalement, le signal est retransposé en HF, en appliquant (2-4) et est additionné au signal BF généré par le décodeur [11].

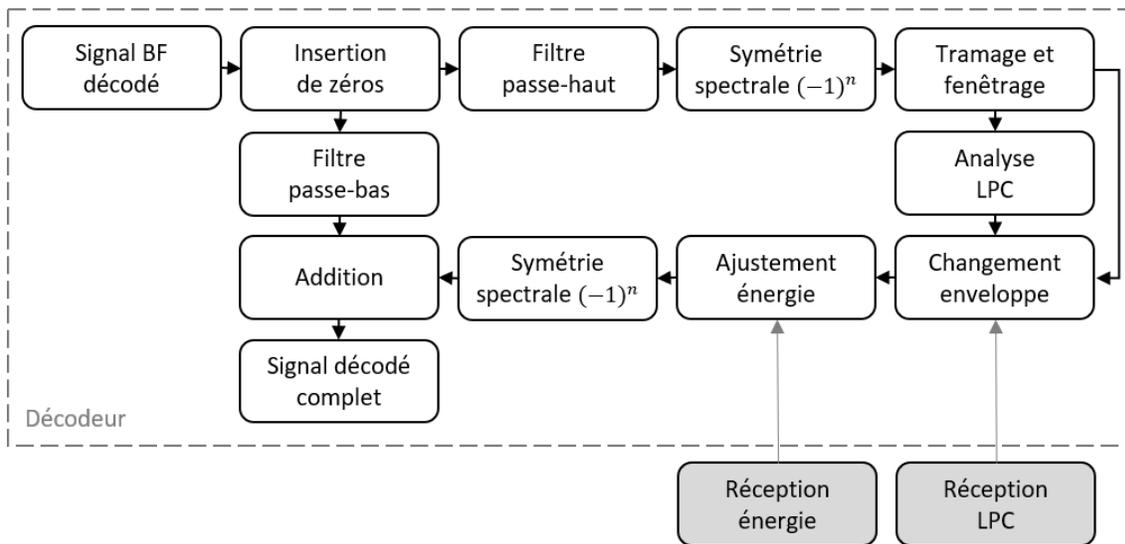


Figure 2-8 : Schéma du décodeur du BWE de l'AMR-WB+

2.6.3 Non-linéarité pour régénérer les HF

Dans la plupart des modèles d'extension de bande, uniquement les paramètres de l'enveloppe temporelle et de l'enveloppe spectrale sont transmis au décodeur. Ceci engendre des artefacts en HF dû au peu d'informations transmises. Il faut trouver des moyens autres que la simple symétrie ou la translation du spectre pour reproduire le signal fin. Il est proposé d'utiliser une non-linéarité sur le signal temporel afin de régénérer les HF [3]. Cette non-linéarité permet d'obtenir un signal cohérent en phase et préserve l'harmonicité du signal. La valeur absolue offre de bonnes performances en comparaison avec la symétrie et la translation puisqu'elle génère un signal HF plus près du signal original HF. Cette caractéristique rend le signal plus naturel [3]. Cette technique est basée autour du filtrage LPC permettant de transmettre et de régénérer l'enveloppe spectrale, voir la figure 2-9.

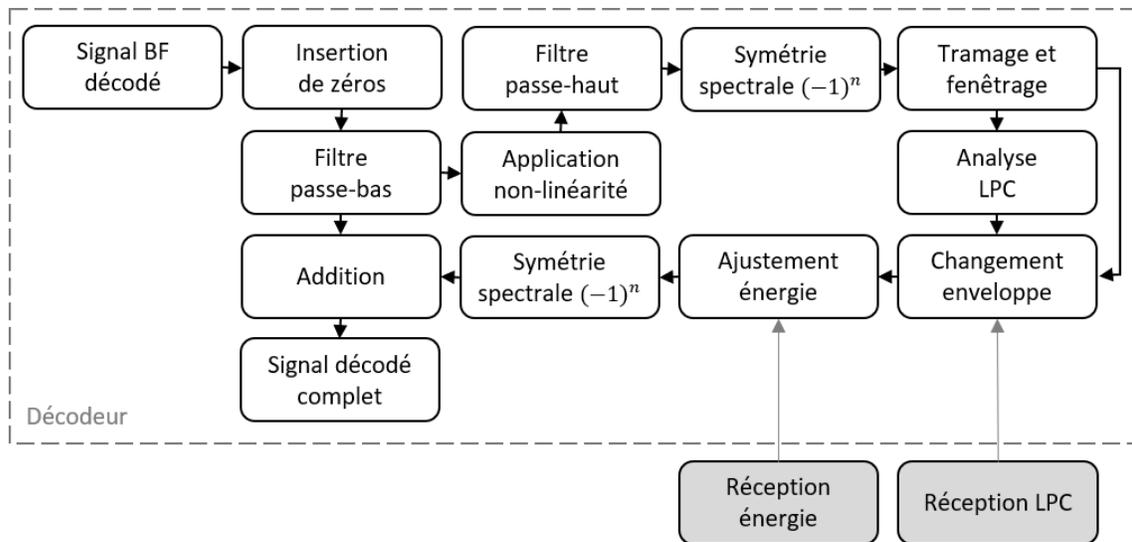


Figure 2-9 : Schéma du décodeur avec non-linéarité pour créer les HF

2.6.4 Estimation de l'enveloppe spectrale

Les extensions de bandes dédiées à certains systèmes téléphoniques n'ont pas l'opportunité de transmettre de l'information supplémentaire. Le spectre est régénéré à partir des BF par symétrie, il ne reste qu'à estimer les enveloppes temporelles et fréquentielles. Pour ce faire, plusieurs recherches ont permis d'estimer l'enveloppe à partir des signaux BF en utilisant des modèles statistiques, associatifs [10], par extrapolation des coefficients LPC [12], extrapolation de l'enveloppe par spline cubique [9] ou par des réseaux de neurones [13]. Les trois dernières méthodes donnent de meilleurs résultats.

2.6.5 Correction de l'enveloppe estimée

Dans le but de limiter le débit utilisé pour transmettre l'enveloppe spectrale, il est aussi possible d'utiliser un modèle prédictif, comme mentionné à la section 2.6.4, puis de transmettre l'erreur de prédiction. Le modèle de classification de l'enveloppe BF est suivi d'un modèle associatif qui propose plusieurs enveloppes HF, voir la figure 2-10. Cette technique alliant tant la quantification que l'estimation permet de diminuer de plus de 1 dB la distorsion spectrale en comparaison avec la quantification vectorielle conventionnelle à débit équivalent [14].

2.6 Extension de bande

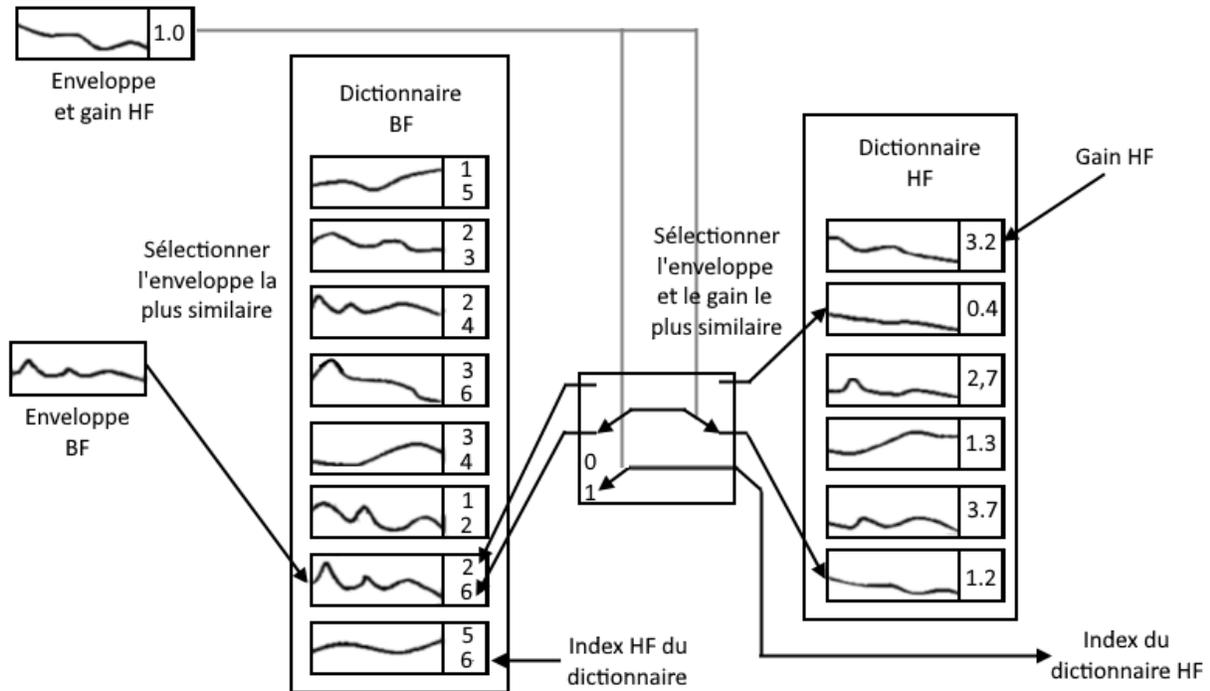


Figure 2-10 : Diagramme de la quantification vectorielle avec classification, 1 bit par trame

2.6.6 Modèle paramétrique sinusoïdale

Il est possible d'utiliser un modèle sinusoïdal jumelé à une enveloppe spectrale et des gains afin de recréer les HF à très faible débit. Le modèle sinusoïdal est basé entièrement sur la fréquence fondamentale (pitch), les informations voisées et l'enveloppe spectrale. Seulement l'enveloppe spectrale et des gains sont encodés et transmis. D'une trame à l'autre, les amplitudes sont interpolées linéairement, les fréquences sont interpolées avec un modèle *multiband excitation* (MBE) et les phases sont interpolées à l'aide de modèle spline cubique. De plus, les parties non voisées sont remplacées par des signaux aléatoires modulés par l'enveloppe spectrale [14].

En conclusion, selon cette méthode d'interpolation, trois critères permettent d'assurer la qualité de l'extension de bande : Les composantes voisées HF doivent être harmoniquement reliées. Le modèle de parole HF doit dépendre du niveau de parole en BF. Les fréquences et phases doivent être interpolées entre les trames consécutives [14].

2.6.7 Spectral Band Replication (SBR)

Le SBR est un traitement à la limite du domaine fréquentiel et temporel, puisqu'il est conçu avec un banc de filtre *Quadrature Mirror Filter-bank* (QMF) à 64 bandes de fréquence [15]. Cette extension de bande permet de quadrupler la bande passante du signal en générant le signal des bandes 17 à 64 du filtre QMF. L'encodeur de cette extension de bande, voir la figure 2-11, calcule l'énergie et le niveau

d'harmonicité pour chaque bande du filtre QMF. De plus, il évalue la nécessité de transmettre un niveau de bruit et/ou des harmoniques supplémentaires [15].

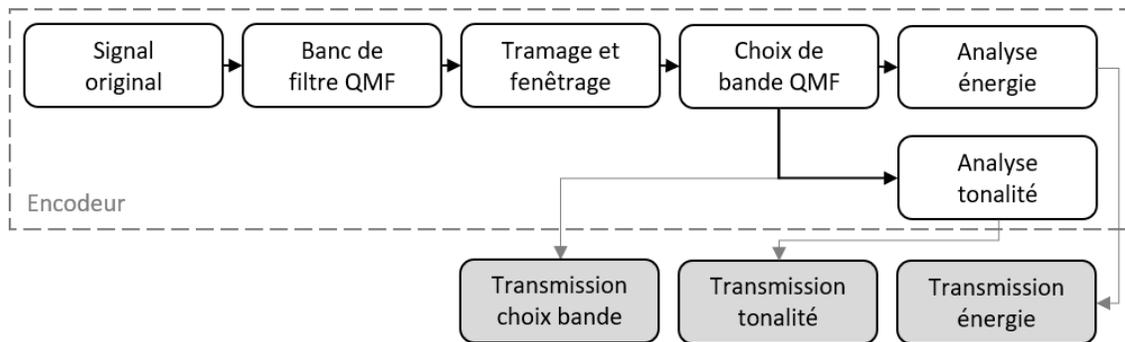


Figure 2-11 : Schéma de l'encodeur SBR

Le décodeur, voir la figure 2-12, réapplique le banc de filtre sur le signal décodé. Par la suite, les 48 bandes hautes manquantes sont générées en utilisant les 16 bandes basses décodées. Les bandes hautes générées sont ajustées en fonction des données reçues : énergie, niveau de bruit et tonalité [2]. Ce modèle permet de conserver les transitoires, la nature du signal, l'harmonicité et la périodicité.

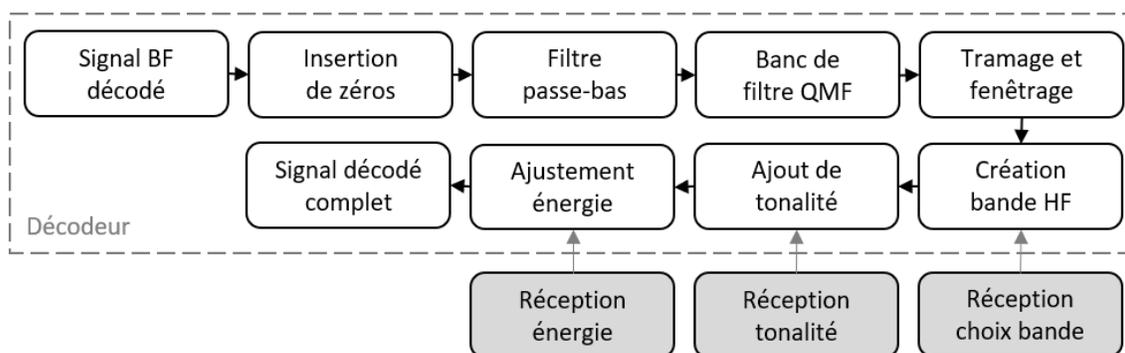


Figure 2-12 : Schéma du décodeur SBR

Puisque ce traitement permet une économie d'environ 30% du débit, la qualité est grandement augmentée comparativement à encoder toute la bande. Ceci est dû au fait que le débit sauvé est récupéré par le codeur de BF [14]. Le SBR est utilisé dans le MPEG-AAC, MPEG-Layer II, MPEG-Layer III, MPEG-4, HE-AAC+ et USAC. Il constitue l'état de l'art dans le domaine de l'extension de bande pour la musique.

2.6.8 Enhanced Voice Service (EVS)

Enhanced Voice Service (EVS) est un standard 3rd Generation Partnership Project (3GPP) est implanté dans la téléphonie cellulaire, les vidéoconférences et la diffusion audio de haute qualité (HD Voice et VoLTE). Ce codec permet de changer de débit et de largeur de bande sur la base d'une trame selon le débit disponible sur le réseau [16]. Il est aussi très robuste aux pertes de paquets de données. Il existe

2.6 Extension de bande

trois extensions de bande selon le mode d'opération choisi : deux modèles paramétriques avec débit et un modèle artificiel sans débit.

Extension de bande artificielle par réseau de neurones de l'EVS

Cette extension de bande permet de doubler la bande passante du signal en créant les signaux HF de 7 à 14 kHz ou 8 à 16 kHz pour les signaux de parole. Le décodeur, voir la figure 2-13, subdivise le signal reçu en trames fenêtrées. Pour chaque trame, le signal est transformé dans le domaine fréquentiel par la TFD. Seules les fréquences positives sont conservées, puis séparées en modules et phases afin de permettre différents traitements. La fréquence d'échantillonnage est doublée par la création des données HF. Les données spectrales sont estimées selon les deux paramètres suivants soit : les phases et les modules, avant d'être recombinaées pour créer le signal fréquentiel. L'estimation des phases est faite par une translation des phases d'un quart du spectre des BF en HF. Les modules sont estimés par un réseau de trois étages de 256 neurones [16]. Par la suite, les données de phases et modules HF sont recombinaées aux BF, puis retransformées dans le domaine temporel par la TFDI.

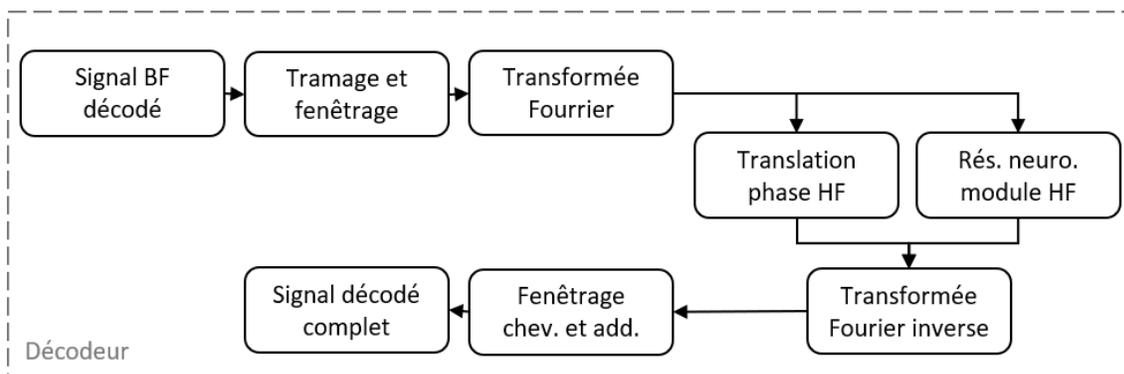


Figure 2-13 : Schéma du décodeur de l'extension de bande du EVS par réseau de neurones

Cette extension de bande est sans débit et permet de doubler la bande passante en effectuant les traitements dans le domaine fréquentiel.

Extension de bande temporelle par TCD

Une des extensions de bande temporelle utilisée dans le codec EVS est conçue autour d'un filtre LPC comme la plupart des extensions de bande pour les signaux de parole. Elle permet d'élargir la bande passante de 6,4 à 8 kHz. En opposition aux autres modèles avec filtre LPC, au décodeur, voir figure 2-14, le signal HF est généré par une translation d'une partie des coefficients de la transformée par cosinus discrète (TCD) faite sur des trames de 5 millisecondes. L'ajout de fréquences permettant d'élargir la bande passante se fait sur les coefficients fréquentiels. L'analyse se fait sur des fenêtres de 256

échantillons. Les coefficients spectraux 240 à 320 générés par l'extension de bande permettent d'effectuer le changement de fréquence d'échantillonnage. Les 81 nouveaux coefficients sont choisis parmi les coefficients en BF afin de permettre de garder les caractéristiques du signal HF. La transformée TCD inverse permet d'obtenir le signal temporel qui est filtré par le filtre LPC d'ordre 2 [17]. Finalement, le signal HF et BF est recombinaé.

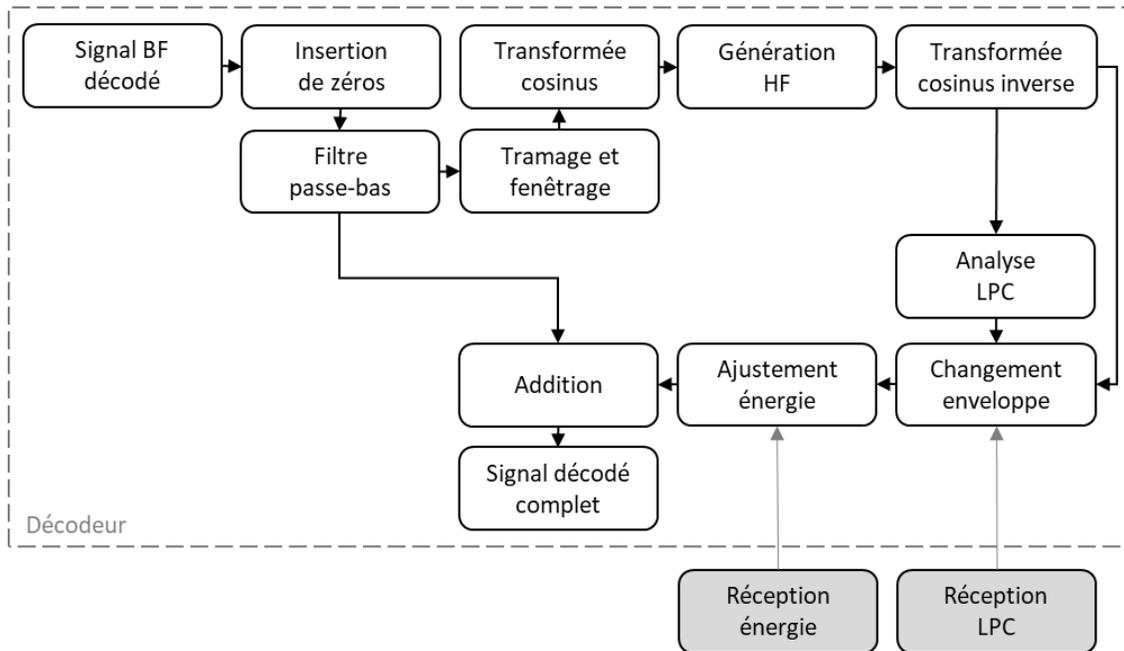


Figure 2-14 : Schéma du décodeur de l'extension de bande du EVS par DCT

Cette extension de bande permet d'étendre la bande de fréquence jusqu'à 8 kHz et innovant par rapport à une translation du spectre fréquentielle dans la TCD.

Extension de bande temporelle avec filtre QMF

Une autre extension de bande temporelle utilisant un banc de filtre QMF, les filtres LPC et une non-linéarité permet de régénérer les signaux HF en deux étapes, soit de 6,4 à 8 kHz et par la suite de 8 à 16 kHz. La première étape se fait à l'aide d'une non-linéarité (2-6) permettant de générer un signal très harmonique, voir la figure 2-15. Deux outils permettent d'adapter le niveau d'harmonicité du signal en fonction du niveau de parole du son. Le premier est un blanchissage adapté du signal par filtrage LPC et le deuxième est l'ajout d'un bruit blanc gaussien adapté au signal par filtrage LPC. Ces outils ne requièrent aucun débit, ils analysent le signal BF pour adapter le blanchissage et le bruit. Comme dans la majorité des modèles, une analyse LPC permet le changement d'enveloppe spectrale et l'enveloppe temporelle est ajustée en fonction de l'énergie du signal. Finalement, le filtre QMF permet de recombinaer les deux parties, BF et HF [18].

modèles sont artificiels, sans transmission d'informations, et d'autres avec un faible débit permettant de transmettre quelques paramètres. Cependant, la majorité des modèles utilisent les mêmes paramètres pour représenter les signaux HF, soit l'enveloppe temporelle et l'enveloppe spectrale. Le chapitre suivant présente plusieurs choix de conception pour le développement du projet de maîtrise.

CHAPITRE 3 MÉTHODOLOGIE

Ce chapitre présente la méthodologie utilisée pour le projet de maîtrise. Quelques choix ont été faits avant de débiter le projet de maîtrise. Parmi ceux-ci : le domaine de traitement, l'extension de bande de référence et les tests subjectifs de validation de la performance.

3.1 Choix préliminaires

L'extension de bande développée se nomme Upper Frequency Reperation (UFR). Trois versions sont développées avec chacune un débit différent afin de valider l'augmentation de la qualité subjective des signaux. Les caractéristiques des trois versions sont spécifiées au chapitre 6. Une des versions permet de comparer la qualité subjective au codec de référence à débit égal.

3.1.1 Choix du domaine de traitement

Le domaine fréquentiel est choisi pour la représentation et le traitement des signaux puisque plusieurs traitements sont simplifiés dans le domaine fréquentiel et impossibles dans le domaine temporel. Pour le traitement, les données sont subdivisées en trames, puis fenêtrées avant d'être transformées dans le domaine fréquentiel. Une fois le traitement effectué, les données sont retransformées dans le domaine temporel, puis les trames sont recombinaées par chevauchement et addition (*overlap and add*).

3.1.2 Codec de référence

L'extension de bande à concevoir utilise la forte corrélation entre les signaux BF et HF. Le codec de référence choisi est l'AMR-WB+, soit de 23 kbits/seconde, avec une bande passante de 16 kHz. Ce codec a été choisi en raison de ses performances subjectives sur la parole et la musique. Au moment de débiter le projet de maîtrise, il s'agissait du codec référence pour les signaux de paroles et de musique.

3.2 Développement

Un codec est un mélange de plusieurs techniques visant à minimiser l'information à transmettre. Toutes les techniques appliquées sur les signaux sont programmées dans un code informatique. Le codec est donc un algorithme mathématique mis en œuvre dans un programme informatique. La conception de codec est basée sur une technique de gestion de projet logiciel nommée *Minimum Viable Product* (MVP). La technique du MVP se base sur plusieurs itérations dans lesquelles des fonctionnalités s'ajoutent à chaque itération. Elle a pour fondement de ne jamais compromettre la fonctionnalité du

3.2 Développement

produit tout au long de son développement [20]. Les différents outils de codec sont développés les uns à la suite des autres afin de solutionner un problème. Les premières versions du codec répondent à la majorité des cas d'utilisation et les dernières versions permettent de résoudre les cas plus particuliers. Ce projet de maîtrise est donc divisé en quatre grandes étapes, lesquelles se subdivisent en sous-étapes afin de correspondre à des fonctionnalités du codec. Voici les grandes étapes du développement du codec.

Générer le traitement de base.

Cette étape permet de mieux comprendre les caractéristiques des signaux et en même temps de se familiariser avec les outils de recherche. Les sous-étapes de cette section sont les traitements de base des techniques d'extension de bande.

- Création des trames et du fenêtrage;
- Régénération des signaux HF;
- Calcul de l'enveloppe spectrale.

Développer des outils permettant d'améliorer les signaux.

Cette étape est de développer des outils permettant d'améliorer le traitement de base. Elle permet de tester de nouvelles techniques maximisant les similitudes entre le signal décodé et l'original. Les sous-étapes de cette section sont des possibilités d'outils à développer.

- Évaluer des méthodes de régénération des HF;
- Corriger la dynamique du signal blanchi;
- Corriger les artefacts.

Quantifier et compresser les données.

À cette étape, les données sont quantifiées et le codec est optimisé pour différents débits.

- Évaluer les corrélations entre les données et leurs sensibilités;
- Trouver une technique de quantification;
- Établir le débit.

Tester et analyser les résultats.

L'étape finale est de tester sur des auditeurs professionnels et des auditeurs du grand public.

- Développer un test MUSHRA et un test A/B;
- Analyser les résultats.

3.3 Test

Les performances des codecs audios sont évaluées par des tests subjectifs. Il existe plusieurs standards de tests d'écoute. Les deux utilisés dans cette maîtrise sont le test AB [21] et le test MUSHRA [22].

3.3.1 Test A/B

Le test AB compare deux codecs semblables : même débit, même bande passante et même qualité. Ce test est simple de compréhension pour l'auditeur et ne requiert pas d'auditeur expert.

L'analyse des résultats démontre quel codec est le plus performant et pour quels types d'échantillons. Chaque échantillon sonore reçoit un vote de préférence des auditeurs envers les codecs A ou B, « je préfère A ou je préfère B ». Le calcul de la moyenne globale en pourcentage des préférences des auditeurs indique le gain de performance du codec A par rapport au codec B.

Le test développé comporte 24 fichiers sonores monos de sept secondes chacun, dont huit fichiers d'entraînement et seize fichiers de comparaison. Le test a été distribué et publicisé via plusieurs groupes *Facebook*, dont quelques groupes d'étudiants de l'Université de Sherbrooke et des groupes spécialisés en audio.

Test A/B sans contrôle de l'environnement

En raison de la pandémie de la Covid-19, le test A/B a été déployé à distance. Ces conditions de test particulières ont nécessité quelques changements au protocole habituel. Les protocoles de test sont adaptés à une opération avec du matériel et un environnement de test contrôlé. Les perturbations de ces deux paramètres ont des conséquences sur les résultats. De plus, les principales caractéristiques des auditeurs ont aussi des effets sur les résultats, tels la langue maternelle, la culture le niveau d'audition et quelques autres [23]. Suivant ces conditions, un nouveau protocole de test a été élaboré en considérant les paramètres.

Le test a été effectué avec le matériel personnel des auditeurs et l'environnement de leur choix. Dans ce contexte, des outils permettent de pouvoir discriminer les résultats des participants. Pour ce faire, les quatre premiers fichiers sont en ordre croissant de difficulté et permettent de faire un entraînement, voir le tableau a-1 de l'annexe A pour les détails. Par la suite, des fichiers repères cachés sont utilisés tous les cinq fichiers sonores afin de confirmer que l'auditeur est toujours en mesure de

3.3 Test

faire le test. L'analyse de ces huit fichiers permet de valider la capacité du participant à reconnaître un fichier original d'un fichier encodé. Dans le cas où un auditeur n'est pas capable de faire la différence, à plusieurs reprises, entre un fichier original et un fichier encodé, ses résultats sont éliminés.

De plus, un questionnaire valide le niveau de difficulté du test. Les participants doivent indiquer s'il est facile, moyen, difficile ou impossible, où impossible est considéré comme un vote aléatoire, de trouver des différences entre les fichiers. Ces résultats permettent d'éliminer les résultats des auditeurs éprouvant trop de difficulté à faire le test.

La distribution de ce test est faite à l'aide des applications web *Google Forms* et *Youtube*. Les fichiers sonores sont transformés en vidéo à l'aide du logiciel *Movie Maker*. Dans l'application *Movie Maker*, le choix du codec de compression audiovisuel est le MP4 avec un niveau de compression audio choisi à 128 kbits/s sur une largeur de bande de 48 kHz. Cette qualité de compression est considérée comme transparente assurant aucune perturbation des signaux à comparer. Du côté de l'application *Youtube*, le niveau de transparence est atteint lorsque la bande passante internet est suffisante, aucune donnée précise n'est fournie à ce sujet autre que 128 kbits/s sur une largeur de bande de 48 kHz. Le protocole de test exige aux participants d'effectuer le test sur une connexion filaire ou sans fil de type *wifi*, excluant ainsi le réseau cellulaire.

Les vidéos contiennent les deux fichiers sonores à comparer comme suit : le premier fichier sonore et l'image de la lettre A pendant les 10 premières secondes, puis le deuxième fichier sonore et l'image de la lettre B pendant les 10 secondes suivantes. Les fichiers sonores, de sept secondes, sont séparés par une pause de trois secondes. La figure 3-1 montre l'interface du questionnaire pour le test A/B pour les votes et figure 3-2, les questions de niveau de difficulté.

Test audio A/B

* Obligatoire

Fichier #1

J Brahms



Entrer votre préférence ci-dessous. *

Je préfère A.

Je préfère B.

Précédent Suivant

Page 3 sur 31

Figure 3-1 : Interface de test A/B pour les votes

3.3 Test

Test audio A/B

* Obligatoire

Question mi-test

Comment avez vous trouvé cette partie du test? *

J'ai trouvé cette partie du test facile à faire, car je détectais des différences facilement.

J'ai trouvé cette partie du test moyennement facile à faire, car je détectais des différences après quelques écoutes.

J'ai trouvé cette partie du test difficile à faire, car je devais écouter à de nombreuses reprises avant de voter.

J'ai trouvé cette partie du test difficile à faire, car je détectais peu de différences.

J'ai trouvé cette partie du test très difficile à faire, car je ne détectais pas de différences (mes votes sont souvent aléatoires).

Précédent Suivant Page 18 sur 31

Figure 3-2 : Interface de test A/B pour les questions

Les fichiers proviennent de quatre catégories différentes, également réparties, soit : parole, musique, parole sur musique et chant sur musique, voir la séquence du test à l'Annexe A.

3.3.2 Test MUSHRA

Le test MUSHRA évalue la performance de codecs par rapport à l'original et à des repères cachés. Ce test est beaucoup plus complexe pour les auditeurs. Statistiquement, on peut démontrer qu'un test comportant une dizaine d'échantillons écoutés par environ dix auditeurs expérimentés est suffisant pour bien évaluer les performances [22]. Plus le nombre d'auditeurs est élevé plus les données convergent et les intervalles de confiance diminuent. L'ajout de repères cachés permet de rendre les écarts-types plus petits, elles sont généralement des copies de l'original filtré passe-bas [22]. Dans le contexte du projet de maîtrise, les deux repères cachés sont :

- L'original filtré passe-bas à 3,5 kHz.
- La combinaison de l'AMR-WB+ en BF et l'original en HF.

L'analyse des résultats démontre quel codec est le plus performant et pour quel type d'échantillon. Cependant, les données sont plus complètes que celles du test A/B, puisque les données en

pourcentage sont comparées au signal original et à des références permettant de situer la note accordée à chaque codec. La moyenne et l'écart type des données recueillies fixent les intervalles de confiance validant les résultats du test.

La base de données de fichiers sonores utilisée est constituée de 43 fichiers sonores monos de MPEG. Cette base est subdivisée en quatre tests de 10 et 11 séquences sonores contenant les quatre types de fichiers : Musique, parole sur musique, chant sur musique et parole. Les tests sont faits par douze auditeurs experts du groupe de recherche (GRPA) dans les conditions habituelles de tests d'écoute subjectifs et suivant le protocole ITU [22]. La figure 3-3 montre l'interface utilisateur du logiciel de test utilisé par le GRPA.

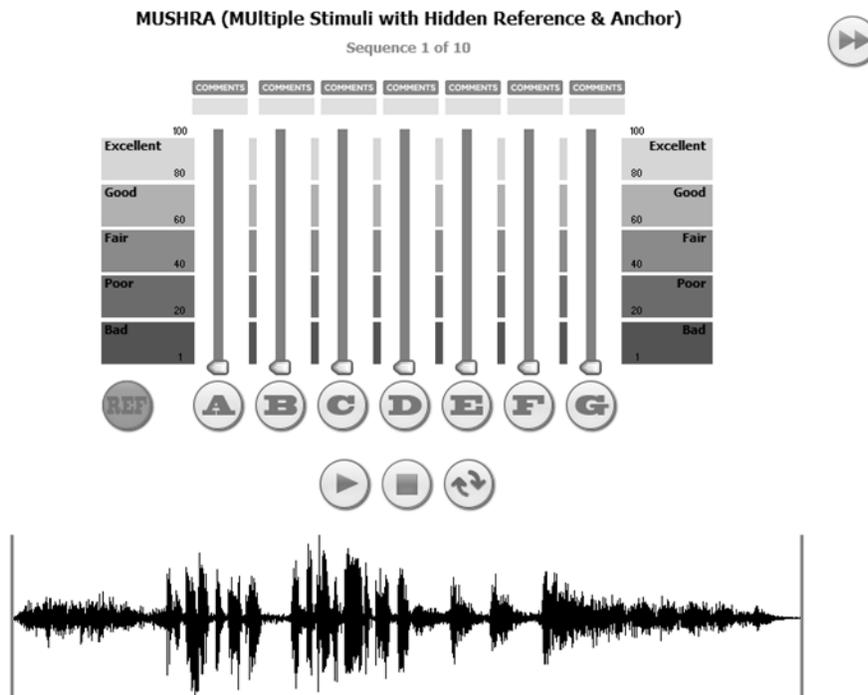


Figure 3-3 : Interface de test MUSHRA

3.3.3 Résumé des requis et caractéristiques des tests subjectifs

Afin de bien évaluer les performances de l'extension de bande développée, quelques requis et caractéristiques sont nécessaires pour la mise en place ainsi que la validité des tests subjectifs, voir le tableau 3-1.

3.4 Support technique disponible

Tableau 3-1 : Tableau des caractéristiques des tests d'écoute

Caractéristiques	Test AB	Test MUSHRA
Base de données	MPEG : 43 échantillons	MPEG : 43 échantillons
Nombre d'échantillons par test	24 échantillons	10 à 11 échantillons
Type de signaux	Parole Parole sur musique Chant sur musique Musique	Parole Parole sur musique Chant sur musique Musique
Débit	12-24 kbits / sec	12-24 kbits / sec
Nombre de canaux	Mono	Mono
Nombre d'auditeurs	Minimum 20	Minimum 8
Signaux de comparaison	AMR-WB+ Codec à tester	Original Référence 3.5kHz AMR-WB+ et original AMR-WB+ Codec à tester

3.4 Support technique disponible

Le groupe de recherche GRPA de l'Université de Sherbrooke possède une expertise en tests audio subjectifs avec plus d'une dizaine d'auditeurs expérimentés pour effectuer les tests. De plus, des logiciels de tests audio et d'analyse des résultats sont disponibles ainsi que des logiciels de développement, voir tableau 3-2.

Tableau 3-2 : Tableau des logiciels disponibles

Type de logiciel	Nom
Test audio	AB MUSHRA
Analyse de résultats	MUSHRA-Stats Excel
Développement	MatLab Spi

En conclusion, les conditions du projet de maîtrise sont définies avec un codec de référence rencontrant les spécifications et objectifs du projet de maîtrise. De plus, le développement du modèle et des outils est fait en suivant une méthodologie de gestion des versions MVP permettant de s'assurer que l'extension de bande développée est toujours fonctionnelle. Finalement, une procédure de

comparaison des performances de l'extension de bande développée avec celle de référence comporte deux tests, A/B et MUSHRA, effectués dans des conditions différentes et sur des auditeurs de compétences différentes. Le chapitre 4 présente le modèle paramétrique de l'extension de bande développée.

CHAPITRE 4 MODÈLE PARAMÉTRIQUE PROPOSÉ

Ce chapitre présente le modèle paramétrique de l'extension de bande développée. Pour chaque partie du modèle, il existe plusieurs techniques et choix de conception, ceux-ci sont présentés dans ce chapitre. Les détails de chaque paramètre de modélisation et les comparaisons des outils de calcul sont présentés dans le chapitre 5.

4.1 Structure du modèle

Les extensions de bandes sont conçues afin d'augmenter la qualité, l'authenticité ou la compréhension du signal. Pour ce faire, les extensions de bande sont implantées dans un système en parallèle, voir la figure 4-1, ou en série, voir la figure 4-2, avec l'encodeur de BF.

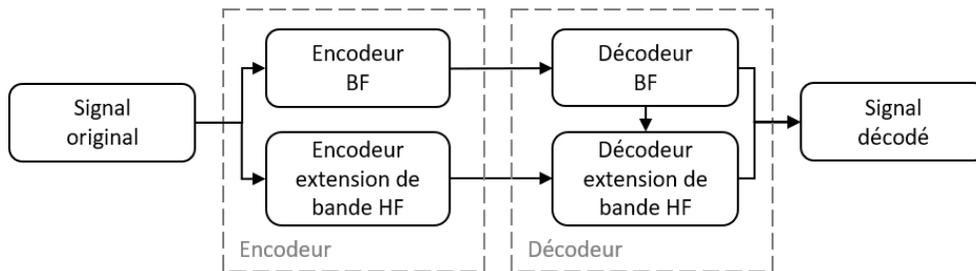


Figure 4-1 : Structure du codec avec extension en parallèle

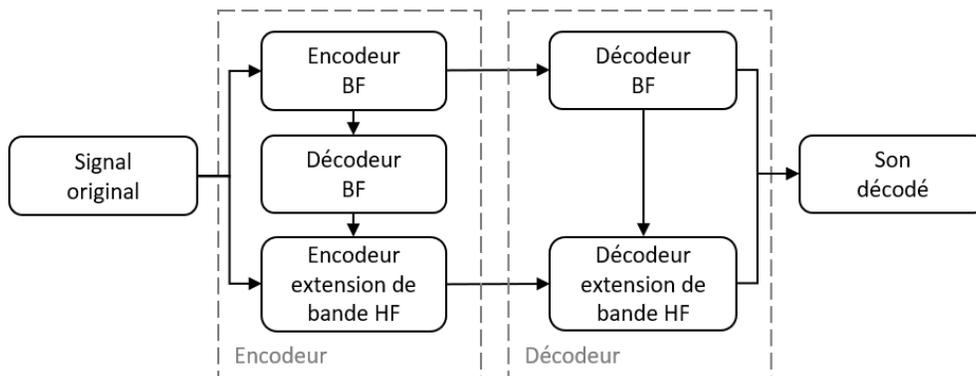


Figure 4-2 : Structure du codec avec extension en série

L'utilisation en parallèle est moins performante, puisque l'extension de bande n'a pas accès à la même information du côté encodeur et décodeur. Le modèle parallèle utilise uniquement le signal original, tandis que le modèle série utilise le signal original et le signal décodé BF. Le modèle choisi pour ce projet de maîtrise est celui en série. Ce choix permet d'éviter les erreurs de calcul ou de décision au décodeur. Cette configuration ajoute un peu de délai et de complexité à l'encodeur.

4.2 Trames, fenêtrage et recouvrement

Les sections suivantes détaillent chacun des modules de l'encodeur, voir la figure 4-3, et du décodeur, voir la figure 4-4, de l'UFR.

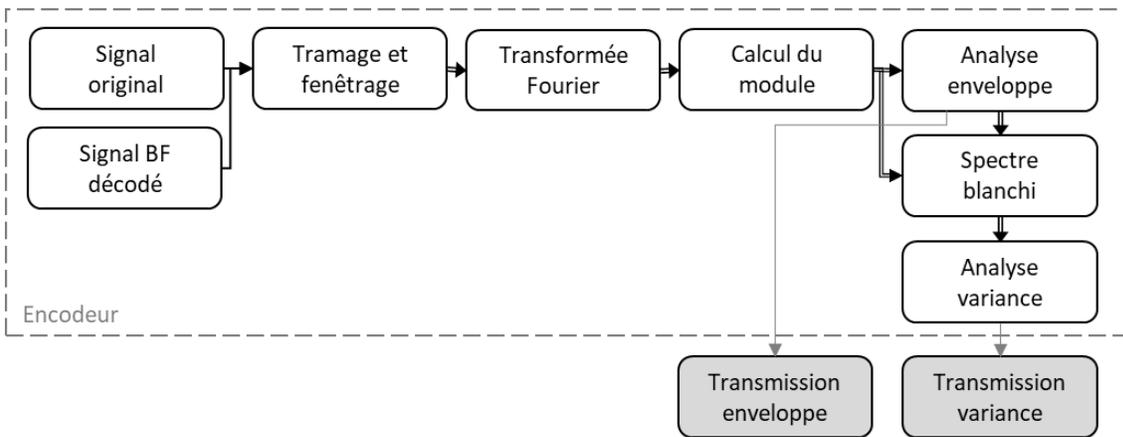


Figure 4-3 : Schéma de l'encodeur de l'UFR

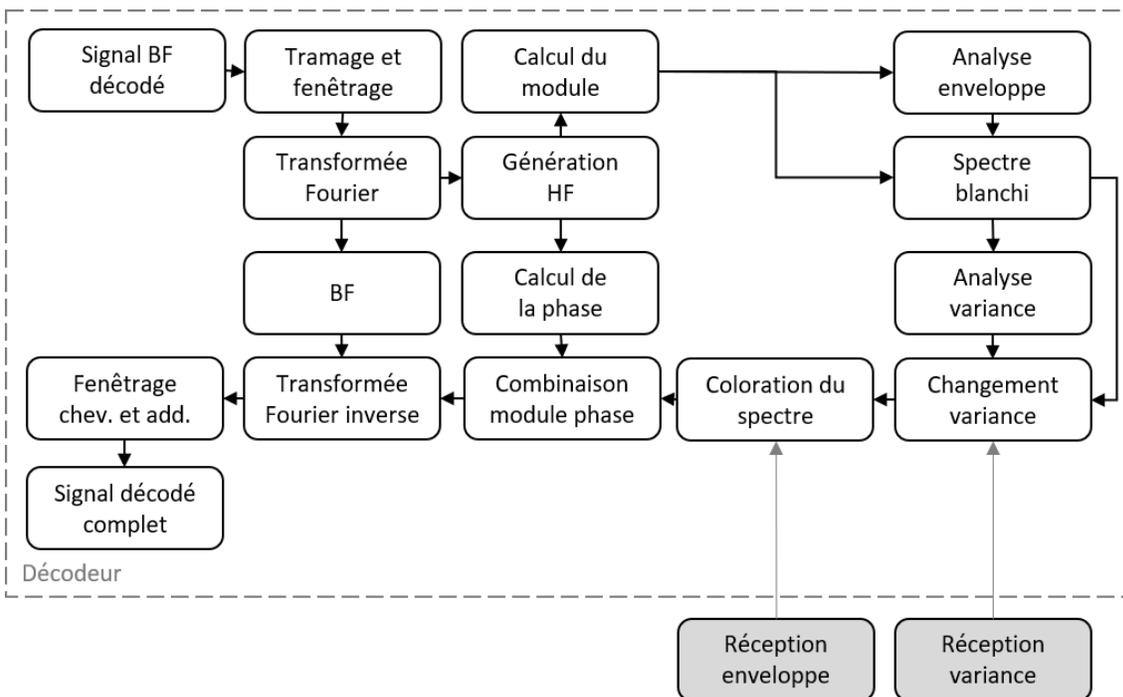


Figure 4-4 : Schéma du décodeur de l'UFR

4.2 Trames, fenêtrage et recouvrement

Pour analyser et reconstruire le signal, l'extension de bande doit diviser le signal en trames et les fenêtrer. Les données de chaque trame sont transposées dans le domaine fréquentiel par la TFD, traitées, puis retransposées dans le domaine temporel par la TFD inverse (TFDI).

4.2.1 Fenêtrage

Il existe plusieurs fenêtres connues dans le domaine du traitement de signal numérique (Hanning, Hamming, carré, triangulaire, etc.), voir la figure 4-5 et une grande variété de possibilités de créer des fenêtres. Chaque fenêtre possède ses propres qualités et limites.

Le fenêtrage d'un signal temporel revient à convoluer la fenêtre avec le signal dans le domaine fréquentiel. Plus une fenêtre possède un étalement temporel étroit et plus l'étalement spectral est grand. Inversement, plus une fenêtre est large, plus l'étalement spectral est faible.

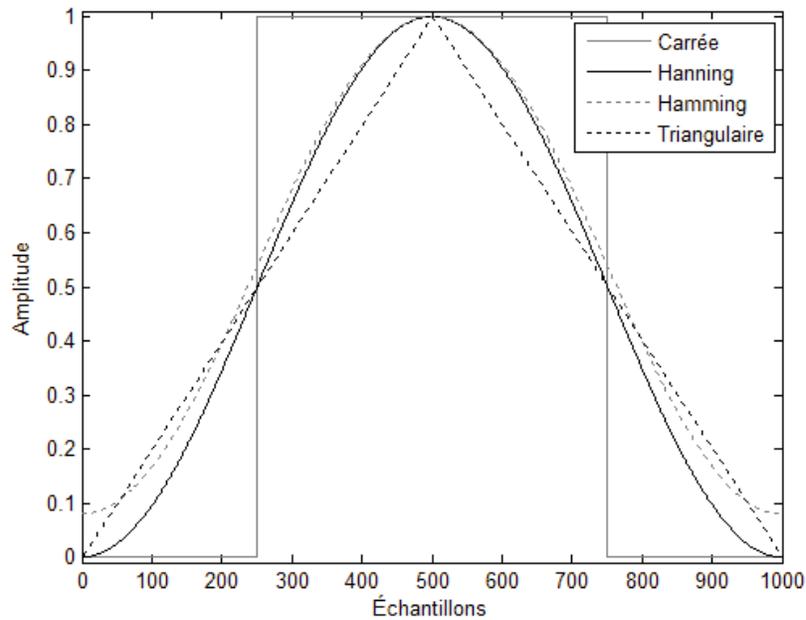


Figure 4-5 : Caractéristiques temporelles des fenêtres

Il est souhaitable de minimiser l'étalement spectral afin de ne pas introduire de bruit dans le signal à analyser. Le choix de la fenêtre est important puisqu'il affecte les analyses par la suite. Les choix de conception priorisent les fenêtres de Hanning et Hamming, puisqu'elles minimisent l'étalement spectral du signal avec des valeurs des modules du spectre en dB largement plus faibles que les fenêtres triangulaires et carrées, respectivement 50 et 100 dB plus bas, voir la figure 4-6. De plus, leur étalement temporel est relativement faible en raison de 81,8% de son amplitude se trouve entre les échantillons 256 et 767, correspondant à la fenêtre carrée sur 1024 échantillons.

4.2 Trames, fenêtrage et recouvrement

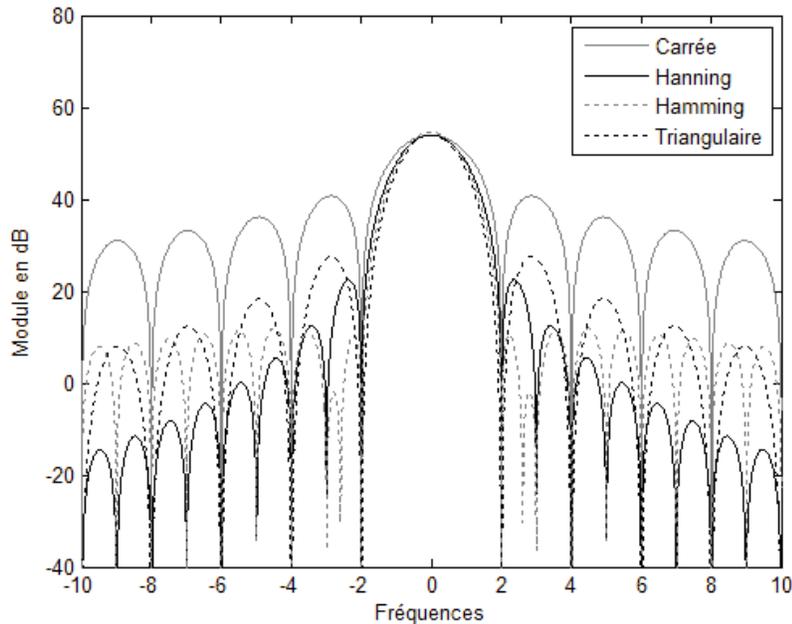


Figure 4-6 : Caractéristiques fréquentielles des fenêtres en dB

4.2.2 Recouvrement

Pour minimiser les effets indésirables des transitions dans l'analyse et la reconstruction du signal, la technique du recouvrement de 50% des fenêtres permet des périodes de transition lentes et continues. L'étalement spectral des fenêtres de Hamming et Hanning est pratiquement identique. Cependant, contrairement à la fenêtre de Hanning, la fenêtre de Hamming n'est pas complémentaire et ne peut être utilisée pour la reconstitution des signaux.

Dans ce projet de maîtrise, la technique de recouvrement des trames est appliquée conjointement avec le fenêtrage de Hanning qui permet une reconstitution parfaite lorsque les signaux de plusieurs trames sont additionnés avec le décalage de 50%, voir la figure 4-7. Les trames de 512 échantillons, soit 16 millisecondes, avec la racine carrée de la fenêtre de Hanning de 1024 échantillons, soit 32 millisecondes offrant un recouvrement de 50%.

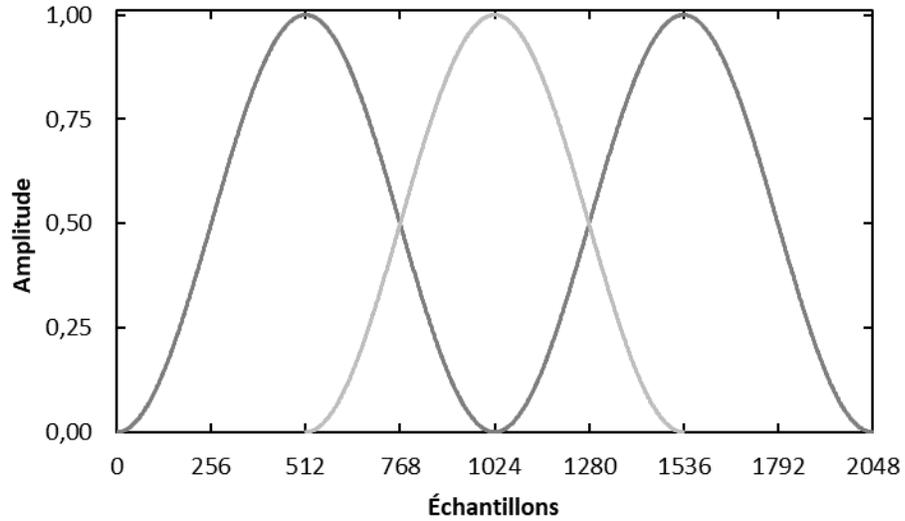


Figure 4-7 : Fenêtre de Hanning avec recouvrement de 50%

Un recouvrement plus faible permet de diminuer le débit, voir la figure 4-8. Les trames ayant un recouvrement de 25% représentent un bon compromis entre la qualité et de débit. Cette technique a été testée, mais non retenue puisqu'elle engendrait trop d'effets de trames, soit des artéfacts aux changements de trames.

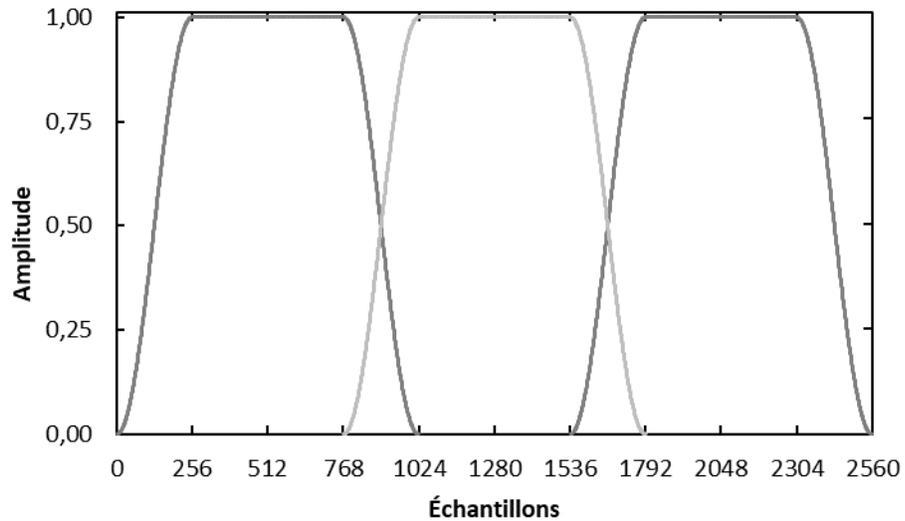


Figure 4-8 : Fenêtre de Hanning modifiée avec recouvrement de 25%

4.2 Trames, fenêtrage et recouvrement

4.2.3 Longueur de trame

Selon la nature des signaux à coder, la longueur de trame optimale peut varier. Pour les signaux plus harmoniques et dont les transitions sont plus lentes, les trames peuvent être plus longues. Pour les signaux dont les caractéristiques changent rapidement, la longueur des trames doit être plus courte. Ceci est aussi directement relié avec la stationnarité du signal. La longueur de la trame ne doit pas excéder la période de stationnarité du signal. Les signaux de parole sont considérés comme stationnaires sur une période d'environ 10 à 30 millisecondes et les signaux de musique sur une période de 20 à 80 millisecondes [1].

Pour analyser et modéliser les signaux de nature plus complexe, le codec fait des transitions entre les trames longues de 16 millisecondes et des trames courtes de 8 millisecondes. La condition pour ces changements de longueurs de trames est d'avoir les mêmes recouvrements entre elles.

La solution est de subdiviser une longue trame en petites trames ayant un recouvrement de 128 échantillons, voir la figure 4-9. Les quatre petites trames combinées sont pondérées par la fenêtrage de Hanning afin d'avoir le même recouvrement avec les longues trames, voir la figure 4-10. Cette technique permet de traiter le signal sans connaître la longueur de la prochaine trame, ce qui diminue le délai de l'encodeur.

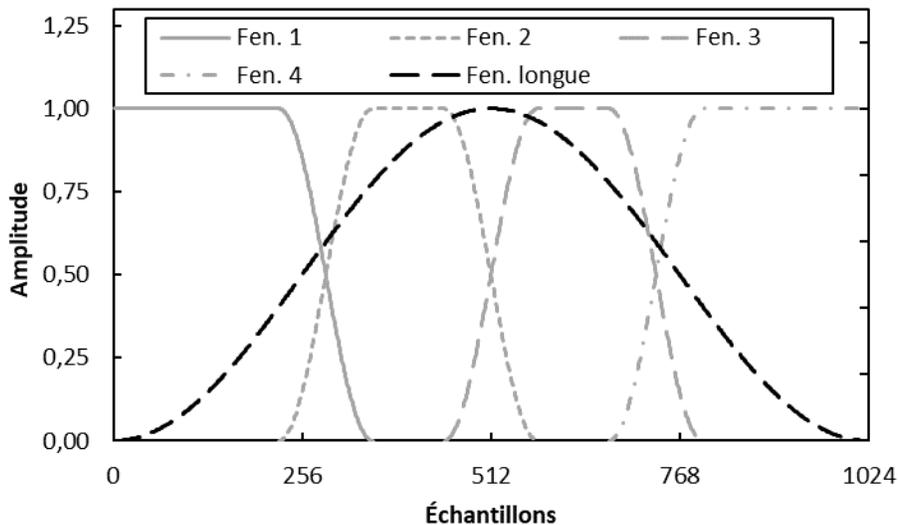


Figure 4-9 : Fenêtres de Hanning courtes modifiées avec recouvrement de 128 échantillons

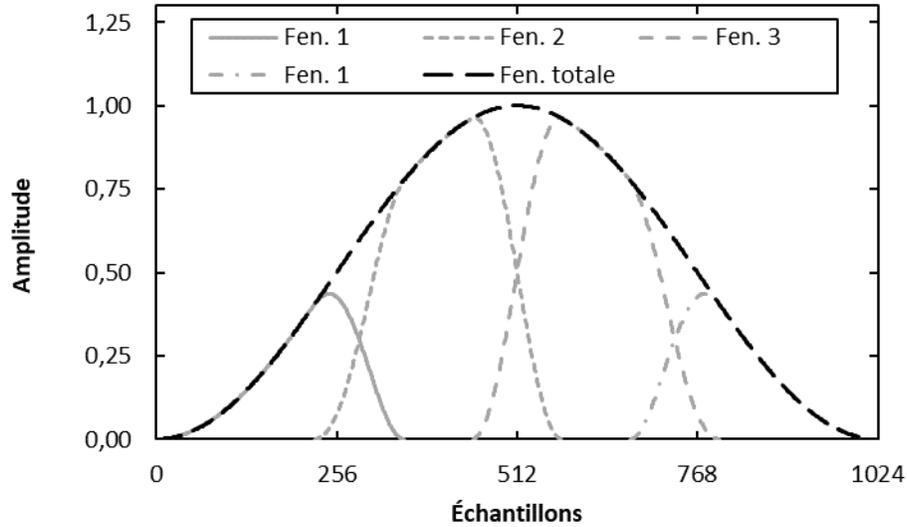


Figure 4-10 : Fenêtres de Hanning courtes modifiées et combinées à recouvrement de 50%

En comparaison, le codec AMR-WB+ a trois longueurs de fenêtres différentes lui permettant de s'ajuster aux différents types de signaux. Les fenêtres longues performant mieux sur des signaux stationnaires comme la musique très harmonique tandis que les fenêtres courtes sont dédiées aux signaux moins stationnaires comme la parole.

4.2.4 Choix de la longueur de trame

Le choix entre les trames courtes et les trames longues est fait par comparaison de l'énergie de chaque trame courte, voir la figure 4-11. Pour ce faire, l'énergie de chaque trame longue est comparée à l'énergie des quatre trames courtes. La trame longue est choisie si l'énergie de la trame longue est trop faible ou si l'énergie est également répartie entre les quatre trames courtes. Dans le cas contraire, l'énergie est potentiellement concentrée dans une trame courte, ce qui nécessite l'utilisation des trames courtes. Les seuils de 0,05 et 0,4 sont choisis selon l'analyse des signaux.

4.3 Générer les HF

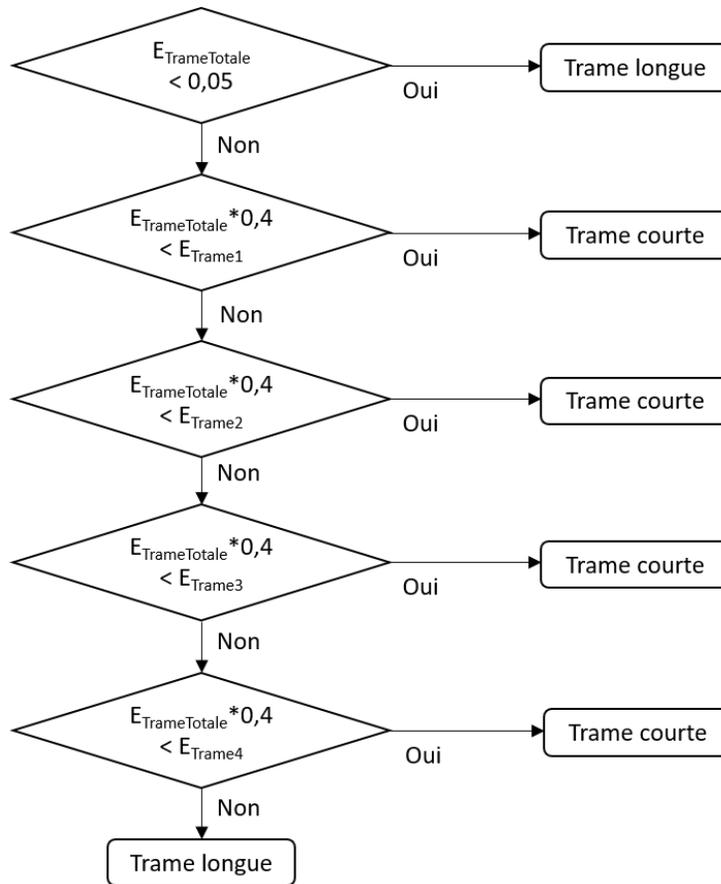


Figure 4-11 : Diagramme du choix de la longueur de trame

4.3 Générer les HF

Pour reproduire le signal HF manquant, le signal BF temporel doit passer par plusieurs modules de modification. Ce modèle est grandement inspiré du HFR, tel que présenté au chapitre 2.

Pour traiter le signal HF les données du spectre fréquentiel, des nombres complexes dans le cas de la TFD, sont représentées sous la forme de modules et phases. Les traitements sont effectués sur les modules du signal tandis que les phases ne sont pas modifiées. Une fois les traitements des modules faits, le signal est remis sous la représentation de nombres réels et imaginaires, puis transformé dans le domaine temporel par la TFDI.

Les traitements en HF se font à partir du signal original de BF et HF. Pour aider la compréhension, les figures suivantes montrent les modules des spectres des signaux où le signal original, considéré comme la référence, est toujours le même. Il s'agit d'un signal artificiel qui se situe à mi-chemin entre un signal de musique et un signal de parole, voir la figure 4-12.

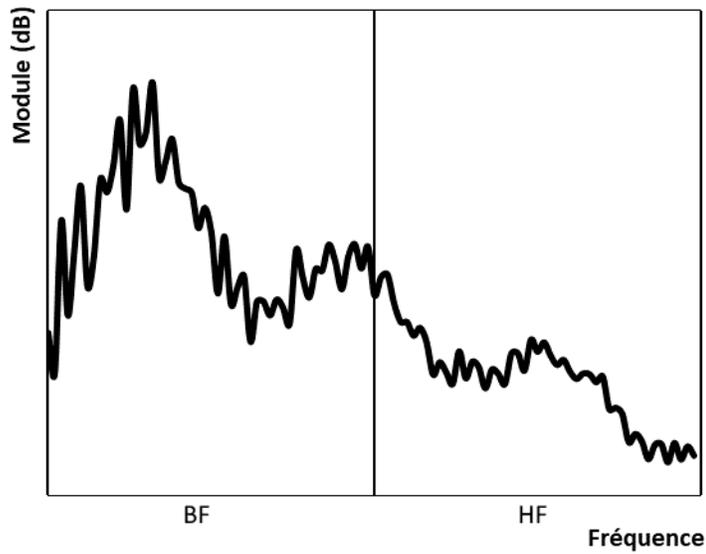


Figure 4-12 : Module du spectre original

Pour reconstruire le signal HF la première étape consiste à créer des HF, et ce à partir du signal BF provenant du décodeur BF. Cette étape génère des signaux de HF et double la bande passante du signal. Plusieurs techniques de base énumérées dans l'état de l'art sont possibles. La plus simple est la symétrie du spectre, voir la figure 4-13. Cette technique est implémentée par suréchantillonnage en y insérant des zéros entre chaque échantillon temporel du signal BF. Les autres techniques sont présentées et comparées au prochain chapitre.

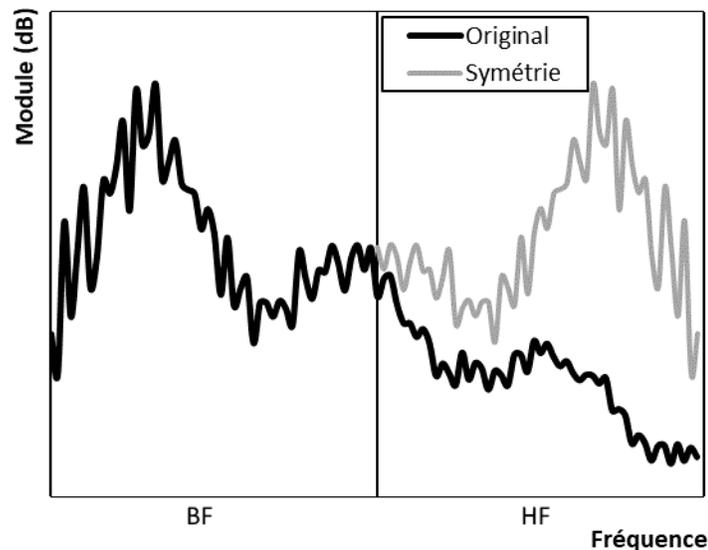


Figure 4-13 : Module du spectre original et généré HF par la symétrie du spectre

4.4 Changement de l'enveloppe spectrale

Cette technique de symétrie génère un module du spectre HF présentant des caractéristiques d'oscillation similaires à l'original, mais dont l'enveloppe spectrale est fautive.

4.4 Changement de l'enveloppe spectrale

Le changement de l'enveloppe spectrale permet de modifier les caractéristiques du signal HF créé par la symétrie du spectre qui ne correspond pas aux caractéristiques du signal original. Considérant que l'oreille est sensible aux amplitudes des fréquences et que le spectre possède déjà les bonnes caractéristiques d'harmonicité en raison de la symétrie; il est primordial de mettre l'emphase sur l'enveloppe spectrale. L'enveloppe spectrale est une méthode d'approximation de l'amplitude du signal qui caractérise l'allure générale du signal. L'enveloppe spectrale HF est calculée par le filtre à moyenne mobile sur dix coefficients spectraux et elle est considérée comme l'enveloppe de référence pour la comparaison avec d'autres techniques de calcul, voir la figure 4-14. Cette technique de calcul d'enveloppe spectrale s'applique aussi au signal HF généré par symétrie, voir figure 4-15.

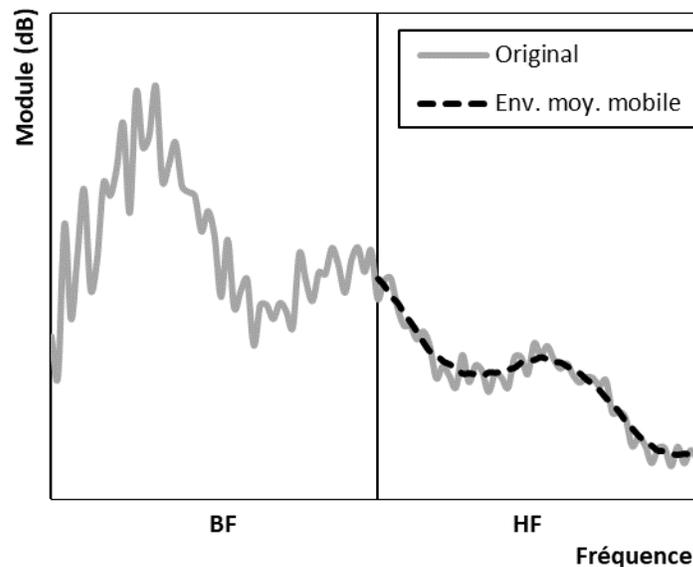


Figure 4-14 : Module du spectre original avec son enveloppe HF

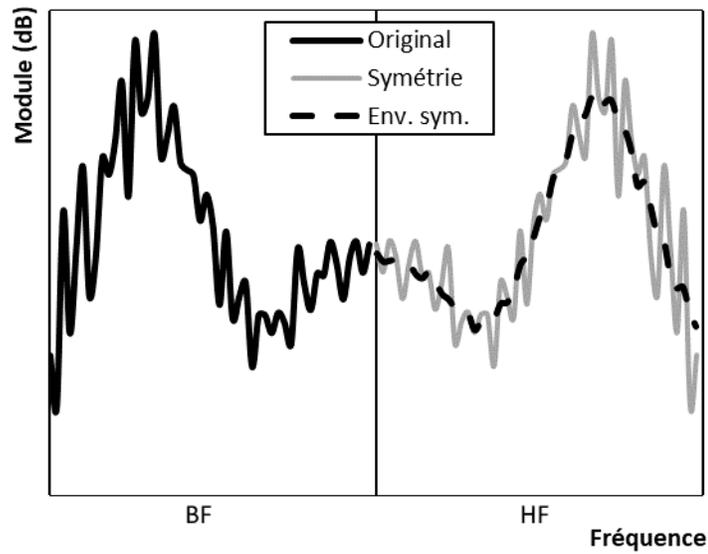


Figure 4-15 : Module du spectre original et généré par la symétrie avec son enveloppe HF

Le signal blanchi HF est le résultat de la division du module du spectre par son enveloppe, voir la figure 4-16. Ce calcul permet d'éliminer les changements d'amplitudes du signal et de garder uniquement les caractéristiques d'harmonicité et des principales fréquences.

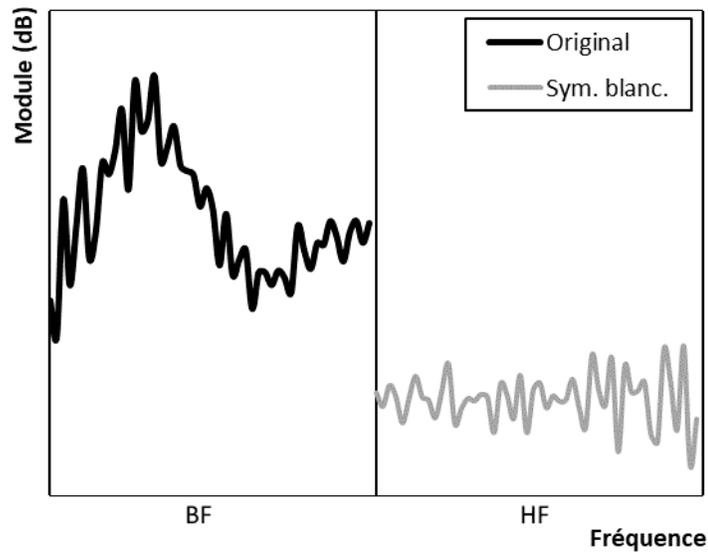


Figure 4-16 : Module du spectre original et du signal blanchi généré par la symétrie

La coloration est l'opération de la multiplication d'un signal blanchi par l'enveloppe spectrale d'un signal. Cette opération permet d'obtenir le signal final généré avec la bonne enveloppe spectrale, voir la figure 4-17. Ce modèle paramétrique permet d'obtenir une bonne approximation du signal original

4.5 Ajout et retrait de tonalité

pour les HF. L'enveloppe spectrale du signal HF est la bonne et les modules du spectre HF généré sont semblables, mais non identiques.

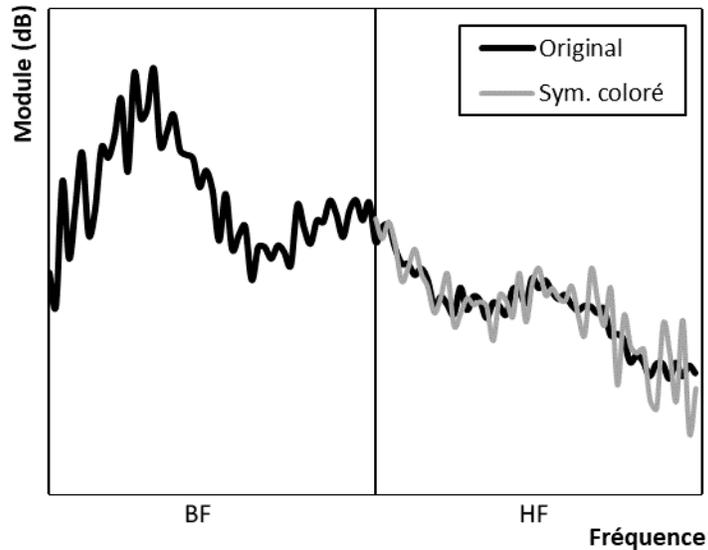


Figure 4-17 : Module du spectre original et coloré généré par la symétrie

Finalement, il faut recombinaison les nouveaux modules avec les phases inchangées afin d'obtenir le signal final sous la forme de nombres complexes. L'avantage de cette technique est de conserver la phase et les principales fréquences porteuses du signal tout en ajustant les modules des HF par le changement de l'enveloppe spectrale. Plusieurs caractéristiques de fréquences et de phases sont semblables en BF et en HF pour la majorité des signaux. Cependant, on y retrouve plusieurs artefacts provenant de la symétrie, notamment pour les plus hautes fréquences de la partie HF où les oscillations sont trop dynamiques. Cette caractéristique est bien visible sur la figure 4-17.

4.5 Ajout et retrait de tonalité

L'ajout et le retrait de pics spectraux permettent d'ajuster l'harmonicité du signal. Il est possible de simplement transmettre l'amplitude et la fréquence des pics. Une fois que le changement de l'enveloppe a été appliqué, le codec génère alors les pics désirés. Cette technique est très onéreuse en quantité d'information et requiert un fort débit. De ce fait, pour les codecs à bas débit, il faut trouver d'autres solutions tels : la translation, l'amplification, le filtrage ou l'ajustement de la variance du signal blanchi.

4.6 Ajustement de la variance

Tel que présenté dans la figure 4-18 une partie des artefacts provient des mauvaises amplitudes du signal blanchi par rapport au signal blanchi original pour les plus hautes fréquences de la partie HF. La modification de l'enveloppe spectrale réajuste l'amplitude des signaux HF générés, mais non les oscillations du signal blanchi. L'ajustement de la variance du signal blanchi généré permet de résoudre ce problème, voir la figure 4-19. L'ajustement de la variance présenté à la figure 5-19 est expliqué la section 5.3.2. Après coloration, le module du spectre obtenu pour les HF générées présente des caractéristiques semblables au module du spectre original avec le traitement proposé sur le signal blanchi, voir la figure 4-20.

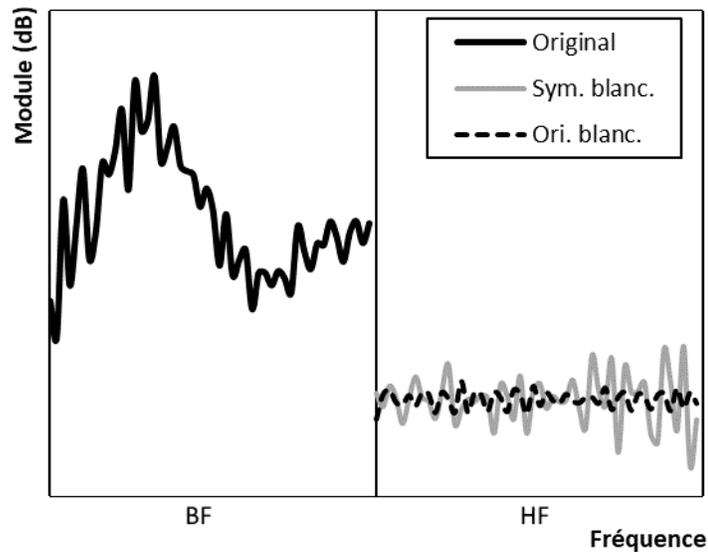


Figure 4-18 : Module du spectre blanchi original et généré par la symétrie

4.6 Ajustement de la variance

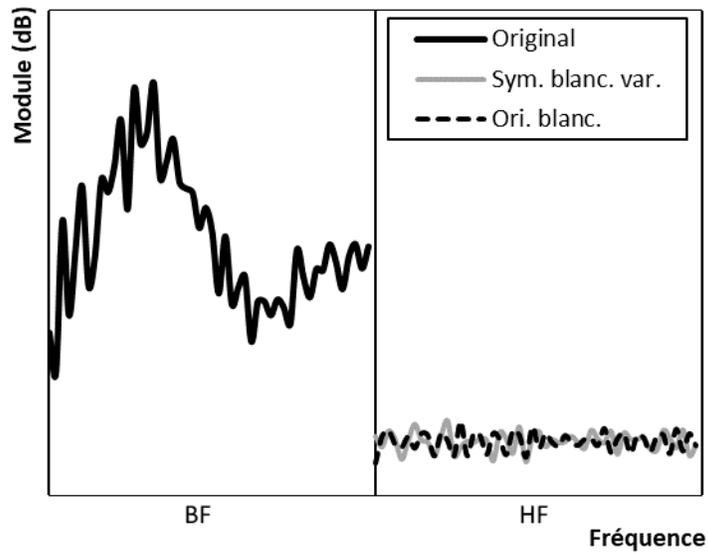


Figure 4-19 : Module du spectre blanchi original et généré par la symétrie avec variance

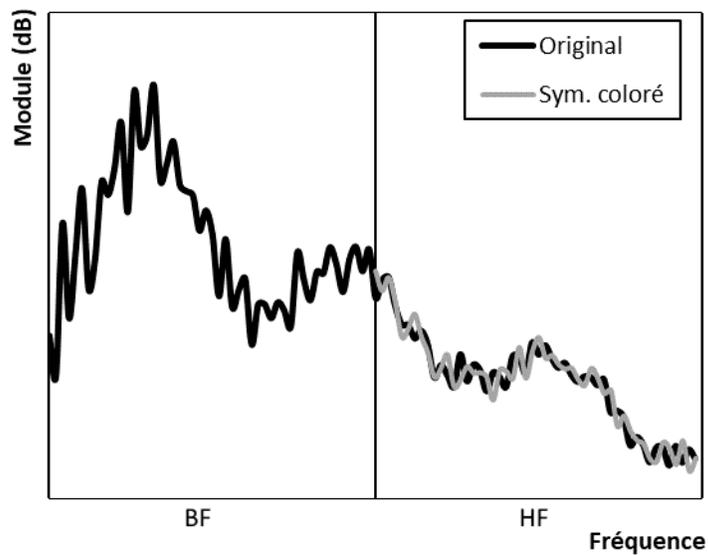


Figure 4-20 : Module du spectre original et généré par la symétrie avec variance

Finalement, l'ajustement du signal blanchi permet d'éliminer les principaux artefacts. Ce changement de variance est un ajout augmentant la qualité du signal comparativement à la méthode utilisée dans le BWE de l'AMR-WB+. Ce paramètre de modélisation supplémentaire doit être transmis. Le chapitre 6 explique en détail comment sont estimés les paramètres du modèle et quel débit est alloué à chacun d'eux.

4.7 Enveloppe temporelle

Une enveloppe temporelle permet d'ajuster les amplitudes du signal temporel, c'est-à-dire les événements ponctuels et courts, ainsi que les changements brusques d'énergie dans le signal. Une manière simple, utilisée dans l'AMR-WB+, est de transmettre l'énergie du signal sur de courtes périodes, 5 millisecondes [1], [11]. Cette méthode peut aussi être appliquée par sous-bandes, cependant ceci requiert un plus grand débit.

4.7.1 Modèle sans enveloppe temporelle

Le modèle paramétrique choisi n'utilise pas d'outil pour transmettre l'enveloppe temporelle puisque l'amplitude de l'enveloppe spectrale est ajustée en amplitude. L'entière reconstruction de l'enveloppe est faite par l'information contenue dans la TFD. Les signaux de HF générés étant grandement corrélés avec les signaux de BF, ils possèdent les mêmes caractéristiques temporelles que les signaux de BF. De plus, les deux choix de longueur de trames permettent d'ajuster l'amplitude de l'enveloppe spectrale selon la distribution de l'énergie du signal BF dans le temps.

4.7.2 Modèle avec enveloppe temporelle

Dans la partie précédente, l'enveloppe fréquentielle est extraite et transmise une fois par trame. En considérant que les coefficients fréquentiels possèdent le bon module et la bonne phase, le signal temporel calculé par la TFDI contient la bonne enveloppe temporelle. Cependant, puisque le signal HF est généré par une symétrie, les modules et les phases sont semblables, mais non identiques. Une autre option est de calculer l'enveloppe temporelle dans le domaine du temps. Cette dernière pourra être beaucoup plus précise en augmentant le nombre de valeurs transmises par trame.

Pour ce faire, il est possible de subdiviser et fenêtrer le signal en blocs et de transmettre l'énergie calculée. Le fenêtrage permet de centrer le calcul d'énergie. Du côté décodeur, le même calcul est appliqué, puis de ce résultat le ratio d'ajustement est calculé. Afin de lisser le signal de sortie, les ratios sont interpolés pour adoucir les transitions dans l'enveloppe temporelle.

L'ajout d'une enveloppe temporelle a été testé pour le développement du codec. Tout d'abord, il a été testé avec un niveau d'énergie pour toute la bande de HF et dans un deuxième temps avec deux et puis quatre sous-bandes sur des périodes de 5 millisecondes. Les résultats obtenus n'apportaient aucune différence notable au signal, mais simplement un grand ajout d'information à transmettre.

4.8 Résumé du modèle proposé

Finalement, les signaux HF sont corrélés avec les signaux BF et la technique de changement d'enveloppe possède une bonne caractérisation de l'enveloppe temporelle. L'utilisation d'une enveloppe temporelle n'est pas nécessaire dans le modèle paramétrique proposé.

4.8 Résumé du modèle proposé

En conclusion, le modèle présenté comporte quelques outils paramétriques pour modéliser les signaux HF, soit : deux longueurs de trame, le changement d'enveloppe spectrale et le changement de variance. Certains outils ont été testés et ne sont pas utilisés en raison de leur niveau élevé de débit nécessaire pour la transmission des paramètres, soit : l'enveloppe temporelle, la synchronisation du signal blanchi par translation et l'ajout/retrait de tonalités. Le débit étant limité, les apports de ces paramètres sont en partie compensés par d'autres paramètres.

CHAPITRE 5 TECHNIQUES D'AMÉLIORATION

Ce chapitre présente et compare les différents outils proposés afin d'augmenter la qualité perceptuelle du signal. Ces outils permettent le calcul et l'estimation des coefficients des paramètres de modélisation du modèle présenté au chapitre 4. Les paramètres de modélisation sont énumérés et justifiés en prévision de leur quantification telle que présentée au chapitre 6.

5.1 Calcul de l'enveloppe spectrale

Cette section compare plusieurs techniques de calcul d'enveloppe afin de prouver que la technique utilisant le cepstre est équivalente aux autres techniques communes. L'objectif principal du projet de maîtrise est de créer une extension de bande dont le traitement se fait dans le domaine fréquentiel. Pour cette comparaison, le nombre de coefficients permettant de modéliser l'enveloppe est fixé à cinq. L'évaluation de la performance d'une enveloppe spectrale se fait selon trois critères :

- Pour avoir une bonne modélisation du signal, l'enveloppe doit suivre les ondulations lentes du module du spectre et se trouver au milieu des oscillations rapides;
- Le signal blanchi produit doit être plat et quasi-symétrique autour de la valeur unitaire (0 dB);
- La technique de calcul de l'enveloppe spectrale doit avoir une représentation compacte sur quelques coefficients.

5.1.1 Calcul d'enveloppe par filtrage à moyenne mobile

Le filtrage à moyenne mobile applique un filtre passe-bas qui lisse le module du spectre et permet le calcul du signal blanchi, voir la figure 5-1. L'enveloppe spectrale obtenue respecte les deux premiers critères. Cependant, cette technique de calcul d'enveloppe ne possède pas de modèle pour transmettre l'enveloppe au décodeur.

5.1 Calcul de l'enveloppe spectrale

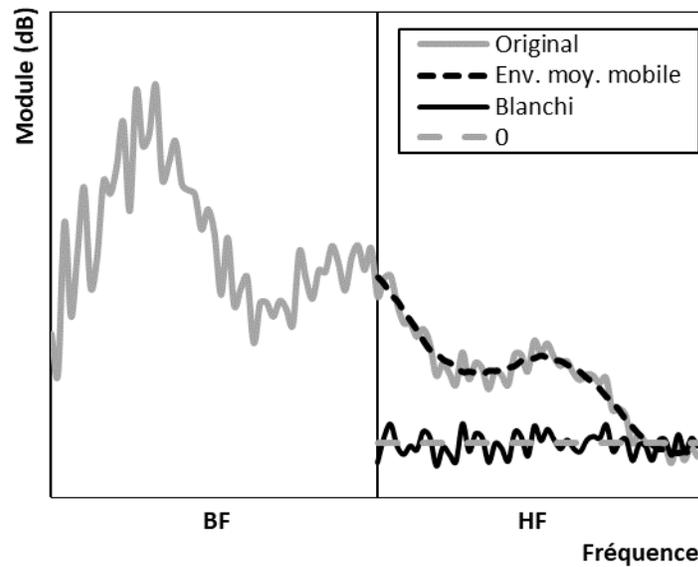


Figure 5-1 : Module du spectre original avec son enveloppe HF par filtrage à moyenne mobile et le signal blanchi HF

Cette technique de calcul d'enveloppe spectrale a été implémentée dans le but de comparer les performances des enveloppes et des signaux blanchis obtenus avec les autres techniques par sous-bandes, filtrage LPC et filtrage du cepstre. Cette technique est considérée comme très performante, mais ne permet pas la représentation de l'enveloppe spectrale sous forme compacte de quelques coefficients.

5.1.2 Calcul d'enveloppe par sous-bandes

L'enveloppe spectrale par sous-bande se calcule en utilisant la médiane ou la moyenne de chaque sous-bande. Le spectre fréquentiel est subdivisé en quelques sous-bandes de largeurs égales ou non. Pour comparaison des méthodes de calcul et de modélisation des enveloppes, la technique utilise cinq sous-bandes de largeurs identiques pour représenter l'enveloppe spectrale HF, voir la figure 5-2.

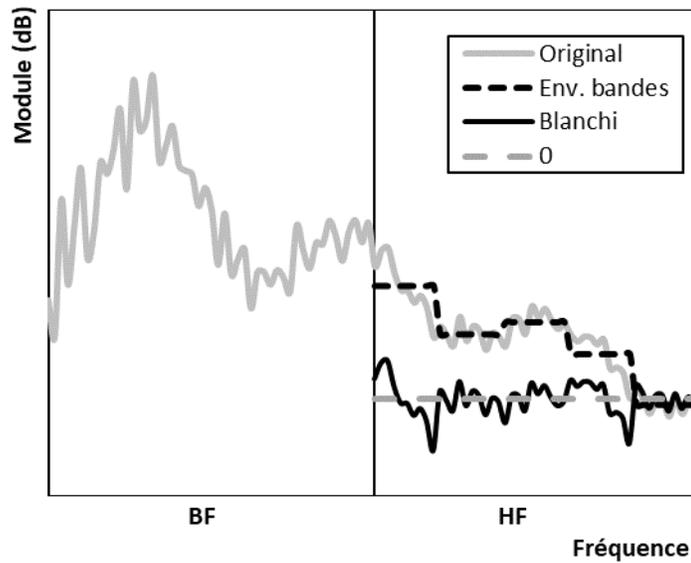


Figure 5-2 : Module du spectre du signal avec son enveloppe HF par 5 sous-bandes et le signal blanchi HF

La méthode par sous-bandes ne satisfait pas les deux critères de performance de l'enveloppe. Les paliers de l'enveloppe ne suivent pas le module du spectre et le signal blanchi n'est pas symétrique. L'interpolation des valeurs crée une enveloppe sans paliers permettant d'améliorer les performances de l'enveloppe spectrale. Les valeurs de moyennes obtenues sont interpolées linéairement, voir la figure 5-3, et par spline cubique, voir la figure 5-4.

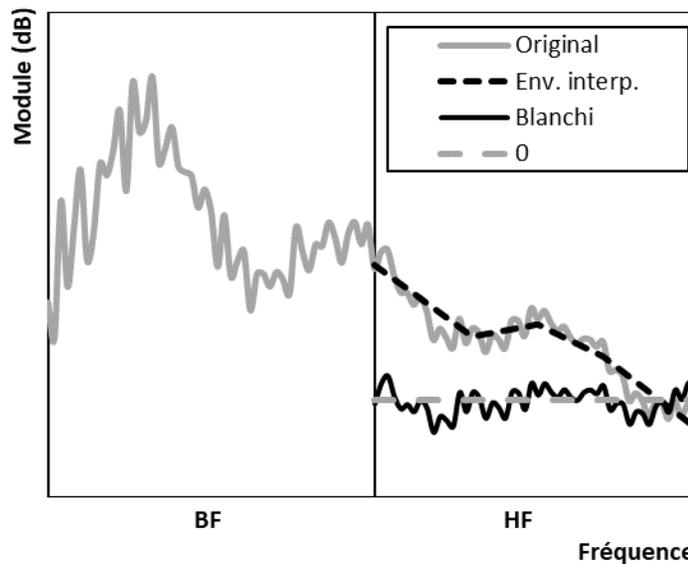


Figure 5-3 : Module du spectre avec son enveloppe HF par 5 sous-bandes interpolées linéairement et le signal blanchi HF

5.1 Calcul de l'enveloppe spectrale

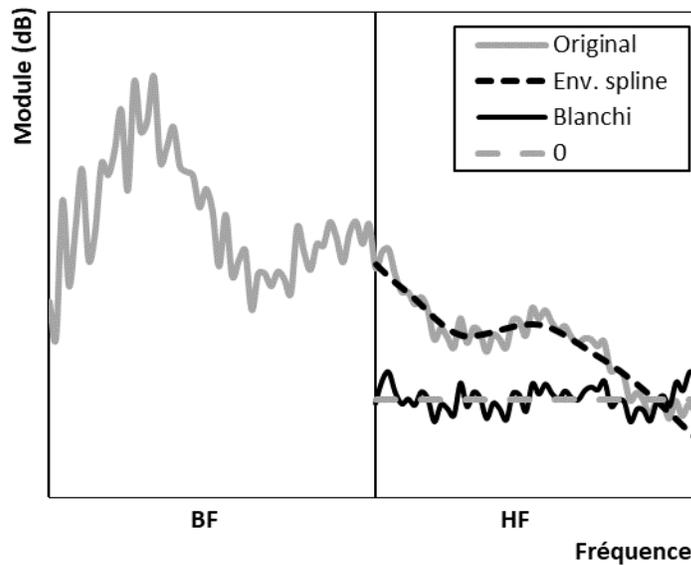


Figure 5-4 : Module du spectre avec son enveloppe HF par 5 sous-bandes interpolées par spline cubique et le signal blanchi HF

Les deux interpolations proposées améliorent grandement le premier critère de performance puisque les enveloppes suivent le signal pour la majeure partie du signal. Le second critère n'est pas entièrement respecté, mais les signaux blanchis sont plus symétriques que sans les méthodes d'interpolation. L'augmentation du nombre de sous-bandes permet de diminuer ce problème, mais augmente le nombre de données de la modélisation de l'enveloppe.

5.1.3 Calcul d'enveloppe par filtrage LPC

La technique de calcul d'enveloppe par LPC permet de calculer le filtre représentant l'enveloppe spectrale. Elle ne requiert aucune transformation dans le domaine fréquentiel. Il s'agit d'un filtre tout pôles permettant de calculer un signal $x[n]$ à partir des échantillons précédents $x[n - k]$ et le signal résiduel (résidu de prédiction) $r[k]$, voir les équations (5-1) et (5-3). Les coefficients $A(z)$ du filtre représentent l'enveloppe spectrale du signal $x[n]$, (5-2). Le filtre inverse permet de restituer l'enveloppe fréquentielle d'un signal résiduel, voir l'équation (5-4).

$$x[n] = \sum_{k=1}^N a_k \cdot x[n - k] + r[k] \quad (5-1)$$

$$A(z) = 1 - \sum_{k=1}^N a_k \cdot z^{-k} \quad (5-2)$$

$$R(z) = A(z) \cdot X(z) \quad (5-3)$$

$$X(z) = \frac{1}{A(z)} \cdot R(z) \quad (5-4)$$

Cette méthode possède plusieurs étapes pour obtenir l'enveloppe spectrale et le signal blanchi du module du spectre, voir la figure 5-5, [3], [2] :

- Calculer, à partir du signal temporel, les coefficients a_k d'un filtre à réponse impulsionnelle infinie (RII) représentant l'enveloppe spectrale par LPC;
- Calculer la TFD des coefficients a_k en ajoutant plusieurs zéros afin d'obtenir un signal de la même longueur que le signal.
- Inverser le module du spectre obtenue $A(z)$, pour obtenir l'enveloppe spectrale $1 / A(z)$;
- Calculer la TFD du signal temporel, puis multiplier par l'enveloppe spectrale pour obtenir le signal blanchi $R(z)$.

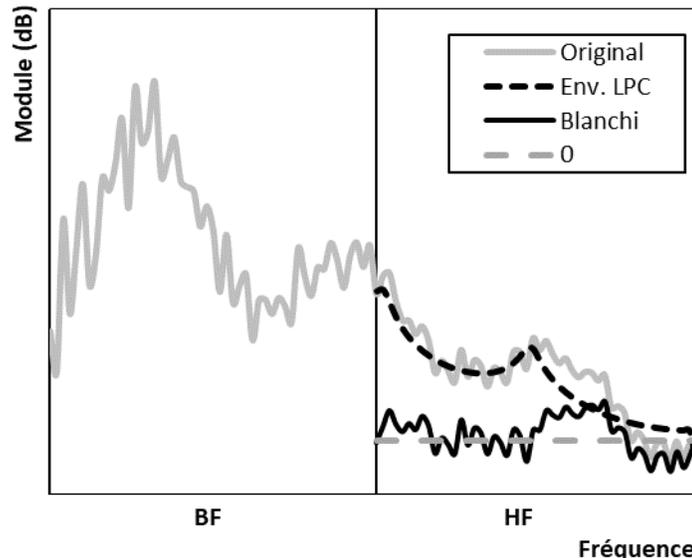


Figure 5-5 : Module du spectre avec son enveloppe HF par LPC et le signal blanchi HF

Cette technique est la plus utilisée afin de modéliser l'enveloppe spectrale pour le codage de la parole et les extensions de bande. Elle respecte généralement bien les trois critères, mais ici en réduisant le nombre de coefficients à cinq, le critère sur le signal blanchi n'est pas respecté. Le calcul des coefficients se fait sur les échantillons temporels ce qui n'est pas un modèle logarithmique et pose problème ici. Il y a donc une prédominance pour BF par rapport aux HF dans le calcul, ce qui tend à mieux présenter ces fréquences. En augmentant le nombre de coefficients, le signal blanchi obtenu respecte le deuxième critère.

Cette technique possède, tout de même, une bonne représentation de l'enveloppe spectrale et une représentation paramétrique compacte. Les paramètres à transmettre sont les coefficients a_k du filtre

5.1 Calcul de l'enveloppe spectrale

LPC. Il existe d'autres représentations de ces coefficients permettant d'assurer la stabilité du filtre face aux erreurs de transmission par exemple les *line spectral frequencies* (LSF).

5.1.4 Calcul d'enveloppe par cepstre

Le domaine cepstral permet l'analyse et le traitement des signaux au même type que les domaines temporel et fréquentiel. Le cepstre $c[\tau]$, équation (5-7), est la représentation fréquentielle de $M_{dB}[k]$, équation (5-5), le module en dB d'un spectre fréquentiel $X[k]$ d'un signal temporel $x[t]$. Il est obtenu par le calcul, (5-6), de la TFDI, noté \mathcal{F}^{-1} , sur le module du signal temporel en dB, $M_{dB}[k]$ de (5-5). L'équation (5-8) permet de calculer le module du spectre en dB $M_{dB}[k]$ par le cepstre inverse.

$$M_{dB}[k] = 20 \cdot \log | X[k] | \quad (5-5)$$

$$c(\tau) = \mathcal{F}^{-1}(M_{dB}[k]) = \mathcal{F}^{-1}(20 \cdot \log | \mathcal{F}(x[t]) |) \quad (5-6)$$

$$c[\tau] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} M_{dB}[k] \cdot e^{\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq \tau \leq N-1 \quad (5-7)$$

$$M_{dB}[k] = \sum_{\tau=0}^{N-1} c[\tau] \cdot e^{-\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq k \leq N-1 \quad (5-8)$$

Le calcul de l'enveloppe spectrale d'un signal par le cepstre est équivalent à la méthode par filtrage à moyenne mobile. Il s'agit d'une application de filtrage par transformée. Le filtrage est appliqué en annulant les coefficients cepstraux $c[\tau]$ des HF et par le calcul de la TFDI (5-8) des coefficients cepstraux $c[\tau]'$ modifiés. Cette technique est donc équivalente au filtre à moyenne mobile avec d'autres coefficients cepstraux.

Un des avantages de cette technique est l'élimination de la réponse transitoire du filtre. Puisque la TFD discrète s'applique sur des signaux de nature périodique, ceci permet de recréer les coefficients de manière périodique.

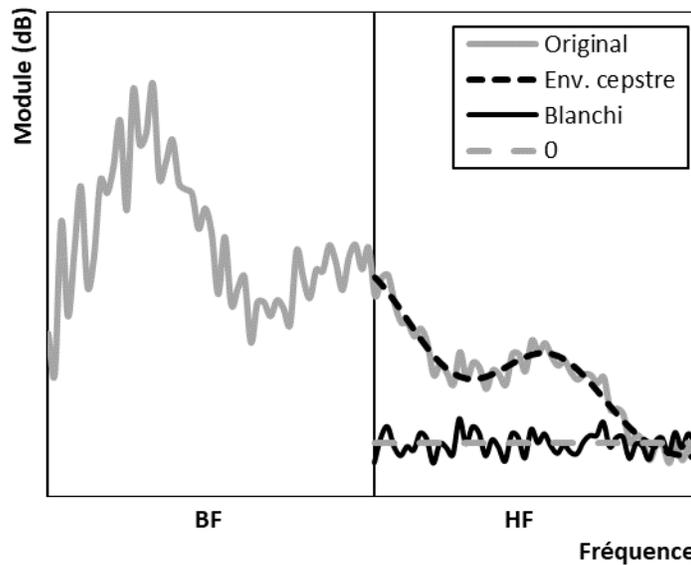


Figure 5-6 : Module du spectre avec son enveloppe HF par cepstre et le signal blanchi HF

La conception du filtre à moyenne mobile, filtre numérique FIR, permet de filtrer passe-bas le spectre et d'arriver aux résultats équivalents. Cependant, les coefficients cepstraux ont l'avantage de permettre de représenter l'enveloppe spectrale de manière très compacte. Les deux autres critères sont respectés afin de bien suivre le spectre et avoir un signal blanchi symétrique autour de zéro. Cette technique est supérieure à la méthode par filtre PLC et aux sous-bandes interpolées par spline cubique.

Symétrie du Cepstre

Puisque les coefficients des fréquences négatives du cepstre sont les conjugués complexes avec leur correspondant en fréquence positive, seuls les coefficients des fréquences positives seront conservés. La quantité d'information à transmettre est diminuée de moitié.

Lors du calcul du cepstre, l'enveloppe désirée est symétrique pour les fréquences positives et négatives. Il devient donc intéressant de seulement extraire les coefficients de la partie des fréquences positives et de faire une symétrie une fois l'enveloppe calculée. Ceci permet de diminuer la complexité des calculs. Cependant, cette opération a pour effet de créer une discontinuité dans le spectre ayant comme conséquence d'engendrer une distorsion au début et à la fin de l'enveloppe, voir la figure 5-7 pour le signal parole Es02 et figure 5-8 pour le signal de musique Brahms. Les deux enveloppes équivalentes en termes de quantité d'informations, soit : un coefficient réel ainsi que deux coefficients complexes sans symétrie ou cinq coefficients réels avec symétrie.

5.1 Calcul de l'enveloppe spectrale

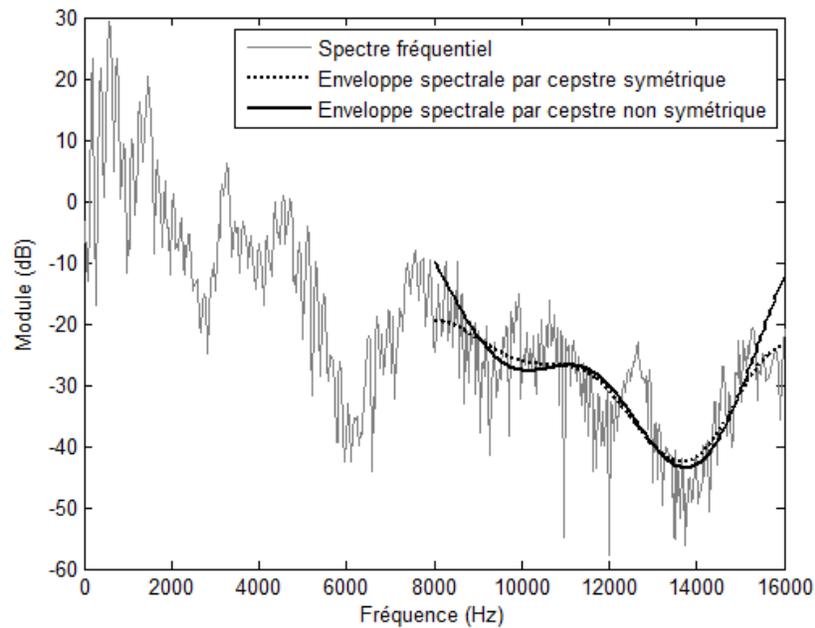


Figure 5-7 : Module du spectre du signal de parole Es02 avec ses enveloppes spectrales avec et sans symétrie

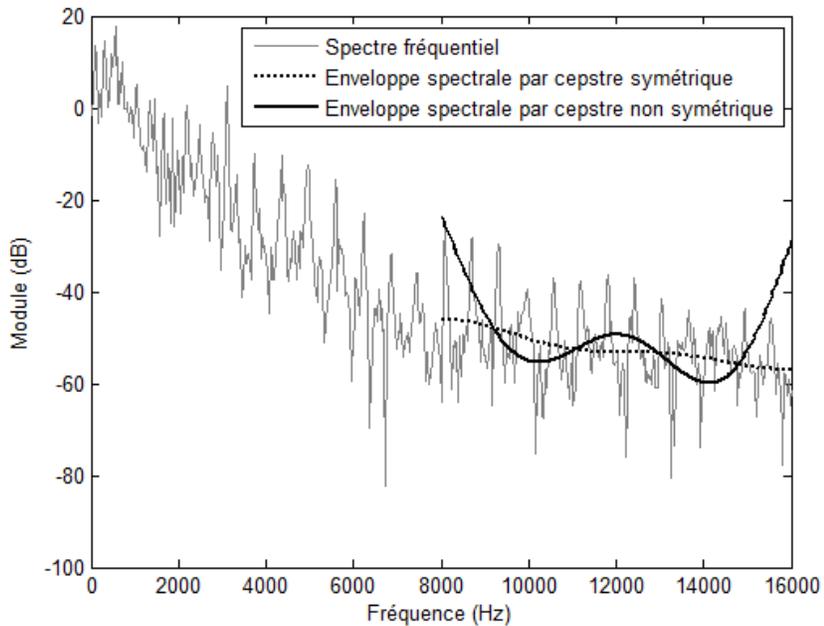


Figure 5-8 : Module du spectre du signal de musique Brahms avec ses enveloppes spectrales avec et sans symétrie

L'avantage d'avoir un spectre parfaitement symétrique est que les parties imaginaires des coefficients cepstraux sont toutes nulles. En conséquence, il faut alors doubler le nombre de coefficients non nuls pour obtenir la même qualité d'enveloppe, mais avec des coefficients tous réels. En conclusion, la

symétrie offre une meilleure représentation des modules du spectre et est équivalente en terme de quantité de coefficients à transmettre.

Nombre de coefficients

Le nombre de coefficients nécessaire à une bonne représentation de l'enveloppe dépend de la nature du signal. À partir de cinq coefficients réels, l'enveloppe suit bien le spectre, voir la figure 5-9 pour un signal parole et figure 5-10 pour un signal de musique. Le nombre de coefficients est aussi relatif au nombre de données. Le calcul de l'enveloppe est fait sur 512 coefficients représentant les HF.

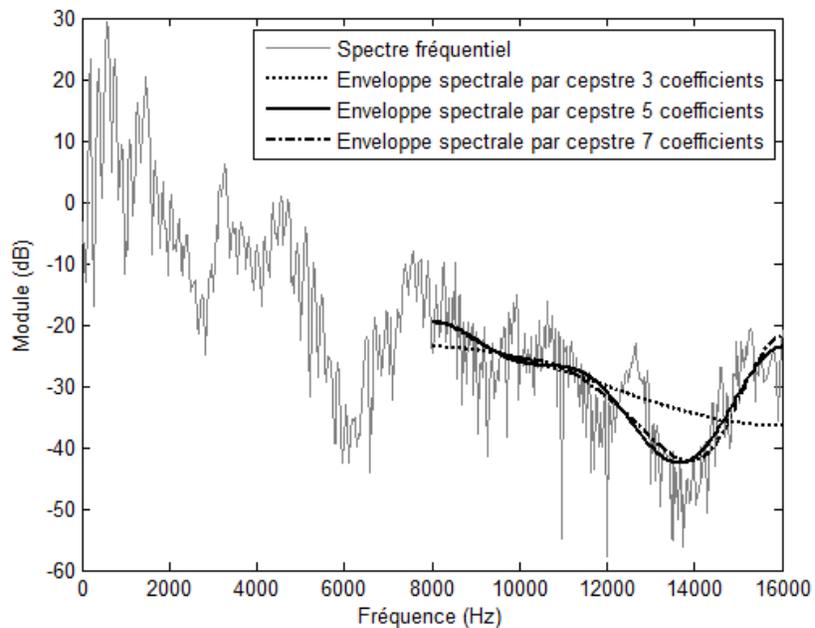


Figure 5-9 : Module du spectre avec les enveloppes spectrales par cepstre avec 3, 5 et 7 coefficients pour le signal de parole Es02

5.1 Calcul de l'enveloppe spectrale

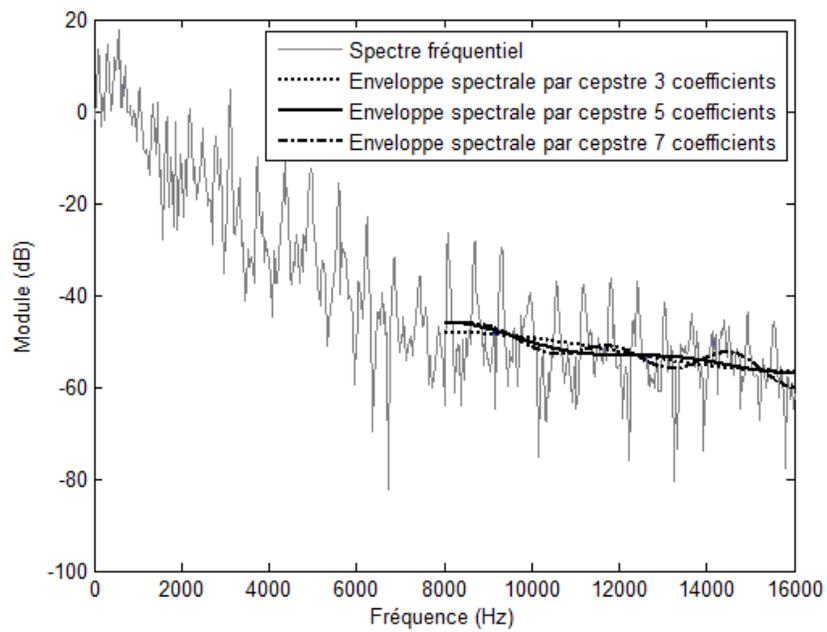


Figure 5-10 : Module du spectre avec les enveloppes spectrales par cepstre avec 3, 5 et 7 coefficients pour le signal de musique Brahms

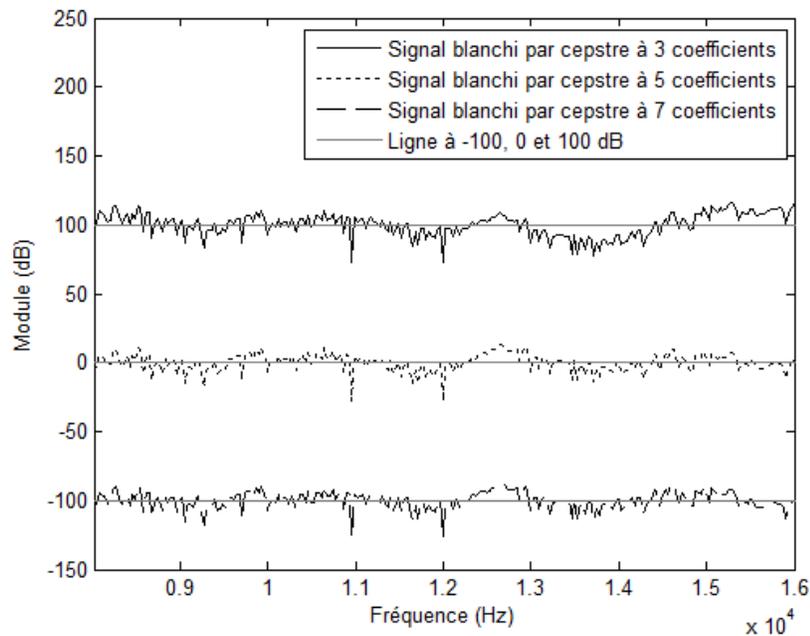


Figure 5-11 : Signal blanchis par cepstre avec 3, 5 et 7 coefficients pour le signal de parole Es02

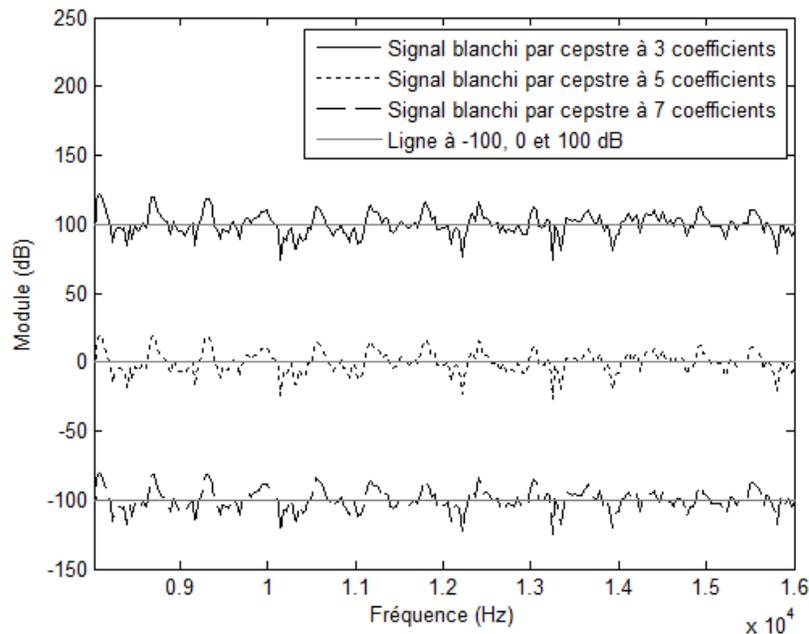


Figure 5-12 : Signal blanchis par cepstre avec 3, 5 et 7 coefficients pour le signal de musique Brahms

La conclusion de ces analyses est d'utiliser au minimum cinq coefficients cepstraux. Le nombre est fixé à cinq en raison du faible débit disponible. Cette méthode est équivalente à l'application d'un filtre passe-bas de type mur de brique avec une fréquence de coupure à $\pi/128$ radians/échantillon. Ce paramètre de conception peut être fixé à d'autres valeurs selon les caractéristiques des signaux à traiter.

Amplitude de l'enveloppe

La technique de calcul et de modélisation de l'enveloppe par le cepstre est aussi très intéressante puisqu'elle permet d'obtenir une enveloppe ajustée selon l'amplitude. Le premier coefficient cepstral représente l'amplitude de l'enveloppe. Il permet de représenter l'enveloppe temporelle du signal. Il est aussi intéressant de constater que le premier coefficient cepstral n'est pas toujours corrélé avec les autres, puisqu'il n'affecte pas la forme de l'enveloppe.

5.1.5 Sensibilité des coefficients au bruit de quantification

Les coefficients représentant l'enveloppe spectrale sont quantifiés afin de diminuer le débit utilisé. En raison du débit, le nombre de niveaux du quantificateur est faible, donc le bruit de quantification considéré non nul à cette étape. Comparons la sensibilité des coefficients modélisant l'enveloppe spectrale pour les trois méthodes présentées.

5.1 Calcul de l'enveloppe spectrale

Sous-bande

Les coefficients de chaque sous-bande sont utilisés directement pour former des paliers, puis sont interpolés pour former l'enveloppe spectrale. Le niveau de sensibilité peut être considéré à un, soit proportionnel. Selon le fait que statistiquement la possibilité d'être plus grand ou plus petit est équiprobable et que les coefficients doivent être interpolés pour recréer l'enveloppe.

LPC

Les coefficients LPC représentent le polynôme d'un filtre tout pôles dont le bruit sur les coefficients a pour effet de déplacer les pôles et de changer rapidement l'enveloppe spectrale résultante. Il est aussi possible et non souhaitable d'obtenir un filtre instable. Il est donc nécessaire de trouver une autre représentation, tels les LSF pour transmettre les coefficients [2], [24].

Cepstre

L'analyse de la sensibilité des coefficients cepstraux a été faite par un test sur le spectre original. Le test a généré 10000 enveloppes à partir des coefficients cepstraux originaux bruités avec un bruit blanc gaussien. La zone quadrillée montre les enveloppes erronées résultantes des coefficients bruités, voir la figure 5-13.

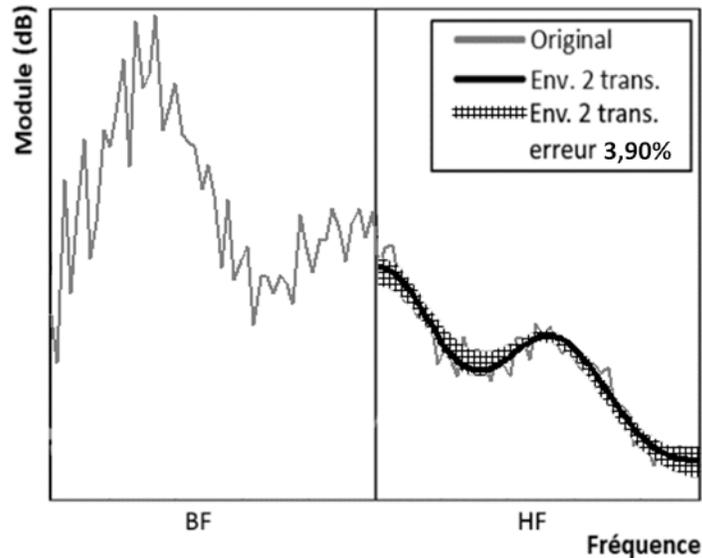


Figure 5-13 : Module du spectre original avec l'enveloppes spectrales originale et les enveloppes spectrales bruitées

Le critère d'évaluation de la distorsion spectrale pour l'enveloppe est l'EQM calculée par l'équation (5-9). Ce calcul est aussi utilisé pour évaluer la distorsion des coefficients cepstraux. Le EQM divisé par sa valeur permet de comparer les deux erreurs sur une base de pourcentage.

$$EQM = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N (x_1 - x_2)^2} \quad (5-9)$$

Les résultats, voir le tableau 5-1, de ce test permettent de conclure que la sensibilité de l'enveloppe est

$S_{c[\tau]}^{Enveloppe} = 0,32$ par rapport aux coefficients cepstraux.

Tableau 5-1 : Comparaison des moyennes de la distorsion des enveloppes et coefficients bruités

Erreurs	Enveloppe spectrale	Coefficients cepstraux	$S_{c[\tau]}^{Enveloppe}$
EQM	0,93	96,23	
EQM/Moyenne	3,90%	12,24%	0,32

Cette valeur de sensibilité est inférieure à l'unité, ce qui permet de conclure que cette représentation de l'enveloppe spectrale est souhaitable pour la quantification des paramètres.

5.1.6 Comparaison de la performance

Le calcul de l'enveloppe par sous-bandes crée des artefacts causés par le fait que l'enveloppe ne suit pas parfaitement le signal. Les bosses et creux ne sont pas bien représentés s'ils chevauchent deux sous-bandes. Il est important de remarquer que même avec une interpolation linéaire ou interpolation par spline cubique ne permet pas une parfaite représentation de l'enveloppe. Il convient de tester ces trois méthodes de calcul de l'enveloppe spectrale avec cinq paramètres de modélisation sur le signal de parole Es02, voir la figure 5-14, et le signal de musique Brahms, voir la figure 5-15.

5.1 Calcul de l'enveloppe spectrale

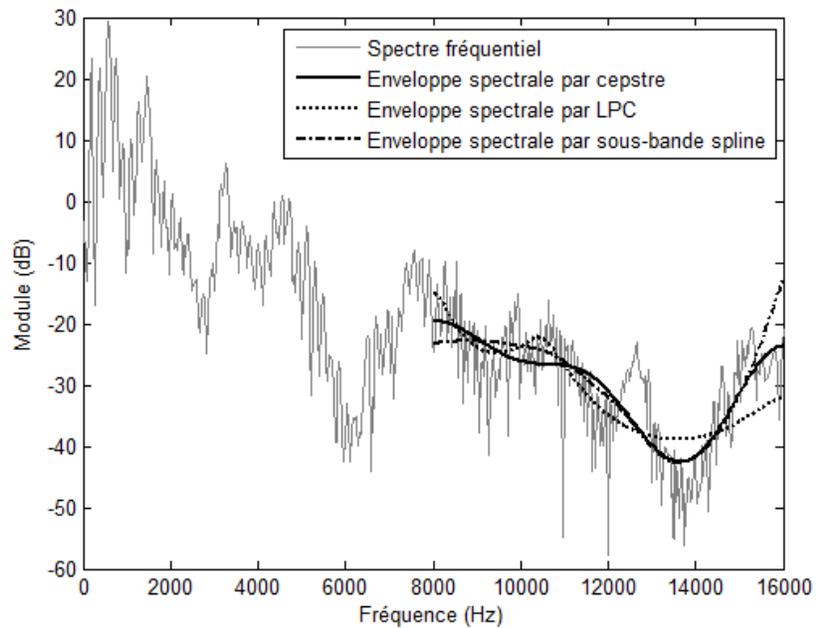


Figure 5-14 : Module du spectre original avec les enveloppes spectrales par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de parole Es02.

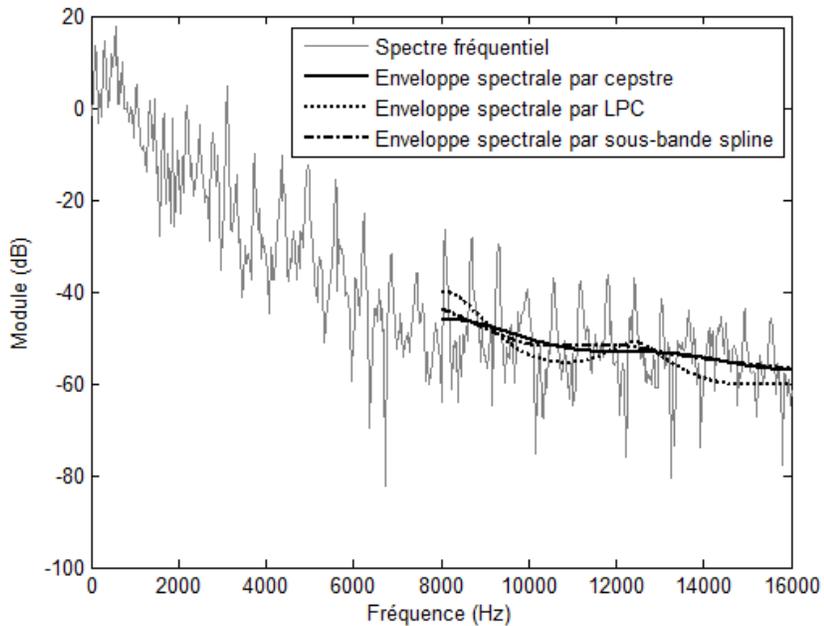


Figure 5-15 : Module du spectre original avec les enveloppes spectrales par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de musique Brahms.

Le signal blanchi offre une autre représentation afin de mieux comparer les enveloppes spectrales pour le signal de parole Es02, voir la figure 5-16, et pour le signal de musique Brahms, voir la figure 5-17. Ils permettent de mieux comprendre l'effet d'une bonne modélisation de l'enveloppe spectrale et de

comparer les techniques. L'enveloppe spectrale doit générer un signal blanchi symétrique horizontalement autour de zéro et ce, sans ondulation lente; tout en gardant la périodicité des harmoniques.

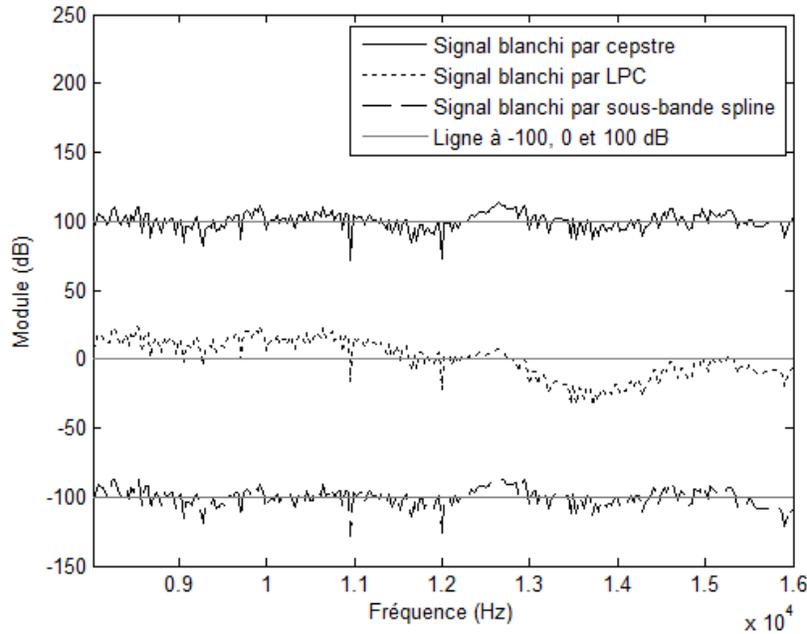


Figure 5-16 : Signaux blanchis par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de parole Es02.

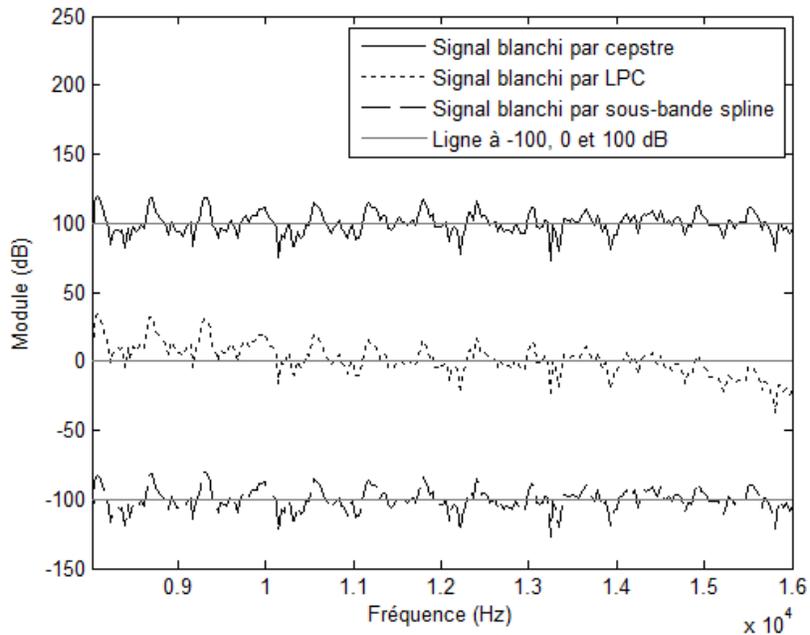


Figure 5-17 : Signaux blanchis par cepstre, par LPC et par sous-bandes interpolées par spline pour le signal de musique Brahms

5.2 Générer les HF

En conclusion, les trois techniques sont semblables. Cependant, la méthode du cepstre suit mieux le module du spectre et génère un signal blanchi plus symétrique et plat. Pour la méthode des sous-bandes interpolées par spline cubique, les artefacts se situent de manière générale au début et à la fin des signaux blanchis. Quant à la méthode par filtrage LPC, es artefacts se situent de manière générale en HF sur les signaux blanchis.

5.2 Générer les HF

L'état de l'art contient plusieurs techniques permettant de générer les HF pour une extension de bande. La technique de symétrie par insertion de zéros publiée [2] est la base de la majorité des modèles d'extension de bande. Comme mentionné au chapitre 4, elle crée quelques artefacts dus à la symétrie. Il existe plusieurs autres techniques pour générer les HF, soit : la translation, de multiples symétries, un agencement de trame, un ajout de bruit, un ajout de tonalités, une application d'une non-linéarité, l'utilisation de changement de fondamentale et plusieurs autres. Les prochaines sections analysent et comparent ces différentes techniques.

5.2.1 Symétrie et translation

Les techniques les plus simples sont la symétrie et la translation, puisqu'elles permettent de conserver l'enveloppe temporelle tout en nécessitant peu de traitement. Pour générer le signal HF, le domaine fréquentiel offre plus de souplesse que le domaine temporel. Afin de réduire les artefacts dus aux formants des BF, la solution proposée est de faire une double symétrie ou double translation. De plus, ces techniques permettent de répliquer deux fois la bande de 4 à 8 kHz dont la corrélation est plus grande avec les bandes de 8 à 12 et 12 à 16 kHz. Les prochaines figures présentent les techniques de symétrie, translation, double symétrie et double translation du spectre.

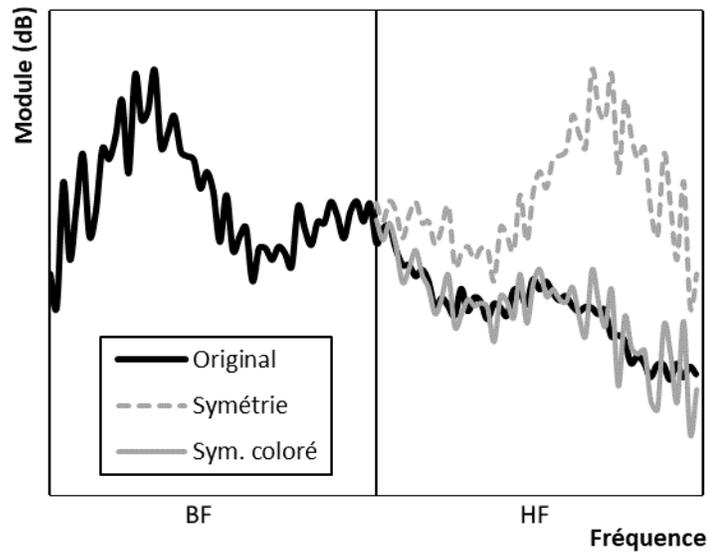


Figure 5-18 : Module du spectre original et généré HF par la symétrie

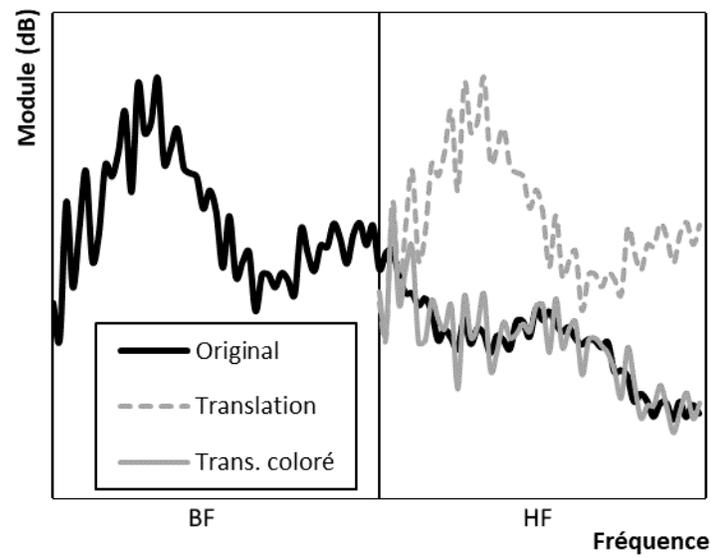


Figure 5-19 : Module du spectre original et généré HF par la translation

5.2 Générer les HF

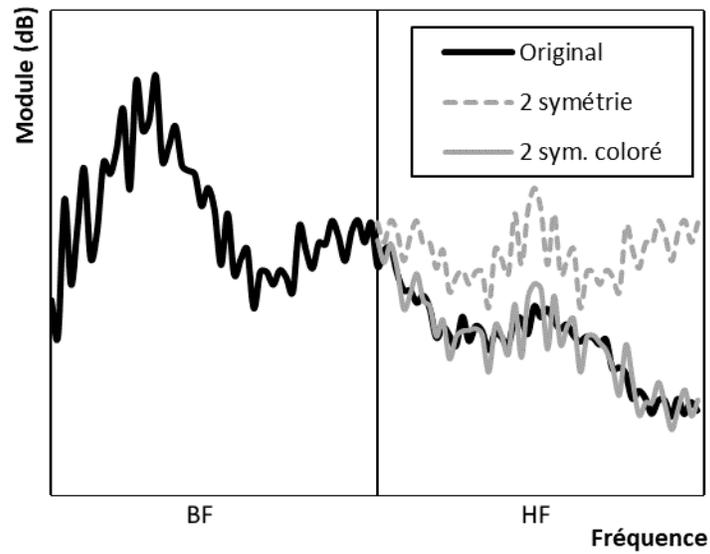


Figure 5-20 : Module du spectre original et généré HF par la double symétrie

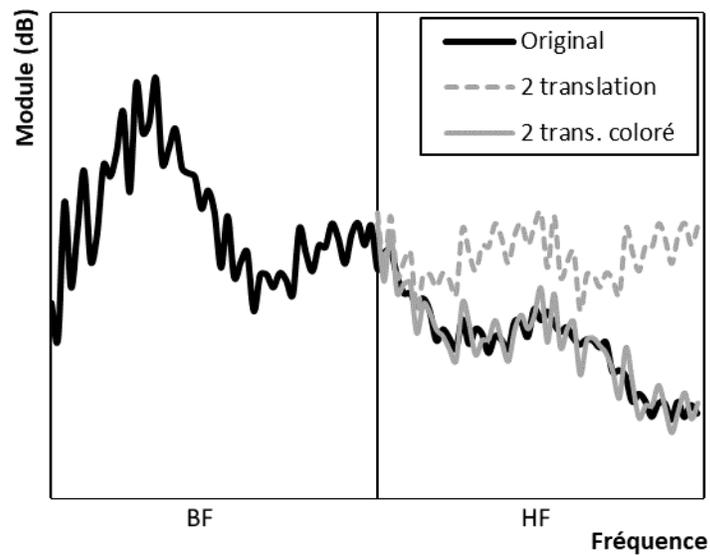


Figure 5-21 : Module du spectre original et généré HF par la double translation

5.2.2 Non-linéarité

L'application d'une non-linéarité sur le signal temporel a pour effet de modifier le spectre fréquentiel et, par le fait même, de générer des HF [3]. L'opérateur non linéaire choisit la carrée en raison de sa simplicité d'exécution dans le domaine fréquentiel, voir la figure 5-22. Selon les propriétés de la TFD, la multiplication point à point de deux signaux revient à faire une convolution circulaire de ces deux signaux dans le domaine fréquentiel, tel que décrit dans l'équation (5-10).

$$x[n] \cdot y[n] \rightarrow X(e^{j\theta}) * Y(e^{j\theta}) \quad (5-10)$$

La convolution retourne le nombre d'échantillons équivalents à la somme de la longueur des deux signaux moins un. Afin d'atteindre le nombre d'échantillons désirés, le dernier échantillon doit être doublé.

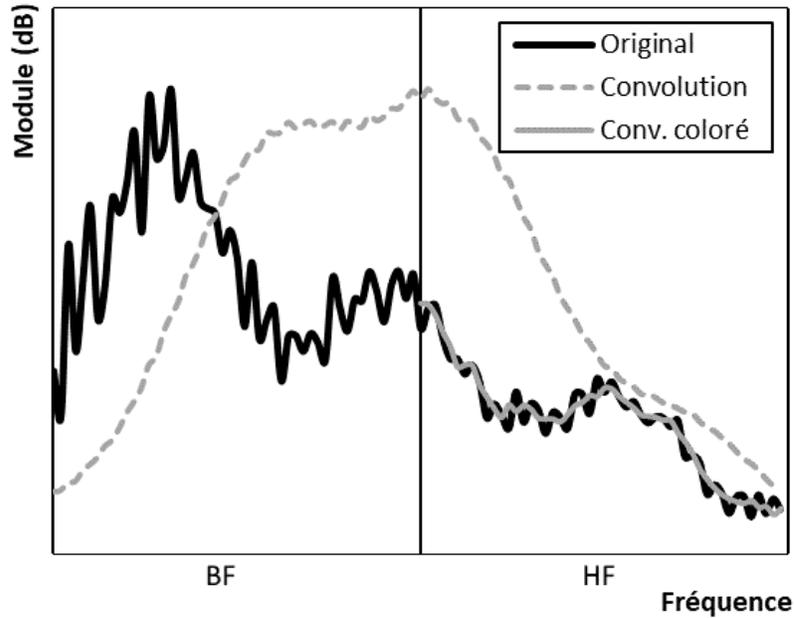


Figure 5-22 : Module du spectre original et généré HF par la convolution

5.2.3 Ajout de bruit blanc gaussien

Le signal blanchi peut être remplacé par du bruit blanc gaussien, voir la figure 5-23. La variance du bruit peut être estimée ou transmise selon le débit disponible. Cette technique donne de bons résultats pour les signaux ayant des caractéristiques de sons fricatifs.

5.2 Générer les HF

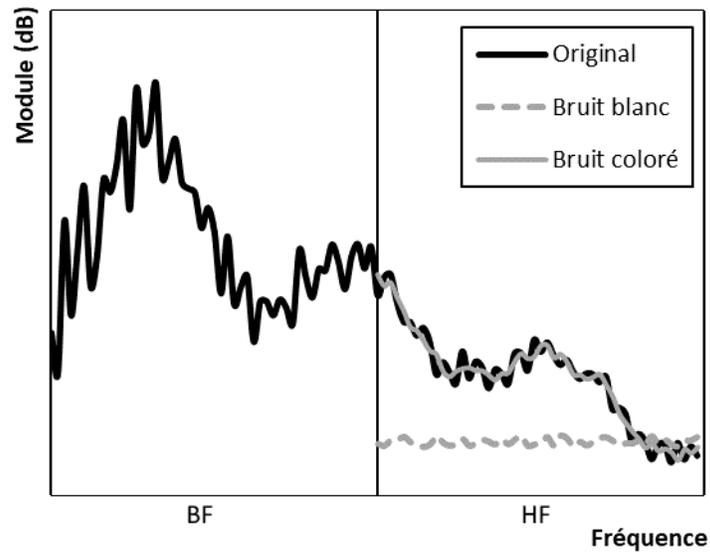


Figure 5-23 : Module du spectre original et généré avec bruit blanc

5.2.4 Somme de suréchantillonnages

Cette technique, inspirée du vocodeur de phase, permet de régénérer des signaux HF par addition de plusieurs spectres suréchantillonnés, voir la figure 5-24. Les différents taux de suréchantillonnage sont des nombres premiers ce qui permet de ne pas superposer les harmoniques, soit : 1, 2, 3, et 5.

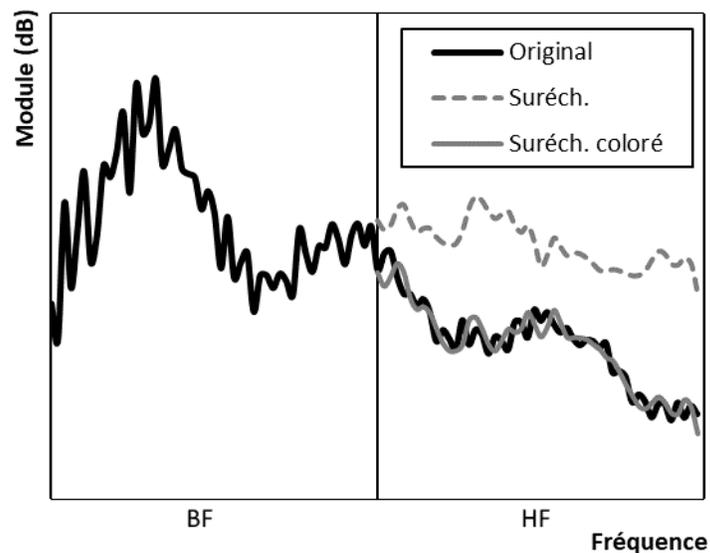


Figure 5-24 : Module du spectre original et généré par suréchantillonnages multiples

5.2.5 Méthode de sélection

Selon plusieurs comparaisons subjectives de fichiers sonores, la méthode la plus performante diffère d'un fichier à l'autre. Les techniques de symétrie double et translation double offrent de meilleurs

résultats subjectifs comparativement aux autres méthodes. Ceci est dû au fait que les phases sont en lien avec les modules générés.

Afin de sélectionner parmi les techniques présentées, il est impératif de créer un outil d'évaluation de la performance des techniques de régénération des signaux HF. Il faut concevoir toutes les méthodes et les comparer à l'aide d'un calcul de performance. L'EQM permet de calculer la distance spectrale du signal blanchi (5-11). Les valeurs calculées indiquent la distorsion entre les deux signaux. Cependant, il faut adapter cet outil de mesure aux notions psychométriques et au codage perceptuel. L'oreille humaine est moins sensible à un léger décalage en fréquence, mais est plus sensible à l'amplitude et à la dynamique des modules du spectre. L'ajout de décalages à l'équation (5-12) inspiré du calcul de la corrélation non biaisée apporte une souplesse dans la synchronisation des signaux. Cette valeur indique aussi pour quel décalage les signaux spectraux sont le plus semblables. Cet outil indique une forte ressemblance avec une valeur faible selon le décalage l . Un deuxième outil de calcul permettant d'évaluer la ressemblance entre deux spectres est la corrélation spectrale non biaisée (5-13). Cet outil indiquera une forte ressemblance avec une valeur élevée selon le décalage l .

$$DistSpec_{blanc} = \sqrt{\frac{1}{N} \sum_{k=0}^{N-1} (X_{blanc}[k] - Y_{blanc}[k])^2} \quad (5-11)$$

$$DistSpec_{blanc}[l] = \sqrt{\frac{1}{N - |l|} \sum_{k=0}^{N-1} (X_{blanc}[k] - Y_{blanc}[k - l])^2} \quad (5-12)$$

$$CorrSpec_{blanc}[l] = \frac{1}{N - |l|} \sum_{k=0}^{N-1} X_{blanc}[k] \cdot Y_{blanc}[k - l] \quad (5-13)$$

Cet outil de performance indique laquelle des sept techniques de génération des HF est la plus performante pour chaque trame. Les calculs de la $DistSpec_{blanc}[l]$ et de la $CorrSpec_{blanc}[l]$ permettent de faire des statistiques sur les performances de chacune des techniques. À la suite de ces analyses, les techniques les plus performantes sont la translation et la double translation selon l'outil de sélection, voir la figure 5-25.

5.2 Générer les HF

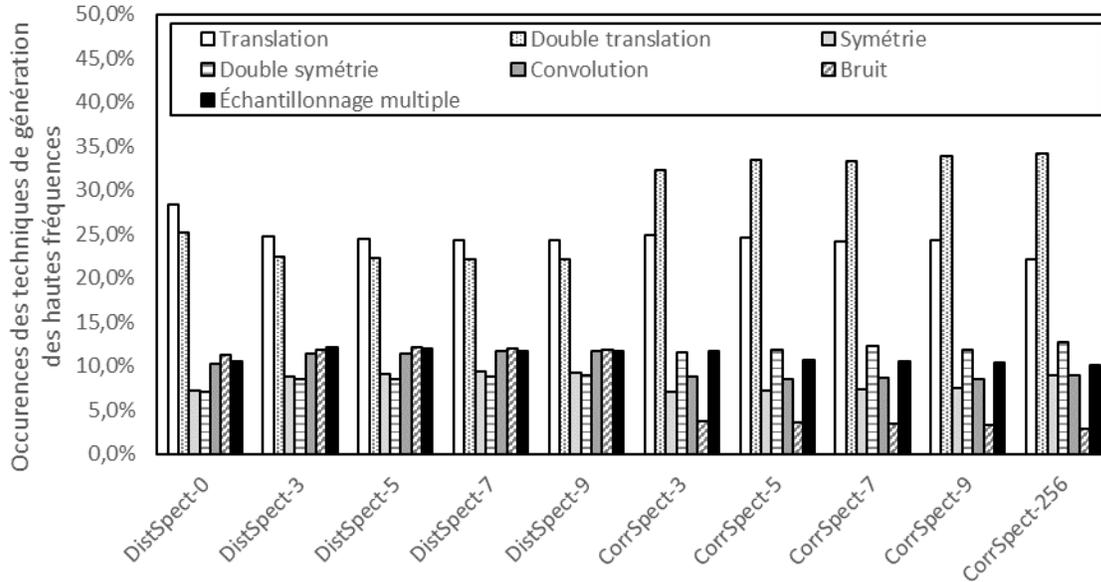


Figure 5-25 : Statistique de la performance des techniques de régénération de signaux HF

À la compilation de ces résultats, nous concluons qu'il est possible de faire trois choix de conception pour le modèle d'extension de bande :

- Transmettre quelle technique doit être utilisée, mais ceci exige un débit supplémentaire;
- Estimer la technique la plus adéquate au récepteur à l'aide d'un outil de décision en fonction du type de signal et de sa catégorie;
- Utiliser la technique qui offre les meilleurs résultats.

La figure 5-26 et la figure 5-27 montrent les pourcentages d'utilisation des techniques de régénération des HF en comparaison avec la symétrie qui est utilisée dans l'AMR-WB+.

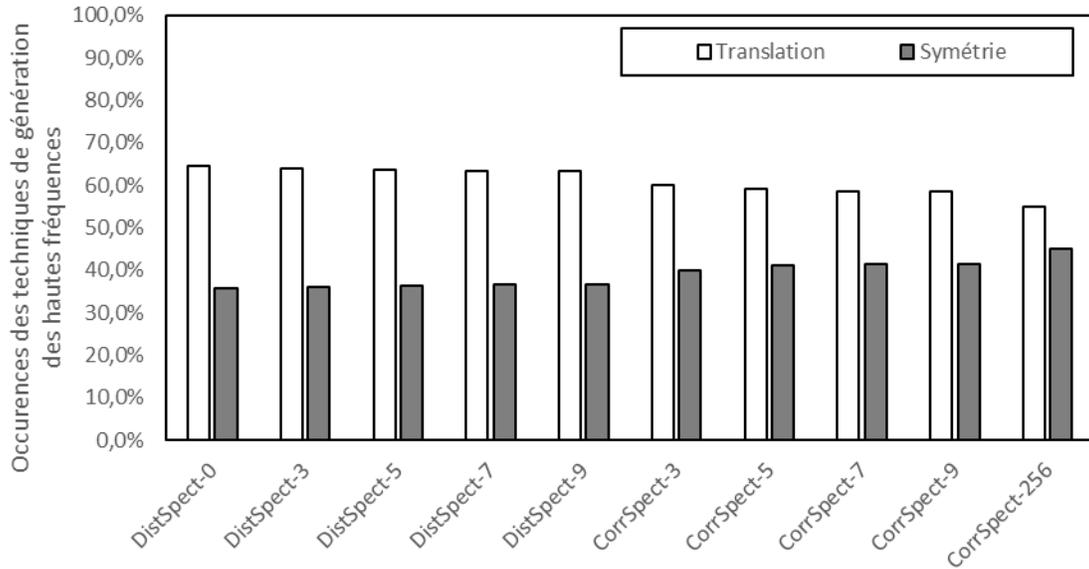


Figure 5-26 : Distribution des techniques entre la symétrie et la double translation

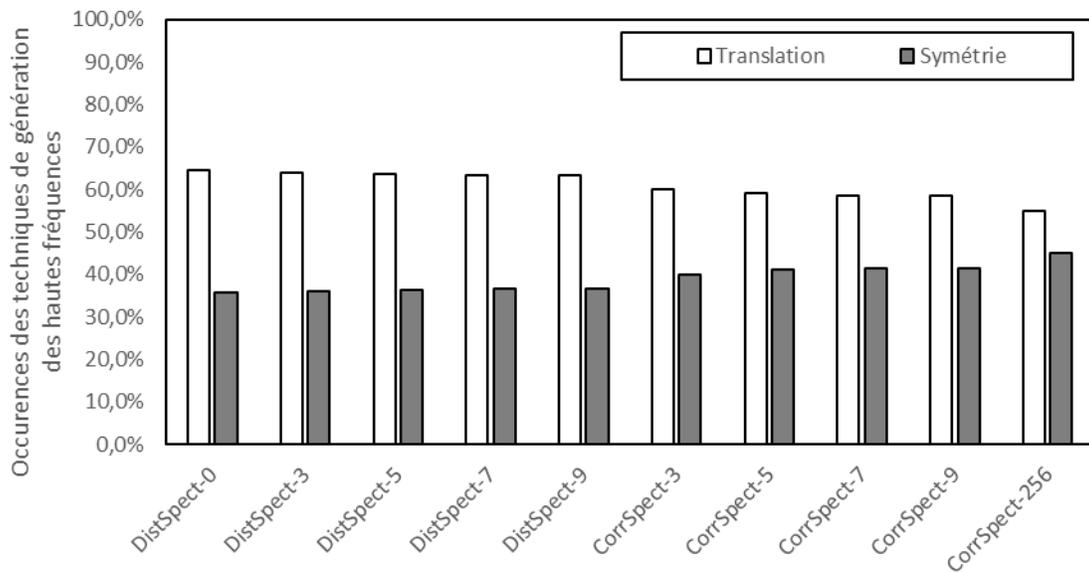


Figure 5-27 : Distribution des techniques entre la translation et la symétrie

Les deux techniques les plus utilisées sont la translation et la double translation. Ces tests montrent aussi que la double translation n'est pas toujours la meilleure, la figure 5-28. Des tests subjectifs montrent qu'il y a quelques artefacts sur certains fichiers de musique avec la méthode de translation. Dans ces conditions, la double translation est choisie et aucune information n'est transmise au décodeur. La double translation n'est pas toujours la meilleure technique, mais elle permet une bonne amélioration face à l'extension de bande de référence.

5.3 Amélioration du signal blanchi

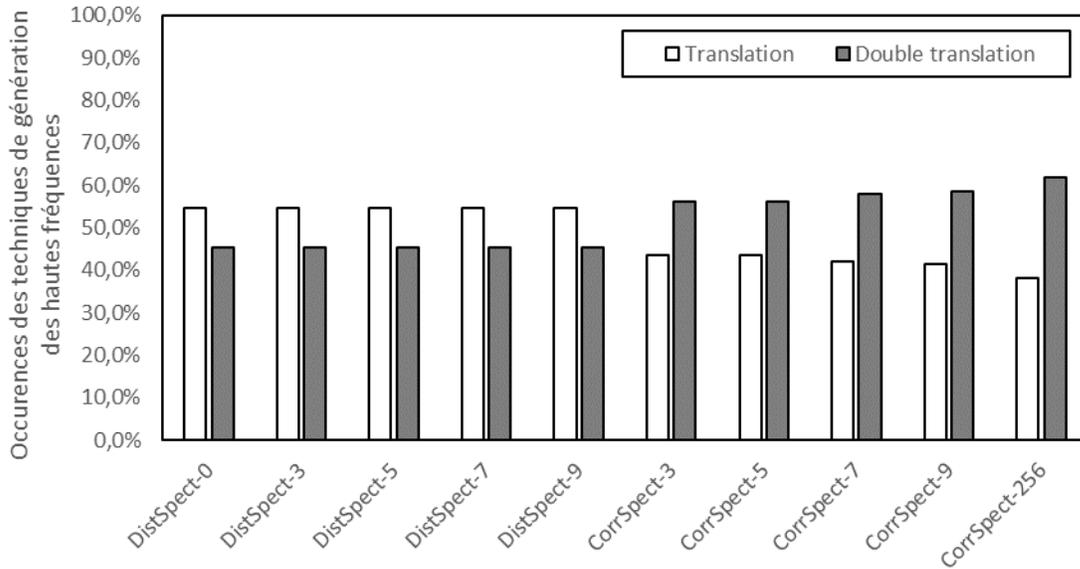


Figure 5-28 : Distribution des techniques entre la translation et la double translation

5.3 Amélioration du signal blanchi

L'enveloppe étant de bonne qualité, la majorité des artefacts restants sont causés par le signal blanchi, voir la figure 5-29. Plusieurs méthodes permettent d'améliorer ce signal. Ce projet de maîtrise propose une technique de changement de la variance du signal blanchi.

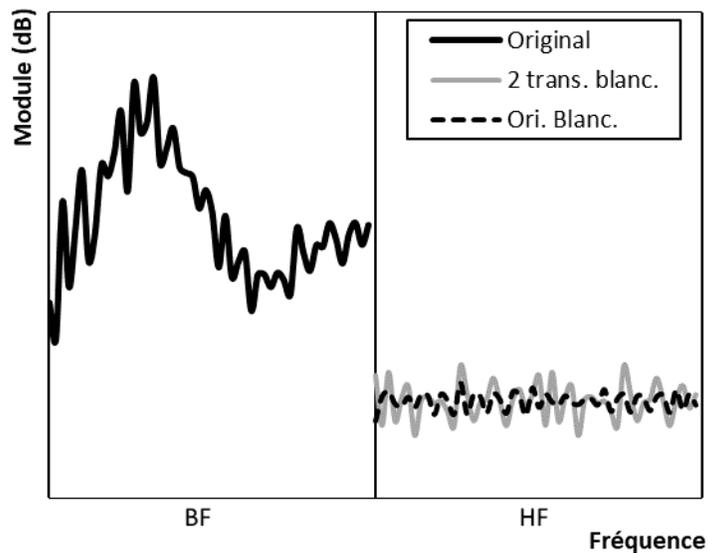


Figure 5-29 : Signal blanchi original et généré sans modification de la variance

5.3.1 Variance

Dans la modélisation des signaux paramétriques, les outils statistiques sont de bons paramètres de modélisation. Pour un signal blanchi, la moyenne se situe autour de la valeur unitaire (0 dB) en raison

de la division du spectre par son enveloppe spectrale ou de la soustraction par son enveloppe spectrale en dB. La moyenne est inutile pour modéliser le signal blanchi puisqu'elle possède toujours la même valeur.

Cependant, la variance est plus utile, puisqu'elle modélise la dynamique du signal blanchi. Cet outil permet de modéliser l'harmonicité d'un signal. Une grande variance signifie que le signal blanchi varie beaucoup, donc le signal sonore est très harmonique. Par opposition, les signaux de bruit possèdent généralement une variance faible. Cette mesure est utile sur tous les signaux, fricatifs, voisés ou harmoniques.

Le calcul de la variance $V[k]$ s'effectue de manière mobile sur 17 coefficients blanchis, soit les huit coefficients précédents et les huit coefficients selon la position évaluée d'indice k . Les valeurs de variances forment alors une courbe de variances où $M = 16$ et $M + 1$ est le nombre de coefficients de la moyenne mobile (5-14) et de la variance mobile (5-15). Pour les extrémités où le calcul est impossible par manque de coefficients, les valeurs des variances sont interpolées linéairement.

$$\bar{X}_{blanc}[k] = \frac{1}{M} \sum_{m=-M/2}^{M/2} X_{blanc}[k+m] \quad (5-14)$$

$$V[k] = \frac{1}{M} \sum_{m=-K/2}^{K/2} (X_{blanc}[k+m] - \bar{X}_{blanc}[k])^2 \quad (5-15)$$

5.3.2 Changement de variance

Le changement de variance se fait sur tous les échantillons HF positifs d'indice k avec l'équation (5-16). Le signal blanchi ajusté présente les caractéristiques de variances du signal blanchi original permettant de corriger les artefacts d'harmonicités, voir la figure 5-30.

$$X_{blanc-var}[k] = X_{blanc}[k] \cdot \frac{V_{Ori}[k]}{V_{Gén}[k]} \quad \text{avec } 256 \leq k \leq 511 \quad (5-16)$$

5.3 Amélioration du signal blanchi

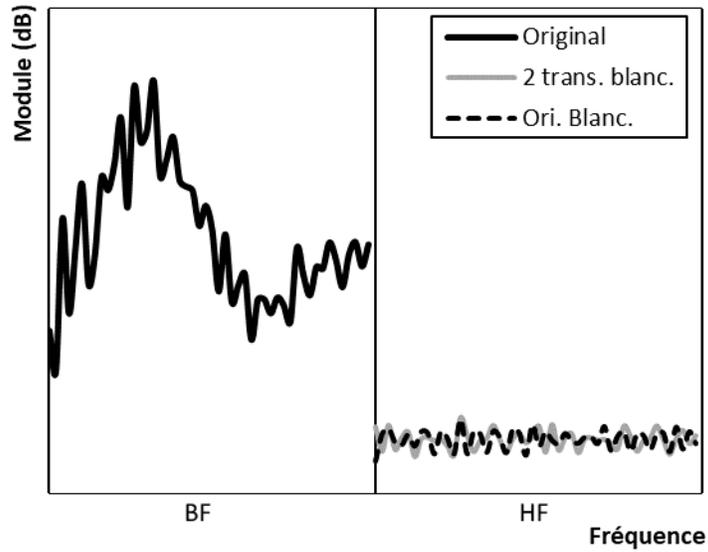


Figure 5-30 : Signal blanchi original et généré avec modification de la variance

5.3.3 Courbe de variances

Le changement de variances permet de résoudre la majorité des artéfacts. Cependant ceci revient à devoir transmettre toutes les valeurs de variances pour les HF. Afin de réduire la quantité d'informations transmises, le modèle proposée est une courbe de variances modélisant la variance mobile. La méthode de calcul de la courbe de variances se fait par sous-bandes et est choisie pour sa simplicité.

Le signal blanchi HF est subdivisé en quatre sous-bandes de 64 coefficients spectraux. Cette quantité de sous-bandes permet un bon compromis entre une bonne représentation de la courbe de variances et la quantité d'informations à transmettre. Pour chaque sous-bande d'indice i , les paramètres statistiques se calculent comme suit : la variance par sous-bande V_i avec l'équation (5-17) et la moyenne par sous-bande \bar{X}_i avec l'équation (5-18) où M est le nombre d'échantillons de la sous-bande. La courbe de variances est formée de paliers associés à chaque sous-bande, voir la Figure 5-31.

$$V_i = \frac{1}{M} \sum_{k=1}^M (X_{blanc}[k] - \bar{X}_i)^2 \quad (5-17)$$

$$\bar{X}_i = \frac{1}{M} \sum_{k=1}^M X_{blanc}[k] \quad (5-18)$$

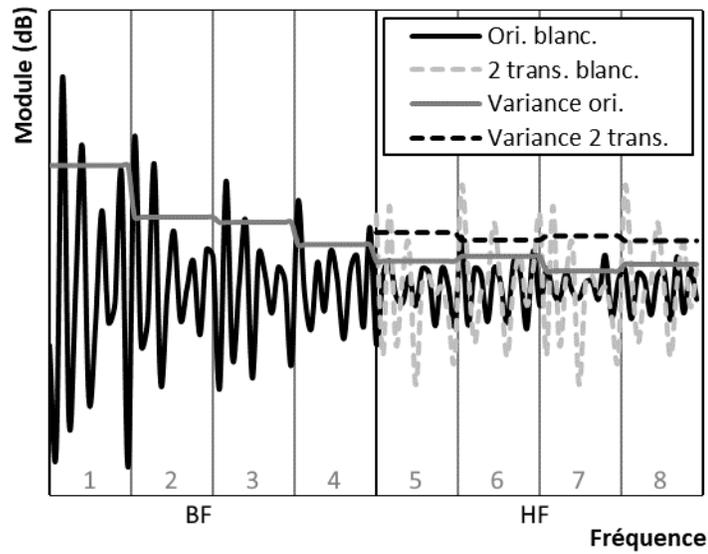


Figure 5-31 : Signal blanchi original et généré sans modification avec courbes de variances

Au décodeur, les variances reçues et calculées sont interpolées linéairement pour retrouver les courbes de variances et ainsi, minimiser les transitions brusques, voir la figure 5-32 sans interpolation et la figure 5-33 avec interpolation linéaire des variances. Le spectre final, voir la figure 5-34, présente des caractéristiques nettement plus proches que celles sans ajustement par les courbes de variances.

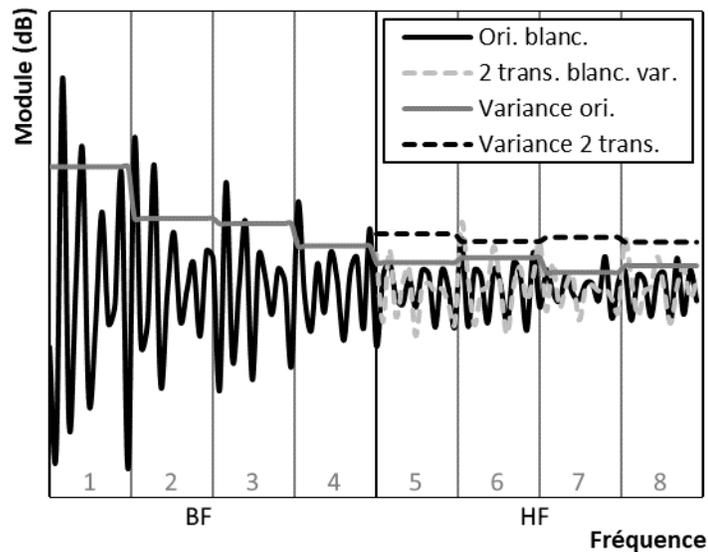


Figure 5-32 : Signal blanchi original et généré modifié par les courbes de variances

5.3 Amélioration du signal blanchi

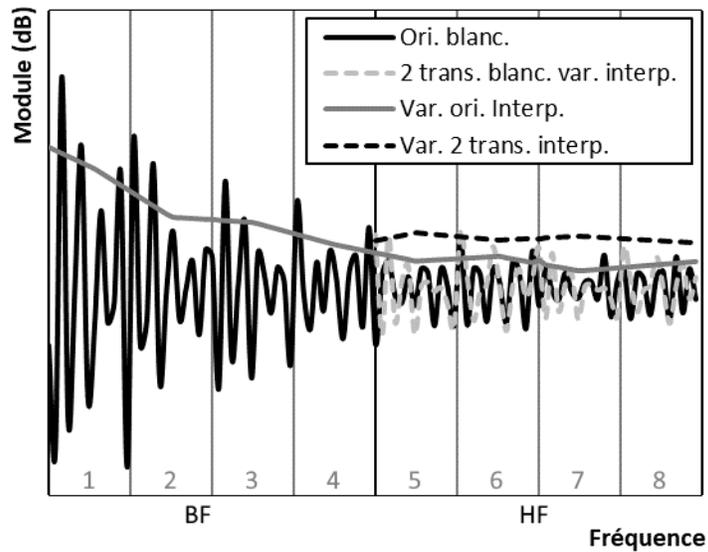


Figure 5-33 : Signal blanchi original et généré modifié par les courbes de variances interpolées

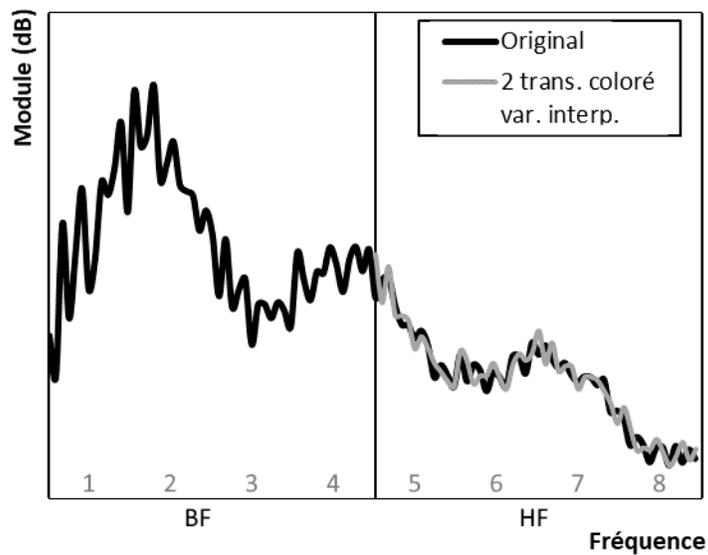


Figure 5-34 : Module du spectre original et final par variances interpolées

5.3.4 Comparaison des performances

L'outil de changement de variances augmente grandement la qualité du signal. En comparaison avec le SBR, le choix de bande à transmettre permet aussi d'ajuster la dynamique du module du spectre. Ce traitement est effectué sur le signal temporel tout en permettant d'obtenir des résultats semblables. En vue d'implémenter un tel système, deux problèmes importants doivent être résolus soit : comment déterminer les bandes à répliquer et quelle quantité d'informations sont à transmettre.

En conclusion, la méthode d'amélioration de la qualité du signal blanchi la plus performante est le changement des variances. La méthode de caractérisation de la dynamique du signal blanchi est le calcul des variances sur quatre sous-bandes permettant de bien représenter les caractéristiques du signal blanchi HF. Par la suite, le changement de variances au décodeur s'effectue à l'aide de courbes de variances créées par l'interpolation linéaire des valeurs de variances reçues et calculées.

5.4 Résumé des outils proposés

Ce chapitre présente les outils proposés pour modéliser les paramètres de l'extension de bande développée. Pour chaque paramètre, quelques outils ont été comparés afin de choisir les plus performants tout en ayant un regard sur le débit nécessaire à la transmission de ces paramètres.

Pour la génération du module du spectre HF, plusieurs méthodes ont été comparées dans le but de trouver la plus performante. Chaque signal étant unique, l'analyse des résultats ne permet pas de conclure sur une méthode plus performante. En raison du faible débit disponible, le choix de la double translation est un compromis entre les méthodes les plus performantes. Cette méthode permet d'optimiser la corrélation des signaux des HF avec la moitié supérieure des BF.

Le paramètre le plus important du modèle d'extension de bande proposé est l'enveloppe spectrale. Elle est un outil utilisé dans la majorité des extensions de bande et plus particulièrement par filtrage LPC. La technique de calcul de l'enveloppe spectrale pour cette maîtrise est d'utiliser le cepstre. Les trois outils comparés permettent d'obtenir une bonne modélisation avec cinq coefficients à transmettre, cependant la technique par le cepstre est supérieure. Elle répond aux trois critères de performance évoqués pour l'évaluation des techniques de calcul de l'enveloppe spectrale. Les traitements de l'extension de bande développée étant faits dans le domaine fréquentiel, le choix du cepstre permet d'obtenir une enveloppe spectrale ayant la bonne amplitude. Cette caractéristique permet de calculer le signal blanchi et remplace l'enveloppe temporelle.

Le nouvel outil proposé, afin d'améliorer la qualité du signal blanchi, est le changement de variance. L'utilisation des outils statistiques pour modéliser un signal est très répandue dans le codage paramétrique des signaux. Inspirée des modèles de l'enveloppe fréquentielle et temporelle, la courbe de variances interpolées linéairement permet d'effectuer une modification du signal blanchi limitant ainsi l'effet des sous-bandes. En augmentant ou en diminuant le nombre de sous-bandes, il est possible de choisir la qualité du signal blanchi. Un nombre élevé de sous-bandes permet d'ajouter ou de retirer des tonalités.

5.4 Résumé des outils proposés

Finalement, ces outils permettent de bien modéliser et de régénérer le module du signal HF. La quantification de ces paramètres demeure un enjeu de taille puisque l'ajout de la variance a pour effet de devoir transmettre un paramètre de plus que le codec de référence.

CHAPITRE 6 QUANTIFICATION ET COMPRESSION DES DONNÉES

Ce chapitre présente les différentes techniques de compression utilisées afin de répondre au besoin de compression des paramètres de l'extension de bande. À 16 bits par trame, le débit disponible est très faible, pour un débit égal avec l'extension de bande de référence.

6.1 Données à compresser

Les données à transmettre, provenant des différents paramètres de l'extension de bande, sont composées de cinq coefficients cepstraux, représentant l'enveloppe spectrale, auxquelles on ajoute quatre valeurs de variances.

Pour les coefficients cepstraux, les valeurs sont corrélées entre elles. Cependant, la première valeur, la fréquence zéro, a un impact plus significatif puisqu'elle donne l'amplitude de l'enveloppe spectrale. Cette fréquence est beaucoup plus sensible sur la distorsion spectrale de l'enveloppe que les autres. Pour la variance, les quatre valeurs sont corrélées et leur sensibilité est équivalente. L'entropie est directement liée au taux de compression du signal. Plus un signal a une entropie forte, plus il requiert de bits pour le compresser [2], [25]. Les distributions des coefficients à quantifier sont tous non uniformes, voir la figure 6-1 pour les fréquences et la figure 6-2 pour les variances. Puisque leur règle de distribution et leur plage dynamique sont inconnues, des quantificateurs avec entraînement permettent de diminuer les erreurs de quantifications.

6.1 Données à compresser

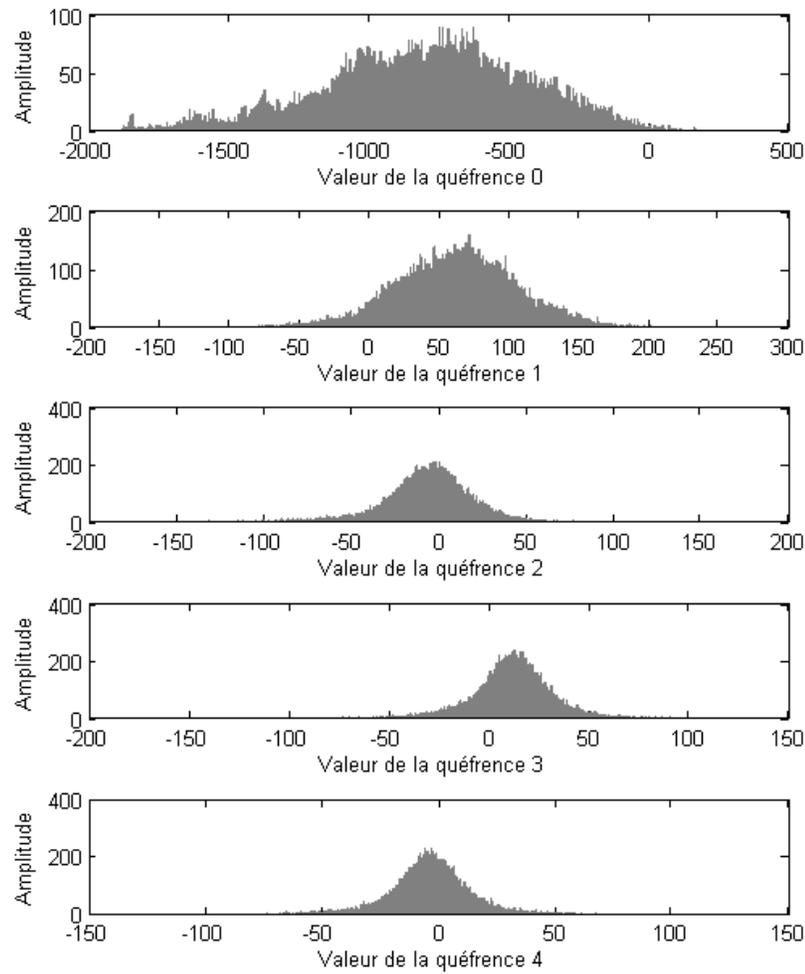


Figure 6-1 : Distributions des quéfrences de 0 à 4

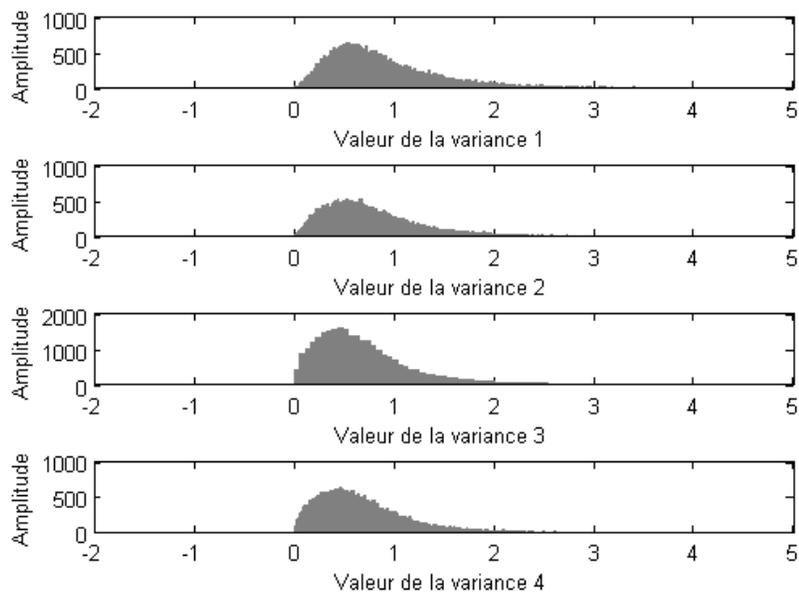


Figure 6-2 : Distributions des variances de 1 à 4

6.2 Quantification de l'enveloppe spectrale

Les coefficients cepstraux permettent de représenter l'enveloppe spectrale de manière compacte en cinq coefficients. Ces coefficients transmis sont les cinq premières quéfrences positives du cepstre. Les quéfrences un à quatre sont répliquées par symétrie afin de créer les quéfrences négatives non transmises.

Le premier coefficient représente l'amplitude moyenne de l'enveloppe spectrale, tandis que les quatre autres représentent les oscillations présentes dans l'enveloppe. L'oreille humaine est plus sensible aux erreurs d'amplitudes que sur la forme exacte de l'enveloppe spectrale. De ce fait, le premier coefficient cepstral est quantifié séparément par un quantificateur scalaire comportant 32 niveaux, soit cinq bits. Les autres coefficients sont quantifiés par un quantificateur vectoriel à quatre dimensions comportant 64 vecteurs, soit six bits.

Les amplitudes des enveloppes peuvent être estimées en joignant la fin de l'enveloppe BF avec le début de l'enveloppe HF. L'estimation se calcule comme l'arrimage entre les extrémités des enveloppes BF et HF. Lorsque la quéfrence zéro est nulle pour l'enveloppe HF, la soustraction des extrémités des enveloppes BF et HF des enveloppes spectrales permet d'estimer le gain de l'enveloppe spectrale, voir la figure 6-3. La valeur de gain estimée est transformée en coefficient cepstral $c[0]$ avec l'équation (6-2), où N est le nombre d'échantillons de la TFD pour le calcul de l'enveloppe spectrale. Étant donné que la méthode de calcul par cepstre considère que l'enveloppe est périodique, la valeur de l'extrémité de chaque enveloppe n'est pas une bonne représentation de l'amplitude pour toutes les trames. Le calcul proposé pour résoudre ce problème est la moyenne des dix coefficients des extrémités de chaque enveloppe (6-1).

$$Amplitude_{estimée} = \sum_{k=1}^{10} \frac{X_{BF}[256 - k]}{10} - \sum_{k=1}^{10} \frac{X_{HF}[k - 1]}{10} \quad (6-1)$$

$$c_{estimé}[0] = N \cdot Amplitude_{estimée} \quad (6-2)$$

Avec cette méthode d'estimation, l'information à quantifier et à transmettre est l'erreur d'estimation $\hat{c}[0]$, voir l'équation (6-3), au lieu du coefficient cepstral zéro $c[0]$.

$$\hat{c}[0] = c_{original}[0] - c_{estimé}[0] \quad (6-3)$$

6.2 Quantification de l'enveloppe spectrale

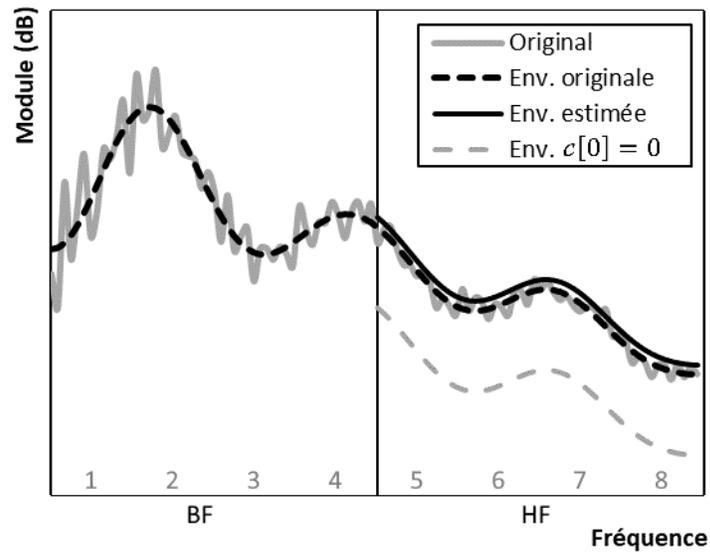


Figure 6-3 : Enveloppe spectrale par cepstre avec estimation de l'amplitude

Les deux quantificateurs sont entraînés avec une base de données constituée des 43 fichiers sonores de MPEG, soit 28035 trames de 16 millisecondes. Pour chaque trame, les coefficients cepstraux des enveloppes spectrales sont extraits.

6.2.1 Recherche du plus proche voisin dans les domaines cepstral et fréquentiel

La recherche du plus proche voisin dans le dictionnaire d'un quantificateur vectoriel se fait en minimisant la distance entre les vecteurs à quantifier et celui du dictionnaire. Cette distance doit permettre de trouver le même indice entre le dictionnaire de coefficients cepstraux et le dictionnaire correspondant composé d'enveloppes spectrales. Le choix de quantifier les enveloppes spectrales dans le domaine spectral permet de diminuer la complexité du calcul du plus proche voisin en comparant les paramètres de modélisation, soit cinq coefficients cepstraux ou 256 coefficients spectraux pour la même enveloppe spectrale.

Soit les vecteurs $c(\tau)$ et $\delta(\tau)$ tous deux composés de coefficients cepstraux (qu'éfrenes) tous réels avec un indice de qu'éfrence τ . Le vecteur $c(\tau)$ représente les coefficients cepstraux de l'enveloppe spectrale $X(k)$ et $\delta_i(\tau)$ représente le vecteur d'indice i de coefficients cepstraux de l'enveloppe spectrale $\mathcal{D}_i(k)$ provenant du dictionnaire du quantificateur vectoriel.

La TFDI du vecteur $c(\tau)$ et $\delta(\tau)$ donnent respectivement les vecteurs de fréquences $X(k)$ et $\mathcal{D}(k)$, soit des enveloppes spectrales :

$$X[k] = \sum_{\tau=0}^{N-1} c[\tau] \cdot e^{-\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq k \leq N-1 \quad (6-4)$$

$$c[\tau] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} X[k] \cdot e^{\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq \tau \leq N-1 \quad (6-5)$$

$$\mathcal{D}_i[k] = \sum_{\tau=0}^{N-1} \delta_i[\tau] \cdot e^{-\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq k \leq N-1 \quad (6-6)$$

$$\delta_i[\tau] = \frac{1}{N} \cdot \sum_{k=0}^{N-1} \mathcal{D}_i[k] \cdot e^{\frac{j2\pi k\tau}{N}} \quad \text{pour } 0 \leq \tau \leq N-1 \quad (6-7)$$

La distance dans le domaine fréquentiel entre deux enveloppes se calcule selon l'équation (6-8). La distance dans le domaine cepstral entre deux vecteurs de coefficients cepstraux se calcule selon l'équation (6-9) entre deux vecteurs de coefficients cepstraux.

$$Distance_{\text{fréquentiel}}^2 = \sum_{k=0}^{N-1} (X[k] - \mathcal{D}[k])^2 \quad (6-8)$$

$$Distance_{\text{cepstre}}^2 = \sum_{\tau=0}^{N-1} (c[\tau] - \delta[\tau])^2 \quad (6-9)$$

En utilisant l'égalité de Parseval, les $Distance_{\text{fréquentiel}}^2$ et $Distance_{\text{cepstre}}^2$ ne sont pas égales, les relations de distance entre le vecteur à quantifier et les vecteurs du dictionnaires sont les mêmes peu importe le domaine de représentation [26], [27]. En comparant les deux méthodes de recherche, spectrale et cepstrale; les résultats des tests faits sur toutes les enveloppes de l'entraînement (28035) sont concluants. Ils ne démontrent aucune différence de choix d'index de dictionnaire en comparant les deux méthodes de recherche, spectrale et cepstrale. Le calcul du plus proche voisin dans le domaine du cepstre permet de réduire la complexité et diminue la taille des vecteurs du dictionnaire.

6.3 Compression de la variance du signal blanchi

Les données de variances sont composées de quatre valeurs corrélées entre elles. De plus, ces valeurs sont aussi corrélées avec les valeurs de variances des BF. Les quatre valeurs à transmettre sont quantifiées vectoriellement afin d'utiliser la caractéristique de corrélation entre les valeurs. Cependant, cette technique ne permet pas d'utiliser la caractéristique de corrélation entre les valeurs de variances des BF et des HF.

6.3 Compression de la variance du signal blanchi

Un estimateur des variances des HF, en fonction des variances des BF, permet de maximiser la corrélation entre les signaux de variances de BF et de HF, voir la figure 6-4. Contrairement au signal blanchi généré, dont la variance est relativement constante; le signal blanchi original présente une dynamique décroissante en fonction des fréquences.

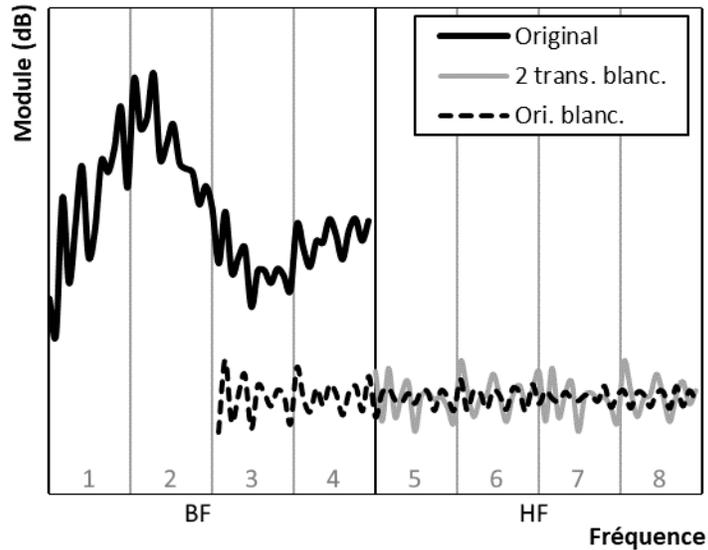


Figure 6-4 : Signal blanchi original et généré

L'estimation des valeurs se base sur les deux dernières sous-bandes des BF. Si la variance diminue de manière décroissante, comme la majorité des signaux, la variance estimée HF $V_{HF-estimé}$ est la valeur interpolée linéairement (6-10), et ce, selon la pente entre les variances des deux dernières sous-bandes.

$$V_{HF-estimé} = V_3 - (V_3 - V_4) \quad (6-10)$$

Les variances des sous-bandes trois à cinq sont décroissantes. La valeur estimée de la variance, pour les sous-bandes cinq à huit, est relativement proche des valeurs originales, voir la figure 6-5.

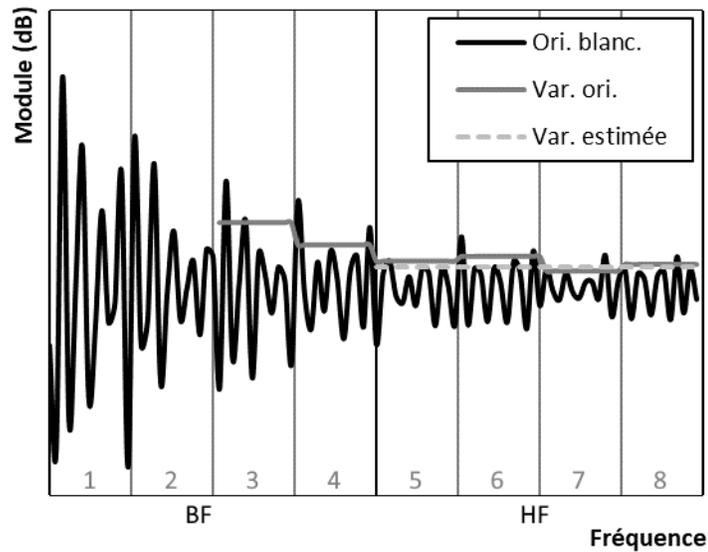


Figure 6-5 : Signal blanchi original et généré avec courbe de variances estimées

Les valeurs à transmettre sont les erreurs d'estimation \hat{V}_l des quatre variances calculées par (6-11), avec l'indice l étant le numéro de la sous-bande. Ces valeurs sont quantifiées par un quantificateur vectoriel à quatre dimensions comportant 32 vecteurs, soit cinq bits. Le quantificateur est entraîné par la base de données constituée des 43 fichiers sonores de MPEG, soit : 28035 trames de 16 millisecondes. Pour chaque trame, les erreurs de variances à transmettre sont utilisées comme vecteur d'entraînement.

$$\hat{V}_l = V_l - V_{HF-estimé} \quad (6-11)$$

6.4 Budget de bits

Plusieurs possibilités d'allocation des bits sont possibles. Le but du projet de maîtrise est de comparer l'extension de bande proposée avec l'extension de bande de référence dont le débit est de 1 kbit/seconde. Les trois versions de l'extension de bande qui sont comparées, soit à débit plus faible, égal et plus grand, voir le tableau 6-1 pour le budget de bit associé aux paramètres des trois débits proposés.

6.5 Résumé de la quantification

Tableau 6-1 : Allocation du budget de bits pour la compression

	Gain enveloppe (bits/trame)	Coefficients cepstraux (bits/trame)	Variance (bits/trame)	Total (bits/trame)	Total (kbits/s)
UFR1	4	4	4	12	0,75
UFR2	5	6	5	16	1,00
UFR3	6	8	6	20	1,25

6.5 Résumé de la quantification

Les neuf paramètres à quantifier possèdent des distributions non uniformes dont les plages dynamiques ne sont pas identiques nécessitant des quantificateurs par entraînement. Ils sont quantifiés par des quantificateurs scalaires et vectoriels dont l'entraînement se fait avec une très grande quantité de vecteurs. Les bits alloués à chaque paramètre varient pour les trois versions de l'extension de bande tels que présentés dans le tableau 6-1.

Les coefficients cepstraux représentant l'enveloppe spectrale sont quantifiés avec deux quantificateurs. Un premier pour l'erreur d'estimation de la fréquence zéro et un deuxième pour les quatre autres fréquences. Cette conception permet de diminuer les erreurs de quantification sur la fréquence zéro, ayant comme conséquence de limiter la distorsion spectrale de l'enveloppe spectrale.

Les quatre coefficients de variance sont estimés en interpolant la pente des variances des deux dernières sous-bandes des BF. Cette estimation permet d'utiliser la corrélation des variances du signal blanchi BF et FH. De plus, elle permet de concentrer la distribution des données limitant les erreurs de quantification. Les erreurs d'estimation sont quantifiées avec un quantificateur vectoriel. Les résultats des tests subjectifs en fonction des trois débits proposés sont présentés au chapitre 7.

CHAPITRE 7 RÉSULTATS ET ANALYSES

Les différentes versions des extensions de bande développées dans le projet de maîtrise sont comparées avec l'extension de bande de référence. Cette comparaison est faite à l'aide de tests subjectifs sur des participants. Les résultats et leurs analyses, pour les deux types de tests, sont présentés dans ce chapitre. L'analyse des résultats vise à vérifier si la qualité subjective perceptuelle de l'extension de bande développée est statistiquement supérieure à celle de l'extension de bande de référence.

Deux types de tests sont faits sur les signaux. Le test MUSHRA exige des participants experts et est beaucoup plus long à faire que le test AB. Le test AB permet de tester une plus grande quantité de fichiers audio en peu de temps et peut être effectué par un large public. Les résultats des deux types de tests possèdent la même validité, cependant le test AB permet de comparer uniquement deux codecs à la fois. En contrepartie, le test MUSHRA permet d'obtenir des résultats plus détaillés quant à la comparaisons des trois extensions de bande développées et celle de référence. De plus, il permet de situer les fichiers sonores comparés par rapport à deux fichiers de repère caché et le signal original. Dans le projet de maîtrise, un premier test MUSHRA a été fait sur des auditeurs experts, par la suite un test AB a été développé dans le but d'obtenir une deuxième validation sur des participants large public.

7.1 Test A/B

Les analyses statistiques du test A/B sont binaires (préférence pour le codec A ou B). Les résultats du test possèdent une distribution binomiale et suivent la loi de Bernoulli. Selon [21], [28], [29], les propriétés statistiques d'une distribution binomiale et de la loi de Bernoulli sont données par les équations (7-1), (7-2), (7-3), (7-4) et (7-5) :

- Support des valeurs x_k

$$x_k \in \{ 0 ; 1 \} \quad (7-1)$$

- Probabilité p que $x_k = 1$ et q que $x_k = 0$

$$p = \mathbb{P}(x_k = 1) \text{ et } q = \mathbb{P}(x_k = 0) = 1 - p \quad (7-2)$$

- Moyenne \bar{x} d'une distribution de valeur x_k ayant N échantillons

7.1 Test A/B

$$\bar{x} = \frac{1}{N} \cdot \sum_{k=1}^N x_k = P(x = 1) \quad (7-3)$$

- Écart-type σ

$$\sigma = \sqrt{p \cdot q} = \sqrt{\bar{x} \cdot (1 - \bar{x})} \quad (7-4)$$

- Intervalle de confiance à 95%

$$interval_{95\%} = \bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{N}} \quad (7-5)$$

Pour calculer le gain de qualité subjective de l'extension de bande proposée UFR2, par rapport à celle de référence, les votes en faveur de l'UFR2 sont mis à un, et ceux de l'AMR-WB+ sont mis à zéro selon l'équation (7-1). Le calcul de la moyenne des valeurs soustrait de 0,5 indique le gain ou la perte. Une valeur moyenne positive indique un gain alors qu'une valeur moyenne négative indique une perte de qualité subjective. Les intervalles de confiance à 95% (7-5) permettent de vérifier si les valeurs obtenues sont statistiquement significatives, soit une chance sur 20 de se tromper. Afin d'avoir des résultats statistiquement significatifs, les intervalles de confiance ne doivent pas croiser l'axe des abscisses.

Les tests d'écoute doivent généralement être faits dans un environnement contrôlé, par exemple une salle d'écoute insonorisée et sans perturbation extérieure. Certains protocoles de test, limitent le temps alloué au test et le nombre d'écoutes de chaque séquence sonore. De plus, le matériel doit être contrôlé afin que tous les auditeurs utilisent le même type de matériel. La qualité du matériel doit aussi être suffisante pour ne pas altérer les signaux sonores et s'assurer, ainsi que les différences soient perceptibles.

Dans le contexte d'un test subjectif, sans contrainte de matériel et de l'environnement des auditeurs, l'analyse doit prendre en compte qu'il y a beaucoup de résultats non utilisables. Ces résultats de ce test risquent donc d'être faussés par plusieurs facteurs, dont le matériel utilisé, la capacité des auditeurs à faire le test, les distractions extérieures et l'environnement dans lequel le test est fait. La sélection des résultats doit permettre de discriminer les auditeurs dont les résultats sont erronés. Pour ce faire, deux critères d'évaluations sont proposés.

- Le taux de réussite suffisant à reconnaître une séquence sonore originale de sa version encodée;

- La réponse à des questions sur le degré de difficulté à reconnaître des différences entre deux versions d’une séquence sonore. La réponse à ces questions se situe sur quatre niveaux de facile à impossible, où impossible est considéré comme un vote aléatoire.

L’analyse des résultats des quatre fichiers d’entraînement et des quatre fichiers de repère cachés, ainsi que les réponses aux questions permettent de déterminer la validité des résultats d’un auditeur. Pour la comparaison des performances entre les deux codecs testés, les quatre groupes d’auditeurs sélectionnés sont marqués par un astérisque (*) dans le tableau 7-1. Ces groupes sont les suivants :

- Tous les auditeurs;
- Les auditeurs ayant réussi à trouver 6/8 fichiers originaux ;
- Les auditeurs ayant réussi à trouver 6/8 fichiers originaux, ayant répondu une fois sur deux « Je ne perçois pas de différence, mes votes sont aléatoires ».
- Les auditeurs ayant réussi à trouver 6/8 fichiers originaux, n’ayant pas répondu « Je ne perçois pas de différence, mes votes sont aléatoires ».

Ces quatre groupes d’auditeurs démontrent des améliorations croissantes de la qualité subjective du signal en fonction des contraintes de sélection ajoutées, mais cependant, au détriment de légères hausses des intervalles de confiance, voir le figure 7-1.

Tableau 7-1 : Groupes d’auditeurs selon les critères de sélections du test A/B

Critères de sélection	Trouvé 2/8 fichiers originaux (tous)	Trouvé 5/8 fichiers originaux	Trouvé 6/8 fichiers originaux	Trouvé 7/8 fichiers originaux	Trouvé 8/8 fichiers originaux
2/2 et moins votes aléatoires (tous)	196*	171	150*	99	32
1/2 et moins votes aléatoires	178	156	131*	92	29
0/2 vote aléatoire	131	114	101*	70	20

7.2 Analyse des résultats du test A/B

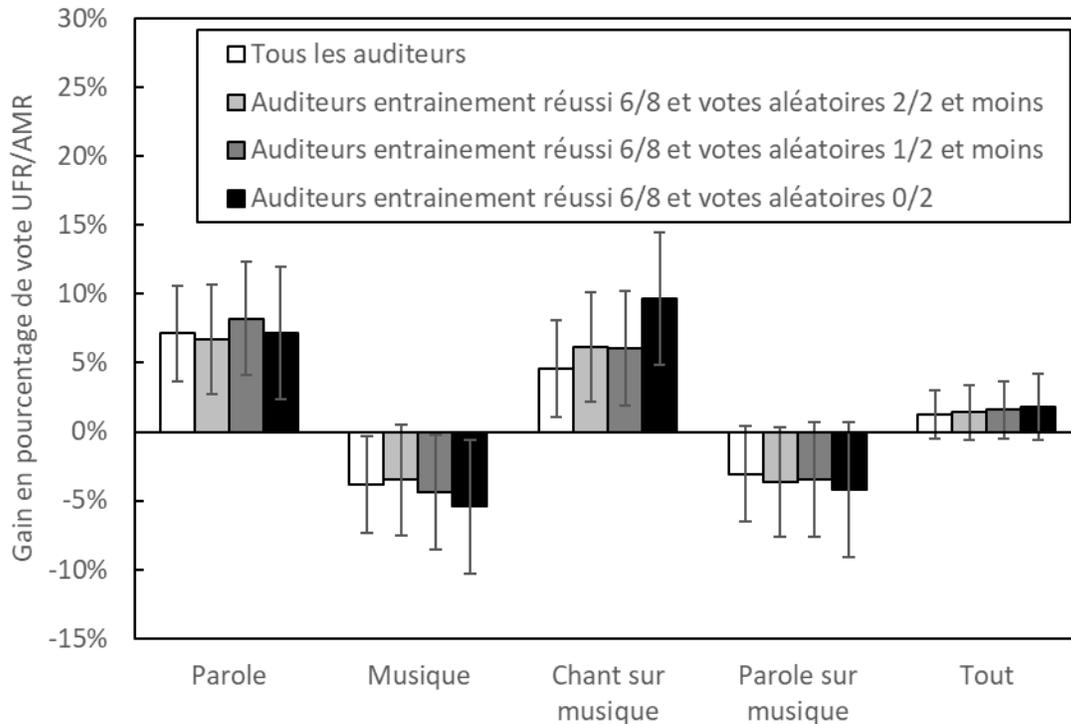


Figure 7-1 : Résultats du test A/B par catégorie de fichiers

7.2 Analyse des résultats du test A/B

L'analyse des résultats montre que la majorité des intervalles de confiance sont trop grands par rapport aux écarts observés. Ceci rend les résultats statistiquement non significatifs. Cependant, la sélection du groupe d'auditeurs ayant réussi à trouver 6/8 fichiers originaux et ayant mentionné que leurs votes n'étaient pas aléatoires, permet d'obtenir des écarts plus élevés dans l'analyse par catégories. Le fait d'éliminer des auditeurs augmente les écarts et augmente légèrement les intervalles de confiance. Les écarts statistiquement significatifs par rapport aux intervalles de confiance à 95% sont notés par un astérisque (*) dans le tableau 7-2.

Les résultats de ce test permettent de conclure que les fichiers de parole et de chant sur musique ont une qualité subjective supérieure, tandis que les fichiers de musique ont une qualité inférieure par rapport à l'extension de bande de référence. Les moyennes des fichiers de parole sur musique ainsi que la moyenne de tous les fichiers sont trop faibles par rapport aux intervalles de confiance. Il est donc impossible de confirmer si la qualité est supérieure ou inférieure.

Tableau 7-2 : Moyennes des écarts par catégorie pour le test A/B pour le groupe d'auditeurs ayant réussi à trouver 6/8 fichiers originaux, n'ayant pas de vote aléatoire

Codec	Type de donnée	Parole	Musique	Chant sur musique	Parole sur musique	Tout
UFR2	Moyenne	7,18%*	-5,45%*	9,65%*	-4,21%	1,79%
	Intervalle confiance	4,83%	4,85%	4,78%	4,86%	2,44%

7.3 Test MUSHRA

Les résultats du test MUSHRA sont des valeurs entières entre 0 et 100. Les propriétés statistiques de cette distribution sont données dans les équations suivantes [22] :

- Support des valeurs x_k

$$x_k \in \mathbb{N} \quad \forall x_k \leq 100 \quad (7-6)$$

- Moyenne \bar{x} d'une distribution de valeur x_k ayant N échantillons

$$\bar{x} = \frac{1}{N} \cdot \sum_{k=1}^N x_k \quad (7-7)$$

- Écart-type σ

$$\sigma = \sqrt{\frac{1}{N} \cdot \sum_{k=1}^N (x_k - \bar{x})^2} \quad (7-8)$$

- Intervalle de confiance à 95%

$$interval_{95\%} = \bar{x} \pm 1,96 \cdot \frac{\sigma}{\sqrt{N}} \quad (7-9)$$

Pour calculer le gain de qualité subjective des extensions de bande UFR(1,2,3) par rapport au BWE de l'AMR-WB+, l'analyse extrait les moyennes (7-7). Une valeur moyenne supérieure par rapport à un autre codec indique un gain de qualité subjective alors qu'une valeur inférieure, indique une perte. Les intervalles de confiance à 95% sont calculés à l'aide de l'équation (7-9). Afin d'avoir des résultats statistiquement significatifs, l'intervalle de confiance d'un codec ne doit pas croiser l'intervalle de confiance d'un autre codec. L'analyse des résultats contient les 43 fichiers de MPEG répartis en quatre tests avec, en moyenne, cinq auditeurs experts par test, et ce, pour un total de vingt auditeurs, voir la figure 7-2. Les critères d'élimination des participants du protocole de test MUSHRA sont très clairs. Les résultats d'un participant doivent être éliminés dans deux cas selon le protocole ITU pour les tests

7.3 Test MUSHRA

MUSHRA [22] : si un participant note plus de 15% des fichiers originaux en dessous de 90 ou s'il note plus de 15% des fichiers compressés à plus de 90. Les analyses du test n'ont rejeté aucun auditeur.

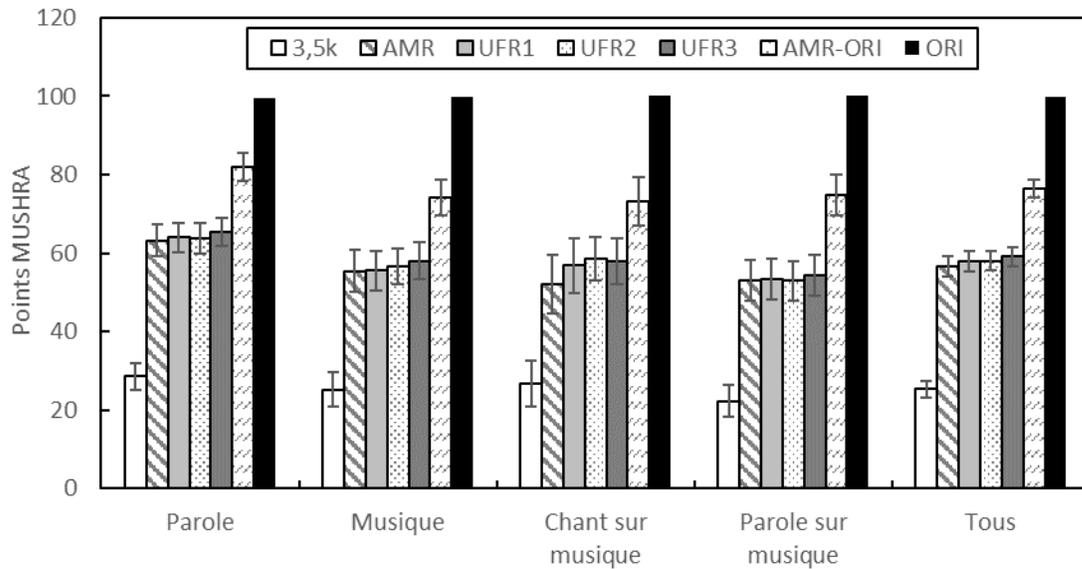


Figure 7-2 : Résultats du test MUSHRA par catégorie par rapport au signal original

La représentation graphique démontrant les différences entre les codecs de référence et les trois versions de l'UFR est plus simple à analyser, voir la figure 7-3. Afin d'avoir des résultats statistiquement significatifs, les intervalles de confiance ne doivent pas croiser l'axe des abscisses. Comme pour le test A/B, une moyenne positive indique un gain et une moyenne négative, une perte de qualité subjective.

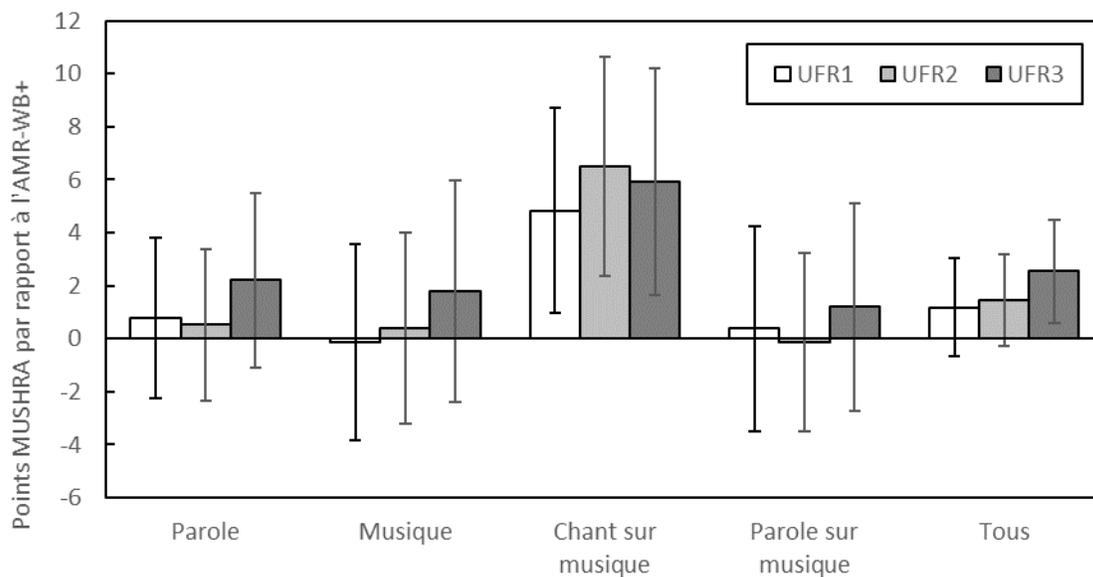


Figure 7-3 : Différences de points du test MUSHRA entre les signaux UFR(1,2,3) et AMR-WB+ par

catégorie par rapport au signal original

7.4 Analyse des résultats du test MUSHRA**Première analyse**

Une première analyse des résultats du test MUSHRA permet de conclure que les fichiers de chant sur musique (UFR1, UFR2 et UFR3) ainsi que la moyenne de tous les fichiers de la version avec un débit supérieur (UFR3) ont une qualité subjective supérieure par rapport à l'extension de bande de comparaison. Les intervalles de confiance des fichiers de parole, de musique et de parole sur musique pour tous les débits (UFR1, UFR2 et UFR3) ainsi que pour la catégorie tous les fichiers (UFR1 et UFR2) croisent l'axe des abscisses; de ce fait, il est donc impossible de confirmer si la qualité est supérieure ou inférieure. L'augmentation du débit permet une augmentation de la qualité du signal. Les résultats sont présentés à l'annexe B, voir le tableau b-2 et le tableau b-3.

Deuxième analyse

Une deuxième analyse est faite supposant cette fois que les signaux comparés ne peuvent dépasser en qualité subjective le signal AMR-ORI, formé des signaux AMR-WB en BF et originaux en HF. Pour chaque séquence, les résultats des votes sont tous divisés par la valeur obtenue pour le signal AMR-ORI, puis multipliés par 100. Cette opération revient à un ajustement proportionnel des résultats plaçant le signal AMR-ORI à 100 points MUSHRA, voir la figure 7-5. À la suite de cette opération, les écarts entre les signaux UFR(1,2,3) et le signal AMR-WB+ sont plus grands et permettent d'obtenir des intervalles de confiance plus faibles, voir la figure 7-5. Les résultats sont présentés à l'annexe B, voir le tableau b-4 et le tableau b-5.

7.4 Analyse des résultats du test MUSHRA

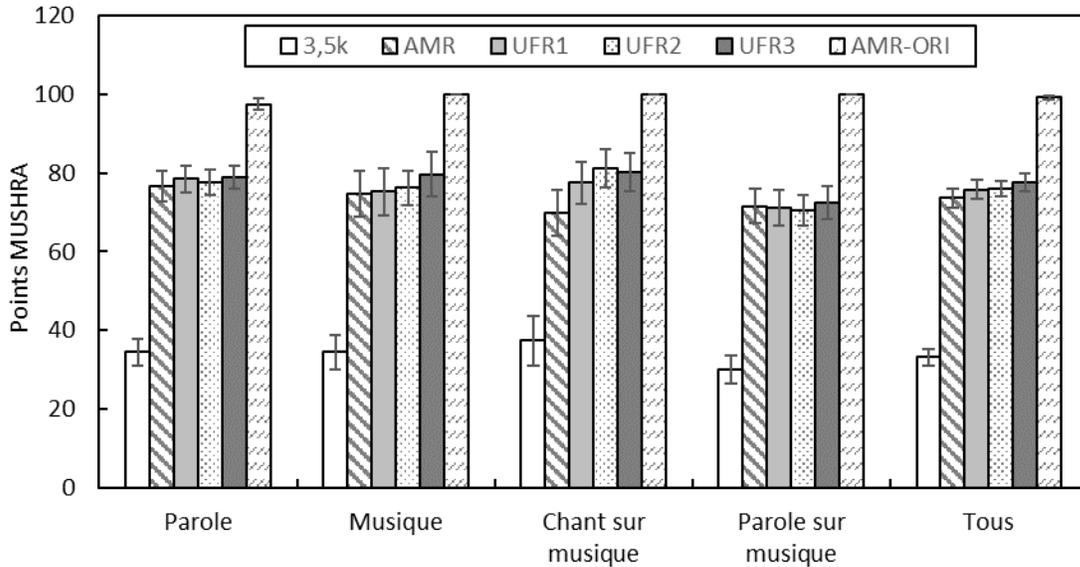


Figure 7-4 : Résultats du test MUSHRA par catégorie par rapport au signal AMR-ORI

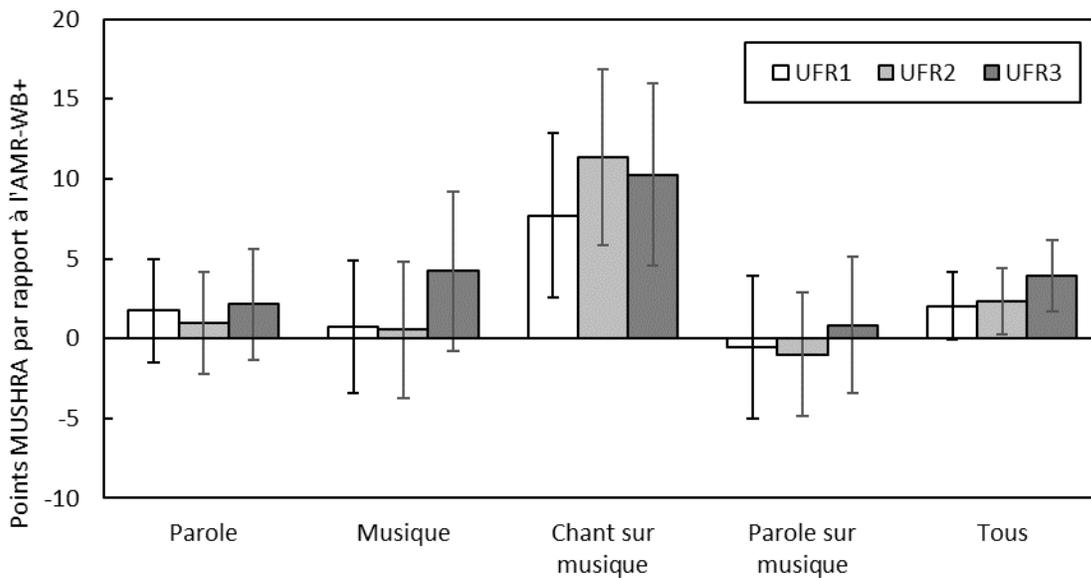


Figure 7-5 : Différences de points du test MUSHRA entre les signaux UFR(1,2,3) et AMR-WB+ par catégorie par rapport au signal AMR-ORI

L'analyse des résultats de cette représentation par rapport au signal AMR-ORI permet de conclure que les fichiers de chant sur musique (UFR1, UFR2 et UFR3) ainsi que la moyenne de tous les fichiers (UFR2 et UFR3) possèdent une qualité subjective supérieure par rapport à l'extension de bande de comparaison. Les intervalles de confiance des fichiers de parole, de musique et de parole sur musique croisent l'axe des abscisses; de ce fait, il est donc impossible de confirmer si la qualité est supérieure ou inférieure. Finalement, la moyenne de tous les fichiers (UFR1) suit la tendance des deux autres débits (UFR1 et UFR2), cependant l'intervalle de confiance est légèrement trop élevé pour confirmer le

gain de qualité subjective avec un débit plus faible, soit un écart de 2,04 et un intervalle de confiance de 2,12.

De manière générale, les moyennes obtenues démontrent une augmentation de la qualité subjective de l'extension de bande en augmentant le débit pour toutes les catégories de fichiers.

7.5 Résumé des analyses des résultats

Les deux tests subjectifs sont très différents tant en termes d'auditeurs et de contrôle du matériel, qu'en terme d'environnement des tests. De plus, la discrimination des auditeurs nécessite l'élimination de près de la moitié des auditeurs du test A/B, mais n'élimine aucun auditeur du test MUSHRA.

Cependant, les analyses concluent d'un gain en qualité subjective pour les fichiers de la catégorie de chant sur musique. De plus, même avec une diminution du débit de 25%, l'écart, de cette catégorie de fichier, est significatif selon l'analyse de résultats du test MUSHRA. Pour les deux tests, les analyses des résultats des versions UFR(1,2,3) pour cette catégorie, permettent de conclure une préférence de l'UFR2 à 59,65% pour le test A/B et des écarts en points MUSHRA variant de 4,84 à 6,48 entre l'UFR(1,2,3) et le signal ARM-ORI, puis une deuxième analyse par rapport au signal AMR-ORI donne des résultats qui varient entre 7,71 et 11,36. Les intervalles de confiance sur ces résultats sont grands, mais garantissent un gain de qualité subjective significatif.

Pour les fichiers de chant sur musique, l'extension de bande proposée à débit égal UFR2, le test A/B montre une préférence de 59,65% et garantit une préférence à 54,78% en raison de l'intervalle de confiance. Pour le test MUSHRA, l'écart est de 6,48 points et l'écart garanti est de 1,59 points de plus que le codec de référence. Pour les autres débits, avec une économie de 25% du débit l'UFR1 dépasse de 4,84 points MUSHRA, mais avec un écart garanti de 0,3 point en raison de l'intervalle de confiance. Quant à UFR3 avec une augmentation de 25% du débit, l'écart est de 5,92 points MUSHRA et garantit un écart de 0,85 point en raison de l'intervalle de confiance.

D'autre part, les analyses des résultats du test A/B montrent un gain pour la catégorie des fichiers de parole et une perte sur la catégorie des fichiers de musique. Cette conclusion n'est pas possible sur les analyses des résultats du test MUSHRA. Pour les fichiers de parole, l'analyse des résultats du test A/B montre une préférence significative à 57,18% en faveur de l'UFR2, dont une préférence assurée à 52,53% selon l'intervalle de confiance, par rapport au codec de référence. Quant à la catégorie des fichiers de parole sur musique, aucune conclusion ne peut être faite à la suite des analyses des résultats de deux tests.

7.5 Résumé des analyses des résultats

En conclusion, sur toutes les catégories de signaux, l'analyse des résultats du test MUSHRA montre des gains significatifs pour les signaux UFR1 et UFR2. Pour UFR1 ayant un débit de 25% inférieur le gain est non significatif en raison de l'intervalle de confiance trop grand. Finalement, la tendance du gain sur le résultat sur le signal UFR2 pour tous les fichiers est aussi présente dans le test A/B, mais dont le résultat est non significatif en raison de l'intervalle de confiance trop grande. Finalement, la catégorie de chant sur musique obtient des gains de qualité subjective significatifs pour tous les signaux dans les deux tests. La discussion de ces analyses est présentée au chapitre 8.

CHAPITRE 8 DISCUSSION

Les analyses des résultats des tests effectués permettent de conclure que l'extension de bande développée offre une qualité subjective supérieure sur certains types de fichiers sonores comparée à l'extension de bande de référence, le BWE de AMR-WB+. Ce chapitre présente la discussion des résultats et les améliorations possibles à l'extension de bande proposée.

8.1 Discussion du modèle paramétrique

Le modèle paramétrique développé représente et effectue tous les traitements de l'extension de bande dans le domaine des fréquences en utilisant la TFD comme outil de calcul de transformation. Ce choix permet d'offrir une plus grande flexibilité de modification du spectre. Il est possible d'ajuster l'amplitude et la phase d'une seule fréquence par une opération simple comme l'addition ou la multiplication. Cette opération dans le domaine temporel exige l'application d'un long filtre avec l'utilisation de plusieurs additions et multiplications, et ce, pour chaque fréquence à ajuster.

Ce domaine permet aussi le calcul de l'enveloppe spectrale par le cepstre qui offre une modélisation performante et facilement paramétrable avec ajustement du liftrage. Les trois méthodes de calcul de l'enveloppe spectrale comparées offrent des performances similaires, mais la méthode par le cepstre est plus performante avec cinq coefficients de modélisation selon les trois critères d'évaluation.

À la suite de cette comparaison, la génération des HF se fait par une double translation qui est le meilleur compromis entre toutes les méthodes comparées par calculs et tests subjectifs. Cette conception minimise le débit en ne transmettant pas le choix de méthode de génération de HF.

De plus, l'ajustement de la variance du signal blanchi en fonction des fréquences s'implémente facilement dans le domaine des fréquences. Cet outil à faible débit permet de compenser l'absence d'outil pour l'ajout et l'élimination de tonalité. L'équivalent du changement de variance sur le signal blanchi, dans le domaine temporel, exige un filtre complexe sur le signal résiduel. Ce traitement dans le domaine temporel est plus complexe. Le développement de cette transformation est préférable dans le domaine des fréquences puisque l'analyse et le traitement s'effectuent en fonction des fréquences.

Le modèle proposé ne contient pas d'enveloppe temporelle, mais la transmission de la fréquence zéro permet d'obtenir une enveloppe spectrale ajustée en amplitude et permettant une reconstruction suffisante de l'amplitude temporelle du signal à chaque trame. Le choix de la longueur de trame

8.2 Discussion des résultats

relativement courte permet de remplacer l'enveloppe temporelle dans le modèle d'extension de bande proposé.

Le modèle proposé est donc complet. Il permet une bonne modélisation du signal original HF et permet une régénération des HF adéquate au décodeur.

8.2 Discussion des résultats

Les deux types de tests d'écoute faits sont très différents par rapport aux contraintes imposées sur les tests. L'un avec un environnement contrôlé avec des auditeurs experts, et l'autre, pratiquement sans contrainte, avec des auditeurs sans expérience. Pour le test A/B, l'ajout de fichiers d'entraînement et de fichiers repères cachés, ainsi que de questions, ayant tous pour but de discriminer les auditeurs, compensent le manque d'uniformité concernant le matériel, l'environnement et les compétences des auditeurs.

Majoritairement, les analyses des fichiers sonores permettent de conclure sur un gain de qualité subjective. De plus, les conclusions des analyses des résultats ne sont pas contradictoires entre les deux tests. Les prochaines sections expliquent, en détail, les résultats obtenus et discutent des raisons potentielles de ces résultats.

8.2.1 Discussion des performances sur les fichiers de musique

Les résultats du test A/B montrent qu'il y a une faiblesse au niveau des signaux de musique. Une des lacunes possibles du modèle de l'extension de bande proposée est la longueur des trames d'analyse et de reconstruction du signal. L'AMR-WB+ utilise des trames de 20, 40 et 80 millisecondes avec une enveloppe temporelle ajustée aux 5 millisecondes. L'extension de bande développée possède des trames de 4 et 16 millisecondes et ne possède aucune enveloppe temporelle. Ce changement de longueur permettrait une amélioration potentielle, mais exigerait peut-être l'ajout d'une enveloppe temporelle.

8.2.2 Discussion des performances sur les fichiers de parole

Dans le test MUSHRA, Les fichiers de parole obtiennent de faibles valeurs positives avec des intervalles de confiance trop grands pour que la différence puisse être statistiquement significative sur. Cependant, cette tendance est confirmée par les analyses du test A/B, lesquelles permettent de conclure un gain, avec une moyenne de préférence variant de 56,67% à 58,21% selon les groupes d'auditeurs. Les artefacts sur le signal blanchi du codec de référence sont notamment plus présents sur les signaux de parole en raison de la symétrie des formants en HF. Ces artefacts se caractérisent par un

grésillement en HF. Il est à noter que ces artefacts ne sont pas perceptibles par tous les auditeurs en raison qu'ils se situent dans la bande de 14 à 16 kHz. Il s'agit d'une partie de la signature du codec AMR-WB+. Cette caractéristique pour avoir pour effet que certains auditeurs ne sont pas incommodés par ces artefacts et jugent les codecs sur d'autres aspects.

8.2.3 Discussion des performances sur les fichiers de chant sur musique

Les analyses des résultats des deux tests, telles que présentées au chapitre 7, montrent de manière plus marquée un gain en qualité subjective pour la catégorie des fichiers de chant sur musique. Il semble que certains paramètres de conception de l'extension de bande favorisent la modélisation des fichiers de chant sur musique. L'ajustement de la dynamique du signal blanchi par la variance en est peut-être l'explication. L'ajout de cet outil de modélisation du signal blanchi aurait dû avoir des résultats plus significatifs sur les signaux de paroles et de parole sur musique. Il est aussi possible que l'extension de bande développée compense une faiblesse du codec de référence. En analysant les résultats du test MUSHRA, les scores obtenus par le codec de référence sont plus faibles pour la catégorie de fichiers de chant sur musique que les autres catégories, voir le tableau b-4 et la figure 7-3. De plus, la comparaison des écarts de points MUSHRA entre le codec de référence et le signal AMR+ORI permet de comprendre la détérioration des HF, voir le tableau b-5 et la figure 7-5. La conclusion de ces données est que le codec de référence performe moins bien pour la catégorie de fichiers de chant sur musique avec 4,54 points en moins pour le MUSHRA avec fichier original, que pour tous les signaux. Le même calcul pour le MUSHRA avec comparaison du signal AMR+ORI donne 3,79 points en moins, que pour tous les signaux. La première valeur compare le codec AMR-WB+ au fichier original et la deuxième valeur compare l'extension de bande WBE à la version HF du fichier original. Ces écarts sont comparables en termes de nombre de points MUSHRA. Pour conclure, ces écarts prouvent que l'extension de bande proposée performe bien sur une catégorie de fichiers où l'extension de bande de référence performe moins bien en comparaison avec les autres catégories de signaux.

8.2.4 Discussion des performances sur les fichiers de parole sur musique

Pour les fichiers de parole sur musique, l'écart entre l'extension de bande de référence et l'AMR-ORI est comparable à celui des fichiers de chant sur musique. Il est difficile de trouver les raisons des performances moindres sur ces fichiers de parole sur musique en comparaison avec les fichiers de chant sur musique. La différence entre ces deux catégories de fichiers est due au fait que la parole est moins harmonique que le chant en HF. L'extension de bande proposée ajuste la variance du signal

8.3 Améliorations possibles

blanchi sans se soucier du type de signal. Si la variance est trop diminuée, cela peut avoir comme conséquence une baisse de la qualité subjective du signal généré. Il y a donc une dualité où la variance des signaux HF de musique harmonique doit être élevée et celle de la parole doit être diminuée pour certaines sections voisées. Le nombre de sous-bandes et les erreurs de quantifications de ce paramètre sont peut-être à revoir.

8.2.5 Discussion des performances pour tous les fichiers

En comparant les trois débits proposés, les résultats démontrent une tendance à augmenter la qualité subjective avec le débit. L'extension de bande UFR2 et UFR3, démontre un gain de qualité subjective supérieure et significative et un gain non significatif pour l'UFR1. Il est donc possible de conclure, avec certitude, que l'extension de bande proposée dépasse la qualité subjective du codec de référence pour le test subjectif MUSHRA. Cependant, l'analyse des résultats du test A/B ne permet pas de valider ce gain, et ce, en raison des intervalles de confiance trop grands. Une des causes de la grandeur de ces intervalles de confiance est le manque d'auditeurs pour ce test, ainsi que les conditions de test sans contrôle du matériel et de l'environnement. Cependant, il importe de noter que les résultats démontrent la même tendance. Une amélioration des performances sur les autres catégories de fichiers permettrait une augmentation des écarts.

8.2.6 Résumé des discussions

Finalement, l'extension de bande conçue possède un outil supplémentaire permettant d'améliorer le signal en ajustant les variances du signal blanchi. Cet outil est très performant, mais une partie du débit alloué à ce dernier est retirée du débit associé à l'enveloppe spectrale. Il devient pertinent de trouver un moyen d'optimiser l'allocation des bits en transmettant moins souvent les paramètres, soit en allongeant les trames, en interpolant les paramètres entre les trames, ou encore, en estimant plus de paramètres.

Finalement, l'extension de bande augmente la qualité du signal HF pour tous les types de fichiers, mais plus particulièrement dans la catégorie chant sur musique; pour laquelle l'extension de bande de référence ne performe pas bien par rapport aux autres catégories de signaux.

8.3 Améliorations possibles

Plusieurs modifications peuvent être apportées à l'extension de bande afin d'améliorer la qualité du signal, soit : optimiser le débit, estimer des paramètres et ajouter des outils de modélisation

paramétriques. La principale contrainte dans le choix des paramètres et leur débit est de concevoir une extension de bande ayant un débit comparable à celui de l'AMR-WB+.

8.3.1 Débit

Tel que mentionné précédemment, le choix des paramètres à transmettre ainsi que le débit associé à chaque paramètre sont des choix ardu. Un exemple de ces choix est le fait de ne pas transmettre d'enveloppe temporelle. De plus, la diminution du débit est permise par l'estimation de paramètres pour l'enveloppe spectrale et la variance. Ces deux économies de débit permettent la transmission des paramètres de la courbe de variances. D'autres choix sont faits afin d'obtenir un débit comparable à celui de l'ARM-WB+, dont le nombre de coefficients représentant l'enveloppe spectrale et le nombre de sous-bandes pour représenter les variances. De plus, le nombre de bits associé à chaque quantificateur a un impact direct sur le bruit de quantification.

Dans les catégories de l'optimisation du débit afin de transmettre les paramètres moins souvent, mais avec plus de précision, plusieurs options sont disponibles. Les principales options sont :

- Optimiser les trames avec des trames plus longues pour les signaux plus stationnaires;
- Diminuer le chevauchement des trames à moins de 50%;
- Interpoler les paramètres entre les trames.

Tous ces éléments permettent de diminuer le débit et d'allouer plus de bits à chaque paramètre qui est transmis moins fréquemment. Une étude de la variation des paramètres permettrait de valider les hypothèses de longueur de trame et d'interpolation des paramètres entre chaque transmission de ceux-ci.

Le codec de référence utilise des trames de 20, 40 et 80 millisecondes avec une fenêtre ayant un très faible pourcentage de recouvrement. Cette configuration permet d'avoir respectivement un budget de bits de 20, 40 et 80 bits par trame. Cette différence avec l'extension de bande conçue est énorme. Le fait d'allonger les trames de traitement, en ayant un recouvrement de 25%, permet une augmentation d'allocation des bits de 25%, soit 20 bits par trame.

L'estimation des paramètres est aussi une option pour diminuer le débit. L'utilisation d'estimateurs d'enveloppes spectrales, à partir des coefficients cepstraux de l'enveloppe en BF, et à partir des coefficients cepstraux des trames précédentes permet aussi de diminuer le débit transmis ou de diminuer les erreurs de quantification. Ces estimateurs peuvent être faits par un réseau de neurones.

8.3 Améliorations possibles

8.3.2 Modèle paramétrique

L'ajout d'outils de modélisation paramétrique à l'extension de bande permet d'améliorer le signal reconstitué. Parmi les outils possibles, l'amélioration du choix de type de génération des HF peut facilement devenir un projet sans fin. L'ajout de plancher de bruit et de tonalité sont deux outils qui permettent d'en augmenter la qualité, mais ces ajouts sont généralement reconnus comme étant très coûteux en débit. Ils sont nécessaires lorsque l'extension de bande vise à reconstruire des signaux dans la bande médiane, et ce, à partir de quelques kilohertz.

L'enveloppe temporelle est aussi un outil indispensable, permettant d'améliorer l'amplitude du signal temporel lorsque l'enveloppe fréquentielle est transmise moins fréquemment. Elle permet en général d'allonger les trames en minimisant l'étalement temporel. Il s'agit donc d'un compromis entre, d'une part, de courtes trames avec une bonne amplitude de l'enveloppe spectrale et, d'autre part, une enveloppe spectrale sans amplitude transmise moins souvent jointe d'une enveloppe temporelle.

De plus, l'ajout d'un outil de translation du signal blanchi afin de repositionner les harmoniques automatiquement ou avec transmission du décalage a été développé. Les résultats étaient non concluants de manière automatique et d'un autre côté, nécessitaient trop de débit pour limiter les effets de trame. Pour un décalage de 16 coefficients fréquentiels, il faut quatre bits par trame. Des développements futurs en ce sens permettraient d'améliorer la qualité du signal régénéré. L'estimation et la transmission de l'erreur d'estimation permettrait de trouver serait sûrement un bon compromis.

Finalement, la question sans réponse: « Comment régénérer la phase de manière plus cohérente avec le module généré? ». Interpoler les phases à partir des phases en BF semble une option non testée pour le moment. À haute fréquence, le système auditif de l'humain est moins sensible aux phases qu'à l'amplitude des signaux sonores, les extensions de bande allouent leur débit aux paramètres ayant le plus d'effet.

CHAPITRE 9 CONCLUSION

Ce chapitre fait le bilan du projet de maîtrise en présentant les résumés de chacun des chapitres de ce document.

Les extensions de bande permettent d'augmenter la largeur de la bande audio avec un très faible débit supplémentaire. L'objectif principal de la maîtrise était de concevoir une extension de bande permettant de corriger des lacunes de celle de l'AMR-WB+ à débit égal.

L'extension de bande a été développée pour les fréquences de 8 à 16 kHz. Le domaine de traitement des signaux choisi est le domaine fréquentiel en utilisant la TFD comme outil de transformation. Les résultats démontrent que ce domaine de traitement est équivalent au domaine temporel pour les signaux de musique et de parole.

Le modèle paramétrique proposé est composé de deux paramètres modélisant les signaux HF. Ces deux paramètres permettent de corriger les principaux artéfacts de l'extension de bande de référence. Les outils de calcul et d'estimation de ces paramètres développés permettent de modéliser, de manière compacte, les données à transmettre, et ce, dans un format minimisant la sensibilité des données.

Le test MUSHRA et ses analyses concluent à un gain de qualité subjective sur tous les fichiers à débit égal et supérieur, mais plus particulièrement sur les fichiers de chant sur musique. Pour les autres catégories de signaux, les écarts de qualité subjective entre les extensions de bande comparées sont faibles, mais non statistiquement significatifs par rapport aux intervalles de confiance. D'un autre côté, les trois débits proposés démontrent une augmentation de la qualité en fonction de l'augmentation du débit. Quant au test A/B, l'analyse des résultats conclue à un gain de qualité subjective pour les fichiers de parole et de chant sur musique, mais une baisse pour les fichiers de musique. Pour tous les fichiers, une hausse de la qualité subjective est aussi présente, mais non statistiquement fiable en raison des intervalles de confiance.

L'analyse de la variation des paramètres dans le temps afin d'ajuster la fréquence de transmission des paramètres permettrait une meilleure allocation du débit sur chaque paramètre. De plus, l'utilisation d'estimateurs de paramètres et le changement de longueur de trame permettraient sûrement d'obtenir de meilleurs résultats.

8.3 Améliorations possibles

De plus, la qualité des signaux compressés est considérée comme suffisante du point de vue de l'intelligibilité des signaux de parole. Il n'est donc pas pertinent pour ce projet de maîtrise d'effectuer des tests d'intelligibilité sur les signaux compressés. Atteinte des objectifs du projet de maîtrise

Le but du projet de maîtrise était de concevoir un nouvel algorithme d'extension de bande pour l'AMR-WB+ dans le domaine fréquentiel dont le débit est paramétrable selon la qualité souhaitée. Le domaine fréquentiel offrant un degré de liberté supérieur comparativement au domaine temporel.

Le premier objectif de conception était l'évaluation des méthodes de créations du signal HF : symétrie, translation, double symétrie, double translation, addition de multiples interpolations, ajout de bruit blanc gaussien paramétrable. Les travaux de recherche ont permis de déterminer de manière qualitative et par calcul que la méthode de génération des signaux HF la plus performante, en moyenne, est la double translation.

Le deuxième objectif était la comparaison des techniques de calcul et de modélisation de l'enveloppe spectrale. La technique de calcul d'enveloppe spectrale choisie est par le cepstre. L'évaluation de la performance de cette technique en comparaison avec les autres méthodes a permis de conclure que les résultats sont comparables au filtre LPC et aux sous-bandes avec interpolation par spline cubique. Cependant, avec seulement cinq coefficients la méthode de calcul par le cepstre est plus performante selon les trois critères d'évaluation de l'enveloppe spectrale. De plus, ce modèle permet de ne pas avoir d'enveloppe temporelle en raison de l'ajustement de l'enveloppe spectrale en amplitude. Le domaine fréquentiel permet aussi l'estimation de l'amplitude de l'enveloppe spectrale transformée par la suite en coefficient cepstral pour le calcul de l'enveloppe au décodeur. Cet estimateur permet de diminuer le débit.

Le troisième objectif était d'améliorer le signal généré en HF; le projet de maîtrise a permis le développement d'un outil d'ajustement de la dynamique du signal. L'outil développé est une courbe de variances du signal blanchi qui permet l'ajustement de l'harmonicité du module du spectre HF. Cet outil augmente et diminue la dynamique du signal blanchi HF en changeant sa variance en fonction des fréquences. En augmentant considérablement le débit de ce paramètre dans un futur projet, cet outil permettrait l'ajout et l'atténuation de tonalité. Pour finir, il permet une bonne amélioration du signal HF en corrigeant les principaux artefacts de l'extension de bande de référence.

En conclusion, le projet de maîtrise avait aussi comme objectif d'évaluer la performance d'un algorithme d'extension de bande dans le domaine fréquentiel dont le débit est paramétrable selon la qualité désirée. Les performances de l'extension de bande ont été évaluées avec deux types de tests :

un test MUSHRA s'adressant à des auditeurs experts et un test A/B large public distribué sans contrainte de matériel, ni de l'environnement de test. Cet objectif a été atteint en raison du développement d'une procédure de test et de l'analyse des résultats. Les analyses des résultats de ces deux tests aboutissent aux mêmes conclusions, ainsi le protocole développé est adapté à la situation du test A/B utilisé.

9.1 Contributions

Les contributions au domaine scientifique sont multiples. Le projet de maîtrise était de concevoir un modèle d'extension de bande dans le domaine fréquentiel. Les contributions scientifiques se trouvent principalement dans les outils et les techniques développées pour mettre en œuvre l'extension de bande.

Les principales contributions se regroupent en quatre éléments : premièrement, le développement de deux nouvelles méthodes, la double translation et de la double symétrie, pour générer les signaux HF; suivi de la comparaison subjective et par critères de performance des méthodes de génération des HF. Deuxièmement, la comparaison des méthodes de calcul et de modélisation de l'enveloppe spectrale. Cette comparaison a permis de déterminer un nombre minimum de coefficients permettant une modélisation adéquate de l'enveloppe spectrale des HF pour les signaux de parole et de musique. De plus, cet outil de calcul d'enveloppe spectrale par le cepstre offre une meilleure modélisation lorsqu'il est appliqué sur un spectre symétrique que non symétrique, donnant des coefficients cepstraux tous réels.

Afin d'améliorer le signal, le développement des outils de la modélisation des variances par une courbe de variances et le changement des variances sur le signal blanchi HF permet d'ajuster l'harmonicité du signal HF. Les principaux artefacts dus à la génération des HF en fonction des signaux BF peuvent être corrigés par ces outils. Selon le nombre de sous-bandes et le débit associé, cet outil permet d'ajout et l'atténuation des tonalités du module du spectre.

Finalement, le projet de maîtrise se déroulant en contexte de pandémie a permis le développement d'une procédure de test et d'analyses des résultats pour un test A/B large public sans contrôle du matériel et de l'environnement. La comparaison des résultats des deux tests effectués conclut que la procédure développée pour le test A/B permet des résultats significatifs au même titre que le test MUSHRA; et ce, avec des critères de réjection des auditeurs en fonction de leur performance sur des

9.1 Contributions

fichiers d'entraînement et des fichiers de repère cachés, ainsi que des questions sur le niveau de difficulté du test.

L'extension de bande développée dans cette maîtrise ne sera surement jamais utilisée dans un standard audio, cependant le développement et la comparaison d'outils permettant la modélisation du signal HF permettront peut-être d'inspirer d'autres travaux de recherche. De plus, le protocole de test développé permettra peut-être d'offrir de nouvelles conditions de test.

RÉFÉRENCES

- [1] G. Fusch, «Codage Audio Hiérarchique à Faibles Débits,» Sherbrooke, 2006.
- [2] E. Larsen et R. M. Aarts, Audio bandwidth extension, Chichester: Wiley, 2004.
- [3] J.-M. Vallin et R. Lefebvre, «Bandwidth Extension of Narrowband Speech for Low Bit-Rate Wideband Coding,» *IEEE Workshop on Speech Coding*, pp. 130-132, 2000.
- [4] R. Drullman, «Temporal envelope and fine structure cues for speech intelligibility,» *The Journal of the Acoustical Society of America*, vol. 9, pp. 585-592, 1995.
- [5] R. Drullman, J. Festen et R. Plomp, «Effect of temporal envelope smearing on speech reception,» *The Journal of the Acoustical Society of America*, vol. 95, pp. 1053-1064, 1994.
- [6] J. Herre et S. Dick, «Psychoacoustic Models for Perceptual Audio - A Tutorial Review,» *Applied Sciences*, vol. 9, pp. 2854-2876, 2019.
- [7] A. Beznosiuk, «De la prise de son à la diffusion: mystères et mécanique de la perspective sonore,» Éditions Téraèdre, 2017.
- [8] J. Makhoul et M. Berouti, «High-Frequency Regeneration in Speech Coding Systems,» *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 428-431, 1979.
- [9] L. Laaksonen, J. Kontio et P. Alku, «Artificial Bandwidth Expansion Method to Improve Intelligibility and Quality of AMR-Coded Narrowband Speech,» *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 809-812, 2005.
- [10] J. R. Epps et W. H. Holmes, «A New Technique for Wideband Enhancement of Coded Narrowband Speech,» *IEEE Workshop on Speech Coding Proceedings. Model, Coders, and Error Criteria*, pp. 174-176, 1999.
- [11] J. Makinen, «AMR-WB+: A New Audio Coding Standard for 3rd Generation Mobile Audio Services Proceedings.,» *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 1109-1112, 2005.
- [12] S. Chennoukh, A. Gerrits, G. Miet et R. Sluijter, «Speech Enhancement Via Frequency Bandwidth Extension using Line Spectral Frequencies,» *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 665-668, 2001.
- [13] J. Kontio, L. Laaksonen et P. Alku, «Neural Network-Based Artificial Bandwidth Expansion of Speech,» *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, 2007.
- [14] J. R. Epps et W. H. Holmes, «A New very Low Bit Rate Wideband Speech Coder with a Sinusoidal Highband Model,» *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 349-352, 2001.
- [15] P. Ekstrand, «Bandwidth extension of audio signals by spectral band replication,» *IEEE Benelux Workshop on Model based Processing and Coding of Audio*, pp. 53-58, 2002.

- [16] J. Abel, E. Seidel et T. Fingscheidt, «Enhancing the EVS Codec in Wideband Mode by Blind Artificial Bandwidth Extension to Superwideband,» *16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 281-285, 2018.
- [17] M. Kaniewska et Al, «Enhanced AMR-WB bandwidth extension in 3GPP EVS codec,» *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 652-656, 2015.
- [18] V. Atti et al, «Super-wideband bandwidth extension for speech in the 3GPP EVS codec,» *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5927-5931, 2015.
- [19] F. Nagel et D. Sasha, «A HARMONIC BANDWIDTH EXTENSION METHOD FOR AUDIO CODECS,» *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 145-148, 2009.
- [20] A. Nguyen Duc et P. Abrahamsson, «Minimum Viable Product or Multiple Facet Product? The Role of MVP in Software Startups,» *International Conference on Agile Software Development*, vol. 251, pp. 118-130, 2016.
- [21] ITU-T, «Méthodes d'évaluation objective et subjective de la qualité de la transmission,» SECTEUR DE LA NORMALISATION DES TÉLÉCOMMUNICATIONS DE L'UIT, 1996.
- [22] The ITU Radiocommunication Assembly, «Method for the subjective assessment of intermediate quality level of audio systems,» RECOMMENDATION ITU-R BS.1534-3, 2015.
- [23] M. H. Pinson et al, «The Influence of Subjects and Environment on Audiovisual Subjective Tests: An International Study,» *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, pp. 640-651, 2012.
- [24] E.-N. Taoufik, «Conversion de voix pour la synthèse de la parole,» Université de Renne, Renne, 2005.
- [25] K. Sayood, *Introduction to data compression*, San Francisco: Morgan Kaufmann, 2006.
- [26] P. Hedelin, «Single stage spectral quantization at 20 bits,» *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 525-528, 1994.
- [27] F. Sadaoki, *Digital Speech Processing, Synthesis, and Recognition*, New York: Marcel Dekker Inc, 1989.
- [28] N. Derobert, «La distribution binomiale» [En ligne]. Disponible : <https://commentprogresser.com/statistique-distribution-loi-bernoulli-binomiale.html>. [Accès le juin 2020].
- [29] S. C. Medin, J. Murray-Bruce, D. Castanon et V. K. Goyal, «Beyond Binomial and Negative Binomial: Adaptation in Bernoulli Parameter Estimation» *IEEE Transactions on Computational Imaging*, vol. 5, pp. 570-584, 2019.

ANNEXE A

Tableau a-1 : Séquence des fichiers sonores et questions du test A/B

Fichiers	Nom	Type de fichier	A	B
1*	Brahms	Musique	Original	Original 3.5k
2*	Dongwoo	Chant sur musique	AMR+UFR2	Original
3*	Arirang_speech	Parole	Original	AMR+Original
4*	Alice	Parole sur musique	AMR+BWE	Original
Question éliminatoire				
5	Green_speech	Parole	AMR+UFR2	AMR+BWE
6	Te16_fe49	Parole sur musique	AMR+BWE	AMR+UFR2
7	Lion	Parole	AMR+UFR2	AMR+BWE
8	Twinkle_ff51	Parole sur musique	AMR+BWE	AMR+UFR2
9*	Dora	Parole	Original	AMR+BWE
10	Music_2	Musique	AMR+UFR2	AMR+BWE
11	Music_5	Chant sur musique	AMR+UFR2	AMR+BWE
12	Louis_raquin_15	Parole	AMR+BWE	AMR+UFR2
13	Music_3	Musique	AMR+UFR2	AMR+BWE
14*	Phi2	Musique	AMR+UFR2	Original
Question mi-test				
15	Music_4	Chant sur musique	AMR+BWE	AMR+UFR2
16	Music_1	Musique	AMR+BWE	AMR+UFR2
17	Noodleking	Parole sur musique	AMR+UFR2	AMR+BWE
18	Phi1	Chant sur musique	AMR+BWE	AMR+UFR2
19*	Phi6	Chant sur musique	AMR+BWE	Original
20	HarryPotter	Parole sur musique	AMR+UFR2	AMR+BWE
21	Phi4	Chant sur musique	AMR+UFR2	AMR+BWE
22	Phi3	Musique	AMR+BWE	AMR+UFR2
23	KoreanM1	Parole	AMR+BWE	AMR+UFR2
24*	SpeechOverMusic_1	Parole sur musique	Original	AMR+UFR2
Question finale				

Les fichiers sonores servant d'entraînement ou de repère cachés sont indiqués par un astérisque (*).

ANNEXE B

Dans les tableaux suivants, les écarts statistiquement significatifs par rapport aux intervalles de confiance à 95% sont notés par un astérisque (*)

Tableau b-2 : Points par catégorie pour le test MUSHRA par rapport au signal original

Codec	Type de donnée	Parole	Musique	Chant sur musique	Parole sur musique	Tous
AMR	Points	63,24	55,40	52,06	53,04	56,60
	Intervalle confiance	3,42	4,69	4,75	4,31	2,15
UFR1	Points	64,01	55,51	56,90	53,42	57,78
	Intervalle confiance	3,77	5,16	6,99	5,25	2,59
UFR2	Points	56,64	58,54	52,91	58,04	56,64
	Intervalle confiance	3,79	4,56	5,54	4,99	2,37
UFR3	Points	58,03	57,98	54,24	59,14	58,03
	Intervalle confiance	3,40	4,33	5,78	3,95	2,11

Tableau b-3 : Points par catégorie pour le test MUSHRA par rapport au signal AMR-ORI

Codec	Type de donnée	Parole	Musique	Chant sur musique	Parole sur musique	Tous
AMR	Points	76,71	74,69	69,85	71,56	73,64
	Intervalle confiance	3,86	5,92	5,91	4,28	2,49
UFR1	Points	78,45	75,27	77,56	71,05	75,69
	Intervalle confiance	3,39	6,02	5,32	4,48	2,43
UFR2	Points	77,69	76,29	81,22	70,58	75,98
	Intervalle confiance	3,22	4,38	4,82	3,95	2,04
UFR3	Points	78,84	79,68	80,12	72,40	77,57
	Intervalle confiance	2,97	5,73	4,89	4,16	2,25

Tableau b-4 : Écarts par catégorie pour le test MUSHRA par rapport au signal original

Codec	Type de donnée	Parole	Musique	Chant sur musique	Parole sur musique	Tous
UFR1	Moyenne	0,77	-0,13	4,84*	0,38	1,18
	Intervalle confiance	3,04	3,40	4,58	3,88	1,84
UFR2	Moyenne	0,53	0,39	6,48*	-0,13	1,45
	Intervalle confiance	2,86	3,30	4,89	3,36	1,74
UFR3	Moyenne	2,20	1,80	5,92*	1,20	2,55*
	Intervalle confiance	3,28	3,86	5,07	3,92	1,94

Tableau b-5 : Écarts par catégorie pour le test MUSHRA par rapport au signal AMR-ORI

Codec	Type de donnée	Parole	Musique	Chant sur musique	Parole sur musique	Tous
UFR1	Moyenne	1,75	0,72	7,71*	-0,51	2,04
	Intervalle confiance	3,23	4,17	5,12	4,47	2,12
UFR2	Moyenne	0,98	0,54	11,36*	-0,98	2,33*
	Intervalle confiance	3,18	4,29	5,51	3,84	2,09
UFR3	Moyenne	2,14	4,22	10,27*	0,83	3,93*
	Intervalle confiance	3,45	4,97	5,74	4,27	2,24